

化合物構造比較法に基づく 酵素反応特有な原子変換パターンの解析

服部正泰、山田拓司、Min-A Oh、五斗進、金久實
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

酵素反応前後の化合物ペア間で原子アラインメントを計算すると、反応前後において変化しない部分、付加または脱離によって変化する部分、前記の二つの境界領域にあつて反応の中心的舞台と考えられる部分、の3種類の情報を抽出できる。これらの情報は酵素反応に特徴的な原子の変換パターンを端的に捉えるものである。我々は、この変換パターンを RPAIR データベースとして構築するとともに、代謝経路の網羅的解析へ応用した。ここでは特に、原子変換パターンで表現された連続する反応モジュールを既存の代謝パスウェイ上で抽出することを行い、また、生分解経路において特徴的に見られる変換パターンの傾向をみた。これらの解析結果は、現在我々が構築を行っている代謝パスウェイ予測システムのキーとなるものである。

Analysis of atom transformation patterns in enzymatic reactions based on the comparison of chemical compound structures

Masahiro Hattori, Takuji Yamada, Min-A Oh, Susumu Goto, Minoru Kanehisa
Bioinformatics Center, Institute for Chemical Research, Kyoto University

After calculating the atom alignment between two chemical compounds, we can elucidate three types of informative atoms through enzymatic reactions, (i) atoms which don't change their physicochemical properties, (ii) atoms which can be changed by addition or deletion, and (iii) atoms which locate between two former parts and seem to be the reaction centers. We have annotated those atom transformation patterns for each enzymatic reaction and developed the RPAIR database in KEGG. Based on these data, here, we analyzed the continuous chemical reaction module along metabolic pathways and demonstrated the distribution of such atom transformation patterns found in degradation pathways.

はじめに

代謝化合物などの低分子化合物とそれらを触媒する酵素反応より構成される代謝反応パスウェイは、代謝化合物をノード、酵素反応をエッジとする巨大なグラフとみなすことが

できる。代謝化合物は、このグラフに示される一連の酵素反応に従って、その構造を少しずつ変化させていく。このような代謝反応グラフのうち、生化学的実験を経て既知となった部分グラフの情報は、たとえば KEGG

PATHWAY[1]に代表されるパスウェイデータベースに収録されており、遺伝子の機能アノテーションや細胞シミュレーションなどのポストゲノム解析に積極的に応用されるようになっている。

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本来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し、植物の二次代謝産物の生合成経路なども未知である場合が多い。このため、代謝反応グラフ、特に生物の持つ酵素反応の化学的特徴をより詳細に記述し、計算機的に取り扱う手法の開発が求められるようになってきている。

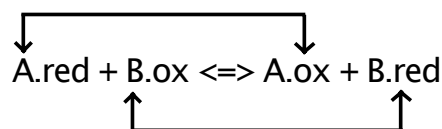
このような背景の下、我々の研究室では、酵素反応前後における化合物構造の変化を「原子変換パターン」として抽出し解析するまったく新しい手法の開発を進めてきた。

手法

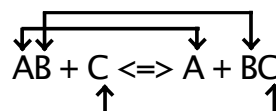
データは、KEGG LIGAND[2]データベースに収録されている、酵素反応 (REACTION) と、代謝化合物 (COMPOUND) を用いた。このうち REACTION データベースに対しては、酵素反応の典型的パターンの知識を基にしたバイナリペアの情報を再定義し拡充している。ここでバイナリペアとは、反応物と生成物のペアのうち、酵素反応の前後における原子の由来関係があるもの、である。(図1)

つぎに、酵素反応のバイナリペアとなる2化合物に対して、我々が新規に開発した化合物構造比較法[3]を用いて原子アラインメントの作成を行う。この化合物構造比較法は、化合物構造を2次元グラフとみなした場合の最大共通部分グラフを抽出するアルゴリズムをベースに、生体内代謝化合物に適合するよ

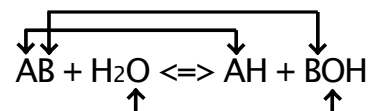
EC1, Oxidoreductases



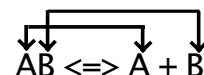
EC2, Transferases



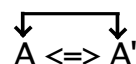
EC3, Hydrolases



EC4, Lyases



EC5, Isomerases



EC6, Ligases

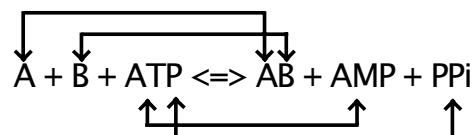


図1 酵素反応前後における化合物ペア
酵素反応カテゴリー毎にみると、反応前後における原子の由来関係により、反応物・生成物ペアを定義可能である。

ういくつかのヒューリスティクスを導入したもの[4]である。そのため得られる解は必ずしも厳密解ではないが、生体内化合物においてはよいパフォーマンスを示すことが分かっている。ここで、導入したヒューリスティクスには、各原子のもつ物理化学的環境を3文字コードで表記する KEGG Atom Type もある。

(表1) これは、原子の周辺環境情報を予めその注目原子の表記として取り込むことによって、前述の構造比較法における計算の一部を事前に行っておくことに相当する。

Description	KEGG	Bond pattern	Description	KEGG	Bond pattern
alkane	C1a	R-CH3	alkyne	C3a	R=-C-H
	C1b	R-CH2-R		C3b	R=-C-R
	C1c	R-CH(-R)-R	aldehyde	C4a	R-CH=O
	C1d	R-C(-R)2-R	ketone	C5a	R-C(=O)-R
cyclic alkane	C1x	ring-CH2-ring	cyclic ketone	C5x	ring-C(=O)-ring
	C1y	ring-CH(-R)-ring	carboxylic acid	C6a	R-C(=O)-OH
	C1z	ring-C(-R)2-ring	carboxylic ester	C7a	R-C(=O)-O-R
alkene	C2a	R=CH2		C7x	ring-C(=O)-O-ring
	C2b	R=CH-R	aromatic ring	C8x	ring-CH=ring
	C2c	R=C(-R)2		C8y	ring-C(-R)=ring
cyclic alkene	C2x	ring-CH=ring	undefined C	C0	
	C2y	ring-C(-R)=ring			

表1 KEGGにおける原子タイプ例 (全 68 タイプから炭素原子のみ抜粋)

全タイプの情報は、<http://www.genome.ad.jp/kegg/reaction/KCF.html> にある。

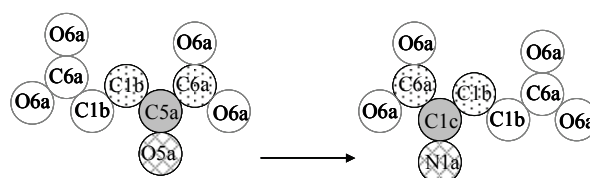
酵素反応前後における化合物の構造は比較的よく似ているので、原子アラインメントの結果、反応前後で保存される部分構造はアラインメントされ、そうでない部分はアラインメントされずに残ることになる。したがって、この情報を基にすると、次の3種類の原子情報を抽出することが可能である。

(1) 酵素反応の前後においてまったく変化していない原子 (Match 領域)。これは基本的には、アラインメントされた原子から後述する R 原子を除いた原子を指す。ここでは特に、後述する R 原子に隣接している原子を M 原子と呼ぶ。

(2) 付加または脱離反応によって原子環境あるいは原子そのものが変化していると考えられる原子 (Difference 領域)。同様に、アラインメントされずに残された原子に相当する。この領域にある原子のうち、後述する R 原子に隣接したものを D 原子と呼ぶ。

(3) 前述の二つの原子の境界領域に位置し、酵素反応機構の中心的舞台と考えられる部分 (Reaction center 領域)。具体的には、Match 領域の含まれる原子のうち、アラインメントされなかった部分と隣接している原子を指しており、R 原子と呼ぶ。

このような R,D,M 原子を用いて、酵素反応における変換規則を、RDM パターンあるいは原子変換パターンとして抽出した。(図2)



- (R)** Reaction center C5a --> C1c
- (D)** Difference region O5a --> N1a
- (M)** Matched region C1b+C6a --> C1b+C6a

図2 RDM パターンの定義

構造アラインメントから、原子環境 (KEGG の原子タイプ) が変化した部分 (D)、変化しない部分 (M)、二つの境界領域 (R)、をそれぞれ同定・抽出する。

```
ENTRY      A00077                               ReactantPair
NAME       C00093_C00111
COMPOUND   C00093_C00111
TYPE       main
RDM        1
           1      C1c-C5a:*-*:C1b+C1b+O1a-C1b+C1b+O5a
ALIGN      10
           1      1:P1b   1:P1b
           2      2:O2b   2:O2b
           3      3:O1c   4:O1c
           4      4:O1c   3:O1c
           5      5:O3b   5:O3b
           6      6:C1b   6:C1b #M1
           7      7:C1c   7:C5a #R1
           8      8:C1b   8:C1b #M1
           9      9:O1a   9:O5a #M1
          10     10:O1a  10:O1a
///
```

図3 RPAIR データベース

RPAIR データベースでは、アラインメントの結果に、RDM 原子の情報を追加している。

結果と考察

化合物ペアとそれから抽出される RDM パターンの情報を KEGG LIGAND データベースの主要コンポーネントの一つである RPAIR データベースとして構築した。[5] (図 3) RPAIR データベースのエントリ数は、現在 5162 個であるが、エントリ数の拡充の作業は、継続的に行われている。

酵素反応前後の化合物ペアのような比較的構造類似性の高い化合物間においては、上述した手法を用いて抽出される原子変換パターンは、酵素反応によって引き起こされる構造変化部位と一致する可能性が高くなる。このため、RDM パターンと酵素番号 (EC 番号) 分類による反応分類との間には、高い相関のあることを示すことが可能である。[6] 一方で、EC 番号における番号付けのあいまい性などの問題から、異なる分類番号を振られた酵素反応が実は同じ原子変換パターンを持っていることを確認することもできた。これは、EC 番号のような配列類似性や番号付けの問題を包含した分類体系を、酵素反応メカニズムの解析に用いることの難しさを示すものである。この点において、RDM による記述は、酵素反応の反応メカニズムのみを原子レベルで抽出することができるため、代謝パスウェイを初めとした生命現象におけるケミカルな解析には、より有効であると思われる。

ここではさらに、代謝パスウェイ上において、酵素反応を RDM パターンで表記したときによく観察される連続反応のセットを、反応モジュールという形で抽出する解析を行った。その結果、連続してみられるパターンの長さや観測頻度の間には、特徴的な分布の傾向が見られた。(図 4) このようなパターンのうち特によく観察されたものは、核酸の生合成経路における連続したリン酸化反応の経路や、タンパク質の生合成経路におけるアミノ基と C α 原子の周辺のような、共通部分構造を合成する経路に見られるものであるが (図 5)、それ以外のパターンも見つかっている。

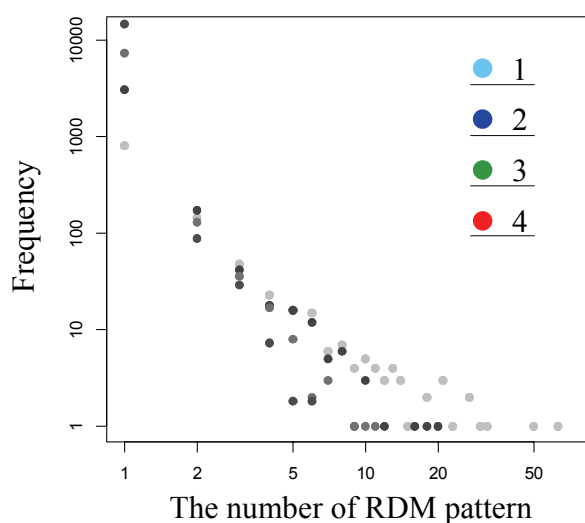


図 4 連続 RDM パターンの頻度分布
横軸は連続する RDM パターンの長さ、縦軸はそれらが観察される頻度を示す

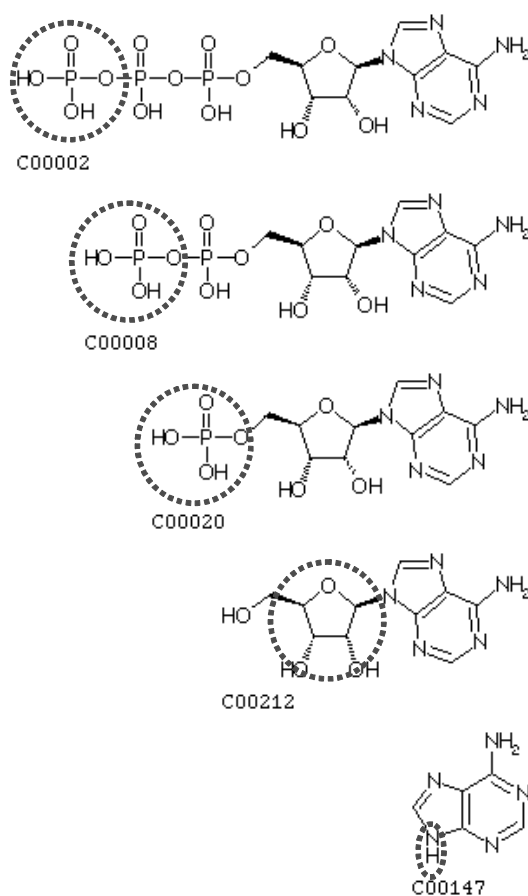


図 5 連続 RDM パターンの例
リン酸化反応が連続するような反応系列が、高頻度に見つかった。

また、生体内で酵素的に行われている環境物質等の生分解経路に特徴的な反応パターンがないかを網羅的に抽出し解析した。(図6) 具体的には、KEGG PATHWAY のほかに、米国ミネソタ大学で構築されている生分解経路データベース UM-BBD から網羅的に収集した 782 個の酵素的生分解反応のグループと、それ以外

の(すなわち生分解的でないと思われる)酵素反応のグループについて、それぞれで特徴的に見られる RDM パターンの分布をみた。その結果、両グループでほぼ変化なく出現する普遍的な反応パターンがある一方、どちらかのグループにのみ特異的に出現する特殊な反応パターンを見出すことができた。

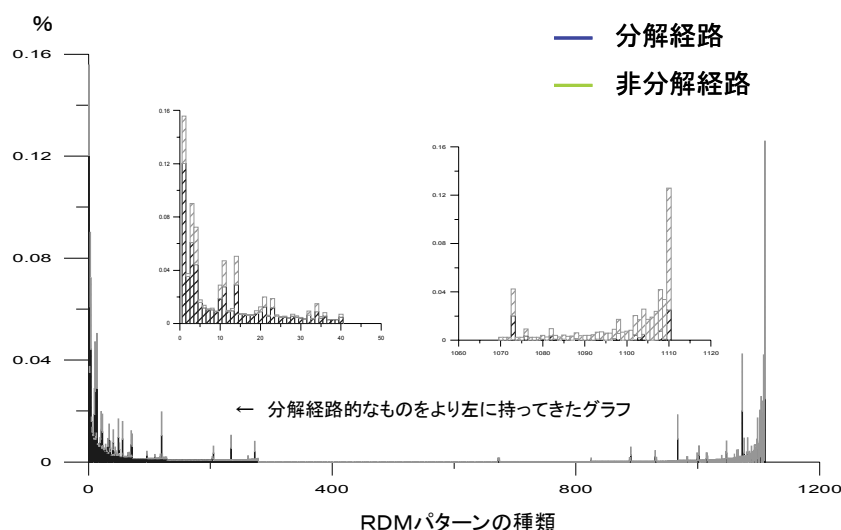


図6 分解・非分解経路における出現 RDM パターンの分布

結び

このような解析結果を踏まえ、現在、与えられた 2 化合物の構造間に起こりうる酵素反応や、薬剤等環境因子の生体内における代謝パスウェイの予測などに取り組んでいる。また、保存されたケミカル・モチーフ構造は、あまたあ

る有機化学反応のうち、生物がより好んで用いる化学反応が何か、それらをどう組み合わせで利用しているのかということを示すものであり、今後、ゲノムの進化との関連性においても非常に興味深いものと考えている。

参考文献

- [1] Kanehisa, M., Goto, S., Kawashima, S., Okuno, Y., and Hattori, M., "The KEGG resource for deciphering the genome", *Nucleic Acids Research*, **32**, D277-280 (2004).
- [2] Goto, S., Okuno, Y., Hattori, M., Nishioka, T., and Kanehisa, M., "LIGAND: database of chemical compounds and reactions in biological pathways", *Nucleic Acids Research*, **30**, 402-404 (2002).
- [3] 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", *Journal of the American Chemical Society*, **125**, 11853-11865 (2003).
- [4] Hattori, M., Okuno, Y., Goto, S., and Kanehisa, M., "Heuristics for Chemical Compound Matching", *Genome Informatics*, **14**, 144-153 (2003).
- [5] Kanehisa, M., Goto, S., Hattori, M., Aoki-Kinoshita, K.F., Itoh, M., Kawashima, S., Katayama, T., Araki, M., and Hirakawa, M., "From genomics to chemical genomics: new developments in KEGG", *Nucleic Acids Research*, **34**, in press (2006).
- [6] Kotera, M., Okuno, Y., Hattori, M., Goto, S., and Kanehisa, M., "Computational assignment of the EC numbers for genomic-scale analysis of enzymatic reactions", *Journal of the American Chemical Society*, **126**, 16487-16498 (2004).