氏　　名　小寺正明（こてら まさあき）
学位(専攻分野)　博　士（理　学）
学位記番号　理　博　第2926号
学位授与の日付　平成17年3月23日
学位授与の要件　学位規則第4条第1項該当
研究科・専攻　理学研究科生物科学専攻
学位論文題目　酵素反応分類アルゴリズムの開発とゲノムスケール解析への応用

論文調査委員　（主査）教授 金久　實　教授 藤吉好則　教授 平野丈夫

論文内容の要旨

ゲノム解析によって得られる遺伝子配列に対するアノテーション（機能注釈付け）は主に配列類似性に基づいており，酵素遺伝子に対してはNC-IUBMB（Nomenclature Committee of the International Union of Biochemistry and Molecular Biology）によるEC（Enzyme Commission）番号がアノテーション情報として与えられている。一方，EC番号は歴史的に酵素反応の分類として用いられており，酵素タンパク質が確実に同定されていない限り，EC番号が正式に与えられることはない。このためゲノムスケールでのシステマティックな配列解析には不適切な点が多い。とくに二次代謝パスウェイなど生物種固有の酵素反応に対しては問題点は特に深刻であり，酵素の機能を的確に定義し分類することの重要性が高まってきている。そこで申請者は，NC-IUBMBにより正式にEC番号が与えられない酵素反応であっても，EC番号と同等の酵素反応分類が行えるようなRC（Reaction Classification）システムの開発を行った。

RCシステムではまず，KEGG/REACTIONデータベース内にある酵素反応を，反応物ペア（基質・生成物ペア）の組み合わせとして表現した。そして各ペアを構成する2つの化合物の構造比較を行うことで，酵素反応により変化した官能基の情報を抽出している。この情報は，構造変化の中心（reaction center）とそのまわりの原子について，酵素反応で切断または生成された結合を表す3種類の原子の変換パターンで表現し，それをRC番号という識別番号で簡便に表した。

次に申請者は，既にEC番号が与えられている酵素反応に対してRC番号を算出し，それとEC番号との対応関係を調べた。そうすることで，新規な（EC番号が未定の）反応に対しても仮のEC番号をアサインすることができた。その信頼度はジャックナイフ交差検証法によって確認され，60%以上の反応に対し約90%の精度でEC番号の第3階層まで予測することができた。なお，開発したプログラムはe-zymeと名づけた予測システムとして，ゲノムネットでサービスを行っている。

さらに申請者はKEGG/SSDBデータベースにある相同遺伝子（Ortholog Cluster，OC）の情報を利用して，配列情報とEC番号およびRC番号との関係を調べた。RC番号とタンパク質配列とを関連づけるためには，EC番号を仲介する必要があるため，これらの間の関係を網羅的に調べることはできなかったが，EC番号では異なるカテゴリーに分類されていながら，反応（RC），配列（OC）ともに一致している例が見られた。このとき反応を構成するRC番号は必ずしも完全に一致する必要はなく，部分的な一致でも関連性を指摘できる例も多数見られた。

論文審査の結果の要旨

ゲノム配列中にコードされたタンパク質の機能を推定するために配列相同性指標が広く利用されているが，そのようにして推定された機能をアノテーションとして記述する際には，どのような記述法を選ぶかが重要である。酵素遺伝子の場合，現在のところEC番号を利用するほかに系統的な方法はない。

申請者は，酵素の分類を見直すことでゲノムのアノテーションや未知酵素の推定に寄与することを考え，RC（Reaction

Classification）法を提案した。この方法は，基質化合物の構造変化のみに基づき，酵素タンパク質の分類や命名等，タンパク質側の情報から完全に独立した分類である。ここではその内容を，EC 番号を予測するアルゴリズムとしての視点と，独自の酵素反応分類法としての視点とに分けて評価した。

申請者が開発した EC 番号予測アルゴリズムは，NC-IUBMB にて定義されたルールに従って酵素反応を仕分けするのではなく，既に KEGG/LIGAND データベース中に登録されている酵素反応の情報から EC 番号と酵素反応式の対応関係を記憶し，それを利用して新規反応に対する EC 番号を予測するものである。RC 法が酵素反応によって切断された結合しか考慮しないため，EC 側にそれ以外のルールが記述されている場合は正しく予測できない。また，RC 分類が細かすぎるため，予測可能な反応が約 6 割に過ぎない。これらの点は新規反応に EC 番号を割り当てるという点で改善すべきところである。しかし予測できた反応の信頼性は90%近くに達すること，ある反応が複数の EC カテゴリーに属する可能性を指摘できることや，EC 分類に新たなカテゴリーを用意する必要性を指摘できることを考慮すると，この分類法は一定の成果をあげていると認められる。

申請者はこの RC 法と EC 番号との関係を調べ，さらに酵素タンパク質のオーソログ遺伝子にあるアノテーション情報を解析することで RC の有用性に関する根拠を与えた。様々な生物種のゲノムが解読され，酵素タンパク質の構造機能相関が研究されるにつれ，EC 番号がタンパク質の配列や立体構造と対応していないと指摘する研究が多く報告されるようになってきた。ところが先行研究はどれもタンパク質の分類に終始しており，酵素反応そのものの分類を見直そうという研究は存在しなかった。申請者が行った解析はタンパク質の分類から完全に独立した酵素反応分類であり，それが反応メカニズムだけでなくタンパク質ファミリーとも関連していることを示唆する結果を得た。ただし申請者が述べているとおり，RC 法とタンパク質との対応関係を調べるには，現時点では EC 番号を仲介せざるを得ないため，網羅的に酵素の構造機能相関を調べるには至らなかった。しかしこの問題は，今後ゲノムデータベースのアノテーション情報を整備していく過程で解決可能であると考えられる。RC 法の強みは，部分的にしか明らかでない反応に対しても類似性を定義できることであり，部分的な情報が多量に得られるゲノムスケールの代謝パスウェイ解析に有効に働くことが期待される。

以上申請者が行った酵素反応分類法の開発は，基礎的な研究成果としても，ゲノム解析の実用的なツールとしても高く評価できる。よって，本論文は博士（理学）の学位論文として価値あるものと認める。なお，論文内容とそれに関連した試問の結果，合格と認めた。