

## 相互情報量と SVM を用いた酵素反応における EC 番号の推定手法の開発

松田祥彦<sup>†</sup> 伊藤将弘<sup>†</sup> 遠里由佳子<sup>†</sup>

EC 番号は酵素反応を分類するための階層的なシステムであり、酵素活性の予測や反応の類似性の指標として用いられる。しかしながら、EC 番号は酵素活性から、手動で割り当てられるため、反応によっては不完全あるいは全く割り当てがなされていないものもある。そこで、本研究では MCS アルゴリズムと相互情報量を用いた EC 番号の推定方法を提案し、さらに機械学習ならびに相関係数で妥当な EC 番号を推定した。本手法はジャック・ナイフ法の結果において最大で感度 (Sensitivity) 83%、精度 (Precision) 84%を示した。また、複数の EC 番号が割り当てられている反応に対しても推定を行い、EC 番号の候補リストの妥当性を確認した。

### Prediction of EC Numbers for Enzymatic Reactions Using Mutual Information and Support Vector Machine

Yoshihiko Matsuta<sup>†</sup>, Masahiro Ito<sup>†</sup> and Yukako Tohsato<sup>†</sup>

The computational prediction of protein catalytic functions is usually based on sequence similarities between enzymes. However, this method is questioned because small changes in key residues may alter the enzyme function. An alternative approach is to use classification systems such as Enzyme Commission (EC) numbers. Although EC numbers represent the hierarchical classification of an enzyme function and a catalyzed reaction, enzymatic reactions have incomplete or no EC numbers owing to the fact that they are manually assigned. In this study, we propose a new method to predict and assign valid EC numbers for unclear enzymatic reactions using a maximal common substructure (MCS) algorithm and mutual information (MI). In addition, we predicted EC numbers by using two methods, support vector machine (SVM) and correlation coefficient. Our proposed method yielded both high performance and high flexibility in predicting the EC numbers, performed sensitivity of 83% and precision of 84%, respectively. Furthermore, we predicted to the reaction assigned multiple EC numbers, and confirmed the validity of the candidate list of EC numbers.

### 1. はじめに

酵素反応とその反応を触媒する酵素には (Enzyme Commission) EC 番号<sup>1)</sup>という分類キーが割り当てられており、酵素機能や反応形式は階層的に分類されている。EC 番号は、反応特異性と基質の違いにより四階層で構成され、第一階層は機能別に大まかな反応特異性で分類される。例えば番号 1 は酸化還元、番号 2 は転移、番号 3 は加水分解であり、第一階層で六つに分類される。第二階層以降は、第一階層によって分類基準が異なるが、一般的には第二階層は酵素の基質特異性や反応特異性、第三階層は基質な結合の詳細な種類で割り振られ、第四階層では特定の酵素を示すシリアル番号となる。また、この EC 番号は酵素の構造や配列との関連性が見出されたことから<sup>2)</sup>、酵素活性の予測や反応の類似性の指標にも用いられてきた。先行研究<sup>3)</sup>では、EC 番号の三階層目が一致する酵素反応の組において、基質と生成物間の化合物変化パターン<sup>4)</sup>の類似性が示唆された。しかしながら、EC 番号は酵素活性から手動で割り当てられるため、酵素活性の知見が実験的に不十分な酵素には EC 番号が割り当てられていない場合がある。そこで代謝データベースに蓄積された知見を活かし、既知の酵素反応の情報から対象の反応における EC 番号を推定する手法がいくつか提案されている。例えば、E-zyme<sup>4)</sup>は、反応を構成する基質と生成物間における化合物変化を RDM パターンで表わし、RDM パターン間の相関を EC 番号ごとの出現数からなるベクトル間の相関として定義する。そして推定対象の反応から得られる RDM パターンと既知データとの相関により EC 番号の推定を行う。しかし、いずれのパターンにも合致しない場合には推定は行えない。また、MOLMAP という、自己組織化マップを利用して、基質と生成物間において生成・消失した結合によって反応を表現し、Random Forests という機械学習アルゴリズムを用いて推定を行う手法<sup>5)</sup>も提案されている。しかし、この手法は複数の EC 番号が割り当てられた反応や基質や生成物の構造の一部が不明瞭な反応は除外している。

そこで本研究では、相互情報量<sup>6)</sup>と SVM (Support Vector Machine)<sup>7)</sup>を用いた酵素反応における EC 番号の推定手法を提案する。提案手法は従来法<sup>3)</sup>と同様に基質と生成物間の化合物の構造変化に着目するが、あらかじめ定義された化合物変化パターンに合致しない場合でも、最も類似する化合物変化パターンを検索し、代用することで推定を可能とした。また、本手法は EC 番号の候補リストを生成するため、複数の EC 番号が割り当てられた反応においても妥当な EC 番号を推定することができ、さらに候補リストを反応類似性の指標として用いることも可能である。

<sup>†</sup> 立命館大学 生命科学部 生命情報学科

Department of Bioinformatics, College of Life Sciences, Ritsumeikan University, 1-1-1 Nojihigashi, Kusatsu, Shiga 525-8577, Japan

## 2. 提案手法

### 2.1 提案手法の手順

あらかじめ既知の反応から部分構造を抽出し、基質側と生成物側の部分構造を組にして、各部分構造の組と EC 番号間の相互情報量を算出し、既知データとして登録する。そして、任意の反応に対する EC 番号の推定を以下の手順で行う。

- (1) 部分構造の抽出  
推定対象の反応から部分構造を抽出する。
- (2) 相互情報量の算出  
基質側と生成物側の部分構造を組にし、EC 番号間の相互情報量を算出する。部分構造が登録されていない場合は最も類似する部分構造を検索し、代用する。
- (3) EC 番号の候補リストの生成  
相互情報量を用いて、対象の反応における各 EC 番号をスコア化し、候補リストを生成する。
- (4) EC 番号の推定  
反応に対する妥当な EC 番号を推定する。
  - 最大スコアによる推定
  - SVM による推定
  - 順位相関係数による推定

### 2.2 部分構造の抽出

酵素反応を構成する基質と生成物の分子構造間の最大共通部分構造を、MCS (Maximum Common Substructure) アルゴリズム<sup>8)</sup>により判定する。MCS は、二つの化合物間において一致する原子の結合の組をノードとしてグラフ上に配置し、隣接するノードを結んで構築される最大の部分構造を再構築する (図 1)。

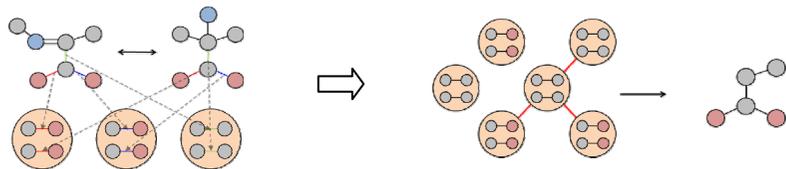


図 1 MCS による原子結合の組の配置と最大共通部分構造の再構築

そこで、MCS アルゴリズムにより、基質と生成物間の最大共通部分構造を「保存された部分構造」、残りの部分を「変化した部分構造」、保存された部分構造と隣接する部分を反応部位における「近傍構造」として抽出した (図 2)。このとき、MCS アルゴリズムでは判定できない、水、酸素、アミノニア、金属イオンなどの小分子の変化も、「小分子」変化として抽出する。以下では、これら保存・変化・近傍・小分子の抽出された部位の違いを「部分構造の種類」と呼ぶ。変化した部分構造は、基質側もしくは生成物側に複数生じる場合がある。

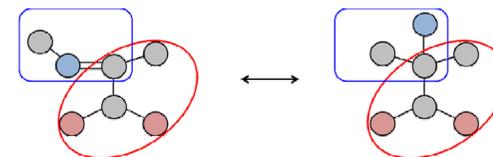
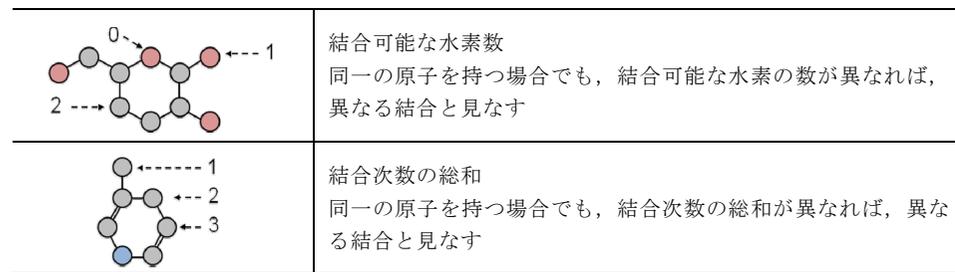


図 2 反応前後における保存された部分構造 (赤) と変化した部分構造 (青)

なお、一般的な MCS アルゴリズムは同じ原子の組み合わせの結合の有無でのみ一致を判定するが、ここでは結合の向きや原子の電荷といった属性を一致の判定に加えている (表 1)。

表 1 MCS の一致判定に加えた原子の結合の属性

	<p>最小環構造の結合数 結合の多重度が異なる場合でも、同じ原子数で構成される環構造に属していれば、同一の結合と見なす</p>
	<p>結合の向き 異性化反応の場合、同一の原子・結合を持つ場合でも、結合の向きが異なれば、異なる結合と見なす</p>
	<p>シス-トランス異性体 異性化反応の場合、同一の原子・結合を持つ場合でも、結合の向きが異なれば、異なる結合と見なす</p>
	<p>原子の電荷 同一の原子を持つ場合でも、形式電荷が異なれば、異なる結合と見なす</p>



### 2.3 部分構造のビット列化と組形成

抽出したすべての部分構造は、CDK (Chemistry Development Kit Ver. 1.2.5) を用いて、MACCS Key に基づき 166 桁のビット列に変換する。そして、基質側と生成物側で対応する部分構造をその種類 (保存・変化・近傍・小分子) 別に組を作成する。EC 番号を推定する対象の反応の部分構造が、既知のデータと一致しない場合は、最も類似する部分構造で代用する。そのため、一致しない部分構造のビット列と既知の部分構造のビット列間の類似度を式(1)の Tanimoto 係数<sup>9)</sup>を用いて算出し、最大値を持つビット列を探し代用する。

$$T(x,y) = \frac{N_z}{N_x + N_y + N_z} \quad (1)$$

ここで、 $N_x$ ,  $N_y$  はビット列  $x$ ,  $y$  における 1 のビット数であり、 $N_z$  は  $x$  と  $y$  がともに 1 のビット数である。定義より、 $T(x,y)$  は 0 から 1 までの値をとり、1 に近づくほど二つのビット列間の類似度は高い。

### 2.4 部分構造の組と EC 番号間の相互情報量の算出

相互情報量は二つの確率変数間の相互依存性を表す値であり、この値を用いて、抽出した部分構造の組と各 EC 番号間の関連性を数値化する。部分構造の組  $S_i$  ( $1 \leq i \leq n$ ) と EC 番号  $E_j$  間の相互情報量  $I(S_i, E_j)$  は式(2)で算出される。 $I(S_i, E_j)$  は非負の実数であり、値が大きいほど対象の部分構造の組  $S_i$  が対象の EC 番号  $E_j$  を推定する上で重要な部分構造の組となる。

$$I(S_i, E_j) = \sum_{e_s \in \{1,0\}} \sum_{e_e \in \{1,0\}} P(S_i = e_s, E_j = e_e) \log_2 \frac{P(S_i = e_s, E_j = e_e)}{P(S_i = e_s)P(E_j = e_e)} \quad (2)$$

### 2.5 相互情報量を用いた EC 番号の候補リストの生成

相互情報量を用いて、反応に割り当てられている全 EC 番号をスコア化する。まず、EC 番号を推定する対象の反応において、その反応を構成する基質と生成物から取得できる部分構造の組  $S = \{S_1, S_2, \dots, S_n\}$  を抽出する。次に、式(3)にしたがって EC 番号ごとに各部分構造の組との相互情報量を加算する。対象の EC 番号  $E_j$  において、 $S_i$  の出現する回数が全 EC 番号の平均値を上回ればスコアをより大きく、そうでない場合はより小さくする重みとして、シグモイド関数を用いた。対象の反応において全 EC 番号をスコア化し、スコアの降順で並べた EC 番号の「候補リスト」を生成する (図 3)。

$$W(S, E_j) = \sum_{S_i \in S} I(S_i, E_j) \times e^{x_i - \bar{x}} \quad (3)$$

#### 2.6 節

SVM で判別する EC 番号

EC 番号	相互情報量 $W$
EC 3.6.1	20.65
EC 3.1.3	5.478
EC 6.3.4	2.161
⋮	⋮
EC 1.6.1	$7.713 \times 10^{-6}$

#### 2.5 節

相互情報量による推定値

#### 2.7 節

順位相関係数で利用する EC 番号の並び

図 3 推定対象の反応に対する EC 番号の候補リストの例

### 2.6 SVM による推定

前節に示した相互情報量による候補リストの作成段階では、EC 番号間の依存性を考慮しない。そこで EC 番号の推定に SVM を導入した。SVM は機械学習の一種であり、既知データを二分類するための境界線を決定できる。具体的には、EC 番号の候補リストから相互情報量が 1 位と 2 位の EC 番号を選択し、その EC 番号に属する他の酵素反応を SVM に学習させる。その際、各反応は、出現する部分構造の組を種類別に論理和をとり連結した長さ 1328 のビット列として与える (図 4)。これにより、反応を構成する基質と生成物から抽出された部分構造の組全体の違いによって、どちらの EC 番号に属するかを判別することができる。



図 4 SVM に学習させる酵素反応におけるビット列表現

SVM プログラムは LibSVM を用いた。そして、式(4)によって定義される制約条件を満たし、この式を最大化するラグランジュ乗数  $\alpha$  を求め、 $\alpha$  を用いて非線形判別関数を求める。

$$\max_{\alpha} L_D(\alpha) = \max \left\{ \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j y_i y_j K(x_i, x_j) \right\} \quad (4)$$

$$\text{制約条件} \quad \sum_{i=1}^n \alpha_i y_i = 0 \quad 0 \leq \alpha_i \leq C \quad i, j = 1, 2, \dots, n$$

$x$  はデータ、 $n$  はデータ数、 $y$  は  $x$  に対するラベル値、 $K$  は学習データを高次元の特徴空間に写像する際に計算を簡略化するためのカーネル関数である。本手法では、式(5)で表わされる RBF (Gaussian radial basis function) カーネルを用いた。

$$K(x_i, x_j) = \exp(-\gamma \|x_i - x_j\|^2) \quad \gamma \geq 0 \quad (5)$$

RBF カーネル中の  $\gamma$  の値が小さいほど境界線の非線形性が増加する(図 5 (a))。また、SVM では線形分離可能でないデータに対して、許容可能な誤差の和が設定されており、パラメータ  $C$  を用いてある程度制約を緩くすることができる。 $C$  の値はマージンの大きさに関与し、大きいほどマージンは狭くなり、誤分類が許されなくなる(図 5 (b))。マージンの大きさと誤分類の確率はトレードオフであり、分類する問題に対して、できるだけマージンを大きくとり、かつ誤分類が少なくなるような  $C$  を探索する必要がある。これらのパラメータは任意の範囲を格子点探索する方法 (grid-search) <sup>10)</sup> によって調整する。

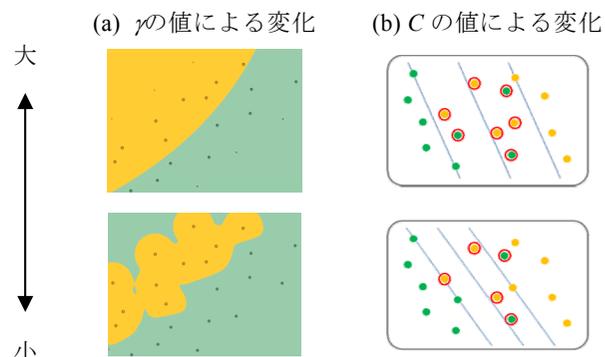


図 5 (a)  $\gamma$  の値による境界線の非線形性の変化と、(b)  $C$  の値によるマージンの大きさと誤分類数の変化

## 2.7 順位相関係数による推定

EC 番号の候補リスト間において、式(6)で定義されるケンドールの順位相関係数を用いて相関を算出する。 $n$  は全 EC 番号の数、 $P$  は二つの候補リストにおいて、大小関係が一致する組の数である。最も相関のある反応に割り当てられている EC 番号を順位相関係数による推定値とする。

$$\tau = \frac{4P}{n(n-1)} - 1 \quad (6)$$

## 3. 実験と結果

### 3.1 実験データ

代謝データベースである KEGG (Kyoto Encyclopedia of Genes and Genomes) <sup>11)</sup> (Release 58.1) の反応データを利用する。表 2 は利用する反応数、化合物数および抽出した一意な部分構造の数を示し、表 3 に取得した反応における EC 番号の一階層目ごとの数を示す。

表 2 実験で用いたデータ数

反応	5644
化合物	4978
保存された部分構造	2404
変化した部分構造	8057
小分子	328

表 3 EC 番号の一階層目別の反応数

EC 1	酸化還元	2015
EC 2	転移	1816
EC 3	加水分解	917
EC 4	付加脱離	535
EC 5	異性化	213
EC 6	合成	202

### 3.2 EC 番号別にみた頻出する部分構造

第一階層別に出現数の最も多い部分構造の組を表 4 に示す。各部分構造の組は EC 番号の定義に即していることが確認できる。

表 4 各 EC 番号において出現数の多い部分構造

EC 番号	部分構造の組	二階層目と三階層目の機能分類
EC 1.1.2	⇔	CH-OH に対し、酸化酵素として働く 酸素を用いる
EC 2.7.1	⇔	リンを含む基を移す アルコールにつなげる
EC 3.1.1	⇔	エステル加水分解 カルボン酸エステル加水分解

EC 4.1.1		C-C リアーゼ カルボキシル基の付加脱離
EC 5.1.3		光学異性の変換 炭水化物及びその類縁体に作用
EC 6.2.1		C-S 結合を形成するもの 酸とチオールを結合するもの

### 3.3 SVMのパラメータ決定

SVM が最適な分類を与えるようにパラメータ $\gamma$ と $C$ を grid-search によって決定した (図 6).  $2^{-15} \sim 2^{-1}$  の範囲で $\gamma$ を,  $2^{-5} \sim 2^{15}$  の範囲で $C$ をとり, 反応データを学習させて各 EC 番号を推定し, 正答率が向上する組み合わせを選択した. grid-search の結果から本手法では $\gamma=0.0004$ ,  $C=128$  とした.

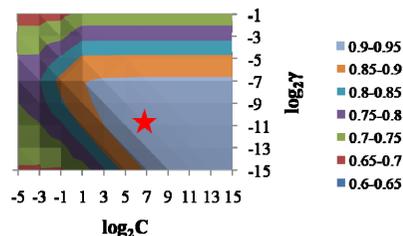


図 6  $\gamma$ と $C$ に対する正答率の変化

### 3.4 EC 番号推定の評価方法

提案手法の評価を行うためにジャック・ナイフ法によって全反応の EC 番号を推定し, 表 5 の定義を用いて感度 (Sensitivity) と精度 (Precision) を算出する (式(7)(8)). 提案手法および既存の推定手法は感度と精度を比較することで評価する.

表 5 TP, FP, FN, TN の定義

	対象の EC 番号に属する	対象の EC 番号に属さない
対象の EC 番号と推定結果が同じ	TP	FP
対象の EC 番号と推定結果が異なる	FN	TN

$$\text{Sensitivity} = \frac{TP}{TP + FN} \quad (7)$$

$$\text{Precision} = \frac{TP}{TP + FP} \quad (8)$$

### 3.5 感度と精度の比較

提案手法の各段階 2.5~2.7 節で求められる EC 番号をそれぞれ  $EC_{mi}$ ,  $EC_{svm}$ ,  $EC_{kdl}$  とし, 各推定結果から 3.4 節の評価方法で感度と精度を算出した. 表 6 では結果を各 EC 番号の一階層目までまとめている. 提案手法において  $EC_{kdl}$  が全体の精度が最も高いが, EC 番号の一階層目ごとに見ると, EC 5 の精度が  $EC_{mi}$ ,  $EC_{svm}$  と比較して大きく減少し

ている. EC 5 には分子内での酸化還元反応や転移反応が存在し, 候補リストの並びが類似していても反応数の多い他の EC 1 や EC 2 に属する可能性が高いため, EC 5 の推定に限っては  $EC_{svm}$  の方が適切であると考えられる.

また,  $EC_{kdl}$  での推定結果と既存の推定手法との比較を行った (表 7). 本手法では登録されていない部分構造の組が反応から得られた場合でも, 類似する部分構造の組を検索し, 代用するため網羅率 (Coverage) は 1 であり, 感度も高いため, より柔軟性があると言える.

表 6 提案手法の感度と精度

推定段階	指標	全体	EC 1	EC 2	EC 3	EC 4	EC 5	EC 6
$EC_{mi}$	感度	0.784	0.797	0.877	0.701	0.688	0.467	0.772
	精度	0.796	0.789	0.844	0.734	0.779	0.861	0.700
$EC_{svm}$	感度	0.817	0.814	0.872	0.816	0.736	0.608	0.796
	精度	0.830	0.814	0.885	0.792	0.802	0.878	0.726
$EC_{kdl}$	感度	0.831	0.843	0.874	0.820	0.742	0.670	0.777
	精度	0.847	0.854	0.901	0.823	0.769	0.670	0.800

表 7 提案手法と既存の推定手法との比較

手法	反応数	網羅率	感度	精度
提案手法	5644	1.000	0.831	0.847
E-zyme	5669	0.934	0.794	0.851
MOLMAP	3671	-	-	0.850

### 3.6 部分構造の組が登録されていない反応に対する評価

EC 番号を推定する反応の部分構造の組が, 既知データに存在しない反応に対しての感度と精度を示す (表 8).

表 8 部分構造の組が登録されていない反応に対する感度と精度

推定段階	指標	全体	EC 1	EC 2	EC 3	EC 4	EC 5	EC 6
$EC_{mi}$	感度	0.709	0.670	0.859	0.584	0.563	0.237	0.733
	精度	0.716	0.628	0.828	0.610	0.657	0.767	0.647
$EC_{svm}$	感度	0.756	0.709	0.861	0.740	0.618	0.330	0.762
	精度	0.764	0.681	0.862	0.709	0.715	0.762	0.635

$EC_{kdt}$	感度	0.749	0.707	0.850	0.734	0.618	0.371	0.695
	精度	0.760	0.709	0.856	0.749	0.642	0.419	0.689

全データの平均である表6と比べても表7の感度と精度は大きく低下していない。EC番号の一階層目ごとに見ると、EC 2 は特に感度と精度が高く保たれている。これはEC 2 の反応数が多いため、様々な炭化水素鎖の転移を表す部分構造の組を利用することができ、類似する転移反応は部分構造の組を見つけやすいためである。一方、EC 5 は著しく感度が低下している。これはEC 5 の反応数が少なく、ビット列内に結合の向きや分子内での変化の情報が含まれていないためである。これは、表1で定義した異性化反応のための属性値と同様に、ビット列内に三次元座標や結合の向きを表すビットを加えることで改善できると考えられる。

### 3.7 複数の EC 番号が割り当てられた反応

取得したデータには、EC 番号の三階層目までの番号が複数割り当てられている反応が85個含まれていた。表9は各段階での推定結果を表し、推定したEC番号が複数割り当てられているEC番号のうちのいずれかである場合と、複数割り当てられているEC番号が過不足なく候補リストのスコア上位に存在する場合の反応数を示している。

表9 複数のEC番号が割り当てられた反応におけるEC番号の推定結果

推定段階	スコア上位に過不足なく存在	いずれかと一致
$EC_{mi}$	42	69
$EC_{svm}$	42	71
$EC_{kdt}$	47	72

いずれかのEC番号が最大のスコアを持つ場合は多いが、複数のEC番号が過不足なくスコア上位を占める割合は半分程度である。これは割り当てられているEC番号に属する反応が推定対象の反応のみといった、特定のEC番号に対するデータ数が少ないことに起因する。

## 4. おわりに

酵素反応におけるEC番号を推定する手法を開発した。提案手法は推定対象の反応から得られる部分構造の組が登録されていない場合でも、類似する構造を検索するため柔軟性が高いと言える。また、複数のEC番号が割り当てられた反応に対しても、候補リストの上位を占める割合が多く、推定が妥当であることが確認できた。しかしながら、他のEC番号と比較すると、異性化反応 (EC 5) に対する感度は低く、変換

するビット列に異性化反応に特化した情報を加えるなどの改善が必要である。今後、誤分類を起こした反応を詳細に調べ、提案手法の精度向上を目指すとともに、EC番号の候補リストを反応類似性の指標として用いることで、酵素反応とその反応を触媒する酵素との関連性を見出すことを目指す。

**謝辞** 本研究の一部は、科学研究補助金 (若手研究(B) 課題番号 23700353) による。

## 参考文献

- 1) IUPAC-IUBMB Joint Commission on Biochemical Nomenclature (JCBN) and Nomenclature Committee of IUBMB (NC-IUBMB), *European Journal of Biochemistry*, Vol. 264, pp. 607–609 (1999). <http://www.chem.qmul.ac.uk/iubmb/enzyme/>
- 2) Tracey Bray, Andrew J. Doig, and Jim Warwicker: Sequence and Structural Features of Enzymes and their Active Sites by EC Class, *J. Mol. Biol.*, Vol. 386, pp. 1423–1436 (2009).
- 3) Yukako Tohsato, and Yu Nishimura: Reaction Similarities Focusing Substructure Changes of Chemical Compounds and Metabolic Pathway Alignments, *IPSJ Transactions on Bioinformatics*, Vol. 2, pp.15-24 (2009).
- 4) Yoshihiro Yamanishi, Masahiro Hattori, Masaaki Kotera, Susumu Goto, and Minoru Kanehisa: E-zyme: Predicting Potential EC Numbers from the Chemical Transformation Pattern of Substrate-product Pairs, *Bioinformatics*, Vol.25, pp.179–186 (2009).
- 5) Diogo A. R. S. Latino, and Joao Aires-de-Sousa: Assignment of EC Numbers to Enzymatic Reactions with MOLMAP Reaction Descriptors and Random Forests, *J. Chem. Inf. Model*, Vol. 49, pp. 1839–1846 (2009).
- 6) Ralf Steuer, Jürgen Kurths, Carsten O. Daub, Janko Weise, and Joachim Selbig: The mutual information: Detecting and Evaluating Dependencies between Variables, *Bioinformatics*, Vol. 18, pp. 231–240 (2002).
- 7) Christopher J. C. Burges: A Tutorial on Support Vector Machines for Pattern Recognition, *Data Mining and Knowledge Discovery*, Vol. 2, pp. 121–167 (1998).
- 8) Christian Tonnelie, Philippe Jauffret, Thierry Hanser and Gérard Kaufmann.: Machine Learning of Generic Reactions: 3. An Efficient Algorithm for Maximal Common Substructure Determination, *Tetrahedron Computer Methodology*, Vol. 3, pp. 351–358 (1990).
- 9) オープンバイオ研究会: オープンソースで学ぶバイオインフォマティクス, 東京電機大学, pp. 160–162 (2008).
- 10) Chih-Wei Hsu, Chih-Chung Chang, and Chih-Jen Lin: A Practical Guide to Support Vector Classification, *Bioinformatics*, Vol. 1, pp. 1–16 (2010).
- 11) Minoru Kanehisa, Susumu Goto, Miho Furumichi, Mao Tanabe, and Mika Hirakawa: KEGG for Representation and Analysis of Molecular Networks Involving Diseases and Drugs. *Nucleic Acids Research*, Vol. 38, pp. D355–D360 (2010). <http://www.genome.jp/kegg/>