

機関番号：10101
 研究種目：若手研究（B）
 研究期間：2008～2010
 課題番号：20700263
 研究課題名（和文） 分子間相互作用の相補性に基づく動径分布関数と自己組織化写像による酵素機能予測
 研究課題名（英文） Predicting enzyme functions based on complementarily in molecular interactions by using radial distribution functions with a self-organizing maps
 研究代表者
 上野 恵介（UENO KEISUKE）
 北海道大学・大学院情報科学研究科・博士研究員
 研究者番号：90435983

研究成果の概要（和文）：本研究では、酵素活性部位の動径分布関数と自己組織化写像に基づく機能予測手法を開発した。動径分布関数は酵素活性部位の構造上の僅かな違いを反映し、自己組織化写像と組み合わせることにより、酵素活性部位の類似性を示した。酵素活性部位の類似性に基づく本手法により、アミノ酸配列や立体構造の類似性では検出できない酵素の機能を予測することができた。また、酵素活性部位の類似性は酵素の立体構造よりも基質反応の類似性をよく反映していることがわかった。

研究成果の概要（英文）：We developed a method to predict enzyme functions by using radial distribution functions of active sites with a self-organizing map. The method outperformed the structural and sequence alignments with their low similarities.

交付決定額

（金額単位：円）

	直接経費	間接経費	合計
2008年度	700,000	210,000	910,000
2009年度	600,000	180,000	780,000
2010年度	500,000	150,000	650,000
年度			
年度			
総計	1,800,000	540,000	2,340,000

研究分野：総合領域

科研費の分科・細目：情報学・生体生命情報学

キーワード：酵素機能予測、動径分布関数、自己組織化写像、分子間相互作用

1. 研究開始当初の背景

ゲノムプロジェクトの進展を背景として、大量の遺伝子配列が得られるようになり、遺伝子配列にコードされているタンパク質の機能を予測することが生物学の重要な課題となっている。

生命情報科学（Bioinformatics）は、これらの課題に対して、機能が既知の遺伝子との相同性を解析するための計算手法を開発してきた。これらの相同性の解析手法には、タンパク質のアミノ酸配列に基づく方法と立体構造に基づく方法がある。タンパク質のアミノ酸配列に基づく方法は、進化的に起源の

異なる類似した機能をもつタンパク質の機能予測には適用できないことが知られている。このような“収斂進化”によって機能を獲得したタンパク質の機能予測ではタンパク質の立体構造を考慮する必要がある。しかし、ラクトアルブミンとリゾチームのように類似した立体構造を持つにもかかわらず、機能が異なるタンパク質が存在することが知られている。これらのタンパク質の機能を予測するには、スプチリシンとトリプシンでみられるような機能を決定づける数アミノ酸残基の同定が重要である。

これら生命情報科学における機能予測の

問題を顕著に表しているのが、オーファン酵素 (orphan enzyme) やオーファン受容体 (orphan receptor) の存在である。これらのタンパク質では実験的に機能がわかっているにもかかわらず、この機能を果たしている遺伝子が同定できていない。

2. 研究の目的

本研究では、機能を決定づける酵素の活性部位に注目し、活性部位の物理化学的な性質を用いた酵素機能の予測方法の開発を目的としている。

機能の構成単位である酵素の活性部位に着目することは、収斂進化により獲得した酵素の機能を予測する上で必須であり、その類似性を測ることは、酵素機能間の機能の関連性を理解する上でも重要である。また、酵素活性部位の物理化学的な性質に着目することは、オーファン酵素の機能予測にとっても重要である。なぜなら、タンパク質の機能は分子間の物理化学的な相互作用に基づいているからである。このため、相互作用する低分子化合物やその反応を触媒する酵素機能の類似性は、酵素活性部位の物理化学的な類似性を反映したものであると考えられる。これら酵素側と基質反応側の2種類の類似性を比較することにより、オーファン酵素の機能予測が可能になると考えられる。

3. 研究の方法

本研究における酵素の活性部位の物理化学的な性質に基づいた酵素機能予測手法は、化学情報科学 (Cheminformatics) の手法を生命情報科学の問題へ適用したものである。生命情報科学がタンパク質のアミノ酸配列や立体構造間の比較を対象に発展してきたのとは対照的に、化学情報科学は分子の構造や物性の表現を中心に発展してきた。

(1) 酵素活性部位の動径分布関数

化学情報科学における動径分布関数 (RDF: Radial Distribution Functions) は、低分子化合物の類似構造の高速な探索に用いられてきた。本研究では RDF を酵素の活性部位の表現として用いた。酵素の活性部位の RDF は反応中心からの距離 r を変数として次式で表される。

$$g(r) = \sum_i^N p_i e^{-B(r-r_i)^2}$$

ここで、 p_i は部分電荷などの原子 i の特性、 B は平滑化因子、 r_i は原子間距離である。

図1はグリセルアルデヒド3リン酸脱水素酵素 (GAPDH) の RDF を示している。Cys 残基は求核基、His 残基は Cys 残基の活性化、Asn 残基と Arg 残基は補酵素や基質の結合にかかわっている。RDF により表現された酵素の活性部位は、適切な距離 r で値を計算すること

により、数十次元の数値データとして、解析に用いられた。

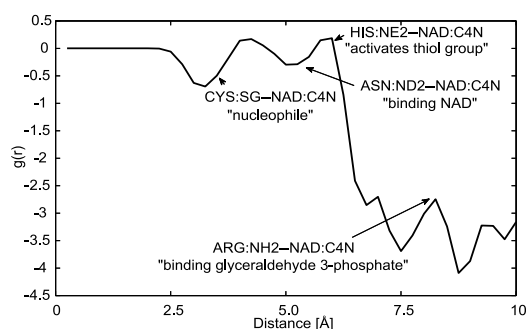


図1 GAPDH の RDF の例

(2) 自己組織化写像 (SOM: Self-Organizing Map)

酵素活性部位の RDF は Kohonen (1984) により提案された教師なし機械学習である SOM により、代表的なパターンに分類された。SOM は分類されたクラス間の類似性を保持する特徴を持ち、類似した酵素活性部位の入力データは近くに配置される。

4. 研究成果

(1) 酵素活性部位の分類

図2は4092個の酵素の活性部位の構造を分類したものである。赤色は GAPDH、緑色はシトクローム c ペルオキシダーゼ (CCP)、青色はその他の酵素、白色は該当する酵素がない場所を示している。これらの酵素が特定の場所に局在していることがわかる。

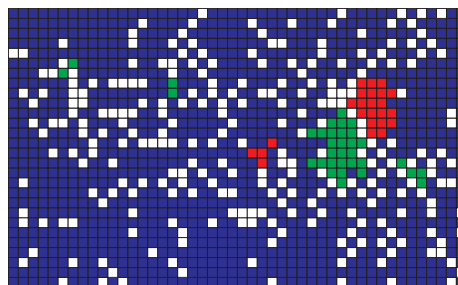


図2 4092個の酵素活性部位で訓練された SOM

(2) 酵素活性部位の照合

RDF で訓練された SOM が酵素の活性部位の類似性をどれだけ反映しているのか調べた。図3は SOM 上で近接して配置された構造を重ね合わせたものである。図3Aの構造は SOM 上で (5, 38) と (9, 38) に配置された GAPDH の2つの構造を重ね合わせたものである。これらの活性化部位の構造上の違いは求核基として働く活性残基 Cys 残基が Ser 残基に変異しているだけである。SOM 上でこれらの活性部位は近傍に配置されているが、やや離れた場所に配置された。図3Bは SOM 上で (8, 36)

と (13, 34) に配置された CCP の 2 つの構造を重ね合わせたものである。これらの活性部位では酵素活性に必要なヘムと結合する His 残基が片方で欠損している。これらの活性化部位も SOM 上で近傍に配置されているが、やや離れたところに配置された。

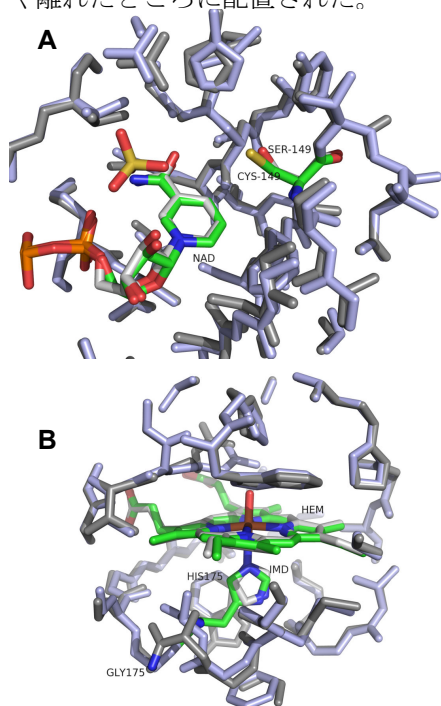


図 3 活性部位の構造 (A)GAPDH, (B)CCP

(3) 野生型と変異型の活性部位の比較

酵素活性部位の照合結果について、より詳細に評価するために、UniProt で共通の ID を持つ野生型と変異型の活性部位の違いを検出できるのか調べた (表 1)。解析には SOM 上の距離をスコアとして数値化したものを用いた。小さい値は活性部位が類似していることを、大きな値は類似していないことを示している。ここで、一致 (野生型-野生型, 変異型-変異型) と不一致 (野生型-変異型) の大小関係に着目すると、一致において、より小さな値であり、野生型と変異型の違いを検出できていることがわかる。

表 1 野生型と変異型の活性部位の距離

UniProt	補酵素	一致	不一致
CCPR	HEM	2.5	2.6
CHOD	FAD	13.1	18.6
FPRA	FAD	3.6	1.8
	NDP	9.0	15.9
FRDA	FAD	1.7	15.1
G3P	NAD	1.0	2.3
IDH	NAP	5.7	4.7
MDH	NAD	26.9	16.1
NIA1	FAD	0	12.2
OYE1	FMN	2.1	2.5

(4) 酵素活性部位と機能の関連性

ここでは、酵素の機能分類として手作業による分類体系である酵素 (EC) 番号と立体構造分類データベース (SCOP) を評価の対象として解析を行った。注意すべきことは、EC 番号は反応、つまり基質 (低分子化合物) 側の分類体系であり、酵素自身の分類体系ではないということである。ここに、酵素自身の情報に基づいた機能予測の困難さの原因がある。これには、分子間相互作用の相補性に基づいた方法が必要である。以下の評価実験では EC 番号を正しく予測することが目標である。

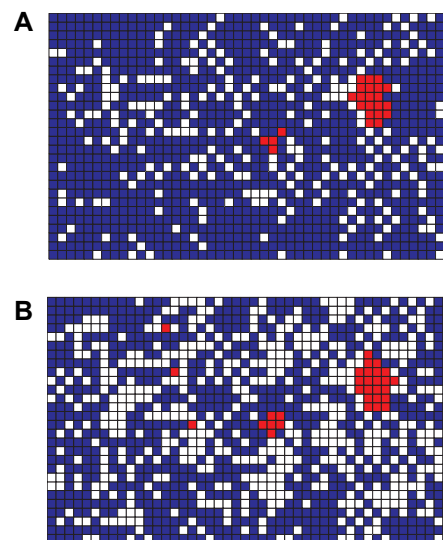


図 4 2 つの分類体系によりラベル付けした SOM (A) EC1.2.1.21, (B) SCOP c81.1.1

図 4 は EC1.2.1.21 と SCOP c81.1.1 の分類結果について赤色で示している。SCOP による分類より EC 番号による分類の方が小さな領域に局在していることがわかる。このことから、酵素活性部位による分類は構造よりも機能をより反映していることが示唆される。

(5) 酵素機能予測

本手法の性能を客観的に評価するため、タンパク質のアミノ酸配列や立体構造が類似していない酵素 (類似度 25%以下) を対象に、SOM 上での分類クラス内の距離と分類クラス間の距離の分離度を調べた。なぜなら、より分離能力が高いものが、より良い予測結果を与えるからである。分離する能力は感度と特異性で表される。感度と特異性はトレードオフの関係にあるため、類似性の判定の閾値により変化し、受信者動作特性 (ROC) 曲線として表される。最適な閾値は傾き 1 の直線との接線で表され、全体の評価は曲線下の面積で表される。

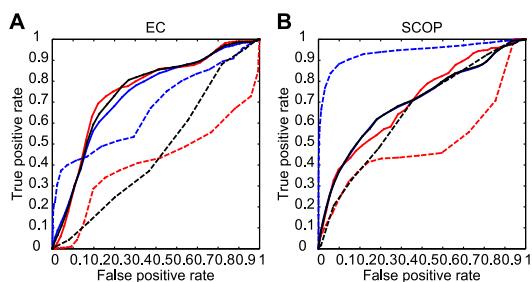


図5 ROC 曲線 (A)EC 番号, (B)SCOP

図5は、EC番号とSCOP分類についての分離度を表すROC曲線を示している。縦軸は感度、横軸は1-特異性を表している。実線はSOMの結果、破線は従来のアラインメントに基づく方法の結果である。青色は、グローバルアラインメント(Needleman-Wunschアルゴリズム)、黒色はローカルアラインメント(Smith-Watermanアルゴリズム)、赤色は構造アラインメント(MAMMOTH)の類似度25%以下のデータを評価に用いたことを示している。図5BのSCOPによる分類では、局所的アラインメントの結果が最も良く、SOMを用いた方法での大きな改善は認められなかった。一方、図5AのEC番号による分類ではSOMを用いた結果が最も良いことが分かった。このことから、SOMはタンパク質のアミノ酸配列や立体構造では類似性が認められない類似した機能の酵素を検出できることが示唆される。また、SOMが構造より機能を良く分離していることが示唆された。

そこで、他の類似性尺度との違いを調べるため、類似性尺度間の相関係数を調べると、SOMを用いる方法は他の方法と弱い相関しか持たないことが分かった(表2)。

表2 スピアマンの順位相関係数

類似度1	類似度2	相関係数
MAMMOTH	Needleman-Wunsch	0.801
MAMMOTH	Smith-Waterman	0.773
MAMMOTH	SOM*	-0.395
Needleman-Wunsch	Smith-Waterman	0.968
Needleman-Wunsch	SOM*	-0.419
Smith-Waterman	SOM*	-0.422

MAMMOTH (構造アラインメント), Needleman-Wunsch (グローバルアラインメント), Smith-Waterman (ローカルアラインメント)

*SOMは非類似性尺度であるので、類似性尺度である他のものとは負の相関を示す。

(6) 総括

酵素活性部位のRDFとSOMを用いた機能予測方法を開発した。本手法はタンパク質のアミノ酸配列や立体構造を用いた機能予測方法では発見することが困難な機能の類似性を検出することができた。今後の発展としては、酵素以外のタンパク質の機能予測への応用が考えられる。

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)

[学会発表] (計3件)

- ① 上野恵介, Enzyme function classified with a catalytic-site similarity, CBI学会2009年大会(CBI-KSBSB JOINT CONFERENCE), 2009年11月5日, 韓国(海雲台グランドホテル)
- ② 上野恵介, Physicochemical Property of Enzyme-Ligand Interface for Molecular Interaction, 2008年日本バイオインフォマティクス学会年会, 2008年12月16日, 大阪(千里ライフサイエンスセンタービル)
- ③ 上野恵介, Landscaping Enzymatic Reactions of Oxidoreductases by a Self-Organizing Neural Network, 第17回国際タンパク質化学構造解析会議, 2008年8月28日, 札幌(北海道大学学術交流会館)

6. 研究組織

(1) 研究代表者

上野 恵介 (UENO KEISUKE)

北海道大学・大学院情報科学研究科・博士研究員

研究者番号: 90435983

(2) 研究分担者

なし

(3) 連携研究者

なし