科学研究費助成事業　　研究成果報告書

平成 ２６ 年　 ６ 月　 ５ 日現在

機関番号：８２６２６
研究種目：基盤研究(C)
研究期間：2011～2013
課題番号：２３５００３７３

研究課題名（和文）補酵素結合様式を考慮した次世代活性部位探索アルゴリズムの開発

研究課題名（英文）Development of novel algorithm to detect enzyme active-sites considering cofactors

研究代表者
長野　希美（Nagano, Nozomi）

独立行政法人産業技術総合研究所・生命情報工学研究センター・主任研究員

研究者番号：７０３５７６４８

交付決定額（研究期間全体）：（直接経費）　　3,900,000円、（間接経費）　　1,170,000円

研究成果の概要（和文）：酵素に分類される蛋白質は活性部位と呼ばれる局所的な部位のみによって機能を果たすため、同じ機能を持つ酵素を見つけるために類似の形状を持つ活性部位を探索する必要がある。従来は、ある機能の活性部位を見つけるために、その機能を持つ既知の活性部位における原子間の二乗距離の重みなし平均距離によって形状の類似性を評価していた。これに対して、本研究では重み係数を導入して原子間の二乗距離の線形結合で評価した。この重み係数の学習にはBregman Divergence Regularized Machine を導入した。このようにして求めた重みつき偏差を使うと良好な探索性能が得られることを実験的に示した。

研究成果の概要（英文）：Prediction of active sites in enzyme proteins is extremely essential not only for protein sciences but also for practical applications. Because enzyme reaction mechanisms are based on the local structures of enzyme active sites, a simple measurement, mean square deviation, has been used to compare such local structures in proteins so far. In order to improve the ability of such a simple measurement, various kinds of template-based methods that compare the local sites have been developed to date. In this work, parameters for the deviation was introduced. Moreover, the Bregman Divergence Regularized Machine was also employed to develop a new machine learning algorithm that determines the parameters of the square deviation. Experimental results showed that the proposed methods possess promising search performance.

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　活性部位　予測法　重みつき偏差　カーネル　補酵素　構造

１．研究開始当初の背景

局所構造の比較法としては、テンプレート解析法(Wallace et al., 1997; Kleywegt, 1999)がある。Wallace らの TESS テンプレート解析法は、原子レベルの局所構造テンプレートを作成し、大規模な構造データベースに検索をかける手法である。Kleywegt の SPASM テンプレート解析法では、アミノ酸残基レベルの局所構造のテンプレートを作成する。しかし、こうした方法は次のような問題点がある：(a)テンプレート構築(どの原子はテンプレートに含め、どの原子を外すかという選別作業)は、職人的手作業で行うことを前提とする。このため、予測性能はテンプレート作成者の能力に大きく依存する；(b) テンプレートに含まれるそれぞれの原子には酵素反応に重要なものと重要でないものがあるにも関わらず、機能未知の蛋白質との重みなし平均原子間距離を距離尺度として評価している；(c) 機能未知の蛋白質の中にはリガンドとの相対的な位置関係も分かっているものもあるが対応していない。以上の欠点により既存のテンプレート解析法は膨大な偽陽性を産むことになる。これまで、このような局所構造比較法に、機械学習を適用している例はなかった。

リガンド化合物が結合することによって、酵素など蛋白質の立体構造は、コンフォメーション変化を起こすことが知られている。実際に 60 種類の酵素の立体構造を解析した結果、触媒残基（もしくは所謂、活性部位）は比較的変動が小さいのに対し、結合残基は、リガンド結合に伴いコンフォメーションが相対的に大きく変動するが、蛋白質によってはその程度も異なる(Gutteridge & Thornton, 2005)。しかも、触媒残基や非機能性残基と比べて、結合残基の主鎖の Cα 原子の変動が大きいことも判明した(Gutteridge & Thornton, 2005)。こうしたことが、リガンド結合部位を予測するのが難しい一因と考えられる。

また､5 種類以上のスーパーファミリーに結合するリガンド分子（ATP、NAD、ヘムなど 9 種類のリガンド）の結合部位を、蛋白質部分だけでなく隣接するリガンドの物理化学的な性質も含めて解析した結果も報告されている（Kahraman et al.,2010)。その報告では、静電ポテンシャルと疎水性スコアを指標にして解析した結果、それぞれのリガンドは、非常に多様な環境に結合していることが判明した（Kahraman et al.,2010)。そうしたリガンドの結合ポケットの形状も多様性がある（Kahraman et al.,2007)。ATP、NAD、FAD の 3 種類のリガンド分子のコンフォメーション自体も、非常に多様性があることが判明している(Stockwell & Thornton, 2006)。

しかしながら、上記のように、対応するリガンド非結合状態を考慮した蛋白質側の詳細な解析は、今のところなされていない。機能予測を行うためには、こうした点も踏まえてリガンドとその周辺のアミノ酸残基の状態を解析・分類する必要がある。

酵素に関する問題点を踏まえて、研究代表者・長野は、酵素触媒機構データベース、EzCatDB を開発し 2004 年 10 月に一般公開を開始している（Nagano, 2005)。この EzCatDB では、従来の EC 分類とは異なる酵素の立体構造や触媒機能を考慮した独自の酵素反応階層分類（RLCP 分類）や活性部位のアノテーションを行ってきた。また、研究代表者（長野）は、131 スーパーファミリーに属し、立体構造が解かれた加水分解酵素や転移酵素を含む 270 種類の酵素の触媒機構を解析した結果、異なるスーパーファミリーに属する酵素でも類似する活性部位を有する例をいくつか見つけた（Nagano et al., 2007)。このような結果から、進化的類縁関係がない蛋白質でも、類似した機能部位を持ちえることが予想される。このような類似機能部位を探索、予測するには、蛋白質立体構造の局所構造解析・比較が重要であると考えられる。そこで、応募者は、従来の局所構造比較法(TESS;Wallace et al.,1997）に学習の技術を取り入れたところ、活性部位予測に機械学習が有効であることを実証した（Kato & Nagano, 2010)。この研究にはまだ洗練する余地が多く残っていた。

２．研究の目的

以下の研究目的がある。

(i)高精度で実用的な次世代活性部位予測算法を構築すべく、汎用機械学習算法を適用できる統一記述表現を通して、大規模データに対応（Kato et al., 2009)、少サンプルでも高精度に学習（Kato et al., 2010a, 2010b)、マルチテンプレート予測への対応を可能にする。

(ii)補酵素結合様式の分類、結合・非結合状態間の変化の解析を通して活性部位周辺における補酵素分子の結合パターンの分析を行い、予測に有用な情報を抽出し、予測モデルを高精度化する。

(iii)開発した手法を応用して、活性部位予測算法を応用して相互連携する補酵素分子の解析を行う。

３．研究の方法

研究代表者・長野は、活性部位における補酵素結合様式の分析を行うために、補酵素としては、ニコチンアミド系の補酵素（NAD(P)+、NAD(P)H）とピリドキサール・リン酸（PLP）を選んで解析を進めることにした。NAD(P)補酵素と PLP の共通点は、いずれもリン酸基を介して、酵素蛋白質と結合していることである。しかしながら、NAD(P)補酵素は、比較的大きな分子で、コンフォメーション変化も大きいのに対し、PLP 補酵素は、分子そのものは小さいのでコンフォメーション変化は小さい。但し、NAD(P)補酵素が担う反応は、ヒドリド転移反応という比較的シンプルな反

応であるのに対して、PLP 補酵素は、複数の反応ステップを担い、酵素蛋白質の活性部位に含まれるリジン残基と共有結合（二重結合、単結合）を形成したり脱離したり、基質化合物のアミノ基と共有結合（二重結合、単結合）を形成したり脱離して、酵素蛋白質に対し、状態を大きく変えることが予想される。NAD(P)補酵素を結合する酵素は、主に、酸化還元酵素であるが、その進化的関連酵素には、異性化酵素やリアーゼ酵素なども含まれている。PLP を結合する酵素としては、アミノ基の代謝を行う転移酵素やリアーゼ酵素が主に含まれている。

研究代表者・長野は、こうした酵素の活性部位テンプレートを作成するために、まず、酵素反応の分類を行う必要があると考え、NAD(P)補酵素が関与するヒドリド転移反応の分類を行った。PLP 結合酵素に関しては、活性部位が反応段階により変化するので、そうした変化を解析するために、アスパラギン酸アミノ転移酵素(EC 2.6.1.1; AAT)の活性部位構造を解析した。これまで、アミノ酸残基の解析用のプログラムしか用意していなかったので、アミノ酸残基以外のヘテロ原子を含む蛋白質構造（PDB）データの解析を行うためのシステムを構築した。このシステムを用い、活性部位の構造類似度として RMSD を指標に、主成分分析を行った。アミノ酸残基に加えて、PLP のリン酸の３つの酸素を除く 12 原子を含んだ活性部位データを用い、主成分分析の比較を行った。

他方で、テンプレートとの局所構造比較のために重みなし RMSD が使われていたが、分担者・加藤は、重みつき RMSD を用いることを考え、重みを自動的に求めるアルゴリズムを開発していた。そのアルゴリズムでは、L∞ノルムと呼ばれる正則化関数を用いていたが、一般には別の正則化関数を用いた方が、予測性能が良くなることが知られており、正則化関数を変更すると汎用線形計画ソルバーでは最適化できなくなる。そこで、分担者・加藤は、正則化関数を変更した学習機械のための専用ソルバーを導出し、予備的実験を行った。

４．研究成果

研究代表者は、40 種類の NAD(P)補酵素結合酵素から 60 種類のヒドリド転移反応の分類を行った結果を、EzCatDB 酵素反応データベース（http://mbs.cbrc.jp/EzCatDB）の RLCP 分類で公開している（http://mbs.cbrc.jp/EzCatDB/RLCP/index.html）。

PLP 結合酵素の解析に関しては、AAT の解析結果では、2 種類の基質の他に、反応過程で少なくとも 8 種類の反応中間体を生成し、そのうち 6 種類は PDB に存在する。PLP を含む活性部位の方が、RMSD のばらつきが小さく、主成分分析によるクラスタリングも変化したが、PLP を入れた効果は予想外に小さかった。また、図１は、AAT 酵素の PLP を結合した活性部位（Asp，Tyr，Lys を含む）の構造比較を主成分分析したものである。図 1 の C00026 は基質のαケトグルタル酸、C00049 は L-アスパラギン酸であり、I00004、I00005、I00006、I00007、I99999 は、反応ステージの異なる PLP を含む中間体である。I00004、I00005 は、PLP と基質のアスパラギン酸が結合した中間体であるのに対し、I00006、I00007 は、産物のグルタミン酸が結合した中間体である。これらの中間体を含む構造は、中間体を含まない PLP のみが結合した活性部位、PLP と基質が結合した活性部位と分離するが、中間体の種類によるクラスタリングは難しいことが判明した。また、PLP のみが結合した活性部位、PLP と基質が結合した活性部位は、部分的に重なるので、分離が難しいことも判明した。テンプレート構造を作ることにより、各反応ステージを分類し予測することが可能であることを期待していたが、むしろ難しいことが判明した。

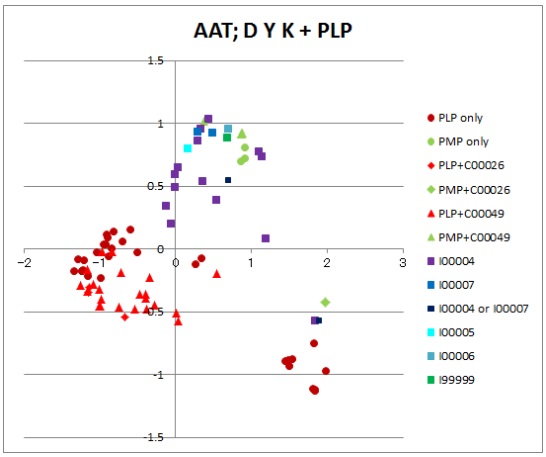

図１．AAT における補酵素 PLP が結合した活性部位の構造比較の主成分分析

他方で、研究分担者・加藤は同じ機能を持つ酵素を見つけるために類似の形状を持つ活性部位を探索するために、新しいアルゴリズムの開発に取り組んだ。従来は、ある機能の活性部位を見つけるために、その機能を持つ既知の活性部位における原子間の二乗距離の重みなし平均距離によって形状の類似性を評価していた。これに対して、本研究では重み係数を導入して原子間の二乗距離の線形結合で評価した。この重み係数の学習には Bregman Divergence Regularized Machine を導入した。重みをつけて計算するアプローチは、Kato & Nagano(2010) によってすでに提案されていたが、近年の機械学習で用いられている正則化損失最小化の原理に則っていなかった。これに対して、本研究では正則化損失最小化の原理に則ってアルゴリズムを再設計した。図 2 に示すように、実験の結果、正則化学習による重みつき偏差(KL)

は、Kato & Nagano(2010) の方法（KND）と comparable な結果が得られ、正則化損失最小化に則らない特殊なアルゴリズムが特に必要ではないことが明らかになった（レラトーレイサ、加藤毅、長野希美、(2013) 電子情報通信学会技術報告書、113 巻、61-66）。

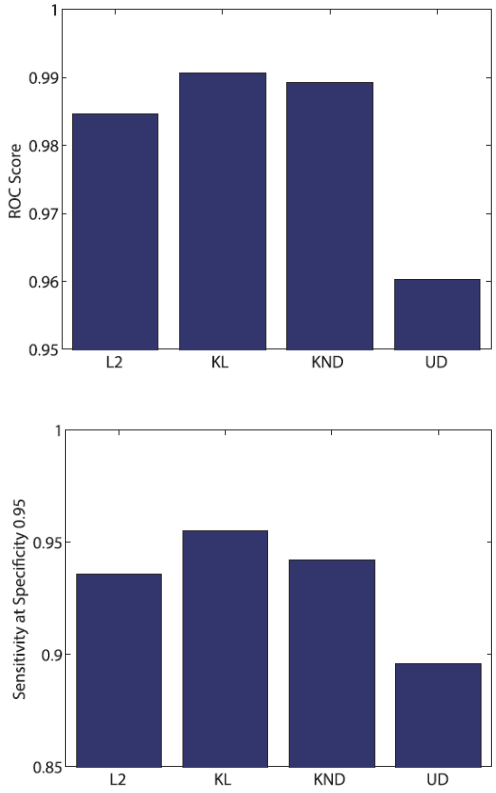

図２. 性能比較：ROCスコア（上）、Specificity 0.95でのSensitivity；l_2正則化学習による重みつき偏差(L2)、KL 正則化学習による重みつき偏差(KL) 、Kato & Nagano による方法(KND) 、従来の重みなし偏差(UD)

このようにして求めた重みつき偏差を使うと良好な探索性能が得られることを実験的に示した。提案法は，損失と正則化の和を最小化するという典型的な方法で設計されているので、機械学習分野で開発されているさまざまな拡張を利用することが可能である。よって、マルチタスク学習、転移学習、構造的学習などの適用の可能性が開かれたことになり、活性部位探索問題のための計算機的手法のさらなる発展が見込まれる。

また、本研究と関連して、識別タスクに関する研究も行った。近年のアプローチは各例題をベクトルの形式とは異なるものを用いていたが、ベクトル系列のための新しいカーネル、平均多項式カーネルを提案した（Raissa Relator, Yoshihiro Hirohashi, Eisuke Ito, Tsuyoshi Kato, IEICE Transactions on Information and Systems, vol.E97-D, 2014, *in press.*）。近年の研究では、各例題を線形部分空間で近似して、グラスマン多様体上の元として扱っているが、提案したカーネルはより一般的な形式でデータを表し、かつ、高速に計算できるという利点がある。

こうした手法は、酵素活性部位予測だけでなく、顔認識など他の技術にも応用できるという利点・側面もある（Tsuyoshi Kato, *et al.*, (2013) IPSJ Transactions on Computer Vision and Applications, vol.5, 85-89.）。

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）

〔雑誌論文〕（計　５件）

① レラトーレイサ、加藤毅、長野希美、Bregman Divergence Regularized Machine による酵素活性部位予測、電子情報通信学会技術報告書、査読無、113 巻、2013、61-66
② Tsuyoshi Kato, Wataru Takei, Shinichiro Omachi, A Discriminative Metric Learning Algorithm for Face Recognition. IPSJ Transactions on Computer Vision and Applications, 査読有、vol.5, 2013, 85-89.
③ Raissa Relator, Tsuyoshi Kato, Richard Lemence, Improved protein-ligand prediction using kernel weighted canonical correlation analysis. IPSJ Transactions on Bioinformatics, 査読有、vol.6, 2013, 18-28.
④ レラトーレイサ、廣橋義寛、伊藤栄祐、加藤毅、平均多項式カーネルと動画像認識およびブレインマシンインタフェースへの応用、電子情報通信学会技術報告書、査読無、113 巻、2014、281-286
⑤ Raissa Relator, Yoshihiro Hirohashi, Eisuke Ito, Tsuyoshi Kato, Mean Polynomial Kernel and Its Application to Vector Sequence Recognition. IEICE Transactions on Information and Systems, 査読有、vol.E97-D, 2014, *in press.*

〔学会発表〕（計　７件）

① 加藤毅、長野希美、Metric-learning based active site prediction for enzyme proteins, BiWO2011, 2012/1/26, 産総研臨海副都心センター（東京）
② Chioko Nagao, Nozomi Nagano, Kenji Mizuguchi, Enzyme function prediction using active sites and ligand binding sites information. 第 12 回日本蛋白質科学会年会、2012/6/21, 名古屋国際会議場（愛知）
③ 長野希美、Analysis of enzyme structures and functions/酵素の構造・機能の解析, BiWO2012, 2012/10/31, 産

総研臨海副都心センター（東京）
④ Tsuyoshi Kato, Wataru Takei, Shinichiro Omachi, A Discriminative Metric Learning Algorithm for Face Recognition. MIRU2013, 2013/7/29～2013/8/1, 国立情報学研究所（東京）
⑤ 長尾知生子、長野希美、水口賢司、A random-forest based method that can predict detailed enzyme functions and also identify specificity determining residues. CBI 学会 2013 年大会―生命医薬情報学連合会―、2013/10/29, タワーホール船堀（東京）
⑥ レラトーレイサ、加藤毅、長野希美、Bregman Divergence Regularized Machine による酵素活性部位予測、PRMU2013, 2013/12/12～2013/12/13, 三重大学（三重）
⑦ レラトーレイサ、廣橋義寛、伊藤栄祐、加藤毅、平均多項式カーネルと動画像認識およびブレインマシンインタフェースへの応用、PRMU2014, 2014/1/23～2014/1/24, 大阪大学（大阪）

〔図書〕（計　0件）

〔産業財産権〕
○出願状況（計　0件）

○取得状況（計　0件）

〔その他〕
ホームページ等
酵素反応データベース
http://mbs.cbrc.jp/EzCatDB/
但し、上記のドメイン名「mbs.cbrc.jp」が「ezcatdb.cbrc.jp」に変更になる可能性もある。

６．研究組織
(1)研究代表者
長野　希美（NAGANO, Nozomi）
独立行政法人産業技術総合研究所・生命情報工学研究センター・主任研究員
研究者番号：７０３５７６４８

(2)研究分担者
加藤　毅　（KATO, Tsuyoshi）
群馬大学・工学研究科・准教授
研究者番号：４０４０１２３６