

令和 4 年 6 月 20 日現在

機関番号：34204

研究種目：基盤研究(C)（一般）

研究期間：2019～2021

課題番号：19K12211

研究課題名（和文）実験的検証のフィードバックを活用した結合補酵素予測法の開発

研究課題名（英文）Development of a prediction method for cofactors binding to uncharacterized enzymes using feedback from experimental validations

研究代表者

塩生 真史（Shionyu, Masafumi）

長浜バイオ大学・バイオサイエンス学部・准教授

研究者番号：30345847

交付決定額（研究期間全体）：（直接経費） 3,400,000円

研究成果の概要（和文）：グラフニューラルネットワークを用いて低分子リガンドの結合部位の構造的特徴を学習することで、低分子リガンドとタンパク質の結合を予測するモデルを構築するProLMS-GNNの開発を行った。また、補酵素の一種であるピリドキサルリン酸（PLP）とタンパク質の結合を検証する実験手法の検討を行い、ビタミンB6微生物定量法により検証可能であることを確認した。ProLMS-GNNにより構築された予測モデルを使って出芽酵母タンパク質のPLP結合能を網羅的に推定し、実験的に検証したところ、少なくとも2つの機能未知タンパク質がPLPと結合することを明らかとした。

研究成果の学術的意義や社会的意義

ゲノムにコードされるタンパク質のうちの約2割は機能の手がかりのない機能未知タンパク質である。本研究で開発したProLMS-GNNは、既存の予測法と比較して同等程度以上の精度でPLPやFADとの結合を予測できることが確認できている。また、PLPと結合することが予測された機能未知タンパク質において、実験的にPLPとの結合が確認できたことから、ProLMS-GNNをさらに多くの補酵素についても結合を予測できるよう発展させることで、機能未知タンパク質の中からそれらの補酵素と関連の深い機能を持つ新奇の酵素が見つけれられる可能性が示された。

研究成果の概要（英文）：We developed ProLMS-GNN, which built prediction models for binding small-molecule ligands to proteins by learning the structural features of the ligand-binding sites using a graph neural network. We also studied the experimental methods for verifying the binding of pyridoxal 5'-phosphate (PLP), an important coenzyme in organisms, to proteins and confirmed that a microbiological assay for vitamin B6 could verify the ability of PLP-binding to proteins. We comprehensively estimated the PLP-binding ability of budding yeast proteins using the prediction model built by ProLMS-GNN and verified that experimentally. As a result, at least two uncharacterized proteins were confirmed to bind PLP.

研究分野：構造情報生物学

キーワード：機能予測 酵素 機械学習 出芽酵母 ピリドキサルリン酸 低分子結合予測

1. 研究開始当初の背景

次世代シーケンサーの普及により、様々な生物種についてゲノム塩基配列の決定が容易に行えるようになり、ゲノムにコードされるタンパク質の配列情報が爆発的に増加している。これらの中には、機能既知のタンパク質と明確な類似性を持たないために機能が予測できない機能未知タンパク質が多く含まれている。機能未知タンパク質の生化学的機能を考える上で、そのタンパク質に結合する低分子リガンドの情報は有用である。特に、結合する低分子リガンドが補酵素であることが分かれば、その機能未知タンパク質が特定の反応を行う酵素である可能性が高くなり、生化学的機能の実験的な検証につながる。

機能未知タンパク質の機能予測の際に立体構造情報が利用できれば、原子の空間配置の類似性に基づくことで、アミノ酸配列の類似性のみでは予測することが困難な低分子リガンドを予測できる可能性がある。しかし、ほとんどの機能未知タンパク質の立体構造は不明であるため、立体構造情報を用いた低分子リガンド結合予測法を適用するには、立体構造の予測が必須となる。一方で、本研究課題の申請時では、最先端のタンパク質モデリング法を使ったとしても、既知構造との配列類似性が低い場合にリガンド結合ポケット構造を十分な精度で予測できないことが多いことが指摘されていた (Liu et al., 2018, Proteins. 86 Suppl 1, 374-386)。そのため、既知のリガンド結合ポケットとの幾何学的な類似性を用いる低分子リガンド結合予測法を用いて機能未知タンパク質に結合する低分子リガンドを予測することは難しく、この問題を解決する新たな手法が必要であった。

本研究課題の申請時において研究代表者は、各低分子リガンドとの結合に使われやすいアミノ酸残基の傾向 (結合傾向値) に基づくスコアを用いた低分子リガンド結合残基予測法である ProLMS (Propensity of Ligand Molecule-binding Site) を開発していた。ProLMS は、原子配置の類似性ではなく、結合に使われやすいアミノ酸残基の空間的な集中度に基づくことで、予測された立体構造を使って予測した場合でも精度があまり低下しないことが強みであった。そこで研究代表者は、ProLMS を改良することで、機能未知のタンパク質にどのような低分子リガンドが結合するかを予測できるのではないかと考えた。

2. 研究の目的

本研究では、研究代表者が開発してきた ProLMS を、アミノ酸の代謝に関わる酵素で使われることの多い補酵素であるピリドキサルリン酸 (PLP) を題材として改良することで、タンパク質の立体構造情報から、PLP との結合を高い精度で予測できるようにすることを目的とする。この目的のために、以下のことを行う。

- (1) 3次元の物体の認識において良い性能があることがわかっている機械学習法である3次元畳み込みニューラルネットワーク (3D-CNN) を取り入れることで ProLMS の改良を行う。
- (2) 改良した ProLMS の予測精度を評価するため、出芽酵母のタンパク質を使って PLP 結合を予測し、実験的な検証を行う。その結果は ProLMS の改善に用いる。

3. 研究の方法

- (1) PLP 結合タンパク質をモデルとした ProLMS の改良と他の補酵素への拡張

ProLMS の改良

本研究課題の申請時の ProLMS は、機械学習の一種であるランダムフォレストを用いて PLP 結合残基およびその周囲の 4 残基の特徴量 (結合傾向値、凹み具合の評価値、相同配列におけるアミノ酸残基の保存性) を学習することで PLP 結合残基の予測を行っていた。この方法は PLP 結合残基かどうかを予測するアミノ酸残基の周囲の残基の情報を取り入れてはいるものの、それらの空間的な位置関係は利用できていない。そこで、タンパク質立体構造を 3次元格子で表現し、そこから一定の大きさの部分格子を取り出して特徴量を割り当てることで、その部分格子が PLP 結合部位と対応するかどうかを 3D-CNN を使って学習する。これにより PLP 結合部位の空間的特徴に基づいて PLP 結合部位を予測する予測モデルを構築できるようにする。この方針に従って、ProLMS に 3D-CNN を取り入れた ProLMS-3DConv を開発し、PLP 結合残基の予測モデルを構築した。部分格子に割り当てる特徴量や 3D-CNN の層のパターンを変えて予測精度を比較したところ、最も良い予測モデルでも改良前の ProLMS の精度と同程度となった。このことから、ProLMS-3DConv によって PLP に結合する残基を予測し、その結果に基づいてタンパク質の PLP 結合能を予測する方法では、PLP 結合能を予測する精度が十分には得られない判断した。

そこで本研究課題申請時の方針を変更し、3D-CNN の代わりにグラフニューラルネットワーク (GNN) を用いて PLP 結合の予測モデルを構築する ProLMS-GNN を開発した。アミノ酸残基をノードとし、近接しているアミノ酸残基どうしをエッジで結んだグラフに基づいてアミノ酸残基に割り当てた特徴量をグラフニューラルネットワークで学習することで、3D-CNN と同様にアミノ酸残基の空間的な位置関係を考慮した学習ができる。さらに、グラフニューラルネットワー

クは、PLP に結合する残基かどうかを判断するように学習する（ノードレベルの学習）だけでなく、PLP と結合するタンパク質かどうかを判断するように学習する（グラフレベルの学習）こともできるという利点がある。これにより、PLP 結合残基を予測した結果を用いてタンパク質の PLP 結合能を予測するという間接的な方法ではなく、タンパク質の立体構造情報から直接 PLP 結合能を予測することが可能となった。また、ProLMS-GNN では、ProLMS で用いていた特徴量を見直し、アミノ酸残基の one-hot 表現や結合傾向値、立体構造的特徴のみを用いるようにし、相同配列におけるアミノ酸残基の保存性を特徴量から外した。これにより、改良前の ProLMS と比べて短時間で特徴量を求めることができ、相同タンパク質の情報に依存せずに PLP 結合を予測できるようになった。さらに、学習に用いるタンパク質の抽出方法を見直し、学習時にデータ拡張が行えるようにした。

他の補酵素への拡張

PLP 以外の補酵素のうち、酸化還元反応に関わる補酵素であるフラビンアデニンジヌクレオチド（FAD）やフラビンモノヌクレオチド（FMN）、ニコチンアミドアデニンジヌクレオチド（NAD）についても、学習用データを収集し、ProLMS-GNN により予測モデルを構築した。

また ProLMS-GNN は、機械学習に基づくため、多様な複合体の立体構造が解明されている補酵素でしか有効な予測モデルを構築できない問題がある。そこで、複合体構造の情報が少ない補酵素でも結合が予測できるよう、タンパク質のポケット同定と低分子化合物のドッキングシミュレーションを組み合わせたリガンド結合予測法の開発を行った。

(2) 出芽酵母プロテオームに対する PLP 結合予測と実験的検証

本研究手法の開発は既知の PLP-タンパク質複合体の立体構造情報を用いて行うが、複合体の立体構造の数は限られている。その中で予測精度を向上させたとしても、立体構造がまだわかっていない PLP 結合タンパク質については十分な予測精度があるか不明である。そこで、モデル生物として様々な分子生物学的手法を適用できる出芽酵母のプロテオームを用いて PLP 結合を予測し、PLP 結合が陽性となったタンパク質について実験的な検証を行うこととした。

PLP 結合アッセイ系の構築

実験的に PLP が結合することが示されている出芽酵母由来の 27 種の酵素から、PLP との複合体の立体構造が決定されているなどして確実に PLP との結合が見込める Yb1036c をポジティブコントロール、PLP 結合がないと見込める Cet1 をネガティブコントロールとして選び、GST との融合タンパク質を大腸菌で発現させて精製した。精製した組換えタンパク質の PLP 結合の検証にはフェニルヒドラジン法を用いる予定であったが、透析などの煩雑な作業を必要としたために、ビタミン B6 を要求する酵母の増殖を指標とした微生物定量法に変更した。具体的には、グルタチオン磁性ビーズに固定した GST 組換えタンパク質を PLP と結合させた後、結合しなかった PLP を洗浄により除去した。組換えタンパク質に結合した PLP を塩酸中で煮沸することによりタンパク質から遊離させ、ビタミン B6 要求性酵母を植菌した培地（ビタミン B6 を含まない）に添加して一晚培養した。培養液の 595 nm の吸光度を酵母の増殖の指標として測定し、ビタミン B6 と吸光度の関係を示す検量線から組換えタンパク質に結合した PLP 量を算出した。

出芽酵母プロテオームの PLP 結合予測と検証

出芽酵母の全タンパク質セットについて網羅的にホモロジーモデリングにより立体構造予測を行い、立体構造が予測できたものに対して改良した ProLMS-GNN を適用した。PLP との結合が知られていないタンパク質の中から PLP 結合予測が陽性となったものを選択して PLP 結合アッセイを行い、新規の PLP 結合タンパク質かどうかを確認した。

4. 研究成果

(1) PLP 結合タンパク質をモデルとした ProLMS の改良と他の補酵素への拡張

ProLMS の改良

ProLMS-GNN の PLP 結合残基の予測精度を、既存の立体構造情報に基づくリガンド結合部位予測法と比較した（図 1A）。その結果、機械学習を用いた方法である p2rank や DeepSurf と比較して F1 値で定量化した予測精度が大きく上回っており、予測精度が良いとされる COACH と比較しても同等の精度で予測できた。さらに、PLP 結合能の予測精度を COACH と比較したところ（図 1B）PLP 結合残基の予測と同様にほぼ同等の精度であった。COACH は既知のリガンド結合部位との構造的類



図 1 ProLMS-GNN と他の手法の精度比較

似を検知することにより結合部位を予測するが、一つのタンパク質の予測にかかる時間が非常に長いという問題があり、ゲノム規模で適用するのは難しい。ProLMS-GNN は一つのタンパク質あたり長くても数分で予測が終了するため、この点について COACH と比較して優位性があると考えられた。

他の補酵素への拡張

PLP に結合するタンパク質と同様に、FAD や FMN、NAD に結合するタンパク質を収集し、ProLMS-GNN による予測モデルを構築した。その結果 FAD については、PLP と同様に、他のリガンド結合部位予測法と比較しても同等以上の予測精度が得られることがわかった。しかし、FMN や NAD については、収集した学習用データに対する負例の割り当てに問題があることが明らかとなったため、今後の研究でデータ収集の方法の見直しが必要となっている。

また、複合体構造の情報が少ない補酵素結合を予測するために、以下のような手法を開発した。まず、タンパク質の立体構造に対して fpocket を適用することでポケット状の表面構造を同定し、それらに対して Glide や AutoDock Vina などの低分子化合物のドッキングシミュレーションを行う。ドッキングシミュレーションにより得られた複合体構造の中で、ドッキングのスコアが高いものにおいて低分子化合物との結合に用いられているアミノ酸残基を求めることによりリガンド結合部位を予測する。開発した手法を用いて、タンパク質との複合体構造がほとんど決定されていないクルクミンについてその結合部位を予測した。その結果、クルクミン、および、その誘導体が複数の活性酸素種 (ROS) 代謝酵素の補酵素結合部位に結合し、ROS の代謝が阻害されることでがん細胞内の ROS が上昇して抗がん活性を発揮するという仮説を提唱した (Molecules, 2019)。また、同手法を用いて新型コロナウイルスの増殖抑制活性を持つセファラインチンおよびその類縁体が結合する可能性のあるタンパク質上の結合部位予測を行い、さらに、予測された複合体構造から推定されるファーマコフォアを提唱した (FEBS Open Bio, 2022)。

(2) 出芽酵母プロテオームに対する PLP 結合予測と実験的検証

PLP 結合アッセイ系の構築

GST タグを付加したタンパク質をグルタチオン磁気ビーズに固定し、PLP 結合アッセイ系の条件検討に用いた。PLP の定量には、ビタミン B6 要求性酵母株を用いた微生物定量法を利用した。その結果、既知の PLP 結合タンパク質である Yb1036c (細胞機能未知タンパク質) において PLP との結合が認められ、PLP と結合しないタンパク質である Cet1 (mRNA キャッピング酵素) や GST タグでは PLP との結合が認められないことを確認した (図 2)。これにより、ビタミン B6 微生物定量法による PLP 結合アッセイ系を構築することができたと判断した。

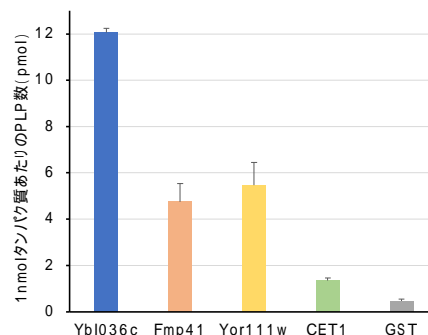


図 2 PLP 結合アッセイの結果

出芽酵母プロテオームの PLP 結合予測と検証

出芽酵母の全タンパク質について、既知立体構造に対して相同性検索を行い、実験により決定された出芽酵母タンパク質の立体構造データ 1,504 個と、MODELLER によるホモロジーモデリングにより予測した立体構造データ 2,495 個について ProLMS-GNN により PLP 結合能を予測した。この時、ProLMS-GNN の学習時に初期パラメータをランダムに変更することで 10 種類の予測モデルを構築し、より多くの予測モデルで PLP 結合能が予測された場合に、その予測の信頼性が高いと順位付けできるようにした。その結果、5 種類以上の予測モデルで PLP 結合能が予測されたタンパク質を見ると、UniProtKB/Swiss-Prot データベースにおいて既に PLP 結合能が指摘されている 57 個のタンパク質の約 9 割に当たる 51 個のタンパク質が含まれており、高い感度で PLP 結合タンパク質を検出することができた。また、5 種類以上の予測モデルで PLP 結合能が予測されており、かつ、これまでに PLP 結合能が指摘されていないタンパク質は 146 個存在した。この中に新規の PLP 結合タンパク質が存在していることが期待される。

予測信頼性の順位の高いものの中から、出芽酵母のデータベースにおいて機能未知タンパク質とされている 5 個を選択して、GST 融合タンパク質を設計し、大腸菌での発現・精製を行った後にビタミン B₆ 微生物定量法を用いて PLP 結合能を確認した。その結果、2 つのタンパク質 (Fmp41 と Yor111w) において予測された通り PLP 結合能を確認することができた (図 2)。これらのタンパク質は新規の PLP 結合タンパク質である。残りの 3 つの GST 融合タンパク質については、大腸菌での発現あるいは可溶化ができておらず、現在、低温誘導による発現系および His タグを使うことでこれらの問題を解決できないかを検討中である。

以上の研究成果により、機能未知タンパク質において酵素活性に重要な補酵素が結合することを予測し、それを検証する手法の基礎が開発できた。今後はこれらの手法を発展させ、ゲノム中に未だ存在する機能未知タンパク質に適用することで、新規の酵素を明らかにすることを進めていく。

5. 主な発表論文等

〔雑誌論文〕 計4件（うち査読付論文 4件/うち国際共著 1件/うちオープンアクセス 4件）

1. 著者名 Hijikata Atsushi, Shionyu Clara, Nakae Setsu, Shionyu Masafumi, Ota Motonori, Kanaya Shigehiko, Shirai Tsuyoshi	4. 巻 18
2. 論文標題 Current status of structure-based drug repurposing against COVID-19 by targeting SARS-CoV-2 proteins	5. 発行年 2021年
3. 雑誌名 Biophysics and Physicobiology	6. 最初と最後の頁 226 ~ 240
掲載論文のDOI (デジタルオブジェクト識別子) 10.2142/biophysico.bppb-v18.025	査読の有無 有
オープンアクセス オープンアクセスとしている (また、その予定である)	国際共著 -
1. 著者名 Hijikata Atsushi, Shionyu Mitsuyama Clara, Nakae Setsu, Shionyu Masafumi, Ota Motonori, Kanaya Shigehiko, Hirokawa Takatsugu, Nakajima Shogo, Watashi Koichi, Shirai Tsuyoshi	4. 巻 12
2. 論文標題 Evaluating cepharanthine analogues as natural drugs against SARS CoV 2	5. 発行年 2021年
3. 雑誌名 FEBS Open Bio	6. 最初と最後の頁 285 ~ 294
掲載論文のDOI (デジタルオブジェクト識別子) 10.1002/2211-5463.13337	査読の有無 有
オープンアクセス オープンアクセスとしている (また、その予定である)	国際共著 -
1. 著者名 Mori Shoko, Nomura Kaoru, Fujikawa Kohki, Osawa Tsukiho, Shionyu Masafumi, Yoda Takao, Shirai Tsuyoshi, Tsuda Shugo, Yoshizawa-Kumagaye Kumiko, Masuda Shun, Nishio Hideki, Yoshiya Taku, Suzuki Sonomi, Muramoto Maki, Nishiyama Ken-ichi, Shimamoto Keiko	4. 巻 17
2. 論文標題 Intermolecular Interactions between a Membrane Protein and a Glycolipid Essential for Membrane Protein Integration	5. 発行年 2022年
3. 雑誌名 ACS Chemical Biology	6. 最初と最後の頁 609 ~ 618
掲載論文のDOI (デジタルオブジェクト識別子) 10.1021/acscchembio.1c00882	査読の有無 有
オープンアクセス オープンアクセスとしている (また、その予定である)	国際共著 -
1. 著者名 Nakamae Ikuko, Morimoto Tsumoru, Shima Hiroki, Shionyu Masafumi, Fujiki Hisayo, Yoneda-Kato Noriko, Yokoyama Takashi, Kanaya Shigehiko, Kakiuchi Kiyomi, Shirai Tsuyoshi, Meiyanto Edy, Kato Jun-ya	4. 巻 24
2. 論文標題 Curcumin Derivatives Verify the Essentiality of ROS Upregulation in Tumor Suppression	5. 発行年 2019年
3. 雑誌名 Molecules	6. 最初と最後の頁 4067
掲載論文のDOI (デジタルオブジェクト識別子) 10.3390/molecules24224067	査読の有無 有
オープンアクセス オープンアクセスとしている (また、その予定である)	国際共著 該当する

〔学会発表〕 計6件（うち招待講演 0件 / うち国際学会 0件）

1. 発表者名 Masafumi Shionyu, Atsushi Hijikata
2. 発表標題 Protein-cofactor binding prediction with graph neural networks
3. 学会等名 第59回日本生物物理学会年会
4. 発表年 2021年

1. 発表者名 塩生真史, 土方敦司
2. 発表標題 グラフニューラルネットワークを用いたタンパク質の低分子リガンド結合予測
3. 学会等名 第21回日本蛋白質科学会年会
4. 発表年 2021年

1. 発表者名 Masafumi Shionyu, Tomohiro Hatta, Atsushi Hijikata
2. 発表標題 Prediction of PLP-binding proteins by using machine learning-based methods
3. 学会等名 第58回日本生物物理学会年会
4. 発表年 2020年

1. 発表者名 Masafumi Shionyu, Atsushi Hijikata
2. 発表標題 Machine learning models for predicting ligand-binding sites using residue-wise features
3. 学会等名 第57回日本生物物理学会年会
4. 発表年 2019年

1. 発表者名 塩生真史, 山崎まど香, 土方敦司
2. 発表標題 結合傾向値を特徴量とした機械学習による低分子リガンド結合残基予測
3. 学会等名 第19回日本蛋白質科学会年会 第71回日本細胞生物学会大会 合同年次大会
4. 発表年 2019年

〔図書〕 計0件

〔産業財産権〕

〔その他〕

Het-PDB Navi2 https://hetpdbnavi.nagahama-i-bio.ac.jp

6. 研究組織

	氏名 (ローマ字氏名) (研究者番号)	所属研究機関・部局・職 (機関番号)	備考
研究分担者	土方 敦司 (Hijikata Atsushi) (80415273)	長浜バイオ大学・バイオサイエンス学部・プロジェクト特任講師 (34204)	
研究分担者	向 由起夫 (Mukai Yukio) (60252615)	長浜バイオ大学・バイオサイエンス学部・教授 (34204)	

	氏名 (ローマ字氏名) (研究者番号)	所属研究機関・部局・職 (機関番号)	備考
研究協力者	八田 友博 (Hatta Tomohiro)		

6. 研究組織（つづき）

	氏名 (ローマ字氏名) (研究者番号)	所属研究機関・部局・職 (機関番号)	備考
研究協力者	岡崎 幸之介 (Okazaki Konosuke)		
研究協力者	鈴木 颯馬 (Suzuki Soma)		
研究協力者	瀧 優介 (Taki Yusuke)		
研究協力者	梅田 知晴 (Umeda Chiharu)		
研究協力者	中島 俊雄 (Nakajima Toshio)		

7. 科研費を使用して開催した国際研究集会

〔国際研究集会〕 計0件

8. 本研究に関連して実施した国際共同研究の実施状況

共同研究相手国	相手方研究機関