

令和 6 年 5 月 22 日現在

機関番号：15401

研究種目：若手研究

研究期間：2021～2023

課題番号：21K14741

研究課題名（和文）ヘムのツボを押さえて使うための構造機能相関解析ツール「PyDISH」の開発

研究課題名（英文）Development of PyDISH, the analytic tool of the structure-function relationship for heme

研究代表者

兼松 佑典 (Kanematsu, Yusuke)

広島大学・先進理工系科学研究科(工)・助教

研究者番号：10765936

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）：ヘムタンパク質の構造機能相関解析アプリケーション「PyDISH」の有用性を向上させるために、データベースの自動更新と、ユーザーがアップロードした構造の解析の機能をそれぞれ実装した。またランダムフォレスト法を用いて機能分類予測について検証し、構造歪みをもとに有意な精度で機能の二値分類が実行可能であることが確認された。さらに与えられたヘムの構造をもとに酸化還元電位と酸素吸着能の近似値を評価する機能も実装した。その一方で、PyDISHの自動更新のアルゴリズムを転用することで、任意の金属補因子について統計解析を行う機能をGoogle Colabを用いて実装・公開し、その有用性を確認した。

研究成果の学術的意義や社会的意義

本研究におけるPyDISHの開発とGoogle Colabを用いた新規解析機能により、誰でもWebブラウザ上で最新のデータを参照しながら特定のヘムや金属補因子の特徴づけを行うことが新たに可能になった。これにより実験やシミュレーションで取得されたヘムの構造歪みに対する意味づけが可能になった。また、タンパク質に結合したヘムの構造を予測する機能を新規実装した構造予測ツール「AlphaFold3」とPyDISHとの併用により、新規タンパク質の活性を予測することも可能になった。ただしその予測精度や適用範囲については今後系統的に検討する必要がある。

研究成果の概要（英文）：To improve the applicability of PyDISH, the application for structure-function relationship analysis of heme proteins, we implemented automatic database update and analytic interface of user-uploaded structures, respectively. We also validated the prediction of functional classification using the random forest method, and confirmed that it is possible to perform binary classification of functions with significant accuracy based on structural distortion of heme porphyrin. We also implemented a tool to evaluate approximate values of redox potential and oxygen adsorption capacity based on the given heme structure. Meanwhile, by diverting the algorithm of PyDISH's automatic update, we implemented and published another application to perform statistical analysis on arbitrary metal cofactors using Google Colab, and confirmed its usefulness.

研究分野：生物物理化学

キーワード：統計科学 機械学習 バイオインフォマティクス 量子化学

1. 研究開始当初の背景

ヘムは鉄イオンとポルフィリンからなる代表的な生命金属補因子である。タンパク質との結合により多様な機能を獲得し、電子伝達、物質運搬、生化学反応触媒などの非常に広範な生命現象を担っている。例えば、真核生物の多種多様な代謝機構の多くは同一祖先由来のヘムタンパク質である cytochrome P450 (CYP) に依存して駆動することが知られている。これにちなんで CYP は "diversozyme" と呼称され、機能の可変性 (多様化可能性) 自体がヘムのアイデンティティだとみなす向きもある。近年では天然に無い機能をヘムタンパク質やヘムの類縁体に持たせバイオ触媒や分子性材料として活用する取り組みも盛んであるが、膨大な試行錯誤に依存しているため、ヘムの機能制御機序解明に基づく開発効率化が望まれている。

ヘムの機能の可変性はその物性が化学的・力学的修飾に対して柔軟に応答することに由来し、タンパク質はヘムに適当な修飾を施して特定の機能を発現させるための制御装置として働いている。ここで化学的修飾とは主にヘム鉄上の軸配位子を指し、力学的修飾とはヘムのポルフィリン環に対してタンパク質や基質によって与えられる構造歪みを指す。特に後者については近年重要性が認識され始めており、例えば、酸化還元電位を変化させ、電子輸送機能の調整に利用されていることが示唆されている「Ruffling 歪み」[1]や、酸素吸着能を変化させる「Breathing 歪み」[2]をはじめとした構造歪みが知られており、それぞれのヘムタンパク質の機能分岐に寄与していると考えられる。

2. 研究の目的

上述のような微細な構造機能相関を解析する機運は、近年の結晶構造解析技術の発展に伴うヘムタンパク質の高解像度結晶構造の PDB 登録数の増加によって高まってきていた。そこで我々は、構造機能相関解析をウェブブラウザ上で実行可能な、解析ツールとデータベースを同梱したアプリケーション「PyDISH」を開発・公開した[3]。本研究では、PyDISH の機能を拡充・発展させることで、ヘムタンパク質の機能分岐を促す「ツボ」となるような構造因子を抽出する手段の確立を目指した。

3. 研究の方法

PyDISH を構成するウェブインターフェース、データ収集、統計解析ツールはすべて Python を用いて実装されている。本研究ではこの Python コードを修正することで新規機能を実装した。また、酸素吸着能と酸化還元電位は密度汎関数法を用いて、PBE0/6-31G(d) レベルで計算した。

4. 研究成果

(1) 自動更新の実装: PyDISH のデータ更新は従来手動で不定期に行われていたが、PyDISH プロジェクトの長期持続化のためにデータ自動更新法を確立させた (図 1)。PyDISH の保有するデータの取得元である日本蛋白質構造データバンク (PDBj) におけるヘムタンパク質結晶構造データの更新を毎週自動的に検出し、PyDISH の公開データを自動的に更新するアルゴリズムを Python によって実装し、PyDISH のサーバ上で稼働させた。具体的には、PDBj の提供する REST API を介して SQL を用いて座標情報以外のデータ (更新の有無を含む) を取得し、更新があったヘムタンパク質についてのみ mmCIF ファイルをパースして軸配位子やポルフィリン環歪みの情報を取得した。これにより、すべての情報を mmCIF ファイルのパースによって取得するよりも大幅に時間短縮されたのに加え、PDBj の中で結晶構造に紐づけられている Gene Ontology などの周辺情報も併せて取得できるようになった。また、4 桁の PDB ID の枯渇が近づいていることから、今後遠からず PDB データの仕様変更が起きることが見込まれるため、それに柔軟に対応できるよう、更新アルゴリズムの簡潔化・汎用化を行った。

(2) 機械学習によるヘムタンパク質の機能予測: ヘムのポルフィリン骨格構造ひずみの程度と軸配位子の分子種を説明変数として、ランダムフォレスト法によるヘムタンパク質の機能分類を行った。PyDISH からヘモグロビンをはじめとする酸素運搬タンパク質と、シトクロム P450 をはじめとする酸化還元酵素の軸配位子とヘムのポルフィリン環



図 1: 自動更新アルゴリズムの概略図

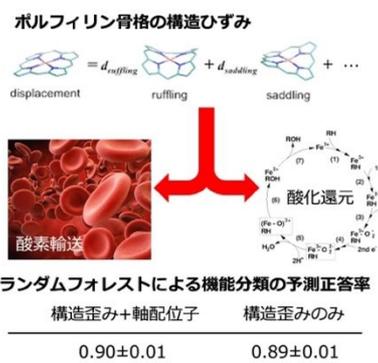


図 2: 機能の二値分類結果

の構造歪みの情報を取得し、それらとタンパク質機能との間に潜在する相関関係をランダムフォレスト分類器に学習させ、学習に用いなかったテストセットに対する予測精度を評価した。その結果、9割の正答率で予測できることが確認され、軸配位子以外のアミノ酸配列を考慮せずとも十分な精度で機能を予測できることが実証された。さらに、説明変数から軸配位子情報を除外した分類でも正答率はほとんど低下せず、ヘムの構造歪みのみでこの2つの機能の分岐は十分に説明付けられることも明らかになった(図2)。

(3) Google Colab を活用した金属タンパク質の統計解析ツールの実装: PyDISH のためのデータベース構築スキームをヘム以外の金属補因子のために流用することで金属タンパク質全般について柔軟にデータ解析を行う手段の確立を試みた。具体的には、Google Colab 上で望みのリガンドについて、上述の PDBj の REST API からデータを取得してオンデマンドで SQLite3 データベースを構築し、データの閲覧、そして解析まで行うスクリプトを公開した[4]。図3にニッケルを含むタンパク質についての解析結果を示す。ユーザーが指定した金属補因子を有するタンパク質のデータセット構築と簡単な統計解析が1分程度で行えることを確認した。指定した金属を含有するタンパク質に紐付けられている機能や配位子の数・種類の分布を確認したり、PDB ID とリガンドの asym ID を指定することでリガンド周りの構造を可視化したりといった一連の操作を全てウェブブラウザ上で完結させることが可能となった。



図3: Google Colab を用いた解析ツールにおける (a) 対象となる金属補因子の指定画面、(b) 金属補因子周りの構造の可視化、(c) 機能の分布比較

(4) 物性値計算の実装: ヘムに固有のポルフィリン環構造歪みに基づいてヘムの酸素吸着能や酸化還元電位といったヘムタンパク質の機能と強い相関を持つことが期待される物性値を評価する方法を実装した。鉄ポルフィリンモデル錯体の平衡構造における酸化還元電位と酸素吸着能とそれらの一次勾配、二次勾配をそれぞれ密度汎関数法によって計算し、構造歪みに基づく各物性値の変位を近似的に求めた。現時点では構造歪みが物性値に与える影響のみ評価可能であるが、今後軸配位子の影響も取り込むことで機能予測の説明変数としての実用性を向上させる。

(5) アップロード解析インターフェースの実装: ユーザーがアップロードしたヘムの構造に対して、PyDISH 上の解析ツールを用いて特長づけを行うためのインターフェースを実装した[5]。これにより、PDB 未登録のヘムやシミュレーションで得られたヘムの構造分布についてもその構造の特徴づけを行うことも可能になった。

最近になって DeepMind 社が公開したタンパク質構造予測ツール「AlphaFold3」は新機能としてタンパク質に結合したヘムの構造も併せて予測する機能を搭載しており、いくつか確認した範囲ではタンパク質ごとで固有のポルフィリン構造歪みを与えることがわかっている。したがって、新規ヘムタンパク質の予測構造を AlphaFold3 で取得し、PyDISH で解析することで機能・活性を定性的、あるいは定量的に評価し、タンパク質の de novo デザインに資する情報を提供できるようになる可能性がある。この様な用途に堪える予測精度を有するかについて今後検討する。

<Reference>

1. Sun Y.; Benabbas A.; Zeng W.; Kleingardner J. G.; Bren K. L.; Champion P. M. *Proc. Natl. Acad. Sci. U.S.A.*, **2014**, 111, 6570-6575.
2. Bikiel D. E.; Forti F.; Boechi L.; Nardini M.; Luque F. J.; Martí M. A.; Estrin D. A. *J. Phys. Chem. B*, **2010**, 114, 8536-8543.
3. Kondo H. X.; Kanematsu Y.; Masumoto G.; Takano Y. *Database*, **2020**, 2023, baaa066.
4. <https://colab.research.google.com/drive/17PSJ4sgSOeEpydZIoEgBUseNhwaBFDa?usp=sharing>
5. <https://pydish.bio.info.hiroshima-cu.ac.jp/upload/>

5. 主な発表論文等

〔雑誌論文〕 計3件（うち査読付論文 3件/うち国際共著 0件/うちオープンアクセス 2件）

1. 著者名 H. X. Kondo, H. Iizuka, G. Masumoto, Y. Kabaya, Y. Kanematsu, Y. Takano	4. 巻 13
2. 論文標題 Prediction of Protein Function from Tertiary Structure of the Active Site in Heme Proteins by Convolutional Neural Network	5. 発行年 2023年
3. 雑誌名 Biomolecules	6. 最初と最後の頁 137
掲載論文のDOI（デジタルオブジェクト識別子） 10.3390/biom13010137	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 Hiroko X. Kondo, Masanori Fujii, Takuma Tanioka, Yusuke Kanematsu, Takashi Yoshida, and Yu Takano	4. 巻 62
2. 論文標題 Global Analysis of Heme Proteins Elucidates the Correlation between Heme Distortion and the Heme-Binding Pocket	5. 発行年 2022年
3. 雑誌名 Journal of Chemical Information and Modeling	6. 最初と最後の頁 775-784
掲載論文のDOI（デジタルオブジェクト識別子） 10.1021/acs.jcim.1c01315	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

1. 著者名 Y. Kanematsu, H. X. Kondo, Y. Takano	4. 巻 24
2. 論文標題 Computational exploration of minimum energy reaction pathway of N2O formation from intermediate I of P450nor using an active center model	5. 発行年 2023年
3. 雑誌名 Int J Mol Sci	6. 最初と最後の頁 17172
掲載論文のDOI（デジタルオブジェクト識別子） 10.3390/ijms242417172	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 -

〔学会発表〕 計7件（うち招待講演 1件/うち国際学会 2件）

1. 発表者名 Y. Kanematsu, H. X. Kondo, Y. Takano
2. 発表標題 Development of the web application for the analysis on the structure-function correlation of hemeproteins and its expansion for metalloproteins
3. 学会等名 10th Asian Biological Inorganic Chemistry Conference (AsBIC10) (国際学会)
4. 発表年 2022年

1. 発表者名 兼松 佑典、石元 孝佳、鷹野 優
2. 発表標題 金属酵素タンパク質解析ウェブアプリケーションの開発
3. 学会等名 第45回ケモインフォマティクス討論会
4. 発表年 2022年

1. 発表者名 兼松 佑典
2. 発表標題 ヘムの構造機能相関解析ウェブアプリ『PyDISH』の開発
3. 学会等名 "医学と数理" (第3回 京大 ハイデルベルク大 理研 ワークショップ)
4. 発表年 2022年

1. 発表者名 近藤寛子, 飯塚博幸, 舩本現, 兼松佑典, 鷹野優
2. 発表標題 ヘムタンパク質におけるタンパク質ポケットとヘムの歪みの相関の解析
3. 学会等名 第47回生体分子科学討論会
4. 発表年 2021年

1. 発表者名 近藤寛子, 飯塚博幸, 舩本現, 兼松佑典, 鷹野優
2. 発表標題 ヘム蛋白質における構造機能相関の解明に向けたヘム結合部位のタンパク質環境とヘムの歪みの相関解析
3. 学会等名 日本コンピュータ化学会2021年秋季年会
4. 発表年 2021年

1. 発表者名 藤川ひな子, 近藤寛子, 兼松佑典, 鷹野優
2. 発表標題 機械学習を用いたヘムの構造にもとづくヘムタンパク質の機能分類
3. 学会等名 第44回日本生体医工学会中国四国支部大会
4. 発表年 2021年

1. 発表者名 Y. Kanematsu, H. X. Kondo, Y. Takano
2. 発表標題 Quantum Chemical and Statistical Analysis of Hemeproteins
3. 学会等名 International Congress on Pure & Applied Chemistry Bali 2023 (招待講演) (国際学会)
4. 発表年 2023年

〔図書〕 計0件

〔産業財産権〕

〔その他〕

PyDISH https://pydish.bio.info.hiroshima-cu.ac.jp Google Colab上での解析の実装 https://colab.research.google.com/drive/17PSJ4sgS0eEpydZ1oe0gBUseNhwa8FDa?usp=sharing PyDISHに新規実装したupload機能 https://pydish.bio.info.hiroshima-cu.ac.jp/upload
--

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

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
---------	---------