

共同利用研究ハイライト

# ヘムタンパク質におけるヘムの多機能性の起源の解明に向けた深層学習によるヘムタンパク質ポケット構造からの機能予測

鷹野 優 広島市立大学大学院情報科学研究科 教授

ヘムは鉄イオンとポルフィリンからなる金属補因子であり (図1)、ヘムを補因子に持つタンパク質をヘムタンパク質という。ヘムタンパク質は多くの生命体に含まれており、電子伝達、物質運搬、生化学反応触媒など極めて広範な生命現象を担っており、心疾患や脳内のコレステロール調整などにも関係することが報告されている。そのため、ヘムタンパク質の機能解明に向けた研究が精力的に進められており、例えば構造解析の分野では約6300ものヘムタンパク質の構造がProtein Data Bank (PDB) に登録されている (2023年11月29日現在)。しかしながら、活性中心のヘムは同一の基本骨格を持つにもかかわらず、取り囲むタンパク質の違いでこれほどまで多様な機能を発現する起源については明快な答えが得られていない。その解明にはヘムの包括的な研究が必要と考え、研究を進めている。

本レターでは、ヘムを取り囲むタンパク質ポケットの構造とタンパク質機能の関係を明らかにするため、深層学習を用いたヘムタンパク質のポケット構造からの機能予測に関する最近の研究成果について報告する<sup>[1]</sup>。まずPDBからヘムa、ヘムb、ヘムc、ヘムoのいずれかを含み、分解能が2 Å以下の構造データを抽出した。これらの構造データから、軸配位子がアミノ酸残基または水分子である鎖を抜き出し、配列類似性に関して重複のないデータセットを作成した。タンパク質ポケットはヘムの座標をもとに決定した立方体領域とし、それを各辺の長さが1 Åの立方体 (ボックス) に分割してポケット構造を表現した。各ボクセル

ルには原子の有無により0または1の数値を割り当てた (図2左)。深層学習手法の畳み込みニューラルネットワーク (Convolutional Neural Network: CNN) を用いて、タンパク質ポケットの立体構造からタンパク質機能の予測を行い、タンパク質環境とタンパク質機能の相関を解析した。タンパク質機能の割り当てには、EC番号、遺伝子オントロジー、PDBファイル中の記述を用いた。機能予測では、酸素結合タンパク質 (OB)、酸化還元酵素 (OR)、および酸素結合と酸化還元反応触媒の二重機能タンパク質 (OB-OR) を解析対象とした。OB, OR, OB-ORについて2レベルでのクラス分類を行い、予測値と観測値の正確度を計算したところ、5回交差検証の平均値として0.959という高い精度が得られ、ポケット構造とタンパク質機能との相関が示唆された。混同行列の結果 (図2右) から、単一

機能のタンパク質については高精度に予測ができていた一方で二重機能 (OB-OR) の予測は難しいことがわかった。入力データの情報を部分的に削除し機能予測を行ったところ、タンパク質ポケットの外側から情報を削除した場合はすぐに予測の正確度が下がってくのに対して (図3A)、内側から削除した場合は20 Åほど削除しても予測の正確度は保たれていた (図3B)。このことから、ヘムの最近傍よりも離れたところの情報が重要であることが示唆された。また、タンパク質の空洞部分の構造の比較から、同一の機能を持つタンパク質でも空洞部分の構造にはばらつきがあることが示された。本研究で構築したCNNモデルでは、空洞部分の構造の類似性だけでなく、タンパク質機能に相関する構造的な特徴を抽出している可能性がある。

以上、ヘムタンパク質におけるヘムの

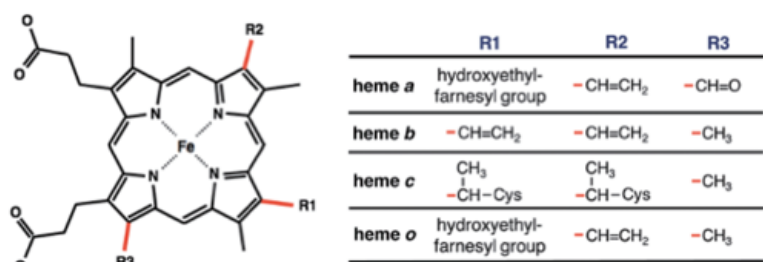


図1 ヘムの化学構造。

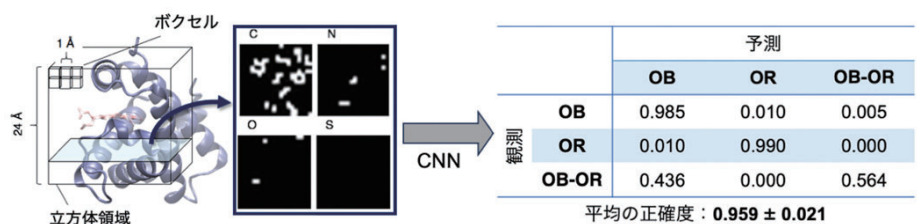


図2 CNNを用いたヘムタンパク質の機能予測。

多機能性の起源の解明に向け、構造データを包括的に取り扱う理論研究を行っている。ほかにも、包括的解析基盤となるヘムの構造データベースと統計解析ツールを統合したwebサービスであるPyDISH (Python-based database and analyzer for **DIS**tortion of **Heme porphyrin**) の開発 (<https://pydish.bio.info.hiroshima-cu.ac.jp/>) [2] や、機械学習や量子化学、分子動力学シミュレーションを併用したヘムタン

パク質の研究も進めている [3-7]。その実施には大量の計算資源を必要とし、計算科学研究センターの整備された計算環境は我々の研究推進の助けとなっている。今後も計算科学研究センターのサポートを受けることができれば幸いである。

本研究は、近藤寛子先生（北見工業大学工学部）、兼松佑典先生（広島大学大学院先進理工系科学研究科）との共同研究として実施された。



たかの・ゆう  
2002年大阪大学大学院理学研究科博士後期課程修了。博士（理学）の学位を取得。2002年日本学術振興会海外特別研究員、2004年日本学術振興会特別研究員、2005年大阪大学蛋白質研究所助教を経て、2015年より広島市立大学大学院情報科学研究科教授。専門は計算化学。

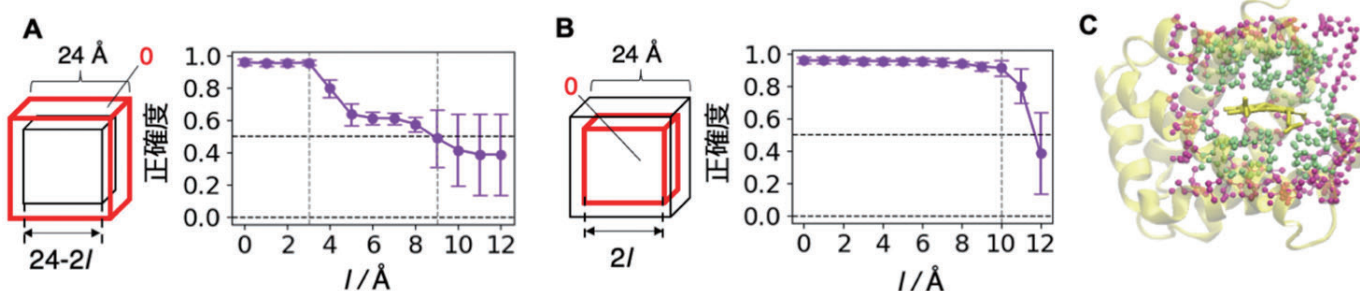


図3 (A) タンパク質ポケットの情報を外側から削除したときの正確度。(B) タンパク質ポケットの情報を内側から削除したときの正確度。(C) タンパク質ポケットを構成する原子 (PDB ID: 1A00)。黄緑の球が18 Åの立方体領域 (タンパク質ポケット) に含まれる原子。黄緑の球とマゼンタの球が24 Åの立方体領域 (タンパク質ポケット) に含まれる原子。

#### 参考文献

- [1] H. X. Kondo, H. Iizuka, G. Masumoto, Y. Kabaya, Y. Kanematsu, Y. Takano, *Biomolecules* **13**, 137 (2023).
- [2] H. X. Kondo, Y. Kanematsu, G. Masumoto, Y. Takano, *Database* **2023**, baaa066 (2023).
- [3] Y. Kanematsu, H. X. Kondo, Y. Imada, Y. Takano, *Chem. Phys. Lett.* **710**, 108–112 (2018).
- [4] H. X. Kondo, Y. Takano, *Life* **12**, 210 (2022).
- [5] H. X. Kondo, M. Fujii, T. Tanioka, Y. Kanematsu, T. Yoshida, Y. Takano, *J. Chem. Inf. Model.* **62**, 775–784 (2022).
- [6] H. X. Kondo, H. Iizuka, G. Masumoto, Y. Kabaya, Y. Kanematsu, Y. Takano, *Biomolecules* **12**, 1172 (2022).
- [7] H. X. Kondo, Y. Kanematsu, Y. Takano, *Chem. Lett.* **51**, 704–708 (2022).

### 共同利用研究ハイライト

## 液体の統計力学理論による 生体分子の機能解析

吉田 紀生 名古屋大学大学院情報学研究科 教授

生命活動はタンパク質やDNAなどの生体分子が織りなす化学過程の連鎖であり、その化学過程は水以外にもイオンや小分子さまざまな成分を含む水溶液の中で起こっている。我々のグループでは、このような溶液中での生体分

子過程を対象とした理論化学手法を開発している。本稿では、分子研共同利用を通じて行った最近の開発事例について紹介する。

まず、溶媒分極を考慮した3D-RISM理論の開発について述べる。3D-RISM

理論は分子性液体の積分方程式理論の一つであり、分子研名誉教授である平田文男先生のグループで精力的に開発が進められてきた理論である（著者も博士研究員・助教として研究に参加していた）。この理論を用いることで生体