

タンパク質の金属イオン結合部位を高精度・高速に予測する深層学習法を新規開発

～金属酵素設計、メタロプロテオーム解析、創薬への貢献に期待～

ポイント

- ・アミノ酸配列と立体構造に基づいて 14 種類の金属イオンの結合部位を予測する「PRIME」を開発。
- ・高精度かつ高速な予測でタンパク質の大規模データベースから金属タンパク質の探索が可能に。
- ・金属タンパク質の機能解明と利用、未知金属タンパク質の発見、創薬応用への貢献が期待。

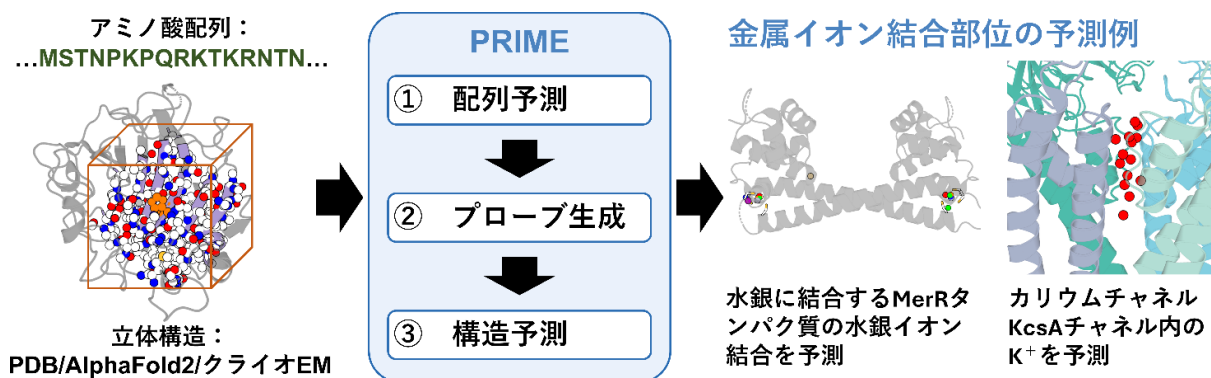
概要

北海道大学大学院地球環境科学研究院の許 仕傑博士研究員、小野田晃教授の研究グループは、金属タンパク質^{*1}の金属イオン結合部位^{*2}を高精度かつ高速に予測する深層学習^{*3}の新手法「PRIME (プライム)」を開発しました。

金属タンパク質は全タンパク質のおよそ 3 分の 1 を占め、物質の変換や、金属イオンの量を一定に保つはたらき（恒常性の維持）など生命活動の根幹を担う機能を有しています。その機能の理解には金属イオンの結合部位を正確に特定することが重要です。しかし、アミノ酸配列や立体構造のどちらか一方の情報のみを使って金属イオン結合部位を予測する従来法は、予測精度・計算速度・予測できる金属イオンの種類という三つの点で課題がありました。

「PRIME」は、アミノ酸配列を扱うタンパク質言語モデル^{*4}と、立体構造を扱う事前学習済みの構造モデルを統合し、両者を橋渡しする独自の「プローブ^{*5}生成アルゴリズム」で金属イオン結合部位の候補点を効率的に絞り込むことで、この課題を解決しました。その結果、遷移金属（亜鉛・鉄・銅など）に加え、結合が弱く従来予測が難しかった非遷移金属（ナトリウム・カリウムなど）を含む 14 種類の金属イオンで、従来法を上回る精度を達成しました。しかも、1 タンパク質鎖あたり約 11 秒という高速で予測できます。クライオ電子顕微鏡などの実験から得られる構造や AlphaFold2^{*6} の予測構造にも有効であり、大規模なメタロプロテオーム^{*7} 解析への応用が期待されます。PRIME は Web サーバー (<https://onodalab.ees.hokudai.ac.jp/prime>) として無償で公開されています。

なお、本研究成果は、2026 年 8 月 11 日（火）公開の Nature Communications 誌にオンライン掲載されました。



本研究の概要図

【背景】

金属タンパク質は、様々な物質変換、金属イオンの恒常性など生命活動の根幹を担う機能を有しており、タンパク質全体のおよそ3分の1は金属タンパク質であると考えられています。金属イオンが結合することで、タンパク質構造が安定化したり、金属酵素の活性中心が形成されます。また、金属イオンの輸送の機能においても結合部位は重要な役割を果たします。金属タンパク質や金属酵素は物質変換で産業に利用されているだけでなく、生命活動に関わることから創薬ターゲットとしても重要です。

金属タンパク質の金属イオン結合部位の精密な構造情報を得るには単結晶 X 線構造解析や核磁気共鳴による実験データが必要でした。このような背景のもと、金属タンパク質の金属イオン結合部位を計算によって予測する手法の開発が進められてきました。

亜鉛・鉄・銅などの遷移金属イオンの結合位置の違いは僅かであり、種類を正確に予測することは困難でした。カルシウムやカリウムでは結合位置の予測がさらに難度が高く解決されていませんでした。AlphaFold2 などの深層学習を用いた構造予測でも、多くの場合は金属イオン結合部位までは高い精度で予測できないことが多いのが現状でした。分子動力学計算や量子化学計算を使った予測も、遷移金属の取り扱いが困難であり、計算負荷も膨大という課題があります。これまでに、タンパク質言語モデルを用いる配列ベースの手法 (LMetalSite、M-Ionic など)、立体構造に基づく深層学習手法 (Metal3D、AllMetal3D など)、テンプレートや幾何学的探索に基づく手法 (MIB/MIB2、BioMetAll など) が開発されてきました。しかし、配列のみを扱う手法は立体情報を扱えないこと、立体構造のみの手法は鎖をまたぐ金属イオン結合部位やデータの少ない金属イオンの予測能力が不十分で、精度と計算速度の両立が大きな課題でした。

【研究手法】

本研究で開発した「PRIME (PRobe-based Identification of MEtal-binding sites)」は、配列情報と立体構造情報の双方を活用して金属イオン結合部位を予測します (図1)。まず、配列予測モジュール (PRIME-seq) が、タンパク質言語モデルによってアミノ酸配列を高次元の数値表現へと変換し、各残基が金属イオンと結合し得る確率を評価します。次に、独自の「プローブ生成アルゴリズム」が、結合する確率が高いと評価された残基の周囲に、立体構造上の幾何学的な配置を考慮しながら金属イオン結合の候補点 (プローブ) を効率的に生成します。その後、構造予測モジュール (PRIME-probe) が、各プローブ周辺の局所的な立体構造を三次元の残差ネットワーク (ResNet3D) で解析し、金属イオンが結合する確率と正確な位置を予測します。この構造モデルは、約1万7千個の重複のないタンパク質立体構造を用いた自己教師あり事前学習を行ったのち、前段で生成したプローブを用いて微調整 (ファインチューニング) しており、予測精度が大きく向上しています。最後に、確率に基づく絞り込みと空間的なクラスタリングによる後処理を行い、最終的な金属イオン結合部位を出力します。PRIME は金属イオンごとに個別に学習・適用を行っているため、金属イオンそれぞれに固有の結合パターンを把握しています。

【研究成果】

PRIME は14種類の金属イオンについて、既存手法 (BioMetAll、Metal3D、LMetalSite、M-Ionic など) を上回る精度・適合率・再現率を達成しました (図2)。アブレーション解析から、立体構造を用いた事前学習が精度向上に必須であることも示されました。また、立体構造に基づく既存手法が1タンパク質あたり1~3分を要するのに対し、PRIME は平均10.7秒で予測を完了し、約10倍の高速化を実現しました (図3)。PRIME は実験構造がない場合にも有効で、AlphaFold2 モデルやクライオ電子顕微鏡

構造でも良好に機能します。例えばカリウムイオンチャンネル KcsA (図 4) ではチャンネル内部のカリウムイオン結合部位を正確に予測することができます。ヒト亜鉛活性化チャンネルでは 10 個中 9 個の亜鉛イオン結合部位を予測できています。さらに水銀結合タンパク質 MerR では、結合部位が亜鉛イオンやカドミウムイオンよりも水銀イオンに対して高い選択性を示すことも予測できており、金属イオンの選択性を予測することも分かりました。

PRIME の拡張性を示すため、Pfam データベースから無作為に選んだ 1,000 種類のタンパク質ファミリーのアミノ酸配列を解析したところ、約 14.3% のタンパク質が高い確率で金属イオン結合部位を持つと予測され、全体の 7.9% は既存データベースでは金属イオン結合が特定されていないものでした。即ち、PRIME は、未知の金属タンパク質や金属酵素の発見に貢献する手法であることを示しています。

なお、PRIME は Web サーバーとして無償で公開されています。

【今後への期待】

本研究成果は、PRIME が高精度と高速性を両立しており、大規模なメタロプロテオーム解析を可能にする強力な手法であることを実証しました。今後、PRIME は未知の金属タンパク質の発見、金属酵素の合理的な設計、金属イオンに関連するタンパク質を標的にした創薬研究への応用が期待されます。

【謝辞】

本研究は、北海道大学 DX 博士人材フェローシップ (JST 次世代研究者挑戦的研究プログラム SPRING、JPMJSP2119)、文部科学省・日本学術振興会科学研究費助成事業 学術変革領域研究 (A) 「化学構造リプログラミングによる統合的物質合成科学の創成 (SReP)」 (JP24A202、JP24H02213) 及び基盤研究 (B) (JP24K01533)、JST/JICA 地球規模課題対応国際科学技術協力プログラム SATREPS 「持続可能な漁業を実現する高付加価値バイオ素材の有効利用 (ReBiS)」 (JPMJSA2206) の支援のもとで行われました。

論文情報

論文名	Probe-Based Identification of Metal-Binding Sites Using Deep Learning Representations (深層学習表現を用いたプローブに基づくタンパク質金属イオン結合部位の同定)
著者名	許 仕傑 ^{1,2} 、小野田晃 ^{1,2} (¹ 北海道大学大学院環境科学院、 ² 北海道大学大学院地球環境科学研究院)
雑誌名	Nature Communications (自然科学全般を扱う総合学術誌)
DOI	10.1038/s41467-026-74657-x
公表日	2026 年 8 月 11 日 (火) (オンライン公開)

お問い合わせ先

北海道大学大学院地球環境科学研究院 教授 小野田晃 (おのだあきら)

T E L 011-706-2257 F A X 011-706-2257 メール akira.onoda@ees.hokudai.ac.jp

U R L <https://onoda-lab.jp>

配信元

北海道大学社会共創部広報課 (〒060-0808 札幌市北区北 8 条西 5 丁目)

T E L 011-706-2610 F A X 011-706-2092 メール jp-press@general.hokudai.ac.jp

【参考図】

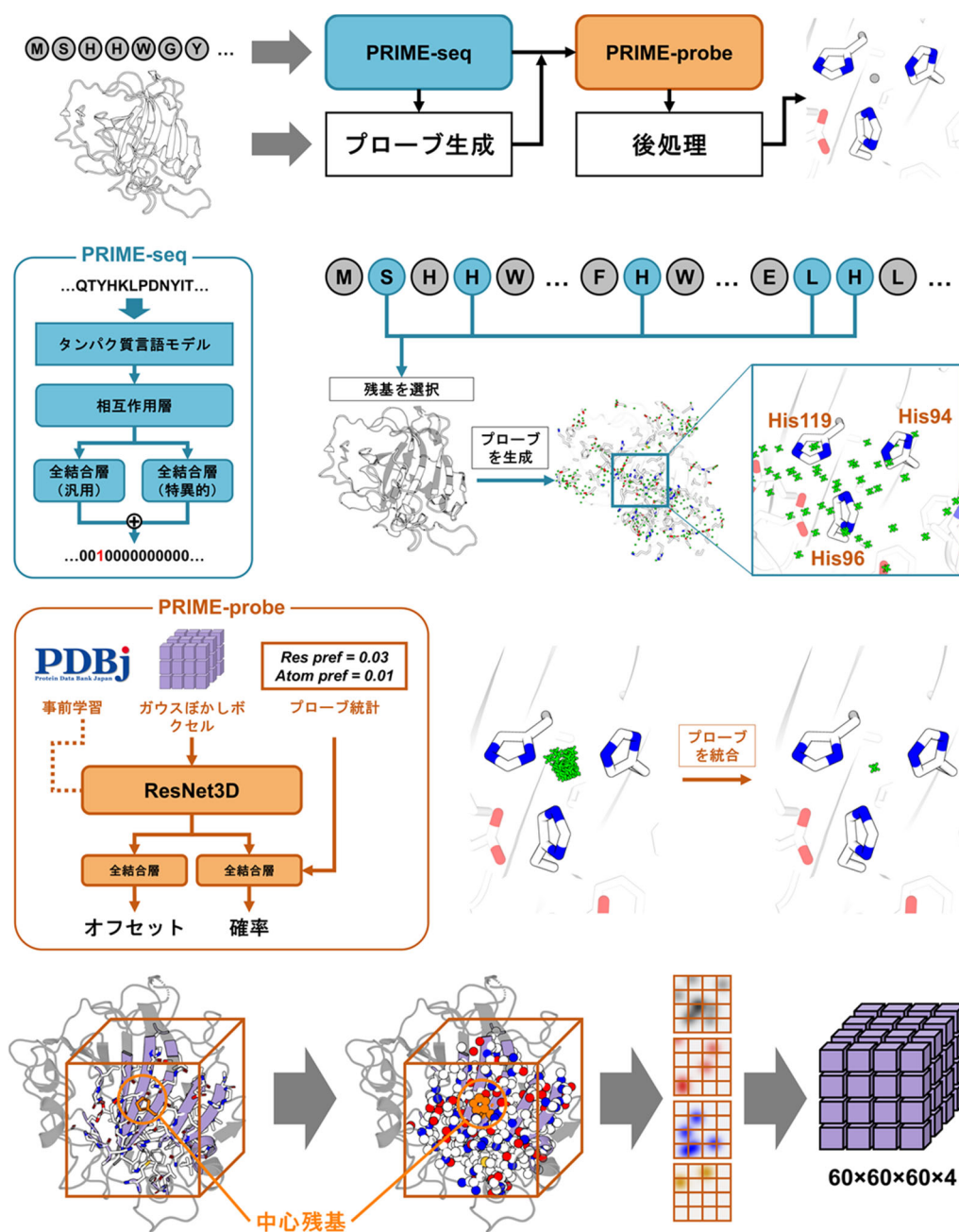


図 1. 深層学習法 PRIME の全体構成。配列を扱う PRIME-seq、立体構造を扱う PRIME-probe、両者を橋渡しするプローブ生成アルゴリズムと後処理から構成される。タンパク質の配列と立体構造を入力として金属イオン結合部位を予測する。

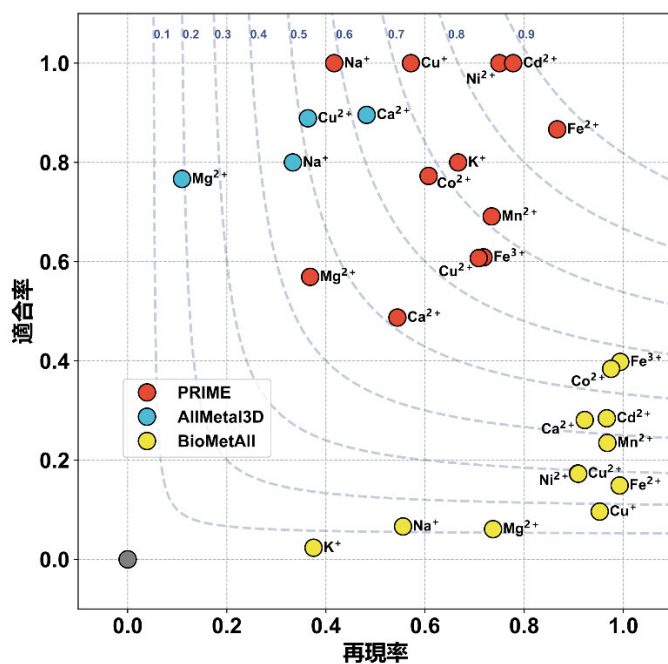


図 2. PRIME と既存手法の予測性能の比較。各点は金属イオンごとの再現率（横軸）と適合率（縦軸）を表し、右上に位置するほど高精度であることを示す。PRIME（赤）は既存手法を上回る性能を示している。

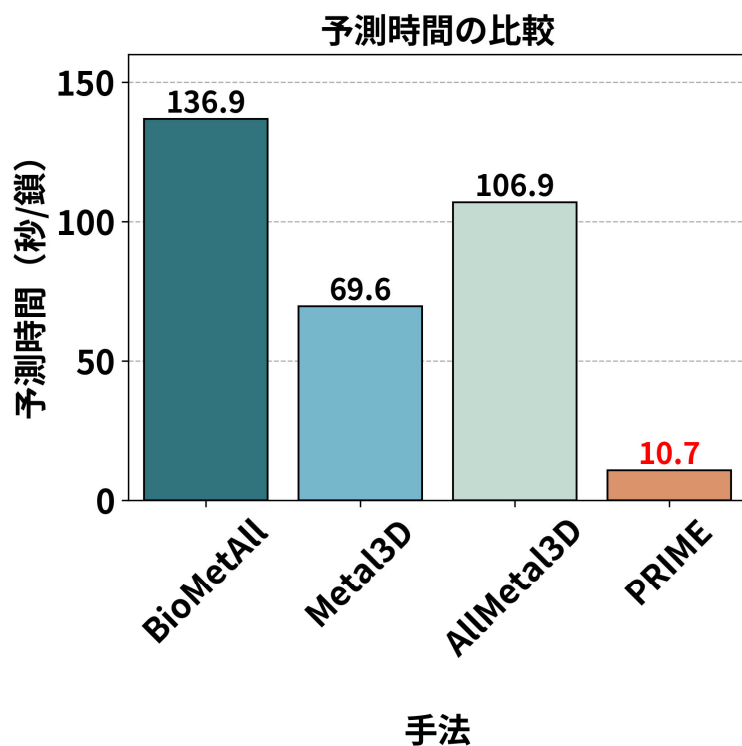


図 3. タンパク質一つあたりの予測時間の比較。PRIME は平均 10.7 秒で予測を完了し、立体構造に基づく既存手法（約 70～137 秒）より大幅に高速である。

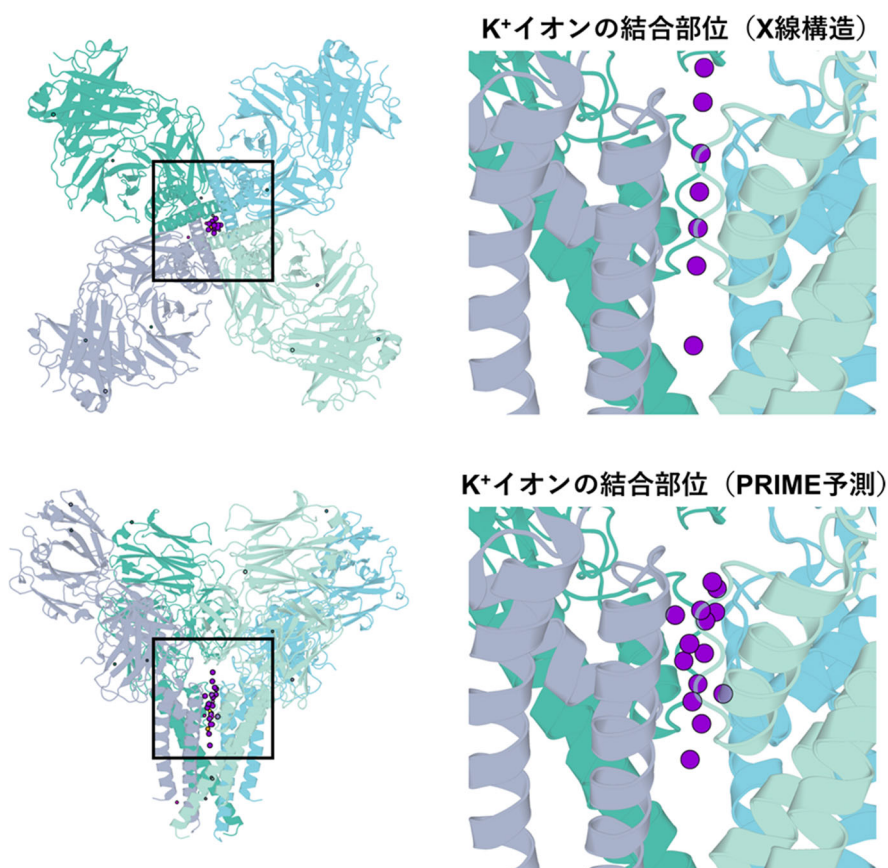


図 4. カリウムイオンチャネル KcsA (PDB ID 1K4C) における金属イオン結合部位の予測。全体の構造（左）。チャンネル内部に並ぶカリウムイオン結合部位（紫）を正確に予測できている（右）。

【用語解説】

- *1 金属タンパク質 … 金属イオンを補因子として持つタンパク質。全タンパク質のおよそ 3 分の 1 を占めるとされ、酵素反応や立体構造の維持などに関わる。
- *2 金属イオン結合部位 … タンパク質が金属イオンと結合する特定の場所。金属イオンは触媒反応・構造維持・電子伝達などに必須であり、その結合部位はタンパク質の機能発現の鍵となる。
- *3 深層学習 … 多層のニューラルネットワークを用いて、大量のデータからパターンや法則を自動的に学習する機械学習の代表的な手法。
- *4 タンパク質言語モデル … 膨大なアミノ酸配列を学習し、配列を高次元の数値表現へと変換する深層学習モデル。配列に潜む進化的・機能的な情報を捉えることができる。
- *5 プローブ … 金属イオンが結合し得る候補点。PRIME は有望な残基の周囲にプローブを生成し、立体構造モデルで結合の有無を判定することで、効率よく結合部位を絞り込む。
- *6 AlphaFold2 … アミノ酸配列からタンパク質の立体構造を高精度に予測する深層学習プログラム。実験構造が存在しないタンパク質の解析を可能にする。
- *7 メタロプロテオーム … ある生物や細胞に存在する金属タンパク質の総体。ゲノムやプロテオームになぞらえ、金属タンパク質を網羅的に解析する対象・分野を指す。