

[受賞者] デイヴィッド・ベイカー (David Baker) 氏 (アメリカ)
デミス・ハサビス (Demis Hassabis) 氏 (イギリス)
ジョン・マイケル・ジャンパー (John M. Jumper) 氏 (イギリス)

「タンパク質の構造予測と人工タンパク質の設計」

東京理科大学 創域理工学部 生命生物科学科 准教授 なかじま 中島 まさひろ 将博

■2024年ノーベル化学賞の受賞のテーマ

2024年のノーベル化学賞は、「タンパク質の驚くべき構造の暗号を解読した」として、全く新しい人工タンパク質の設計に対しワシントン大 (アメリカ) の David Baker 氏、高精度のタンパク質構造予測を可能にしたことに対し DeepMind 社 (イギリス) の Demis Hassabis 氏と John M. Jumper 氏に授与されました。

■タンパク質の構造予測は50年来の大きな課題

タンパク質は酵素として化学反応を触媒するだけでなく、抗体としての生体防御、物質輸送、組織構造の骨格として機能するなど生体における重要な鎖状分子です。タンパク質は、ポリペプチド鎖が一定の形に折り畳まれて初めて機能を発揮します。そのため、その機能を知るにはタンパク質の3D構造を知る必要があります。3D構造は、主に結晶化したタンパク質にX線を照射して得られるデータから個別に解析されてきました。しかし、実験的に3D構造を得るには多くの時間と労力を要します。

1972年のノーベル化学賞受賞者のアンフィンセン氏がタンパク質の構造はアミノ酸配列によって決まる、という仮説を提唱して以来、タンパク質のアミノ酸配列から3D構造を予測する試みが続けられてきました。しかし、タンパク質構造の自由度は理論的には天文学的な数字であり、その予測は極めて困難でした。その後、実際のタンパク質構造のデータが蓄積すると、その構造を参照することで、参照タンパク質とアミノ酸配列が類似しているタンパク質の3D構造を精度よく予測できるようになりました。しかし、参照となる構造がない場合、タンパク質構造の正確な予測は依然と

して困難でした。

Baker 教授は人工タンパク質の設計の先駆的な研究者の一人であり、亜鉛結合タンパク質、ステロイドに高い結合能を示すタンパク質、当時知られていた構造のパターンにはない新規な構造のタンパク質、いくつかの人工酵素の作製に成功していましたが、構造予測が困難な中でのこのような設計はきわめて高度な職人技であったといえます。

■AlphaFold2の衝撃

タンパク質構造予測精度の向上を促進する目的で、その精度を競う CASP (Critical Assessment of protein Structure Prediction) という大会が1994年以降2年ごとに開催され、Baker 教授のグループも3回目の大会から参加し、常に好成績を収めていました。ところが、2018年の大会で DeepMind 社の AlphaFold というグループがトップの成績を叩き出しました。DeepMind 社は2015年に囲碁のトップ棋士を破った AlphaGo を開発し、人工知能の有用性を広く世界に知らせた会社です。この大会ではまだ大きな話題にはなりませんが、2020年の大会では AlphaFold2 として参加し、圧倒的な大差でトップの成績を収め、しかも、出題されたタンパク質の半分以上でほぼ実験的に得られた3D構造とほぼ同一ともいえる予測構造を出力し、驚くべき高さの予測精度を示しました。

その後、2021年7月には *Nature* 誌に AlphaFold2 の詳細が掲載され¹⁾、AlphaFold2 のソースコードも公開されました。さらに、これをベースとした ColabFold²⁾ などの平易に使えるオンラインのツールも登場しました。これにより、実験で得たタンパク質構造で未公開のものを持っている研究者からはその予

測精度の高さに驚いたという話が頻出しました。また、実験で得たデータはあるものの、そのデータから目的の構造を得られない場合があります。しかし、目的のタンパク質の予測構造を用いることで、長年得られなかった3D構造が解析不能であったデータからあっさり得られた、と驚く研究者も少なからずいたと思われます。

■ AlphaFold2の原理

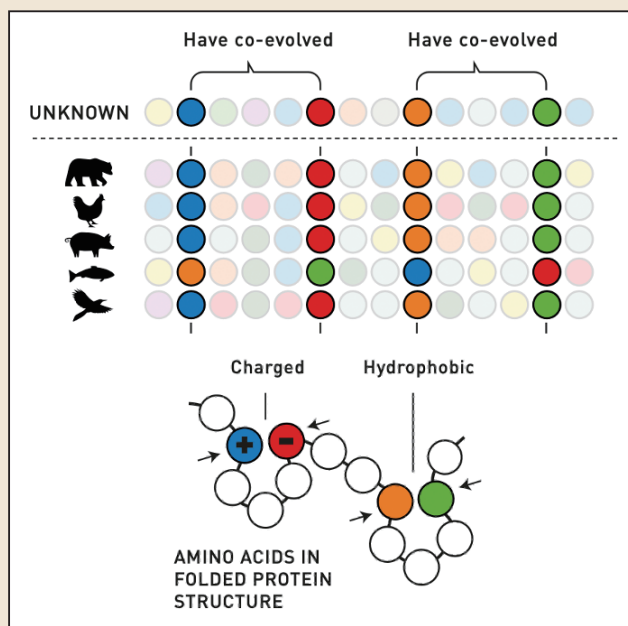
AlphaFold2の成果は驚異的ですが、基本的な部分では従来の構造予測の方法論に基づいています。アミノ酸配列が類似したタンパク質との配列比較を行うと、ペアでアミノ酸残基が変異している箇所が見出されます。これを共進化といい、このペアは空間的に近傍に存在する傾向があることが知られています【図】。AlphaFold2でもこの配列比較により共進化と推定されるアミノ酸残基のペアを抽出します。そして、Protein Data Bankという3D構造の良質のデータベースで学習させたアルゴリズムでこのペアの位置関係を推定し、最終的な予測構造を出力します。したがって、多くの研究者が解いた3D構造の蓄積の上に、トランスフォーマーなどの近年の主流の機械学習の技術が融合してもたらされたのがAlphaFold2のブレークスルーといえるでしょう。

■ AlphaFold2後の展開と展望

AlphaFold2の登場後ほどなくして、まずヒト、マウスなどの生物由来のタンパク質のAlphaFold2による予測構造がデータベースに収録されました。現在は、微生物を含む様々な生物種から2億個を超える予測構造が公開されています。実験的に解析され登録、公開された3D構造の数が(執筆時点で)20万個程度です。その数は膨大です。誰もがおよそ正しい3D構造に容易にアクセス可能になり、タンパク質構造の利用の裾野が一気に広がりました。

タンパク質構造予測は進化を続けています。最近ソースコードが公開されたAlphaFold3では、画像データがノイズの追加により破壊される過程を学習して画像生成に用いる拡散モデルという仕組みを新たに導入しています³⁾。タンパク質以外の化合物と結合した構造や糖鎖修飾などを含むタンパク質構造の予測が可能になり、構造予測の適用範囲が広がりました。

高精度の構造予測が可能になったことで人工タンパ



【図】 共進化によるアミノ酸残基の変異

THE NOBEL PRIZE IN CHEMISTRY 2024 (Popular Science Background © Johan Jarnestad/The Royal Swedish Academy of Sciences) を一部抜粋

生物種ごとにペアとなって置換している青と赤または黄と緑の○で示されたアミノ酸残基(上図)が空間的に近い位置に存在する傾向があることを表している(下図)。

ク質の設計における展開も加速しています。Baker教授のグループは、AlphaFold2登場後の短期間にモーターを模したタンパク質、補因子であるヘムを結合したタンパク質やタンパク質に結合するタンパク質、化学発光を触媒する酵素(ルシフェラーゼ)を人工的に作成しています。このように、作成できるタンパク質のレパートリーは増えつつあり、応用利用を目指してタンパク質を人工的に設計する流れは今後盛んになっていくでしょう。

参考文献

- 1) *Nature* 596 (2021) Jumper, J., et al.
- 2) *Nature Methods* (2022) Mirdita, M., et al.
- 3) *Nature* 630 (2024) Abramson, J., et al.