

発表論文

Abramson, J., Adler, J., Dunger, J. *et al.* Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature* (2024). <https://doi.org/10.1038/s41586-024-07487-w>

研究目的・概要

AlphaFold2 の登場によりアミノ酸配列からタンパク質の立体構造を高精度で予測することが可能になり、生命科学の研究に大きな影響を与えた。しかし生体内においてタンパク質は単体ではなく他のタンパク質、核酸、低分子リガンドなどの分子と相互作用し、複合体を形成することでその機能を発揮する。これまで生体分子の複合体の立体構造を予測する計算手法がいくつか開発されてきたが、深層学習をベースにした手法はドッキングシミュレーション手法と比較して精度が劣る場合が多かった。そのような状況のもとで、John Jumper 率いる Google Deepmind 社の研究チームはあらゆる生体分子複合体の構造を高精度で予測する AlphaFold 3 を開発した。

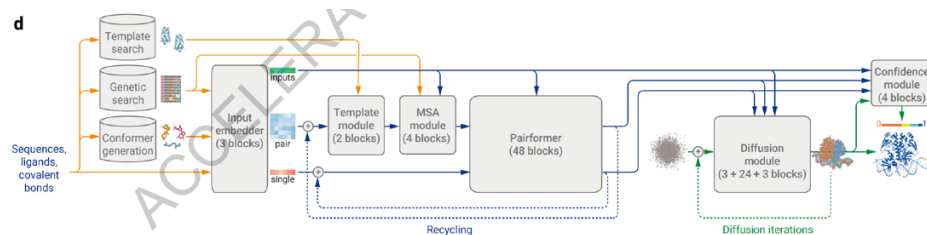


図 1. AlphaFold 3 モデルの構造

AlphaFold 3 はタンパク質のアミノ酸配列、核酸配列、リガンド情報を元に複合体の立体構造を推論する。まず、入力情報をもとに既知のフラグメント構造検索、MSA による相同配列検索、リガンドのコンフォーマー生成を行う。これらの結果はトークン化され AtomTransformer と呼ばれるネットワークにより原子レベルの情報を含む表現へと変換される。得られた表現は複数の処理ブロックを経て特徴量化され、最後に拡散モデルによる構造生成と、予測結果に対する確信度の定量化が行われる (図 1)。

先行研究と比べて何がすごい？技術やアプローチのキモはどこ？

- AlphaFold2 の構造をベースにタンパク質以外の分子に対応できるようにモデル構造の大幅な改良を行った。
- MSA 表現に対する処理を削減し、原子レベルの情報を保持したペア表現に対する処理を重視したネットワークを構築した。
- 構造生成に拡散モデルを導入し、全原子の絶対座標を推論することを実現した。
- 拡散モデルを導入したことにより生じた hallucination への対策として、AlphaFold Multimer により得られるタンパク質構造を蒸留データセットとして使用した。

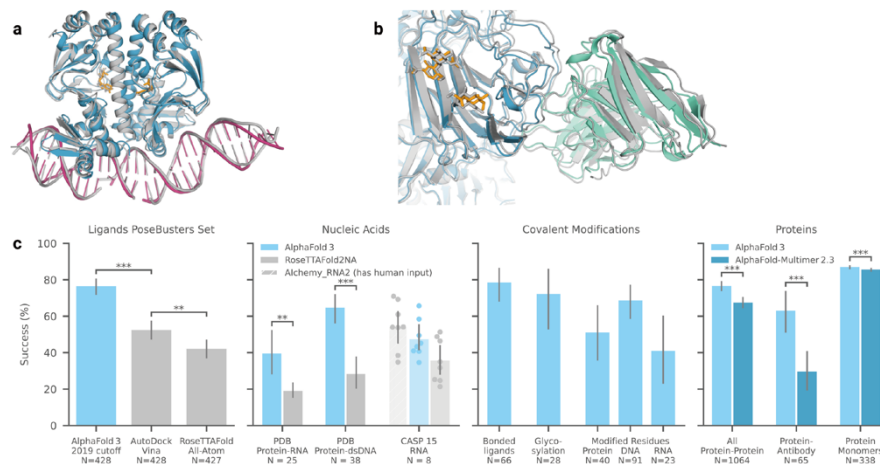


図 2. AlphaFold 3 による複合体立体予測結果、他モデルとの性能比較

どうやってこの手法・仮説の有効性を検証したか。

- タンパク質—核酸複合体や糖鎖修飾を受けたタンパク質—抗体複合体など多様な生体分子の立体構造を高精度で予測することができることを示した (図 2a, b)。
- Ligands PoseBusters ベンチマークデータセットを使用した検証により、AlphaFold 3 の予測性能がドッキングシミュレーションや既存の深層学習モデル (RoseTTAFold All-Atom) と比較して優れていることを示した (図 2c、左端)。
- AlphaFold Multimer と比較を行い、タンパク質単体や抗体との複合体の構造予測精度が向上していることを示した (図 2c、右端)

その他議論した内容

- AlphaFold 3 のソースコードは公開されておらず、使用は多くの制限がかけられたウェブサーバーを通じたアクセスに限られている。AlphaFold2 は発表当初からソースコードが公開されており、世界中の研究者が実験や工夫を通して見つけた活用法や拡張を共有したことで生命科学研究のさらなる進展に繋がった。もし今後も AlphaFold 3 のソースコードが公開されないのであれば、それは科学にとって大きな損失である。
- 技術の進歩によりタンパク質単体のみならず、あらゆる生体分子の複合体の立体構造が高精度で予測できる時代に突入しようとしている。今後は、生体分子のスナップショットではなく、立体構造の変化の過程などのダイナミクスを予測する手法の開発が目標となることが考えられる。また、タンパク質のアミノ酸配列から詳細な機能の予測を予測する手法が開発されれば大きなインパクトを与えるだろう。

この研究をさらに発展させるとしたら

- AlphaFold 3 を活用した創薬を目指す。特定のターゲットタンパク質に高い結合親和性を持つことが予測されるリガンドを探索し、その結果を実験的に検証する。