

AI 特許紹介(78)
AI 特許を学ぶ！究める！
～AlphaFold 3 特許～

2025 年 7 月 10 日
河野特許事務所
所長 弁理士 河野英仁

「AI 特許紹介」シリーズは、注目すべき AI 特許のポイントを紹介します。熾烈な競争となっている第 4 次産業革命下では AI 技術がキーとなり、この AI 技術・ソリューションを特許として適切に権利化しておくことが重要であることは言うまでもありません。

AI 技術は Google, Microsoft, Amazon を始めとした IT プラットフォーマ、研究機関及び大学から毎週のように新たな手法が提案されており、また AI 技術を活用した新たなソリューションも次々とリリースされています。

本稿では米国先進 IT 企業を中心に、これらの企業から出願された AI 特許に記載された AI テクノロジー・ソリューションのポイントをわかりやすく解説致します。

1.概要

特許出願人 Deepmind

出願日 2024 年 5 月 21 日

公開日 2024 年 11 月 28 日

公開番号 WO2024240774

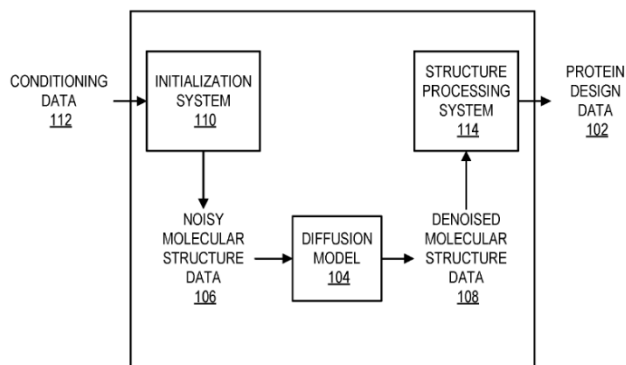
発明の名称 完全な原子表現で動作する拡散モデルを用いたタンパク質設計

774 特許は、ノイズ除去ニューラルネットワークを含む拡散モデルを用いることにより、特定の機能を持つタンパク質を設計することが可能な Alphafold3 技術に関する。

2.特許内容の説明

2021 年にタンパク質の折り畳み構造を予測する Alphafold2¹が公表されたが、2024 年にその改良版である Alphafold3 が公表された。Alphafold3 は、原子の空間位置を出力するために拡散モデルを用いている点が特徴である。以下に詳細を説明する。図 1 はタンパク質設計システム 100 の一例を示す。

¹ <https://knpt.com/contents/ai/2021.09.10.pdf>



タンパク質設計システム 100 は、拡散モデル 104 を用いて、設計されたタンパク質を特徴付けるタンパク質設計データ 102 を生成する。システム 100 は、拡散モデル 104 を用いて、特定の機能を持つタンパク質を設計することができる。例えば、システム 100 は、特定のリガンド（例えば、低分子リガンドまたはタンパク質リガンド）と結合するようにタンパク質を設計する。さらに別の例として、システム 100 は、特定のリガンドの特性を示すデータを受信し、受信した特定のリガンドの特性を示すデータに基づいて、設計されたタンパク質のタンパク質設計データ 102 を生成する。別の例として、システム 100 は、特定の標的タンパク質と結合するようにタンパク質を設計する。

タンパク質を設計するために、システム 100 は、拡散モデル 104 を用いてノイズを含む分子構造データ 106 を処理し、設計されたタンパク質を特徴付けるノイズ除去済み分子構造データ 108 を生成する。ノイズを含む分子構造データ 106 およびノイズ除去済み分子構造データ 108 は、設計されたタンパク質のアミノ酸配列内の各アミノ酸について、対応するアミノ酸埋め込みを含む。各アミノ酸埋め込みは、タンパク質の対応するアミノ酸内に含まれる可能性のある、定義済みの原子セット（すなわち、集合）のそれぞれについて、一連の原子埋め込みを含む。定義済みの原子セットは、同じタイプの原子（すなわち、元素）の複数のインスタンスを含む。各原子埋め込みは、対応する原子の空間位置を特徴付ける。

ノイズを含む分子構造データ 106 は、設計タンパク質のアミノ酸に含まれる可能性のある各原子の初期空間位置を指定する。拡散モデル 104 は、ノイズ除去処理後にノイズを含む分子構造データ 106 を処理し、設計タンパク質を特徴付けるノイズ除去済み分子構造データ 108 を生成する。ノイズ除去処理は反復処理であり、拡散モデル 104 は、ノイズを含む分子構造データ 106 を一連のノイズ除去反復にわたって反復的にノイズ除去することにより、ノイズ除去済み分子構造データ 108 を生成する。ノイズ除去処理後、拡散モデル 104 は、設計タンパク質の各アミノ酸について、(i)アミノ酸に含ま

れる可能性のある原子の位置と、(ii)アミノ酸に含めるべき原子の両方を決定する。

設計されたタンパク質内の特定のアミノ酸は、そのアミノ酸に含まれる可能性のある定義済みの原子集合の全ての原子を含まない可能性がある。拡散モデル 104 は、設計されたタンパク質のアミノ酸に含めるべき原子を決定し、ノイズ除去された分子構造データ 108 を生成することで、設計されたタンパク質の各アミノ酸に含まれる原子をエンコードする。特に、拡散モデル 104 は、生成された原子の空間位置を用いて、設計されたタンパク質に含まれる原子をエンコードする。例えば、設計されたタンパク質の各アミノ酸は、そのアミノ酸の「廃棄 throw-away」位置（空間位置）に関連付けることができ、拡散モデル 104 は、特定の原子をアミノ酸の廃棄位置に配置することで、その原子が設計されたタンパク質から除外されることをエンコードする。

タンパク質設計システム 100 は、初期化システム 110 を含む。初期化システム 110 は、ノイズを含む分子構造データ 106 を生成する。特に、初期化システム 110 は、設計されたタンパク質内に含まれる可能性のある原子の初期空間位置の一部またはすべてをノイズ分布（例えば、ガウス分布、一様分布など）からサンプリングすることにより、ノイズを含む分子構造データ 106 を初期化する。初期化システム 110 は、コンディショニングデータ 112 を受信し、コンディショニングデータ 112 に基づいてノイズの多い分子構造データ 106 を生成する。

コンディショニングデータ 112 は、設計されたタンパク質の 1 つ以上の原子について、事前定義された空間位置を指定し、初期化システム 110 は、ノイズを含む分子構造データ 106 を生成して、それらの原子を、拡散モデル 104 によって空間位置が変更されない静的原子として指定する。これにより、タンパク質設計システム 100 は、静的原子を含むようにタンパク質を設計する。

別の例として、コンディショニングデータ 112 は、標的分子（例えば、設計タンパク質の標的小分子リガンド、標的タンパク質リガンド、標的タンパク質など）を特徴付けることができ、初期化システム 110 は、ノイズを含む分子構造データ 106 を生成して、標的分子の構造を定義する。一般に、ノイズを含む分子構造データ 106 は、標的分子の各原子の初期空間位置を特定する。特に、ノイズを含む分子構造データ 106 は、標的分子の原子の空間位置を特徴付ける 1 つ以上の埋め込みを含む。

ノイズを含む分子構造データ 106 に標的分子の埋め込みが含まれる場合、標的分子の埋め込みは、設計タンパク質のアミノ酸埋め込みと同じ次元を共有することができる。例えば、標的分子の埋め込みは、設計タンパク質に含まれる可能性のある事前定義され

た原子集合からの原子埋め込みとすることができる。

初期化システム 110 は、標的分子の構造を定義するためのノイズ分子構造データ 106 を生成する際に、標的分子内の 1 つ以上の原子を、拡散モデル 104 によって変更されない所定の空間位置を有する静的原子として指定するためのノイズ分子構造データ 106 も生成する。例えば、初期化システム 110 は、標的分子のすべての原子を、拡散モデル 104 によって変更されない所定の空間位置を有する静的原子として指定するためのノイズ分子構造データ 106 を生成することができる。これにより、タンパク質設計システム 100 は、標的分子に結合するタンパク質を設計することができる。

タンパク質設計システム 100 は、構造処理システム 114 を含む。構造処理システム 114 は、拡散モデル 104 によって生成されたノイズ除去された分子構造データ 108 を処理し、タンパク質設計データ 102 を生成する。

上述のように、ノイズ除去済み分子構造データ 108 は、設計されたタンパク質の各アミノ酸に含まれる原子をエンコードする。構造処理システム 114 は、ノイズ除去済み分子構造データ 108 を処理して、設計されたタンパク質から除外された原子を除去する。

構造処理システム 114 は、ノイズ除去された分子構造データ 108 に対して様々な演算処理を実行し、設計タンパク質のアミノ酸配列中のアミノ酸の同一性を決定する。例えば、構造処理システム 114 は、原子とアミノ酸の集合間の事前定義されたマッピングを用いて、設計タンパク質のアミノ酸配列中のアミノ酸の同一性を決定する。

タンパク質設計システム 100 は、タンパク質を設計した後、設計されたタンパク質を合成するために、タンパク質設計データ 102 を外部システムに出力する。例えば、タンパク質設計データ 102 は、設計されたタンパク質のアミノ酸配列を指定し、外部システムは、指定されたアミノ酸配列を有するタンパク質を合成する。

3.クレーム

774 特許のクレーム 1 は以下の通りである。

1. 1 台以上のコンピュータでタンパク質を設計する方法において、

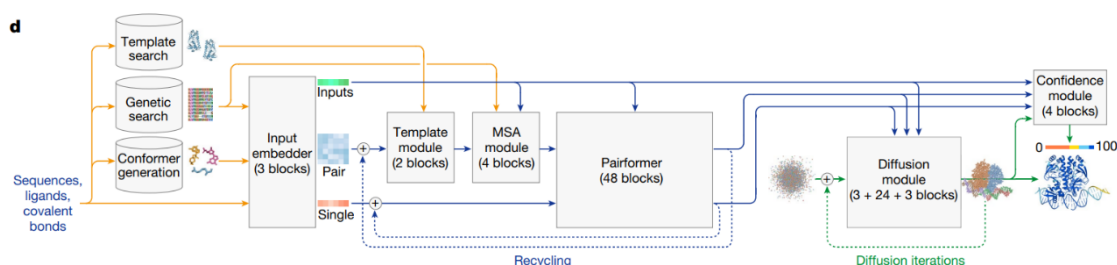
タンパク質のアミノ酸配列内の各位置について、可能な原子の事前定義されたセット内の各原子の対応する初期空間位置を定義するノイズを含む分子構造データを生成し、ノイズを含む分子構造データ内の初期空間位置の少なくともいくつかは、ノイズ分布からサンプリングされ、

ノイズの多い分子構造データのノイズ除去バージョンを定義するノイズ除去分子構

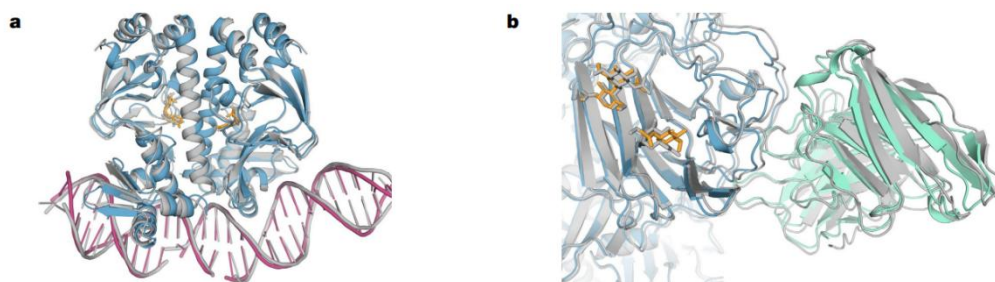
造データを生成するために、ノイズ除去ニューラルネットワークを含む拡散モデルを使用してノイズの多い分子構造データを処理する。

4. 本特許に関連する論文

本特許に関する論文 “Accurate structure prediction of biomolecular interactions with AlphaFold 3”²が、Deepmind の Josh Abramson 氏らにより nature 誌に公表されている。2024 年ノーベル化学賞を受賞した Demis Hassabis 氏及び John Jumper 氏も本論文の発表者に含まれている。下記図は Alphafold3 のパイプラインを示す説明図である。右側に本稿で説明した拡散モデルが設けられている。



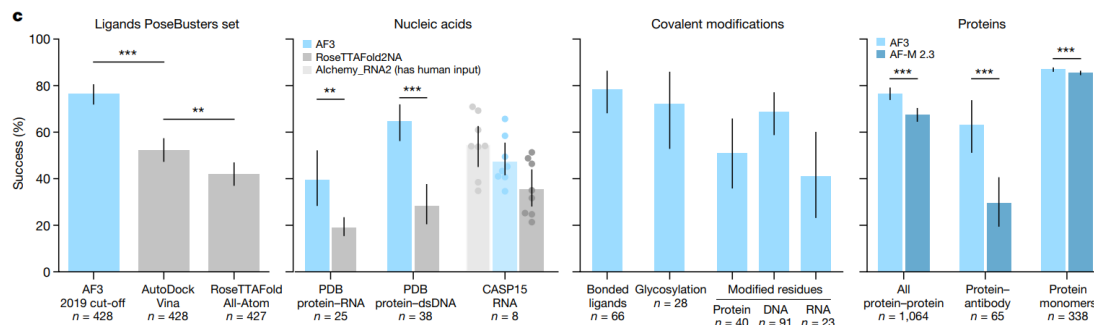
下記図 a 及び b は Alphafold3 により予測された構造例を示す。a は DNA 及び cGMP に結合した細菌性 CRP/FNR ファミリー転写調節タンパク質 (PDB 7PZB; 完全複合体 LDDT47, 82.8; 全体距離テスト (GDT) 48, 90.1) であり、b はヒトコロナウイルス OC43 スパイクタンパク質 (4,665 残基) である。



下記グラフは PoseBusters、最新の PDB 評価セット、CASP15 RNA での Alphafold3 のパフォーマンスを示す。AlphaFold3 は、従来の多くの専用ツールに比べて大幅に改善された精度を示している。最先端のドッキングツールと比較して、タンパク質-リガンド相互作用の精度がはるかに高く、核酸固有の予測子と比較して、タンパク質-核酸

² Josh Abramson, et al. “Accurate structure prediction of biomolecular interactions with AlphaFold 3” <https://doi.org/10.1038/s41586-024-07487-w>

相互作用の精度がはるかに高く、AlphaFold-Multimer v.2.3 と比較して、抗体-抗原予測の精度が大幅に高くなっている。



以上

著者紹介

河野英仁

河野特許事務所、所長弁理士。立命館大学情報システム学博士前期課程修了、米国フランクリンピアースローセンター知的財産権法修士修了、中国清华大学法学院知的財産夏季セミナー修了、MIT(マサチューセッツ工科大学)コンピュータ科学・AI 研究所 AI コース、生成 AI ビジネスコース修了。

AI 特許コンサルティング、医療 AI 特許コンサルティングの他、米国・中国特許の権利化・侵害訴訟を専門としている。著書に「世界のソフトウェア特許(共著)」、「FinTech 特許入門」、「AI/IoT 特許入門 3」、「ブロックチェーン 3.0(共著)」がある。