

ゲノム情報を用いた構造生物学： ポスト AlphaFold2 の環境毒性学

Structural biology using genomic information:
exploring the post-AlphaFold2 era

武田 一貴*
Kazuki TAKEDA*

北里大学 獣医学部獣医学科 毒性学研究室
Laboratory of Toxicology, School of Veterinary Medicine, Kitasato University

摘 要

AlphaFold2 は深層学習による高精度タンパク質立体構造予測手法であり、新奇性の高いアミノ酸配列に対しても高精度に予測できるうえ、タンパク質複合体の構造予測が可能なることから注目を集めている。Google colabatory 上でアミノ酸配列を入力するだけで予測構造が得られる ColabFold の実装や Uniprot に登録されている全アミノ酸配列の予測構造を収載した AlphaFold Protein Structure Database の公開により専門知識抜きに予測構造を入手できることから爆発的に利用が進んだ。本稿では AlphaFold2 の利用法の概説とともに AlphaFold 予測構造と分子ドッキング・分子動力学シミュレーションを組み合わせた構造解析ベースでの非モデル動物種への毒性評価の利用例を紹介する。

キーワード：AlphaFold2, 環境汚染物質, タンパク質,
分子ドッキングシミュレーション, 野生動物

Key words：AlphaFold2, environmental pollutant, protein,
molecular docking simulation, wildlife

1. 序文

人工知能(AI)技術の発展は目覚ましく、一行の入力文章から実写さながらのリアルな画像を作成する生成 AI が出現したかと思えば(Podell *et al.*, 2023)、文章を基に 1 分程度の映画のようなクオリティの動画を出力する生成 AI(Sora)すら開発されている(OpenAI, 2024)。生物学分野においても例外ではなく、種々の AI 手法が研究に活用されている。その中でも 2020 年代に最も生物学分野全体へ衝撃を与えた AI 技術と言われると、間違いなく AlphaFold2 が挙げられるであろう。AlphaFold2 はアミノ酸配列(のみ)からそのタンパク質の立体構造を超高精度で予測する深層学習手法である(Jumper *et al.*, 2021a)。AlphaFold2 の原著論文は 2021 年 7 月の公開からわずか 3 年足らずで 20,000 回以上引用されており、後述するウェブブラウザ版の ColabFold も 3,700 回もの引用数を有する(2024 年 4 月現在)。AlphaFold2 は何故これほど「バズった」のか? 本稿は、その利用法や原理を概説するとともに環境毒性学研究への応用事例を紹介し、読者の皆様が AlphaFold2 を自身

の研究に取り入れるための導入となることを目的としている。本稿で使用している略称については、表 1 にまとめた。

2. タンパク質立体構造とは

この文脈で述べるタンパク質の立体構造とは、生化学用語におけるタンパク質の三次構造(単一のポリペプチド鎖が折りたたまれ機能的な高次構造を形成した状態)を指す。タンパク質構造解析の歴史は 90 年前の 1934 年に遡り、J. D. Bernal と D. C. Hodgkin らがペプシンの結晶を用いて X 線回折写真の撮影に成功したのが始まりだとされている(Bernal & Crowfoot, 1934)。その後、蛋白質結晶構造解析に最初に成功したのは J. C. Kendrew で、1958 年に分解能 0.6 nm によるミオグロビンの立体構造を発表した(Kendrew *et al.*, 1958)。それに引き続き 1960 年に M. F. Perutz がヘモグロビンの分解能 0.55 nm の解析に成功(Perutz *et al.*, 1998)、これらの功績により J. C. Kendrew と M. F. Perutz は 1962 年にノーベル化学賞を受賞している。その後、X 線結晶構造解析

受付：2024 年 5 月 7 日，受理：2024 年 7 月 23 日

* 〒034-0021 青森県十和田市東 23 番町 35-1, E-mail : takeda@vmas.kitasato-u.ac.jp

表 1 略語一覧.

Abbreviation	Full Name
AI	Artificial Intelligence
AhR	Aryl Hydrocarbon Receptor
bHLH	Basic Helix-Loop-Helix
CASP	Critical Assessment of Structure Prediction
chatGPT	Chat Generative Pre-Trained Transformer
CYP1A	Cytochrome P450 1A
GDT-TS	Global Distance Test-Total Score
GHECOM	Grid-based Hecomi Finder
GPU	Graphics Processing Unit
LBD	Ligand-Binding Domain
MSA	Multiple Sequence Alignment
PAE	Predicted Aligned Error
PAS domain	Per-ARNT-Sim Domain
PDB	Protein Data Bank
RCSB	Research Collaboratory for Structural Bioinformatics
pLDDT	Predicted Local Distance Difference Test
RMSD	Root Mean Square Deviation
SD rat	Sprague-Dawley rat
VKOR	Vitamin K Epoxide Reductase
Database/Method	解説
AlphaFold Database	AlphaFold2 で予測されたタンパク質構造データベース
Autodock	自動分子ドッキングのためのソフトウェア
ColabFold	Google Colaboratory 上での AlphaFold2 タンパク質構造予測
Fpocket	タンパク質の高速ポケット検出ソフトウェア
Google colaboratory	クラウドベースのプログラミングプラットフォーム
RCSB PDB	生体分子の立体構造データベース
ROSETTA	タンパク質のモデリングと設計のためのソフトウェア
SWISS Model	タンパク質構造ホモロジーモデリングのウェブツール
Uniprot	タンパク質配列と機能情報の包括的なデータベース
Vina	Autodock ベースの分子ドッキングソフトウェア

に続き核磁気共鳴法(NMR)やクライオ電子顕微鏡による構造解析も行われるようになり, これら実験的観察により得られたタンパク質立体構造は Protein Data Bank (PDB) に収載されデータベース化されている (Berman *et al.*, 2000)。その一方で, これら実験的なタンパク質結晶構造解析は大型の実験機器を必要とするうえに, 一つ一つの構造解析に際し高度な実験手法と時間を必要とする。このため, RCSB (Research Collaboratory for Structural Bioinformatics)

PDB に収載されている実験的観察立体構造は指数的に増加しているものの 2024 年現在で約 21 万 8 千件と, Uniprot に収載されている総アミノ酸配列 2 億 4 千万件と比較してわずか 1/1000 程度でしかない (PniProt, n.d.; RCSB PDB, 2024)。

3. タンパク立体構造の予測と AlphaFold2

タンパク質立体構造が生物学的にどのように決定されるかに対する仮説として, 生理的環境ではタンパク質立体構造はアミノ酸配列のみに基づいて決定されるというアンフィンセンのドグマ(熱力学的仮説)が広く支持されている (Anfinsen, 1973)。この仮説に従うと, 理論上アミノ酸配列のみを入力としてそれが取りうる立体構造を予測できるということになる。アンフィンセンのドグマの提唱以降, タンパク質構造の予測技術は進化し続け, SWISS-MODEL に代表されるホモロジーモデリングがその中心的手法となった (Waterhouse *et al.*, 2018)。このアプローチでは, 既に解明された類縁タンパク質構造を鋳型にして新しいタンパク質の構造を推定する。これはアミノ酸配列の類似性と構造の類似性が相関するという原理に基づいている。次いで, フラグメントアッセムブリの技術が開発された。この方法では, 小さなタンパク質片を再利用して全く新しいタンパク質構造を構築する。ROSETTA が代表的な手法であり, 既存のタンパク質データベースから抽出された構造断片を組み合わせて, 新しいタンパク質の構造を予測する (Simons *et al.*, 1997)。しかしながら, これらを用いて立体構造予測を実施するには相溶性の高い既知構造に依るところが大きく, 新奇構造に対する予測精度は高くなかった。タンパク質構造の予測技術の進歩を評価するために 1994 年から 2 年ごとに Critical Assessment of protein Structure Prediction (CASP) と呼ばれる国際コンペティションが開催されている。2016 年 CASP12 までは, GDT-TS スコア(実験的決定構造との類似性に基づく評価法, 90 以上が実験的観察と同等とされる)が最高点をえたアルゴリズムでも約 40 点であり, 予測精度は実験的観察に及ばなかった (Moult *et al.*, 2018)。AlphaFold2 は, 2020 年開催の CASP14 でその圧倒的な性能を示し, タンパク質構造予測の分野における新たな基準を確立した。AlphaFold2 は CASP の GDT-TS スコア(実験的決定構造との類似性に基づく評価法, 90 以上が実験的観察と同等とされる)で中央値が 92.4 という驚異的な精度を示した (Jumper *et al.*, 2021b)。

4. AlphaFold2 の予測原理概説

AlphaFold2 による超高精度の構造予測の原理をごく簡単に説明する (図 1)。まず, AlphaFold2 の予測

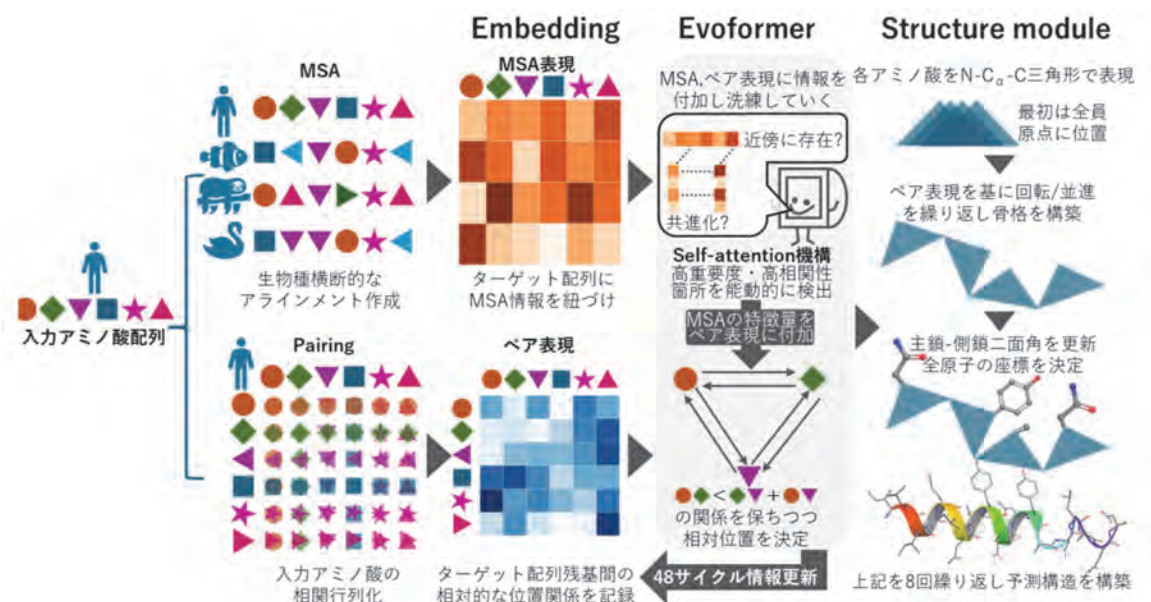


図1 AlphaFold2 の予測原理。

AlphaFold2 原著論文 (Jumper *et al.*, 2021a) の図1 と図3 を基に簡略化, 日本語注釈を追加した。

プロセスにおいて重要な初期段階となるのが多重配列アラインメント (MSA) である。この段階では、ターゲットタンパク質のアミノ酸配列を基に検索された多数の類似配列がアラインメントされ、配列間の進化的関連性が捉えられる。MSA から得られる情報は、タンパク質のどの部分が相互作用を持ちうるか、または独立して動くかの手がかりを提供し、タンパク質の構造予測に不可欠なデータとなる。次に、ターゲットタンパク質の配列を相関行列化したペア表現が導入される。ここではターゲットタンパク質内の各アミノ酸残基間の相互作用をモデル化し、全ての残基ペア間の相互作用を表現する。ペア表現は各アミノ酸ペア間の距離と角度を予測する際に重要であり、タンパク質の立体構造を形成するための空間的な制約を提供する。これら MSA 表現、ペア表現をもとにタンパク質の立体構造が組み立てられる。この予測プロセスを統括するのが Evoformer という MSA とペア表現の情報を統合して処理するディープラーニングモジュールである。Evoformer は Transformer ベースのアーキテクチャを採用しており、タンパク質の構造を決定するための詳細な特徴を抽出する。Transformer とは近年の自然言語処理分野に大きな影響を与えた深層学習手法で、入力データに含まれる情報の関連性を効率的に捉えるために設計されており、chatGPT 等の大規模言語モデルにも不可欠な手法である (Vaswani *et al.*, 2017)。Transformer の主要機構が self-attention である。セルフアテンションは与えられた入力の各単語 (トークン) と他の単語との関係性を評価、その重要度を算出して適切に加重しながら情報を取得する。Evoformer でもこの self-attention 機構により、MSA 表現から得られる共進化情報やアミノ酸ペア

の長距離の依存関係と局所的な特徴の両方を効果的に捉え、高度な予測が可能になる。Evoformer による MSA、ペア表現の特徴量抽出作業は 48 サイクル実施され、特徴量を洗練・高精度化していく。この特徴量を基に Structure module でタンパク質の立体構造の予測が行われる。本モジュールではタンパク質主鎖中のアミノ酸を N-Ca-C からなる三角形として表現し、ペア表現で得られた位置関係に基づきそれぞれ回転・並進させ骨格を形成する。その後主鎖と側鎖の二面角を更新しタンパク質を構成する全原子の座標を決定する。「拡がる環境ゲノム研究」特集号の記事として特筆すべき点として、AlphaFold2 の予測ではヒトのアミノ酸配列のみならずあらゆる生物種のアミノ酸配列から作成した MSA が予測に重要だという点だろう。この比較生物学の知識と Transformer に代表される最先端の深層学習手法の融合により超高精度な予測が可能になった。

5. AlphaFold2 の利用法と予測精度の評価法

AlphaFold2 が生物学全体を巻き込んだ一大ムーブメントになった理由にはその高精度さだけではなく利用の簡便さも大きく貢献している。AlphaFold2 をローカル PC で使用するには 2 TB 以上の空き容量を持つ Nvidia GPU 搭載 Linux 機が必要であり、インフォマティクスに明るくないバイオ系研究者にとって使用のハードルは低くない。これに対し AlphaFold2 の爆発的な利用促進の立役者となったのが, ColabFold と AlphaFold Protein Structure Database である (Mirdita *et al.*, 2022)。ColabFold は Google が提供するウェブブラウザ上での python スクリプト実行サービスである Google Collaboratory 上で動作

する AlphaFold2 であり、Google アカウントを有してれば無料で実行することが可能である(図 2)。使用法も基本的には予測したタンパク質のアミノ酸配列を入力するだけのワンクリックであり簡便なため爆発的に利用が進んだ。また、予測精度自体は本家 AlphaFold2 とほぼ同等であるのに対し、予測速度は 50 倍程度高速であり、ローカル PC へのインストールの必要がなく、ストレージ容量も 1.4 GB と少ないのも利点である。また、DeepMind 社が Uniprot に掲載されている全アミノ酸配列 2 億種超に対し網羅的に予測を実施し、AlphaFold Protein Structure Database として公開したことから、単量体構造については予測を実施せずとも入手可能になった(Varadi *et al.*, 2022)。

AlphaFold2 予測構造を研究に用いる際に、予測構造の信頼性を事前に確認するのは重要である。予測精度の指標として pLDDT(predicted Local Distance Difference Test)と PAE(Predicted Aligned Error)という 2つのスコアがある。pLDDTは、各アミノ酸の位置がどれだけ正確に予測されているかを表す指標である(Mariani *et al.*, 2013)。値は 0 から 100 までのスコアで表されスコアが高いほど予測の信頼性が高いことを意味し、タンパク質の各アミノ酸に対しその三次元座標の予測精度が計算され、それらの平均値が pLDDT スコアとなる。AlphaFold Protein Structure Database ではタンパク質構造の各部分の信頼度を色分けして可視化している(図 3)。PAE は、タンパク質のアミノ酸間の距

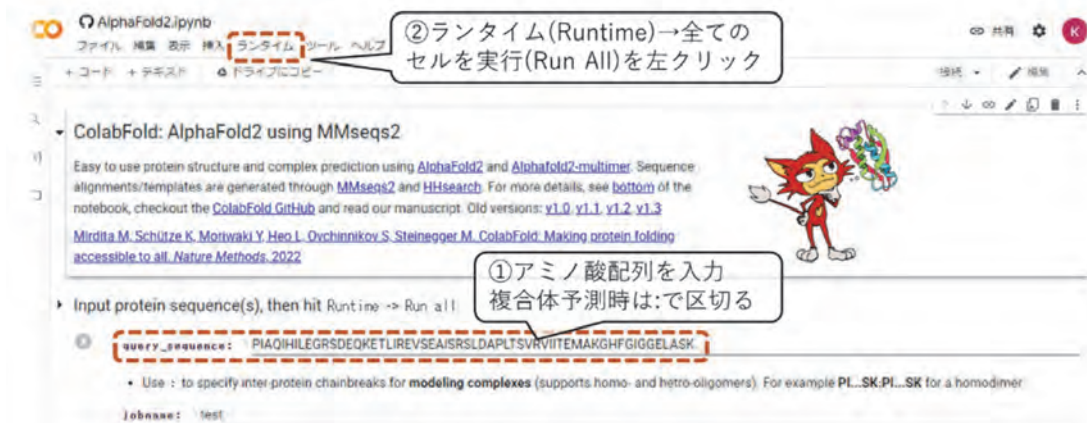


図 2 ウェブブラウザ上での ColabFold の使用法.

①の query_sequence に予測したいアミノ酸配列を入力し、その後上部タブのランタイムから「全てのセルを実行」を左クリックするだけで予測～結果のダウンロードが進行する。複合体構造を予測する際は入力アミノ酸配列を:で区切り入力する。1つの Google アカウントにつき同時に実行可能なジョブは1つのため、複数タブを開いて多数の予測を実施することはできない。

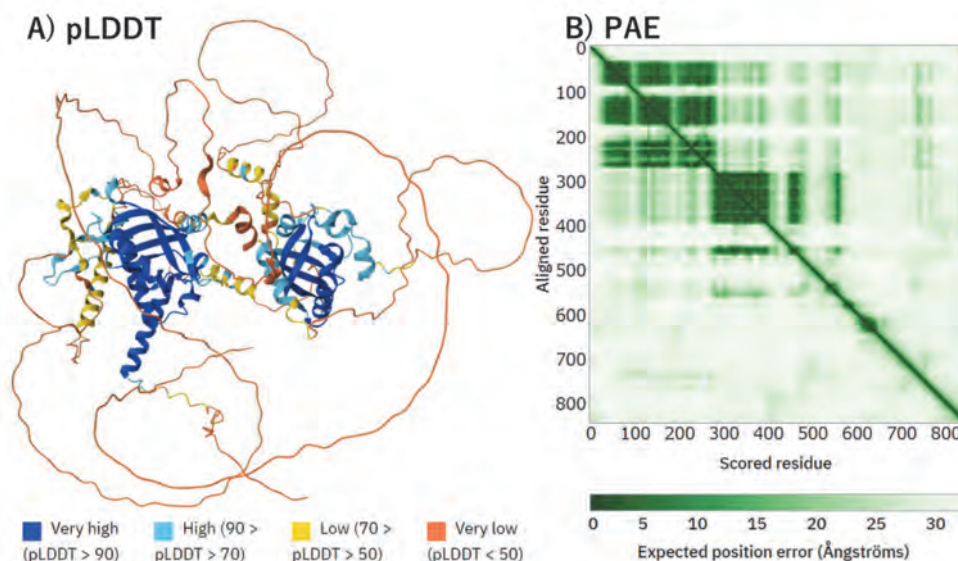


図 3 AlphaFold2 予測構造の評価方法.

シロイロカ(ペルーガ)の芳香炭酸化水素受容体(AhR)を一例に示した(<https://alphafold.ebi.ac.uk/entry/Q95LD9>)。A) Predicted local distance difference test(pLDDT)は1アミノ酸残基ごとの予測信頼性を示している。中心部のリガンド結合部位の予測精度は比較的高く青色～黄色で表示されているが、前後の天然変性領域の予測精度は低くオレンジ色で表記されている。B) Predicted Aligned Error (PAE)は予測構造のアミノ酸配列をX、Y軸にプロットした相関行列により各アミノ酸ペアの位置関係の信頼性を示している。緑色が濃いほど相対位置が確かであることを示す。

離の予測誤差を表す指標である (Zhang & Skolnick, 2004)。具体的には、タンパク質の二次元の距離マップを元に、各アミノ酸対の予測された距離と実際の距離との間の誤差が計算される。この値が小さいほど、タンパク質の予測構造が実際の構造に近いことを示す。PAE は、特にタンパク質のドメイン間の位置関係の信頼性を把握するために利用される。これら pLDDT と PAE は AlphaFold2 本体と同様に大規模データセットでの学習を行っており、新規構造についてもスコアリングを行うことが可能である。図 3 では主要な環境汚染物質受容体である Aryl hydrocarbon receptor の予測構造を取り上げた。芳香族炭化水素受容体 (AhR) は、主に外来性及び内因性の多環芳香族化合物に対する細胞の応答を調節する役割を果たす核内受容体である (Larigot *et al.*, 2018)。AhR にこれらの化学物質が結合すると、核内へ移行し、DNA 上の特定の応答配列に結合し、解毒酵素 (主に CYP1A) の発現増加や免疫応答、細胞増殖、分化代謝経路の変化などの遺伝子発現の調節を行う。図 3 を見ると AhR では高度に折り畳まれた領域と Unfolded な領域が二分化されている。Fold されている領域はバシック・ヘリックス・ループ・ヘリックス (bHLH) ドメイン (DNA 結合に関与する一般的に転写因子に見られる特徴的な構造) や PAS ドメイン、リガンド結合ドメイン (LBD、低分子化合物の結合サイト) 等 AhR の機能に重要な部分である。Unfolded な部分の多くは天然変性領域と呼ばれる生理的に不安定であり特定の安定構造を取らない部分や C 末端、N 末端の領域である。このように AlphaFold2 も全ての配列に対し高精度な予測ができるわけではなく、天然変性領域などの一意な構造を取らない構造の予測は不得手としている。実際にヒト全タンパク質配列に対する予測精度を評価した研究では、pLDDT 値 50% 以上の「折り畳まれている」予測が出来たのは全配列 (10,537,122 残基) 中 75% であった (Tunyasuvunakool *et al.*, 2021)。

6. 環境毒性学研究への応用性：

分子ドッキングの標的としての利用

環境毒性学研究における AlphaFold2 の最も簡単かつ有効な利用法が分子ドッキングシミュレーションの標的タンパク質としての利用である (図 4)。環境汚染物質の多くは低分子化合物であり、毒性学研究ではこれら低分子化合物 (リガンド) とタンパク質の結合性を評価することで毒性ターゲットの同定や結合力の評価を実施する。一方で、AlphaFold2 自体は低分子化合物の結合像やその影響は予測しない。AlphaFold2 構造を用いての毒性評価に際して、分子ドッキングシミュレーションとの併用が有効な手法となりえる。分子ドッキングは、リガンドとタンパク質 (レセプター) の立体構造を用い、最も安定して

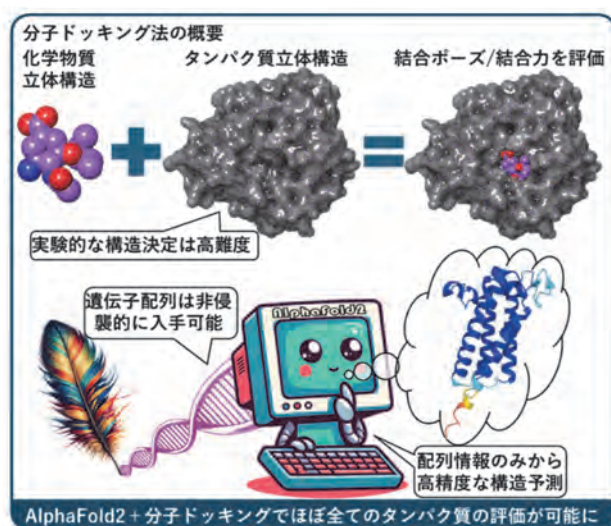


図 4 分 (窪み) に対し最も安定するリガンドのポーズを計算する。

AlphaFold2 により高品質なタンパク質立体構造の取得が容易になったため、分子ドッキング法を用いることで幅広い種類の化学物質-タンパク質結合予測が可能になった。

結合する際のリガンド構造と結合サイト、その結合親和性を予測する計算アプローチである (Fan *et al.*, 2019)。このプロセスは、リガンドとレセプターの自然な相互作用を以下のプロセスを経て模倣する。

1. リガンドの最適化：リガンド分子の 3D 構造は、結合長、角度、及び原子電荷を割り当てることでドッキング計算に適した形に最適化される。
2. レセプターの準備：レセプターの 3D 構造も、同様の手法で下処理される。
3. グリッド生成：レセプター内における探索空間を定義するために、受容体の結合サイトの周囲にグリッドを生成する。結合サイトの中心座標の指定は知識ベースで既知の結合サイトを指定するほかに、Fpocket や GHECOM 等のタンパク質内結合ポケット検出手法で探索的に検出することができる (Kawabata, 2010; Le Guilloux *et al.*, 2009)。
4. リガンドポーズ探索：リガンド分子はグリッド内に配置され、系のエネルギーを最小化することで最適な結合ポーズを探索する。この際計算量と精度のトレードオフからレセプター原子は剛体として扱いリガンドは可動性があるものとして扱うリジッドドッキング法が一般的に用いられる。
5. スコアリング関数評価：上述のリガンド-受容体複合体の結合親和性の計算は、水素結合、静電気力、及びファンデルワールス力などの様々な相互作用を考慮に入れたスコアリング関数を使用して評価される。スコアリング関数はドッキング計算手法ごとに異なるが学術利用で無償使用できる Autodock や Autodock vina が汎用される (Eberhardt *et al.*, 2021)。

分子ドッキングの結果からは以下のような知見が得られる。

・結合親和性の予測：分子ドッキングはリガンドの受容体への結合親和性を予測，潜在的なリード化合物の同定と結合力によるランキングが可能になる。

・重要な相互作用の特定：リガンドと受容体の間の重要な相互作用，特に受容体内の結合サイトとその結合の種類を同定することで新規分子の設計に情報を提供する。

・生物学的応答のシミュレーション：分子ドッキングである化合物が結合しうる複数のタンパク質への結合性を評価することで標的タンパク質の同定及び結合により生じる生物学的応答を予測することができる。

ドッキング計算は比較的計算負荷の小さいシミュレーション手法であり，家庭用のノートPCでも1リガンド1レセプター1セットの計算で数十秒～数分，GPU計算を併用することで1セット数秒で計算を完了できる。このため数万種類の創薬シード化合物の薬効評価に対するバーチャルスクリーニングとして適用されている。一方で，化学物質の構造は容易に用意できるのに対し，先述のようにタンパク質立体構造実験的観察の困難さや構造予測の精度への信頼性によりタンパク質立体構造の入手性が課題であった。タンパク質構造の実験的観察はヒトやウイルスといった医学的に重要な生物種に対し重点的に実施されており，環境科学が対象とする野生動物のタンパク質を対象とした研究はごく限られていた。これに対し，AlphaFold2があればアミノ酸が入手できる動物種のほぼ全てのタンパク質に対し化学物質の結合性を評価できるようになったと言えることができる。

7. 環境毒性学研究の応用事例：天然記念物オガサワラオオコウモリの殺鼠剤感受性評価

AlphaFold2を用いた希少種における環境毒性学研究の応用事例として，オガサワラオオコウモリの殺鼠剤感受性評価を一例に挙げる(Takeda *et al.*, 2022)。小笠原諸島は固有の生態系からユネスコ世界自然遺産に登録されているが，外来種の移入による生態系の攪乱が問題となっている。特に繁殖力の強いイエネズミ(ドブネズミ・クマネズミ)による固有植生・海鳥の卵の食害は深刻な問題である(Yabe *et al.*, 2009)。そこで殺鼠剤(抗血液凝固作用を有するダイファシノン製剤バイト)の空中散布が環境省主導で実施されており一定の成果を上げている。一方で，殺鼠剤＝殺哺乳類剤でもあり，ネズミ以外の哺乳類に対しての中毒事例も数多く報告されている(Morita *et al.*, 2018)。そこで，筆者らは小笠原諸島唯一の固有哺乳類であるオガサワラオオコウモリのダイファシノン感受性評価を実施した。しかしながら，オガサワラオオコウモリの推定生息数は100頭～200頭であり絶滅危惧種に登録されており，これらを用い

ての致死的な毒性試験は不可能である。このような希少な野生動物に対しての感受性評価は以下の2つのアプローチから実施される。

(1)実験可能な近縁種を用いた動物実験結果の外挿

(2)遺伝子情報を用いた予測

本研究では(1)として近縁種で日本で商用入手可能なエジプシャンルーセットオオコウモリを用いた *in vivo* PK/PD 解析及び(2)としてオガサワラオオコウモリの遺伝子情報から構築したダイファシノン標的分子ビタミンKエポキシド還元酵素(VKOR)立体構造を用いた分子シミュレーションを実施した。(1)は本特集の本題ではないためごく簡単に結果を述べると，エジプシャンルーセットオオコウモリとSDラットの殺鼠剤投与後の血液凝固時間を比較したところ，エジプシャンルーセットオオコウモリはワルファリンに対してはSDラットより血液凝固時間が遅延し辛く感受性が低いことが示唆されるのに対し，ダイファシノンに対してはSDラットよりも早期に血液凝固時間が遅延し高い感受性を示すという同じ標的分子VKORに結合する2薬剤で感受性の動物種差が異なるという興味深い結果となった。

上述の傾向はVKORへの結合性によるものか，また，オガサワラオオコウモリも同様の傾向を有するか検証するためにAlphaFold2予測構造を用いた分子シミュレーションを実施した。事故死したオガサワラオオコウモリの肝臓から抽出したRNAを用いた次世代シーケンズ解析からオガサワラオオコウモリのVKOR配列を同定し，AlphaFold2で立体構造を構築，ワルファリン及びダイファシノンの結合性をGlide((Schrödinger, New York, NY)を用いた分子ドッキング法で比較した(図5)。その結果，ワルファリンの結合性はオガサワラオオコウモリ・エジプシャンルーセットオオコウモリのVKORに対し比較的低く(結合親和性値：7.15～7.25 -kcal/mol)，ダイファシノンの方が結合しやすい(8.24～8.59 -kcal/mol)という知見が得られた。更なる検証にあたり，分子ドッキング法より生理的環境に近いシミュレーションである分子動力学シミュレーションをこれらの複合体を用いて実施した。分子ドッキング法では本来可塑性のあるタンパク質を剛体として扱うが，分子動力学シミュレーションではタンパク質－リガンド複合体とその周囲に配置した脂質二重膜・水分子全原子からなる全原子の運動方程式を複数回計算する。これにより経時的かつタンパク質の可動性も含めた動的なシミュレーションでより生理的環境に近い計算結果を得ることができる(De Vivo *et al.*, 2016)。分子動力学シミュレーションで得られる結果の1つとしてRMSD値(Root Mean Square Deviation, 平均二乗偏差)を示した。これはシミュレーション中に，開始時のリガンドの構造からどれだけ構造が変化したかを示した偏差である。オガサワラオオコウモリ(Bat)のVKOR中に配位したワ

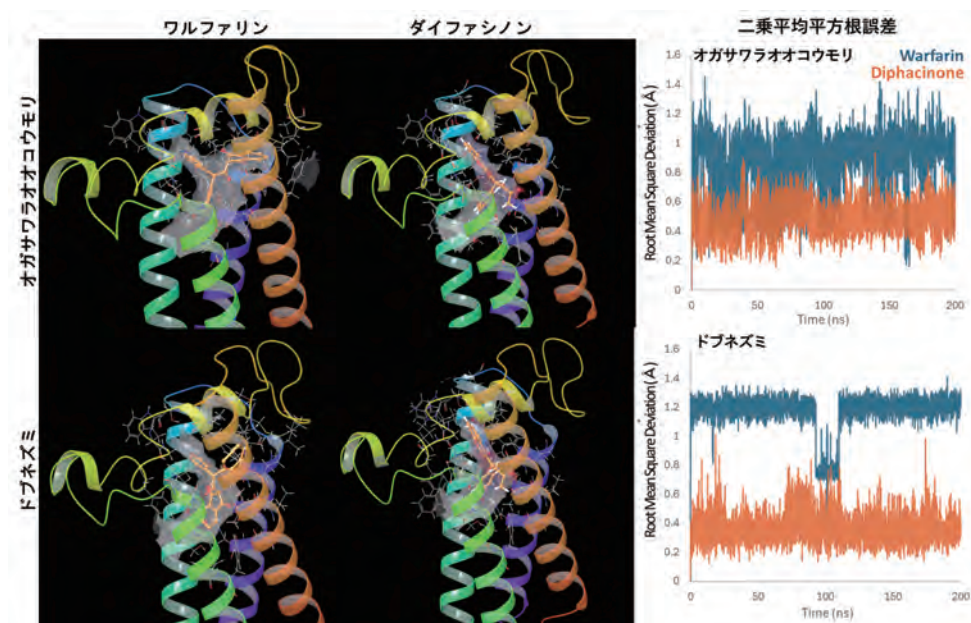


図5 AlphaFold2 予測構造を用いた VKORC1 と殺鼠剤 (ワルファリン・ダイファシノン) の分子ドッキング・分子動力学シミュレーション結果。先行研究 (Takeda *et al.*, 2022) で得た予測構造を基に再解析し作図。

ルファリンは 200 ns のシミュレーションを通して RMSD 値の振れ幅が大きく、結合が不安定であることが示唆される。一方で、ラット VKOR 中のワルファリンは 100 ns 前後で一時的に変動するも RMSD 値がほぼ一定であり高度に安定している、つまり、VKOR に強固に結合することが示唆される結果となった。一方で、ダイファシノンに関してはオガサワラオオコウモリ VKOR 中での RMSD 値 (平均 0.31) よりもラット VKOR 中 RMSD 値 (平均 0.36) が小さい傾向にあり、ドッキング計算を支持する結果となった。以上から、同じ標的タンパク質 (VKOR) に結合し同じ薬効を示す化合物 (ワルファリンとダイファシノン) でも薬効に種差があることが *in vivo*, *in silico* 解析双方から明らかになった。

8. 結語

以上のように、AlphaFold2 による高精度タンパク質立体構造予測と分子シミュレーションの組み合わせで、動物実験を行うことができない希少種においても化学物質の影響評価が可能なが示された。一方で、AlphaFold2 を分子ドッキング法の鋳型とする研究は複数実施されているが、同じタンパク質でも実験的観察で得られた構造と比較すると結合力が低く算出される傾向がある (Scardino *et al.*, 2023)。実験的観察構造ではリガンドが結合した状態で観察されることが多いのに対し、AlphaFold2 では側鎖構造が結合ポケットとして最適化されていない場合が多いことが原因であると考えられる。これに際し、上述の分子動力学シミュレーションはタンパク質構造の最適化による AlphaFold2 を用いた分子ドッ

キング精度向上に対し有効な手法である。また、AlphaFold2 はタンパク質-タンパク質複合体の構造予測も可能なことが知られている (Evans *et al.*, 2022)。ColabFold も複合体予測に対応しており、図 2 で示した Colaboratory 上では入力アミノ酸配列を「:」で区切ることで複合体の予測ができる。これを利用すれば、低分子化合物のみならず、ペプチド毒や自然毒といった高分子毒性物質の結合性評価も可能になると予想される。

このように、AlphaFold2 の登場以降、その応用例や AlphaFold2 超えを目論むタンパク質構造予測法が日進月歩的に報告されている…と記載した本稿の査読期間中 (2024 年 5 月 8 日)、遂に AlphaFold3 が報告された (Abramson *et al.*, 2024)。AlphaFold3 は構造予測に画像生成 AI で汎用される拡散モデルを採用することで、タンパク質構造のみならず低分子化合物・核酸・金属イオンも含む“全原子”の予測を実現した。今のところそのソースコードは公開されておらず、ウェブブラウザ上でのトライアルのみ可能となっている。このウェブブラウザでのトライアルではアミノ酸配列や核酸配列は自由に入力できる一方で、低分子化合物については ATP やヘム等の内因性化合物 18 種の中から選択するしかなく、環境汚染物質との結合性の計算には用いることができない。また、AlphaFold3 で生成したタンパク質/核酸構造についても、分子ドッキングやその他計算手法へ利用は利用規約で禁止されているため、残念ながら本稿で述べた毒性評価手法に際しては現状 AlphaFold2 の構造を用いるほかない。しかしながら、タンパク質単量体構造の予測精度については AlphaFold3 で性能の改善は見られたものの 2 と 3 の

間に劇的な差はないことがAlphaFold3 原著論文でも報告されており、AlphaFold2 予測構造と分子ドッキングの組み合わせによる評価手法は今もなお有効な選択肢である。以上のように、構造生物学は今、最も研究の進展が目覚ましい学術領域であり、これらの手法を取り入れることでポスト AlphaFold2 のゲノム研究はあらゆる生物種において更に加速していくと期待される。

謝 辞

本研究は、環境省・(独)環境再生保全機構の環境研究総合推進費(JPMEERF20214R02, JPMEERF20184R02)及びJSPS 科研費(23K14091, 21K14988)の助成を受け実施した。

引用文献

- Abramson, J., Adler, J., Dunger, J., ... and Jumper, J. M. (2024) Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature*, 630(8016), 493–500. <https://doi.org/10.1038/s41586-024-07487-w>
- Anfinsen, C. B. (1973) Principles that govern the folding of protein chains. *Science*, 181(4096), 223–230. <https://doi.org/10.1126/SCIENCE.181.4096.223>/ASSET/1A648F0F-E322-4108-98DD-B0857103738D/ASSETS/SCIENCE.181.4096.223.FP.PNG
- Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N. and Bourne, P. E. (2000) The Protein Data Bank. *Nucleic Acids Research*, 28(1), 235–242. <https://doi.org/10.1093/NAR/28.1.235>
- Bernal, J. D. and Crowfoot, D. (1934) X-Ray photographs of crystalline pepsin. *Nature*, 133(3369), 794–795. <https://doi.org/10.1038/133794b0>
- De Vivo, M., Masetti, M., Bottegoni, G. and Cavalli, A. (2016) Role of molecular dynamics and related methods in drug discovery. *Journal of Medicinal Chemistry*, 59(9), 4035–4061. <https://doi.org/10.1021/ACS.JMEDCHEM.5B01684>
- Eberhardt, J., Santos-Martins, D., Tillack, A. F. and Forli, S. (2021) AutoDock Vina 1.2.0: new docking methods, expanded force field, and python bindings. *Journal of Chemical Information and Modeling*, 61(8), 3891–3898. <https://doi.org/10.1021/ACS.JCIM.1C00203>
- Evans, R., O'Neill, M., Pritzel, A., ... and Hassabis, D. (2022) Protein complex prediction with AlphaFold-Multimer. *BioRxiv*, 2021.10.04.463034. <https://doi.org/10.1101/2021.10.04.463034>
- Fan, J., Fu, A. and Zhang, L. (2019) Progress in molecular docking. *Quantitative Biology*, 7(2), 83–89. <https://doi.org/10.1007/S40484-019-0172-Y>
- Jumper, J., Evans, R., Pritzel, A., ... and Hassabis, D. (2021a) Highly accurate protein structure prediction with AlphaFold. *Nature*, 596(7873), 583–589. <https://doi.org/10.1038/s41586-021-03819-2>
- Jumper, J., Evans, R., Pritzel, A., ... and Hassabis, D. (2021b) Applying and improving AlphaFold at CASP14. *Proteins: Structure, Function, and Bioinformatics*, 89(12), 1711–1721. <https://doi.org/10.1002/PROT.26257>
- Kawabata, T. (2010) Detection of multiscale pockets on protein surfaces using mathematical morphology. *Proteins: Structure, Function, and Bioinformatics*, 78(5), 1195–1211. <https://doi.org/10.1002/PROT.22639>
- Kendrew, J. C., Bodo, G., Dintzis, H. M., Parrish, R. G., Wyckoff, H. and Phillips, D. C. (1958) A three-dimensional model of the myoglobin molecule obtained by x-ray analysis. *Nature*, 181(4610), 662–666. <https://doi.org/10.1038/181662a0>
- Larigot, L., Juricek, L., Dairou, J. and Coumoul, X. (2018) AhR signaling pathways and regulatory functions. *Biochimie Open*, 7, 1–9. <https://doi.org/10.1016/J.BIOPEN.2018.05.001>
- Le Guilloux, V., Schmidtke, P. and Tuffery, P. (2009) Fpocket: An open source platform for ligand pocket detection. *BMC Bioinformatics*, 10(1), 1–11. <https://doi.org/10.1186/1471-2105-10-168>/TABLES/1
- Mariani, V., Biasini, M., Barbato, A. and Schwede, T. (2013) IDDT: a local superposition-free score for comparing protein structures and models using distance difference tests. *Bioinformatics*, 29(21), 2722. <https://doi.org/10.1093/BIOINFORMATICS/BTT473>
- Mirdita, M., Schütze, K., Moriwaki, Y., Heo, L., Ovchinnikov, S. and Steinegger, M. (2022) ColabFold: making protein folding accessible to all. *Nature Methods*, 19(6), 679–682. <https://doi.org/10.1038/s41592-022-01488-1>
- Morita, A., Nakayama, S. M. M., Mizukawa, H., Ishizuka, M. and Ikenaka, Y. (2018) A review: poisoning by anticoagulant rodenticides in non-target animals globally. *Journal of Veterinary Medical Science*, 81(2), 298–313. <https://doi.org/10.1292/jvms.17-0717>
- Moult, J., Fidelis, K., Kryzhtafovych, A., Schwede, T. and Tramontano, A. (2018) Critical assessment of methods of protein structure prediction (CASP) -Round XII. *Proteins*, 86 Suppl 1(Suppl 1), 7–15. <https://doi.org/10.1002/PROT.25415>
- OpenAI. (2024) *Video Generation Models as World Simulators*. <https://openai.com/index/video-gene>

- ration-models-as-world-simulators(2024 年 5 月 7 日 確認)
- Perutz, M. F., Wilkinson, A. J., Paoli, M. and Dodson, G. G. (1998) The stereochemical mechanism of the cooperative effects in hemoglobin revisited. *Annual Review of Biophysics and Biomolecular Structure*, 27, 1–34. <https://doi.org/10.1146/ANNUREV.BIOPHYS.27.1.1>
- PniProt (n.d.) *UniProtKB | Statistics*. 2024. <https://www.uniprot.org/uniprotkb/statistics>(2024 年 5 月 7 日 確認)
- Podell, D., English, Z., Lacey, K., Blattmann, A., Dockhorn, T., Müller, J., Penna, J. and Rombach, R. (2023) *SDXL: Improving Latent Diffusion Models for High-Resolution Image Synthesis*. <https://arxiv.org/abs/2307.01952v1>(2024 年 5 月 7 日 確認)
- RSCB PDB (2024) *PDB Statistics: Overall Growth of Released Structures Per Year*. <https://www.rcsb.org/stats/growth/growth-released-structures>(2024 年 5 月 7 日 確認)
- Scardino, V., Di Filippo, J. I. and Cavasotto, C. N. (2023) How good are AlphaFold models for docking-based virtual screening? *IScience*, 26(1), 105920. <https://doi.org/10.1016/J.ISCI.2022.105920>
- Simons, K. T., Kooperberg, C., Huang, E. and Baker, D. (1997) Assembly of protein tertiary structures from fragments with similar local sequences using simulated annealing and Bayesian scoring functions. *Journal of Molecular Biology*, 268(1), 209–225. <https://doi.org/10.1006/JMBI.1997.0959>
- Takeda, K., Manago, K., Morita, A., ... and Nakayama, S. M. M. (2022) Toxicokinetic analysis of the anticoagulant rodenticides warfarin & diphacinone in Egyptian fruit bats (*Rousettus aegyptiacus*) as a comparative sensitivity assessment for Bonin fruit bats (*Pteropus pselaphon*). *Ecotoxicology and Environmental Safety*, 243, 113971. <https://doi.org/10.1016/j.ecoenv.2022.113971>
- Tunyasuvunakool, K., Adler, J., Wu, Z., ... and Hassabis, D. (2021) Highly accurate protein structure prediction for the human proteome. *Nature*, 596(7873), 590–596. <https://doi.org/10.1038/s41586-021-03828-1>
- Varadi, M., Anyango, S., Deshpande, M., ... and Velankar, S. (2022) AlphaFold protein structure database: massively expanding the structural coverage of protein-sequence space with high-accuracy models. *Nucleic Acids Research*, 50(D1), D439–D444. <https://doi.org/10.1093/NAR/GKAB1061>
- Vaswani, A., Brain, G., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł. and Polosukhin, I. (2017) Attention is all you need. *Advances in Neural Information Processing Systems*, 30.
- Waterhouse, A., Bertoni, M., Bienert, S., ... and Schwede, T. (2018) SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Research*, 46(W1), W296–W303. <https://doi.org/10.1093/NAR/GKY427>
- Yabe, T., Hashimoto, T., Takiguchi, M., Aoki, M. and Kawakami, K. (2009) Seabirds in the stomach contents of black rats *Rattus rattus* on Higashijima, the Ogasawara (Bonin) Islands, Japan. *Marine Ornithology*, 37(3), 293–295. <http://www.marineornithology.org/content/get.cgi?rn=863>(2024 年 5 月 7 日 確認)
- Zhang, Y. and Skolnick, J. (2004) Scoring function for automated assessment of protein structure template quality. *Proteins: Structure, Function, and Bioinformatics*, 57(4), 702–710. <https://doi.org/10.1002/PROT.20264>



武田 一貴 / Kazuki TAKEDA

北里大学獣医学部獣医学科毒性学研究室講師。2020 年 3 月北海道大学大学院獣医学研究科修了(博士(獣医学))、東京工業大学(現東京科学大学)情報理工学院特別研究員(2020 年 4 月–9 月日本学術振興会特別研究員 SPD として)を経て 2020 年 9 月現所属助教、2023 年 4 月より現職。専門は獣医毒性学及び計算科学。構造科学、分子シミュレーション手法を用いた毒性標的発見及び野生動物への非侵襲的毒性評価に関する研究を実施している。