

AlphaGo seals 4-1 victory over Go grandmaster Lee Sedol. ディープマインド社が開発したアルファ碁は 2015 年秋に欧州の王者ファン・ウェイに対し互先なしで全勝（5 勝）、返す刀で当時の最強棋士・李世との五番勝負を 4 勝 1 敗で切って落とした。冒頭の一文は英ガーディアン紙の見出しである。アルファ碁が世間を席巻したこの年、ディープマインド社は AlphaFold（AF）の開発を本格化した。

### 1. AF 以前のタンパク質構造予測（計算的手法）

AF 登場以前から、タンパク質構造を計算的に予測する手法は数多く提案され、応用されてきた [1]。これらの手法の多くは、物理法則や統計的推論に基づいており、AF 開発の礎ともなった。表 1 に代表的な予測アプローチを示す。

#### (1) 全原子シミュレーション

分子動力学法（Molecular Dynamics; MD）は、1990 年代以降に飛躍的に発展し、タンパク質の折りたたみ過程の動的解析にも広く用いられている。その原型は 1950 年代末、理論物理学の分野で提案されたものであり、原子や分子の運動を時間軸に沿ってモデル化する。各原子の時間発展は、経験則または量子力学に基づいて定義されたポテンシャル関数（力場）に従い、運動方程式を用いて計算される。タンパク質の折りたたみに加え、変性、分子間相互作用、結合・解離などの多様な動的イベントを、理論上、原子単位で追跡することができる。

一方で、正確な予測を実現するには、高精度な力場の使用が不可欠であり、その妥当性の検証が難しいことや、膨大な計算資源を必要とすることが、現在なお課題とされている。

MD を活用したタンパク質構造解析に関する論文数は、1990 年頃から右肩上がりの傾向を示しており（図 1）、その背景には関連技術の革新も影響しているとみられる。2000 年代後半には、画像処理装置（Graphical Processing Unit; GPU）を科学計算に活用するためのライブラリ（AMBER, GROMACS など）が登場し、並列計算能力が飛

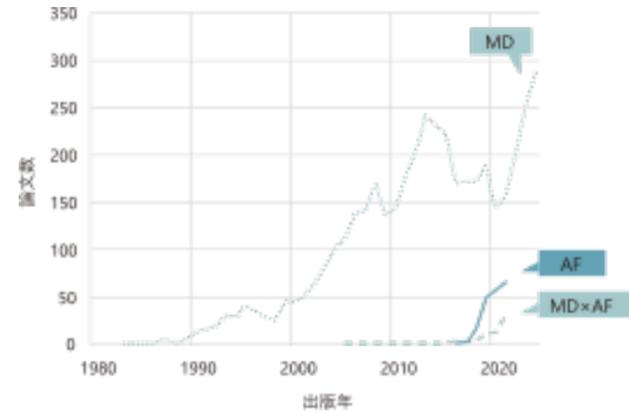


図 1 タンパク質構造解析における MD 論文数の年次推定 Scopus 2024.10  
 検索式：  
 • TITLE-ABS-KEY ("protein structure") AND TITLE ("molecular dynamics" OR MD)  
 • TITLE-ABS-KEY ("protein structure") AND TITLE (alphafold)  
 • TITLE-ABS-KEY ("protein structure") AND TITLE ("molecular dynamics" OR md) AND TITLE-ABS-KEY (alphafold OR "machine learning" OR "deep learning")

躍的に向上した。また、2008 年には MD 計算用に設計されたスーパーコンピュータ Anton が開発された。こうしたハードウェアおよびソフトウェア技術の進歩により、2010 年代の後半には長時間スケール（ミリ秒オーダー）のタンパク質折りたたみシミュレーションが可能となった [2, 3]。

さらに 2020 年代に入ると、AF を含む AI モデルとのハイブリッド統合や実験手法との相補的活用が進み、タンパク質折りたたみ過程や受容体 - 抗体間の結合動態といった複雑な生体反応の理解が大きく進展している [4, 5]。

#### (2) ホモロジーモデリング法（比較モデリング法）

データベースから検索した近縁のアミノ酸配列を鋳型（テンプレート）にして、未知のタンパク質立体構造を予測する。 「相同なタンパク質の立体構造は類似する」という経験則に基づいたこの手法は、一次配列の相同性（ホモロジー）が高い既知構造を出発点として、未知のタンパク質の構造を予測することから「ホモロジーモデリング法（比

較モデリング法）」と呼ばれている [6]。立体構造データベース（主に Protein Data Bank; PDB）から、クエリ配列に最も適合する構造（テンプレート構造）を探し（ステップ 1）、見出したテンプレート構造とそのアラインメントに対して、側鎖原子や挿入ループ部を構築し、全体の立体構造モデルを作成する（ステップ 2）（図 2）。

この手法の有用性は、ミオグロビンとヘモグロビンという古典的なタンパク質においても示されている。ミオグロビンは鉄を含むヘム基とグロビンタンパク質から成り、筋肉内で酸素を貯蔵する役割を果たす。一方、ヘモグロビンは赤血球中で酸素を全身に運搬するタンパク質であり、ミオグロビンと構造的に類似したα鎖とβ鎖がそれぞれ 2 分子ずつ組み合わさった四量体構造をもつ<sup>1</sup>（図 3）。ホモロジーモデリングでは、ミオグロビンの既知の立体構造を鋳型にして、ヘモグロビンα鎖の立体構造を予測することができる。

ホモロジーモデルは、SWISS-MODEL や Modeller といったツールを用いて構築されることが多く、AF 登場以前は実験構造が未解明のタンパク質構造予測モデルとして広く利用されてきた。得られた構造は、X 線結晶構造解析における分子置換法<sup>2</sup>において、初期構造モデルとしてしばしば活用されてきた [8]。本モデルの予測精度は配列相同性に依存しており、相同性が概ね 40% 以上あれば、比較的精度よく予測できるとされている。

<sup>1</sup> X 線構造解析法によって、1958 年にケンブリッジ大学のジョン・ケンドリューらはミオグロビンの構造（分解能 6.0Å）を、1960 年には同大学のマックス・ヘルツらがヘモグロビンの構造（分解能 5.5Å）を、それぞれ明らかにした。これらの功績により両者は 1962 年にノーベル化学賞を受賞している。

<sup>2</sup> X 線結晶構造解析単体では立体構造は決定できない。回折データから得られるのは振幅（回折強度）のみであり、構造解明に不可欠な位相情報（phase informa-

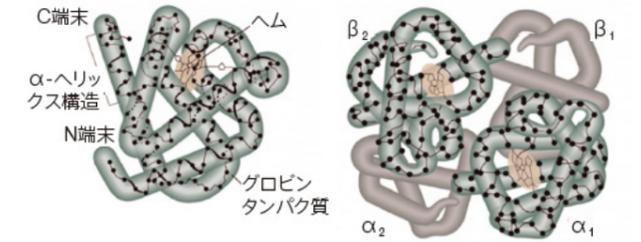


図 3 ミオグロビンとヘモグロビンの立体構造  
 （浜島書店『生物図表オンライン』(hamajima.co.jp/rika/bio/protein) を基に作成）

tion) は失われている（いわゆる位相問題; phase problem）。この問題を解決する代表的手法が分子置換法（Molecular Replacement; MR）であり、類似タンパク質の立体構造やホモロジーモデルを鋳型として用いることで、理論的に位相を推定し、電子密度マップの再構成を可能にする。

#### (3) フラグメントアセンブリ法

1997 年にペイカーらが提唱したフラグメントアセンブリ法（図 4）[9] は、平たくいえば、既知のタンパク質構造を部分ごとにつなぎ合わせ、全体として最も安定な構造を予測するというものである。

予測対象となるタンパク質配列をアミノ酸フラグメント（3～9 残基程度）に分割し、それぞれに対応する膨大な構造候補をデータベースから取得する。これらのフラグメントをランダムに組み立てて全体構造を構築する（フェーズ I）。構築された多数の構造候補に対して、物理化学的なエネルギーや立体配置に基づくスコア関数を用いて評価し、最も安定と見なされる構造を選出する（フェーズ II）。

この手法は、テンプレートに依存しない「ab initio」型の予測手法

表 1 代表的なタンパク質の予測手法（計算的手法）

手法	概要	特徴	課題
全原子シミュレーション (ab-initio法, 分子動力学など)	• 電子～原子レベルで各原子間相互作用を計算 • 運動方程式に基づいて原子移動を反復	• 原子レベルでの精密構造や動的挙動の解析が可能	• 原子間相互作用の精度確認が困難 • 大規模リソースが必要
ホモロジーモデリング法 (比較モデリング法)	• 構造データベースから検索した近縁（一次配列の相同性が高い）アミノ酸配列を鋳型（テンプレート）にして未知タンパク質の構造を予測	• 既知構造に近い場合は高精度な予測が可能	• テンプレートが存在しない場合や相同性が低い場合に精度が低下しやすい
フラグメントアセンブリ法	• フラグメント（3～9残基程度）を予測したいアミノ酸配列に基づいてランダムに組み立て • スコア関数を用いて最適構造を選択	• テンプレート不要 • 高自由度のタンパク質構造予測	• 計算時間やリソースの消費 • 最適構造への収束が困難な場合がある

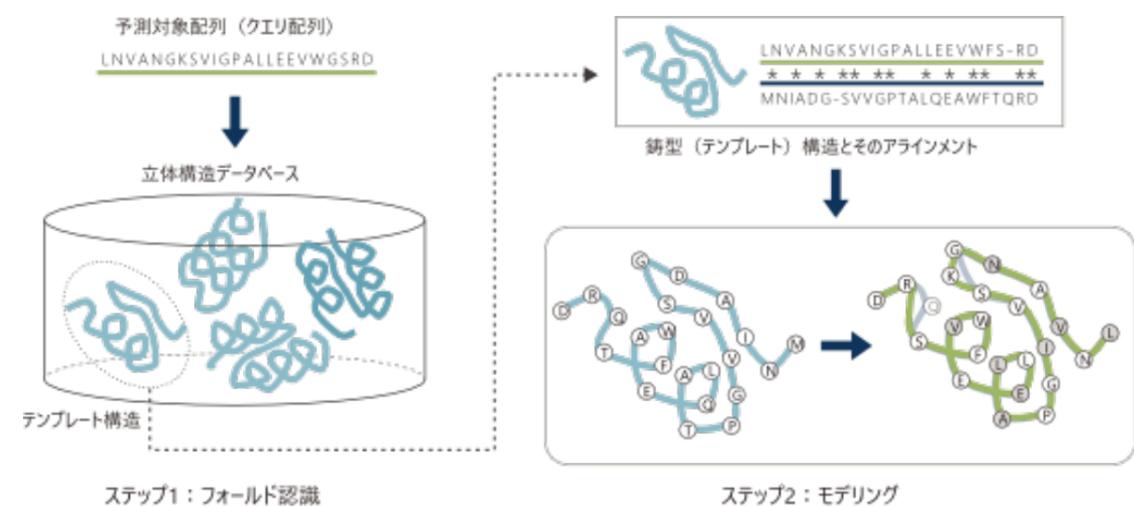


図 2 ホモロジーモデリング法 [7]

# No. 26 AI × 医療シリーズ

## AlphaFold

であることから、従来のホモロジーモデリングでは扱いきれない自由度の高い構造予測にも対応可能とされる。しかし、残基数が増えるとともに二面角や立体配置の自由度が増加し、構造候補の組み合わせ数は指数関数的に増大するため、計算時間や大規模なリソース消費がボトルネックとなる。

バイカーはフラグメントアセンブリ法に基づくソフトウェア“Rosetta”を開発、CASP3（1998年）、CASP4（2000年）<sup>3</sup>で優れた成績を収め、「未知なるフィールドの予測（類似構造がPDBに存在しない）」に有効なアプローチであると位置づけられた[10, 11]。一方で、大規模タンパク質や特殊な局所構造を持つターゲットに対しては、まだ予測精度が十分ではないとの指摘もあった[11]。

<sup>3</sup> CASP (Critical Assessment of protein Structure Prediction) は、米国国立総合医学研究所 (NIGMS) の支援のもと、1994年から隔年で開催されているタンパク質立体構造予測に関する国際的な評価実験。参加者は、まだ公開されていないタンパク質の立体構造について、その一次配列の情報のみを基に予測を行い、後に実験により明らかにされた構造との一致度を競う。1990年代からの大規模な技術革新により、ゲノムシークエンシングの効率と精度が飛躍的に向上し、タンパク質をコードするアミノ酸配列の情報が爆発的に蓄積されていった。一方で、「対応する立体構造情報が十分に得られていなかったこと」が、CASP設立の背景にある。

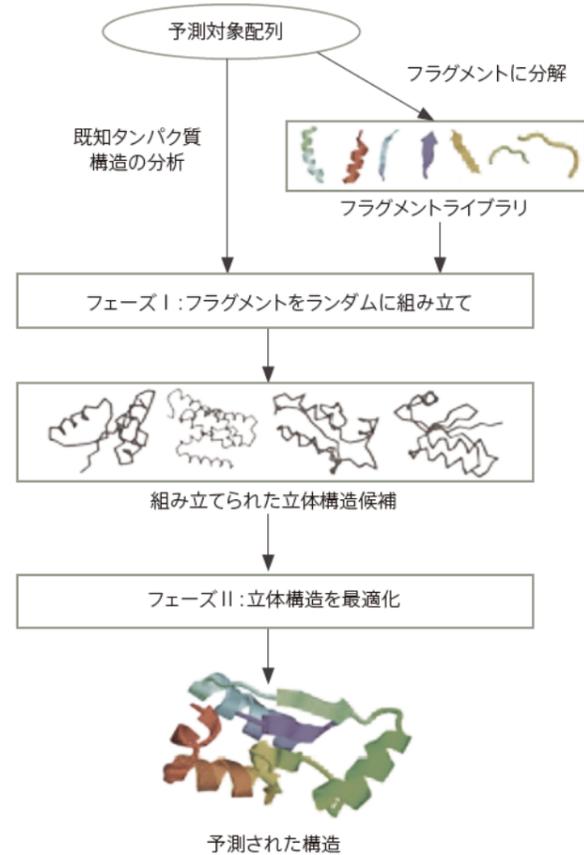


図4 フラグメントアセンブリ法（文献[12]を基に作成）

## 2. AlphaFold 小史

2018年にAFが登場してから、タンパク質立体構造予測技術は急速に進化を遂げた。それから6年間の主な出来事を表2にまとめる。

### (1) 初代 AlphaFold

アルファ碁がもたらした一種の狂騒。その余韻覚めやらぬ2018年、タンパク質立体構造予測コンペ CASP13[7]に初名乗りを上げたAFは、いきなりGDTスコア（「CASPの主要指標“GDT\_TS測定値”」）約60を達成した（図5濃灰色棒）[9, 10]。CASP12までのGDTは40程度であったことから、大きく記録を更新した形となり、「構造予測コミュニティに少なからぬ動揺を与えた[13]」。その一方で、「その絶対値としての予測精度は、生物学者が研究室で日常的なルーチンとして使うにはまだ足りないレベル[14]」という認識、「既存手法の延長線上にあり、予測に失敗したタンパク質構造も依然として多く存在していた[1]」という評もあり、この時点では来たるべき変革の予兆にすぎなかった。

### (2) AlphaFold 2

#### CASP14（2020）を席巻したAF2

予兆が現実となった。2020年11月、AFの改良版であるAlphaFold 2（AF2）を擁したディープマインド社はCASP14でGDTスコア87（テンプレート依存なしの未知構造に限定）を達成し連覇を果たした（図5青色棒）。

きわめて大きなタンパク質などの難題が含まれていたにもかかわらず（CASPでは約100種類のタンパク質アミノ酸配列が「お題」として提示される）、AF2は実験で決定された結晶構造とほぼ変わらないという驚異的な予測精度を達成した。「予測構造の信頼性が高く、様々な研究にそのまま利用できる[15]」など、研究コミュニティにも大きな反響があった。

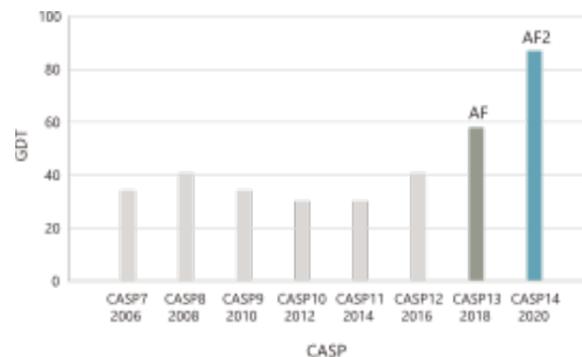


図5 CASPにおけるGDTスコアの変遷[16]

表2 AF小史

年	月	事例
2016		ディープマインド社がAF開発を開始
2018	12	CASP13においてAFが圧勝
2020	11	CASP14において、AF2が実験的手法に匹敵する精度を実現
2021	7	AF2の技術的詳細と成果をまとめた論文がネイチャー誌に掲載（2025年1月時点での被引用数が2万回超） AF2がオープンソース化 「AlphaFold DB」を公開（約35万のタンパク質構造予測データを無償提供）
	10	「AlphaFold-Multimer」を発表（複数鎖の予測に拡張）
	12	「ネイチャーの10人(Nature's 10)」にジャンパーが選出 「ブレイクスルー・オブ・ザ・イヤー」にAFが選出
2022	7	「AlphaFold DB」の登録データ数が2億超に
2023	7	「AlphaFill」を発表（リガンドやコファクターを自動的に補充）
2024	5	AF3登場（核酸、リガンドなどの相互作用を予測可能に） 9 AF3がオープンソース化（商用利用は限定的） 10 ハサビス、ジャンパーがノーベル化学賞を受賞

#### 様々な指標からみたAF2

GDTスコア以外の指標からもAF2の圧巻ぶりが見えてくる。CASP14で出題された87タンパク質に対して、他グループのRMSD値<sup>4</sup>が3.0Å前後だった中、AF2は中央値0.9Åを記録、構造類似性で他を大きく引き離す結果となった（図6）[17]。また、図に示されるバーの長さ（信頼区間）から、AF2は、全体的に予測誤差のばらつきが少なく、どのターゲットに対しても安定して高精度であったことがわかる。

#### CASPの主要指標“GDT\_TS測定値”

グローバル距離テスト（Global Distance Test - Total Score, GDT\_TS）は、実験で明らかになった構造と予測モデルとの類似性を測定する尺度であり[18]、CASPにおける主要な評価基準の一つである。この指標は、「実験構造の対応原子からの距離が1Å, 2Å, 4Å, 8Å以内にある各残基の割合」の平均値を、0～100の範囲で正規化している（100は完全一致、0は完全不一致）。スコアが90以上で、予測構造は実験構造とほぼ一致しており、信頼性が非常に高いと見なされる。CASP14でAF2はGDT\_TS中央値92.4（全課題を対象）を達成した。この他、CASPではLDDT（Local Distance Difference Test）やTM-score（Template Modeling score）などの評価基準もあり、それぞれ「局所的な構造精度」「グローバルな構造類似性」の指標となる。

<sup>4</sup> Root Mean Square Deviation. 実験構造と予測構造の対応する原子（通常はCα原子）間の距離差を二乗し、その平均をとった後に平方根をとる指標。値が小さいほど予測構造が実験構造に近いことを示す。一般に、RMSDが2.0Å以下であれば構造類似性が高いと評価される。

また、CASPで用いられる9種類のZスコア（平均よりも上であればZは大きくなる）に重み付けて総和したS値の中央値で比較したものが図7。2位（バイカーら）の約2倍のスコアを記録し、圧倒的な総合力を示した。

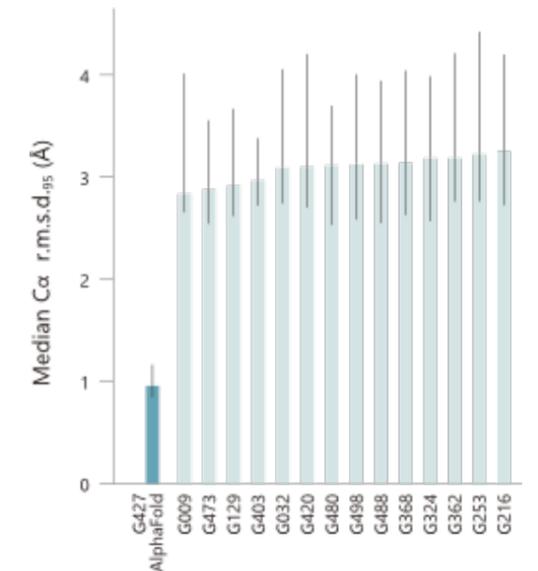


図6 CASPにおけるAF2の予測精度  
CASP14データセット（87個のタンパク質ドメイン）でのAF2のパフォーマンスを上位15エントリー（146エントリー中）と比較。横軸G+3桁はCASPが各エントリーに割り当てたグループ番号。棒グラフは中央値を、エラーバーは中央値の95%信頼区間を示す（ブートストラップ10,000回による推定）。（[17]Fig.1より引用）

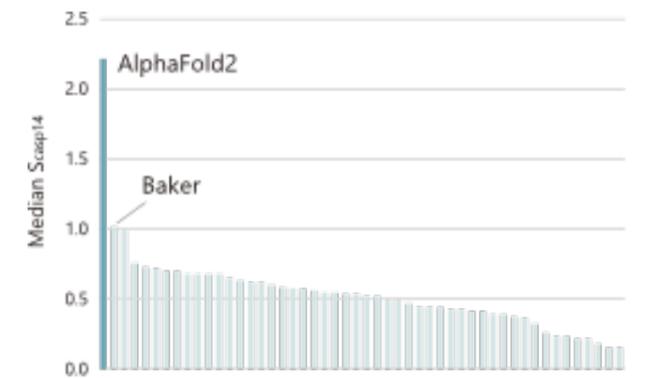


図7 CASP（High Accuracy Modeling Category）の上位50グループ（文献[19]を基に作成）

## 自然言語処理技術を応用したアルゴリズム「エボフォーマー」導入

驚異的な精度を達成した要因の一つが、**エボフォーマー (Evoformer)**・モジュールの導入である。これは、自然言語処理分野の飛躍的な進展を支えた深層学習手法である**トランスフォーマー・アーキテクチャ**（特に自己アテンション機構）を、タンパク質構造予測の文脈に最適化したアルゴリズムである。アミノ酸配列間や立体構造における相互作用を効率的に抽出するために設計されており、AF2 に実装することで、配列上では遠く離れていても空間的に近接するアミノ酸残基間の相互作用を適切に捉え、構造予測の精度を大きく高めた。

## AF2 における MSA・テンプレート検索

タンパク質のアミノ酸配列を入力すると、AF2 がデータベースから類似する配列や既知のタンパク質構造を探し出す。AF2 における構造予測（推論）は、① MSA（多重配列アラインメント）およびテンプレート構造検索、②学習済みモデルを用いた構造推定、③構造緩和（分子力場による立体構造最適化）という三段階で構成される [22]。ここでは、①の検索過程について、ノーベル賞発表時の資料を参照しながら簡潔に紹介する。

## 共進化情報を基に構造を推定する

相互に作用する2つのタンパク質（Protein-Protein Interaction; PPI）は「共進化 (Co-evolution)」すなわち、一方の変化が他方にも影響を与える進化的関係にあると考えられている [23, 24]。AF2 では、この共進化のシグナルを構造予測に活用しており、とりわけ重要な情報源となっている。ただし、実際の配列変化は独立して生じる挿入・欠失・置換を含むため、単純に比較しても元となった配列を推定することは難しい。

## 多重配列アラインメント (MSA)

そこで、未知のアミノ酸配列と類似する既知の配列を多数並べ、進化の過程で保存されている領域を特定する手法が適用される。この手法を**多重配列アラインメント (Multiple Sequence Alignment; MSA)**と呼ぶ。異なる生物種も対象にしなが保存領域を抽出し、その情報をもとにアミノ酸残基同士の空間的な関係性、すなわち共進化的な相互作用を推測する (図 8)。

## 距離マップ

次に、この配列情報に基づいて、タンパク質内で各アミノ酸残基がどれくらい近接しているかを見積もる「距離マップ」を作成する (図 9)。距離マップは、各残基間で予測された「距離」や「角度」を定量化したもので、立体構造を構築する際の指標として用いられる。

## ネイチャー誌、サイエンス誌から 2021 年の注目技術に選出

2021 年末、ネイチャー誌とサイエンス誌からそれぞれ注目技術として紹介された。「驚くほど正確にタンパク質構造を予測する AF を公表し、構造生物学分野を震撼させた」としてジャンパーは「**ネイチャーの 10 人 (Nature's 10)**」の一人に選出された [20]。また、サイエンス誌による「**ブレイクスルー・オブ・ザ・イヤー**」では「すべてのタンパク質の構造を明らかに：AI を用いた予測がタンパク質の形状の秘密を明らかにする (原題：Protein structures for all: AI-powered predictions show proteins finding their shape)」として AF がブレイクスルー技術の一つに選出された [21]。



図 8 多重配列アラインメント [25] より抜粋  
© Johan Jarnestad/The Royal Swedish Academy of Sciences

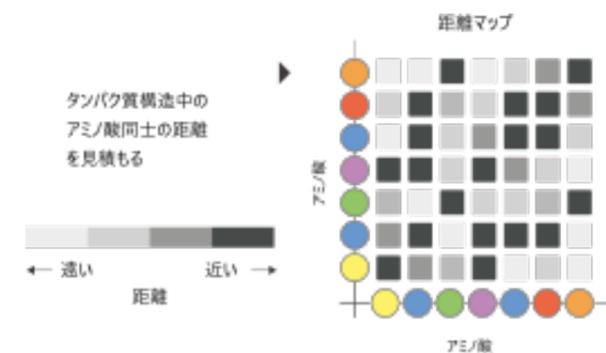


図 9 距離マップ [25] より抜粋  
© Johan Jarnestad/The Royal Swedish Academy of Sciences

## References

- [1] 森脇由隆, JSBi Bioinformatics Review, 3, 2, pp. 47-60, 2022
- [2] Klepeis, J.L. et al., Curr Opin Struct Biol, 19, 2, pp. 120-127, 2009
- [3] Shaw, D.E. et al., Science, 330, 6002, pp. 341-346, 2010
- [4] Son, A. et al., Int J Mol Sci, 25, 17, p. 9725, 2024
- [5] Raisinghani, N. et al., Viruses, 16, 9, p. 1458, 2024
- [6] Ginalski, K., Curr Opin Struct Biol, 16, 2, pp. 172-177, 2006
- [7] 川端猛, PDBjng& 創薬等情報拠点講習会, 2015, <https://pdbj.org/cms-data/workshop/20150613/kawabata.pdf>
- [8] McCoy, A.J., Acta Crystallogr D Biol Crystallogr, 63, Pt 1, pp. 32-41, 2007
- [9] Simons, K.T. et al., J Mol Biol, 268, 1, pp. 209-225, 1997
- [10] Simons, K.T. et al., Proteins, Suppl 3, pp. 171-176, 1999
- [11] Bonneau, R. et al., Proteins, Suppl 5, pp. 119-126, 2001
- [12] Lee, J. et al., Ab Initio Protein Structure Prediction, Springer Netherlands, pp. 3-25, 2009
- [13] Kryshchuk, A. et al., Proteins, 87, 12, pp. 1011-1020, 2019
- [14] 寺師玄記, 木原大亮, 日本結晶学会誌, 65, pp. 3-9, 2023,
- [15] 中川敦史, 化学, 79, 12, pp. 12-15, 2024
- [16] Google DeepMind, AlphaFold: a solution to a 50-year-old grand challenge in biology, 2020, <https://deepmind.google/discover/blog/alphafold-a-solution-to-a-50-year-old-grand-challenge-in-biology/>
- [17] Jumper, J. et al., Nature, 596, 7873, pp. 583-589, 2021
- [18] Zemla, A., Nucleic Acids Res, 31, 13, pp. 3370-3374, 2003
- [19] Pereira, J. et al., Proteins, 89, 12, pp. 1687-1699, 2021
- [20] Nature's 10, Ten people who helped shape science in 2021 <https://www.nature.com/immersive/d41586-021-03621-0/index.html>
- [21] 2021 Breakthrough of the year, <https://www.science.org/content/article/breakthrough-2021>
- [22] 理化学研究所, 「富岳」実装版「OpenFold」の提供開始 - タンパク質構造の大規模高速推論でゲノム医療や創薬に貢献 -, 2023, [https://www.riken.jp/press/2023/20230621\\_1/index.html](https://www.riken.jp/press/2023/20230621_1/index.html)
- [23] Pazos, F. et al., J Mol Biol, 271, 4, pp. 511-523, 1997
- [24] Hopf, T.A. et al., Elife, 3, 2014
- [25] Scientific Background on the Nobel Prize in Chemistry 2013 DEVELOPMENT OF MULTISCALE MODELS FOR COMPLEX CHEMICAL SYSTEMS, <https://www.nobelprize.org/uploads/2018/06/advanced-chemistryprize2013.pdf>