

2.3.10 タンパク質設計

(1) 研究開発領域の定義

本研究開発領域は、タンパク質分子の構造・機能に関わる情報科学や物理化学にもとづき、自然界のタンパク質とは異なるタンパク質を創出することで、産業、医療、細胞の制御設計に貢献する新規タンパク質を合理設計する基盤技術を構築することを目的とするものである。

(2) キーワード

タンパク質、立体構造予測、アミノ酸配列解析、進化学、合理設計、深層学習（ディープラーニング）、Rosetta、AlphaFold2

(3) 研究開発領域の概要

[本領域の意義]

タンパク質はウイルスを含むほぼすべての生命体において生命維持に不可欠な物質であり、様々な機能を担っている。ほぼすべてのタンパク質は各生物固有のDNAの塩基配列の並び順を元にポリペプチド鎖（アミノ酸配列）へと「翻訳」され、それが折りたたまって立体構造を形成することで本来の機能を発揮する。そこで、天然に存在するタンパク質よりも望ましい機能を持ったタンパク質を創出しようとする際に必要となるのがタンパク質設計技術である。酵素や抗体の改変に代表されるように、これまでにタンパク質工学は産業や医療に貢献するタンパク質を生み出してきた。本領域を推進することで、これらの分野で有用なタンパク質の創出が加速される。具体的には、種々のウイルスに結合し細胞侵入を阻害するタンパク質、ドラッグ・デリバリーシステムとして機能するタンパク質、などの創出が考えられる。さらに将来的には、超安定かつ高活性な酵素の創出にもつながると考えられる。また、望みのタンパク質を自在に設計することが可能になれば、新規に設計したタンパク質を細胞に組み込むことで、生命を制御・設計することや、さらにこれらを通じた生命の理解につながる。

タンパク質の機能はその立体構造に基づいて発揮されるため、タンパク質の立体構造をX線回折などによって解明すること、そしてタンパク質を構成するアミノ酸の配列から、タンパク質の立体構造を予測することはタンパク質設計において、極めて重要な技術の一つである。そこで本稿ではまず、タンパク質の立体構造の予測技術について、2021年に発表されて大きな話題となったAlphaFold2を中心に解説する。タンパク質設計の方法論は、天然に存在するタンパク質のアミノ酸配列に変異を導入したタンパク質を大量に準備し、その中からより望ましい機能を持つタンパク質を選抜していく「進化学」的手法と、タンパク質分子の立体構造と機能に関わる理論、仮説、データに基づいて、新規タンパク質を創出しようと試みる「合理設計」の二つに大別できる。「進化学」的手法は比較的長い歴史を持ち、既に多くのスタートアップ企業などによって社会実装されているが、「合理設計」は今まさにその学理の発展途上であり、極めて競争の激しい分野であるため、本稿では、タンパク質の「合理設計」に関わる新知見と技術について概説する。

[研究開発の動向]

【タンパク質の立体構造予測】

DNAの塩基配列を決定することは比較的容易な作業であり、そこから決まるタンパク質のアミノ酸配列はUniProtデータベース上に2022年までに2億以上蓄積されているのに対し、そのタンパク質の安定な立体構造情報を実験的に決める工程は、今なお非常に時間と労力がかかる作業であり、Protein Data Bankデータベースには2022年9月時点でも19.5万件（重複を含む）しか存在していない。したがって、50年以上もの間、アミノ酸配列からその安定な立体構造を高精度で予測する技術の開発は生命科学に携わる研究者にとって非常に大きな関心の問題の一つであった。

タンパク質構造予測は1994年から隔年で行われている世界的なコンペティション：Critical Assessment of protein Structure Prediction (CASP) によって技術が競われ続けてきた。1998年に登場した米国ワシントン大学のDavid Bakerらの研究グループが発表したRosetta¹⁾は、当初、タンパク質構造予測を行うためのソフトウェアであり、「未知のタンパク質であっても、すでにその時まで実験的に決定されているタンパク質構造の断片をうまくつなぎ合わせることで、その中で最もエネルギー的に安定なタンパク質が、自然の最も安定なタンパク質構造と一致するはず」というフラグメントアセンブリ法による構造予測手法をこれ以降展開した。この手法はおよそ100アミノ酸残基数以下で構成される類型のない新規なタンパク質構造の予測については非常に高い性能を示したが、天然のタンパク質は数百アミノ酸以上で構成される物も多いため、アミノ酸残基数が増えると組み合わせが爆発的に増えることが難点として挙げられていた。近年の構造予測は、2009年頃に発見された類縁配列を並べたもの（Multiple sequence alignment: MSA）の中に潜む共進化残基ペアの発見手法^{2, 3, 4)}、また主鎖・側鎖二面角の予測手法の発達によって、その組み合わせを大きく減らすことができ、その限界を大きく伸ばすことができるようになってきた。さらに2010年代中盤からは、米国トヨタ技術研究所シカゴのJinbo Xuらによって深層学習を取り入れることでその予測精度が伸びることが示された⁵⁾。こうした流れをくみ、2018年に米国Google社傘下のDeepMind社が第13回CASPにおいてAlphaFoldバージョン1（AlphaFold1）を発表した^{6, 7)}。このときのAlphaFold1は当時最高の性能を誇ったものの、それでもなお正解できない構造が多く存在しており、正解と判定されている予測構造の中でもまだ改善可能な点が多く残る性能だった。用いられている手法に特筆すべき目新しい技術はあまり導入されておらず、またAlphaFold1がライセンス上公開することに問題があるRosettaを用いていたため、そのソフトウェアを他の人間が使用・調査することは不可能だった。しかし、2020年にDeepMind社はAlphaFold2（AF2）⁸⁾を公開し、わずか数時間程度の計算時間で95%以上のタンパク質に対し実験的に決定された結晶構造とほぼ同じ精度の立体構造予測結果を返すという驚異的な性能を示した。タンパク質はあらゆる生命体の生体制御に大きく関わっているが、AF2の登場以前は、タンパク質の構造を決定するために必要な人的リソースや研究資金は、主にヒトの生活に関わる生命体、すなわちヒト・病原菌・ウイルス・モデル生物が持つタンパク質の構造決定に優先的に充てられており、その他の生物についての同研究分野は遅れていた。そんな中、AF2が発表されたことで、あらゆる天然に存在するタンパク質の構造を高速かつ精度良く予測できることになったことは、これまで研究が困難だった微生物・植物・様々な希少生物のタンパク質の働きを大幅に加速させることができると容易に想像できる。

【タンパク質の論理設計】

・ 20^N のアミノ酸配列空間の中から未踏のタンパク質を探索する理論・技術の開発

タンパク質は、自身のアミノ酸配列に従って形成する三次構造に基づき機能を発現する。タンパク質のアミノ酸配列空間は果てしなく広大であり、自然がこれまでに探索した配列空間は極わずかであり、この配列空間の外側には広大な未開のアミノ酸配列空間が広がっている。100残基のアミノ酸配列を持つ小型タンパク質を考慮しただけでも、アミノ酸配列の組み合わせは 20^{100} （べき乗）、つまり約 10^{130} 通り（130桁の数字）の組み合わせを考えることができる。もし地球上に1000万種類の生物が存在し、それらが10万種類の遺伝子を持っていると仮定したときのアミノ酸配列の通りは 10^{12} （12桁の数字）であり、これは生物界に存在しないかもしれないアミノ酸配列をも含めた 10^{130} （130桁の数字）通りの数に比べて非常に小さな数である。この広大なアミノ酸配列空間の中に、人類にとって役立つ、タンパク質が眠っている可能性は非常に大きい。しかしながら、これまでのタンパク質工学では、自然がこれまでに探索した配列空間の周囲のみが探索されてきた。本領域を推進することで、未踏のアミノ酸配列空間を広範に探索し、人類にとって役立つタンパク質を見つけ出すための理論と技術が構築される。

タンパク質の合理設計を行う際に、アプローチとして「自然界のタンパク質を改変すること」と「自然界のタンパク質とは全く異なるタンパク質をゼロから創出すること」で大きく分けられる。前者の場合、自然

界のタンパク質の主鎖はそのままに側鎖を変異させることで（近年は部分的に主鎖構造を変えることもなされる）、後者の場合は、タンパク質の主鎖構造から側鎖構造まで全てがゼロから設計される。本稿では後者に焦点を当てて記述する。

近年、新規タンパク質を主鎖を含めてゼロから新規設計する技術が発展し、自然が見出したアミノ酸配列空間とは全く異なる配列空間を探索することが可能になっている。その背景には、タンパク質デザインソフトウェア Rosetta に代表されるタンパク質三次構造を発生させ、かつ構造安定性を計算機上で評価する技術の発展、分子科学研究所の古賀らによる主鎖構造に関するデザインルールのような、タンパク質人工設計に関する知見の蓄積、DNA合成技術の発展、といった技術を組み合わせることができるようになってきたことが挙げられる。なかでも、主鎖構造のデザインルールについては、半世紀に及ぶ人類の試行錯誤が結実しつつある研究領域である。1973年、Anfinsenは「タンパク質は自由エネルギー最小に対応する立体構造に折りたたむ」というタンパク質熱力学原理を提唱した⁹⁾。しかし、1983年、タンパク質を設計するためには、この熱力学原理に加えて、日本原子力研究所の郷らによる整合性原理が重要となることが示された¹⁰⁾。この原理は「タンパク質が折りたたんだ後の構造は、自由エネルギー最小に対応するのみならず、アミノ酸配列上近い残基間に働く相互作用と、遠い残基間に働く相互作用が矛盾なく折りたたみ後の構造を安定にするよう設計されている」というものである。

(4) 注目動向

[新展開・技術トピックス]

【タンパク質の立体構造予測】

AF2の公開後からおよそ1週間もしない間に、現韓国のソウル大学に所属する Mirdita、Steinegger と米国ハーバード大学の Ovchinnicov が、AF2を Google Colaboratory 上で動かすための改造を施した ColabFold プロジェクトを始動した。Mirdita と Steinegger らが以前から開発を続けてきた MSA を高速に作成するためのソフトウェア HHblits¹¹⁾ は AF2 の構造予測工程の一部として組み込まれている。Ovchinnicov の提示したメタゲノムデータベースを用いた MSA の作製方法は AF2 にも取り入れられている。ColabFold プロジェクトはその後 Twitter 経由で集まった研究者を開発者・著者に迎え、AF2 にない独自の機能（複合体予測機能のプロトタイプや高速化など）を取り入れ、論文を発表した¹²⁾。

AF2 はその驚くべき性能と自由なライセンス性から、すでに多くの派生ソフトウェアや改善方法が発表されている。中国の北京大学と民間企業 DP Technology 社は AF2 を倣って自ら再実装し直し、AF2 以上の精度が出ていることを示す構造予測ソフトウェア Uni-Fold を発表した¹³⁾。米国の Helixon 社とマサチューセッツ工科大学、中国杭州の西湖大学による研究グループは、MSA を必要としない構造予測ソフトウェア OmegaFold を開発した。また、Facebook AI 研究所（現 Meta AI 研究所）の生命科学部門に所属する Alexander Rives らの研究グループも以前から MSA におけるディープラーニングの研究を発表し続けており¹⁴⁾、ESMFold と呼ばれる構造予測ソフトウェアを発表した¹⁵⁾。

AF2 が依然精度良く構造を決定できないタンパク質対象の傾向として、本質的に複数の安定な状態を取ることができる膜タンパク質や、天然に類似したアミノ酸配列が少ない（または見つからない）タイプの希少なタンパク質ファミリー、そして病原体に対するヒトを含む生物の抗体タンパク質などが挙げられる。AF2 が公開された後であってもこれらのタンパク質の構造を精度よく予測できることは生命科学または創薬の観点から重要であるため、さらに予測精度を向上させるための研究開発が続けられている。

AF2 はアミノ酸配列からタンパク質構造を高精度で予測することができるが、その予測が深層学習によるものであり、ネットワークを逆向きに遡ることで、ある望ましい立体構造を持つようなアミノ酸配列を設計できる。すなわち AF2 はタンパク質デザインに直接応用できる可能性がある。AF2 公開後に登場した米国ハーバード大学の Ovchinnicov らが開発する AFDesign¹⁶⁾ や、ProteinMPNN¹⁷⁾ では、AF2 の機能を取り入れることで、配列予測と構造予測のサイクルを繰り返すことが可能になりつつある。こうした試みは、望ましい立体

構造を取るようなアミノ酸配列を精度良く設計できるようになることを意味し、対象タンパク質と結合できるタンパク質の設計、すなわち創薬的な使い方への応用も可能となる。

【タンパク質の論理設計】

タンパク質の合理設計における生合成原理では、設計ターゲットとなる立体構造を安定化するだけでなく、その立体構造を不安定化する要素が含まれないよう設計することが重要となることが示された。古賀らは、これらの知見を踏まえ、あるタンパク質のトポロジーを設計するときに、整合性原理をみたすような局所的な主鎖構造 (α ヘリックスや β ストランドの長さおよびループの長さや形状) を記述する、タンパク質構造のデザインルールを構築してきた¹⁸⁾。このデザインルールをもとに、様々な $\alpha\beta$ 型タンパク質の主鎖構造設計図を描き、この設計図をもとに主鎖構造を構築し、この主鎖構造を安定にするような側鎖構造を設計すると、100残基を超える様々なトポロジーの $\alpha\beta$ 型タンパク質構造について、原子レベルの正確さで、主鎖を含めてゼロから構造を設計することができる^{19, 20)}。興味深いことに、これら設計したタンパク質のほとんどは、変性温度が100°Cを超える超安定構造を形成しており、このことは設計したタンパク質が機能設計の鋳型構造として利用できることを示唆している²¹⁾。

また、米国ワシントン大学のBakerらは、機能性タンパク質の主鎖構造を含めたゼロの設計に成功している。一例として、彼らはSARS-CoV-2のスパイクタンパク質結合タンパク質の設計に成功した¹⁹⁾。この研究では、SARS-CoV-2スパイクタンパク質へ結合可能な側鎖構造様式を計算機上で発生させ、次にその計算機上で大量に創り出した主鎖構造ライブラリーの中から、その側鎖構造様式を挿入可能な鋳型主鎖構造を探し出し挿入する。さらに、挿入した側鎖構造様式以外の側鎖構造をデザインすることで、スパイクタンパク質に結合するタンパク質を創り出すことに成功した。あるタンパク質に結合するタンパク質が必要な場合、人工設計したタンパク質は抗体と比べて、小型であること、耐熱性に優れていること、免疫システムに感知されにくいという利点がある。また、特殊ペプチドに比べると、人工設計したタンパク質は、大腸菌等を用いた生合成が可能のため大量生産が容易である点で優位である。

一方、AF2が、自然界のタンパク質のアミノ酸配列からの構造予測に成功したことを受け、AF2を応用してタンパク質を設計する研究がなされている。これらの研究では、「AF2は正しく折りたたみ構造を予測できる」という前提のもと、アミノ酸配列をランダムに発生させ、どのような構造に折りたたむのかをAF2で予測させることを繰り返すことで、新規タンパク質を発生させる。この技術は、単量体タンパク質の設計、タンパク質-タンパク質複合体の設計や、回転対称多量体タンパク質の設計に応用されている²³⁾。

【注目すべき国内外のプロジェクト】

【国内】

タンパク質分子の立体構造と機能に関わる理論、仮説、データに基づいた「合理設計」という観点を含みうる関連プロジェクトとして、以下のものが存在する。

• CREST・さきがけ：

「自在配列システム」原子・分子の自在配列と特性・機能

「バイオDX」データ駆動・AI駆動を中心としたデジタルトランスフォーメーションによる生命科学研究の革新

「ゲノム合成」ゲノムスケールのDNA設計・合成による細胞制御技術の創出

「高次構造体」細胞の動的な高次構造体

• 学術変革と新学術領域：

生物を凌駕する無細胞分子システムのボトムアップ構築学

発動分子科学 - エネルギー変換が拓く自律的機能の設計

【国外】

- ・米国ワシントン大学のBakerらが主導するInstitute for Protein Design (タンパク質デザイン研究所)。2012年設立。
- ・米国民間財団のOpen Philanthropyがワシントン大学のタンパク質デザイン研究所を2017～2021年の5年で1,100万ドル(約14億円)支援。
- ・TED講演会を展開するTEDが主催する米国民間財団The Audacious Projectがワシントン大学のタンパク質デザイン研究所を2019～2028年の10年で9億ドル(約1,190億円)支援。
2021年、米国マイクロソフト社がワシントン大学タンパク質デザイン研究所に500万ドル(約6.6億円)を寄付。
- ・European Research Council (ERC: 欧州研究会議)の「ProCovar」プロジェクトが英国ユニヴァーシティ・カレッジ・ロンドン(UCL: ロンドン大学)において実施された。このプロジェクトではde novoタンパク質設計技術を利用しつつ、タンパク質の構造と機能の相関について解析が進められた。2016～2022年で243万ユーロ(約3億4,000万円)の助成。

(5) 科学技術的課題

今後取り組むべき課題は、1つにAF2によって予測されたタンパク質に結合し、その機能を制御することができるような小分子を精度良く予測できるようなソフトウェアの開発が挙げられる。これは従来の創薬の場面でも使われてきた*in silico*化合物スクリーニングと似たような趣旨であるが、この化合物スクリーニングがタンパク質構造予測のときと同様に物理化学法則ベースの計算で行われていたものを、AF2の成功も後押しする形で深層学習によってこれを置き換えていく試みが、まだ性能が良いとは言い難いものの、ソフトウェアEquibind²⁴⁾など近年のトレンドとして見受けられる。

近年はAF2、囲碁のソフトウェアAlphaGOなどをはじめとして、深層学習が人間の能力を上回る性能を叩き出す例がいくつか知られているが、すべてに共通する背景として、①学習に必要なデータが大量に公開され、利用可能であること、②AIが目指すべきスコアが計算機上で明確に定義可能であること(AF2であれば立体構造の正解構造とのずれ、AlphaGOであれば単純に囲碁の勝利条件)、そして③非常によく訓練された人間ならば、その勝ち筋を言語化・理論化することができずとも、高いスコアを出して勝つことのできるゲームのルール設定であること、がある。Rosettaを用いたタンパク質構造のデザインは、かなり職人芸的な側面が強い。また、創薬の1つの手法としては、タンパク質構造の座標情報を利用し、その空隙に挟まるような化学小分子の骨格をうまく作成するStructure-based Drug Designが昔から存在する。この手法についても、よく訓練された人間であれば成功率が上がるといった性質があり、裏を返せば、それをAIが学習できれば自動化でき、かつ高い精度を出せる余地が十分あると言える。これを達成するためには、生命科学と情報学の融合を日本の大学ならびに民間の企業においても促進すること、またそのような人材を育成できるような環境を整えること(生命科学を専攻する研究科で情報科学を学んでも良いし、その逆でも良い)、そして製薬企業などが持つ膨大な治験データを大学の研究機関と提携して解析できる環境が必要となってくるかもしれない。

一方で、タンパク質合理設計の技術が進展しているものの、機能性タンパク質をゼロから自由自在かつ高精度に設計できる技術力はまだない。機能性タンパク質を設計するためには、サブÅの精度で立体構造を設計する必要があるが、現在の合理設計技術では、良く設計できた場合においてもその精度は1～2Åである。従って、計算機を用いて、いわばシードとなるタンパク質を設計したのちに、進化学の技術を用いて、そのシードタンパク質の機能活性を向上させることが必要となる。そのため、合理設計技術を向上させるとともに、機能の高いタンパク質を実験的にスクリーニングする技術を向上させることも重要な課題である。また、現在設計可能な機能性タンパク質としては、あるターゲットとなるタンパク質(例えば上記のSARS-CoV-2のスパイクタンパク質)に「結合可能」なタンパク質を設計できるのみであり、酵素活性を持つタンパク質や、モータータンパク質のように構造変化することで機能するタンパク質をゼロから設計することは非常に困難な課題

である。また、RNAやDNAに結合するタンパク質の合理設計は未だなされていない。さまざまな生体小分子に結合するタンパク質の合理設計も進んでいるとは言い難い。他にも、非天然アミノ酸を用いたタンパク質の設計や、望みの位置に望みの糖鎖を付加するような合理設計技術についてはまだ進展が見られていない。

(6) その他の課題：

AF2に代表される計算機によるタンパク質構造予測法の進歩により、自然界に存在するタンパク質であればその立体構造を高い精度で予測することが可能になった。すなわち、地球上の生物が持つ遺伝子のタンパク質構造情報が簡単に手に入る時代になったと言っても過言ではない。そのため今後は、タンパク質の立体構造を決定するだけの従来の構造生物学を刷新し、自然界のタンパク質立体構造を基にこれらを改変することで、産業・医療や生命の制御・設計に貢献するようなタンパク質の創出を試みるような分野の確立が期待される。自然界のタンパク質の現在の姿は進化の結果として偶然生じたものであり、これらタンパク質は人間が必要とする文脈において最適化できる余地がある。よって立体構造をもとに自然界のタンパク質を最適化することで、自然界のタンパク質に“付加価値”を加えようとする試みが多くなされるべきである。これを実現するためには、タンパク質科学（構造生物学）、生物学、医学が、お互いの分野の垣根を超えて連携を密にすることを促すファンディングや政策などが望まれる。

人材育成の面からは、AF2を始めとして、これからのタンパク質研究、創薬、化学、生命科学について深層学習的なアプローチが大きな影響を持っていることは必至であるが、日本において、生物・化学と深層学習を組み合わせる精力的に行っている研究チームが数えるほどしか存在していない。というのも、世界的・歴史的に見てもタンパク質の構造生物学についての発展が、生命科学を学んできた者たちによってもたらされたというよりは、物理学・情報学・統計学などの数理的分野を専門に学んできた者が、あるとき生命科学に興味を持って参入するという形で生命科学を発展させてきたことが多いためである。かつて日本において1990年代後半～2000年前半代のゲノム解読などバイオテクノロジーブームのときには、大学院の選択のときに物理学系・情報学系の専攻から生命科学系に参入した人間が多かったと聞く。2020年代では世界的な深層学習の発展に伴いそちらの分野は非常によく人材が活発であるが、現在は日本において生命科学のブームはやや下火になっており、情報系の学生が生命科学に来るということは純粋な興味を持つ希少な一部を除いてほとんどいない。

多くの大学において生命科学を専門とする学部において数理的な教育も行えるところは、教員が存在しないという背景もあり、まだ少ない。生命科学の専攻向けの情報系の講義が行われるような仕組み作りを急ぐ必要がある。

産学連携という観点からも、製薬企業、食品の酵素材などのバイオインダストリー系の会社が、深層学習や計算科学を扱い研究できるような人材を多く採用し、AIによる創薬が盛んになれば、これを受けて世間の認知も広まり、大学側では情報系学科と生命科学系学科の交流が増えていくかもしれない。

ただ、いずれにせよ当該分野がこれから大きく成長することが予想される分野であるにも関わらず、それを遂行できるアカデミック人材（非学生）が現状ほとんどいないことが最大の問題である。「選択と集中」による研究開発予算の投下はたいてい一過性であり、レベルの高い専門性を持った人材が「選択と集中」によって育成されたとしても、そうした人材は割の良い民間企業へ流れることが容易に想像される。この分野に限ったことではないが、官民間問わずレベルの高い専門知識を持った研究者を輩出するためには、大学においてそれを教え、研究開発できる人材を確保し続けることが大前提となる。

(7) 国際比較:

国・地域	フェーズ	現状	トレンド	各国の状況、評価の際に参考にした根拠など
日本	基礎研究	△	↗	タンパク質の立体構造と深層学習を組み合わせた研究は産総研・九州工業大学・東京工業大学など一部の研究者で、Rosettaを用いたタンパク質デザインは分子科学研究所、名古屋大学・東京大学教養学部の一部の研究者周辺で、それぞれ見られるが、いずれも中国・米国と比べると発達がかなり遅れている。
	応用研究・開発	△	↗	Rosettaを用いたタンパク質デザインの応用例は分子科学研究所のグループからいくつか発表されている程度。製品化されたものはほぼないが、特殊ペプチドなどの大学発スタートアップが存在する。
米国	基礎研究	◎	↗	タンパク質の立体構造の深層学習とワシントン大学のBakerを中心とするRosettaのタンパク質デザインの本場であり、20年以上に渡って非常に盛んに基礎研究が行われている。上述のように大学だけでなく、Meta社、DeepMind社などの民間企業による基礎研究も非常に盛んである。
	応用研究・開発	◎	→	ワシントン大学のBakerらはマウスでアポトーシスを誘導する人工タンパク質の医学応用 ²⁵⁾ を2014年ですでに手掛けており、2020年にはSARS-CoV-2の活動を阻害する人工タンパク質のデザインも達成している ²²⁾ 。ここから派生した民間のタンパク質デザインベンチャーも、ワシントン大学のあるシアトルを中心としていくつか誕生している。
欧州	基礎研究	○	↗	米国ほどではないが、計算によるタンパク質研究は昔から盛んに行われている。University College London (UCL) のDavid Tudor Jonesが主導するUCL Bioinformatics Groupはこの中でも精力的に計算科学によるタンパク質構造予測を行い1999年にタンパク質二次構造予測ツールPspred ²⁶⁾ を発表、2019年にはDeepMetaPSICOV ²⁷⁾ を開発している。近年はそして深層学習を用いたアミノ酸配列の生成器作製を開発している ²⁸⁾ 。
	応用研究・開発	○	↗	スイス・ドイツの研究機関から、RSウイルス感染症に対するヒトの中和抗体産生を誘導する人工タンパク質をデザインした論文が2020年Science誌で発表されている。これはRosettaだけでなく彼ら独自の支援ツールTopobuilderソフトウェアを開発しているという点でも計算科学と生命科学の融合が見られる ²⁹⁾ 。
中国	基礎研究	○	↗	北京大学& DP Technology社など、AF2を模しているとは言え類似のソフトウェアを自前で開発できるだけの能力と体制は整っているようである。また、米国やドイツのMPIなど他国の研究機関に渡っていた中国系研究者が中国に戻り研究を続ける体制が整っている。
	応用研究・開発	△	→	製品化という意味ではまだこれらの分野における目立った応用研究は少ない。
韓国	基礎研究	△	↗	タンパク質構造予測・タンパク質デザインに直接関係する研究の例は少ないが、AF2の構造予測とColabFold開発に関わったMartin SteineggerとMilot Mirditaらが現在ソウル大学で研究室を主宰し、タンパク質デザインを追求している。
	応用研究・開発	△	→	現在のところ目立った応用研究は少ない。

(註1) フェーズ

基礎研究：大学・国研などでの基礎研究の範囲

応用研究・開発：技術開発（プロトタイプの開発含む）の範囲

(註2) 現状 ※日本の現状を基準にした評価ではなく、CRDSの調査・見解による評価

◎：特に顕著な活動・成果が見えている

○：顕著な活動・成果が見えている

△：顕著な活動・成果が見えていない

×：特筆すべき活動・成果が見えていない

(註3) トレンド ※ここ1~2年の研究開発水準の変化

↗：上昇傾向、→：現状維持、↘：下降傾向

参考文献

- 1) Simons KT, Bonneau R, Ruczinski I, Baker D. *Ab initio* protein structure prediction of CASP III targets using ROSETTA. *Proteins-Structure Function and Bioinformatics*. 1999 ; 171-6.
- 2) Weigt M, White RA, Szurmant H, Hoch JA, Hwa T. Identification of direct residue contacts in protein-protein interaction by message passing. *Proceedings of the National Academy of Sciences of the United States of America*. 2009 ; 106 (1) : 67-72.
- 3) Morcos F, Pagnani A, Lunt B, Bertolino A, Marks DS, Sander C, et al. Direct-coupling analysis of residue coevolution captures native contacts across many protein families. *Proceedings of the National Academy of Sciences of the United States of America*. 2011 ; 108 (49) : E1293-E301.
- 4) Balakrishnan S, Kamisetty H, Carbonell JG, Lee SI, Langmead CJ. Learning generative models for protein fold families. *Proteins-Structure Function and Bioinformatics*. 2011 ; 79 (4) : 1061-78.
- 5) Wang S, Sun SQ, Li Z, Zhang RY, Xu JB. Accurate De Novo Prediction of Protein Contact Map by Ultra-Deep Learning Model. *Plos Computational Biology*. 2017 ; 13 (1) : e1005324.
- 6) Senior AW, Evans R, Jumper J, Kirkpatrick J, Sifre L, Green T, et al. Protein structure prediction using multiple deep neural networks in the 13th Critical Assessment of Protein Structure Prediction (CASP13). *Proteins-Structure Function and Bioinformatics*. 2019 ; 87 (12) : 1141-8.
- 7) Senior AW, Evans R, Jumper J, Kirkpatrick J, Sifre L, Green T, et al. Improved protein structure prediction using potentials from deep learning. *Nature*. 2020 ; 577 (7792) : 706-10.
- 8) Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, et al. Highly accurate protein structure prediction with AlphaFold. *Nature*. 2021 ; 596 (7873) : 583-9.
- 9) Anfinsen, C B Principles that govern the folding of protein chains. *Science* (1973) 181 : 223-230.
- 10) Go, N. 1983. "Theoretical Studies of Protein Folding." *Annual Review of Biophysics and Bioengineering* 12: 183-210.
- 11) Steinegger M, Meier M, Mirdita M, Vohringer H, Haunsberger SJ, Soding J. HH-suite3 for fast remote homology detection and deep protein annotation. *BMC Bioinformatics*. 2019;20(1) : 473.
- 12) Mirdita M, Schütze K, Moriwaki Y, Heo L, Ovchinnikov S, Steinegger M. ColabFold: making protein folding accessible to all. *Nature Methods*. 2022.
- 13) Li Z, Liu X, Chen W, Shen F, Bi H, Ke G, et al. Uni-Fold: An Open-Source Platform for Developing Protein Folding Models beyond AlphaFold. *bioRxiv*. 2022 : 2022.08.04.502811.
- 14) Rao R, Liu J, Verkuil R, Meier J, Canny JF, Abbeel P, et al. MSA Transformer. *bioRxiv*. 2021 : 2021.02.12.430858.
- 15) Lin Z, Akin H, Rao R, Hie B, Zhu Z, Lu W, et al. Language models of protein sequences at the scale of evolution enable accurate structure prediction. *bioRxiv*. 2022 : 2022.07.20.500902.
- 16) Sergey Ovchinnikov, & Justas Dauparas. (2022). AfDesign - Partial Hallucination with sidechain constraints (v0.0.0). Zenodo. <https://doi.org/10.5281/zenodo.6803187>
- 17) Dauparas J, et al. Robust deep learning-based protein sequence design using ProteinMPNN. *Science* (2022) 378: 49-56.

- 18) Koga N. et al., “Principles for Designing Ideal Protein Structures.” *Nature* (2012) 491: 222-27.
- 19) Lin Y-R et al., “Control over Overall Shape and Size in de Novo Designed Proteins.” *PNAS* (2015) 112: E5478-5485.
- 20) Koga N. et al., “Role of Backbone Strain in de Novo Design of Complex α / β Protein Structures.” *Nature Communications* (2021) 12: 3921.
- 21) Koga R. et al., “Robust folding of a de novo designed ideal protein even with most of the core mutated to valine.” *PNAS* (2020) 117: 31149-56.
- 22) Cao, L. et al., “De novo design of picomolar SARS-CoV-2 miniprotein inhibitors.” *Science* (2020) 370: 426-31.
- 23) Wicky B I M et al., “Hallucinating symmetric protein assemblies.” *Science* (2022) 378 : 56-61.
- 24) Stärk H et al., “EquiBind: Geometric Deep Learning for Drug Binding Structure Prediction” Accepted at the ICLR 2022 Workshop on Geometrical and Topological Representation Learning. Available from [<https://openreview.net/pdf?id=Brxey2E-plq>] 2023年2月アクセス
- 25) Procko E et al., A Computationally Designed Inhibitor of an Epstein-Barr Viral Bcl-2 Protein Induces Apoptosis in Infected Cells. *Cell*. 2014 ; 157 (7) : 1644-56.
- 26) Jones DT. Protein secondary structure prediction based on position-specific scoring matrices. *Journal of Molecular Biology*. 1999 ; 292 (2) : 195-202.
- 27) Kandathil SM, Greener JG, Jones DT. Prediction of inter residue contacts with DeepMetaPSICOV in CASP13. *Proteins-Structure Function and Bioinformatics*. 2019 ; 87 (12) : 1092-9.
- 28) Greener JG, Moffat L, Jones DT. Design of metalloproteins and novel protein folds using variational autoencoders. *Scientific Reports*. 2018 ; 8.
- 29) Sesterhenn F, Yang C, Bonet J, Cramer JT, Wen XL, Wang YM, et al. De novo protein design enables the precise induction of RSV-neutralizing antibodies. *Science*. 2020 ; 368 (6492) : 730

2.3