

In silico スクリーニングを用いて、毒性ウィルスタンパク質と宿主タンパク質との相互作用を予測する

河口 真一

大阪大学 大学院生命機能研究科

1. はじめに

細胞内で生命現象の物理化学的反応を担っているのは、多くの場合、タンパク質である。タンパク質は、20 種類のアミノ酸が連なったポリマーであり、アミノ酸配列に応じた立体構造に折り畳まれる。生物のゲノム DNA にコードされているタンパク質の種類は、ヒトで約 20,000 種類も存在する。これらのタンパク質が相互作用し、複合体を形成することによって、複雑な生体反応が制御されている。従って、タンパク質の機能を研究する場合には、それと相互作用するタンパク質を探索することがよく行われる。相互作用するタンパク質を実験的に同定する研究も広く行われているが、時間と労力が多いため、迅速なスクリーニング手法が望まれている。

タンパク質間の相互作用は、各タンパク質の立体構造表面の特性によって決まる。タンパク質の立体構造を決定するには、当該タンパク質を単離精製する必要があるが、多くのタンパク質については、実験的に立体構造を決定することは行われていない。一方で、最近、DeepMind 社によって開発された AlphaFold2 という AI プログラムは、アミノ酸配列の共進化情報を利用して、タンパク質の立体構造を高精度で予測できるとして注目されている¹。さらに、タンパク質の複合体構造も予測できることが期待されている²。そこで、AlphaFold2 を利用して、迅速なタンパク質複合体のスクリーニングを行うことが試みられている³。このような in silico のスクリーニング系は、実験上の制約がある強毒性タンパク質の場合に、特に有用だと考えられる。AlphaFold2 を用いて、強毒性タンパク質と複合体を形成する可能性のあるタンパク質を絞り込むことによって、実験の負担を軽減し、毒性タンパク質の病理学的メカニズムの研究を促進できる。

2. 研究方法

2.1 クリミア・コンゴ出血熱ウイルス (CCHFV) の NP タンパク質

本研究では強毒性ウイルスであるクリミア・コンゴ出血熱ウイルス (CCHFV) のタンパク質に着目した。CCHFV は、アフリカ、アジア、ヨーロッパの一部で発生しており、主にマダニを介して感染する。感染した場合には非常に重篤な症状を示し、致死率は 10~40% にもなる⁴。CCHFV のゲノムにコードされているタンパク質のうち、NP タンパク質は、ウイルスの RNA に結合して、ウイルス粒子内にパッキングする役割を担っている。さらに、ヒトの細胞内では、アポトーシスという細胞死の誘導に関わることが報告されている⁴。NP タンパク質は、482 アミノ酸残基からなり、単独の立体構造はすでに明らかにされている⁵。N 末領域と C 末領域によって形成される Head ドメインと、中央領域によって形成される Stalk ドメインに分けられる (図 1)。ドメインとは、タンパク質の立体構造における部分的なユニットを意味する。

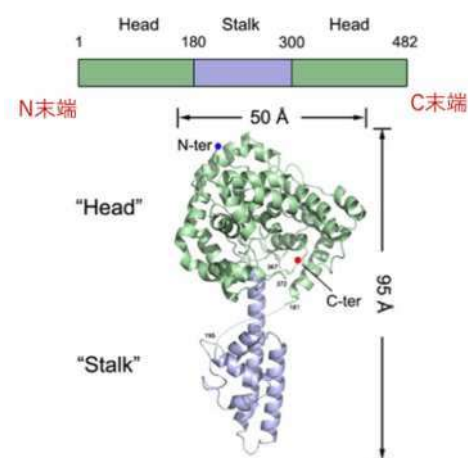


図 1 CCHFVのNPタンパク質の立体構造
NPタンパク質は、2つのドメインから成っている。

2.2 AlphaFold2 プログラムを用いたタンパク質複合体の構造予測

AlphaFold2 プログラム^{1,2}は、結合を予測したいタンパク質と似た配列をもつタンパク質を、多くの生物種から集め、保存性に従ってアライメントする。これによって、保存されているアミノ酸残基と、生物の進化に伴って変化している残基にわけける。NP タンパク質の変化と、ターゲットとなるタンパク質との変化に対応がみられる場合には、それら残基間の結合が保存されている可能性が考えられる（共進化）。アライメントの作成は、大阪大学 D3 センターにあるスーパーコンピュータ、SQUID の CPU ノードを用いて計算した。

AlphaFold2 では、上述のアライメント情報を基に、初期パラメータが異なる 5 つの予測構造を出力する。オリジナルの計算フローでは、1 つずつ順に計算していたが、本研究では、5 つの GPU を使って並列に計算し、予測時間を短縮している³。出力された 5 つの構造のうち、最もスコアの高いものを予測構造として採用した。1 つのタンパク質ペアの予測構造を計算する時間は、1~3 時間程度であり、タンパク質の長さと同タンパク質の数に依存していた。SQUID の計算資源を効果的に活用するため、タンパク質の長さが 1,000 アミノ酸残基以下のものに対してスクリーニングを行った。

3. AlphaFold2 を用いたタンパク質間相互作用の in silico スクリーニング

3.1 信頼性スコアの分布

本研究では、CCHFV 由来の NP タンパク質とヒトタンパク質とが、1 : 1 で結合すると仮定して、相互作用の可能性を、AlphaFold2 を用いてスクリーニングした。AlphaFold2 は、予測した複合体に対して、信頼性スコアを 0 から 1 の範囲で与える。先行研究から、スコアが 0.6 以下の場合には、計算毎にスコアが大きく異なる場合があり、信用性は低い³。

約 15,000 種類のヒトタンパク質について、NP タンパク質との二量体構造を予測させた結果、信

頼性スコアが 0.7 以上であるものが、93 種類ほど見つかった（図 2）。スコアが高いほど信頼性が高いと考えられるが、偽陽性も多い³。従って、さらに精度を高めることを目的として、以下の二次スクリーニングを行った。



図 2 AlphaFold2による信頼性スコアの分布

CCHFV-NPタンパク質と15,000種類のヒトタンパク質との結合予測スクリーニングを行った。信頼性スコアは、0から1の範囲で与えられ、大きいほど信頼性が高い。

3.2 複合体予測の二次スクリーニング

3.2.1 AlphaFold3 の利用

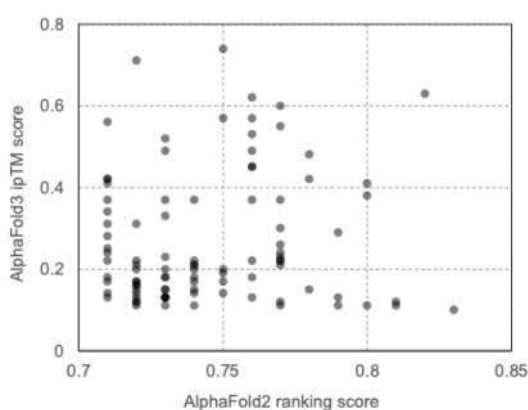


図 3 : AlphaFold Version 2 と 3 の比較

AlphaFold2 を用いた複合体予測を行っている最中に、改訂版である AlphaFold3 が発表・公開された⁶。予測精度も向上していることが期待される。そこで、AlphaFold2 で信頼性スコアが高い（0.7 以上）候補について、AlphaFold3 による複合体予測を行った（alphafoldserver.com/）。その結果、AlphaFold3 の予測では多くの候補が低いスコアを示し、2 つバージョンから得られる予測スコアには相関が見られなかった（図 3）。このことは、AlphaFold2 の予測精度が低いことを反映していると考えられるが、今後、AlphaFold3 における予測精度の実験的検証も必要である。

3.2.2 結合強度の見積もり

AlphaFold は複合体の構造を予測するプログラ

ムであるが、その複合体の結合強度を予測することはできていない。そこで、AlphaFold2によって予測された複合体の立体構造を基に、物理化学的に結合強度を見積もることを行った⁷。2つのタンパク質間の接触面にある荷電残基や疎水性相互作用を考慮して、解離定数を見積もられる。その結果、解離定数が1から1000 nMの範囲に予測されるタンパク質が多くみられた(図4)。

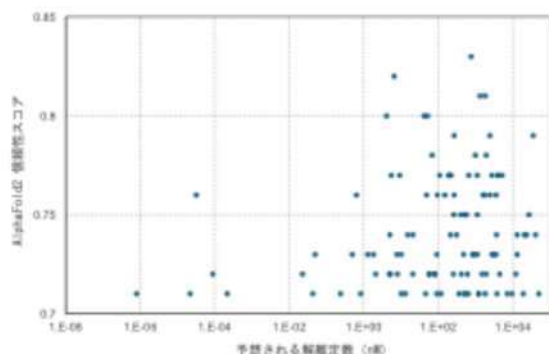


図4 複合体の解離定数の見積もり

AlphaFold2によって予測された複合体の立体構造を基にして、その解離定数を見積もった。

タンパク質間の結合を検出する目的には、共免疫沈降法がよく用いられるが、結合が強い(解離定数が小さい)ことが前提となる。一般的に、強い結合とされる解離定数が10 nM程度以下であれば、共免疫沈降法を用いて、2つのタンパク質間の結合を検出できると期待される。本スクリーニングにおいて、AlphaFold2の信頼性スコアが、0.7以上、見積もり解離定数が10 nM以下のタンパク質は、24個見られた。

4. 予測された複合体

二次スクリーニングの結果、CCHFV-NPタンパク質と複合体を形成する可能性が高い候補を得ることができた。AlphaFold2が予測した複合体構造を精査した。

4.1 #7052 タンパク質

#7052 タンパク質とNPタンパク質の予測複合体構造を図5に示す。#7052は、低分子量Gタンパク質を活性化するTBCドメインとキナーゼ様ドメインを併せ持つタンパク質であり、細胞

質に存在する。細胞内のmRNA輸送に関わるFERRY複合体に結合することが報告されているが、分子機能の詳細は不明である。

予測された複合体構造では、2つのタンパク質ともに、フレキシブルなループ領域は少なく、安定した構造であると考えられる。#7052のTBCドメインが、NPのHeadドメインと相互作用していることから、NPが#7052による低分子GTPase活性化を阻害しているのかもしれない。

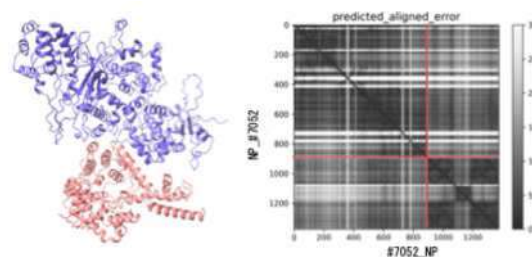


図5 AlphaFold2が予測したタンパク質複合体の立体構造(#7052_NP)

(左) タンパク質複合体の立体構造をリボンモデルで示す。#7052タンパク質は紫色、NPタンパク質は、オレンジ色で示す。(右) PAEプロットでは、#7052とNPタンパク質のアミノ酸残基を、X軸では左から、Y軸では上から並べている。アミノ酸残基が交差する点が黒色の場合、タンパク質複合体において、それら残基の相対的位置が確からしいことを表す。赤色の線は、タンパク質の境界を示す。

4.2 #7030 タンパク質

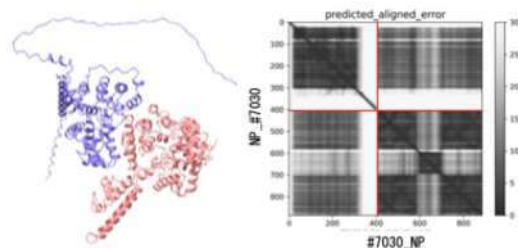


図6 AlphaFold2が予測したタンパク質複合体の立体構造(#7030_NP)

(左) タンパク質複合体の立体構造をリボンモデルで示す。#7030タンパク質は紫色、NPタンパク質は、オレンジ色で示す。(右) PAEプロットでは、#7030とNPタンパク質のアミノ酸残基を、X軸では左から、Y軸では上から並べている。アミノ酸残基が交差する点が黒色の場合、タンパク質複合体において、それら残基の相対的位置が確からしいことを表す。赤色の線は、タンパク質の境界を示す。

#7030 タンパク質も、TBCドメインを持つタンパク質である(図6)。TBCドメインを有するタンパク質は他にも数多く知られており、NPタンパク質は、それらのTBCドメインとも結合するかもしれない。本スクリーニングで示した2つのTBCドメインタンパク質(#7052と#7030)の他、さらに2つのTBCドメインタンパク質が、NPと結合する可能性があることが予測されている(AlphaFold2の信頼性スコアが0.68と0.65)。#7030は、C末端近くに1つの膜貫通領域があり、ERに局在する。TBCドメイ

ンは細胞質側に位置しており、ER 膜近傍で NP タンパク質と結合すると考えられる。

4.3 #9673 タンパク質

#9673 タンパク質は、カルボキシペプチダーゼのサブユニットとの 1 つであり、細胞外に分泌されて、基質タンパク質の分解に寄与する。C 末の小さいドメインを除いた大部分の領域には、ロイシンリッチリピートがあり、17 本のベータストランドが蹄鉄型フォールドを形成している。NP タンパク質は、蹄鉄型の内側と相互作用することが予測された (図 7)。これは、#9673 が、カルボキシペプチダーゼの他のサブユニットと結合する場所でもある。即ち、NP タンパク質が細胞外に存在する様な条件下では、カルボキシペプチダーゼの複合体形成が阻害され、その活性も損なわれる可能性がある。

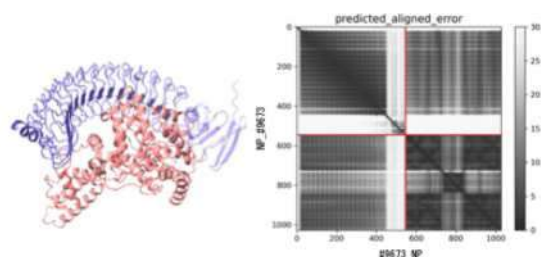


図 7 AlphaFold2 が予測したタンパク質複合体の立体構造 (#9673_NP)

(左) タンパク質複合体の立体構造をリボンモデルで示す。#9673 タンパク質は紫色、NP タンパク質は、オレンジ色で示す。(右) PAE プロットでは、#9673 と NP タンパク質のアミノ酸残基を、X 軸では左から、Y 軸では上から並べている。アミノ酸残基が交差する点が黒色の場合、タンパク質複合体において、それら残基の相対的位置が確からしいことを表す。赤色の線は、タンパク質の境界を示す。

今後、これらの相互作用について、実験的検証を行い、結合を確認できれば、CCHFV がもたらす疾患のメカニズムを明らかにできると期待される。

5. おわりに

現在、AlphaFold2 のような AI プログラムの開発が盛んに行われており、生命科学の発展に大きく貢献することが期待されている。しかし、そのような計算機科学と、それを利用する実験科学の間には、まだ隔たりがあるのが実情である。両者の隔たりを埋めて、最新の計算機科学を、実験科学が迅速に取り込むことができる環境が望まれる。AlphaFold によるタンパク質間複

合体の予測は、未だ精度が低いのが現状であるが、二次スクリーニング法の開発や、相互作用実験からのフィードバックを学習することによって、予測精度の向上が期待される。その結果、タンパク質の機能解析が効率的に推進されることが期待される。

このプロジェクトを進めるにあたり、サイバーメディアセンター応用情報システム研究部門の伊達進教授、高性能計算・データ分析融合基盤協働研究所の曾我隆特任准教授に大変お世話になりました。感謝の意を表します。

参考文献

1. Jumper, J. *et al.* Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).
2. Bryant, P., Pozzati, G. & Elofsson, A. Improved prediction of protein-protein interactions using AlphaFold2. *Nat Commun* **13**, 1265 (2022).
3. Shinichi, K. *et al.* In silico screening by AlphaFold2 program revealed the potential binding partners of nuage-localizing proteins and piRNA-related proteins. *eLife* **13**, (2025).
4. Hawman, D. W. & Feldmann, H. Crimean–Congo haemorrhagic fever virus. *Nat Rev Microbiol* **21**, 463–477 (2023).
5. Guo, Y. *et al.* Crimean–Congo hemorrhagic fever virus nucleoprotein reveals endonuclease activity in bunyaviruses. *Proceedings of the National Academy of Sciences* **109**, 5046–5051 (2012).
6. Abramson, J. *et al.* Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature* **630**, 493–500 (2024).
7. Xue, L. C., Rodrigues, J. P., Kastitis, P. L., Bonvin, A. M. & Vangone, A. PRODIGY: a web server for predicting the binding affinity of protein-protein complexes. *Bioinformatics* **32**, 3676–3678 (2016).