

特集「AIにより加速する生体分子デザイン」によせて タンパク質分子デザインの進歩を振り返る

中村 春木

(大阪大学蛋白質研究所 名誉教授)

タンパク質デザインの源流とも思われる活動として、1990年9月末から2週間にわたり行われたドイツ・ハイデルベルクの欧州分子生物学研究所 (European Molecular Biology Laboratory; EMBL) における ProDes 90 というタンパク質デザインのワークショップ¹⁾を紹介する。主催者は C. Sander と G. Vriend であり、参加者 21 名が 3-4 名ずつ 6 つのグループに分かれ、J. Moult (米国)、A. Tramontano (イタリア)、A. Lesk (英国)、A. V. Finkelstein (ロシア)、T. Hubbard (英国) と筆者 (日本) がグループリーダーを務め (括弧内は当時の所属)、各々テーマを決めてタンパク質デザインを行った。これらの研究者の多くがその後の構造バイオインフォマティクス分野の研究を牽引した。

ここで、当時のレポートを基に具体的な設計手順を記載しておく。

①天然タンパク質中の進化的に近縁でないが構造型が類似している Superfold のテンプレートを選択する、②天然タンパク質におけるアミノ酸配列と立体構造の関係性を調べる、③主鎖立体構造を構築する、④ペプチド鎖上にアミノ酸を並べる、⑤上記で選んだアミノ酸配列を評価する、⑥側鎖構造も加えた立体構造を構築し最適化する、⑦様々な指標によってタンパク質らしさを評価する、⑧より良い評価が得られるまで上記の③～⑦を繰り返す。

具体的な Superfold としては、 α バンドル、 β サンドイッチ、抗体、mini-NAD 結合モチーフ、mini-MHC クラス I 抗原 headpiece が考えられた。この設計手順は、当時の最先端のコンピューターとコンピューターグラフィックスを用いてなされてはいたが、多くが手作業的に行われた。上記の設計タンパク質は、その後ペプチド合成などがなされたものもあったが、そのままでは設計どおりに fold するものは得られなかった。

ProDes 90 で実施された設計手順は、現在では AI やエネルギー計算によって客観的・網羅的で精度もはるかに向上してはいるが、本質的には大きく変わっていない。決定的に異なるのは情報量であり、当時の蛋白質立体構造データバンク (後述する Protein Data Bank; PDB) にはわずか 500 程度の立体構造しか登録されておらず、細菌などのゲノム配列も未解読の状況であった。

2000 年以降には、ProDes 90 主催者の C. Sander は構造ゲノム科学の旗振り役を務め、筆者も日本国内でタンパク 3000 プロジェクトの振興に関わり、世界中でタンパク質立体構造解析手法の進展とその結果としての立体構造データの急増がなされた (現在は 23 万件超)。PDB はその後に筆者も設立に関わった国際組織 wwPDB

(Worldwide PDB) となった²⁾。米国で wwPDB の運営を行ってきた H. M. Berman 教授は、2024 年のノーベル化学賞に AlphaFold2 とタンパク質デザインが対象となったことについて、正確性と網羅性がある PDB データと透明性が高く、IT 技術をいち早く取り入れてきた wwPDB の果たしてきた役割を強調している³⁾。

タンパク質の立体構造予測を行う CASP (critical assessment of protein structure prediction) コンテストは 1994 年から ProDes 90 のグループリダーであった J. Moult と T. Hubbard によって創設され、前者は開始以来今に至るも CASP の主催者であり、AlphaFold も CASP で生まれてきた。ちなみに、ProDes 90 主催者の C. Sander は、この AlphaFold 開発の端緒となるゲノム配列から得られる共進化情報に基づくアミノ酸残基間コンタクト予測法 (間接相関を省いて直接相関のみを抽出) によるタンパク質の高精度立体構造予測⁴⁾に寄与しており、膨大なゲノム配列情報も必須であったことが理解される。

一方、タンパク質の分子設計は、タンパク質立体構造予測の表裏一体としての逆問題であり、現在では D. Baker らによる ProteinMPNN も開発・公開されている。しかし、タンパク質立体構造予測では天然タンパク質の立体構造を予測する一方、タンパク質デザインは逆問題ではあるものの、相違点として構造のユニークさの問題がある。すなわち、天然タンパク質の多くは一意的な立体構造をとるが、設計タンパク質の多くは核磁気共鳴 (nuclear magnetic resonance ; NMR)、円偏光二色性 (circular dichroism ; CD) などの分光学的には設計どおりの二次構造を含むものの結晶化ができず複数の構造の混合状態になってしまう問題があり⁵⁾、目的とした立体構造だけが安定で他の立体構造が不安定となるための“ネガティブな設計”が必要とされる⁶⁾。具体的には、その後の S. Honda⁷⁾や D. Baker らの研究から、ある特定の構造のみを自由エネルギー的に安定化させる設計指針が考えられる。

以上のように、タンパク質デザインは単に AI の活用によってのみ実現されたのではなく、タンパク質構造情報とゲノム配列情報の集積と整備、およびデザインの失敗から学んだ新たなアルゴリズム、手法の開発と国際的な競争・連携によるものと言えよう。

●文 献

- 1) Sander C, Vriend G, Bazan F et al : *Proteins*. **12** : 105-110, 1992
- 2) 中村春木, 栗栖源嗣 : *生物物理*. **58** : 71-77, 2018
- 3) Callaway E : *Nature*. **634** : 1028-1029, 2024
- 4) Marks DS, Colwell LJ, Sheridan R et al : *PLoS One*. **6** : e28766, 2011
- 5) Tanaka T, Kuroda Y, Kimura H et al : *Protein Eng.* **7** : 969-976, 1994
- 6) Furukawa K, Oda M, Nakamura H : *Proc Natl Acad Sci U S A*. **93** : 13583-13588, 1996
- 7) Honda S, Akiba T, Kato YS et al : *J Am Chem Soc*. **130** : 15327-15331, 2008