

特集「AIにより加速する生体分子デザイン」によせて AIによる生体分子デザインの動向

坪山 幸太郎

(東京大学生産技術研究所生体分子設計工学分野)

タンパク質は、生体にとって必須の分子であり、多様かつ精緻な生体機能を担っている。生体内の天然タンパク質が担っている機能の例として、太陽エネルギーを化学エネルギーに変換すること(ルビスコなど)、小さな分子や光を超高感度に検出すること(嗅覚受容体やロドプシン)、pH勾配を化学エネルギーへと変換すること(ATP合成酵素)、化学エネルギーを仕事へ高効率に変換すること(アクチンおよびミオシン)などが挙げられる。また、RNAなどの核酸も、タンパク質へ翻訳される情報担体としてだけでなく、タンパク質の合成やスプライシングなどに酵素的に働くことが示されている。過去数百年の技術の進歩にもかかわらず、人間がつくった機械は、ナノスケールの生体分子の精密さや効率の良さに太刀打ちできていない。ナノマシンとして多彩で高性能な機能を発揮できる生体分子を自在に設計できるようになれば、医学・工学分野における様々な技術的課題を解決する可能性を秘めている。

特にタンパク質に注目すると、驚くべきことに、30-40億年にわたる進化の過程で生命は、タンパク質の可能性を余すことなく使いこなしているとはいえない。200アミノ酸から成る典型的な長さを持つタンパク質について、あり得る配列数は 20^{200} (それぞれの位置について、20種類のアミノ酸があり得るため)であるが、現存する生物が利用しているタンパク質の数はわずか 10^{12-16} 個程度である。また、自然界に存在するタンパク質は進化により形成されるため、全配列空間に一樣に広がっているわけではなく、それぞれのファミリーがある程度かたまっている。つまり、アミノ酸配列空間のごくわずかな領域を進化は探索したにすぎず、これまでの進化の過程で試されていない可能性の高い、残りの広大なアミノ酸配列空間が、新奇人工タンパク質設計の舞台となる。したがって、進化の過程はその探索の助けにはなるが、絶対的な指標とは言えない。そのため、進化の過程から得られる情報に加えて、タンパク質構造、性質、機能に関する生物物理学における原理を理解することによって、“新奇タンパク質”を人工的に設計することができるはずである。

これまでも実際に、様々な新奇タンパク質が設計されてきている。例えば、新型コロナウイルスの感染を強力に阻害できるブロッカーのような人工タンパク質、全く異なる基質を利用する発光タンパク質、イオン濃度やpHにより構造などが変化するスイッチのようなタンパク質などが挙げられる。しかし、その設計は現時点では“運と職人技”に依存していることが、医学・工学分野における人工タンパク質の

より積極的な応用の妨げになっている。その理由は、タンパク質のアミノ酸配列-構造-機能という“タンパク質の基本原則 (Anfinsen のドグマ)”を、現時点では、正確に理解できていないためである。Anfinsen のドグマは約半世紀前に提案されたものの、タンパク質のアミノ酸配列空間があまりにも広大であるため、その全容はいまだに明らかではない。核酸についても、同様に、どのような配列がどのような構造、機能を示すのかという全容は明らかではない。

その全容の解明の鍵を握っているのは、AI を含む機械学習モデルを構築するための情報科学であろう。例えば、米国 Alphabet 傘下の Google DeepMind 社が開発した AlphaFold は、約半世紀の課題となっていたタンパク質構造の正確な予測を可能にし、世界中を驚嘆させた。このような情報科学を駆使することが、有用な生体分子の設計には必須になると考えられる。ちなみに、2024 年のノーベル化学賞は、人工タンパク質の大家である David Baker 氏、AlphaFold 開発元の Google DeepMind 社の Demis Hassabis 氏、John Jumper 氏の 3 名へ授与されている。

今回、特集「AI により加速する生体分子デザイン」について執筆を依頼した先生方のご専門は、生体分子デザインの基盤となる AI や機械学習技術から、その技術を使った生体分子デザインの実践例まで広く網羅するように工夫した。読者の方々が「このようなことまで、計算機でデザインできるようになったのか!」と驚かれることを期待している。

最後に筆者の私見を申し上げると、生物学は現在大きな過渡期にあると感じている。仮説を基に数個のタンパク質・核酸の性質や機能を検証する研究から、大規模データの取得・解析から始まるデータ駆動型の研究へと移行しつつある。このような新たな生物学分野には、大規模データ取得とその解析技術の双方が必要となる。情報科学の進歩に伴い、データ解析技術は飛躍的に向上しているものの、一方で、大規模データの取得が困難な領域においては、データの生成が全体の律速段階となってしまっている。したがって、次世代の生物学分野を切り拓くなかで、情報科学技術はもちろん、生体分子の大規模データの取得を可能とする wet 実験の重要性が今後更に増すと考えられる。大規模データ取得により、天然生体分子の性質や機能だけでなく、全く新奇な生体分子についてもより正確に構造、性質や機能を予測できるモデルの構築につながるであろう。そのようなモデルを用いることによってよりデザインされた人工生体分子を活用することで、多様なデータ取得が更に加速し、より高精度なデザインが可能となる、というようなポジティブフィードバックを達成することが期待される。本特集がそのようなポジティブフィードバックの第一歩となれば、幸甚である。