

AlphaFold Shock

——構造生物学の大激震

永田和宏

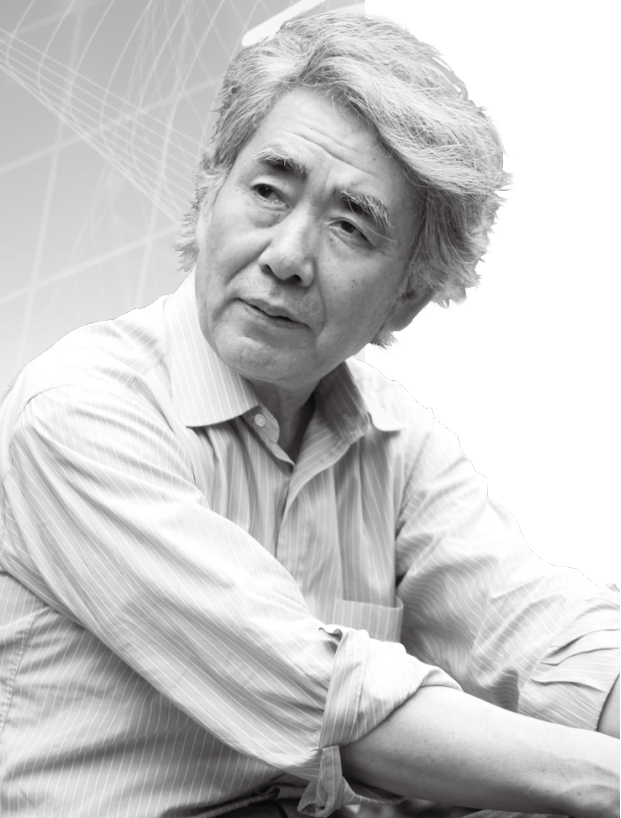
2021年という年は、私たち生命科学の研究に携わっている者にとって、後年、あれが大きな転換点だったと思い出されるような特別な年になるだろう、という予感がある。

生命活動の根幹を担っているのはタンパク質である。一般には「生命」というと、すぐさまDNAとかゲノムとかいった言葉を思い出す方が多いだろうが、それらは基本、タンパク質を作るための情報を格納しているものであり、細胞の中で実際に働いているのはタンパク質である。DNA上には約2万数千種類のタンパク質の情報がコードされているが、実際にはそれらを切ったり貼ったりして、私たちのなかでは10万種ほどのタンパク質が働いていると考えられている。

遺伝情報に従って、20種類のアミノ酸を一列に並べたものがポリペプチドと呼ばれるが、これは単なるヒモであり、タンパク質としての機能を果たすことができない。このヒモが正しく折り畳まれて、空間的な構造を作ったものがタンパク質なのである。二つとして同じ構造を持ったタンパク質は存在しない。この構造の特異性が、タンパク質の個々の機能を規定する基盤となっている。生命活動を、それを担うタンパク質の機能を、タンパク質の構造をもとに理解し説明しようとするのが構造生命科学であるが、そのためにはタンパク質の構造がわかっていることが必須である。

それでは、どのようにして構造は作られるのか。約50年前、C. Anfinsenは、いったん構造を壊したタンパク質を、ある条件下に置くと、再び正しい構造を取って機能を発揮できることを示し、ノーベル化学賞を受賞した。このことは、タンパク質はアミノ酸の並び方さえ決まれば、もっとも安定な形に自然に折り畳まれていくことを示唆している。それなら、熱力学的な計算によって、一次構造から高次構造が求められるはずである。コンピューターを使って、(専門的には分子動力学計算というが) ガチガチ計算をさせればすべてのタンパク質の構造は解けるはずである。

誰もがこう考える。しかし、実際にはこれはほとんどない計算量であり、スーパーコンピューターの京や



プロフィール

永田和宏 (ながた かずひろ)

JT生命誌研究館館長、京都大学名誉教授、

京都産業大学名誉教授

専門：細胞生物学

富岳を使っても、できる計算ではなかった。

タンパク質の構造解明への挑戦は、先のAnfinsenの研究と同時期に始まっていた。ミオグロビンやヘモグロビンを結晶化して、それにX線を当て、その回折像から計算によって構造を求めるといふものである。これらの研究もノーベル賞に輝いたが、それ以来構造研究は、端的に言って、まずは結晶化との戦いであった。

種々の条件を試して結晶化を試みる。それでも成功するとは限らない。一つの構造を解くのに、二年、三年かかるのはザラであった。私たちが発見したタンパク質の場合も、構造生物学の専門家と共同研究を組み、二年近くかかってやっと構造に辿り着いたものだ。これなどまだいい方で、どうしても結晶ができず、遂にギブアップというケースは数限りなくあったはずだ。

近年はこの方法よりずっと簡便で効率よくタンパク質の構造を求められる手段として、クライオ電子顕微鏡による方法が開発されて一大転機を迎えたが、このクライオ電子顕微鏡は高額であり、どこでも揃えられるものでもない。かつて、日本全体で4台しかない時期に、中国では一研究所で4台持っているなどと聞き、大きな衝撃を受けたものであった。

そのような状況のなかで、今年7月、ScienceとNatureという二つのトップジャーナルに、AIを駆使した、RoseTTAFoldとAlphaFold2という二つの方法によって、タンパク質の構造をいともたやすく求める方法が報告されたのである。ネット上に公開されているプログラムにアミノ酸の一次配列情報を入れるだけで、誰でもすぐに高い精度で構造を得ることができる。

これには驚愕という言葉しか浮かばなかった。大変な時間と金をかけて、まさにコツコツとやっていたこれまでの努力と労力は何だったのだ。もちろん科学の進歩によって得られたブレイクスルーであり、科学の発展のためには貴重な一歩であるには違いないが、どこかに人の営みの徒労感とでもいったものを感じさせもしたものだ。

構造を決めるためにだけ費やされてきた膨大な労働と人手をスキップして、ネット上で簡単に得られた構造情報を、生命活動そのものにぶつけて生命機能を構

造の側面から理解できるのは、素晴らしい時間の節約であり、圧倒的な研究速度の加速である。構造を決定することが大きな目標であり、成果であった時代は去り、文字通り構造は、機能理解に奉仕するものとして誰にも初めから与えられる情報になろうとしている。

そんな時代に居合わせたことを、思ってもみない幸運だと思いつつ、しかし、若干の危惧も覚えるのである。それはAIを駆使したコンピューターの計算によって教えられるタンパク質の構造が、本当に正しいとどのようにして確認するのかという問題である。構造既知のタンパク質と、この新しい方法で求めた構造を比較してみると、驚くほど正確であり、これまでのところ、この方法の信頼性は担保されている。しかし、だからと言って、それが他の未知のタンパク質すべてに当てはまると考えてしまっているのか。

専門家に聞くとところによると、AlphaFold2は、構造既知のタンパク質をdeep learningに利用するだけでなく、自らが計算によって解いた構造そのものをも参照しつつ、さらに次のタンパク質の構造決定に利用するのだという。その過程で小さな誤差が蓄積していつ、ついには本物とはまったく違った構造を予測してしまうといった危険性はないのだろうか。

コンピューターが出した構造を、実際の実験によって検証しつつ、次のプロセスに進むのが科学的には望ましい姿だが、X線結晶解析を用いるにせよ、クライオ電子顕微鏡を用いるにせよ、単にコンピューターの結果が正しかったということを証明するためだけに、労多くして得るものの少ないそんな実験に取り組もうと考える研究者はまずいないだろう。

構造決定という最大の難所をいとも易々とスキップし、すぐさま生命活動の本質にアプローチできるようになるという、夢のような方法を手に入れたのが2021年という年であったのかもしれない。一方でそれが、一つ一つの構造を実験的に確認する作業をスキップしてなされることに、一抹の不安を感じることも事実なのである。これからの生命科学、構造生命科学においては、実験計画やロジックの組み方において、劇的な変化が予想されるが、その動向から目が離せない。