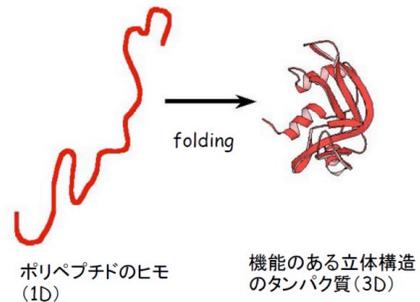


革命：タンパク質の立体構造が計算だけでわかる

私が生きている間は実現するまい、と思っていた夢のような科学の発見が最近、発表された。

地球の生命は、遺伝情報の伝達やコピーには DNA (RNA) を採用し、個体の活動にはタンパク質を採用している。DNA は、4 種のヌクレオチドという単位分子があれこれの順番で直鎖状につながって意味のある文章の役割を果たす分子である。分岐や枝がない文字列は、コピーや操作が簡単なのである (パソコンと同じ)。DNA の情報をもとに合成されるのは 20 種のアミノ酸があれこれの順番で直鎖状につながったポリペプチドのヒモである。ここまでは、1次元(1D)のヒモの世界である。ヒモのままではなにもできない、情報はあれども機能・行動がない。そこでポリペプチドのヒモは誰の助けも借りずに、みずから、曲がりくねり折りたたみ (folding)、複雑な 3D 立体構造を形成する (右図)。3D をまたヒモ (1D) にして再び 3D にすることもできる。その立体構造によってタンパク質は機能・行動する分子となる。タンパク質が違えば (=アミノ酸配列が違えば) 違った立体構造となる。ヒトの体の中では、約 5 万種の立体構造の異なったタンパク質が存在し、からだのほとんどすべての作業をおこなっている。



約 50 年前、Anfinsen により、アミノ酸配列が決まれば立体構造も決まるという実例が示された。x が決まれば y も決まる、これは数学的に言えば関数である。

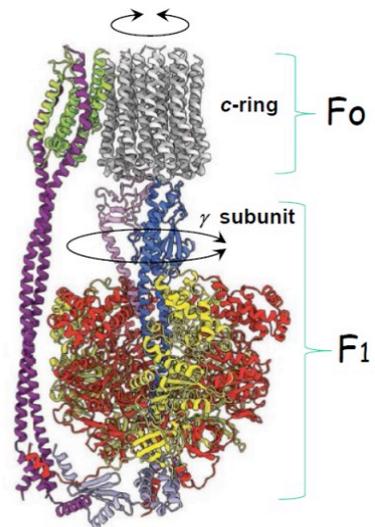
$$y = f(x) \text{ ただし } y: \text{立体構造、} f: \text{folding 関数、} x: \text{アミノ酸配列}$$

今までに数十万の x と y のセットが知られている。しかし、f はわからないままだった。問題(x)と答え(y)がわかっているのに、解き方(f)がわからないのである。ヒントはあった。だれの助けも借りずに自発的に folding できるのだから、これはエネルギー的に (熱力学的に、統計力学的に) もっとも安定な存在状態としてその立体構造に落ち着くのだ。それならポリペプチドのエネルギー最低の存在状態をもとめる計算でよいはず、その計算が f である。というわけで、このノーベル賞確実の課題に、我こそはと、腕に (頭か) 自信のある俊英が挑戦した。しかし、f の解明は超難問だった。この半世紀、満足に成功した者はいなかった。

fがわからないゆえに、バイオ研究の世界は多大な才能と労力の奉仕を強いられた。知りたいのは、xではなくγ、すなわち生命活動のあらゆる場面に登場する機能するタンパク質それぞれの立体構造である。それは、一つ一つ実験で決めなければならない。そのためにはタンパク質を結晶化する必要があるが、これまたタンパク質の一つ一つについて、成功の保証のない手探りのほとんど運任せの長期の作業である。これに研究人生の大半をささげる者も多数であった。私の場合も、ATP合成酵素の結晶化に幾多の年月を費やしたが、ついに結晶を得ることはなかった。これに従事した研究員には成果がなく気の毒な事だった。この酵素の半分の部分構造の結晶解析に成功した者はノーベル賞に輝いたのだが。

立体構造決定の難しさは、この10年で大いに緩和された。それは低温電子顕微鏡による構造決定である。もともと電子顕微鏡は原子レベルの分解能がある。ただ、電子線により構造が破壊される、超拡大した場合に視野がどうしてもずれる動いてしまう、タンパク質はコントラストが弱い、したがってノイズが大きい、などの問題があって、特殊な繰り返し構造のあるものを除いては、ごくおおざっぱな構造しかわからなかった。それを、超低温で電子線障害を最少にして、1電子検出のアレイを開発し高速で多数の連続撮影をおこなって視野ドリフトを補正し、何十万個のタンパク質の像を計算機で向きをそろえて積算しノイズを除く、などの工夫を重ねてついに結晶解析に匹敵する高分解能の立体構造を得ることに成功した研究グループが現れ、瞬く間にこの分野を席卷した。私たちを20年間悩ましたATP合成酵素の構造決定も、低温電子顕微鏡により、あっという間に構造が決定されてしまった(右図)。

回転モーター ATP合成酵素の全貌が低温電子顕微鏡ではじめて見えた



c-ringとγ subunitが回転する

先ほどのfの話に戻ると、fは何か、実は早くから見当はついていたのである。それは分子動力学(MD, molecular dynamics)つまりニュートン力学による計算であって、1つ1つの原子に化学結合の拘束と原子間力と電荷を与えてタンパク質の全原子と付近の水分子(全部で数万原子以上)の運動をフェムト秒(1000兆分の1秒)刻みで計算し追跡すればいい。多くの星がお互いに重力を及ぼしながら運動する天文学の多体問題と似て、これは解析的(一般解)に解けないので、数値をいれて計算を繰り返すしかない。ただ、生物的に意味のある時間は、大体マイクロ秒からミリ秒なので、この計算を1000万回以上繰り返す必要がある。

これはスパコン京だろうが富岳だろうが、できる計算ではない。だから、みんな粗視化と言って変数を桁違いに減らす操作を行っていた。当然のこと、結果はあまりあてにならない。

ところが10年前くらいに、これに真正面から挑む研究者が現れた。計算機を使って株で大富豪になった米国の計算科学者 D. Shaw である。彼は私費 1000 億円以上を費やして研究所を建てて、優秀な計算科学者を集めて分子動力学に特化した CPU チップを作製したタンパク質構造予測専用の計算機 Anton を組み上げた。Anton による手抜きなしの分子動力学計算の威力は相当なもので、分子量 1 万程度のタンパク質ならアミノ酸配列から立体構造を正確に予測できた。Folding の中間状態の構造もわかる。分子量がさらに大きいものでも、初期構造を与えると、構造の揺らぎ、低分子化合物との結合などを予測できた。科学にとっての問題は、この計算は Anton しかできず、Anton は Shaw の私物であり、得られた情報や計算機の使用は、外部に対して閉ざされていることである。この分野は大きな実用的な需要がある。独占的な新薬の開発である。Shaw は米国の薬会社と 1000 億円超の契約を結んでいるようで、外部のアクセスを遮断する理由がある。科学界が今まで経験しなかった資本主義的な科学知識の独占である。

しかし、状況を一変させる発見がつい最近、2つ報告された。AI（人工知能）による3D構造予測である。この場合、AIとは要するに既知のxとyの膨大なデータベースからの類推である。MDはもういらぬ（ところどころで補助的に使っているかもしれない）。

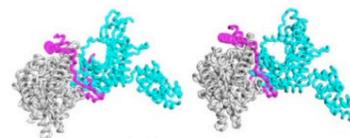
まず、7月15日、米国ワシントン大学のグループは RoseTTAFold（ロゼッタフォールド）を開発したと報告した（Baek, et al, Science, July 15, 2021）。目を見張ったのは、本当にアミノ酸配列だけから立体構造を予測できることを、多くの例で示していることである。もっと驚いたことに、2つ（あるいは3つ）のタンパク質のアミノ酸配列から両者（あるいは3者）の結合した複合体の構造までほぼ正確に予測した（右図）。一本のポリペプチドの鎖に切れ目を入れた、として難なく計算できるらしい。

この驚きがさめやらぬうちに、今度は Google 傘下の英国の DeepMind 社が AlphaFold2 を発表した

（Tunyasuvunakool et al, Nature, July 22 on-line）。これでまた仰天。AlphaFold2 は 2020 年の国際構造予測コ

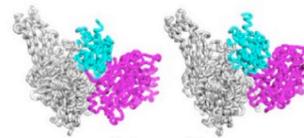
AI の驚くべきパワー

tRNA-dependent amidotransferase



TM-score: 89

4-hydroxybenzoyl-CoA reductase



TM-score: 90

左、実験

右、予測

アミノ酸配列だけから3種のタンパク質の複合体の構造を予測

ンテストでもっとも実際の構造に近い予測をして1位となった(2位は上記のロゼッタグループだった)。その後、AlphaFold2 はなりを潜めていたが、実はその間に、ヒトのタンパク質5万種のほとんど全て(!!)を構造予測していたのだ。そうしたら、5万のうち、予測構造に確信のもてるものが約3万、さらにそのうちもう間違いないと言える構造が約2万予測できたという(<http://alphafold.ebi.ac.uk>)。そして、構造がはっきりしないもののうち相当部分は本当に構造のない天然変性部分、あるいは他のタンパク質と結合して複合体を作った時に初めて構造を持つ部分ではないか、という。AlphaFold2 は、ヒトに限らず大腸菌、カビ、ハエ、ネズミ、コメ、など20種の生物の利用できるアミノ酸配列30万種を全部、立体構造に変換している。

RoseTTAFold も AlphaFold2 とほぼ同等の性能で、したがって、膜タンパク質でも支障なく構造が求まる。さらにアミノ酸配列の似たタンパク質がない場合でも予測ができた。AI法のいいところは、だれでもアミノ酸配列を入力すればすぐに予測結果が得られることである。アミノ酸配列が3D構造を決定するという Anfinsen の発見以来50年、バイオ研究の超難問がついに解かれた・・・ノーベル賞確実の偉業だと思う。

ただ、正攻法の分子動力学(MD)は不要か、ということそうではない。AIで求めた立体構造は folding の完成形であって、どうやって folding するか、には全く答えない。そこは MD が必要である。また、完成した立体構造も実は構造がゆらいでいる。その揺らぎが生物学的に重要なことはわかっている。それも MD でしか計算できない。

今、私が考えるのは、 f の逆関数 g を求める研究である。できれば、本当の「タンパク質工学」が開かれる(一時期、工学系のバイオ研究者がタンパク質工学を唱えたが、設計もできないのに工学と言えるか、と私はこの誇大宣伝を笑っていた)。タンパク質は万能であるので、それを設計し動かせれば、いまの化学工業は大きな変革を経験するだろう。

吉田賢右

だれかにすぐに始めてほしい研究

1年以内にできたらひよっとすると来年の3人目のノーベル賞かも

$y = f(x)$ は解けた、ならば

逆関数 $x = g(y)$ を解いて欲しい

Y (立体構造)を与えて x (アミノ酸配列)を求める

y: 立体構造
f: folding 関数
x: アミノ酸配列
g: 逆folding 関数

つまり、デザインした立体構造にfoldingするアミノ酸配列を見出すこと

あるタンパク質によく結合する別のタンパク質の3D構造をデザインし、そのアミノ酸配列を見出すこと

本当の「タンパク質工学」の始まり となる