

## 巻頭言「幼年期の終わり」

タンパク質動態研究所所長 遠藤斗志也

アーサー・C・クラークの「(地球) 幼年期の終わり」は、人類を圧倒的に凌駕する知的生命との出会いを描いた SF の古典的名作だ。「上主 (Overlord)」と呼ばれる彼らの出現により、地球上から無智、疾病、貧困、恐怖は事実上姿を消し、人類ははじめてほぼ完全な状態で安全と平和と繁栄を手に入れる。圧倒的な力を持つ彼らの前では国家間の戦争は無意味であり、困っている人間がなくなれば、物を盗む意味もなくなる。潜在的に犯罪者の素質を持っている人間も上主の監視から逃れることができないことを皆が知っている。誰もが好きな職業に就き、労働時間は短くなり、あまった余暇では好きなだけ教育を受けられる。

ただ、この与えられた微温的な至福の世界は、ある種の喪失感を伴っている。あらゆる葛藤、闘争が消滅したことで、優れた創作芸術は途絶えてしまう。さらに科学も斜陽を迎える。「好奇心は残っている。好奇心にふける余暇もある。だが基礎科学の研究に捧げる情熱がなくなったのである。上主たちが何時代も前に発見してしまった秘密をさぐるために一生をついやすなど、むなしく思えるだけだった。」

そう、これはもはや遠い未来の話ではない。既に現実起こりつつあることなのかもしれない。ただし「上主」は遠い宇宙から来た知的生命ではなく、AI である。

ご存じのように 7 月に生命科学分野に激震が走った。かつて AI を使った囲碁プログラム AlphaGo をつくって囲碁の世界チャンピオンを打ち負かしてしまった Google 傘下の英国 DeepMind 社が、今度は AI を使ってきわめて精度の高いタンパク質の立体構造予測プログラム「AlphaFold2」を作ってしまったのである。ほぼ同時期に米国の David Baker らのグループも同様のプログラムを発表した。DeepMind 社はソースコードを公開して誰でも「AlphaFold2」を利用できるようにしただけでなく、簡易版を web 上に公開し、アミノ酸配列を投げれば 20 分とかで立体構造の原子座標が返ってくるようになった。すでにデータベース上にはヒト、酵母、ショウジョウバエ、シロイヌナズナ、大腸菌などモデル生物の予測構造が 7 月末の時点で 35 万種類も登録され、誰でも閲覧することができる。このあたりのことは、本年報でもすでに吉田さんと岸川さんが書かれている。

それにしても、AI がこんな短期間で生命科学最大の難問といってもいい、タンパク質の立体構造予測問題を解決してしまうとは思わなかった。プロテインデータバンクには X 線、NMR、クライオ電顕等で決定された 18 万を越えるタンパク質の立体構造が登録されているが、それでもそれが AI のトレーニングデータとして十分だとは思いつかなかった。そこで思い出されるのは、かつてこの国で動いていた「タンパク 3000」なる総額 500 億円以上が費やされた大型プロジェクトのことである (2002~2006)。タンパク質の折れたたみパターン (フォールド) 数は限られている (Cyrus Chothia により 1000 種類程度と予想された) ので、1 万~1 万 2 千種類のファミリーを代表するタンパク

質の構造を決定すればすべてのタンパク質の立体構造を推定できるということで、わが国はその 1/3 にあたる 3000 種類のタンパク質の立体構造決定をめざすという触れ込みであった (米国でも同様のプロジェクトが走っていた)。そして実際に 2700 種類以上のタンパク質の立体構造が決定され、そのうち 900 以上は新規構造 (既知の構造と 30% 以下の相同性) であった。一方でプロジェクト自体が立体構造決定の数値目標を設定していたことから、構造決定されたタンパク質には機能不明のタンパク質が数多く含まれていたし、各国のプロジェクトが終了しても結局新規にタンパク質の立体構造を高い精度で予測することはできなかった。しかし、このとき決定された数多くの新規構造が今回の AlphaFold2 のトレーニングデータに役立ったことは、疑う余地がない。タンパク 3000 の成果は、終了後 15 年を経てプロジェクト自体が皆の記憶の彼方に遠のいたころ、AlphaFold2 という形で実を結んだということだろうか。

短期的には、AlphaFold2 は構造生物学の研究者に恩恵をもたらすだろう。少なくとも正解に近い構造があるわけだから、クライオ電顕の単粒子解析における粒子画像の分類、X 線構造解析における位相決定のための分子置換法、結晶化しやすいコンストラクト作成などを助け、構造決定の効率をあげるだろう。それ以上にメリットがあるのは、一般の生命科学研究者だ。これまでのように構造を手にした一部の研究者のみが、機能研究の美味しいところを独占してしまうということはなくなる。精度の高い予測構造に基づいて、ただちに変異体解析や機能解析ができるようになる。貴重なタンパク質の構造情報を一般の生命学者に広く開放することにより、真の意味で構造情報入手の格差がなくなる。

一方、中長期的な展望は、問題をはらむ。現時点での AlphaFold2 の構造予測には苦手なものがいろいろあるようだが、そうした問題は早晚解決できそうな気がする。タンパク質複合体、タンパク質と核酸や小分子との複合体の予測やサーチも時間の問題だろう。吉田さんが書かれている、立体構造からのアミノ酸配列の予測 (真の意味でのプロテインデザイン) も、AI が予測と逆予測を繰り返していけば、正しい答えに到達してしまいうさだ。そうなったときに、構造生物学はどこに向かうのだろうか。現在の構造生物学研究は時間も労力もコストもかかりすぎる。AI があつという間に立体構造をかなりの精度で予測できるようになったら、膨大な時間と労力とコストをかけて精密な立体構造を決める意義があるタンパク質のターゲットはかなり限られてくるはずだ。囲碁や将棋の世界では、AI が人間に勝ったところでリーグ戦には参加できないわけで、むしろ人間が AI を使ってトレーニングすれば良いという棲み分けができたが、生命科学の世界ではそうはいかない。Science 誌も「In the short term, it will boost structure determination efforts, . . . And over time it will also slowly replace experimental structural determination efforts」と書いている。

「地球幼年期の終わり」には続きがある。上主の目的は単に人類にユートピアを与えることではなかった。上主のさらに上の存在である「主上心 (Overmind)」は、人類の子供たちを自分たちと一体化させた存在に進化させる、「トータルブレイクスルー」を目指していた。進化する人類=子供たちの世代は、これまでの人類=親たちの世代とは全く異質の存在であり、両者は希望や求

めるものを分かち合うことはできない。これまでの人類=親たちの世代は、そして地球という存在そのものが終焉を迎えていく。AIとともに歩いていくことが具体的に見えてきた生命科学の、そして私たちの未来には、そのような終焉が来ないことを望むのだが・・・。

\* \* \* \* \*

2020 年度はコロナの影響で、本研究所としてのアウトリーチ活動はできず、学生の入構制限により、研究に少なからず遅れも生じました。そんな中で、今回は研究所の活動報告に加えて、2020 年度でオフィシャルに本研究所を去られた近藤寿人さん、本研究所 OB の伊藤維昭さん、吉田賢右さん、そして本研究所から大阪大学に転出された若手の岸川淳一さんに、ご自身の研究のこと、あるいはサイエンスに関連して思うこと、などの文章をお寄せいただきました（ご寄稿ありがとうございました）。コロナからプラズモン共鳴、クライオ電顕、AlphaFold2 まで興味深いトピック満載で、インスパイアされるところの多い研究紹介やエッセイかと思います。お楽しみいただければと思います。

(2021.8.13)