

まえがき

ゲノムの情報が公共のデータベースに大量に発表され、誰でも利用できるようになったのは、21世紀になってからである。それ以前にも配列情報の大きなデータベースはあったが、一つの生物体を持つ完全な配列情報は、単なる大量の配列情報以上のものである。ゲノムの配列情報は、一つの生物体を作る設計図であり、自然が長い時間をかけて進化というプロセスによって作ってきたもので、全体として非常に調和のとれたものであるに違いない。そういう意味でゲノムの完全な情報が得られるようになったことは、生物を理解するうえで革命的なことである。しかし、得られた情報を利用してゲノム全体の調和についてしっかりした議論が行われたかという点、残念ながらほとんど行われてこなかったと思う。20世紀には個々の遺伝子、個々のタンパク質の情報しか手に入らなかった。したがって、21世紀に入りゲノムの全配列情報が手に入るようになってきても、多くの研究者がこれまでの延長線上でゲノム情報を考えてきたのもやむを得ないことだったと思う。しかし、21世紀最初の10年が経ち、第3世代シーケンサと言われる桁違いに高性能の配列解析装置が登場した。今こそ生物の原理に基づく（つまりゲノム全体の調和を考慮した）情報解析手法についてしっかり考えるべきではないだろうか。

生物の設計図であるゲノムには、情動的側面と物理的側面の二面性があるのだが、従来の情報解析手法はゲノムの情動的側面に注目しすぎた傾向がある。個々の遺伝子の情報を知るために、配列の類似性検索などのいわば力技で機能の情報を抽出してきた。しかし、大量の配列情報に対して、従来の情報解析手法には限界がある。常に大量の未知配列が発生し、ゲノムの生物学的意味が曖昧になってしまうからである。これに対して、ゲノムの配列には物理的実体があ

るということを本当に理解すれば、それによって生物ゲノム全体の調和について情報解析を行うことが可能である。自然はゲノム（とその産物であるタンパク質集団）の配列情報の物理的性質を巧妙に利用し、ランダムな変異によってより高い秩序を生み出す方法を編み出してきたのである。本書では、生物を理解するために必要なゲノム配列情報の物性分布の解析手法を展開したのが本書である。詳細は本文を読んでいただく必要があるが、自然が生物を進化させるときに行ってきたことのエッセンスは、以下のとおりである。

生物は、一次元の配列情報で設計され、それによって三次元の生物体が実現されるという、実に複雑な構造を持った物質（物体）である。しかも、生物の配列情報はランダムな遺伝子変異の集積だけで進化してきたということがわかっている。本書では、生物の進化や多様化という不思議な現象が、3つの概念で説明できるということを示している。わかりにくいかもしれないが、まとめておきたいと思う。

配列情報に対する遺伝子変異は、的のあるランダム過程によって駆動されている。自然は遺伝子変異というサイコロを振っているのだが、これはフェアなサイコロではない。

生物体を作るタンパク質の構造は、配列の物性分布の粗視化によって大枠が決められている。多くのタンパク質について結晶構造解析によって原子分解能で立体構造が明らかにされているが、構造を作る物理的ユニットは個々のアミノ酸よりはるかに大きな配列断片なのである。

生物ゲノムの遺伝子集団は、すでに遺伝子変異の平衡分布に達している。自然現象を見てみるとあらゆるところで平衡分布が見られるが、その共通的な特徴は平衡状態の周りで分布が非常にロバストだということである。生物、生物種、生物界が示すロバスト性もこの平衡分布によるものである。これらの概念は、いくつかの計算科学的研究によって裏打ちされており、本書の各章で解説されている。

ゲノム系計算科学は、新しい学問分野である。第1章では、この新しい学問分野を定義する疑問について説明した。その疑問は「生物とは何か？」である。これは生物に関する究極の疑問である。一次元の生物情報であるゲノム配列情報については、1000をはるかに超える生物種についてゲノム情報がすでに得ら

れ、10万人を超えるパーソナルゲノムの解析が計画されている。「生物とは何か?」という疑問に答えることなく、きわめて大量の配列情報から生物あるいはヒトの生物学的意味を抽出することはほとんど不可能である。第2章では、ゲノムの配列情報からのアミノ酸配列集団を物理的な考え方で解析することで、進化の重大事件において何が起こったかを示した。このとき、配列の物性分布の粗視化が本質的であることがわかる。第3章は、ランダムな遺伝子変異でなぜ生物という秩序構造が生まれることができたかを議論した。遺伝子変異はランダムだが、完全に無秩序へのランダム過程ではなく、的のあるランダム過程である。第4章は、個別の問題としてアミノ酸配列からの膜タンパク質の予測システムについて詳しく議論した。疎水性と両親媒性というアミノ酸の物性を、粗視化することで高精度の予測システムができること、それをを用いると配列の変異シミュレーションが可能となることを示した。第5章では、ゲノムからの全アミノ酸配列を、最もよくわかっている物性である電荷の分布に変換し、自己相関関数を計算することで、脊椎動物の誕生と同時に一群のDNA結合タンパク質が増えたことを示した。第6章では、タンパク質の活性部位で起こる電子移動について量子力学的な解析を行い、電子移動経路を解明できることを示した。その電子移動経路のキーとなるアミノ酸が、多様な生物の間で完全に保存されていることを示した。第7章は、生物システムに対する統計力学的なアプローチについて議論した。生物システムは、大きな揺らぎに対しても頑丈で安定な振る舞いをするが、それを統計力学によって示した。第8章では、ゲノム科学を学ぶ若い学生や研究者に向けて、生物科学の研究で陥りやすい誤解や偏見について述べた。

本書の内容は、生物および生物ゲノムを、どのように物理的に考えることができるかを議論したもので、ほかの類書には無いものである。できるだけ多くの人たちに、この考え方が伝わっていくことを切に願っている。

2012年12月

編者 美宅成樹