

生命を計算して理解する

Computational Solutions for Complexity of Life

内藤 泰宏

慶應義塾大学環境情報学部准教授

Yasuhiro Naito

Associate Professor, Faculty of Environment and Information Studies, Keio University

現代の生命科学は生命を分子機械と捉え、そのリバースエンジニアリングによって生命現象の動作原理の解明に取り組んでいる。本稿ではまず、生命学者が生命を分子機械と捉えるに至った過程を辿る。その上で、一斉大量測定や1分子レベルの顕微鏡観察といった測定技術の革新を経て、コンピュータ上で生命情報を計算することによって、生命についていかなる理解がもたらされつつあるかを検討する。

Biologists have been investigating general design and operating principles by reverse engineering the molecular machinery of biological systems. Here I illustrate how biologists have been convinced to consider life as molecular machinery first. Then, I discuss what kind of insights about the nature of life can be provided by computation on vast amounts of biological data, produced by innovative technologies such as high-throughput measurements and single-molecule microscopy.

Keywords: バイオインフォマティクス、オミクス科学、1分子計測、細胞シミュレーション

自然科学による解明は、対象の計測から始まる

自然科学の何よりの特徴は、自然を観察・測定し、得られた知見から自然の本性を理解しようとする点にある。自然科学では、対象を測り、定式化し、検証する。測ることのできない対象を、自然科学の枠組みに取り込むことはできない。対象の観察・測定には客観性と再現性が求められる。ここでの客観性とは、観察・測定の結果を他者と共有可能であることを指すといって概ね差し支えない。客観性を確保するために、測定は、測定者の五感の介在を排除した装置を用いておこなわれる場合が多い。再現性と

は、ある観察・測定を繰り返し実施した際に、一定の誤差の範囲内で同じ結果を得ることができることを指す。客観的な観察・測定の結果に再現性が伴うとき、自然科学はそれを事実と判定する。

したがって、自然科学によって自然の本性をどこまで明らかにできるかは、測定装置によって何をどこまで測れるかに大きく依存している。自然科学の黎明期には、レンズの精度向上によって、はるか彼方の天体や、マイクロメートルサイズの微生物を拡大して観察することが可能となり、それぞれの対象に関する理解が一新した。その後現在にいたるまで、ヒトの五感だけでは経験しえない物体や現象を検出

する技術が次々と開発され、人類の経験世界を押し広げている。

また、自然科学がもたらした知識を、人々の暮らしに還元する際にもさまざまな技術が介在している。私たちの身のまわりの物品をひとつずつ手に取ったとき、その多くには、なにがしかの自然科学の成果が応用されている。同時にその多くでは、用いられている自然科学の知見は露わにはなっておらず、技術やデザインによって隠蔽されている。自然科学と技術の相互依存は時とともに深まり、融合して判別が難しくなりつつある。

一方、自然科学の一部の領域は、現在の技術では観測困難な自然の本性に挑戦しようとしている。たとえば、ひも理論など最先端の理論物理学では、観測によってその正しさを厳密に検証できる見込みのない領域で、数学的な合理性を頼りにした検討・議論が繰り返されている。

検証困難であることが予めわかっている領域に踏み込むことは、自然科学からの逸脱を意味するのだろうか？この問いにはさまざまな回答が投げられている。自然科学は目的ではなく手段である。どういった目的のために自然科学を用いるのかによって、回答は異なるだろう。

生命の2つの特性が、生命科学の進展を阻んできた

自然科学の中であって生命科学はとりわけ幼い領域である。なぜか。生命現象は古代より人類の関心を集めてきた対象のひとつであり、神話、宗教、哲学といった知識体系の中で、生命とは何かの説明が試みられてきた。それにも関わらず、生命科学の進展が物理学をはじめとする他領域の後塵を拝している最大の理由は、計測の難しさだろう。

現在、生命現象の本質をなす素過程は、ナノメートルスケールの分子間相互作用であると考えられている。これらの分子を個々に同定し、その性質を調べるための技術が調うのは20世紀以降のことである。古典物理学が成熟期を迎えた19世紀、生物学領域では、生命には非生命にはない特別な力が宿っていると考える生氣説が優勢だったのである。

生命現象の素過程が非常に小さいスケールで起こっていることに加え、生物が「生きている」という事実そのものも、生命の解明を阻む大きな障壁となってきた。「生きている」とは、変化しつづけているということである。どんな生物であれ、生まれてから死ぬまで、たゆみなく変化しつづける。そして、変化する対象の計測は、静止した対象の計測に比べ、格段に難しい。

たとえば、動物の体内で心臓がポンプの役目を果たし血液が全身を循環することは、今日では中学生が学ぶ知識である。この血液循環の“仮説”が初めて提案されたのはようやく17世紀を迎えてからである。解剖の記録は紀元前3000年より古いものもあるので、人類が身体の内部構造の観察を始めてから、心臓をポンプとして血液が循環しているという基本的な事実に気づくまで、4千年以上を要したことになる。動きを止めた死体をいくら丹念に解剖しても、すでに停止して“そこにはない”血液循環に気づくのは困難だったのである。

今日最先端の生命科学を以てしても「生きている」とは何かは解明し尽くされてはおらず、その科学的定義も確立していない。

遺伝と進化の理論が、生命科学の近代化を先導した

生命科学の近代化に中心的な役割を果たした指導原理は、遺伝と進化である。これらは、それぞれメンデルとダーウィンによって19世紀中葉に相次いで最初のアイディアが提案された“若い”概念である。それ以前の生命科学は、博物学的な生物学や実践的な医学として存在し、発展してきた。遺伝と進化という概念が与えられたことによって始めて、地球上のすべての生物は共通の祖先を持つ一つの「系統」であり、あらゆる生命に共通する基本的なメカニズムが存在するという、統合的理解がもたらされる可能性が示されたのである。

プランクが量子のアイディアを発表し、当時すでに完成の域に達しつつあると考えられていた物理学の知識体系がまるごと“古典”へと追いやられる革命が始まった1900年、生命科学の領域では、35年

間忘れ去られていたメンデルの遺伝に関する理論がようやく再発見された。

メンデルは有名なエンドウの交配実験の結果から数理法則を導出した。メンデルが示したのは、生物個体はその機能を裏打ちする設計情報を体内に持っており、その情報が世代間を離散的に伝達していく、というモデルである。子が親に似る遺伝現象は古くから世界中で知られていたが、情報伝達の様式と、情報から機能が発現する機構について、実験結果に基づいて明確な数理モデルが提案されたのはこれが最初だった。

ただし、当時の技術的限界にも阻まれ、メンデルの唱えた遺伝情報は理論上の存在に止まり、それが生物個体のどこに、どんな物質として存在しているかはまったく解らなかった。そもそも1900年当時は、物理学者でさえその大部分が、物質の構成要素としての原子の存在を信じていなかった。

その後「遺伝子」と名付けられたその遺伝情報の実体が、二重らせん構造を持つデオキシリボ核酸(DNA)であることが突きとめられるまでに、さらに50年余を要した。DNAの塩基配列(文字列)としてコード(暗号化)されている遺伝情報を解読する技術が確立したのは1970年代で、まとまった遺伝情報が初めて解読されてから^(注1)、まだ40年すら経過していない。

分子生命科学のセントラルドグマ

遺伝情報がDNAにコードされていることが発見されると、科学者の関心は情報が機能として実体化する過程の解明へと向けられた。ワトソン、クリックらによるDNAの二重らせん構造の発見(1953年)からわずか5年後、クリックが、DNAにコードされた情報は、まずRNAに写しとられ、つづいてRNAの配列情報に基づいて機能分子であるタンパク質が合成されるというシナリオを提案し、自らセントラルドグマ(central dogma、中心教義)と名付けた^(注2)。これまでにいくつかの例外事象が見つかっているが、現在に至るまで、このアイディアは分子生命科学の中心教義として君臨している。

遺伝子の大部分はタンパク質の設計情報をコード

している。タンパク質はそれぞれに固有の機能を持ち、それらが触媒する化学反応の多くは、タンパク質の存在抜きに自然界で進行することはない。古来より、生物には非生物には見られない“何か”が宿っていると考えられてきたが、その“何か”は神秘的な属性ではなく、タンパク質をはじめとする生命分子に由来していたのである。

DNAの二重らせん構造の発見とセントラルドグマの確立によって、生命科学者たちは、DNAを調べれば、細胞の中に存在する生命分子を書き下すことができ、個々の分子の機能を調べ尽くすことで細胞ひいては生物個体全体を理解できるという期待を抱き、分子生物学と呼ばれることになる新領域を切り拓いた。DNA構造の発見から間もなくしてDNA配列決定技術が開発され、タンパク質のアミノ酸配列決定技術の開発がこれにつづいた。また、生物のDNA配列を改変する遺伝子工学技術も飛躍的に進歩した。

これらの技術を用いることで、細胞が自身の遺伝情報を複製し次世代に伝えるメカニズムが明らかになり、また、免疫や発癌といった複雑な生命現象のしくみが解明された。分子生物学は一気に生命科学を牽引する立場に躍りでた。

現代の生命科学は何を解明しているのか

現代の分子生命科学は何をどのように解明しているのだろうか。それは、現代社会が生命科学に求めるものと密接に関係している。生命科学に限らず、現代の自然科学は、公共の財源によって支えられており、その成果を出資者である社会に還元することが強く求められている。

生物は、太陽光から光合成を行うなどして、非生物を糧に活動のためのエネルギーを産みだせる独立栄養生物と、他の生物(あるいはその派生物)を摂取しなければ生きられない従属栄養生物に二分される。ヒトは従属栄養生物であり、他の生物を摂取しなければ生きていけない。つまり、われわれ人類もその食糧も、ことごとく生命科学の対象であり、事実、古来より医学・農学によって探究され、得られた知識は体系化され実践に役立てられてきた。それ

らは今も生命科学と社会の重要な実学上の接点となっている。これに加え、近年は土中の微生物によって完全に分解される生分解性プラスチックなど、生命科学の工業への応用も進展している。

生命科学にとって、社会实践に資する知見とは何か。分子生物学によって、生命は分子を部品とする複雑な機械であるとする「分子機械論」が広く浸透し、現在に至っている。分子機械論に基づく研究は、一言でいえば生命システムのリバースエンジニアリングである。現代の生命科学は、莫大な数の研究者を擁する巨大科学であり、多様な手法を用いて多彩な対象に対する研究が展開されているが、その大部分は「生命システムを構成している分子は何か」「ある生命分子と相互作用する生命分子は何か」「ある生命分子あるいは複数の生命分子からなるネットワークはどのように状態遷移していくか」「分子ネットワークの相互作用の総和は、より巨視的な階層でどのような現象に帰結しているのか」といった生命システムのリバースエンジニアリングの観点に立っている。

分子生物学の黎明期には、未知の生命分子を探しだし、生命システムの部品リストを完成することに研究の重点が置かれていた。近年は、リスト上の分子の相互作用ネットワークやダイナミクスの研究の重要性が認識され「生命システムの動作原理」という表現が頻繁に用いられている。分子機械論的アプローチにより、生命システムを構成する“回路図”が、部分的ながら大量に明らかになり、その一部を制御することも可能になった。分子標的薬などの新世代の薬剤、害虫に強い遺伝子組換え農作物などは、生命のリバースエンジニアリングの典型的な成果といえる。

生命情報解読の加速が、バイオインフォマティクスの確立を促した

遺伝情報はすべて DNA の二重らせん構造にコードされている。遺伝情報の一次産物の大部分はタンパク質であり、DNA の配列情報を遺伝コードと呼ばれる暗号解読表に従って変換したアミノ酸配列を持つ。DNA の塩基配列やタンパク質のアミノ酸配

列を同定する技術を開発すれば、あらゆる生物種に適用することができる。超多様な機能を持つ生物を、分子レベルで単純な要素の集まりとして捉えられる。こうした発想に基づき、20 世紀後半の分子生命科学は飛躍的に進歩した。1990 年代には、DNA 配列解読技術の進歩と、国際協力による研究資源の集中的投下によって、ある生物種のまるごとの遺伝情報（ゲノム）の全解読が実現し、2000 年にはヒトゲノムの解読がほぼ完了しその概要が発表された。

従来の生命科学の研究スタイルは、科学者が着想した作業仮説を実験・観察によって検証する仮説検証型を主流としていた。ゲノム解析技術の進歩とそこから生まれた成果は、まず測定できるものをすべて測ってデータ化し、それを足がかりに有益な知見を発見しようとする新しい研究スタイルをも生み出した。

1970 年代の DNA 配列解読（シーケンシング）技術の黎明期にはすでに情報科学技術が利用されていたが、爆発的に蓄積された生命情報が、科学者がヒューリスティックに知識発見できる容量を超えると、情報解析技術の全面的な活用が研究遂行に不可欠となった。生命科学を補佐する情報科学はバイオインフォマティクス（生命情報科学、bioinformatics）として存在感を増し、独立した学問領域として整備が進んだ。

今世紀に入り、ナノテクノロジーの導入などによって DNA 配列解読の原理そのものが刷新されるとともに、検出器の小型化やイメージングの活用により、超並列の同時大量高速測定が実現した。「次世代シーケンシング」と呼ばれるそれらの技術は、現在もさらなる革新を競っている^[4]。最近の 20 年間で、DNA シーケンシングのコストは 1 万分の 1 以下に低下し、人類が解読した DNA 配列の総量は千倍以上に増加している。

オミクス科学がもたらすデータとその解析

ゲノム (genome) とは、ある生物の持つすべての遺伝情報のセットを指し、遺伝子 (gene) に、総体を表す接尾語 -ome を合わせた造語である。そしてゲノムを対象とする科学をゲノム科学（あるい

はゲノミクス、genomics) と呼ぶ。ゲノム科学の成功に導かれ、細胞内のその他の生命分子に注目し、対象と定めた分子種を網羅的かつ高速に測定する技術が数多く開発された。こうした一斉・大量・高速・網羅的な測定は、ハイスループット (high-throughput) 計測と呼ばれている。

ゲノムに倣い、細胞内のタンパク質 (protein) の総体を指すプロテオーム (proteome)、代謝物質 (metabolite) の総体を指すメタボローム (metabolome)、遺伝子発現に際して DNA から写しとられた RNA の総体を指すトランスクリプトーム (transcriptome) など、さまざまな“ーム”が提唱され、そのそれぞれに対してハイスループット計測技術が開発され、プロテオミクス、メタボロミクス、トランスクリプトミクスといった新しい研究領域が生まれた。それらは総じてオミクス科学と呼ばれている。

オミクス科学では、最先端の物理化学計測技術を基盤に、大量の生命情報が取得・蓄積されている。DNA の核酸塩基配列やタンパク質のアミノ酸配列のような定性情報 (文字列情報) に加えて、どういった分子が、どういった組織・細胞に、どの程度存在しているかという定量情報も大量にもたらされるようになった。

バイオインフォマティクスは、こうした生命情報の量的爆発と多様化に、データの管理・解析手法の提供、アクセシビリティ向上などの諸側面で応えている^[1]。生命情報に関するデータベースは、*Nucleic Acids Research* 誌が運営する *Molecular Biology Database* に収録されているものだけでも 2012 年中に 1500 を超えた。これらが提供するデータの形式は統一されておらず、複数のデータソースを用いて横断的な解析を行うのは必ずしも容易ではない。こうした現状を改めるべく、複数の国際プロジェクトが、オミクスデータの標準化、収集、保存、統合的提供に取り組んでいる。

データへのアクセスと同様に、解析手法へのアクセスも難度を増している。大規模かつ複雑なデータセットを解析するには、適切なアルゴリズムを選択し、必要なソフトウェアを導入し、その上で複数

の解析を連結したパイプラインを構成する必要がある。これらの工程をスムーズに進めるための統合的なソフトウェアプラットフォームの開発も複数の研究プロジェクトによって進められている。

解析手法の試行錯誤も目まぐるしく、新旧さまざまなデータマイニング手法、ネットワーク解析手法など、ありとあらゆる大規模データ解析技術の適用が試みられている。複雑に相互作用する遺伝子ネットワークの構造を明らかにするために、マルコフネットワークやベイジアンネットワークといった確率分布のグラフィカルモデルが用いられ、部位・条件・時間毎の遺伝子発現データセットから条件特異的に発現している要素群を抽出するために、線形混合モデル、行列因子分解といったデジタル信号処理技術が利用されるなどしている。圧縮センシングのように比較的新しい技術の応用例もある。生命情報特有の構造に特化した新たな解析アルゴリズムも毎年多数考案されている。

相関関係による理解と因果関係による理解

ゲノム科学をはじめとするオミクス科学が生命科学の本流に加わったことにより、生命科学による対象の理解の形式も変貌しつつある。理解の形式には、相関関係による理解と因果関係による理解がある。たとえば、キイロショウジョウバエの眼は正常では赤いが、稀に白眼になる変異が見られる。遺伝学研究によって、ある遺伝子に変異が起こると、白眼変異が表れることが突きとめられ、この遺伝子は *white* と名づけられた。これは相関関係による理解である。*white* 遺伝子がコードするタンパク質を解析すると、赤色素を眼の細胞に輸送する機能に関わっていることがわかった。変異によって *white* 遺伝子の機能が損なわれると、色素が眼に運ばれなくなり、白眼となるのである。ここまで解明されると、*white* 遺伝子の変異 (遺伝子型の変化) が白眼という表現型の変化に帰結する過程が因果関係によって理解できたといえる。そこには、*white* 遺伝子の変異がタンパク質の構造にどういった変化をもたらし、どういったメカニズムで色素を輸送する能力が損なわれているのか、という新たな因果関係の問い

が生じているが、因果関係による説明がまったくない状況は脱したといえる。生命科学による生命現象の理解は、遺伝学などがもたらす相関関係による理解が先行し、タンパク質などの生命分子の機能を丹念に同定する生化学などによって、因果関係による理解が付加されていくという流れで進んできた。

ゲノム科学を嚆矢とするオミクス科学の台頭と、バイオインフォマティクスによる知識の抽出は、こうした理解の形式をも変えつつある。オミクスデータに限らず、ビッグデータの解析によってもたらされる知識は、その大部分が統計的な相関関係である。したがって、オミクス科学がもたらす理解のほとんどは相関関係によるものである。それを足がかりに因果関係の理解へと進もうとした途端、ハイスループット計測の適用が困難な、対象ごとに特化した解析が要求される。因果関係の発見も加速しているが、急激に膨張する相関関係の知見の蓄積には到底追いつかない状況である。

また、生命現象の理解を実学に応用し、社会還元するという視点に立てば、相関関係の抽出で充分である場合も多い。社会還元の効率を優先するなら、個別事象の因果関係の解明に注力するよりも、新たな相関関係を広範に抽出の方がより確実に、高い生産性を期待できるのかもしれない。

もうひとつの測定の革命：1分子計測

生命を理解する上での大きな障壁が（１）非常に小さい要素の集合であること、（２）生きている＝動いていることであるのはすでに述べた。オミクス科学は、測定技術ごとに対象を限定して精度を高めることで（１）に、瞬間冷却や薬剤の添加によって測定対象を“凍結”することによって（２）に対応している。これらの障壁に別のアプローチで挑戦する新しい測定技術がある。ひとつひとつの生命分子の振る舞いを顕微鏡下で追跡しようとする「1分子計測」である。

顕微鏡は、17世紀中葉にフックがコルクを観察し、その構造が小部屋（cell＝細胞）に分画されていることを発見して以来、生命科学研究に不可欠の観察装置である。20世紀に登場した電子顕微鏡は、

ナノメートルスケールの物質の構造を観察することを可能にし、ウイルスの存在の確認などに大きく貢献した。しかし、高分解能の電子顕微鏡観察のためには試料を高真空条件に置く必要があり、生物を生きたまま観察するのは困難である^{（注3）}。光学顕微鏡を使えば対象を生きたまま観察できるが、その分解能は可視光線の波長である数百ナノメートル程度にとどまる。細胞内分子としては巨大なタンパク質ですらその大きさは数ナノメートル程度であり、光学顕微鏡で直接観察することはできない。

1980年代、この困難を克服し間接的にはあるが光学顕微鏡で1分子のタンパク質の振る舞いを観察できる走査型近接場光顕微鏡が開発された。試料表面のごく狭い領域に発生させた近接場光によって、予めタンパク質に付加した蛍光色素を励起し、生じた蛍光を観察することで1分子のタンパク質の位置を同定する技術である。

ほぼ同時期に、蛍光タンパク質の遺伝子が発見・解析され、遺伝子組換え技術を用いてタンパク質そのものが蛍光を発するような改変を遺伝子レベルで加えられるようになり、細胞や組織を生かしたまま1分子レベルで観察できる可能性を拓いた。

1990年代後半には、蛍光共鳴エネルギー移動（Förster resonance energy transfer, FRET）を利用した分子間相互作用の可視化技術が開発された。FRETは、ごく近傍に存在する2つの色素分子間で、励起エネルギーが直接移動する現象である。2種類のタンパク質に異なる蛍光タンパク質を組み込み、FRETを観察することで、両者がごく近傍に接近するような相互作用が起こっているかを間接的に検出することができる。生きた細胞内でのFRETを顕微鏡下でリアルタイムに観察する技術も開発され、1分子レベルの分子間相互作用あるいは分子内相互作用のライブイメージング技術が複数派生した^{〔5〕}。ゲノム科学の最新技術である次世代シーケンシングも、ライブイメージングでDNAの分解・合成を1分子計測する技術を利用している。

1分子計測技術を用いた測定では（１）非常に小さい要素そのものを（２）生命活動を維持したまま非破壊的に測定する技術を開発することで上述の障

壁を乗り越えようとしているといえる。

細胞は狭く、分子の数は少ない

1 分子計測によって、細胞内の分子を概算によってではなくひとつずつ数えることができるようになった。これに関連して“分子の数”に着目することで拓かれた視点がある。タンパク質をはじめとする生命分子の化学的性質を詳細に解明する生化学は、20 世紀初頭より膨大な知見を蓄積してきた。その手法は、充分量の試料を精製あるいは合成し、試験管内で化学的性質を解析するものだった。

試験管は細胞よりはるかに広大な空間である。微生物 1 個体の容積が μl (フェムトリットル、千兆分の 1 リットル) スケールであるのに対し、一般的な試験管は ml (ミリリットル、千分の 1 リットル) スケールなので、実に 1 兆倍の差がある。試験管内で生命分子の反応を観察する際には、分子が充分大量に存在するため、個々の分子の軌跡を考慮する必要はなく、分子間相互作用の平均像を測定すれば充分である。一方、 $1 \mu\text{l}$ という超狭小な容積しか持たない微生物では、状況が一変する。容積 $1 \mu\text{l}$ の微生物細胞内の酸性度が $\text{pH } 7$ (中性) のとき、細胞内に存在する水素イオンの数はたった 60 個しかなく、“十分に大量”にはほど遠い状態にある。

1 細胞内で発現するタンパク質分子の数は、1 遺伝子あたり数個～数千個程度である。遺伝子をコードする DNA は、無性生殖する微生物の細胞には 1 組程度、ヒトのように有性生殖する生物の細胞には 2 組程度しかない。細胞は、少数の分子が超狭小な空間に詰め込まれ、ひしめきあって相互作用する、試験管内とはまったく異なる空間なのである。試験管内と細胞内の環境の違いについては古くから認識されていたが、環境の差によって生じる影響を測定する手段がなかったこともあり、長らくほとんど手つかずのまま棚上げにされてきた。

2002 年、2 つのグループがほぼ同時に、遺伝子発現制御が、大容量の試験管内で測定されるような再現性の高い決定論的事象ではなく、大きなゆらぎを伴う確率事象であることを明らかにした。それぞれの遺伝子は、1 細胞あたり 1 ～数コピーしか存在

しない。それらが RNA に転写されるかどうかは、確率的なゆらぎの影響を受け、同じ環境下にある細胞集団内でも、細胞ごとに大きくばらついていることが、測定技術の進歩にも助けられ実際に確認されたのである^[2]。

こうした確率的なゆらぎは、細胞の容積が小さいことに由来しており、生命が進化してきた 30 億年以上の間つねに存在してきたと考えられる。したがって、あらゆる細胞システムは、確率的なゆらぎの存在下で、その生命活動を安定に維持できるように進化してきたであろうと結論づけられる。

モデル化による理解：生命現象のコンピュータシミュレーション

人間が対象を理解する際、必ず何らかのモデル化を行う。「理解する＝わかる」とは何かを定義するのは非常に難しいが、そこには、人間の知的能力（演算能力）で把握可能な形式に対象を整理しなおす工程が必ず含まれる。モデル化とは、対象に恣意的な抽象・捨象を施し、自らの能力で取り扱い可能な形式に再構成する行為といえる。数理モデルは、対象を数学のことばで表現したモデルである。

生命科学と数理モデルの関わりは古く、メンデルの遺伝理論もその一例である。20 世紀に入ると、タンパク質の機能を定量する酵素学が発展し、個々の研究成果はタンパク質の触媒機能を再現する微分方程式に集約された。第二次世界大戦後、プログラム可能なコンピュータの市販がはじまり、価格が急速に低下してくると、個々のタンパク質の機能を記述した酵素反応速度論モデルを連成した代謝反応ネットワークモデルが構築され、そのコンピュータシミュレーションが試みられた。このころの数理モデルのほとんどは、試験管スケールの測定結果に基づく微分方程式で記述されており、数理モデルに確率的なゆらぎは含まれていなかった。こうした研究は 1960 年代終盤までに一定の成果を生んだが、シミュレーションによって新規に得られた知見は乏しく、また、個々の酵素反応については再現性の高い数理モデルであっても、それらを連成したモデルが必ずしも反応ネットワーク全体の挙動を正

確に再現できないといった事例も多く、1980 年ごろまでには大規模な細胞シミュレーション研究は下火になった。

その間もその後もムーアの法則に則って、コンピュータはより速く、より安価になっていった。そして 1990 年代後半、ゲノム科学の台頭がはじまると、要素還元によって素過程を丹念に解析するだけでなく、蓄積した知見を統合し、素過程の集合として機能する細胞内の分子間相互作用ネットワーク全体を把握したいという気運が高まった。生命科学の知識の統合と、生命現象の大局的俯瞰のための手段のひとつとして、再び数理モデル化とコンピュータシミュレーションが日の目を見ることになった。

細胞シミュレーションの現在

今日、コンピュータシミュレーションは、安価に高速な計算が可能となったことにも助けられ、自然科学はもちろん、航空機の設計や、映像作品などのエンターテインメントまで、ありとあらゆる分野で活用されている。

その中で、分子レベルの細胞シミュレーションは、シミュレーション科学にとって最も困難な対象のひとつである。生命システムが持つ際だった特徴として、構成要素と要素間相互作用の甚だしい多様性がある。コンピュータシミュレーションが対象としている巨大システムには、細胞以外にも、銀河の衝突や大規模集積回路の設計、原子レベルで表現した分子の挙動（分子動力学）などがある。銀河に含まれる恒星の数や、集積回路に含まれるトランジスタ素子の数、分子を構成する原子の数は、細胞に含まれる分子の数を凌駕することもある。しかし、シミュレーション対象となっている大規模システムの多くは、構成要素と要素間相互作用の多様性が著しく低い、あるいは多様性が低くなるように捨象した上で数理モデル化することが可能である。

これに対して、細胞では構成要素の多様性を縮減することが非常に難しい。遺伝子に直接コードされているタンパク質に限っても、ヒトでは数万、大腸菌のような微生物でも数千種を有しており、機能的に類似する分子種をグループ化するなどして工夫し

ても、銀河の衝突や半導体設計程度にまで単純化することは到底できない。よりミクロな階層に移行して原子レベルにモデル化の粒度を設定すれば、要素の多様性は縮減でき、タンパク質であればおよそ 20 種類のアミノ酸の鎖として表現できるようになる。ただし、アミノ酸そのものは生命分子としてのまとまった機能情報をほとんど持たず、アミノ酸配列からタンパク質の機能を高精度に予測できるアルゴリズムは未確立である。タンパク質の機能は、数百から数千のアミノ酸から成る鎖状分子が、立体的に折りたたまれた構造によって決まる。現時点では、最先端の分子動力学と、世界最速のスーパーコンピュータを投入しても、たかだか数個のタンパク質が相互作用する様子をリアルタイムに計算することすら不可能である。また、実験によってタンパク質の立体構造を決めるには高度な技術とコストが必要で、オミクス科学のように大量かつ高速に立体構造を同定できる状況にもない。

単位要素の機能が多様であるため、それらの相互作用は組み合わせの爆発によってさらに多様になり、それぞれの相互作用の結果を推定するのも至極困難である。多様な単位要素が、さらに多様な相互作用をし、それら膨大な素過程が、細胞という狭小な空間で同時並行に進行していることが、細胞システムを精密かつ効率的にシミュレートする上での障壁となっている。

このように、細胞内の生命分子ネットワークの動態は、シミュレーション科学にとって非常に困難な対象だが、この複雑な対象を人間が理解するにあたって採りうる最良の手段のひとつが、数理モデル化とシミュレーションであることも確かだろう。2012 年、スタンフォード大学のカーらは、既知の生物中最小のゲノムを持つ寄生細菌（*Mycoplasma genitalium*）を対象に、525 の遺伝子すべての機能を実装した“全細胞モデル”を構築した（図 1）。このモデルは、各分子の機能の計算結果を統合して細胞分裂に要する時間すなわち増殖速度をシミュレートするものである。生物の生存に欠かせない遺伝子を必須遺伝子と呼ぶが、カーらは 525 の遺伝子のそれぞれについて遺伝子欠損株モデルを作成してシ

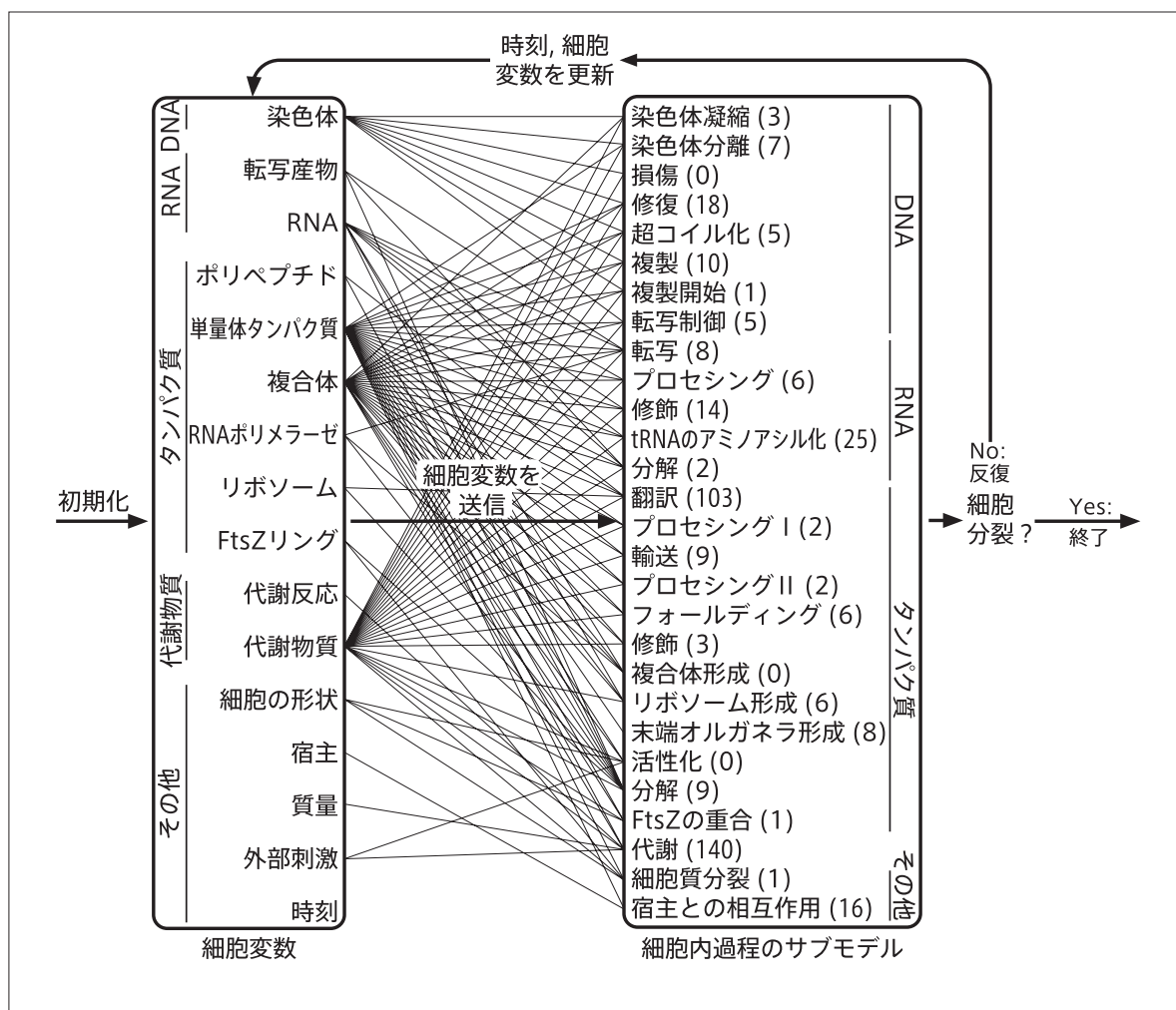


図1 *Mycoplasma genitalium* の全細胞モデルおよびシミュレータの構成 (参考文献3より許諾を得て翻訳、転載)

モデルは900以上の文献、1900以上の実測されたパラメータに基づいている。細胞と周辺環境に関わる状態変数は16の細胞変数にグループ化されている。細胞内で起こるさまざまな過程は28個のサブモデルに分けられ、サブモデルごとに異なる数理表現(質量保存則、代謝流束均衡解析、マルコフ連鎖、ポアソン過程、ブーリアンネットワークなど)を採用してモデル化されている。サブモデルに付したカッコ内の数字は、関与する遺伝子数である。シミュレーションは、1ステップを1秒とし、各サブモデルが必要とする細胞変数を参照し、その変化量を計算する。シミュレータはサブモデルの計算結果を統合し、1秒間の細胞変数の変化を決定し一斉に更新する。これを1ステップとして、細胞分裂が完了するまで繰り返す。

シミュレーションを行い、必須遺伝子であるかを判定した。その結果は、細胞を用いた実験結果と79%一致した^[3]。この成果は、オミクス科学や1分子計測がもたらす知見を活用することにより、高精度の全細胞シミュレーションが実現する可能性と課題を示すマイルストーンとなった。

われわれはどこから来たのか：分子系統学
ここまで駆け足で俯瞰してきたオミクス科学、1

分子計測、細胞シミュレーションが目指しているのは、いずれも生命システムの動作原理 — 今ここにある生命システムは分子機械としてどのようなメカニズムで動いているのか — の解明である。現生生物のリバースエンジニアリングから得られる機械論的理解は、分子機械のデザインの由来を問わず、デザインの洗練過程を説明しない。

分子機械のデザインの洗練過程は、分子レベルの生命進化の歴史である。ゲノム科学から得られる知

見により、生命進化に関しても分子レベルの理解が進んでいる。

生命科学の近代化を先導した原理が遺伝と進化であることはすでに述べた。両者が説明しようとするのは、生命が30億年以上の時間をかけて今日のような複雑華麗な生態系に至った過程である。ダーウィンの進化理論は、少しずつ変化しつつ世代を重ねていく生物集団と周辺環境の相互作用によって、より環境に適応した個体がより多くの子孫を残すことになり、結果として新たな形態や機能を持つ生物が繁栄し、この過程を連綿と繰り返すことで、太古より生物は漸進的にその姿と能力を変えてきたのだと説明する。このダーウィン進化を中核とする進化理論は、ダーウィン以降に提案された中立進化説などの有力理論を取り込みつつ、数学的な枠組みも整え、今日では進化の総合説と呼ばれている。

一方、増殖時に起こる変化の実体が、ゲノム複製の際にDNAに生じる塩基配列の突然変異であることが解明されるなど、遺伝情報の世代間伝達と、その際に突然変異が発生するメカニズムの詳細が分子レベルの生命科学によって解明された。偶発的に生じる突然変異には、生物の機能にまったくあるいはほとんど影響を与えない中立変異も多い。中立変異

は大きな時間スケールで見れば安定した速度で蓄積していく。これを「分子時計」として利用し、生命進化の系統樹を再構成する取り組みが、DNAシーケンシング技術が急速に進歩した1980年代以降盛んになり、分子系統学という進化生物学の新たな領域を生み出した。

分子系統学では、複数の生物種間で解読した遺伝情報を比較し、その差分を統計学的に解析することで、相互の系統関係（共通祖先からどのような分岐を経て現生種に到達したか）を推定する。従来の系統学では、形態学・解剖学的知見や、卵や胚からの発生過程の比較、代謝物質などの生化学的解析に基づく系統分類が行われてきたが、分子系統学は突然変異が起こる場である遺伝情報そのものを用いる点で強い説得力を備えている^[6]。

現実起こった自然選択は知り得ないのか

分子系統学により、従来より信頼性の高い進化系統樹を推定できるようになった。初期の生命進化では、樹木が分枝するような分岐だけでなく、生物種間で遺伝情報が激しくやり取りされ、網目状の系統樹としてしか描画できないような進化が起こっていたらしいという、従来の定説を全面的に刷新する鮮

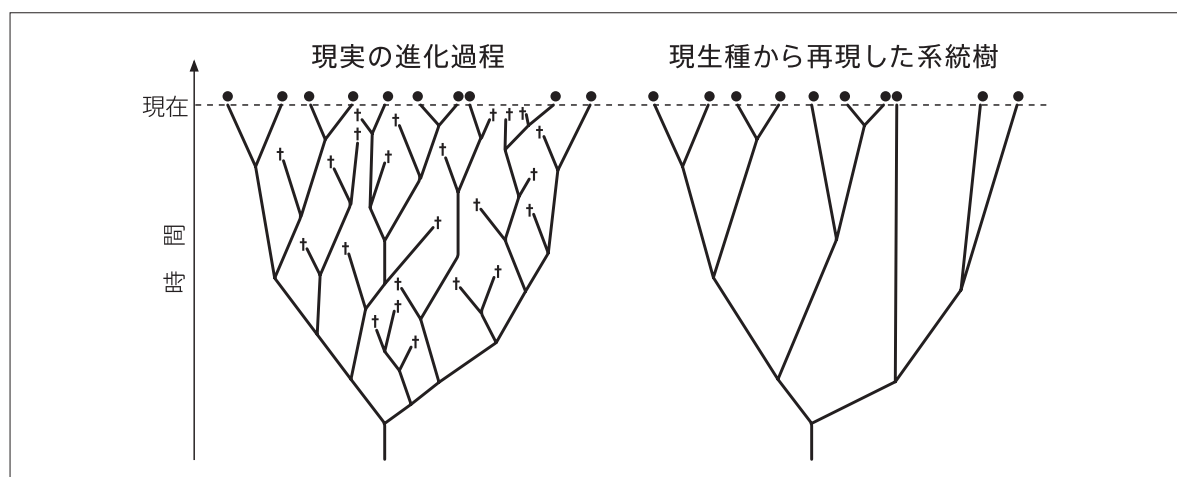


図2 分子系統学によって求められる系統関係

図は、進化系統樹のごく一部の模式図である。現実の進化過程（左）では、現在まで子孫を残す系統（●）はごく一部で、大部分の系統はある時点で子孫が途絶え、絶滅する（+）。現生生物の遺伝情報に基づき、分子系統解析によって再現される系統関係を右に示す。推定されるのは現生生物間の分岐が起こった年代に止まり、現生生物の祖先が過去にどのような生物と競合し、いかなる自然選択を生き抜いてきたかについての情報がもたらされることはない。

やかな知見も得られた。

しかし、地球上で展開されてきた生命進化の歴史のすべてが明らかになったわけではない。また、この研究プログラムを突き詰めていくことで、いずれすべてが明らかになることもない。分子系統学が明らかにするのはあくまでも現生生物の系統関係である（図2）。古生物学的試料から奇跡的に抽出されたごく稀な例外を除き、すでに絶滅してしまった生物のDNAを手に入れる機会はなく、系統関係の推定に供されるDNA配列情報は現生生物のものに限られる。現実には、現在まで子孫を残すことに成功した現生種より、はるかに多くの生物種が絶滅し地球上から消え去っている。現生種の祖先が、過去に地球上で自然選択の競争を生き抜き、子孫を残したことは間違いない。しかし、わかっているのは、競争に勝利したことだけであり、どんな相手と何を巡って競争し、なぜ生き抜くことができたかについて、人類が知り得たことはごくわずかである。

ダーウィン進化は、遺伝子型の突然変異と、表現型に対する自然選択を両輪として、地球上の全生物個体が死滅するまで、淡々とつづいていく。突然変異は、それぞれの生物個体内で完結したメカニズムであり、その源となるDNA複製機構は、ひとつひとつの細胞の中にひと揃い内蔵されている。一方、自然選択は、生物個体の機能と周辺環境の相互作用によるものであり、その解明には、生物だけでなく周辺環境に関する情報も必要である。周辺環境と生物個体の相互作用は非常に多様であり、ハイスループット計測の対象とはなりにくい。現在精力的に進められている分子レベルの生命進化の解明は、突然変異の発生機構および蓄積の歴史の解明であり、進化の過程で起こった具体的な自然選択の解明はほとんど手つかずの状態で棚上げされている^(注4)。

計算による自然選択の探究

過去の地球環境の詳細な変遷を知ることは難しい。また、すでに絶滅した生物に関する情報は、そうした生物の遺骸の大部分がすでに分解してしまっている以上、今後も潤沢に手に入れられる可能性はほとんどない。実際に起こった自然選択の復元は困

難な課題である。

筆者は、数理モデルとコンピュータシミュレーションによって、この問題に新たな光を投げかけることができるのではないかと考えている。細胞内の事象を分子レベルで定式化した数理モデルの中には、部分的にはあるが高精度に細胞機能を再現できるものがある。これらのモデルを用いて、突然変異による分子レベルの微細な機能変化が生じたとき、細胞システム全体にどういった影響が及ぶかをシミュレートすることも可能である。

過去に存在したかもしれない現生種と近縁の生命システムを、数理モデルであれば、計算機資源の許す限り大量かつ自在につくりだすことができる。そうした“こうであったかもしれない”「可能な」デザインの細胞モデルと、現実に存在している細胞のモデルの機能を比較することによって、現生する細胞のデザインの何が優れているのかを評価することができるのではないか。

地球上からすでに失われてしまった生命システムを、コンピュータ上の仮想的な可能世界につくり出すことによって、現生生物のデザインの必然性あるいは偶有性を検討する足がかりとすることができるのではないか。不可知の領域にある対象を経験の地平に引きずり込むためにコンピュータシミュレーションを利用するという発想は、蛮勇に過ぎ、自然科学の枠組みを逸脱しているだろうか。

むすび

本稿では、分子レベルの生命科学の展開と、その中で情報科学技術が果たす役割を概観した。現在の生命科学は、生命を分子機械と捉え、リバースエンジニアリングによって動作原理を解明するプログラムに研究資源を集中しており、生命情報のハイスループット計測、データの管理、可視化、解析、アクセシビリティの提供などのあらゆる段階で、情報科学技術が大きな役割を果たしている。

すでに計測された、あるいは今後計測される大量の生命情報から、有益な科学的知見を抽出するためには、今後も最新の情報科学技術の投入が不可欠であり、より多くの情報科学者の生命科学領域への合

流ならびに生命科学者に対する情報科学教育のさらなる充実が求められる。

注

- 1 1975 年、微生物に感染するバクテリオファージ ϕ X174 の全遺伝情報 (5386 塩基対) が解読された。
- 2 セントラルドグマのアイディアには、逆向きの反応、すなわちタンパク質のアミノ酸配列情報に基づく RNA あるいは DNA の合成は起こらないという提案が含まれる。
- 3 2013 年に針山 (浜松医科大学) らのチームが、高真空下で生物を保護し、生きたまま高分解能で電子顕微鏡観察することに成功した。
- 4 自然選択を対象とした研究は、工夫を凝らして行われている。たとえば、極端な環境変化に遭遇した現生する野生生物集団に起こる変化の観察事例は数多くある。混合培養した微生物に人為的に選択圧を与えて個体数の推移や遺伝情報の変化を追跡するなど、人工的な進化系を構築し自然選択の一般的な特徴を探る実験室内進化の研究も成果を挙げている。また、人工生命を用いたコンピュータ上での進化実験の研究事例も多い。にもかかわらず、地球上で過去に現実起こった個別具体的な自然選択の「歴史」の解明は依然として困難であり、その大部分は手つかずのままである。

参考文献

- 1 Berger B, Peng J, Singh M. “Computational solutions for omics data”, *Nat. Rev. Genet.* 14(5), 2013, pp.333-346.
- 2 Kaern M, Elston TC, Blake WJ, Collins JJ. “Stochasticity in gene expression: from theories to phenotypes”, *Nat. Rev. Genet.* 6(6), 2005, pp.451-464.
- 3 Karr JR, Sanghvi JC, Macklin DN, Gutschow MV, Jacobs JM, Bolival B Jr, Assad-Garcia N, Glass JI, Covert MW. “A whole-cell computational model predicts phenotype from genotype”, *Cell.* 150(2), 2012, pp.389-401.
- 4 Metzker ML. “Sequencing technologies – the next generation”, *Nat. Rev. Genet.* 11(1), 2010, pp.31-46.
- 5 Swedlow JR. “Innovation in biological microscopy: current status and future directions”, *Bioessays.* 34(5), 2012, pp.333-340.
- 6 Yang Z, Rannala B. “Molecular phylogenetics: principles and practice”, *Nat. Rev. Genet.* 13(5), 2012, pp.303-314.

〔受付日 2013. 10. 3〕