

# 数学協働プログラム提言

## 「数理生命科学」

### 1. 始めに

本提言の目的は、数学が生命科学の様々な分野において既に多岐にわたって利用され、分野の基礎となっている現実を紹介すると同時に、これから分野の発展に寄与すると考えられる数学的手法や視点について情報を提供することにある。生命科学の歴史において、数学は手法として用いられてきた一方で、時に生命科学との相互作用から新たなアイデアやブレイクスルーを見出してきた。生命科学が急速に発展し、その内容が複雑化してくる中で、幅広い領域を俯瞰して将来の発展の糸口を見いだせる新たな人材が求められている。この提言をきっかけに、分野間での対話が始まり、数学を橋渡しとした新しい分野の創出や発展につながることを期待する。

まず初めに、生命科学の持つ特徴としての複雑な階層性と、それに対する共通概念としての数学について、理解しておく必要がある。生命科学の対象は生体分子から生物個体・環境まで幅広い時空間スケールにまたがっており、現象の階層ごとに異なる学問分野を生み出してきた歴史がある。ごく簡単にまとめるならば、遺伝子や酵素といった生体分子とその働きを研究する分子生物学や生化学、生体分子が調和して作り出す細胞機能を研究する細胞生物学、細胞集団の作る組織や形態形成を扱う発生生物学、個体としての（或いはその部分系の）生理現象を対象とする生理学や医科学、個体のレベルを超え集団および群集レベルの現象を扱う集団遺伝学、進化生物学そして生態学などが、それぞれに独自に発達してきた。実際には、それぞれの系で見られる現象は階層を超えて互いに深くかかわりを持っている。しかし、特定の階層に注目することで現象が捉えやすくなり、学問として発展してきた点は注目に値する。例えば形態形成を考える際に、細胞を単位として考えることは、細胞集団である組織の形を理解する上で助けとなる。

分子、細胞、個体、集団など、生物学に存在する異なる空間スケールのそれぞれにおいて、しばしば共通の数学が用いられる。例えば、生命現象を時間発展のダイナミクスとして捉える視点はいずれの階層でも重要である。生物集団というマクロなレベルに対して用いられた数理理論が、そのまま一遺伝子の振る舞いというミクロなレベルの現象に適用された場合もある。別の表現をするならば、実験生物学者が生命現象に様々な視点で取り組んできた中から、幅広い現象に共通する概念または知識を取り出す試みが数理生命科学といえる。

生物現象に対する数理的手法はしばしば、大きく「モデル」的アプローチと「データ解析」的アプローチに分けられる。生命科学は「現象の背後に潜むメカニズムを一つ一つ明らかにする実証科学」として発展してきており、顕微鏡技術など「事実を発見する実験手法の力」に大きく依存してきた。現在でも研究の対象として注目を集める生命現象の多くは、メカニズムが不明ではあるが興味深いふるまいを示す、ブラックボックスの系として観測されている。このような実証科学に対し理論が貢献するためには、「既知の情報から未知の機構の可能性をいかに絞り込むか」、という思想が重要である。この絞り込みのために、仮説としてのモデルをトップダウンに提案して整合性を議論するのが「モデル科学」、膨大な実験データからボトムアップに可能性を絞り込んでゆくのが「データ科学」である。絞り込み作業という観点からは、いずれも生命現象を記述する「モデル」の導出が目標となる。つまり、モデル的アプローチとデータ解析的アプローチとは、理論の組み立て方に関する大まかな分類にすぎず、これら二つに明確な境界があるわけではない。現在の生物学

においては、あらかじめ関数形を仮定しないモデル理論や、大量データからモデルを生成する理論などが活躍しており、これら二つの性質を併せ持つような理論の発展が期待されている。

これからの数理生命科学に期待されるモデルとは何か。複雑な現象の理解、新しい視点の提示、検証可能な予測提示、など、モデルには様々な役割がありうる。ただ現在において、生命現象を数理的に表現するだけでは新規性に欠け、単に現象を再現するだけのモデルに対しては厳しい評価がなされる。むしろ実験系に対して予測を与える理論、以前にない新しい考え方の提唱や、複数のモデル候補を挙げそれらのいずれが正しいか実験生物学に選ばせる理論などが期待されており、実際にそのような実証的で力強い理論が現れつつある。

また現代の生物学実験の特徴として、実験の精度が上がりノイズの少ない定量データが得られるようになってきていること、また多数の遺伝子の振る舞いを同時に捉えるなどの大規模化が進んでいることが挙げられる。実験的に得られた知見を様々な形式でまとめたデータベースも世界中に多数存在している。分野によっては、個人の力で精査して理解することが到底不可能な量に達している場合もある（例えばゲノム科学）。膨大なデータの中から、有意な変化や特徴を抽出する広い意味でのデータ科学と、単なる統計処理に終わらないでそこからモデルを抽出する試みが、今後重要になるだろう。

## 2. 生命科学の礎となる数学

以下では、生物学の歴史において様々な形で重要な役割を果たした数理的手法を概観する。古典的なものから近年の流行までを、大きく8項目にまとめている。

- (1) 時間発展ダイナミクス（常微分方程式）
- (2) 時空間ダイナミクス（偏微分方程式）
- (3) エネルギー、ポテンシャル
- (4) 確率モデル
- (5) 制御、最適化、ゲーム理論
- (6) 情報処理
- (7) 統計と機械学習
- (8) ネットワークとヒューマンインターフェース

各項目には以下の書式で分野のキーワードを並べ、その下に簡単な解説をつけた。

書式：（数字） 大項目 （手法として重要なもの）  
・ 中項目 （大項目に分けられる分野や概念）  
小項目 （具体的な研究や方法）

### (1) 時間発展ダイナミクス（常微分方程式）

- ・ 個体群動態
  - 捕食者と被食者のロトカー-ヴォルテラモデル（生態学）
  - 感染症のSI(R)モデル（疫学）
- ・ 進化動態
  - 適応進化モデル、プライス方程式、アダプティブダイナミクス

- ・分子反応系

  - 酵素反応のミカエリスメンテンモデル

  - ニューロンのホジキンハックスレーモデル

- ・大規模ネットワーク（遺伝子、代謝、シグナルなど）

  - 高次連立微分方程式系、分子（particle）シミュレーション

生命現象をダイナミクスとして捉える研究は、生物学における数理科学としてもっとも初期から継続的に成功してきた分野である。野外の生物の個体数が時間とともに変動することは、20世紀初頭から知られていた。これを捕食者数と被食者数との相互作用が作り出す振動として説明した常微分方程式モデルが、提案者である二人の数学者の名を冠したロトカ-ヴォルテラ捕食系である。類似の数理モデルに、「食う-食われる」の関係ではなく、同じ資源をめぐる競争する生物種についてのダイナミクスを研究し、種内競争、種間競争の激しさを表す係数などによって競争ダイナミクスを記述したロトカ-ヴォルテラ競争系がある。これらの数理モデルはとても単純なものであるが、進化と個体群動態の相互作用の関係を調べたり、多種から構成される群集の安定性を調べたりするための基礎モデルとして今なお幅広く用いられている。

数理モデルの価値の一つに、新たな考え方の提唱があることは序段で述べた。ロバート・メイはロトカ-ヴォルテラ方程式を用いて生物群集を記述し、力学的安定性を用いた議論によって、群集内における生物種数が大きいほど一般にその系が不安定となることを示した。それ以前に実験生態学においてナイーブに信じられていた定説「多くの種をその系に含むことで生態系は安定する」という理解を覆したこの理論によって、群集生態学の問題設定は「種数が多いシステムがいかにか安定に維持されているか」を解明する方向へと向かうことになった。これは数理理論が分野全体に問題意識を与えた好例だと言える。

このような価値に加えて、ロトカ-ヴォルテラモデルには、変動する生物個体数に対してダイナミクスの観点から初めて合理的な説明を与えたという歴史的意義がある。ミクロなレベルでの生物種間の相互作用の記述により、マクロなレベルでの個体数の変動がダイナミクスとして理解できるという考え方は、数理生態学という分野を切り開くこととなった。

より実用的なモデルとして、感染性の病気が集団中に侵入してきた後の時間変化を考える SIR モデルがある。このモデルは未感染集団の個体数密度がある閾値を超えるか否かで、大規模感染が起きるか否かが決まるという興味深い予測を導く。SIR モデルは感染症の基本モデルとして、感染拡大予測や、各種疫学パラメータの推定の目的で幅広く用いられる。

ダイナミクスを通して生命現象を理解するという考え方は、何もマクロな対象に限らない。同様の考え方は生物学の階層を超えて様々な対象に適用されてきた。例えば、細胞内では非常に多数種の化学反応が起こる。現在では、細胞内の多数の化学反応を取り入れた計算機シミュレーション研究もなされている。実はこれら生体内の化学反応は、一つ一つが酵素タンパク質によって引き起こされる触媒反応である。酵素反応は、反応の材料となる物質（基質）が酵素と可逆的に結合する反応と、基質が化学反応を起こして別の物質に変わる反応との組み合わせからなる。この多段階反応を近似的に一つの式にまとめたものがミカエリスメンテン式であり、これにより生体内の化学反応ネットワークはずっと考えやすくなった。

## (2) 時空間ダイナミクス（偏微分方程式）

- ・空間分布、時間分布

  - 生物の侵入、年齢構造モデル

- ・反応拡散系

- 発生／形態形成、非対称性の数理

生物集団の空間分布を考えた時、或いは生物の体作りや形作りを考えた時、扱うべきダイナミクスは時間発展だけでなく、空間構造も含んだ時空間ダイナミクスとなる。生態学においては、生物の空間的伝搬や侵入が、拡散方程式などを用いて研究されてきた。ほぼ同様の取り扱いが後述の集団遺伝学において、集団中の遺伝子頻度分布のダイナミクスを記述する方程式として用いられていることは興味深い。また、個体の空間的位置の代わりに個体の齢を考えれば、どの齢の個体がどれだけ多くいるかを記述する齢構造モデルとなる。齢構造モデルは人口論の分野において安定齢分布（人口ピラミッド）の導出や年金・保険に関する基礎データの導出に使われるなど、実用的側面が強い。また感染症は接触率等の違いにより齢毎に感染率が異なることが多く、そこでも齢構造モデルは応用上重要である。

個体より小さなスケールでは、特に形態形成、発生生物学において、時空間ダイナミクスとしての理解は、重要な役割を果たしてきた。反応拡散方程式系は、発生生物学における基本的理解を大きく進めたのみならず、現在でも実験結果の解釈などに幅広く用いられている。特筆すべきものとして、計算理論の開祖でもあるアラン・チューリングによる拡散不安定性の発見が挙げられる。チューリングは2種の物質が相互に影響しながら拡散する方程式系を考えた。このとき空間構造を省いた化学反応だけの系が安定な平衡解を持つとしても、そこに拡散項が加わるだけで不安定化しうること、さらに特定の波数を持った周期分布が現れることを示した。ごく簡単な解釈を加えるならば、拡散する分子どうしが相互作用することで、秩序や構造の無い状態から縞や斑点のような周期的パターンを作り出せる、というメッセージである。

チューリングの拡散不安定性理論が、発生生物学に与えた影響には特別なものがある。生物は単細胞の受精卵から、多細胞からなる秩序と機能を持った構造を作り出すが、そもそもどのようにして秩序の無い状態から構造を作り出すのか。形態形成の担い手は体を構成する多数の細胞であるが、個々の細胞は組織全体の中での位置を、どのように知るのか。これらの問いは発生生物学の本質である。チューリングの理論は、これらの問いを具体化すると同時に、一つの数学解を与えている。この理論は生物学者を刺激し、様々な解釈を加えながら検証実験がなされている。

形態形成現象に対しては偏微分方程式に限らず、セルオートマトン、粒子多体系などもパターンを生成する数理として発生生物学に貢献してきた。また、古典的に用いられてきた連続体力学（座屈など）も近年になって復権しつつある。

### (3) エネルギー、ポテンシャル

- ・分子の力学系

- 蛋白質フォールディング、リガンドドッキング

- ・細胞／組織の力学系

- 細胞運動ダイナミクス、コンパートメント化された統合モデル

- ・バイオメカニクス

- 生体医工学

- 運動、歩行パターン

現在の多くの生物理論は物理法則を元に作られているのではなく、注目するそれぞれの系の素子（形態形成ならば細胞、個体群動態ならば個体）の振る舞いを仮定した数理モデルが基になってい

る。一方で、物理学に基づいて生物現象を捉えようとする試みも多数あり、もっとも成功したものがタンパク質フォールディングなど、生体分子の構造を研究する分子科学である。このように生体分子の構造やダイナミクスを計算する理論には複数の立場があり、近似の度合いにより、第一原理計算、全原子分子動力学計算、粗視化分子動力学などに、分類される。この分野は高速化を目指す計算科学としての側面もあり、計算統計物理学と連携しながら発展してきた。ただし以下に示すように、大きな課題もある。

物理学の原理に基づいてボトムアップ的に現象を理解できることは、自然科学の一つの理想かもしれない。しかしその方法は生物学に限らず、しばしば実際的ではないことがある。タンパク質フォールディング問題は、その象徴的な例である。タンパク質はアミノ酸の配列が折りたたまれて、機能を持った3次元構造となる。折りたたみの過程で働く作用は、基本的には静電相互作用であり、物理学としての原理は分かっている。にもかかわらず、アミノ酸配列だけからタンパク質の3次元構造を理論的に決定する、もしくは計算することは、100残基以下の小さなタンパク質のみでしか実現できていない。

ここには複数の問題があると考えられている。第一に計算機コストの問題である。すなわちアミノ酸の構造について、その原子や電子の確率密度まで含め、さらに重要な役割をされると思われる周辺の水分子まで含めて、折り畳みの全行程を計算することは、現在の最先端の計算機の能力をもってしても困難である。一方でタンパク質の構造を単純化し、一つのアミノ酸を一つの粒子として扱うなどの粗視化（モデル化）を行った計算方法もある。しかし、そのような場合には、モデルの信頼性に起因する多数の局所安定状態が現れるという別種の問題が現れる。つまりアミノ酸配列からの折りたたみを計算した場合に、（別の実験で明らかにされた）目的のタンパク質とは異なる様々な構造が作られ、もっとも安定と思われる真の構造になかなか至らない。これは現実に観察されるタンパク質の折り畳み現象とは異なる結果である。つまり、アミノ酸配列に対してタンパク質構造を与えるポテンシャル関数は、まだ解明されていない。

物理学に基づいてダイナミクスを考える数理は、分子構造以外に細胞や組織の変形や形態形成など、様々なスケールの現象に対して研究が進められており、これらはバイオメカニクスと呼ばれる。例えば、生物の歩行パターンに対する運動力学に基づいた研究は、多くの実績がある。

#### (4) 制御・最適化・ゲーム理論

- ・進化生態学、保全生態学  
生活史戦略／社会生物学／ESS
- ・バイオエンジニアリング  
代謝工学、
- ・最適化アルゴリズム  
遺伝的アルゴリズム

生物の形態や行動などの形質（表現型）の多様性は、進化の過程で作られてきたものである。場合によっては一見奇妙に見える生物の行動や生態は、その生物が置かれた環境への最適化、あるいは他個体との相互作用における最適化が実現した姿であるとして理解がなされてきた。例えば、植物個体が自己の持つ栄養資源をいつ成長に投資しいつ繁殖に投資するかといった生活史戦略の研究は、最適化という視点の導入によって飛躍的に進展を遂げた例である。また社会性昆虫における性比の制御や魚類などで見られる性転換など、様々な生物で見られる多様な性システムは、他者との相互作用における最適化を解析するゲーム理論の導入によって大きく成功した。この分野には、

ポントリャーギンの動的最適化や、ゲーム理論におけるナッシュ均衡など、工学や経済学における数理的アイデアが積極的に導入された。特にメイナード-スミスが提唱した ESS（進化的に安定な戦略）の概念は、ゲーム理論を生み出した経済学でもその後用いられるようになるなど、生物学のみならず諸学問分野の発展にも大きく貢献した。

生物の進化には、環境や競争に対する表現型の適応という側面と、その表現型を作り出す遺伝的なメカニズムの側面とがあるが、後者を意識せずに表現型の最適性だけで多くの現象が理解されてきたことは特筆に値する。このようなアプローチは表現型ギャンビットと呼ばれ、70年代以降幅広く用いられてきた（なお、進化を生み出す二つの原動力「突然変異」と「自然選択」のうち、前者を強調する立場を **mutationist**、後者を強調する立場を **selectionist** と呼ぶ表現がある）。

性に関わる多様な生物現象が数理的に解明された一方で、性システムそのものの進化については、いまだに決定的な理論が無いことは興味深い。交配によってそれぞれの持つ遺伝情報を半分ずつ次世代に受け渡すという繁殖方法は、全ての遺伝情報を次世代に受け渡す単為生殖に比べて、伝達効率から単純に計算すると 2 倍のコストがかかる。幾つかの説明は存在するものの、有性生殖の 2 倍のコストを説明する決定打となる理論はまだ存在しない。つまり有性生殖を前提とした性システムの進化生態学が大きく成功した一方で、そのシステムの根幹に対する説明はいまだなされていないのである。

制御工学を生体機能に適用した分野として、代謝工学が挙げられる。バクテリアなどの微小生物を対象とし、これらの生体の代謝反応を制御することで、有用な産物を効率よく得ようとする研究がなされている。また生物の進化過程に着想を得て考案された最適化の手法として、遺伝的アルゴリズムが挙げられる。

## (5) 情報処理

- 配列解析  
ゲノムアセンブリ、遺伝子機能予測
- 画像解析  
輪郭抽出、トラッキング
- テキストマイニング  
自然言語処理
- メタアナリシス  
医療、遺伝子データベース

DNA の塩基配列やタンパク質のアミノ酸配列を記号配列としてとらえ、情報科学的に解析する配列解析は、分子生物学の発展に大きく寄与してきた。生命情報の分野でもっとも多用されるソフトウェアは配列相同性検索に用いられる **BLAST** だろう。このツールは（統計）数学が生命科学にもたらしたパラダイムシフトの典型例である。それまでの配列相同性解析は、配列進化のモデルを組み立てて「より正確に進化を反映した相同性検索」を目指してきた。これに対して、**BLAST** の出力は与えられたデータベースの中身がランダムと仮定した時に偶然ヒットする期待値(**E-value**)である。つまり進化という概念から離れ、純粹に統計的な指標に基づく検索ツールが **BLAST** である。同様に **RNA** の二次構造もエネルギー最小化という物理的指標に基づいて計算される。このように、生命情報をあくまで情報と捉え、数理科学的に処理することで得られる結果が、分子生物学に役立ってきた。

このような立場は、近年の情報爆発によってより重要になってきている。その典型例がゲノムアセンブリーと呼ばれる処理である。ゲノムのサイズは数百万塩基の微生物から六千億塩基を超えるアメーバまで様々だが、それらを機械的に読み取れる単位は数百塩基にすぎない。つまり、もとのゲノムをランダムに数百塩基毎の断片に裁断して読み取ってから、それらを再びつなぎ直し、もとの長大な配列を再構築する処理が必要になる。この処理がアセンブリーである。その際に、読み取るべき断片数や最適な断片の長さに対する、確率論的指針が必要となる。また配列の再構築のための高速アルゴリズムも必要になる。

同じことは画像処理・解析についてもいえる。これまで生命科学者が自らの眼で観察し自然言語で記述してきた現象を、今や画像や動画としてデジタル化できる。その大量デジタルデータから人間の理解する内容を数値として取り出すためには、輪郭の抽出から動く要素のトラッキング、色調変化の測定など様々な処理を施す必要がある。ここで役立つのが主に工学系で培われてきた情報処理技術である。

解析された生命現象やデータは、最終的に論文の形で出版される。現在は論文も全てデジタル化されており、生命科学系論文のほぼ半数が収録される PubMed データベースには毎分 1 本以上のペースで論文が登録される。2000 万件を超える文献はキーワードですら人手で処理しきれず、自然言語処理技術の力を借りて維持・管理されている。また大量の論文に記載される情報を組み合わせることで新しい生命知識の獲得を目指す分野も生まれている。

報告される大量の論文の中には、結果が矛盾するものや重複するものも多く含まれる。とりわけ医療の分野では、薬剤の有効性を検証する試験が異なる国や地域で複数行なわれるため、その結果を統合して統計解析をやり直すメタ・アナリシスが重要になる。得られたメタ・アナリシス結果もまた論文として発表されデータベースに登録もされる。こうして何度も検証される過程で、生命科学の知識とよべるものが選びぬかれ、確立してきた。

このように生命現象に直接結びつく情報から、二次的三次的情報まで、様々な抽象化段階において情報処理は生命科学に重要な役割を果たしている。

## (6) 確率モデル

- ・集団遺伝学、分子進化学
  - コアレッセント過程、遺伝的浮動、種の絶滅
- ・確率微分方程式
  - 細胞運動など、拡散方程式

生物現象に対する初めてのそしてもっとも成功した数学・数理モデルは、メンデルの遺伝モデルだといえるかもしれない。染色体や DNA などの遺伝の分子メカニズムが全く分かっていなかった時代に、単純な交配実験と統計的解析だけから、遺伝を司る粒子としての実体、すなわち遺伝子の存在を予測し、そして後年その通りの存在が見つかった、これがメンデルの業績である。高等学校で習う 1 遺伝子座 2 対立遺伝子のメンデル遺伝が簡単な数学で理解できる一方で、これを染色体上の多数の場所（遺伝子座、サイト）、それぞれの遺伝子座における多数の対立遺伝子（アレル）に拡張し、集団中の頻度を考えると、たちまち困難な数学的問題となる。この問題は集団中の遺伝子頻度のダイナミクスを捉える集団遺伝学として発展した。集団遺伝学は確率論と同時進行で発達しており、集団遺伝学の基礎を築いたライト、フィッシャー、ホールデンの 3 人は、同時に確率論の発展にも大きく貢献してきた。

遺伝子の実体としての DNA 分子の、塩基配列の変化（置換率）が研究対象となった結果、染色体上の塩基配列の高い割合が選択の対象ではなく、時間（進化時間）あたり一定頻度で置換されると考えられるようになった（分子進化の中立説）。これをむしろ前提とおくことにより、生物種間の距離を測る方法として使われている。関心のある複数種の生物における共通遺伝子の配列の違いを図ることによって、それら生物種の系統樹を構築できる。

遺伝子頻度のダイナミクスを扱う集団遺伝学に限らず、生物学には確率論を用いて理解すべき現象が多数ある。例えば種の絶滅率などは、よりマクロなレベルでの確率的挙動を考える数理である。広く確率モデルとしては、遺伝にのみでなく細胞の運動などマクロなレベルでの確率的振る舞いに対しても、確率微分方程式などが用いられてきた。

## (7) 統計と機械学習

- スパース推定  
ゲノムデータ解析、グラフィカルモデル
- ベイズ推定（最近は特に Approximation Bayesian Computation (ABC) )  
データ統合、分子系統解析、因果解析、バイオイメージング
- ニューラルネットワーク

統計学は、データを解析する様々な手法を提案してきた。生命科学における代表例としては、説明変数と応答変数の間の関係を記述するための回帰モデル、二つの母集団の平均が同じであるかどうかを検証するための平均の同等性検定、二つの変数に関わりがあるかどうかを検証するための独立性の検定などがあるだろう。ただし、統計学は、もともと、データ数が小さい時により効率的にデータを解析することに主眼を置いてきた傾向がある。しかし、昨今のデータ数の巨大化に対応するために、違う観点も生まれてきている。モデルを丁寧に吟味して精緻化するという観点よりも、モデルを自動的に高速に学習していく観点なども必要になってきた。その観点を重要視しているのが機械学習である。

生命科学においては、21 世紀以降、マイクロアレイや高速シーケンサーなど、革新的なバイオテクノロジーの誕生が相次いだ。生命科学が解析対象とするデータは急速に大規模かつ多様になって来ている。特に、標本数( $n$ )よりも変数の次元( $p$ )の方が大きくなることがあり、古典的な方法が使えなくなって来た。そのような問題に対処するために、パラメータの多くを 0 と推定するスパース推定や、仮説の数が多すぎるときに使いにくかった  $p$  値に代わる指標としての FDR (False Discovery Rate) などが、近年になって提案された。

また、ある生命現象に対し、異なる階層のオミックスデータを統合的に解析することが一般的になっている。加えて、実験データと数理モデル（シミュレーション）を統合的に解析することもある。このような統合的なデータ解析方法も発展してきている。そこでは、事前情報をうまく組み込んだベイズ統計が使われることがある。特に、実験と数理モデルの統合の流れにおいて、興味深い発展の遂げ方をしたものとして、集団遺伝学における近似ベイズ計算 (ABC: approximate Bayesian computation) が挙げられる。ABC は、複雑なモデルを用いてベイズ推定を行うとき、複雑な尤度計算をうまく回避して近似的にベイズ推定を行うテクニックである。現在は、集団遺伝学以外にも、システム生物学・生態学・疫学などの分野で応用が進んでいる。

脳のニューロン特性を数学的に抽象化して作られたモデルとして、ニューラルネットワークモデルがある。複雑な情報処理を実行できる脳の構造を模倣することによって、様々な問題に対して柔軟に適用することが可能であり、多くの実応用で用いられている。しかし、モデルが複雑であるた

め高精度のモデルを安定して得ることが難しかったが、最近になって、深層学習（ディープ・ラーニング）の手法が急速に発達し、パラメータ推定法に工夫が行われ、また大規模データの効率的な処理が可能になることによってあらためて脚光を浴びている。

## (8) ネットワーク解析とヒューマンインターフェース

- ・ネットワーク視覚化  
オミックス解析（遺伝子発現）
- ・クラスタリング  
大規模情報の視覚化

生命科学が対象とする多くの系に、多数の生体分子が関わっており、それら分子間の関係性が重要と考えられている。多数の因子の関係性を視覚的に表現する目的で、グラフ、或いはネットワークを用いた表現が、実験生物学の現場においても多用されている。グラフとは、頂点（ノード）の集合と、それらの二者間を結ぶ辺（エッジ）の集合からなる構造であるが、生命科学においては、グラフと呼ぶよりは、ネットワークと呼ぶことの方が一般的である。生命科学では、大きく分けて3種類の意味の異なるネットワークが用いられる：(1)生体分子が（化学反応により）構造変化をおこす過程を辺で表現した反応ネットワーク、(2)生体分子が伝える情報の伝達経路を辺で表現した制御ネットワーク、(3)生体分子の活性相関など、物理的な相互作用と無関係でありうる概念を辺で表現した相関ネットワーク、である。

このように、生命科学で用いられるネットワークは、数学的に定義されるグラフと様々な意味で異なっている。例えば、生化学反応には、2種の生体分子から1種の生体分子が作られる複合体形成反応や、1種から2種が作られる分解反応が多く含まれる。これらの反応は習慣的に、先が2つに分かれる辺や、2つの根が1つの先に合流する辺として表現されることが多く、「2つの頂点を結ぶ」というグラフの定義にあてはまらない。生物学或いは化学において習慣的に用いられる反応ネットワークを数学的に表現するためにはハイパーグラフという概念が必要だが、計算処理の複雑さから用いられることは稀である。

また、ネットワークで表現された内容を数理的観点から処理するには、各頂点における接続関係が他の頂点とは独立に定まることが重要である。しかし生命科学の分野ではこの独立性が担保されない場合が多い。つまり、2頂点間の関係性が、第3の分子や条件に依存して変わりうる。例えば組織によってタンパク質の相互作用相手が変化していたり、観測法によって振る舞いの判断が難しい、といったことがある。

生命科学におけるネットワークの扱いは、人間が把握しやすい情報表現法として、ヒューマンインターフェースに近い利用法となっているのが、現状である。ネットワークによる高度な表現法が登場すれば、システムの振る舞いを予測したり、性能を評価するような解析が期待できる。

同様に、大規模情報を理解する手段として用いられる代表的手法がクラスタリングである。クラスタリングは生命現象にビルトインされているものを見出すというよりは、人間の理解の手助けとして用いられる。そのため最適なクラスタリング結果が生命現象の中に存在すると考えるのではなく、あくまで状況に応じて、便利なものを使い分けることが重要である。実はクラスタリングは情報科学の分野で昔から研究されてきたが、近年の生命科学から生みだされる多様な超大規模データの集合は、これまでとはスケールの異なる対象として新たな注目を集めている。視覚化だけでなく、超大規模データの利用法としても注目すべき分野である。

### 3. 重要な未解決問題（オープンプロブレム）

以上の数理的手法または新たなアプローチにより、今後解決すべき主要な未解決問題や興味深い課題の例をいくつか挙げておきたい。もちろんここに挙げられてない重要問題も数多くあることに留意してほしい。前章であげた項目とおおよそ一致する形式で紹介するが、前章（5）～（7）で挙げた確率モデル、統計と機械学習、ネットワーク分野は最近になって多くの進展がみられる分野であるため、コンセンサスとしての未解決問題を特に挙げないことにする。

- (1) ダイナミクス
- (2) 形態形成・進化
- (3) 脳神経
- (4) エネルギー、ポテンシャル
- (5) 制御、最適化
- (6) 情報処理

#### (1) ダイナミクス

##### ・ノイズ、分子ゆらぎ、細胞間個体差

生体分子の濃度や活性レベルを高時間解像度で計測すると、時間に対して大きく揺らいでいることや、細胞間で大きな差があることが分かってきた。これらの揺らぎやノイズが生物としての振る舞いにどのような効果を持っているのか、十分に理解されていない。分子レベルでノイズがあったとしても、それをうまく解消する機構がある、とする見解や研究がある一方で、むしろ行動決定や分子反応などにおいて、ノイズは生命現象を維持するのに有利に働くとする研究もある。

##### ・情報の不十分性

細胞の振る舞いは、本質的に生体分子のダイナミクスに帰着されると考えられている一方で、ダイナミクスを決定するための情報は、ほとんどの場合十分に与えられていない。例えば、天体の運行のように基礎方程式が与えられており、原則的には時間発展を計算することで、状来（じやうらい）の天体の位置を予測できる、といった状況とは異なっている。実験的に得られる限られた情報から、内部のメカニズムやダイナミクスを推定する、といった方向がありうるが、そもそも情報が不十分な状況で確実な帰結を導く新たな方法論は有り得ないか、問われている。

##### ・種多様性の起源

ダイナミクスの項で詳述したとおり、多数の種を含む生態系がなぜ安定に維持されるのか、本質的な解は与えられていない。これは言い換えるならば、高自由度力学系の解の安定性に関する問題だといえる。

#### (2) 形態形成・進化

##### ・発生／形態形成の安定性

生物の形態形成は、体を構成する細胞がその場所に応じて、「正しく」増殖や移動や変形を行うことでなされる。その過程を明らかにする学問が発生生物学であるといえる。形態形成の様々な局面における細胞レベルの振る舞いが明らかにされてきた。しかし、マク

ロなレベルである個体の形態形成を安定に実現する、マイクロレベルである細胞の振る舞いに関する「一般的条件」がどのようなものであるか、理論的にはまだ理解が進んでいない。

- ・「個体発生が系統発生を繰り返す」理由

「個体発生が系統発生を繰り返す」とは、博物学者ヘッケルの言葉であり、生物の発生過程において、その生物種が進化の中で歩んできた形態変化を観察できる、という主張のことを言う。進化学発生生物学の進んだ現在では、必ずしもこの言説は正しくないとも言われる。また「生物の形態の進化は、前段階の生物の形態形成の後に新規な形質が付加することによってなされる」と考えることで、この傾向は理解できるとも言われる。しかし、なぜ形態進化がそのようなパターンを示すのかに対する本質的な回答はなされていない。

### (3) 脳神経系

- ・記憶のメカニズム

現在の脳科学では、脳の活動の単位となる神経の振る舞いや、その回路（ニューラルネットワーク）について多くのことが分かっている。しかし、我々の記憶が脳のどの部位にどのような形で保持されているのか、全容はまだ分かっていない。

- ・rate/temporal コーディング

脳の活動の単位である神経が発火する機構や、それがシナプスを経て他の神経に伝達される仕組みは良く理解されている。しかし神経の発火パターンにどのように情報が埋め込まれているのか、実はいまだに決着がついていない。発火頻度に情報が込められているとする **Rate coding** と発火の時系列そのものに込められているとする **Temporal coding** が、二つの大きな対立仮説である。

- ・意識の起源

無機質である脳からどのように意識が生まれるのか、意識とは何であるのか。哲学の問題ではあるが、自己を認識するきっかけとなるメカニズムや分子の実体を知りたいと考える生命科学研究者は多い。

### (4) エネルギー、ポテンシャル

- ・タンパク質、DNA の立体構造予測

前章で詳しく説明したが、これは計算速度の問題としての側面と、アミノ酸配列情報からポテンシャル関数を決定する問題としての側面とがある。また **DNA** の核内における折り畳みも、蛋白質構造予測と同様に理論的に大きな課題である。

- ・Drug design

薬効のある新しい化学物質を理論的に予測することが、医科学分野から求められている。分子の構造に基づいた計算を行う分子科学的方法や、過去の知見やデータに基づく情報科学的方法など、複数のアプローチの可能性はある。

- ・力を感じての形態形成制御

上述の発生生物学と関連した課題であるが、骨格や血流など力学的負荷のかかる組織の形成においては、力学的負荷に依存した形態形成がなされることが分かっている。しかしそれを実現するメカニズムや、その機構がどの程度形態形成やその安定性に寄与しているかなど、ほとんど分かっていない。

## (5) 制御・最適化

- ・性の起源（有性生殖の2倍のコスト）

前章で詳述したとおり、意味が明確で分かりやすい一方で、解くことが難しい課題である。

- ・幹細胞の動的制御

多細胞生物では体の体制を維持するために、最終分化する前段階の分裂能を備えた細胞（幹細胞）が、様々な組織に用意されている。例えば皮膚が傷ついてもいずれは治癒するのは、上皮幹細胞が皮膚の奥に存在し、絶えず皮膚の細胞を供給しているからである。この幹細胞は分裂により、分化細胞を供給する役割を持つ一方で、自身が失われることがあってはならない。生体内では組織の状況に応じて、幹細胞の維持や数の調節がされていると考えられるが、そのメカニズムはまだ明らかではない。また現在の医科学における主要な課題は再生医療であるが、この実現のためには幹細胞を人為的に作り、維持する技術が必要となる。生物学的メカニズムの解明と、医療への応用とが密接に結び付いた課題である。

- ・組織のホメオスタシス

多細胞生物の多くは、環境要因たとえば栄養条件の変動により体のサイズが大きく変化しても、各組織のサイズは一定の比例関係を保ちつつ変化する。我々の体を構成する組織や細胞の性質の違いは、細胞分化というローカルでマイクロな過程により作り出されるが、一方で個体全体としてバランスを保つためのグローバルな調節メカニズムが存在することを示唆している。このメカニズムや原理の実体については、関与するシグナル伝達経路が部分的に報告されているものの、全体像の解明にはほど遠い。

- ・人の意思決定と保全生態学

保全生態学は野外の生物種多様性の維持を目的とする分野であるが、現在では生物学の枠を超えて人の意思決定過程を対象として含む、社会学との融合分野へと発展している。ある生態系を保全の目的とした場合に、どのようなインセンティブを与えることが、人間社会にも負担をかけず実用的になるかという、政策決定とも関わる問題となっている。

## (6) 情報処理

- ・一分子計測からの内部ダイナミクスの推定

一分子イメージング解析技術は、生きた細胞内における生体高分子の時空間動態（分子内部動態）を計測するための強力な道具である。データとしては、多くの場合、個々の分子の位置変化（一分子の運動の軌跡）の情報が得られる。このような時系列データから分子運動の動的特性（拡散係数、状態間の遷移速度）を推定するために、新たなデータ解析手法の開発に期待が高まっている。また、大量の顕微鏡動画から定量的な情報を効率的に

取得することが求められるが、バイオイメージインフォマティクスの解析手法の開発ならびに普及が遅れていることも問題視されている。

・GWAS (ゲノムワイドアソシエーションスタディ)

ゲノムデータを網羅的に採取して、疾患などに関連する遺伝子を見つけようとする研究が盛んに行われている。そのとき、データ数は見かけ上は非常に多いのだが、結果の信頼性を高めるための標本数よりも説明変数の数が多すぎるデータであるが故に、信頼性のあるデータ解析ができにくいという問題が生じている。特に、遺伝子間の交互作用まで見る解析は、計算量などの問題から難しい問題として残っている。

・ビッグデータ解析

次世代シーケンサによって次々と生み出される膨大な配列データを、リアルタイムで高精度に処理し、疾患等との関連付けを行うための解析技術はいまだ確立されておらず、一層の発展の余地がある。また、配列データに限らず、病理画像などの画像理解・検索技術もいまだ完全ではなく、深層学習等技術の一層の発展を要する。

以上、生物学の理論に関心のある方、特にこれからそれらを学ぼうと考える方の参考になれば幸いです。

2015年4月16日

数学協働プログラム「数理・生命科学」WG

有田正規 (国立遺伝学研究所/理研)

上村匡 (京都大学)

大槻久 (総合研究大学院大学)

鹿島久嗣 (京都大学)

小林亮 (広島大学)

柴田達夫 (理化学研究所)

藤澤洋徳 (統計数理研究所)

間野修平 (統計数理研究所)

藤森俊彦 (基礎生物学研究所)

望月敦史 (理化学研究所) (代表)

吉田亮 (統計数理研究所)