

(原研CCSE量子生命情報解析グループ<sup>1</sup>、原研NSRC生体物質機能解析グループ<sup>2</sup>、奈良先端大バイオインフォマティクス人材養成ユニット<sup>3</sup>) 郷 信広<sup>1, 2, 3</sup>

【ポストゲノムの生物学】 ヒトをはじめとするかなりの数の生物のゲノム情報が読み取られ、その数は尚急速に増えつつある。ゲノムとは、ある生物の全遺伝情報、具体的には全DNA塩基配列情報、を指す。その塩基配列には、その生物に関する総ての情報が書き込まれているはずなので、それが生物学にもたらす影響の広さ、深さは計り知れない。

現在、塩基配列が文字列として読み取られはしたが、その文字列の持つ意味を読み解く問題は、ほとんど今後に残された(ポストゲノムの)問題である。

ゲノムを構成する全塩基配列上には遺伝子と呼ばれるまとまりがあり、各遺伝子は転写、翻訳されて細胞内で特定の要素機能を担う蛋白質分子(まれにはRNA分子)となる。ゲノム上には各遺伝子の発現(転写、翻訳によって機能分子を細胞内で合成すること)の制御に関する情報も書き込まれている。

高度な制御の下で合成された蛋白質分子、RNA分子は、細胞内においてしばしば特定の複合体を形成し、要素機能を統合した高次の機能を発揮する。細胞は複雑な分画構造を持っている。各分画には多種類の蛋白質分子、RNA分子、さらにはそれらからなる複合体が存在するが、それらは特定の相手分子を探して相互作用しつつ、複雑に階層構造を持つ「生きている状態」を作り出す。

ゲノム情報の読み取りが可能となった現在の生物学の課題は、ゲノム上に要素機能を同定し、それら要素機能の統合としての生物を理解することである。

【生体高分子の立体構造に基づく生物学の位置付け】 上記の背景にある生物の基本的な論理構造を「セントラルドグマ」と言う。基礎に遺伝情報があり、ゲノムは要素機能を担う遺伝子を要素単位として構成されている。遺伝子は細部内で機能を担う実体としての生体高分子(蛋白質やRNA)に翻訳され、それら生体高分子は細胞内で特定の相手分子と相互作用することによって、要素機能を高次機能へと統合していく。

ポストゲノム生物学の大きな課題は、要素機能の統合としての生物を理解することである。そのための研究の方法論は、理念上次の二つの流れがある。(現実の個々の研究は、しばしば両方の側面を持つ。)

第1の方法においては、各遺伝子はその塩基配列をラベルとして持つ操作概念で、塩基配列に変異を加えた時の統合された高次機能への影響を分析することによって、各遺伝子の担う要素機能を同定し、さらに、機能統合の際にいろいろな遺伝子がいかに関わりあうかを解明しようとする。

第2の方法においては、各遺伝子をその産物である実体としての生体高分子と捉え、要素機能をその分子の物理化学的振る舞いで、また機能統合を分子間相互作用として、理解しようとする。細胞内の生体高分子は通常、そのアミノ酸配列、塩基配列によって定まる特異的立体構造を平均構造としてとり、平均構造の周りの微小な、しかし極めて複雑な動きによって機能を発揮する。生体高分子はあくまでも分子であるから、物理学や化学等の物質科学の法則にしたがって振舞う。実体としての生体高分子に基礎を置く第2の流れの生物学においては、配列によって定まる特異的立体構造形成の仕組み、その構造を平均とする構造ダイナミクスとそれに基づく分子間相互作用を始めとする機能発現の仕組みを、物質科学に立脚して解明しようとする。

これら二つの流れの研究手法は、共に、生物個体が多く要素機能の統合体であることに基礎を置いている。これに対して、すべての生物は進化の産物であることに基礎を置く博物学的手法が、第3の研究手法として重要性を増しつつある。

【バイオインフォマティクス研究】 現在、生物を構成している要素とそれらの高次機能への統合に関する膨大な量の情報が蓄積しつつある。もっとも基礎的となるのが、ゲノムを構成するDNAの塩基配列情報で、ヒトをはじめとする多くの生物種のデータが蓄積されつつある。1生物種でも、個

体間に塩基配列の変異がありそれが個性の遺伝的基礎であるが、特にヒトに関して、配列の変異に関する情報が蓄積しつつある。配列情報と共に基礎的なのが、蛋白質の立体構造に関する情報で、これも急速に蓄積情報量が増大しつつある。また、いろいろな条件下の細胞において発現されている mRNA や蛋白質に関する実験データを取りまとめたデータベースも、機能統合に向けた研究推進のため、整備のための努力が払われている。

バイオインフォマティクス研究の目的は、これら蓄積しつつある膨大な情報にアクセスし、博物学的立場で解析することにより、そこに経験則を見出し、さらには、そのような経験則を用いることによって、新たな知見を生み出していくことである。例えば、アミノ酸配列の似ている蛋白質は、その立体構造も機能も似ていることが、経験的に知られている。この経験則は次のように利用できる。新しい生物種のゲノム解析がなされた場合、そこから導かれる新しい蛋白質のアミノ酸配列が、立体構造や機能が既知の蛋白質のアミノ酸配列と似ていれば、立体構造や機能も似ていると推定できる。

基本的にはこのような簡単な論理で、いろいろな生物のゲノム上に同定された遺伝子の機能推定がなされている。しかし、現在は、ゲノム上に同定される遺伝子の約半数は機能がまったく未知である。この状態を改善するためにすべきことは、蓄積された情報の中に、有用な経験則をより多く見出していくことである。

そのような経験則は一般的に「あるデータベース D を構成する要素の中で F と言う(立体構造上の、あるいは機能上の)性質を持つものの中には、P と言う特徴を持つ一群のものがある」という形で表現される。ここで P と言う特徴は、一般には F と言う性質の必要条件でも、十分条件でもない。そのような法則は、弱い法則と呼ぶべきであろう。

経験則が一般に弱い法則であるのには、主に 2 種類の原因がある。第 1 は、F と言う性質を持つものが、いくつかのタイプに分類できる場合が多いと言う事情。第 2 は、個々のタイプの境界が明確にしにくいと言う事情。この 2 種類の事情共に、集合の要素が進化の産物であることに起因している。

経験則の弱さは、それ自体問題ではない。弱い経験則が得られた場合、「F と言う性質を持つもので、P 以外の特徴を持つものには、どのようなものがあるか」を問い、その中から特徴抽出を試み、その結果をまとめて新しい経験則が得られれば、性質 F のための条件の必要性が向上する。十分に性を向上するためには、複数の法則を同時に満たすものを検索する。経験則を集積すれば、個々の経験則の弱さは克服できるであろう。

これがバイオインフォマティクスの研究の基本戦略である。そのためには上記の意味の弱い経験則を網羅的体系的に収集し、計算機上で操作性の良いアルゴリズムの集積として実現していくことが必要である。われわれはそのための統合ソフト BAAQ (Bioinformatics: Ask Any Questions) を開発している。

【生体高分子系のシミュレーション研究】 物理学、化学等の物質科学に立脚して遺伝子産物である生体高分子の機能を解明していく研究において、計算機シミュレーション法の有用性は、計算機のハード、ソフトの進歩の結果、最近著しく向上した。

ポストゲノム生物学の時代に入った現在、シミュレーションによる機能研究のフロンティアの一つは、生体高分子複合体や、生体高分子間相互作用を計算機内の論理空間に構築し、その時間的振る舞いを古典力学に基づく分子動力学法を用いて調べることである。計算機の能力向上の結果、より大きな系のシミュレーションが可能になったので、現実化したフロンティアであるが、それは細胞内の置ける要素機能の統合を調べるためにちょうど必要とされる大きさである。このようにシミュレーション法は、その有用性向上の結果、もっとも最先端の生物学的課題追求の有力な研究手段になったのである。われわれはそのような(溶媒分子を含めて数百万原子からなる)大きな生体高分子系のシミュレーションを遂行するための並列処理に基礎を置いたシステムの開発に取り組んでいる。

構造生物学においても一つのフロンティアは、細胞内に存在する生体高分子複合体の立体構造を決定することであるが、現在可能性が世界的に追求されている一つの方向は、電子顕微鏡で得られる低分解能の構造情報と計算的手法を組み合わせる可能性である。われわれはこれにバイオインフォマティクス手法で得られる経験則も持ち込むことを考えている。