

バイオインフォマティクスの過去，現在， そして未来へ向けて

岡 本 尚

名古屋市立大学大学院医学研究科細胞分子生物学

皿 井 明 倫

九州工業大学情報工学部生命情報工学科

はじめに

政治の世界でも「2大政党」論を巡る議論がかまびすしいが、科学の世界においても二元論的対立がしばしば起こってきた。よく知られているものでは、科学史における経験論と演繹論の相克である。すなわち、技術発展に伴って科学が「進化」を遂げてきたが、その際にはしばしば新しい概念を加えた上での概念的な枠組みの再構成が起こることが契機になってきた。これを「パラダイムシフト」と呼ぶ。しかし、科学の進歩は単に技術の進歩に伴う付随的なものではなく、すぐれて哲学的な要素を持っている。このような「科学革命」の際には、新しい科学の枠組みの構成と展開の仕方を巡って経験論と演繹論との間で議論がらせん状に錯綜し、厳しい批判や試行錯誤が繰り返される。このような激しいやりとりを通し、その苦しみの中から次の世代の新しい学問が誕生するのである。私達もバイオインフォマティクスというまだ揺籃期にある新しい学問の誕生に立ち会っており、この事実は現代に生きる我々の僥倖ともいふべきものである。

新しいイギリス経験論の祖でありシェークスピアの同時代人でもあったフランシス＝ベーコン(1561-1626)は、「知は力なり *Ipsa scientia potestas est*」と述べ、スコラの議論のように一般的原理から結論を導くのではなく、厳密で周到な実験的観察に基づいた認識とその知的な解析および解釈こそが科学認識の根底にあらねばならないと主張した。これに対してルネ＝デカルトは、人間の五感是不確かなものであり、見えているものや聞こえて来るものが本当にそこにそ

ういう形で存在しているかは疑わしい、と考えた。そして、あらゆる経験されて認識されるものを否定し去って行った後に残るものこそが哲学の原点であると考えた後、後世に残る「我れ思う、故に我れ有り *Cogito ergo sum*」ということばを遺した。このような論争は、今日の医学生物学の中にも受け継がれてきている。バイオインフォマティクスの誕生も、このような議論の中から生まれてきたわけである。

今回は、平成20年8月28日の夏休みの中、バイオインフォマティクスの旗手のひとりでもある皿井明倫教授(九州工業大学情報工学部生命情報工学科)に名市大までお越し願ひ、「バイオインフォマティクスの過去・現在・未来」と題して90分にわたる講演を名古屋市立大学医学研究科で開催した。以下は、そのサマリーである。なお、本文の内容は参加者のひとり(岡本)が講演を聴きながら、さらに考えを加えたものも含まれている。セミナーの報告であるので、原則として引用文献は付さないが、この分野に興味を抱いた読者のために、文末に参考図書に記載したのでご利用いただきたい。さらに詳細な情報や出典などを知りたい読者が居れば、個別に著者までご連絡をいただきたい。

ゲノム科学の進歩に伴う情報量の激増とバイオインフォマティクスの誕生

バイオインフォマティクスという学問が誕生した客観的背景として、生物科学の著しい発展に加えて情報科学のめざましい発展があったが、実はそれ以外にも数学、物理学、化学、工学などとも密接につながっている(図1)。す

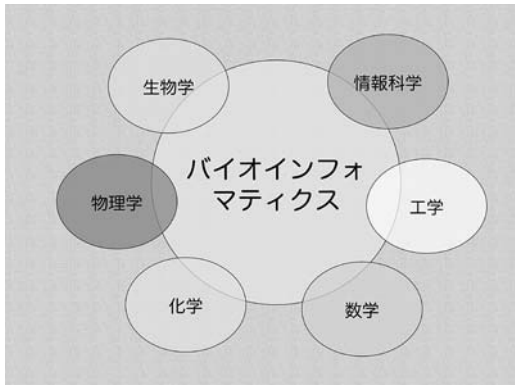


図1 「学際」科学としてのバイオインフォマティクス。

なわち、バイオインフォマティクスとは、計算機科学の著しい工学的発展を技術的な背景として、特に生物の持つ諸現象に対して従来の諸科学を組み合わせることで適用し、生命現象の解明、疾患の解析や治療薬の開発などのために誕生した学問である、と便宜上定義することが可能である。そのため、バイオインフォマティクスは、生物学を中心に、物理学、化学、数学などを有機的に統合するという性格を持つ。また、現時点ではその定義は確立されたものではなく、研究者集団ごとによって異なった概念が存在する、のが現実であろう。

しかし、その成立の背後には、近年のゲノム科学の進展によって得られるデータ量が膨大となり、知識にまでなかなか行き着かないという差し迫った事情があった。図2は、さまざまな生物の遺伝子DNA配列の情報量の年々増大してゆく様子を示している。ゲノム科学によってもたらされるデータを合理的かつ効率よく統合し、解析することによって創薬など疾患の治療に役立つ情報が得られるはずであるが、このままでは雑然と散らかった机の上のような状態であり、なかなか有用な情報(=知識)にたどり着けそうにない。そこで、これらの様々な雑多とも言える情報を統合するために、様々なエビデンスは「情報」という共通のプラットフォーム上に集約され、実験的に得られたある種の法則性のもとに統合される(図3)。



図2 DNA配列情報の年ごとの増加。横軸に観察年、縦軸にはエントリー数(右)と塩基配列数(左)を示した。

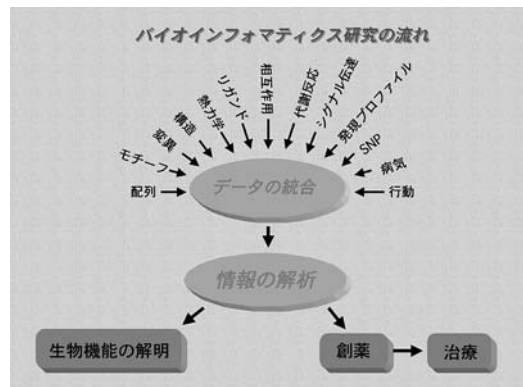


図3 様々な生物情報は、「情報」という共通語に集約され、有用な情報へと変換されてゆく。

新しい「知のプラットフォーム」とバイオインフォマティクス

従来であれば、データから論文ができ、さらにそれが集まって教科書に記載できるような確立された「知識」となり、この新しい知識に触発されて新しい研究が始まる、という知識のスパイラル(図4の知識のスパイラル1)があったが、近年では膨大な情報量のためにこの古典的な知的生産のスパイラルが成り立ちにくくなった。すなわちここにも新しい枠組みが必要となったのである(図5)。この際にツールとして予想以上の成果を挙げたのが「データベース」の構築であった。また、これらの膨大な情報を「合理的」かつ「効率よく」分析する、というところで物理学、数学、化学などの基礎科

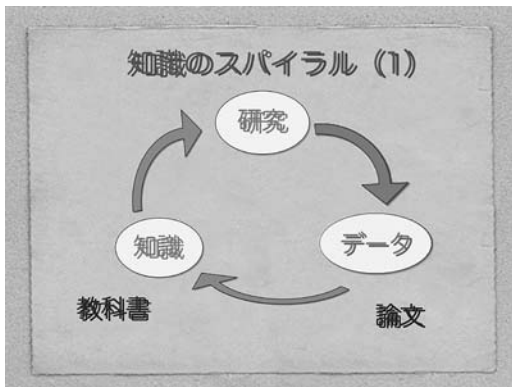


図4 従来の知識のスパイラル。
「研究」結果が蓄積されて「知識」に到ることによって、また、新たな研究の萌芽が出現して、それがやがて集積されて新たな知識となっていく。

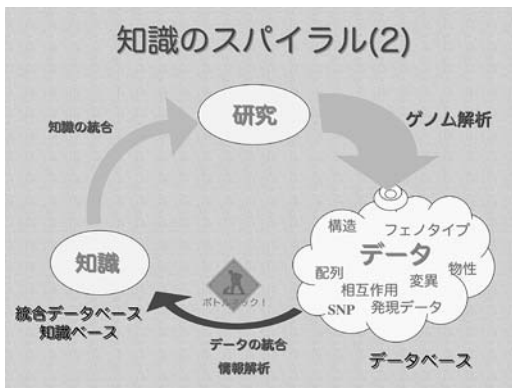


図5 ポストゲノム時代の新しい知識蓄積のためのプラットフォーム。
何故、データベースが必要となったか。また、どのように使うべきか。情報科学の必要性。

学に加えて情報科学に依存する部分が出てくることになる。このように膨大な情報を統合し、多面的な角度から解析することから、複雑に見えた生物機能の解明が進み、これを使った病態解析が可能になる。さらにもうひとつの帰結として、創薬が可能になり、新しい治療法が開発されるのである（詳細は後述する）。

例えば、ある実験から未知の遺伝子の配列を見つけたと仮定しよう。次に行うべき事は、その機能を推定することである（図6）。既知の遺伝子との相同性がBLASTやFASTAサーチ

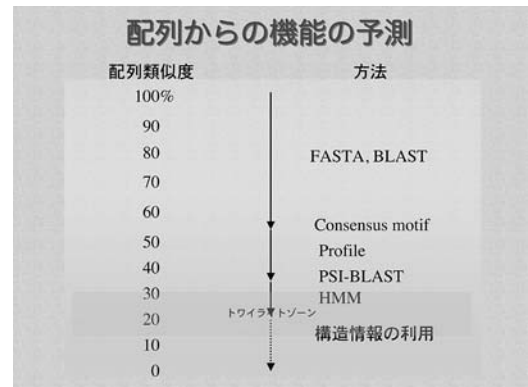


図6 新しい「知のプラットフォーム」の適用の仕方。未知の遺伝子の配列から機能を予測する。

などで70-90%もあれば、その推定はほぼ確実であるが、問題となるのは相同性が20-30%前後の場合（「twilight zone」）である。この課題をクリアするのに導入されたのが、生体高分子をいくつかの断片に分けて考える「モチーフ」レベルでの検索である（図7）。モチーフは類似アミノ酸配列の断片を相互に比較することによって得られ、蓄積されたものであるが、図7に示すように様々なモチーフ検索方法があり、またモチーフ自体の存在様式も一様ではない。さらに、モチーフの検出は分子機能と密接な関係にある構造予測にも繋がり、構造を前提とした分子間認識すなわち分子間結合へと展開して行くことになる（図8）。バイオインフォマティ

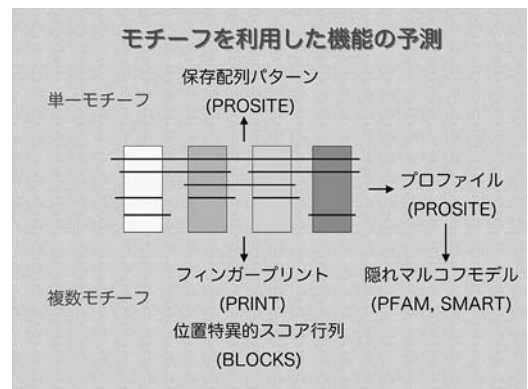


図7 蛋白構造における「モチーフ」。
モチーフの検索から蛋白質分子の機能を予測する。そのための種々の技法を記載した。

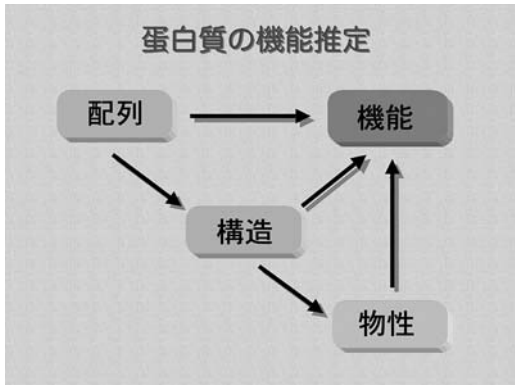


図8 アミノ酸配列から高次構造さらに機能推定へ。

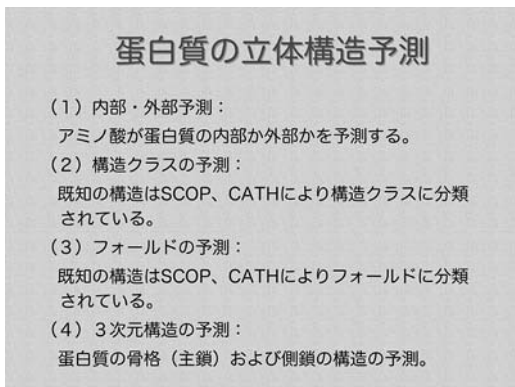


図9 蛋白質の分子立体構造の予測。

クスによる蛋白の立体構造予測の進め方には図9に示したようなものがあるが、その精度はまだ一定しておらず開発の余地が残されている。

蛋白質分子の配列（1次構造）から立体（3次元）構造および物性や機能へ

蛋白を含む生体高分子構造の問題は生体高分子の「物性」の追究へと展開される。やがて「配列」→「構造」→「物性」へと理解を深める過程で、生物種間の比較を行うことにより「進化」の問題を扱うことになる。蛋白質分子を中心にした生体高分子の立体構造解明から分子間相互作用の問題へと移行することになるのであるが、皿井らは通常の物理化学的解析法を越えてニューラルネットワークを用いて計算機に学習させ、既知の蛋白分子の情報をもとに未知の蛋白分子の構造や相互作用分子を合理的に予測す

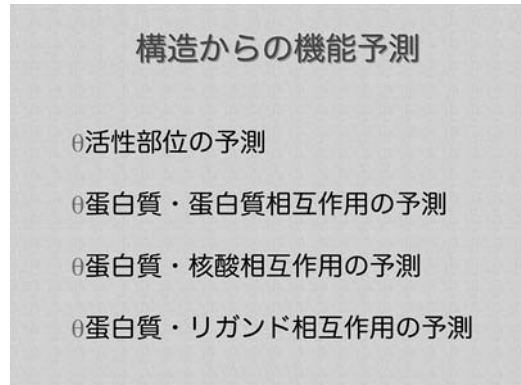


図10 構造から機能へ。


立体構造決定から機能を予測するためには、活性部位や相互作用分子およびその作用の予測などを含めて探索する必要がある。

る方法論も開拓した。しかしながら、あくまで「構造は映画の一コマに過ぎない」と考えられ、生体高分子の構造特異的な相互作用の後に起こる化学反応、リガンド結合、アロステリック効果、induced fitといったより高次の現象に対してチャレンジして行かねばならない（図10）。

システムバイオロジーとの融合

今後このバイオインフォマティクスがより拡張性の高い学問として進むべき方向性としては、システムバイオロジーとの学問的融合ということが挙げられる（図11）。具体的には以下のような項目が挙げられる。（1）ファジー理論や自然言語を用いて遺伝子発現論理の解析を膨大化したマイクロアレーデータの解析に適用して新しい生物原理を追究する（例えば、図12）、（2）ゲノミクス（genomics）からプロテオミクス（proteomics）、生体分子構造の全体（structurome）、相互作用の全体（interactome）、生体反応の全体（metabolome）、表現型の全体（phenome）へと高次生命現象に到る対象を正しく表現する方法を構築する、（3）すでに明らかにされているおよそ800種以上にわたる各種生物のゲノム情報をもとに、進化の過程、生物種の違いなどを明らかにする比較ゲノム解析、などが提唱された（図13）。（3）の比較ゲ

システムバイオロジー



- システムの構造
- システムの特性 (柔軟性と頑強性)
- システムのふるまい
- 生命システムの理解
- 創薬、病気の治療などへの応用

図11 システムバイオロジーとの融合。
バイオインフォマティクスの技術や知識の統合の結果、生命システムを理解するための学問的「枠組み」として新しい生物学が構築されつつある。

マイクロアレイデータのデータマイニング

- ◆ 遺伝子制御モチーフの同定
 - ▼ HMMなど
- ◆ 制御ロジックの解明
 - ▼ ファジィロジック
 - 自然言語処理など




図12 マイクロアレイデータから医学・生物学に有意な情報や概念を抽出する。これを「データマイニング」と呼ぶが、新しい法則性や独特な数学的形式が発見される可能性がある。

ノム解析からは、ある特定の遺伝子が通常の垂直伝搬ではなく水平伝搬することの発見から新しい「病原体学」の可能性も見て取れる。さらに、ある特定の遺伝子が特定の種にだけ存在することをもとに特異的治療が可能になる場合もあるだろう。また、(4) 遺伝子型から表現型の推定 (図14)、すなわち今日盛んに行われている SNP (一塩基多型) 解析の結果から病気の解析さらには特異的 (tailor made) 治療の開発への展開も可能となるであろう。

比較ゲノム解析

異なる生物のゲノム全体を比較して進化的な関係を解析する

- 配列の比較
- 同じ機能をもつ対応する遺伝子 (オースログ) が存在するかどうかを調べる
- 遺伝子構成の比較

図13 比較ゲノム解析から進化の仕組みを知る。生物学的及び生化学的な情報と照合・統合することが必要。
また、遺伝子ファミリーが順次誕生して行く状況を明らかにすることで、個々の遺伝子の担う生物機能をよりよく理解できる。

Genotype → Phenotype

(遺伝子型) (表現型)

Database of gene knockout mouse




図14 遺伝子型と生物学的な表現型の関係。遺伝子ノックアウトやトランスジェニック動物のデータベースも整備されつつある。

ゲノム創薬から理論医学へ

ゲノム創薬、ということばは医学薬学業界にとっては魔法の言葉、夢の言葉であるが、ここでのべた「テーラーメイド医療」という目標は、バイオインフォマティクスにとっては夢に到るまでの途中で実現可能な目標のひとつである。そして、その究極目標は、図15に掲げたように、ゲノム情報をもとに様々な病気の分子標的を明らかにし、標的分子の構造や相互作用分子の情報をもとにして効率的に薬剤を設計することである。

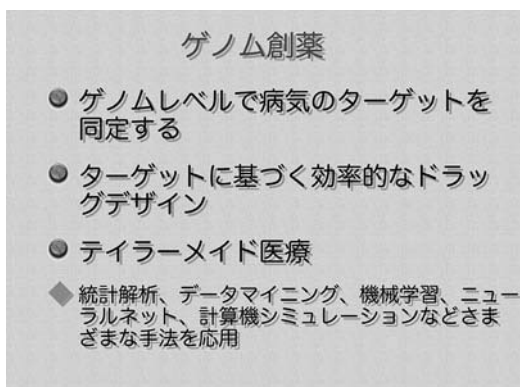


図15 ゲノム情報に基づく創薬。

生化学的およびシステム生物学的な解析から病気の分子標的を導くことや、一塩基多型 (SNP) 解析から責任遺伝子を見つけたたり、個別に薬剤感受性等を推定することによって、より効率的に有効な治療法を組み立てることができるようになった。

図16は、近い将来の薬剤設計の基本的な考え方を示したものである。低分子化合物の集合としての「薬剤」空間と生体高分子およびそれらの相互作用の集合としての「生体蛋白質」空間とを想定し、ふたつの数理的な空間の間に存在する複雑な写像関係を解明することによってゲノム創薬が可能になる。さらに、個々の「空間」を構成する要素と要素間の内部的な関係、また両空間の各要素間の関数関係を明らかにできれば

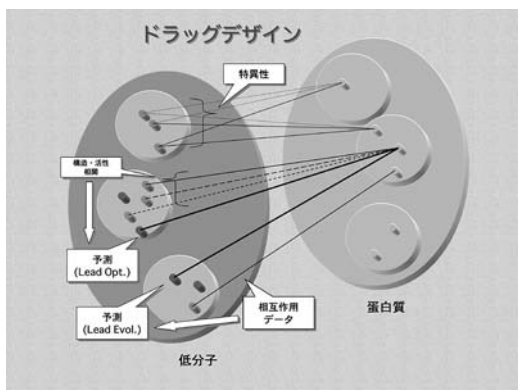


図16 薬剤設計の基本となる考え方。

小分子化合物の集合と蛋白質分子の有機的な集合からなる集合との間で「写像関係」を想定し、その関数関係を求めることで、方程式を解くように治療法を導き出す。

ば、医学の大系を大きく演繹的に再構成することとなるであろう。また、ここにおいて初めて医学における「経験論」と「演繹論」の発展的調和が見られることとなろう。

最後に

このように、バイオインフォマティクスはすでに単なる膨大なゲノム情報の情報処理の手段ではなくなり、生命現象を追究するためのより先鋭的で有力な手段となりつつある。以上、皿井教授が名市大で行った講演のエッセンスを述べると共に著者らの提唱するバイオインフォマティクスおよびシステム生物学を用いた新しい医学の進むべき道を論述してきた。なお、この中で述べたバイオインフォマティクス・ツールは、インターネットを通じて利用することが可能である (BioInfo Bank, [http:// dna01.bse.kyutech.ac.jp/jouhou/jouhoubank.html](http://dna01.bse.kyutech.ac.jp/jouhou/jouhoubank.html))。読者諸賢が、時間のある時に興味のある遺伝子もしくは蛋白質などの情報を入力して、最新のバイオインフォマティクスの進歩を体感してみられることをお勧めしたい。

参考文献

- 1) バイオインフォマティクス David W. Mount 著, メディカル・サイエンス・インターナショナル, 2002
- 2) Bioinformatics and Functional Genomics, J.Pevsner, John Wiley & Sons, Inc. 2003
- 3) Bioinformatics — A Practical Guide to the Analysis of Genes and Proteins—, Eds, A.D. Baxevanis and B. F. F. Ouellette, John Wiley & Sons, Inc.
- 4) 実践バイオインフォマティクス Gibas & Jambeck 著 オライリー・ジャパン, 2002
- 5) 分子生物学のためのバイオインフォマティクス入門 Setubal & Meidanis 著, 共立出版, 2001
- 6) バイオインフォマティクス・コンピューティング B. Bergeron 著, オーム社, 2004
- 7) 実験医学26 (増刊)「バイオデータベースとソフトウェア最前線」森下真一, 阿久津達也 (編) 2008年