



NIRF研究会報告「情報生物学への招待」：学術活動

メタデータ	言語: Japanese 出版者: 福島県立医科大学看護学部 公開日: 2022-04-13 キーワード (Ja): キーワード (En): 作成者: 森, 努 メールアドレス: 所属:
URL	https://fmu.repo.nii.ac.jp/records/2000650

学 術 活 動

NIRF 研究会報告「情報生物学への招待」

森 努（看護学部生命科学部門）

中心性遺伝子の機能解析

私たちの NIRF 研究会は、私が cDNA cloning した NIRF/UHRF2 の機能解析を目標に始められた。NIRF はマルチドメイン型のクロマチン蛋白であり、細胞周期やエピジェネティクスの制御に関わる構造を持っている。従来の分子生物学に従えば、ノックアウトマウス解析によって、分子機能や病態関連性を追求して行くべきところである。しかし私は NIRF のネットワーク機能に注目することを優先すべきと判断した。NIRF に関しては、余りにも多くの重要な、しかも互いに相反する生理機能を持つ遺伝子との相互作用が予想されることから、ミクロレベルでの個別的な分子機能よりも、マクロ的で総合的なネットワーク機能の方が本質的と考えたからである。かくて NIRF を特徴付けるネットワーク中心性の生理的意義を解明することを目標に、システム生物学的方法論の開発を手掛けることになった。

研究会設立の当時、Cell 誌上でシステム生物学に関する議論が交わされていた。いわく、「個々の構成因子の性質をいくら解析しても、システムレベルの性質を明らかにすることは不可能である。」というものであった。例を挙げれば、砂丘に見られる”砂紋”の波のような形状は、砂粒の性質を幾ら綿密に調べても予測することは出来ない。同じように、個々の遺伝子をどれほど厳密に解析しても、生体レベル・社会レベルにまで及ぶ表現型のすべてを予言することは不可能である。

一方、従来のミクロ的解析法とは逆向きの、システムレベルでの現象解析については、「個々の遺伝子解析からは予想も出来ない結論、すなわちパラダイムシフトをもたらすに違いない。」というのがオピニオンリーダーたちの統一見解であった。この議論は、ヒトゲノム解析プロジェクトの終了後に到来する、新しい生物学の時代の到来を予言するものであった。私はこの観点を取り入れ、中心性を持つ遺伝子の機能解明を、研究会として指向することにした。

それではシステムレベルで遺伝子機能解析を行うためには、如何なる手段を採用すれば良いのだろうか？システム生物学の歴史は浅く、目標すら十分に明確ではな

い。そこで私は微視的世界と巨視的現象を繋ぐことで浮かび上がる動的生命現象を正確に描出することを目標に定め、そのための方法論を開発することが私の最初の課題となった。その出発点となるのは、現代分子生物学の常識を問い直すことである。

遺伝情報の複雑さ

遺伝情報の巨大さをご存じだろうか？私たち医療従事者が持つべき基本常識のひとつに、「生体の持つ複雑さの認識」を私は挙げたい。身体を構成する数十兆もの細胞により、私たち人間の肉体は精密に構成されている。たった1個の受精卵に発した細胞群は発生過程で数多くの系譜に分化し、形態的にも機能的にも甚だ大きなバリエーションを形成する。それら多種多様な細胞が緻密な相互作用によって複雑な臓器を構築し、さらに緊密に協調することで恒常性が維持される。特に人間の場合、多彩な細胞間ネットワークが進化し、その果てに中枢神経系と免疫系が産み出された。これら2大ネットワークシステムの構造と機能は、他の臓器組織とは比較にならないほど高度なものである。

細胞組織が構築する人体の仕組みを知れば、誰もが人体の神秘に驚嘆するところであるが、ここで驚くのは余りにも早すぎる。細胞の形態と機能を裏で支える遺伝情報の複雑さは、細胞自体の複雑さを遙かに超えるからである。私たちが意識しないうちにも、個々の細胞中では遺伝子が統括するプランの下に、物質とエネルギーと情報が、厳密な規則と広範な相互関係を維持しつつ稼働している。それら細胞間の情報伝達もまた極めて複雑であって、その多彩で幅広い形式の情報交換を可能とし、全般動態を経時的かつ緻密に監視しているのは、遺伝情報である。

一見シンプルに動作するプログラムの陰に隠れて存在する、外部からは見えにくい機構の複雑さを比喩的に理解するために、パソコン画面上でマウスカーソルを動かすことを考えてみよう。表面的には、手の動きに合わせてカーソルが動くだけの単純な機構に見える。しかしそれを可能とするソフトウェアは BIOS、オペレーティングシステム、アプリケーション、デバイスドライバ等、

多くのレイヤーとモジュールに跨がって構築されており、互いに連携する多くのソースコードを数十万～数百万行も正確に記述することで初めて動作する。他方、ハードウェアにおいても、顕微鏡レベルで設計された微細構造を持つVLSIや記憶媒体、位置情報を検出する光学マウスの介在などがあって初めて、実用可能なレベルで動作する。これら精巧な電子機器を製造する技術の高さと、それを基礎から支える物理工学の奥深さは、ともに最先端を行く現代科学技術の粋である。

ところがパソコンの構造やその上で動作するソフトウェアに比較した場合、細胞の構造機能の複雑さはそれらを圧倒的に上廻るばかりか、そのサイズは遙かに小さい。生体は分子レベル、いや正確には分子下レベルで設計構築がなされている。分子「下」の意味は、遺伝子による生体分子の構造設計と機能支配が、分子を構成する各原子の電子軌道と、それらの量子力学的効果にまで及ぶ様相を指している。ここまで考えれば、細胞の恒常性を支える遺伝情報の量と質がいかに莫大なものであるか、容易に想像が付くであろう。さらに、中枢神経系と免疫系を支える遺伝子群の機能は、生体が生涯のうちに遭遇するであろう、無限に近いほど多様で、しかも新規な外界事象・外来病原体に関する情報を、正確かつ効率的に処理すべく、圧倒的に高機能な複雑システムであり、40億年に渡る自然淘汰を経て精緻に進化した遺伝情報の威力を見せつけている。これら2大ネットワーク組織の出現を可能ならしめた遺伝情報の巨大さに、私は畏敬の念を憶える。

しかしながら以上の認識は、現代生物学では初歩的常識の範疇に入るものであって、それだけでは研究は始動しない。遺伝子の微細機能についての知見を記した文献は既に膨大な量であり、なお日進月歩で増えつつある。それにも関わらず、マクロ的現象を予言することに成功していないではないか。こうした場合、発現する遺伝情報から生じる個別の現象に注目するよりも、すべての遺伝情報に普遍的に共通する性質を見抜くことが肝要であろう。すなわち、遺伝情報の抽象化を最大に押し進めることで明確となる、一般的かつ究極的な原理に依拠することで、生命現象を総体的に予見可能とすることが必要である。こうした方針の行き着く先では、中枢神経に関わる新しい知見に基づいて精神構造への情報学的視点を提供することから、「人間とは何か」という哲学的問いかけに対する解答を、将来的には用意することにも繋がるだろう。

複雑系の統計力学

上述のような、システムを構成する微視的な世界に適

用される法則に基づいて、システム全体の巨視的な特性を導き出す方法は、物理学の主要分野である統計力学に先例を求めることが出来る。統計力学は統計物理学とも呼ばれ、物質を構成する素粒子・原子・分子など無数の粒子が従う法則を確率的に扱うことで、巨視的な諸現象を統計的・確率的に説明しようとする学問である。マクスウェルの気体分子運動論に始まり、ボルツマンとギブズが体系を確立した。統計力学は物理学で最もロバスト（頑健）とされる体系であって、情報理論との結びつきが深いなど、その基本原理が成り立たない系は存在しないと考えられている。

しかし純粋物理学である統計力学や応用数学としての情報理論を、複雑系である遺伝子や人間の特性解明に活用することは、果たして可能なのだろうか？アシモフは統計力学にヒントを得て、その解答をノベルの形で提案した。個々の分子運動は予測不能だが、膨大な分子集団である気体の平均運動は計算可能である。彼は分子を人間に、気体を人間集団に置き換えた。そして人間の感情と反応に規則性を見だし、精神を数学的に厳密な形で記述できれば、膨大な数の人間から成る集団の行動を確率的に予見し得ると考えた。このアイデアがノーベル賞経済学者クルーグマンを魅了し、自らが経済学を志す動機を提供したという逸話は広く知られている。事実、統計力学の適用範囲は極めて広範であり、物理・化学・生物の全ての系で成立する。つまり統計力学的原理は遺伝情報の発現様式や人間の精神構造、ひいては社会構造の形成過程においても、大きく貢献しているのである。ゆえにアシモフの提案にも一考の余地はあり、これら大規模システムの基盤をなす数理法則を見出すことは、幅広い学問分野を前進させるものと期待される。

加えて前述の通り、目に見える現象の陰には遙かに複雑な深層システムが介在することを考慮すれば、1個の遺伝子あるいは多数の遺伝子間相互作用の複雑さは、1人の人間あるいは人間社会のそれを、圧倒的に凌駕するに違いない。人間中心の世界に生きる私たちにとって、にわかには信じ難いことではあるが、生命を稼働させる巨大情報処理システムである遺伝子系の複雑さに比べれば、それがコードする神経組織の構造と機能、人間の精神構造、社会歴史現象などは、いずれも遙かに単純なシステムなのである。したがって、遺伝子系を余すことなく記述できるような統計力学・情報理論を開発することは、人間とその相互関係、さらには人間社会の本質までも理解可能とする手段を創出するための必要条件であると同時に、意外な近道でもあると捉えることができよう。

だがこうした希望的なストーリーは統計力学・情報理論の普遍性を踏まえた論考ではあるものの、これらが生

物学・精神心理学・社会科学の発展に本質的に貢献した事例は乏しいのではあるまいか。この疑問を考察する過程で私は、既に2世紀の歴史を持ち、広範な系に応用されている統計力学が、いまだ未完成であることに気が付いた。複雑な情報を持つ因子間で交換される情報量、及びその帰結についての議論が不十分なのである。この問題の本質は少々難解であるから、まずは遺伝子同士あるいは人間同士の相互関係を例に、背景の説明をさせて頂きたい。ついで従来の統計力学／情報理論の要点と、その問題点、および解決策について述べる。

遺伝子情報の二面性

遺伝子解析の常套手段は、「遺伝子単独の機能」および「遺伝子間相互作用」の両面に着目することである。遺伝子 G_i の機能を知るためには、まず G_i 自身がどのような機能を発揮するかを知らねばならないが、 G_i が他の如何なる遺伝子 G_j と相互作用するかも併せて知る必要がある。後者は、 G_i が別の遺伝子 G_j の機能に何らかの影響を与え、間接的に G_i 自身の機能が発揮されることを意味している。このような二面的な認識手段は、私たちが他人の特徴を知るために日頃から採用しているはずである。ある人物を知る際には、彼・彼女本人の身体的・人格的・精神的・行動様式的な特性に着目するだけでは十分でなく、他の誰とどの様な交流を行い、如何なる社会的役割を果たしているかを把握しなくてはならない。遺伝子の場合も同様であって、多くの遺伝子集団の中で G_i が果たす、一種の社会的機能を同定することが肝心なのである。私は以上の認識の下に、遺伝子が発現する情報を数理的に表現するための、一般性ある手法を開発することに着手した。

遺伝子情報の進化

ゲノム情報に情報理論を適用するに当たって私が先ず取り組んだのは、ヒト染色体上に存在する約20,000種のタンパク遺伝子 (protein-coding gene) と、タンパクを産生しない数万種の non-coding RNA (miRNA, long non-coding RNA) に共通する情報発現様式を抽象化し、その一般的性質を定式化することだった。現在主流の分子生物学では、ゲノム DNA の塩基配列に基づいて遺伝子産物 (タンパクや RNA) が合成され、生成された遺伝子産物が生化学的機能を発揮することで生命現象が進行し、恒常性が維持されることになっている。しかし遺伝子が発揮する情報は、量が巨大であるばかりか、発現様式が極めて多彩である。それらの情報は時間的・空間的に変動するのみならず、量的変化・構造変化・分解制御・

局在変化・複合体形成・クロマチン制御等の幅広い様式で発現し、別々の臓器組織中での相異なる種類の細胞中で、連続的または離散的な形式で同時変化する結果として、形態学的・生理学的・社会生物学的な表現型を形成する。この複雑極まりない遺伝情報に数理解析を適用することは、当然ながら極めて難しい課題である。

私たちはこの難題に対して、遺伝情報の取り扱いを可能とする数理的手法を開発することで解決した。詳細は論文をご覧くださいとして、ここでは要点のみ記載したい。遺伝情報というものは、たとえ僅か一つの遺伝子から発現される情報であっても、従来の情報理論が扱う対象とするには余りに複雑かつ莫大に過ぎる。情報理論を適用して遺伝情報の総体を把握することに成功した例を、寡聞にして私は知らない。

そこで私は、遺伝子が発現する情報を無次元に近い高次元確率変数として扱う方法を考案し、この高次元確率変数 $\mathbf{X} = (X_1, X_2, \dots, X_m)^T$ が表現する情報量を計算する手法を開発した。まず \mathbf{X} の自己情報量を物理系のエネルギーレベルに対応する離散的な n 個の情報レベルと見なす。これらの情報レベルを生起確率順にソートして $n \rightarrow \infty$ の極限を取ると、情報レベルは連続変数 X を用いた表現が可能となる。このとき X の実現確率 $p(X)$ は情報カノニカル分布

$$p(X) = \frac{1}{I} e^{-\frac{X}{I}}$$

に従う。ここで I は X の期待値 (有効自己情報量) である。この I の増加は、基準状態では発生しにくい情報が、より発生し易くなる変化を表現する。野生型の I を 1 と置いたとき、生物に有利な突然変異に伴う I の増分は、1千万分の1から1億分の1程度に過ぎないので、情報レベルの順位変化はほぼ無視できる。野生型遺伝子と変異型遺伝子の情報比較を目的とした場合、これは無次元の複雑な遺伝子情報を、僅か1次元の確率変数として扱うことが可能となったことを意味する。

次いで、遺伝子情報が生物の生存確率、すなわち適応度 (fitness) に影響する様式を定式化した。適応度とは「ある個体が産んだ子供のうち、繁殖年齢まで成長できた子供の数」を意味している。野生型遺伝子と変異型遺伝子が与える fitness の差 (selection coefficient) を s とすると、遺伝的に安定な生物集団では、

$$s = \Delta I$$

が成り立つ。ここでは統計力学のエルゴード仮説に倣い、遺伝子の「1個体における (野生型 or 変異型の) 状態確率が、生物集団中の個体頻度に等しくなる」と想定した。この等式は遺伝子情報と適応度の等価性を示すもので、遺伝子は生き残りに貢献する情報を生み出すことで

進化してきたことになる。

最後に私たちは、日本が誇る「中立説」の木村資生と「ほぼ中立説」の太田朋子が開発した集団遺伝学を、情報理論に適用して微分方程式を作ることで、複雑な遺伝情報が生物進化をもたらす過程を導くことが出来た。この微分方程式の設定には研究会の波江野洋先生に、方程式を解く際には河村隆先生にお世話になった。私自身は理論面の作業だけでなく、プログラムを書くことで、遺伝子進化の過程を可視化する作業も行った。かくて、遺伝子が発現する深淵極まりない情報に対して、正面から情報理論と進化生物学を適用したのは、私たちが世界初である。

遺伝子間相互作用の進化

しかし私たちが得た最大の成果は、個々の遺伝情報の数理解析ではない。ひとつの遺伝子が発現し得る情報量には限界があり、厳しい生存競争を勝ち残るのに十分な適応度を、容易には獲得できないからだ。「分子進化のほぼ中立説」によれば、新たに出現する突然変異の殆どは有害変異 (deleterious mutation) であり、生存に有利な変異 (advantageous mutation) はほとんど発生しない。例外となる場合を除いて、有害変異は遺伝子プールから短期間で排除されるため生物進化には寄与しないのである。それならば、生物進化を促す情報とは一体何なのだろうか。

実はこの問題は「分子進化と表現型進化のギャップ」として、現代生物学上の大問題だったのである。ギャップを示す例を2つ示したい。まずひとつ目は、古生代初頭のカンブリア紀 (5億4200万年前から5億3000万年前) に生じた爆発的生物進化 (カンブリア爆発) である。生命の長い歴史から見れば一瞬とも言える短期間に、今日見られる動物のすべての「門」(つまり身体構造の共通形式) が出揃った。生物を規定するのが遺伝子ならば、生物進化は遺伝子進化によって生じると考えるのは自然であろう。にも関わらず、爆発的遺伝子進化はカンブリア爆発の3億年も前に生じており、カンブリア紀と同時期ではないことが知られるようになった。

もう一つの例は、人類の文化が約5万年前以降に急激に進歩した「大躍進」とも呼ばれる現象である。その詳細についての議論は現在も続いているが、ホモ・サピエンスの登場が僅か25万年前であることも合わせて考えると、ヒト脳神経の機能が、ごく最近になって劇的に進化したことは疑いようがない。脳神経系の複雑極まりない構造を思えば、細胞間ネットワークを構築する分子的基盤に著しい変動が生じたことは間違いないだろう。しかるにタンパク遺伝子の進化では、100アミノ酸あたり1

千万年に1個程度のアミノ酸置換しか生じない上に、神経系遺伝子の進化速度は一般に遅いことが知られている。結論として、これら2つの例に代表される通り、従来は遺伝子進化で生物進化を説明することは非常に困難だった。だが進化に貢献する結果として遺伝情報が実在するのであるから、この困難さは、従来の遺伝子概念に重大な問題が隠れていたことを示唆している。

ここで私たちが見出した解答は、遺伝子同士に**自発的な相互作用**が発生し、従来知られていなかった莫大な情報量をコードするという知見だった。これも論文中に記載したが、単一遺伝子が発現し得る情報量を遙かに超える規模の情報があり、進化の過程で獲得される。私たちが最終的に到達した結論は、遺伝子間相互作用の進化こそが、生物表現型進化の主たる要因であるというものだった。

生物学のパラダイムシフト

私たちは生物進化を遺伝情報の進化とみなすことにより、従来知られて来なかった巨大情報量の存在を明らかにした。この情報の基本単位は、遺伝子間相互作用に起因する**相互情報量** (mutual information, MI) である。遺伝子 G_i と G_j から発現される情報を高次元確率変数 X , Y と表現するとき、それぞれの情報レベルが X_k, Y_l となる実現確率を $p(X_k)$, $p(Y_l)$, それらの同時確率を $p(X_k, Y_l)$, 各々の情報レベル総数を m, n とすると、 G_i と G_j 間の相互情報量は、

$$MI(G_i; G_j) = \sum_{k=1}^m \sum_{l=1}^n p(X_k, Y_l) \log \frac{p(X_k, Y_l)}{p(X_k)p(Y_l)}$$

で表現される。重要なのは、遺伝子間相互情報量 MI が、各遺伝子の自己情報量 I と同様に fitness に貢献することである。2遺伝子間相互作用による fitness の増分 (epstatic fitness) を ε_{evo} とすると、

$$\varepsilon_{evo} = \Delta MI$$

が成立する。この等式は、2遺伝子間相互作用が、個々の遺伝子情報と同じように、適応度に貢献する情報を生み出すことを示している。具体例として、あるリン酸化酵素とそのリン酸化標的となるタンパクが、生体内で時間的・空間的に同期して発現することは、生理的に意味のある情報伝達を促すことで、個体生存に有利となる状況が生じ得るだろう。すなわち**生物進化の過程では、遺伝子だけではなく遺伝子間相互作用も進化する。**

しかし染色体上には数万種の相異なる遺伝子が存在するから、遺伝子間 $MI(G_i; G_j)$ は高次元空間において、複雑なネットワークを形成する。このとき数万に上る遺伝子集団からは、高次元空間における空間配置のバリ

ーションに基づいて、さらに桁違いに大きなネットワーク情報量

$$H(N_i) = \sum^W + \sum^{sgn} + \sum^{rank}$$

が発生する。ここで \sum^W , \sum^{sgn} , \sum^{rank} はそれぞれ、ネットワーク情報量を特徴付けるエッジ重みの情報量総和、方向性情報量の総和、順列情報量である。結果として、遺伝子 G_i を中心とするネットワーク情報量 $H(N_i)$ は、 G_i 自身が単独でコードする情報量よりも遙かに巨大となる。

私たちの数理モデルによれば、タンパクをコードする個々の遺伝子は殆ど進化するにも関わらず、高次元ネットワークは個別の $MI(G_i; G_j)$ の増大に伴って必然的に複雑さを増し、淘汰圧との兼ね合いで生物の表現型進化が促される。かくして生物進化の過程で遺伝子 G_i が獲得する情報のほとんどは、ネットワーク構造を規定するネットワーク情報量 $H(N_i)$ であり、遺伝子産物の分子構造を規定する分子情報量 $H(M_i)$ ではない。 簡単な数値計算を行えば、ネットワーク情報量が分子情報量を遙かに凌駕することを容易に示すことができる。すなわち、

$$H(N_i) \gg H(M_i)$$

である。結論として、現存生物の遺伝子が発現する情報の大部分はネットワーク情報量であり、その基本単位は遺伝子間 MI であることが示された。これは核酸配列を情報の基本単位と見なす、現代分子生物学の遺伝子概念に革新をもたらすものであり、本稿の最初に記した「パラダイムシフト」に相当するものであろう。

Ab initio 遺伝子軌道法

さて、以上の知見を実証し、応用を図る手段として、私は ab initio 遺伝子軌道法 を開発した。これはハイスループットなデータから遺伝子間 MI を効率的かつ定量的に数値計算し、それに基づいて遺伝子のネットワーク機能を実験無しに算出する方法である。この方法は、従来は約10年もの歳月と、大きな人的・財政的投資を必要とした遺伝子機能解析を、ごく短時間の計算で完了するばかりか、原理的に実験の難しい遺伝子についても機能計算を可能とするものである。なおこの名称は、量子化学において、実験値を全く使用せずに分子軌道を計算する *ab initio* 分子軌道法と、同じ数式になることから名付けられた。つまり遺伝子 G_i のネットワーク機能は、 G_i 以外の全遺伝子に与える確率論的作用の総和で表現でき、

$$\varphi[\vec{r}(G_i)] = \sum_{j \neq i} c_{ij} \chi_{ij}(r)$$

となる。この *ab initio* 遺伝子軌道法を応用することで、私は重要な疾患関連遺伝子を同定することに成功し、既に4つの研究拠点と共同研究を進めている。また、本法の計算技法は2件の知財として出願し、先行1件については特許査定が完了している（特許第6820621号）。さらに上記疾患遺伝子に関連して1件の知財を共同出願している。

遺伝情報と統計情報力学

生命は40億年もの地質時代を経て進化してきた。故に遺伝子の機能は深遠であり、容易に人知の及ぶものではない。しかし *ab initio* 遺伝子軌道法の成功により、遺伝子間相互作用が遺伝情報の実質であって、これが生命を規定する情報の一般的特性でもあることが確実となった。翻って、遺伝情報を扱うために開発された高次元情報理論は、高度に複雑なシステムを数理的に解析するために有効な手段であることが示された。将来的には、人間同士の相互関係やマクロ的社会現象のメカニズムを明らかにする学問領域への応用も期待される。しかし私は、遺伝子間相互情報量 MI の性質を調べる過程で、前にも触れた統計力学の問題点を認識するに至ったのである。

情報理論で扱われる情報とは、情報エントロピー H である。情報エントロピー H は、統計力学の熱力学的エントロピー S とともに、両体系を結ぶカナメとなる状態量である。近年、応用数学である情報理論で扱われてきた問題に、物理学である統計力学の手法でアプローチしようとする、情報統計力学と呼ばれる学問分野が形成されてきた。しかし私は逆に、統計力学的現象に確率論・情報理論を適用することによって、物理学をさらに抽象化する方針を採用した。これを統計情報力学と呼称することとする。これは遺伝情報によって起こる現象と、物理学的な現象とを、同等の土俵で議論することを目指したからである。その結果、遺伝子間の相互作用を「力」と表現できるようになった。同様の議論を進めることで、人間同士の相互作用も力と見なすことが可能となるだろう。詳細は論文をご参照頂きたい。

統計力学・情報理論の要点

エントロピーは統計力学と情報理論の双方に登場する指標である。例として気体のような物理系において、構成因子（例えば気体分子）の持つエネルギー E_i が取り得る状態確率を p_i としたとき、熱力学的エントロピー S は、

$$S = -k_B \sum_i p_i \log p_i$$

で表現される。ここで k_B はボルツマン定数である。 S は

物理系の微小構成要素が持つエネルギーを情報量に変換したものである。等温条件下の物理系では、一般に構成因子の持つエネルギー E_i の状態確率 p_i がカノニカル分布 $p_i = \exp(-\beta E_i)/Z$ に従うことから、 S は分子エネルギーの確率分布を反映する。ここで絶対温度を T とすると、 $\beta = 1/k_B T$ であり、また、 $Z = \sum_i \exp(-\beta E_i)$ は分配関数と呼ばれる。

一方、情報理論においては情報エントロピー H が用いられる。これは

$$H = -\sum_i p_i \log p_i$$

と表現される。ある事象 i の生起確率を p_i とすると、自己情報量 $I_i = -\log(p_i)$ は事象 i の起こりにくさを表現する。このことから、情報エントロピー H は事象の生起確率分布に基づいて、「事象の起こりにくさ」を総計した指標と見なすことができるだろう。その一方で、気体分子のエネルギー E_i について、その「起こりにくさ」を総計したものが熱力学的エントロピー S であると考えれば良い。

なお本稿では、情報エントロピー H が、何らかの観測可能な現象と結びつけられる場合に限定して記述するものとする。このとき物理系では、熱力学的エントロピー S が情報エントロピー H と $S = k_B H$ で結ばれる。さらに系の内部エネルギーを U 、ヘルムホルツの自由エネルギーを F とすると、 $U = F + TS = F + k_B TH$ であるから、エントロピーは一般物理学と情報理論を結びつける役割も担うことが判る。

古典統計力学における情報交換

従来の統計力学の問題点とは何だろうか。古典物理学における情報交換の例として、等温条件下での不活性ガスの混合を例に説明したい。2つの不活性ガス X と Y を混合すると、暫くして、ちょうど大気中の窒素と酸素のように、あたかも一種類のガスであるかのように振る舞うようになる。等温だから、 X と Y 分子の平均的な運動速度は混合前後で一定に保たれる。だが大切なのは、 X と Y の間の平均的な相対速度が減少することである。つまり運動の方向が一致するようになる。これは衝突が非弾性的であって、衝突熱が発生し、一部のエネルギーが外界に放出される過程で、平均として相対運動エネルギーが減少することに他ならない。

私たちはこの過程に数理解析を適用し、以下のような結論を得た。この現象では、衝突に際し X と Y の分子間で位置情報・エネルギー情報が $MI(X; Y)$ として交換され、その結果として相対運動エントロピーが減少することが判明した。この時、 X と Y の平均運動エントロピー

$H(X)$ と $H(Y)$ が維持されるにも関わらず、合成系の運動エントロピー総量 = 結合エントロピー $H(X, Y)$ が減少する。 重要なのは、この $H(X, Y)$ の減少が、有名なエントロピー増大則（熱力学第二法則）に一見反するように見えることである。もちろん外界まで含めて考えればエントロピーは増大するのであるが、2つのガスに焦点を当てた際には、局所的にエントロピーが減少するのである。私たちは量子系においても、物体間の情報交換が普遍的に生じることを見出した（論文参照）。すなわち2体 X 、 Y の作る閉鎖系（物質の出入りは無いが、エネルギーの出入りは許され、等温が維持される）においては、2体間相互作用は $MI(X; Y)$ による情報交換を伴い、結合エントロピー $H(X, Y)$ が減少傾向を示す。結論として、私たちが物理的相互作用に情報理論を適用した結果として、従来見逃されてきた $MI(X; Y)$ 生成とエントロピー減少が発見されたことになる。

統計力学・情報理論の問題点

上述した一連の、物理化学系および遺伝子系に関する知見から、従来の統計力学に関して私が着目した問題点は以下の通りである。

① エントロピーが減少する場合を考慮して来なかったこと

統計力学は熱力学と深い繋がりがあり、エントロピー増大則として知られる熱力学第2法則は、その最も重要な原理である。19世紀中半頃にクラウジウスによって導入されたエントロピーは、宇宙全体で増大の方向に向かい、減少することはない。局所的にエントロピーが減少する場合があっても、あくまで例外である。これは現代人の知るべき一般常識でもあるから、物理学では当然のこと、最大限優先される法則であることは言うまでもない。しかしながらこの常識を尊重するあまり、2体閉鎖系での情報交換のような、エントロピーが減少する場合を考慮対象から除外してきたのではなからうか。閉鎖系においては、外界にエネルギーが逃げていく方向に反応が生じる場合があり、このときは2体系のエントロピーが局所的に減少する。（ただし代償的に、エントロピー増加方向への反応も生じることを付け加えておく。論文を参照されたい）

② 情報交換に関する統計力学が整備されてこなかったこと

熱力学的エントロピーと情報エントロピーの間には、常に等価の規則が成立するわけではない。しかし生体とは一種の物理化学的システムであるから、エントロピー増大が宿命付けられているのは当然である。ところがそれに反して、遺伝情報は生体を構築する物理化学的シス

テムの統括を通じて生体に秩序をもたらし、外部環境と内部環境の変動に拮抗することで恒常性維持に働く。情報理論の観点では、秩序形成とは生体エントロピーの減少そのものである。これは具体的には適時適切な生体反応の制御を意味しており、生物個体の生存確率すなわち適応度を向上させる効果を発揮する。

遺伝子による秩序形成の基本は、遺伝子間相互情報量 $MI(G_i; G_j)$ であり、結合エントロピー $H(G_i, G_j)$ を減少させることで、2 遺伝子間の相互関係に秩序をもたらしている。すなわち、

$$H(G_i, G_j) = H(G_i) + H(G_j) - MI(G_i; G_j)$$

さらに一般的には、遺伝子情報は生体エントロピーを減少させる。生物身体の情報エントロピーを $H(B)$ 、遺伝子 G の情報エントロピーを $H(G)$ 、2 者間の相互情報量を $MI(G; B)$ とすると、結合エントロピー $H(G, B)$ は

$$H(G, B) = H(G) + H(B) - MI(G; B)$$

であり、遺伝子と身体の情報交換によって生体エントロピーが減少することを表現できる。これが情報理論に基づいた遺伝子の本質の描出であり、遺伝子に関する従来のパラダイムを根底から書き換える結論であることは言うまでもない。

シャノンの情報理論においては、基本的に1因子のエントロピーが先行して扱われており、2因子間の相互作用を確率論的に扱う手法は、どちらかと言えば後回しにされて来た。私は遺伝子間相互作用を解析する過程で、遺伝子間相互情報量 $MI(G_i; G_j)$ が自発的に増大する傾向を持つことに気付いた。重要なのは、 $MI(G_i; G_j)$ の自発的増大は結合エントロピー $H(G_i, G_j)$ の自発的減少を導くことである。この $\Delta MI(G_i; G_j) > 0$ 、すなわち遺伝子間相互作用の発生が、一種の物理化学的システムである生物系で普遍的に観察されることを考慮すると、自然界でも一般的に自発的に生じているのではなかろうか。私はこの問題が生物学に留まらず、物理化学系でも、さらには冒頭に述べた多くの複雑システムでも重要であることを認識した。そこで論文では $MI(G_i; G_j)$ の増大と、それに伴う結合エントロピー $H(G_i, G_j)$ の減少が様々な系で生じることを示した。得られた知見を総合して判断すると、情報を持つ因子間の相互作用を扱う統計情報物理学を、新規の学問分野として開拓する必要がある。この必要性は上に示した古典物理系だけではなく、量子物理系でも同様である。量子力学の応用である量子統計物理学に足りない部分があり、2量子間相互作用を精密に扱う量子統計情報物理学を新たに開拓しなければならない。論文ではこの点で特に重点的な検討を加えた。

③ 複雑系を扱う統計力学が整備されてこなかったこと

従来の統計力学においては、物理系を構成する微粒子が均一な場合を主に想定している。例えば電子の場合を考えてみると、これは素粒子であるから、宇宙の何処で測定しても、同一の大きさ・同一質量・同一電荷を持っている。これら電子同士は区別が不可能である。しかし生体を構成する遺伝子は、すべて互いに異なる構造の遺伝子産物を、互いに異なる時間的空間的条件下で産生する。人間個体同士を比較しても、遺伝的特性・身体的特徴・人格的特性・居住環境・従事する仕事など、まったく異なるものである。

均一な構成因子レベルの議論に留まるならいざ知らず、私たちに必要なのは複雑系を支配する一般的な数理法則である。そのためには、個々の因子自身が複雑かつ巨大な情報量を持つ場合を想定した上で、それらの因子間に働く、さらに格段に複雑な相互作用を数学的に厳密に記述する必要がある。それに対して、従来の統計力学と情報理論は、①それ自身が複雑な情報を持つような因子同士の相互作用を扱うことができず、②自発的信息交換に因って生じる結合エントロピー減少に関する議論が欠けており、③多体系の議論には適さない。そこで私は複雑系を構成する多数の、それ自身が複雑な情報因子間での情報交換を扱う数理体系の構築を目指し、さらにその実証手段の開発を試みることにした。

私は以上の観点を踏まえ、従来の統計力学を超えた統計情報物理学を創始する必要があると判断した。さらに物理学領域への応用を通じて、その正統性を確認する方針を採用した。得られた知見の詳細については、またの機会にお伝えしたい。

情報物理学への招待

以上、私が初期に手掛けた情報生物学から得られた知見を中心に紹介し、実験の困難さゆえに *ab initio* 計算によって初めて発見された疾患遺伝子とその病態形成機序、informatics の理論体系と実践方法および実施結果、多次元情報理論で見出された多くの新定理などは論文中にのみ記載した。しかし本質的に重要な発見は情報物理学で得られたものであり、本稿は論文中の約40分の1程度の内容しか紹介していない。私自身は情報生物学の正確な構築を意図したが、理論体系の正統性を示すには、物理系に情報理論を適用した上での validation が望ましい。そこで敢えて生物学のみでは完結させず、既存物理学体系に存在する問題を解決することを通して、理論体系の正しさを証明しようと試みた。物理学には多くの難問が未解決のまま残されている。宇宙全体の物質エネルギーのうち、説明が可能な「物質」は僅か5%に満たず、

残り95%以上は未知のままである。何十年も掛けて準備され、国際協力により莫大な経費を投じて建設された、LHCでの超対称性粒子検出実験や、3代に渡るカミオカンデでの陽子崩壊検出実験などがいずれも現時点では成功せず、現代物理学は壁に突き当たっている。「物理学史上、最も混乱」した「まさに暗黒の時代」とまで言われる危機的状况である。しかし私たちの統計情報力学を適用することで、多くの問題が解決可能ではないかと考えている。

最後に

ともあれ、これまでに行った多角的な検討の結果、遺伝子機能の数理的解析法の正統性はほぼ確定した模様である。NIRFの機能も、*ab initio*法による理論計算と実験結果の一致が見られ、満足の行く結論が得られた。ご協力頂いた研究会の皆様には、長期に渡る研究活動をご支援下さったことに、心から感謝を申し上げたい。研究会発足時からの主要メンバーである池田大祐先生とはSkypeを通して議論してきた。*ab initio*計算を行うにあたっては、データセットの選択やp値計算の妥当性、遺伝子解析結果の解釈などについて、あらゆる角度から激論を重ねたのが懐かしい思い出である。物理学では河村隆先生が、私の理論を理解できる世界唯一の人物であり、アインシュタインの相談相手だったベッソンの役割を果たして下さいました。彼には特に数式計算を担当して頂いた。数式が書き込まれた膨大な計算用紙がテーブル上に^{うずたか}に堆く載っている様子を是非ご覧頂きたいと思う。波江野洋先生とは、がんや生物の遺伝子進化について沢山お教え頂き、また多くの議論を重ねてきた。上述した通り、遺伝子進化が1億分の1程度という僅かなfitness変化で生じることを酒の席で教えて頂いたときは衝撃を受けた。合山進先生は白血病研究の権威である。彼は*ab initio*計算が実験データに良く合ったことから、本法の意義を誰よりも早く認識して下さいました。「分子間相互作用のみに注目するという発想が斬新ですね」とのお言葉を頂き、私の心の支えになった。それ以来、長く共同研究を続けて下さっている。池田和彦先生は血液学の大家であり、優秀な研究者であると同時に笑顔の素敵な臨床医でもある。医師でありながら宇宙論にもご興味をお持ちで、早くから拙論に耳を傾けて下さり、余所との共同研究でも陰に陽にお力添え頂いた。以上ご紹介した方々以外でも研究会メンバーは全員が素晴らしい才人ばかりである。彼らに敬意を持って接して来られたことが、上手く行った理由と振り返っている。少々辛かったのは、研究過程で得られた多くの発見について、インパクトが大き過ぎると判断されたことから、条件が揃うまで公開

できなかったことである。今後の研究会活動は、再び以前のようにオープンな交流に戻して行うこととし、現在も残された数多くの未検討遺伝子の機能解析と病態関連性・治療標的の検索を続けるとともに、物理学史上初となる単一かつ単純な数式を採用した情報統一理論の議論を展開していきたいと願っている。