

## 解説

## 生命科学が情報科学に期待するもの

—ヒト・ゲノムプロジェクトを契機とする  
新しい学問の流れ—†

五條堀 孝†† 内藤公敏††† 郷 通子††††

生物の遺伝情報は一般に DNA という生体高分子によって担われている。この DNA は、A, T, G, C の 4 種類のアルファベットで記述される塩基の配列であり、長大な文字列と情報的には同等なものといえる。特に、ヒトの DNA の文字列を完全に決定してしまおうという計画が「ヒト・ゲノムプロジェクト」である。実際には、ヒトだけでなくマウスやショウジョウバエ・酵母・線虫そしてある種の高等植物（恐らくアラビドプシス）などの遺伝情報も完全に決定しようという壮大なプロジェクトである。ヒトの DNA は、総計  $3 \times 10^9$  個の塩基対からなる巨大な情報分子であり(表-1)、数万個にのぼる遺伝子が含まれていると考えられている。この中にはアルツハイマ型老人性痴呆症、ダウン症、筋ジストロフィなどをはじめとする疾病に関係した遺伝子だけでも 4000 個以上も存在し、医学的にもその診断や治療に役立てるために、その情報の解読が期待されている。さらに、生物個体の発生や分化、そして脳や神経などの高次機能の解明や生命の誕生及び生物進化の理解などについても、現在の知識をはるかに超えた研究の進展やそれに基づく応用分野としての新しい基幹産業の創成が予想されている。

一方、DNA データバンクという国際的な協力機関のコンピュータに蓄積されている全生物の塩基数は

表-1 各種生物のゲノム DNA に含まれる塩基対

大腸菌	$4.7 \times 10^6$
酵母	$14 \times 10^6$
アラビドプシス	$70 \times 10^6$
線虫	$80 \times 10^6$
ショウジョウバエ	$105 \times 10^6$
マウス	$2000 \times 10^6$
ヒト	$3000 \times 10^6$

$10^7$  のオーダーに過ぎず、現在よりも 2~3 桁以上も大量な情報を自由に取り扱える情報処理系の開発が急務となっている。すでに、アメリカ合衆国では、パラレルコンピュータやスーパーコンピュータを用いて、これらの巨大な DNA 情報のどの部分に生物機能を有するいわゆる「遺伝子」が存在するか、あるいはそれらをコントロールする調節領域がどこにあるかなどを理論的に予測するソフトウェアの開発や専用計算機の開発に着手しつつある。とくに、ソフトウェアの開発では、人工知能やニューロ・コンピュータの活用及び異種間データベースのクロス・レファレンスシステムの応用が、このプロジェクトへの参加を目的として具体的に議論されている。

たとえば、筆者らが参加した国際シンポジウム(「巨大分子・遺伝子・コンピュータ」1989年8月米国ニューハンプシャー州ウォータービルバレーで開催)は、コンピュータサイエンス、数学、遺伝学、医学、分子生物学、生化学にわたる境界領域での、研究者間の情報交換を活発化することを目的に開催された。約 280 名の参加があったが、日本からは著者たち 3 名が参加した。ヒト・ゲノムプロジェクトが、種々の分野の若手研究者から強い関心を得ていることが、とくに印象に残った。若手の参加者は相当数であったが、数人ごとに分かれてディナを食べながらの討論の場で、彼らはこのプロジェクトの意義について熱心に質問し、議論していた。参加者の専門分野は多岐にわたっていた。分子生物学、生物物理学、分子系統学、生化学、統計学の分野からの参加は期待されていたとおりであったが、予想外にも参加者のほぼ 1/3 がコンピュータ分野からであった。この事実からも、多くのコンピュータ科学者がすでに、生命科学に向かって、動きだしていることがうかがえた。シンポジウムでは、①分子間認識、②調節シグナル・トランスダクション・ネットワーク、③配列比較解析、④機能パターン相互関係と同定、⑤ゲノム構造、⑥データベース:

† Contribution of Information Science is Expected in Human Genome Project by Takashi GOJOBORI (National Institute of Genetics), Kimitoshi NAITOU (Facom-Hitac Limited) and Mitiko GO (Nagoya University, Faculty of Science, Department of Biology).

†† 国立遺伝学研究所進化遺伝研究部門

††† ファコム・ハイタック(株)

†††† 名古屋大学理学部生物学科

表-2 シンポジウムの主なテーマと内容

セッション	内 容
分子間認識	抗原抗体反応, 転移 RNA の分子認識, ヴィリンへの F アクチン結合部位の同定, タンパク質間とタンパク質と基質間の分子認識モデル
調節シグナル・トランスダクション・ネットワーク	タンパク質による遺伝子発現制御コードの認識, ペプチドフラグメントのコンフォメーション, 転写制御因子 AP-2 の制御と発現, エイズウィルスの LTR の構造と機能
配列比較解析	配列のアラインメントと統計的有意性, アミノ酸配列上での電荷分布と機能予測, アミノ酸配列から構造パターンの認識, 塩基の配列に特異的な化学プローブによるリボソーム RNA の構造と機能の解析, ヒトの進化, 塩基組成を考慮した節約法による DNA 系統樹
機能パターンの相互関係と同定	タンパク質の2次構造予測法の進展, タンパク質1次構造からの機能パターンの同定, タンパク質複合体の解析法の技術, タンパク質のおれたたみと配列パターン, 配列パターンに基づくタンパク質立体構造の予測
ゲノム構造	配列解析のための特殊なコプロセッサ, PCR 使用による遺伝解析, 比較遺伝子マッピングとゲノム構成と進化, タンパク質のモジュール構造と進化, ヒト遺伝子のリンケージマッピングの5年間の成果, 真核生物ゲノム配列内で新しい遺伝子発見のためのアルゴリズム
データベース: 相互接続, ユティリティとマネージメント	大腸菌ゲノムのデータベースの展望, 塩基配列データベースの解析手続きの自動的軽減化, タンパク質の配列と構造の特徴: 統合的データベース, 1次構造モチーフの構築と評価
将来の方向	触媒作用をもつ抗体: 展望と期待, タンパク質モデリングの応用: 配列から細胞へ, タンパク質構造のリレーショナルデータベースと予測のための利用, ヒト・ゲノム計画の終了後の生物学への展望

相互接続, ユティリティとマネージメント, ⑦将来の方向, の7つのセッション別に(表-2), 総勢約40名の著名な研究者が講演した。この中に, DNA 塩基配列決定法の開発でノーベル賞を受賞した W. ギルバート博士(ハーバード大)をはじめ, 免疫グロブリンの遺伝子構造の研究で多くの業績をあげている L. フッド博士(カリフォルニア工科大), 分子遺伝学の S. プレナー博士(英国 MRC) などの実験研究者の講演もあった。著者の一人(郷)は⑥で「タンパク質のモジュール構造と進化」について講演した。また, ワークショップは, 「コンピュータ研究資源」と題して, 会場に多数用意されたワークステーションを用いて, 解析用ソフトウェアのデモンストレーションを中心に, 約30のグループが参加して行われた。このほか, ポスタによる一般発表も行われ, 連日活発な情報交換や熱心な議論が, 会場のどこでも見られ, 非常に活気に満ちたシンポジウムであった。

このように, ヒト・ゲノムプロジェクトを契機として, 生命科学の挑戦的要請と情報科学のそれに応じられる知的資産の蓄積とが見事にかみ合せて, 両分野の融合が今まさに大規模なスケールで起ころうとしている。つまり, 大容量の生物情報を収集, 処理, 解析する「ニュー・バイオ・インフォメーション・サイエンス」とか「バイオ・インフォーマテクス」とでも呼ぶべき新しい研究分野が誕生しつつあるのである。

ヒト・ゲノムプロジェクトが, アメリカ合衆国の国家プロジェクトとして巨額な予算の下にすでに始動しており, HUGO (Human Genome Organization) と

いう世界的な協力組織もその活動を充実しつつある現在, 日本の生命科学と情報科学の両分野の研究者が有機的に連携・協力して, この分野の発展を緊急に推進することがきわめて重要となるであろう。そうでなければ, 世界における日本の役割が, 基礎科学の分野においても, 厳しく問われかねないのではないだろうか。

「生物界のもつ大量の情報を定量化し, それに必要な処理と判断を加え, もし必要なら生物界にフィードバックするという一貫した情報処理システムの基盤を, 生物学と非生物学の科学者が協力して早く確立しなければならない。」と10余年も前に和田昭允先生(東大理学部, 現在相模中研)が「ライフサイエンスの振興とコンピューター」と題して, 一般の人々に広く訴えられた(1978年3月30日朝日新聞夕刊)言葉が, 今まさに新鮮によみがえる。塩基配列の決定を自動的に行う装置開発を, ヒト・ゲノム計画に先駆けて進められたのが, 和田先生であった<sup>1)</sup>ことを, 指摘しておきたい。

## 参 考 文 献

- 1) Wada, A.: Automated High-Speed DNA Sequencing, Nature, Vol. 325, pp. 771-772 (1987).
- 2) 「塩基配列とコンピューター」に関連した参考書として, 著者の目にとまった2冊をあげておく。Bell, G. I. and Marr, T. G. (eds.): Computers and DNA, Addison-Wesley Pub. Comp. (1990), Weir, B.S.: Gene Data Analysis, Sinauer Assoc. Inc., Sunderland, MA (1990).

(平成2年5月7日受付)