

ゲノム情報科学における我が国の戦略について

平成12年11月17日

科学技術会議

ライフサイエンス部会

ゲノム科学委員会

目 次

はじめに

1. ゲノム情報科学の現状と将来展望について

- 1-1. ゲノム情報科学とは何か
- 1-2. 将来展望
- 1-3. 基本的な認識（論点の整理）

2. 人材養成について

- 2-1. 「バイオのための情報科学」の人材養成について
 - 2-1-1. 即戦力の養成
 - 2-1-2. 研究交流、試行錯誤の「場」としての拠点形成
 - 2-1-3. 大学等における中長期的視点に立った人材の育成
- 2-2. 「情報論的解釈による生命科学」の人材養成について
 - 2-2-1. ゲノム情報科学の拠点の整備
 - 2-2-2. 大学等における人材養成
-大学院の専攻や学部の学科の整備等-

3. データベース構築の体制について

4. 情報解析技術の開発について

5. まとめ

おわりに

- (別添1)大学等における「バイオのための情報科学」の人材養成に関する私案
- (別添2)大学等における「情報論的解釈による生命科学」の人材養成に関する私案
- (別添3)国家的に支援すべきデータベースとその規模に関する私案
- (別添4)国家的に支援すべき情報解析技術とその規模に関する私案

はじめに

ヒトゲノムの概要配列決定を受け、また、その他数多くの生物の配列決定を受け、遺伝子発現解析、タンパク質の構造決定、プロテオーム解析、分子間相互作用解析等をはじめとした、いわゆるポストゲノムシーケンス研究が激しい国際競争の中で本格化してきた。

このようなゲノム研究の急激な進展及びそれを取り巻く学問的社会的環境の大きな変化の中で、膨大で多種多様な生物情報を効率よく整理・解析し、その生物学的医学的意味を明らかにすることを目的とするゲノム情報科学の重要性がますます高まってきている。

「ゲノム科学に関する研究開発についての長期的な考え方（ゲノム科学委員会、平成10年6月29日策定）」にも明確に述べられているように、ゲノム情報科学はこのようなポストゲノムシーケンス研究を効率よく進めるための道具としての役割だけでなく、ゲノム科学ひいては将来の生命科学の中核を担う存在としても大きな期待が寄せられている。

しかしながら、その重要性にもかかわらず、これまで、ゲノム情報科学における大局的長期的戦略が我が国には欠如していた。

ゲノム情報科学とは具体的にどのような研究分野であるか、また、今後どのように展開して行くか等の詳細については、次節以下で述べることとするが、ゲノム情報科学の研究開発課題はおおよそ以下の2つに分けることができる。

- ・ゲノム配列データを始めとする多種多様で膨大な生命情報をどのように整理統合しデータベース化するかという課題（データベース構築）
- ・そのデータベースから生物学医学に関する有用な知識を導き出すとともにそれを検証するための情報解析技術をいかにして開発するかという課題（情報解析技術開発）

このような2つの課題のいずれに対しても、わが国では、これまで基本的には研究者個人の創意工夫や努力に頼る形で研究開発が進められてきた。

しかしながら、近年の生命科学における情報の種類の多様化とデータ量の爆発的増加は一個人や一組織でこれらの課題に取り組むことを困難にしつつあり、早急に国家的な支援体制を検討すべき状況が生じてきた。

なぜなら、世界的に競争力のある大規模なデータベースの構築や維持管理には大型の計算機システムはもちろんのこと、多くの研究者やデータ作成者がチームを組んでこれに取り組むことが欠かせないからである。また、先進的な情

報解析技術の開発においても大規模な並列コンピュータと多くのソフトウェア技術者を必要とする場合が多いからである。

ところで、ゲノム情報科学において、データベース構築を行うにせよ情報解析技術の開発を行うにせよ、最も重要なことはそれらを担う人材（研究者、技術者）をどのようにして確保するかという点である。

ゲノム情報科学ではどのような課題に取り組む場合においても生命科学と情報科学との両方の知識や技術を備えた人材を数多く必要とする。このような人材がいなければ、データベース構築や情報解析技術開発に対してどれだけ国家的支援策を講じても意味がない。

しかし、残念ながらゲノム情報科学を遂行できる人材はわが国にはほとんどいないのが現状である。早急に対策を講じる必要がある。

冒頭に述べたように、ゲノム情報科学がゲノム科学ひいては将来の生命科学の中核を担う存在になって行くものとするれば、この問題はゲノム情報科学だけに留まらない。

ゲノム情報科学の人材養成は今後の生命科学およびそれを基盤としたバイオ産業の発展にとって最も重要で本質的な問題と言えよう。すでに、ゲノム配列決定やタンパク質立体構造決定の現場における人材不足は大きな社会問題化しつつあることからこのことは明らかである。人材養成には一般に長い時間を必要とすることを考慮すると、すぐにでも国家的な視野で人材養成の戦略を立てる必要がある。

さて、これまで述べてきたような状況認識を背景として、我が国におけるゲノム情報科学に係る戦略、特に上記の3点（人材養成、データベース構築、情報解析技術開発）について、今後の在り方や支援方策を議論するために、ゲノム科学委員会にゲノム情報科学ワーキンググループが設けられた。

本報告書は、約4ヶ月にわたるゲノム情報科学ワーキンググループの議論を踏まえ、我が国におけるゲノム情報科学の今後の在り方及びそのために取るべき方策をまとめたものである。

1. ゲノム情報科学の現状と将来展望について

1-1. ゲノム情報科学とは何か

ゲノムインフォマティクス及びその日本語訳であるゲノム情報学は、ゲノム情報（主にDNA配列やアミノ酸配列）の構造や機能を効率よく整理し、解析するためのデータベースやソフトウェアツールの開発研究から派生して成長発展してきた研究分野である。

「インフォマティクス」という言葉は、情報の単なる整理のイメージが強いことから、また、単なるツールとしてだけでなく、ゲノムや生命を理解する方法論や基礎理論であるという側面を強調する意味から、最近では、「ゲノム情報科学」という言葉が使われるようになってきた（ただし、これに対応すると思われるgenome information scienceという言葉はまだ外国では定着していない）。

ゲノム情報科学に類する言葉として、先に述べたゲノム情報学の他に、バイオインフォマティクス、バイオコンピューティング、コンピューショナルバイオロジー、インシリコバイオロジー、システムバイオロジー、生物情報学、生命情報学、情報生物学、計算生物学、ゲノム生物学等がある。

ゲノム情報科学は、これまで主に配列情報をその研究対象としてきたが、最近では、遺伝子発現や分子間相互作用等の細胞レベルへとその対象を急速に拡大しつつある。さらに、まだ緒についたばかりとはいえ発生のシミュレーション等個体レベルへもその対象を広げつつある。したがって、現在、ゲノム情報科学という研究分野は多くの場合、生物（バイオ）情報科学、生命情報科学を包含していると考えらるべきである。

このような生命の階層における対象の広がりとは別に、従来どちらかといえば独立に研究が進められてきたタンパク質立体構造の理論研究や進化研究との融合も図られようとしている。また、SNPと疾患や薬剤感受性との関係解明等医学的な視点からの研究も始まりつつある。

これらをまとめると、ゲノム情報科学とは、ゲノムの配列情報、遺伝子発現情報、プロテオーム情報、タンパク質立体構造情報、分子間相互作用情報、ゲノムの多型情報等の空間的にも時間的にも多様かつ膨大なデータの情報科学的解析を通して、個々の遺伝子やタンパク質の構造や機能を解明するとともに、最終的には生命全体をそれらの遺伝子やタンパク質が織りなすシステムとして理解することを目指す研究分野と捉えることができよう。

生命をシステムとして理解するということは、ゲノム情報科学においては、（生命現象には数多くの階層性があるがそれらの）階層性に応じて、生命系を

計算機の中に再構築しシミュレーションするという他にない。また、進化のミッシングリングを予見するような生物進化のシミュレーション等もこの範疇に含まれよう。

このことを踏まえてもう一度まとめると、ゲノム情報科学とは多様性、階層性に富む生命現象を様々な視点から情報科学的アプローチで整理解析し、その中から共通的あるいは個別的な特徴（知識）を抽出して行く研究分野と言えよう。

さて、このようにいろいろな意味で急速に拡大・変貌を遂げつつあるゲノム情報科学の戦略を考えるには、目先の需要や技術だけに捕らわれることなく、10年後にこの研究分野及びこれを取り巻く状況や学問的社会的要請がどのようになっているのか、そのイメージをできるだけ具体的に描き、それに沿って骨太の戦略を立てることが重要である。

本報告書では、ゲノム情報科学の長期的なビジョンを示した上で、今すぐに何をなすべきか、数年後にはどのような手を打つべきか、各提案毎にその意図するところ明らかにしつつ、今後の方向性を示そうとするものである。

1-2. 将来展望

「ゲノム科学に関する研究開発についての長期的考え方（ゲノム科学委員会、平成10年6月29日策定）」に述べられているとおり、将来の生命科学は、機能予測を始め、その多くがゲノム情報科学に含まれる可能性がある。しかし、それはすぐに到達できることではなく、まず、5～10年先にはおおよそ以下に示す方向に研究が展開して行くものと思われる。

まず、生物学の視点で考えると、配列レベル、分子レベルでは、*ab initio*な（実際の生体内の分子メカニズムに基づく）遺伝子構造予測及びそれに基づく遺伝子発現の時間的・空間的特異性の予測、自由なタンパク質酵素反応の設計等の研究が活発化するであろう。また、細胞レベルでは、細胞全体の計算機シミュレーションが個体レベルでは発生や分化の計算機シミュレーションが大きな研究課題となっているであろう。また、進化の過程を計算機で再現する試みも活発化するであろう。このような研究の潮流を一言で言えば、ゲノム情報科学は生物学そのものとほとんど一体化したものになるであろうと予測されるということである。

また一方で、脳研究との融合も大幅に図られるであろうし、地球環境の保全の点からも、工業プロセスにおいて生体反応のシミュレーションに基づいたバ

イオプロセスの開発と導入が進むであろうし、生態系のシミュレーション等も現実のものとなろう。このように地球環境の研究とも大きな接点が出てくるものと思われる。

このような流れの中で重要なポイントは、現在のゲノム配列やタンパク質立体構造のような、これを決定すればよいという自明の目標は将来はないであろうということである。高次の生命現象の解明には、ゲノム情報科学と生物学が一体となって実験や計測をデザインしその結果をデータベース化して行くことが不可欠である。またこの際、単に生化学的存在として生命を捉えるのではなく、遺伝子発現の時期やタンパク質の局在部位を、つまり生命としての拘束条件を十分に考慮した形で研究を展開することが必要である。

そして、このような研究のスタイルが実現している状況では、生物学の知識を背景として持つ研究者と情報科学を背景とする研究者とが分かち難く共存し、また、ゲノム情報科学という枠組みでしか分類できない人材が大量に存在しているはずである。

さて、次に、医学的な視点で眺めると、免疫系等の計算機シミュレーションや計算機デザイン、ゲノム情報に基づく診断、治療、創薬、つまり、医療のデザインが一般化するであろう。ここでもゲノム情報科学は不可欠の存在になってくることが予想される。

また、当然のことながら、ゲノム情報科学の応用範囲は医学に留まらず、ゲノム情報を応用した農業、畜産等も活発化するであろうし、バイオセンサーやバイオマシン等工学との接点に位置する研究も花開くであろうと思われる。

最後に、情報科学、計算機工学の視点から眺めると、大規模なデータベース（現在の数百倍から数千倍の大きさ、つまり、ペタバイト級）を処理して、その中から各種の有用な知識を発見する情報技術、また、そのための並列処理アルゴリズムや並列コンピューターの研究は依然として重要な課題として存在するであろう。また一方、生命系のような複雑な系を記述しシミュレーションする技術が大きな研究テーマになっているであろう。また、ゲノムを一種の言語として捉え、その文法を記述し解き明かす研究や複雑な生命現象を視覚化するための研究も重要な課題となっているであろう。

このような研究を展開するには、現在のコンピューターの処理能力の数千倍から数万倍の能力をもった並列コンピューター（少なくとも数千台から1万台程度のプロセッサからなる並列機で数十テラから数百テラFLOPS）が必要であり、そのようなコンピューター及び活用技術の開発研究もゲノム情報科学の中で大きな位置を占めているであろう。

1-3. 基本的な認識（論点の整理）

ここまで述べてきたように、ゲノム情報科学の将来は非常に大きな広がりをもった存在になるものと予想される。これをゲノム情報科学と呼ぶかどうかは別にして、少なくとも道具としても、方法論やものの見方としても、生命科学及びバイオ産業と情報科学との関わりは今後ますます重要になってくることは疑いの余地のないところである。

しかしながら、これだけ大きな広がりを持つものであるから、ゲノム情報科学の将来像の意味するところをできるだけ整理して、それらをいくつかの柱に分け、それぞれについて有効な戦略を立てることが肝要である。

ゲノム情報科学の将来像を整理する上で、以下のような視点や切口が考えられる。

- 生命科学としての側面と情報科学としての側面
- 「情動的解釈のための基礎理論」としての役割と「研究の効率的道具」としての役割
- 基礎研究としての側面と産業や医療への応用のための側面

この他にも、研究対象（ゲノムかネットワークか；配列か立体構造か）、研究目的（生物学か医学か）、情報技術（知識発見かシミュレーションか）等によってより詳細な分類も可能であろう。

上記の視点や切口で形成される多次元空間内のそれぞれの課題について、きちんとした戦略を立てることも大切であろうが、今回検討すべき課題である人材養成、データベース構築、情報技術開発等の観点からは共通する部分も多いためここでは大胆に以下の二つの柱にそれらを集約して今後の戦略を考えることにする。

A. バイオのための情報科学

上記の分類で言えば、情報科学としての側面、「研究の効率的道具」としての役割、産業応用・医療応用のための側面が主にこの中に入る。

B. 情報論的解釈による生命科学

ここには、生命科学としての側面、「情動的解釈のための基礎理論」としての役割、基礎研究としての側面が主に入る。

以下、「バイオのための情報科学」、「情報論的解釈による生命科学」それぞれについて、

イ. 人材養成

ロ. データベース構築

ハ. 情報解析技術開発

の戦略を提案することとする。

2. 人材養成について

前節に示した、ゲノム情報科学の2つの柱「A. バイオのための情報科学」「B. 情報論的解釈による生命科学」は、どちらも重要であることはこれまで述べてきた通りであるが、現在、特にゲノム配列決定やタンパク質立体構造決定の現場あるいは産業界から緊急に人材養成が求められているのは、前者の「A. バイオのための情報科学」のための即戦力である。

そこで以下ではまず、「バイオのための情報科学」のための人材養成を最初に取り上げる。特に、その中でも即戦力養成の方策を先に示すことにする。

ただしこれは、中長期的な視点での人材育成を後回しにしてよいということでは決してなく、その間に並行して中長期的視点に立って人材養成を進めることこそが最も大切であることを強調しておく。

2-1. 「バイオのための情報科学」の人材養成について

2-1-1. 即戦力の養成

即戦力を養成するには、既存のバイオ系研究者・技術者や情報系研究者・技術者の活用を図るかしか手がない。そのためには、それぞれに研修、訓練のプログラムと機会とを提供するとともに、いわゆるインセンティブを与える必要がある。そのための方策としてここでは以下のものを提案する。

(研修プログラムの開発)

- (1) 既存のバイオ系研究者・技術者のための「バイオのための情報科学」習得のための塾（計算機、インターネットの設備必要）や研修セミナーを早急に設ける。そのためのカリキュラムを学会等の協力のもとに整備する。
- (2) またその逆に、情報系研究者・技術者のための生物実験入門コースを設け人材の発掘を図る。

(研修の機会及びインセンティブの提供)

- (3) 上記の塾や研修セミナーへの参加、国内国外の「バイオのための情報科学」の教育研究を実施している機関への留学やサマースクールへの参加のための費用を援助する。

(訓練プログラムの開発)

- (4) ゲノムプロジェクトやデータベース開発プロジェクトを遂行している既存の研究機関を人事面予算面で拡充し、そこに実地訓練の場を設ける。

(訓練の機会及びインセンティブの提供)

(5) 上記(4)の受入れを積極的に進める研究機関に対して、そのための教官枠やポスドク枠を用意したり、企業からの人材受入れを容易にする仕組みを作ったり、そのための雇入れ費用を別途用意する。

(外国人の登用)

(6) 上記の施策と並行して外国人の積極的な登用も図るべきである。現在、外国人の雇用には制度上の障害はほとんどないが、現実的には依然として障害があることも事実である。そこで、外国人技術者向けの語学研修の制度導入や予算面での手当等でより積極的に受け入れるための手だてを講じることが重要である。

(専門学校等の活用と国家資格)

(7) 専門学校等では短期間で技術者を養成できるというメリットがある。そこで、専門学校等に「バイオのための情報科学」習得コースの設置を促進する。また、これに関する国家資格を設けることにより、受講者のインセンティブを向上させる。

なお、上に示した(1)から(7)の施策は、それぞればらばらに進めたのでは効果が薄い。これらは相互に連携させながら、また、以下に示す中核拠点と併設、連動させながら進めることが肝要である。

なお、この連携や連動を具体的にどのように実現するかについては、後ほど述べることとする。

2-1-2. 研究交流、試行錯誤の「場」としての拠点形成

急速な発展途上にある科学及び技術の分野では、学際的というにはあまりにも創造的な空間であるため、最も先端的な部分では問題設定だけがあるという状況が生まれる。ゲノム情報科学がまさしくそうであろう。このような状況においては、それを解くためのコンピューター等の道具が準備されているところに、自由に若者が出入りしながら新たな問題解決を試行錯誤させることができるような「場」が必要ではないかと考えられる。

このような「場」は、これまで大学や大学院が提供することを期待されたものであったが、その実現のための道具立てが極めて大がかりであり（例えばこれに必要なコンピューターは数百台から数千台のプロセッサがつながったもので、価格的にも各研究室や各大学で準備できるものではない）、しかも、多くの異なる分野の人々が共有できることが不可欠であることを考慮すると、日本に

このような「場」に相当する拠点を置くことが重要であると考えられる。

このような「場」では、現在ゲノム情報科学を担っている、情報系、生物系以外からの、例えば、物理系、複雑系等からの人材参入も期待できよう。これにより現在では想像できないような、全く新しい発想をもった人材や研究テーマが生まれる可能性がある。

また、前項の最後に述べたように、この「場」と研修や訓練とを連動させ人材の交流を図ることにより、このような「場」は5年後10年後の研究の進展やそのための創造的人材の誕生に重要なだけでなく、短期の即戦力の養成にも大いに寄与すると期待される。

さて、このような「場」は、物理的な空間、大規模な並列コンピュータをはじめとするいくつかの道具立て、および、これらを管理運営する最小限の事務局があれば、あとはこのような活動に熱意のあるボランティアの参加により実現はある程度可能であろう。

しかしながら、このような場を定常的に維持し、多くの人々がわざわざ時間を割いてまで集まりたいと望むような、活気と魅力にあふれた「場」を実現することは容易ではない。

いろいろな人々が簡単に集まれるような交通至便な場所にあること、そこに行けば、各種のデータベース、CPUパワーなどの豊富な計算資源、高速なコンピュータネットワークが利用できることは「場」を構築する上で大変重要なことではあるが、これだけでは十分ではない。なぜなら、コンピュータネットワークを利用すれば、わざわざそのような「場」に足を運ぶ必要はないからである。

もちろん、ここでいう「場」はそのようなコンピュータネットワークを介した研究交流も含む広い概念のものではあるが、やはり、直接的に会い相互に教え合ったり自由に意見を交換するような「場」を実現することが大切である。

そのためには「場」を物理的に作るだけでは不十分であろう。「場」が自立的に動き出すまで、当面の間、先に述べたように「場」と研修や訓練のプログラムと連動させたり、つまり各種講習会を開催したり、ゲノム情報解析の相談室を設けたり、何人かのコーディネータを決めて定期的な勉強会（生物系向け、情報系向け両方）や研究セミナーを開いたりするようなことを積極的に行うことが必要と思われる。このような活動を中核として、はじめてその周りに自由な研究空間が広がって行くであろう。

つまり、ゲノム情報科学の初心者から最先端の研究者まで、それぞれの興味や知識のレベルに応じた、魅力ある各種プログラムを「場」の中に設け、それへの参加を通して、研究交流、共同研究、試行錯誤が図られ、それが将来の新たな研究分野を切り開くような仕組みを備えた「場」にしなければならないと思われる。繰り返しになるが、日本にこのような「場」およびそれを実現実施するための組織を早急に作らなければならない。

なぜなら、先に述べたように、このような「場」を大学や大学院の枠組みの中で実現することはなかなか困難だからである。現在の制度では大学／大学院は一様な資格審査に基づく、ある程度の年齢層に対する、ある確立された学問の教育実践の場となっていることから、ゲノム情報科学のような急速に発展する研究分野にとっては古い革袋としてしか機能し得ない可能性が高いからである。したがって、受身の教育実践の場ではない、創造的実践の場たりえる「場」を新しい革袋の中に実現することが望まれるのである。

2-1-3. 大学等における中長期的視点に立った人材の育成

これまで、即戦力の養成及び拠点形成の重要性について述べてきたが、「2. 人材養成について」節の冒頭に触れたように、人材の養成に関してはやはり中長期的視点に立った戦略が不可欠である。

中長期的視点に立った人材育成を担う存在として当然のことながら、大学及び大学院がある。これらについて、以下の施策を進めることを提案する。

なお、繰り返しになるが、中長期的視点に立つということは、今すぐに手を打たなくてよいという意味では決してない。逆に今すぐに手を打たないと将来立ち行かなくなってしまう。是非とも以下に例示するような施策の実現が望まれる。

(a) 既存の学科や専攻（情報系及び生命系それぞれ）に「バイオのための情報科学」を指向した教育研究組織を整備する。その規模等については既存の組織の配置転換等も考慮して適切なやり方を工夫すること。

(b) いくつかの大学に「バイオのための情報科学」教育センターの教育研究組織を整備する。その規模等については既存の組織の配置転換等も考慮して適切なやり方を工夫すること。

なお、上に示した教育研究教育組織の具体案については、別添1に参考資料としてゲノム情報ワーキンググループの私案を添付しておく。

さて、このような講座の設置は迅速かつ機動的に行うことが最も重要である。

また、学問や社会状況の変化に柔軟に対応できることが望ましい。そのため、この実現に当っては寄付講座受託講座等も十分に活用すべきである。

また、上記の「バイオのための情報科学」を指向した講座は将来的にはほとんどの大学に必要となるものであるが、当面は指導力のある人材を整備した大学等から順次整備して行く必要がある。

2-2. 「情報論的解釈による生命科学」の人材養成について

先に述べたように、現在必要とされている人材は、前項に示した「バイオのための情報科学」を推進できる人材であろう。しかしながら、1-2. 将来展望で描いたように将来的にはゲノム情報科学が生命科学そのものと分ちがなくなっている状況を考えると、「情報論的解釈による生命科学」を遂行できる人材の養成を並行して進めることは欠かせないと考えられる。

この「情報論的解釈による生命科学」のための人材養成については、以下の施策を進めることが望ましい。

2-2-1. ゲノム情報科学の拠点の整備

「情報論的解釈による生命科学」のための人材養成及び研究の発展のためには、将来の生命科学や脳科学について、様々な分野の研究者が交流し議論できるようなゲノム情報科学の拠点を整備することが重要である。

なぜなら、ゲノム情報科学の大きな特徴の一つは、演繹的な生物学を展開することにあるが、一方、実際にデータを産生する生物諸科学は帰納的な研究スタイルを旨とする。この2つは前節で述べたように、密接に連携しやがて融合に向かうに違いない。そのことを予見するなら今から様々な分野の研究者が日常的に交流し、研究の展開を図ることの出来る拠点を整備する必要がある。

この研究拠点は、当面前項で述べた「場」と表裏一体として運営する。つまり、この研究拠点のメンバーは「場」のコアメンバーとしての役割も担うこととする。

当面、この研究拠点は小規模な組織とし徐々に大きくする。ただし、研究の柔軟性機動性を確保するため、将来的に大きくなっても基本的には一部のコアメンバーを除いて客員研究員や任期付き研究員で構成する。

この研究拠点で取り組むべき具体的な研究テーマとしては、前節「将来展望」で予見したような研究を先取りするような先進的なものが望まれる。

2-2-2. 大学等における人材養成 - 大学院の専攻や学部の学科の整備等 -

大学院及び学部レベルにおいて「情報論的解釈による生命科学」を指向した専攻や学科を整備する。学生は生物系、情報系、医学系、物理系、農学系、薬学系等から幅広く受け入れられるよう配慮する。なお、その規模等については既存の組織の配置転換等も考慮して適切なやり方を工夫すること。

このような教育研究組織は、全く新しいセンスを身につけた学生を養成するには不可欠のものである。

なお、上に示した専攻や学科の具体案については、別添2に参考資料としてゲノム情報ワーキンググループの私案を添付しておく。

3. データベース構築の体制について

データベース構築体制の充実を図るという観点では、「バイオのための情報科学」と「情報論的解釈による生命科学」をあえて区別する必要はない。以下では二つの共通の戦略として述べることにする。

まず、日本でもこれまで様々な視点から数多くのデータベース構築がなされてきたが、いくつかの例外を除いてそれらはいずれも小粒で、外国から頻繁にアクセスされる、あるいは、コピーされるようなデータベースは少ないのが現状である。

その理由はいろいろあるが、大きな要因としては、データベース構築に関する国家戦略の欠如、予算面人事面組織面での支援体制の不備、データベース作りに対する社会的学問的評価の低さ、データベース構築に関するチェック機構のなさ等が挙げられよう。

今後はデータベース構築に関して国家的視野に立って戦略を立案し、それを支援して行く体制を整備する必要がある。その際留意すべきは、戦略にそってデータベース作りがきちんに行われ、しかもそれが世界的視野で見て競争力のあるものかどうか、例えば、外国から多くのアクセスがあるかどうか等をきちんと評価する体制も合わせて整備することが重要である。

また、このような戦略の検討やその実施に関しては、平成13年1月から発足する科学技術政策全般の総合的な司令塔である総合科学技術会議とそこで決定された総合戦略に基づいて、より具体的な事項を審議する科学技術・学術審議会の役割が重要であると考えられる。すなわち、後ほど述べる情報解析技術の開発に関する戦略のほか、先に述べた研修プログラムや講座等の人材養成施策、バイオのための情報科学振興の「場」の設定、情報論的解釈による生命科学のためのゲノム情報科学の研究拠点等に関する具体的な戦略等について、十分な検討ができるような体制を整えることが必要である。

ただし、委員会形式だけでは、戦略立案その他に際して世界的動向を調査したり、具体的な計画作りを行ったり、細部にわたる評価を行ったりすることが困難なため、戦略の検討やその実施を支える組織を設けることとする。

さて、データベースの構築を国家的に支援する際に重要なことは、データベース構築の予算措置等に際して競争原理を働かせることである。データベース構築を希望するいくつかの研究機関からデータベース構築に関する提案を出してもらい、それを審査した上で支援を行うことが大切であろう。もちろん、定

期的にその提案が確かに実行されているかどうかをチェックする仕組みも合わせて持つ必要がある。

データベースには、その構築や保守に多くの人員や大規模な計算機を必要するものから、研究者個人あるいはその少数の仲間で作る小規模なものまで、様々なものがあるが、大規模なものに関して言えば当面別添3に参考資料として例示するような生物学医学の基盤として不可欠の情報に対応した複数のデータベースを国家的に支援すべきであろう。

なお、上記のデータベースの構築に当っては、データの表示方式や検索条件の記述方式等の利用者インタフェースの統一化、標準化等にも十分配慮すべきである。また、これを具体化するためにデータベースのコンテンツ作成だけでなく、このような利用者インタフェースの開発にも予算措置を講じることが重要である。

実際にデータベース構築を実施する組織として、既存の大学や研究所を活用することがふさわしいと考える。その理由は後にも述べるように、データベース構築は最先端の研究と緊密な連携を保つことが必要不可欠であり、それには活発に研究を進めている大学や研究所の中で行うことが最善だからである。

そこで、既存のデータベース構築機関を大学や研究所の付属施設ではなく、できるだけ独立した形に組織替えし、かつ、人事面や予算面で充実させるとともに、それらを連携させ、さらに、その周りに企業を巻き込んだコンソーシアム等も形成し、全体の調和を図って行くことがよいと考えられる。

このようなデータベースの戦略立案や拠点形成に際しては、以下の点に十分に配慮すべきである。

- 国家支援のデータベースは、データの公開が大原則であるということ
- データベースはそこに入れるべきデータそのものが重要である。そのようなデータを具体的にどのように産出して行くかという議論及びそれに基づく予算面の配慮とセットにして考えるべきである
- 従来、データベースというとゲノム配列データやタンパク質立体構造データ等の1次データの集積のイメージが強いが、もっと付加価値のついた統合化されたデータベースの開発についても十分な配慮を払うべきである
- 本報告書の冒頭にも触れたが、これからは、ゲノム情報科学、生物学、計測技術研究が一体となって実験や計測をデザインしその結果をデータベース化して行くことが不可欠である。そこで、計測装置とのインタフェースの標準化等にも十分な配慮が必要である。
- 優れたデータベースは、優れた研究との緊密な連携があってはじめて生ま

れるものである。そうでなければ、世界的に競争力のあるデータベースは作れない。ここでいう優れた研究は「バイオのための情報科学」と「情報論的解釈による生命科学」との両方を意味する。常にこれらの研究との連携を視野に入れた体制作りが望まれる。

- また、人材養成で述べたが教育機関との緊密な連携も人材育成の観点から重要である。

4. 情報解析技術の開発について

この問題は、主に「バイオのための情報科学」に関係した問題である。

ゲノム配列をはじめとした生命情報の情報解析技術の開発に関しては、データベース構築に比べてマンパワー的にも予算的にも小規模で済むため、これまで正面からこの問題について議論されることは少なかった。

しかしながら、データ量の爆発的急増やデータの種類の多様化によって、研究者個人の努力や計算機能力の自然増に任せておく訳には行かなくなってきた。

ゲノムの配列データを始めとするゲノム解析関連のデータは、ここ半年で2倍という爆発的な伸びを示しており、計算機の処理能力の伸びを遥かに凌いでいる。このままでは現在行われているホモロジー検索等のデータベース解析サービスは計算機の処理能力の点で早晚破綻を来たすことが予想される。

これを回避するには、

- (a) 生命科学のための並列処理アルゴリズムや高速コンピューターの研究開発
- (b) (ホモロジー検索に頼らなくて済むような) 画期的な情報解析技術の研究開発

等が不可欠であり、これを国家的に支援することが重要である。

つまり、データベース構築の課題と同様に、情報解析技術(知識発見やシミュレーション等)の開発やその基盤となる並列処理アルゴリズムや並列コンピューターの開発に関する戦略を国家的な視野に立って立案し、それに従って予算面人事面で支援して行く体制を整備する必要がある。

その際留意すべきは、やはりデータベース構築の場合と同様に競争原理を働かせること、その成果をきちんと評価する体制も合わせて整備することが重要である。

具体的な推進方策(研究開発課題等)については、データベース構築と同様にその検討を行う体制を整えるべきであるが、当面少なくとも以下の別添4の参考資料に例示するような研究開発課題は国家的規模で支援すべきである。

なお、データベース構築と同様に、国家的に支援する情報解析技術の研究開発課題の数は研究の発展とともに増やすことが必要である。また、ソフトウェアのインターフェースの標準化や共有化等にも配慮すべきである。

実際に情報解析技術の開発を遂行する組織としては、これもデータベース構築と同様な理由から、最先端の研究を行っている大学や研究所がふさわしい。その受け皿を、これも繰り返しなるが、できるだけ独立した形に組織替えし、

かつ、人事面や予算面で充実させるとともにそれらを連携させ、また、その周りに企業を巻き込んだコンソーシアム等も形成し、全体の調和を図って行くことがよいと考えられる。

ただし、並列コンピューターの開発等は、予算的にも非常に大規模なものを要することであり、また、大規模な並列計算環境を使いこなすこと自体が極めて先端的な技術であるため、「バイオのための情報科学」の中だけで閉じて論じることはできない。「場」との連携はもちろんのこと、情報科学研究そのものとの緊密な連携を図って進めて行くことが肝要である。

5. まとめ

これまで、人材養成、データベース開発、情報技術開発の3つの観点から、我が国におけるゲノム情報科学に係る戦略を述べてきた。これらを簡単にまとめると、以下のようになる。

(バイオのための情報科学のための人材養成)

- (1) 即戦力を養成するために、研修セミナーの開講、ゲノムプロジェクトやデータベース開発プロジェクトにおける訓練プログラムの提供、外国人研究者技術者の積極的登用、専門学校を活用とそのための国家資格の制定等を行う。
- (2) 研究交流、試行錯誤の「場」としての拠点形成を図る。
- (3) 大学等における教育研究組織を整備する。

(情報論的解釈による生命科学のための人材養成)

- (4) ゲノム情報科学の研究拠点を整備する。
- (5) 大学院の専攻や学部の学科の整備を図る。

(データベース構築)

- (6) 科学技術・学術審議会等に、我が国のデータベース構築の在り方等に関する戦略を検討する場を設ける。
- (7) 戦略の検討やその実施を支える組織を設ける。

(情報解析技術開発)

- (8) 我が国の情報解析技術開発の在り方等に関する戦略を検討する体制を整える。

(全体の連携調整)

- (9) 上記の施策の全体の連携調整を図る体制を整える。

おわりに

これまで、人材養成、データベース構築、情報技術開発の3つの観点から、我が国におけるゲノム情報科学の戦略についてあるべき姿を述べてきた。本報告書で提案している施策はどれをとっても当たり前のことであり突飛なものは何もない。速やかに着実に実行されることを望む次第である。なお、提案の中に一部具体策に欠けるところもあると思われるが、これは急速な勢いで発展しつつある研究分野ゆえということかどうかお許しいただきたい。これについては今後の議論に委ねたい。

(別添1) 大学等における「バイオのための情報科学」の人材養成に関する私案

バイオ情報科学を指向した講座や「バイオのための情報科学」教育研究組織の整備

最低限以下に示すような2小講座程度が必要

データベース講座：生物知識表現、文献からの情報抽出、オントロジー、シミュレーション、計測技術開発等の研究教育

情報解析講座：配列解析、知識発見、進化解析、ネットワーク解析、立体構造解析・予測等の研究教育

(別添2) 大学等における「情報論的解釈による生命科学」の人材養成に関する私案

当面(5年ほど)、5小講座程度(修士30名、博士15名程度)、将来的には10小講座程度(修士50名、博士20名程度)に増やす。なお、講座の構成はおおよそ以下のようなものが考えられる。

- 高次生物知識の計算機上での再構築
- ゲノム構築原理の解明
- 比較ゲノム解析
- タンパク質立体構造の予測と解析
- 生命系シミュレーション

教育研究の進展に応じて、将来的には、規模の拡大等も視野に入れるものとするが、その場合は、以下の分野を加えることが考えられる。

- 脳研究
- 発生分化研究
- 疾患や免疫のメカニズム等の研究
- バイオコンピューター、バイオプロセス、バイオセンサーの開発
- 人工生命の設計

(別添3) 国家的に支援すべきデータベースとその規模に関する私案

以下に例示するような生物学医学の基盤として不可欠の情報に対応した複数のデータベースを国家的に支援すべきであろう。

- 配列情報

- 多型情報
- 発現情報
- 遺伝子ネットワーク情報
- 分子間相互作用情報
- タンパク質立体構造情報
- 遺伝子、タンパク質及びその分類等の用語辞書

なお、支援すべき金額は一概には言えないが、おおよそ1つのデータベース当り、コンピューターの費用以外に1億円から3億円程度必要と思われる。なお、国家的に支援するデータベースの数は研究の発展とともに増やすことが必要である。

(別添4) 国家的に支援すべき情報解析技術とその規模に関する私案

以下に例示するような情報解析技術の開発は国家的規模で支援すべきであろう。

- 数百生物種規模のゲノム配列比較技術
- 大規模なデータベース（テラバイト級）からの知識発見技術
- 大規模な遺伝子タンパク質系（数百以上の要素からなる）系のシミュレーション技術
- また、上記の可視化技術
- 大規模（現在の規模の100倍から1000倍程度）のタンパク質の分子動力学計算

なお、上記に示したような情報解析技術の研究開発には、ソフトウェアの開発要員の確保が不可欠であり、最低でも1課題当り、年間数千万から1億円程度の支援が必要であると思われる。もちろんこれとは別に、先の並列コンピューター（共同利用可）を準備する必要がある。

科学技術会議ライフサイエンス部会ゲノム科学委員会
ゲノム情報科学ワーキンググループ委員

- (主 査) 高木 利久 東京大学医科学研究所教授
- (委 員) 秋山 泰 電子技術総合研究所生命情報科学ラボ主任研究官
- 安西祐一郎 慶応義塾大学理工学部長
- 勝木 元也 東京大学医科学研究所教授
- 金久 實 京都大学化学研究所教授
- 五條堀 孝 国立遺伝学研究所教授
- 小長谷明彦 北陸先端科学技術大学院大学知識科学研究科教授
- 小原 雄治 国立遺伝学研究所教授
- 坂田 克己 農業生物資源研究所ゲノム構造研究室主任研究官
- 根本 靖久 山之内製薬(株)創薬研究本部研究推進部主任研究員
- 山口 直人 国立がんセンター研究所がん情報研究部長
- 横山 茂之 東京大学理学部教授