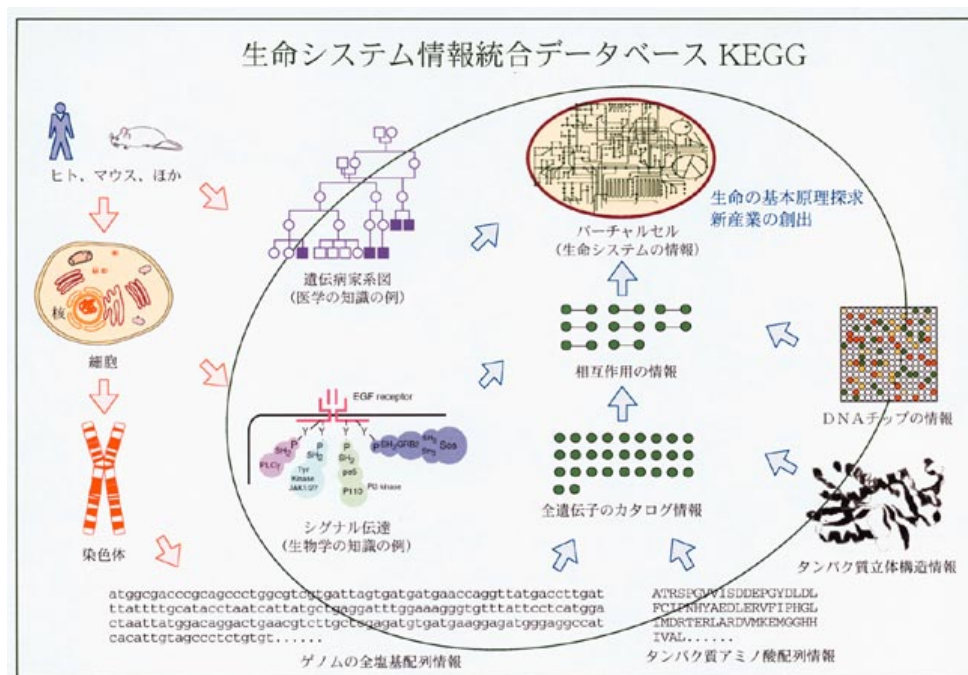


Biological Systems Database
and Genome Information Science
生命システム情報統合データベースの
構築とゲノム情報理学の創成

プロジェクトリーダー 金 久 實

京都大学 化学研究所 教授



1. 研究の目的

ヒトゲノムの大雑把な配列 (ドラフトシーケンス) 決定は 2000 年 6 月に終了し、完全な配列も 2003 年には読み取り作業が完了する。21 世紀のバイオサイエンスは、ヒトをはじめとした数多くの生物種の膨大なゲノム情報をもとにした基礎研究、ならびにそこから有用性を見いだす応用研究が主流となるだろう。しかしながら、配列だけではゲノムに書かれた生命のはたらきの情報を読み取ることはできず、本当のゲノム解読競争はまだ始まったばかりだといっても過言ではない。とくに、現在用いられている個々の分子レベルではたらきを解読する方法の限界を乗り越え、細胞レベル、個体レベルといった高次の生命システムのはたらきを解読する方法が確立されなければならない。

本研究は、これまでに蓄積された高次の生命のはたらきに関する膨大な知識を、分子間相互作用ネットワークシステムの情報としてコンピュータ化し、ゲノム解析がもたらす遺伝子カタログの情報、DNA チップやプロテオーム解析による遺伝子間または分子間の関係情報、さらに生体内化合物と化学反応に関する情報を統合した「生命システム情報統合データベース」を構築することを目的とする。そして、生命科学、情報科学、物理学、化学の分野を包含する「ゲノム情報理学」の研究領域を創成し、ゲノムから生命のシステムを、さらには自然界のダイナミクスを理解することを目的とする。一方、生命システム情報統合データベースは、ゲノムの配列から生命システムのはたらきに関する情報を解読するためのレファレンスデータベースでもあるので、ゲノム情報を医学・薬学・農学・工学等の分野で有効利用するための技術開発も行う。

2. 研究の内容

20世紀後半の生命科学は分子生物学全盛の時代であった。その根底には「生命は分子や遺伝子のはたらきとして理解できる」とする還元論の考え方があつた。還元論の究極の形として、遺伝子の全セットすなわちゲノムを明らかにするため、1980年代の終わりにヒトゲノム計画が始まった。その結果、ヒトをはじめ数多くの生物種におけるゲノムの情報が大量に蓄積されつつある。21世紀の生命科学はこのゲノム情報を基盤とし、還元論と反対のアプローチ、すなわち個々の部品(遺伝子)の集まりからシステム全体(細胞あるいは生物個体)を再構築し、「生命のはたらきをシステムのはたらきとして理解する」合成論のアプローチが主流になると考えられる。還元論が主に実験科学として行われてきたのに対し、合成論は基本的に情報科学の問題である。つまりゲノムの情報から生命の情報システムを再構築することは、コンピュータの中で行われる。

これを実現するため、本研究では多様な情報を統合して解析する高度な情報処理技術の開発を行う。その最も基盤となるのが、これまで文部省ヒト・ゲノムプログラムにより構築してきた生命システム情報統合データベース KEGG (Kyoto Encyclopedia of Genes and Genomes) である。KEGG は4つのコンポーネントから構成され、それぞれ以下のように構築している。

(1) 生命システムのはたらきに関する情報

代謝、シグナル伝達、細胞周期、発生をはじめ、細胞レベル、個体レベルでの生命システムのはたらきに関する知識を、分子間相互作用のネットワークとして表現し、PATHWAYデータベースとして体系化する。膨大な文献から知識を集約したり、ゲノムの比較解析を行ったり、最も高度な知的作業である。

(2) 生命システムの部品に関する情報

ゲノム解析がもたらす遺伝子カタログは生命システムの部品のカタログであり、生物種ごとに遺伝子カタログ情報をGENESデータベースとして蓄積する。各遺伝子の機能アノテーションは独自に追加・更新と標準化を行い、常に最新の情報を提供する。

(3) 生命システムの配線に関する情報

部品と部品のつながり(配線)を二項関係と呼び、遺伝子または遺伝子産物に関するさまざまなタイプの二項関係をBRITEデータベースとして蓄積する。文献から手作業で抽出するものと、DNAチップやプロテオーム等のシステムティック機能解析によるデータがある。

(4) 生命システムの環境に関する情報

生命システムとその環境である自然界とは物質としての連続性がある。生体内・生体外の有機化合物の知識と、生体内の有機反応の知識をLIGANDデータベースに登録する。

本研究ではKEGGを飛躍的に発展させ、ゲノムと生命システムの統合解析によりゲノム情報理学の研究領域創成を行うものである。

3. 研究の体制

機関：2000年6月～2005年3月

構成：プロジェクト・リーダー

コアメンバー1名

研究協力者5名

実施場所：京都大学化学研究所

