ツールが含まれる大規模なシステムとなっており、ここでその全体像を紹介することはできない。WWW (<http://www.genome.ad.jp/kegg>) 経由で一般に公開しているのを、ぜひ一度使ってみていただきたい。

さて次に、こうして蓄積したパスウェイ情報をどのように解析に利用するかである。KEGG/PATHWAY データベースでは、パスウェイの情報をおもに可視化を目的として描かれた図で蓄積しているが、こういったパスウェイは分子と分子の二項関係の組み合わせと考えることができる。二項関係によるパスウェイの表現は、通常グラフを記載するために使われるリスト表現と本質的に等価なので、パスウェイをグラフととらえることができる。そうすることで種々のグラフアルゴリズムの応用によって、さまざまな解析が行えるようになる。KEGG/PATHWAY のパスウェイマップを二項関係として表現しなおしたデータは、別途分子間相互作用のデータベースのBRITeに格納されている。BRITe データベースには、これ以外にも、文献から抽出したシグナル伝達関連の分子間相互作用や、酵母の2- ハイブリッド実験によるタンパク質間相互作用、マイクロアレイによる発現類似の関係など、あらゆる分子間相互作用を、二項関係として格納していく予定である。

パスウェイをグラフとして表現する話をしてきたが、同様にゲノムも一次元のグラフとしてとらえることができる。そうすることにより、ゲノムとパスウェイを直接比較することができるようになる。例えば、ゲノム上近傍にコードされる酵素遺伝子の一群が、代謝経路近傍上で機能することはあるのだろうか。機能的に関連した遺伝子群がゲノム上で近傍に集まっているものを、我々は遺伝子クラスターと呼んで注目している。グラフ比較を利用した遺伝子クラスター検出法により、大腸菌においては多くのオペロンが代謝経路上近傍で機能していることが明らかになった。同様の戦略で、配列類似の遺伝子群とパスウェイの関係や、発現パターンが類似した遺伝子クラスターが、ゲノム上でもクラスターを形成しているか、などさまざまな解析の可能性が考えられる。このように、ゲノム、パスウェイ、配列類似性、発現類似性など、生物学的には異なるさまざまなレベルで形成されるクラスターを、グラフとして抽象化し、それらの相関を解析することで、ゲノムから未知の情報が引き出せるものと考えている。

文 献

- 1) 川島秀一, 金久實 (1999) bit **31** [9], 63-70.

生物物理 40 (5), 316-318 (2000)

特集

—分子ネットワーク/シミュレーション—

遺伝子発現制御ネットワーク Gene Network

九州大学大学院農学研究院 久原 哲

Satoru KUHARA

1. 要約

ゲノムスケールで全遺伝子の発現パターンを同時に知するというトランスクリプトーム解析が可能となり、これらのデータを解析し、遺伝子発現の制御ネットワーク (分子ネットワーク) を一挙に明らかにしていくプロジェクトが開始されている。このプロジェクトは従来の研究戦略とは異なる、システム論的研究戦略を可能にしている。

2. 新展開

20世紀の最も偉大な分子生物学プロジェクトの一つと考えられるゲノムプロジェクトの成果として、多くのモデル生物のゲノム構造がすでに明らかにされ、ヒトゲノムの構造までもが決定されようとしている。こ

のゲノム構造決定の急速な進展により、従来の生化学的、遺伝学的手法ではなしえなかった生物機能の解明が可能となってきている。特に、従来個々の酵素反応あるいは遺伝子の機能を研究し、その結果を積み上げる還元論的研究方法が主流を占めていた研究戦略を根本から覆し、システム論的研究法をその研究の主流に据える可能性が出てきた。その最右翼にあるのが全遺伝子の発現を調べるトランスクリプトームであり、全タンパク質の発現を調べるプロテオームである。これらに加えて代謝系全体を調べるメタボロームへも研究が進められて、これらのデータを解析し、生物を構成する全遺伝子の発現ネットワークを調べ上げて細胞機能の全体像を明らかにすることも現実のものとなってきている。これらの研究では、対象となる遺伝子の数が膨大となるだけでなく、各遺伝子の性質も単純ではない。例

図1 分子ネットワークの推定概念図

例えばコーディング領域と制御領域、種特異的あるいは亜種におけるポリモルフィックな変異等のDNA配列レベルから、時空間における遺伝子発現の違い、あるいは翻訳産物としてのタンパク質の局在化および相互作用までの多くの次元が存在する。これらのすべてを観測し、最終的には観測したデータを統合して生物のシステム解明に結び付けなければならない。単独の遺伝子、タンパク質を解析していく従来の手法では、このような多次元情報を収集、解析することは不可能に近い。

その第一歩がトランスクリプトーム解析を可能にした、多種類のDNAあるいはRNAの変化量（発現プロファイル）を同時に測定できるマイクロアレイ¹⁾あるいはGeneChip²⁾法の出現である。すでに*Nature Genetics*が1999年の1月に遺伝子チップの特集³⁾を組んでいる。この手法は現在、生物学・医学の分野ですでに広く用いられ始めており、世界的にはベンチャー企業も数多く設立され、チップあるいはアレイのトータルシステムの販売も開始されている。

これらの方法から得られる実験結果は従来の方法と比較して桁違いの量である。このため新規のモデルおよび解析手法が必要となってくる。加えて大規模データから一挙にモデルの構築を行うことを考えなければならない。例えば、さまざまな条件下で生物の全遺伝子の発現パターンを解析することにより、遺伝子発現の制御ネットワークを構築する必要がある。網羅的な手法で検索することにより、疾病の原因遺伝子を探索し、疾病モデルを構築することなども望まれる。特に、図1に示すようなシステム論的研究が必要であり、入力と出力と外乱のデータからシステムの内部構造を推定するという研究が必要な状況になってきている。つまり、外乱のある遺伝子の破壊（その遺伝子の発現がなくなる）と見たときに、入力を正常の全遺伝子の発現状態

とし、出力を遺伝子が破壊されたときの全遺伝子の発現状態とみなせば、遺伝子の発現のシステムが推定可能であるかというシステム論の問題となる。

では、そのようなことが実際に可能であろうか、高等生物では前に述べているように複雑さの次元が多くなり、容易ではないが、出芽酵母などの微生物であれば可能だと思われる。また、数多くの実験結果が必要となるが⁴⁾、多次元のデータが容易に測定できる環境ができ、ゲノムレベルでの研究が可能になってきた現在、多くの研究プロジェクトが発現プロファイルの収集を開始し、すでに公開している⁵⁾⁻¹⁰⁾。

3. 今後の課題

最も大きな問題として、データに含まれる誤差が挙げられる。現在の実験技術では30～40%程度の誤差が存在している。このような誤差を前提とした情報処理技術の開発を行わなければならない。現状ではこのように大きな誤差を取り扱う方法は十分ではない。逆に実験的にも誤差を少なくする方法論の開発あるいは誤差をある程度見積もれる方法論の導入が期待される。

システム論的観点からは、一般的には、プロジェクトから出てくる遺伝子発現のパターンから、クラスタリングを用いて遺伝子発現のパターンが異なる遺伝子群を明らかにする方法がとられている¹¹⁾⁻¹³⁾。また、遺伝子発現制御のネットワークを推定する研究もすでに世界的に開始されている¹⁴⁾⁻¹⁸⁾。基本は全遺伝子の発現データから遺伝子発現の制御ネットワークの構成図を作り上げるための計算手法を開発することであり、今後この問題の解決のための有効なアルゴリズムの開発が望まれる。

また、膨大な数の遺伝子を取り扱うためには、結果の表示方法の開発が重要である。出芽酵母にしても6000の遺伝子のネットワークを表示しなければならないわけであり、単純な表示法ではほとんど対処できない。表示機能の開発は第一に望まれるところである。

最終的に、これらのシステムを総合して、細胞あるいは個体全体を計算機上でシミュレートできる日が、すぐそこにきている。

参考文献

- 1) Schena, M. *et al.* (1995) *Science* **270**, 467-470.
- 2) Lockhart, D. J. *et al.* (1996) *Nature Biotechnol.* **14**, 1675-1680.
- 3) The chipping forecast (1999) *Nature Genet.* **21** (1).
- 4) Akutsu, S. *et al.* (1999) Pacific Symposium on Bio-computing, 17-28.
- 5) <http://cmgm.Stanford.EDU/pbrown/explore/>

- 6) <http://rsb.info.nih.gov/mol-physiol/PNAS/GEMtable.html>
- 7) <http://genomics.stanford.edu/yeast/celcycle.html>
- 8) <http://cmgm.Stanford.EDU/~kimlab/>
- 9) <http://www.mcb.arizona.edu/wardlab/microarray.html>
- 10) <http://www.grt.kyushu-u.ac.jp/mgt/>
- 11) Eisen, M.B. *et al.* (1998) *Proc. Natl. Acad. Sci. USA*, **95**, 14863-14868.
- 12) Toronen, P. *et al.* (1999) *FEBS Lett.* **451**, 142-146.
- 13) Arkin, A. *et al.* (1997) *Science* **277**, 1275-1279.
- 14) Akutsu, S. *et al.* (1998) In Ninth Annual ACM-SIAM Symposium on Discrete Algorithms, ACM-SIAM.
- 15) Chen, T. *et al.* (1999) In Proc. 3'rd Annual Int. Conference on Computational Biology (RECOMB).
- 16) Somogyi, R. *et al.* (1996) In The Second World Congress of Nonlinear Analysis (WCNA).
- 17) Weaver, D. *et al.* (1999) Pacific Symposium on Bio-computing, 112-123.
- 18) Liang, S. *et al.* (1998) Pacific Symposium on Bio-computing, 18-29.

生物物理 40 (5), 318-319 (2000)

特集

—進化/進化工学—

情報生物学の発展へ向けて

Toward the Development of Information Biology

国立遺伝学研究所生命情報研究センター 五條堀 孝, 新村 芳人

Takashi GOJOBORI and Yoshihito NIIMURA

ゲノムに関する情報や解析が生命科学の基盤を構築していることは言うまでもない。したがって、ゲノム情報科学は、生命情報科学あるいはバイオインフォマティクスと呼ばれる新しい分野の主要な部分を占めている。

ゲノム情報科学あるいは生命情報科学は、2つの側面をもっている。1つの側面は、情報科学や情報技術 (IT) を方法論として生命科学に導入することである。いわば「バイオ-IT」とか「B-IT」などと呼ぶにふさわしい技術的側面である。もう1つは、時空的な情報の流れやシステムとして生命を捉えて、生命現象を理解しようとする情報生物学とでも言うべき側面である。ここで時空的というのは、ナノ秒の生体現象から数十億年の生物進化にわたる幅広い時間スケールや、細胞内分子の生体反応の場から生物集団や生態系に至る広範な空間スケールを含むという意味である。

情報生物学が科学である以上、そこには背景となるフィロソフィーが必要である。筆者らは、先述の「時空的」という意味において、「生命進化」という概念こそが、このフィロソフィーに値すると考えている。つまり、生命進化の原理や過程をさまざまな生物階層のレベルで解き明かすことができれば、生命現象の完全理解に至ったと考えることができる。情報生物学が「時空的」にすべての生命現象を包括することから、「これからの生命科学」という意味で「ネオ生物学」と呼ぶ人々もいるが、要は「生命の多様性と進化」という視点が根底になければならないということである。ゲノム情報科学あるいは生命情報科学がもつこれら2つの

側面は、両者が存在して初めてこの分野の発展が見込まれるものであり、ともに非常に重要である。

すでに30を超える各種の完全ゲノム配列データが発表され、ヒトを含む他の多くの生物種がゲノムプロジェクトの対象となっている。これらのプロジェクトの進展は、ゲノム情報を異種間で比較する「比較ゲノム学」を誕生させた。今後は、遺伝子やゲノムの構造・機能の予測を超えて、それらと深く結びついている生命進化を強く意識した「ゲノム進化学」へと発展していくだろう。また、マイクロアレイやDNAチップなどの先端技術の発展により、遺伝子発現やそのネットワーク、さらには発現現象の解明が進んでくる。それと同時に、それらの種間比較も可能となり、ここでも生命進化の研究へとつながっていくように思われる。

情報科学的手法によって予測された遺伝子ネットワークや発現制御は、タンパク質の立体構造のみならず細胞内外の構造的な研究を全般的に行う「構造生物学」のさらなる発展により、実験的に検証されていくだろう。その際、必要な要素としての遺伝子産物を取り出し、それらの反応の場としての最小状況を生体外で再現して、予測を実験的に検証する「生命工学」の研究が必要になってくると思われる。さらに、転写因子やシスエレメント、あるいは遺伝子ネットワークに関与する遺伝子群を異なるモデル生物間で交換して、その機能を調べる「進化工学」がきわめて重要になるものと考えられる。

具体的な1つの例として、完全ゲノム配列から染色体構造を予測し、その機能を解明できるかという問題が