

# 大規模遺伝子ネットワークの相互作用推定

●岡本 正宏<sup>1)</sup> ◆小野 功<sup>2)</sup>

1) 九州大学大学院農学研究院 2) 徳島大学工学部

## ＜研究の目的と進め方＞

大規模な遺伝子発現プロファイル解析により、遺伝子間相互作用のネットワークが解明されると期待される。現在、ブーリアンネットワークモデル（離散モデル）でさかんに行なわれているが、各遺伝子の発現を 0, 1 に単純化することから、より詳細な相互作用推定には不向きであるとの批判がある。それに対して、連続値を取り扱う解析では微分方程式でモデル化（連続モデル）を行なう方法があるが、現在のところ遺伝子間の詳細な相互作用に関する知見が不足していることから推定が困難である。一般に、観測されるシステム要素の動的挙動（タイムコース）からシステム要素間の相互作用を推定することは、一種の逆問題（inverse problem）である。相互作用ネットワークを連立微分方程式でモデル化する方法が一般に用いられるが、現段階では、遺伝子間の詳細な相互作用に関する知見が十分でなく、遺伝子ネットワークを構成する物質の生成過程や分解過程がそれぞれいくつのパス（経路）から成るのか特定できないため、質量作用則による表記は不相当である。我々は、これまで逆問題解決のための革新的な突破口として、微分方程式の立式に、べき乗則（power-law formalism）に基づいた S-system モデルを、観測データを再現する多数の内部パラメータの自動推定法に進化アルゴリズムを適用する方法を提案してきた。S-system モデルは次のようなものである。n 個のシステム構成要素（状態変数） $X_i$  ( $i=1,2,\dots,n$ ) の値（濃度、発現量に相当）が時間的に変動し、 $X_i$  同士が相互作用しているネットワークシステムを考える。

$$\frac{dX_i}{dt} = \alpha_i \prod_{j=1}^n X_j^{g_{ij}} - \beta_i \prod_{j=1}^n X_j^{h_{ij}} \quad (i=1, 2, \dots, n) \quad (1)$$

式（1）において、 $g_{ij}$  は、状態変数  $X_i$  の生成過程に関与する状態変数  $X_j$  の相互作用係数であり、同様に  $h_{ij}$  は、 $X_i$  の分解過程（消費過程）に関与する  $X_j$  の相互作用係数である。たとえば、 $g_{ij}$  が正の値なら、 $X_i$  の生成過程に対し  $X_j$  は + の作用を及ぼし、同様に  $h_{ij}$  の値が負なら、 $X_i$  の分解過程に対し  $X_j$  は - の作用を及ぼすことになる。 $\alpha_i$ ,  $\beta_i$  は、それぞれ  $X_i$  の生成項、分解項に乗じる係数である。式（1）は、状態変数  $X_i$  の生成過程（右辺第 1 項）と分解過程（右辺第 2 項）にシステムを構成しているすべての状態変数  $X_j$  ( $j=1,2,\dots,n$ ) が関与していると仮定する全結線モデルである。もし、 $X_i$  の生成過程（あるいは分解過程）に  $X_j$  が関与していない（相互作用がない）場合、 $g_{ij}$ （あるいは  $h_{ij}$ ）の値はゼロということになる。しかし、生成過程、分解過程がそれぞれ 1 つの項で表現されているため、生成項、分解項が複数の経路で構成されている場合は、一般質量作用則（generalized mass action law (GMA)）を近似した表現になる。現在のところ、それぞれの遺伝子の mRNA の生成過程、分解過程の詳細な機構は明らかになっておらず、式（1）の近似表現法は有効なものと思われる。つまり、 $g_{ij}$ ,  $h_{ij}$  の値を推定することで、相互作用ネットワークが推定できる。野生型株およ

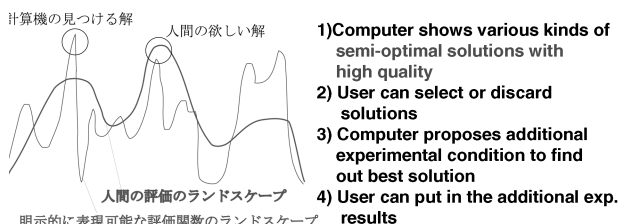
び遺伝子破壊株を想定した複数の発現タイムコースデータを推定システムに与えて、それらの発現タイムコースを再現するような  $\alpha_i$ ,  $\beta_i$  および  $g_{ij}$ ,  $h_{ij}$  の値を推定しなければならないが、推定すべきパラメータの総数は、 $2n(n+1)$  である。観測データセット数は数多く用意することは困難なので、実験タイムコースを再現しうる  $\alpha_i$ ,  $\beta_i$  および  $g_{ij}$ ,  $h_{ij}$  のパラメータ値の組み合わせは、一意に決定できない。組み合わせ数は数多く存在する可能性が高いことから、開発する推定システムは、正解を含めてなるべく多くのパラメータセットを探索できるようにし、ユーザがその解候補から絞り込むことができるような対話型システムにする必要がある（図 1 参照）。また、その解候補から、相互作用に関する有益な知識も抽出できるようにしなければならない。

## Explore the solutions with high quality

Man's evaluation ..... explicit evaluation represented by quantity  
+ implicit evaluation (apriori knowledge, sense etc.)

Computer's evaluation ..... only by explicit evaluation  
based on the value of objective function

Computer's evaluation does not coincide with man's evaluation



## Interactive inference system of genetic networks

図 1 対話型遺伝子相互作用推定システム概念図

本研究では、数 10 の構成要素から成る遺伝子ネットワークを S-system 表記に基づいてモデル化し、次に、観測可能な物質の量のタイムコースデータを再現しうるように S-system パラメータの最適化を行い、遺伝子間相互作用を推定するシステムを構築する。さらに、並列クラスターシステムを用いた進化アルゴリズムを基にした高速かつ高精度な多変数非線形数値最適手法の開発を行う。

## ＜研究開始時の研究計画＞

1) 高速かつ高精度な多変数非線形数値最適手法の開発

これまでに遺伝的アルゴリズムを基本とする手法を開発しているが、ブーリアンネットワーク解析後、いくつかグループ化（クラスタリング）された遺伝子ネットワークシステムをとりあげても、物質種が多い場合、例えば 10 を超えることも想定しなければならない。この場合、S-system のパラメータは 210 個で、そのうち多くのパラメータの値が 0 となるが、そのパラメータを特定し、その他のパラメータの値を高速かつ高精度に最適化するアルゴリズムを新たに開発しなければならない。基本方針としては、大域探索、局所探索を同時に行わせる（並

列化手法)、また最適化するパラメータのうち、0になりうるものをいかに早い段階で見つけるか(カットオフ手法)を設計する。我々は、これまでに多変数非線形数値最適化手法として、遺伝的アルゴリズムを採用し、その中でも、単純遺伝的アルゴリズム(Simple GA)を用いていたが、解の初期収束性のため最適化の効率化が悪かった。Simple GAに替わる新たなGA法を用いて最適化効率の上昇をねらう。

## 2) 小規模な遺伝子ネットワークシステムでの詳細な相互作用推定

観測タイムコースを再現しうるS-systemのパラメータ推定を上記の新たな手法を用いて行い、推定した相互作用係数から遺伝子間の相互作用を明らかにする。

## 3) 離散モデルと連続モデルの統合法の確立

まずブーリアンネットワークの解析によって大分類された数10の構成要素からなる遺伝子ネットワークをS-system表記に基づいてモデル化し、次に、観測可能な物質の量のタイムコースデータを再現しうるようにパラメータの最適化を行い、最終的に遺伝子間相互作用を推定するという、離散モデルと連続モデルの統合システムを構築する。

## <研究期間の成果>

### 1) 高速かつ高精度な多変数非線形数値最適化手法の開発

遺伝子ネットワークの場合、全ての要素間(遺伝子間)に直接的な相互作用があるとは考えにくい。最適解のモデルでは、S-systemの $g_{ij}$ や $h_{ij}$ の行列の多くの要素(パラメータ)の値がゼロであると予想される(分子機構に関する知見から明らかにゼロである要素については、はじめからゼロに固定し最適化の対象からはずす)。GAの多くの手法は親個体を選択圧をかける(優秀な親個体ほど次世代に残す確率を高くする)いわゆる、単純GAであるが、我々は実数値GAを用いて、新しい世代交代モデルとしてMGG(Minimal Generation Gap)を採用し、交叉法としてUNDX(Unimodal Normal Distribution Crossover)を採用したGA手法を設計・開発した[1, 4]。MGGのアルゴリズムをまとめると、図2、3のようになる。図3の番号は図2のアルゴリズム番号に相当する。

begin t = 0

1) randomly create initial population  $P(t)$

2) evaluate each individual in  $P(t)$

while (not satisfy the terminal condition) do t = t + 1

3) randomly select two parents  $C(t)$  from the  $P(t-1)$

4)  $P(t-1) = P(t-1) \cdot C(t)$

5) create children  $C'(t)$  from the  $C(t)$

6) evaluate each individual in  $C'(t)$

7) select one elite individual and randomly select one individual among the  $C(t)$  and  $C'(t)$

8) add one elite one individual and one randomly selected individual by the above procedure to the  $P(t-1)$  and set  $P(t)$

end

end.

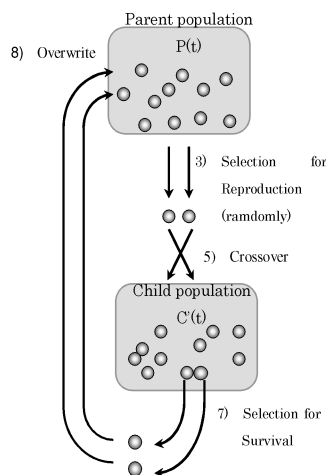


図3 MGGのアルゴリズム

評価値(適応度)には無関係にランダムに選んだ親1と親2のコードを交叉させ、子個体を生成する(図2、3のステップ5)方法には、図4に示すUNDX法を用いた。図4において、親3は親集団からランダムに選んだ親個体であり、探索初期においては、親1、2、3の解はそれぞれかけ離れているので、かなりバラエティーにとんだ解をもった子個体が、探索後期ではかなり収束された子個体が生成されることになる。UNDX法は、親個体の解構造を継承しつつ、子個体の解空間が自動調整される効率的な交叉方法である。

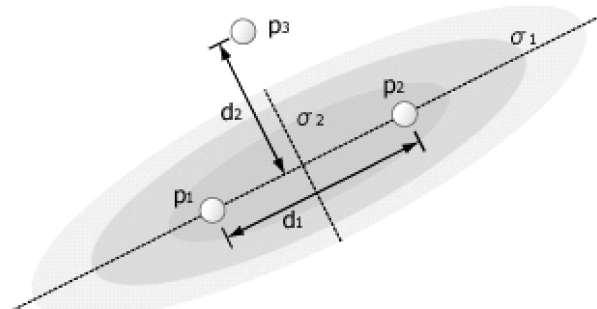


図4 交叉法UNDXを用いた子個体の解空間; 親個体1および2( $p_1$ ,  $p_2$ )とランダムに選んだ親個体3( $p_3$ )で作られる解空間内(図は2個のパラメータ探索の場合)に、 $p_1$ と $p_2$ を交叉させた子個体が生成される。

UNDXは実数ベクトル上の交叉演算子として次の設計指針を考慮している。

1. 統計量の遺伝
2. 多様な解の生成
3. ロバスト性の確保

設計指針1は交叉により生成される子の分布が親の分布の平均値ベクトルや分散・共分散行列を継承するためのものである。特に共分散の継承は変数間の依存関係の強い非分離型の関数の最適化において重要である。S-systemモデルでは、変数間の依存関係が非線形な挙動を記述する上で十分に考慮しなければならないことから、設計指針1は極めて有効である。設計指針2は設計指針1の拘束条件を持ち、初期収束や進化停滞を防ぐことを目的としている。設計指針1と2は最適解を含む領域に母集団が分布しており、選択によりその領域へ個体群の集中化が行われることを仮定しているが、実際には選択が不適切に行われることも予想されることから設計指針3が必要となる。設計指針3では探索をよりロバストにするため、

子の分布は設計指針1を満たす分布よりも若干広いものであることが望ましい。UNDXでは図4に示すように、両親を結ぶ直線上およびその近傍に、両親と第3の親によって決まる正規分布にしたがって子cを生成する。

$$c = m + z_1 e_1 + \sum_{k=2}^l z_k e_k$$

$$z_1 \sim N(0, \sigma_1^2), z_k \sim N(0, \sigma_2^2)$$

$$(k = 2, 3, \dots, l) \quad (2)$$

ここでlは探索空間の次元 ( $2n(n+1)$ )、mは親p1, p2の midpoint、e1はp2p1 (主探索直線) の単位ベクトル、ekは主探索直線に直交する部分空間の正規直交基底ベクトル、 $N(0, \cdot)$  は平均0、分散 $\cdot$ の正規分布を表す。また $\sigma_1$ 、 $\sigma_2$ はそれぞれ主探索方向と、第3の親から主探索直線に直交する直線 (副探索直線) 方向の分布を決めるパラメータで、経験的な推奨値は、

$$\sigma_1 = 0.5d_1$$

$$\sigma_2 = 0.35d_2 / \sqrt{l} \quad (3)$$

である。ここでd1は主探索直線の長さ、d2は副探索直線の長さである。このため、UNDXによって生成された子は座標系のスケールに依存する。

最適化対象となる $2n(n+1)$ 個のS-systemパラメータのうち、相互作用係数gij, hijは遺伝子ネットワークの性質からスパース性が高いと考えられる。つまり、状態変数Xiの合成過程、分解過程にすべてのXjが関与しているわけではなく、大部分のgij, hijは0の値を持つはずである。したがって、生成された子個体の相互作用係数がある閾値を下回る場合、相互作用なしとみなして値を0にする手法を構造骨格化と呼んでいる。開発した手法には、この構造骨格化を導入している。閾値の設定基準は今のところ経験に頼らざるを得ないが、この手法は非常に重要である。実数値GAを使う限り、相互作用係数がびったり0になる可能性は限りなく低い。感度の高い係数は非常に小さな値でも、状態変数の挙動に大きな影響を及ぼすので、0になるべき係数が値を持ち続けることは最適化の成功率自体を著しく低下させてしまう。システム同定における相互作用は、その大きさに関わらず「存在する」と「存在しない」の違いはキーポイントである。

研究の目的でも述べたように、観測データセット数は数多く用意することは困難なので、実験タイムコースを再現しうる $\alpha_i, \beta_i$ およびgij, hijのパラメータ値の組み合わせは、一意に決定できない。組み合わせ数は数多く存在する可能性が高いことから、開発する推定システムは、正解を含めてなるべく多くのパラメータセットを探索できるようにしなければならない。しかし、遺伝的アルゴリズムの性質として、探索終盤においては解候補が急速に絞られてしまう。S-Systemのシステムパラメータの探索手法は、図1のような実験生物学者支援の観点から、次のような要件を満たすべきであると考えた。1) 満足構造 (実験タイムコースを再現できるネットワーク構造) 集合の効率的探索: 互いに異なる複数の満足構造からなる満足構造集合を重複することなく効率よく探索できる。2) 生物学的知見を満たす構造の探索: ネットワーク構造に関する生物学的知見を探索過程に反映することができる。3) 真の構造の探索: 実際の生体内における遺伝子ネットワーク構造である真の構造を含んだ満足構造集合を探索できる。従来の遺伝的アルゴリズムの問題点は、構造探索において用いられている構造突然変異が、形質

遺伝をまったく考慮していない破壊的なオペレータである点である。タイムコースの形状の微調整が必要になる探索終盤において、構造突然変異はタイムコースの形状を大きく変えてしまう可能性が高いため、探索効率が悪いと考えられる。そこで、我々は、実数値GAに、新たに、物質間の相互作用を考慮した構造探索オペレータを考案した [8]。つまり、探索終盤において、物質間の相互作用を考慮することにより、タイムコースの形状を大きく変えずに、かつ、ネットワークの構造を変化させることを考えた。S-Systemは、式 (1) から明らかなように、物質種 (遺伝子種) Xiの生成・分解へのXjの相互作用の種類 (生成を促進、生成を抑制、生成へ影響せず、分解を促進、分解を抑制、分解へ影響せず) を変化させたとしても、Xiの増減へのXjの影響さえ保存できれば、Xiのタイムコースの形状は、ある程度保存されると考えられる。Xiの増減への影響の観点から、Xiの生成・分解へのXjの相互作用を変化させる以下のような4つのオペレータを考案した。(1) 合併オペレータ: このオペレータは、XjからXiへの作用として、「Xiの生成を促進 (抑制)」と「Xiの分解の抑制 (促進)」がともに存在するとき、「Xiの生成を促進 (抑制)」または「Xiの分解の抑制 (促進)」のどちらか一つの作用にまとめても、タイムコースの形状は大きくは変化しないという考えに基づいている。すなわち、gij, hijの値の符号をそれぞれ、s (gij)、s (hij) とした場合、(適用の前提条件):

s (gij) = '+' かつ s (hij) = '-'、または、s (gij) = '-' かつ s (hij) = '+'、(操作): s (gij)、s (hij) のどちらかの要素を '0' に変更する (図5参照)。(2) 分配オペレータ: このオペレータは、XjからXiへの作用として、「Xiの生成を促進 (抑制)」または、「Xiの分解の抑制 (促進)」のどちらかが存在するとき、「Xiの生成を促進 (抑制)」と「Xiの分解の抑制 (促進)」の両者に分配しても、タイムコースの形状は大きく変化しないという考えに基づいている。(適用の前提条件): s (gij) = '+' かつ s (hij) = '0'、または、s (gij) = '-' かつ s (hij) = '0'、または、s (gij) = '0' かつ s (hij) = '+'、または、s (gij) = '0' かつ s (hij) = '-'、(操作): s (gij) = '+' かつ s (hij) = '0' の場合、s (hij) = '-' とする。s (gij) = '-' かつ s (hij) = '0' の場合、s (hij) = '+' とする。s (gij) = '0' かつ s (hij) = '+' の場合、s (gij) = '-' とする。s (gij) = '0' かつ s (hij) = '-' の場合、s (gij) = '+' とする (図6参照)。(3) 置換オペレータ: このオペレータは、XjからXiへの作用として、「Xiの生成を促進 (抑制)」または、「Xiの分解の抑制 (促進)」のどちらかが存在するとき、「Xiの生成を促進 (抑制)」を「Xiの分解の抑制 (促進)」で置き換えたとしても、また逆に、「Xiの分解の抑制 (促進)」を「Xiの生成を促進 (抑制)」で置き換えたとしても、タイムコースの形状は大きく変化しないという考えに基づいている。(適用の前提条件): s (gij) = '+' かつ s (hij) = '0'、または、s (gij) = '-' かつ s (hij) = '0'、または、s (gij) = '0' かつ s (hij) = '+'、または、s (gij) = '0' かつ s (hij) = '-'、(操作): s (gij) = '+' かつ s (hij) = '0' の場合、s (gij) = '0' かつ s (hij) = '-' とする。s (gij) = '-' かつ s (hij) = '0' の場合、s (gij) = '0' かつ s (hij) = '+' とする。s (gij) = '0' かつ s (hij) = '+' の場合、s (gij) = '-' かつ s (hij) = '0' とする。s (gij) = '0' かつ s (hij) = '-' の場合、s (gij) = '+' かつ s (hij) = '0' とする (図7参照)。(4) 反転オペレータ: XjからXiへの作用として、「Xiの生成を促進 (抑

制)」と「Xiの分解の抑制（促進）」がともに存在するとき、これらの作用はお互いに打ち消しあうと考えられるので、「Xiの生成過程および分解過程への影響なし」としても、タイムコースの形状は大きく変化しないと考えられる。逆に、XjからXiへの作用として、「Xiの生成過程および分解過程への影響がない」とき、「Xiの生成を促進（抑制）」および「Xiの分解の抑制（促進）」を導入しても、タイムコースの形状は大きく変化しないと考えられる。反転オペレータは、XjからXiへの作用の有無を反転する構造探索オペレータである。（適用の前提条件）： $s(g_{ij}) = '+'$  かつ  $s(h_{ij}) = '+'$ 、または、 $s(g_{ij}) = '-'$  かつ  $s(h_{ij}) = '-'$ 、または、 $s(g_{ij}) = '0'$  かつ  $s(h_{ij}) = '0'$ 、（操作）： $s(g_{ij}) = '+'$  かつ  $s(h_{ij}) = '+'$  の場合、 $s(g_{ij}) = '0'$  かつ  $s(h_{ij}) = '0'$  とする。 $s(g_{ij}) = '-'$  かつ  $s(h_{ij}) = '-'$  の場合、 $s(g_{ij}) = '0'$  かつ  $s(h_{ij}) = '0'$  とする。 $s(g_{ij}) = '0'$  かつ  $s(h_{ij}) = '0'$  の場合、 $s(g_{ij}) = '+'$  かつ  $s(h_{ij}) = '+'$ 、または、 $s(g_{ij}) = '-'$  かつ  $s(h_{ij}) = '-'$  とする（図8参照）。

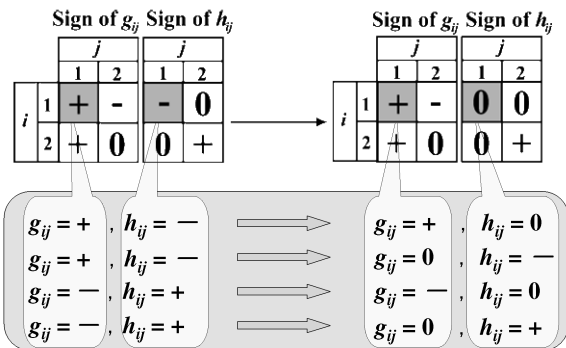


図5 合併オペレータの操作

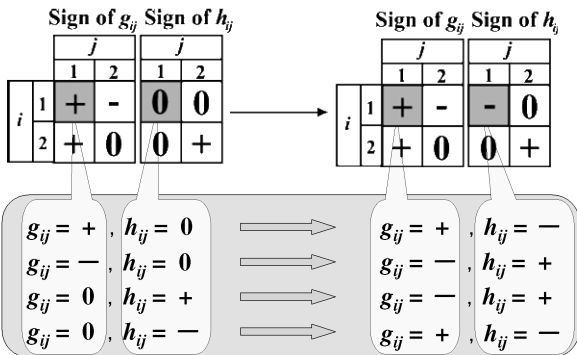


図6 分配オペレータの操作

次に、実数値GAで用いている世代交代モデルMGGを並列実装するために、MGGの個体の評価値計算を複数のワーカー・ノードで実行し、その他の処理をマスター・ノードで実行するマスター・ワーカーモデルを設計・開発した[7, 10]。設計したマスターの概念図を図9に示す。マスターにおいては、図に示すように、メイン・スレッド、世代交代スレッド、クライアント・スレッドの3種類のスレッドが協調しながら動作している。メインスレッドは、他のスレッドの初期化、世代交代スレッドとクライアント・スレッドの通信用キューなどデータ構造の初期化、初期集団の生成などの処理を行う。また、ユーザからのワーカーの追加・削除の要求を処理する。世代交代スレッドは、複製選択、子の生成、生存選択を行うスレッドである。世代交代スレッドの主な処理の流れを

以下に示す：

- (1) 集団から親個体を選択する。
- (2) 選択された親個体に交叉および突然変異を適用して複数個の子個体を生成する。
- (3) 評価値計算のため、生成した子個体をクライアント・スレッドとの通信用のキューに登録する。
- (4) 全ての子個体の評価値計算が正常に終了するか、ユーザが定めた時間を経過するまで待機する。
- (5) 評価値計算が正常に終了した子個体と両親の中から最良2個体を選択して、集団中の両親と入れ替え、世代数カウンタを増す。
- (6) 打ち切り世代数に達した場合、世代交代スレッドに打ち切り世代数に達したことを通知して終了する。打ち切り世代数に達していない場合はステップ1へ。

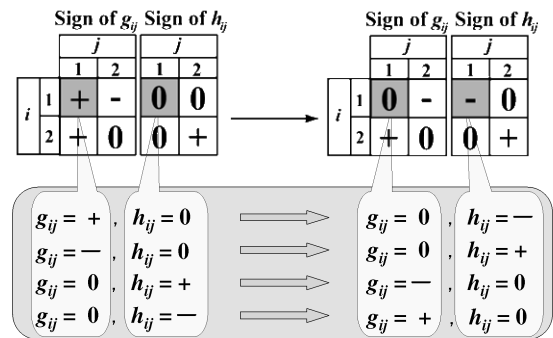


図7 置換オペレータの操作

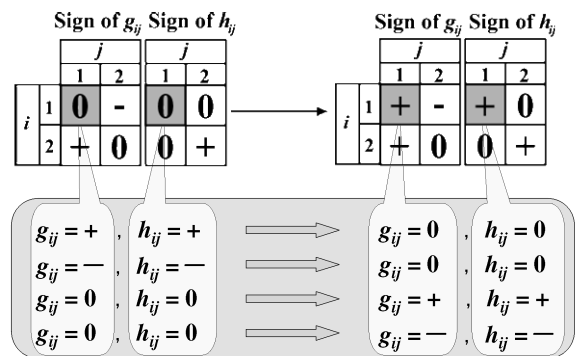


図8 反転オペレータの操作

キューを通じてクライアント・スレッドに個体を絶えず供給するために、図9に示すように、複数の世代交代スレッドが動作している。そのため、集団中の親個体は複数の世代交代スレッドに選択されないようにロック機構を有している。また、世代交代スレッドは全ての生成個体の評価が終了するまで新たにキューに個体を供給できないため、効率性の観点から、同時に複数のスレッドがステップ1~3を実行できないようになっている。クライアント・スレッドは、ワーカーに一对一に対応して動作するスレッドである。クライアント・スレッドの主な処理の流れを以下に示す：

- (1) ユーザが指定した数の個体をキューから取り出す。指定された数の個体がキューに供給されていない場合は、ある一定時間、個体の供給を待つ。
- (2) ステップ1において、キューから取り出せた個体の数が0であり、かつ打ち切り世代数に達している場合、終了する。
- (3) Javaのソケットを用いて、キューから取り出され

た個体群の評価値計算をワーカー側で行う。

- (4) 世代交代スレッドに個体の評価値計算が終了したことを通知する。
- (5) ステップ1へ。

ステップ3が異常終了した場合、メインスレッドにその旨を通知して終了することにより、対応するワーカーに異常が起きても、全体の計算は止まらないようになっている。

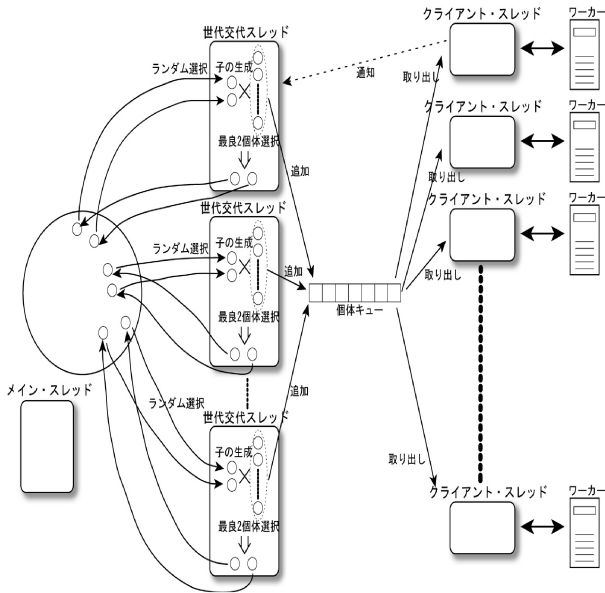


図9 MGGの並列実装： 個体キューへはS-systemの各パラメータの符号(+、ゼロ、-)のみのセットが送られ、クライアント・スレッドを介して各ワーカー上で、符号を考慮した探索範囲内(実数値)で、UNDX+MGGを用いて探索を行い、最良解の評価値を該当する世代交代スレッド内の子個体に通知する。

図1でも示したように、我々は、コンピュータが唯一無二の最適解を提示するのではなく、分子生物学実験を行っている生物学者と対話しながら、計算と実験をフィードバックしながらネットワークの推定を行っていく、対話型遺伝子相互作用推定システムを開発した。ネットワークに関する生物学的知見を考慮しながら解候補を絞り込むことが可能になる。そのためには、パラメータ推定については、コンピュータは限られた実験観測データを再現しうるパラメータセットをできるだけ多く提示できるようなアルゴリズムを用いる必要があり、クラスターコンピュータやグリッドコンピュータが必須となる。さらには、最適化の進行状況を描写したり、最終結果を解析するGUIも必要である。図10は、20プロセス(20CPU)を用いて5つの遺伝子間相互作用の推定を行っているシステムを管理するGUIのスクリーンショットである。図の上半分は、20プロセスのそれぞれの最適化の進行状況を表しており、縦軸は、観測データ1点あたりの計算値との誤差(%)を示している(横軸は世代数)。下半分は、あるプロセスの世代ごとの推定したネットワーク構造を表している。

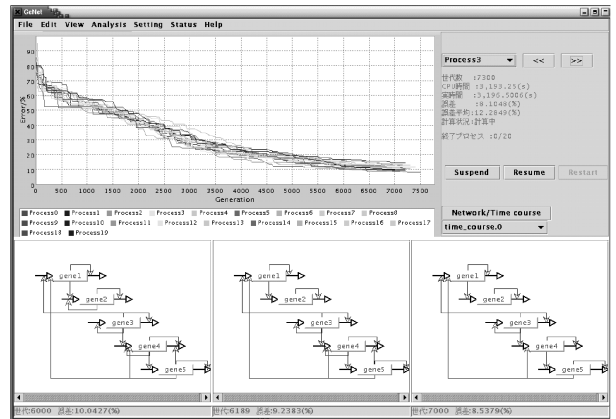


図10 並列コンピュータを用いた遺伝子ネットワーク推定のためのGUI画面(1):全プロセスの時々刻々のネットワーク推定状況を描写できる。

また、図11に示すGUI画面は、実験観測タイムコースと計算タイムコースを世代ごとに描写している。

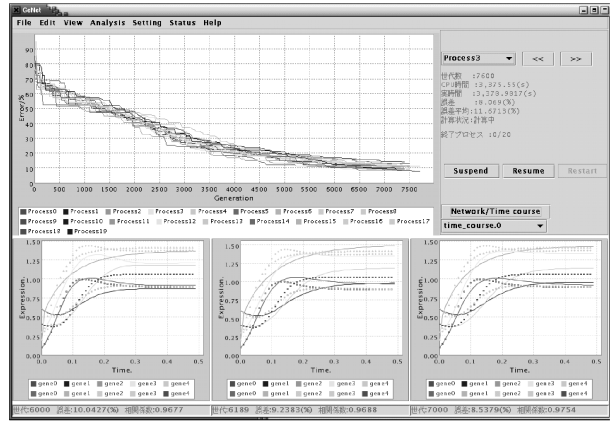


図11 並列コンピュータを用いた遺伝子ネットワーク推定のためのGUI画面(2):全プロセスの時々刻々の推定結果(実験観測タイムコースと計算タイムコースの比較)を描写できる。

## 2) 小規模な遺伝子ネットワークシステムでの詳細な相互作用推定

1で開発した遺伝子相互作用推定システムの有用性を検証するために5遺伝子から構成される図12のネットワークの推定問題に適用した。また、図12ネットワークをS-System表記で表したときの60個のキネティックパラメータの値を表1に示す。

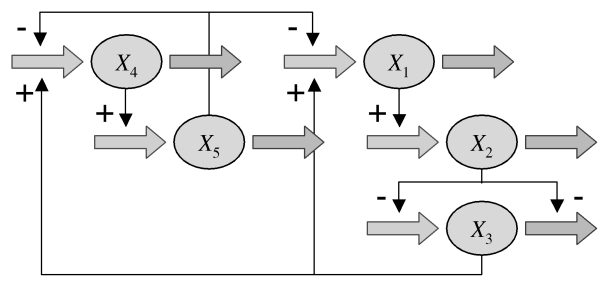


図12 遺伝子ネットワークモデル:それぞれの遺伝子の合成過程、分解過程に他の遺伝子から、活性化(induce)、抑制(suppress)の相互作用があると仮定している。

表1 図12のネットワークを表すS-Systemのパラメータ値

$\alpha$	$g_1$	$g_2$	$g_3$	$g_4$	$g_5$	$\beta$	$h_1$	$h_2$	$h_3$	$h_4$	$h_5$
6.0	0.0	0.0	1.0	0.0	-1.0	10.0	2.0	0.0	0.0	0.0	0.0
8.0	2.0	0.0	0.0	0.0	0.0	8.0	0.0	2.0	0.0	0.0	0.0
10.0	0.0	-1.0	0.0	0.0	0.0	5.0	0.0	-1.0	2.0	0.0	0.0
8.0	0.0	0.0	2.0	0.0	-1.0	8.0	0.0	0.0	0.0	2.0	0.0
6.0	0.0	0.0	0.0	2.0	0.0	10.0	0.0	0.0	0.0	0.0	2.0

次に、野生型株および1遺伝子破壊株を想定した6種の発現タイムコースデータを準備した(図13参照)。これらのタイムコースデータを開発した遺伝子相互作用推定システムに入力として与える。システムは、それらのタイムコースデータを再現しうる60個のS-Systemパラメータの値を推定し、その結果からネットワーク構造を同定するわけであるが、真の構造は図12および表1である。図13(A)は、すべての遺伝子が働いている野生型株(wild type)、(B)~(F)は、それぞれ遺伝子1、2、3、4、5が欠損している1遺伝子破壊株(single gene disrupted strain)を想定している。具体的には、たとえば遺伝子1が破壊された場合(B)は、 $\alpha_1 = 0$ として数値計算を行う。その結果、表2に示すように、5回の試行において、すべて真の構造(図12)を見つけることに成功した。なお、表2において、満足構造とは、前にも述べたように、図9の実験タイムコースを再現できる、図12の真の構造を含むネットワーク構造のことである。つまり、それらはすべて、図13を再現する解構造である。最終的に絞り込むためには新たな実験タイムコースデータが必要となる。

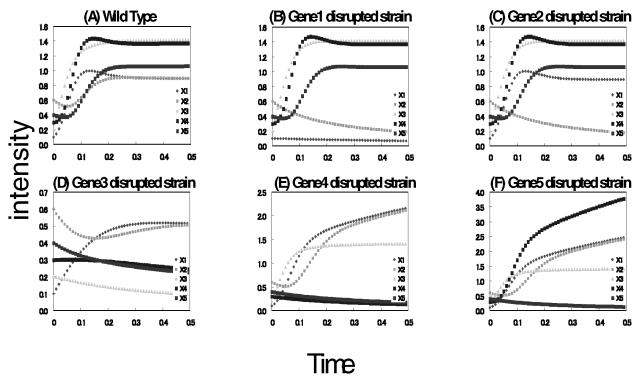


図13 推定システムに与えるタイムコースデータセット;これらのデータセットを用いて、表1に示す60個の実数値パラメータを推定させる。

表2 5遺伝子ネットワーク構造推定問題の結果

試行	1	2	3	4	5
探索構造数	10825	9526	14030	14934	10384
満足構造数	50	67	70	62	82
図8の構造を含む?	Yes	Yes	Yes	Yes	Yes

### 3) 離散モデルと連続モデルの統合法の確立

大規模系を解析する場合、これまでに、我々は、離散モデルと連続モデルの組み合わせ手法によって遺伝子ネットワークの推定を行うことを提案してきた。推定のための戦略を図14に示す。その中で、S-Systemを用いることで、フィードバックのようなループ構造を推定できることを検証した。その手法は以下のようにまとめられる[9]。

## 遺伝子発現制御ネットワーク推定のための戦略

発現プロファイルデータに応じてモデルを組み合わせ段階的にネットワークを推定

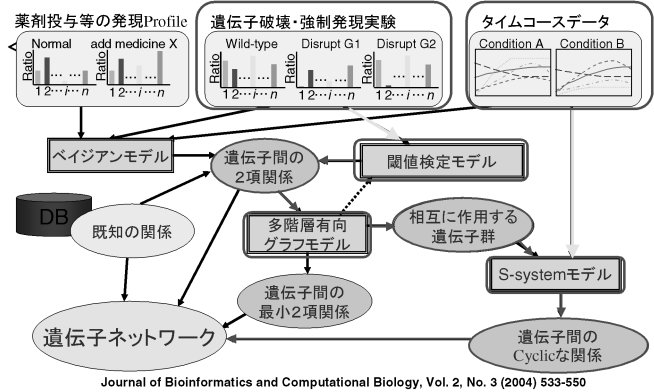


図14 離散モデルと連続モデルの統合:遺伝子発現ネットワーク推定のための戦略

(1) まず、用意した解析対象遺伝子数+1セットのタイムコースデータの定常状態の値(発現量)に着目して、閾値検定モデルの解析手順に従って、二項関係を導出する。この場合、過渡状態の発現量データは用いない。(2) 次に、閾値検定モデルで導出した二項関係を用いて、多階層有向グラフモデルによりグループ化を行う。(3) 次に、S-Systemを用いた推定を行う。まず、同値類でない遺伝子間の相互作用について、正か負かは定常状態の発現量データより確定できる。同値類の遺伝子間相互作用については、同値類に属する遺伝子群と同値類に影響を与えている遺伝子の自然状態の発現量のタイムコースデータと関係する遺伝子一つずつ破壊した場合の発現量のタイムコースデータを与えて、S-Systemモデルのパラメータの最適化を行う。この手法におけるキーポイントはデータセット数である。実験上のコストから多セットの破壊実験のタイムコースデータを収集することは困難である。そこで、同値類に属する遺伝子群の相互作用推定に、どの程度の実験データセット数が必要かを検証した。その結果、解析に用いるデータセット数は解析対象遺伝子数に対して60%程度必要であった。しかしながら、この60%(解析対象遺伝子数に対する実験数の比率)という数値は、解析対象遺伝子数が多くなるほど実験側に多くの実験を要求するものであり、数千、数万オーダーの遺伝子を対象とした場合、求められる実験数は60%程度には減るが、その数は数百、数千のオーダーとなるため非現実的な数値である。また、解析対象とする遺伝子群の選択によってはエリート個体のみでは正しいネットワーク構造を見出すことは難しい。そこで、導出したネットワーク構造について、有意なネットワーク構造を見出すために、全試行(50回分)における最終世代全ての個体を対象に遺伝子間の二項関係を抽出した。その結果、WildTypeのタイムコースデータのみでは、正解ではない二項関係を抽出する可能性があるが、WildTypeの他の破壊実験のデータセットを1セット以上追加することで、間違ったネットワーク(二項関係)を導出しないことが明らかになった。

### 4) 推定した相互作用ネットワーク構造からの知識抽出

図15に示すように、複数の遺伝子発現プロファイルの時系列データセットを基に、それらを再現しうる遺伝子間相互作用ネットワークを推定することは、数学的には、逆問題であり、解を一意に決定することはできない。そこで、開発した推定システムを用いて、推定試行を繰

り返し、実験データを再現する多数のネットワーク候補を見出すことに成功した(表2参照)。これらの解候補から相互作用に関する重要な知識抽出ができるか検討した。

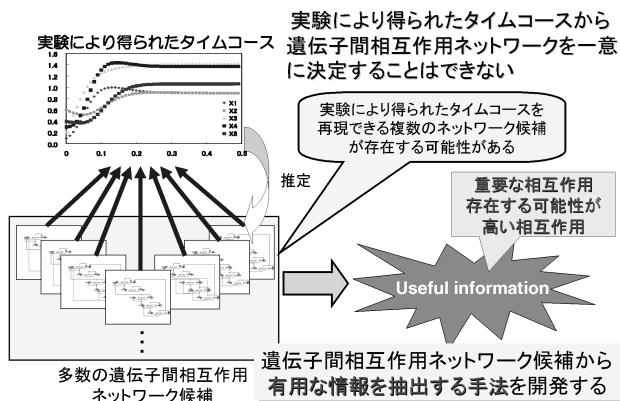


図15 推定した相互作用ネットワーク構造からの知識抽出

注目したのは、共通構造である。つまり、各ネットワーク候補において、状態変数同士の相互作用の符号(十なら活性化、一なら抑制、ゼロなら無関係)に着目し、全構造ですべてに共通に現れる共通構造とその他の構造(非共通構造)に分類し、共通構造のもつシステム解析的解釈を検討した。つまり、実験データである時系列データを再現する上で、どの相互作用が重要か(感度解析)調べ、その結果が、共通構造とどのような関係があるのか考察した。手法としては、各ネットワーク候補の各相互作用係数の値を標準値から1%変化させ、変化前と変化後で、タイムコースデータがどの程度変化したかを定量的に調べた(図16参照)。もし、ある相互作用係数のわずかな変化が、タイムコース全体の形状に大きく変化を与えるのであれば、その相互作用は、実験タイムコースを再現する上で重要かつ不可欠なものといえる。逆に、係数の変化がタイムコースの形状にほとんど影響を及ぼさないのであれば、その相互作用は、不可欠なものであるとはいえず、システムの冗長性(ロバストネス)に関わっている可能性が高いことになる(図17参照)。図12で与える5遺伝子のネットワーク構造をケーススタディーとし、WildTypeのタイムコースと1遺伝子破壊のタイムコースを与え、推定誤差範囲(3%から30%まで)を設定して、ネットワーク推定を各推定誤差範囲で20回試行し、20個のネットワーク候補を抽出し、共通構造と非共通構造の感度値の累計を調べた。感度評価式は、図18で表される。

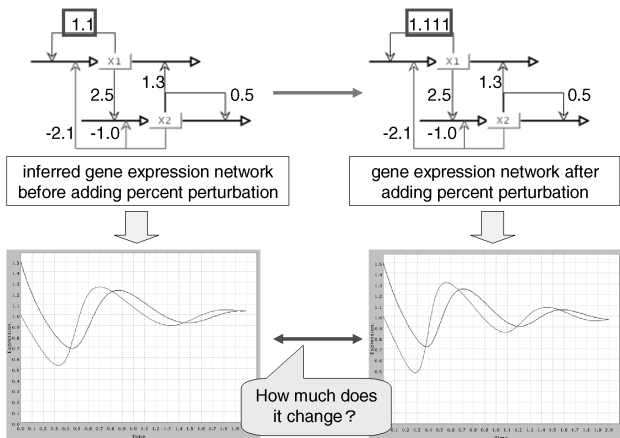


図16 相互作用係数の感度解析。各ネットワーク候補の各相互作用係数に1%の摂動を与え、得られるタイムコースの変化から相互作用係数の感度を調べる。

## Method : Sensitivity Analysis

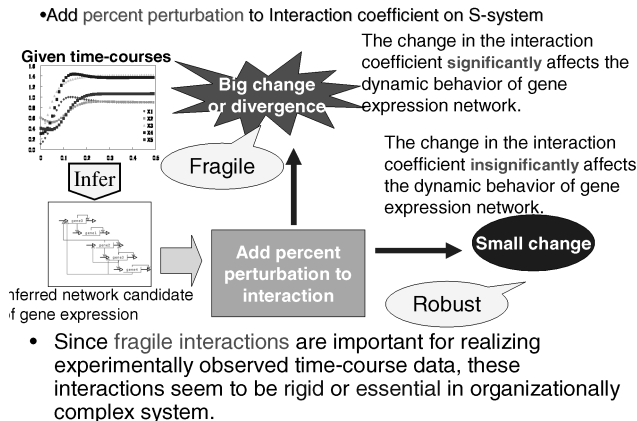


図17 相互作用の感度解析

$$S_{g_{kl} \text{ or } h_{kl}} = \sum_{d=1}^D \sum_{i=1}^n \sum_{t=0}^T \left( \frac{PER_{dit} - CAL_{dit}}{CAL_{dit}} \right)^2$$

Total relative error per sampling point of each time-courses

$g_{kl}$  or  $h_{kl}$ : interaction coefficient which is added percent perturbation

$PER_{dit}$ : gene expression level at time  $t$  of the  $i$ -th gene on time-course  $d$ , which is calculated from the gene expression network after adding percent perturbation.

$CAL_{dit}$ : gene expression level at time  $t$  of the  $i$ -th gene on time-course  $d$ , which is calculated from the gene expression network before adding percent perturbation.

$D$ : number of time-courses  $n$ : number of genes

$T$ : number of sampling points on a time-course

図18 相互作用の感度の評価式

なお、感度値は、1ネットワーク構造において、各相互作用の感度値の総計が1.0になるように正規化した。したがって、各推定誤差範囲において、20個のネットワーク候補が抽出されるので、感度値の累計は20.0になる。この20.0が、共通構造と非共通構造でどのように分配されるか調べた。以上の手順を図19にまとめる。その結果、共通構造および非共通構造1本あたりの感度値は図20のようになり、共通構造は、感度の高い相互作用であり、実験タイムコースデータを再現する上で、重要不可欠なものであることが示唆された。

We obtained 20 network candidates which were inferred from each time-course combination of Wild Type & 1 gene disrupted strain.

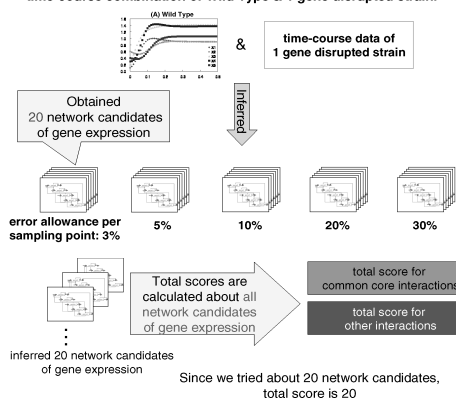


図19 図12のネットワーク構造をケーススタディーとして、WildTypeのタイムコースと1遺伝子破壊のタイムコースを与え、ネットワーク推定を各20回行う。その後、各相互作用係数の感度解析を行い、共通構造と非共通構造の感度値の累計を調べる。

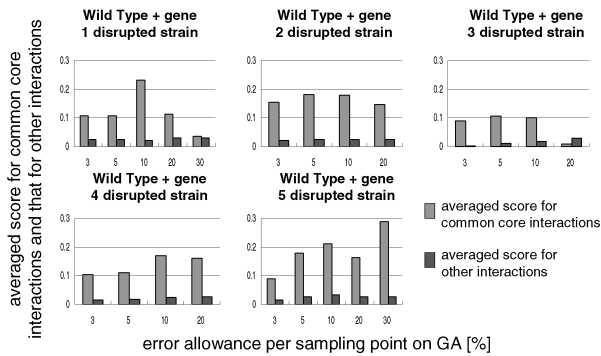


図2 共通構造および非共通構造1本あたりの感度値。各グラフの横軸は実数値GAで用いる推定誤差範囲。グラフは、上段左より、WildTypeと遺伝子1(X1)破壊株の実験タイムコース(図13(A)+13(B))を用いた場合、(図13(A)+13(C))を用いた場合、(図13(A)+13(D))を用いた場合、下段左より、(図13(A)+13(E))を用いた場合、(図13(A)+13(F))を用いた場合。

以上のことから、ネットワーク候補すべてに現れる共通の相互作用の抽出は、実験タイムコースを再現する必要不可欠な相互作用を探索する、第1スクリーニングの手順であることが示唆された(図21参照)。

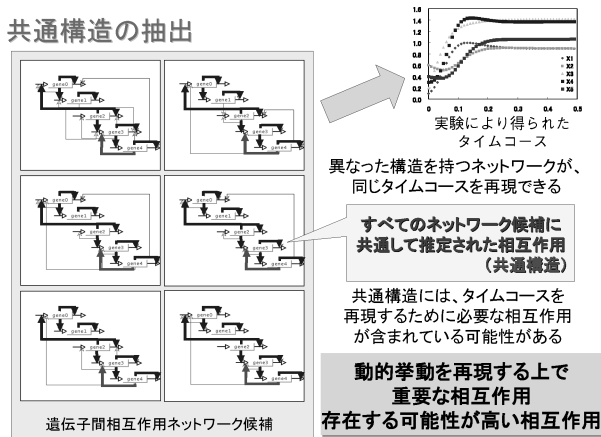


図21 共通構造抽出の意義

しかし、この段階では、共通構造の相互作用の種類(活性化、あるいは抑制)のみの論議であって、強さについての情報は含まない。つまり、実際に存在する可能性の高い二項関係の集合である。できるだけ多くの共通構造(相互作用)を見出すにはどのようにすればよいのだろうか。図13に示すWildType(図13(A))とm個の1遺伝子破壊株のタイムコース(m=1,2,...,5)の組み合わせを用いて、推定の目標誤差ごとに200回の遺伝子間相互作用ネットワーク推定を行った。そして、mや推定の目標誤差によって、共通構造(相互作用)の数がどのように変化するかを検討した。なお、mを決定した場合、試行回数をmごとに200回とした。つまり、m=1の場合、WildTypeとの組み合わせは、5種類考えられ(図13(A)+(B)、(A)+(C)、(A)+(D)、(A)+(E)、(A)+(F))、各組み合わせで200/5=40回試行するm=2の場合は、10種類組み合わせが考えられ、各組み合わせごとに20回試行する。推定の目標誤差は、3%、5%、7%、10%、20%とした。各推定の目標誤差において、mを固定し、200回の推定後に、すべてのネットワーク構造で共通に存在する構造(共通二項関係)の数を、表3にまとめる。

表3 ネットワーク推定後に抽出された共通構造の数

ケース	推定の目標誤差				
	3%	5%	7%	10%	20%
図8の二項関係の数	8				
WildTypeのみの場合	0	0	0	0	0
m=1(Wild+1set)の場合	0	0	0	0	0
m=2(Wild+2set)の場合	0	0	0	0	0
m=3(Wild+3set)の場合	0	0	0	0	0
m=4(Wild+4set)の場合	1	2	0	0	0
m=5(Wild+5set)の場合	4	4	4	4	1

\*数字は自己分解以外の二項関係数

表3において、数字は自己分解以外の共通二項関係の数であり、オリジナルの図12のネットワークでは、8個存在する。ほとんどの試行において、抽出された共通二項関係の数は、0であり、すべてのタイムコースデータを用いても、オリジナルの半分しか抽出できなかった。そこで、抽出される共通二項関係の数を多くするように、抽出方法の検討を行った。その結果、図18に示す抽出方法を考案した。図22は、m=4の場合を想定したもので、

### タイムコースの組み合わせごとに共通構造を抽出 タイムコース数:5(野生株+1遺伝子破壊株×4)の場合

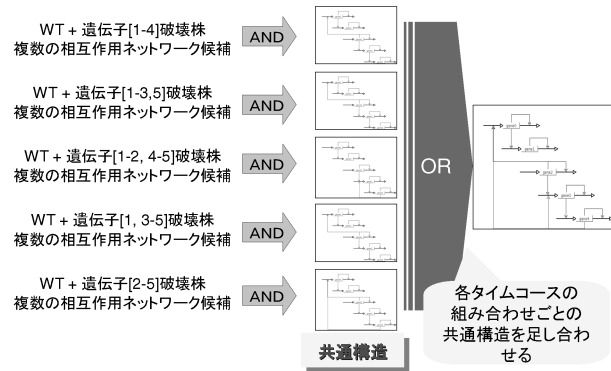


図22 共通二項関係を抽出する改良法

手順は次のようにまとめられる。1) m=4の場合、WildType(WT)との組み合わせは5通り考えられる。その組み合わせごとに、200/5=40回ネットワーク推定を行う。2) それぞれの組み合わせごとに、共通構造を抽出する。3) 各組み合わせにおいて抽出された共通二項関係をすべて足し合わせる。それを最終的な共通二項関係とする。この抽出手法を用いて、表3に相当する検証を行った。その結果を表4に示す。表3と表4を比較すると、従来の抽出法では、すべてのタイムコースデータを用いて8個中4個の相互作用を抽出できたのに対して、図22の方法では、少ないタイムコースデータを用いて、しかも推定の目標誤差をあまりしても同様の結果を得ることができることがわかる。特に、m=4の目標誤差20%で、8個中4個を抽出でき、目標誤差を厳しくして3%すれば、8個中6個(75%)を抽出できている。なお、表4において括弧の中の数字は、図12のネットワークにはない二項関係の個数であって、本来+の相互作用をーに、あるいは、ーの相互作用を+と間違えたものではない。つまり、括弧の数の相互作用は、false-positiveな相互作用といえる。



表3 ネットワーク推定後に図22で抽出された共通構造の数

ケース	推定目標誤差				
	3%	5%	7%	10%	20%
図8の二項関係の数	8				
WidTypeのみの場合	0	0	0	0	0
m=1 (Wild + 1set)の場合	5(1)	5(1)	4	3	0
m=2 (Wild + 2set)の場合	9(3)	5(1)	6(2)	5(1)	5(1)
m=3 (Wild + 3set)の場合	7(1)	4	6(2)	5(1)	5(1)
m=4 (Wild + 4set)の場合	6	4	4	4	4
m=5 (Wild + 5set)の場合	4	4	4	4	1

\* 数字は自己分解以外の二項関係数

\*\* ( )の数は図8にない二項関係数

m=4で、目標誤差3%の場合は、8個中6個を抽出しており、残りの2個の相互作用は、図12において、遺伝子2 (X2)が遺伝子3 (X3)の合成過程を阻害する相互作用 ( $g_{32} < 0$ )と、遺伝子2 (X2)が遺伝子3 (X3)の分解過程を阻害する相互作用 ( $h_{32} < 0$ )である。これら2つの相互作用は、先に反転オペレータで述べた、互いに打ち消しあう相互作用なので、抽出できなかったのも仕方ないことであろう。そのことを考慮すると、図22の抽出方法は、図12の自己分解以外のすべての遺伝子二項関係を抽出したことになる。

### 5) 大規模系解析への方策の検討

3)の離散モデルと連続モデルの統合法の確立でも述べたが、大規模な遺伝子ネットワークの相互作用を推定するためには、いきなり最初からS-System法を導入するわけではなく、まず、定常状態での遺伝子の発現量に着目して、遺伝子破壊、強制発現等の実験で、それぞれの遺伝子の発現量がどのように変化したかの発現プロファイルを用意する。次に、閾値検定モデルの解析手順に従って、二項関係を導出し、多階層有向グラフモデルによりグループ化を行う(図23参照)。そして、グループ内において、遺伝子発現の時系列変化(タイムコース)を用いて、S-Systemを用いた推定を行う。しかし、たとえグループ化(クラスタリング)しても、グループを構成する遺伝数は2桁のオーダーであろう。

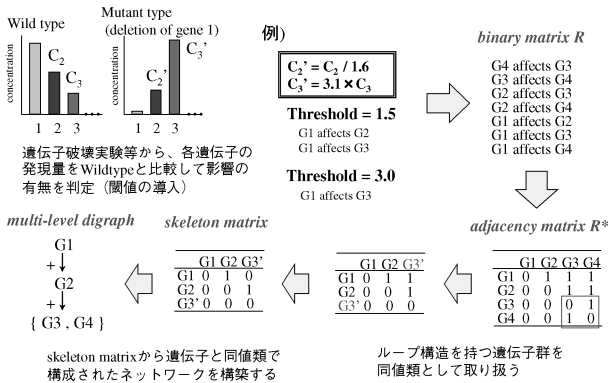


図23 多階層有向グラフ法の解析フロー

S-Systemの構造は、図24のようなマトリックスで表記することができ、その要素数は、 $2n(n+1)$ 個 ( $n$ は遺伝子数)である。 $n$ の増加に対して、 $n$ の2乗オーダーで推定するパラメータの数が増えていくので、パラメータの推定法を検討した。その結果、新たに、Step-by-Step strategyを開発した[2](図25参照)。従来は、図24のマトリックス全体要素の値の探索を同時に行っていたのに対して、図25の方法は、マトリックスの1行の要素の値のみの探索を行う。たとえば、遺伝子1の発現タイムコースに着目し、このタイムコースを再現する遺伝子1に関わる相互作用(図24の1行目の要素)の値を探索する。そ

の際、遺伝子1以外の遺伝子の発現値は、実験データより読み取り、その時刻での値として固定する。そして、1行目の探索が終われば、2行目、3行目と探索を続け、 $n$ 行目まで行う。

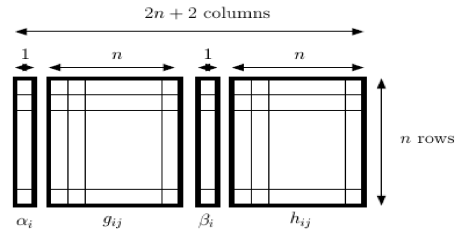


図24 S-Systemのパラメータのマトリックス表記

For the estimation of more large number of parameters...

◆ Step-by-step strategy [Maki, 02]

This strategy is summarized as follows;

- 1) Focused on one temporal profile of gene expression ( $i$ ), and it is supposed that other temporal profiles are treated as known and fixed data without noise.
- 2) Estimate the parameter value in row  $i$ , which realizes the temporal profile of gene expression ( $i$ ).
- 3) Repeat above procedures ( $n-1$ ) times.

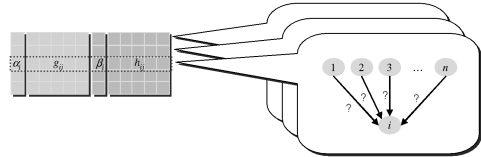
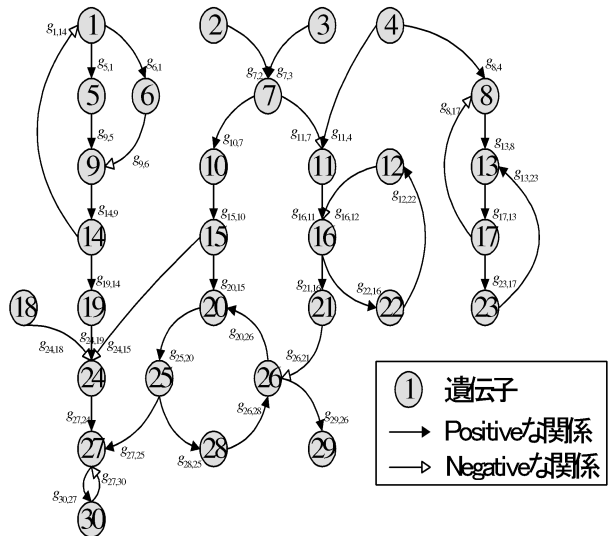


図25 Step-by-Step strategy の手順

図25の手法の有用性を検証するため、図26に示す30個の遺伝子から成る相互作用ネットワークを人為的に構築し、図25の手法を導入した推定システムに、WildTypeおよび30個の遺伝子それぞれの1遺伝子破壊株を想定した31種類のタイムコースデータ(図27)を入力として与え、図26のネットワーク構造が再現できるか調べた。



26 30 遺伝子ネットワーク: closed arrow および open arrow は、それぞれ、正および負の相互作用を表す。

推定すべきパラメータの数は960個である。その結果、最終的に図28に示すネットワーク構造を得ることができ、これは図26に示すネットワーク構造と同一であっ

た。

以上のことから、大規模系解析の手順の1つとして、Step-by-Step strategyが有効であることが明らかとなった。

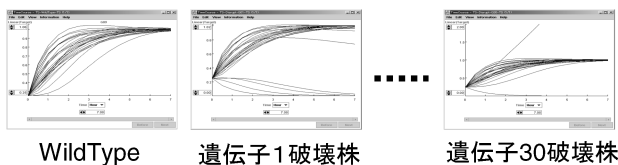


図27 図26のネットワークの推定のために与えるタイムコースデータ。野生株と30個の1遺伝子破壊株を想定した31種類のタイムコースデータ。

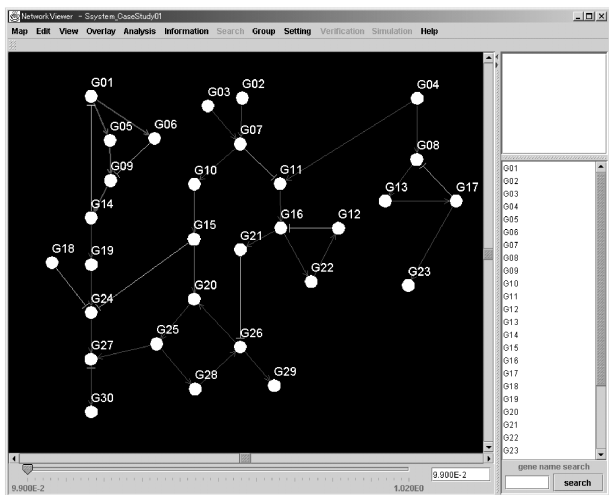


図28 Step-by-Step strategy を用いて推定した最終ネットワーク構造

### 〈国内外での成果の位置づけ〉

この計画研究がスタートした2000年当時は、連続モデルといえば、線形多項式で表された微分方程式であった。つまり、遺伝子間の相互作用を線形と近似する、いわば、線形全結線型ニューラルネットの微分方程式であった。我々は、遺伝子発現量の時間的変化の非線形性を表現でき、かつ、モデル内のキネティックパラメータの生物物理学的解釈が可能な連続モデル表記法として、S-System表記を提案し、モデル内のパラメータの最適化したその結果から相互作用ネットワーク構造を明らかにする方法を世界に先駆けて提案した。5年を経て、国内外の遺伝子ネットワーク解析の手法は、離散値手法としての、ブーリアンモデル、多階層有向グラフ、連続値手法としての、ベージアンモデル、常微分方程式モデル、統計的手法としての、グラフィカルガウシアンモデルに分類されるが、S-Systemは、常微分方程式モデルに属し、時系列データからループ構造を含む非線形の相互作用を推定できる唯一の手法として広く使われている。現在のところ、マイクロアレイを用いて時系列データをとることは経済的にもコスト高であることから、未だに、定常状態での遺伝子発現量を用いて、離散値手法、統計的手法で解析する研究、または、時系列データが得られたとしても、クラスタリング手法を用いた解析が主流であるが、様々な時系列データを数多くとることができれば、S-Systemを用いた解析がさらに普及するものと思われる。

5年前、工学領域における最適化問題の研究は、予め最適解が設定され、初期設定値から如何に効率よく最適解まで探索できるかが主体であった。様々なベンチマークテストを経て、多変数の数値最適化手法として、遺伝的アルゴリズムが普及しているが、本研究のような、生物

現象の問題にはそのまま適用できなかった。つまり、マイクロアレイなどで得られる出力から、システムの内部構造(相互作用ネットワーク)を予測することは、逆問題であり、不良設定問題(ill-posed problem)に属し、何らかの拘束条件を与えなければ、解を求めることはできない。その拘束条件として、実験から得られた定常状態での発現データあるいはタイムコースデータであるが、前にも述べたように、パラメータを決定できるほど、十分な実験データを獲得することはできないのが現実である。したがって、我々は、解候補となりうるものをできる限り拾い上げ、ユーザに示すことで、ユーザが解を絞り込んだり、遺伝子破壊等の新たな実験計画を考案できるような推定システムを開発してきた。このような試みは、遺伝的アルゴリズムなどを取り扱っている進化アルゴリズムの国際会議(Congress on Evolutionary Computation)でも注目を浴びてい[8]。

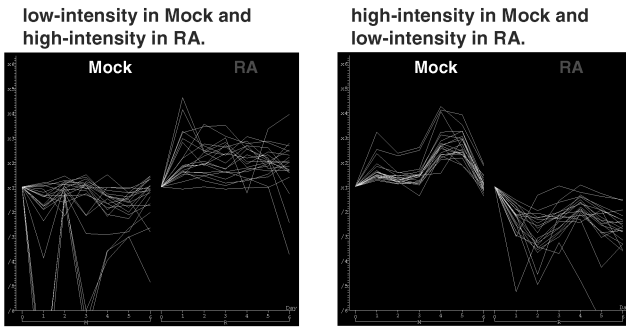
### 〈達成できなかったこと、予想外の困難、その理由〉

開発したアルゴリズムを適用する実験データの入手が非常に少なく、あくまでもケーススタディーとして実験データを仮想的に作成し、アルゴリズムの有効性を確かめることができなかった。

しかし、筑波大学医学部の内田和彦研究グループとの共同研究で、マウスP19細胞を用いた神経細胞の分化に関わる遺伝子群の相互作用推定を、遺伝子発現(mRNA)の時系列データを用いて行った[2, 6]。P19細胞は、薬剤、レチノイン酸(Retinoic acid)を添加することで、7日後にはニューロンへと分化する。薬剤を添加しなければ、筋肉細胞になる。薬剤添加によって、発現が変化する(up-regulationあるいはdown-regulation)遺伝子をつとめ、それらの相互作用を推定しようとした。

ターゲット遺伝子を11に絞り込み、それらに、薬剤を加えたとき(RA)と加えないとき(Mock)の各遺伝子の発現の時系列データをとった(図2.9参照)。図において、横軸は時間で、0日目から6日目までの発現量の変化を、MockとRAのそれぞれに関してup-regulation、down-regulationで分けて描画している。この時系列データを推定システムに与え訳であるが、薬剤を加えることで、ネットワーク構造が変化する可能性は否定できない。もし構造が変化するのであれば、加えたときと加えないときの両方のタイムコースを再現するようなパラメータ推定を行うことができず、加えたとき、加えないときのそれぞれで推定しなければならない。これまで開発してきた推定システムは、すべての実験タイムコースデータを再現するネットワーク構造を探索するものであり、ネットワーク構造は実験によって変化しないという前提に立っている。したがって、実験タイムコースのセット数が多いほど、ネットワーク構造の解候補が絞られてくるようになっていく。実験などで、ネットワーク構造が変化するならば、推定のための拘束条件を増やすことができず、推定システムの設計そのものを再考する必要がある。

- 回 Genes related to the translation and signal transfer
- 回 High expression & Welch t-test (P=0.01)



K. Uchida, N. Uematsu, University of Tsukuba

図2.9 レチノイン酸を添加したとき(RA)、添加しないとき(Mock)の遺伝子発現の時間的変化。4つに分類している。

### 〈今後の課題〉

システム生物学の研究では、これまで、システムの構成因子の意味づけとその相互ネットワーク構造を推定・同定するための情報科学的手法の構築（システム同定）とシステムの構成要素の時間的変化を解析・シミュレーションを行うための情報・数理科学的手法の構築（システム解析 [3, 5]）の2点に重点が置かれていた。しかしこの後、システムの制御方法の理解を目指したシステム制御と、その設計方法の確立（システム設計）に関する諸問題が残されている。現在まで開発されているほとんどの代謝シミュレータは、シミュレーション、パラメータ推定までの機能しか備えておらず、たとえばある代謝物質の生産を最大化するための制御方策や代謝物質の特定の挙動を再現するためのシステム設計などを行うことはできない。システム同定、解析、制御、設計を統合したシステム生物学プラットフォームの構築が今後の課題として残される。現在の代謝パスイ解析は主に物質の流れの解析であり（event driven）、フィードバック制御等の情報が十分に組み入れられていない。動的最適化あるいは最適制御の理論・アルゴリズムを導入することで、ゲノム・メタボロームシステムの構成要素の単なる時間挙動シミュレーションが制御方策提案のためのシミュレーションへ拡張できると考える [11]（図3.0参照）。

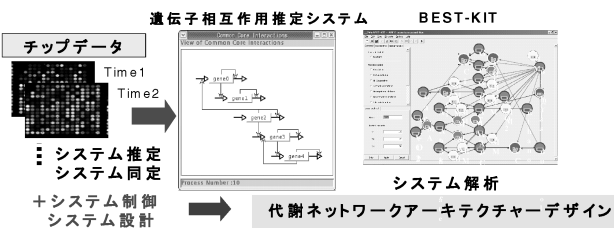


図3.1 システム生物学プラットフォーム

### 〈研究期間の全成果公表リスト〉

- [1] 305121556  
T. Ueda, I. Ono, M. Okamoto: Development of System Identification Technique Based on Real-coded Genetic Algorithm, *Genome Informatics*, 13, 386-387 (2002) .
- [2] 305121607  
Y. Maki et al.: Inference of Genetic Network Using the Expression Profile Time Course Data of Mouse P19 Cells, *Genome Informatics*, 13, 382-383 (2002) .

- [3] 30828133  
岡本正宏: 情報科学的手法を用いた生命ソフトウェアの解明, *生物物理*, 43 (4) , 192-197 (2003) .
- [4] 404081347  
岡本正宏, 小野 功: 実数値GAのバイオ分野への応用-大規模遺伝子ネットワークの相互作用推定, *人工知能学会誌*, 18(5), 502-509 (2003).
- [5] 404081408  
J. Yoshimura, T. Shimonobou, T. Sekiguchi, M. Okamoto: Development of the Parameter-fitting Module for Web-based Biochemical Reaction Simulator BEST-KIT, *Chem. Bio. Inform. J.*, 3 (3) , 114-129 (2003) .
- [6] 404081431  
N. Uematsu, Y. Maki, M. Okamoto, K. Uchida: Analysis of Genetic Networks Using Time-course Data of Gene Expression Profiling, *Cytometry Research*, 13 (2), 45-53 (2003) .
- [7]  
H. Imade, R. Morishita, I. Ono, N. Ono, M. Okamoto: A Grid-Oriented Genetic Algorithm Framework for Bioinformatics, *New Generation Computing*, 22, 177-186 (2004) .
- [8]  
I. Ono, Y. Seike, R. Morishita, N. Ono, M. Nakatsui, M. Okamoto: An Evolutionary Algorithm Taking Account of Mutual Interactions Among Substances for Inference of Genetic Networks, *Proc. of 2004 Congress on Evolutionary Computation (CEC2004)* , 2060-2067 (2004) .
- [9]  
Y. Maki, Y. Takahashi, Y. Arikawa, S. Watanabe, K. Aoshima, Y. Eguchi, T. Ueda, S. Aburatani, S. Kuhara, M. Okamoto: Integrated Comprehensive Workbench for Inferring Genetic Network: VOYAGENE, *J. Bioinformatics and Computational Biology*, 2 (3) , 533-550 (2004) .
- [10] 601281523  
H. Imade, R. Mizuguchi, I. Ono, N. Ono, M. Okamoto: Gridifying: An Evolutionary Algorithm for Inference of Genetic Networks Using the Improved GOGA Framework and its Performance Evaluation on OBI Grid, *Lecture Notes in Bioinformatics*, 3370, 171-186 (2005) .
- [11] 601281530  
岡本正宏: バイオインフォマティクスおよびシステム生物学の最近の動向, *生物工学ハンドブック*, コロナ社, 337-347 (2005) .