

# イン・シリコ生命ネットワーク構築のための遺伝子ネットワーク推定とシミュレーション

●宮野 悟<sup>1)</sup> ◆松野 浩嗣<sup>2)</sup> ◆井元 清哉<sup>1)</sup> ◆長崎 正朗<sup>1)</sup> ◇坂内 英夫<sup>3)</sup>

1) 東京大学医科学研究所 2) 山口大学大学院理工学研究科 3) 九州大学大学院システム情報科学研究院

## <研究の目的と進め方>

計測技術の発明・改良により、遺伝子情報に関して、DNA 配列や mRNA 発現データをはじめとして、蛋白質間の相互作用データやメタボローム解析データなど多種多様な情報がゲノムワイドかつハイスループットに生産されるようになってきた。また、細胞内骨格、細胞内輸送系、さらには多細胞系における遺伝子ネットワークについての知識とデータも着実に蓄積されている。これらのいわゆるオーミックデータ及び細胞構造と機能についての情報は、生命をシステムとして理解していくための重要なインフラ情報である。このインフラ情報を生命の理解へと繋げるための情報技術を研究・開発することは、バイオインフォマティクス研究の重要な方向であり、このために本研究では、次の2つの目標を立て、研究を展開する。

(1) 遺伝子ネットワークの推定技術の開発及び遺伝子ネットワーク情報の整理

(2) 生命ネットワークのシミュレーション技術の開発

本研究で追求する2つの目的は、本特定領域内の他の研究によって得られるデータ及び知識と有機的に結合するもので、互いに強いシナジーを産みだすことを可能にするものである。したがって、生命システムの解明のために、新たな研究方法を作り出すことが期待される(図1参照)。

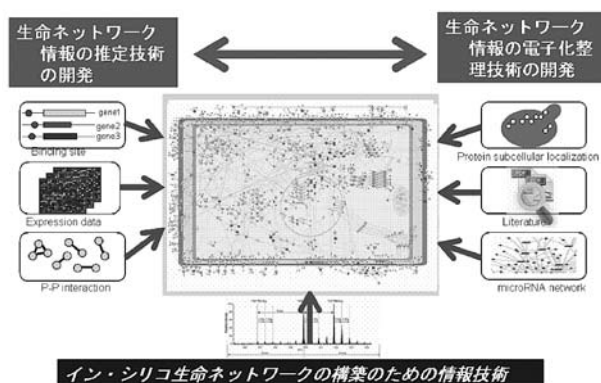


図1：研究方法の概略

## <研究開始時の研究計画>

(1) の目標に対する計画：遺伝子破壊株及びショック等に基づくマイクロアレイデータを基本にして、蛋白質の相互作用情報、局在情報、蛋白質・DNA 相互作用情報、文献情報などを用いて、ネットワークの推定法を開発を行う。特に、これまでの研究で開発してきたベイジアンネットワークと非線形回帰を組み合わせた遺伝子ネットワークの推定法を基にして、複数ソースの情報を同時に活用する方式を開発する。また、生物知識等を制約条件として利用することにより、計算量的な困難を解消し、最適なネットワークを推定する方式を開発する。そして、最終的には、生物的事実を最も反映できる遺伝子ネットワーク推定法を追求する。

(2) の目標に対する計画：これまで開発してきたハイブリッドペトリネットを拡張したアーキテクチャを見直すことにより、細胞内構造や確率的な現象及び多細胞系にも対応できる生命ネットワークのモデル化とシミュレーションのための技術を開発する。また、生命ネットワークモデルを様々なソフトウェアで共有できるように、相互作用情報、動的構造、アノテーション情報などを含めた XML による表現形式を改良・開発する。そして、この形式を用いて、(1) の遺伝子ネットワーク情報の整理を行う。

これらの目標を達成することにより、イン・シリコで生命ネットワークを構築する情報技術を創出する。

## <研究期間の成果>

### (I) 生命ネットワーク情報の推定技術に関する研究成果

#### (a) ベイジアンネットワークに基づいた方法

マイクロアレイデータとタンパク質間相互作用データから、遺伝子制御ネットワークとタンパク質ネットワークを同時に推定するための方式を開発した(0601301544)。遺伝子制御ネットワークのモデリングにはこれまで用いていた Bayesian ネットワークを使い、タンパク質ネットワークのモデリングには Markov ネットワークを適用し、これらを統合した統計モデルを定義し、それぞれのデータの信頼性に基づいて相互に補いあうことができるようにしたものである。ネットワークを探索するアルゴリズムを実装し、パン酵母の細胞周期に関するマイクロアレイデータとタンパク質相互作用データを用いて、遺伝子制御ネットワークとタンパク質ネットワークの同時推定計算機実験を行った結果、既知のものに関しては信頼性の高いネットワーク情報を推定できていることを確認した。また、この方式は、Y2H によって得られたタンパク質相互作用データにおけるフォールスポジティブ情報の検出にも応用できるものである。

Bayesian ネットワークに基づくマイクロアレイデータからの遺伝子ネットワーク推定において、生物種間の進化情報を用いる方法を開発した(0601301551)。この方式を実装し、細胞周期に関する遺伝子ネットワークに対して、ヒト及びパン酵母の細胞周期に関係している遺伝子のタンパク質アミノ酸配列のホモロジー情報を進化情報として用いて、ヒト及びパン酵母の細胞周期マイクロアレイ時系列データからヒトのネットワークとパン酵母のネットワークを同時にかつ相互に推定する計算機実験を行った。

その結果、ヒトのネットワーク推定は精度がよくなるがパン酵母のネットワーク推定にはそれほど貢献しないことを確認した。

最適ベイジアンネットワーク探索によるネットワークの構造モチーフの解析を行った(0601301557)。マイクロアレイデータから最適ベイジアンネットワークを推定するアルゴリズムをこれまで開発してきたが、数学的に最適なものでいくつかの間違ったエッジを含んでいることが判明している。そこで、最適なネットワークから準最適なものまで枚挙するアルゴリズムにより、多数のネットワークを計算し、それに基づいてネットワークモチーフという概念を定義した。計算機実験によると、最適なベイジアン

ネットワークよりも、本研究で定義したネットワークモチーフとよばれるネットワークがより生物情報を正確に反映していることがわかった。

マイクロアレイデータと生命知識ベースを利用した遺伝子ネットワークの推定法を開発した(0601301555)。非常に多くの知識が生命知識ベースには収録されてきているが、様々なエラーが含まれているため、そのためこうした情報を鵜呑みにして利用することは危険である。そこで、我々は、マイクロアレイデータからの遺伝子ネットワーク推定において、知識ベースの修復とダイナミックベイジアンネットワーク推定を合わせて行う確率的フレームワークを構築した。パン酵母の細胞周期マイクロアレイデータ及びKEGGの対応する知識ベースを用いて、この方式の有効性を検証した。

ベイジアンネットワークの最適解を探索する問題はNPハードな問題であることが知られており、そのためいくつかのヒューリスティックアルゴリズムが提案されてきたが、K2アルゴリズムとよばれるベイジアンネットワークの探索アルゴリズムを拡張することにより、より効率よく探索できるアルゴリズムを開発した(0801272026)。モンテカルロシミュレーションにより、他のアルゴリズムと比較して、このアルゴリズムの有効性を、実データを使って示した。順序制約を用いたベイジアンネットワークの構造学習アルゴリズムは多々あるが、これまでに開発した順序間に距離を定めて効率的な学習を行うKNアルゴリズムを拡張し、半順序制約を用いたベイジアンネットワークの構造学習アルゴリズムeKNアルゴリズムを開発した(0901161749)。

#### (b) 状態空間モデル及びARモデルに基づいた方法

状態空間モデルを用いて時系列で得られるマイクロアレイ遺伝子発現データから遺伝子モジュールネットワークの推定法を開発した(0702121642)。時系列マイクロアレイデータの特徴は時点数が極端に少なく(数個~数十個)、次元数(遺伝子数に対応)が大きい(数千~数万)ことである。この特徴から multivariate autoregressive モデルのような通常の統計モデルを用いるとオーバーフィットした結果をもたらすことになり、都合がわるい。状態空間モデルに次元縮小法を用いた方式を開発することによりこの困難を解決できることを理論的にまた、パン酵母の細胞周期に関する時系列データを用いて示した。次に、時系列遺伝子発現データを分類するための状態空間モデルの混合モデルを用いた新たな方式を開発した(0801272018)。この方式は、(1) 遺伝子発現の時間パターンのクラスタリング、(2) クラスタを分類する遺伝子の自動検出、(3) 各クラスターに対して制約付き自己回帰係数の推定、を行うもので、53人の動脈硬化症の患者へのインターフェロン $\beta$ を用いた組合せ治療における長期的な遺伝子発現データに適応してその有効性を示すことができた。

時系列マイクロアレイデータから大規模な遺伝子ネットワークを推定する際の特有の困難さは、典型的な時系列マイクロアレイデータの遺伝子が数万に達する一方で観測時点数は僅かに10程度であるという次元と標本数の不均衡に起因している。このようなデータに対し、多変量時系列データを解析するための標準的な手法であるベクトル自己回帰モデルを仮定すると、パラメータ数が観測時点数よりも大きくなるため解の不定性の問題が起こる。この問題を、次元圧縮を考慮した状態空間モデルから次元制約付きの自己回帰過程を導出することで克服し大規模な遺伝子ネットワーク推定を実現した。状態空間モデルによる次元圧縮は、転写因子が多くの遺伝子群のタンパク質を制御するという知見と適合しており、このことはさらに、転写因子が制御する遺伝子群の抽出および転写因子タンパク質間ネットワークの同時推定を可能にした。提案手法を細胞周期データとヒト血管内皮細胞データに応

用し大規模かつ高精度の遺伝子ネットワークを推定することに成功した(805171951)。

時系列遺伝子発現データから遺伝子ネットワークを精度よく推定するためには、リプリケートデータが有用である。しかし、これまでの推定方法では、生物学的リプリケートと技術的リプリケートは、異なった特徴をもっているが、これらをきちんと区別して扱っていない。そこで、状態空間モデルに基づいた統計的モデルを構築し、生物学的/技術的リプリケートをうまく使えるようにし、それに基づいた遺伝子ネットワーク推定のアルゴリズムを提案した。ヒトT細胞遺伝子発現データを用いた解析を通して、この方法の有効性を示した(806231520)。

時系列遺伝子発現データから Granger 因果律を概念とするベクトル自己回帰モデルによる遺伝子制御ネットワーク推定を基底関数展開によるノンパラメトリック回帰の枠組みで非線形構造を捉えると、基底関数の数に比例してパラメータが増加するため、ネットワーク推定に困難が生じる。そこで、L1正則化ノンパラメトリックベクトル自己回帰モデルによる遺伝子ネットワーク推定手法を開発した。L1正則化法では真のネットワーク構造が疎であるとき、パラメータ数がデータサンプル数を超過した場合でも、パラメータ推定が可能となる。これにより、既存の手法では不可能なサイズの遺伝子制御ネットワークを推定することが可能となった。非線形構造を持つ人工ネットワークによるシミュレーション実験を通して、L1正則化を用いないモデルとの比較を行うことで、提案手法の有効性を示した。また、ヒトのHeLa細胞からの時系列マイクロアレイデータに提案手法を適用することで遺伝子ネットワークの推定を行った(0901161745)。

さらに recursive elastic net を用いた方法を考案し、時系列遺伝子発現データから擬陽性の少ない遺伝子ネットワークを推定する技術を開発した(0912040923)。

時系列マイクロアレイデータから遺伝子ネットワークを効率的に推定するために、ベクトル自己回帰モデルを状態空間モデルで表現し、ネットワーク構造を表すパラメータをL1正則化にもとづく sparse 学習により推定する手法を開発した(0912040905)。この方法はこれまで遺伝子ネットワークの推定に用いられてきたベクトル自己回帰モデルや状態空間モデルをもとにした手法の以下の問題点を克服するものである：(i) ベクトル自己回帰モデルでは時系列データが等間隔であるという仮定や、システムノイズと観測ノイズを区別することができないという問題点がある。(ii) また状態空間モデルでは推定を行うネットワークにおいてモジュール性が仮定されるため、実際のネットワークがモジュール性を仮定することができないような状況では推定精度が落ちるおそれがある。この方法を EGF 受容体刺激と抗がん剤である Gefitinib が投与されたヒトの肺の細胞からの時系列マイクロアレイデータに適用することで遺伝子ネットワークの推定を行い、通常の肺の細胞からのデータから推定されたネットワークとの比較を行った。Gefitinib は EFG の選択的な抑制因子として知られていることから、この比較により Gefitinib の off-target 遺伝子の同定が期待される。比較の結果、Gefitinib により動かされた遺伝子が検出され、推定されたネットワークにおいて Gefitinib により動かされた遺伝子の上流または下流における遺伝子が Gefitinib による副作用に関連しているのではないかと推測される。

また、状態空間モデルを *Aspergillus fumigatus* のヒートショックによる時系列遺伝子発現データを用いて、その転写ネットワーク解析に応用している(0912040901, 0912040857)。

#### (c) 生命システムのためのデータ同化技術の開発

地球科学の領域でモデル化技術として注目されているデータ同化の考えを生命システムのモデル化に適用する枠組みを構築した

(0702121658). これはシミュレーションモデルと観測データを融合する技術で、この研究では状態空間モデルによって表現されたHFPNeで表現されたシミュレーションモデルと時系列遺伝子発現マイクロアレイデータを想定した計算実験を行い、その有効性と限界、及び今後の可能性について議論した。そして、epidermal growth factor 受容体 (EGFR) シグナル伝達パスウェイに対して、データ同化技術を用いて、SILAC 法によるタンパク質の量的計測データをシミュレーションモデルに同化させ、新たな仮説を合理的に探索する方式を開発した (0702121646)。まず、Hybrid Functional Petri Net with extension (HFPNe)を用いて、文献からの知識とデータを整理することによりシミュレーション可能なEGFR モデルを作成した。次に、そのモデルと Blagoev et al. (Nat. Biotechnol.,22:1139.1145, 2004) によるタンパク質の時系列量的データを用いて、データ同化技術を用いて新たなモデルの構築に成功した。

#### (d) その他の方法及び解析

ヒトの遺伝子発現データとゲノムデータを用いて共発現している遺伝子のプロモーター領域に保存されている様々の転写因子結合部位 (TFBSs) について報告した (0901161835)。また、枯草菌及びその関連種について、そのターミネーターを予測する方法を開発した (0601301606)。統計的メタアナリシスにより、がんの転写ネットワークの解析を行った (0912040920)。

混合因子分析法による遺伝子発現パターン解析ソフトウェアツール ArrayCluster を開発した (0608061253)。

エクソアレイ GeneChip Human Exon 1.0 ST Array による計測データから、スプライスバリエントを同定する方法を開発した。この方法は、(1) 特定のスプライスバリエントを表すバイオマーカーを同定するための ANOVA 解析と (2) サンプルにおける臨床的不均一性に起因する非特定のスプライスバリエントを区別するためのメタアナリシスとから構成されている。この方式の有効性を大腸癌のエクソアレイデータを用いて示した (0801272034)。そして、これを実装した web サービス ExonMiner を開発した (0901161840)。ExonMiner はエクソアレイデータの解析に特化した統合解析サイトである。データの正規化や two-way ANOVA, meta analysis などの統計的解析モジュールをもち、また、疾患特異的な発現を示す遺伝子の発見や、エクソレベルの発現の様子を可視化するモジュールを備えている。

ブーリアンネットワークモデルにより時系列マイクロアレイデータから遺伝子ネットワーク推定する問題において、パン酵母の転写因子結合部位データを補足的に用いることにより、より精度の高い遺伝子ネットワークを推定する方式を開発した (0601301553)。

LC-MS/MS を SILAC 法と組み合わせることによりプロテオームの相対的量的変化を計測することができるようになったが、現在、使われている GUI アプリケーションでは大量のプロテオーム解析が困難な状態にある。そのため、様々の自動化により、非常に効率よくデータ解析を行うことができるソフトウェア AYUMS を開発した (0801272029)。

また、BONSAI という正負の文字列データからそれらを区別するパターンを発見するツールを開発してきたが、これに新たに “insignificant indexing symbol” というものを導入することでタンパク質の N ミリストイル化に必要なルールを発見することができた (0801272032)。

## (II) 生命ネットワーク情報の電子化整理技術の開発に関する研究成果

### (a) Cell System Markup Language (CSML 3.0) 及び Cell System Ontology (CSO 3.0) の開発

Hybrid Functional Petri net with extension に基づいて (i) モデルの静的な構造に関する記述、(ii) モデルの動的な構造 (ダイナミクス) に関する記述、(iii) モデルの描画に関する記述、(iv) モデルの生物学的な情報の記述、及び (v) シミュレーション実行結果を保存する部分の 5 つから構成するようにした (0601301604)。CSML に関する情報は <http://www.csml.org/> (0702121830) に公開している。これにより、CSML は SBML (Systems Biology Markup Language) 及び CellML を完全に包含したものとなり、SBML 及び CellML から CSML への変換プログラムを開発し、CSML のドキュメンテーションと共に上述の web サイトから公開している。システムダイナミクスのための Cell System Ontology (CSO) を開発した。このオントロジーは Web Ontology Language (OWL) を用いて実装されており、そのためパスウェイモデルについての知識についてセマンティックパリレーションや自動推論を可能としている。CSO は、(i) パスウェイを様々なレベルで扱える、(ii) hybrid functional Petri net with extension (HFPNe) を用いることによりパスウェイの量的及び質的な機能を捉えることができる、(iii) ダイナミクスが入った動的モデルの知識だけでなく、可視化やシミュレーションの情報を取り込むことができる、といった特徴をもっている。CSO が、パスウェイのモデル化、可視化、シミュレーションにおいて高いポテンシャルをもっていることを *C. elegans* における ASEL/ASER 遺伝子制御ネットワークを HFPNe モデルを用いて実証した (0801272023)。BioPAX を CSO に変換する方式を開発した (0801272021)。この変換により、分子のインタラクションのダイナミクスの解析が可能となるだけでなく、BioPAX の枠組みではできなかった解析をさらに行うことができるようになった。さらに、BIOBASE 社の Transpath を CSML 形式に変換する方式を構築した (0901161740)。これによりダイナミクスや可視化を含む生命システムの記述を統合的に扱えるようになった。動的パスウェイに対して、オントロジーを用いた論理推論などの知識処理の枠組みを構築した (0901161753)。

Petri Net の概念でのパスウェイのモデル化チュートリアル web site を整備した (0702121834)。また、パスウェイモデルライブラリーを整備した (0702121836)。

### (b) パスウェイレイアウトアルゴリズムの研究

エッジ交差、ノード交差、ノード間距離、Gene Ontology による細胞内局在情報の 4 つを考慮したバイオパスウェイの自動レイアウトアルゴリズムを開発した (0601301558)。このパスウェイのレイアウトを、(1) 初期レイアウトのための新たな force-directed アルゴリズム、(2) 同じ生物学的属性のノードをアラインメントするためのスコアを定義し、それにより理解しやすいレイアウトを作り出す方法、(3) 計算速度を落とすことなくノードの swapping を効率よく行う探索法、により改良した (0801272024)。大規模な血管内皮細胞のパスウェイモデルを例にとり、これまでに研究されてきたパスウェイのレイアウト結果及び人手によるレイアウト結果と比較した結果、人手によるものと同じぐらいの精度のものが得られた。大きなサイズのパスウェイを高速にレイアウトする sweep calculation 法というグリッドレイアウトアルゴリズムを開発した。TRANSPATH から CSML 形式に変換した 95 のパスウェイモデルを用いて、アルゴリズムの効率を計算機実験によっても確認した (806231447)。また、“betweenness” をグラフィックレイアウトに融合した実用的なアルゴリズム (BFL) を開発し、より高い betweenness をもつノードがグラフィックレイアウトにおいて最適な場所に配置できるようになった (0903120802)。また、パスウェイの構造解析のために、betweenness centrality を用いた最大ウエイトの連結成分を探索する効率のよい発見的アルゴリズムを開発した

(0912040926).

#### (c) 生命システムモデリングとその解析法の研究

「ある遺伝子 A が上昇したらある遺伝子 B が発現する」, 「あるタンパク質 A がある一定濃度である一定時間キープされれば, ある遺伝子 B が発現する」といったパスウェイの動的規則について, モデル検証の枠組みを, 量的な関係を含むパスウェイの枠組みで利用できるようにする方法を開発した (0912040907). これにより細胞運命決定など, 今までの枠組みではできない規則のチェックが可能になった.

遺伝子 p53, MDM2, p19ARF の制御関係についての情報とデータを文献に基づいて抽出し, ハイブリッド関数ペトリネットを用いてそれをモデル化した (0601301600). シミュレーションの結果から, p53, MDM2, p19ARF がコンプレックスをつくると, それは転写活性をもっていることが示唆された. CDK 依存リン酸化及び ARF 依存安定化パスウェイのモデル化とシミュレーションによる p53 の転写活性の解析を行った (0702121636). タンパク質 p53 は, CDK7 などのプロテインキナーゼによりリン酸化され, またタンパク質 ARF によって安定化され, p53 のリン酸化と安定化は, その転写活性を促進した, それらは同時におこることが知られている. Hybrid Functional Petri net (HFPN) のモデルを用いて, 論文からの知識とデータを整理することにより, これらの CDK 依存のリン酸化パスウェイ及び ARF 依存のパスウェイのモデルを構築し, それらを組み合わせたシミュレーションモデルを作成し, シミュレーションにより安定化についてのリン酸化の影響を再現した. また, リカレントニューラルネットワークを用いてハイブリッドペトリネットで記述されたシステムの瓦解部位の予測法を開発した (0601301602). シミュレーションによるサーカディアン制御ネットワークにおける新たな仮説の生成を行った (0606181600). *C. elegans* の gustatory neurons (ASEL と ASER) の細胞運命の決定には, Johnson 等によると *lsy-6* と *mir-273* microRNAs を含んだ double-negative feedback loop が関与していると考えられているが, それに関する知識とデータを整理してそのシミュレーションモデルを HFPN を用いて構築した (0702121633). 彼等の提示していた仮説は, そのシミュレーションで観察することができた. さらに, よく知られているミュータントについても, このシミュレーションモデルにより再現することができた. Petri net に基づいたシグナル伝達系のモデル化とその解析法を開発した (0702121654). シグナル伝達系における酵素活性プロセスを記述するために, "activation transduction component" という概念を新たに導入し, Petri net 理論における elementary T-invariant との関係付けをおこなった. この方法により, elementary T-invariant を探索するアルゴリズムにより, 酵素活性プロセスを探索する方法を構築した. シグナル伝達パスウェイにおける分子間相互作用やメカニズムを分類し, それぞれに対応する Petri Net 表現をシステムティックに構成する方法を与えた (0702121703). 次に, Petri Net のトランジションの遅延時間を決定するための方法を開発した. そして, アポトーシスパスウェイを例に用いて, その方式の有効性を示した. その他, アポトーシスパスウェイのモデル解析を行っている (0912040917).

外部からの光がサーカディアンリズムに与える影響を, マウスの 5 つの時計遺伝子からなるモデルをハイブリッド関数ペトリネットに構築し, Cell Illustrator を用いてシミュレーション実行することによって調べた. その結果, これら 5 つの時計遺伝子からなるフィードバックループ機構が光による時刻依存的な位相シフトを起こさせていることを見出した. さらに, 光同調は 24 時間に近い時間周期に対してしか働かないという生物学的事実を同じ結果をシミュレーションによって確認した. また応用として,

東京からニューヨークに航空機で移動した場合の時差ボケシミュレーションを実行し, 時差ボケ解消には 1 週間必要なことを示した. これはヒトの感覚と一致していることから, 分子レベルで時差ボケ解消が行われていることを示唆している (0901161757). その他, サーカディアンリズムのモデルの解析を行っている (0912040911).

リカレントニューラルネットワークによる細胞内反応経路の脆弱・頑健部位の推定法を開発した (0901161800).

また, シグナル伝達経路上の物質がシグナル開始から発現するまでのデータがある場合, これらのデータを用いてシグナル伝達経路上の全ての物質の発現タイミングを決定するアルゴリズムを開発した (0901161803).

#### (d) 大腸菌のイン・シリコパスウェイ構築の事例研究

森研究室と合同セミナーをテレビ会議を利用して密接に行い, Ecocyc や iAF1260 などのデータの利用可能性や処理方法について検討を行ったのちに, その情報を CSML に統合するソフトウェアを開発し, Cell Illustrator で取り扱い可能なパスウェイ構造ファイルを構築した. さらに, 森研究室から提供される, 代謝物と酵素量の時系列データに適合するダイナミクスモデルを Cell Illustrator を用いて作成した. これにより, 統合 CSML ファイルを基盤データとし, ダイナミクスモデル構築法を基盤手法として, Cell Illustrator を用いて, 大腸菌代謝パスウェイの大規模シミュレーションモデルの構築に展開する基盤ができた.

#### <国内外での成果の位置づけ>

状態空間モデルによる動的な遺伝子ネットワーク推定におけるいくつかの困難を解決し, 実用レベルの方式を開発することができた. Cell System Markup Language (CSML) 及び Cell System Ontology (CSO) が開発され, それに基づいた動的パスウェイのバリデーションや知識処理が可能になり, そのための情報技術の開発を進められるようになった. このように, 動的モデルの推定からシミュレーションまでを一貫して研究しているものは世界に類があまりなく, 生命システム研究の情報基盤技術が構築されたと考える.

本研究で開発した CSML 並びに CSO はパスウェイのモデリングとシミュレーションソフトウェア Cell Illustrator Online 4.0 に用いられ, 国内だけでなく, 米国の製薬会社などで商用利用されている. また, Biobase 社のオンライン解析ソフトウェア EXPLAIN のパスウェイ情報の外部出力形式として CSML が用いられている.

#### <達成できなかったこと, 予想外の困難, その理由>

時系列データをパスウェイモデルのパラメータチューニングに用いる生命体データ同化技術を開発したが, 本格的な研究には現在のスパコン (十数テラフロップス) では不十分であることが判明し, この研究を次世代スーパーコンピュータ開発プロジェクトへ移すこととなった.

CSO 及び CSML を利用した生体ネットワークの知識処理では, バイオインフォマティクスとしての先駆的な研究成果は出すことができたが, 現場での利用レベルで実現するためには研究費と時間が不足した. また, 大腸菌に関するパスウェイモデル作成は, 当初の計画より限定された部分のパスウェイモデルしかできなかった.

#### <今後の課題, 展望>

本研究で開発したプログラム群は, 別プロジェクトで開発した, 分子ネットワーク推定・可視化・シミュレーションなどのデー

タ解析の流れをグラフィカルに自在に組み立てることができるソフトウェア CSML Pipeline に組み込むことにより、データ解析から生命ネットワークシステム解析を一気通貫に行えるシステムの実現を目指している。また、化学物質などの環境因子情報に対応した CSML の開発も今後の課題として明確になった。

#### <研究期間の全成果公表リスト>

##### 1) 論文 / プロシーディング

1. 0912040857 Do, J.H., Miyano, S., Choi, D.K. Statistical approaches to genome-wide biological networks. *Biochip J.* 3(3):190-202 (2009).
2. 0912040901 Do, J.H., Yamaguchi, R., Satoru Miyano, S. Exploring temporal transcription regulation structure of *Aspergillus fumigatus* in heat shock by state space model. *BMC Genomics.* 10:306 (2009).
3. 0912040905 Kojima, K., Yamaguchi, R., Imoto, S., Yamauchi, M., Nagasaki, M., Yoshida, R., Shimamura, T., Ueno, K., Higuchi, T., Gotoh, N., Miyano, S. A state space representation of VAR models with sparse learning for dynamic gene networks. *Genome Informatics.* 22:56-68 (2009).
4. 0903120802 Hashimoto, T.B., Nagasaki, M., Kojima, K., Miyano, S. BFL: a node and edge betweenness based fast layout algorithm for large scale networks. *BMC Bioinformatics.* 10:19 (2009).
5. 0912040907 Li, C., Nagasaki, M., Ueno, K., Miyano, S. Simulation-based model checking approach to cell fate specification during *Caenorhabditis elegans* vulval development by hybrid functional Petri net with extension. *BMC Systems Biology.* 3:42 (2009).
6. 0912040911 Matsuno, H., Mito, N., Inouye, S.T., Simulation of mammalian circadian clock for the analysis of interlocked feedback loops in robust oscillation. *Proc. ITC-CSCC2009.* 264-267 (2009).
7. 0912040917 Miyakawa, C., Sugii, M., Matsuno, H., Miyano, S. Computational predictions for functional proteins working after cleaved in apoptotic pathway. *Proc. Second International Workshop on Intelligent Informatics in Biology and Medicine.* 807-812 (2009).
8. 0912040920 Niida, A., Imoto, S., Nagasaki, M., Yamaguchi, R., Miyano, S. A novel meta-analysis approach of cancer transcriptomes reveals prevailing transcriptional networks in cancer cells. *Genome Informatics.* 22:121-131 (2009).
9. 0912040923 Shimamura, T., Imoto, S., Yamaguchi, R., Fujita, A., Nagasaki, M., Miyano, S. Recursive regularization for inferring gene networks from time-course gene expression profiles. *BMC Systems Biology.* 3:41 (2009).
10. 0912040926 Yamamoto, T., Miyano, S., Nagasaki, M., Bannai, H. Better decomposition heuristics for the maximum-weight connected graph problem using betweenness centrality. *Lecture Notes in Artificial Intelligence.* 5808: 465-472 (2009).
11. 0901161835 Hatanaka, Y., Nagasaki, M., Yamaguchi, R., Obayashi, T., Numata, K., Fujita, A., Shimamura, T., Tamada, Y., Imoto, S., Kinoshita, K., Nakai, K., Miyano, S. A novel strategy to search conserved transcription factor binding sites among coexpressing genes in human. *Genome Informatics.* 20: 212-221 (2008).
12. 0901161803 Hioka, R., Miwa, Y., Li, C., Ge, Q.-W., Matsuno,

- H., Miyano, S. An algorithm to estimate delay times in Petri net models of signaling pathways with experimental data. *Proc. 3rd International Conference on Bio-Inspired Models of Network, Information, and Computing System.* CD-ROM, 8 pages (2008).
13. 806231520 Hirose, O., Imoto, S., Miyano, S., Analyzing time course gene expression data with biological and technical replicates to estimate gene networks by state space models, *Proc. 2nd Asia International Conference on Modelling & Simulation,* 940-946 (2008).
14. 0805171951 Hirose, O., Yoshida, R., Imoto, S., Yamaguchi, R., Higuchi, T., Charnock-Jones, D.S., Print, C., Miyano, S., Statistical inference of transcriptional module-based gene networks from time course gene expression profiles by using state space models, *Bioinformatics,* 24(7). 932-942 (2008).
15. 0901161753 Jeong, E., Nagasaki, M., Miyano, S. Rule-based reasoning for system dynamics in cell systems, *Genome Informatics,* 20, 25-36 (2008).
16. 0901161800 Kitakaze, H., Kanda, M., Nakatsuka, H., Ikeda, N., Matsuno, H., Miyano, S. Prediction of fragile points for robustness checking of cell systems. *IEICE TRANSACTIONS on Information and Systems D,* J91-D(9), 2404-2417 (2008).
17. 0901161745 Kojima, K., Fujita, A., Shimamura, T., Imoto, S., Miyano, S. Estimation of nonlinear gene regulatory networks via L1 regularized NVAR from time series gene expression data. *Genome Informatics,* 20. 37-51 (2008).
18. 0806231447 Kojima, K., Nagasaki, M., Miyano, S. Fast grid layout algorithm for biological networks with sweep calculation, *Bioinformatics,* 24(12), 1433-1441 (2008).
19. 0901161757 Mito, N., Ikegami, Y., Matsuno, H., Miyano, S., Inouye, S.T. Simulation analysis for the effect of light-dark cycle on the entrainment in circadian rhythm. *Genome Informatics,* 21, 212-223 (2008).
20. 0901161740 Nagasaki, M., Saito, A., Chen, L., Jeong, E., Miyano, S. Systematic reconstruction of TRANSPATH data into Cell System Markup Language. *BMC Systems Biology,* 2, 53 (2008).
21. 0901161749 Numata, K., Imoto, S., Miyano, S. Partial order-based Bayesian network learning algorithm for estimating gene networks. *Proc. IEEE Bioinformatics and Biomedicine.* 357-360 (2008).
22. 0901161840 Numata, K., Yoshida, R., Nagasaki, M., Saito, S., Imoto, S., Miyano, S. ExonMiner: Web service for analysis of GeneChip exon array data. *BMC Bioinformatics.* 9, 494 (2008).
23. 0801272018 Hirose, O., Yoshida, R., Yamaguchi, R., Imoto, S., Higuchi, T., Miyano, S. Clustering with time course gene expression profiles and the mixture of state space models. *Genome Informatics.* 18, 258-266 (2007).
24. 0801272021 Jeong, E., Nagasaki, M., Miyano, S. Conversion from BioPAX to CSO for system dynamics and visualization of biological pathway. *Genome Informatics.* 18, 225-236 (2007).
25. 0801272023 Jeong, E., Nagasaki, M., Saito, A., Miyano, S. Cell System Ontology: Representation for modeling, visualizing, and simulating biological pathways. *In Silico Biology,* 7, 0055 (2007).

- 26.0801272024 Kojima, K., Nagasaki, M., Jeong, E., Kato, M., Miyano, S. An efficient grid layout algorithm for biological networks utilizing various biological attributes. *BMC Bioinformatics*. 8:76 (2007).
  - 27.0702121703 Li, C., Ge, Q.-W., Nakata, M., Matsuno, H., Miyano, S. Modeling and simulation of signal transductions in an apoptosis pathway by using timed Petri nets. *J. Biosciences*, 32, 113-125 (2007).
  - 28.0801272026 Numata, K., Imoto, S., Miyano, S. A structure learning algorithm for inference of gene networks from microarray gene expression data using Bayesian networks. *Proc. IEEE 7th International Symposium on Bioinformatics & Bioengineering*. 1280-1284 (2007).
  - 29.0801272029 Saito, A., Nagasaki, M., Oyama, M., Kozuka-Hata, H., Semba, K., Sugano, S., Yamamoto, T., Miyano, S. AYUMS: an algorithm for completely automatic quantitation based on LC-MS/MS proteome data and its application to the analysis of signal transduction. *BMC Bioinformatics*. 8:15 (2007).
  - 30.0801272032 Sugii, M., Okada, R., Matsuno, H., Miyano, S. Performance improvement in protein N-myristoyl classification by BONSAI with insignificant indexing symbol. *Genome Informatics*. 18:277-286 (2007).
  - 31.0702121642 Yamaguchi, R., Yoshida, T., Imoto, S., Higuchi, T., Miyano, S. Finding module-based gene networks with state-space models? Mining high-dimensional and short time-course gene expression data. *IEEE Signal Processing Magazine*. 24(1), 37-46 (2007).
  - 32.0801272034 Yoshida, R., Numata, K., Imoto, S., Nagasaki, M., Doi, A., Ueno, K., Miyano, S. Computational discovery of aberrant splice variations with genome-wide exon expression profiles. *Proc. IEEE 7th International Symposium on Bioinformatics & Bioengineering*. 715-722 (2007).
  - 33.0601301600 Doi, A., Nagasaki, M., Matsuno, H., Miyano, S. Simulation based validation of the p53 transcriptional activity with hybrid functional Petri net. *In Silico Biology* 6, 0001 (2006).
  - 34.0702121636 Doi, A., Nagasaki, M., Ueno, K., Matsuno, H., Miyano, S. A combined pathway to simulate CDK-dependent phosphorylation and ARF-dependent stabilization for p53 transcriptional activity. *Genome Informatics*, 17(1), 112-123 (2006).
  - 35.0601301555 Imoto, S., Higuchi, T., Goto, T., Miyano, S. Error tolerant model for incorporating biological knowledge with expression data in estimating gene networks. *Statistical Methodology*, 3(1), 1-16 (2006).
  - 36.0702121654 Li, C., Suzuki, S., Ge, Q.-W., Nakata, M., Matsuno, H., Miyano, S. Structural modeling and analysis of signaling pathways based on petri nets. *J. Bioinf. Comput. Biol.*, 4(5), 1119-1140 (2006).
  - 37.0606181600 Matsuno, H., Inouye, S.-T., Okitsu, Y., Fujii, Y., Miyano, S. A new regulatory interactions suggested by simulations for circadian genetic control mechanism in mammals. *J. Bioinformatics and Computational Biology*, 4(1), 139-154, (2006).
  - 38.0702121658 Nagasaki, M., Yamaguchi, R., Yoshida, R., Imoto, S., Doi, A., Tamada, Y., Matsuno, H., Miyano, S., Higuchi, T. Genomic data assimilation for estimating hybrid functional Petri net from time-course gene expression data. *Genome Informatics*, 17(1), 46-61 (2006).
  - 39.0702121633 Saito, A., Nagasaki, M., Doi, A., Ueno, K., Miyano, S. Cell fate simulation model of gustatory neurons with microRNAs double-negative feedback loops by hybrid functional Petri net with extension. *Genome Informatics*, 17(1), 100-111 (2006)
  - 40.702121646 Tasaki, S., Nagasaki, M., Oyama, M., Hata, H., Ueno, K., Yoshida, R., Higuchi, T., Sugano, S., Miyano, S. Modeling and estimation of dynamic EGFR pathway by data assimilation approach using time series proteomic data. *Genome Informatics*, 17(2), 226-228 (2006).
  - 41.0608061253 Yoshida, R., Higuchi, T., Imoto, S., Miyano, S. ArrayCluster: an analytic tool for clustering, data visualization and module finder on gene expression profiles. *Bioinformatics*, 22(12), 1538-1539 (2006).
  - 42.0601301606 De Hoon, Michiel J. L., Makita, Y., Nakai, K., Miyano, S. Prediction of transcriptional terminators in *Bacillus subtilis* and related species. *PLoS Computational Biology*, 1(3), e25 (2005).
  - 43.0601301553 Hirose, O., Nariai, N., Tamada, Y., Bannai, H., Imoto, S., Miyano, S. Estimating gene networks from expression data and binding location data via Boolean networks. *Lecture Notes in Computer Science*. 3482, 349-356 (2005).
  - 44.0601301558 Kato, M., Nagasaki, M., Doi, A., Miyano, S. Automatic drawing of biological networks using cross cost and subcomponent data. *Genome Informatics*, 16(2), 22-31 (2005).
  - 45.0601301602 Kitakaze, H., Matsuno, H., Ikeda, N., Miyano, S. Prediction of debacle points for robustness of biological pathways by using recurrent neural networks. *Genome Informatics*, 16(1), 192-202 (2005).
  - 46.0601301604 Nagasaki, M., Doi, A., Matsuno, H., Miyano, S. Petri net modeling of biological pathways. *Proc. Algebraic Biology 2005* (Universal Academy Press), 19-31 (2005).
  - 47.0601301544 Nariai, N., Tamada, Y., Imoto, S., Miyano, S. Estimating gene regulatory networks and protein-protein interactions of *Saccharomyces cerevisiae* from multiple genome-wide data. *Bioinformatics*, 21, ii206-ii212 (2005).
  - 48.0601301557 Ott, S., Hansen, A., Kim, S.-Y., and Miyano, S. Superiority of network motifs over optimal networks and an application to the revelation of gene network evolution. *Bioinformatics*, 21(2), 227-238 (2005).
  - 49.0601301551 Tamada, Y., Bannai, H., Imoto, S., Katayama, T., Kanehisa, M., Miyano, S. Utilizing evolutionary information and gene expression data for estimating gene networks with Bayesian network models. *J. Bioinformatics and Computational Biology*, 3(6), 1295-1313 (2005).
- 2) データベース／ソフトウェア
1. 0702121830 Cell System Markup Language (CSML)  
<http://www.csml.org/>
  2. 0702121834 Petri Net Pathways:  
<http://genome.ib.sci.yamaguchi-u.ac.jp/~pnp/>
  3. 0702121836 Pathway Database for Cell Illustrator  
<http://genome.ib.sci.yamaguchi-u.ac.jp/~gon/index.html>