

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¹⁾ ◆松野 浩嗣²⁾ ◆井元 清哉¹⁾ ◆長崎 正朗¹⁾ ◇坂内 英夫³⁾

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

<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)。

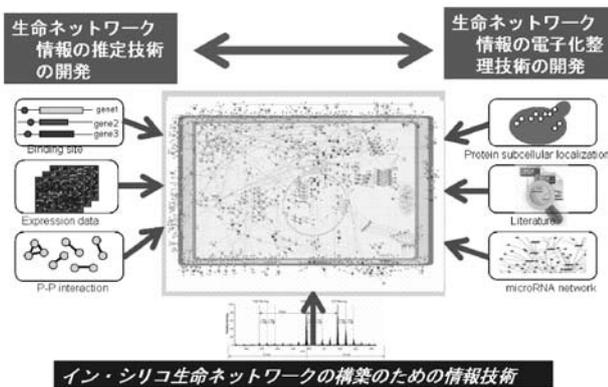


図1：研究方法の概略

<2008年度の研究の当初計画>

1. 生命システム情報の融合技術の研究：状態空間モデルにより推定されたネットワーク構造、およびダイナミクスの情報をシミュレーションモデルに変換する情報技術を開発する。具体的には、これまでの研究成果において導いた状態空間モデルの自己回帰モデル (AR(1)) 表現を利用し、推定された自己回帰係数を用いてシミュレーションモデルの反応速度を設定できることが基本的なアイデアとなっている。

2. 生命システム情報の電子化整理技術の高度知識処理への対応：平成19年度までに、CSML3.0及びCSO3.0という生体内の動的な現象を記述できるOWLを利用したオントロジー言語の開

発、BIOBASE社のTranspathのデータを動的なパスウェイモデルに変換する手法の開発、及び静的な生体内パスウェイ構造を記述できるオントロジー言語BioPAXで記述されたモデルを動的なモデルへと変換する技術の開発をした。この成果に基づいて、Transpath, Reactome, INOH, BioCycなどのデータベースを1つのオントロジーデータベースとして統合できるようにする。この統合化により、SPARQLやSWRLを用いることが可能になり、効率のよい知識検索と抽出法の実装を開始する。さらに、これに基づいて、同義、推移、サブクラスを用いた高度の知識推論技術の開発を進める。

3. 大腸菌のイン・シリコパスウェイ構築の事例研究：SBML2CSML, CSML parserなどのツール群が整備され、これを応用して大腸菌の大規模パスウェイモデルを作る。EcoCycデータベースの情報を基にして、大腸菌の代謝経路を上位、遺伝子制御ネットワークを下位とする2階層モデルをCSMLで作成する。Cell Illustrator (CI)上で上位層と下位層を区別して表示できるようにし、EcoCycのSBML形式の大腸菌のパスウェイ情報を、SBML2CSMLを用いて変換し、CIで表示しシミュレーション可能なものとして充実させる。

<2008年度の成果>

(a) 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(80517195)。

(b) 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)。

(c) 時系列遺伝子発現データからGranger因果律を概念とするべ

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)。

- (d) 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)。
- (e) 本研究で開発した Cell System Ontology に基づいて記述された動的パスウェイを妥当なものであるかを検証するための方式を、モデルが、(i) ペトリネットとして構造的に正しいか、(ii) 生物学的な意味を正しく捕えているか、(iii) 生物学的な動的挙動を正しく反映しているか、という3つの観点から開発した。これにより、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)。
- (f) 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。(0901161740)で開発した技術を用いて 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)。
- (g) 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で構築し、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と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)。
- (h) 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)。
- (i) 班員間の共同研究として、森浩禎研究室 (奈良先端) と、大規模大腸菌パスウェイモデルの構築を目指し、森研究室から生産されている大腸菌の中央代謝経路の酵素発現の過渡的タイムコースデータのデータの時間変化に合うハイブリッド関数ペトリネットモデルを作成した。このモデル化により、実験方法に依存するデータのゆれが見出され、実験方法の改善が検討された。また、大腸菌のパスウェイデータを EcoCyc と iAF1260 の2つのデータベースから得て、CSML ファイルとして Cell Illustrator に統合するための手法について継続して

開発した。

<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 及び Cell System Ontology 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。

<達成できなかったこと、予想外の困難、その理由>

大腸菌のイン・シリコパスウェイ構築は来年度に完成する。

<今後の課題>

1. CSML3.0 と CSO3.0 の拡張及びそれに基づいた生命システムの知識処理技術のさらなる開発。
2. CSML/CSO で記述されたパスウェイにモデルチェックの考えを導入し、生命システムモデルのデバッグを可能とする研究。

<成果公表リスト>

1) 論文/プロシーディング (査読付きのものに限る)

1. 805171951 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)
2. 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)
3. 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)
4. 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)
5. 0901161753 Jeong, E., Nagasaki, M., Miyano, S. Rule-based reasoning for system dynamics in cell systems, *Genome Informatics*, 20, 25-36 (2008)
6. 806231447 Kojima, K., Nagasaki, M., Miyano, S., Fast grid layout algorithm for biological networks with sweep calculation, *Bioinformatics*, 24(12), 1433-1441 (2008)
7. 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)
8. 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)
9. 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)