

# 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

## <2007年度の研究の当初計画>

目標(1)の研究計画としては、状態空間モデル、ベイジアンネットワークモデル、およびハイブリッドペトリネットモデルを用いて時系列データから生命ネットワークを推定・精緻化する方式を改良する。また、最適及び準最適なネットワークを推定する方式の改良を行う。

目標(2)の研究計画としては、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CSMLを改良・発展させるとともに、パスウェイ情報を記述するためのほかの形式(XML形式など)から、CSMLへの変換法の研究を行う。パスウェイのレイアウトアルゴリズムの改良を行う。また、いく

つかのパスウェイモデルを具体的に構築することを通して、目標(2)の研究へフィードバックを図る。

## <2007年度の成果>

(a) 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。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(0801272018)。

(b) 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。この論文では、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。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(0801272026)。

(c) 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デーショニングや自動推論を可能としている。CSOは、(1) パスウェイを様々なレベルで扱える、(2) 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、(3) 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)。

(d) システムダイナミクスと可視化を含む形でパスウェイの知識を記述するためのフレームワークCell System Ontology (CSO)を開発してきた。この研究ではBioPAXをCSOに変換する方式を開発した。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。具体例によってその方式および結果を説明している(0801272021)。

(e) 生命システムシミュレーションのためのモデル記述XML Cell System Markup Language Ver.3 (CSML 3.0)及びオントロジーCell System Ontology 3.0 (CSO 3.0)を完成させ(0702121830)、BIOBASE社のTranspathをCSML形式に変換する方式を構築した。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。

- (f) パスウェイのレイアウトを行う新しいグリッドレイアウトアルゴリズムを開発した。これは、(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、により実現されている。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(0801272024)。
- (g) 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、我々は、様々な自動的に、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)。
- (h) エクソンアレイ 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ンアレイデータを用いて示した(0801272034)。
- (i) 前年度に、シグナル伝達経路の離散ベトリネットモデルを用いてシミュレーションを実行するため、必要な反応時間を半自動的に決定するアルゴリズムを開発した。今年度はシグナル伝達経路上の物質について濃度変化などのデータがある場合、それも取り入れて反応時間を決定するようにアルゴリズムの拡張を行なった。マウスの概日リズムでは、Per 遺伝子と Bmal 遺伝子の産物の振動は逆位相になることが知られている。この逆位相をもたらす遺伝子は Rev-Erb と思われているが(N. Preitner et al., Cell, 110:251-260, 2002 に記載) Cell Illustrator を用いたシミュレーションによって解析した結果、Rev-Erb 遺伝子では逆位相関係にすることは不可能であり、Ror 遺伝子では可能であることが分かった。
- (j) 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 ミリスチル化に必要なルールを発見することができた(0801272032)。

#### <国内外での成果の位置づけ>

これまでに開発してきた CSML3.0 は、“Bench to Computer”の精神のもとで、SBML 2.0 及び CellML 1.0 を包含するものとなり、2007年度は、商用データベース BIOBASE 社の合意のもと、パスウェイデータベース Transpath の CSML 化に成功し、シミュレーションのための生命システム記述の de facto standard 化がいつそう進んだ。マイクロアレイデータからの遺伝子ネットワーク推定技術開発においては、世界最高水準のものを継続的に開発してきた。

#### <達成できなかったこと、予想外の困難、その理由>

データ同化技術については、通常のスパコンでの研究開発が困難であることが次第に判明し、この研究は、本特定研究の範疇を越えることになるため、データ取得も含め別プロジェクトで今後展開することになった。

#### <今後の課題>

1. CSML3.0 及び CSO3.0 を拡張し、生命システム全体を取り扱うことができるようにするための研究及びその応用について
2. CSML/CSO で記述されたパスウェイにモデルチェックングの考えを導入し、生命システムモデルのデバッグを可能とする研究
3. 時系列データから推定したモデルにより適切に動的構造を導入するための研究

#### <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. 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).
  2. 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).
  3. 0801272023 Jeong, E., Nagasai, M., Saito, A., Miyano, S. Cell System Ontology: Representation for modeling, visualizing, and simulating biological pathways. *In Silico Biology* 7:0055 (2007).
  4. 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).
  5. 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).
  6. 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).
  7. 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).
  8. 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)
- 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>