

平成 22 年 6 月 10 日現在

研究種目：基盤研究（A）
 研究期間：2007～2009
 課題番号：19201039
 研究課題名（和文） 細胞内ネットワークモデルと分子計測データとの整合性評価法の開発
 研究課題名（英文） Development of method for evaluating the consistency between network model in the cell and the molecular measured data
 研究代表者
 堀本 勝久（HORIMOTO KATSUHISA）
 独立行政法人産業技術総合研究所・生命情報工学研究センター・研究チーム長
 研究者番号：40238803

研究成果の概要（和文）：既知ネットワーク構造についてデータがどの程度整合性を示すかを推定する2つの手法、統計アプローチと記号-数値計算アプローチ、の手法を開発した。これらアプローチにより、特定の条件下で計測されたデータと複数のネットワーク構造との整合性を推定することで、有意な整合性を示すネットワーク構造、すなわち条件に応じた活性化ネットワーク群の推定により、ネットワーク構造変化が追跡できる。

研究成果の概要（英文）：We have developed two methods for evaluating the consistency between the given network structures and the measured data of constituent molecules, based on the statistical and symbolic-numeric approaches. By using these methods, the active networks can be detected under a particular condition, and furthermore, the changes of network structures can be traced, depending on the environmental changes.

交付決定額

（金額単位：円）

	直接経費	間接経費	合計
2007年度	10,400,000	3,120,000	13,520,000
2008年度	11,400,000	3,420,000	14,820,000
2009年度	12,600,000	3,780,000	16,380,000
年度			
年度			
総計	34,400,000	10,320,000	44,720,000

研究分野：複合新領域

科研費の分科・細目：ゲノム科学・基礎ゲノム科学

キーワード：システム生物学、ネットワーク解析、グラフィカル・モデル、記号計算

1. 研究開始当初の背景

（1）過去に発見・検証された生体分子間の相互作用に関する実験結果が膨大な文献に蓄積され、これらの情報に基づいてネットワークモデルを構築する様々な方法が報告されている(Nikitin, A *et al. Bioinformatics*, **19**, 2155-2157, 2003)。また数社の企業が文献情報データベースとそれに基づくネットワ

ーク構築を行うソフトウェアを販売している。実際、文献情報に基づくネットワークモデルは、それぞれの分子の関係性が実験結果に裏付けられているため信用度が高く、ゲノムワイドな分子間の関係性に関する網羅的なネットワーク構築に広く利用されている(Calvano, SE *et al. Nature*, **437**, 1032-1037, 2005)。

(2) 一方、細胞内において分子間の関係性が環境にตอบสนองして時空間で変化するものであることは周知の事実である。文献情報に記載されているそれぞれの関係性は、ある特定の環境下で関係性があることは実験的な裏付けが存在するが、すべての関係性を統合したネットワークモデル全体が成り立つとは限らない。このような状況を鑑みるに、文献情報に基づくネットワークモデルと計測データの整合性を慎重に評価する必要があると考える。

(3) ネットワークモデルと計測データとの整合性を見積もりは、統計学においては観察データによる因果性の仮説検定問題として知られている。既に1934年にWrightによって考えられたパス解析(Wright, *S Ann.Math. Statist.* 5, 161-215, 1934)があるが、不幸にもその後永く研究がなされなかった。1960年代頃から社会科学者を中心にパス解析が研究され、構造方程式モデル(SEM: Structural Equation Model) (Joreskog, *KG Biometrika*, 57, 239-251, 1970)として発展し、現在ではコンピュータ性能の向上に伴い市販の統計ソフトウェアにも組み込まれるほど普及している。ただし、生体ネットワークに適用される潜在変数を仮定しないSEMは、尤度の計算の際にノイズを含むデータや類似なパターンを含むデータに対して脆弱性を示すことが多い。SEMのデータに関する脆弱性を克服するために、1980年代後半に主に人工知能研究者によって開発された因果推定法であるグラフィカルモデルの中心的概念である、d-separationの概念(Pearl, *J Probabilistic Reasoning in Intelligent Systems*. Kaufmann, 1988)を利用して、d-sep testが開発された(Shipley, *B Structural Equation Modeling* 7, 206-218, 2000)。この方法は偏相関係数による条件付独立関係に基づく方法でSEMの脆弱性を克服しているが、グラフにおいて辺の無いことを検定する方法である。この方法の生体ネットワークへの適用は数例知られるだけで(Bisits AM, *etal. PLoS Comput. Biol.* 1, 132-136, 2005)、積極的な適用が行われていない。これはグラフにおいて辺の無いこと、すなわちネットワークにおいて関係性の無いことを検定するアプローチであることが原因と考えられる。

(4) 以上のように、数理科学において、グラフの辺が存在することを評価する、すなわち変数間の関係性を積極的に評価することで、因果グラフと計測データとの整合性を見積もる方法は開発されていない。したがって、過去に蓄積された膨大な生物学的知見の宝庫である文献情報と近年新規開発された実験技術によって得られた網羅的な計測データとを、ネットワークモデルの評価に関して

融合する技術が無いのが現状である。

2. 研究の目的

(1) 本研究では、環境により変化するネットワーク構造を解明するため、膨大な実験解析に関する文献情報に基づき構築されたネットワークモデルとモデルを構成する分子に関する計測データとの整合性を数量的に評価する技術を開発する。詳細は、以下の通りである。

- ・新規な整合性評価法を開発する。
- ・新規開発の方法に相補的な整合性評価のために、既存の評価法を実装する。
- ・実験研究との融合を促進するために、解析用ウェブサーバを構築する。

(2) まず新規に、ガウシアン・ネットワークと一般化極値分布の統計学に基づく方法とコンパートメントモデルの計算機代数に基づく方法を開発する方法とする。統計学に基づく方法の新規な点は、従来ガウシアン・ネットワークに基づく数量の大小関係で整合性を表現している点を、一般化極値分布を利用することで有意確率として評価できることである。さらに本研究では、動態解析において用いられるコンパートメントモデルに関して、計算機代数によるモデルと計測データとの一致を見積もる Identifiability Algorithm (Buchberger, *B Aequationes Mathematicae*, 4, 45-50, 1988)に基づいた評価法を開発する。評価するネットワークの構造特性から、前者は遺伝子制御ネットワーク、後者はタンパク質相互作用ネットワークの評価にそれぞれ適用する。

(3) 次に、様々な生物学的な問題意識の基で計測されたデータと文献情報に基づくネットワークモデルとの適合度を評価するウェブサーバを構築する。ウェブサーバでは、入力・出力の形式に関して簡便な方法を工夫し、コンピュータ操作に不慣れた研究者にもストレス無く利用できるようにする。

3. 研究の方法

(1) 新規方法として、ガウシアン・ネットワークと一般化極値分布の統計学に基づく方法とコンパートメントモデルの計算機代数に基づく方法を開発する。

(2) 統計学に基づく方法は、所与のネットワークモデルをガウシアン・ネットワーク(Whittaker, *J Graphical Models in Applied Multivariate Statistics*. John Wiley and Sons, 1990)に基づき計測データから対数尤度として数量化する部分、ランダムに生成されたネットワークに關数する一般化極値分布(Coles, *S An Introduction to Statistical Modeling of Extreme Values*. Springer-Verlag, 2001)に基づいて、所与のネットワークの対数尤度を確率として評価する部分、の

2つの部分から成る。

まず、ネットワークと同等な同時確率密度関数を、因数分解定理に基づき条件付き確率密度関数の積に分解する。次に、分解されたそれぞれの条件付き確率密度関数を計測データ間の線形式で表し、その線形式を最小二乗法によって求める。最後に、線形式で表現された条件付き確率密度関数の積の対数をとることで、ネットワークと同等な同時確率密度関数に対応する対数尤度を求める。このようにして、ガウシアン・ネットワークに基づいて、所与のネットワークはその構造に応じて計測データから対数尤度として表される。

次に所与のネットワークに対する対数尤度を確率で評価する。まず、所与ネットワークと辺と節が同数で辺の結合具合が異なるネットワークを n 個生成する。上記ガウシアン・ネットワークに基づき生成された n 個の対数尤度を求める。 n 個の対数尤度の最大値を求める。上記のステップを m 回繰り返す。こうして生成された m 個の対数尤度が一般化極値分布に従うかを調べる。 m 個の対数尤度が一般化極値分布に従う場合、その分布に基づき所与のネットワークの対数尤度の出現確率を求める。上記の手続きにおいて、適切なパラメータ n と m の設定によって m 個の対数尤度が一般化極値分布に従うことが期待される。このように統計学に基づく方法の新規な点は、ガウシアン・ネットワークに基づいて算出される対数尤度は従来大小関係で整合性を表現しているが、一般化極値分布を利用することで出現確率として評価できることである。

(3) さらに本研究では上記の欠点を補うために、計算機代数に基づいた評価法を開発する。動態解析において用いられるコンパートメントモデルで、計算機代数によるモデルと計測データとの一致を見積もる

Identifiability Algorithm (Buchberger, *B Aequationes Mathematicae*, 4, 45-50, 1988) を適用する。すなわち、

- ・所与のネットワークをコンパートメントモデルに基づき微分方程式系に定式化する。

微分方程式系をラプラス変換し、代数方程式系として再度定式化する (Identifiability Algorithm の適用)。

- ・計測データについて指数多項式による近似式を求め、それら式をラプラス変換して代数式にする。

- ・所与のネットワークの代数方程式系と代数式で表された近似式が等しいとして、最小二乗法を適用する。

- ・最小二乗法によるパラメータ推定において代数算法による解の推定法を適用する。

- ・所与のネットワークについて求められたパラメータを用いて残差平方和を見積もる。

- ・上記手続きを、所与のネットワークと同数の辺と節をもち結合パターンの異なるように生成したネットワークについて行う。

- ・生成したネットワークの残差平方和の分布から、所与のネットワークの残差平方和の出現確率を見積もる。

以上の手続きにより、所与のネットワークと計測データとの整合性が残差平方和の出現確率の形で見積もることが可能になる。このアプローチの利点は、DAG を含む多様なネットワーク構造について柔軟に適用できることである。このため、多様なネットワーク構造が混在するタンパク質相互作用ネットワーク (代謝ネットワーク) に適用する。一方、この方法の欠点は、

- ・評価法が誤差を含まないデータを仮定しているため、揺らぎの大きな計測データでは計算不能な状況が起る。

- ・整数演算を行うため計算量が大きい。第一の欠点については、近年誤差を含むデータについて開発した計算機代数手法 (Anai, H et al. *J. Bioinfo. Comput. Biol.* in press) によって解消が可能である。また、第二の欠点については、評価対象ネットワークの分割と計算の並列化とによって困難を解消できると考えている。

(4) 2つの新規開発法について、整合性評価の性能とネットワーク構造特性との関連は発見論的に探索すべき課題である。様々な構造をもつネットワークに適用しそれぞれの構造特性を表す数量 (クラスター係数など) を算出することで、評価性能を構造特性の観点から把握する。

4. 研究成果

(1) 既知ネットワーク構造についてデータがどの程度整合性を示すかを推定する2つの手法、統計アプローチと記号 - 数値計算アプローチ、の手法を実応用に向け改良した。これらアプローチにより、特定の条件下で計測されたデータと複数のネットワーク構造との整合性を推定することで、有意な整合性を示すネットワーク構造、すなわち条件に応じた活性化ネットワーク群の推定により、ネットワーク構造変化が追跡できる。

(2) 統計アプローチは、参照する制御ネットワークに関して知データを用いて整備すると共に、ChIP-seq データ等の転写因子結合情報に関する実験データから参照制御ネットワークを導出するシステムを開発した。これにより、実験計測データの入力のみから活性化ネットワーク群の出力が可能になり、利便性を向上した。

(3) 記号 - 数値計算アプローチについては、微分方程式系をラプラス変換による代数方程式系へ変換するアプローチを採用で問題

である、ラプラス空間の極の存在のため数値計測データとの照応に際して極めて不安定であり、また定式化において仮定が必要という点に関し改良した。代数算法の一つである Differential Elimination を利用して、微分方程式系と同等の方程式系を導出し、この系を評価関数の束縛条件の一つとして採用する方法を考案した。

5. 主な発表論文等

〔雑誌論文〕(計 25 件)

1. Nakatsu, M., Horimoto, K., Lemaire, F., Urguplu, A., Sedoglavic, A. and Boulier, F.: Brute force meets Bruno force in parameter optimization: Introduction of novel constraints for parameter estimation by symbolic computation. *IET Sys. Biol.*, 査読有, in press.
2. Nakatsui, M., Horimoto, K., Okamoto, M., Tokumoto, Y. and Miyake, J.: Parameter Optimization by Using Differential Elimination: a General Approach for Introducing Constraints into Objective Function. *BMC Sys. Biol.* 査読有, in press.
3. Otsuji, H., Honda, M., Takamura, M., Usui, S., Okajima, M., D, Okada, H., Sakai, Y., Takamura, T., Horimoto, K. and Kaneko, S.: Altered Hepatic Gene Expression Profiles Associated with Myocardial Ischemia. *Circ Cardiovasc Genet*, 査読有, 3, 68 - 77, 2010.
4. Morioka, R., Arita, M., Sakamoto, K., Kawaguchi, S., Tei, H. and Horimoto, K.: Period-Phase Map: Two-dimensional Selection of Circadian Rhythm-Related Genes. *IET Sys. Biol.*, 査読有, 3, 487-495, 2009.
5. Tokumoto, Y., Horimoto, K. and Miyake, J.: TRAIL inhibited the cyclic AMP responsible element mediated gene expression. *Biochem. Biophys. Res. Commun.*, 査読有, 381, 533- 536, 2009.
6. Ura, S., Honda, M., Yamashita, T., Ueda, T., Takatori, H., Nishino, R., Sunagozaka, H., Sakai, Y., Horimoto, K. and Kaneko, S.: Differential miRNA expression between hepatitis B and hepatitis C leading disease progression to HCC. *Hepatology*, 査読有, 49, 1098-1112, 2009
7. Saito, S., Honda, M., Kaneko, S., and Horimoto, K.: Detection of network structure changes by graphical chain modeling: a case study of hepatitis C virus-related hepatocellular carcinoma. *Proceedings of 48th IEEE Conference on Decision and Control*, 査読有, pp. 5624-5630, 2009.
8. Saito, S. and Horimoto, K.: Co-Expressed Gene Assessment Based on the Path Consistency Algorithm: Operon Detention in *Escherichia coli*. *Proceedings of IEEE International Conference on Systems, Man and Cybernetics*, 査読有, pp. 4280-4286, 2009.
9. Nakatsui, M. and Horimoto, K.: Parameter Optimization in the network dynamics including unmeasured variables by the symbolic-numeric approach. *Proceedings of the Third International Symposium on Optimization and Systems Biology (OSB '09)*, 査読有, pp. 245 253, 2009.
10. Akutsu, T., Tamura, T. and Horimoto, K.: Complementing Networks Using Observed Data. *Lecture Notes in Artificial Intelligence 5809*, 査読有, pp. 126-140, 2009.
11. Yoshida, H., Horimoto, K. and Anai, H.: Inference of Probabilities over a Stochastic IL-System by Quantifier Elimination. *Math.Comput.Sci.*, 査読有, 1, 473 485, 2008.
12. Nishino, R., Honda, M., Yamashita, T., Takatori, H., Minato, H., Zen, Y., Sasaki, M., Takamura, H., Horimoto, K., Ohta, T., Nakanuma and Y., Kaneko, S.: Identification of novel candidate tumour marker genes for intrahepatic cholangiocarcinoma. *J. Hepatol.*, 査読有, 49, 207 216, 2008.
13. Hayashida, M., Sun, F., Aburatani, S., Horimoto, K. and Akutsu, T.: Integer Programming-based Approach to Allocation of Reporter Genes for Cell Array Analysis. *Int. J. Bioinformatics Research and Applications*, 査読有, 4, 385-399, 2008.
14. Saito, S., Aburatani, S. and Horimoto, K.: Network evaluation from the consistency of the graph structure with the measured data. *BMC Sys. Biol.* 査読有, 2, 84, 2008.
15. Tominaga, D., Tokumoto, Y., Nakatsui, M., Sun, F., Miyake, J. and Horimoto, K.: Analysis of network dynamics including hidden variables by symbolic-numeric approach. *Proceedings of the Second*

- International Symposium on Optimization and Systems Biology (OSB '08)*, 査読有, pp. 242-248, 2008.
16. Morioka, R., Arita, M., Sakamoto, K., Kawaguchi, S., Tei, H. and Horimoto, K.: Phase Shifts of Circadian Transcripts in Rat Suprachiasmatic Nucleus. *Proceedings of the Second International Symposium on Optimization and Systems Biology (OSB '08)*, 査読有, pp. 109-114, 2008.
 17. Nakatsui, M., Yoshida, H. and Horimoto, K.: An Algebraic-Numeric Algorithm for the Model Selection in Network Motifs in *Escherichia coli*. *Proceedings of the Second International Symposium on Optimization and Systems Biology (OSB '08)*, 査読有, pp. 257-264, 2008.
 18. Zhang, Z.Y., Horimoto, K. and Liu, Z.: Time Series Segmentation for Gene Regulatory Process with Time-Window-Extension Technique. *Proceedings of the Second International Symposium on Optimization and Systems Biology (OSB '08)*, 査読有, pp. 198-203, 2008.
 19. Wu, Z.K., Zhang, Z.Y., Zhang, L.W. and Horimoto, K.: Revealing Disease Related Interactions by Correlation Analysis. *Proceedings of the Second International Symposium on Optimization and Systems Biology (OSB '08)*, 査読有, pp. 341-349, 2008.
 20. Aburatani, S., Sun, F., Saito, S., Honda, M., Kaneko, S. and Horimoto, K., Gene systems network inferred from expression profiles in hepatocellular carcinogenesis by graphical Gaussian model. *EURASIP J. Bioinfo. Systems Biol.*, 査読有, 47214, 2007.
 21. Yoshida, H., Anai, H. and Horimoto, K., Derivation of rigorous conditions for high cell type diversity by algebraic approach. *BioSystems*, 査読有, 90, 486-495, 2007
 22. Sato, T., Yamanishi, Y., Horimoto, K., Kanehisa, M. and Toh, H., Inference of Protein-Protein Interactions by Using Co-evolutionary Information. In Anai, H., Horimoto, K. and Kutsia, T. (eds), *Algebraic Biology 2007, Lecture Notes in Computer Science 4545*, 査読有, p.322-333, Springer, Heidelberg, 2007.
 23. Yoshida, H., Nakagawa, K., Anai, H. and Horimoto, K. Exact parameter determination for Parkinson's disease diagnosis with PET using an algebraic approach. In Anai, H., Horimoto, K. and Kutsia, T. (eds), *Algebraic Biology 2007, Lecture Notes in Computer Science 4545*, 査読有 p.110-124, Springer, Heidelberg, 2007.
 24. Yoshida, H., Nakagawa, K., Anai, H. and Horimoto, K., An Algebraic-Numeric Algorithm for the Model Selection in Kinetic Networks. *Proceedings of 10th CASC, LNCS 4770*, 査読有, p.433-447, Springer, Heidelberg, 2007
 25. Hayashida, M., Sun, F., Aburatani, S., Horimoto, K. and Akutsu, T., Integer Programming-based Approach to Allocation of Reporter Genes for Cell Array Analysis. *Proceedings of 10th OSB, Lecture Notes in Operation Research*, 査読有, p.288-301, World Publishing Corporation, Beijing, 2007.
- [学会発表](計 11 件)
1. Saito, S., Honda, M., Kaneko, S., and Horimoto, K.: Detection of network structure changes by graphical chain modeling: a case study of hepatitis C virus-related hepatocellular carcinoma. *48th IEEE Conference on Decision and Control*, 2009/12/17, China.
 2. Saito, S. and Horimoto, K.: Co-Expressed Gene Assessment Based on the Path Consistency Algorithm: Operon Detention in *Escherichia coli*. *IEEE International Conference on Systems, Man and Cybernetics*, 2009/10/11, USA.
 3. Nakatsui, M. and Horimoto, K.: Parameter Optimization in the network dynamics including unmeasured variables by the symbolic-numeric approach. *The Third International Symposium on Optimization and Systems Biology (OSB '09)*, 2009/9/20, China.
 4. Tominaga, D., Tokumoto, Y., Nakatsui, M., Sun, F., Miyake, J. and Horimoto, K.: Analysis of network dynamics including hidden variables by symbolic-numeric approach. *The Second International Symposium on Optimization and Systems Biology (OSB '08)*, 2008/10/18, China.
 5. Morioka, R., Arita, M., Sakamoto, K., Kawaguchi, S., Tei, H. and Horimoto, K.

- K.: Phase Shifts of Circadian Transcripts in Rat Suprachiasmatic Nucleus. *The Second International Symposium on Optimization and Systems Biology (OSB '08)*, 2008/10/18, China.
6. Nakatsui, M., Yoshida, H. and Horimoto, K.: An Algebraic-Numeric Algorithm for the Model Selection in Network Motifs in Escherichia coli. *The Second International Symposium on Optimization and Systems Biology (OSB '08)*, 2008/10/18, China.
 7. Zhang, Z.Y., Horimoto, K. and Liu, Z.: Time Series Segmentation for Gene Regulatory Process with Time-Window-Extension Technique. *The Second International Symposium on Optimization and Systems Biology (OSB '08)*, 2008/10/18, China.
 8. Wu, Z.K., Zhang, Z.Y., Zhang, L.W. and Horimoto, K.: Revealing Disease Related Interactions by Correlation Analysis. *The Second International Symposium on Optimization and Systems Biology (OSB '08)*, 2008/10/18, China.
 9. Yoshida, H., Nakagawa, K., Anai, H. and Horimoto, K., An Algebraic-Numeric Algorithm for the Model Selection in Kinetic Networks. *10th CASC*, 2007/9/16, Germany
 10. Sato, T., Yamanishi, Y., Horimoto, K., Kanehisa, M. and Toh, H., Inference of Protein-Protein Interactions by Using Co-evolutionary Information, *Algebraic Biology 2007*, 2007/7/2, Austria
 11. Yoshida, H., Nakagawa, K., Anai, H. and Horimoto, K. Exact parameter determination for Parkinson's disease diagnosis with PET using an algebraic approach, *Algebraic Biology 2007*, 2007/7/2, Austria

[図書](計5件)

1. Saito, S. and Horimoto, K.: Network Screening: Detection of Active Regulatory Networks. In Emmert-Streib, F., Dehmer, M., Salvador, A. and Graber, A. (eds): *Applied statistics for Network Biology: Methods in Systems Biology*, Wiley, in press.
2. Aburatani, S., Saito, S., and Horimoto, K.: ASIAN: Automatic System for Inferring A Network. In Krawetz, S. (ed): *Bioinformatics for Systems Biology: Second Edition, Introduction to Informatics*, p.563-577, Humana

Press, New Jersey, 2009.

3. Anai, H. and Horimoto, K. (eds): *Symbolic Computation in Biology*, Math. Comput. Sci., 2, 399-556, 2009.
4. Horimoto, K., Regensburger, G., Rosenkranz, M. and Yoshida, H. (eds): *Algebraic Biology, Lecture Notes in Computer Science 5147*, Springer, Heidelberg, 2008.
5. Anai, H., Horimoto, K. and Kutsia, T. (eds) *Algebraic Biology, Lecture Notes in Computer Science 4545*, Springer, Heidelberg, 2007.

[その他]

・ネットワーク推定サーバ"ASIAN":
<http://eureka.cbrc.jp/asian/>
 ・研究会「生体モデリングの新展開 - 分子からネットワークへ」開催(産業技術総合研究所臨海別館 11F 第三会議室、2009年11月26日)発表者:岡本正宏(九大院農学研究科) Francois Lemaire、Alexandre Sedoglavic (Computer Science Lab., Lille Univ.) 阿久津達也(京大化学研究所) 今田友樹、永持仁(京大院情報学研究科) 堀本勝久(産総研 CBRC)

6. 研究組織

(1)研究代表者

堀本 勝久 (HORIMOTO KATSUHIISA)
 独立行政法人産業技術総合研究所・生命情報工学研究センター・研究チーム長
 研究者番号: 40238803

(2)研究分担者

油谷 幸代 (ABURATANI SACHIYO)
 独立行政法人産業技術総合研究所・生命情報工学研究センター・研究員
 研究者番号: 10361627
 富永 大介 (TOMINAGA DAISUKE)
 独立行政法人産業技術総合研究所・生命情報工学研究センター・研究員
 研究者番号: 20357578
 穴井 宏和 (ANAI HIROKAZU)
 株式会社富士通研究所・ITコア研究所・研究員
 研究者番号: 20417520
 吉田 寛 (YOSHIDA HIROSHI)
 九州大学・数理科学研究院・特任助教授
 研究者番号: 60401262
 (2007年度のみ)