

機関番号：20103

研究種目：若手研究（B）

研究期間：2009～2010

課題番号：21700327

研究課題名（和文） 二分グラフによる転写調節ネットワークの数理モデル化

研究課題名（英文） Mathematical modeling of transcriptional regulatory networks based on bipartite graphs.

研究代表者

ホセ ナチエル（JOSE NACHER）

公立はこだて未来大学・システム情報科学部・准教授

研究者番号：60452984

研究成果の概要（和文）：本研究では、ncRNA 分子とそれらに制御されるタンパク質との相互作用についてネットワーク分析する。全体的構造と局所的構造の分析し、特定された構造特性を再現する進化モデルを提案する。ネットワークの構造と頂点を持つ分子機能の分布の関係についての分析を行う。最適なアルゴリズムを用い、ネットワークが機能クラスに属するときの頂点の分布を特定した。この結果、ncRNA 分子が転写制御機能に関係している可能性が考えられる。

研究成果の概要（英文）：We construct the ncRNA-protein interaction network and analyze its large-scale organization as well as the local interaction structure. In addition, an evolutionary model for bipartite graphs that includes several biological mechanisms is presented. We also investigate the distribution of molecular functionalities in these networks by using optimization algorithms. Our findings show that both transcriptional and ncRNA-interactions exhibit characteristic couplings between network and functionality that deviate from random expectations, supporting growing evidences for a ncRNA regulatory capability.

交付決定額

（金額単位：円）

	直接経費	間接経費	合計
2009年度	700,000	210,000	910,000
2010年度	800,000	240,000	1,040,000
年度			
年度			
年度			
総計	1,500,000	450,000	1,950,000

研究分野：総合領域

科研費の分科・細目：情報学・生体生命情報学

キーワード：生体生命システム情報学, 複雑ネットワーク科学, 数理モデル, データ分析

1. 研究開始当初の背景

本研究の分析対象の ncRNA 分子とはタンパク質に翻訳されない RNA 分子のことを指す。この分子は、最近研究により真核細胞の転写制御機能において重要な役割を担う可能

性が示され、注目されている分子である。ncRNA が転写制御に係わる可能性は、高等生物になるに従い ncRNA 分子の割合が高くなるという研究結果からも示唆される。この研究では、ヒトにおける ncRNA 分子の割合は 98.5%以上となることが報告されている。

高等生物において ncRNA が多く転写されているというこの発見は、細胞機能に関する疑問と、現在考えられている遺伝情報の伝達経路が不完全である可能性を示す。現在の分子生物学では、遺伝情報の伝達は「遺伝子から mRNA に転写され、タンパク質に翻訳される」と考えられる。この考えセントラルドグマと呼ばれ分子生物学の絶対的な原理とされていた。したがって、ncRNA 分子が転写制御の重要な機能に係わっている場合、現在のセントラルドグマには無い、新しい遺伝情報の伝達経路が存在する可能性が示される。しかし、このような重要な可能性が示されているが、ncRNA 分子が関与するネットワークの局所的な構造分析はもちろん、全体的構造に対する研究もおこなわれていない。本研究ではこのような背景をもとに ncRNA 分子に係わる分子間相互作用データに対してネットワーク分析をおこなう。

2. 研究の目的

我々は ncRNA 相互作用を含む転写調節ネットワークの形成および進化のモデル構築に焦点を当てる。この種類のネットワークは、転写因子 TFs をコードするオペロンとその標的遺伝子の mRNA の二種類の頂点を組み合わせた二部グラフとなっている。我々は生物学的な複雑ネットワークの中で観察される、次数分布からモチーフのような局所的相互作用パターンまで、大域的、局所的な統計的特徴を再現する、新しい二部グラフによる成長の数理モデルを発展させることを目的とする。また、転写制御ネットワークと ncRNA 相互作用において、構造と頂点を持つ分子機能の関係についての分析を行う。分析の対象は有機体 *S. cerevisiae* の転写ネットワーク、および同有機体の ncRNA 相互作用ネットワークとする、対象のネットワークは公開されているデータベースより作成する。転写制御ネットワークは遺伝子発現の制御を行う転写因子 (TFs) と制御される標的遺伝子によって定義される転写制御のデータから作成した。ncRNA 相互作用ネットワークは、ncRNA とタンパク質の機能的相互作用データを利用している。本研究では、転写制御ネットワークと non-coding RNA が関与する相互作用ネットワークの構造と頂点を持つ分子機能の分布の関係についての分析を行う。分析には 2 つのパラメータと最適なアルゴリズムを用い、ネットワークが 2 つの異なる機能クラスに属するときの頂点の分布を特定した。この結果、転写制御ネットワークと ncRNA 相互作用ネットワークの双方において、ネットワーク構造と機能の関係がランダムな分布とは大きく異なることが示された。こ

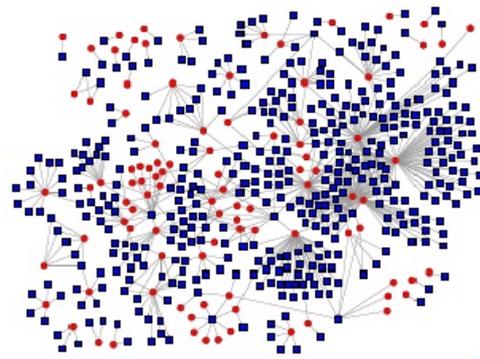


図 1. ncRNA-タンパク質ネットワークの例

のことから、ncRNA 分子が転写制御機能に関係している可能性が考えられる。

3. 研究の方法

(1) データベース

研究対象とする ncRNA 相互作用ネットワークは、NPInter database で公開されている ncRNA 分子とタンパク質分子の関係データを用いる。利用したデータは、6 種類の有機体 *Escherichia coli*, *Saccharomyces cerevisiae*, *Caenorhabditis elegans*, *Drosophila melanogaster*, *Mus musculus*, *Homo sapiens* のものである。データベースより得たこれらのデータから必要情報の取得、整理を行い、6 種類の全ての有機体データを含むネットワーク all RP と *Homo sapiens* のデータによるネットワーク Hs RP を作成する。また、ncRNA 分子と転写制御機能との関連を調べるための比較対象として、有機体 *Saccharomyces cerevisiae* と *Escherichia coli* K12 の転写制御データからネットワークを作成する。*S. cerevisiae* は U.Alon の公開研究サイト、*E. coli* K12 は RegulonDB からのデータを用いネットワークを作成する。

(2) ネットワーク解析

ネットワーク構造の分析は、累積次数分布とネットワークモチーフの特定によっておこなう。累積次数分布の特定からは、ncRNA 相互作用ネットワークがスケールフリー性を持つか否かという点と調べる。またネットワークモチーフの特定によって、ネットワークに頻出する局所的な構造特徴となる部分構造を特定する。

(3) 数理モデルとシミュレーション

本研究では ncRNA-タンパク質ネットワーク

と転写制御ネットワークの構造特性を再現する成長モデルとして、三つの成長過程によって構成されるモデルを提案した。数理解析とシミュレーション結果によると、このモデルではそれぞれの成長過程が起こる確率の違いによって、次数分布が異なるネットワークを生成することが分かる。具体的には、①二つの頂点の次数分布が共にベキ乗則に従う、②一つの頂点の次数分布はベキ乗則に従い、もう片方の頂点の次数分布は指数関数に従う、③両頂点の次数分布が共に指数関数に従う、という三つの場合があることが示された。実際の ncRNA 相互作用ネットワークや転写制御ネットワークの構造分析からは N 頂点と M 頂点で次数分布が異なる結果が確認された。これは、提案モデルの解析結果とも一致している。また、実際のネットワークで見つかった次数分布の特徴は、数理解析とシミュレーション解析の結果から継承複製過程が関与していることが分かる。この結果から、ncRNA 相互作用ネットワークの ncRNA 頂点の次数分布がベキ乗則に従うことは、タンパク質頂点において継承複製過程(遺伝子の複製)が多く起きていることが予想される。また、タンパク質頂点の次数分布が指数関数に従っていると考えると、ncRNA 頂点では継承複製過程はあまり起きず、次数複製過程や追加過程(遺伝子の突然変異)が多く起きていることが予想される。

(4) ネットワーク構造と機能の分析方法

本研究では、ネットワークの構造だけでなく、構造と機能という二つの観点からネットワーク分析をおこなう。この分析では、頂点(分子)が持つ生体機能の情報をもとに、頂点に対応する生体機能クラスに分類する。その後、同じ生体機能クラスに属する頂点同士の繋がりや違う生体機能クラスに属する頂点同士の繋がりについて分析する。各生体機能クラスごとの頂点の繋がりの特徴について調べることにより、特定の機能に属する頂点のネットワーク分布とネットワーク構造の関係を明らかにする。その結果から「ncRNA 分子が類似した生体機能を持つタンパク質との間に辺を持つ可能性は、ランダムな場合と比べ高くなるのか」という問いに対する答えを提示する。また、同様の分析を転写制御ネットワークに対しても行い、転写制御ネットワークと ncRNA 相互作用ネットワークの構造と機能の関係を比較し、ncRNA が制御機能を持つ可能性について考察する。

ここでは、特定の生体機能クラスに属する頂点分布の特徴を表すパラメータ D , H を計

算し、各ネットワークの構造と機能の関係について調べる。この手法は過去に携帯電話サービスとタンパク質の相互作用の機能的な特徴についての研究に用いられおり、頂点を持つ機能とネットワーク構造の関係について、その特徴を示すことに成功している。以後の各小節ではパラメータ D , H の計算方法とそれらが示すネットワーク構造と機能の関係について説明する。

(5) 境界面の計算方法

生体機能に属する頂点 n_1 と属さない頂点 n_0 をネットワーク上に配置し、その配置の組合せを全て計算することによってパラメータ D , H の値が取り得る範囲(境界面)を調べることができる。本研究では Ant Colony Optimization (ACO) と Fiduccia-Mattheyses (FM) アルゴリズムを利用し D , H の境界面を特定する

4. 研究成果

我々は転写因子-遺伝子相互作用および ncRNA-タンパク質相互作用のデータを収集し、これらのデータをもとに制御ネットワークを構築しました。特に、私たちはヒトを含む6種のモデル生物それぞれについて、ncRNA タンパク質相互作用ネットワークを構築しました。

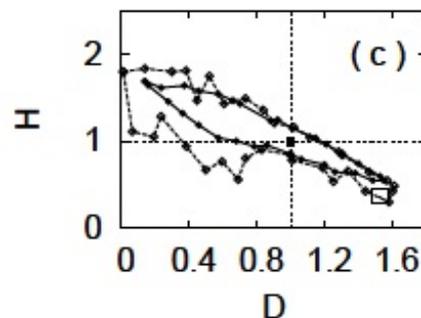


図2. ncRNA 相互作用ネットワークの D , H の境界面図. 破線(白)はF-M アルゴリズム, 実線(黒)はACO アルゴリズムによる境界面図. 図中の黒四角はランダムな場合の値, 白四角は実際の ncRNA 相互作用ネットワークの (D , H) の値. (c) は機能クラス4 の RNA processing の境界面である。

この ncRNA タンパク質相互作用ネットワークは、大域的なネットワーク分析によって、次数分布について転写調節ネットワークと高い類似度があることを明らかにしました。また、これらのネットワークの局所的な相互作用構造を豊富なモチーフに基づいて特徴付けすることで、ncRNA タンパク質ネットワークと転写因子-遺伝子調節ネットワークの間

のいくつかの重要な類似性を明らかにしました。

さらに、ネットワーク構造と生体機能の関係について調べるために、特定の機能に関する頂点のネットワーク分布について分析した。この分析では、特定の機能に関する頂点の分布を示す二つのパラメータを計算し、その結果から各ネットワークの頂点分布の特徴を特定した。最後に、ncRNA 相互作用と転写制御ネットワークの頂点分布とその分布範囲について調べた。ここでは、最適化アルゴリズムによって二つのパラメータの値が取り得る範囲を計算し、境界面図として表した。境界面図と実際のパラメータの値の比較から、対象とした生体機能に属する頂点の分布は境界面図の中の特定の領域に集中していることが分かった。また、頂点の分布は ncRNA 相互作用ネットワークと転写制御ネットワーク共に同じ領域に集中していることを確認した。この結果から、分析した生体機能に属する頂点のネットワーク分布はランダムなものではなく、生体機能と関係していることが考えられる。

本研究では、ncRNA 相互作用と転写制御ネットワークの頂点分布に類似点が見つかることより、ncRNA 分子が転写制御機能に関与している可能性が考えられる。本研究では ncRNA 相互作用と転写制御ネットワークに対して、全体的構造、局所的構造、構造と頂点機能との関係についてネットワーク分析をおこない、二つのネットワークの特徴を特定した。また、新しいネットワークの成長モデルの提案と解析をおこなった。これらの結果は、生体細胞内において ncRNA 分子が転写制御機能に関与している可能性が高いことを示している。

さらに、関連研究では、1,000以上個の有機体以上のマルチドメインタンパク質における情報解析と数理モデル化が研究された。21年度に開発した類似数理モデルを用い、マルチドメインタンパク質における内部重複ドメイン過程を研究しました。最終的に、病原体分子パターンによって誘導される遺伝子を同定するため新たな方法が発見された。

5. 主な発表論文等

(研究代表者、研究分担者及び連携研究者には下線)

[雑誌論文] (計 2 件)

①J.C. Nacher and N. Araki,
Structural characterization and modeling

of ncRNA-protein interactions, BioSystems Elsevier, 査読有, 101, 10-19, (2010).

②J.C. Nacher, M. Hayashida, T. Akutsu,
The role of internal duplication in the evolution of multi-domain proteins, BioSystems Elsevier, 査読有, 101, 127-135, (2010)

[学会発表] (計 1 件)

①J.C. Nacher and N. Araki,
On the relation between structure and biological function in transcriptional networks and ncRNA-mediated interactions, In Proc. of 2011 IEEE International Conference on Bioscience, Biochemistry and Bioinformatics, 査読有, pp. 348-352, Singapore, Feb. 2011.

6. 研究組織

(1) 研究代表者

ホセ・ナチエル (JOSE NACHER)

公立ほこだて未来大学・システム情報科学部・准教授

研究者番号: 60452984

(2) 研究分担者

なし

(3) 連携研究者

なし