

ncRNA-タンパク質相互作用ネットワークの 構造特性とモデル化

荒木 直 央^{†1} Jose Nacher^{†1}

真核生物の細胞内の機能制御において non-coding RNA(ncRNA) が重要な働きをする可能性が考えられている。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を行う。具体的には、ncRNA-タンパク質ネットワークと転写制御ネットワークの全体的構造と局所的構造の分析とその比較を行う。また、分析で特定された構造特性を再現する進化モデルを提案する。

Structural characterization and modeling of ncRNA-protein interactions

NAO ARAKI ^{†1} and JOSE NACHER^{†1}

Recent works have suggested possible regulatory roles to non-coding RNA molecules (ncRNA) in eukaryotic cells. Here, 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 that considers several mechanisms for the growth of the interaction network is discussed.

1. はじめに

哺乳類の体は約数兆個の細胞から成り立っている。さらに、それぞれの細胞は、数十万の DNA, RNA, 代謝物質, タンパク質分子などから構成されている。この膨大な分子間の相互作用によって、ゲノムから個々の機能タンパク質まで遺伝情報が伝達される。近年、分子間相互作用の関係を表した生体内ネットワークの研究に注目が集まっており、タンパク質相互作用ネットワークや代謝制御ネットワークなどの研究が進められている^{1),2)}。

^{†1} 公立はこだて未来大学 Future University-Hakodate

遺伝情報の伝達は「遺伝子 → RNA → タンパク質」という経路を経て、最終的にタンパク質として生体内で機能する。遺伝子から RNA が造られる過程は転写、RNA からタンパク質が造られる過程は翻訳と呼ばれている。生体内の機能分子であるタンパク質の発現量や発現時期は、転写によって制御されている。従って、転写は生体内の情報処理における重要な機能の一つと言える³⁾⁻⁶⁾。転写による制御は、転写因子 (Transcription Factor: TF) によって行われる。転写因子は標的遺伝子の転写を促進または抑制する。転写因子と標的遺伝子の間の相互作用をネットワークによって表現したものが転写制御ネットワークである。近年、転写の重要性から、転写制御ネットワークの研究が進められている^{7),8)}。

最近の研究により、non-coding RNA (ncRNA) 分子の転写過程への関与が示唆されている⁹⁾⁻¹¹⁾。ncRNA 分子とはタンパク質へ翻訳されない RNA 分子のことを指す。最近まで、ncRNA 分子はタンパク質の輸送や結合、合成などの特別な機能に関わる分子を除き、生命活動に関与しないと考えられていた。しかし、タンパク質をコードしているゲノムの割合が高等生物になるに従い高くなること、高等真核生物の細胞では転写される RNA の大体は、ncRNA であることなどが明らかになり、ncRNA 分子に対する考え方は変わりつつある¹¹⁾⁻¹³⁾。実際、特別な機能を持っている ncRNA 分子として、タンパク質の合成を行う rRNA やアミノ酸の運搬を行う tRNA などが知られている。最近では、snoRNAs や microRNAs などの small RNA 分子が見つかっており、これらの分子が遺伝子制御の機能を持つことが期待されている。

ncRNA を基盤とした転写制御という仮説は、発見されていない遺伝情報の伝達経路の存在を示唆する^{12),14),15)}。従って、ncRNA の転写機能への関与が認められれば、遺伝情報の処理や多細胞生物の制御に関する現在の知見の大きな変化が考えられる。しかしながら、ncRNA 制御ネットワークに関する全体的構造、局所的構造の研究はどちらも進んでいない。本研究では、ncRNA-タンパク質ネットワークの全体的構造と局所的構造の分析、転写制御ネットワークとの比較を通して ncRNA ネットワークの構造特性を調べる。また、特定した構造特性を再現するネットワーク進化モデルを提案し、数理モデルによる解析を行う。

2. ネットワーク解析

2.1 データセット

本研究では、NPInter database 16) で公開されている ncRNA 分子とタンパク質分子の相互作用データを用いる。利用したデータは、Escherichia coli, Saccharomyces cerevisiae, Caenorhabditis elegans, Drosophila melanogaster, Mus musculus, Homo sapiens の計 6

体の有機体のものである。また、今回の研究では、ncRNA 分子の転写機能への関連を調べるため、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比較を行う。比較には、研究が進んでおり、データが公開されている *Saccharomyces cerevisiae* と *Escherichia coli* K12 の転写制御ネットワークを用いる^{3),7),17),18)}。

2.2 ネットワーク構造

ncRNA-タンパク質ネットワーク、転写制御ネットワークは二部ネットワーク (二部グラフ) の構造をとる。二部ネットワークは二種類の頂点によって構成されており、同種の頂点間には辺が存在しない構造をとる。ncRNA-タンパク質ネットワークの場合、ncRNA 頂点とタンパク質頂点によって構成され、頂点間を結ぶ辺は ncRNA からタンパク質に対する生体機能的な作用を意味する。ここでは、ncRNA 頂点と繋がっている辺の本数を出次数、タンパク質頂点の辺の本数を入次数と考える。転写制御ネットワークの場合、二種類の頂点は転写因子 (TF) 頂点と標的遺伝子頂点となる。二頂点間にある辺は、TF 頂点が標的遺伝子頂点を制御するという意味を持つ。

今回の研究では、6 有機体の全てのデータを合わせた ncRNA-タンパク質ネットワーク (all RP ネットワーク) と *H. sapiens* のネットワーク (Hs RP ネットワーク) の分析を行う。比較に用いる転写制御ネットワーク (TG ネットワーク) は、*E. coli* K12 (Ec) と *S. cerevisiae* (Sc) の転写因子 (TF) と標的遺伝子のデータを用いる。ここでは、TF 頂点は ncRNA 頂点と対応し、標的遺伝子頂点はタンパク質頂点と対応する。

2.3 全体的構造 - 次数分布 -

ここでは、ncRNA-タンパク質ネットワークの全体的構造を知るために ncRNA 頂点の次数 (出次数) とタンパク質頂点の次数 (入次数) の分布について調べる。次数の分布は次数分布 $P(k)$ と累積次数分布 $P(> k)$ によって特定する。次数分布 $P(k)$ とは、あるネットワーク内における、次数が k の頂点の割合、累積次数分布 $P(> k)$ とは、ある次数 k よりも次数が多い頂点の割合を求めたものである。累積次数分布 $P(> k)$ が指数 γ' のべき乗則に従うとき、対応する次数分布 $P(k)$ は指数が $\gamma = \gamma' + 1$ のべき乗則に従う。次数分布が指数 γ のべき乗則に従うとは、次数分布の傾きが $P(k) \propto k^{-\gamma}$ の式に従うことを意味する。

図 1 に示すのは、6 有機体全てのデータを合わせた累積次数分布と、*H. sapiens* の累積次数分布の結果である。図 1(a) が 6 有機体の ncRNA 頂点、(b) がタンパク質、図 1(c) が *H. sapiens* の ncRNA 頂点、(d) がタンパク質頂点の分布である。図中の直線は、次数分布の傾きをべき乗則で近似したものである。図 1 の (a),(c) がべき乗則に従うと考えると、(a) は $\gamma' = 0.88 \pm 0.09$ ($\gamma = 1.88$)、(c) は $\gamma' = 0.97 \pm 0.06$ ($\gamma = 1.97$) となる。

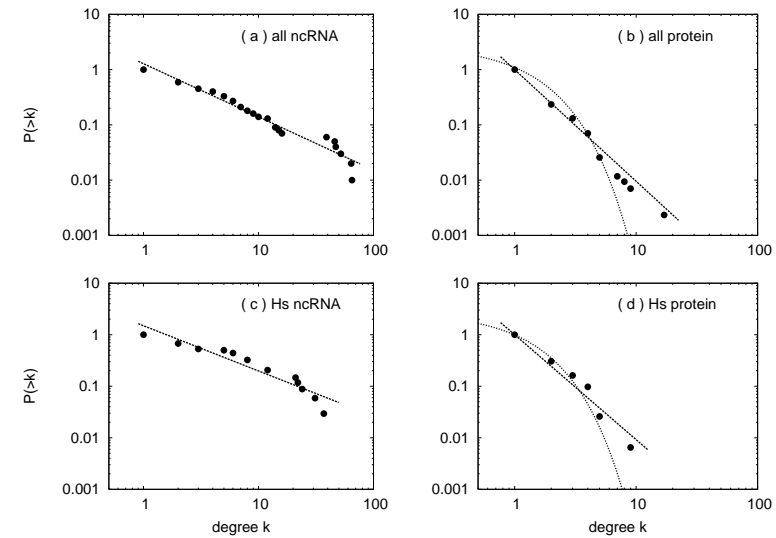


図 1 ncRNA-タンパク質ネットワークの累積次数分布

Fig. 1 cumulative degree distribution of ncRNA-protein networks

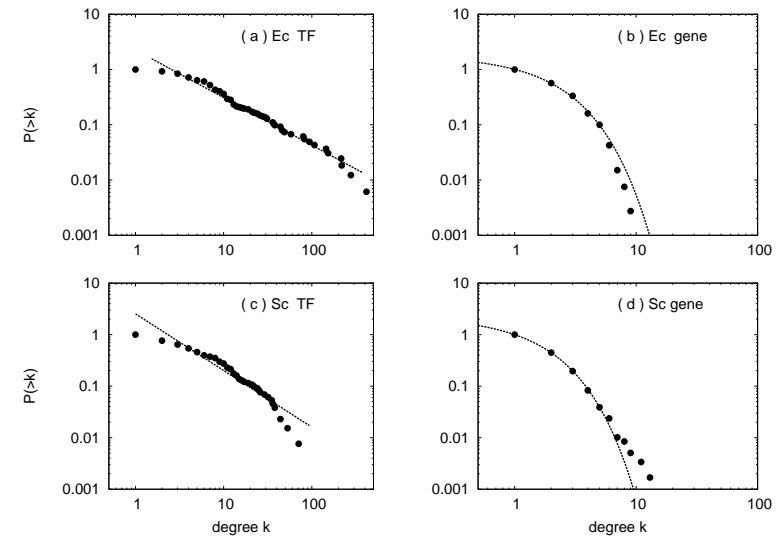


図 2 転写制御ネットワークの累積次数分布

Fig. 2 cumulative degree distribution of transcriptional regulation networks

ncRNA-タンパク質ネットワークの ncRNA 頂点の度数分布とタンパク質頂点の度数分布の図より、1. ncRNA 頂点の度数分布は負の指数のべき乗則に従う。2. タンパク質頂点の度数分布はべき乗則に従っているが、 e を底に持つ負の指数関数との関連も見られる、という二点を読み取れる。さらに、1, 2 の結果より、ncRNA 頂点とタンパク質頂点が異なる性質の度数分布を持つことが推測される。

同様に転写制御ネットワークの累積度数分布の結果は図 2 の様になった。図 2(a) が E. coli K12 の TF 頂点、(b) が標的遺伝子 (gene) 頂点、(c) が S. cerevisiae の TF 頂点、(d) が標的遺伝子頂点の分布である。転写制御ネットワークでは、TF 頂点がほぼべき乗則に従っている。図 2(a) は $\gamma' = 0.86 \pm 0.02$ ($\gamma = 1.86$)、(c) は $\gamma' = 1.11 \pm 0.07$ ($\gamma = 2.11$) となる。また、標的遺伝子の度数分布は指数関数に従っていることが分かる。この結果から、ncRNA-タンパク質ネットワークと転写制御ネットワークが類似した度数分布をとることが分かる。

2.4 局所的構造 - ネットワークモチーフ -

ここでは、ネットワークの局所的構造を特定するためにネットワークモチーフについて調べる。ネットワークモチーフとはネットワークに頻出する部分構造 (部分グラフ) のことを指す。モチーフの特定はランダムネットワークとの比較によって行われる。具体的には、実際のネットワークのある部分構造の出現頻度がランダムネットワークの出現頻度と比べ、統計的に有意な場合、その部分構造はモチーフと呼ばれる^{18),19)}。本研究では、式 1 より求められる Z 値によってモチーフの特定、比較を行う。

$$Z = \frac{N_{real} - N_{rand}}{\sigma} \quad (1)$$

ここでは、 N_{real} は実際のネットワークで出現する部分構造の数、 N_{rand} と σ はランダムネットワークで出現する部分構造の平均数と標準偏差である。 Z が高い値を示すとき、部分構造は統計的に有意なパターンである。本研究は Milo のスイッチングアルゴリズム¹⁹⁾ に従ってランダムネットワークを生成し、 Z 値を求めた。

調査の対象とする部分構造は、bi-fan モチーフと呼ばれる 4 つの頂点からなる部分構造 (完全二部グラフ K_{22}) と 5 つの頂点から成る部分構造 (完全二部グラフ K_{23} , K_{32}) である。各ネットワークにおけるそれぞれの部分構造の Z 値の結果を図 3 に示す。

3. 進化モデル

ここでは、入次数と出次数の度数分布が異なる (べき乗則と指数関数に従う) ネットワー

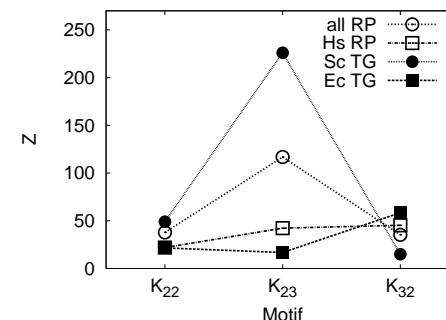


図 3 完全二部グラフ K_{22} , K_{23} , K_{32} の Z 値
Fig.3 Z of complete bipartite graph K_{22} , K_{23} , K_{32}

ク構造を再現する進化モデルを提案する。過去研究において、ネットワークの進化モデルや対応する数理モデルは、いくつか提案されている²¹⁾⁻²³⁾。本研究の進化モデルでは、生体内でおこなわれている遺伝子の複製をネットワーク上でモデル化する。具体的には、遺伝的形質を伴った複製と複製の過程で遺伝的形質が変化した複製についてのモデル化を行う。ここでは、遺伝的形質を伴う複製を継承複製と呼ぶ。継承複製では、ある頂点を複製した場合、その頂点を持つ度数と辺のリンク先の情報が、複製された頂点に継承される。また、遺伝的形質が変化した複製は、ある頂点を複製する場合、その頂点を持つ度数のみが継承され、辺のリンク先はランダムに選択する。この複製のことは今後、度数継承複製と呼ぶ。提案する進化モデルでは、これら二つの複製過程の他に頂点の追加による過程も考える。この追加過程では、既存のネットワークに一つの新しい頂点が追加される。このとき、新しい頂点と辺によってつながる頂点はランダムに選択される。

したがって、本研究では、継承複製と度数継承複製の二つの頂点の複製過程と一つの頂点の追加過程によるネットワーク進化モデルを提案する。また、これらの進化過程は二種類の頂点それぞれに対して起こる。このことより、本研究のモデルでは、二種類の頂点それぞれに対して数理モデルを立て解析を行う²⁴⁾。今後 ncRNA, TF に対応する頂点を N 頂点、タンパク質、標的遺伝子に対応する頂点を M 頂点と呼ぶ。具体的なネットワーク進化モデルのアルゴリズムは次に記す。

3.1 アルゴリズム

提案するネットワーク進化モデルのアルゴリズムを以下に示す。

ステップ 1 N 頂点と M 頂点を l 個ずつ用意する. 各 N 頂点は一つの M 頂点と辺で繋がっている. (このときの N 頂点, M 頂点の次数は全て 1 である)

ステップ 2.1 時刻 $t = 1$ において, α_N の確率で N 頂点が複製される. 複製される N 頂点の選択はランダムである. 頂点の複製方法は二つあり確率 $\hat{\alpha}_N$ で継承複製が, 確率 α_N^* で次数継承複製が選ばれる. これらの確率は $\alpha_N = \hat{\alpha}_N + \alpha_N^*$ となる.

継承複製

選択された N 頂点とその N 頂点の辺が複製される. 複製された新たな N 頂点は, もとの N 頂点と同じ次数となり, 同じ M 頂点との間に辺を持つ.

次数継承複製

選択された N 頂点とその N 頂点の次数 (k^0) が複製される. 複製された新たな N 頂点は, もとの N 頂点と同じ次数 k^0 となる. k^0 本の辺と繋がる M 頂点はランダムに選択される.

ステップ 2.2 同時刻 $t = 1$ において β_N の確率で, 一つの新しい N 頂点が追加される. このとき, 追加される N 頂点は, ランダムに選ばれた M 頂点 (n_0 個) と辺で繋がれる. 時刻 t におこる進化過程は $\alpha_N + \beta_N = 1$ の確率で選択される.

ステップ 3 ステップ 2.1, 2.2 と同じ時刻において, M 頂点も同様にステップ 2.1 の複製過程またはステップ 2.3 の追加過程が起こる (このときの過程は N と M を逆転させたものとなる). 複製過程の確率は $\alpha_M = \hat{\alpha}_M + \alpha_M^*$, 追加過程の確率は β_M となり, $\alpha_M + \beta_M = 1$ である.

ステップ 4 ステップ 2.1 からステップ 2.2 が, ネットワークのサイズが望まれる大きさに成長するまで t 回繰り返される. よって, ネットワークは $t+l$ 個の N 頂点と同じく $t+l$ 個の M 頂点から成る.

3.2 解析的導入

前節のアルゴリズムから次のような方程式を得る. これは次数 k である頂点 N_k, M_k の時間変化を数式化したものである.

$$\begin{aligned} \frac{dN_k}{dt} = & \hat{\alpha}_M \left(\frac{k-1}{M(t)} N_{k-1} - \frac{k}{M(t)} N_k \right) + \alpha_M^* \left(\frac{k^0}{N(t)} N_{k-1} - \frac{k^0}{N(t)} N_k \right) \\ & + \alpha_N^* \frac{N_k}{N(t)} + \hat{\alpha}_N \frac{N_k}{N(t)} + \beta_M \left(\frac{m_0}{N(t)} N_{k-1} - \frac{m_0}{N(t)} N_k \right) + \beta_N \delta_{kn_0} \end{aligned} \quad (2)$$

$$\begin{aligned} \frac{dM_k}{dt} = & \hat{\alpha}_N \left(\frac{k-1}{N(t)} M_{k-1} - \frac{k}{N(t)} M_k \right) + \alpha_N^* \left(\frac{k^0}{M(t)} M_{k-1} - \frac{k^0}{M(t)} M_k \right) \\ & + \alpha_M^* \frac{M_k}{M(t)} + \hat{\alpha}_M \frac{M_k}{M(t)} + \beta_N \left(\frac{n_0}{M(t)} M_{k-1} - \frac{n_0}{M(t)} M_k \right) + \beta_M \delta_{km_0} \end{aligned} \quad (3)$$

ここでは, $N(t) = M(t) = t+l$ は時刻 t における N 頂点 (M 頂点) の全頂点数である. また, δ_{kn_0} は新しい N 頂点からの既存の n_0 個の M 頂点への新たな繋がりを, δ_{km_0} は新しい M 頂点からの既存の m_0 個の N 頂点への新たな繋がりを意味している. また, k^0 はネットワークの頂点の平均次数を示している. 頂点が次数継承複製 (確率 α_N^*, α_M^* の過程) によって起こるとき, 複製された新しい頂点は元の頂点の次数を継承し, 新しく繋がる k^0 個の頂点をランダムに選ぶ.

式 2 に確率分布 $n_k = N_k/N(t)$ を導入し, $t \rightarrow \infty$ の定常分布を考えることにより次のような解得る.

$$n_k \propto (\hat{\alpha}_M k + \beta_M m_0 + \alpha_M^* k^0)^{-\frac{1-(\hat{\alpha}_N + \alpha_N^*) + \hat{\alpha}_M}{\hat{\alpha}_M}} \quad (4)$$

ここで, k の極限を考えると式 4 は次のようになる.

$$n_k \propto k^{-\frac{1-(\hat{\alpha}_N + \alpha_N^*) + \hat{\alpha}_M}{\hat{\alpha}_M}} \quad (5)$$

同様に, 式 3 に確率分布 $m_k = M_k/M(t)$ を導入する.

$$m_k \propto (\hat{\alpha}_N k + \beta_N n_0 + \alpha_N^* k^0)^{-\frac{1-(\hat{\alpha}_M + \alpha_M^*) + \hat{\alpha}_N}{\hat{\alpha}_N}} \quad (6)$$

特に, $\alpha \rightarrow 0$ を考えると式 3 の解として次を得る.

$$m_k \propto e^{-\frac{\beta_M}{n_0 + \alpha_N^* (k^0 - n_0)} k} \quad (7)$$

3.3 平均次数

ここでは, 平均次数 k^0 の時間変化について述べる. 提案したネットワーク進化モデルでは, 平均次数は N 頂点, M 頂点共に同じ値 $k^0 = \langle k_m \rangle = \langle k_n \rangle$ をとる. ネットワークの全ての辺の本数 L , 全頂点数 $N(t) = M(t) = t$ とすると, 平均次数を次のように表せる.

$$k^0 = L/t \quad (8)$$

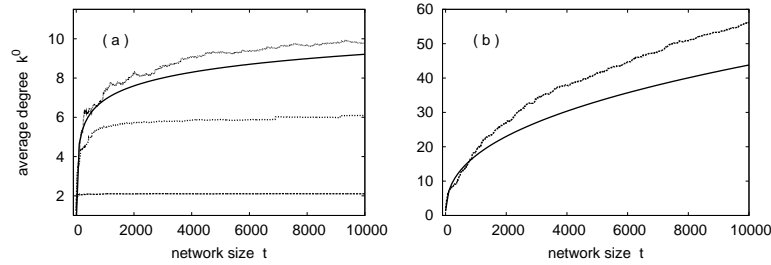


図4 平均時数 k^0 の時間変化のシミュレーション結果
Fig. 4 The simulation result of the time change of the mean degree k^0

また、辺の増加量は次のような式になる。

$$\frac{dL}{dt} = \beta_N + \beta_M + (\alpha_N + \alpha_M)k^0 \quad (9)$$

これらの方程式に対して、三つの確率の場合分けを行うことにより、次のような時間経過における平均次数 k^0 が導かれる。 $\alpha_N + \alpha_M < 1$ の場合、 k^0 は定数 $(\beta/(1-\alpha))$ となり、時間による変化はない。ここでは、 $\beta = \beta_N + \beta_M, \alpha = \alpha_N + \alpha_M$ とする。 $\alpha_N + \alpha_M = 1$ の場合、 k^0 は $\beta \ln t$ に従い、時間経過と共に緩やかに増加する (t は全 $N(M)$ 頂点数)。 $\alpha > 1$ の場合、平均次数は $t^{\alpha-1}$ に従って増加する (図4)。

図4は進化モデルのアルゴリズムに基づいたシミュレーションによる平均次数の時間変化を示す。図4(a)の破線は、上から順に $\alpha_N = 0.05, \alpha_M = 0.8, \alpha_N = \alpha_M = 0.5, \alpha_N = \alpha_M = 0.05$ の結果である。また、上から二番目の実線は解析結果 $\beta \ln t$ の時間変化を示す。図4(b)の破線は $\alpha_N = \alpha_M = 0.7$ 、実線は $t^{\alpha-1}$ の結果である。

ネットワークサイズを10000とし、シミュレーションと同じ確率を解析結果に代入すると $\alpha_N = 0.05, \alpha_M = 0.8$ は $k^0 = 7$ 、 $\alpha_N = \alpha_M = 0.5$ は $k^0 = 9.2$ 、 $\alpha_N = \alpha_M = 0.05$ は $k^0 = 2.1$ となる。この結果はシミュレーションの結果とほぼ一致している。

3.4 シミュレーション結果

ここでは、提案したネットワーク進化モデルの次数分布のシミュレーション結果について述べる。図5は $\alpha_N = \alpha_M = 0.5$ の次数分布である (左: N 頂点, 右: M 頂点)。このとき、確率 $\hat{\alpha}_N = 0.5 (\alpha_N^* = 0)$ 、 $\hat{\alpha}_M = 0.5 (\alpha_M^* = 0)$ では、両頂点共にべき乗則に従う分布となる (図5(a),(b))。しかし、 $\hat{\alpha}_N (\hat{\alpha}_M)$ が0に近付くと、次数分布は次第に指数関数に従うようになる。図5(c)は $\hat{\alpha}_N = 0.35 (\alpha_N^* = 0.15)$ 、(d)は $\hat{\alpha}_M = 0.35 (\alpha_M^* = 0.15)$ 、(e)は

$\hat{\alpha}_N = 0.25 (\alpha_N^* = 0.25)$ 、(f)は $\hat{\alpha}_M = 0.25 (\alpha_M^* = 0.25)$ 、(g)は $\hat{\alpha}_N = 0.15 (\alpha_N^* = 0.35)$ 、(h)は $\hat{\alpha}_M = 0.15 (\alpha_M^* = 0.35)$ の結果である。

$\alpha_N = \alpha_M = 0.05$ の次数分布は図6となる。確率が $\hat{\alpha}_N = 0.05$ と $\hat{\alpha}_M = 0.05$ のとき、つまり、複製が継承複製のみで行われた場合、次数分布は両頂点共に指数関数に従う (図6(a),(b))。ここでは、 $\alpha_N (\alpha_M)$ の確率が小さいため、 $\hat{\alpha}_N, \alpha_N^* (\hat{\alpha}_M, \alpha_M^*)$ の確率も共に小さくなる。したがって、それらの確率を変化させても、次数分布は指数関数に従うままで変化しない。図6(c)は $\hat{\alpha}_N = 0.015 (\alpha_N^* = 0.035)$ 、(d)は $\hat{\alpha}_M = 0.015 (\alpha_M^* = 0.035)$ の結果である。

最後に、 $\alpha_N = 0.05, \alpha_M = 0.8$ の場合を考える (図7)。 $\hat{\alpha}_N = 0.05, \hat{\alpha}_M = 0.8$ のとき、つまり複製が継承複製のみで行われた場合、 N 頂点の次数分布はべき乗則、 M 頂点の分布は指数関数に従う (図7(a),(b))。ここで、次数継承複製の確率 $\alpha_N^* (\alpha_M^*)$ を増やし、継承複製 $\hat{\alpha}_N (\hat{\alpha}_M)$ を減らした場合、両頂点の次数分布は指数関数へと変化する。図7(c)は $\hat{\alpha}_N = 0 (\alpha_N^* = 0.05)$ 、(d)は $\hat{\alpha}_M = 0 (\alpha_M^* = 0.8)$ の結果である。この結果より、提案した進化モデルでは、異なる次数分布 (べき乗則と指数関数) を再現するためには複製過程と追加過程の割合が重要となることが分かる。さらに、べき乗則に従う分布を再現するためには、継承複製が重要であり、次数継承複製はあまり関与していないと考えられる。

4. 結果

4.1 次数分布

次数分布の結果より、ncRNA-タンパク質ネットワークでは、ncRNA頂点とタンパク質頂点が異なる次数分布を持つことが推測される。ncRNA頂点のべき乗則に従う次数分布から、少数のncRNA頂点が多くタンパク質頂点との間に辺を持つこと、つまり、ncRNA頂点には次数が高いハブ頂点が存在することが確認できる。また、この次数分布は、ほとんどのncRNA頂点は1、2個のタンパク質頂点とのみ作用していることも表している。これらの結果から、ncRNA頂点はスケールフリー性を持つ可能性が考えられる。タンパク質頂点の次数分布も、べき乗則に従っている様に見える。しかし、タンパク質頂点では、最大次数が小さく、最大次数へ向かう分布の減少が速い。このことから、ほとんどのタンパク質頂点は1、2個のncRNA頂点とのみ作用しており、ハブ頂点は存在しないことが分かる。従って、ncRNA頂点とは異なり、タンパク質頂点の分布は指数関数に近い性質があると推測される。

転写制御ネットワークの次数分布は、TF頂点がべき乗則、標的遺伝子が指数関数に従う。

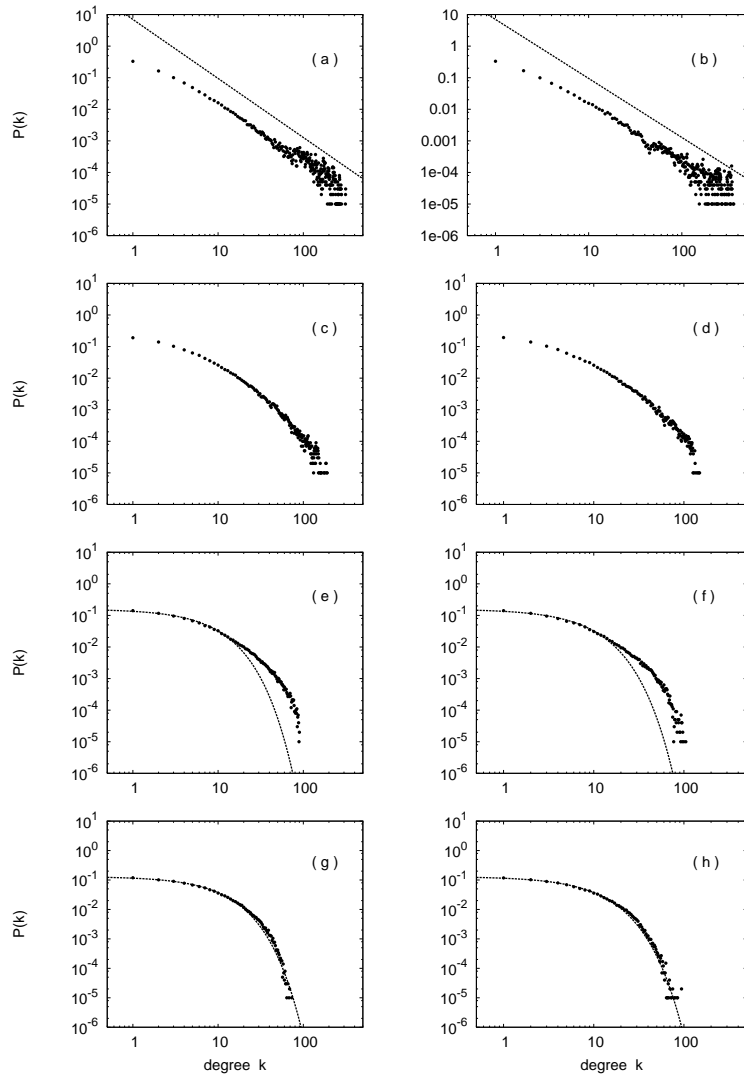


図 5 $\alpha_N = \alpha_M = 0.5$ の次数分布のシミュレーション結果
 Fig. 5 The simulation result of the degree distribution of $\alpha_N = \alpha_M = 0.5$

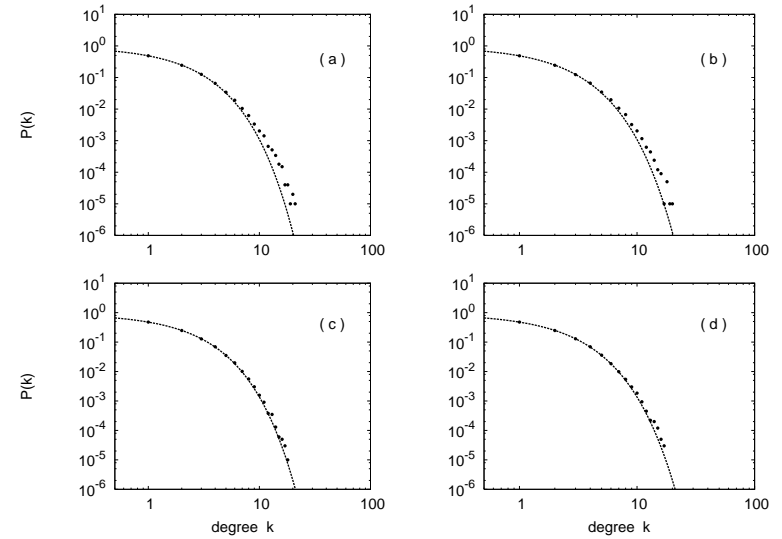


図 6 $\alpha_N = \alpha_M = 0.05$ の次数分布のシミュレーション結果
 Fig. 6 The simulation result of the degree distribution of $\alpha_N = \alpha_M = 0.05$

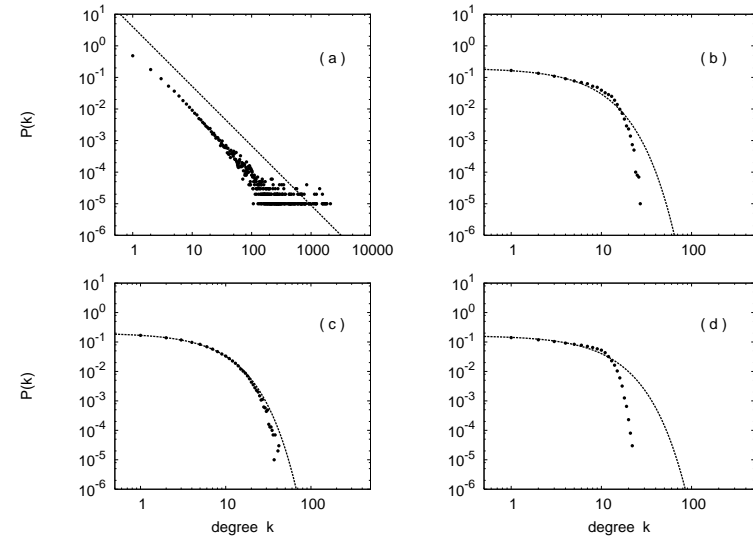


図 7 $\alpha_N = 0.05, \alpha_M = 0.8$ の次数分布のシミュレーション結果
 Fig. 7 The simulation result of the degree distribution of $\alpha_N = 0.05, \alpha_M = 0.8$

これは、ncRNA-タンパク質ネットワークの次数分布の結果とほぼ一致している。このことより、ncRNA-タンパク質ネットワークと転写制御ネットワークの全体的構造が類似していることが分かる。

4.2 ネットワークモチーフ

ncRNA-タンパク質ネットワークと転写制御ネットワークでは、部分構造 K_{22} (bi-fan), K_{23} , K_{32} が統計的有意性を持ち、ネットワークモチーフとなる。また、それぞれの Z 値が高いことより、これらの部分構造はネットワークの局所的構造の特徴と考えられる。

また、bi-fan モチーフの頂点の一つが継承複製されることにより発生する K_{23} , K_{32} モチーフの Z 値が高いことが分かる。これはネットワークの進化に継承複製の過程が関わっている可能性を示す。さらに、真核生物である *S. cerevisiae* の転写制御ネットワークと ncRNA-タンパク質ネットワークの Z 値の分布には相関が見られる。

4.3 モデルとシミュレーション

提案モデルを数理解析を行った結果、次数分布がべき乗則指と指数関数に従うことが分かった。この結果はネットワーク構造分析で特定した次数分布とも一致している。

また、提案モデルをもとにシミュレーションを行った結果、複製過程(継承複製と次数継承複製)と追加過程の確率を変化させることにより、1. 次数分布が両頂点共にべき乗則に従う、2. べき乗則と指数関数に従う、3. 両頂点共に指数関数に従う、という三つ分布をとる確率の組合せがあることが分かった。また、これらの結果は、実際の ncRNA-タンパク質ネットワークや転写制御ネットワークでは、ncRNA 頂点(TF 頂点)の複製過程の確率が低く、タンパク質頂点(標的遺伝子頂点)の複製過程の確率が高い可能性を示している。また、次数継承複製による複製の確率は低いことも分かる。

これらの結果より、ncRNA-タンパク質ネットワーク、転写制御ネットワーク次数分布の特徴を再現するためには、複製もとの頂点の情報(次数と辺のリンク先の情報)が継承される、継承複製の過程が重要であることが分かる。

5. 考 察

ncRNA-タンパク質ネットワークの分析によって、転写制御ネットワークとの次数分布に類似性がある可能性が示された。特に、ncRNA と TF の次数分布がスケールフリー性を持つことから、ncRNA 分子が転写制御に関連している可能性が考えられる。一方、タンパク質頂点の分布は、指数関数的な減少ではなく、べき乗則に従う減少のように見える。しかし、タンパク質頂点の最大次数が 20 程度であることから、ハブ頂点を持たず、ネットワーク構

造は指数関数のそれに近いことが読み取れる。また、この次数最大値は、制御ネットワークの標的遺伝子の値とほぼ一致している。non-coding 分子による相互作用の実験結果は少なく、大半のデータは不完全である。したがって、より多くの ncRNA 分子とタンパク質分子の相互作用のデータが実験により集まれば、タンパク質の次数分布は指数関数の減少に従うようになることが予想される。

ncRNA-タンパク質相互作用ネットワークの局所的構造に関する分析では、部分構造 K_{22} (bi-fan), K_{23} , K_{32} が、これらのネットワークのモチーフであることが分かった。 K_{23} , K_{32} モチーフの構造は、bi-fan モチーフの頂点の一つが継承複製された構造と同じである。このことより、高い Z 値を持つ K_{22} (bi-fan), K_{23} , K_{32} モチーフの存在は、ネットワークの進化に頂点の複製過程(継承複製)が関わっている可能性を示す。

ネットワーク成長の数理モデルの解析とシミュレーション結果によると、このモデルでは、それぞれの過程(頂点複製、頂点追加)が起こる確率の組合せを変えることにより、次数分布の特徴が異なることが明らかになった。これにより、次数分布がべき乗則と指数関数に従うネットワーク構造を再現するためには、頂点の複製過程と追加過程が異なることが重要となることが分かる。また、次数継承複製は、次数分布のべき乗則の構造には、あまり関係していないと考えられる。さらに数理モデルの解析には、平均次数の値が関与し、次数分布に関係していることが分かった。

本研究では生体分子の機能関係をネットワークとして表現し、ネットワーク分析と数理モデルによる解析を行った。この研究では、それぞれの頂点特定の機能、つまり ncRNA 分子に期待される制御機能の存在を確かめることはできない。しかし、生体内の複雑なネットワークの研究では、モデルを単純化することが重要となる。なぜなら、複雑なモデルの場合、解が求まる数理モデルを得ることが困難になるためである。本研究で行った二部ネットワークのモデル化は、あまり研究が進んでいない。また、二種類の頂点の個々の解を求めるため、解析は複雑になる。このためのモデルの単純化は重要となる。数理モデルの解析による解の導出は、何千もの分子からなるネットワーク構造や、それらの確率分布の予測に有用である²¹⁾⁻²⁴⁾。

ncRNA が関わる実験データは少なく、ncRNA の制御機能についての分析は不完全である。しかし、本研究の結果は ncRNA 分子の制御機能への関連性を示している。また、最近の研究では、癌の生成に関与する ncRNA 分子の新たに機能が見つかり、頸動脈疾患や糖尿病などの複雑な病気への関与の可能性も示唆されている¹⁵⁾。これらの研究の結果は、ncRNA 分子と制御機能との関係を明らかにすること、つまり、新しい ncRNA 分子の機能

の発見を目的とした実験とその進歩の重要性を示している。数理モデルによる解析とデータベースに蓄積されているデータの分類と解析、さらに実験によるデータの収集が組み合わされることにより、ncRNA分子が持つ機能の解明が期待される。

謝辞 筆者(荒木直央)は公立はこだて未来大学特別研究費重要区分(平成21年度)の助成を受けている。支援して頂いた皆様に深い感謝の意を示す。

参 考 文 献

- 1) Hartwell, L.H., Hopeld, J.J., Leibler, S. and Murray, A.W.: From molecular to modular cell biology, *Nature*, Vol.402, pp.C47-C52 (1999).
- 2) Bray, D.: Molecular Networks: The Top-Down View, *Science*, Vol.301, pp.1864-1865 (2003).
- 3) Guelzim, N., Bottani, S., Bourguin, P. and Kepes, F.: Topological and causal structure of the yeast transcriptional regulatory network, *Nature Genetics*, Vol.31, pp.60-63 (2002).
- 4) McAdams, H.H., Srinivasan, B. and Arkin, A.P.: The evolution of genetic regulatory systems in bacteria, *Nature Review Genetics*, Vol.5, pp.169-178 (2004).
- 5) Harbison, C.T., Gordon, D.B., Lee, T.I., Rinaldi, N.J., Macisaac, K.D., Danford, T.W., et al.: Transcriptional regulatory code of a eukaryotic genome, *Nature*, Vol.431, pp.99-104 (2004).
- 6) Baraábsi, A.-L. and Oltvai, Z.N.: Network biology: understanding the cell's functional organization, *Nature Review Genetics*, Vol.5, pp.101-103 (2004).
- 7) Luscombe, N.M., Babu, M.M., Yu, H., Snyder, M., Teichmann, S.A., and Gerstein, M.: Genomic analysis of regulatory network dynamics reveals large topological changes, *Nature*, Vol.431, pp.308-312 (2004).
- 8) Teichman, S.A., Madan Babu, M.: Gene regulatory network growth by duplication, *Nature Genetics*, Vol.36, pp.492-496 (2004).
- 9) Mattick, J.S. and Makunin, I.V.: Small regulatory RNAs in mammals, *Human Molecular Genetics*, Vol.14, pp.R121-R132 (2005).
- 10) Shimoni, Y., Friedlander, G., Hetzroni, G., Niv, G., Altuvia, S., Biham, O. and Margalit, H.: Regulation of gene expression by small non-coding RNAs: a quantitative view, *Molecular Systems Biology*, Vol.3, pp.1-9 (2007).
- 11) Zhao, Y, et al.: MicroRNA regulation of messenger-like noncoding RNAs: a network of mutual microRNA control, *Trends in Genetics*, Vol.24, pp.323-327 (2008).
- 12) Mattick, J.S.: RNA regulation: a new genetics, *Nature Review Genetics*, Vol.5, pp.316-322 (2004).
- 13) Seitz, H., Youngson, N., Lin, S., Dalbert, S., Paulsen, M., Bachelier, J., et al.: Imprinted microRNA genes transcribed antisense to a reciprocally imprinted retrotransposon-like gene, *Nature Genetics*, Vol.34, pp.261-262 (2003).
- 14) Mattick, J.S., Gagen, M.J.: Accelerating networks, *Science*, Vol.307, pp.856-858 (2005).
- 15) Amaral, P.P., Dinger, M.E., Mercer, T.R., Mattick, J.S.: The eukaryotic genome as an RNA machine, *Science*, Vol.319, pp.1787-1789 (2008).
- 16) Wu, T., Wang, J., Liu, C., Zhang, Y., Shi, B., Zhu, X., et al.: NPInter: the non-coding RNAs and protein related biomacromolecules interaction database, *Nucleic Acid Research*, Vol.34, pp.D150-D152 (2006).
- 17) Regulon Database version 6.3. <http://regulondb.ccg.unam.mx/>
- 18) Shen-Orr, S., Milo, R., Mangan, S. and Alon, U.: Network motifs in the transcriptional regulation network of Escherichia coli, *Nature Genetics*, Vol.31, pp.64-69 (2002).
- 19) Milo, R., Shen-Orr, S., Itzkovitz, S., Kashtan, N., Chklovskii, D. and Alon, U.: Network Motifs: Simple Building Blocks of Complex Networks, *Science*, Vol.298, pp.824-829 (2002).
- 20) Krapivsky, P.L. and Redner, S.: Rate equation approach for growing networks, *Lecture notes in Physics*, Vol.625, pp.3-20 (2003).
- 21) Foster, D.V., Kauman, S.A. and Socolar, J.E.S.: Network growth models and genetic regulatory networks, *Physical Review E*, Vol.73, pp.031912 (2006).
- 22) Newman, M.E.J., Barabási, A.-L. and Watts, D.J.: *The structure and dynamics of networks*, Princeton University Press, Princeton (2007).
- 23) Dorogovtsev, S.N. and Mendes, J.F.F.: *Evolution of networks: From biological nets to the Internet and WWW*, Oxford University Press, Oxford (2003).
- 24) Nacher, J.C., Ochiai, T., Hayashida, M. and Akutsu, T.: A bipartite graph based model for protein domain networks, Springer Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering (Complex '09), Vol.4, pp.525-535 (2009).