生体分子ネットワークのトポロジー解析

京都大学化学研究所　バイオインフォマティクスセンター
山田拓司、五斗進

はじめに

遺伝子やタンパク質といった生体分子は生体内で複雑な相互作用をしており、巨大なネットワークを形成している。バイオインフォマティクスの分野では、このような生体ネットワーク自体が解析の対象になっており、様々な解析が行われている。特に、ネットワークトポロジーに関する解析は盛んであり、生体ネットワークと、それ以外の様々な大規模ネットワーク（インターネットや社会ネットワーク）の間に共通する多くの特徴が見出されたことから、非常に注目を集めている。

ここでは最初に、解析の対象として扱われている生体ネットワークとはどのようなものかということを紹介する。次に、それらのネットワークを用いたいくつかの研究例を、ネットワークのトポロジー解析を中心に解説していきたい。タンパク質間相互作用ネットワークなどの大規模ネットワークを用いたトポロジー解析は近年非常に盛んな研究領域であり、多くの解析例があるが、ここでは特にスケールフリー性に関わる解析を紹介する。最後に、我々の研究室で実際に行っている、代謝ネットワークについての解析例を紹介する。

生体分子ネットワークとは

生体分子の相互作用ネットワークとして、タンパク質、遺伝子、低分子化合物のネットワークなどがあげられる。遺伝子間、タンパク質間、化合物間それぞれの相互作用の例を表１に示した。この中から今回は遺伝子間の相互作用として発現パターン、タンパク質間相互作用として物理的結合、化合物間相互作用として酵素反応による構造変化（代謝系）を取り上げ、次節以降で紹介する。

表１ 生体ネットワークの例

- 遺伝子－ 遺伝子
  - 発現パターンの変化
  - アミノ酸配列類似性
- タンパク質－ タンパク質
  - シグナル伝達ネットワーク
  - 物理化学的結合
- 化合物－ 化合物
  - 代謝系

タンパク質間相互作用ネットワーク

酵母ツーハイブリッドシステム

生体ネットワークの代表例として、まずタンパク質間相互作用ネットワークを取り上げる。タンパク質は生体内で物理的に結合することで、複雑な機能体を形成しているが、酵母ツーハイブリッドシステムという方法が確立されたことにより、高速にその相互作用（二つのタンパク質が結合するかどうか）を検出できるようになった。酵母ツーハイブリッドシステムでは、相互作用の有無を確認したい二つのタンパク質の一方にDNA結合ドメインを、もう一方に転写活性化ドメインをフュージョンさせる。二つのタンパク質が結合すると、DNA結合ドメインが認識した塩基配列の下流にある遺伝子が発現するので、下流の遺伝子の発現情報から結合の有無を確認できる[1]。

トポロジー解析～その①：次数分布とべき乗則

このように大量の相互作用データが蓄積されることで、数千単位のタンパク質からなる大規模なネットワークを用いた解析が可能になった。特に２０００年頃にネットワークのトポロジーに関する多くの報告がなされている。トポロジーとはネットワークにおけるノード間のつながり構造の特徴を表す概念である。その基本となるものがノードの次数（ノードから出るエッジの数）分布であり、ネットワークを"グラフ"とみなして、その大域的な構造をノードの次数によって特徴付けるという方法である。ノードとはネットワークを構成する点であり、エッジとはノード間の関係性を表す線のことである。タンパク質間ネットワークの場合、ノードはタンパク質であり、エッジはタンパク質間の物理的結合に相当する。どのタンパク質がどれくらい他のタンパク質に結合し得るかという分布情報から、スケールフリーという概念が提唱された。すなわち、次数 k を持つタンパク質の数 $P(k)$ は $k^{-\gamma}$ であるというものである[2]。図１にはランダムなネットワークと実際に得られる大規模ネットワークの間に見られる次数分布の違いを示している。ここで得られた特徴的な性質は、実際のネットワークには"ハブ"と呼ばれる、非常に多くの関係性を持つノードが少数存在する反面、大多数のノードはごく少数の関係性しか持たないということである。このような性質はタンパク質間相互作用ネットワークのみならず、インターネット、送電線ネットワーク、人間関係のネットワークなど、一見非常に異なって見える様々な相互関係ネ

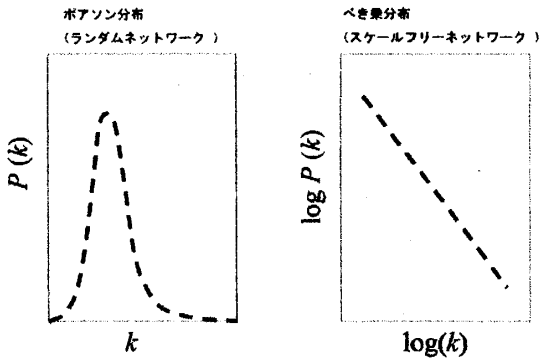

図１　トポロジーと次数分布

ランダムネットワークとスケールフリーネットワークの比較。横軸は次数ｋ、縦軸は次数ｋを持つノードが出現する頻度 P(k)。ランダムネットワークにおける次数分布は正規分布、スケールフリーネットワークの次数分布はべき乗則に従い、$P(k)=k^{-\gamma}$ となる。

ットワークにも共通して確認され、非常に注目を集めた。タンパク質間相互作用ネットワーク以外の生体ネットワークも例外ではなく、代謝系と呼ばれる低分子化合物のネットワークでも同様にスケールフリー性が確認されている。次節ではその代謝ネットワークに関わる解析について、詳細を解説する。

低分子化合物ネットワーク（代謝系）

代謝系とは

生体内に存在する多くの化合物は酵素反応によって、代謝、合成されている。代謝ネットワークとは、酵素反応をエッジとした生体内低分子化合物のネットワークであり、生物学的にも古くから研究されていることから、生体ネットワーク解析の対象によく用いられている。図２に示すように、化合物ネットワークにおける次数分布がべき乗則に従い、スケールフリー性を持つことは多くの生物種で確認されている。

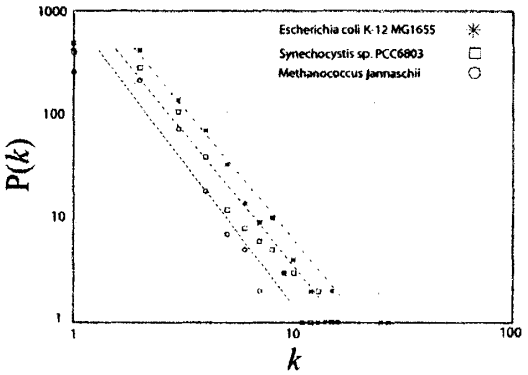

図２　生物種毎の代謝ネットワーク

それぞれの生物種において、スケールフリー性（べき乗則）が確認できる。

トポロジー解析～その②：流量解析

一つの化合物は異なる酵素反応により様々な形に変化していく。このような反応の流れを調節しているのは酵素タンパク質であり、そのタンパク質の性能、または変化する化合物の特性によってその流量は異なる。流量（flux）とは単位時間当たりの化合物の生成・代謝量と定義され、次数分布同様に、べき乗則に従うことが近年の解析で明らかにされた。Almaas らは大腸菌における 537 の化合物と 739 の酵素反応を用いて、栄養源としてコハク酸とグルタミン酸を用いた場合の各酵素反応の代謝流量の分布を調べている[3]。その結果、化合物ネットワークの"ハブ"の存在と同様に、非常によく利用される代謝流量の大きな酵素反応が少数存在し、残りの大部分は流量の小さな酵素反応であることが示された。代謝流量の分布はこれまでのネットワークトポロジー解析とは少々異なる解析であったが、代謝系に関する詳細な解析を行う場合に極めて重要である。べき乗則に従うという共通の性質が見出されたことから、ネットワーク解析と密接に関係した多くの解析が現在も進められている。次節では遺伝子間相互作用の解析と遺伝子発現量の分布についての新たな解析を紹介する。

遺伝子間相互作用ネットワーク

DNA マイクロアレイ

遺伝子はゲノム上に塩基配列として暗号化された領域であり、ゲノム配列（DNA）->

mRNA -> タンパク質という流れ（セントラルドグマ）に従い遺伝情報が生体に反映されている。DNA から mRNA へ、その遺伝情報が転写されることにより、生体内で様々な機能をはたすタンパク質が作られる。遺伝子がその機能を担うためには mRNA への転写が必須である。DNA マイクロアレイ（図３）は数万遺伝子の発現情報を一度に測定することができる。各遺伝子の発現量は細胞の状態や時期によって様々で、この発現時期や発現量が類似している遺伝子間には機能的な相関があると考えられる。ここで紹介するのは、遺伝子の発現量の分布がスケールフリー性を持っているという報告である。

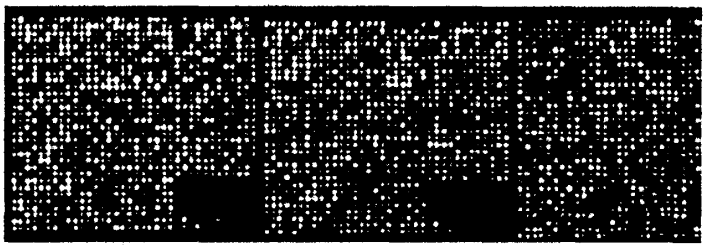

図３　DNA マイクロアレイ

mRNA の量を細胞全体で見る。各スポットに各遺伝子の cDNA が貼り付けてあり、発現している遺伝子は光る。

トポロジー解析～その③：rich-travel-more

上田らは複数の生物種の遺伝子発現量（細胞内における mRNA 量）を測定し、同程度に発現している遺伝子（種類）数の分布を調べた。その結果、どの生物種においても遺伝子発現量に基づく遺伝子数の分布がべき乗則に従うことが示された[4]。上田らは、さらに遺伝子発現量の変化についても調べた。任意の遺伝子が mRNA へと転写される発現量は細胞の状態によって変化する。例えば、細胞が熱にさらされたときや癌化した場合などには普段は全く機能していない遺伝子が発現する。同じ遺伝子の発現量の変化を二つの状態における発現量（k1、k2）で比較すると、もともと発現量の大きい遺伝子ほど、その変化も大きいことが示された。

一般的なネットワーク解析において、その状態変化は優先的選択によって説明されてきた。優先的選択とは、ネットワークの変化が起こるとき（ノードやエッジが増えるとき）、エッジを多く持つノードが新たなエッジを獲得しやすいと定義されている。このようなモデルは rich-get-richer モデルとも呼ばれる。それに対して上田らが明らかにしたのは、発現量の多い遺伝子（この場合は発現量であり、厳密にはネットワークのエッジではない）ほど、その量の変動が激しいという点である。彼らはこれを rich-travel-more と呼び、お金持ちはよりお金を持つようになる一方で、消費する金額も多くなることに例えた。

代謝ネットワーク解析

この節では、実際に我々の研究室で行っている代謝ネットワークの進化的挙動についての解析の一部を紹介する。一つは代謝ネットワークをタンパク質間相互作用ネットワークとみなして、そこに見られるモジュール構造を抽出するという試みである。もう一つは代謝ネットワークを化合物のネットワークとみなし、そのネットワークが進化過程でどの

ように生じてきたかを化合物のエッジ数の変化という観点から解析したものである（図４）。二つの解析ともにこれまで明らかにされているゲノム配列情報から得られる酵素タンパク質の系統プロファイルという情報量を用いている（図５）。

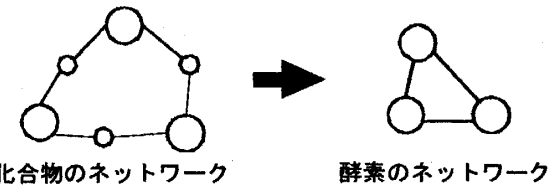

図４　代謝ネットワークの表現

代謝ネットワークは①化合物をノード、その間の酵素反応をエッジとする場合と、②酵素をノード、反応の連続性をエッジとする場合がある。

	organism 1	organism 2	organism 3	organism 4	organism 5	organism6
gene A	0	0	1	1	1	1
gene B	1	1	1	1	1	1
gene C	1	1	1	0	0	1
gene D	1	0	1	0	0	0

図５　系統プロファイル

ひとつの遺伝子に対して定義される指標の一つで、その遺伝子がどの生物種に存在するかを表したもの。存在すれば１、しないならば０。

代謝ネットワークのモジュール構造とそのトポロジー

様々な生物はそれぞれ固有の遺伝子セットを持っている。酵素反応を担うタンパク質も生物種固有であるために、酵素反応のつながりである代謝ネットワークも生物種それぞれが独自のネットワークを持っている。植物に見られる複雑な化合物を生合成する二次代謝経路などはその代表例と言える。一方で、解糖系のようにほとんどの生物種で保存されているような経路も存在する。代謝ネットワークとは生体内の化合物変換のネットワークであるが、生物種によってどの部分構造を保持しているかが異なる。この解析では、174生物種のゲノム情報を用いた酵素タンパク質の系統プロファイル（図５）を構築し、その類似度と代謝ネットワーク上でのつながり関係を考慮して代謝ネットワークを進化的保存度の類似な部分構造に分割した。系統プロファイルとはどの生物がどの酵素を持っているかという情報であるが、それは同時に進化過程における酵素遺伝子の欠失や挿入の情報が反映されている。したがって、このような代謝ネットワーク上で系統プロファイルが類似の酵素群によって構成されている部分構造（モジュール構造）は、進化的には一つの単位になっていると考えることができる。ここでは、さらに代謝ネットワークを抽象化し、モジュールをノードとしたネットワークとみなしそのトポロジーを調べた[5]。図６にその結果を示す。化合物をノードとした場合にそのトポロジーがスケールフリー性を示すことは先に示したが、酵素反応をノードとしたネットワークにおいてもスケールフリー性を示すことが知られている（化合物をノードとした場合と比べて、ちょうどノードとエッジ入れ替えた関係に当たる）[6]。図６は、進化的に関連性の強い遺伝子群によって構成される酵素モジュールをノードとした場合でも、同様にスケールフリー性を持つことを示している。

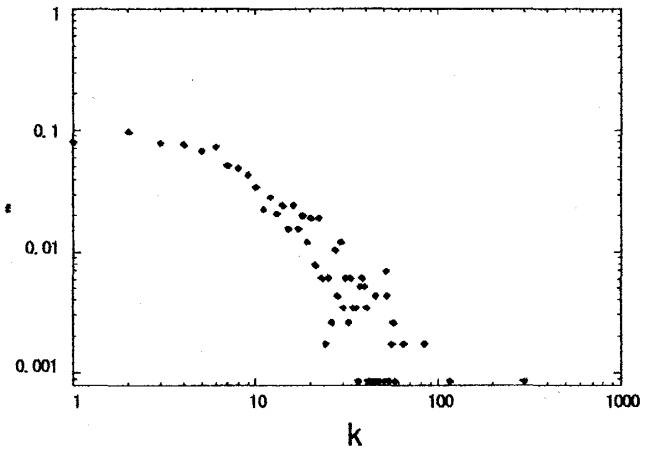

図６　モジュールネットワークのトポロジー

横軸は次数、縦軸はその頻度。モジュールをノードとしてみた代謝ネットワークもスケールフリー性を示す。

化合物ネットワークの次数変化による進化解析

代謝ネットワークの進化過程において、同じ化合物の利用法も変化していくと考えられる。この解析では、その利用法がどのように変化しているのかということを、ネットワーク情報を利用して検出するために、その次数の変化量を利用する。まず、系統プロファイルと最尤推定法を用いて、系統樹上の祖先生物種における酵素遺伝子セットを推定することにより、祖先生物種における代謝ネットワークを再構築する。次に同じ化合物の次数がどのように変化してきたのかを|k2-k1|プロットすることでその変化を追っていくことができる。ここで、k1 と k2 はそれぞれ、ある化合物について祖先生物種における代謝ネットワーク上の次数、および現存する生物種における代謝ネットワーク上の次数である。したがって、|k2-k1|は進化過程におけるある化合物の次数変化の絶対値を示す。

ここでは、|k2-k1|を以下に示したβ関数に拡張し利用している。

$$\beta = \frac{\sqrt{\sum_{i=1}^{N_k}(K-K_i)^2}}{N_k}$$

β関数は同じ次数 K を持つノードにおける次数変化の平均である。図７は祖先生物種（最終共通祖先 LUCA: Last Universal Common Ancestor）と現存する生物種の間でその比較をした結果である。遺伝子発現量と同様に、ここでも rich-travel-more の傾向が現れている。すなわち、非常に次数の多い化合物（よく利用される化合物）ほどその変化量も大きくなっており、その利用のされ方が進化過程でよく変化するということを示唆している[7]。

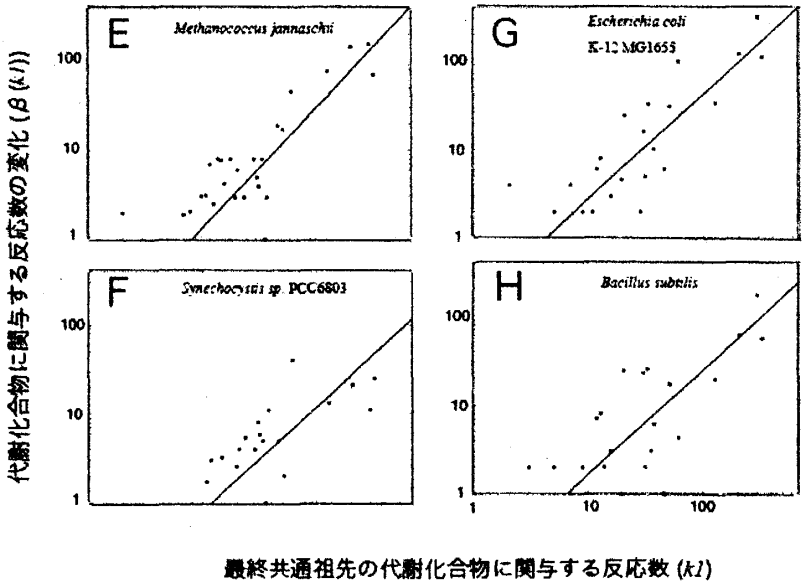

図７　β関数を用いた次数変化の検出

横軸に祖先生物種におけるエッジ数（ｋ）、縦軸にβ(k)。異なる生物種においても、右上がりの傾向は変わらない。

終わりに

生体分子間の様々な相互作用ネットワークはほとんどスケールフリー性を示す。さらに、ネットワーク情報だけでなく、流量や発現量といったものまで類似の傾向を示している。これらの知識から直接に遺伝子機能が推定できるわけではないが、生体システムとしての興味深い性質であり、代謝経路予測や創薬に利用されている遺伝子発現調節などに必須な知識として、今後もネットワーク解析は進められていくと考えられる。

参考文献

[1] Van Criekinge W & Beyaert R, *Bio. Proced. Online,* 2:1-38 (1999).

[2] Jeong H, *et al.*, *Nature* 407:651-654 (2000).

[3] Almaas E, *et al.*, *Nature* 427:839-843 (2004).

[4] Ueda HR, *et al.*, Proc. Natl. Acad. Sci. U. S. A., 101:3765-3769 (2004).

[5] Yamada T, *et al.*, *BMC Bioinformatics*, 7:130 (2006).

[6] Nacher JC, *et al.*, *Physica A*, 349:349-363 (2005).

[7] Tanaka M, *et al.*, *Genome Informatics*, 17(1):230-239 (2006).