

平成 15 年度知能システム科学専攻修士論文

代謝ネットワーク解析に基づく形成モデルの提案

松田 大典

Constructive Models for Metabolic Networks

Daisuke Matsuda

提出年月日 2004 年 2 月 24 日

主査教官 山村雅幸教授

審査教官 新田克己教授

審査教官 樺島祥介教授

代謝ネットワーク解析に基づく形成モデルの提案

松田 大典

Constructive Models for Metabolic Networks

Abstract

In this paper, we proposed constructive models that recreate metabolic networks. To this aim, we placed the restriction based on organic chemical properties, and broke the metabolic evolution into three phases: primitive network, primary network, secondary network. We considered their biological backgrounds, and then proposed each constructive rule. To check appropriateness of our models, we reconstructed primary and secondary networks. Using network measures (degree distribution, AL, diameter, clustering coefficient, betweenness centrality distribution), we characterized these networks. As a result, the primary model was well accorded with primary network, and the ability of the restriction was shown.

1 はじめに

20世紀に飛躍的な発展を遂げた分子生物学は、生物の構成要素や機能に関する豊富な知見を我々にもたらした。タンパク質やゲノムなどに代表されるそれらの個々の知見はデータベース上で整理され、容易に利用することができる。しかしながら、生体内で働く機能モジュールが個々の構成要素と直接対応付けられることはほとんどない。生物が実現している多くの特性は、DNAやRNA、タンパク質、生体内化合物による複雑な相互作用によって引き起こされている現象である。従って、生体内の相互作用によって形成される複雑なネットワークの構造やダイナミクスを理解することが、今世紀に残された重要な課題である。

この課題に対して多くの試みが行われている中、代謝ネットワークは近年注目されている研究分野である。代謝ネットワークでは、生体化合物によって構成されるネットワークを扱う。生物が引き起こす様々な現象は、最終的には生体化合物によって実現されている。従って、代謝ネットワークの構造やダイナミクスは、生物が持つシステムとしての特性を最も忠実に示していると考えられる。

また、生物が活動に必要なエネルギーや、生体の構成物質となる化合物を得るために利用する物質変換過程である代謝ネットワークは、生物の進化と強い相関がある。代謝ネットワークの担い手であるタンパク質とDNAの関係はニワトリと卵の関係であり、生物の

進化がどのようなであったかは未だ解明されていない。しかしながら、代謝ネットワークの形成原理のモデル化が可能であれば、生物の進化について大きな示唆を与えることができると期待される。特に、代謝ネットワークの特徴と他の人工、非人工的なネットワークの特徴(インターネットや人間関係ネットワーク)を比較することによって得られる知見は、代謝ネットワークの設計原理や機能特性に関する重要な示唆を与えるものとなるだろう。

代謝ネットワーク研究で現在主流となっているのは代謝ネットワークの解析である。Joengらは生体化合物をノード、生体内反応をエッジ(またはアーク)に対応させることで、作成したネットワークに対して純粋なグラフ理論解析を適用することに成功した[1]。この手法を利用して、代謝ネットワークがScale-Freeと呼ばれる特徴を持ち[1]階層的なモジュール構造を持つこと[2, 3]が示唆されており、現在も活発に研究されている。

一方、代謝ネットワークのもう一方の柱である形成モデルについて研究した論文は殆どない。古典的なネットワークの研究から得られたモデルの概念を利用して代謝ネットワークの形成モデルを示した研究[2]が報告されているが、生物学的な根拠に基づいたモデルではない。

本研究では、生物学的な考察から代謝ネットワークの形成モデルを提案することで、代謝進化に対して示唆を与えることを目的とする。代謝ネットワークの進

化を、3つの段階に分けて捉え、各段階における形成モデルを提案する。また、提案した形成モデルの妥当性を実際の代謝ネットワークの特徴量との比較によって検討する。この際、既存研究で利用されている特徴量の他に、新たに媒介中心性分布を利用してネットワークを特徴付ける。また、既存の代謝ネットワーク解析では、代謝ネットワーク全体に対して解析を行っていたのに対し、本研究では、1次代謝ネットワークと2次代謝ネットワークに分けて解析を行う。以下、2章で代謝ネットワークの形成モデルを提案し、3章で実際の代謝ネットワークと古典的なネットワークモデル、代謝ネットワークモデルの特徴量を比較する。

2 代謝ネットワークの形成モデルの提案

先行研究によってモデル化手法が提案されていないため、代謝ネットワークのモデル化を以下の2つの観点からモデル化することを提案する。

- 代謝ネットワークが持つ有機化学的な制約
- ネットワークの形成ルール(どのようにリンクが張られるか)

代謝ネットワークは、生体内に構築された有機化合物による化学反応ネットワークである。つまり、代謝ネットワークは化学反応ネットワークのサブネットワークであり、化学反応ネットワークが持つ特徴を継承していると考えられる。そこで、代謝ネットワークが形成される際の制約として有機化学的な制約条件を形成モデルに組み込む。

有機化学的な制約条件の下、何らかの形成ルールに従って代謝ネットワークは形成されたと考えられる。ネットワークの形成ルールとは、どのようにノードやリンクが張られるのかを決める方法であり、代謝ネットワークの進化と深い関わりがあると推察できる。そこで、代謝ネットワークの進化を以下の3つの段階に分けて考えた。

- 原始代謝ネットワークの形成

- 1次代謝ネットワークの形成
- 2次代謝ネットワークの形成

この章では、有機化学的な制約条件とネットワークの形成ルールをモデル化し、原始代謝、1次代謝、2次代謝の形成モデルを提案する。

2.1 有機化学的な制約

有機化学的な背景

代謝ネットワークが化学反応ネットワークの特徴を継承していることは既に述べた。有機化合物は、反応性に富む官能基と呼ばれる部分構造と反応性の乏しい骨格構造を持つ。これらの構造はそれぞれ局所的な構造と大局的な構造と言い換えることができる。有機化学的な反応は、分子間、分子内の局所的な構造が相互作用することによって引き起こされる。反応によって、局所的な構造、または大局的な構造の変化が起こるが、極端な大局的な構造の変化が起こることは稀である。例えば、直鎖化合物か環状化合物へと変換される反応は限られている。この事実は、有機化合物がその大局的な構造によってグループ化されていることを示唆している。同じ大局的な構造を持つ化合物間の変換は比較的自由に行われるのに対し、異なる大局的な構造を持つ化合物間の変換は限られた反応を介してのみ実現される。Fig2.1Aは、実際の生体内反応の一部分を抜粋したものである。芳香族化合物から直鎖状化合物への分解過程がカテコールを介して行われていることが観察できる。

一方、複雑な構造を持つ生体内化合物に注目すると、複雑な構造を持つ化合物の大局的な構造は、より複雑でない化合物が持つ大局的な構造の組み合わせによって構成されていることが分かる。例えば、ナフタレン環はベンゼン環が2つ組み合わさった構造である。Fig2.1Bは、大局的な構造間の関係を模式的に表したものである。大局的な構造間の関係は、より単純な構造を親に持ち、親の構造を部分的に持つ構造が子を持つような階層的な構造を形成していることが見て取れる。

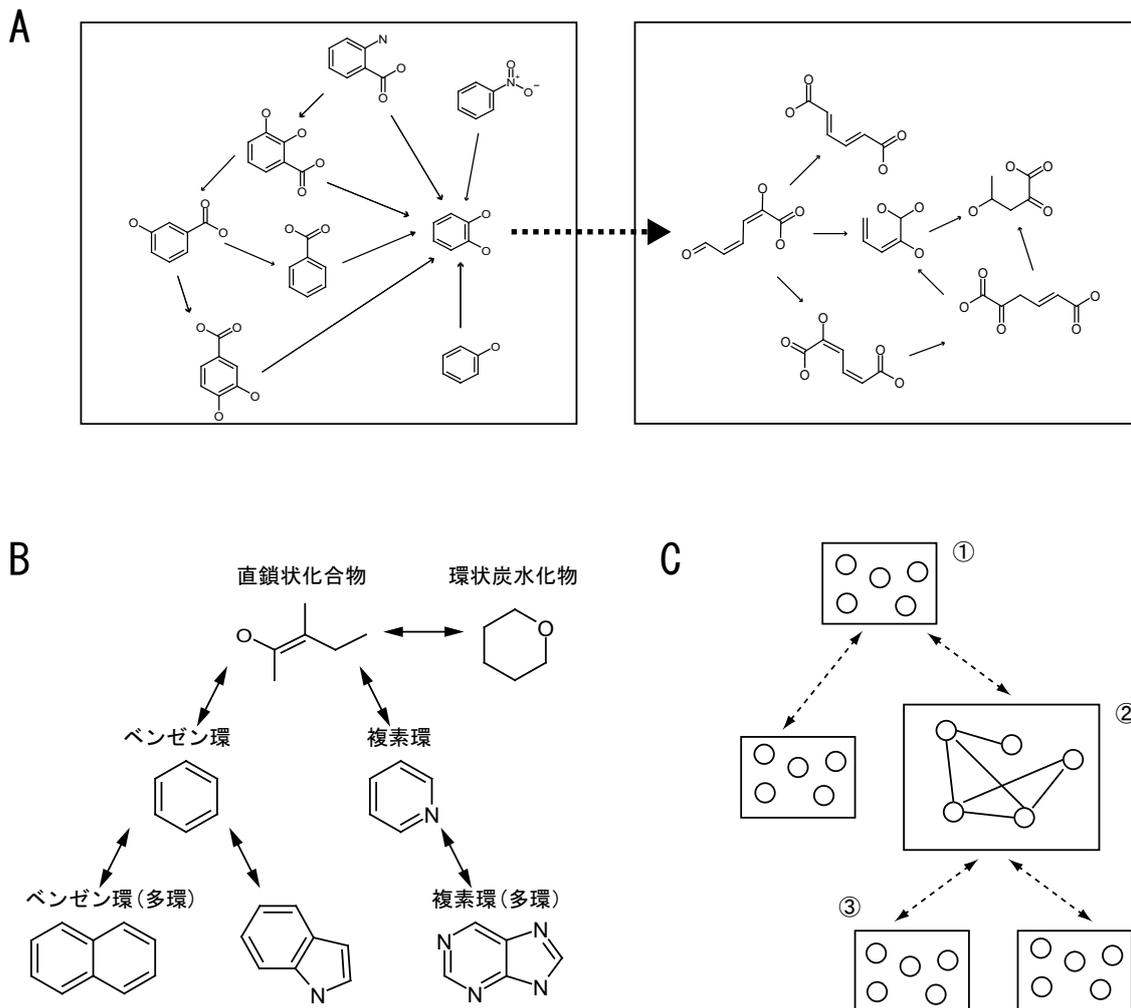


Fig. 2.1 大局的な構造と反応：(A) 生体内での化学反応の例、左のボックス内にはベンゼン環を持つ化合物、右のボックス内には直鎖状の化合物が含まれている。破線矢印はボックス間の反応を表しており、カテコールを解して直鎖状の化合物と繋がっていることが分かる。(B) 大局的な構造間関係 (C) 大局的な構造間関係をモデル化したもの

階層モデルによる制約表現

以上の事柄から、生体内化合物は大局的な構造によってグループ化が可能で、大局的な構造が持つ階層的な関係性から、グループ間関係、つまり、構造的な類似性は階層構造によって表現できる。また、離れたグループ間の生体化合物の変換反応は稀であると考えられる。そこで、代謝ネットワークのノードに対して Fig2.1C のような2分木構造を仮定し、有機化学的な制約を以下のようにモデル化する。

- 各ノードはそれぞれグループに属する
- グループ間関係を2分木で表現する

- ノード i, j 間にリンクが張られる確率を以下の式で与える

$$P_{ij} = p^{l_{ij}} \quad (2.1)$$

グループ間の距離 (Step 数) は、大局的な構造の類似性を表現しており、Step 数が小さい程、類似性が高いとする。従って、Step 数が小さい程、グループ間のリンク確率は高いと考えられ、ノード i, j 間のリンク確率 P_{ij} を式 2.1 で与える。この時、 p は、 $(0 < p < 1)$ となる定数、 g_i, g_j は、 i, j が属するグループ、 l_{ij} はグループ g_i, g_j 間の距離 (Step 数) とする。例えば、Fig2.1C のグループ①に含まれるノードからグループ

```

• initialize(){ //初期化
  M; ネットワークに含まれるノード集合 ( $|M| = n$ )
   $k_r \leftarrow$  ネットワークに張られるリンク数
   $g \leftarrow$  ノードが属するグループ数
}

• makeTree(){ //ノードの属するグループ構造の生成
  各ノードをグループに振り分ける
   $g$  個のノードを持つ2分木を生成し、木のノードをグループと対応させる
}

• main(){ //メイン
  initialize()
  makeTree()
  while( $k < k_r$ ) //リンクの生成
    ノード  $i, j$  をランダムに選択 ( $i, j \in M, i \neq j$ )
     $i, j$  が含まれるグループ間の距離  $l_{ij}$  を計算する
     $probability \leftarrow$  乱数 ( $0 < probability < 1$ )
    if ( $probability < P_{ij}$  (式 2.1)) then
       $i, j$  間にリンクを張る
       $k++$ 
  endwhile
  生成されたネットワークを保存
}

```

Alg.1: 階層モデル (有機化学的な制約+ランダム選択) によるネットワーク生成アルゴリズム

③に含まれるのノードへのリンク確率 P_{ij} は $l_{ij} = 2$ であるため p^2 となる。また、グループ①とグループ②に含まれるノード間のリンク確率とグループ②とグループ③に含まれるノード間のリンク確率は共に p となる。この制約を以下、階層性による制約と呼ぶ。階層性による制約下で、形成ルールをランダムとして作成した形成モデル (階層モデル) を Alg.1 に示す。階層モデル ($n = 1000$, $g = 2^4 - 1$, $p = 0.1$) によって作成したネットワークを Fig2.2 に示す。根 (階層 1) に位置するグループのノードを中心に各グループのノードが階層的に分布している様子が見て取れる。

2.2 ネットワークの形成ルール

代謝ネットワークの進化

古典的な生物学的な分類に従うと、代謝経路は1次代謝と2次代謝の2つに分けて考えられる。1次代謝は、エネルギー代謝やアミノ酸、核酸などの生合成など基本的な代謝ネットワークから構成される。一方、2次代謝は1次代謝から派生した代謝ネットワークを指す。一般的に2次代謝ネットワークによって合成される生体化合物は特殊な化学構造を持ち、多くの場合、ある種の生物が限定して持つネットワークである。1次代謝ネットワークが生体の維持、2次代謝ネットワークが生体防御機構や誘引作用などを担っていると考え

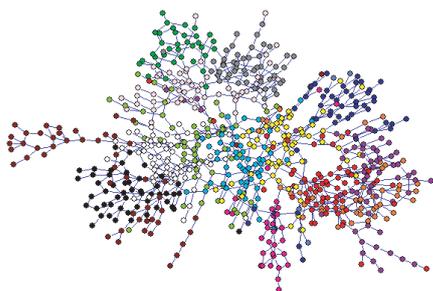


Fig. 2.2 階層モデルによるネットワーク (n=1000, 階層数=4):グループによってノードを色分けしている (階層 1=水色)

られている。これらの事柄から、1次代謝ネットワークと2次代謝ネットワークでは経路の果たす役割 (1次代謝ネットワークは生存に必須となる機能、2次代謝ネットワークは生存戦略に関わる機能) や形成時期が大きく異なっていると考えられる。

原始的な生物が出現した当時、原始的な環境 (原始スープ内) は生物と同じ組成であったと考えられている。従って、生物は構成物質を環境から取り込むだけで生体を維持することができた。しかしながら、時間経過と共に環境の組成が変化し、生物は構成物質を直接環境から取り込むことができなくなると、構成物質の前駆体を変換することで目的の構成物質を獲得する変換ネットワークの必要性が生じた。このとき利用されたのが原始的な代謝ネットワーク (原始代謝ネットワーク) である。環境中の前駆体が不足するに従い、原始代謝ネットワークは徐々に変化し、最終的には、現在のように限られた栄養源から構成物質へ変換する複雑な代謝ネットワーク (1次代謝ネットワーク) が形成されたと考えられる。また、複雑なネットワークにより、安定して構成物質を確保できるようになると、付加的な化合物を作り出すネットワーク (2次代謝ネットワーク) が誕生したと考えられる。このように代謝ネットワークの進化は3段階に分けて考えることができる。

以下、各代謝ネットワークの形成ルールをモデル化し、有機化学的な制約と組み合わせることで形成モデルを提案する。

原始代謝ネットワークの形成モデル

原始的な環境下では、原始生物の必要とする構成物質は、原始スープ内に含まれている化合物そのものであったことは既に述べた。原始代謝ネットワークは環境の変化によって膜内の化合物と膜外の化合物の組成に差が生じたことで形成されたと考えられる。この時、膜外から取り入れることができなくなった化合物を前駆体から直接変換することで原始生物は構成物質を補給していたと考えられる。原始ネットワークが形成された状況を以上のように仮定すると、原始代謝ネットワークは、有機化学的に可能な反応経路をランダムに選択して形成されるネットワークと捉えることができる。その理由として、

- 外界から取り込むことのできる他の構成物質が豊富にあったため、どのような経路でも要求を満たせた
- 生体化合物は全て平等であり、重要な化合物が存在していなかった
- セントラルドグマが確立されておらず、ネットワークの担い手である酵素はランダムに繋がれたポリペプチド鎖として存在していた

と考えられるからである。従って、原始代謝ネットワークは、有機化学的な制約とランダムな形成ルールとしてモデル化することができる。この形成モデルは、Alg.1で既に定義している階層モデルと同じネットワークである。このモデルを以下、原始代謝モデルと呼ぶ。

1次代謝ネットワークの形成モデル

1次代謝ネットワークは原始代謝ネットワークによって形成された物質変換ネットワークを継承し、限られた供給化合物から生体の構成物質を効率よく獲得するために形成されたと考えるのが自然である。この時期になるとセントラルドグマが確立し、遺伝子にコードされた酵素によって代謝ネットワークが運営されるようになったと考えられる。遺伝子は、自らのコピーを作ることで、既存の酵素を保存しながら新しい反応を触媒する酵素を獲得することを可能とした (遺伝子重複)。コピー先の遺伝子とコピー元の遺伝子の非対称

な変異によって、新しい機能を持った酵素が誕生し、代謝ネットワークの構造が変化したと考えられる。遺伝子重複を階層的な制約下で、以下のようにモデル化する。

- 原始代謝ネットワーク上のノードをランダムに選択 (Figure2.3 A)
- 選択されたノードを中心としたサブネットワークをコピー (Figure2.3 B)
- コピーしたサブネットワークの各ノードを確率 p_d で変更 (Figure2.3)
- 変更されたサブネットワークを元のネットワーク上に重ね合わせる (Figure2.3)

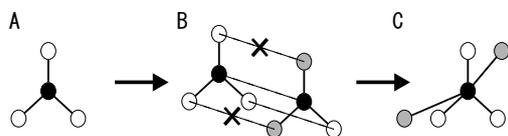


Fig. 2.3 遺伝子重複による形成モデル (1次代謝モデル): 選択されたノード (黒)、変更されたノード (灰色)

ノードを中心としたサブネットワークをコピーした際、コピー先のサブネットワークのノードを一定の確率 p_d で変更する。これは、変異によって酵素の認識部位または、反応部位が変化したことを意味する。Figure2.3

Bで示した灰色のノードが変更されたノードである。その後、変更前のネットワークに変更後のネットワークを重ね合わせることで新しいネットワークが作られている。1次代謝ネットワークの形成モデル (以下、1次代謝モデル) を Alg.2 に示す。

2次代謝の形成モデル

これまでの形成モデルでは、ノードが既存で、形成ルールによってリンクの張り方が決定された。しかし、2次代謝ネットワークは、生体の維持に直接関係のない2次的な産物を生成するネットワークであり、1次代謝ネットワークから分岐して形成されたネットワークである。従って、ノードが既存ではなく、新しいノードへ至るリンクを次々に張ることで、生体内で何らか

の役割を果たす化合物となるノードまでのパスを生成するネットワークであると考えられる。つまり、1次代謝ネットワーク上に存在する既存ノードを出発点として、生体内で有用な化合物を探索する過程で形成されたネットワークと考えられる。また、探索の過程で多くの意味を持たない化合物へ至る経路が発生する。これらの経路は時間と共に除去または機能しなくなると考えられる。これらの仮定から、2次代謝ネットワークの形成モデルを、新規ノードの探索と有用ノード以外へのパスの淘汰によって以下のようにモデル化する。

- 既存ノードから出発 (Figure2.4 A)
- 探索 $\times k_q$ (Figure2.4 B)
 - 既存ノードから新規ノードへのリンクの形成
- 淘汰
 - 新規ノードの中から有用なノードを選択 (Figure2.4 C)
 - 有用なノードへのパス以外の新規パスは淘汰 (Figure2.4 D)

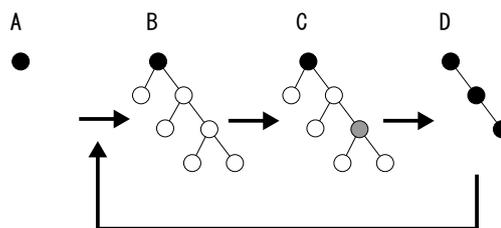


Fig. 2.4 探索と選択による形成モデル (2次代謝モデル)

```

• initialize(){
   $k_d \leftarrow$  サブネットワークのコピーを行う回数
   $p_d \leftarrow$  ノードが置き換わる確率 }

• make_primitive_network(){ //原始ネットワークの生成
  Alg.1.main() }

• copy_subnetwork(){ // サブネットワークのコピー
   $M_i = \{\emptyset\}$ 
  ノード  $i$  をランダムに選択 ( $i \in M$ )
   $i$  を中心とした距離  $r$  までのサブネットワークを選択
  サブネットワークに含まれるノードを  $M_i$  に追加
   $probability\_1 \leftarrow$  乱数 ( $0 < probability\_1 < 1$ )
  for( $\forall j, j \in M_i$ )
    if( $probability\_1 < P_d$ ) then
      while(true)
        ランダムにノード  $k$  を選択 ( $k \in M$ )
         $i, k$  が含まれるグループ間の距離  $l_{ik}$  を計算する
         $probability\_2 \leftarrow$  乱数 ( $0 < probability\_2 < 1$ )
        if( $probability\_2 < P_{ij}$ ) then
           $j$  を  $k$  に置き換える
          break
      endwhile
    endifor
  サブネットワーク上で新たに生成したリンクを元のネットワークに追加
  ネットワークの更新 }

• main(){
  initialize()
  make_primitive_network()
  while( $k < k_d$ )
    copy_subnetwork()
     $k++$ 
  endwhile
  ネットワークの保存 }

```

Alg.2; 1次代謝モデルによるネットワーク生成アルゴリズム

```

• initialize(){
   $M_f$ ; ネットワークに含まれるノード集合 ( $M_f = \{\emptyset\}$ )
   $k_q \leftarrow$  探索を行う回数、 $k_s \leftarrow$  選択を行う回数、 $p_s \leftarrow$  ノードが有用である確率
  初期ノードを  $M_f$  に追加し、ノード数 =1 のネットワークを作成
  makeTree() }

• search(){ //探索
  while(true)
    ノード  $i$  をランダムに選択 ( $i \in M_t$ )
    新規ノード  $j$  の属するグループをランダムに選択 ( $j \notin M_f$ )
     $probability\_1 \leftarrow$  乱数 ( $0 < probability\_1 < 1$ )
    if( $probability\_1 < P_{ij}$ )
       $i$  から  $j$  へ一時ネットワーク上でリンクを張る break
  endwhile
   $M_t, M_n$  に  $j$  追加 }

• select(){ //淘汰
  for( $\forall i, i \in M_n$ )
     $probability\_2 \leftarrow$  乱数 ( $0 < probability\_2 < 1$ )、 $M_p = \{i\}$ 、 $target \leftarrow i$ 
    if( $probability\_2 < p_s$ )
      while(true)
        一時ネットワーク上で  $target$  へのリンクを持つノード集合の中からノード  $j$  をランダムに選択
        if( $j \notin M_p$ )  $target \leftarrow j$ 
        if( $target \in M_f$ )
           $target$  から  $i$  へと至る経路をネットワークに追加
           $M_p$  を  $M_f$  へ追加、break
        else  $target$  を  $M_p$  へ追加
      endwhile
    endif }
  endfor }

• main(){
  initialize(), makeTree()
  for( $k < k_s$ )
     $M_t \leftarrow M_f$ 、 $M_n = \{\emptyset\}$ 、一時ネットワーク  $\leftarrow$  ネットワーク
    for( $k < k_q$ )
      search()
    endfor
  select()
  ネットワークの更新
  endfor}

```

Alg.3: 2次代謝モデルによるネットワーク生成アルゴリズム

3 代謝ネットワーク解析

提案したモデルの有用性を確認するために、実際のデータから構築した代謝ネットワークの解析を行い、ネットワークの特徴量を比較する。

3.1 従来の代謝ネットワーク解析

Joeng 等の研究によって、グラフ理論に基づいたネットワーク解析手法が提案され、代謝ネットワーク解析に対する新しい側面が示された [1]。

この手法を用いた解析によって WWW のような人工的なネットワークと同様に代謝ネットワークが Scale-Free ネットワークの特徴を持つことが示された [1]。Scale-Free ネットワークでは、殆どのノードが低い次数 (ノードが持つリンク数) を持つのに対し、少数の非常に高い次数を持つノードが存在し、ネットワークの次数分布が冪法則に従う。高い次数を持つノードはハブと呼ばれ、ハブによって殆どのノードが短いステップ数で繋がる [4, 5]。Joeng 等は、43 種の代謝ネットワークに対して、各ノード間の最短経路長の平均 (平均最短距離: AL) と次数分布を計算し、AL が種によらず小さな値 (約 3.2) で、次数分布が冪法則に従うことを示した [1]。彼等の解析では、ATP や ADP といった物質変換の際に共役的に働く化合物を 1 つのノードとして扱っていた。そのため、それらの共役的な化合物がハブとなり AL が非常に小さな値となっていた。Ma 等はこの点を指摘し、物質変換の主役となる化合物のみで代謝ネットワークを再構築し、代謝ネットワークが Scale-Free 特性を持つことを再び示した [6]。一方、Ravasz らはクラスタリング係数を利用して代謝ネットワークが階層的なモジュール構造を持つことを示した [2]。

従来の代謝ネットワーク解析では、代謝ネットワーク全体に対して特徴量を算出している。本研究では、代謝ネットワークの進化的な背景から、3 種類の形成モデルを提案した。提案した各形成モデルと対応する天然の代謝ネットワークの特徴量を比較することで、提案したモデルの妥当性を検討する。ただし、原始代謝モデルに対応する天然の代謝ネットワークを得るこ

とが不可能であるため、1 次、2 次代謝ネットワークのみの解析に留まる。

3.2 代謝ネットワークの準備

生物の代謝ネットワークを構築するためには生体化合物や生体内反応、種のゲノム情報など多くの情報が必要となる。これらの情報を公開しているデータベースとして WIT[7] や EcoCyc[8]、KEGG[9] が有名である。本研究では、KEGG データベースを利用して、1 次代謝ネットワークと 2 次代謝ネットワークを作成した。原始代謝は、1 次代謝によって上書きされているため、現在、生物から構築することができるのは 1 次代謝ネットワークと 2 次代謝ネットワークのみである。

まず、KEGG/LIGAND データベース上の生体内化合物と生体内反応のデータから、生体内化合物の隣接行列を作成した。この隣接行列は KEGG に含まれる全生物種の可能性のある生体内反応のネットワーク (reference ネットワーク) を表現している。KEGG/PATHWAY データベースから、生物の代謝経路を生物学的な知見に基づいて分類したデータが得られる。このデータを利用して reference ネットワークから 1 次代謝ネットワーク (1 次 reference ネットワーク) と 2 次代謝ネットワークを取り出した。1 次代謝ネットワークは、多くの種で共通したネットワークであるが、1 次 reference ネットワークは、複数の種のネットワークの重ね合わせによるネットワークである。そのようなネットワークでは、実際の代謝ネットワークのトポロジーが再現できない。そこで、KEGG/GENOME データベースを利用して、1 次 reference ネットワークから 35 種の生物種の代謝ネットワークを作成した。一方、2 次代謝は種特有のネットワークであり、上記で挙げたような重ね合わせは見られなかった。

3.3 代謝解析

得られた 1 次、2 次代謝ネットワークと既存のネットワークモデル、本研究で提案した形成モデルの 3 種類の特徴量によって比較する。本研究では、先行研究で利用されている特徴量に加え、媒介中心性を利用し

て特徴付けた。

既存のネットワークモデル（以下、典型モデル）として

- ランダムネットワーク [10]
- Scale-Free ネットワーク ($m_0=3, \alpha=0.6$)[11]

を利用する。また、提案した形成モデルは原始代謝モデル、1次代謝モデル、2次代謝モデルである

ネットワーク比較の際、ノードの数を基準とする。1次代謝ネットワークの平均 $n = 735$ 、2次代謝ネットワークの平均 $n = 105$ に合わせ、典型モデルと形成モデルをそれぞれ作成した。形成モデルのパラメーターは以下の通りである。

共通パラメーター： $p = 0.08, g = 2^4 - 1$

1次代謝モデルのパラメーター： $p_d = 0.3$

2次代謝のパラメーター： $p_s = 0.01$

媒介中心性

ネットワークは、互いに連結することで関係を形成している。ネットワーク内部に複数のサブネットワークが存在し、サブネットワーク内のノードが特定のノードを経由して他のサブネットワークと連結しているような場合、この特定のノードを切断点 (cutpoint) と呼ぶ。媒介中心性 (betweenness centrality : Cb) は、切断点に代表されるように、ノード間の連結構造上で重要な意味を持つノードを特徴付けるための指標である [12]。ノード i の媒介中心性 $Cb(i)$ は、ネットワーク上に存在するノードペアの最短経路がどの程度ノード i を通過しているかの比で表現され、式 3.1 で定義される。

$$Cb(i) = \frac{\sum_{j < k} g_{jk}(i) / g_{jk}}{\frac{1}{2}N(N-1)} \quad (3.1)$$

この時、 $g_{jk}(i)$ は、 i を通る j, k 間の最短経路の数、 g_{jk} は j, k 間の最短経路の総数、 N は、ネットワークの総ノード数である。

Figure3.1 は、媒介中心性分布を片対数グラフにプロットしたものである。Figure3.1:A、B はそれぞれ1次代謝、2次代謝ネットワークと典型モデルを比較したものである。この結果から、1次、2次代謝ネットワーク共に典型モデルに比べて大きな媒介中心性を持

つノードが段階的に含まれていることが分かる。一方、Figure3.1:C、D にそれぞれ1次、2次代謝モデルと1次、2次代謝ネットワークの比較を示した。1次代謝モデルでは、多少媒介中心性分布が代謝ネットワークよりも大きくなっているが、1次、2次代謝モデル共に天然のネットワークと分布が一致していることが分かる。

度数分布

Figure3.2 に、代謝ネットワーク (左)、典型的モデル (中)、形成モデル (右) の度数分布を両対数グラフにプロットした。1次代謝ネットワークの度数分布が Scale-Free 特性に従うことが分かる。一方、1次代謝モデルの度数分布は、ランダムグラフで見られる分布と冪分布との中間的な分布をしているものの、次数の高いハブの存在が冪分布と同様に見て取れる。

AL、Diameter、クラスタリング係数

- Diameter

各ノード間の最短経路長の内、最大の値

- クラスタリング係数

あるノードと繋がった他のノードが、どの程度密に連結しているかを示す量

代謝ネットワークの特徴を表現する特徴量として AL、ネットワークの直径 (Diameter)、クラスタリング係数 (C) を算出した (Table3.1)。代謝ネットワークに共通する特徴として AL、Diameter の値が典型モデルよりも大きいことが分かる。また、1次代謝ネットワークと2次代謝ネットワークそれぞれの特徴として、1次代謝ネットワークではクラスタリング係数が高く、2次代謝ネットワークは AL、Diameter の値が特に大きいことが分かる。

一方、形成モデルを見てみると、1次代謝モデルと1次代謝ネットワークはよく一致するのに対し、2次代謝モデルは、AL、Diameter 共に典型モデルに比べて大きな値をとるが、天然の2次代謝ネットワークのものよりかなり小さな値となっていることが分かる。

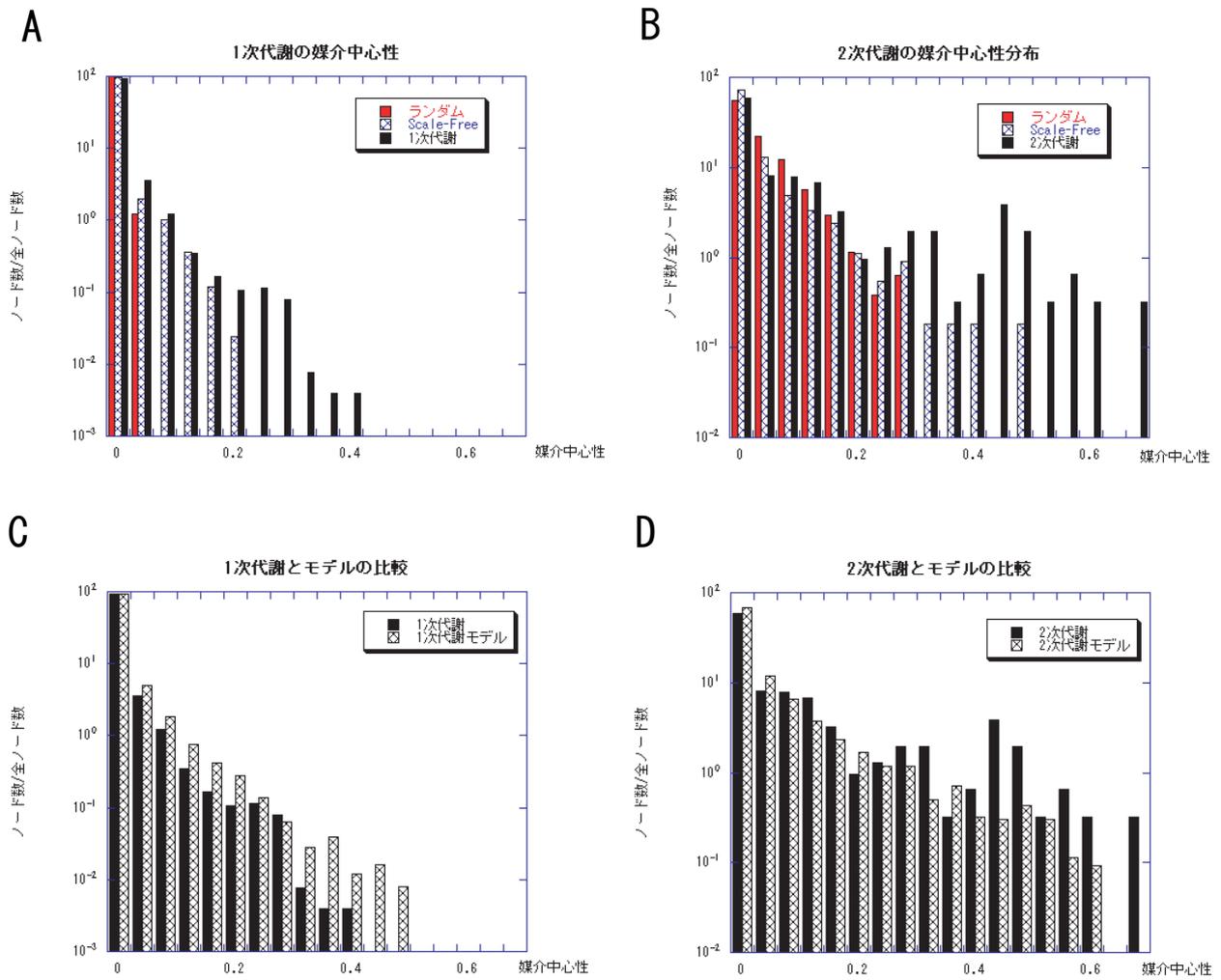


Fig. 3.1 媒介中心性分布 : (A)1次代謝ネットワークと典型モデル (B)2次代謝と典型モデル (C)1次代謝モデルと1次代謝ネットワーク (D) 2次代謝モデルと2次代謝ネットワーク

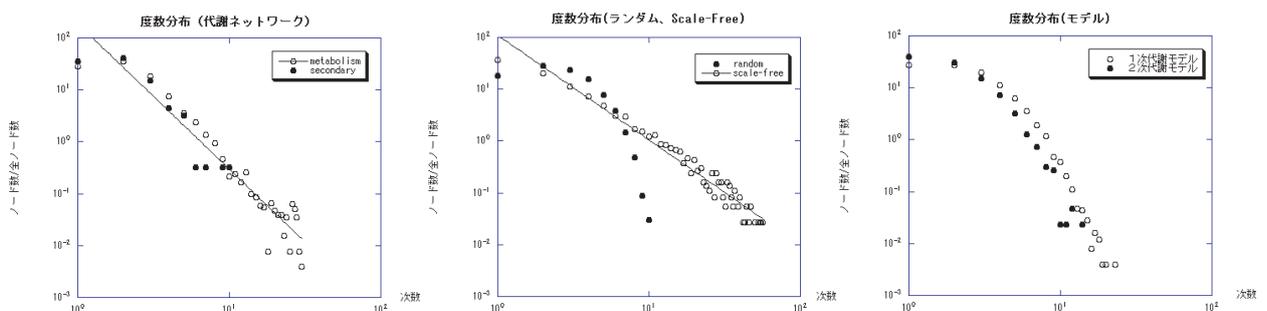


Fig. 3.2 次数分布 : 代謝ネットワーク (左)、典型モデル (中央)、形成モデル (右)

Network	N	AL	Diameter	C
ランダム	105	5.2	11.9	2.2×10^{-3}
	735	6.6	15.6	2.6×10^{-4}
Scale-Free	105	3.0	6.8	2.0×10^{-2}
	735	3.8	9.8	2.4×10^{-2}
Secondary Primary	105	10.4	30.0	6.2×10^{-3}
	735	8.2	26.6	2.8×10^{-2}
model sec	105	7.8	19.7	3.6×10^{-3}
model pri	735	10.7	27.8	1.3×10^{-2}
model hie	735	13.2	33.7	0.7×10^{-2}

Table 3.1 ネットワークの特徴量：典型モデルと代謝ネットワーク、形成モデルの比較 (model_hie=階層モデル、model_pri=1次代謝モデル、model_sec=2次代謝モデル)

階層モデルの解析

形成モデルが示す特徴が有機化学的な制約に基づくものか、形成ルールに基づくものかを調べるために、原始代謝モデル ($n = 735$) に対しても同様のネットワーク解析を行った (Figure 3.3, 3.4)。大きな媒介中心性を持つノードが段階的に存在しており、次数分布がランダムネットワークと同じ分布であることが分かる。

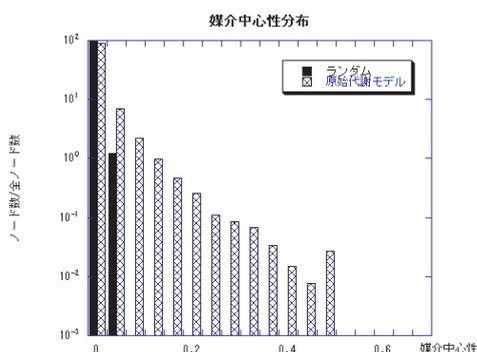


Fig. 3.3 媒介中心性分布:階層モデル

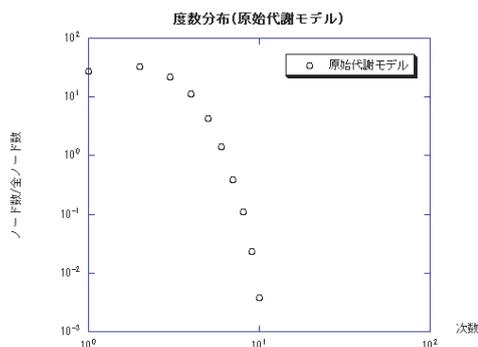


Fig. 3.4 次数分布:階層モデル

4 考察

代謝ネットワークの特徴について

天然の代謝ネットワークを1次、2次代謝ネットワークに分類してネットワーク解析を行ったところ、天然の1次代謝ネットワークにおいて明らかな Scale-Free 性が観察されたのに対し、2次代謝ネットワークでは観察されなかった (Figure 3.2)。一般的に、AL、Diameter の値は、ネットワークのノード数に従って大きくなる傾向があるにも関わらず、2次代謝ネットワークでの値は、1次代謝ネットワークでの値より大きくなっていることが分かる。この逆転現象は、典型モデルの AL、Diameter の値とを比較することによって更に強調される (Table 3.1)。また、クラスタリング係数に注目すると、天然の1次代謝ネットワークでは高い値をとっており、クラスタリングされたネットワークであることを示唆している (Table 3.1)。これら事柄から、天然の1次代謝ネットワークと2次代謝ネットワークのネットワークの特徴は異なっていると結論付けられる。つまり、この2つのネットワークの構造が異なること示唆しており、ネットワークの形成ルールが異なっているという我々の仮説を強く支持していると言える。

一方で、天然の代謝ネットワークで共通した特徴も観察されている。1つは、AL、Diameter の値が典型モデルの値に比べて共通して大きいことである (Table 3.1)。もう1つは、媒介中心性分布を観察した時に媒介中心性が高いノードの存在が認められる点である。ランダムネットワークでは、殆どが極めて小さな媒介中心性

を持ち、大きな値を持つノードは存在しない。一方、天然の代謝ネットワークでは、1次、2次に限らず、大きな媒介中心性を持つノードが存在しており、その割合が段階的に減少している。このことは、ネットワークに切断点が存在し、ネットワークの段階的な切断点によってネットワークが階層的に組織されている可能性を示唆している (Figure3.1)。

形成モデルについて

本研究では、代謝ネットワークのモデル化を有機化学的な制約と形成ルールという視点から行った。前節で挙げた代謝ネットワークに共通する特徴を1次代謝モデルと2次代謝モデルともによく再現しているといえる。最も単純な階層モデルですらこの特徴を備えている (Figure3.3, 3.4, Table3.1)。従って、この結果に最も貢献した要素は有機化学的な制約をモデル化した点にあると考えられる。つまり、天然の代謝ネットワークが有機化学的な制約によって階層性や広がりを持っていることをこの結果は示唆している。一方、Scale-Free 特性はネットワークの形成ルールによって獲得されるものであり、特に遺伝子重複によって獲得された可能性が示唆されている。

5 まとめと今後の課題

本研究では、これまであまり扱われていなかった代謝ネットワークの形成モデルを提案することを目的とした。また、形成モデルの妥当性を検討するために天然の1次、2次代謝ネットワークと形成モデルを特徴付けて比較した。形成モデルでは生体内において重要な機能を果たす分子や機能モジュールという前提を置かず、原始代謝ネットワークにおいては有機化学的な関係のみが支配するネットワークであると仮定した。そもそも、重要な機能を果たす分子とは何なのか、機能モジュールとは何なのかといった疑問に答えることはできない。しかしながら、代謝ネットワークの進化の過程で、偶然、切断点の位置にあった分子がそのネットワーク構造によって重要な位置付けを得るに至ったのではないかとこの研究から導き出される。

また、機能モジュールも同様に化学構造が似た化合物のモジュール構造がその原型であったのではないかと考えることができる。つまり、生物においては、偶然獲得したネットワーク構造を活かす形でその利用法が決定されたと考える方が自然であるのではないだろうか。本研究の意図するところは、まさにその点にある。既存の研究で当然のように扱われている化合物の特性とは生物の自己組織化によって創発された後天的な産物に過ぎない可能性を示唆している。

また、生物における代謝ネットワークの形成モデルは、人間関係や、会社内の組織におけるネットワークの形成過程に対する示唆を与えてくれる。今回モデルに取り入れた階層性の制約は、会社内における役割に置き換えることができる。組織内のネットワークを改善するにはどのような要因を加えればよいのか、ネットワークがどのように変化するのかといった実社会へのフィードバックが可能となると、この研究の意義が高まると考えられる。

今後の課題として、有機化学的な制約を表現した階層構造と、より詳細なネットワーク解析が挙げられる。本研究では、有機化学的な制約を便宜上2分木でモデル化したが、反応ネットワークの階層構造が2分木である根拠はない。それどころか、実際には2分木のような平等な構造ではなく、もっと偏った構造であることは容易に想像がつく。従って、反応ネットワークの階層構造は、どのようなものが最適であるかを考える必要がある。また、ネットワーク解析についてもより詳細な解析が求められる。例えば、大きな次数を持つノードと大きな媒介中心性を持つノードに相関が認められるかどうか、もし相関が認められるならば形成モデルがその関係を再現しているかどうかを確認する必要があると考えられる。

参考文献

- [1] Joeng H., Tombor B., Albert R., Oltvai Z.N., and Barabasi A.L. The large-scale organization of metabolic networks. *Nature*, Vol. 407, pp. 651–654, 2000.

- [2] E.Ravasz, A.L.Somera, D.A. Mongru, Z.N.Oltvai, and A.L.Barabasi. Hierarchical organization of modularity in metabolic networks. *Science*, Vol. 297, No. 30, pp. 1551–1555, 2002.
- [3] Petter Holme, Mikael Huss, and Howoong Joeng. Subnetwork hierarchies of biochemical pathways. *Bioinformatics*, Vol. 19, pp. 532–538, 2003.
- [4] Albert-Laszlo Barabasi and Reka Albert. Emergence of scaling in random networks. *Science*, Vol. 286, , 1999.
- [5] Reka Albert and Albert-Laszlo Barabasi. Topology of evolving networks: local events and universality. *Physical Review Letters*, Vol. 85, pp. 5234–5237, 2000.
- [6] Hongwu Ma and An-Ping Zeng. Reconstruction of metabolic networks from genome data and analysis of their global structure for various organisms. *Bioinformatics*, Vol. 19, pp. 270–277, 2003.
- [7] Overbeek R., Larsen N., Pusch G.D., D’Souza M., E. Selkov, Jr., Kyrpides N., Fonstein M., Maltsev N., and Selkov E. Wit: integrated system for high-throughput genome sequence analysis and metabolic reconstruction. *Nucleic Acids Res*, Vol. 28, pp. 123–125, 2000.
- [8] Karp P.D., Riley M., Saier M., Paulsen I.T., Collado-Vides J., Paley S.M., Pellegrini-Toole A., Bonavides C., and Gama-Castro S. The ecocyc database. *Nucleic Acid Res*, Vol. 30, pp. 56–58, 2002.
- [9] Goto S., Bono H., Fujibuchi W., Nishioka T., Sato K., and Kanehisa M. Organizing and computing metabolic pathway data in terms of binary relations. *Pacific Symposium on BioComputing '97*, pp. 175–186, 1997.
- [10] Reka Albert and Albert-Laszlo Barabasi. Statistical mechanics of complex networks. *Reviews of Modern Physics*, Vol. 74, pp. 47–97, 2002.
- [11] *Winner don't take all: Characterizing the competition for links on the web*, 2002.
- [12] Freeman and Linton C. Centrality in social networks: I a conceptual clarification. *Social Networks*, Vol. 1, pp. 215–239, 1979.