

## 部分代謝構造を用いた代謝パスウェイの比較解析

芦田 優太<sup>†</sup> 尾崎 知伸<sup>††</sup> 大川 剛直<sup>†</sup>

<sup>†</sup> 神戸大学大学院工学研究科 <sup>††</sup> 神戸大学自然科学系先端融合研究環

代謝パスウェイは生物の機能に関する情報を含んでおり、近年様々な解析手法が提案されている。本研究では、代謝パスウェイ上の重要な酵素である choke point や、代謝の定常状態を実現する最小の部分構造である elementary flux mode を用い、生物にとって重要な部分代謝構造に着目した代謝パスウェイの比較解析を行う。古細菌、真正細菌、真核生物の3つのドメインに含まれる生物に対して上記手法を適用し、部分構造の比較、距離算出を行う。その後、系統樹を構築し、その結果に関して考察した。

## A Comparative Analysis of Metabolic Pathways based on Reaction Substructures

Yuta Ashida<sup>†</sup> Tomonobu Ozaki<sup>††</sup> Takenao Ohkawa<sup>†</sup>

<sup>†</sup> Graduate School of Engineering, Kobe University

<sup>††</sup> Organization of Advanced Science and Technology, Kobe University

We propose a new method for comparing the metabolic pathways with reaction substructures. In this paper, we use subgraphs which include important enzymes, or elementary flux modes which are minimal sets of enzyme reactions that can each generate valid steady states as feature. By applying the proposed method to the metabolic networks of organisms selected from archaea, eubacteria and eukaryote, we succeed in reconstructing a phylogenetic tree, and confirm the effectiveness of the method.

### 1 はじめに

生物は物質を食物等の形で体内に取り入れる。これらの物質は体内の様々な化学変化を通してエネルギーや生命維持に必要な化合物へと変換される。このような化学変化の総体が代謝と呼ばれ、代謝パスウェイとしてグラフデータ化されている。

代謝パスウェイは生物の機能に関する重要な情報を含んでいる。さらに、異種生物種間での代謝パスウェイ比較解析は生物機能における類縁関係についての重要な情報を得る有効な手段となる。

本研究では、代謝パスウェイ上の重要な酵素である choke point<sup>3)</sup> 周辺の代謝構造や、代謝の定常状態を実現する最小の部分構造である elementary flux mode<sup>4)</sup> を用いる。

choke point 周辺の構造を用いる手法<sup>1)</sup> は範囲指定のためのパラメータを必要とするが、この値は必ずしも自明ではない。加えて、choke point を中心としているとはいえ、抽出される部分構造が全て生物学的な重要性を持つとは限らない。そこで、様々な範囲で choke point 周辺の構造を抽出

し、パスウェイ比較解析を行う。

一方、elementary flux mode は生物学的に意味のある構造として抽出される。これによりパラメータ設定を必要とせず、様々な大きさの構造で生物の比較が可能となるため、elementary flux mode を用いたパスウェイ比較手法を提案し、それぞれの手法による違いを比較することを目的とする。

本論文は以下の構成となっている。2章で choke point を用いた代謝パスウェイ比較手法を紹介し、3章で elementary flux mode を用いた比較手法を提案する。そして4章で提案手法による代謝パスウェイ比較解析結果を示し、5章で結論と今後の課題について述べる。

### 2 choke point 周辺の代謝構造を用いた代謝パスウェイ比較解析

本論文では代謝パスウェイ上の重要な酵素を含む部分グラフを反応構造と呼ぶ。反応構造抽出の例を Fig. 1 に示す。本論文では反応構造を、酵素をノードとする無向グラフとして扱う。Fig. 1 中

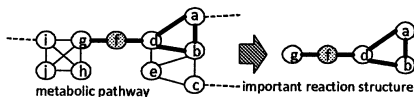


Fig. 1 反応構造の例

の代謝パスウェイは基準点酵素 (ノード f) の周辺を表しており、破線は図中にないノードに接続しているエッジである。太線で接続されている部分が反応構造の一例である。実際には一つの基準点酵素から複数の反応構造が抽出される。

まず、生物種間の距離を求めるために重要な反応構造を抽出し、反応構造プロファイルを作成する。反応構造プロファイルは異なる生物の代謝パスウェイ間でどれだけ重要な反応構造が重複して存在するかを表し、生物種間の距離は反応構造プロファイルを元に Jaccard 係数法を用いて算出される。この操作を生物の全組み合わせについて行い、階層的クラスタリングに用いる距離行列を作成する。

## 2.1 基準点酵素探索

反応構造プロファイル作成のため、まず反応構造抽出の起点となる酵素、すなわち基準点酵素の決定が必要である。基準点酵素決定には Syed らによる 'load point', 'choke point' 探索手法<sup>3)</sup>を用いた。

'load point' は代謝パスウェイ上の全酵素の組み合わせから得られた最短経路が集中している酵素であり、代謝中で頻繁に用いられる酵素を表す。'load point' を決定する 'load value' は隣接している酵素数と自身を通る最短経路の数から算出され、高い 'load value' を持つ酵素が 'load point' となる。'load value' は以下の式で求められる。

$$L_m = \ln \left[ \frac{p_m/k_m}{\sum_{i=1}^M p_i / \sum_{i=1}^M k_i} \right], (-\infty < L_m < \infty) \quad (1)$$

ここで  $M$  は代謝パスウェイ上の酵素数を表し、 $p_m$  は酵素  $m$  を通る最短経路の数を、 $k_m$  は  $m$  と隣接している酵素数を意味する。

一方、'choke point' とはその酵素の欠損が生命の維持に大きな影響を与えるような酵素のことである。ある酵素を通る最短経路に多くの 'load point' が含まれる場合、その酵素を 'choke point' とする。

酵素を通る最短経路上の 'load point' の数によってその酵素の 'choke value' が求められ、高い 'choke value' を持つ酵素が 'choke point' となる。本論文では Pathway Hunter Tool<sup>1</sup>を用い、高い 'choke value' を持つ酵素上位 5 個を 'choke point' とした。

## 2.2 反応構造プロファイルの作成

まず choke point を探索し、その後、基準点酵素から特定範囲内で、ある一定のノード数の反応構造を抽出する。これはノード数が小さすぎれば抽出される反応構造が多くなりすぎ、大きすぎれば反応構造の重要性が薄まるからである。また、基準点酵素からの範囲が広すぎれば基準点酵素から離れすぎてしまい、重要性が薄まってしまう。

生物種  $O$  と  $O'$  から抽出した重要な反応構造の集合をそれぞれ  $G_O$  と  $G_{O'}$  とする。そして、 $G_O$  と  $G_{O'}$  の和集合を  $R\{O, O'\} = G_O \cup G_{O'} = \{r_1, r_2, \dots, r_n\}$  とする。

生物種  $O$  の  $O'$  に対する反応構造プロファイルは、 $R\{O, O'\}$  に含まれる反応構造が  $O$  の代謝パスウェイ上にいくつ存在するかを表し、以下の式で求められる。

$$P_o\{O'\} = [p_{o1} \ p_{o2} \ \dots \ p_{on}] \\ \begin{cases} p_{oi} = 1 & (r_i \leq N_O) \\ p_{oi} = 0 & \text{otherwise} \end{cases} \quad (1 \leq i \leq n)$$

$P_o\{O'\}$  は  $O$  の  $O'$  に対する反応構造プロファイルを表し、 $r_i \leq N_O$  は反応構造  $r_i$  ( $1 \leq i \leq n$ ) が  $O$  の代謝パスウェイ  $N_O$  上に含まれることを意味する。

## 3 elementary flux mode を用いたパスウェイ比較解析

### 3.1 elementary flux mode

生体において代謝流量は動的に変化しているが、定常状態となるパスウェイを抽出することで、生体における長期的観点での代謝の特徴を得ることができる<sup>4)</sup>。elementary flux mode とは定常状態を満たす最低限のパスウェイであり、それ以上分割すると非定常状態になるようなパスウェイである。定常状態における全ての代謝活動は elementary flux mode から成る。代謝パスウェイが  $\mathbf{0} = \mathbf{N}\mathbf{r}$  を満たす時、そのパスウェイは定常状態である。ここで  $\mathbf{N}$  は  $m \times q$  の行列であり  $m$  は化合物の数、 $q$  は酵素反応の数を表す。また、 $\mathbf{r}$  は反応率を表す

<sup>1</sup> <http://pht.tu-bs.de>

$q \times 1$  のベクトル,  $\mathbf{0}$  は  $m \times 1$  の零ベクトルである.

elementary flux mode を用いて代謝パスウェイを比較することで, 定常状態における代謝構造での生物種比較が可能となる. elementary flux mode の算出には CellNetAnalyzer を用いた.

### 3.2 elementary flux mode を用いた比較解析

算出した elementary flux mode を用いて生物種間の距離を算出する. 距離算出には José らの提案した代謝パスウェイ間の類似度を求める pseudo-alignment<sup>2)</sup> を生物種間の類似度を求めるように拡張し, 使用した. 2つの生物種  $X = (\mathbf{P})$  と  $Y = (\mathbf{Q})$  の類似度は以下の式で求められる.

$$\begin{aligned} \text{sim}(X, Y) = & \frac{1}{\|\mathbf{P} \cup \mathbf{Q}\|} \left( \|\mathbf{P} \cap \mathbf{Q}\| + \sum_{P \in \mathbf{P} \setminus \mathbf{Q}} \max_{Q \in \mathbf{Q}} \text{sim}(P, Q) \right) \\ & + \sum_{Q \in \mathbf{Q} \setminus \mathbf{P}} \max_{P \in \mathbf{P}} \text{sim}(P, Q) \end{aligned}$$

ここで  $\mathbf{P}$  と  $\mathbf{Q}$  は elementary flux mode の集合である. 2つの elementary flux mode  $\mathbf{P} = (\mathbf{R})$  と  $\mathbf{Q} = (\mathbf{S})$  の類似度は以下の式で求められる.

$$\begin{aligned} \text{sim}(\mathbf{P}, \mathbf{Q}) = & \frac{1}{\|\mathbf{R} \cup \mathbf{S}\|} \left( \|\mathbf{R} \cap \mathbf{S}\| + \sum_{R \in \mathbf{R} \setminus \mathbf{S}} \max_{S \in \mathbf{S}} \text{sim}(R, S) \right) \\ & + \sum_{S \in \mathbf{S} \setminus \mathbf{R}} \max_{R \in \mathbf{R}} \text{sim}(R, S) \end{aligned}$$

ここで  $\mathbf{R}$  と  $\mathbf{S}$  は酵素反応の集合を表している. また, 酵素反応  $R = (\mathbf{C}, \mathbf{E})$  と  $S = (\mathbf{D}, \mathbf{F})$  の類似度は以下の式で求められる.

$$\begin{aligned} \text{sim}(\mathbf{R}, \mathbf{S}) = & \frac{1-\alpha}{\|\mathbf{C} \cup \mathbf{D}\|} \left( \|\mathbf{C} \cap \mathbf{D}\| + \sum_{C \in \mathbf{C} \setminus \mathbf{D}} \max_{D \in \mathbf{D}} \text{sim}(C, D) \right) \\ & + \sum_{D \in \mathbf{D} \setminus \mathbf{C}} \max_{C \in \mathbf{C}} \text{sim}(C, D) + \frac{\alpha}{\|\mathbf{E} \cup \mathbf{F}\|} \left( \|\mathbf{E} \cap \mathbf{F}\| \right) \\ & + \sum_{E \in \mathbf{E} \setminus \mathbf{F}} \max_{F \in \mathbf{F}} \text{sim}(E, F) + \sum_{F \in \mathbf{F} \setminus \mathbf{E}} \max_{E \in \mathbf{E}} \text{sim}(E, F) \end{aligned}$$

ここで  $\mathbf{C}$ ,  $\mathbf{D}$  は化合物,  $\mathbf{E}$ ,  $\mathbf{F}$  は酵素の集合を表し,  $\alpha (0 \leq \alpha \leq 1)$  は酵素の類似度に対する化合物の類似度の重みを表す. 化合物の類似度は 1 か 0 どちらかの値しかとらない. つまり, 同一であれば 1, そうでなければ 0 となる. 酵素間類似度は hierarchical similarity<sup>5)</sup> を用いた.

## 4 実験

提案手法の有効性を確認するため, 複数の生物種に対して二つの手法により生物種比較を行い系統樹を構築した. rRNA 配列解析による従来の系統分

類学は誤りを含む可能性があるが, 大部分は確立されている. そのため, 本論文では NCBI taxonomy による既存の系統分類に近い系統樹の構築に価値があるとした.

### 4.1 各手法による生物種比較

反応構造抽出を抽出する際, 反応構造のノード数と基準点酵素からのステップ数の設置が必要となる. そこで, ノード数とステップ数を 1~7 の様々な値に設定し, 実験を行った. 実験には archaea, eubacteria, eukaryota の 3つのドメインに属する生物種の中から合計 39種の生物を用いた.

ステップ数, ノード数の変化による傾向として, ノード数は大きくなるにつれてドメインを分類できるようになっていったが, ステップ数は 5 までは大きくなるにつれて NCBI taxonomy による分類に類似していき, 5 より大きくなると類似しなくなっていった. 結果, 基準点酵素からの範囲が 5 ステップ, ノード数が 7 個の設定で NCBI taxonomy による分類と最も近い系統樹が得られることがわかった. 結果を Fig. 2(A) に示す.

Fig. 2(A) では全ての生物種がそれぞれ archaea, eubacteria, eukaryota のクラスタに分かれている. この結果は C. R. Woese による 3ドメイン説を代謝機能の観点から再現しており, NCBI taxonomy に近い分類結果が得られた.

次に, elementary flux mode を用いた生物種比較を行った. 実験には archaea, eubacteria, eukaryota の 3つのドメインに属する生物種の中から合計 8種の生物を用いた. 結果を Fig. 2(B) に示す.

Fig. 2(B) では (A) と同じく全ての生物種がそれぞれのドメインのクラスタに分類されている. また, NCBI taxonomy において近隣である生物種同士がほとんど近接しており, 非常に類似した系統分類結果が得られた.

### 4.2 2手法間の比較

更に, これら二つの手法を比較するため, 同じ 8種の生物種を用いて choke point を含む代謝構造を用いた生物種比較実験を行った. 結果を Fig. 2(C) に示す.

Fig. 2(C) では NCBI taxonomy とは大きく違った結果となった. これは比較に用いる生物種数の減少により, 特徴抽出が困難になったことが理由だと考えられる. また, この実験では Fig. 2(A) と同じパラメータを使用しているが, 使用する生物種数

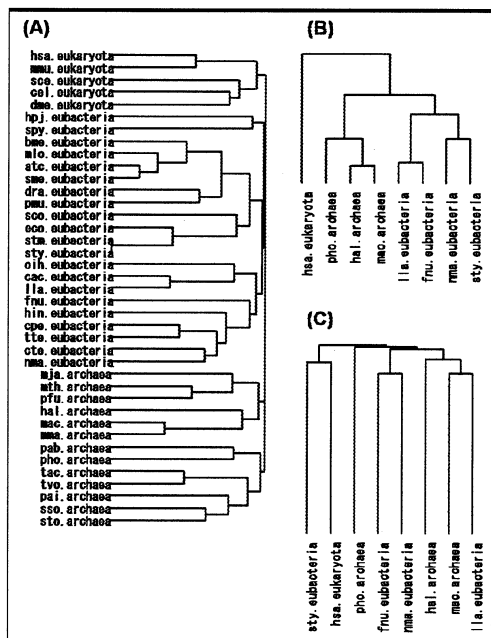


Fig. 2 それぞれの手法を適用した生物種比較による系統樹.

に適したパラメータ設定が必要であることが考えられる. この結果から, choke point 周辺の代謝構造による比較は多数の生物のドメイン分類に有用であるが, 使用する生物種数に応じたパラメータ設定が問題となることが分かる. 一方, elementary flux mode を用いた比較はドメイン以下の詳細な分類に対して有用であることを示唆している.

choke point 周辺の代謝構造を用いる手法では, 生物種比較に用いられる部分構造の大きさが一定である. また, 中心となる choke point は生物学的重要性を持つが, 抽出される周辺構造が必ずしも意味があるとは限らない. 更に, 比較に用いる生物種数に応じて一方, elementary flux mode は様々な大きさの部分構造を用いて比較を行っており, 部分構造全てが意味を持つ. これにより, choke point 周辺の一定の大きさの代謝構造では, 近隣の生物間の比較において特徴抽出が困難になると考えられる.

## 5 おわりに

本論文では代謝パスウェイ上の部分代謝構造に着目することで, 効果的に生物の機能的特徴を抽出する二つの手法を提案し, 比較を行った. 結果, 相対的にはあるが, choke point 周辺の代謝構造は多数の生物に対するドメイン分類に, elementary flux mode は少数の生物の詳細な分類に対しても有用であることを示した. このことから, choke point 周辺の代謝構造は生物種の属するドメインの特徴を多く含み, elementary flux mode は生物種固有の特徴を有していると考えられる. 本論文では NCBI taxonomy を指標としたが, 本手法の目的は従来の系統樹の再現ではなく, 異種生物種間の特徴的な代謝構造による類似度を算出し, 従来の系統分類を代謝の観点から補完することである. 提案手法により既存の系統樹と類似した系統樹が構築されたことにより, 提案手法は従来の系統樹を補完することが可能であると考えられる.

今後の課題として, 多数の生物に対するより詳細な分類を代謝の観点から可能とすることがある. 対策としては choke point 周辺の代謝構造と elementary flux mode 両者の特徴を使った生物種比較手法が考えられる.

## 参考文献

- 1) Ashida Y., Ozaki T. and Ohkawa T.: Reaction Structure Profile: A comparative analysis of metabolic pathways based on important substructures, *IPSP Transaction on Bioinformatics*, Vol. 49, No. 5, pp. 36-45 (2008).
- 2) Clemente J. C., Satou K. and Valiente G.: Phylogenetic reconstruction from non-genomic data, *Bioinformatics*, Vol. 23, pp. e110-e115 (2006).
- 3) Rahman S. A. and Schomburg D.: Observing local and global properties of metabolic pathways: 'load points' and 'choke points' in the metabolic networks, *Bioinformatics*, Vol. 22, No. 14, pp. 1767-1774 (2006).
- 4) Schuster S., Dandekar T. and Fell D. A.: Detection of elementary flux modes in biochemical networks: a promising tool for pathway analysis and metabolic engineering, *Trends in Biotechnology*, Vol. 17, pp. 53-60 (1999).
- 5) Tohsato Y., Matsuda H. and Hashimoto A.: A multiple alignment algorithm for metabolic pathway analysis using enzyme hierarchy, *In Proceedings of the 8th International Conference on Intelligent Systems for Molecular Biology*, pp. 376-383 (2000).