

Elementary flux mode 型代謝経路の化学量論的に釣り合った 代謝産物レベルネットワーク構造の算出

太田潤^{†1,a)}

概要：代謝ネットワークにおける“経路”として、*in silico* atomic tracing により基質-生成物の二項関係を連ねたタイプの“経路”以外に、化学量論的に釣り合った個々の反応の集合としての“経路”を考えることができる。これまで、反応の集合としての“経路”を elementary flux mode 型経路 (EFM 型経路) と呼び、通常は原子レベルで考えることのない EFM 型経路に原子レベルの情報を与える (原子レベルマッピングを行う) ためのアルゴリズムを報告してきた。EFM 型経路の原子レベルマッピングでは、経路の化学量論的に釣り合った代謝産物レベルネットワーク構造も算出される。本稿では、原子レベル情報を用いることなく EFM 型経路の化学量論的に釣り合った代謝産物レベルネットワーク構造を算出するための基本アルゴリズムを述べる。基本アルゴリズムにおいて、ネットワーク構造が複数算出されそれらの異同判定が必要な場合がある。本稿では、EFM 型経路の化学量論的に釣り合った代謝産物レベルネットワーク構造の異同判定アルゴリズムについても述べる。

キーワード：代謝ネットワーク, EMU tracing, elementary flux mode

Calculation of stoichiometrically balanced metabolite-level network structure of elementary flux mode type metabolic pathway

JUN OHTA^{†1,a)}

1. はじめに

生理活性物質の生成は、代謝ネットワークの主要な機能であり、生理活性物質の生成経路に関する知識・情報は、医学・生物学的に重要である。これまで、代謝ネットワークにおける経路 (代謝経路) として研究されてきたものには 2 種類の考え方の“経路”, 原料代謝産物に含まれる原子が基質 - 生成物の二項関係の連なりにより目的代謝産物に含まれる原子に移動する道筋としての“経路” [1-6] と化学量論的に釣り合った個々の反応の集合としての“経路” [7,8] が含まれる。前者の“経路”の算出を *in silico* atomic tracing と呼ぶことがある。また、後者の“経路”の代表例は, elementary flux mode [7] と extreme pathway [8] であり、本稿ではこれらの例を含む後者の“経路”を elementary flux mode 型経路 (EFM 型経路) と呼ぶ。

特定の目的代謝産物 B の生成原料として代謝産物 A が適しているか判断する際に、通常、*in silico* atomic tracing により A の分子を構成する原子のうちより多くをより少ない数の反応で B までトレースできる linear な経路が代謝マップ上で検討されるが、A から B に至る linear な経路の存在は、A が B の原料であるための必要条件であって十分条件ではない。一方、反応の集合である EFM 型経路においては、反応の集合から求まる経路の収支式左辺に現れる原料分子群のすべてが揃った場合に収支式右辺の目的分子群が生成すると言えるが、反応の集合そのものは、原料分子群から目的分子群に至る具体的な道筋 (EFM 型経路に含まれ

る反応がどのような順序で働くか) を明示しない。

通常は原子レベルで考えることのない EFM 型経路について、私は、“EFM 型経路における原料分子群と目的分子群は化学量論的に釣り合っており、基本的には 1 セットの原料分子群の各原子が 1 セットの目的分子群の各原子に取り込まれる原子レベルの道筋が存在する筈である”と考え、EFM 型経路に原子レベルの情報を与える (原子レベルマッピングを行う) ためのアルゴリズムを研究してきた [9,10,11]。その過程で、EFM 型経路の原子レベルマッピングの副産物として経路の代謝産物レベルの代謝ネットワーク構造が求まることを経験した。一般的な代謝マップが、化学量論を考慮していない (解糖系では化学量論的に 1 分子の glucose から 2 分子の lactate が生成するが代謝マップ上には glucose ノードと lactate ノードがそれぞれ 1 つしかない等) ことが多いのに対して、EFM 型経路の原子レベルマッピングの副産物として求まる代謝産物レベルの代謝ネットワーク構造は、経路の化学量論を正確に表現する、化学量論的に釣り合った代謝ネットワーク構造であった。また、1 つの EFM 型経路に対して可能な化学量論的に釣り合った代謝産物レベルネットワーク構造は 1 種類とは限らなかった。ペントースリン酸経路の反応により 2 分子の fructose 6-phosphate と 1 分子の glyceraldehyde 3-phosphate から 3 分子の ribose 5-phosphate が生成する EFM 型経路の完全原子レベルマッピングにおいては、1 つの EFM 型経路に対して 2 種類の化学量論的に釣り合った代謝産物レベルネットワーク構造 (cycle を含まない経路と cycle を含む経路) が算出された [9,10],

^{†1} 岡山大学 大学院医歯薬学総合研究科 (医) 生化学分野
a) jo25@md.okayama-u.ac.jp

EFM 型経路の化学量論的に釣り合った代謝産物レベルネットワーク構造は、原料分子群から目的分子群に至る具体的な道筋 (EFM 型経路に含まれる反応がどのような順序で働くか) を示す、EFM 型経路の理解に不可欠なものである。EFM 型経路の化学量論的に釣り合った代謝産物レベルネットワーク構造を EFM 型経路の原子レベルマッピングの副産物として求めるには、EFM 型経路を構成する反応の原子レベル情報が必要であるが、原理的に、代謝産物レベルのネットワーク構造は原子レベル情報に依存しない。そこで、原子レベルの情報をを用いることなく、EFM 型経路を構成する反応の代謝産物レベルの化学量論情報のみから EFM 型経路の化学量論的に釣り合った代謝産物レベルネットワーク構造を算出する方法の研究を行った。

これまでの代謝産物レベルの生合成代謝経路に関する情報学的研究には、目的代謝産物の生成に必要な原料代謝産物の集合を求める手法の研究 [12] や与えられた代謝ネットワークに存在する生合成経路の列挙を試みる研究 [13] などがある。前者の手法では生合成経路の算出はできない。後者の手法では生合成経路が産出されるが、算出されるのは、化学量論を考慮しない 1 代謝産物 (種) 1 ノードの代謝産物レベルネットワーク構造である。

本稿では、EFM 型経路の化学量論的に釣り合った代謝産物レベルネットワーク構造を求めるための基本アルゴリズム (基本アルゴリズム) と EFM 型経路の化学量論的に釣り合った代謝産物レベルネットワーク構造の異同判定アルゴリズム (異同判定アルゴリズム) およびそれらを用いた計算実験の結果を述べる。異同判定アルゴリズムは、基本アルゴリズムによりネットワーク構造が複数算出された場合に、それらの異同判定を行うためのものである。

2. 方法

基本アルゴリズムと異同判定アルゴリズムを述べる。基本アルゴリズムで直接に算出される“化学量論的に釣り合った代謝産物レベルネットワーク構造”は、2.1 に述べる変換後の EFM 型経路の、代謝産物と反応をノードとするネットワーク構造であり、すべての代謝産物ノードが、そのノードが示す代謝産物分子を生成する反応の反応ノードと利用する反応の反応ノードの計 2 ノードのみとつながりを持つ。本稿では、代謝産物に対応するノードを“代謝産物ノード”、反応に対応するノードを“反応ノード”と呼ぶ。以下において“化学量論的な代謝ネットワーク構造”は“化学量論的に釣り合った代謝産物レベルネットワーク構造”と同義である。

2.1 与えられた EFM 型経路の、単一の仮想原料反応から単一の仮想目的反応に至る EFM 型経路への変換

構成反応の集合として与えられた EFM 型経路から、経路の原料代謝産物 source(s) と目的代謝産物 target(s) の情報を得る。原料代謝産物と目的代謝産物は何れも複数個の分

子である場合がある。この原料代謝産物と目的代謝産物に対して、原料代謝産物を何もないところから生成する仮想原料反応 source reaction と目的代謝産物を除去するが何も生成しない仮想目的反応 target reaction を定義して反応の集合に追加し、与えられた EFM 型経路を単一の仮想原料反応 source reaction から単一の仮想目的反応 target reaction への経路に変換する。

source reaction: no metabolite(s) → source(s)

target reaction: target(s) → no metabolite(s)

この変換により、原料代謝産物と目的代謝産物の代謝産物ノードを含むすべての代謝産物ノードが例外なくそのノードが示す分子を生成する反応の反応ノードと利用する反応の反応ノードの計 2 ノードのみとつながりを持つ EFM 型経路のネットワーク構造を考えられるようになる。2.2, 2.3, 2.4 における“EFM 型経路”は変換後の EFM 型経路を指す。

2.2 EFM 型経路の反応構成の反応番号による記述および反応番号と代謝産物番号による反応の記述

EFM 型経路に現れる反応、代謝産物に対して、反応番号、代謝産物番号を付す。同じ反応・同じ代謝産物が経路に複数現れる場合、同じコンパートメントにあるものであれば、複数のそれらに同じ反応番号・代謝産物番号を付す。

EFM 型経路の反応構成を反応番号の配列 (反応番号配列) として表現する。EFM 型経路に反応番号 r の反応が n 回現れる場合は、反応番号配列に r が n 回現れるようにする。各反応の反応番号を反応番号配列に格納する順序は任意であるが、いったん格納したら番号の順序を変更しない。

EFM 型経路に現れる反応はそれぞれ、反応番号、基質の代謝産物番号の配列、生成物の代謝産物番号の配列の組み合わせとして表現する。ある反応の反応式の基質側の辺に代謝産物番号が m で係数が n の代謝産物分子があれば、その反応の基質の代謝産物番号の配列に m が n 回現れる。生成物についても同様である。仮想原料反応 source reaction の基質の代謝産物番号の配列、仮想目的反応 target reaction の生成物の代謝産物番号の配列は空配列である。

2.3 EFM 型経路の化学量論的な代謝ネットワーク構造算出のための基本アルゴリズム

EFM 型経路の化学量論的な代謝ネットワーク構造に存在する反応ノードは、EFM 型経路の反応番号配列の要素と 1 対 1 に対応する。各反応ノードは、その反応の基質に相当するすべての代謝産物ノードそれぞれからのエッジが入る地点であるとともに、その反応の生成物に相当するすべての代謝産物ノードそれぞれに向けてのエッジが出る地点でもある。すなわち、各反応ノードには、代謝産物 (基質) が入るエッジと (エッジを介して入る) 代謝産物 (基質) の組み合わせの集合と代謝産物 (生成物) が出るエッジと (エッジを介して出る) 代謝産物 (生成物) の組み合わせの集合が対応する。EFM 型経路の化学量論的なネットワーク構造に含まれるすべての代謝産物ノードはそれぞれが、1

エッジを介して1反応ノードから出て、別の1エッジを介して別の1反応ノードに入る。したがって、EFM型経路の化学量論的なネットワーク構造において、エッジを介して反応ノードから出る代謝産物ノードの集合をP、エッジを介して反応ノードに入る代謝産物ノードの集合をSとすると、集合Pと集合Sは代謝産物ノードの集合としては同一の集合でなければならない。すなわち、集合Pの要素(どの反応ノードから出るかにより区別される代謝産物ノード)は集合Sの要素(どの反応ノードに入るかにより区別される代謝産物ノード)と同一ノードとして1対1に対応付けられる。Pの特定の要素とSの特定の要素は、両方が同じ(種類の)代謝産物のノードである場合に限り対応付けできる。基本アルゴリズムにおいては、ネットワークに含まれる代謝産物(の種類)ごとに、その(種類の)代謝産物の代謝産物ノードを要素とするPの部分集合とSの部分集合を作成し、両部分集合の要素を任意の組み合わせで対応付ける(同一とみなす)ことを行い、EFM型経路の、すべての可能な、化学量論的な代謝ネットワーク構造を算出する。ネットワーク内に代謝産物が m_1 から m_k のk種類あり、代謝産物番号が m_j (jはk以下の自然数)の代謝産物ノードの個数が $n(j)$ であるとすると、代謝産物番号が m_j の代謝産物ノードを要素とするPの部分集合とSの部分集合の要素間の対応付けは $n(j)!$ 通りある。したがって、PとSの要素間の可能な対応付けは全部で $n(1)! \times n(2)! \times n(3)! \times \dots \times n(k)!$ 通り存在する。

基本アルゴリズムでは、EFM型経路の反応番号配列R、EFM型経路に含まれる各反応の情報(反応番号、基質の代謝産物番号の配列、生成物の代謝産物番号の配列)から、 cCM_s (EFM型経路の化学量論的な代謝ネットワーク構造におけるノード間のつながりを示すエッジリストCM_sを要素とする配列)とnode_list(各ノードのnode number(ノード番号)、node content(代謝産物番号または反応番号)、node type(1: 代謝産物ノード, 2: 反応ノード)を格納した行列)を得る。 cCM_s を空配列、nR, nM, snr, rpnを空行列、iを1として計算を開始する。以下の記述では、配列Aのj番目の要素を“A[j]”, 配列Aの要素数を“length(A)”と表す記法を用いる。

Step 1. node_number = i, 反応R[i]の基質の代謝産物番号の配列をs, jを1とする。snrに[s[j], node_number]を新たな行として追加してjを1増加させる操作を、操作前のjがlength(s)以下である間、繰り返す。

Step 2. node_number = i, 反応R[i]の生成物の代謝産物番号の配列をp, jを1とする。rpnに[node_number, p[j]]を新たな行として追加してjを1増加させる操作を、操作前のjがlength(p)以下である間、繰り返す。

Step 3. node_number = i, node_content = R[i], node_type = 2としてnRに[node_number, node_content, node_type]を新たな行として追加する。

Step 4. 引き続き、Step 1, Step 2, Step 3の操作を、2からlength(R)まで1ずつiを増加させながら繰り返す。

Step 5. snrの行を、1列目の代謝産物番号が小から大に並ぶようにソートする。rpnの行を、2列目の代謝産物番号が小から大に並ぶようにソートする。ソート後のsnrの1列目とソート後のrpnの2列目が一致することを確認し、snrをソート後のsnrで、rpnをソート後のrpnで置き換える。その後にM = (snrの1列目), i = 1とする。

Step 6. node_number = length(R) + i, node_content = M[i], node_type = 1としてnMに[node_number, node_content, node_type]を新たな行として追加する。

Step 7. 引き続き、Step 6の操作を、2からlength(M)まで1ずつiを増加させながら繰り返す。最終的に得られるnMの上にnRを積み重ねて得られる行列(行数がlength(R)+length(M), 列数が3)をnode_listとする。

Step 8. snrの1列目(代謝産物番号の列)をnMの1列目(node numberの列)で置き換えたものをsrとする。rpnの2列目(代謝産物番号の列)をnMの1列目(node numberの列)で置き換えたものをrpとする。Step 8で用いるnMは、Step 7で最終的に得られるnMである。

Step 9. Mに格納されている代謝産物番号をunique化した配列をuMとする。cvniを配列とし、代謝産物番号uM[k]と等しいMの要素の要素番号すべてをcvni[k]に格納する。cvni[k]に格納されるMの要素番号は、uM[k]がsnrの1列目(=rpnの2列目)に現れるsnr(rpn)の行の行番号、node contentがuM[k]である代謝産物ノードがsrの1列目(=rpの2列目)に現れるsr(rp)の行の行番号と一致する。

Step 10. 1以上length(cvni)以下のすべての整数kに対してsrのcvni[k]行1列の要素の置換(恒等置換を含む)を行いsrからsrwを生成し、srwをrpの上に積み重ねて得られる行列CM_sは、EFM型経路の化学量論的な代謝ネットワーク構造におけるノード間のつながりを示すエッジリストとなっている。逆に、EFM型経路の化学量論的な代謝ネットワーク構造におけるノード間のつながりを示すエッジリストで上記の方法で作成できないものはない。このような置換で作成できるCM_sをすべて作成しcCM_sに格納する。

2.4 EFM型経路の化学量論的な代謝ネットワーク構造の異同判定アルゴリズム

2.3の基本アルゴリズムでは、EFM型経路に対して、複数の化学量論的な代謝ネットワーク構造が算出される場合がある。この場合には、得られたネットワーク構造に同じものがないかネットワーク構造の異同判定を行うことが必要である。一般に、ノード数とノード内容の構成が同じ2つのネットワーク構造a, bは、aのノードをノード内容が同じbのノードと1対1に対応付けbのエッジリスト上のbのノードのノード番号を対応付けられたaのノードのノード番号で置き換えて得られるエッジリストがaのエッジリストと同じになれば同じ構造であると判定される。異なる

構造であると判定するために、**a** のノードと **b** のノードのすべての対応付けに対する確認を行えば、計算量は膨大となる。しかし、変換後の EFM 型経路に対して得られる化学量論的な代謝ネットワーク構造は、入るエッジ（“入り”）のないノードが仮想原料反応 **source reaction** のノードのみ、出るエッジ（“出”）のないノードが仮想目的反応 **target reaction** のノードのみの有向グラフであり、比較する 2 つのネットワーク構造における仮想原料反応 **source reaction** のノード間の対応付けと、仮想目的反応 **target reaction** のノード間の対応付けが固定されているので、ネットワーク構造の異同判定を以下のアルゴリズムにより比較的小さい計算量で行うことができる。このアルゴリズムでは、EFM 型経路に対して得られる化学量論的な代謝ネットワーク構造 **a**、**b** の **CM_s** と **node_list** に含まれる情報から、ネットワーク構造 **a** とネットワーク構造 **b** の異同を判定する。

Step 1. ネットワーク構造 **a** において、“入り”がなく“出”があるノードのリスト **s**，“入り”はあるが“出”がないノードのリスト **d** を作成し、**s** に含まれるノードを始点、**d** に含まれるノードを終点とする経路をすべて算出し、配列 **xa** に要素として格納する。

Step 2. ネットワーク構造 **a** の部分ネットワーク構造であって、Step 1 の経路計算で使われていないエッジからなるネットワーク構造 **ar** を得る。ネットワーク構造 **ar** において、“入り”がなく“出”があるノードのリスト **s**，“入り”はあるが“出”がないノードのリスト **d** を新たに作成し、**s** に含まれるノードを始点、**d** に含まれるノードを終点とする経路をすべて算出し、配列 **xar** に要素として格納する。

Step 3. ネットワーク構造 **b** において、“入り”がなく“出”があるノードのリスト **s**，“入り”はあるが“出”がないノードのリスト **d** を新たに作成し、**s** に含まれるノードを始点、**d** に含まれるノードを終点とする経路をすべて算出し、配列 **xb** に要素として格納する。

Step 4. ネットワーク構造 **b** の部分ネットワーク構造であって、Step 3 の経路計算で使われていないエッジからなるネットワーク構造 **br** を得る。ネットワーク構造 **br** において、“入り”がなく“出”があるノードのリスト **s**，“入り”はあるが“出”がないノードのリスト **d** を新たに作成し、**s** に含まれるノードを始点、**d** に含まれるノードを終点とする経路をすべて算出し、配列 **xbr** に要素として格納する。

Step 5. 配列 **xa** の要素である経路を経路長ごとに分類し、 $S(xa, k) = \{xa \text{ の要素 (経路) で経路長が } k \text{ に等しいもの}\}$ とする。配列 **xar**、**xb**、**xbr** についても同様に、 $S(xar, k) = \{xar \text{ の要素 (経路) で経路長が } k \text{ に等しいもの}\}$
 $S(xb, k) = \{xb \text{ の要素 (経路) で経路長が } k \text{ に等しいもの}\}$
 $S(xbr, k) = \{xbr \text{ の要素 (経路) で経路長が } k \text{ に等しいもの}\}$ とする。集合 **A** の要素数を $|A|$ と表すとき、すべての自然数 **k** に対して $|S(xa, k)| = |S(xb, k)|$ かつ $|S(xar, k)| = |S(xbr, k)|$ が成り立つ場合に、次の Step に進む。成り立たない場合は、ネ

ットワーク構造 **a** とネットワーク構造 **b** が異なる構造であると判定される。

Step 6. $|S(xa, k)| = |S(xb, k)|$ が 0 でないような **k** すべて ($k_j = k_1, k_2, k_3, \dots$) に対して、次の条件を満たす $S(xa, k_j)$ の要素（経路）と $S(xb, k_j)$ の要素（経路）の 1 対 1 の対応付けをすべて含む集合 $C(k_j)$ を作る。 $S(xa, k_j)$ の要素（経路）**p** と $S(xb, k_j)$ の要素（経路）**q** が対応付けできるための条件は、経路（代謝産物ノードと反応ノードの配列）上の代謝産物と反応の配列が **p** と **q** で一致していることである。また、 $|S(xar, k)| = |S(xbr, k)|$ が 0 でないような **k** すべて ($k_{r_i} = k_{r_1}, k_{r_2}, k_{r_3}, \dots$) に対して、次の条件を満たす $S(xar, k_{r_i})$ の要素（経路）と $S(xbr, k_{r_i})$ の要素（経路）の 1 対 1 の対応付けをすべて含む集合 $Cr(k_{r_i})$ を作る。 $S(xar, k_{r_i})$ の要素（経路）**p** と $S(xbr, k_{r_i})$ の要素（経路）**q** が対応付けできるための条件は、経路上の代謝産物と反応の配列が **p** と **q** で一致していることである。得られた $C(k_j)$ と $Cr(k_{r_i})$ に空集合が含まれない場合に、次の Step に進む。空集合が含まれる場合は、ネットワーク構造 **a** とネットワーク構造 **b** が異なる構造であると判定される。

Step 7. ネットワーク構造 **a** とネットワーク構造 **b** が同じ構造である場合には、ネットワーク構造 **a** のノードとネットワーク構造 **b** のノードの間に、互いのノードの内容 **node content** と各所属ネットワーク内でのトポロジカルな位置が一致するような 1 対 1 の対応付けが存在する。Step 6 で得られるすべての $C(k_j)$ と $Cr(k_{r_i})$ から 1 つずつ経路の対応付けを選び組み合わせたものの集合を **Q** とすると、**Q** の各要素から、ネットワーク構造 **a** のノードとネットワーク構造 **b** のノードの間のノードの内容 **node content** が一致する対応付けが生成する。**Q** の各要素から生成するノード間の対応付けの集合を **P** とする。**P** に含まれる対応付けの中に、対応付けられた 2 つのノードの間で各所属ネットワーク内でのトポロジカルな位置が一致するような 1 対 1 対応が存在する場合に限って、ネットワーク構造 **a** とネットワーク構造 **b** が同じ構造であると言える。この Step では、このような対応付けの有無を確認する。ネットワーク構造 **b** の **CM_s** のノード番号を、対応付けられたネットワーク構造 **a** のノード番号で置き換えて得られる行列を行ソートしたものがネットワーク構造 **a** の **CM_s** を行ソートしたものと同じになれば、その対応付けは、対応付けられた 2 つのノードの間で各所属ネットワーク内でのトポロジカルな位置が一致している **a** のノードと **b** のノードの間の 1 対 1 の対応付けである。

3. 計算実験の結果と考察

糖質代謝モデルネットワーク [6] のペントースリン酸経路に含まれる 2 種類の **transketolase** 反応・**transaldolase** 反応・**ribulose-5-phosphate 3-epimerase** 反応の計 4 反応（文献 [6] のプロセス番号 26, 28, 29, 30）の逆反応と

ribose-5-phosphate ketoisomerase 反応 (文献 [6] のプロセス番号 27) の組み合わせ (計 5 反応, 延べ 7 反応) により化学量論的に 2 分子の fructose 6-phosphate と 1 分子の glyceraldehyde 3-phosphate から 3 分子の ribose 5-phosphate を生成する反応収支式ができる。これは 5 反応 (延べ 7 反応) からなる EFM 型経路である。1 で述べたように, この EFM 型経路に炭素原子部分の完全原子レベルマッピングを行うと, cycle を含まない経路と cycle 含む経路に対応する 2 種類のマッピングが得られたことを報告している [9,10]。既報では原子レベルの情報を用いて, 本稿の“化学量論的に釣り合った代謝産物レベルネットワーク構造” (“化学量論的な代謝ネットワーク構造”) である “cycle を含まない経路” と “cycle 含む経路” を算出した。本項の計算実験の結果としては, 2.3 で述べた EFM 型経路の化学量論的な代謝ネットワーク構造算出のための基本アルゴリズム (基本アルゴリズム) を用いて, 原子レベルの情報を用いることなく, 既報と全く同じ反応の集合である EMU 型経路の化学量論的に釣り合った代謝産物レベルネットワーク構造 (化学量論的な代謝ネットワーク構造) を算出し, 算出されたネットワーク構造に 2.4 で述べた EFM 型経路の化学量論的な代謝ネットワーク構造算出のための異同判定アルゴリズム (異同判定アルゴリズム) を適用した結果を述べる。計算は Matlab と GNU Octave 上で行った。

上述の EFM 型経路を 2.1 の方法で変換して基本アルゴリズムを行うと Step 7 で下記の node_list が算出された。

1	31	2
2	32	2
3	-26	2
4	-26	2
5	27	2
6	27	2
7	-28	2
8	-29	2
9	-30	2
10	3	1
11	3	1
12	5	1
13	5	1
14	28	1
15	28	1
16	29	1
17	29	1
18	29	1
19	30	1
20	31	1
21	32	1
22	32	1

node_list の 1, 2, 3 列は, それぞれ, node number (ノード番号), node content (代謝産物番号・反応番号), node type (1: 代謝産物ノード, 2: 反応ノード) である。node_list を参照すると, ノード番号 (node number) から, そのノードが代謝産物ノードと反応ノードのどちらであるか (node type) を含む, そのノードの内容 (node content) を知ることができる。仮想原料反応 source reaction はノード番号 (node number) が 1, 反応番号 (node content) が 31, 仮想目的反応 target reaction はノード番号 (node number) が 2, 反応番号 (node content) が 32 である。

基本アルゴリズム Step 8 では下記の rp (左) と sr (右) が算出された。

1	10	10	8
1	11	11	9
1	12	12	7
8	13	13	9
3	14	14	5
4	15	15	6
5	16	16	2
6	17	17	2
7	18	18	2
8	19	19	7
9	20	20	8
7	21	21	3
9	22	22	4

rp と sr の要素はノード番号である。rp と sr の各行は, rp の場合は反応ノード (1 列目) から代謝産物ノード (2 列目) に至る, sr の場合は代謝産物ノード (1 列目) から反応ノード (2 列目) に至るつながりを示す。rp の 2 列目と sr の 1 列目は, どちらもノード番号 10 から 22 であり, 内容と順序が完全に一致している。ノード番号 10 から 22 は, node_list に記されているように, ノード番号 10 と 11 (2 ノード) が代謝産物番号 3 (fructose 6-phosphate), ノード番号 12 と 13 (2 ノード) が代謝産物番号 5 (glyceraldehyde 3-phosphate), ノード番号 14 と 15 (2 ノード) が代謝産物番号 28 (ribulose 5-phosphate), ノード番号 16, 17, 18 (3 ノード) が代謝産物番号 29 (ribose 5-phosphate), ノード番号 19 (1 ノード) が代謝産物番号 30 (sedoheptulose 7-phosphate), ノード番号 20 (1 ノード) が代謝産物番号 31 (erythrose 4-phosphate), ノード番号 21 と 22 (2 ノード) が代謝産物番号 32 (xylulose 5-phosphate) である。括弧内に記した各代謝産物のノードの個数から, 基本アルゴリズム Step 10 では $2! \times 2! \times 2! \times 3! \times 1! \times 1! \times 2! = 96$ の構造が算出されることが予想されるが, 実際に算出された構造も 96 であった。基本アルゴリズムで 96 の構造を算出するのに要した時間は Matlab で 0.02 秒以下, GNU Octave で 0.3 秒程度であっ

た (Intel(R) Core(TM) i7-8565U CPU @ 1.80GHz 1.99GHz).

基本アルゴリズムで算出された 96 の構造を異同判定アルゴリズムで調べると、2 種類の構造がそれぞれ 48 ずつあることがわかった。2 種類の構造は cycle を含まない構造 (経路) と cycle 含む構造 (経路) であり、既報 [9,10] の結果と一致した。異同判定アルゴリズムを行った回数は、96 構造のうち 1 構造からそれと同じ 47 構造を見つけるために 95 回、残りの 48 構造がすべて同じであることを確認するための 47 回、計 142 回であった。この 142 回の異同判定アルゴリズムの実行に要した時間は、Matlab で 0.6 秒程度、GNU Octave で 5 秒程度であった (Intel(R) Core(TM) i7-8565U CPU @ 1.80GHz 1.99GHz)。

反応の集合としての経路 (EFM 型経路) の例として挙げた elementary flux mode [7] と extreme pathway [8] は、数学的には、出現回数付きの反応リストとして理解されるベクトルであり、通常、それが具体的にどのような経路を意味するかを知るために既存の代謝マップを参照することが必要であった。本稿のアプローチは、出現回数付きの反応リストとして理解されるベクトルである elementary flux mode [7] と extreme pathway [8] から、それが意味する (本来の意味での) 経路・道筋・代謝マップを純粋に数学的に生成しようとするものであり、一定の価値があると考えられる。本稿のアルゴリズムを、同じ代謝産物が多数回現れる、あるいは、複数回現れる代謝産物が多い EFM 型経路に適用すると、算出される構造が多いため、基本アルゴリズムの Step 10 と異同判定において計算時間の問題を生じる可能性がある。これを避けるための高速アルゴリズム (基本アルゴリズムの Step 10 に異同判定を組み込んだもの) の開発を進めており、別に報告する予定である。

参考文献

- [1] Arita, M. *In silico* atomic tracing by substrate-product relationship in Escherichia coli intermediary metabolism. *Genome Res.* **2003**, *13*, 2455-2466. DOI: 10.1101/gr.1212003
- [2] Latendresse, M.; Krummenacker, M.; Karp, P.D. Optimal metabolic route search based on atom mappings. *Bioinformatics* **2014**, *30*, 2043-2050. DOI: 10.1093/bioinformatics/btu150
- [3] Heath, A.P.; Bennett, G.N.; Kavasaki, L.E. Finding metabolic pathways using atom tracking. *Bioinformatics* **2010**, *26*, 1548-1555. DOI: 10.1093/bioinformatics/btq223
- [4] Pitkänen, E.; Jouhten, P.; Rousu, J. Inferring branching pathways in genome-scale metabolic networks. *BMC Syst. Biol.* **2009**, *3*, 1-22. DOI: 10.1186/1752-0509-3-103
- [5] Ohta, J. Connectivity matrix method for analyses of biological networks and its application to atom-level analysis of a model network of carbohydrate metabolism. *IEE Proc. Syst. Biol.* **2006**, *153*, 372-374. DOI: 10.1049/ip-syb:20060018
- [6] Ohta, J. Single-atom tracing in a model network of carbohydrate metabolism and pathway selection. *IPSI Transactions on Bioinformatics* **2018**, *11*, 1-13.
- [7] Schuster, S.; Hilgetag, C. On elementary flux modes in biochemical reaction systems at steady state. *J. Biol. Syst.* **1994**, *2*, 165-182.
- [8] Schilling, C. H.; Palsson, B.O. The underlying pathway structure of biochemical reaction networks. *Proc. Nat. Acad. Sci. USA* **1998**, *95*, 4193-4198.
- [9] 太田 潤 代謝ネットワークにおける elementary flux mode 型経路の完全原子レベルマッピング. 情報処理学会研究報告, **2019**, Vol. 2019-BIO-59, No. 2, 1-6.
- [10] 太田 潤 Elementary flux mode 型代謝経路の完全原子レベルマッピングの高速化. 情報処理学会研究報告, **2019**, Vol.2019-BIO-60, No. 6, 1-7.
- [11] 太田 潤 対称性のある代謝産物を通る elementary flux mode 型代謝経路の完全原子レベルマッピング. 情報処理学会研究報告, **2020**, Vol.2020-BIO-61, No. 9, 1-8.
- [12] Acuña, V.; Milreu, P.V.; Cottret, L.; Marchetti-Spaccamela, A.; Stougie, L.; Sagot, M.F. Algorithms and complexity of enumerating minimal precursor sets in genome-wide metabolic networks. *Bioinformatics* **2012**, *28*, 2474-83.
- [13] Ravikrishnan, A.; Nasre, M.; Raman, K. Enumerating all possible biosynthetic pathways in metabolic networks. *Sci. Rep.* **2018**, *8*, 9932. DOI: 10.1038/s41598-018-28007-7