

Elementary flux mode 型経路における 代謝産物分子の生成・利用状態のスナップショットとしての 化学量論的な代謝ネットワーク構造

太田潤^{†1,a)}

概要：代謝ネットワークにおける“経路”として、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”である elementary flux mode 型経路 (EFM 型経路) を考えることができる。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 (化学量論的な代謝ネットワーク構造) のそれぞれを EFM 型経路の variant と呼び、EFM 型経路に含まれる反応に関する情報から、EFM 型経路の unique な variant すべてを求めるアルゴリズムを開発してきた。開発したアルゴリズムにおいては、存在可能な1つの variant に対する置換として他の異なる variant を算出する。本稿では、各 variant が何種類の異なる置換から生成し得るかを高速に計算する方法を述べる。EFM 型経路の化学量論的な代謝ネットワーク構造を、EFM 型経路における代謝産物分子の生成・利用状態のスナップショット (ある瞬間に存在する EFM 型経路上の各代謝産物分子を直接生成した反応と次の瞬間に直接利用する反応を示す) として理解するとき、各置換には、variant に加えて、代謝産物分子の生成・利用状態が対応する。各 variant が何種類の異なる置換から生成し得るかの情報から、各 variant に対応する型の、EFM 型経路における代謝産物分子の生成・利用状態の起こりやすさがわかると考えられる。

キーワード：代謝ネットワーク, elementary flux mode

Stoichiometric metabolic network structure as a snapshot of the state of formation and utilization of metabolite molecules in elementary flux mode type pathway

JUN OHTA^{†1,a)}

1. はじめに

代謝ネットワークの主要な機能の1つである生理活性物質生成のための“経路”に関する知識・情報は、医学・生物学的に重要である。代謝ネットワークにおける“経路” (代謝経路) 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” [1,2] がある。その代表例は、elementary flux mode [1] と extreme pathway [2] であり、本稿ではこれらの反応の集合としての“経路”を elementary flux mode 型経路 (EFM 型経路) と呼ぶ。

EFM 型経路においては、経路を定義する反応の集合に含まれる反応の情報から経路の収支式が求まり、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、反応の集合自体は、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) を示さない。

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) とき、代謝産物ノードそれぞれが2つの反応ノード (代謝産物ノードに対応する代謝産物分子を直接生成する反応と直接利用する反応に対応する) のみとつながりを持つ、代謝産物レベルの代謝ネットワーク構造が副産物として得られることを私は経験した [3,4,5]。こ

で、“代謝産物ノード”は、代謝産物分子に対応するノード、“反応ノード”は、反応に対応するノードのことである。一般の代謝マップには、化学量論を考慮していない (たとえば、1分子の glucose から2分子の lactate が生成する解糖系において glucose ノードと lactate ノードがそれぞれ1つしかない) のものがあるが、EFM 型経路の原子レベルマッピングの副産物として得られた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もので、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) を示していた。本稿における“化学量論的な代謝ネットワーク構造”とは、EFM 型経路の原子レベルマッピングの副産物として得られたような“代謝産物ノードそれぞれが2つの反応ノード (代謝産物ノードに対応する代謝産物分子を直接生成する反応と直接利用する反応に対応する) のみとつながりを持つ代謝ネットワーク構造”のことである。

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 [6] 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 [7] などがあるが、前者の手法では生合成経路の算出はできず、後者の手法で算出される生合成経路は、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

^{†1} 岡山大学 学術研究院医歯薬学域 (医) 生化学分野
a) jo25@md.okayama-u.ac.jp

ーク構造である。

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化学量論的な代謝ネットワーク構造が得られることを経験したが、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。私はこのことに着目し、特定の EFM 型経路を定義する、経路に含まれる反応に関する代謝産物レベルの情報(どの反応がいくつ含まれるか)から、その EFM 型経路の化学量論的な代謝ネットワーク構造を求めるためのアルゴリズムを開発してきた [8,9]。

1つの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1種類の場合もあるが2種類以上の場合もある [3,4]。そこで、EFM型経路に対して可能な化学量論的な代謝ネットワーク構造のそれぞれを EFM 型経路の variant と呼ぶこととした。2.2 で導入する“ n に対する置換” (n は variant を含む化学量論的な代謝ネットワーク構造) という用語法によれば、EFM 型経路のすべての variant は、任意の1つの variant に対する置換として生成する。これまでに開発したアルゴリズムは何れも、EFM 型経路のある1つの variant に対する置換として生成する化学量論的な代謝ネットワーク構造から、異同判定により EFM 型経路の unique な variant をすべて求めるものであった。異なる複数の置換から同じ化学量論的な代謝ネットワーク構造が生成する場合があるので異同判定が必要であった [10,11]。

これまでのアルゴリズム開発における関心は、EFM 型経路に含まれる反応に関する情報から EFM 型経路の unique な variant すべてをどれだけ高速に求められるかにあったが、本稿では、各 variant が何種類の異なる置換から生成し得るかを高速に計算する方法を述べる。また、EFM 型経路の化学量論的な代謝ネットワーク構造を、EFM 型経路における代謝産物分子の生成・利用状態のスナップショット(ある瞬間に存在する EFM 型経路上の各代謝産物分子を直接生成した反応と次の瞬間に直接利用する反応を示す)として理解する考え方を説明し、各 variant が何種類の異なる置換から生成し得るかを計算することの意味について考察する。

2. 方法

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、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 unique な variant 算出のためのアルゴリズム A ならびにアルゴリズム B、EFM 型経路の unique な variant に対応する置換の数の計算、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について述べる。

基本アルゴリズム、アルゴリズム A、アルゴリズム B が直接に算出するのは、2.1 に述べる変換後の EFM 型経路の化学量論的な代謝ネットワーク構造である。アルゴリズム

A では、基本アルゴリズムによりすべての置換に対応する化学量論的な代謝ネットワーク構造を生成した後に異同判定を行い unique な variant を算出する。アルゴリズム B では、基本アルゴリズムにおける置換を互換の繰り返しとして行い、互換により化学量論的な代謝ネットワーク構造が生成するごとに異同判定を行うことにより unique な variant を算出する。

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 [8,9]

基本アルゴリズム、アルゴリズム A、アルゴリズム B の何れにおいても、反応の集合として与えられた EFM 型経路全体の収支から、経路の化学量論的な原料代謝産物 source(s) (原料分子群) と目的代謝産物 target(s) (目的分子群) を求める。原料代謝産物と目的代謝産物は何れも複数個、複数種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 を定義する。

source reaction: no metabolite(s) \rightarrow source(s)

target reaction: target(s) \rightarrow no metabolite(s)

この2つの仮想反応を反応の集合に追加すると、与えられた EFM 型経路は、単一の仮想原料反応から単一の仮想目的反応への経路に変換される。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それぞれが例外なく2つの反応ノード(その代謝産物ノードに対応する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ることができるようになる。ただし、同じ反応が経路に v 回現れる場合はそれらに v 個の反応ノードを対応させ、反応を定義する反応式で基質または生成物の係数が v である場合はその基質または生成物に v 個の代謝産物ノードを対応させるものとする。2.2 から 2.6 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 [8]

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 型経路に現れる反応と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。EFM 型経路の化学量論的なネットワーク構造の各代謝産物ノードは、反応ノードからその代謝産物ノードに入るエッジの数、その代謝産物ノードから反応ノードに向けて出るエッジの数ともに1である。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ある。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同一ノードとして対応付けできる（“対応付け条件”）。化学量論的な代謝ネットワーク構造はこの“対応付け条件”を満たす P の要素と S の要素の対応付けにより生成する。すなわち、“対応付け条件”を満たす P の要素と S の要素の対応付けには、それに対応する化学量論的な代謝ネットワーク構造が存在する。以下、“ P の要素と S の要素の対応付け”の意味で“ P - S 間対応付け”を用いる。

P と S を、要素（代謝産物ノード）の代謝産物種（代謝産物の種類）ごとに部分集合に分割し、 P と S の同じ代謝産物種の部分集合間で要素を対応付ける（同一とみなす）ことを考える。このような対応付けは“対応付け条件”を明らかに満たす。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 m_i (i は k 以下の自然数) の代謝産物ノードの個数が $d(m_i)$ であるとする、代謝産物 m_i 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 $d(m_i)!$ 通りある。したがって、“対応付け条件”を満たす P - S 間対応付けは、全部で $r = d(m_1)! \times d(m_2)! \times d(m_3)! \times \dots \times d(m_k)!$ 通り存在する。それらは、“対応付け条件”を満たす P - S 間対応付け x により対応付けられた S の要素の、代謝産物種ごとの置換と 1 対 1 に対応する。

基本アルゴリズムにおいては、“対応付け条件”を満たす任意の 1 つの P - S 間対応付け x とそれに対応する化学量論的な代謝ネットワーク構造を求め、“対応付け条件”を満たす x の置換（“ x により対応付けられた S の要素の置換により得られる P - S 間対応付け”の意）とそれに対応する化学量論的な代謝ネットワーク構造をすべて求める。

本稿では、“対応付け条件”を満たす P - S 間対応付け x に化学量論的な代謝ネットワーク構造 n が対応するとき、“対応付け条件”を満たす x の置換に対応する化学量論的な代謝ネットワーク構造について“化学量論的な代謝ネットワーク構造 n に対する置換”または“ n に対する置換”と呼ぶことがある。この用語法によれば、基本アルゴリズムによる EFM 型経路の化学量論的な代謝ネットワーク構造算出は、EFM 型経路の可能な化学量論的な代謝ネットワーク構造の 1 つである n の算出に引き続く、 n に対する置換としての化学量論的な代謝ネットワーク構造すべての算出である。

2.3 EFM 型経路の unique な variant 算出のためのアルゴリズム A [8,9]

基本アルゴリズムにより、“対応付け条件”を満たす P - S 間対応付けの数と同数の EFM 型経路の化学量論的な代謝ネットワーク構造を得る。得られた化学量論的な代謝ネットワーク構造の異同判定を異同判定アルゴリズムにより行い、unique な化学量論的な代謝ネットワーク構造を求める。得られる unique な化学量論的な代謝ネットワーク構造が、EFM 型経路の unique な variant である。

アルゴリズム A は、EFM 型経路の unique な variant の算出のために、すべての置換に対応する化学量論的な代謝ネットワーク構造の異同判定を行うので、アルゴリズム A では、異同判定の結果を記録すると、EFM 型経路の unique な variant に対応する置換の数が直接得られる。

2.4 EFM 型経路の unique な variant 算出のためのアルゴリズム B [9]

P - S 間対応付け x に対して、本項でも 2.2 と同様に、 x により対応付けられた S の要素の置換を“ x の置換”と呼び、さらに、 x により対応付けられた S の要素の互換を“ x の互換”と呼ぶ。また、“ x の置換”と“ x の互換”は、それぞれ、“ x により対応付けられた S の要素の置換により得られる P - S 間対応付け”、“ x により対応付けられた S の要素の互換により得られる P - S 間対応付け”の意味でも用いる。互換が置換に含まれることから、“ x の互換”は“ x の置換”でもある。

アルゴリズム B では、最初に“対応付け条件”を満たす P - S 間対応付けの 1 つ x を求め、 x に対応する化学量論的な代謝ネットワーク構造を空集合 U に要素として入れる。2.2 と同様に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 m_i (i は k 以下の自然数) の代謝産物ノードの個数が $d(m_i)$ であるとする、 x の置換それぞれに対して可能な、“対応付け条件”を満たす P - S 間対応付けを生成する互換が全部で $c = d(m_1)C_2 + d(m_2)C_2 + d(m_3)C_2 + \dots + d(m_k)C_2$ 通り存在する。数学的に、任意の置換は、互換の組み合わせで表現できるので、“対応付け条件”を満たす P - S 間対応付けを生成する x の互換を繰り返し (x に可能な、“対応付け条件”を満たす P - S 間対応付けを生成する 1 回の互換をすべて行い、得られた x の置換それぞれに対して可能な、“対応付け条件”を満たす P - S 間対応付けを生成する 1 回の互換をすべて行い、さらに、得られた x の置換それぞれに対して可能な、“対応付け条件”を満たす P - S 間対応付けを生成する 1 回の互換をすべて行うことを繰り返し)、“対応付け条件”を満たす x の置換 (P - S 間対応付け) を生成する。その際、各互換で x の置換 y が得られるごとに、その x の置換 y (P - S 間対応付けである) に対応する化学量論的な代謝ネットワーク構造 n を求めて、 U の要素の化学量論的な代謝ネットワーク構造との異同判定を行い、次のルールに従って、 x の置換 y の互換をやめるか y に次の互

換を行うかを決める。

- n が U の要素であれば, x の置換 y の互換をやめる.
- n が U の要素でなければ, n を U に要素として入れ, x の置換 y に次の互換を行う.

互換により U に含まれない新たな化学量論的な代謝ネットワーク構造 n が生成されなくなるまで互換を続けると, 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が U の要素として得られる. U に含まれる化学量論的な代謝ネットワーク構造が, EFM 型経路の unique な variant である.

U の要素に前述の手順で得られる順に番号を付し, i 番目に得られた U の要素を $U[i]$ と表す. $U[i]$ は, x の特定の置換に対応する化学量論的な代謝ネットワーク構造として求めたものである. この, $U[i]$ に対応する ($U[i]$ を生成した) x の特定の置換を $x(i)$ で表す. また, x の置換 y に対応する化学量論的な代謝ネットワーク構造を $n(y)$ と表す. これらの記法では, $U[i] = n(x(i))$, $U[1] = n(x(1))$, $x = x(1)$ である.

x の置換 a , b に対して, $n(a) = n(b)$ のとき, $n(a)$ と $n(b)$ の同型条件を満たす全単射 f [11] により, $n(a)$ のノードと $n(b)$ のノードを 1 対 1 に対応付けできる. $n(a) = n(b)$ のとき, $n(a)$ の 2 ノード a_i, a_j に f により対応付けられた $n(b)$ の 2 ノードを b_i, b_j とし, x の置換 a における a_i, a_j に対応する S の要素の互換を $a(a_i, a_j)$, x の置換 b における b_i, b_j に対応する S の要素の互換を $b(b_i, b_j)$ と表すと, $n(a(a_i, a_j)) = n(b(b_i, b_j))$ であるので, $n(a) = n(b)$ であるときには $\{n(z) | z \text{ は } a \text{ の互換}\}$ の要素と $\{n(z) | z \text{ は } b \text{ の互換}\}$ の要素が, 同一 (の構造) であるという関係により 1 対 1 に対応付けできる. したがって, 一般に, x の置換 a , b に対して,

$$n(a) = n(b) \text{ ならば } \{n(z) | z \text{ は } a \text{ の互換}\} = \{n(z) | z \text{ は } b \text{ の互換}\}$$

が成り立つ.

よって, 前述の手順で最終的に得られる U と x の置換 y に対して $n(y) \in U$ ならば, $n(y) = U[m] = n(x(m))$ となる m が存在して,

$$\{n(z) | z \text{ は } y \text{ の互換}\} = \{n(z) | z \text{ は } x(m) \text{ の互換}\}$$

が成り立つ. $U[m]$ を生成した x の置換が $x(m)$ であるので, $x(m)$ の互換はすべて U に含まれ,

$$\{n(z) | z \text{ は } x(m) \text{ の互換}\} \subset U$$

が成り立つ. したがって,

$$\{n(z) | z \text{ は } y \text{ の互換}\} \subset U$$

が成り立つ. すなわち, x の置換 y に対応する化学量論的な代謝ネットワーク構造が U に含まれていれば, y の互換に対応する化学量論的な代謝ネットワーク構造すべてが U に含まれる.

したがって, $x = x(1)$ (x の恒等置換) に対応する化学量

論的な代謝ネットワーク構造 $U[1]$ が U に含まれていることと合わせて, 数学的帰納法により, x の任意の置換 (互換の組み合わせで表される) に対応する化学量論的な代謝ネットワーク構造はすべてが U に含まれることが証明される.

アルゴリズム B は, アルゴリズム A と異なり, すべての置換に対応する化学量論的な代謝ネットワーク構造を算出することなく EFM 型経路の unique な variant を算出する. アルゴリズム B で得られる EFM 型経路の unique な variant に対応する置換の数は, 次項 2.5 に示す計算により間接的に求められる.

2.5 EFM 型経路の unique な variant に対応する置換の数の計算

EFM 型経路の unique な variant が variant 1 から variant λ まで λ 種類あるとき, variant α (α は λ 以下の自然数) に対応する置換の数を v_α と表し, $v = (v_1, v_2, \dots, v_\lambda)$ とする. 求める未知数は $v = (v_1, v_2, \dots, v_\lambda)$ である.

2.2 と同様に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 m_i (i は k 以下の自然数) の代謝産物ノードの個数が $d(m_i)$ であるとする. “対応付け条件” を満たす P - S 間対応付け x に対して, “対応付け条件” を満たす P - S 間対応付けを生成する x の置換 (x により対応付けられた S の要素の置換) が全部で, $r = d(m_1)! \times d(m_2)! \times d(m_3)! \times \dots \times d(m_k)!$ 通り存在する. これらの r 個の x の置換を $x_1 = x, x_2, x_3, \dots, x_r$ と表し, $X = \{x_1, x_2, x_3, \dots, x_r\}$ とする.

P - S 間対応付けにおいて P の要素 p_i, p_j に対応する S の要素を互換することを P 対応互換と呼ぶと, $x_1 = x$ に対する, “対応付け条件” を満たす P - S 間対応付けを生成する P 対応互換が全部で, $c = d(m_1)C_2 + d(m_2)C_2 + d(m_3)C_2 + \dots + d(m_k)C_2$ 通り存在する. これらの c 個の P 対応互換を $h_1, h_2, h_3, \dots, h_c$ と表す. 一般に, x_i (i は r 以下の自然数) に対する, “対応付け条件” を満たす P - S 間対応付けを生成する P 対応互換も $c = d(m_1)C_2 + d(m_2)C_2 + d(m_3)C_2 + \dots + d(m_k)C_2$ 通り存在し, これらは $h_1, h_2, h_3, \dots, h_c$ である. $H = \{h_1, h_2, h_3, \dots, h_c\}$ とする.

i を r 以下の自然数, j を c 以下の自然数とするとき $x_i h_j$ により x_i の P 対応互換 h_j により得られる x の置換を表し, $Xh_j = \{x_i h_j | x_i \in X\}$, $x_i H = \{x_i h_j | h_j \in H\}$ とする.

x の置換 $x_i h_j$ を i 行 j 列のセルの内容とする表 A (行数 r 列数 c) と $n(x_i h_j)$ (x の置換 $x_i h_j$ に対応する化学量論的な代謝ネットワーク構造) を i 行 j 列のセルの内容とする表 B (行数 r 列数 c) を考え, EFM 型経路の variant の種類別に各 variant が入っている B のセルの個数を 2 つの方法で数える. 以下, $x_i h_1, x_i h_2, \dots, x_i h_c$ が左から順に入っている A の i 行目を A の x_i の行, $n(x_i h_1), n(x_i h_2), \dots, n(x_i h_c)$ が左から順に入っている B の i 行目を B の x_i の行と呼ぶ. また, variant α が入っている B のセルの個数 (variant α に対応する x の置換が入っている A のセルの個数でもある) を u_α と表し, $u = (u_1, u_2, \dots, u_\lambda)$ とする.

まず, 列ごとに考えると, X の要素 (x の置換) x_i, x_j が

$x_i \neq x_j$ のとき任意の $h \in H$ に対して x_{ih}, x_{jh} はともに X の要素であり $x_{ih} \neq x_{jh}$ なので, $X = Xh_1 = Xh_2 = Xh_3 = \dots = Xh_c$ が成り立つ. X に等しい $Xh_1, Xh_2, Xh_3, \dots, Xh_c$ はそれぞれ, A の $1, 2, 3, \dots, c$ 列目のセルの内容の集合と一致するので, $u_\alpha = c \times v_\alpha$ であり, E を λ 行 λ 列の単位行列とすれば,

$${}^t u = c \times E \times {}^t v$$

が成り立つ.

次に, 行ごとに考える. variant α に対応する v_α 個の x の置換を $x_\alpha(1), x_\alpha(2), \dots, x_\alpha(v_\alpha)$ とする. A の $x_\alpha(1)$ の行に対応する B の行の variant 1, variant 2, \dots , variant λ が入っているセルの個数をそれぞれ $v_1(\alpha), v_2(\alpha), \dots, v_\lambda(\alpha)$ とする.

$v_1(\alpha), v_2(\alpha), \dots, v_\lambda(\alpha)$ は, $x_\alpha(1)$ に可能な, “対応付け条件” を満たす P - S 間対応付けを生成する 1 回の互換をすべて行い, 得られた $x_\alpha(1)$ の互換に対応する化学量論的な代謝ネットワーク構造の異同判定を行うことにより求まる. このとき, A の $x_\alpha(2), x_\alpha(3), \dots, x_\alpha(v_\alpha)$ の行に対応する B の行についても, 各行の variant 1, variant 2, \dots , variant λ が入っているセルの個数はそれぞれ $v_1(\alpha), v_2(\alpha), \dots, v_\lambda(\alpha)$ である.

2.4 に述べたように, 一般に, x の置換 a, b に対して,

$$n(a) = n(b) \text{ ならば } \{n(z) | z \text{ は } a \text{ の互換}\} = \{n(z) | z \text{ は } b \text{ の互換}\}$$

が成り立つからである. B の $x_\alpha(1), x_\alpha(2), \dots, x_\alpha(v_\alpha)$ に対応する v_α ある行全体では, variant 1, variant 2, \dots , variant λ が, それぞれ, $v_\alpha \times v_1(\alpha), v_\alpha \times v_2(\alpha), \dots, v_\alpha \times v_\lambda(\alpha)$ の個数のセルに入っている.

variant α に対する $v_\alpha \times v_1(\alpha), v_\alpha \times v_2(\alpha), \dots, v_\alpha \times v_\lambda(\alpha)$ を, λ 種類の unique な variant すべてに対して求めると, B 全体の variant 1, variant 2, \dots , variant λ が入っているセルの個数, それぞれ $u_1, u_2, \dots, u_\lambda$ は,

$$\begin{aligned} u_1 &= v_1 \times v_1(1) + v_2 \times v_1(2) + \dots + v_\lambda \times v_1(\lambda) \\ u_2 &= v_1 \times v_2(1) + v_2 \times v_2(2) + \dots + v_\lambda \times v_2(\lambda) \\ u_3 &= v_1 \times v_3(1) + v_2 \times v_3(2) + \dots + v_\lambda \times v_3(\lambda) \\ &\vdots \\ u_\lambda &= v_1 \times v_\lambda(1) + v_2 \times v_\lambda(2) + \dots + v_\lambda \times v_\lambda(\lambda) \end{aligned}$$

と計算できる. β 行 α 列の成分が $v_\beta(\alpha)$ であるような λ 行 λ 列の行列を M とすれば, 計算式をまとめて

$${}^t u = M \times {}^t v$$

が成り立つ.

列ごとに考えて得られた ${}^t u = c \times E \times {}^t v$ と行ごとに考えて得られた ${}^t u = M \times {}^t v$ を合わせた ${}^t u = M \times {}^t v = c \times E \times {}^t v$ から

$$(M - c \times E) \times {}^t v = \text{零行列}$$

が得られる. この式から $(M - c \times E)$ (計算できる) の null space を求め $v_1, v_2, \dots, v_\lambda$ の値の比を得る. さらに,

$$v_1 + v_2 + \dots + v_\lambda = r$$

が成り立つことを利用して $v_1, v_2, \dots, v_\lambda$ の値を確定する.

アルゴリズム B では, 各置換に対して可能な, “対応付け条件” を満たす P - S 間対応付けを生成する 1 回の互換すべてを行い, 得られる P - S 間対応付けに対応する化学量論的なネットワーク構造すべての中で異同判定を行うので, その際に各 unique な variant を生成する互換の数 (行列 M の作成に必要) を記録すれば, 行列 M の算出とアルゴリズム B を同時に行うことが可能である.

2.6 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

2.3, 2.4, 2.5 における 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は, グラフ同型判定アルゴリズム $+EC_pGC_{Core}$ [11] により行う.

2.3, 2.4, 2.5 における 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においては, 3 つ以上のグラフ構造間の異同判定が必要となる場合がある. グラフ構造が 3 つ以上ある場合は, 1 つのグラフ構造を, 残りの 2 つ以上のグラフ構造と, 1 つ 1 つ異同判定しなければならない.

$+EC_pGC_{Core}$ では, グラフ構造の number_of_Map 次までの隣接行列, グラフ G のグラフ ${}_pG$ への変換, EC 値 [12,13] の計算が必要であるが, 2.3, 2.4, 2.5 における異同判定においては, それぞれの化学量論的な代謝ネットワーク構造であるグラフ構造のこれらの計算結果を保持しておき, 3 つ以上のグラフ構造の異同判定のための一連の過程のなかで繰り返し利用する.

3. 計算実験の結果と考察

EFM 型経路の unique な variant と各 variant に対応する置換の数を 2 つの方法 (異同判定の結果の記録を伴うアルゴリズム A, および, 行列 M の算出を伴うアルゴリズム B + 行列 M を用いる 2.5 の計算) により求める計算実験を行った. 本項では, EFM 型経路の unique な variant と各 variant に対応する置換の数の算出を, 括弧内の前者の方法で行った場合をアルゴリズム A, 括弧内の後者の方法で行った場合をアルゴリズム B と記す. 行ったすべての計算実験において, アルゴリズム A とアルゴリズム B で算出される, EFM 型経路の unique な variant と各 variant に対応する置換の数は完全に一致した. 1 つの EFM 型経路で, unique な variant に対応する置換の数が variant ごとに異なる場合があった. また, アルゴリズム B による計算はアルゴリズム A による計算よりも高速であった. 以下に, 計算実験の結果の例を示す. 計算はすべて Matlab 上で行った (Intel(R) Core(TM) i7-8565U CPU @ 1.80GHz 1.99GHz).

反応 R1 を $M1 + M2 \rightarrow M3 + M4$, 反応 R2 を $M3 + M4 \rightarrow M5 + M6$ とするとき, 2 つの R1 と 2 つの R2 により $2M1 + 2M2$ から $2M5 + 2M6$ が生成する EFM 型経路に対しては,

アルゴリズム A, アルゴリズム B ともに, 2 つの unique な variant をそれぞれ 32, 32 と同数の置換に対応するものとして算出した. 計算時間は, アルゴリズム A, アルゴリズム B それぞれ 0.30 秒, 0.067 秒であった. 3 つの R1 と 3 つの R2 により 3M1+3M2 から 3M5+3M6 が生成する EFM 型経路に対しては, アルゴリズム A, アルゴリズム B ともに, 3 つの unique な variant をそれぞれ 7776, 23328, 15552 の置換に対応するものとして算出した. 計算時間は, アルゴリズム A, アルゴリズム B それぞれ 404 秒, 0.482 秒であった.

糖質代謝モデルネットワーク [14,15] 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 [14,15] 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 EFM 型経路を EFM1 と呼び, EFM1 に含まれる反応を, それぞれの数を 2 倍にして組み合わせた, 化学量論的に 4 分子の fructose 6-phosphate と 2 分子の glyceraldehyde 3-phosphate から 6 分子の ribose 5-phosphate を生成する EFM 型経路を EFM2 と呼ぶ [11] とし, EFM1 に対しては, アルゴリズム A, アルゴリズム B ともに, 2 つの unique な variant をそれぞれ 48, 48 と同数の置換に対応するものとして算出した. 計算時間は, アルゴリズム A, アルゴリズム B それぞれ 0.36 秒, 0.070 秒であった. EFM2 に対しては, アルゴリズム B は, 8 つの unique な variant を, そのうち 4 つはそれぞれ 79626240 の置換に対応するものとして, 残りの 4 つはそれぞれ 159252480 の置換に対応するものとして算出した. 計算時間は 2.52 秒であった. アルゴリズム A については, 955514880 の置換に対応する, 異同判定前の化学量論的なネットワーク構造の 0.01% に相当する 100000 構造を算出した時点で計算時間が 209 秒に達したので, 計算実験を途中で停止した.

ところで, ここまで述べてきた EFM 型経路の unique な variant に対応する置換の数の計算にはどのような意味があるのだろうか. その問いに対する 1 つの回答として, 各 variant が何種類の異なる置換から生成し得るかの情報から, 各 variant に対応する型の, EFM 型経路における代謝産物分子の生成・利用状態の起こりやすさがわかるとする考え方を説明する. その考え方は以下の枠組みに基づく.

まず, 2.1 の変換後の仮想原料反応・仮想目的反応を含む 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各代謝産物ノードを 1 個の代謝産物分子, 各反応ノードを 1 個の酵素分子とみなす (仮想原料反応・仮想目的反応の反応ノードも 1 個の酵素分子とみなす). そして, 系内には, 与えられた EFM 型経路の 2.2 でのべた集合 P に含ま

れる代謝産物ノードに対応する代謝産物分子がちょうど 1 セットと反応ノードに対応する酵素分子がちょうど 1 セット (セット内の酵素分子は EFM 型経路を構成する反応と 1 対 1 に対応) 存在し, 系は, 単位時間に 1 セットの代謝産物分子が, EFM 型経路を構成する反応 (1 セットの酵素分子) により基質として 1 回利用されると同時に生成物として 1 回生成 (再生) されることを繰り返す定常状態にあるものとする. 代謝産物ノードに対応する, 系内の, ある瞬間の 1 セットの代謝産物分子のうちのそれぞれは, 反応ノードに対応する系内の特定の酵素分子から生成したものである. それらは, 次の瞬間には, 反応ノードに対応する系内の特定の酵素分子の基質として利用される. この枠組みにおいて, EFM 型経路の化学量論的な代謝ネットワーク構造は, ある瞬間に存在する EFM 型経路上の各代謝産物分子を直接生成した反応 (酵素分子) と次の瞬間に直接利用する反応 (酵素分子) を示すもの, すなわち, EFM 型経路における代謝産物分子の生成・利用状態のスナップショットとして理解できる. ここで, 2 つの化学量論的な代謝ネットワーク構造が異同判定により同一構造であると判定されるとき, それらは, 1 つの同じ型の, 代謝産物分子の生成・利用状態を示すと定めれば, EFM 型経路の unique な variant を, それぞれが, 異なる 1 つの, 代謝産物分子の生成・利用状態の型を示すものと理解して, 代謝産物分子の生成・利用状態の型と EFM 型経路の unique な variant を 1 対 1 に対応付けることができる.

“対応付け条件”を満たす P - S 間対応付けは, EFM 型経路の 1 variant の置換である. 置換それぞれは, どの反応ノードに対応する系内の特定の酵素分子の生成物であるかが与えられている各代謝産物分子の利用状態 (各代謝産物分子が, 次にどの反応ノードに対応する系内の特定の酵素分子の基質として利用されるか) に対応し, 代謝産物分子の生成・利用状態を決定する. 各代謝産物分子と各酵素分子の遭遇しやすさが同じであると仮定すれば, EFM 型経路の 1 variant の置換それぞれに対応する, 各代謝産物分子の利用状態は, 同じ確率で生起すると考えられる. すなわち, EFM 型経路の 1 variant の置換それぞれに対応する, 代謝産物分子の生成・利用状態は, 等しい確率で生起すると考えられる. したがって, 各 variant が何種類の異なる置換から生成し得るかの情報から, 各 variant に対応する型の, EFM 型経路における代謝産物分子の生成・利用状態の起こりやすさがわかる (より多くの置換から生成し得る variant に対応する型の, 代謝産物分子の生成・利用状態ほど起こりやすい) と考えられる. たとえば, EFM2 の 8 つの unique な variant に対応する型の代謝産物分子の生成・利用状態のうち, それぞれ 159252480 置換に対応する 4 つの variant に対応する型の状態は, それぞれ 79626240 置換に対応する残りの 4 つの variant に対応する型の状態よりも起こりやすいという解釈が可能である.

参考文献

- [1] Schuster, S.; Hilgetag, C. On elementary flux modes in biochemical reaction systems at steady state. *J. Biol. Syst.* **1994**, *2*, 165-182.
- [2] Schilling, C. H.; Palsson, B.O. The underlying pathway structure of biochemical reaction networks. *Proc. Nat. Acad. Sci. USA* **1998**, *95*, 4193-4198.
- [3] 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.
- [4] 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.
- [5] 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.
- [6] 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.
- [7] 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
- [8] 太田 潤 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. 情報処理学会研究報告, **2020**, Vol.2020-BIO-62, No. 3, 1-6.
- [9] 太田 潤 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算出の高速化. 情報処理学会研究報告, **2020**, Vol.2020-BIO-63, No. 1, 1-7.
- [10] 太田 潤 Elementary flux mode 型代謝経路の化学量論な代謝ネットワーク構造の算出: 算出されたネットワーク構造間の異同判定の高速化. 情報処理学会研究報告, **2020**, Vol.2020-BIO-64, No. 5, 1-7.
- [11] 太田 潤 Elementary flux mode 型代謝経路の化学量論な代謝ネットワーク構造の算出: 算出されたネットワーク構造間の異同判定のためのアルゴリズムの最適化. 情報処理学会研究報告, **2021**, Vol.2021-BIO-65, No. 4, 1-8.
- [12] Gasteiger, J.; Engel, T.; 船津公人; 佐藤寛子; 増井秀行 ケモインフォマティクス - 予測と設計のための化学情報学. 丸善株式会社, **2005**.
- [13] Morgan, H. L. The generation of a unique machine description for chemical structures – A technique developed at Chemical Abstracts Service. *J. Chem. Doc.* **1965**, *5*, 107-113.
- [14] 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
- [15] Ohta, J. Single-atom tracing in a model network of carbohydrate metabolism and pathway selection. *IPJS Transactions on Bioinformatics* **2018**, *11*, 1-13.