

博士論文

確率的手法に基づいた大規模
代謝シミュレーション手法の構築

桂樹 哲雄

2014年9月17日

奈良先端科学技術大学院大学

情報科学研究科

本論文は奈良先端科学技術大学院大学情報科学研究科に博士(工学)授与の要件として提出した博士論文である。

桂樹 哲雄

審査委員:

| | |
|----------------------|---------|
| 金谷 重彦 教授 | (主指導教員) |
| 安本 慶一 教授 | (副指導教員) |
| Md.Altaf-Ul-Amin 准教授 | (副指導教員) |
| 杉浦 忠男 准教授 | (副指導教員) |
| 小野 直亮 助教 | (副指導教員) |
| 佐藤 哲大 助教 | (副指導教員) |

確率的手法に基づいた大規模 代謝シミュレーション手法の構築

桂樹 哲雄

内容梗概

実験で質量分析装置から得られる生体内の代謝物の濃度は通常相対値であり、量的計量を行うためには、得られた結果をもとに、シミュレーションによって個々の代謝の挙動を明らかにする手法が有効である。本研究では、植物の大規模代謝シミュレーションシステムの開発を目指し、以下の3つの課題、すなわち、(1) シミュレーションに必要なパラメータの数が膨大であること、(2) 決定論的定式化を行うとシステムの確率依存的ふるまいを記述できないこと、(3) 質量分析装置から得られる代謝量は絶対定量ではなく、また全ての代謝物を含んでいるわけではないこと、に着目する。これらの課題について、(1) 着目すべきサブネットワークのみを抽出すること、(2) 確率的シミュレーション手法を用いること、(3) 遺伝的アルゴリズムおよび分散遺伝的アルゴリズムによるパラメータ推定を行うこと、によってそれぞれ対処することとし、3つのツール(SS-mPMG、SS-GA、SS-dGA)を開発した。SS-mPMGは、ゲノム全体の代謝ネットワークから部分ネットワークを抽出し、代謝モデルの複雑性を軽減して動的ネットワークシミュレータを自動で構築するツールである。SS-GA、SS-dGAは、実験で観測された代謝の挙動を再現できるようなパラメータを、それぞれ遺伝的アルゴリズムまたは分散遺伝的アルゴリズムによって推定し、それらの推定したパラメータを用いて確率的シミュレーションを実行するツールである。これらのツールにより、シロイヌナズナの実験的に観測されたアミノ酸生成を再現できることを示した。これらのツールが代謝実験データを解析する一助となることを期待する。

キーワード

シロイヌナズナ、バイオインフォマティクス、分散遺伝的アルゴリズム、代謝、確率的シミュレーション

*奈良先端科学技術大学院大学情報科学研究科情報科学専攻学位論文, NAIST-IS-DD1161201, 2014 年 9 月 17 日.

Dynamic simulation of metabolic network based on stochastic approach

Tetsuo Katsuragi

Abstract

Metabolite concentrations are generally obtained as relative-levels with mass spectrometry. For getting further information of dynamic behavior of metabolites, it is useful to create simulation tools that can reproduce the experimental data. In this study, we focus on the truth that (1) the number of the parameters needed for the simulation is huge, (2) stochastic behaviors of metabolites cannot be described by a deterministic approach, and (3) the amount of the metabolites obtained from experiments are generally in relative levels. These are solved by considering that (1) extracting a subsystem of the metabolic network from the genome-scale pathway maps, (2) using a stochastic simulation, and (3) the parameters needed for the simulation are determined using a genetic algorithm and a distributed genetic algorithm. In this study, we propose three tools; the minimum pathway model generator tool for simulating the dynamics of metabolite concentrations (SS-mPMG) and two other tools for parameter estimation by genetic algorithm (SS-GA) and distributed genetic algorithm (SS-dGA). In order to reduce the complexity of the simulation model, SS-mPMG can extract a subsystem of the metabolic network from the genome-scale pathway maps, and it automatically constructs a dynamic simulator to evaluate the observed behavior of metabolites. Results show that stochastic simulation can reproduce dynamics of amino acid biosynthesis observed in *Arabidopsis thaliana*. These software will be helpful for researchers to create relevant metabolic networks and carry out simulations of metabolic reactions derived from metabolomics data.

Keywords:

Arabidopsis thaliana, Bioinformatics, Genetic algorithm, Metabolism, Stochastic simulation

*Doctoral Dissertation, Department of Information Systems, Graduate School of Information Science, Nara Institute of Science and Technology, NAIST-IS-DD1161201, Sep 17, 2014.

目次

| | |
|---|-----|
| 目次..... | iii |
| 図目次..... | v |
| 表目次..... | vi |
| 第1章 はじめに..... | 1 |
| 第2章 サブネットワーク抽出による代謝を再現するモデルの自動構築..... | 7 |
| 2.1 序論..... | 7 |
| 2.1.1 代謝とは..... | 8 |
| 2.1.2 メタボローム解析結果をシミュレーションに用いる場合の問題 点..... | 9 |
| 2.1.3 代謝シミュレーション研究の現状と課題..... | 12 |
| 2.1.4 代謝パスウェイ・データベース..... | 14 |
| 2.2 解析対象に用いた代謝物の情報..... | 15 |
| 2.3 方法..... | 16 |
| 2.3.1 Dijkstra アルゴリズム..... | 17 |
| 2.3.2 最短経路探索によるサブネットワークの選択方法..... | 18 |
| 2.3.3 Gillespie アルゴリズムに基づいた確率的シミュレーション..... | 19 |
| 2.4 結果と考察..... | 22 |
| 2.4.1 大規模代謝ネットワークシミュレータプログラム SS-mPMG..... | 22 |
| 2.4.2 実験で得られた代謝の挙動..... | 24 |
| 2.4.3 代謝パスウェイ・サブネットワークの抽出..... | 28 |
| 第3章 代謝シミュレーションのためのパラメータ推定手法の開発..... | 31 |
| 3.1 序論..... | 31 |
| 3.1.1 遺伝的アルゴリズム..... | 32 |

| | | |
|-------|---------------------------------|-----|
| 3.1.2 | 分散遺伝的アルゴリズム | 34 |
| 3.2 | 方法 | 36 |
| 3.2.1 | 遺伝的アルゴリズムを用いた推定方法 | 37 |
| 3.2.2 | 分散遺伝的アルゴリズムを用いた推定方法 | 41 |
| 3.3 | 結果と考察 | 43 |
| 3.3.1 | 大規模代謝シミュレータ・プログラム SS-GA, SS-dGA | 43 |
| 3.3.2 | シロイヌナズナの代謝データを用いたモデルの評価 | 45 |
| 3.3.3 | 計算によって示された代謝実験の解釈 | 51 |
| 第4章 | おわりに | 67 |
| | 謝辞 | 69 |
| | 参考文献 | 72 |
| | 付録 | 78 |
| | 実験の詳細 | 78 |
| | 付図 | 79 |
| | 付表 | 87 |
| | SS-mPMG の詳細 | 99 |
| | SS-GA, SS-dGA の詳細 | 106 |
| | 関連業績 | 113 |
| | 参考業績 | 114 |

図目次

| | |
|---|----|
| 図 1 本研究で開発したプログラムの機能と本論文の全体の流れ..... | 6 |
| 図 2 遺伝子から表現型までの情報の流れとオーム解析..... | 11 |
| 図 3 SS-mPMG および SS-GA で用いられる確率的シミュレーションの概要 | 21 |
| 図 4 SS-mPMG による代謝ネットワークのモデル化の概要..... | 23 |
| 図 5 PCA 解析による代謝の全体的挙動..... | 26 |
| 図 6 実験結果に基づき選ばれた代謝ネットワーク..... | 30 |
| 図 7 本研究で用いる遺伝的アルゴリズムの手順を表す概念図..... | 39 |
| 図 8 適応度の計算方法..... | 40 |
| 図 9 本研究で用いたアルゴリズムの概念図 (a) 遺伝的アルゴリズムと (b) 分散遺伝的アルゴリズムの違い..... | 42 |
| 図 10 SS-GA および SS-dGA による代謝ネットワークのモデル化の概要..... | 44 |
| 図 11 適応度の分布の比較..... | 48 |
| 図 12 適応度の履歴の比較..... | 49 |
| 図 13 移住率と最終的な適応度の分布..... | 50 |
| 図 14 反応係数の変化の分布 (A) 遺伝的アルゴリズムによる結果 (B) 分散 遺伝的アルゴリズムの結果..... | 58 |
| 図 15 反応係数が変化した反応..... | 60 |
| 図 16 シミュレーションにより再現された各代謝物の挙動の比較..... | 61 |
| 図 17 5'メチルチオアデノシンと α ケトグルタル酸の予測値が実験値と離 れた理由..... | 65 |
| 図 18 実験でデータが得られなかった代謝物について、シミュレーションに よって予測された時間変化..... | 66 |

表目次

| | |
|--|----|
| 表 1 実験結果に基づいて選ばれた始端化合物と終端化合物..... | 27 |
| 表 2 初期相と後期相で顕著に変化の見られた反応 (A) 遺伝的アルゴリズムによる結果 (B) 分散遺伝的アルゴリズムによる結果 | 56 |

第1章 はじめに

地球上の生物は、多種多様な形態を持っているが、その多様性にも関わらず、全ての生物は細胞という共通の構造体を基本単位とする。細胞の形態もまた多種多様であるが、細胞の構造と活動の仕組みは非常に似通っている。全ての細胞は生命活動を行うにあたって多様なタンパク質を生合成して利用している。タンパク質は、分子生物学上の重要な概念であるセントラルドグマの仕組みによって遺伝子から生合成される (Crick 1958)。すなわち、デオキシリボ核酸 (DNA) を鋳型として遺伝情報がメッセンジャーRNA (mRNA) に転写され、そこからアミノ酸からなるタンパク質へと翻訳される。この意味で、遺伝情報の総体であるゲノムは生命の設計図といえる。

生命の構築原理を解明することを目的として、セントラルドグマのそれぞれの過程において得られる分子情報を、生命の設計図であるゲノムとその活動の写像であると捉え、それぞれの階層で網羅的に解析するオーム解析が盛んに行われてきた (Joyce & Palsson 2006; Fiehn et al. 2001)。DNA の塩基配列の網羅的解析 (genome、ゲノム解析)、遺伝子転写産物 mRNA の網羅的解析 (transcriptome、トランスクリプトーム解析)、タンパク質の網羅的解析 (proteome、プロテオーム解析) である。これらのオーム解析の次の段階として、細胞の活動によって生じるアミノ酸、有機酸、脂肪酸などの低分子化合物をはじめとした代謝物の網羅的解析 (metabolome、メタボローム解析) が近年盛んに行われるようになってきた (Hirai et al. 2005)。ゲノム解析、トランスクリプトーム解析、プロテオーム解析が対象とする分子は、基本骨格が同一であるのに対し、メタボローム解析の対象である代謝物質は、その形態が多様であるために、これらを網羅的に検出することが可能となったのは、質量分析装置とそれに関連する技術が発展したごく最近のことである (Sumner et al. 2003; Bedair & Sumner 2008; Sawada et al. 2009)。

現在では、高スループットの機器によって、これらの濃度プロファイルを時系列に解析することも可能である (Buchholz et al. 2002; Sato et al. 2008)。これらの化合物が含まれる代謝経路の情報は、KEGG (Kyoto Encyclopedia of Genes and Genomes) (Kanehisa & Goto 2000) や MetaCyc (<http://metacyc.org/>) (Caspi et al. 2012) などのデータベースにおいて公開され、また、実験により得られる代謝に関する情報の質が向上したため、これらの化合物が関わる代謝パスウェイの多くについて、情報付けを行うことができる。

一般的に、実験から得られる代謝物の量は、質量分析装置から得られる代謝物ごとの量の相対値、つまり、代謝物スケールの時系列データであるため、代謝物の絶対量を得ることはできない。また、実験で得られるデータは代謝ネットワーク全体の代謝物のデータを含んでいるわけではないため、システム全体の挙動を実験から直接観察することはできない。そこで、実験によって代謝物のデータをもとに、何らかのリバースエンジニアリングによって、これらの未知のデータが潜んでいる代謝システムの詳細を明らかにすることが重要な課題である (Tomita et al. 1999; Edwards et al. 2000; Feist & Palsson 2008; de Oliveira Dal’Molin et al. 2010)。これを達成するためには、実験データを再現できる数理モデルを構築することが必要である。

本研究では、代謝実験データを再現する大規模代謝パスウェイ・ネットワークモデルを構築することを目標として、実験から得られたシロイヌナズナの代謝反応の挙動をシミュレーションで再現し、代謝のダイナミクスを明らかにすることを旨とする。その際の問題点として以下の3つの点に着目し、これらを解決する。

まず、はじめに、代謝シミュレーションに必要なパラメータの数が膨大であるという点に着目する。これまで、主にバクテリアのデータを用いて代謝システム

をモデル化する試みが多くの研究で行われてきた (Tomita et al. 1999; Ishii et al. 2004; Karr et al. 2012)。生体内すなわち細胞の全ての代謝物の動的な挙動をモデル化するには、関わるすべての反応の反応係数と化合物の初期濃度が必要である。バクテリアのような単純な生物においても、関わる代謝物の数は約 1000 種といわれており (Feist & Palsson 2008)、必要とされるパラメータの数が非常に大きいという問題がある。植物の代謝を対象とした場合、バクテリアに比べて反応の数はさらに膨大となり、システムはさらに複雑となる (Morgan & Rhodes 2002; Schwender et al. 2004; Collakova et al. 2012)。そこで、巨大な代謝パスウェイ・ネットワークのうち、反応が大きく変化している一部の代謝パスウェイのみをモデル化する試みがなされている。本研究においては、代謝サブネットワークの抽出を行ってモデルを構築する。すなわち、ゲノムにコードされている酵素遺伝子情報における全ての代謝物の反応を元にして、代謝データの解析により得られる代謝変化の情報から、変化に寄与したと考えられる代謝パスウェイに焦点を当てたサブネットワークを対象を絞り込んで数理モデルを構築し、細胞全体の代謝の動的挙動を理解することを目的とする。

第 2 番目に、代謝システムの確率過程的な振る舞いについて着目する。これまで、よく混合された反応系の反応過程のモデル化を行う場合には、連立した常微分方程式を解く方法が多く用いられてきた。このような決定論的なアプローチに加え、生体内のわずか数個の代謝物が不連続かつ確率的なふるまいを引き起こすような、確率過程的システムをとらえるアプローチも必要となってきた (McAdams & Arkin 1997; Arkin et al. 1998)。また、このような確率過程的システムは、今挙げたような代謝物の数が少ない場合にのみ見られるものではなく、生体内の代謝物の量が多い場合にもみられることが報告されている (Kummer et al. 2005; Pahle et al. 2012)。生化学システムに内在する、これらの確率過程的システ

ムの変動は、確率的シミュレーションの Gillespie アルゴリズム (Gillespie 1976; Gillespie 1977) を含む、マルコフ過程の仕組みを用いることによってシミュレートすることが可能である。決定論的なシミュレーションと確率的シミュレーションを組み合わせることによって代謝の動的な振る舞いを調べるのが簡単になるようになれば、生体における代謝をシステムの流れとしてより深く理解することができる。そのためには、決定論的ならびに確率的なシミュレーションの両方の発展が必要である。そこで、本研究では、確率的なシミュレーション手法として、Gillespie のアルゴリズムを基礎においた代謝シミュレーションシステムを開発した。

第 3 番目に、シミュレーションに必要な全てのパラメータが既知とは限らないという点に着目する。代謝の挙動をシミュレートするためには、全ての反応の反応定数と全ての代謝物の初期濃度が必要である。しかし、反応定数は必ずしも全て既知というわけではない。また、質量分析装置から得られる代謝物の量は絶対量ではなく、また、全ての代謝物の量を計測できるわけではないことから、代謝物の初期濃度についても同様に既知ではない。このため、シミュレーションの結果を実験結果に合わせられる最適なパラメータセットを設定する必要がある。つまり、相対値で表された実験結果に合うように、計算結果と実験結果におけるそれぞれの代謝物の初期濃度からの相対変化を近づけるように、シミュレーションに必要なパラメータを全て設定する必要がある。しかしながら、代謝ネットワークの抽出によってモデルを簡略化した場合でも、必要なパラメータの組み合わせの数は非常に大きいため、これらを手動で設定するのは困難である。そこで、パラメータ設定を行うため、進化的アルゴリズムの一種である遺伝的アルゴリズムおよび、分散遺伝的アルゴリズムに基づいたパラメータ推定法を検討した。

本研究では、これらの3つの点を考慮した代謝解析シミュレータを研究開発するとともに、実際の代謝データを用い、代謝の流れのシミュレーションを行い、分子生物学による代謝の解釈を行った。本論文の構成は以下のとおりである。全体の流れを図1に示す。まず、与えられた生物種の全てのパスウェイから、実験に関連する適切な代謝パスウェイのみからなるサブシステムを抽出し、さらにそのサブシステムを対象として確率的シミュレーションを行うソフトウェア SS-mPMG (Stochastic Simulator with minimum Pathway Model Generator) を開発した(第2章)。抽出したサブネットワーク内のそれぞれの反応定数が既知の場合に、SS-mPMGによって、代謝反応の振る舞いをシミュレートし、実験データと比較することによってモデルを評価することができる。次に、実験で得られたシロイヌナズナの代謝データを再現するパラメータを、遺伝的アルゴリズムおよび分散遺伝的アルゴリズムによって推定し、代謝の挙動をシミュレーションにより明らかにするソフトウェア SS-GA (Stochastic Simulator with Genetic Algorithm) と SS-dGA (Stochastic Simulator with distributed Genetic Algorithm) を開発した(第3章)。これらの進化的アルゴリズムに基づいたパラメータ最適化が、代謝シミュレーションに必要なパラメータ評価に適用可能であることと、開発したモデルが観測された反応経路の途中にある代謝物の振る舞いを量的に再現することが可能であることを示す。

以上の結果から、ゲノム全体の反応リストからパスウェイ・ネットワークのサブセットを生成する、シミュレーションに必要なパラメータを設定する、得られた代謝ネットワークのシミュレーションを行うという操作が可能となった。さらに、遺伝的アルゴリズムおよび分散遺伝的アルゴリズムがこれらのシミュレーションに必要なパラメータを設定するのに有効であることを示した。

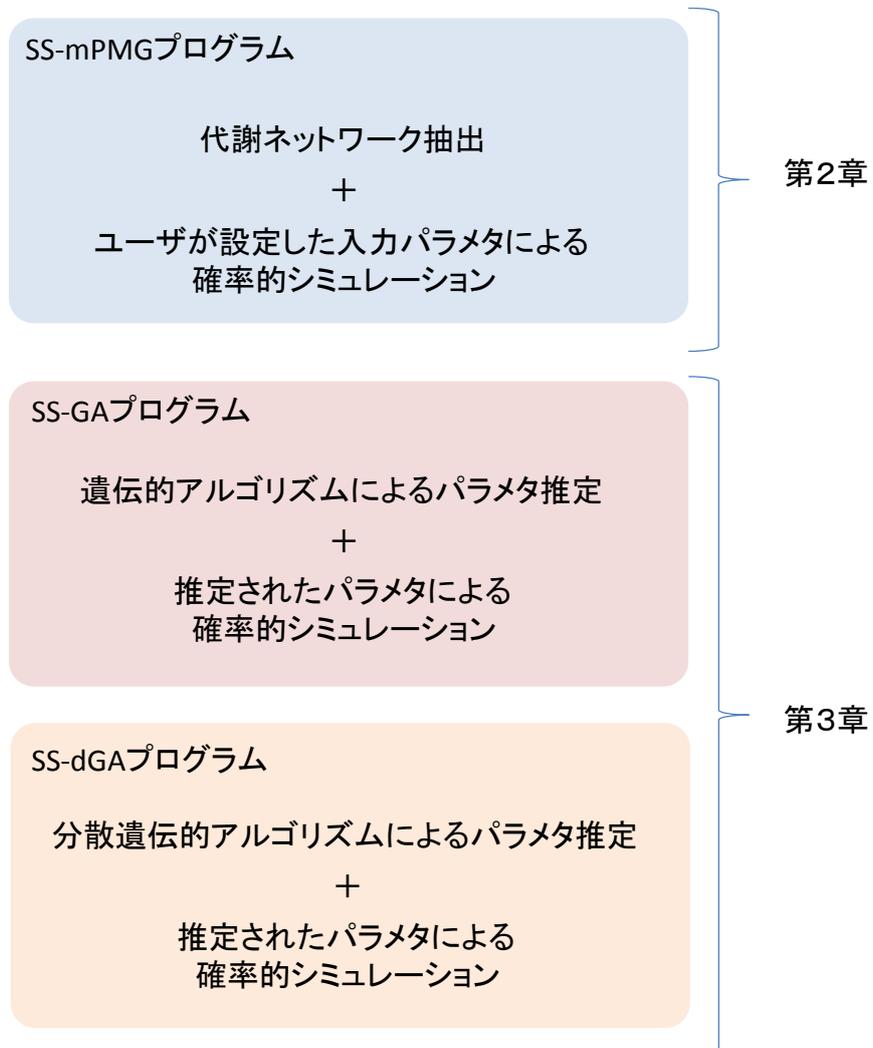


図 1 本研究で開発したプログラムの機能と本論文の全体の流れ

第2章 サブネットワーク抽出による代謝を再現するモデルの自動構築

2.1 序論

実験で得られたシロイヌナズナの代謝ダイナミクスを表現しうるサブネットワークの自動構築を行い、システムの確率過程的な挙動を再現できるシミュレーションツール SS-mPMG を開発した。第2章では、この SS-mPMG の詳細について述べる。

本手法の背景として、代謝について解説した後、代謝シミュレーションを行うにあたって、その元となるデータとしてメタボローム解析のデータを用いることの問題点を示し、代謝シミュレーションの現状を解説した後に、代謝パスウェイ・データベースについて述べる。2.2では、解析に用いる代謝実験の情報を述べる。2.3では、本章で用いる方法の戦略、最短経路探索と Gillespie のアルゴリズムについて述べ、2.4では、開発したツールによるパスウェイ探索の結果とその考察を述べる。

2.1.1 代謝とは

生物の体内では、外部から取り込まれた物質が新たな別の物質へと変化する化学変化が連鎖的に起こっている。このような化学変化は代謝反応と呼ばれ、代謝反応に関わる物質を代謝物と呼ぶ。代謝の主な役割は二つある。一つ目は、外部から吸収した食物からエネルギーを得ること（エネルギー代謝）であり、二つ目は、外部から取り込まれた物質から生物が必要とする物質を合成すること（物質代謝）である。生物は、代謝によって運動に必要なエネルギーや成長、増殖に必要な物質を生産しているため、代謝の挙動、すなわち代謝物量の変化を調べることは、生命活動を理解する上で大変重要である。

生物は、体内の代謝物量の調節を、酵素の触媒作用を調整することで行っている。触媒作用とは、酵素によって反応速度が桁違いに高速化される作用であり、一般的に、代謝反応の速度は触媒作用なしでは極端に遅く、触媒がないか、あるいは十分には存在しない場合、対応する反応はほとんど起きない。生物の体内で触媒作用を担う酵素は、ほぼ全てタンパク質であり、そのタンパク質は、遺伝子に含まれる遺伝情報がメッセンジャーRNA (mRNA) に転写され、さらに翻訳されるという分子生物学上の重要な概念であるセントラルドグマに沿って作られる (Crick 1958)。生物の体内において、代謝反応の速度は、この一連の流れの中で、転写や翻訳の効率を変化させ、酵素の量を変化させることで調整されている。本研究では、この調整の結果として実験で観測された代謝物量の変化をもとに、その代謝物量の変化を再現する代謝シミュレーションツールを開発する。

2.1.2 メタボローム解析結果をシミュレーションに用いる場合の問題点

オームとは、総体を現す言葉であり、オーム解析とは、生体内のあらゆる分子情報を網羅的に解析する手法である。生体内では、遺伝子に蓄えられた遺伝情報がRNAに転写され、さらに翻訳されてタンパク質が作られる。細胞の機能を包括的に理解することを目的として、これらの分子情報を網羅的に解析するオーム解析が行われてきた。すなわち、ゲノム解析、トランスクリプトーム解析、プロテオーム解析である。これらの次の段階として、セントラルドグマによって作成されたタンパク質による代謝の変化を捉えるために、代謝物（メタボライト）のオーム解析であるメタボローム解析が行われるようになった (Joyce & Palsson 2006; Ryan & Robards 2006) (図 1)。メタボローム解析が対象とする代謝は、ゲノム情報が転写され、翻訳されてできあがる酵素に基づいた小分子の化学変化であるので、細胞の成分と直結する。つまり、ゲノム情報と表現型をつなげる有力な情報がメタボロームであり、メタボロミクスは、生物学の分野だけでなく、医薬品、食料品など、様々な分野で応用されている (Rochfort 2005)。

メタボローム解析において、代謝物の分離と同定には主にクロマトグラフィーを伴った質量分析が用いられる。質量分析装置 (mass spectrometer, MS) は、分子の質量を測定する装置であり、その名のとおり、質量 (マス) を分けて (スペクトル) 計測する (メトリー) 装置である。質量分析装置により、分子の質量は非常に正確に測定できるが、全く同じ分子量をもつ物質を分けることができないため、通常、クロマトグラフィーなどの装置を組み合わせる。質量分析装置による測定では、抽出とイオン化の効率が化合物によって異なる。そのため、ターゲットが決まっている場合には、ターゲットの定量用標準試料を用いるこ

とで、厳密な定量分析が可能である。しかしながら、網羅的解析を目指すメタボローム解析においては、従来の定量用標準試料を用いた定量は困難である。そのため、通常メタボロームデータは代謝物スケールの相対値でしか得られないことが問題となる。そこで、本研究では、第3章において、代謝シミュレーションに必要な各代謝物の初期濃度とその他のパラメータを遺伝的アルゴリズムあるいは分散遺伝的アルゴリズムを用いて推定する。

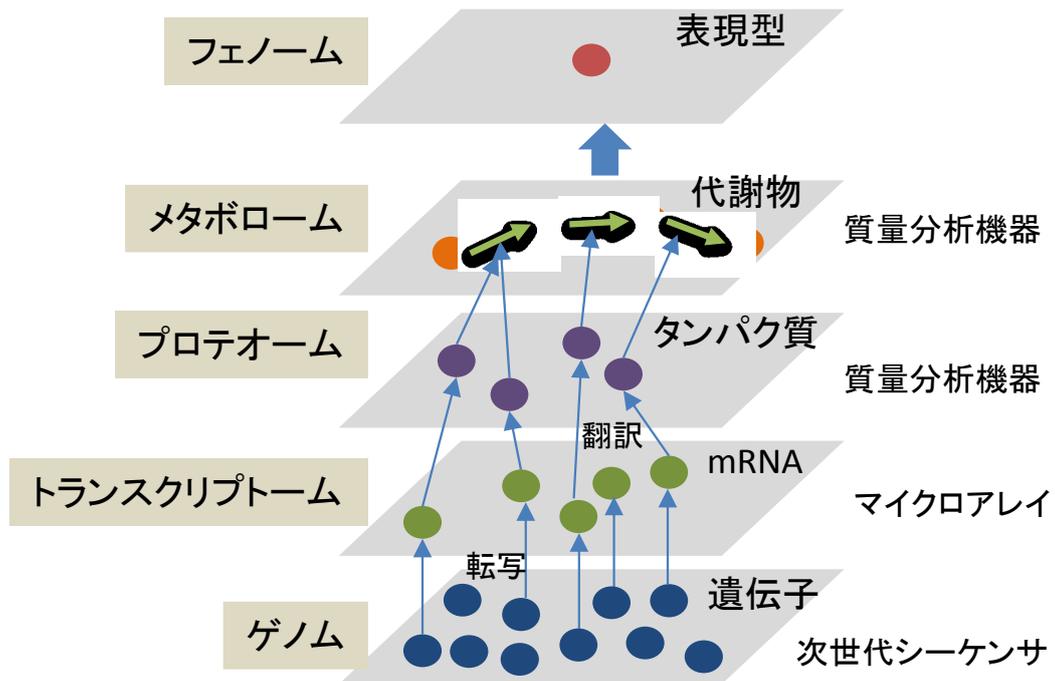


図 2 遺伝子から表現型までの情報の流れとオーム解析

遺伝子の情報は、セントラルドグマに基づいて、図の下の遺伝子の階層から上のタンパク質の階層に向かって伝達される。遺伝情報をもとに作成されたタンパク質は触媒として代謝反応を制御し、最終的に遺伝情報が表現型として現れる。オーム解析とは、この遺伝情報の流れに沿って、それぞれの階層の情報を図右側に挙げた装置を用いて網羅的に解析する手法である。

2.1.3 代謝シミュレーション研究の現状と課題

実験によって得られたメタボロームデータは、生体内の代謝産物の変化として解釈する必要がある。しかし、実験によって全ての代謝物の変化が測定できるわけではなく、また、質量分析装置によって得られるデータは通常代謝物スケールの相対値で、絶対量ではない (第 2.1.2 節参照)。そこで、実験の裏に潜む代謝の挙動を量的にモデル化するツールが必要となる。

第 2.1.1 節で述べたように、生物は、転写、翻訳といったセントラルドグマの段階、タンパク質-タンパク質相互作用など、様々な階層において代謝反応の調整を行っている。これら全ての階層の調整を再現したセルスケールの予測モデルを構築することは、植物代謝シミュレーションにおける究極の目標であるが、植物の生命システムは巨大であり、現状では困難である。セルスケールの予測モデルとしては、ゲノムサイズが最小である 525 個の遺伝子を持つマイコプラズマを対象として、細胞内で起きているイベントを 28 個のモジュールに分け、それらをそれぞれに最適な数理モデルで計算して統合するという方法で全細胞シミュレーションを行っている (Karr et al. 2012)。しかしながら、顕花植物では最小程度の遺伝子数をもつシロイヌナズナにおいても、遺伝子数は約 2 万 6000 個と非常に多く (The Arabidopsis Genome Initiative, 2000)、植物においてこれら全ての階層をモデルで再現するのは依然として難しい。このため、これらの代謝のみを適切に考慮したモデルが十分有効であり、実験データを計算機内で精度良く再現する代謝シミュレーション・プログラムは大変有用であると考えられる (Buchanan et al. 2010)。実験において発現したと考えられる代謝反応をつないだ代謝パスウェイは、実験で得られた代謝物を代謝パスウェイ・データベースの情報を用いてつなぐことで予測できる。このパスウェイ・ネットワークを対象として精度の良い代謝シミュレーションが可能となれば、実験では測定できない現

象、あるいは再現が難しい条件下での現象を、実験を行わず予測することが可能となるだけでなく、実験で測定できなかった化合物の挙動についても予測することが可能となる。

これまで、既知の代謝パスウェイの情報をもとに、代謝の過程を計算機内で再現する代謝シミュレーションが数多く行われてきた。たとえば、ペントースリン酸経路と解糖系の反応を用いて、大腸菌のホストトランスフェラーゼ系のモデルを構築し、実験による裏づけを行った研究 (Chassagnole et al. 2002)、ブドウ糖上の好気条件下での大腸菌の中央代謝系について、転写と翻訳を含んだモデルを構築してピルビン酸ノックアウト株の計算を行い、欠損の結果起こった代謝のリプログラミングを再現できることを示した研究 (Varner 2000) など。このほかにも、出芽酵母の解糖系のモデル (Hynne et al. 2001) など、主にバクテリアに関する試みが行われている。また、全細胞シミュレーションを目的とした E-Cell (Tomita et al. 1999) や、COPASI (Hoops et al. 2006) などのモデルも開発されている。いずれのモデルを用いる場合でも、計算に必要な全てのパラメータを実験的に得ることは難しいことから、遺伝的アルゴリズムや、代謝流束解析などの補助的な技術がパラメータ推定に応用されている。本研究においては、シミュレーションに用いるパラメータについて、第2章では、ユーザの設定したパラメータを用いることとし、第3章では、遺伝的アルゴリズムと分散遺伝的アルゴリズムによって、実験で観測された代謝物の挙動を再現するために必要なパラメータを推定することとする。

2.1.4 代謝パスウェイ・データベース

第2.1.1節で述べた代謝反応の連鎖を代謝パスウェイという。これらの代謝物を頂点、代謝反応を代謝物同士をつなげる枝と見立てると、これらが大規模なネットワークを構成していると考えられる。これを代謝パスウェイ・ネットワークという。これまで、複数のモデル生物が選ばれて、重点的に研究が行われてきた。生体内の代謝物の数は生物種によって大きく異なり、大腸菌では、約 1000 種 (Orth et al. 2011)、植物では、モデル植物の一つである *Arabidopsis thaliana* (シロイヌナズナ) の場合に、数千種に上る (Saito & Matsuda 2010)。これまでの研究により、代謝パスウェイに関連する代謝物、酵素などの膨大な情報が得られている。代謝反応は、生物間で似通っている部分が多いが (Pace 2001)、全く同じという訳ではない。これらの生化学的代謝反応パスウェイの情報を収集、共有するために、あるいは、生物間の比較を行うために、いくつかの代謝パスウェイ・データベースが開発されてきた。現在よく使われているものとして、京都大学化学研究所の金久研究室により構築された代謝パスウェイのデータベース KEGG (Kyoto Encyclopedia of Genes and Genomes) (Kanehisa & Goto 2000; Ogata et al. 1999)、MetaCyc (<http://metacyc.org/>) (Caspi et al. 2012)、ヒトに関連した代謝反応のデータベース Reactome (Croft et al. 2011) などがある。植物については、モデル植物のひとつであるシロイヌナズナの代謝データベースとして、AraCyc (Mueller et al. 2003) が The Arabidopsis Information Resource (<http://www.arabidopsis.org>) により開発、公開されている。AraCyc には、予測された代謝パスウェイと、人手により集められた代謝パスウェイが収録されている。最新の AraCyc version 11.5 (2013 年 6 月公開) には、合計 597 個のパスウェイ、3490 個の反応と 2613 個の化合物が登録されている。本研究では、AraCyc データベースの代謝反応リストを用いる。

2.2 解析対象に用いた代謝物の情報

本研究では、有機硫黄代謝を対象とする。アミノ酸のひとつであるメチオニンは、有機硫黄化合物で、生物の成長に必須のものであり、メチオニン回路は、エチレンや高分子アミンのような生物活性物質を S-アデノシルメチオニンを通して生成するのに重要である。本研究では、メチオニンとその他のアスパラギン酸ファミリーのアミノ酸、特にアスパラギン酸、リジン、トレオニンの代謝に焦点を当てる。これらのアミノ酸は、その合成過程で共通の前駆体であるアスパラギン酸 4-セミアルデヒドを共有しており、生体にリジンとトレオニンを添加した場合、フィードバック阻害によって生体内のメチオニンの濃度が低下することが知られている (Buchanan et al. 2007)。これらのフィードバックと制御を評価するため、本研究の共同研究先である北大および理研において、リジンとトレオニンをシロイヌナズナのカルスに外部添加し、UPLC-タンデム四重極型質量分析装置によるワイドターゲットメタボロミクス解析が行われた (実験条件の詳細は付録参照)。ここでは、268 個の代謝物質 (21 個のアミノ酸を含む。付表 1 参照) の濃度の時間変化 (Sawada et al. 2009) が測定されている。これらの代謝物が含まれる経路としては、アミノ酸合成経路だけでなく、解糖系やクエン酸回路、ペントースリン酸経路、グルコシノレート生成経路、プリンおよびピリミジン代謝経路が挙げられる。本研究では、このメタボロミクス解析の結果を用いる。

2.3 方法

質量分析装置によって観測された代謝物は、その生体内で反応が起きている代謝パスウェイ・ネットワーク上に存在する代謝物の一部のみであり、全ての代謝物を含んでいるわけではない。また、代謝物同士をつなぐ代謝パスウェイは自明ではない。このため、実験データから実験内で発現した代謝パスウェイを推測する必要がある。一方、代謝パスウェイ・ネットワーク全体に関わる代謝物数、反応数は膨大であるが、そのうちの顕著に変化している部分のみからなる代謝ネットワークを抽出し、その動的挙動を明らかにすることは十分に有用である。そこで、観測された代謝の動的振る舞いから統計的に有意に変化した代謝物を選び、それらの構成要素に関連する代謝ネットワークを対象としてモデル化する。代謝ネットワークの抽出にあたり、興味を持っているパスウェイに始端と終端の化合物があると仮定する。この始端と終端の化合物を用いて、ゲノム全体の反応リスト（例えば代謝パスウェイ・ネットワークなどから得られた反応リスト）から、始端から終端へとつなぐ最短経路を選び出し、代謝ネットワークを構築する。最短経路の探索には、Dijkstraの方法 (Dijkstra 1959) を用いる。これらの最短経路からなる代謝パスウェイ・サブネットワークを、評価するための「標的」代謝ネットワークとし、代謝予測のシミュレーションを行う。

本研究では、シミュレーションの方法として、確率過程的なシステムの変動を捉えるために、Gillespie (Gillespie 1976; Gillespie 1977) による確率的シミュレーションのアルゴリズムを用いる。これらの手法を実装したシミュレータ SS-mPMG を開発する。

2.3.1 Dijkstra アルゴリズム

グラフ理論の分野に、最短経路を求めるアルゴリズムとして Dijkstra のアルゴリズムがある (Dijkstra 1959)。この方法は、最短経路の探索中に、一度に複数点に対して距離を求めることができるため、ある一点から複数の点に対して最短経路が必要な場合に有効である。(ただし、全てのノード間距離が非負であるという条件を満たす必要がある。) 以下にゴール地点を G とし、そこへの最短経路を探索するという問題を解く場合のアルゴリズムの説明を記す。

ステップ 1 各頂点に対し、ゴール地点からの最短距離を入れる「コスト」を設定する。初期値として、 G に 0、その他には ∞ を入れる。また、各頂点に対し、見つかったかどうかを入れる「済」フラグを設定し、全て「未」に設定する。

ステップ 2 「未」の頂点の中で、最もコストが最小の点 D を見つける。次に点 D につながる全ての点 S について、「 S から D への距離」+「 D のコスト」<「今の S のコスト」を満たす場合、 S のコストを「 S から D への距離」+「 D のコスト」で更新し、 D のフラグを「済」にする。

ステップ 3 ステップ 2 を全てのフラグが「済」になるまで繰り返す。

以上により、最短経路を探索する。本研究では、この Dijkstra のアルゴリズムを用いて最短経路を探索する。

2.3.2 最短経路探索によるサブネットワークの選択方法

ここでは、反応リストからサブネットワークを抽出する方法を述べる。まず、対象とする生物の代謝パスウェイ上の全ての反応リストを代謝パスウェイ・データベースなどから得る。この反応リストを元に、含まれる化合物の全ての組み合わせについて、互いを最短で結ぶパスウェイを Dijkstra のアルゴリズム (Dijkstra 1959) (第 2.3.1 節) を用いて求め、データベースを作成する。次に、興味のある代謝パスウェイの始端化合物と終端化合物を決める。その後、作成された最短パスウェイ・データベースから、始端化合物から終端化合物までの最短経路を抽出してつなぎ、代謝ネットワークを生成する。

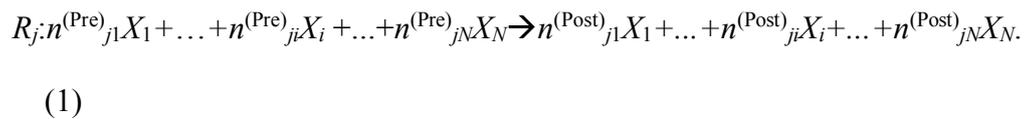
これらの機能を SS-mPMG に実装する。本研究で対象とする生物はシロイヌナズナなので、シロイヌナズナに特化した AraCyc から代謝パスウェイ上の全ての反応リストを得る。代謝物は複数の名前で知られているため、実験で得られた代謝物が必ずしも同じ名前で代謝パスウェイ・データベースに登録されているとは限らない。これを解決するために、データベースを用いた名寄せを行う。これには、当研究室で運用している代謝データベース、KNApSAcK Core データベース (http://kanaya.naist.jp/knapsack_jsp/top.html) (Afendi et al. 2012) 内の識別番号 C_ID、アメリカ化学会 (the American Chemical Society) の一部会である CAS (Chemical Abstracts Service) による CAS 登録番号 (the CAS number)、および MetaCyc データベースの同義語エントリーのデータを用いる。

2.3.3 Gillespie アルゴリズムに基づいた確率的シミュレーション

ヨン

生体内の不連続かつ確率的な振る舞いが、生化学システム全体に影響を及ぼすような、システムの確率過程的な振る舞いは、確率的シミュレーションにより記述可能である。対象とする代謝ネットワークについて、代謝物の初期濃度と反応速度係数を元に、確率的シミュレーション手法の一つである、Gillespie アルゴリズム (Gillespie 1976; Gillespie 1977) を用いたシミュレーションを行う。Gillespie アルゴリズムにおいて、代謝物の濃度は離散的な集合として扱われ、個々の化学反応は物質の濃度と反応係数に比例した確率でランダムに選ばれる事象として扱われる。SS-mPMG に実装された Gillespie アルゴリズムは、以下の5つのステップから成り立つ。図 3A にフローチャートを示す。

まずステップ 1 について記述する。 M 個の化学反応と N 個の代謝物からなる代謝ネットワークを考える。このとき、 j 番目の反応式は式 (1) のようにあらわされる。



ここで、 X_i は、 i 番目の代謝物、 $n^{(\text{Pre})_{ji}}$ と $n^{(\text{Post})_{ji}}$ は、それぞれ基質と生成物における化学量論的反応係数である。 j 番目の化学反応が起こると、 i 番目の代謝物が $n^{(\text{Post})_{ji}} - n^{(\text{Pre})_{ji}}$ 変化する。以下の値、計算の最大時間 T 、 i 番目の代謝物 ($i = 1, 2, \dots, N$) について、代謝物の初期濃度 $X_i^{(\text{init})}$ 、代謝物量の境界を $X_i \in [X_i^{\min}, X_i^{\max}]$ のように決める境界値 X_i^{\min}, X_i^{\max} 、反応 R_j に対する反応速度係数 c_j は、ユーザによって決められる。

ステップ 2 では、全ての代謝物濃度が $[X_i]$ に代入される。ステップ 3 では、

j 番目の反応における hazard 関数を式 (2) のように計算する。

$$h_j = c_j [X_i] \quad \text{for } R_j: X_i \rightarrow X_i' \quad (2a)$$

$$h_j = c_j [X_i]([X_i]-1)/2 \quad \text{for } R_j: 2X_i \rightarrow X_i' \quad (2b)$$

$$h_j = c_j [X_i][X_i] \quad \text{for } R_j: X_i + X_i \rightarrow X_i'' \quad (2c)$$

この hazard 関数の合計 ($H = \sum_{j=1}^M h_j$) を用いて、式 (3) のように時間ステップ Δt を計算し、 j 番目の反応の起こる確率は、式 (4) のようにあらわされる。

$$\Delta t = -(1/H) \log(\tau) \quad (3)$$

$$p_j = h_j/H \quad (4)$$

ここで、 τ は、0 から 1 までの一様乱数から得られるランダムな値である。

ステップ 3 では、 j 番目の反応は確率分布 $P(X=j) = p_j$ となる j からランダムに選ばれる。これにより、次に起こる反応をランダムに選ぶことができる。

ステップ 4 では、分子の濃度が選ばれた反応式に従って式 (5) のように更新される。

$$[X_i]' \leftarrow [X_i] - n^{(\text{Pre})}_{ji} + n^{(\text{Post})}_{ji}. \quad (5)$$

ここで、 $[X_i]'$ が全ての i について $X_i^{\min} \leq [X_i] \leq X_i^{\max}$ を満たす場合には、代謝物の濃度が更新され、時間 t が $t+\Delta t$ へと進み、アルゴリズムはステップ 2 に戻る。そうではない場合には反応は否定され、もう一度ステップ 3 によって別の反応が選ばれる。これらのステップが、時間が最大時間 T を超えるまで繰り返される。

Gillespie アルゴリズムの利点は、タンパク質と遺伝子の発現抑制による代謝調節を反応式のセットで表すことによってシミュレートできる点である。たとえば、自己調節 (図 3B) やフィードバック阻害 (図 3C) は反応式のセットで表現でき、これらは本ソフトウェアで動的にシミュレートできる。

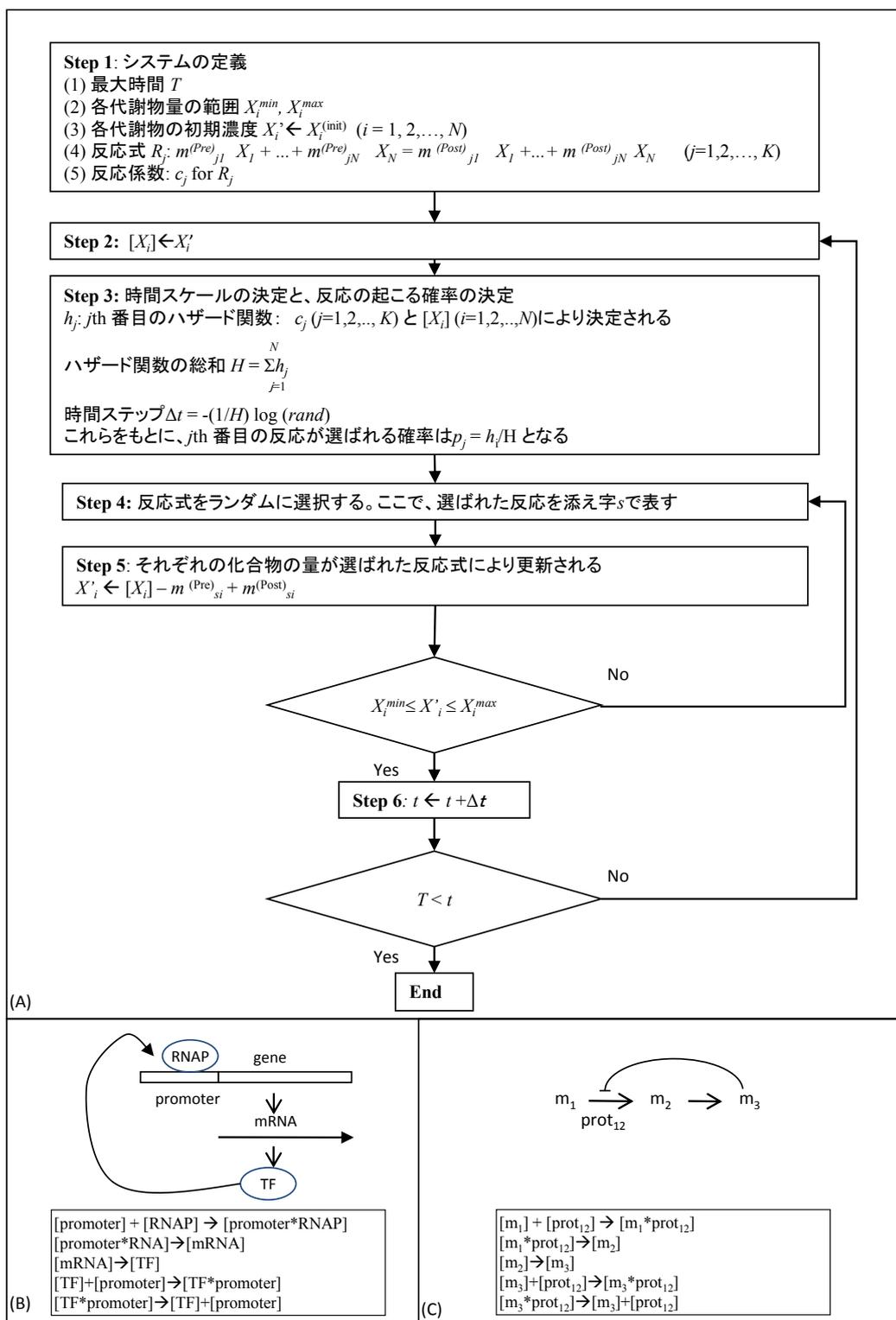


図 3 SS-mPMG および SS-GA で用いられる確率的シミュレーションの概要

Gillespie アルゴリズムのフローチャート、(B) 遺伝子の自己制御機構を記述した反応の例、(C) フィードバック阻害を記述した例。

2.4 結果と考察

2.4.1 大規模代謝ネットワークシミュレータプログラム SS-mPMG

対象とする代謝物に関連する代謝パスウェイを探索し、それらの代謝ネットワークの挙動のモデルを構築するためのツール SS-mPMG を開発した (Katsuragi et al. 2013; Katsuragi et al. 2012)。このツールは、全代謝物のシミュレーションを行う代わりに、パスウェイ・データベースから代謝ネットワークのサブセットを抽出することを目的に、Dijkstra アルゴリズム (Dijkstra 1959) を用いて、始端化合物から終端化合物までの最短経路を抽出し、関連する代謝ネットワークを生成する。さらに、SS-mPMG は、選ばれた代謝ネットワークを対象として、ユーザが設定したパラメータを用いてその挙動をシミュレートすることができる。このソフトウェアパッケージはインターネット上にダウンロード可能な状態で公開している (<http://kanaya.naist.jp/mPMG>)。図 4 にソフトウェアの概要を示す。SS-mPMG は、2つの部分からなる。すなわち、(1) 最短経路による反応の選択、(2) 確率的シミュレーションである。なお、SS-mPMG の詳細については、付録に記載した。

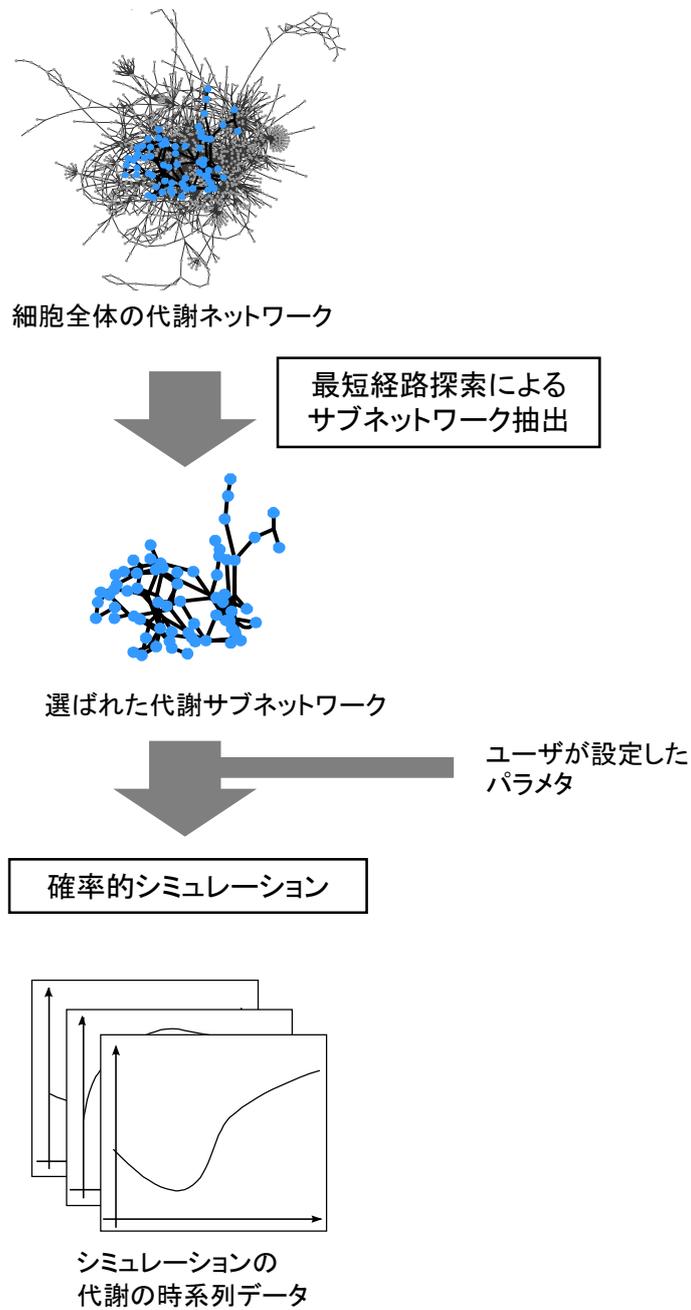


図 4 SS-mPMG による代謝ネットワークのモデル化の概要

システムの複雑さを軽減するため、SS-mPMG は、与えられた始端と終端に基づいて全代謝ネットワークからその一部を抽出し、シミュレーションの対象とする

2.4.2 実験で得られた代謝の挙動

本研究では、代謝反応のダイナミクスがアミノ酸の添加によって実験内でどのように変化するかを調べる。ここでは、第2.2節の実験で得られた268個の代謝物質の実験開始時からの2, 6, 12, 24, 36, 48, 60, 72, 84 h後の濃度変化から、代謝反応の全体的挙動の特徴を調べ、その結果をもとに、代謝ネットワークを抽出するための始端化合物と終端化合物を設定する。

実験で得られた代謝反応について、なるべく少ない情報損失で縮約して全体の挙動をみるために、実験で得られた全ての代謝物の濃度の相対的な時系列変化について、主成分分析を行った。得られた第一主成分、第二主成分の寄与率はそれぞれ85.88%、寄与率は7.41%であり、第二主成分までの累積寄与率は93.29%であったため、第二主成分までで十分に累積寄与率が高いと判断し、第二主成分までを解析することとした。図5は、第一主成分と第二主成分の散布図を示す。図からは、実験群について0-24 hの間に減少していた第二主成分が、24-84 hの間で増加している傾向が見られた。この結果から、代謝産物の濃度は、実験の前半期間(0-24 h)と後半期間(24-84 h)で大きく変動しており、背後にある代謝システムに何らかの変化があったことが示唆された。そこで、本研究では、0-24 hを初期相、24-84 hを後期相と定義し、初期相、後期相それぞれの代謝のダイナミクスを再現するモデルを構築し、それぞれのモデル内における代謝反応の反応係数を比較することによって、代謝システムにどのような変化が生じたのかを調べる。

モデルの対象となる代謝ネットワークは以下のように構築する。代謝ネットワーク上の反応を通じて、濃度が顕著に減少した代謝物から顕著に増加した代謝物が生成されたと考え、前者を始端化合物、後者を終端化合物として選び、その間を最短経路探索でつなぐことによって代謝ネットワークを構築する。具体

的な始端化合物、終端化合物の選び方は以下のとおりである。初期相と後期相の2つの期間 (0-24 h, 24-84 h) と、全体の期間 (0-84 h) の3つの期間において、顕著に増えた、あるいは減った代謝物を特定した。すなわち、これらの3つの期間において、全ての代謝物の始めの時刻から終わりの時刻への相対的な変化量を調べ、5%の有意差で変化が認められた代謝物を選んだ。これらの顕著に減少した代謝物を始端化合物、増加した代謝物を終端化合物とした (表 1)。始端化合物として選ばれたのは、マニトール、ラフィノースであり、終端化合物として選ばれたのは、プロリン、サッカロピン、ホスホリルコリン、シチジン二リン酸コリン、トレオニン、シキミ酸、5'-メチルチオアデノシン、ホモシステイン、S-アデノシル-L-メチオニンである。これらの化合物を次節の代謝パスウェイ・サブネットワークの抽出に用いる始端化合物と終端化合物とした。

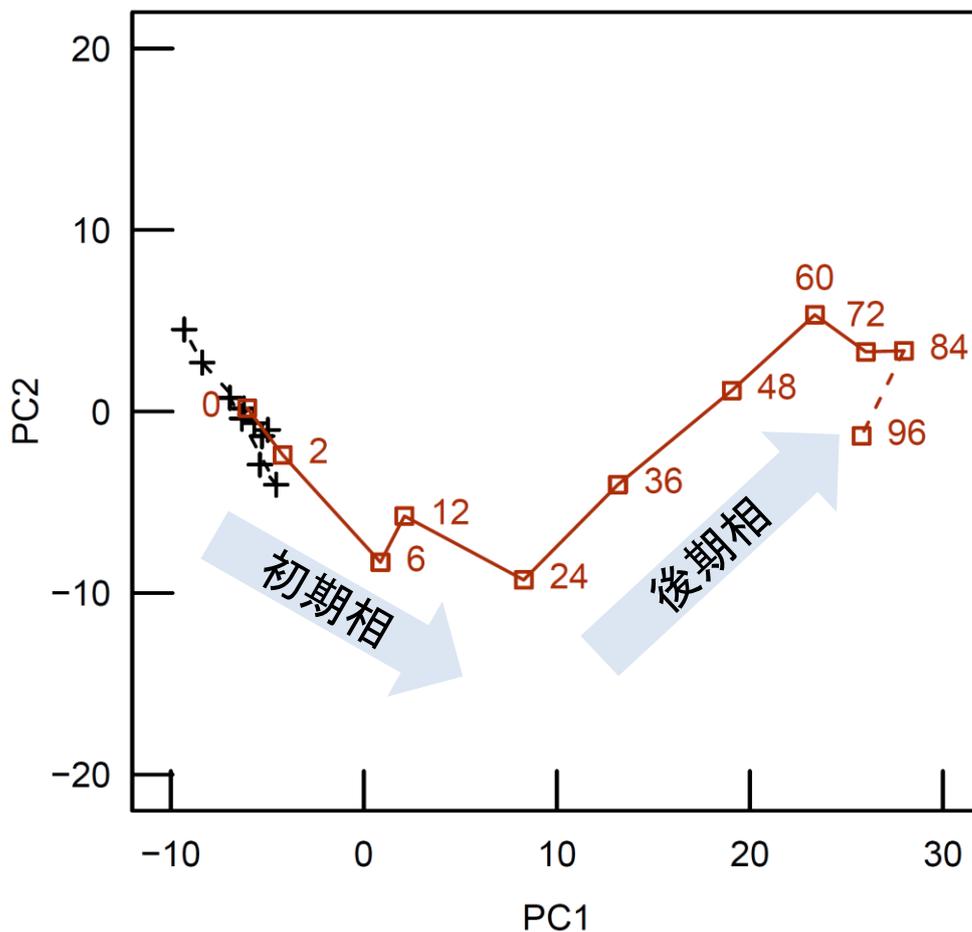


図 5 PCA 解析による代謝の全体的挙動

x 軸と y 軸はそれぞれ、PCA 解析の第一主成分、第二主成分をあらわす。黒の点は統制群、赤の点は 10mM のリジンと 1mM のトレオニンを追加した実験群の結果を表す。図中の数字は、サンプリング時刻を追加実験開始からの時間経過として時間で表している。実験群において、96 h の値を外れ値とみなすと、0-24 h において減少していた第二主成分が、24-84 h で増加するという変化が見られる。そこで、0-24 h を初期相、24-84 h を後期相と定義した。

表 1 実験結果に基づいて選ばれた始端化合物と終端化合物

| Period | Metabolites | Type of Compound | p-value |
|-------------|-------------------------|------------------|----------|
| 0-84 hours | raffinose | Start | 7.00E-03 |
| | L-proline | End | 2.52E-03 |
| | saccharopine | End | 2.27E-02 |
| | phosphoryl-choline | End | 1.29E-04 |
| | CDP-choline | End | 3.29E-03 |
| 0-24 hours | mannitol | Start | 7.03E-03 |
| | saccharopine | End | 1.17E-04 |
| | CDP-choline | End | 7.84E-06 |
| | L-threonine | End | 9.80E-03 |
| 24-84 hours | raffinose | Start | 1.65E-02 |
| | shikimate | End | 3.79E-04 |
| | 5'-methylthioadenosine | End | 2.85E-03 |
| | L-homocysteine | End | 1.56E-03 |
| | S-adenosyl-L-methionine | End | 8.15E-03 |

2.4.3 代謝パスウェイ・サブネットワークの抽出

代謝パスウェイ・データベース AraCyc より得た全反応リストから、本研究の対象外となる H₂O やその他の物質を除外し、これを初期代謝ネットワークとした。この初期代謝ネットワークから、SS-mPMG を用いて興味の対象となるべき部分ネットワークを抽出した。このとき、第 2.4.1 節で得られた始端化合物から、いくつかの代謝経路を通して終端化合物が生成されると考え、計算対象とする代謝ネットワークを、以下の 3 つの代謝ネットワークの結合とした。すなわち、0-84 h, 0-24 h, 24-84 h の 3 つのそれぞれの期間において、始端化合物から終端化合物への最短経路上にある反応と化合物により構成される代謝ネットワークを SS-mPMG を用いて抽出し、それらの結合とした。このとき選ばれた反応群は、クエン酸回路の一部を含んでいた。そこで、クエン酸回路を完全に包含させるために、いくつかの自動では選ばれなかった反応を手動で加えた。

以上の結果として、1250 個の化合物と 1440 個の代謝反応からなる初期代謝ネットワークから、59 個の代謝物を含む 56 個の反応が選ばれた。(付表 2, 付表 3)。この中には、実験でデータが得られている 268 個の代謝物のうち、26 個の代謝物が含まれていた。選ばれた 59 個の反応と 56 個の代謝物からなる代謝パスウェイ・ネットワークを図 6 に示す。図に示すように、選ばれた代謝パスウェイ・ネットワークは糖類ラフィノースまたはマニトールから始まる。その後、解糖系とクエン酸回路を通り、アミノ酸トレオニン、プロリン、セリン、またはメチオニンで終わる。

次に、対象とする代謝ネットワークを抽出したことによって、元の初期代謝ネットワークを対象とした場合と比べて計算時間がどの程度減るかを調べた。Intel(R) Xeon(R) CPU E5-2690 v2 (3.00GHz) の CPU を搭載したクラスタ計算機上で SS-mPMG による初期代謝ネットワークと、サブネットワーク抽出により得

られた代謝ネットワークとのそれぞれを対象とした 50 回の代謝シミュレーションを試行し、それぞれの試行が終了までにかかった時間を計測した。このとき、代謝物の初期濃度、または反応の反応係数の k 番目の値が以下を満たすように生成した：

$$v_k = \begin{cases} v_{0k}r & (\text{if } p < 0.7) \\ v_{0k} & (\text{if } p \geq 0.7) \end{cases}, \quad (6)$$

ここで、 v_{0k} は、初期濃度と反応係数として、それぞれ仮に 1000 と 0.001 とした基本値、 p は 0 から 1 までの一様乱数から得られるランダムな値、 r は対数正規分布から得られるランダムな値である。

抽出元である初期代謝ネットワークを対象とした 50 回の試行にかかった 1 回あたりの平均実行時間は 491 ± 40.1 秒であった。一方、サブネットワークの抽出により得られた代謝ネットワークを対象とした場合は 3.5 ± 0.8 秒であった。この結果から、サブネットワークの抽出によって、平均実行時間で見ると約 140 倍の高速化ができた。以上より、SS-mPMG により、代謝パスウェイ・データベースの反応リストと代謝実験の結果からサブネットワークを抽出することができた。

第3章 代謝シミュレーションのためのパラメータ推定手法の開発

3.1 序論

本章では、実験で観察された代謝の挙動をシミュレーションによって再現する際に、設定が必要となるパラメータのセット（それぞれの代謝反応の反応係数とそれぞれの代謝物の初期濃度）について、実験データをよく再現できるように、進化的アルゴリズムを用いて推定できるツールを開発する。進化的アルゴリズムは幅広い問題に適用できるが、その一つである遺伝的アルゴリズムと、その応用である分散遺伝的アルゴリズムを用いてパラメータを推定する。

3.1.1 遺伝的アルゴリズム

遺伝的アルゴリズムは、組み合わせ最適化問題に使われる、自然界の進化を模倣した手法である (Holland 1975)。遺伝的アルゴリズムでは、対象問題の探索空間中の一点を一つの個体として表し、解は染色体と呼ばれる文字列に対応する。一つの個体は、染色体ならびに評価関数により与えられる適応度を一つずつ持っている。まず、遺伝的アルゴリズムでは、個体群と呼ぶ個体のセットをランダムに生成する。個体群は、評価、選択、交叉、突然変異を繰り返すことによって進化し、その結果、準最適な解が得られる。一般的な遺伝的アルゴリズムの特徴としては、(1) 最適化するパラメータセットをコード化して扱うこと、(2) 一点ではなく多点で探索すること、(3) 導関数やその他の情報を用いず、目的関数値のみを用いること、および、(4) 確率的探索法であることが特徴である (Goldberg 1989)。遺伝的アルゴリズムの利点は、解が広域に広がっている問題に対して、実行時間に比して良質な解を得られることと、定式化ができない、あるいは性質がわかっていない問題についても探索可能なことが挙げられる。遺伝的アルゴリズムは、多分野にわたって応用されており (Goldberg 1994)、化学や生物学の分野でも用いられてきた (Yen et al. 1995; Morbiducci et al. 2005)。遺伝的アルゴリズムによって、トリペプチド・ライブラリを構築するための準最適なアミンを瞬時に選べることを示した研究 (Sheridan & Kearsley 1995) や、タンパク質の部分的な柔軟性によって、リガンドの立体配座の全柔軟性を探索した研究 (Jones et al. 1995) などがある。

古くからある遺伝的アルゴリズムでは、上で述べた特徴 (1) のように、最適化されるべき実数値のベクトルをビット列にコード化して用いるのが一般的である。さらに、実数値ベクトルのまま扱う方法もある。このタイプの遺伝的アルゴリズムは、実数値遺伝的アルゴリズムと呼ばれ、様々な最適化問題に用いられ

てきた (Corcoran & Sen 1994; Huang & Huang 1997; Wu et al. 2007)。この場合、遺伝子型は実数値の配列であり、突然変異はひとつあるいは数個の値がランダムに連続値として変化するように実装される。本研究では、最適化するパラメータセットは代謝物の初期濃度と反応係数であり、計測されたデータを精度良く再現するために、実数値遺伝的アルゴリズムを採用する。

3.1.2 分散遺伝的アルゴリズム

遺伝的アルゴリズムによる探索では、解がしばしば大域的最適解ではなく、局所的最適解に集まる傾向がみられる。これは、ある世代に局所的最適解が見つかった場合、その解が次の世代に優先的に多く選ばれ、かつ交叉によって、母集団全体に似たような解が広がってしまうことによる。全体に局所最適解が広まってしまうと交叉や突然変異が起きたとしても、簡単にはその局所解を越えることができず、大域的最適解にたどり着くことが難しくなる。この問題を解決するためには、解の多様性を維持する仕組みが必要である。

ひとつの解決策として、分散遺伝的アルゴリズム (Distributed Genetic Algorithm) (Tanese 1989; Belding 1995) を用いる方法がある。分散遺伝的アルゴリズムは、島モデル (Island Model) と呼ばれる最適化手法である。遺伝的アルゴリズムでは解探索に用いる母集団はひとつであるが、分散遺伝的アルゴリズムは母集団を複数のサブ母集団に分けて、それぞれで独立に進化操作を行い、その過程で、ある確率でサブ母集団内の個体を別のサブ母集団内に「移住」させるというモデルである。それぞれのサブ母集団に、異なった最適値が保持されていることが期待でき、適度な移住操作を行うことによって、他の最適値を導入することができるため、より良い個体が生成されると期待される (Whitley et al. 1998)。実際、このモデルは解の多様性を保つのに有効であるといわれており (Alba & Troya 1999)、局所解を避けるための有効な手段と期待できる。

この手法では、それぞれのサブ母集団において、評価関数を変える、交叉率、突然変異率等の進化パラメータを変える、あるいは、交叉方法などの進化方法を変えるとといったことも可能である。遺伝的アルゴリズムが自然界の選択、淘汰、突然変異の仕組みを模倣したと考えると、分散遺伝的アルゴリズムはさらにそこに地域性あるいは環境多様性といった概念を導入したものとも考えられる。

分散遺伝的アルゴリズムには、複数の母集団と、その間を個体に移住するという操作が必要なため、遺伝的アルゴリズムの場合と比べて、母集団数、移住操作回数、移住率、母集団のトポロジといったパラメータが加わる。Cantu-Paz は、これらのパラメータの探索への影響について調べている (Cantu-Paz 2000)。本研究では、推定すべきパラメータが 171 個の実数であり、その解空間が非常に広大である。そのため、分散遺伝的アルゴリズムによる解の多様性の維持が有効であると期待できる。そこで、遺伝的アルゴリズムに加えて、分散遺伝的アルゴリズムをプログラム SS-dGA に実装した。

3.2 方法

実験を再現する代謝のダイナミック・シミュレーションを行うためには、それぞれの代謝物の初期濃度と、それぞれの代謝反応の反応係数があらかじめ設定されている必要がある。しかしながら、一般的に、質量分析装置を用いた実験の場合、抽出とイオン化の効率が各代謝物の特性に依存するために代謝物スケールの計測しかできず、絶対量を得ることができない。このため、モデルに用いる代謝物の初期濃度を実験から得ることは難しい。同じ理由により、たとえモデル上で代謝物の濃度が明確に計算できたとしても、それを直接実験結果と比較することができない。

前者の問題に対しては、実験で得られた代謝濃度の振る舞いを再現する最も良い値のセットを探す数値的最適化を用いることで対処する。進化的アルゴリズムは幅広い問題に適用できることから、本研究では、例として、遺伝的アルゴリズムと分散遺伝的アルゴリズムを、プログラム **SS-GA** と **SS-dGA** にそれぞれ実装する。後者の問題に対しては、実験結果と計算結果を比較する際に、それぞれの代謝反応の絶対量の変化ではなく、相対変化率に注目することで対処する。代謝物の濃度のスケールに関しては任意性が存在するものの、生成物量の基質量に対する比のような相対的な値は、動的モデル内で適切に再現される。第2.2節で得られた質量分析装置によって観測された代謝濃度の振る舞いを計算機上で再現するために、本研究で開発したソフトウェア (**SS-GA**, **SS-dGA**) を用いて、それぞれの代謝物の初期濃度とそれぞれの反応の反応係数のセットを最適化する。

3.2.1 遺伝的アルゴリズムを用いた推定方法

本研究ではパラメータの推定に実数値遺伝的アルゴリズムを用いるので、図 7 に示されるように、一つの個体の染色体は一つの実数値の配列として表現される。これらは代謝物の濃度と反応係数を一つの配列としてつなげたものである。本研究では、59 個の代謝物と 56 個の反応を用いる。初期相と後期相の反応係数を独立して計算するため、反応 1 つあたりに 2 つの推定すべき反応係数がある。よって、推定すべきパラメータの数は合計 171 (=59+56*2) 個であり、染色体は 171 個の実数値をもつ配列となる。

以下に、遺伝的アルゴリズムを用いたパラメータの推定方法を述べる (図 7)。

まず、初期値の生成を行う。最初の世代の母集団として、 K 個のランダムに生成された個体群を、遺伝子上の h 番目の値が以下を満たすように生成する：

$$v_h = \begin{cases} v_{0h}r & (\text{if } p < 0.7) \\ v_{0h} & (\text{if } p \geq 0.7) \end{cases} \quad (7)$$

ここで、 v_{0h} は、初期濃度と反応係数として、それぞれ仮に 1000 と 0.001 とした基本値、 p は 0 から 1 までの一様乱数から得られるランダムな値、 r は対数正規分布から得られるランダムな値である。

次に、母集団から個体が 2 つ選択される。このとき、個体はルーレット選択によって適応度 (後述) の値に基づいて個体を選ばれる。すなわち、 k 番目の個体が、その適応度 F_k に比例する確率で選ばれる。次に、選ばれた 2 つの個体間で、 $p_x=0.8$ の割合で、交叉が起きる。すなわち、正規分布からランダムに選ばれた交叉点を境目に、それぞれの染色体が交換される。選択と交叉の後、染色体の保持するそれぞれの値が、確率 $p_{mu}=0.005$ で突然変異する。このとき、新しい値 (v'_h) を式 (8) のように定義する。

$$v'_h = \begin{cases} v_h r & (\text{if } p < 0.005) \\ v_h & (\text{if } p \geq 0.005) \end{cases} \quad (8)$$

ここで、 v_h は h 番目の個体の変異前の値であり、 p は区間 $[0, 1]$ の一様分布からランダムに得られた値、 r は対数正規分布からランダムに選ばれた値である。このように選択、交叉、突然変異を経て生成された新しい個体が、次の世代の個体の一つとなる。

次に、適応度の計算方法を示す (図 8)。適応度は、計算で得られた代謝の挙動と実験で観測された代謝の挙動を比較した評価値であり、以下のように計算する。それぞれの世代において、ひとつの染色体から与えられるパラメータを元にシミュレーションを行い、代謝の動的変化を求め、それぞれの代謝物の相対変化を、実験から観測された値と比較する。ある個体 k の適応度 F_k は、下式のように、観測された代謝物の変化とシミュレートされた代謝物の変化から得られる平均残余残差の逆数とする：

$$q_{ki} = \sqrt{\frac{\sum_s (\log_{10} \frac{X_{kis}/X_{ki0}}{Y_{is}/Y_{i0}})^2}{N_s}} \quad (9)$$

$$R_k = \frac{\sum_i 10^{q_{ki}}}{N} \quad (10)$$

$$F_k = 1/R_k, \quad (11)$$

ここで、 s は、代謝の観測時間点 ($s = 0, 2, 6, 12, 24, 36, 48, 60, 72, 84$ h)、 X_{kis} と Y_{is} は、 i 番目の代謝物の s 時間点での計算と実験の代謝物量をそれぞれ表し、 N_s は、時間点の総数、 N は代謝物数を表す。これにより、適応度は 0 から 1 までの実数値となり、1 が最も高い適応度となる。本研究では、はじめに、事前探索として、個体数 $K=5000$ での世代数 $G=10$ の計算を行い、そこで得られた最も良い個体から、再度個体数 $K=100$ 、世代数 $G=500$ の探索を行うという、2段階の探索を行った。これらの探索について、探索時に確率的選択を行うため、探索結果は一意に定まらないことを考慮し、乱数に与える初期シードを変更した複数の試行を行う。

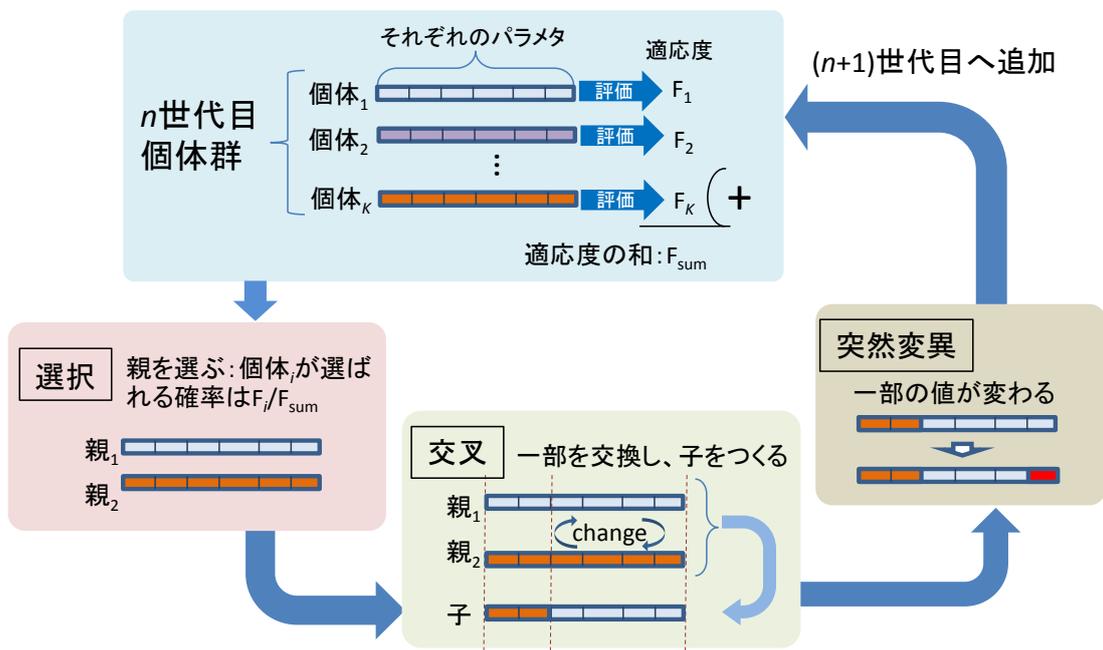


図 7 本研究で用いる遺伝的アルゴリズムの手順を表す概念図

一様なパラメータをもつ染色体を元に、初期の染色体がランダムに生成される。それぞれの世代において、新しい染色体が適応度に比例して選ばれ、交叉または突然変異による値の操作が行われる。 G 世代後の最も適応度の高い値を示した染色体が選ばれる。

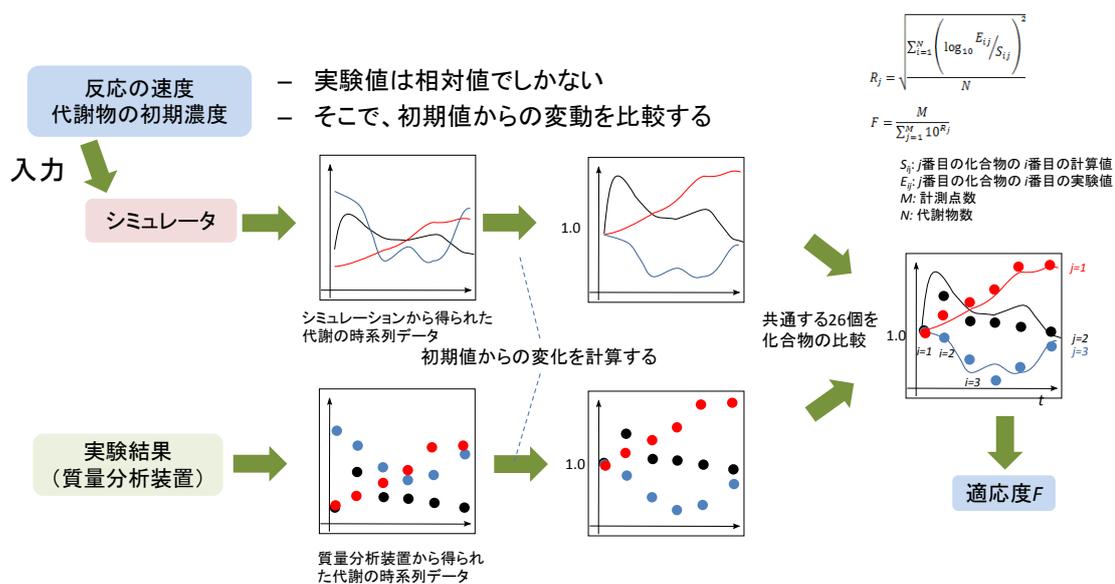


図 8 適応度の計算方法

実験と計算双方のそれぞれの化合物の量の時間変化をもとに初期濃度からの変化量を求め、それらを化合物ごとに比較して適応度を計算する。

3.2.2 分散遺伝的アルゴリズムを用いた推定方法

分散遺伝的アルゴリズムにおいては、母集団が複数の母集団に分けられる。本研究においては、それぞれの母集団内における個体の進化方法（選択、交叉、突然変異）は、第3.2.1節で述べた遺伝的アルゴリズムを用いた方法と同じである。異なるのは、複数ある母集団それぞれにおいて独立に進化したパラメータをもつ個体が、ある一定間隔ごとに移住操作によって他の母集団に移動することである。ここで、移住操作が起きてから、次に移住操作が起きるまでの間を1ステージと数えることとする。すると、通常の遺伝的アルゴリズムに用いるパラメータに加えて、以下のパラメータが増えることになる。すなわち、母集団同士の接続形態（トポロジ）、母集団数、ステージ数、移住率である。

分散遺伝的アルゴリズムの方法について述べる。母集団を複数個用意し、それぞれを遺伝的アルゴリズムの場合と同様に進化させる。一定期間ごとに、移住操作を行う。移住操作は、簡単のため、母集団をランダムに選んで移住させるという方法を用いる。すなわち、全母集団から母集団2つを選ぶ全順列について、区間 $[0, 1]$ の一様分布からランダムに得られた確率 p が移住率 p_{mi} よりも小さいときに移住が起きることとし、選ばれた対の一つ目の母集団で最も適応度の高い個体によって、もう一方の母集団の最も適応度の低い個体を置き換える。

これらの値が実際に代謝シミュレーションのためのパラメータセットの探索にどのような影響を与えるかをみるために、総計算試行回数（適応度を計算する延べ個体数）を同じにした場合の、母集団ごとの個体数、世代数、母集団数、ステージ数、移住率の組み合わせをいくつか試す。分散遺伝的アルゴリズムの探索では解の多様性が保持されていることが期待できるため、遺伝的アルゴリズムを用いた場合とは異なり、1段階のみの探索を行った。

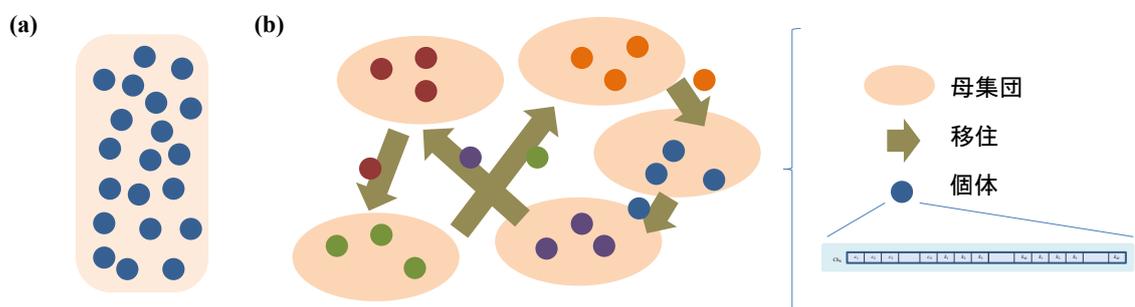


図 9 本研究で用いたアルゴリズムの概念図 (a) 遺伝的アルゴリズムと (b) 分散遺伝的アルゴリズムの違い

通常、遺伝的アルゴリズムは母集団が一つである。分散遺伝的アルゴリズムでは、母集団が複数存在する。それぞれの母集団は独立にパラメータを進化させる。ある一定の確率で移住操作が起こり、ランダムに選ばれた2つの母集団間で染色体の移動が行われる。

3.3 結果と考察

3.3.1 大規模代謝シミュレータ・プログラム SS-GA, SS-dGA

与えられた代謝実験データおよび代謝パスウェイを元に代謝データに合うようなパラメータを推定して確率的シミュレーションを行うツール、SS-GA と SS-dGA を開発した (Katsuragi et al. 2013; Katsuragi et al. 2014)。パラメータ推定には、前者は遺伝的アルゴリズム、後者は分散遺伝的アルゴリズムをそれぞれ用いる。パラメータ推定時、代謝実験データが相対値であることを考慮し、シミュレーションの結果と代謝実験の結果のそれぞれについて、各化合物の初期値からの相対変化量を用いて適応度を求め、パラメータの評価を行う。代謝パスウェイのデータは、SS-mPMG と共通なので、SS-mPMG を用いて作成することができるが、他のオミクスデータから得られたものを用いることもできる。これらのソフトウェアは、SS-mPMG と同様に、インターネット上にダウンロード可能な状態で公開している (<http://kanaya.naist.jp/mPMG>)。図 10 にソフトウェアの概要を示す。SS-GA および SS-dGA は、2つの部分からなる。すなわち、(1) 確率的シミュレーション、(2) 実験値に基づいたパラメータ推定である。なお、SS-GA、SS-dGA の詳細については、付録に記載した。

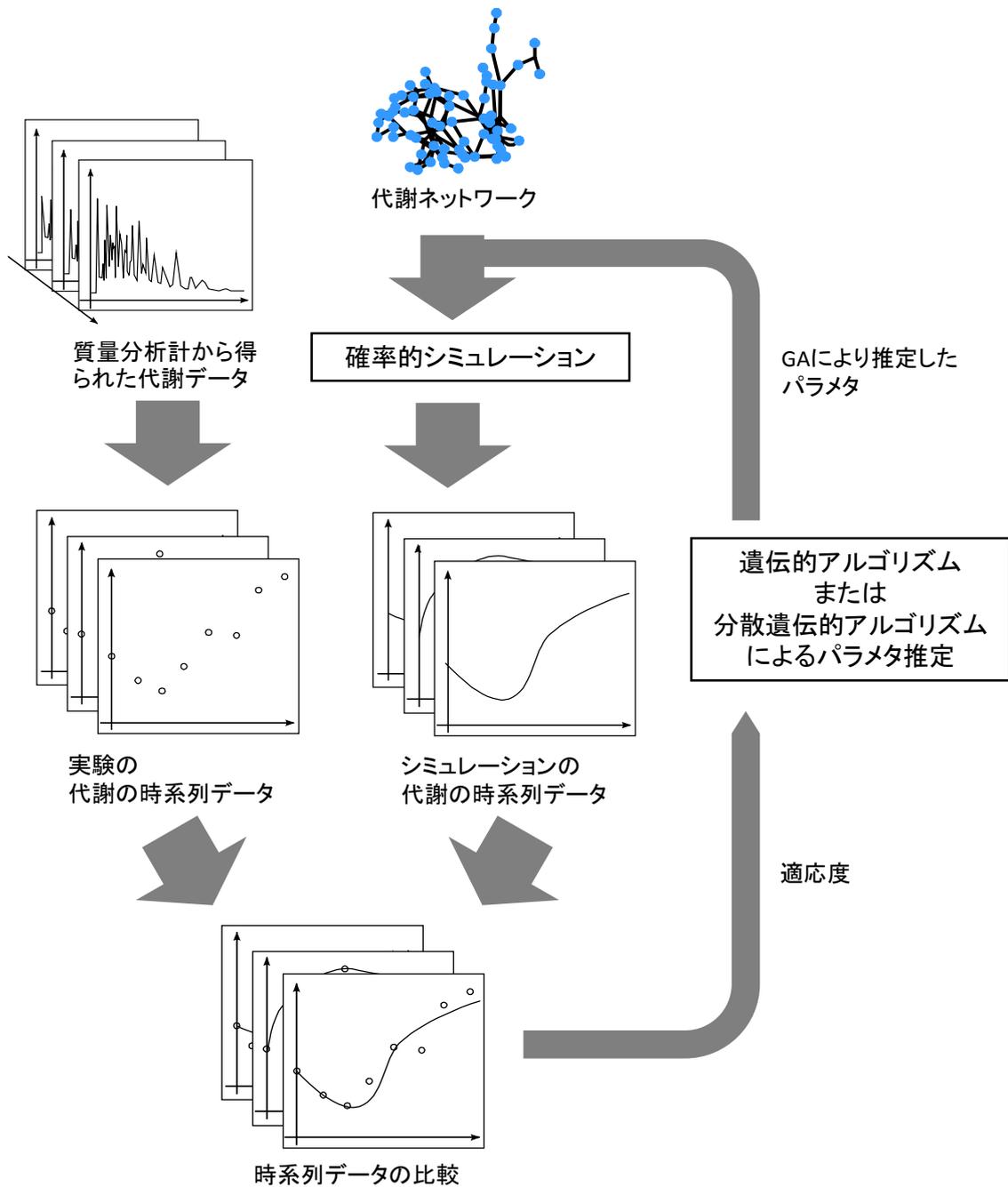


図 10 SS-GA および SS-dGA による代謝ネットワークのモデル化の概要

代謝データと対象の代謝ネットワークから実験を再現するのに必要なパラメータの推定を行って確率的シミュレーションを行うことができる。

3.3.2 シロイヌナズナの代謝データを用いたモデルの評価

171 個のパラメータを推定するため、遺伝的アルゴリズムと分散遺伝的アルゴリズムを用いて、シミュレーションの結果が実験で得られたデータに合うように、パラメータの推定を行った。前述のとおり、実験でデータが得られた 268 個の代謝物のうち、第 2 章で選ばれたサブネットワークには 26 個の代謝物が含まれていた。それぞれの代謝物の量は代謝物スケールの相対値であるため、これらの代謝物について、実験開始後から 0, 2, 6, 12, 24, 48, 60, 72, 84 h 後の 9 個の時刻において、初期濃度 (0 h) からの相対変化を見た。つまり、26 個の代謝物について、実験における 9 個の欠損値を除いた合計 225 個のデータ点を用いて、初期濃度からの相対変化を比較した。

これらの実験で得られた値とシミュレーションで得られた値とを比較するために、1 つのパラメータセットに対して 1 回のシミュレーションを行い、それぞれの代謝物の初期濃度からの変化を計算した。そののち、それぞれの時刻における計算での代謝物量の変化と実験での代謝物量の変化の残差をとり、適応度を計算した (式 9-11)。

SS-GA または SS-dGA によるシミュレーションおよびパラメータ推定は、確率過程に基づいているため、複数試行を行った場合に確率過程的要因によって、結果にばらつきが生じる。そこで、結果のばらつきを評価する。まず、 t 検定を用いて、遺伝的アルゴリズムの結果と、分散遺伝的アルゴリズムのそれぞれの結果との、互いの分布を比較したところ、全ての分散遺伝的アルゴリズムの適応度の分布が遺伝的アルゴリズムの適応度の分布とは有意に異なることがわかった (閾値 0.01 以下)。図 11 は、遺伝的アルゴリズムと分散遺伝的アルゴリズムを用いた場合の計算について、終了時の適応度のヒストグラムを表す。全ての場合で、総計算試行回数 (適応度を計算する延べ個体数) を 1000000 回とした結果で

ある。遺伝的アルゴリズムを用いた場合の代謝物の変化について、実験結果と計算結果の間で得られた相関は $r=0.9657$ であった。また、分散遺伝的アルゴリズムを用いた場合には、最も高い相関は $r=0.9677$ であった。図 11 を見ると、分散遺伝的アルゴリズムを用いた場合、遺伝的アルゴリズムの場合と比べて適応度が高い傾向があり、また、ばらつきが小さくなっている。分散遺伝的アルゴリズムの方が、適応度が高い傾向と、少ないばらつきを示したことから、より良い解を見つけやすい傾向にあることがわかった。

次に、分散遺伝的アルゴリズムの移住率と母集団数の違いが適応度探索にどのような影響を示すかを調べた。図 12 は、図 11 と同条件の場合について、それぞれの条件で最も適応度が高かった試行結果における、母集団ごとの最も高い適応度の履歴の変化を表す。遺伝的アルゴリズムの場合、150 世代あたりから急激に進化が遅くなっているが、分散遺伝的アルゴリズムの場合には、その変化はゆるやかである。移住率が高い場合 (図 12B) には、他のパラメータが同じであるが移住率が低い場合 (図 12C) と比べて、母集団ごとの適応度のばらつきが全体的に小さい傾向にある。この場合、過度に移住が起き、母集団の多様性が失われていると考えられる。一方、移住率が低い場合 (図 12F) には、他のパラメータが同じであるが移住率が低い場合 (図 12E) と比べ、母集団ごとの適応度のばらつきが大きい傾向にある。この場合は、移住の効果が十分に発揮されず、各母集団が独立に進化しており、その結果、最適な解が全体に行き渡っていないことを示している。この結果から、移住率が大きすぎる場合には母集団の多様性が失われ、移住率が小さすぎる場合には移住による最適解の交換が十分になされないという傾向がみられた。

そこで、分散遺伝的アルゴリズムに用いる移住率が最終的な適応度に与える影響を調べる。図 13 は、分散遺伝的アルゴリズムの場合の 50 回の試行につい

て、移住率ごとに最終的な適応度の分布を表す。分散遺伝的アルゴリズムの場合の全ての移住率において、適応度の中央値は遺伝的アルゴリズムの場合よりも大きい。このことは、図 11 から得られた、分散遺伝的アルゴリズムを用いた方が遺伝的アルゴリズムの場合と比べて適応度が高い傾向にあるという結果と一致する。分散遺伝的アルゴリズムの場合の適応度の中央値は、移住率が 10^{-5} の場合に最も小さく、移住率が大きくなるにつれて大きくなるが、移住率が $10^{-1.5}$ の場合に最大となり、それより大きい移住率の場合にはまた小さくなる。この傾向は、移住率が小さすぎても大きすぎても適応度の中央値が小さくなるという図 12 の結果の傾向と同様である。また、移住率が $10^{-0.5}$ の場合には、他の場合に比べて最終的な適応度の分散が大きくなっている。この場合、移住率が大きいため、ある母集団で得られた局所最適解が一気に他の母集団の最適解を駆逐して多様性が失われたために、進化を重ねても局所最適解を抜け出すことができなかった試行が多いからと考えられる。

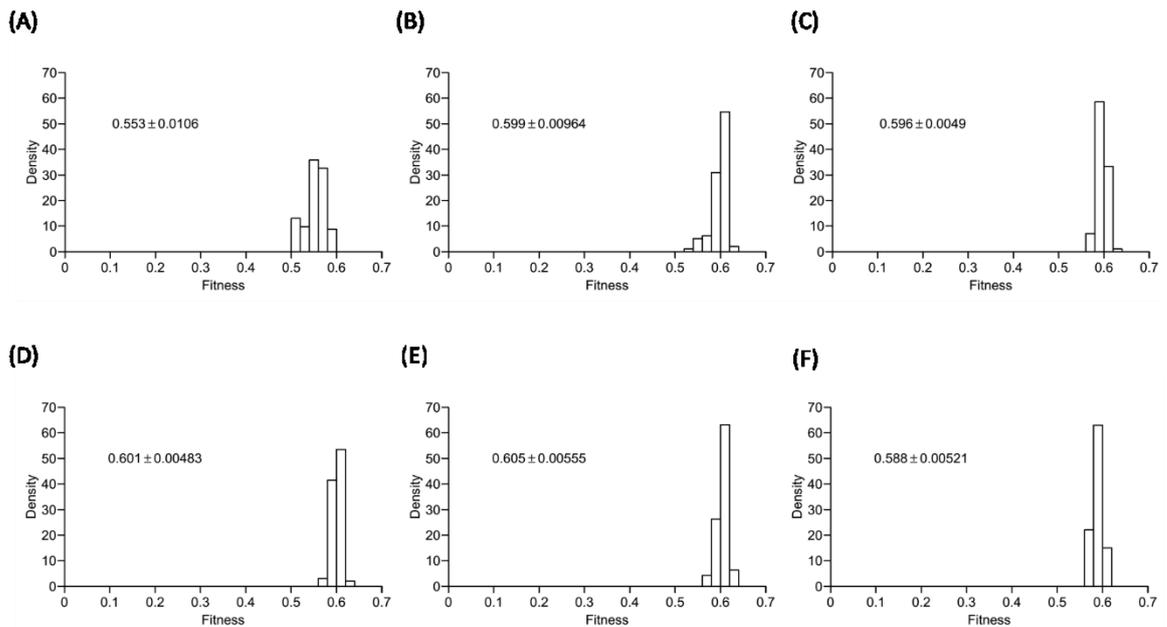


図 11 適応度の分布の比較

全ての場合において、適応度の計算回数を同じ 1000000 とした場合に、分散遺伝的アルゴリズムに用いるパラメータの違いを調べた。(A) 比較のため、遺伝的アルゴリズムを用いた場合。(B) 50 個体 50 世代 20 島 20 ステージ、移住率 0.25。(C) 50 個体 50 世代 20 島 20 ステージ、移住率 0.025。(D) 50 個体 20 世代 50 島 20 ステージ、移住率 0.025。(E) 20 個体 50 世代 50 島 20 ステージ、移住率 0.025。(F) 20 個体 50 世代 50 島 20 ステージ、移住率 0.005。分散遺伝的アルゴリズムを用いた全ての条件において、遺伝的アルゴリズムを用いた場合に比べて適応度が高くなる傾向があり、また、ばらつきが少なくなる傾向がある。

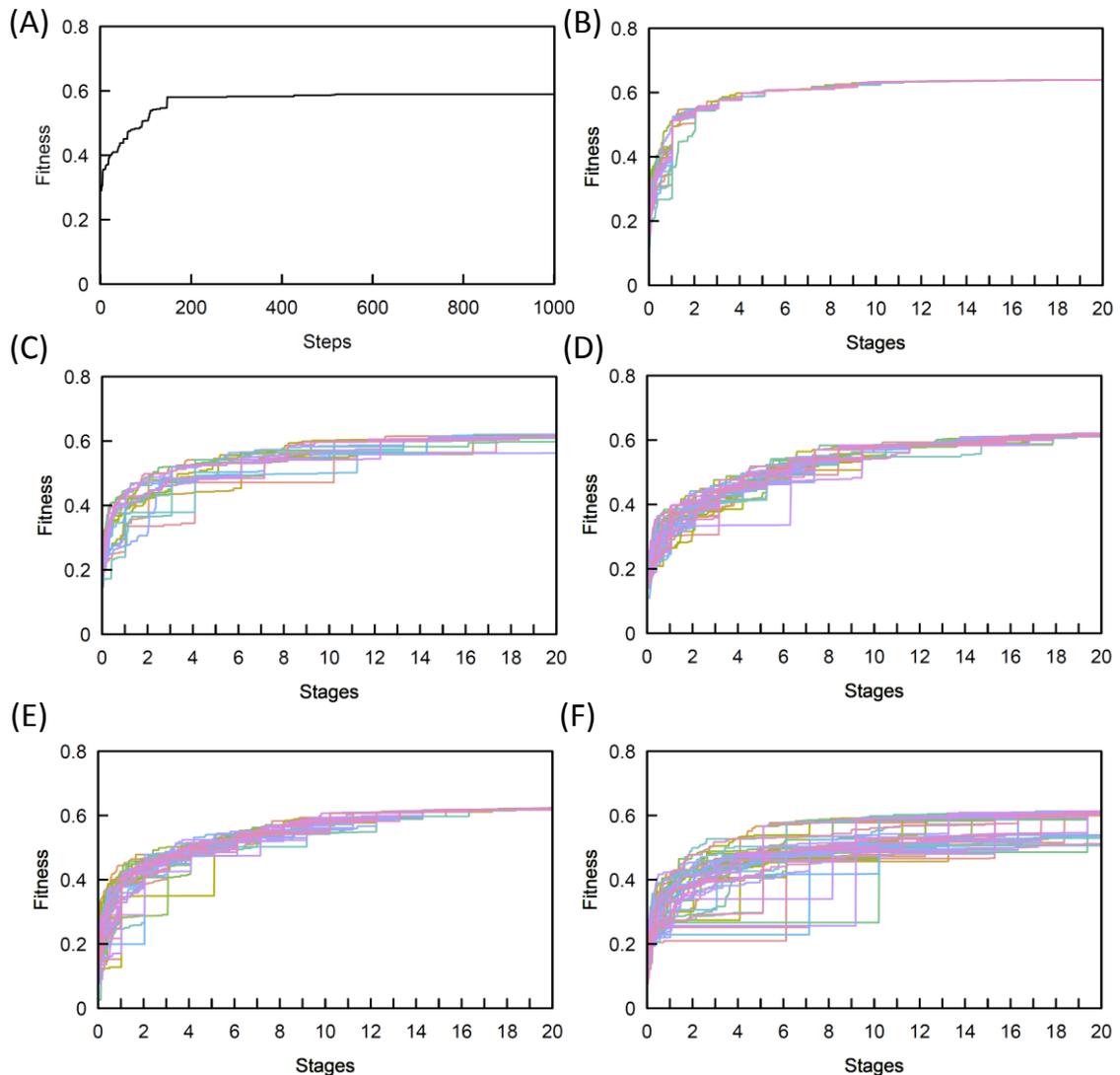


図 12 適応度の履歴の比較

進化のある時点で最も適応度の高い適応度を母集団ごとに記録した履歴を表す。パラメータは図 11 の場合と同じである。(A) 比較のため、遺伝的アルゴリズムを用いた場合。(B) 50 個体 50 世代 20 島 20 ステージ、移住率 0.25。(C) 50 個体 50 世代 20 島 20 ステージ、移住率 0.025。(D) 50 個体 20 世代 50 島 20 ステージ、移住率 0.025。(E) 20 個体 50 世代 50 島 20 ステージ、移住率 0.025。(F) 20 個体 50 世代 50 島 20 ステージ、移住率 0.005。x 軸は、(A) については世代数、(B-F) についてはステージ数を表す。遺伝的アルゴリズムの場合、150 世代目あたりから、進化の仕方が急激に遅くなっているように見える。分散遺伝的アルゴリズムの場合にも、徐々に進化の仕方がゆるやかになっているが、遺伝的アルゴリズムの場合ほど遅くはなっていない。移住率以外の全てのパラメータが同じ (B) と (C) の場合、および (E) と (F) について比較すると、移住率が高い場合には値が収束し、移住率が低い場合には値が分散する傾向がみられた。

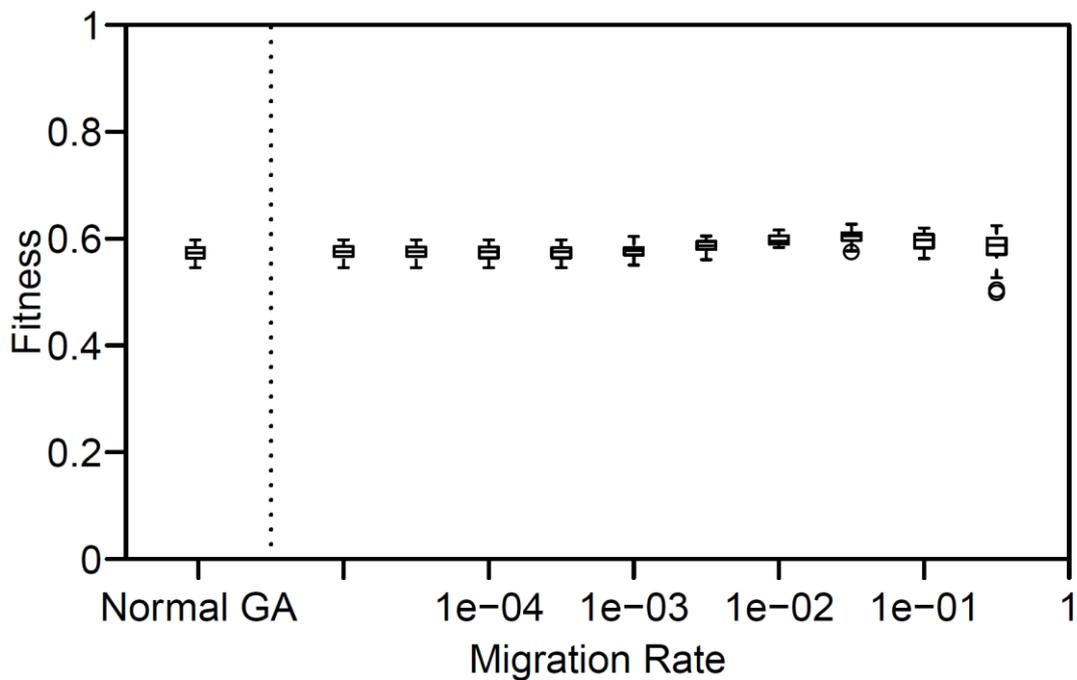


図 13 移住率と最終的な適応度の分布

分散遺伝的アルゴリズムの場合の 50 回の試行について、移住率ごとに適応度の分布を表す。全て 20 個体 50 世代 50 島 20 ステージの場合であり、移住率を、 10^{-5} から $10^{-0.5}$ まで $10^{0.5}$ 倍刻みに変化させた。比較のため、遺伝的アルゴリズムの場合として、同条件で移住率を 0 にした結果を左に示す。分散遺伝的アルゴリズムの場合の全ての移住率において、適応度の中央値は遺伝的アルゴリズムの場合よりも大きい。適応度の中央値は、移住率が $10^{-1.5}$ の場合が最も大きい。移住率が $10^{-0.5}$ の場合には、他の場合に比べて適応度の分散が大きくなっている。

3.3.3 計算によって示された代謝実験の解釈

本手法では、確率的なモデル評価方法を用いているため、最適化の結果は一意に決まるとは限らない。そこで、以下では遺伝的アルゴリズムと、最もよい適応度の分布を示した分散遺伝的アルゴリズムの場合 (20 個体 50 世代 50 島 20 ステージ、移住率 0.025) について、100 回の試行中で適応度が高かった 10 回の試行の結果を用いて、再現性を検証する。初期相 (0-24 h) と後期相 (24-84 h) で得られた代謝の挙動の変化を理解するため、反応係数の初期相の値からの相対的变化を以下のように定義する。

$$\begin{aligned} v_g &= (v_{g1}, \dots, v_{g171}) \\ &= (v_{g1}^{\text{conc}}, \dots, v_{gN}^{\text{conc}}, v_{g1}^{\text{for}}, \dots, v_{gM}^{\text{for}}, v_{g1}^{\text{lat}}, \dots, v_{gM}^{\text{lat}}) \end{aligned} \quad (12)$$

$$w_{gj} = \log_{10}(v_{gj}^{\text{lat}}/v_{gj}^{\text{for}}), \quad (13)$$

ここで、 v_{gj}^{conc} 、 v_{gj}^{for} 、 v_{gj}^{lat} はそれぞれ、 g 回目の最適化計算についての初期濃度、初期相の反応係数、後期相の反応係数を表している。また、比 w_{gj} は、反応係数の変化を表す。10 回の試行について、 t 検定 (閾値 0.05) によって、反応係数が有意に変化したものを調べた。

遺伝的アルゴリズムと分散遺伝的アルゴリズムによる結果からは、それぞれ、8 個、11 個の反応の反応係数が有意に変化したと認められ (表 2)、このうち、7 個の反応は、遺伝的アルゴリズムと分散遺伝的アルゴリズムの両方の結果に認められた。これらの分布を図 14 に示す。選ばれた全ての反応において、反応係数の変化は、増加あるいは減少のどちらかにほとんど偏って分布していることがわかる。これらの反応係数の前半と後半における変化が代謝の挙動の変化を引き起こしたと考えられる。これらの反応は、図 15 において赤と青の矢印で色づけされている。特筆すべきことは、マニトールとラフィノースを消費する反応の係数の変化が、実際にマニトールの大半が初期相で消費され、ラフィノースの

大半が後期相で消費されているという事実と一致することである。この結果から、材料の消費によって、これらの反応に対してなんらかのフィードバックがかかったということが示唆される。しかしながら、その他のほとんどの反応の係数については、有意な変化が見られなかった。この結果から、ある環境変化がおきたとき、それに対して代謝ネットワークが適応しようとした際に、代謝経路上の大部分の活動はほとんど変化せず、非常に限られた経路のみが変化していることが示唆される。これは、局所的な代謝ネットワークの乱れに対して、代謝物質レベルの全体的な安定性は維持される傾向にあるという研究結果 (Ishii et al. 2007) と同様である。Segrè らは、摂動下にある代謝ネットワークの流束を予測するモデルを提案する際に、代謝変化の最小化を行う方法を用いた (Segrè et al. 2002)。これらの研究によると、細胞が環境変化に適応するとき、自身の代謝状態を維持するために必要な最小限の変化を好む傾向にある。

次に、最適化計算の中で最も適応度が高かった場合のパラメータを用いてシミュレーションを行った結果について詳しく調べる。図 16 は、適応度が最も高かったパラメータによるシミュレーションの結果から得られた、各代謝物濃度の初期値からの時間変化量を実験結果と比較したものを示す。5'メチルチオアデノシンと α ケトグルタル酸以外においては、シミュレーション結果の代謝の傾向は実験の結果と非常によく一致した。5'メチルチオアデノシンと α ケトグルタル酸の場合には、計算と実験の結果は大きくずれている。ここで、この理由について考える。図 17 は、この二つの化合物と、一つの反応でつながっている代謝物の濃度変化を図 16 から選び、代謝ネットワークの図に描いたものである。まず、5'メチルチオアデノシンについて考える。図 16 によると、実験において、5'メチルチオアデノシンは 36-60 h にかけて大きく減った後にほぼ初期の量まで戻っている。しかしながら、5'メチルチオアデノシンが途中で減少するという事

象を今回使用した代謝ネットワークを用いて再現することは原理的にできない。なぜなら、図 17 の代謝ネットワークを見ると、5'メチルチオアデノシンは、本モデルにおいて代謝ネットワークの最下流に位置する最終生成物だからである。このため、実験結果と大きく異なる結果となった。次に、 α ケトグルタル酸について考える。図 17 の代謝ネットワーク内において、 α ケトグルタル酸は 3 箇所に現れている。 α ケトグルタル酸の隣接する化合物の実験における挙動を見ると、サッカロピンの増加が顕著であるが、その他の隣接化合物である、イソクエン酸、グルタミン酸、3-ホスホセリン、アスパラギン酸の量はほぼ一定である。このことから、隣接する化合物の量が一定であることを再現するために、計算結果において、 α ケトグルタル酸の生成がほとんど行われなかったものと考えられる。すなわち、 α ケトグルタル酸はサッカロピンの生成に使われて減少したものの、その後、隣接する化合物からの α ケトグルタル酸の生成がないために、 α ケトグルタル酸が減少したままとなったと考えられる。以上をまとめると、ここで挙げた 2 つの化合物について、実験と計算の結果が一致しなかったことには理由があり、それ以外については、計算結果は実験結果と非常に良く一致している。本研究では、相対値の変化のみを元にパラメータを推定しているために、代謝物の初期濃度には任意性が存在するため、推定されたパラメータが唯一の解という訳ではないが、一部を除いて計算結果と実験結果が良く一致したことから、代謝物量の相対変化を再現するパラメータを推定することができたといえる。

さらに、実験でデータが得られなかった代謝物の挙動について、シミュレーションによる予測結果を調べる。図 18 は、代謝シミュレーションで用いた全代謝物の濃度の時間変化（付図 1 参照）を、図 15 の代謝ネットワークの図に加えたものである。粒子数が多くかつ粒子数の変動が大きい代謝物として、実験期間を

通した粒子数の平均が 100 以上かつ粒子数の最大値と最小値の差が 100 以上の代謝物を選んだ。選ばれた代謝物と、それらの代謝物同士を結ぶ反応を、シミュレーションによって予測できたと考え、図 18 に赤色で示した。選ばれた代謝物とそれらをつなぐ反応は、図中の 4 箇所に着目してサブネットワークを形成している。そこで、それぞれのサブネットワーク内の代謝物の動的挙動を詳細に見る。図 18 内の A のサブネットワークにおいて、実験データが存在している代謝物はラフィノースとスクロースであるが、実験データが存在しない下流の 3 つの代謝物の挙動を予測できた。糖であるラフィノースを材料として、その下流のスクロース、フルクトース、フルクトース-6-リン酸、フルクトース-1-6 リン酸が生成されている。図 18 内の B のサブネットワークでは、メチオニンと、ホスホリルコリン、シチジン二リン酸コリンの間の 2 つの化合物を予測できた。最終生成物であるシチジン二リン酸コリンの予測された粒子数は、初期状態の 10 から 84h 後には 216 に増えた。図 18 内の C のサブネットワークにおいては、アスパラギン酸とトレオニンの間の 4 つの代謝物についての動的挙動が予測された。トレオニンの粒子数は 744 から 3347 に増えているが、それに対し、前駆物質である O-ホスホ-L-ホモセリンが 2609 から 6 に減っている。一方、そのさらに上流のホモセリンの粒子数は 5 から 411 に増えている。これらのことから、予測されたモデルでは、トレオニンの増加は直近の前駆物質からの生成によって再現されていることがわかった。図 18 内の D のサブネットワークについては、含まれる 9 個の代謝物のうち、6 個の代謝物について実験データが存在し、3 個の代謝物の挙動が予測できた。

これらの予測には、解の任意性が存在する。たとえば、図 18 内の C のサブネットワークにおいて、アスパラギン酸とトレオニンの間の 4 つの代謝物は分岐なしで連なっており、これらについての関連する拘束条件はアスパラギン酸と

トレオニンの相対量の変化のみであることを考えると、この2つの代謝物を結び一連の5つの反応の反応係数と4つの代謝物の初期濃度からなるパラメータを一意に決めるには条件が少なすぎる。このため、同じ相対変化を示す別のパラメータセットが存在し得る。しかしながら、これらの任意性があるものの、相対値の変化が実験値に合うようにシミュレーションのパラメータを推定することができた。今後、さらに実験データを増やすことによって拘束条件を増やすことができれば、解の任意性を減らすことができる。

質量分析装置によるメタボローム実験によって得られたデータが相対値である以上、初期濃度の任意性は避けられない。また、実験によって全ての代謝物のデータを得られるわけではない。本章では、このような条件下でも、実験で得られた代謝濃度の相対変化をもとに、代謝物の初期濃度と代謝反応の反応係数をSS-GA および SS-dGA を用いて適切に予測することによって、実験で観測された相対変化をよく再現するだけでなく、実験でデータが得られなかった代謝物についての挙動も予測できることが示された。

表 2 初期相と後期相で顕著に変化の見られた反応 (A) 遺伝的アルゴリズムによる結果 (B) 分散遺伝的アルゴリズムによる結果

(A)

| EC Number | Mean Ratio | P-value | Function |
|-----------|----------------------|----------------------|--------------------------------------|
| 2.4.1.- | 8.6×10^3 | 9.2×10^{-5} | Ajugose biosynthesis II |
| 1.1.1.255 | 1.1×10^{-2} | 3.5×10^{-4} | Mannitol dehydrogenase |
| 2.7.7.15 | 2.4×10^{-1} | 2.4×10^{-3} | Choline-phosphate cytidyltransferase |
| 4.2.1.3 | 8.7×10^{-1} | 5.5×10^{-3} | Aconitase |
| 4.2.3.1 | 3.0×10^{-1} | 5.8×10^{-3} | Threonine synthase |
| 4.4.1.8 | 2.7×10^6 | 1.7×10^{-2} | Cystathionine β -lyase |
| 2.1.1.10 | 5.4×10^0 | 2.0×10^{-2} | Homocysteine S-methyltransferase |
| 4.2.1.2 | 2.0×10^0 | 2.2×10^{-2} | fumarase A |

表 2 (続き)

(B)

| EC Number | Mean Ratio | P-value | Function |
|-----------|----------------------|----------------------|--|
| 1.1.1.255 | 9.1×10^{-3} | 3.6×10^{-5} | Mannitol dehydrogenase |
| 1.1.1.25 | 1.2×10^3 | 4.8×10^{-5} | Shikimate dehydrogenase |
| 2.6.1.52 | 9.5×10^{-2} | 2.1×10^{-4} | 3-phosphoserine aminotransferase |
| 4.2.3.1 | 4.4×10^{-1} | 9.8×10^{-4} | Threonine synthase |
| 2.4.1.- | 1.6×10^{-7} | 1.7×10^{-3} | Ajugose biosynthesis II |
| 2.7.7.15 | 1.1×10^{-1} | 3.9×10^{-3} | Choline-phosphate cytidyltransferase |
| 2.1.1.10 | 3.0×10^{-1} | 4.1×10^{-3} | Homocysteine S-methyltransferase |
| 4.4.1.8 | 1.0×10^6 | 4.7×10^{-3} | Cystathionine β -lyase |
| 4.2.1.3 | 5.0×10^{-1} | 1.6×10^{-2} | Aconitase |
| 2.5.1.54 | 4.7×10^0 | 4.5×10^{-2} | 2-dehydro-3-deoxyphosphoheptonate aldolase |
| 4.1.2.13 | 3.8×10^4 | 4.7×10^{-2} | Fructose-bisphosphate aldolase |

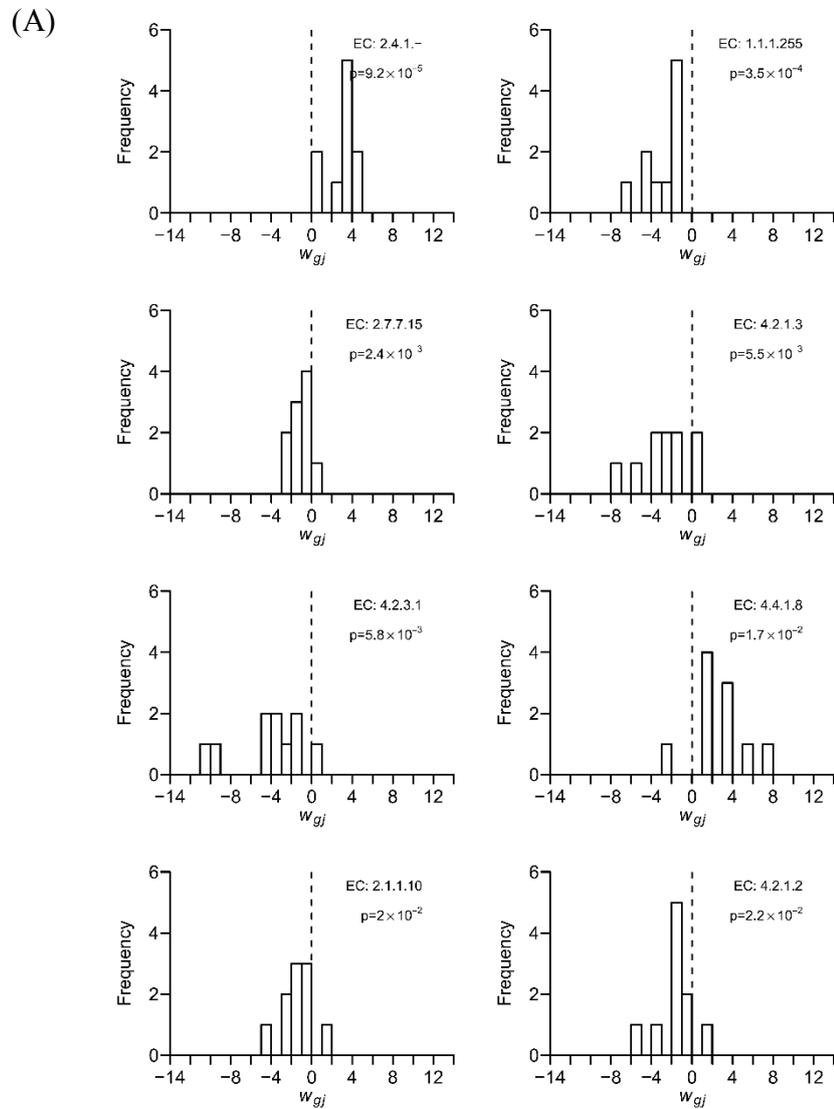


図 14 反応係数の変化の分布 (A) 遺伝的アルゴリズムによる結果 (B) 分散
遺伝的アルゴリズムの結果

最も適応度が高かった分散遺伝的アルゴリズム (20 個体 50 世代 50 島 20 ステージ、
移住率 0.025) の 100 回の試行について、その結果から適応度が上位 10 回に入った計算
について、反応係数が初期相と後期相で顕著に変化した反応を選んだ。ほとんどの反応
について、反応係数の変化の傾向が増加あるいは減少のどちらかに偏っていることがわ
かる。

(B)

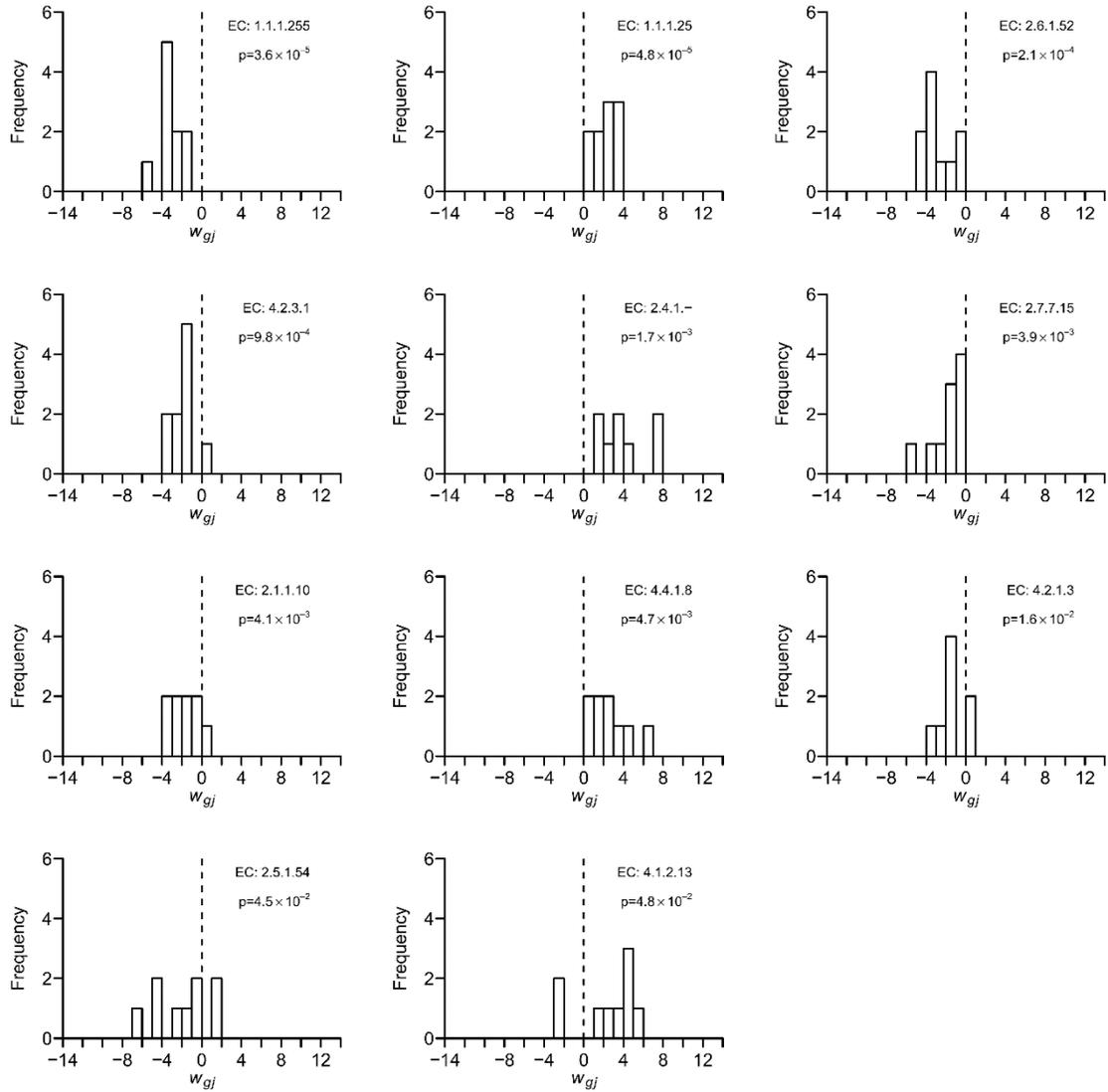


図 14 (続き)

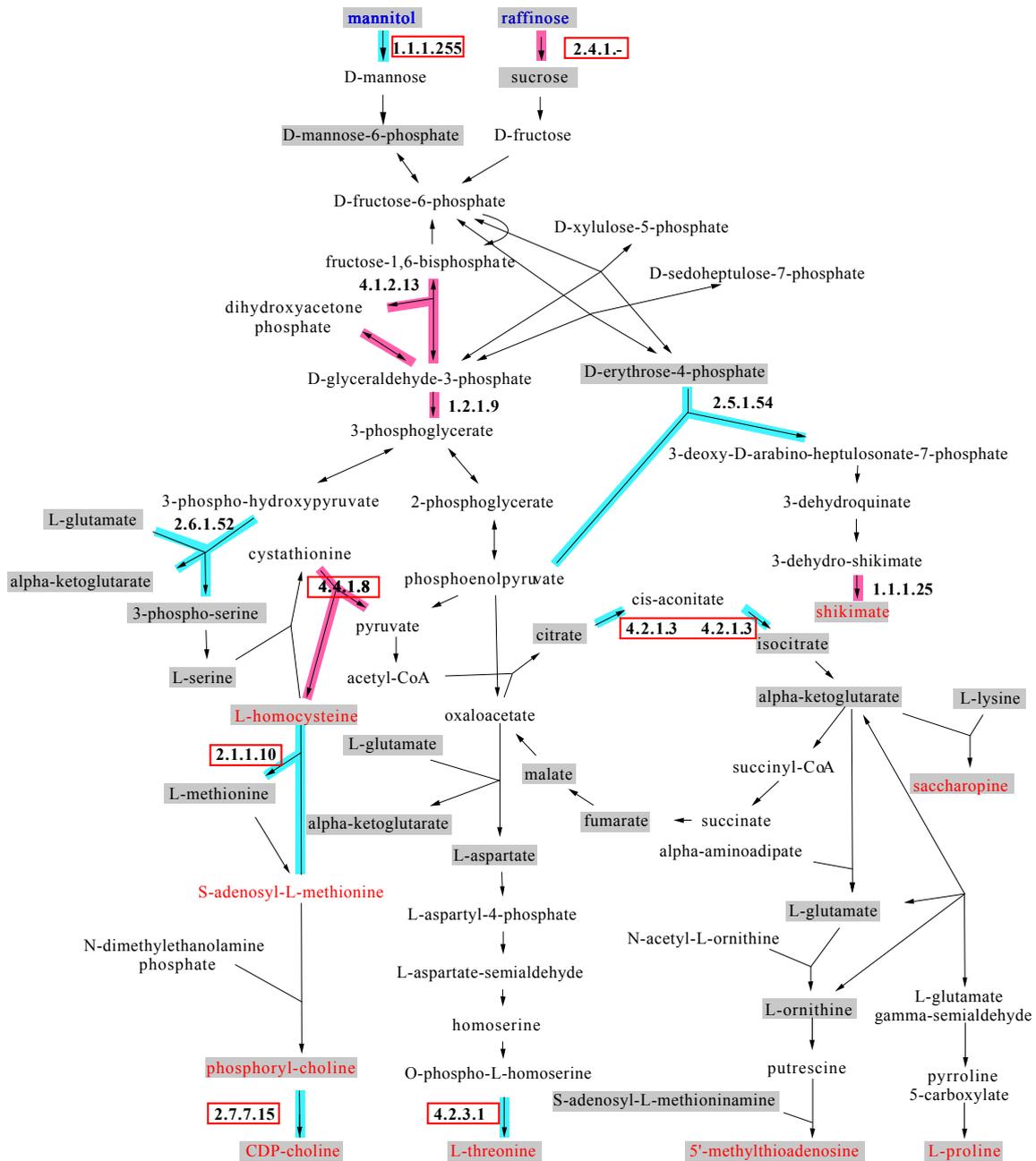


図 15 反応係数が変化した反応

最も適応度が高かった分散遺伝的アルゴリズム (20 個体 50 世代 50 島 20 ステージ、移住率 0.025) の 100 回の試行について、その結果から適応度が上位 10 回に入った計算の結果において、反応係数が初期相と後期相で顕著に変化した反応を選び、図 6 の代謝マップに色付けした。再現計算の結果、反応係数が初期相 (0-24 h) と後期相 (24-84 h) で顕著に変化した反応を赤 (増加) と青 (減少) の矢印で示す。また、遺伝的アルゴリズムの場合においても同様の方法で選ばれた 7 個の反応の反応名を赤枠で囲んだ。

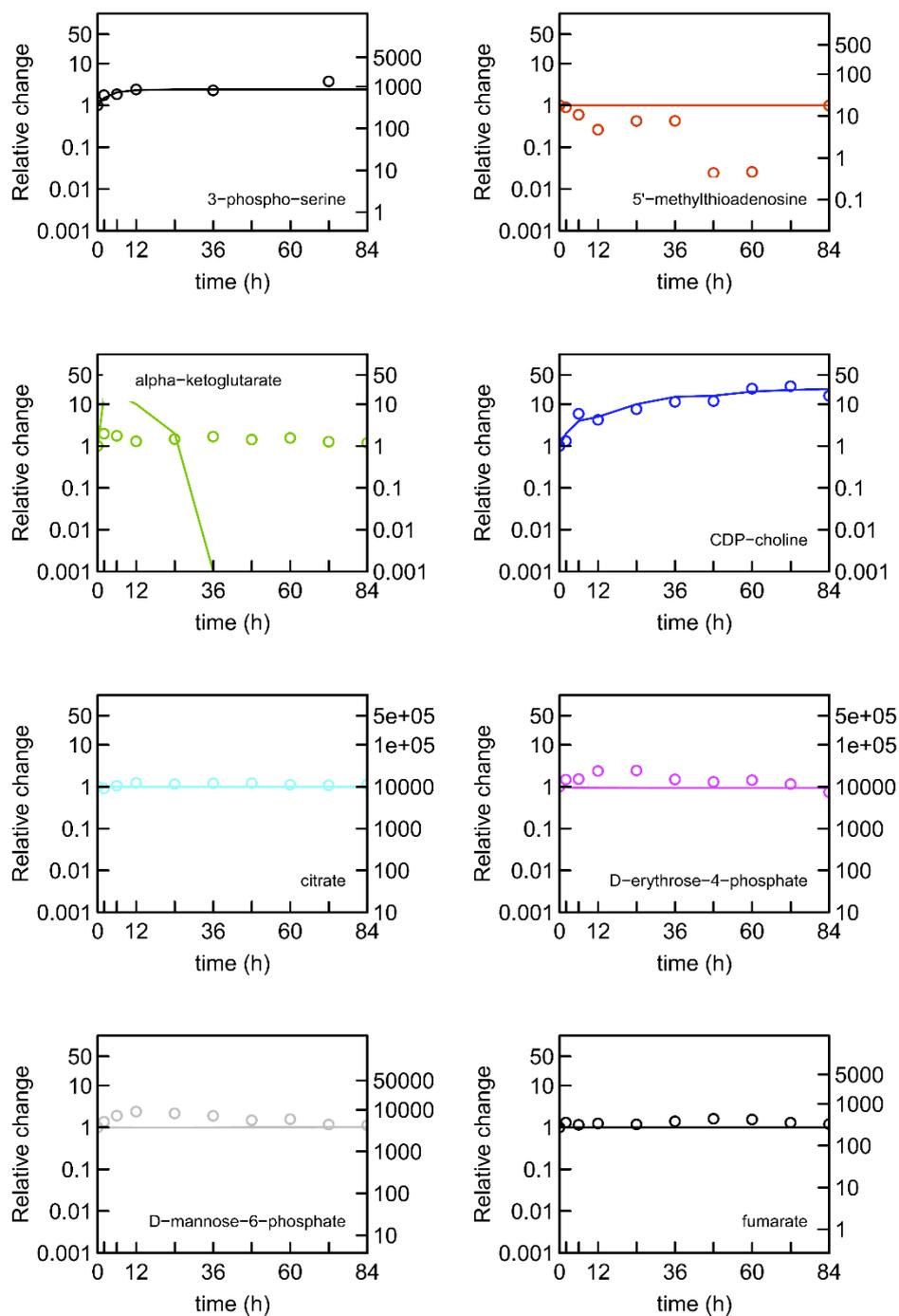


図 16 シミュレーションにより再現された各代謝物の挙動の比較

最も適応度が高かった分散遺伝的アルゴリズム (20 個体 50 世代 50 島 20 ステージ、移住率 0.025) の場合の各代謝物の初期濃度からの変動の履歴を示す。線がシミュレーション結果を、点が実験結果を表す。左の軸は初期値からの変化を、右の軸はシミュレーションにおける分子の個数を表す。5'メチルチオアデノシンと α ケトグルタル酸を除き、非常によい一致が得られた。

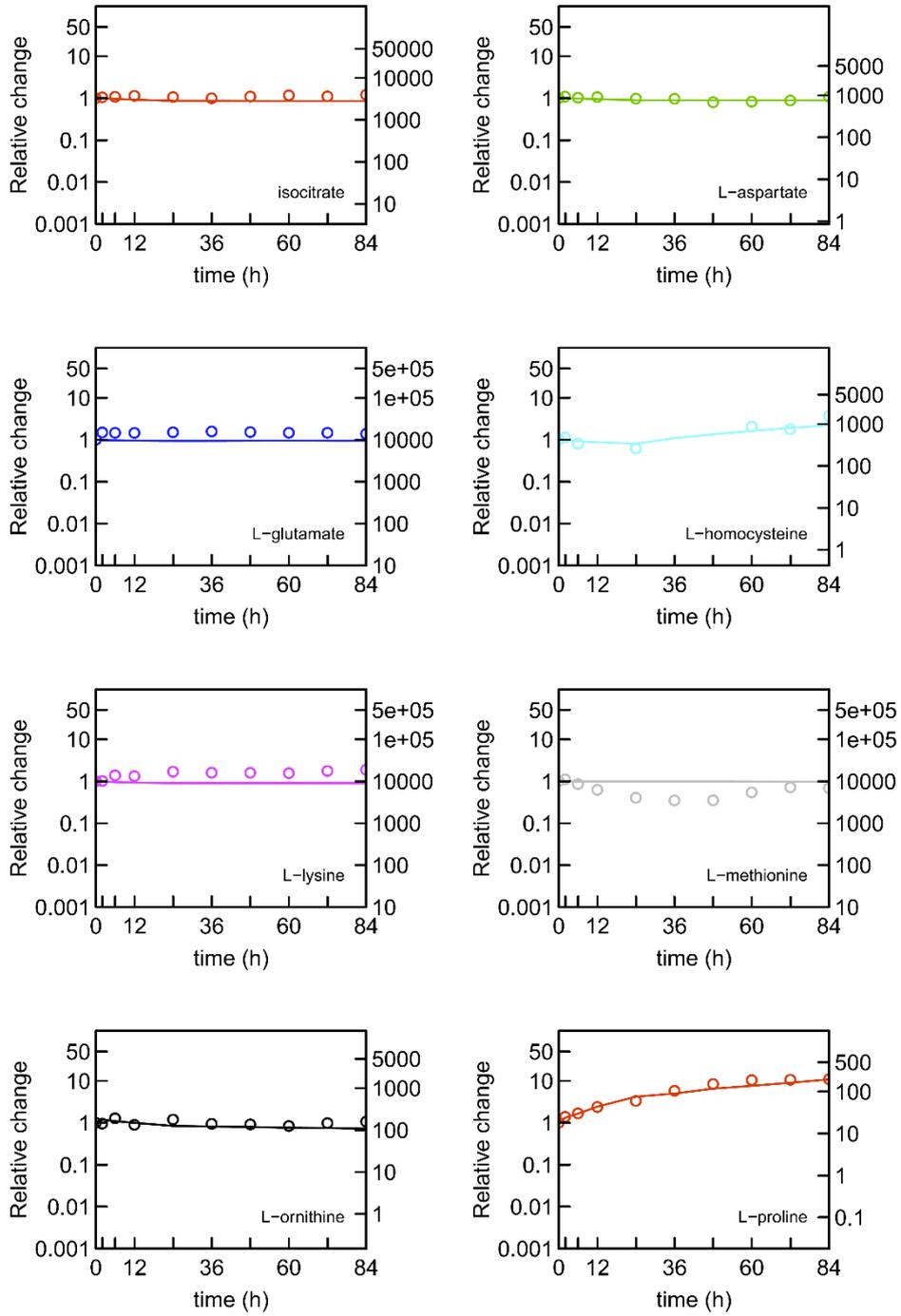


図 16 (続き)

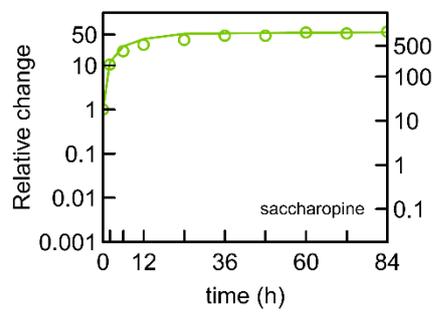
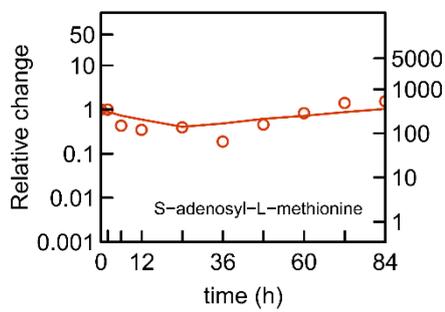
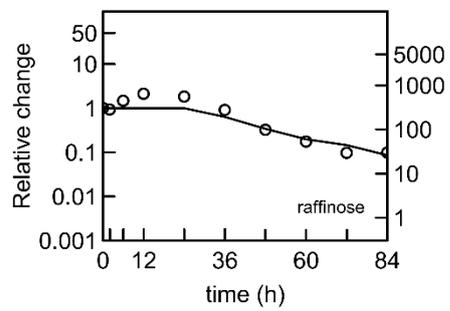
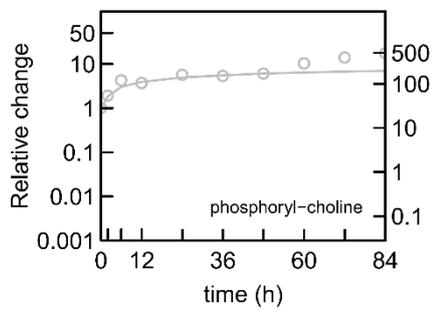
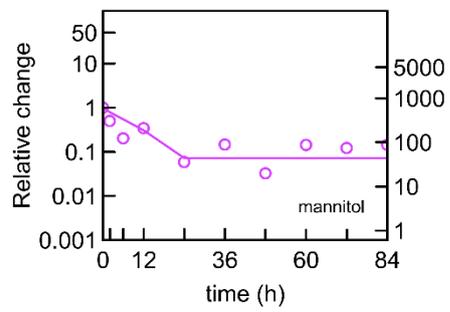
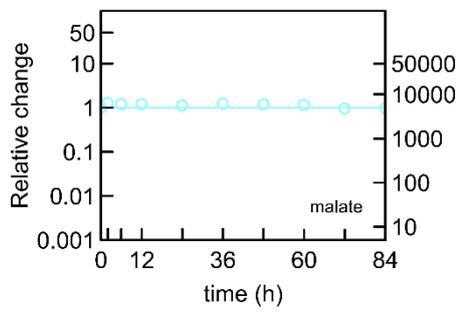
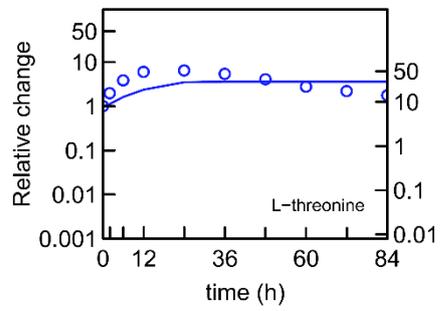
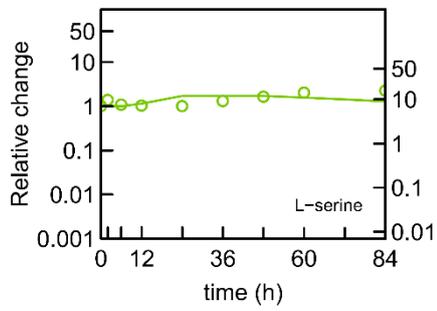


図 16 (続き 2)

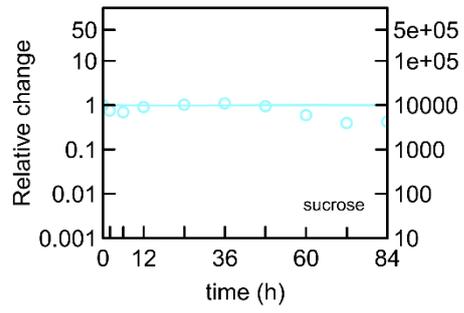
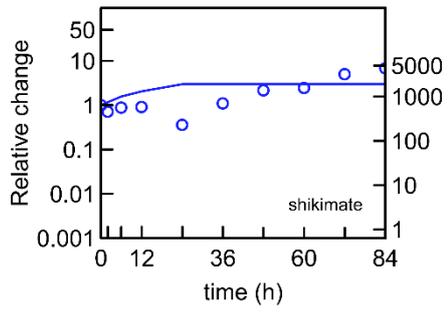


図 16 (続き 3)

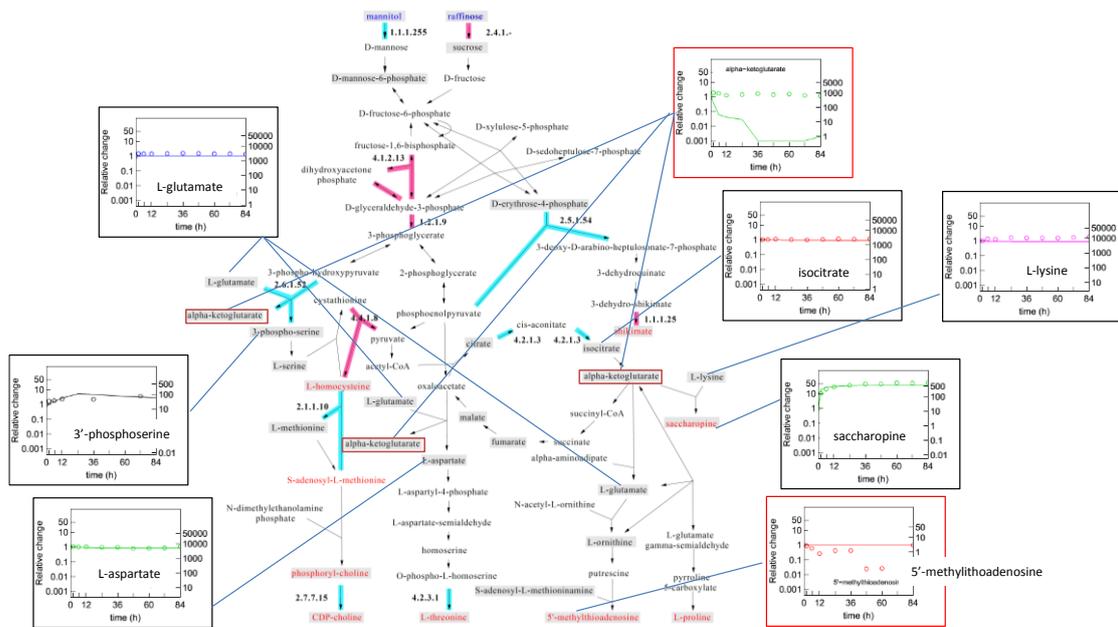


図 17 5'メチルチオアデノシンと α ケトグルタル酸の予測値が実験値と離れた理由

図 15 の結果に、5'メチルチオアデノシンと α ケトグルタル酸について、それ自身 (赤枠) とその周辺の代謝物 (黒枠) の動的挙動の変化のグラフを描いた。これらの2つの代謝物の予測値が実験値と離れた理由はそれぞれ異なると考えられる。前者については、対象として設定した代謝ネットワークでは、5'メチルチオアデノシンを減少させる反応は存在しないために、実験値の減少を再現できなかったからである。後者については、下流側の代謝物の生成に α ケトグルタル酸が使われたにも関わらず、上流側の代謝物からの供給がなかったため、初期の濃度を維持できなかったからと考えられる。

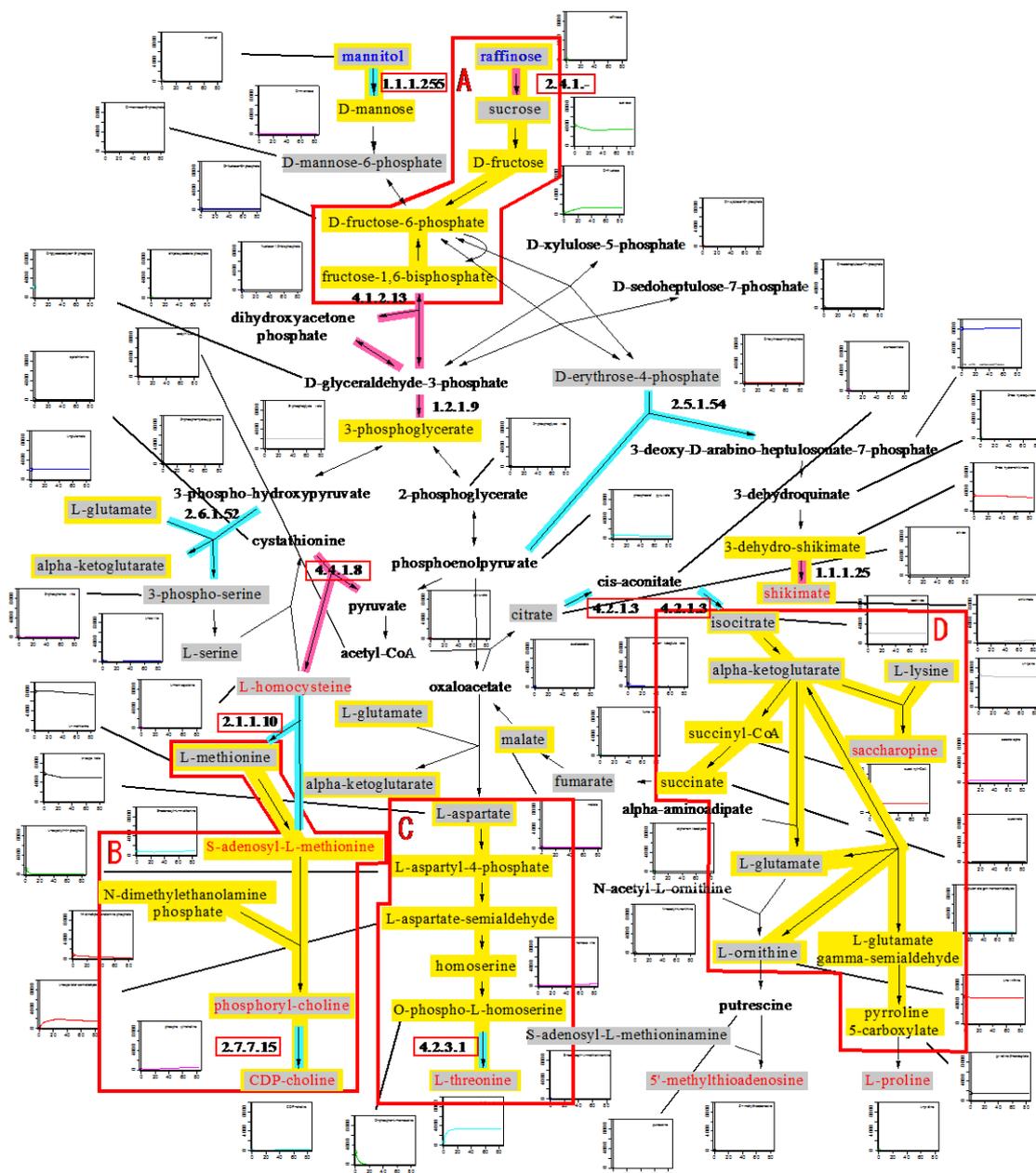


図 18 実験でデータが得られなかった代謝物について、シミュレーションによって予測された時間変化

対象の代謝ネットワーク内の全代謝物濃度の時間変化（付録参照）を代謝ネットワーク上に示す。粒子数の平均値と、最大値と最小値の差が 100 以下の化合物とそれらをつなぐ代謝物を黄色で表す。これらをつないだ 4 つのサブネットワークが得られた (A-D)。

第4章 おわりに

一般に、大規模な代謝ネットワークを量的にモデル化するのは、質量分析装置から得られる大量のデータがあったとしても、なお難しい。本論文では、その理由を (1) モデルに必要なパラメータ数が膨大である、(2) 決定論的な手法では、確率過程的な振る舞いを再現できない、(3) 実験データを再現するために必要な入力値が実験から全て手に入るわけではない、の3点に整理し、研究を進めた。これらの3点について、それぞれ以下の方法で対処した。すなわち、(1) については、興味の対象を含むサブネットワークのみを抽出することによって、推定すべきパラメータ数を減らした。(2) については、確率的シミュレーション方法を用いた。(3) については、パラメータ最適化により実験を再現できるパラメータを推定した。

第2章で、始端化合物と終端化合物を指定すると、最短経路探索によって全代謝リストから部分代謝ネットワークを抽出し、ユーザが与えたパラメータを入力値とした Gillespie アルゴリズムに基づく確率的シミュレーションを行うことが可能なツール SS-mPMG について述べた。

第3章で、遺伝的アルゴリズムおよび分散遺伝的アルゴリズムに基づくパラメータ最適化によって、実験で得られた代謝データを再現できるパラメータセットを、シミュレーション結果と実験結果における各代謝物の初期値からの相対変化量に着目して推定し、そのパラメータを用いて確率的シミュレーションによって実験データを再現することができるツール SS-GA, SS-dGA を開発した。

これらの開発したツール群をシロイヌナズナの代謝実験の再現に用いた。まず、実験結果に基づいて始端化合物と終端化合物を選び、AraCyc パスウェイ・データベースの全代謝反応リストからパスウェイ抽出を行った (第2章)。次に、抽出されたパスウェイを対象とし、SS-GA, SS-dGA を用いて実験で計測された

シロイヌナズナの代謝量の変化を再現した。シミュレーションにより再現された代謝物量の初期値からの相対変化の時系列は、実験結果のそれとよく一致した。さらに、シミュレーションによって再現された代謝反応は、環境摂動に対して代謝ネットワーク中の一部の反応のみを変化させるという挙動を示しており、代謝ネットワークのロバスト性を示した結果 (Ishii et al. 2007) と同様の挙動であることが示唆された。また、シミュレーションによって、実験で計測できなかった代謝物の動的挙動について予測することが可能であることが示された (第3章)。

質量分析装置等の計測技術の発達により、得られる代謝データの質と量が急激に増えている現在において、実験結果からは直接見ることができない代謝の詳細な挙動を、シミュレーションによって明らかにすることは、ますます重要となっている。本研究では、モデル化に関して隘路となっているモデルの巨大さ、代謝の確率的挙動、代謝実験データの不足という課題を、サブネットワークの抽出、確率的シミュレーション、パラメータ推定によって解決した。

生物の生命活動を支える代謝反応には酵素が不可欠である。多くの酵素は生体内で生成されるタンパク質を基にして構成されるため、生命活動は、タンパク質とそれを作るための設計図である遺伝情報によって支えられているといえる。本研究では、代謝動態のみに焦点を絞って実験で得られた代謝物の挙動を再現する手法を提案した。今後、計算機の性能がさらに上がり、また、実験技術の発展によって未知パラメータの同定が進めば、遺伝子発現をも含めたモデル、すなわち、DNA から mRNA への遺伝情報の転写、さらにタンパク質への翻訳を含むセントラルドグマ全体を包含した多階層のモデルの構築が可能となり、さらには、細胞全体の挙動、細胞同士の相互作用、あるいは細胞外の環境の影響等を考慮に入れたモデルの構築が可能となると期待している。

謝辞

本研究を遂行するにあたり、奈良先端科学技術大学院大学情報科学研究科 金谷 重彦 教授から、主指導教員としてご指導をいただきました。初めて研究室に伺ったころ、私は何年も前に遭った交通事故の後遺症と思われる症状が原因で、それまで所属していた環境での研究継続を諦めることとなって、途方に暮れていました。そんなとき、情報生命とは全く異なるバックグラウンドをもつ私を博士後期課程の学生として快く受け入れてくださいました。入学後は、研究に関することに限らず本当に様々な助言をいただきました。特に、ユニークな発想と、新規の事柄を始める際の取り組み方、研究遂行の仕方について、今後の人生において大いに参考にさせていただこうと思います。近くで研究されている姿を拝見するだけでも、勉強になることがたくさんありました。このようなすばらしい環境で研究を行う機会を頂いたことに、厚く御礼申し上げます。

奈良先端科学技術大学院大学情報科学研究科 安本 慶一 教授には、遺伝的アルゴリズムについて、研究室のプログラムを参考として快く提供していただきました。また、ミーティング等でお会いした際には、毎度温かいお言葉をいただきました。奈良先端科学技術大学院大学情報科学研究科 杉浦 忠男 准教授には、休日や、夜遅くまで淡々と研究されている姿に刺激を頂きました。奈良先端科学技術大学院大学情報科学研究科 Md. Altaf-Ul-Amin 准教授には、英語の指導、論文レフリーの仕方等の指導を頂きました。奈良先端科学技術大学院大学情報科学研究科 小野 直亮 助教には、研究の進め方から投稿論文作成に関することまで、ありとあらゆることのご指導を頂きました。また、しばしば夜遅くまで、研究に関することや、その他の大変興味深いお話をさせていただきました。さらに、研究に関する進め方、論理の構築方法についても、いつもの的確な助言を頂きました。奈良先端科学技術大学院大学情報科学研究科 佐藤 哲大 助教には、研究で

行き詰っているとき等、私が困っていると、いつもさりげなく的確なご助言を頂きました。前橋工科大学生命情報学科 中村 建介 教授には、当研究室在籍当時に、様々な助言を頂きました。また、たくさんの楽しいお話をして頂きました。千葉大学真菌医学研究センター 高橋 弘喜 テニユアトラック准教授には、研究室在籍時、未だ情報生命学分野について無知だった研究室配属当初の私に、分野のいろはを教えてくださいました。以上の先生方のご指導があつて、ここまで研究を進めることができました。ここに厚く御礼申し上げます。

この他、奈良先端科学技術大学院大学情報科学研究科計算システムズ生物学研究室の皆さんからは多大なるご助力を頂きました。中村 由紀子 博士研究員には、しばしば、弱っているときや行き詰っている、ここぞというときに温かい励ましの言葉を頂きました。いつも研究室を明るい雰囲気にしてくださっている森田 晶 研究員には、研究だけでなく、多岐に渡る方面で支えて頂きました。特に、席が近くなつてからは、私のつまらない与太話にもいつも笑顔で答えてくださり、たくさんの元気を頂きました。本研究室卒業生の和田 眞昌さんには、研究室配属当初からたくさんの助言を頂きました。また、研究以外に関しても夜遅くまで様々な議論につきあつてくださいました。同じく、本研究室卒業生の池田 俊さんにも、研究室配属当初から、研究だけではなく、様々なことについて助言を頂きました。また、同じ博士後期課程の学生として、たくさんのことについて教えていただきました。Nelson Kibinge さんには、いつも楽しい話をして頂きました。結婚式に出席するためにケニアまで行ったのはいい思い出です。また、発表等において、英語の指導を度々して頂きました。Li Donghan さんには、一緒に旅行に行くなど、楽しい時間を共有して頂きました。この他にも、研究室在籍中、研究室の皆さんには多くのご助力を頂きました。ありがとうございます。

また、私がこの学校に在籍した間、たくさんの方々にご助力いただきました。

視覚情報メディア研究室博士後期課程に在学中の青砥 隆仁さん、ユビキタスコンピューティングシステム研究室博士後期課程に在学中の柏本 幸俊さんには、所属研究室が異なるにもかかわらず、いつも様々な議論につきあっていただきました。その他、奈良先端科学技術大学院大学のスタッフの皆様にも、顔を覚えていただき、たくさんの励ましを頂きました。感謝いたします。

最後になりますが、本当に長年にわたって、辛抱強く私を学校に通わせてくださった私の両親に感謝いたします。

参考文献

- Afendi, F.M. et al., 2012. KNApSAcK family databases: integrated metabolite-plant species databases for multifaceted plant research. *Plant Cell Physiol.*, 53(2), e1.
- Alba, E. & Troya, J.M., 1999. A survey of parallel distributed genetic algorithms. *Complexity*, 4(4), pp.31–52.
- Arkin, A., Ross, J. & McAdams, H.H., 1998. Stochastic kinetic analysis of developmental pathway bifurcation in phage lambda-infected *Escherichia coli* cells. *Genetics*, 149(4), pp.1633–48.
- Bedair, M. & Sumner, L.W., 2008. Current and emerging mass-spectrometry technologies for metabolomics. *TrAC Trends Anal. Chem.*, 27(3), pp.238–250.
- Belding, T., 1995. The distributed genetic algorithm revisited. In *Proceedings of the 6th International Conference on Genetic Algorithms*. pp. 114–121.
- Buchanan, B.B., Gruissem, W. & Jones, R.L., 2007. *Biochemistry and Molecular Biology of Plants*, I.K. International.
- Buchanan, M. et al. eds., 2010. *Networks in Cell Biology*, Cambridge: Cambridge University Press.
- Buchholz, A. et al., 2002. Metabolomics: quantification of intracellular metabolite dynamics. *Biomol. Eng.*, 19(1), pp.5–15.
- Cantu-Paz, E., 2000. *Efficient and Accurate Parallel Genetic Algorithms*, Norwell, MA, USA: Kluwer Academic Publishers.
- Caspi, R. et al., 2012. The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of pathway/genome databases. *Nucleic Acids Res.*, 40(Database issue), pp.D742–53.
- Chassagnole, C. et al., 2002. Dynamic modeling of the central carbon metabolism of *Escherichia coli*. *Biotechnol. Bioeng.*, 79(1), pp.53–73.

- Collakova, E., Yen, J.Y. & Senger, R.S., 2012. Are we ready for genome-scale modeling in plants? *Plant Sci.*, 191-192, pp.53–70.
- Corcoran, A.L. & Sen, S., 1994. Using real-valued genetic algorithms to evolve rule sets for classification. In *Proceedings of the First IEEE Conference on Evolutionary Computation. IEEE World Congress on Computational Intelligence*. IEEE, pp. 120–124.
- Crick, F., 1958. On protein synthesis. *Symp. Soc. Exp. Biol.*, 12, pp.138–63.
- Croft, D. et al., 2011. Reactome: a database of reactions, pathways and biological processes. *Nucleic Acids Res.*, 39(Database issue), pp.D691–7.
- Dijkstra, E.W., 1959. A note on two problems in connexion with graphs. *Numer. Math.*, 1(1), pp.269–271.
- Edwards, J.S., Ø, B. & Papin, 2000. The *Escherichia coli* MG1655 *in silico* metabolic genotype: its definition, characteristics, and capabilities. *Proc. Natl. Acad. Sci. U. S. A.*, 97(10), pp.5528–33.
- Feist, A.M. & Palsson, B.Ø., 2008. The growing scope of applications of genome-scale metabolic reconstructions using *Escherichia coli*. *Nat. Biotechnol.*, 26(6), pp.659–67.
- Fiehn, O., Kloska, S. & Altmann, T., 2001. Integrated studies on plant biology using multiparallel techniques. *Curr. Opin. Biotechnol.*, 12(1), pp.82–86.
- Gillespie, D., 1977. Exact stochastic simulation of coupled chemical reactions. *J. Phys. Chem.*, 93555(1), pp.2340–2361.
- Gillespie, D.T., 1976. A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. *J. Comput. Phys.*, 22(4), pp.403–434.
- Goldberg, D.E., 1989. *Genetic algorithms in search, optimization and machine learning* 1st ed., Boston, MA, USA: Addison-Wesley Longman Publishing Co., Inc.
- Goldberg, D.E., 1994. Genetic and evolutionary algorithms come of age. *Commun. ACM*, 37(3), pp.113–119.

- Hirai, M.Y. et al., 2005. Elucidation of gene-to-gene and metabolite-to-gene networks in arabidopsis by integration of metabolomics and transcriptomics. *J. Biol. Chem.*, 280(27), pp.25590–5.
- Hoops, S. et al., 2006. COPASI--a COmplex PATHway SIMulator. *Bioinformatics*, 22(24), pp.3067–74.
- Huang, Y.-P. & Huang, C.-H., 1997. Real-valued genetic algorithms for fuzzy grey prediction system. *Fuzzy Sets Syst.*, 87(3), pp.265–276.
- Hynne, F., Danø, S. & Sørensen, P.G., 2001. Full-scale model of glycolysis in *Saccharomyces cerevisiae*. *Biophys. Chem.*, 94(1-2), pp.121–63.
- Initiative, T.A.G., 2000. Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature*, 408(6814), pp.796–815.
- Ishii, N. et al., 2007. Multiple high-throughput analyses monitor the response of *E. coli* to perturbations. *Science*, 316(5824), pp.593–7.
- Ishii, N. et al., 2004. Toward large-scale modeling of the microbial cell for computer simulation. *J. Biotechnol.*, 113(1-3), pp.281–94.
- John Henry Holland, 1975. *Adaptation in natural and artificial systems*, MIT Press.
- Jones, G., Willett, P. & Glen, R.C., 1995. Molecular recognition of receptor sites using a genetic algorithm with a description of desolvation. *J. Mol. Biol.*, 245(1), pp.43–53.
- Joyce, A.R. & Palsson, B.Ø., 2006. The model organism as a system: integrating “omics” data sets. *Nat. Rev. Mol. Cell Biol.*, 7(3), pp.198–210.
- Kanehisa, M. & Goto, S., 2000. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.*, 28(1), pp.27–30.
- Karr, J.R. et al., 2012. A whole-cell computational model predicts phenotype from genotype. *Cell*, 150(2), pp.389–401.
- Katsuragi, T. et al., 2014. Dynamic simulation of metabolic network in *Arabidopsis thaliana* using parameters estimated by a distributed genetic algorithm. Poster

- presented at the 10th International conference of the Metabolomics Society. P311, June 25, 2014, Tsuruoka, Japan.
- Katsuragi, T. et al., 2013. SS-mPMG and SS-GA: tools for finding pathways and dynamic simulation of metabolic networks. *Plant Cell Physiol.*, 54(5), pp.728–39.
- Katsuragi, T. et al., 2012. Stochastic simulation of metabolic network of *Arabidopsis thaliana* using experimental data. Poster presented at XV International Congress on Molecular Plant-Microbe Interactions. PS15–582, July 31, 2012, Kyoto, Japan.
- Kummer, U. et al., 2005. Transition from stochastic to deterministic behavior in calcium oscillations. *Biophys. J.*, 89(3), pp.1603–11.
- McAdams, H.H. & Arkin, A., 1997. Stochastic mechanisms in gene expression. *Proc. Natl. Acad. Sci.*, 94(3), pp.814–819.
- Morbiducci, U., Tura, A. & Grigioni, M., 2005. Genetic algorithms for parameter estimation in mathematical modeling of glucose metabolism. *Comput. Biol. Med.*, 35(10), pp.862–74.
- Morgan, J.A. & Rhodes, D., 2002. Mathematical modeling of plant metabolic pathways. *Metab. Eng.*, 4(1), pp.80–9.
- Mueller, L.A., Zhang, P. & Rhee, S.Y., 2003. AraCyc: a biochemical pathway database for *Arabidopsis*. *Plant Physiol.*, 132(2), pp.453–60.
- Murota, K. et al., 2011. *Arabidopsis* cell-free extract, ACE, a new in vitro translation system derived from *Arabidopsis* callus cultures. *Plant Cell Physiol.*, 52(8), pp.1443–53.
- Ogata, H. et al., 1999. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res.*, 27(1), pp.29–34.
- De Oliveira Dal’Molin, C.G. et al., 2010. AraGEM, a genome-scale reconstruction of the primary metabolic network in *Arabidopsis*. *Plant Physiol.*, 152(2), pp.579–89.
- Orth, J.D. et al., 2011. A comprehensive genome-scale reconstruction of *Escherichia coli* metabolism—2011. *Mol. Syst. Biol.*, 7, p.535.

- Pace, N.R., 2001. The universal nature of biochemistry. *Proc. Natl. Acad. Sci. U. S. A.*, 98(3), pp.805–8.
- Pahle, J. et al., 2012. Biochemical fluctuations, optimisation and the linear noise approximation. *BMC Syst. Biol.*, 6(1), p.86.
- Rochfort, S., 2005. Metabolomics reviewed: a new “omics” platform technology for systems biology and implications for natural products research. *J. Nat. Prod.*, 68(12), pp.1813–20.
- Ryan, D. & Robards, K., 2006. Metabolomics: The greatest omics of them all? *Anal. Chem.*, 78(23), pp.7954–8.
- Saito, K. & Matsuda, F., 2010. Metabolomics for functional genomics, systems biology, and biotechnology. *Annu. Rev. Plant Biol.*, 61, pp.463–89.
- Sato, S. et al., 2008. Time-resolved metabolomics reveals metabolic modulation in rice foliage. *BMC Syst. Biol.*, 2, p.51.
- Sawada, Y. et al., 2009. Widely targeted metabolomics based on large-scale MS/MS data for elucidating metabolite accumulation patterns in plants. *Plant Cell Physiol.*, 50(1), pp.37–47.
- Schwender, J., Ohlrogge, J. & Shachar-Hill, Y., 2004. Understanding flux in plant metabolic networks. *Curr. Opin. Plant Biol.*, 7(3), pp.309–17.
- Segrè, D., Vitkup, D. & Church, G.M., 2002. Analysis of optimality in natural and perturbed metabolic networks. *Proc. Natl. Acad. Sci. U. S. A.*, 99(23), pp.15112–7.
- Sheridan, R.P. & Kearsley, S.K., 1995. Using a genetic algorithm to suggest combinatorial libraries. *J. Chem. Inf. Model.*, 35(2), pp.310–320.
- Sumner, L.W., Mendes, P. & Dixon, R. a, 2003. Plant metabolomics: large-scale phytochemistry in the functional genomics era. *Phytochemistry*, 62(6), pp.817–836.
- Tanese, R., 1989. Distributed Genetic Algorithms. In *Proceedings of the 3rd International Conference on Genetic Algorithms*. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., pp. 434–439.

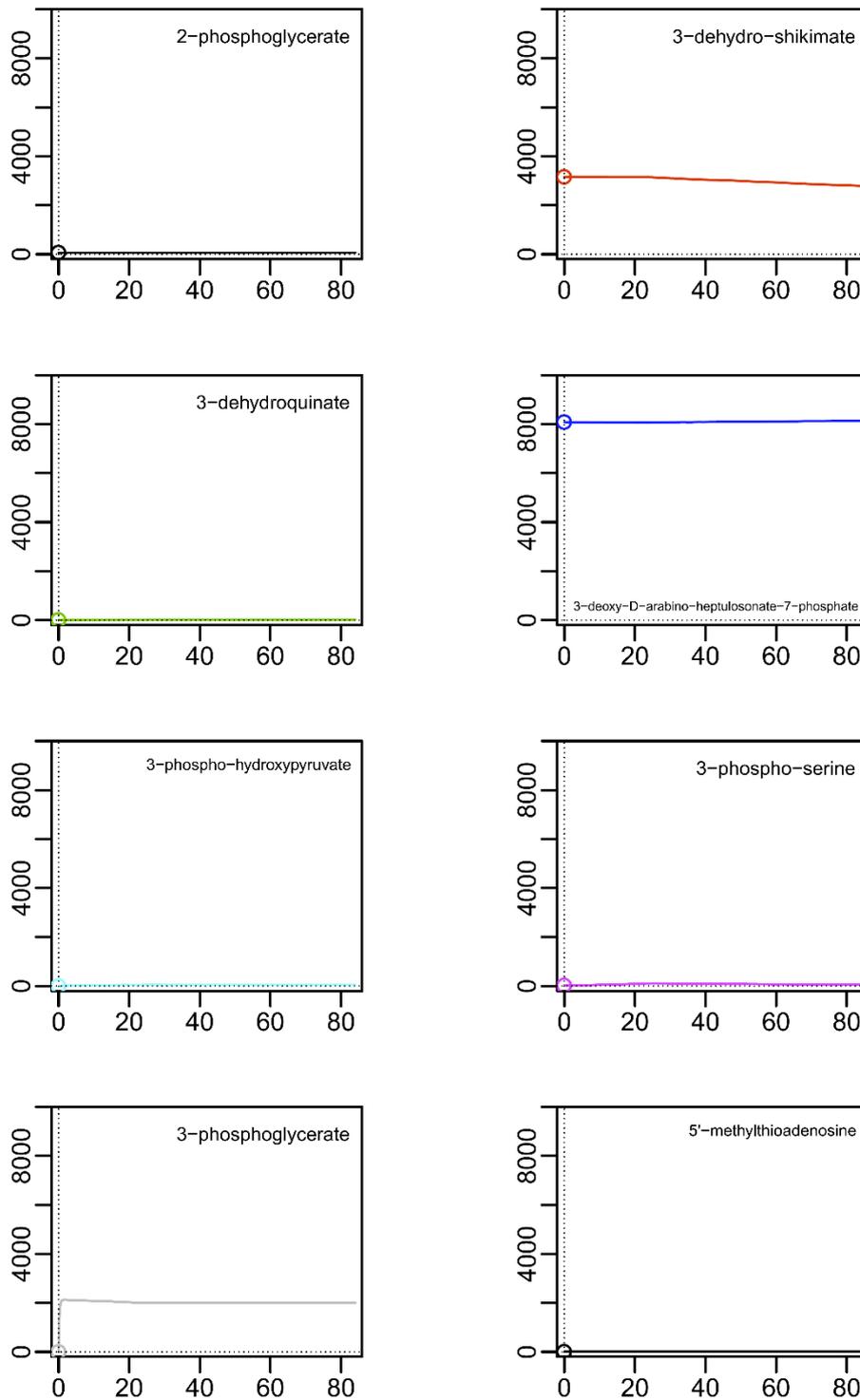
- Tomita, M. et al., 1999. E-CELL: software environment for whole-cell simulation. *Bioinformatics*, 15(1), pp.72–84.
- Varner, J.D., 2000. Large-scale prediction of phenotype: concept. *Biotechnol. Bioeng.*, 69(6), pp.664–78.
- Whitley, D., Rana, S. & Heckendorn, R.B., 1998. The island model genetic algorithm: on separability, population size and convergence. , pp.1–17.
- Wu, C.-H. et al., 2007. A real-valued genetic algorithm to optimize the parameters of support vector machine for predicting bankruptcy. *Expert Syst. Appl.*, 32(2), pp.397–408.
- Yen, J. et al., 1995. Approach to modeling metabolic systems algorithm and simplex method.

付録

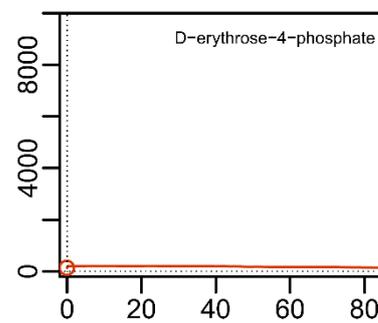
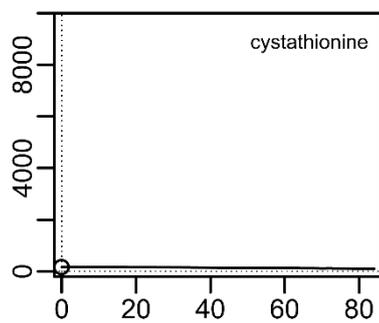
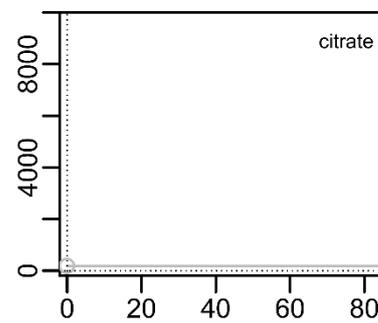
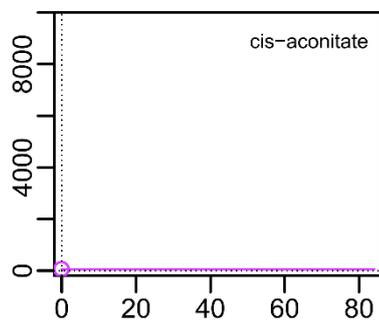
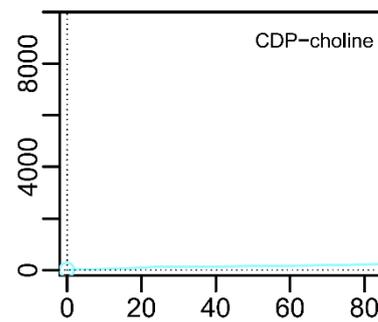
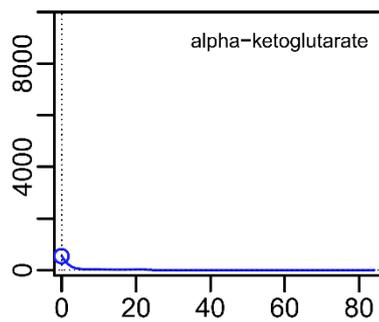
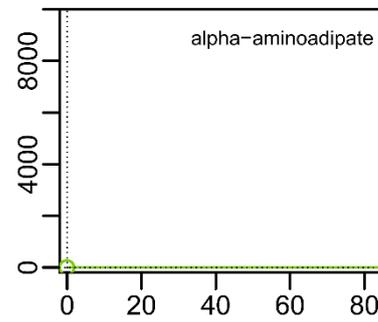
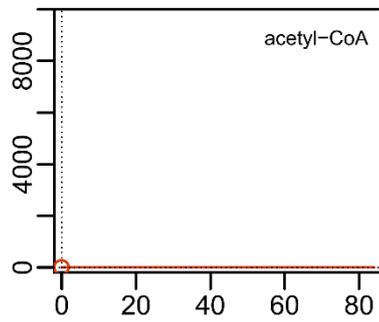
実験の詳細

本研究の共同研究先である、理研で行われた実験結果 (Sawada et al. 2009) を用いた。Murota ら (2011) に記述された方法を修正して用意したアクセシオン Col-0 から得られたシロイヌナズナの液体カルス培地を用いた。カルスの誘導として、RM28 培地内で切り刻んだ苗木を一定の光量下で育てた。6 日ごとに培地を交換し、3 度目の交換後に、カルスにリジン 10mM とトレオニン 1mM を添加した場合 (実験群) と添加しなかった場合 (統制群) を作成した。リジンとトレオニンの添加時 (0 h) と、添加後 2, 6, 12, 24, 36, 48, 60, 72, 84 h にカルスを集め、液体窒素により冷凍し-80 度で使用まで保管した。それぞれのサンプルに対し、3 セットの解析を UPLC-TQMS (UPLC-タンデム四重極型質量分析装置) により行った。細胞粉碎機を用い、500 μ l, 80% の MeOH 緩衝液内で 2mg の植物細胞標本を抽出した。250 μ l の抽出液を 96 のプレートに分割して乾燥し、再度 120 μ l の H₂O に溶かした。最後に、1 μ l のサンプルを UPLC-TQMS 装置にかけた。3 回の反復実験を行った。

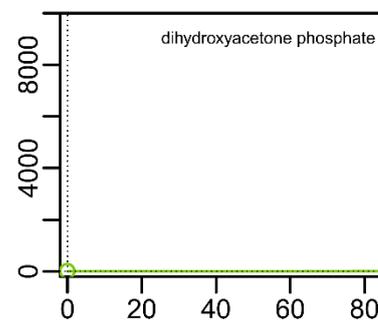
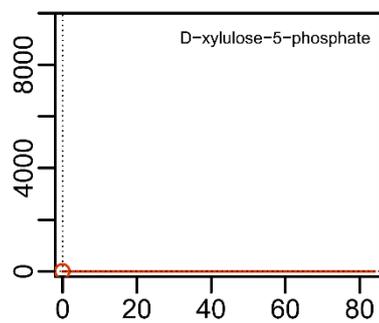
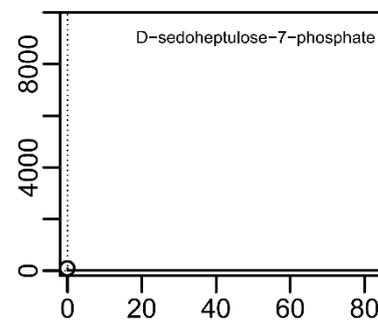
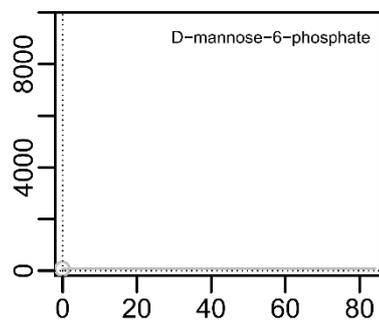
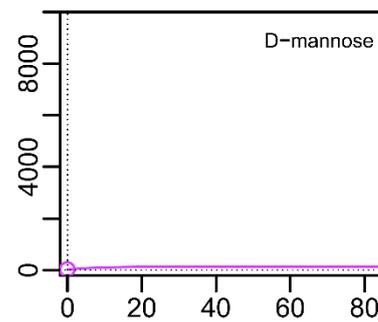
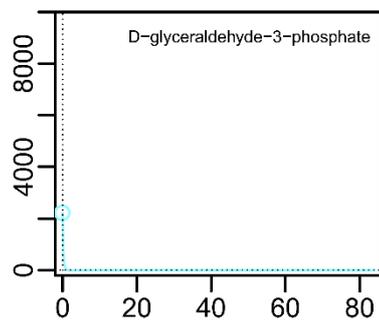
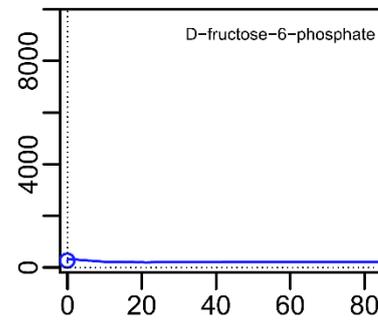
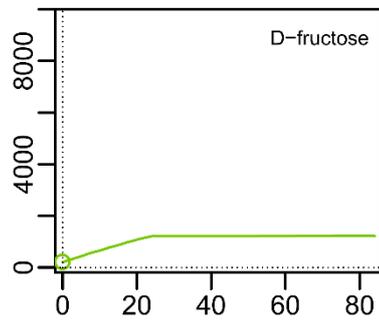
付図



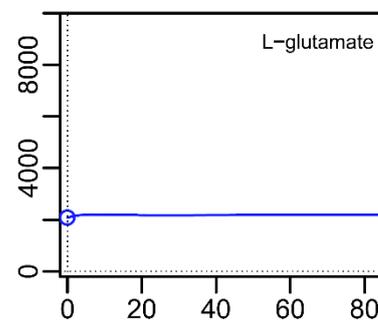
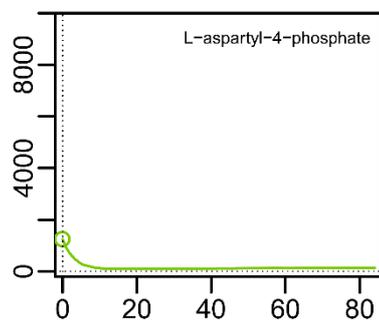
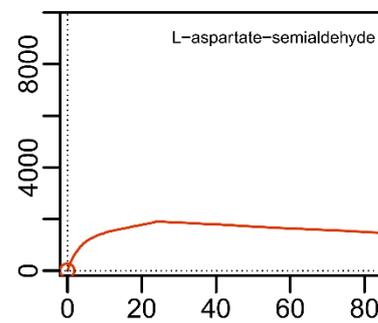
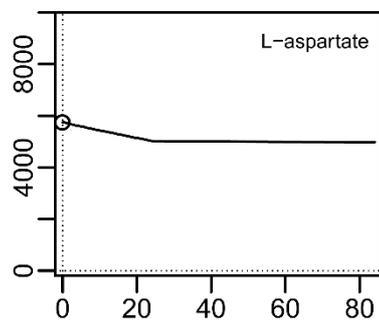
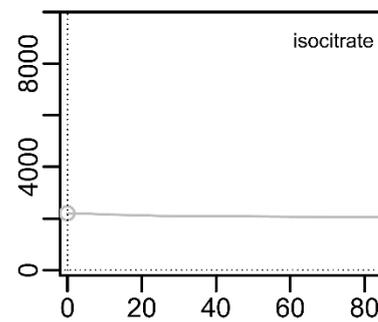
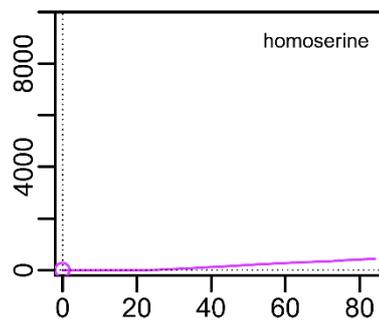
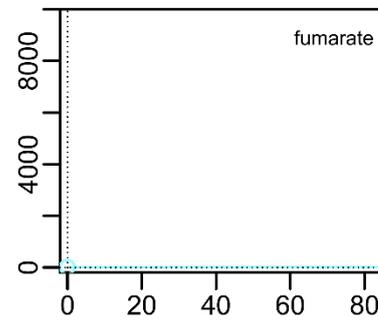
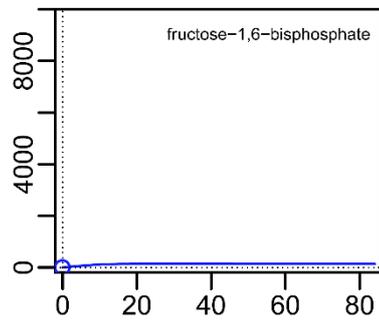
付図 1 シミュレーションによって予測された代謝物濃度の時間変化
縦軸横軸を全ての図でそろえ、代謝物濃度の時間変化を示す。



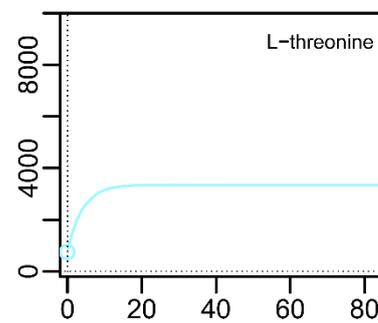
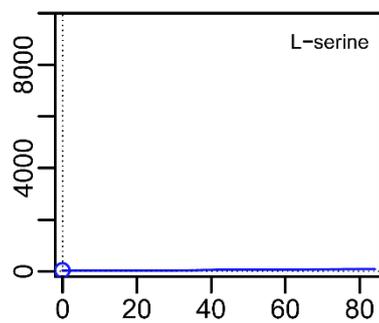
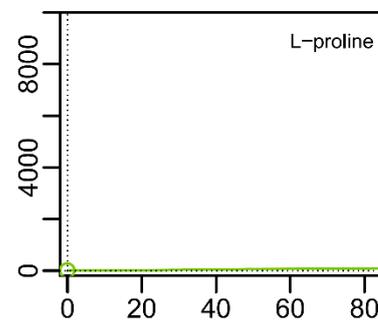
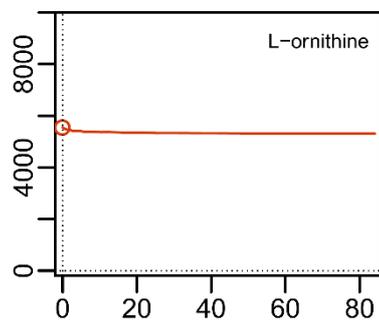
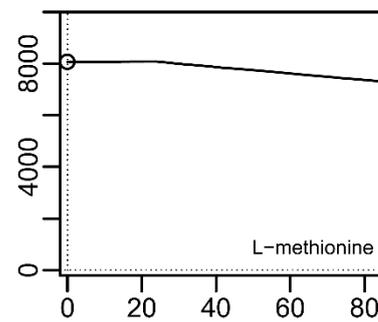
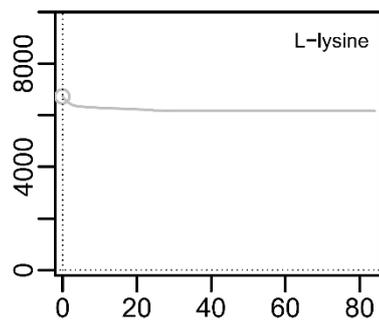
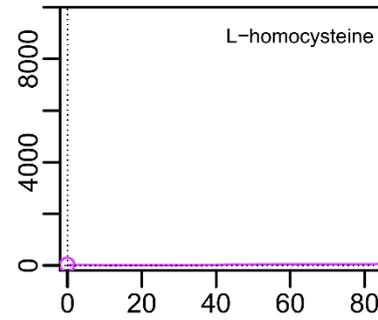
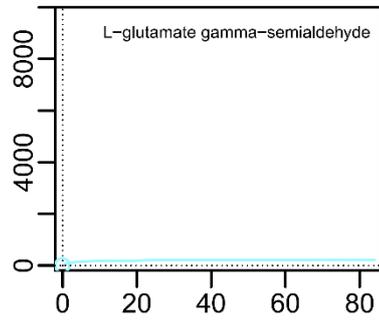
付図 1 (続き)



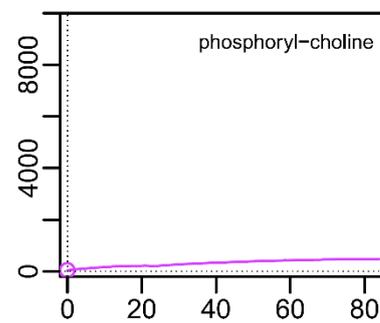
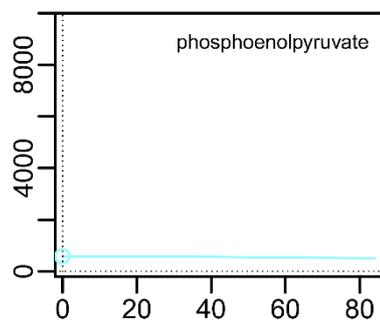
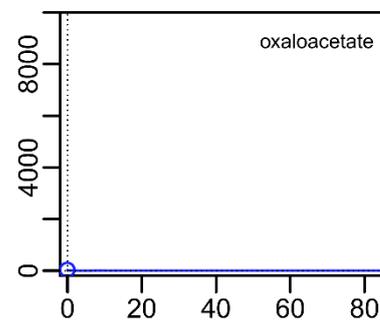
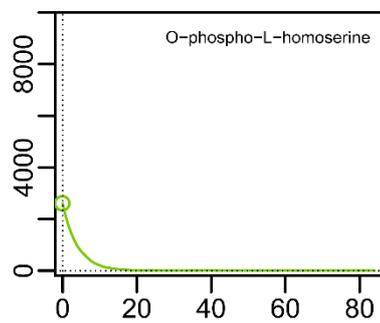
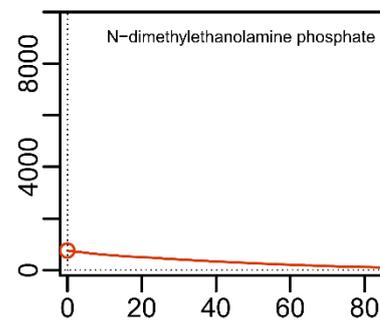
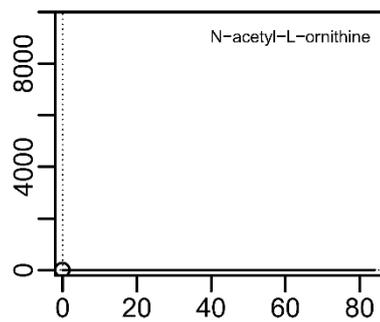
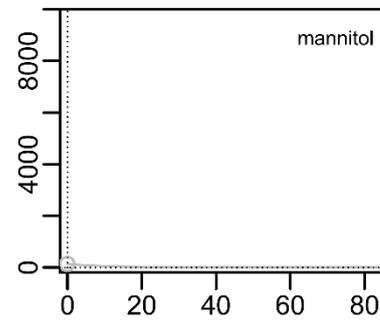
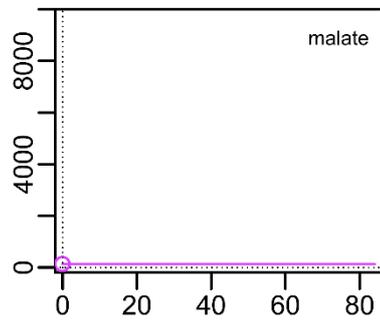
付図 1 (続き 2)



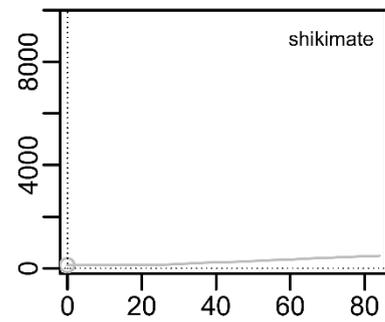
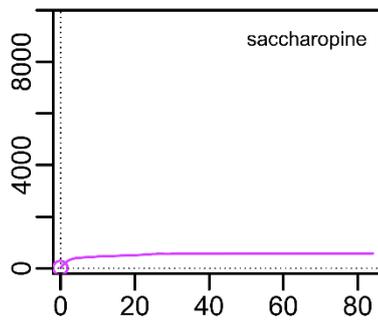
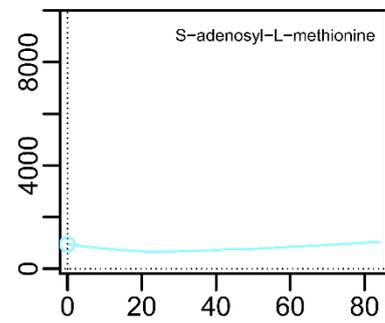
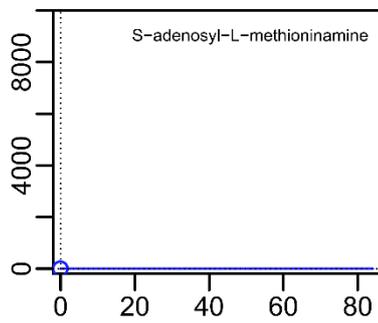
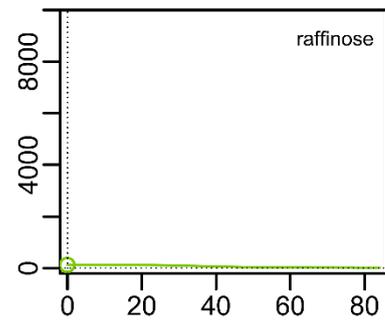
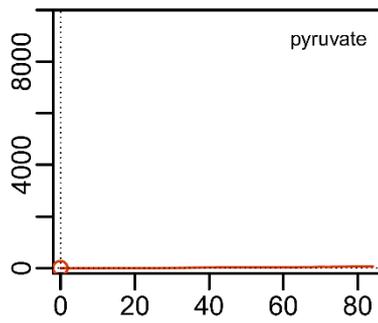
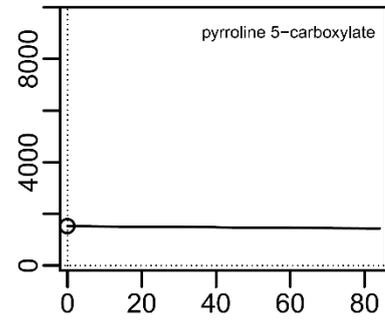
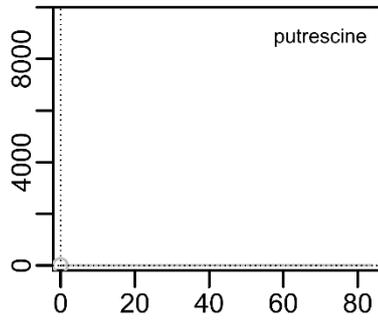
付図 1 (続き 3)



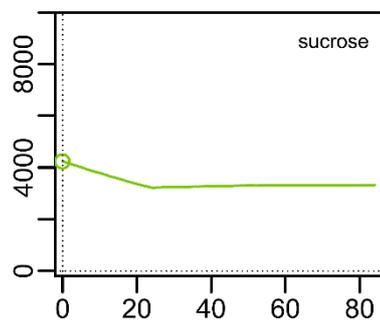
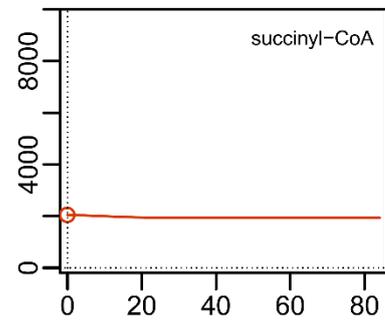
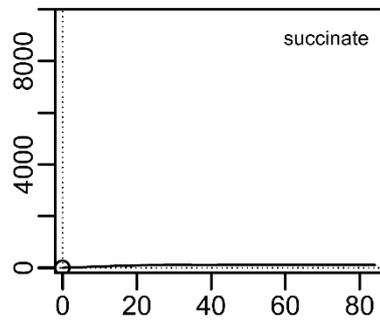
付図 1 (続き 4)



付図 1 (続き 5)



付図 1 (続き 6)



付図 1 (続き 7)

付表

付表 1 代謝実験で得られた化合物の一覧

| |
|---|
| (-)-Norepinephrine |
| (-)-Riboflavin |
| (-)-Shikimic acid |
| (+)-2'-Deoxyuridine |
| (+)-Jasmonic acid |
| (2R)-2-Hydroxy-2-phenethylglucosinolate |
| 1,1-Dimethylbiguanide hydrochloride |
| 10-camphorsulfonic acid |
| 11i_D-(-)-Fructose_D-(+)-Galactose_D-(+)-Mannose,from wood_D-(+)-Glucose_D-(+)-Sorbitol_D-(+)-Maltodextrin_D-(+)-altrose_D-(+)-Allose_D-Psicose,approx_D-(-)-Tagatose |
| 1-Amino-1-cyclopentanecarboxylic acid |
| 1-Aminocyclopropane-1-carboxylic acid |
| 1-Isothiocyano-6-(methylsulfinyl)-hexane |
| 1-Isothiocyano-6-(methylsulfonyl)-hexane |
| 1-Methoxyindole-glucosinolate |
| 1-Methylguanidine hydrochloride |
| 1-Methylhistamine dihydrochloride |
| 1-O-beta-D-glucopyranosyl sinapate |
| 2-Aminoethylphosphonic acid |
| 2'-Deoxyadenosine monohydrate |
| 2'-Deoxyadenosine-5'-monophosphate |
| 2'-Deoxycytidine |
| 2'-Deoxyguanosine monohydrate |
| 2-Deoxyribose-5-phosphate sodium salt |
| 2i_1,3-Diaminopropane dihydrochloride_1,3-Propanediamine |
| 2i_Adipic acid_2-Methylglutaric Acid |
| 2i_alpha-D-Galactose-1-phosphate dipotassium salt pentahydrate_alpha-D-glucose-1-phosphate dipotassium salt dihydrate |
| 2i_alpha-Ketoglutaric acid disodium salt_2-Oxoglutaric Acid |
| 2i_alpha-Lipoamide_DL-Thioctamide |
| 2i_cyanidin-3,5-di-O-glucoside chloride_Cyanidin-3-O-(2"-O-beta-glucopyranosyl-beta-glucopyranoside) |

2i_D-(-)-Ribose_D-Xylulose
2i_D-(+)-Cellobiose_Lactulose
2i_D-(+)-Melezitose monohydrate_1-Kestose
2i_D-Mannose-6-phosphate barium salt hydrate_D-Mannose 6-phosphate mono sodium salt
2i_eriodictyol-7-O-glucoside_Marein
2i_Ideain chloride_cyanidin-3-glucoside chloride
2i_Indole-3-carboxyaldehyde_Indole-3-aldehyde
2i_Kaempferol-3-Glucoside-2-Rhamnoside,-7-Rhamnoside_Kaempferol-3-O-alpha-L-rhamnopyranosyl(1-2)-beta-D-glucopyranoside-7-O-alpha-L-rhamnopyranoside
2i_L-Alanine_Sarcosine
2i_L-Norleucine_D-Alloisoleucine
2i_L-Norvaline_L-Valine
2i_Maritimein_luteolin-7-O-glucoside
2i_Nicotinic Acid_Isonicotinic acid
2i_Quercitrin_Quercetin-7-O-rhamnoside
2i_Rutin_Quercetin-3-O-b-glucopyranosyl-7-O-a-rhamnopyranoside
2i_Sarsapogenine_Smilagenin
2i_Theophylline,anhydrous_1,7-Dimethylxanthine
2i_Uridine-5'-diphospho-N-acetylgalactosamine disodium salt_Uridine-5'-diphospho-N-acetylglucosamine sodium salt
2-Isopropylmalic acid
3,5-Dimethoxycinnamic acid (predominantly trans)
3-Benzoyloxy-n-propyl-glucosinolate
3-Chloro-L-tyrosine hydrochloride
3-cyanopyridine
3'-Dephosphocoenzyme A
3-hydroxy-3-methylbutanoic acid
3-Hydroxy-3-methylglutarate
3i_Ethylmalonic acid_Glutaric acid_Methylsuccinic acid
3i_Glycolaldehyde dimer,mixture of stereoisomers_D-Erythrose_D(-)-Threose
3i_Isorhamnetin-3-Glucoside-6-Rhamnoside_Isorhamnetin-3-O-rutinoside_Isorhamnetin-3-Galactoside-6-Rhamnoside
3i_L-2-Aminobutyric acid_N,N-Dimethylglycine hydrochloride_N-Methyl-DL-Alanine
3i_L-Iditol_D-Sorbitol_D(-)-Mannitol
3i_L-Threonine_-Methyl-DL-serine_L-Homoserine
3i_Sodium pantothenate_D-Pantothenic acid hemicalcium salt_Calcium (+)-pantothenate

3i_Uridine-5'-diphosphogalactose disodium salt_UDP-glucose Disodium Salt_UDP-Galactose
disodium salt
3-Methyl-L-histidine
3-Methylsulfinyl-n-propyl-glucosinolate
3-Methylthio-n-propyl-glucosinolate
3-ureidopropionic acid
4-(Methylsulfinyl)but-3-enylglucosinolate
4-Hydroxybenzoate
4i_D-Fructose-6-phosphate disodium salt hydrate
4i_Kaempferol-3-Rhamnoside-7-Rhamnoside_vitexin-2"-O-rhamnoside_Kaempferol-3,7-O-bis-
alpha-L-rhamnoside_Kaempferol-3,7-O-di-rhamnopyranoside
4i_L(+)-ArginineJHCl_Nalpha-Acetyl-L-ornithine_L-Citrulline
4-Methoxyindole-glucosinolate
4-Methyl-5-thiazoleethanol
4-Methylsulfinyl-n-butyl-glucosinolate
4-Methylthio-n-butyl-glucosinolate
4-Nitrophenol
4-Pyridoxate
5,6-Dihydrouracil
5-Aminovaleric acid
5'-Deoxy-5'-Methylthioadenosine
5-Hydroxyindole-3-acetate (5-HIAA)
5-Methylcytosine hydrochloride
5-Methylsulfinyl-n-pentyl-glucosinolate
5-Methylthio-n-pentyl-glucosinolate
6-Aminohexanoic acid
6i_Melibiose hydrate_D-(+)-Turanose_Isomaltose_Gentiobiose_MelibiosePalatinose
Monohydrate
6-Methylsulfinyl-n-hexyl-glucosinolate
6-Methylthio-n-hexyl-glucosinolate
7i_kaempferol-3-O-rutinoside
7-Methylsulfinyl-n-heptyl-glucosinolate
7-Methylthio-n-heptyl-glucosinolate
8-Methylsulfinyl-n-octyl-glucosinolate
8-Methylthio-n-octyl-glucosinolate
Acetaminophen

Acetylcholine chloride
Adenine
Adenosine
Adenosine 3'-monophosphate From Yeast
Adenosine-5'-monophosphate sodium salt
Adenosine-5'-phosphosulfate sodium salt
Agmatine sulfate salt
Allantoic acid
Allantoin
alpha-Lactose monohydrate
alpha-Methyl-DL-histidine dihydrochloride
alpha-Tocotrienol
Amantadine hydrochloride
Benzylglucosinolate
Betaine
beta-Nicotinamide adenine dinucleotide hydrate
beta-Nicotinamide mononucleotide
Callistephin Chloride
Carbamoyl-DL-aspartic acid
Choline chloride
cis or trans-Cinnamyl alcohol
cis-Aconitic Acid
Citramalic acid
Citric acid, Anhydrous
Creatinine, anhydrous
Cyanidin-3-O-(2-O-beta-xylopyranosyl-beta-glucopyranoside)
Cyanidin-3-O-(2-O-beta-xylopyranosyl-beta-glucopyranoside)-5-O-beta-glucopyranoside
cyanidin-3-O-rhamnoside chloride
Cystathionine
cysteinylglycine
Cytidine, cell culture tested
Cytidine-3'-monophosphate
Cytidine-5'-diphosphocholine sodium salt dihydrate
Cytidine-5'-monophosphate disodium salt
D-(-)-Quinic acid
D-(+)-Raffinose pentahydrate

D-Ala-D-ala
Daphnetin
Delphinidin-3-O-(6-O-alpha-rhamnopyranosyl-beta-glucopyranoside)
D-Erythrose-4-phosphate sodium salt
D-Glucoheptose
Diethanolamine
DL-2,3-Diaminopropionic acid monohydrochloride
DL-2-Aminoadipic Acid
DL-beta-Hydroxybutyric acid
DL-Cysteine
DL-Glyceric Acid calcium Salt
DL-homocysteine
DL-homocystine
DL-Isocitric acid trisodium salt
DL-Malic acid
DL-Pipecolinic acid
DL-threo-beta-Methylaspartic acid
D-Panose
D-Ribose-5-phosphate disodium salt hydrate
dUDP
Fumaric acid
gamma-Amino-n-butyric acid
Gluconasturtiin
Glutathione (oxidized form)
Glycine
Glycoeyamine
Glycyl-L-proline
Guanine
Guanosine
Guanosine-3',5'-cyclic monophosphate
Guanosine-5'-diphospho-beta-L-fucose
Guanosine-5'-diphosphoglucose sodium salt
Guanosine-5'-monophosphate Jdisodium salt hydrate
Hesperetin
Hesperidin
Hypoxanthine

Indol-3-ylmethyl-glucosinolate
Inosine
Inosine-5'-monophosphate
Isoguvacine hydrochloride
isorhamnetin-3-O-glucoside
Itaconic acid
Kaempferol-3-Galactoside-6-Rhamnoside-3-Rhamnoside
Kaempferol-7-O-alpha-L-rhamnoside
Keracyanin Chloride
L-(-)-Phenylalanine
L-allo-threonine
L-Anserine nitrate salt
L-Asparagine
L-Aspartic acid
L-beta-homoglutamine-HCl
L-beta-homoleucine-HCl
L-beta-homolysine-2HCl
L-beta-homomethionine-HCl
L-beta-homoproline-HCl
L-beta-homothreonine
L-Carnosine
L-Cystine
L-Glutamic acid
L-Glutamine
L-Glutathione (reduced form)
L-Histidine
Lidocain
Lignoceric Acid
L-Isoleucine
L-Kynurenine
L-Leucine, (Cell Culture Reagent, Crystalline)
L-Lysine
L-Methionine
L-Ornithine monohydrochloride
L-Proline
L-Pyroglutamic acid

L-saccharopine
L-Serine
L-Threonic acid hemicalcium salt
L-Tryptophane
L-Tyrosine
luteolin-4'-O-glucoside
Maleic acid
Maltitol
Methionine sulfoxide
Methylmalonic acid
m-Hydroxycinnamic acid
mucic acid
N-6-(delta-2-Isopentenyl)adenosinehemihydrate
N-acetyl putrescine hydrochloride
N-Acetyl-DL-aspartic acid
N-Acetyl-DL-glutamic acid
N-acetyl-DL-serine
N-acetyl-D-mannosamine
N-Acetylglycine
N-acetylneuraminic acid, Type IV-S, Synthetic
naringenin-7-O-glucoside
Neoeriocitrin
Neohesperidin
Nicotinic acid mono nucleotide
Nystose Trihydrate
O-Acetyl-L-homoserine hydrochloride
O-acetyl-L-serine hydrochloride
o-Anisic Acid
O-Phosphocholine
o-Phospho-L-serine
O-Succinyl-L-homoserine
Pimelic acid
Puerarin
pyridoxal hydrochloride
pyridoxal-5'-phosphate hydrate
Pyridoxamine dihydrochloride

Pyridoxine
Quercetin-3,7-O-alpha-L-dirhamnopyranoside
Quercetin-3-O-alpha-L-rhamnopyranosyl(1-2)-beta-D-glucopyranoside-7-O-alpha-L-rhamnopyranoside
Robinin
S-(5'-Adenosyl)-L-methionine chloride
S-Adenosyl-L-homocysteine
Safranine
Salicylic Acid
Sinapic acid
Sinapine
Sinapoyl malate
S-Methyl-L-cysteine
S-Methylmethionine
sn-Glycero-3-phosphocholine 1:1 cadmium chloride adduct
Sodium phenylpyruvate
Suberic acid
Sucrose
Thiamine hydrochloride
Thioglycolic acid solution
Thymidine
trans-Zeatin-riboside
Trimethylamine N-oxide dihydrate
UDP-beta-L-rhamnose
UDP-xylose
Uridine
Uridine-5'-monophosphate
Vanillin
Xanthine
Zeatin-9-glucoside

付表 2 SS-mPMG により全反応リストから抽出された反応の一覧

| 反応名 | 反応式 |
|--------------------------|---|
| 2.7.1.40 | [pyruvate] + ATP < ADP + [phosphoenolpyruvate] + 2 H+ |
| 1.2.1.--004 | [pyruvate] + NAD+ + coenzyme A > [acetyl-CoA] + CO2 + NADH |
| 1.3.5.1 | [succinate] + a ubiquinone > a ubiquinol + [fumarate] |
| 1.4.1.4 | NAD+ + [L-glutamate] + H2O < ammonia + NADH + [alpha-ketoglutarate] + 2 H+ |
| 5.3.1.1 | [D-glyceraldehyde-3-phosphate] = [dihydroxyacetone phosphate] |
| 4.1.2.13 | [fructose-1,6-bisphosphate] = [dihydroxyacetone phosphate] + [D-glyceraldehyde-3-phosphate] |
| 3.1.3.11 | [fructose-1,6-bisphosphate] + H2O > [D-fructose-6-phosphate] + phosphate |
| 2.7.1.11 | ATP + [D-fructose-6-phosphate] > ADP + [fructose-1,6-bisphosphate] + 2 H+ |
| 2.6.1.1 | [oxaloacetate] + [L-glutamate] > [L-aspartate] + [alpha-ketoglutarate] |
| 4.2.3.1 | [O-phospho-L-homoserine] + H2O > phosphate + [L-threonine] |
| 6.2.1.5 | [succinate] + ATP + coenzyme A < [succinyl-CoA] + ADP + phosphate + H+ |
| 4.2.1.2 | [malate] < [fumarate] + H2O |
| 2OXOGLUTARATEDE H-RXN | [alpha-ketoglutarate] + NAD+ + coenzyme A > [succinyl-CoA] + CO2 + NADH |
| 1.1.1.41 | NAD+ + [isocitrate] > NADH + CO2 + [alpha-ketoglutarate] |
| 3.2.1.26 | [sucrose] + H2O > [D-fructose] + (_alpha-D-glucose_) |
| 2.7.1.4 | [D-fructose] + ATP > [D-fructose-6-phosphate] + ADP + 2 H+ |
| 2.5.1.16 | [putrescine] + [S-adenosyl-L-methioninamine] > (_spermidine_) + [5'-methylthioadenosine] + H+ |
| 3.1.3.3 | [3-phospho-serine] + H2O > [L-serine] + phosphate |

| | |
|-------------|---|
| 2.6.1.52 | $[3\text{-phospho-hydroxypyruvate}] + [\text{L-glutamate}] > [3\text{-phospho-serine}] + [\alpha\text{-ketoglutarate}]$ |
| 1.1.1.95 | $[3\text{-phosphoglycerate}] + \text{NAD}^+ = [3\text{-phospho-hydroxypyruvate}] + \text{NADH} + \text{H}^+$ |
| 2.5.1.6 | $\text{ATP} + [\text{L-methionine}] + \text{H}_2\text{O} > \text{phosphate} + \text{diphosphate} + [\text{S-adenosyl-L-methionine}] + \text{H}^+$ |
| 4.1.1.17 | $[\text{L-ornithine}] + \text{H}^+ > \text{CO}_2 + [\text{putrescine}]$ |
| 2.2.1.2 | $[\text{D-glyceraldehyde-3-phosphate}] + [\text{D-sedoheptulose-7-phosphate}] = [\text{D-fructose-6-phosphate}] + [\text{D-erythrose-4-phosphate}]$ |
| 2.2.1.1 | $[\text{D-erythrose-4-phosphate}] + [\text{D-xylulose-5-phosphate}] = [\text{D-fructose-6-phosphate}] + [\text{D-glyceraldehyde-3-phosphate}]$ |
| 4.4.1.8 | $[\text{cystathionine}] + \text{H}_2\text{O} > \text{ammonia} + [\text{pyruvate}] + [\text{L-homocysteine}] + \text{H}^+$ |
| 2.7.1.39 | $[\text{homoserine}] + \text{ATP} > [\text{O-phospho-L-homoserine}] + \text{ADP} + 2 \text{H}^+$ |
| 2.1.1.10 | $[\text{L-homocysteine}] + [\text{S-adenosyl-L-methionine}] > _(_ \text{S-adenosyl-L-homocysteine} _)_ + [\text{L-methionine}] + \text{H}^+$ |
| 5.3.1.8 | $[\text{D-mannose-6-phosphate}] = [\text{D-fructose-6-phosphate}]$ |
| 2.7.1.7 | $[\text{D-mannose}] + \text{ATP} > [\text{D-mannose-6-phosphate}] + \text{ADP} + 2 \text{H}^+$ |
| 1.1.1.255 | $\text{NAD}^+ + [\text{mannitol}] > [\text{D-mannose}] + \text{NADH} + \text{H}^+$ |
| 2.6.1.39 | $[\alpha\text{-ketoglutarate}] + [\alpha\text{-aminoadipate}] > [\text{L-glutamate}] + _(_ \alpha\text{-keto adipate} _)_$ |
| 1.5.1.8 | $\text{NADP}^+ + [\text{saccharopine}] + \text{H}_2\text{O} < \text{NADPH} + [\alpha\text{-ketoglutarate}] + [\text{L-lysine}] + 5 \text{H}^+$ |
| 2.7.2.4 | $[\text{L-aspartate}] + \text{ATP} > [\text{L-aspartyl-4-phosphate}] + \text{ADP} + \text{H}^+$ |
| 1.2.1.11 | $\text{NADP}^+ + \text{phosphate} + [\text{L-aspartate-semialdehyde}] < \text{NADPH} + [\text{L-aspartyl-4-phosphate}] + \text{H}^+$ |
| 1.1.1.3 | $[\text{L-aspartate-semialdehyde}] + \text{NAD(P)H} + \text{H}^+ > [\text{homoserine}] + \text{NAD(P)}^+$ |
| 4.2.1.22 | $[\text{L-homocysteine}] + [\text{L-serine}] > [\text{cystathionine}] + \text{H}_2\text{O}$ |
| 4.2.1.3-001 | $[\text{citrate}] > [\text{cis-aconitate}] + \text{H}_2\text{O}$ |
| 4.2.1.3 | $[\text{cis-aconitate}] + \text{H}_2\text{O} > [\text{isocitrate}]$ |
| 2.3.3.1 | $[\text{oxaloacetate}] + [\text{acetyl-CoA}] + \text{H}_2\text{O} > [\text{citrate}] + \text{coenzyme A} + \text{H}^+$ |

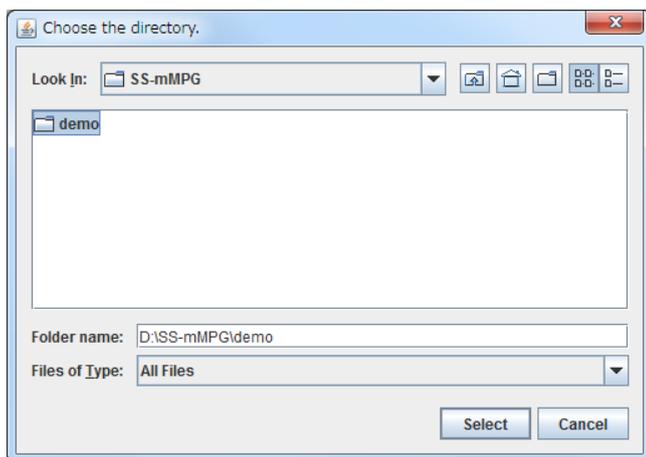
| | |
|--------------|--|
| 1.1.1.37 | $[\text{malate}] + \text{NAD}^+ > [\text{oxaloacetate}] + \text{NADH} + \text{H}^+$ |
| 5.4.2.1 | $[\text{3-phosphoglycerate}] + \text{H}^+ = [\text{2-phosphoglycerate}]$ |
| 4.2.1.11 | $[\text{2-phosphoglycerate}] = [\text{phosphoenolpyruvate}] + \text{H}_2\text{O} + \text{H}^+$ |
| 1.2.1.9 | $\text{NADP}^+ + [\text{D-glyceraldehyde-3-phosphate}] + \text{H}_2\text{O} > [\text{3-phosphoglycerate}] + \text{NADPH} + 2 \text{H}^+$ |
| 4.1.1.31 | $[\text{phosphoenolpyruvate}] + \text{CO}_2 + \text{H}_2\text{O} > \text{phosphate} + [\text{oxaloacetate}] + \text{H}^+$ |
| 4.2.3.4 | $[\text{3-deoxy-D-arabino-heptulosonate-7-phosphate}] > \text{phosphate} + [\text{3-dehydroquinate}]$ |
| 4.2.1.10 | $[\text{3-dehydroquinate}] > [\text{3-dehydro-shikimate}] + \text{H}_2\text{O}$ |
| 2.5.1.54 | $[\text{phosphoenolpyruvate}] + [\text{D-erythrose-4-phosphate}] + \text{H}_2\text{O} > [\text{3-deoxy-D-arabino-heptulosonate-7-phosphate}] + \text{phosphate}$ |
| 1.1.1.25 | $[\text{3-dehydro-shikimate}] + \text{NADPH} + \text{H}^+ > \text{NADP}^+ + [\text{shikimate}]$ |
| 2.7.7.15 | $[\text{phosphoryl-choline}] + \text{CTP} + \text{H}^+ > [\text{CDP-choline}] + \text{diphosphate}$ |
| 2.1.1.103 | $[\text{N-dimethylethanolamine phosphate}] + [\text{S-adenosyl-L-methionine}] > [\text{phosphoryl-choline}] + \text{[_S-adenosyl-L-homocysteine_]} + \text{H}^+$ |
| 2.6.1.1 | $[\text{oxaloacetate}] + [\text{L-glutamate}] > [\text{L-aspartate}] + [\text{alpha-ketoglutarate}]$ |
| SPONTPRO-RXN | $[\text{L-glutamate gamma-semialdehyde}] > [\text{pyrroline 5-carboxylate}] + \text{H}_2\text{O} + \text{H}^+$ |
| 2.6.1.13 | $[\text{L-ornithine}] + [\text{alpha-ketoglutarate}] = [\text{L-glutamate}] + [\text{L-glutamate gamma-semialdehyde}]$ |
| 1.5.1.2 | $[\text{L-proline}] + \text{NAD(P)}^+ < \text{NAD(P)H} + [\text{pyrroline 5-carboxylate}] + 2 \text{H}^+$ |
| 2.3.1.35 | $[\text{L-glutamate}] + [\text{N-acetyl-L-ornithine}] > \text{[_N-acetyl-L-glutamate_]} + [\text{L-ornithine}]$ |
| 2.4.1.--003 | $2 [\text{raffinose}] > \text{[_stachyose_]} + [\text{sucrose}]$ |

付表 3 選ばれた代謝ネットワークに含まれる代謝物のリスト

| | |
|---|----------------------------------|
| 2-phosphoglycerate | L-aspartyl-4-phosphate |
| 3-dehydroquininate | L-glutamate gamma-semialdehyde |
| 3-dehydro-shikimate | L-glutamate |
| 3-deoxy-D-arabino-heptulosonate-7-phosphate | L-homocysteine |
| 3-phosphoglycerate | L-lysine |
| 3-phospho-hydroxypyruvate | L-methionine |
| 3-phospho-serine | L-ornithine |
| 5'-methylthioadenosine | L-proline |
| acetyl-CoA | L-serine |
| alpha-aminoadipate | L-threonine |
| alpha-ketoglutarate | malate |
| CDP-choline | mannitol |
| cis-aconitate | N-acetyl-L-ornithine |
| citrate | N-dimethylethanolamine phosphate |
| cystathionine | O-phospho-L-homoserine |
| D-erythrose-4-phosphate | oxaloacetate |
| D-fructose | phosphoenolpyruvate |
| D-fructose-6-phosphate | phosphoryl-choline |
| D-glyceraldehyde-3-phosphate | putrescine |
| dihydroxyacetone phosphate | pyrroline 5-carboxylate |
| D-mannose | pyruvate |
| D-mannose-6-phosphate | raffinose |
| D-sedoheptulose-7-phosphate | saccharopine |
| D-xylulose-5-phosphate | S-adenosyl-L-methioninamine |
| fructose-1,6-bisphosphate | S-adenosyl-L-methionine |
| fumarate | shikimate |
| homoserine | succinate |
| isocitrate | succinyl-CoA |
| L-aspartate | sucrose |
| L-aspartate-semialdehyde | |

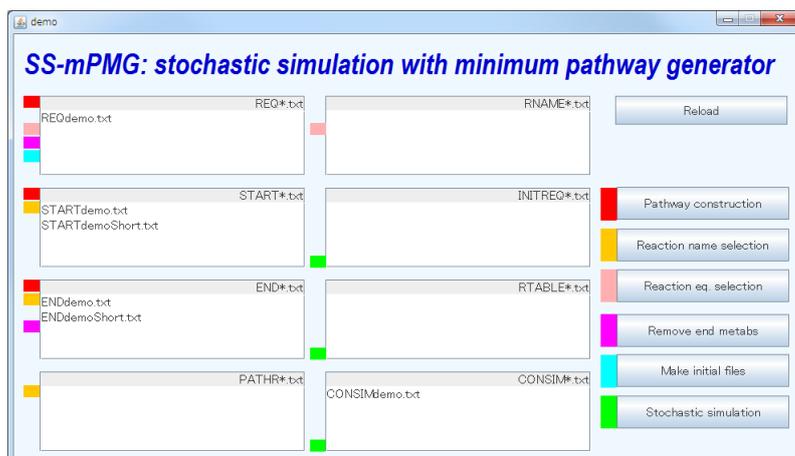
SS-mPMG の詳細

SS-mPMG の使い方の手順を、例を示しながら説明する。例では、いくつかの始端化合物と終端化合物をつなぐ反応により構成されるネットワークを構築し、確率的シミュレーションを行う。使用条件として、j2sdk-1.6.X かそれ以上がインストールされている必要がある。初めに、<http://kanaya.naist.jp/mPMG/> から、SS-mPMG.zip をダウンロードする。SS-mPMG フォルダの下には、demo フォルダがあり、REQ*.txt, START*.txt と END*.txt が含まれている。正しく設定してあれば、SS-mPMG は、SS-mPMG.jar をダブルクリックすると実行できる（画面 1）。



画面 1 実行フォルダを指定する画面

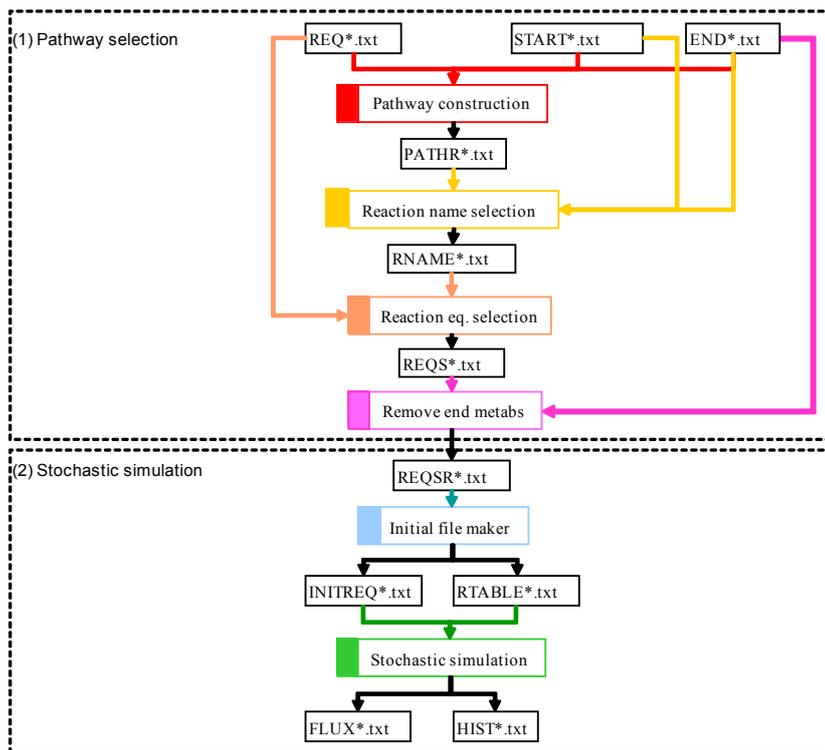
実行フォルダ（この例では demo フォルダ）を選び、select をクリックすると、SS-mPMG のメイン画面が表示される（画面 2）。実行フォルダには、インプットファイル群が含まれている必要がある。



画面 2 SS-mPMG のメイン画面

SS-mPMG は、始端化合物と終端化合物を選ぶことにより、代謝パスウェイのシミュレーションを行うことが可能である。メイン画面（付図 1）は、(1) パスウェイ選択と (2) 確率的シミュレーションからなる。

パスウェイ選択では、代謝反応のリスト内から、始端化合物から終端化合物までをつなぐ最短経路上の全反応が選ばれる。確率的シミュレーションでは、代謝の挙動を Gillespie アルゴリズム (Gillespie 1977) によって、予測することができる。付図 A は、それぞれの過程における入出力ファイルを表す。



付図 SS-mPMG のそれぞれの過程を表す模式図

シミュレーション設定ファイル (CONSIM*.txt) は、実行フォルダ（例では demo フォルダ）に入っている必要がある。ここで、"*"は、ユーザにより決められる任意の文字列である。CONSIM*.txt のフォーマットは以下のとおりである。入力ファイルはタブ区切りを想定している。すなわち、列は、一つのタブ文字で区切られている。1 列目は変数を表し、t0, t1, t2, seed はそれぞれ、計算の開始、終了、間隔、Java の Random クラスの乱数に与えるシード値を表し、対応する値がそれぞれ 2 列目に入っている。

CONSIM*.txt のフォーマット

| | |
|------|------|
| t0 | 0.0 |
| t1 | 12.0 |
| t2 | 0.1 |
| seed | 0 |

(1) パスウェイの選択

ターゲットとなる代謝パスウェイは、以下の 4 ステップで選ばれる ; (1) パス

ウェイの構築、(2)反応名の選択、(3)反応式の選択、(4)不要末端物質の除去。

(i) パスウェイ・データベースの構築

パスウェイ・データベースの構築では、始端化合物から終端化合物までの最短経路が全反応式のリストから選ばれる。始端化合物と終端化合物はそれぞれ、START*.txt and END*.txt に入れておき、反応式のリストは REQ*.txt に入れる。ファイル REQ*.txt には、対象とする生物の全ての反応式を入れておく。様々な生物の代謝反応の包括的なリストは、PlantCyc (<http://plantcyc.org>) や KEGG (<http://www.kegg.jp>) といった公開データベースから得ることができる。REQ*.txt は、例に見られるように、反応名、反応係数、反応式からなっている。ここで、計算対象とする代謝物は、カギ括弧 “[” と “] ” で囲まれており、他の化合物は無視される。代謝反応中では、矢印とイコールの記号が、正反応、逆反応、可逆反応をそれぞれ表す。ファイル START*.txt と END*.txt は、カギ括弧 “[” と “] ” で囲まれた代謝物のリストからなる。メイン画面上で、REQ*.txt, START*.txt, END*.txt を選んで、“Pathway construction”をクリックすると、START*.txt 内に記述された化合物から、END*.txt に記述された化合物までの最短系路上にある全ての反応が PATHR*.txt に得られる。SS-mPMG プログラム上では、ボタンの横のリボンの色は、その過程に必要なファイル名を選ぶボックスの上のリボンの色と一致している。PATHR*.txt には、1行目に最長パスウェイの長さ、2行目以降に始端化合物と終端化合物とともに、最短系路上の反応名が挙げられている。

REQ*.txt のフォーマット

| | | |
|-----------|-----|--|
| 4.2.3.48 | 1.0 | [trans, trans-farnesyl]diphosphate + H ₂ O > [(3S,6E)-nerolidol] + diphosphate |
| RXN-10441 | 1.0 | [geranylgeranyl]diphosphate + H ₂ O > [(E,E)-geranyl]linalool + diphosphate |
| 4.1.3.36 | 1.0 | [4-(2'-carboxyphenyl)-4-oxobutryl-CoA] + H ⁺ > [1,4-dihydroxy-2-naphthoyl-CoA] + H ₂ O |
| 4.2.1.113 | 1.0 | [(1R,6R)-6-hydroxy-2-succinylcyclohexa-2,4-diene-1-carboxylate] > [O-succinylbenzoate] + HO |
| ... | | ... |

START*.txt と END*.txt のフォーマット

| |
|--|
| [alpha-aminoadipate] [L-proline] [L-valine] [5-oxoproline] ... |
|--|

PATHR*.txt のフォーマット

| | |
|---|----------------------------------|
| 11 | |
| [5'-methylthioadenosine] to [L-glutamine] | 3.2.2.16;2.4.2.7;3.5.4.6;6.3.1.2 |
| [L-methionine] to [L-glutamine] | 4.4.1.11;6.3.1.2 |
| ... | |

(ii) 反応名の選択

“Reaction name selection”を押した後、START*.txt, END*.txt および PATHR*.txt に基づいて反応名が RNAME*.txt に保存される。ユーザは、START*.txt と END*.txt から代謝物を取り除くことで始端と終端に選ぶ化合物を選ぶことができる。RNAME*.txt には、反応名とどの始端化合物と終端化合物に関連しているかが記録される。次の反応式選択では RNAME*.txt の1カラム目のみが使われる。したがって、REQ*.txt 反応式データベース内の他の反応名をここに加えることによって、最短経路上以外の反応も加えることが可能である。

Format of RNAME*.txt

| | |
|----------|--|
| 3.2.2.16 | [5'-methylthioadenosine] to [alpha-aminoadipate] |
| 1.5.1.2 | [5'-methylthioadenosine] to [L-proline] |
| ... | |

(iii) 反応式を選択

REQ*.txt から、RNAME*.txt 内の反応名の反応式のみが抽出され、REQS*.txt に保存される。REQS*.txt のフォーマットはREQ*.txt と同一である。

(iv) 不要終端物質の削除

END*.txt に含まれる以外の終端化合物は、シミュレーションには必要ではないので、“Remove end metab”で削除することができる。削除された化合物は、REQS*.txt 内において“_(5-methylthioribose)_”のように名前が変更される。こ

これらの化合物を再びシミュレーションに含みたい場合には、“_(”と“_)_”を “[”と“]”に変更することで簡単に戻すことができる。

(2) 確率的シミュレーション

SS-mPMGにおいて、確率的シミュレーションは以下の2ステップで行われる:

(i) 初期値作成 (2) 確率的シミュレーション。

(i) 初期値作成

初期値作成プログラムにより、INITREQ*.txt と RTABLE*.txt に、それぞれの代謝物の初期濃度とそれぞれの反応の反応係数が保存される。INITREQ*.txt には、それぞれの行に代謝物名と、その初期濃度、最小濃度、最大濃度が保存され、RTABLE*.txt には、反応名と、反応係数 (c_i) と、反応に関わる代謝物の数が化合物ごとに保存される。ユーザは、代謝物の初期濃度と c_i で定義される反応係数を変更することで、簡単に様々な条件における代謝物の挙動をシミュレーションすることができる。

INITREQ*.txt のフォーマット

| | | | |
|----------------------|-------|-----|---------|
| [3-ureidopropionate] | 100.0 | 0.0 | 50000.0 |
| [ammonia] | 100.0 | 0.0 | 50000.0 |
| [L-tyrosine] | 100.0 | 0.0 | 50000.0 |
| | | | |

RTABLE*.txt のフォーマット

| Reaction Parameter | c_i | [3-ureidopropionate] | [3-ureidopropionate] | ... |
|--------------------|-------|----------------------|----------------------|---------|
| 3.5.1.6-001 | 1.0 | 1 | 0 | 0 1 ... |
| 3.5.3.6 | 1.0 | 0 | 0 | 0 1 . |
| ... | | | | |

(ii) 確率的シミュレーション

確率的シミュレーションは、以下の3つのファイルを用いて行われる: CONSIM*.txt、INITREQ*.txt、RTABLE*.txt。結果は以下の2つのファイルに保存される: HIST*.txt、FLUX*.txt。それぞれ、代謝物の濃度の時系列変化と、酵素のハザード関数の時系列変化である。代謝の挙動は、HIST*.txt 内の代謝の挙動

の時間変化をプロットすることで観察することができ、反応の安定性は、
 FLUX*.txt.内のハザード関数の時間変化をプロットすることで調べることが
 できる。

HIST*.txt のフォーマット

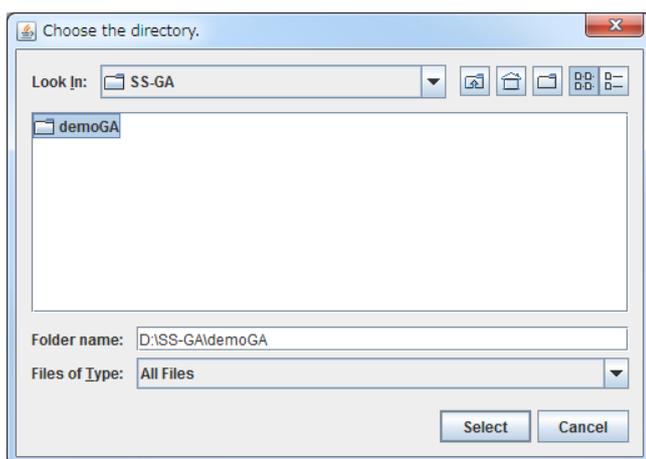
| time | [3-ureidopropionate] | [ammonia] | [L-tyrosine] | ... |
|-------------|----------------------|-----------|--------------|-----|
| 0 | 100 | 100 | 100 | 100 |
| 0.01000043 | 99 | 27 | 59 | 17 |
| 0.020042799 | 97 | 13 | 56 | 5 |
| 0.030095405 | 97 | 11 | 53 | 3 |
| 0.040175784 | 95 | 8 | 53 | 0 |
| 0.050175481 | 94 | 6 | 53 | 0. |
| ... | | | | |

FLUX*.txt のフォーマット

| time | 3.5.1.6-001 | 2.6.1.5-001 | 4.2.3.1 ... |
|-------------|-------------|-------------|-------------|
| 0 | 0 | 0 | 0 ... |
| 0.01000043 | 99 | 1003 | 98 ... |
| 0.020042799 | 97 | 280 | 98 ... |
| 0.030095405 | 97 | 159 | 97 ... |
| 0.040175784 | 95 | 0 | 96 ... |
| 0.050175481 | 94 | 0 | 95 ... |
| 0.050175481 | 94 | 0 | 95 ... |
| | | | |

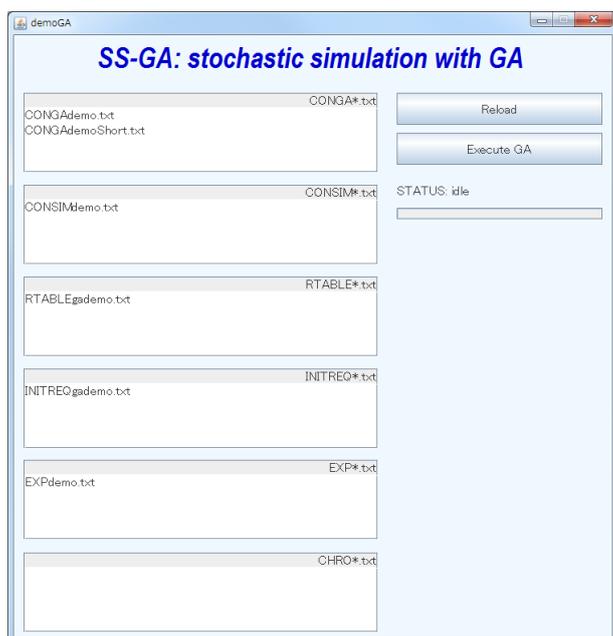
SS-GA, SS-dGA の詳細

SS-GA および SS-dGA を用いて Gillespie のアルゴリズムに基づいた確率的シミュレーションを行う方法を記述するために、実験値に合うように代謝物の初期濃度と反応の反応係数のセットを最適化する例を示す。SS-GA と SS-dGA は、入力ファイルが異なるだけなので、ここでは例として SS-GA の使い方を説明する。使用条件として、j2sdk-1.6.X かそれ以上がインストールされている必要がある。初めに、<http://kanaya.naist.jp/mPMG/>から SS-GA.zip をダウンロードする。SS-GA フォルダには、demoGA というデモ用のフォルダがある。正しく設定してあれば、SS-GA.jar をダブルクリックすると、SS-GA が立ち上がる（画面 3）。



画面 3 実行フォルダの選択画面

フォルダを選択し (demoGA)、“select” をクリックすると、SS-GA のメイン画面が表示される。(画面 4).



画面 4 SS-GA の画面

5つのファイル (CONGA*.txt, CONSIM*.txt, RTABLE*.txt, INITREQ*.txt および EXP*.txt) が SS-GA の実行に必要である。このほかに、もし必要ならば、初期パラメータとして CHRO*.txt を使うことが可能である。まず、これらのファイルについて説明する。

(1) 入力ファイルのフォーマット

CONGA*.txt のフォーマット

CONGA*.txt は、CONGA*.txt のフォーマットに示したように、以下の遺伝的アルゴリズムに関する 5つのパラメータからなる：1行目の個体数、2行目の世代数、3行目の交叉率、4行目の突然変異率、5行目の Java の Random クラスの乱数に与えるシード値。SS-dGA の場合、これらに、ステージ数、母集団数、移住率をそれぞれ表す値、era1、nisland、pmigration が加わる。

CONGA*.txt のフォーマット

| | |
|----------------|------|
| populationsize | 20 |
| generation | 30 |
| pcrossover | 0.8 |
| pmutation | 0.05 |
| seed | 0 |

CONSIM*.txt

CONSIM*.txt は、シミュレーション用のパラメータを格納する。一回分の計算を、`_-BEGINCONSIM-_*` と `_-ENDCONSIM-_*` で囲まれた部分に示す。例では、2回の計算が記述されており、1回目は、`_-BEGINCONSIM-_1` から `_-ENDCONSIM-_1` まで、2回目は、`_-BEGINCONSIM-_2` から `_-ENDCONSIM-_2` までである。時間のパラメータ `t0,t1,t2` は、それぞれ、開始、終了、間隔を表す。開始と終了の時間はEXP*.txt内の実験値のものと同じ必要がある。 `seed` は Java の `Random` クラスの乱数に与えるシード値である。

2回目の計算である `_-BEGINCONSIM-_2` から `_-ENDCONSIM-_2` には、`iscontinue` が `TRUE` の場合が含まれている。この場合、一つ前の計算で得られた濃度が初期濃度として使われ、続きの計算を行うことができる。`FALSE` の場合には使われない。この例では、`t0,t1,t2` は、それぞれ `6.0, 12.0, 0.1` となっており、これの意味するのは、時刻 `6.0` から `12.0` まで、`0.1` おきに `EXP*.txt` との比較がなされるということである。

CONSIM*.txt のフォーマット

| | |
|-------------------------------|-------------------|
| <code>_-BEGINCONSIM-_1</code> | |
| <code>t0</code> | <code>0.0</code> |
| <code>t1</code> | <code>6.0</code> |
| <code>t2</code> | <code>0.1</code> |
| <code>seed</code> | <code>0</code> |
| <code>_-ENDCONSIM-_1</code> | |
| <code>_-BEGINCONSIM-_2</code> | |
| <code>iscontinue</code> | <code>TRUE</code> |
| <code>t0</code> | <code>6.0</code> |
| <code>t1</code> | <code>12.0</code> |
| <code>t2</code> | <code>0.1</code> |
| <code>_-ENDCONSIM-_2</code> | |

RTABLE*.txt と INITREQ*.txt のフォーマット

INITREQ*.txt と RTABLE*.txt のフォーマットは、SS-mPMG のものと同じである。このため、SS-mPMG と同じファイルを用いることができる。INITREQ*.txt と RTABLE*.txt は、初期値生成プログラムによって生成された、それぞれ代謝物の初期濃度と反応係数のファイルである。INITREQ*.txt には、それぞれの行に代謝物名と、その初期濃度、最小濃度、最大濃度が保存され、RTABLE*.txt には、反応名と、反応係数 (c_i) と、反応に関わる代謝物の数が化合物ごとに保存される。ユーザは、代謝物の初期濃度と c_i で定義される反応係数を変更することで、簡単に様々な条件における代謝物の挙動をシミュレーションすることができる。

INITREQ*.txt のフォーマット

| | | | |
|----------------------|-------|-----|---------|
| [3-ureidopropionate] | 100.0 | 0.0 | 50000.0 |
| [ammonia] | 100.0 | 0.0 | 50000.0 |
| [L-tyrosine] | 100.0 | 0.0 | 50000.0 |
| | | | |

RTABLE*.txt のフォーマット

| Reaction Parameter | c_i | [3-ureidopropionate] | [3-ureidopropionate] | ... |
|--------------------|-------|----------------------|----------------------|---------|
| 3.5.1.6-001 | 1.0 | 1 | 0 | 0 1 ... |
| 3.5.3.6 | 1.0 | 0 | 0 | 0 1 . |
| ... | | | | |

EXP*.txt

EXP*.txt は実験データを表す。時間スケールが最初の行に示されている。最初の行の最初のカラムは読み飛ばされる。最初の行は、サンプリング時間が 2 つ目のカラムから最後のカラムまで入っている。EXP*.txt のフォーマットでは、0, 2, 6, 12 時間後の 4 回の計測が行われたことを示す。2 行目から最終行までには、それぞれの代謝物のそれぞれの 4 回の計測結果が示されている。2 行目から最終行までの 1 カラム目の代謝物の名前は、部分的に INITREQ*.txt 内に同一の名前で含まれている必要がある。

EXP*.txt のフォーマット

| | | | | |
|-------------------------|-------------|-------------|-------------|-------------|
| time | 0 | 2 | 6 | 12 |
| [citrate] | 2.026272438 | 1.830223712 | 2.13013707 | 2.493230809 |
| [isopentenyl adenosine] | 0.00068112 | 0.000534926 | 0.001242609 | 0.00029025 |
| ... | | | | |
| ... | | | | |

CHRO*.txt

CHRO*.txt は出力ファイルである。ユーザが一つ前の条件で最適化を行うと、CHRO*.txt のファイルを選び、計算に使うことができる。GA によるパラメータ最適化が上の CONSIM*.txt の例にあるように 2 つの計算で定義されている場合、1 セットの初期濃度と 2 セットの反応係数を設定する必要がある。代謝物名と反応名が、1 行目で与えられたものと同一の順番になるように定義され、それぞれに該当するパラメータは 2 行目に記述される。CHRO*.txt のフォーマットに示される例では、代謝物の初期濃度に相当するのは、[phosphoenolpyruvate] から [raffinose] までであり、2 つの反応係数のセットは、2.7.1.40 から 2.4.1.--003 の部分に該当する。

CHRO*.txt のフォーマット

| | | | | | | | |
|-----------------------|-----|-------------|----------------|----------|-------------|----------|----------|
| [phosphoenolpyruvate] | ... | [raffinose] | 2.7.1.40 | ... | 2.4.1.--003 | 2.7.1.40 | |
| | ... | | 2.4.1.--003 | | | | |
| 1992.13693 | ... | 43.00273128 | 0.002131599... | 2.59E-04 | 0.002131599 | ... | 2.59E-04 |

(2) SS-GA の実行

CONGA*.txt, CONSIM*.txt, RTABLE*.txt, INITREQ*.txt, EXP*.txt と、必要ならば CHRO*.txt を選び、“Execute GA”をクリックする;すると、パラメータ最適化が実行され、結果が 4 つのファイルに格納される (HIST*.txt, FLUX*.txt, CHRO*.txt, HISTFitness.txt)。HIST*.txt と FLUX*.txt は、代謝物濃度と反応の hazard 関数の時間変化である。代謝の挙動は、HIST*.txt 内の代謝の挙動の時間変化をプロットすることで観察することができ、反応の安定性は、FLUX*.txt 内のハザード関数の時間変化をプロットすることで調べることができる。

CHRO*.txt は (1) で説明した。実行の状況が “Execute GA” ボタンの下に表示される。

(3) 出力ファイル

HISTFitness.txt のフォーマット

HISTFitness.txt には、それぞれの個体 (0 から 20) の CONGA*.txt に設定された世代 (0 から 30) での適応度が保存される。適応度は $F_k = 1/R_k$ で表される。こ

ここで、 $F = \frac{M}{\sum_{j=1}^M 10^{R_j}}$, $R_j = \sqrt{\frac{\sum_{i=1}^N (\log_{10} E_{ij}/S_{ij})^2}{N}}$, であり、 s は代謝物計測の時間点であり、

X_{kis} と Y_{is} は計算と実験の i 番目の時間 s での代謝物量を表し、 N_s は総時間点数、 N は代謝物数を表す。もし $\frac{X_{kis}}{X_{kio}}, \frac{Y_{is}}{Y_{io}}$ のどちらかが 0 ならば、 10^{-3} で置き換えられる。

それぞれの世代 (0 から 30) において、個体群の中で最も良い最適値と、それぞれの個体 (1 から 20) の最適値が HISTFitness.txt に保存される。

HISTFitness.txt のフォーマット

| | | | | |
|-----|-------------|-------------|-----|-------------|
| | bestfitness | 1 | ... | 20 |
| 0 | 2.627314532 | 6.083680122 | ... | 45.41337197 |
| 1 | 2.734235272 | 2.734235272 | ... | 42.9211501 |
| 2 | 2.337839361 | 2.627684036 | ... | 53.83355098 |
| ... | ... | ... | ... | ... |
| 30 | 1.777618006 | 1.937976983 | ... | 1.915398967 |

HIST*.txt のフォーマット

| | | | | |
|-------------|----------------------|-----------|--------------|-----|
| time | [3-ureidopropionate] | [ammonia] | [L-tyrosine] | ... |
| 0 | 100 | 100 | 100 | 100 |
| 0.01000043 | 99 | 27 | 59 | 17 |
| 0.020042799 | 97 | 13 | 56 | 5 |
| 0.030095405 | 97 | 11 | 53 | 3 |
| 0.040175784 | 95 | 8 | 53 | 0 |
| 0.050175481 | 94 | 6 | 53 | 0. |
| ... | ... | ... | ... | ... |

FLUX*.txt のフォーマット

| | | | | |
|-------------|-------------|-------------|---------|-----|
| time | 3.5.1.6-001 | 2.6.1.5-001 | 4.2.3.1 | ... |
| 0 | 0 | 0 | 0 | ... |
| 0.01000043 | 99 | 1003 | 98 | ... |
| 0.020042799 | 97 | 280 | 98 | ... |
| 0.030095405 | 97 | 159 | 97 | ... |
| 0.040175784 | 95 | 0 | 96 | ... |

| | | | | |
|-------------|------|------|------|------|
| 0.050175481 | 94 | 0 | 95 | ... |
| 0.050175481 | 94 | 0 | 95 | ... |
| | | | | |

関連業績

査読付き論文

1. Tetsuo Katsuragi, Naoaki Ono, Keiichi Yasumoto, Md. Altaf-Ul-Amin, Masami Y. Hirai, Kansuporn Sriyudthsak, Yuji Sawada, Yui Yamashita, Yukako Chiba, Hitoshi Onouchi, Toru Fujiwara, Satoshi Naito, Fumihide Shiraishi and Shigehiko Kanaya. “SS-m PMG and SS-GA: Tools for Finding Pathways and Dynamic Simulation of Metabolic Networks,” *Plant Cell Physiol.*, **54**, 728-739, 2013.第 2、3 章に関連

査読付き国際会議発表

1. Tetsuo Katsuragi, Shun Ikeda, Md. Altaf-Ul-Amin, Aki Hirai-Morita, Kansuporn Sriyudthsak, Yuji Sawada, Yui Yamashita, Yukako Chiba, Hitoshi Onouchi, Toru Fujiwara, Satoshi Naito, Fumihide Shiraishi and Shigehiko Kanaya, “Stochastic simulation of metabolic network of *Arabidopsis thaliana* using experimental data,” *XV International Congress on Molecular Plant-Microbe Interactions*, PS15-582, Kyoto, Japan, July 31, 2012. 第 2 章に関連
2. Tetsuo Katsuragi, Naoaki Ono, Tetsuo Sato and Shigehiko Kanaya “Dynamic simulation of metabolic network in *Arabidopsis thaliana* using parameters estimated by a distributed genetic algorithm”, *the 10th International conference of the Metabolomics Society*, P311, Tsuruoka, Japan, June 25, 2014. 第 3 章に関連、The Best Plant Metabolomics Poster Awards 受賞

参考業績

査読付き論文

1. Tetsuo Katsuragi, Naoaki Ono, Aki Hirai-Morita, Yukiko Nakamura, Md. Altaf-Ul-Amin and Shigehiko Kanaya, “Cuisine Omics: Fundamental Structures of Zouni and Retortable Pouched Pack of Curry Unveiled by Multivariate Analysis Based on Food Ingredients,” *Foods Food Ingredients J of Jpn*, **218**, 43-60, 2013.
2. Farit Mochamad Afendi, Tetsuo Katsuragi, Akira Kato, Noritaka Nishihara, Kensuke Nakamura, Yukiko Nakamura, Ken Tanaka, Aki Hirai-Morita, Md. Altaf-Ul-Amin, Hiroki Takahashi, and Shigehiko Kanaya, “Systems Biology Approaches and Metabolomics for Understanding Japanese Traditional Kampo Medicine,” *Current Pharmacogen Personalized Med*, **10**, 111-124, 2012.
3. Md. Altaf-Ul-Amin, Tetsuo Katsuragi, Tetsuo Sato, Naoaki Ono, and Shigehiko Kanaya, “An Unsupervised Approach to Predict Functional Relations between Genes Based on Expression Data,” *BioMed Research International*, **2014**, 154594, 8 pages, 2014.
4. 池田 俊, 桂樹 哲雄, 小野 直亮, 中谷 淳至, 中村 由紀子, 森田 晶, 金谷 重彦, “オミックス・プラットフォーム：バイオ・ビッグ・データに挑む,” *生物工学会誌*, **90**, 777-781, 2012.

査読付き国際会議発表

1. Yukiko Nakamura, Farit Mochamad Afendi, Tetsuo Katsuragi, Shun Ikeda, Ken Tanaka, Aki Hirai-Morita, Md. Altaf-Ul-Amin and Shigehiko Kanaya, “KNApSAcK Family Databases connect biological activities of metabolites and plants with microorganisms,” *XV International Congress on Molecular Plant-Microbe Interactions*, PS15-579, Kyoto, Japan 30, July, 2012