

代謝物質のマスペクトルデータベース

西岡 孝明

和漢医薬学総合研究所 情報機能解析部門 教授

1. はじめに

ポストゲノム科学を特徴づける言葉の1つは"omics"であろう。これは、ゲノム解析された生物種について遺伝子発現、タンパク質、代謝物質という3つの階層ごとに、網羅的かつ定量的に化学分析することによって生命活動を理解しモデル化しようとするものである。

遺伝子発現、タンパク質と代謝物質はシグナル伝達ネットワークと代謝反応ネットワークを介して関係づけられている(図1)。3つの階層に介在している2つのネットワークのうち、これまでの生化学研究の成果が最も数多く凝縮しているのは代謝反応ネットワークである。それに対して、データは得られるものの生物学的意義の解明がほとんど進展していないのが、シグナル伝達ネットワークである。メタボローム解析への期待の1つは、代謝反応ネットワークやその産物である代

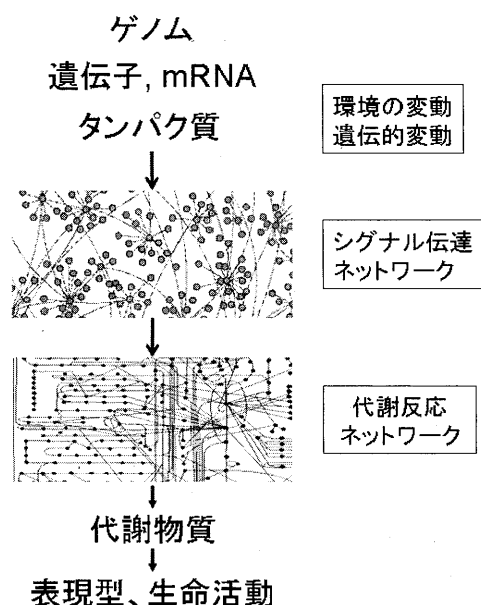


図1. Omics 解析における遺伝子、タンパク質、代謝物質の関係。

謝物質からシグナル伝達ネットワークの解釈をおこなうことであろう。

3つの階層を互いに関連づけて統合しているのが代謝反応ネットワークと代謝物質ではあるが、3つの階層のうちで最も難しいのが代謝物質の化学分析である。ある生物種のゲノム解析がおこなわれて、遺伝子がわかると、その生物がつくりだしている mRNA とタンパク質を容易に推定することができる。また、ほとんどの生物種に共通に存在している一次代謝物質については、BRENDA や KEGG などの代謝データベースを参照することによって推定することができる。しかし、生物種の生命活動を特徴づけている2次代謝物質はほとんどのものが未知であり、それらの代謝反応ネットワークもまた未知である。

メタボローム解析の目的も、二次代謝物質の同定と計量、それらの代謝反応ネットワークの構築である。この課題にとって、マスペクトルによる化学分析は不可欠な研究手段となっている。本総説では、マスペクトルによる化学分析の現状を概観し、私たちが構築しているマスペクトルデータベース MassBank について紹介する。

2. 質量分析の進歩とメタボローム解析

近年の質量分析 (MS) の進歩は著しく、短期間で開発と商品化がおこなわれ、生命科学研究への導入もまた速やかである。(注: 本文では MS を質量分析 mass spectrometry と質量分析計 mass spectrometer の2つの意味で用いている。) メタボローム解析に用いられる質量分析には、磁場型 (magnetic sector) MS をはじめ FT-ICR (Fourier transform ion cyclotron resonance) -MS や TOF (time-of-flight) -MS のようにイオンの高質量分解能測定, また MS/MS (tandem MS) や ITMS (ion trap MS) のように衝突誘起解離 (collision-induced dissociation: CID) によって生成する product イオンの測定や、precursor イオンと

product イオンの関係を分析する (MS)ⁿ スペクトルの測定、など多様である。これらの異なる質量分析を組み合わせた hybrid MS¹⁾では、フラグメントイオンの化学組成や precursor-product イオンの関係を高感度・高精度で測定することができるので、未知代謝物質の化学構造を推定するために必要な化学情報が得られるようになった。

・測定技術の進歩はメタボローム解析の進歩に反映していない

このような hybrid MS を高速・高分離能の分離法 GC や LC、CE (capillary electrophoresis) と組み合わせた GC-MS, LC-MS, CE-MS は hyphenated method と呼ばれ、メタボローム解析にも導入されている。しかし、このような化学分析の高性能化は、残念ながらメタボローム研究に革新的な進歩をもたらしているとは言い難い。

例えば、枯草菌の細胞抽出液を CE-MS で分析すると、1692 個のイオン性代謝物質を検出できる。それらのうち同定できたのは約 170 代謝物質にすぎない²⁾。この分析では、あらかじめ代謝物質を化学合成した 859 標準物質を用いて、CE の泳動時間と ESI-MS に現れる分子イオン(M+1)⁺の *m/z* 値と相対強度を測定している。試料の分析においては、泳動時間と *m/z* 値の2つのパラメータによって代謝物質を同定している。泳動時間と *m/z* 値を測定してある代謝物質のうち、約 170 しか検出できなかったことになる。残り約 690 代謝物質は

<i>m/z</i>	Composition	CAS	KEGG	DNP
180.0423	C9H8O4	588	7	51
180.0437	C10H4N	24	0	0
180.0535	C8H8N2O3	594	1	7
180.0563	C11H6N3	2	0	0
180.0575	C13H8O	237	1	12
180.0634	C6H12O6	945	41	109
180.0647	C7H8N4O2	428	3	13
Number of Hits		2,818	53	192

(1) 10 possible chemical compositions composed of C,H,N,O (2) Search CAS, KEGG & DNP

図 2. 測定誤差 50 ppm の精度で測定した *m/z* 180.0534 から 7 つの化学組成式がえられ、CAS を検索すると 2,818 化合物がヒットする。この中から二次代謝物質だけを探し出すことは容易ではない。

細胞内量が少ないため、あるいは抽出効率が低いため、MS の検出限界量以下となって検出されなかったのであろう。

検出されたものの同定できなかった約 1520 代謝物質は、それらの標準物質が入手できなかったことによるものである。動物の生体組織の抽出物を高分解能の FT-ICR-MS で測定すると、数千から数万の代謝物質を検出できるが、それらのうちわずかに数百が同定されているにすぎない、とされている。

このように生体成分を高感度・高分解能 MS で分析しても、high-throughput なメタボローム解析を実現できていない現実がある。

標準物質が無くても代謝物質を同定するための方法は、(1) (M+H)⁺ や (M+Na)⁺、(M-H)⁻ などの分子イオンの *m/z* 値を精密に測定することによって化学組成式を決定し、化合物データベースを参照して特定する、(2) (MS)ⁿ スペクトルを測定し、既知のマススペクトルと一致するものを探して特定する、(3) 化学構造とマススペクトルとの関係を利用して、(MS)ⁿ スペクトルから化学構造式を推定する、の3つであろう。MassBank は (2) と (3) に寄与することを目的としている。(3) については実用になるものはまだ無い。

3. 二次代謝物質のデータベース

分子イオンの *m/z* 値を精度よく測定できると、二次代謝物質を同定する手順は次のようになる。測定した *m/z* 値からその値を与える化学組成式を推定する。論理的に推定される化学組成式は極めて多いので、同位体ピークの *m/z* 値と相対強度、構成元素の種類など、を考慮して組成式を絞り込むことになる。

例えば、あるイオンについて測定誤差 50 ppm と 10 ppm で *m/z* 180.0534 がえられたとき、CHNO の 4 原子から構成される可能な化学組成式はそれぞれ 7 つ、1 つである (図 2)。これらの化学組成式を用いて Chemical Abstracts Service (CAS) を検索すると、誤差が 50 ppm, 10 ppm のとき、それぞれ 2,818 化合物、594 化合物がヒットする。それらから二次代謝物質をリストアップする作業は人手に頼らざるをえない。このようなイオンが数千から数万検出される FT-ICR-MS 分析では、二次代謝物質を短期間に手作業で選び出すことは

不可能であろう。

そこで、代謝物質だけを集めたデータベースがあれば、分子イオンの m/z 値や化学組成式から代謝物質を自動的にリストアップすることができるかと期待される。

代謝に関する総合的なデータベース KEGG^{3,4)} は様々な生物種に共通に存在している基礎代謝物質に重点をおいて収集している。生物種に固有な二次代謝物質をほとんど収集していないので、 m/z 値や化学組成式で検索してもヒット率は極めて低い (図 2)。

フラボノイドなど化合物グループごとに収集した書籍類はこれまでに数多く出版されている。Dictionary of Natural Product (DNP)⁵⁾ は二次代謝物質を網羅的に収集したデータベースである。2007 年 6 月版には 20 万 7 千化合物が収集されている。先ほどの化学組成式を用いて検索すると、192 化合物がヒットする (図 2)。

ここで日本で構築しているユニークな二次代謝物質のデータベースを 3 つ紹介しておこう。

1 つは、奈良先端科学技術大学院大学・金谷重彦教授のグループが作成している KNApSAcK^{6,7)} である。これにはフラボノイド (既知のものほぼ全て)、テルペノイド類 (全体の約 3 割)、その他にグルコシノレート、アルカロイドなど計 21,009 化合物が登録されている (2008 年 1 月現在)。

KNApSAcK には、これまでの二次代謝物質のデータベースにはなかった、2 つの特徴がある。(i) 従来の天然物関連データベースでは、その二次代謝物質が最初に報告された論文やその後の構造決定に関する論文を収集している。KNApSAcK では、このような論文だけでなく、その後、各二次代謝物質が他の生物種について見つけられたことを報告する論文があれば、これらをもれなく収集している。(ii) これらの論文を整理することによって、代謝物質とそれが見つけられた生物種との関係 (代謝物質-生物種の関係) を収集している。20,500 代謝物質と 12,419 生物種による 41,530 代謝物質-生物種ペアが登録されている (2008 年 1 月現在)。

この 2 つの特徴は二次代謝物質をゲノム情報とリンクするために不可欠のものである。例えば、3 つの近縁生物種のうち、a 生物種で二次代謝物質 D の生合成中間体と推定される A と C が、b 生物種で B が、c 生物種で B と C が見つかり

るならば、それらを総合して D の生合成経路として $A \rightarrow B \rightarrow C \rightarrow D$ を予測できる。この 3 つの生合成反応に関与する酵素遺伝子をゲノムから見つけることも容易になるであろう。

2 つめの二次代謝物質データベースは、東京大学大学院・有田正規准教授が作成している Flavonoid viewer^{8,9)} である。これは KNApSAcK との共同研究の成果が集約されたものであり、約 6,800 のフラボノイド、1 万 4 千のフラボノイド-生物種ペアが登録されている (2008 年 1 月現在)。Flavonoid viewer の特徴は、フラボノイドの化学構造式を詳細に調べて分類し、化学構造から生合成反応の経路を推定していることである。これは、化学構造式の誤りを徹底的に取り除くとともに関連論文を収集し、生体内反応に関する豊富な知識を蓄積した成果にもとづくものである。

3 つめの二次代謝物質データベースは日本脂質学会で作っている LipidBank¹⁰⁾ である。これには 7009 化合物が収集されている (2008 年 1 月現在)。このデータベースも有田らが収集と維持にあっている。

これら 3 つのデータベースに共通していることは、化学構造式や文献情報が正確で、誤りがないうように細心の注意のもとに収集、維持されていることである。

4. 化学構造式を推定するために

・(MS)ⁿ の測定

二次代謝物質のデータベースが充実してくると、図 2 で示したように分子イオンの m/z 値や化学組成式からかなりの代謝物質がヒットすることがわかる。

しかし、このようにしてヒットした代謝物質は、MS で測定した分子イオンを与えている代謝物質そのもの、と判定できない。それが文献に記載されていない、新規な代謝物質である可能性を否定できないからである。

MS を用いて化学物質の化学構造を解析するためには、一般に、次のような測定がおこなわれる。質量分析計の中で物質を高エネルギーの窒素ガスで壊す (CID) と、分子は一度に小さな断片に粉々に壊れるのではなく、大きな断片から小さな断片へと、順番に壊れると推定される。ある分子の中で最も不安定な化学結合で先ず切断して大

きなフラグメントイオンが生成し、次にそのフラグメントイオンの中で最も不安定な化学結合が切断してさらに小さなフラグメントイオンを生成する、を繰り返している。このように大きなフラグメントイオン (precursor ion) から小さなフラグメントイオン (product ion) が生成する関係 (precursor-product イオンの関係) は化学構造に依存することが知られている。

この関係は、MS/MS や (MS)ⁿ で測定できる。(MS)ⁿ は次のような測定法である。分子イオンから生じたフラグメントイオンのうち1つを選んで、それを precursor イオンとしてさらに CID によって壊すと、いくつかの product イオンを生ずる。この product イオンのうちから1つを選んで precursor イオンとしてさらに CID によって壊すと product イオンを生ずる、この過程を n 回繰り返して測定することを (MS)ⁿ とあらわす。MS/MS は n が 2 の測定であり、(MS)² と同じである。ITMS を用いると n が 2 以上の (MS)ⁿ 測定ができる。

同じ化学組成式で表される化学物質でも、化学構造が異なると、(MS)ⁿ で測定したマススペクトルが異なることになる。

新規な二次代謝物質を (MS)ⁿ で測定したマススペクトルから、その化学構造を推定するためには、化学構造の特徴がどのようなマススペクトルの特徴として現れるのか、という関係 (化学構造とマススペクトルの関係) をあらかじめ知っておく必要がある。このような関係を収集して解析するためには、多様な化学構造の代謝物質を (MS)ⁿ で測定したマススペクトルを収集したデータベースが必要である。

・有機化合物のマススペクトルデータベース

マススペクトルデータベースとして最大のものは、米国 National Institute of Standards and Technology (NIST) が測定し、販売している NIST/EPA/NIH Mass Spectral Library である¹¹⁾。2005年版には、16.3万化合物について測定した約19万 EI-MS スペクトルと約5千 MS/MS スペクトルデータが収録されている。また、日本では産業技術総合研究所が測定・収集している有機化合物のスペクトルデータベース SDBS がある¹²⁾。これには 23.5 千件の EI-MS で測定したマススペクトルデータが収集されている。

テルペン類など揮発性の二次代謝物質は

GC-EI-MS を用いて測定しているので、得られたマススペクトルを NIST や SDBS を検索・参照して、二次代謝物質を同定することができる。

ドイツ Max Planck 植物分子生理学研究所では、植物の抽出物を TMS 誘導体化した試料を GC-EI-TOF-MS で測定した 855 マススペクトルタグ (MST) を収集して、CSB.DB として公開している^{13,14)}。タグと命名している理由は、GC の保持時間で分割して得られたマススペクトルであるので、GC ピークの重なり由来するノイズが含まれるからである。重なりを除くことができた 632MST のうち、標準物質のマススペクトルと比較して代謝物質を同定できたものは 299MST である。

残念ながらこれらのデータベースには、メタボローム解析によく用いられているイオン化法である electrospray ionization (ESI) や fast atomic bombardment (FAB), matrix-assisted laser desorption ionization (MALDI) を用いた MS や hybrid MS を用いて測定したマススペクトルは収集されていない。また、イオンの *m/z* は実数値ではなく整数値で表現されているので、イオンの化学組成式を推定したり、化学構造とマススペクトルの関係を収集することは難しい。

5. データベース MassBank

・様々な条件で測定したマススペクトルの収集

MassBank (www.massbank.jp) はメタボローム解析に用いられているイオン化法と hybrid MS を用いて測定したマススペクトルのデータベースであり、科学技術振興機構のバイオインフォマティクス事業 (JST-BIRD) の支援を受けている (平成 18-22 年度)。この事業には KNApSACk と Flavonoid viewer グループも参加して、密接に連携をしている。

これまで、メタボローム解析に用いられているイオン化法と hybrid MS を用いて測定したマススペクトルには、次のような問題があることが指摘されていた。そのために、このようなデータベースは無かったし、それを構築しようとする意欲も打ち砕かれてきた。

ESI などのイオン化法では、イオン化の条件や CID の条件によって、さらには、装置によっても得られるマススペクトルが大きく異なる。例えば、

同じイオン化法を使用しても、装置や hybrid MS の組み合わせによって、出現するフラグメントイオンの種類や相対強度も大きく変化することも知られている。

これに対して、約 70 eV の熱電子を用いてイオン化する EI 法では、装置にかかわらず、得られるマススペクトルはほぼ同じである。従って、自分で測定したマススペクトルがデータベースにあるマススペクトルと同じかどうか、照合することによって化合物を同定することができる。

MassBank では、このような問題点があるにもかかわらず、異なる分析条件で測定したマススペクトルをともかく収集することにした。

イオン化の化学的原理が同じあるいは似たものであれば、解裂を誘導する電荷中心 (charged site) や切れやすい化学結合はほぼ同じなので、生ずるフラグメントイオンの種類にはそれほど違いはないと期待される。また、 m/z の値とその相対強度の 2 つを重視した従来の検索アルゴリズムを修正すれば、異なるマススペクトルであっても検索できる、との期待もある。

・ MassBank の現状

これまでに 1,174 代謝物質の 10,344 マススペクトルを収集した (2008 年 1 月)。1 つの代謝物質あたり、平均 9 個の異なる測定条件で測定したマススペクトルを収集していることになる。例えば、695 代謝物質について、各代謝物質を 5 段階の異なる CID 衝突エネルギーで壊した MS/MS を、QqQ (Q は quadrupole、四重極型 MS の略) や高分解能の QqTOF-MS で測定した 8,785 マススペクトルがある。また、収集している代謝物質は、基礎代謝物質、植物ホルモンなどの植物が生産する二次代謝物質、動物組織の脂質、カロチノイドなど、多様である。

従来のデータベースは低分解能マススペクトルだけを収集していたので、 m/z はユニットマス (整数に丸められた m/z) で扱われているが、MassBank では実数値 m/z で検索できるようにした。

・ スペクトル検索

スペクトル検索は、あるマススペクトルをクエリ (質問) として、MassBank に蓄積されたスペクトルを検索するサービスである。たとえば、ユーザが自ら測定したスペクトルをクエリとして検索することができる。

2 つのマススペクトルが似ていることを判定する検索アルゴリズムは、NIST で採用しているもの^{15,16)}をそのまま使ってみた。それでも検索してみると、化学構造が似た代謝物質がヒットすることがわかった。化学構造が似ているにもかかわらず、ヒットされないマススペクトルや、全く似ていないにもかかわらずヒットされるものもあったが、実用に値するデータベースにできるという感触をえた。

検索アルゴリズムは後に改良するとして、検索結果の表示とマススペクトルの統合化、の 2 つについて検討した。

・ 検索結果の表示

MassBank には 1 つの代謝物質に複数のマススペクトルが収集されているので、1 つのマススペクトルをクエリとして検索すると、クエリに最も似ているマススペクトル 1 つが選ばれる (図 3)。それと同時に、そのマススペクトルを与えている代謝物質について登録されているマススペクトルを全て同時に三次元的に表示するようにした。このように 1 つの化合物について測定された全てのマススペクトルを見ると、その代謝物質 (図 3 の例で示すと S-adenosylmethionine) のマススペクトルの特徴をととてもよく理解できることがわかった。

・ マススペクトルの統合化

S-adenosylmethionine の例で示したように、異なる条件で測定したマススペクトルを重ね合わせた「統合マススペクトル」(図 4) は、この代謝物質を代表するマススペクトルであるとみなすことができる。各代謝物質ごとに統合マススペクトルをつくっておいて、統合マススペクトルに対して検索すれば、各マススペクトルを対象にして検索するよりもヒット率やその正答率が向上すると期待される。

また、MassBank では、数段階の異なる CID 衝突エネルギーで壊して得られるマススペクトル

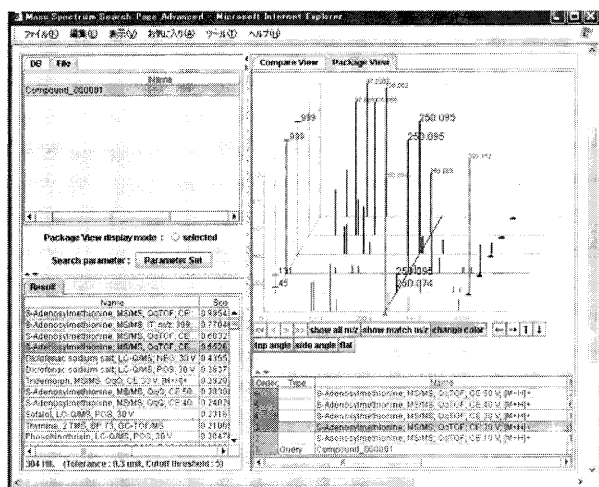


図3. 検索結果の多重表示。左上のウィンドウにはクエリとして入力されたマススペクトルのファイル名 **compound_00001** が表示され、左下のウィンドウには検索されてヒットしたマススペクトルのリストが、一致度の高いものから順に出力されている。右上のウィンドウには最も似ていたマススペクトルを与えた **S-adenosylmethionine** について MassBank に登録されている全てのマススペクトルが自動的に三次元表示される。右上のウィンドウで、一番手前にあるのがクエリのマススペクトルである。

を、自動的に積算した ramp と呼ばれるマススペクトルも収集している。QqTOF-MS で測定した ramp の MS/MS データが 127 件収集されている。Ramp のデータは今後増えていく予定である。

・分散型データベース

GenBank や PDB で代表される生命科学関連の学術関連データベースは、いずれもデータを一ヶ所に集めて公開している。集中させることによって、データ形式の統一をはかり、データやサーバの集中管理をおこなうためである。

それに対して、MassBank は、マススペクトルを測定した研究者がそれぞれ所属する研究室や研究機関においてサーバーを設置して、データを公開する分散型データベースである。MassBank はデータを収集しているのではなく、検索ツールを開発しているにすぎない。本総説では収集という言葉を使っているが、実際には検索しているという意味である。

現在、8 大学・研究所（うち1つはドイツ）に MassBank をインストールしたサーバが設置されている。そのうちデータを公開しているのは4大学・研究所であり、残りはデータの準備中である。検索コマンドを入力すると、4ヶ所に分散しているデータを検索しに行くことになるが、応答が極めて速いのであたかも全てのデータが一ヶ所に集中しているように感ずる。

分散型データベースの利点は、マススペクトル

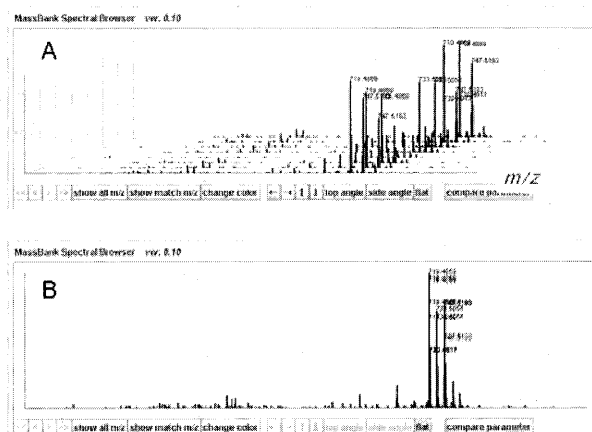


図4. 統合マススペクトル。A は1つの代謝物質を異なる 8 つの条件で測定したマススペクトル。B は 8 つのマススペクトルを重ね合わせたもの。これを統合マススペクトルとする。

を測定した研究者や研究機関がデータの著作権を有する、データに責任をもって公開する、自分たちが公開しているデータの数や質を他のサイトと比較して質や量を競うことができる、などである。MassBank で検索するごとに全てのサイトを検索するので、アクセス数が増えるという利点もある。

・データの共有や交換を可能にするために

データを共有するためには、レコード形式を標準化して、共通にすることが必須である。MassBank の各データは、大きく分けて次の5つから構成されている。データの識別番号や著作権に関する情報、代謝物質の化学情報、クロマトグラフィーによる分離に関する情報、マススペクトルの測定に関する情報、マススペクトルのデータである。異なる質量分析や装置で、異なる測定条件で測定しているので、それらについて詳細に記述しておくことができるように配慮している (図5)。

プロテオーム解析やメタボローム解析のデータを論文などに公表する場合にも、これらの項目を詳細に記述することが必要であるということが、共通の認識になっている¹⁷⁻¹⁹。生命科学研究において測定されたマススペクトルデータを交換・検索することを可能にするためには、これらの項目を表現するデータ形式が標準化され・国際的に共通になることが必要である。

6. おわりに

・MassBank が実証したこと

MassBank には現在のところ、12の異なる質量分析装置を使って、異なる条件で測定したマススペクトルを検索することができる。まだデータの数は少ないが、このように様々なマススペクトルを分散型データベースとしてデータを交換したり、検索することができることを実証したことは質量分析を生命科学に利用している研究者に大きな希望を与えている。近い将来、データ提供をする研究機関が日本だけでなく世界に広がることを期待するものである。

文献

1. Gross JH, Mass Spectrometry. A Textbook, Springer-Verlag Berlin Heidelberg, pp 518, 2004.
2. Soga T, Ohashi Y, Ueno Y, Naraoka H, Tomita M & Nishioka T. Quantitative metabolome analysis using capillary electrophoresis mass spectrometry. *J. Proteome Res.*, 2, 488-494 (2003).
3. Kanehisa M, Goto S. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucl. Acids Res.*, 2000, 28: 27-30.
4. KEGG database: <http://www.genome.ad.jp/ja/>
5. Dictionary of Natural Products. Chapman and Hall/CRC. 16:1 版 (June 2007) .
6. Shinbo Y, Nakamura Y, Altaf-Ul-Amin M, Asahi H, Kurokawa K, Arita M, Saito K, Ohta D, Shibata D, Kanaya S. KNApSAcK: A comprehensive species-metabolite relationship database. In Saito, K., Dixon, R. and Willmitzer, L. (eds.), *Plant metabolomics*. Springer, New York, Vol. 57, pp. 165-181, 2006.

Tag	Mandatory / Optional	Unique / Iterative	Single line / Multi line	Subsection	Description
Record Specific Information					
ACCESSION	M	U	S	2.1	Record identifier
RECORD TITLE	M	U	S	2.2	Short title
DATE	M	U	S	2.3	Date of creation
AUTHORS	M	U	S	2.4	Name and affiliation of authors
PUBLICATION	O	I	S	2.5	Bibliographic information of reference
COMMENT	O	I	S	2.6	Comments
Compound Information					
CH\$NAME	M	U	S	2.7	Name of chemical compound
CH\$FORMULA	M	U	S	2.8	Chemical formula
CH\$EXACT MASS	M	U	S	2.9	Exact mass
CH\$SMILES	M	U	S	2.10	SMILES code
CH\$INPAC	M	U	S	2.11	InChI code
CH\$LINK	O	I	S	2.12	Identifier of external databases
Analytical Information					
ACS\$INSTRUMENT	M	U	S	2.13	Name of equipment and its manufacturer
ACS\$ANALYTICAL CONDITION	O	I	S	2.14	Analytical conditions
Spectral Information					
MSS\$FOCUSED ION	O	I	S	2.15	Precursor ion and m/z
MSS\$PROFILE	O	U	M	2.16	Profile data information
MSS\$BINDATA	O	U	M	2.17	Raw binary data information
MSS\$DATA PROCESSING	O	I	S	2.18	Data processing method
Peak Information					

図5. MassBank のレコード形式。データのレコード番号や著作権に関する情報、代謝物質の化学情報 (CH\$)、クロマトグラフィーによる分離条件に関する情報 (ACS)、マススペクトルの測定条件に関するデータ (MSS)、マススペクトルのデータ (PK\$) の5つから構成されている。

7. KNApSAcK:
<http://kanaya.naist.jp/KNApSAcK/KNApSAcK.php/>
8. 時松敏明、有田正規、代謝マップビューワで見るフラボノイド、細胞工学、**25**(12), 1388-1393, 2006.
9. FlavonoidViewer:
<http://www.metabolome.jp/software/FlavonoidViewer/viewer/>
10. LipidBank: <http://www.lipidbank.jp/>
11. NIST/EPA/NIH Mass Spectral Library:
<http://www.nist.gov/srd/nist1a.htm/>
12. SDBS:
http://riodb01.ibase.aist.go.jp/sdbs/cgi-bin/cre_index.cgi?lang=jp/
13. Wagner C, Sefkow M, Kopka J (2003). Construction and application of a mass spectral and retention time index database generated from plant GC/EI-TOF-MS metabolite profiles. *Phytochemistry* **62** 887-900 (2003).
14. CSB.DB:
<http://csbdb.mpimp-golm.mpg.de/index.html/>
15. Stein SE (1994). Estimating Probabilities of Correct Identification from Results of Mass Spectral Library Searches. *J. Am. Soc. Mass Spectrom.*, **5**, 316-323.
16. Stein SE. Scott DR (1994). Optimization and Testing of Mass Spectral Library Search Algorithms for Compound Identification. *J. Am. Soc. Mass Spectrom.*, **5**, 859-866.
17. HUPO Proteomics Standard Initiative:
<http://www.nature.com/nbt/consult/index.html/>
18. Lindon JC. (2005). Summary recommendations for standardization and reporting of metabolic analysis. *Nature Biotechnology*, **23**(7), 833-838.
19. Sansone SA. et al (2007). The metabolomics standards initiative. *Nature Biotechnology*, **25**(8), 846-848.