

解説

マススペクトルのデータベース MassBank が 世界標準になるまで

西岡 孝明*

京都大学大学院農学研究科名誉教授

〒 606-8502 京都市左京区北白川道分町

* takaaki.nishioka.23u@st.kyoto-u.ac.jp

(2017年8月21日受理; 2017年9月27日掲載決定)

マススペクトルのデータベース MassBank について、それを必要とした質量分析の背景、設計コンセプト、メタデータの内容、データの冗長性、レコード作成を支援するツールの提供、データ検索機能、レコードの著作権とコピーの利用、日本質量分析学会による支援、データの社会的貢献、国際連携、データの品質、新しい MassBank の開発、マススペクトルデータベースの将来、MassBank のさらなる発展、について解説する。

Progress of MassBank towards the Global Standard of Mass Spectral Database

Takaaki Nishioka*

Professor Emeritus, Graduate School of Agriculture, Kyoto University,

Oiwake-cho, Kitashirakawa, Sakyo-ku, 606-8502 Kyoto, Japan

* takaaki.nishioka.23u@st.kyoto-u.ac.jp

(Received: August 21, 2017; Accepted: September 27, 2017)

This article describes the development of MassBank in terms of the historical background why mass spectrometry has required MassBank, concepts, metadata, data redundancy, tools helping data contributors who construct MassBank records, data search services, copyright and Creative Commons License of Mass Bank records, contribution from the Mass Spectrometry Society of Japan, Open Science of mass spectral data, international collaborations, data quality of public repository, new MassBank, and future direction of mass spectral databases.

1. はじめに

本解説では分子の質量が約 1,500 Da までの有機化合物のマススペクトルを収集したデータベース MassBank [1] が世界標準とみなされるに至った背景を紹介する。最初にマススペクトルのデータベースの歴史を簡単に紹介しておこう。1950 年には原油に含まれる炭化水素化合物の組成を調べるために電子イオン化質量分析 (electron ionization mass spectrometry: EI-MS) が利用され、約 1,500 の炭化水素化合物を EI-MS で測定したマススペクトルのカタログが作られていた[2]。その後、ガスクロマトグラフィーと EI-MS を組み合わせた GC-MS が普及す

るとともに、分析対象も有機化合物一般へと拡大した。いろいろな有機化合物を EI-MS で測定した大規模なデータベース (spectral library) が有料で提供されてきた。これが今日の NIST/EPA/NIH Mass Spectral Library [3]や Wiley Registry of Mass Spectral Data [4] へと継承されている。このように大規模で長い歴史を有するマススペクトル DB が構築された背景には、EI-MS の測定条件が標準化されていたことと、分析機器に大きな変化がなかったことを挙げることができる。

1990 年代に LC-MS が実用化されると GC-MS とともに、環境科学をはじめ細胞内にある全ての代謝

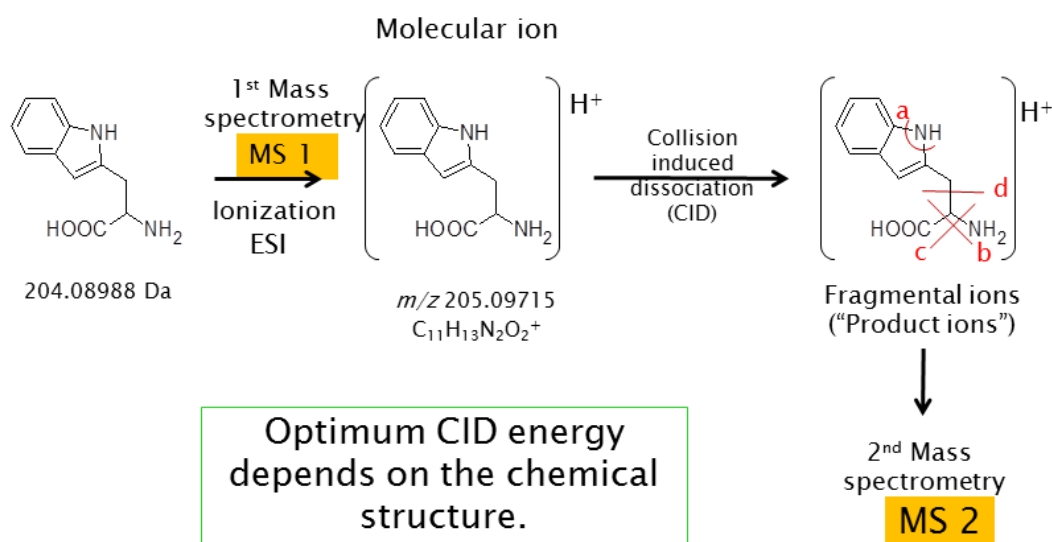
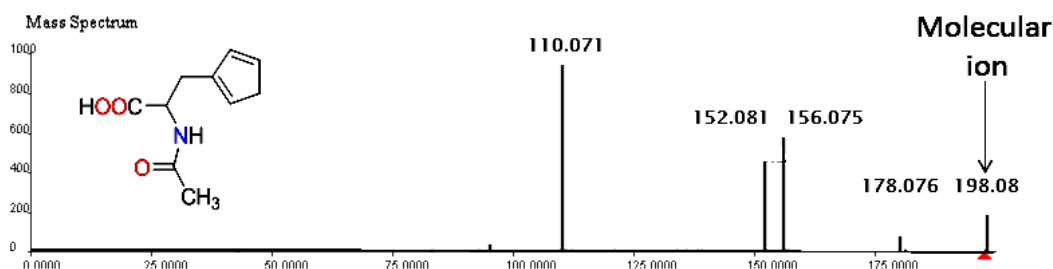


Figure 1. In the electrospray ionization-tandem mass spectrometry (ESI-MS/MS), the first mass spectrometry (MS1) ionizes the neutral molecule to molecular ion and selects the ion. Collision induced dissociation (CID) of the selected molecular ion produces fragmental ions, which are analyzed by the second mass spectrometry (MS2).

A) Collision Dissociation Energy = 20 V.



B) Collision Dissociation Energy = 50 V.

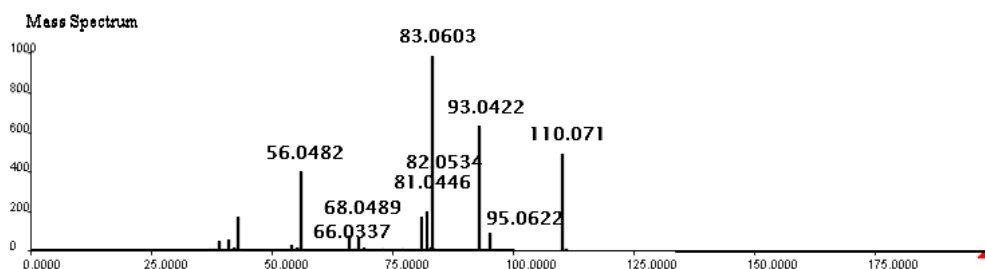


Figure 2. Optimum energy of collision induced dissociation (CID) depends on the chemical structure of the molecular ion. (A) Lower CID energy produces larger fragment ions. (B) Higher CID energy gives smaller fragment ions.

物を分析しようとするメタボロミクス (metabolomics) などの生物科学への利用がはじまった。これらの分析には主にエレクトロスプレーイオン化 (electrospray ionization; ESI) が利用される。ソフトイオン化である ESI は安定な分子イオンを生成するので、最初の質量分析 (M1) で分子イオンを検出する。次に分子の化学構造情報を得るために、

分子イオンを衝突誘起解離 (collision-induced dissociation; CID) させて、化学構造に特徴的な断片化したイオン (fragment ion) を生成して質量分析 (M2) をする。このように2つの質量分析を連結したタンデム質量分析 (ESI-MS/MS) で測定している (図 1)。このとき解離エネルギーが分子イオンの構造エネルギーと比べて、大きいと質量が小さな

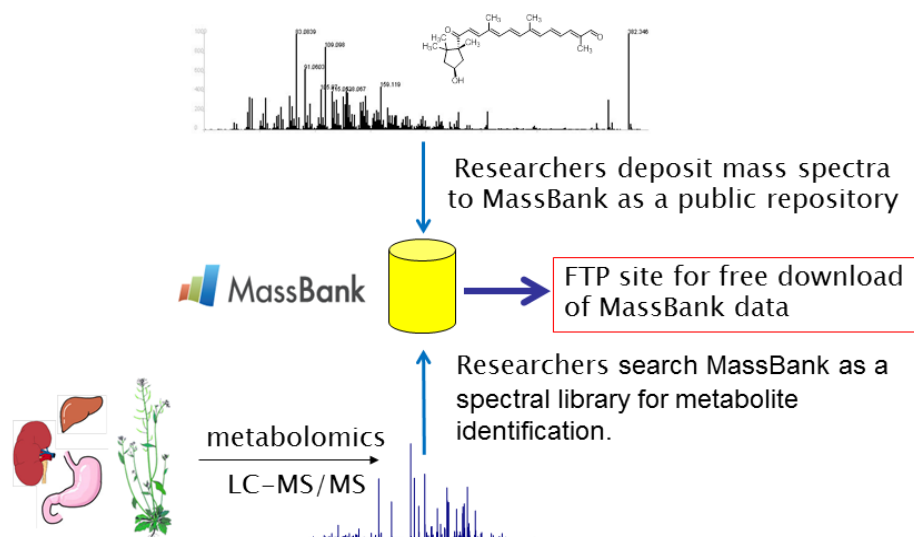


Figure 3. MassBank is a public repository for publishing mass spectra as experimental data and a spectral library for the chemical identification. All the mass spectral data deposited on MassBank are freely downloadable from the ftp site.

fragment ion が得られるし、小さいと質量が大きな fragment ion が得られる (図 2)。分子を化学的に信頼性よく同定するためには、質量が小さい fragment ion から大きなものまでが観測される「適度な解離エネルギー」で測定をしたマスペクトルを spectral library として収集しなければならない。このエネルギーは分子の化学構造に依存するので、化合物ごとに異なるのと、具体的に定義することが難しい、などの理由によって、ESI-MS/MS の分析条件を標準化することはできなかった。このことが ESI-MS/MS で測定した本格的な spectral library の構築を制約していた。

さらにもう 1 つの制約要因があった。このころ様々なタイプのイオン分析計、例えば 4 重極 (quadrupole; Q) をはじめイオントラップ (ion trap; IT), 飛行時間 (time of flight; TOF), フーリエ変換 (Fourier transform; FT) イオン分析計、などが一斉に実用化されたことである。MS1 と MS2 をそれぞれ異なるイオン分析計で構成したハイブリッド型のタンデム質量分析、例えば IT-FT や Q-TOF, も急速に普及した。これらのイオン分析計は生成する fragment ion や測定の分解能が異なっている。これらの装置を買い揃えた分析センターを設置して、spectral library を作成することはもはや不可能であった。さらに世界の研究グループは、経済的事情が許す範囲で、それぞれ異なるタイプのタンデム質量分析を購入してメタボロミクス解析をおこなっていた。これらのうちどれか 1 つを対象として spectral library を構築することもできなかった。

2. MassBank のコンセプトとサービスの提供

ESI-MS/MS では「分析条件を標準化できない」、 「多様なタイプの分析装置が利用されている」という 2 つの困難をクリアして、有機化合物の spectral library を構築する方法として、世界中の研究者からマスペクトルを提供してもらって、それを spectral library として研究者に提供しよう、と著者は考えるにいたった。2006 年、JST から分子生物学関連のデータベース化事業の公募があった。応募したところ採用されたので、この考えを実行することにした。

MassBank のコンセプト[5]。次の 4 つである。(1) MassBank プロジェクトはデータの入れ物を開発するだけである。自ら質量分析をおこなうことはしない。Spectral library にマスペクトルを提供することはしない。(2) MassBank は研究者コミュニティでマスペクトルを共有する public repository であり、(3) マスペクトルを提供する研究者が自らデータと分析条件などを定められたレコード形式にして提供、公開する、(4) MassBank は public repository に提供されたマスペクトルの全てをそのまま spectral library として提供する (図 3)。

各マスペクトルのメタデータは必須。 実験データとしてのマスペクトルの再現性を担保するために、マスペクトルを提供する研究者にはピークデータ (m/z と相対強度) だけでなくメタデータを

記述することを必須条件とした。ここでメタデータとは、測定に用いた質量分析装置についての情報はじめイオン化やCIDエネルギーなどの分析条件、分析した化合物の化学構造情報、データを提供した研究者に関する情報、などである。実際の MassBank レコードでピークデータとメタデータがどのように記述されているのかを実例で見ていただくのがよいだろう (Browser は Internet Explorer を使用のこと)。

(<https://massbank.eu/MassBank/jsp/RecordDisplay.jsp?id=AC000107&dsn=AAFC>)。

論文の Materials and Methods の内容と似ているが、質量分析の分析パラメータをより詳述したものである。メタデータの用語や記述を規定した MassBank Record Format 定義書 (英語版) を作成し、公開している [6]。定義書は新しい分析方法が開発されると、日欧の MassBank プロジェクトメンバーから構成される MassBank Committee で協議して随時改訂をしている。

マススペクトルの冗長性が重要。 メタボロミクスでは様々な代謝物の混合物を、解離エネルギーを固定して ESI-MS/MS で分析するので、想定外のマススペクトルが得られる。そこで MassBank では1つの化合物を段階的に異なる5つ以上の解離エネルギーで測定したマススペクトルを提供、公開することを研究者に推奨している。例えば caffeine は38件

のマススペクトルが登録されている。ユーザがどのような分析条件で測定した caffeine のマススペクトルであっても、そのどれかと一致するか、極めて似ているマススペクトルが MassBank で見つかることが期待できる。さらに、この化合物が与えることができるフラグメントイオンとその測定条件を網羅的かつ系統的に知ることができるので、マススペクトルの化学的注釈には極めて有用でもある [7]。これはマススペクトルが冗長に公開されていることと詳細なメタデータがあつてこそ、の利用法である。

マススペクトルの提供と公開を支援するツールの提供。 質量分析ではたえず新しいイオン化や高精度なイオン分析計が開発、実用化されている。新しい機器を購入した研究グループは、彼らの装置で測定したマススペクトルを速やかに MassBank から公開することが期待されている。MassBank プロジェクトでは Excel をベースとして、マクロを利用したレコード編集ツールを開発して研究者に提供した [6]。それでもメタデータの作成は手作業が多く、レコード作成に時間と人手を必要とするので、公開する研究者の負担となっていた。マススペクトルを測定すると、質量分析計から自動的にピークデータとメタデータを取り出し、化合物 DB である PubChem [8] や KEGG [9] などから化合物構造情報を取得して、MassBank レコードを自動生成するツールを提供してほしい、という強い要望がマススペ

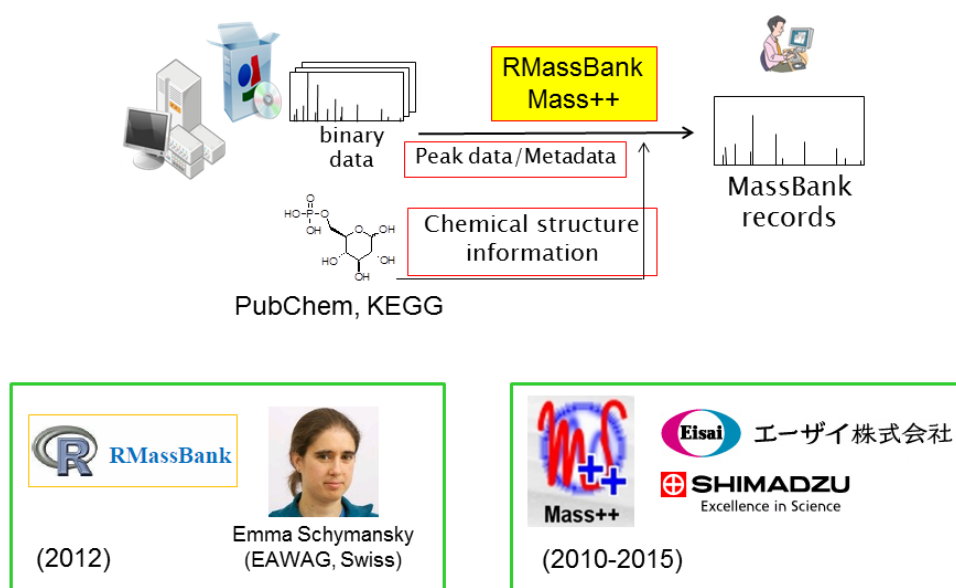


Figure 4. RMassBank and Mass++ are the fully automatic tools for preparing MassBank records. Both of the tools fetch the peak data with its metadata from the raw data of mass spectrometry instruments and the chemical structure information from the internet websites of chemical compound databases such as PubChem and KEGG.

クトルの提供者からいくつも寄せられるようになった。しかし、分析計はメーカーごとに異なるバイナリ形式でピークデータやメタデータを保存、出力しているため自動生成ツールの作成は困難であった。

2013年、EU参加国の環境研究グループの連合体である NORMAN Association は、最も高分解能で利用者が多い質量分析 Orbitrap に特化したものではあるが、MassBank レコード完全自動生成ツール RMassBank を開発し、無料提供した [10]。国内では、エーザイが開発した LC-MS/MS 用ピーク検出ツール Mass++ をベースとして、島津・田中耕一グループが島津製品だけでなく主要な分析機器メーカーの質量分析計に対応した MassBank レコード完全自動生成機能を付与した Mass++ 改定版を開発して無料提供した [11] (図 4)。

研究者の協力とこのようなレコード作成ツールの配布の結果、現在 MassBank から公開されているマスペクトルは質量分析 27 機種で測定した 48,252 レコードに達している (図 5)。

データ検索機能の充実。 MassBank がユーザに好評であった理由の 1 つは豊富な検索ツールを日本語、英語 2ヶ国語の利用マニュアルとともに提供したことである。マスペクトルの類似性を利用した化合物同定用の検索ツールや、ある特定の部分化学構造、例えばピリミジン骨格、を有している化合物のマスペクトルだけを抽出する部分化学構造検索、ある特定のピークセットが観察されているマスペクトルだけを串挿し検索するピーク検索、ある特定のイオン化や質量分析計タイプを対象とした検索、などのツールを提供している [6]。またメタボロミクスでは、1 試料を LC-MS/MS すると数百以上のマスペクトルが得られる。これらをクエリとして一挙に検索するバッチ検索など、いずれも他のマスペクトルの DB が提供していないユニークな検索ツールを開発して、提供している。ユーザが Python などの言語で記述したツールを用いて MassBank 検索をおこなうための application program interface (API) も提供した [6]。

MassBank レコードの著作権、コピーとその利用。 研究者が学術雑誌の論文で公表したマスペクトルを MassBank に提供するとき、研究者は雑誌の出版元に著作権の許可を得る必要があるのだろうか？ 答えは No である。なぜなら、マスペクトルはイオン化で生じたイオンの物理量 (m/z 値と強度) を

Ionization	Data	Ionization	Data	Ionization	Data
EI-B	11,810	ESI-IT	664	APCI-ITTOF	1
EI-EBEB	12	ESI-ITFT	10,429	APCI-QTOF	633
EI-Q	45	ESI-ITTOF	266	APPI-QQ	287
EI-QQ	19	ESI-Q	2,720	CI-B	796
EI-TOF	944	ESI-QFT	4,592	FD-B	41
EI Total	12,830	ESI-QIT	378	FI-B	1
		ESI-QQ	5,175	Others Total	1,759
FAB-B	26	ESI-QTOF	7,453		
FAB-BE	1	ESI-TOF	446	MALDI-QIT	1
FAB-EB	5	ESI Total	32,123	MALDI-TOF	17
FAB-EBEB	172			MALDI-TOFTOF	45
FAB Total	204			MALDI Total	63

(14 Feb 2017)

Figure 5. Ionizations and types of tandem mass spectrometry of MassBank data are summarized.

あらかず数値データであり、物理量には著作権は及ばないからである。これに対して MassBank ではマスペクトルとメタデータを併せて 1 レコードと定義している。研究者は自らの実験方法をもとにメタデータを記述しているため、MassBank レコードは提供者である研究者の著作物とみなすことができる。各レコードには提供者の氏名と所属機関を著作権者として明記している。MassBank はいつも最新のデータを誰でも無料でダウンロードすることができる ftp サイトを設けている (図 3)。MassBank ユーザはダウンロード (コピー) したレコード (著作物) の改変やそれらの再配布、有料での配布を、本来ならばレコードの提供者の許可を得ておこなうことになるが、これではとても不便であるのと提供者の負担になるため、マスペクトルの提供者にはあらかじめ各レコードにこのような取扱いを Creative Commons License (CC ライセンス) [12] に従って表示することをお願いしている。情報システム研究機構ライフサイエンス統合データベースセンター (DBCLS) が推奨する学術データベースの CC ライセンスは、コピーしたデータを有料で配布することも認める CC BY である。しかしマスペクトルが販売されている現状を考慮して、企業が提供する MassBank レコードについては第 3 者が改変し、有料で配布 (販売) することを禁止する表示 (CC BY-NC) を例外的に認めている。2018 年度からは全て CC BY に統一する予定である。

製薬企業などは自社内で測定したマスペクトルをクエリとして、社外にある MassBank を検索することを禁止している。新薬候補化合物のマスペクトルが MassBank サーバで収集されている可能性を懸念するからである (MassBank はクエリを収集して

いないことを Web home page で表明しているが、収集することは技術的に可能である)。MassBank システムのコピーを企業内に構築したい、という要望が国内外からいくつも寄せられた。そこで MassBank システムを MS Windows 版 PC 上で構築することができる installer を開発して提供した。企業内 LAN 上に MassBank システムを構築すると、自家版 spectral library を作成することや、社内にある異なるメーカーの質量分析装置で分析したマスペクトルを一括してデータベース化することが可能になった(図 6)。CC ライセンスと MassBank システムの配布によって、ダウンロードした MassBank データのコピーはまたたくまに国内外に拡散した。

日本質量分析学会による支援。 著者は質量分析の経験はほとんどないので、専門家の協力を必要とした。日本質量分析学会に入会して、データベースの必要性を訴えた。幸いなことに当時の学会長をはじめマスペクトルデータ部会長や会員の理解を得て、2008年3月に MassBank は学会の公式データベースとして認められ、それ以降、様々な支援を受けることができた(2017年度からは「学会が authorize するデータベース」と変更になった)。

学会が主催する講演会やシンポジウムの機会に MassBank の利用法を説明する講習会を開催した。これらの講習会には初心者だけでなく MassBank のヘビーユーザも参加していたので、彼らからいろいろなアドバイスをいただいて、MassBank の改善に役立てることができた。分析機器の製造や販売業者に集まってもらって MassBank についての説明もおこなった。これは分析機器から共通のデータフォーマットで分析パラメータを取出すことができるように要望をすることと、機器を購入した研究者に MassBank へマスペクトルを提供してもらうことを期待したからである。プロテオミクス関係者の努力もあって、数年後には分析パラメータが機器メーカーとも共通のデータフォーマットで出力されるようになった。

このような活動によって MassBank の知名度は徐々に上昇し、MassBank の使い勝手も改良されていった。

MassBank の社会的貢献。 Web サイトへのアクセス数(32万件/2016年)と MassBank を引用した論文件数(623件/2016年末)は年々増加している。MS 利用者の総数は他の分析手段と比較するとそれほど

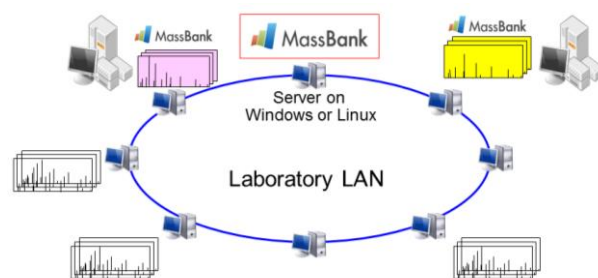


Figure 6. Laboratory MassBank database for sharing mass spectral data analyzed on various mass spectral instruments is constructed by installing a Windows or Linux version of the MassBank system on LAN.

多くないことを考慮すると、この 2 つの統計値は MassBank が世界標準であることを示している。引用論文は薬学関係、生物科学、環境科学など多様な研究分野にわたっている。

国際連携。 MassBank.jp のミラーサイトには Norman MassBank [13]があり、subversion のソースコード管理システムを利用して日欧間でデータの同期をとっている。Norman MassBank を運営する NORMAN Association は EU と北米にある約 70 の国公立環境科学研究機関が参加する世界最大の連合組織である。2012, 2014, 2017 年に Norman MassBank Workshop (2017 年からは Norman Databases Workshop に統合された)を開催している。その他に協力組織として MassBank of North America (MoNA) [14]がカリフォルニア大学デービス校ゲノムセンターから提供されている。

MassBank をはじめ NIST/EPA/NIH Mass Spectral Library や Wiley Registry of Mass Spectral Data などのデータベースはそれぞれ独自のレコード形式を採用しているために、これらデータベース間を串刺しにマスペクトルの検索をすることができなかった。MassBank.jp と Norman MassBank, MoNA のメンバーが協力して、各 MassBank レコードのピークデータ (m/z と相対強度)だけをハッシングしたアルゴリズム SPLASH を開発した[15]。SPLASH は MassBank をはじめ世界の主要なマスペクトルデータベースで採用された。これによって Google を利用して、SPLASH を検索キーワードに入力するだけで、容易に異なるデータベースにある同一のマスペクトルを検出することが可能になった。その結果、MassBank のピークデータがレコード形式を変えて、さまざまなデータベースにコピーされている

ことがわかった。MassBank が世界標準となっていることを実感するできごとであった。

3. データの品質

MassBank は誰でも測定したマススペクトルを公開することができる。当初は QQ-MS/MS などピークの m/z を整数値 (nominal mass) で測定するイオン分析計を用いたマススペクトルがほとんどであった。しかし、2010 年ごろから TOF や FT などの高分解能イオン分析計が普及するとともに m/z 実測値と理論値との差を 10~1 ppm 以下の誤差で測定することが可能になった。理論値から極端に外れたマススペクトルは、提供した研究者の許可を得て MassBank プロジェクトが削除しているが、まだ多くの低品質なマススペクトルがそのままになっていると推定される。日本質量分析学会スペクトルデータ部会では、2011 年に MassBank のマススペクトルの品質チェック (監修) をおこなったことがある。しかしこのチェックはとても手作業でできることではないことを実感して 1 日で頓挫した。当時は MassBank プロジェクトがマススペクトルの品質を定義することができていなかったため、データを自動的にチェックするツールを開発することができなかった。提供者がデータを公開するときにチェックをすることの重要性を認識したできごとであった。

日本質量分析学会では「MassBank データベース構築委員会」を 2016 年に設置して、マススペクトルの品質に関する基準を定める議論をしているところである。MassBank public repository にマススペクトルの提供をうけるにあたってはチェックを自動的におこなって、信頼性あるマススペクトルをだけを MassBank spectral library として提供する予定である。

4. MassBank の危機と第 2 世代 MassBank の開発

MassBank がユーザに好評であった理由の 1 つは豊富な検索ツールを日本語、英語 2 ヶ国語の利用マニュアルとともに提供したことである。例えば、マススペクトルを自由に拡大、縮小して表示することができる、などである。これらのツールのソースコードは Java applet で記述されているが、2012 年に Java applet に致命的なセキュリティホールが見つかった。2017 年春には、ほとんどのインターネットブラウザが Java applet をサポートしなくなったので、MassBank の検索ツールは機能不全に陥った。2013 年から日本と EU で Java applet を使用しない第 2 世

代 MassBank の開発が始まり、今年度内に新しい MassBank が稼動する予定である。

5. マススペクトルデータベースの将来

生物種全体が作る代謝物は約 20 万化合物以上と推定されている。現在、MassBank に収集されている ESI-MS/MS データは約 3,000 化合物である。メタボロミクスの目的である網羅的な化学分析にはほど遠い数である。その対策として現在までに提供されているマススペクトルについて化合物の化学構造とマススペクトルとの関係」を解析して、与えられた化合物の化学構造に基づいて人工的なマススペクトルが作られている。MoNA のマススペクトル約 22 万件のうち実測したデータは MassBank.jp のマススペクトル約 2 万件にすぎない。残りはカリフォルニア大学デービス校ゲノムセンターで測定したマススペクトルをもとにして人工的に作られたマススペクトル (in silico spectral library) である[16]。脂質には共通な骨格がグリセロールエステル、ステロイドのように特徴的で互いに区別できるものが多い。さらにメチレン基 (-CH₂-) の数だけが異なる脂肪酸エステルや環状の骨格に水酸基 (-OH) の数や置換位置が異なる、など「数え上げ」することができることも in silico spectral library の作成を容易にしている。

脂質以外の化合物については人工的なマススペクトルを提供することはとても難しい。多様な化合物を測定したマススペクトルが数多く提供、公開されること必要である。第 2 世代 MassBank では public repository としてのサービスと spectral library としての機能を分離する予定である。これによって in silico spectral library を spectral library に加えることが可能になるだろう。

6. MassBank のさらなる発展のために

これまで MassBank にマススペクトルを提供してきた研究者は、比較的大規模に質量分析をおこなっている、研究資金が潤沢な研究グループであった。彼らは市販の標準化合物を測定したマススペクトルを提供してきた。これらはデータの冗長性という点で MassBank に大きな貢献をしてきたが、化学構造の多様性という面ではこれ以上期待することは難しい。すなわち彼らに新しい化合物を分析したマススペクトルの提供を期待することは難しいと思われる。

むしろ天然物化学や有機合成化学の分野で活躍す

る個人研究者の協力をえて、多様な化学構造をした化合物を高分解能で測定した品質の良いマススペクトルを収集、公開することに MassBank の将来はかかっているだろう。日本質量分析学会 MassBank データベース構築委員会の目的は、マススペクトルを利用した論文を探し出し、その著者にマススペクトルとメタデータの提供を依頼することである。提供されたマススペクトルは提供者に代わって委員会が MassBank レコードを作成して、MassBank から公開することである[17]。

質量分析は分子の質量電荷比 m/z を高精度に測定する、というシンプルな分析法のゆえに、その応用範囲はますます拡大している。実験データとしてのマススペクトルの数は年々増大していると推定される。それにたいして学術雑誌は論文あたりの紙面を制約しているので、論文で実験データとしてのピークデータをすべて掲載する機会は (supplementary も含めて) 失われている。だからこそマススペクトルを MassBank に提供して、公開すべきであるが、MassBank データベース構築委員会の依頼に対して、提供されるマススペクトルは極めて少なく、苦戦をしいられている。

著者に問い合わせた明らかになった理由は次のようなものであった。天然物化学者や有機合成化学者は新規化合物の高分解能質量分析を分析センターなどへ依頼しているが、彼らにとって興味あるデータは MS/MS 測定したマススペクトルではなく、分子イオンの高精度な m/z 値だけである。質量分析を元素分析のかわりに利用しているにすぎないので、MS/MS 測定したマススペクトルは化学的解釈もされないまま、分析装置のハードディスクに保存されたままになっているとのことである。実験をしてから論文が出版されるまでに2,3年はかかるとして、その間に分析依頼をした学生や院生が卒業してしまうと、保存されたマススペクトルを見つけることはとても難しい。このような事例が多いと思われる。

そこで対応策として、質量分析のオペレータ (技術員) や分析依頼した教員に次のような協力をお願いしている。「マススペクトルは分析依頼した学生院生だけでなく担当教員にもデジタルデータとして email の添付ファイルにして渡す」、受け取った教員は「自分の PC 上でマススペクトルを整理、保存をしておいて、論文掲載が受理されると MassBank に提供する」という協力である。

「論文を書いたらマススペクトルを MassBank に提供する」ことを研究者としての義務と理解してほ

しい。これまで利用されなかった貴重な実験データとしてのマススペクトルが有効に活用されることを期待したい。

7. 謝辞

本解説は第 48 回表面分析研究会 (2017 年 2 月 16 日, 大阪大学中之島センター) でおこなった講演の要旨に、最近の MassBank プロジェクトの研究成果を加筆したものである。日本質量分析学会 MassBank データベース構築委員会 (荒川隆一委員長) に感謝する。本解説で紹介した研究成果の一部は科学研究費補助金 (研究成果公開促進費) 「データベース」17HP8021 で実施したものである。

8. 参考文献

- [1] MassBank <http://www.massbank.jp> (2017 年 8 月 14 日閲覧)
- [2] P. D. Zemaný, *Anal. Chem.* **22**, 920 (1950).
- [3] <https://www.nist.gov/srd/nist-standard-reference-database-1a-v17> (2017 年 8 月 21 日閲覧)
- [4] <http://www.sisweb.com/software/wiley-registry.htm> (2017 年 8 月 21 日閲覧)
- [5] H. Horai, M. Arita, S. Kanaya, Y. Nihei, T. Ikeda, K. Suwa, Y. Ojima, K. Tanaka, S. Tanaka, K. Aoshima, Y. Oda, Y. Kakazu, M. Kusano, T. Tohge, F. Matsuda, Y. Sawada, M. Y. Hirai, H. Nakanishi, K. Ikeda, N. Akimoto, T. Maoka, H. Takahashi, T. Ara, N. Sakurai, H. Suzuki, D. Shibata, S. Neumann, T. Iida, K. Funatsu, F. Matsuura, T. Soga, R. Taguchi, K. Saito, and T. Nishioka, *J. Mass Spectrom.*, **45**, 703 (2010). DOI: 10.1002/jms.1777
- [6] MassBank レコードフォーマット定義書や検索ツールの利用マニュアル, データ公開マニュアルなどのダウンロードサイト <http://www.massbank.jp/ja/manual.html> (2017 年 9 月 20 日閲覧)
- [7] 西岡孝明, 蓬萊尚幸, *J. Computer Aided Chem.*, **18**, 15 (2017).
- [8] PubChem <https://pubchem.ncbi.nlm.nih.gov/search/> (2017 年 8 月 14 日閲覧)
- [9] KEGG <http://www.genome.jp/kegg/> (2017 年 8 月 14 日閲覧)
- [10] M. A. Stravs, E. L. Schymanski, H. P. Singer, and J. Hollender, *J. Mass Spectrom.*, **48**, 89 (2013). DOI: 10.1002/jms.3131
- [11] S. Tanaka, Y. Fujita, H. E. Parry, A. C. Yoshizawa, K. Morimoto, M. Murase, J. Yamada, S.

Utsunomiya, S. Kajihara, M. Fukuda, M. Ikawa, T. Tabata, K. Takahashi, K. Aoshima, Y. Nihei, T. Nishioka, Y. Oda, and K. Tanaka, *J. Proteome Res.*, **13**, 3846 (2014). DOI: 10.1021/pr500155z

[12] Creative Commons License <https://creativecommons.jp/licenses/%20> (2017年8月14日閲覧)

[13] NORMAN MassBank <http://massbank.normandata.eu/MassBank/index.html> (2017年8月14日閲覧)

[14] MassBank of North America <http://mona.fiehnlab.ucdavis.edu/> (2017年8月14日閲覧)

[15] G. Wohlgemuth, S. S. Mehta, R. F. Mejia, S. Neumann, D. Pedrosa, T. Pluskal, E. L. Schymanski, E. L. Willighagen, M. Wilson, D. S. Wishart, M. Arita, P. C. Dorrestein, N. Bandeira, M. Wang, T. Schulze, R. M. Salek, C. Steinbeck, V. C. Nainala, R. Mistrik, T. Nishioka, O. Fiehn, *Nature Biotech.*, **34**, 1099 (2016). DOI:10.1038/nbt.3689

[16] T. Kind, K. H. Liu, D. Y. Lee, B. DeFelice, J. K. Meissen, O. Fiehn, *Nat Methods*, **10**, 755 (2013). DOI:10.1038/nmeth.2551

[17] T. Nishioka, and R. Arakawa, *J. Mass Spectrom. Soc. Jpn.*, **64**, 271 (2016).

査読コメント, 質疑応答

査読者 1 阿部 芳巳 (三菱ケミカル)

マススペクトルのデータベースを構築した背景から設計コンセプト, 将来展望まで, 幅広い視点で解説されており, データベースのあるべき論を考える上でたいへん示唆に富んだ解説記事です. 表面分析研究会でも, スペクトルやイオンエッチングプレートなどのデータベースの構築に関わっているため, 本文ではなく, 査読の質疑応答の中で以下の点にコメントいただくと JSA 読者にさらに有益になるものと思われます.

[査読者 1-1]

3章で記述されている通り, データベースを構築する上では, データ提供のハードルをなるべく下げ多くのデータを収集する環境を整えることも大事にはなりますが, データベースとしての信頼性を維持するためには, 低品質なデータを排除したり品質を高めたりする工夫が必要になると思われます. そのためにスペクトルの品質をチェックする基準として, 縦軸(強度), 横軸(質量)に関し, いまどのような議論がなされているのか, 可能な範囲で教えていただくと幸いです.

[著者]

マススペクトルは実験データです. 一般に科学の実験データを発表するにあたって満たすべき要件は, 第3者がそのデータを再現することができるように, 実験条件を記述することです. この実験条件としてどのような項目があるかについて, 日本質量分析学会 MassBank データベース構築委員会[17]で議論をしているところです. MassBank レコード定義書に記述すべき分析条件のリストでほぼつくされていますが, 次の2点についてはこれまで記述項目がありませんでした.

(1) 質量分析装置が正しく校正されていることを示す.

(2) 分子イオンの横軸の値 (m/z) の「精度 (precision; 標準偏差 σ) と確度 (accuracy; 平均実測値と理論値との差)」を記述して, 確度は 2-3 σ の範囲にあることを示す.

になります. これまで, この範囲を超えた(= 確度が劣る) マススペクトルが MassBank にはいくつか見つかりました. そこで, これらの値を表記することを MassBank 提供の条件にすることを検討して

います。

縦軸（強度）については、まだ議論が煮詰まっていませんが、例えば（１）S/N 比がいくら以上のピークまたは、最大強度ピークの強度の 0.5 %以上のピークをデータとする、（２）最大強度のピークが検出感度の飽和状態に達していない試料濃度で測定したことを示す、というようなことになるとと思います。

適切なご質問をありがとうございました。

査読者 2 飯田真一（アルバック・ファイ）

本解説記事は、マススペクトルデータベース「MassBank」について、開発に至った背景から、コンセプト、データベースの内容等について詳細にまとめられており、JSA 誌に掲載の価値があると考えます。掲載に当たり、以下の点をご検討頂きますようお願い致します。

[査読者 2-1]MassBank のコンセプト

本文中に「タイプの異なる分析計で測定したスペクトルは、フラグメントイオンや質量分解能が異なるため、これが本格的なスペクトルライブラリの構築を制約していた。これを解決するため、測定者がデータと分析条件などを定められたレコード形式にして提供するようにした」とありますが、具体的にはスペクトル提供者はどのような形式で測定パラメータやスペクトルデータを提出しているのでしょうか？ 実際のデータ入力の画面例などがあればイメージが付きやすいかと考えます。御検討頂ければと思います。

[著者]

スペクトルの提供者が提供したピークデータとメタデータは、数十行からなる MassBank レコード（テキストファイル形式）として提出しています。これはそのまま MassBank レコードとして MassBank の web サイトから見る事ができます。それを見ていただくとピークデータとメタデータからなるレコードの全容を理解していただくことができます。論文の文中に実例の URL を記載しました。

<https://massbank.eu/MassBank/jsp/RecordDisplay.jsp?id=AC000107&dsn=AAFC>

Web 上ではマススペクトルの絵と分析をした化合物の化学構造の絵が表示されます。マススペクトルの絵はピークデータを読み込んで描画しています。閲覧する度に描画されます。ピークの上にカーソル

をのせると、 m/z と強度が数値で表示されます。拡大、縮小もできますが Internet Explorer でしか正しく作動しないと思います。化学構造の絵は別途入力した mol ファイルから生成して表示しています。画面の化学構造の絵をクリックすると、別ウィンドウに拡大した化学構造の絵が表示されます。MassBank は全て Java で記述され、GitHub から download できるはずで

MassBank に登録するとき提出するテキストファイルは、Internet Explorer などの Web browser から（許可された）提供者が自分で、MassBank サーバに登録するようになっています。これらのファイルを作成するレコードエディターやそれを利用したレコード作成法、入力画面の例などの詳細については、データ公開マニュアルの第 4 分冊「データ作成編」に詳しく説明がされていますのでそれをご覧ください。これらマニュアル類のダウンロードサイトの URL を本文の「8. 参考文献」[6]として新たに追加をいたしましたので、そちらをご覧ください。

[査読者 2-2]データ検索機能の充実

こちらにも具体的なスペクトル検索事例を載せて頂ければ理解しやすいかと考えます。

[著者]

本文「4. MassBank の危機と第 2 世代 MassBank の開発」で述べていますように、現在、多くの検索機能が利用できない状態にありますので、現在、利用できます機能 2 つだけを紹介します。

まず、ご質問 1 に対する回答でお答えしました URL にアクセスしていただきます。

レコード AC000107が表示されます。この表示の上から 2 行目に

Home | Quick Search | Peak Search | Record Index | Statistics | MassBank ID: Go

が表示されます。これが現在利用できる機能です。

Quick Search

Search by keyword は化合物名の一部（例：adenine）を入力すると、その化合物に関するレコードのリストが出力されます。

Search by Peak は入力したピークが観察されているマススペクトルが表示されます。

入力 box の下にあります Example1 または Example 2 ボタンをクリックすると例が入力されま

す.

いちばん下にあります Search ボタンをクリックすると検索がはじまります.

Peak Search

Quick Search の Search by Peak と機能は似ていますが, 次の違いがあります.

Quick Search では入力したピーク (例 4つのピーク) のうちいくつかを観察されているマススペクトルが表示されます. それに対して, Peak Search では入力したピークが全て観察されているマススペクトルが表示されます. 例: C6H12NO2

Peak Differences は2つのピークの m/z の差が入力した値になるマススペクトルが表示されます.

例: C6H12NO2

本来ですと, 2つのピークがマススペクトルの上に異なる色で表示されるはずですが, 未完成のようです.

Record Index

マススペクトルを提供している研究グループごとにレコードのリストが表示されます.

例: JEOL Ltd をクリックすると, 日本電子が提供した Spiral MS/MS 測定したマススペクトルが化合物ごとに表示されます. 右上の Open All Tree ボタンをクリックしていただきますと, レコードのリストが表示されます.

まだ, 完全にJavaの書き換えが終了していませんので, いずれの検索機能も完全な作動からはほど遠いです. あと1年ほど開発にかかるかもしれません.

[査読者 2-3]

また, 参考文献の欄に MassBank のマニュアル

<http://www.massbank.jp/ja/manual.html>

を追加してはいかがでしょうか?

[著者]

参考文献 [6] に MassBank のマニュアル <http://www.massbank.jp/ja/manual.html> を追加いたしました.

適切なお助言をいただきましてありがとうございました.