

一般社団法人日本環境化学会 質量分析インフォマティクス研究会 共催

「環境化学を加速する質量分析インフォマティクス」

日時： 平成 30 年 5 月 25 日（金） 9:00～12:35

場所： 沖縄県市町村自治会館 4F 第 6 議室

プログラム

8:45 受付開始

9:00 開場

9:10 1. 開会のあいさつ 早川 英介（沖縄科学技術大学院大学）

2. 口頭発表 <セッション 1> 座長：山本 淳史

9:15 演題 1 「GCxGC-ToFMS によるノンターゲット環境モニタリングの試みと課題」  
橋本 俊次（国立環境研究所）

9:35 演題 2 「LC/QTOF hybrid MS で環境試料ノンターゲット一括分析」  
山本 敦史（鳥取環境大学）

9:55 演題 3 「CASMI2016 から学ぶ環境化学物質の同定戦略」 津川 裕司（理研 CSRS・IMS）

10:15 演題 4 「環境化学にプロテオミクスは貢献できるか？」 吉沢 明康（京都大学）

10:35 <休憩>

3. 口頭発表 <セッション 2> 座長：橋本 俊次

10:50 演題 5 「ケモインフォマティクスとネットワークが魅せる未知化合物の世界」  
早川 英介（沖縄科学技術大学院大学）

11:10 演題 6 「質量分析データベースの世界」 奥田 修二郎（新潟大学）

11:30 演題 7 「構造推定の精度向上を目指した化合物データセットの選択」  
山本 博之（ヒューマン・メタボローム・テクノロジーズ株式会社）

11:50 演題 8 「Mass++ ver. 4: オープンソース質量分析データビューア開発の現状と今後の展望」  
田中 聡（Trans-IT）

12:10 4. パネルディスカッション：有志パネラーと会場参加者による自由討論  
司会：早川 英介、橋本 俊次

12:30 5. 閉会のあいさつ 柴田 康行（日本環境化学会会長）

12:50 閉場

## GCxGC-ToFMS によるノンターゲット環境モニタリングの試みと課題

橋本 俊次 (国立環境研究所)

化学物質による環境汚染問題がマスコミに取り上げられることは珍しくなり、あたかも我が国の環境は近代工業が発展する以前の状態に戻ったかのような錯覚に囚われがちであるが、インドや中国、ベトナムといった国々では環境汚染は深刻であり、PCB や DDT、HCH による汚染は地球規模で拡大し、すっかり定着してしまっているようである。それに加え、PFOS、PFOA、PBDE、HBCD、HCBd、塩素化パラフィンなどによる環境汚染が次々と報告され、トリクロロエチレンやテトラクロロエチレンなどは不検出となる地域を見つけることが困難なほどである。

ケミカルアブストラクトに登録された化学物質は優に1億を超え、日常に流通している化学物質は10万種とも100万種ともいわれているが、その実態は明らかではない。化学工業は少量多品種の生産モデルを加速しており、多様化する化学物質への対応は、環境化学分野においても最優先課題の一つと言える。一方、計測技術の発展も目覚ましく、高精細、高感度、高精度、高頻度の情報が以前より容易に得られるようになってきている。この多情報計測を環境化学分析に取り込むことで、多様化する化学物質への対応を急ぐ必要がある。我々は十数年来、多次元ガスクロマトグラフィ飛行時間型質量分析法 (GCxGC/ToFMS) の開発から環境試料分析への適用について研究を行ってきた。この技術を環境モニタリングに応用し、従来の対象物質数や時間的に限定的なモニタリングではカバーできない化学物質相の変化を検知可能な、化学物質の包括的かつ高精細な監視手法の開発に取り組んでいる。今回はその初段的な検討と課題について紹介する。

<内容>

- ・ 攪拌子型固相抽出-加熱脱着-GCxGC/ToFMS による少量河川水中のノンターゲット測定
- ・ 1件/日×6日分の試料を n=5 測定し、再現性 (測定法変動) と日間変動を統計解析した結果
- ・ 課題と今後の計画
- ・ 精密質量を活用した GCxGC/ToFMS ノンターゲット測定データからの情報抽出法の照会 (時間が許せば)

環境中で見いだされる生体リスク物質の分析に質量分析は欠くことのできない要素になっている。近年はこの分野の分析でもフーリエ変換型や飛行時間型質量分析計といった高分解能質量分析計 (HRMS) のユーザーが増えつつある。質量分解能, 精密質量の正確さ, 測定の高速度性, 微量イオンの検出能力等, HRMS の性能は年々向上している。このため, Suspect screening や Non-target analysis といった新しい手法が環境分析の分野でも可能となっている。Non-target analysis は検出されたイオンの精密質量, フラグメンテーション, 保持時間, 類縁物質の検出状況等から構造を読み取っていく。環境中に存在する商業的に生産されている化学物質は明確な発生源があるものが多いため, 十分な量があり質の高いデータが取れることが多い。しかしながら, HRMS のイオン化として良く用いられるエレクトロスプレーイオン化では単純にプロトン付加やプロトン脱離によりイオン化するだけでない。付加イオン形成や量が多い場合は二量体やクラスターイオンも生じるために, まずは観測されたイオンがどのようにイオン化されたものであるかから考える必要がある。今回の発表では SCIEX の四重極と飛行時間型のハイブリッド型質量分析計である X500R を用いて, 河川中に含まれる物質を Non-target 分析により見いだした例について報告する。Sequential window acquisition of all theoretical fragment-ion spectra (SWATH) mode は取りこぼし無く試料中に含まれる成分について, 解析に耐えうるデータを取るのに最も適した方法である。河川試料から検出されるイオンはその多くが商業的に生産されている化学物質に由来するものとして同定できた。解析の自動化をどのように行っていくか, また大気試料等に見られる低分子や二次生成物質をどのように扱うかが将来的な課題であり, インフォマティクスの研究者のお知恵を是非お借りしたい。

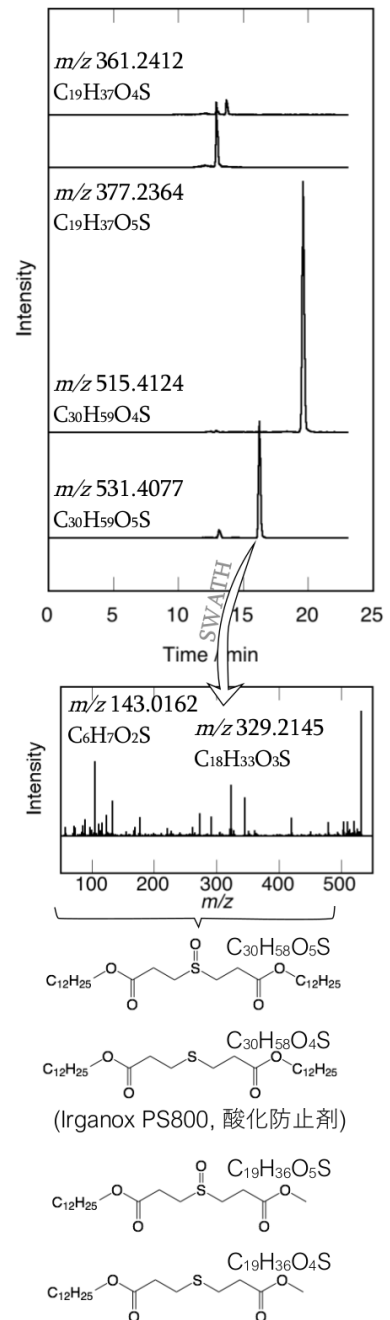


図 SWATH データを用いた河川試料中酸化防止剤とその類縁化合物の同定

## CASMI2016 から学ぶ環境化学物質の同定戦略

津川 裕司（理化学研究所環境資源科学研究センターメタボローム情報研究チーム・  
理化学研究所生命医科学研究センターメタボローム研究チーム）

質量分析は試料中の低分子化合物を包括的に捉えるために必須の分析装置である。中でも、液体クロマトグラフィータンデム質量分析（以下 LC-MS/MS）はメタボローム解析の対象である代謝産物だけでなく、天然物化合物の探索や薬物動態、更には環境化学物質の測定に頻用されている。本発表では、発表者がこれまで開発してきた質量分析インフォマティクスの技術を、CASMI2016 に出題された化合物構造推定問題を題材として紹介する。

CASMI とは、Critical Assessment of Small Molecule Identification の略であり、2012 年より続いている「マススペクトルから構造予測」を行う大会のようなものである。2016 年からは出題方式が大きく変わり、これまで 10–20 題だった問題数が 200–300 題のマススペクトルが出題され、「構造推定プログラムの機能を競う」趣旨も盛り込まれてきた。発表者は、2016 年と 2017 年の両方に参加したが、2016 年の出題された化合物が主に環境化学物質より出されたものであった。

ここで、マススペクトルから化合物構造を推定する大まかな流れは、

1. 着目するプリカーサーイオンがインソースフラグメントや同位体イオンではないかという判定を行う。
2. ポジティブイオンモードとネガティブイオンモードのデータを見比べてプリカーサーイオンのアダクトタイプ ( $[M+H]^+$ ,  $[M+Na]^+$ ,  $[M+NH_4]^+$  など) の決定を行う。
3. (構造推定プログラムを使いながら) 精密質量, 同位体比, MS/MS スペクトルの情報や原子価則等を用いて組成式の決定を行う。
4. 予測組成式と合致する構造をデータベースから探索する。
5. 得られた構造式候補の中から (構造推定プログラムを駆使し) 最も妥当性の高い化合物構造を提案する。

という流れとなるが、上記 5 つの行程でも重要なのは構造式を検索するデータベースの内容である。たとえば、環境化学物質を対象とするならば、水中に含まれる有機化合物や水中微生物により代謝された化合物、そして既知もしくは予測環境化学物質を集積したデータベースである STOFF-IDENT を用いるべきである。また、ハロゲンの中でも Cl や Br に特徴的な同位体パターンに注意しながら構造推定プログラムに供することが重要である。2016 年の CASMI では、この STOFF-IDENT をデータベースに用い、構造推定プログラムとして MS-FINDER を用いたチームが 208 問中 159 問を正解するという結果となった。本発表では、上記のことを踏まえながら「マススペクトルから標準品化合物が無い既知の環境化学物質を同定するための方法」を議論したい。

## プロテオミクスは環境化学に貢献できるか？

吉沢 明康（京都大学大学院薬学研究科）

タンパク質は細胞が生産するため、環境中に直接存在する物質という印象は弱い。しかし実際には、多くの細菌が（生理活性のある）ペプチドを菌体外に分泌することや、個体から剥落したタンパク質断片（ペプチド）が環境中に存在することが知られている。

このようなペプチドは、それ自体が本来の目的を超えて周囲に影響を与えている可能性がある。また測定という面からも、環境試料の質量分析法を用いた測定に於いて、このようなペプチドが化合物測定上での夾雑物となる可能性が高い。

一般に、質量分析計で得られた情報からタンパク質を同定する手法は、タンパク質のアミノ酸配列が遺伝子としてゲノム中にコードされていることをフル活用しているため、他の物質同定の手法、例えば化合物の検出方法とは異なっている。本講演では、タンパク質の同定方法について概観しながら、タンパク質に比べてより困難であるペプチドの同定方法や、ゲノム未知の生物種に由来する混合試料の同定方法など、最新的话题を交えて紹介する。

## ケモインフォマティクスとネットワークが魅せる未知化合物の世界

早川 英介（沖縄科学技術大学院大学）

世の中に存在する膨大な種類の化合物を定性的に分析することは、質量分析のもっとも重要な役割のひとつである。液体クロマトグラフィーと連携した高感度質量分析装置により膨大な数の質量スペクトル情報を得ることが容易になった一方で、質量ピークとして検出した物質、特に低分子化合物の同定はいまだに困難な場合が多い。質量分析による定性分析では通常化合物を開裂したフラグメントイオンのスペクトルを解析に用いることが多いが、低分子化合物では開裂パターンの予測が困難であることや、得られるフラグメントイオンの数が少ないということが長年ボトルネックとなってきた。

現在までに低分子化合物の定性分析において質量分析データを様々なインフォマティクスの手法により解析するアプローチが多数報告されている。近年、膨大な質量分析データをネットワーク的アプローチで解析する手法が複数報告されている。ネットワーク解析はデータの分類、階層化、可視化など多くの利点があり、特に大量の未知化合物のスペクトルデータの解析に適している。さらにケモインフォマティクスや他の情報科学との連携により、従来とは全く異なったアプローチでの定性分析のためのデータ解析が可能が示されてきている。本発表では低分子化合物にフォーカスしたネットワークとケモインフォマティクスを活用したデータ解析分野の研究を紹介し、さらに我々の最近の取り組みも紹介したい。

## 質量分析データベースの世界

奥田 修二郎（新潟大学大学院医歯学総合研究科）

質量分析法が利用される分野として、プロテオミクスやメタボロミクス等のオミクス分野がある。タンパク質、糖鎖、脂質、低分子化合物と言った構造分類の違いは関係なくイオン化される分子であれば質量分析の対象となるため、一口に質量分析と言う場合にはあらゆる意味を包含している。そのため、質量分析の解析手法は、対象とする分子の種類、機器の違い、サンプル調製などあらゆる条件に応じた最適化が実施されている。質量分析のデータベースには、これらの多様な条件から得られる情報をまとめたものであることが期待されるが、現状ではそれぞれ別のデータベースとして情報がまとめられている。ここでは、質量分析に関するデータベースの現状を概観した後、プロテオミクスに焦点を当て、質量分析データベース構築における課題について議論したい。

## 構造推定の精度向上を目指した化合物データセットの選択

○山本 博之 (ヒューマン・メタボローム・テクノロジーズ株式会社 研究開発本部)  
小寺 正明 (東京大学 工学系研究科)

低分子の代謝物を包括的に解析する研究分野であるメタボロミクスにおいて、質量分析データの解析によって得られたピークの中には、代謝物名が付与されるものだけでなく、化学構造が未知のピークが数多く含まれるのが現状である。このような背景から、ケモインフォマティクスと機械学習を用いた未知ピークの構造推定法が様々提案されており、さらに質量分析から得られる限られた情報から構造推定を行い、その正解率を競う CASMI コンテストが実施され、構造推定法の精度が競われている。

代謝物の構造推定の手順は、ある特定の分子式を持つ代謝物群(化合物データセット)をデータベースから検索し、それらの代謝物群の化学構造からコンピューター上で結合の開裂を行って得られた理論 MS/MS スペクトルと、実測値を照合することで行われる。構造推定の精度を向上させるためのポイントとしては、適切な化合物データセットの選択、もしくは結合開裂の計算精度が考えられるが、我々は前者の化合物データセットの選択に着目した。

構造推定における化合物データベースとして、網羅的な化合物データベースである PubChem が一般的に用いられるが、例えばヒト由来サンプル中の未知ピークの構造推定を行う際には、ヒト由来代謝物のデータベースである Human Metabolome DataBase(HMDB)を用いることで推定精度が改善することが報告されている。しかしながら、現時点での HMDB はヒト由来代謝物が全て網羅されているとはいえ、構造候補がヒットしないことがある。

そこで我々は、HMDB に登録されていないヒト由来化合物候補を PubChem から取得することで、構造推定の精度向上の可能性について検討した。具体的には、構造類似度を利用したナイーブな方法、主成分分析、機械学習の一手法であるランダムフォレストを適用し、化合物データセットの選択方法について検討を行った。

本発表では、ヒト由来サンプル中の代謝物の構造推定を目的としたが、環境サンプル中の未知ピークの構造推定を行う際には、環境化学物質のデータベースを利用することで、本手法をそのまま環境化学分野での構造推定に適用可能である。最後に、本研究は 2017 年度の質量分析インフォマティクス研究会のハッカソン、国内版バイオハッカソンにおいて、東京大学の小寺正明氏と行った共同研究の成果である。



## Mass++ ver. 4: オープンソース質量分析データビューア開発の現状と今後の展望

田中 聡 (Trans-IT [トランジット])

質量分析においてソフトウェアは、機器制御、データ管理、自動化、解析、表示といった役割を持つが、特に機器により得られたデータを効率的、効果的に解析するには洗練されたソフトウェアが必要である。そこで、我々は 2006 年よりエーザイ(株) に拠点を置く CREST プロジェクトにより、質量分析データ解析・表示ソフトウェア Mass++ を開発した。

Mass++ は、複数ベンダー対応、プラグイン構造、豊富なビューアや解析機能といった 3 つの特徴を持つ。

Mass++ は、2010 年より開発拠点を FIRST プロジェクト (株式会社島津製作所 内) に拠点を移したが、プロジェクト終了後 2015 年 3 月にオープンソースソフトウェアとして公開し、以後は有志による開発メンテナンスが続けられていた。

2006 年に最初の Mass++ を開発してから、現在まで約 12 年の月日が経っているが、質量分析におけるソフトウェアにおける現状も当時と比べていろいろと変化している。

例えば、当初の Mass++ はいろいろなベンダーのデータを読み込めるようになっていたが、現在においては ProteoWizard の様なソフトウェアが普及して、さまざまなデータを一括で mzML の様な共通フォーマットに変換できる様になっている。

また、MaxQuant や OpenMS, MZmine 等といった優れたオープンソースソフトウェア、フリーウェアというのでも普及する様になってきた。この様な状況においては、マルチベンダー対応や、豊富な機能というのは開発コストの割に大きな恩恵は得にくくなっている。従って、我々は今回方針を大きく転換し、マルチベンダー対応ではなく共通フォーマット mzXML/mzML のみに対応。また、いろいろな事ができるソフトウェアではなく、解析機能は他のオープンソースソフトウェアやフリーウェアに任せ Mass++ はその効果的な表示やビューア つまり可視化に集中する事にして、再度 0 から開発しなおす事とした。

今回は、そういった新たな方針で開発している Mass++ の開発の現状と今後の展望について説明する。