

厚生労働科学研究費補助金（食品の安全確保推進研究事業）
「新たなバイオテクノロジーを用いて得られた食品の安全性確保と
リスクコミュニケーションのための研究」
分担研究報告書（令和2年度）

質量分析インフォマティクスによる未知化合物データ解析

研究分担者 早川英介（沖縄科学技術大学院大学）

研究要旨：

本年度は従来の標準品に依存したスペクトルライブラリに加え、食品・植物などから内在性の化合物を組織的に抽出した試料を LC-MS で分析し、構造上情報を付与することでライブラリの大幅な拡張を行った。一方のデータ解析システムでは従来の化合物クラス情報に加え、毒性情報を統合することで、未知化合物の毒性予測にもつながり得る新規の解析システムを作成した。さらにこのデータ解析システムを広範囲の研究者が使用できるように Web ベースのアプリケーションとして GUI 等の整備を行った。

A. 研究目的

ゲノム編集作物はその作成過程で想定外の代謝の変化とそれによる未知の化合物や毒性の増加など安全性の問題が危惧されている。食品分析における質量分析等の従来の分析化学では特定の化合物のみを分析する“ターゲット分析”が主流であるが、ゲノム編集で危惧される未知化合物等を検出することはできない。そこで本研究では液体クロマトグラフィーと質量分析をもちいたハイスループットな分析データから未知化合物の迅速な検出と構造推定を可能にするデータ解析プラットフォームを開発している。令和2年度では前年度までに構築したデータ解析システムの拡張と汎用化を目的とした。

B. 研究方法

本研究では試料中の化合物の質量スペクトル（フラグメントスペクトル）の類似度をもとに食品中の未知化合物の検出と構造推定を行うシステムの拡張を行った。スペクトル類似度計算および部分構造推定に用いるフラグメントスペクトルライブラリとしては、パブリックなスペクトルライブラリに加え、in house で構築した標準物質のスペクトルライブラリと食品試料や植物体から抽出した試料由来のスペクトルをライブラリに加えた（図1）。

スペクトルライブラリに含まれる全化合物に関して化合物クラスアノテーションアルゴリズム Cclassyfire を適用することで、化合物クラス情報をライブラリに付与した。さらに、各化合物のフラグメントスペクトルのピークに関して、ケモインフォマティクスツールキット RDKit で網羅的に共有結合を切断した仮想フラグメント構造のアサインを行った。毒性情報に関しては、網羅的な毒性物質データベースである T3DB を XML 解析することでスペクトルライブラリ内の化合物に毒性および毒性のカテゴリ情報の付与を行った（図2）。

データ解析の実行環境として各種パラメータをインタラクティブに反映させることが可能な GUI を、Plotly Dash を利用して Web ブラウザ上で実行可能なシステムとして実装した。

C. 研究結果

本年度では、前年度までに構築した基本的なデータ解析システムを拡充することで、未知化合物に関する豊富な情報を迅速に出力するシステムへと発展させることができた。食品から抽出した未知化合物を含む試料の LC-MS データを本解析プラットフォームで解析した場合、未知化合物であっても「化合物クラス・構造的に類似した化合物・部分構造・構造類似性に基づく毒性」の情報が自動的に得られる。未知化合物の LC-MS デー

タからこのような豊富な情報を見出すデータ解析系はこれまでに存在せず、本データ解析プラットフォームは極めて独自性の高いものと言える。

一般的に高度な質量分析データ解析アルゴリズムはインストールや操作が難解なものが多く、非インフォマティシヤンの一般の分析化学の研究者に実際に利用されることが少ないというケースが多々あった。それに対して本研究のデータ解析プラットフォームは Web ブラウザ上で GUI によりインタラクティブに操作が可能となり、より広範囲の研究者に使用されることが期待される (図 3)。

D. 考察

本年度はスペクトル類似度による構造類似性と化合物クラスを見出すという昨年までの基本的なデータの解釈に加え、毒性や部分構造の情報を付加するより高度で豊富な化合物情報を提示する機能を加えた。毒性に関しては構造類似性が毒性に直接反映されないケースもあるが、リスクの迅速な検出という意味では有効であると考えられる。また、スペクトル類似度に関連した部分構造を得る機能に関しては、部分構造から候補構造を構造データベースから見出すための情報として用いることや、全く新規の構造推定する際に構造創出アルゴリズムのシードにするなど、他のケモインフォマティクスのツールと連携させることでこれまでにない高度な構造推定に繋がる。

このような煩雑なデータ解析は実際の分析を行っている研究者・技術者は扱うことが困難であり、アルゴリズムとして存在していても使用されることがないというケースも想定された。それに対し、本研究では Web ブラウザ上で解析に必要なパラメータを入力してインタラクティブで容易な GUI を実装した。現在は配布可能な python ファイルであるが、サーバーを設置してオンラインで利用可能なプラットフォームとすることで、広範囲な研究者が容易に本データ解析システムを利用可能になることが期待される。

E. 研究発表・業績

1. 論文発表

該当なし

2. 学会発表

- 1) 日本食品化学学会 第26回総会・学術大会 (2020.5.28)、書面開催「多層質量スペクトルネットワークによる食品中の未知化合物の可視化」
- 2) 日本毒性学会 第47回日本毒性学会学術年会 (2020.6.29)、オンライン開催「An analytical framework using multilayer mass spectral similarity network to detect unknown compounds for toxicological analysis.」

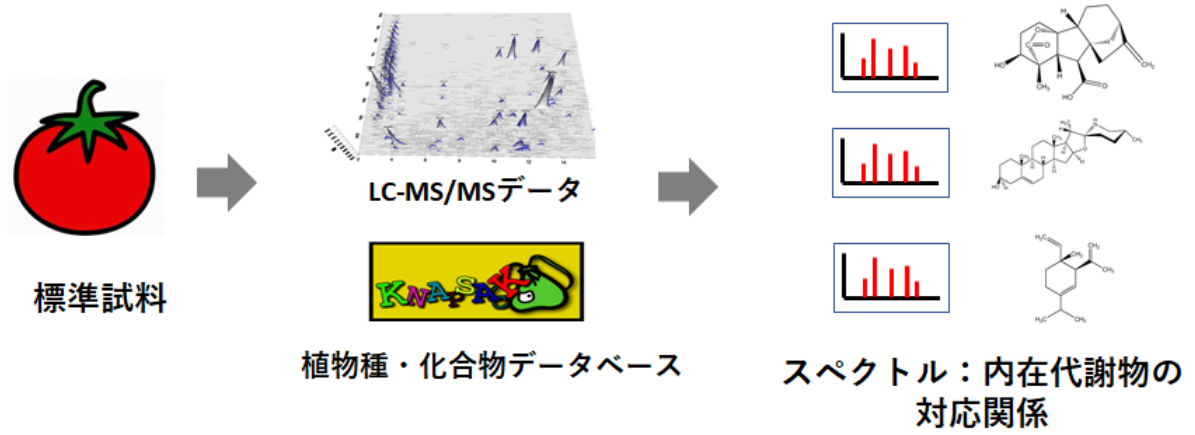


図 1

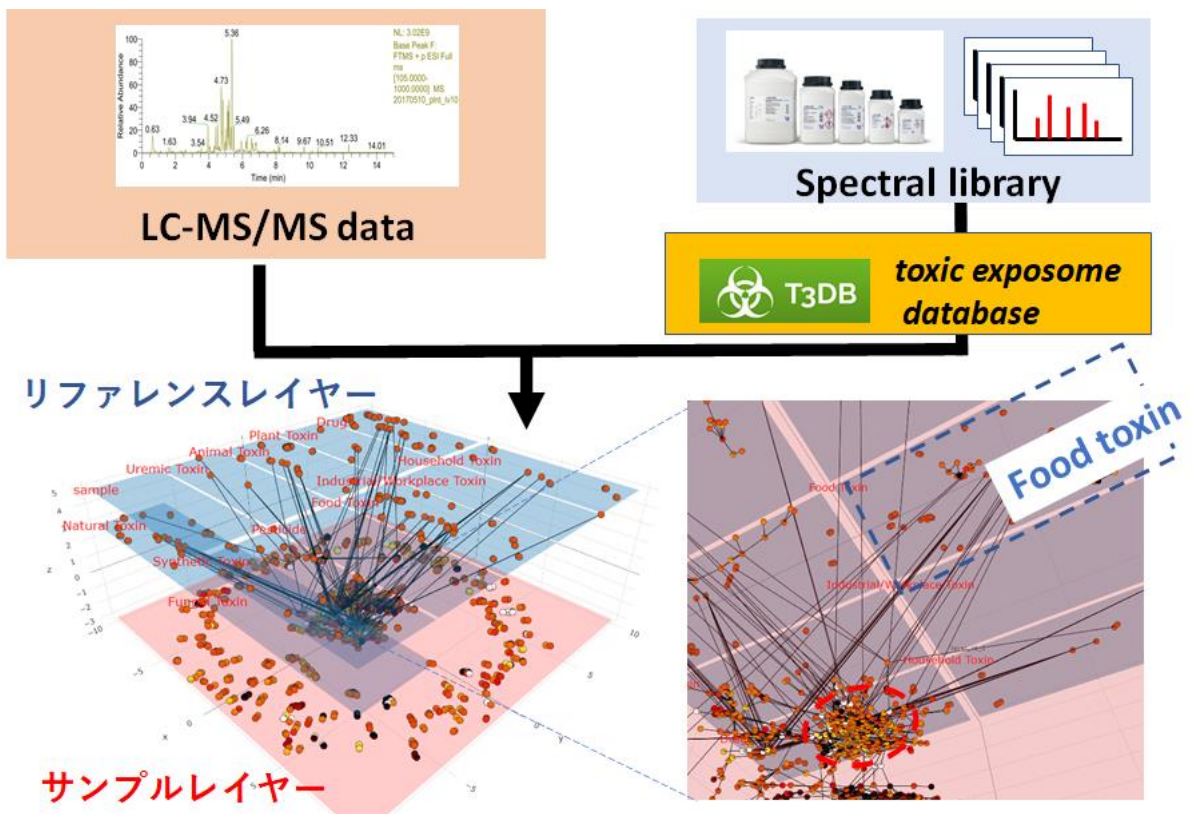


図 2

