

メタボロームインフォマティクスによる未知化合物推定

研究分担者 早川英介（沖縄科学技術大学院大学）

研究要旨：

本研究は食品試料の内在性低分子化合物の質的な変動を迅速に検出するデータ解析システムの開発を目的とし、質量スペクトルの類似度をもとにした食品中の未知化合物の検出と構造推定のためのデータ解析システムの開発を行った。スペクトルライブラリ構築、アルゴリズム開発、可視化法の開発を通して、スペクトル類似度計算を核とした迅速な未知化合物の検出と構造推定などが行われるシステムを構築した。

A. 研究目的

本研究はゲノム編集作物のバイオテクノロジーにより作成された食品試料の内在性低分子化合物の質的な変動を迅速に検出するデータ解析システムの開発である。

従来の分析化学ではあらかじめ分析対象の化合物を決めて行う“ターゲット分析”が行われることがほとんどであった。しかし、ゲノム編集食品のように想定外の代謝の変化により未知化合物が懸念されている場合、そのような“ターゲット外”の化合物は従来の化学分析では見過ごされてしまう。本研究では構造的に類似した化合物の質量スペクトル（フラグメントスペクトル）が類似するという特性に着目し、質量スペクトルライブラリやケモインフォマティクスによる化学情報を統合したこれまでにないデータ解析プラットフォームを構築し、食品試料中の未知化合物の迅速な検出と構造推定を可能にすることを目的とした。（図1）

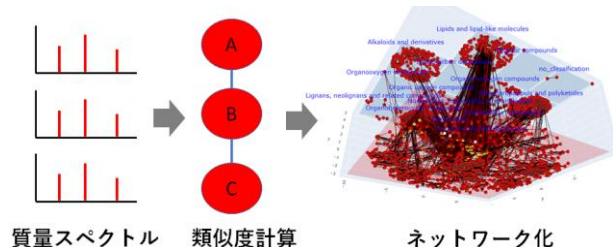


図1

B. 研究方法

本研究では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構築を行った。

データ解析のフローとしては、LC-MS から得られる比較定量情報を伴った質量スペクトルを入力とする。安全と考えられる食品試料（例：非ゲノム編集体）と分析対象の試料（例：ゲノム編集体）の比較定量により、注目している試料で濃度が増加している未知化合物を見出すことができる。

さらに、検出された全質量スペクトルに関して、試料内および質量スペクトルライブラリの全スペクトルに対して総当たりでスペクトル類似度計算を行うことで、構造的に類似している化合物を見出す。この質量スペクトルライブラリに関してはパブリックな質量スペクトルライブラリに加え、in houseで作成した化合物標準品と、内在性化合物情報が既知である代表的な食品・植物の抽出物を分析したことで得たライブラリを統合したものを構築した。

ライブラリ内の全化合物に関して、Classyfireにより化合物クラスをsuper class、class、subclassの各階層で付与し、さらに外部毒性情報データベースから抽出した情報を付与した。各化合物の構造とフラグメントスペクトルを解析し、フラグメントピークに相当する部分構造を推定し、データベースに統合した。（図2）

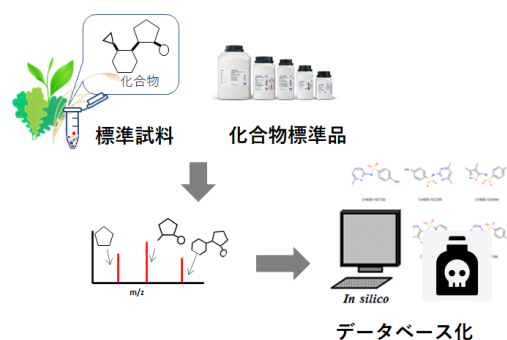


図2

スペクトル類似度計算とこの統合質量スペクトルライブラリを連携することで、試料中の未知化合物の質量スペクトルから化合物クラス、毒性情報、部分構造を得るシステムを構築した。このデータ解析システムの実行環境として、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

本研究では「B. 研究方法」で述べたデータ解析ワークフローに関し、プログラミング言語：Pythonを用い、ケモインフォマティクスライブラリ：RDKit、データ可視化ライブラリ：Plotly等と連携させることでデータ解析環境の構築を行った。入力データは一般に広く用いられる質量分析データ処理フリーソフト（Mzmine2）の出力形式に対応させた。

フラグメントスペクトルライブラリはMassBankに加え、天然化合物を多く含む GNPS、Respectを加えて特に植物二次代謝物に対応できるよう整備した。加えて独自で500種類の低分子化合物を分析しライブラリに追加した。さらに代表的な食品・モデル植物等（ダイズ、トマト、ジャガイモなど）50種類以上から抽出した低分子化合物に関してもスペクトルライブラリに追加した。

この膨大なスペクトルライブラリに対して、「B. 研究方法」で述べた化合物クラス・ピーク部分構造・毒性情報を付与することで、網羅的かつ極めて情報が豊富な独自性の高いスペクトルライブラリが構築された。スペクトル類似度計算とこのライブラリをデータ解析プラットフォーム上で統合することで、スペクトル類似度に基づいた未知化合物の構造解析を迅速に行うシステムが実現された。

さらに、このデータ解析システムをWebブラウザベースの実行環境で動作させることで、プログラミングなどの専門知識がない一般的な研究者でも容易にデータ解析とデータ可視化ができるようになった。

D. 考察

本研究の目的はゲノム編集作物等で想定外の質・構造的変化を生じる化合物の迅速な検出と構造の推定である。従来の質量分析のデータ解析では困難だった未知化合物の検出と構造推定をスペクトル類似性に着目した独自の手法を採用することで実現した。

手法の特徴として、試料中の未知化合物に関して化合物クラスや部分構造、毒性情報など多面的な情報を出力するという点がある。これは単に構造を推定するという点以外にも、変動した化合物“群”の

クラスや共通構造を見出すことであり、分析対象の食品の未知化合物グループを包括的に明らかにすることにもつながるなど、応用範囲は広い。単に試料中の未知化合物を明らかにするだけではなく、変動した代謝パスウェイの推定など、様々な利用法も考えられる。

本解析プラットフォームのもう一つの特徴は、上記のような複雑なデータ解析を容易に行うことが可能という点である。Webブラウザ経由でインタラクティブに出力結果をみながら操作を行うことで、データ解析に不慣れな研究者も操作が可能であり、幅広く普及されることが期待できる。（図3）



図3

E. 結論

本研究により、様々な食品試料の内在性化合物の質的・量的な変動を迅速に解析できるプラットフォームが確立された。この手法はゲノム編集作物のバイオテクノロジーにより作成された食品試料中の未知化合物の分析に極めて有効であり、化合物クラス、部分構造、毒性予測などの豊富な化学情報は、単一の化合物の構造解析にとどまらず、未知化合物群とその代謝パスウェイを包括的に理解することにつながる。

本データ解析システムは一般の研究者でも容易に利用可能なプラットフォームであり、広範囲の食品分析に今後利用されることが期待される。

F. 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.」

G. 健康危険情報

各年度の分担研究報告書を参照

H. 知的財産権の出願・登録状況

各年度の分担研究報告書を参照