

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」

質量分析インフォマティクスによる化合物同定 ～解析フレームワークの確立・スペクトルライブラリの拡充～

研究分担者 早川 英介 （沖縄科学技術大学院大学）

研究要旨：

本研究では化合物の質量スペクトルの類似度をもとに試料中の未知化合物の検出と構造推定を行う解析システムの構築を行う。本年度は既存のスペクトルライブラリに加え標準品・標準試料から大量の質量スペクトルライブラリを取得すること統合スペクトルライブラリの拡充を行うとともに、比較定量情報を反映した可視化機能の実装など解析ツールの高機能化を行った。本年度の開発より試料間比較による代謝物の変動の検出と連携した、広範囲の化合物のクラスの迅速な推定・可視化という解析フレームワークの根幹が確立されたと言える。

A. 研究目的

ゲノム編集作物はターゲット遺伝子以外にも想定外の代謝への影響など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。従来の質量分析による分析ではターゲットを特定の化合物に絞った“ターゲット分析”が一般的であるが、想定外の代謝経路の変動やその結果生じる未知の化合物を分析するためには有効な手法ではない。近年、分析ターゲットを限定しない“ノンターゲット分析”が提唱されているものの、未知の化合物を迅速に検出・構造推定することは未だに困難である。そこで本研究では、化合物の質量スペクトルの類似性に着目したデータ解析法により、「想定外の未知化合物」の迅速な検出と構造推定を可能にすることを目的としている。

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。データ解析のワークフローとしては安全と考えられる食品試料（非ゲノム編集体）と分析対象の試料（ゲノム編集体等）の液体クロマトグラフィー質量分析データから得られる化合物イオンの比較定量値およびフラグメントスペクトルを解析データとして用いる。

試料から検出された化合物の構造・クラス推定のために、標準物質のフラグメントスペクトルの組織的な取得を行った。一般に公開されているフ

ラグメントスペクトルデータベースとともに、標準物質を我々が実際に分析したデータ、さらに食品試料から抽出した代謝物のスペクトルデータも追加した統合フラグメントスペクトルデータベースを構築した。

試料由来のデータと統合フラグメントスペクトルデータベースを用い、網羅的にフラグメントスペクトル類似度を計算することで化合物の類似度を反映したネットワーク構造を創出するデータ解析フレームワークの構築を行った。（図1）

C. 研究成果

本研究では「B. 研究方法」で述べたデータ解析フレームワークに関し、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。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)からのフラグメントスペクトルおよび比較定量データのインポートを実装し、より広範囲の研究者が容易に利用可能になるデータ解析環境を実現した。

フラグメントスペクトルデータベースとしては従来のMassBankに加え、天然物に特化したGNPS、理研ReSpecライブラリを追加し、各化合物の化合物クラス情報等構造情報を付加したうえで統合データベース化した。さらに、独自で代謝物の標準品300種類をQ-exactive HF (Thermo Fischer

Science) で分析し in house ライブラリを作成した。

さらに、代表的な食品・モデル植物 30 種(大豆、トマト、ジャガイモ等) に関しては低分子化合物の抽出と Q-exactive HF によるスペクトルデータ取得を行った。生物種-代謝物関係データベース (KNAPsACK) から試料ごとの代謝物情報を抽出し、スペクトルデータと照合することで標準試料由来のスペクトルデータベースを構築した。

この大規模なフラグメントスペクトルライブラリにより、全化合物スーパークラスを網羅し、広範囲の化合物クラスとその類似物質の検出が可能となった。(図 2) この標準物質及び標準試料の質量スペクトルデータ取得は今後も継続することで、スペクトルデータベースの拡充を図る。

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。本年度では解析フレームワークの根幹の確立のためにフラグメントスペクトルライブラリの拡充にとくに力を入れることで、広範囲な化合物クラスの検出が可能となった。さらに質量分析定量データ解析ツール Mzmine2 と連携することで、試料間での代謝物の量的変動と統合した上でのデータ可視化が可能となり、他に類をみないユニークな解析ワークフローが確立できた。

検出された化合物が未知の化合物であることを想定した場合、実際の構造推定を行う上で構造類似性やフラグメントの構造の推定など、より詳細な構造情報をケモインフォマティクスとの連携により実現する必要がある、これは次年度の課題のひとつである。

また、本年度では一般的に広く使われている質量分析データ解析ツール Mzmine からの出力ファイルを読み込む機能を実装したことで、ユーザーが利用する上での技術的ハードルは低くなった。一方でデータの可視化と化合物クラス推定等の部分は Python スクリプトの編集が必要となるなど、現状では一般ユーザーには困難な部分が多い。次年度では GUI を実装するなど、より容易に利用可能なツールとして確立する予定である。

E. 結論

本年度の研究により、試料間比較による代謝物の変動と連携した化合物のクラス推定・可視化と

いう解析フレームワークの根幹が確立されたと言える。本解析法では標準物質のフラグメントスペクトルライブラリが量・質的にも重要であり、今後も継続して拡充を進めていく。次年度の課題としては未知化合物の構造推定のための厳密な構造推定のためのアルゴリズムの先鋭化と利用が容易になるユーザーインターフェースの開発である。

F. 健康危険情報

なし

G. 業績

1. 論文発表

なし

2. 学会発表・講演

なし

H. 知的財産権の出願・登録状況

該当なし

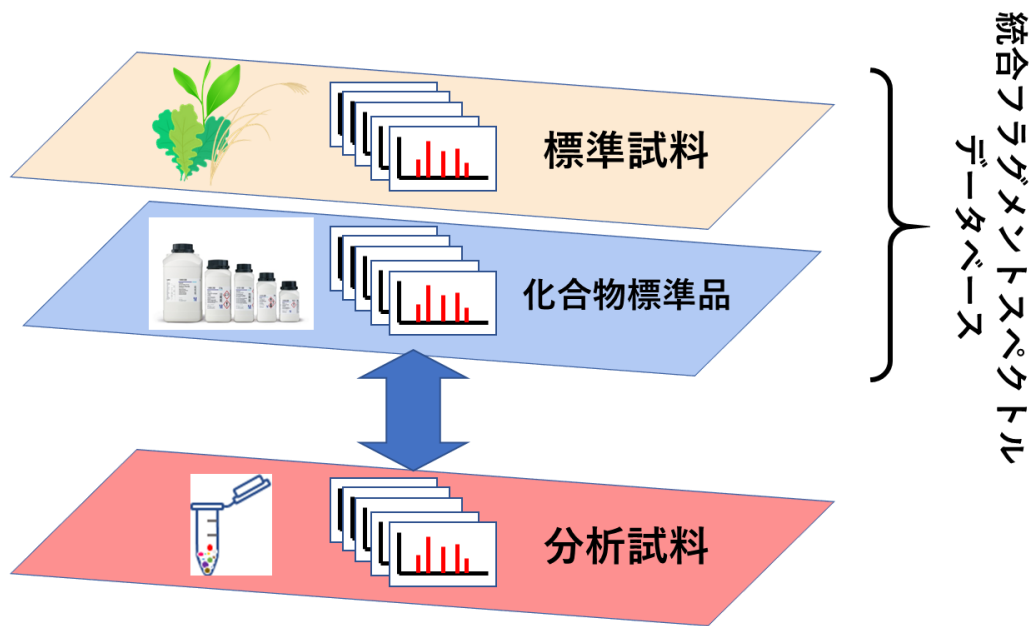


図1 解析データ構造の模式図

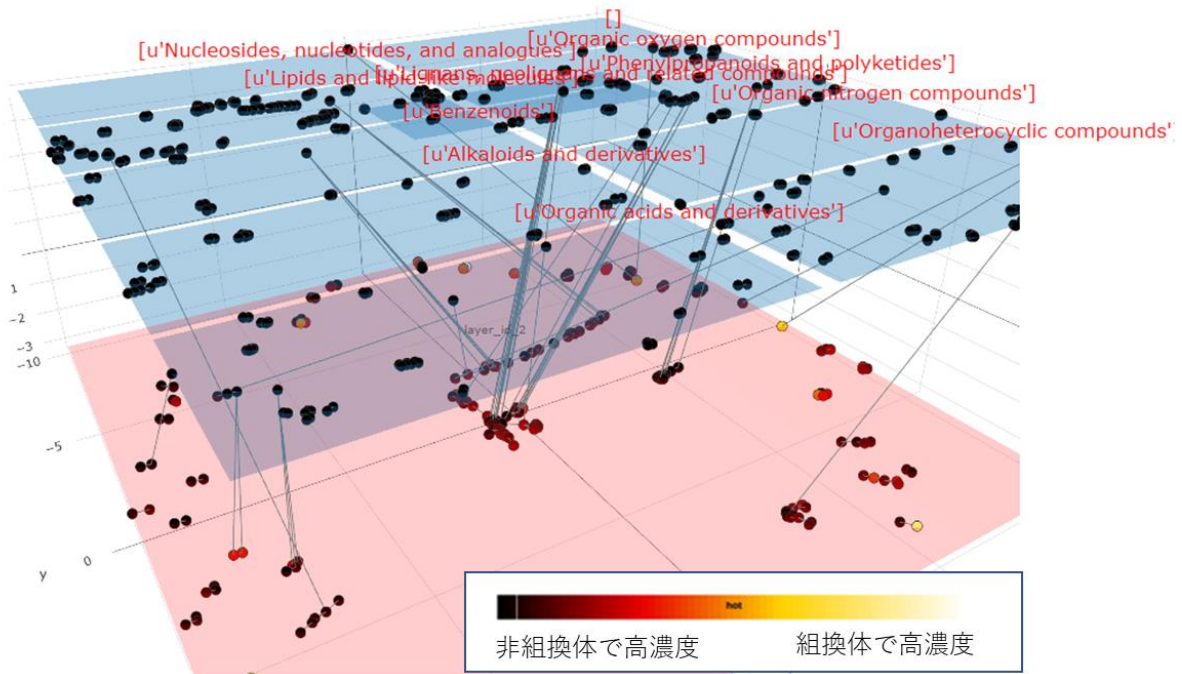


図2 標準物質の大豆の組み換え体・非組み換え体の比較定量（下層）と標準物質のフラグメントスペクトル（上層）