



Title	GC/MSメタボロミクスにおける未同定代謝物同定の効率化および品質向上に向けた保持指標予測モデルの構築とその応用
Author(s)	松尾, 晃子
Citation	大阪大学, 2018, 博士論文
Version Type	VoR
URL	https://doi.org/10.18910/69533
rights	
Note	

The University of Osaka Institutional Knowledge Archive : OUKA

<https://ir.library.osaka-u.ac.jp/>

The University of Osaka

論文内容の要旨

氏名 (松尾 晃子)

論文題名

GC/MSメタボロミクスにおける未同定代謝物同定の効率化および品質向上に向けた保持指標予測モデルの構築とその応用

論文内容の要旨

【第一章】ノンターゲットメタボロミクスは生体サンプル中の代謝物を一斉に測定・解析する、オミクス解析の一種である。この手法では測定対象とする代謝物を事前に決定しないため、代謝物同定が重要な手順となる。GC/MS (gas chromatography/mass spectrometry) は堅牢かつ高感度な分析法であり、サンプルの測定で数百から千近くのピークが得られる。代謝物同定では標準物質の測定結果を格納したデータベース (以下、ライブラリと呼ぶ) から、候補化合物 (同定したいピークと類似した測定結果を示す代謝物) を探索し、純粋な標準品によるバリデーションを行う。GC/MSは、各ピークの保持指標 (基準物質によって標準化された保持時間) とEIスペクトル (electron ionization) に基づいて代謝物を同定する。EIスペクトルの再現性は高く、ライブラリには数十万種のスペクトルが登録されている。一方、保持指標ライブラリの登録数は一万種ほどにとどまっている。EIスペクトルの類似性のみに基づいた代謝物同定では、多くの偽陽性の候補化合物が現れることが知られている。測定で得られる数百のピークに対し、それぞれの候補化合物の標準品を揃え、測定するには多大な労力が必要となる。以上の現状を踏まえ、同定すべきピークの選択法の確立と、候補化合物の絞り込みへの保持指標情報の活用が必要であると考えた。

本研究ではQC (quality control) サンプルの希釈系列に基づくピークピッキング法と、保持指標予測モデルの構築および代謝物同定への活用を提案し、これらの組み合わせによる代謝物同定のスループットの向上を試みた。

【第二章】約四百種の親水性代謝物の誘導体化合物が登録されたライブラリから、保持指標予測モデルを構築した。まず化学構造の数量的な表現である、記述因子を計算した。ライブラリに登録されている保持指標との相関係数 (R^2) が0.8以上となった記述因子を説明変数、純粋な標準品による保持指標の実測値を目的変数として、重回帰分析を行った。ライブラリに登録されている代謝物を保持指標の分布が等しくなるように二群に分け、それぞれトレーニングセット、テストセットとした。その結果、以下の保持指標予測式を得た。

$$\text{Predicted RI} = -495.2 \cdot \text{ATSc1} + 29.8 \cdot \text{topoDiameter} + 101.4 \cdot \text{MLFER_L} + 27.9 \cdot \text{ETA_Beta} + 388.7$$

Predicted RIは保持指標の予測値を示す。記述因子のATSc1は代謝物内の繰り返し構造、topoDiameterは分子の大きさ、MLFER_Lは固定相と移動相の相関、ETA_Betaは分子内の電荷の分布を表している。予測された保持指標値と実測値の R^2 は0.9となり、高い直線性を示した。以上の結果から、EIスペクトルの類似性に基づいた候補化合物へのさらなる絞り込みに活用可能な、保持指標予測モデルの構築に成功したと結論付けた。

【第三章】本章では、QCサンプルによるピーク選択と第二章で構築した保持指標予測モデルを、生薬の一種である川芎 (*Conium officinale Makino*, *Ligusticum chuanxiong* Hort) に適用した。全六種類の川芎を等量比で混合したものをQCサンプルとし、希釈系列を作製した。これをサンプルと同じ分析バッチで測定し、ピークごとに検量線を作成した。代謝物に由来するピークの強度値は希釈系列に沿って増加すると考えられることから、検量線の直線性 ($R^2 > 0.9$) を基準としたピークピッキングを行った。この結果、クロマトグラム上の1974ピークから457ピークが選択された。

未同定ピークの定性では、まず類似したEIスペクトル (Similarity > 0.8) を持つ候補化合物を検索した。次にこれらの候補化合物の保持指標値を第二章の式によって計算した。未同定ピークの保持指標値 (実測値) と、候補化合物の保持指標値 (予測値) の差分が100未満となった候補化合物を選び出した。標準品によるバリデーションを経て、代謝物を同定した。Butane-1,2,3-triolの定性では、EIスペクトルから得られた45種類の候補化合物が予測モデルによって6種類へ絞り込まれた。3-Deoxy glucosoneでは5種類から2種類へ、Palatinitolでは63種類から6種類に、候補化合物が削減された。以上の結果から、標準品による確認が必要となる候補化合物を大幅に削減することに成功したと結論付けた。

【第四章】本研究では未同定代謝物の効率的な同定法に向けた、保持指標予測モデルの構築とQCサンプルによるピークピッキングを提案した。さらに生体サンプルへこの戦略を応用し、生薬川芎に含まれる3種類の未同定代謝物を新たに定性した。本手法によって従来、必要とされていた化合物同定への労力が低減されるため、GC/MSノンターゲットメタボロミクス普及への一助となることが期待される。

論文審査の結果の要旨及び担当者

氏 名 (松 尾 晃 子)			
	(職)	氏	名
論文審査担当者	主 査	教授	福崎 英一郎
	副 査	教授	村中 俊哉
	副 査	教授	藤山 和仁
	副 査	教授	仁平 卓也
	副 査	教授	永井 健治
	副 査	教授	大政 健史
	副 査	教授	渡邊 肇
	副 査	教授	紀ノ岡 正博
	副 査	教授	松田 史生

論文審査の結果の要旨

ノンターゲットメタボロミクスは生体サンプル中の代謝物を一斉に測定・解析する。代謝物同定では標準物質の測定結果を格納したデータベース（以下、ライブラリと呼ぶ）から候補化合物を探索し、純粋な標準品によるバリデーションを行う。GC/MS(gas chromatography/mass spectrometry)はノンターゲットメタボロミクスでよく使用される機器の一種であり、各ピークの保持指標（基準物質によって標準化された保持時間）とEIスペクトル(electron ionization)に基づいて代謝物を同定する。EIスペクトルの再現性は高く、ライブラリには数十万種のスペクトルが登録されている。一方、保持指標ライブラリの登録数は一万種ほどにとどまっている。EIスペクトルの類似性のみに基づいた代謝物同定では、多くの偽陽性の候補化合物が現れることが知られている。測定で得られる数百のピークに対し、それぞれの候補化合物の標準品を揃え、測定するには多大な労力が必要となる。本研究ではQC(quality control)サンプルの希釈系列に基づくピークピッキング法と、保持指標予測モデルの構築および代謝物同定への活用を提案し、これらの組み合わせによる代謝物同定のスループットの向上を試みた。

第二章では約四百種の親水性代謝物の誘導体化合物が登録されたライブラリから、保持指標予測モデルを構築した。まず各代謝物の記述因子を計算し保持指標との相関係数(R^2)が0.8以上となったものを説明変数とし、保持指標の実測値を目的変数として、重回帰分析を行った。予測された保持指標値と実測値の R^2 は0.9となり、高い直線性を示した。このことから未同定代謝物の定性に使用可能なモデルの構築に成功したと結論付けた。

第三章ではQCサンプルによるピーク選択と保持指標予測モデルを、生薬の一種である川芎(*Conium officinale* Makino, *Ligusticum chuanxiong* Hort)に適用した。全六種類の川芎を等量比で混合したものをQCサンプルとし、この希釈系列からピークごとに検量線を作成した。検量線の直線性($R^2 > 0.9$)を基準としたピークピッキングから代謝物に由来するピークを選択し、クロマトグラム上の1974ピークのうち457ピークを選択した。未同定ピークの定性では、まず類似したEIスペクトルを持つ候補化合物の保持指標値を予測し、実測値との差が100未満となったものを選び出した。標準品によるバリデーションを経て、代謝物を同定した。Butane-1,2,3-triolの定性では、EIスペクトルから得られた45種類の候補化合物が予測モデルによって6種類へ絞り込まれた。3-Deoxy glucosoneでは5種類から2種類へ、Palatinitolでは63種類から6種類に、候補化合物が削減された。以上の結果から、標準品による確認が必要となる候補化合物を大幅に削減することに成功したと結論付けた。

本研究では未同定代謝物の効率的な同定法に向けた、保持指標予測モデルの構築とQCサンプルによるピークピッキングを提案した。この戦略を応用し、生薬川芎に含まれる3種類の未同定代謝物を新たに定性した。本手法によって従来、必要とされていた化合物同定への労力が低減されるため、GC/MSノンターゲットメタボロミクス普及への一助となることが期待される。よって本論文は博士論文として価値あるものと認める。