

キー成分の効率の良い見つけ方/データ解析の工夫

メタボロミクス研究成功の鍵はデータ解析

(大阪大学大学院工学研究科生命先端工学専攻) ○馬場健史 (ばんばたけし)

E-mail: bamba@bio.eng.osaka-u.ac.jp

多成分の一斉分析をベースとするメタボロミクスのデータ解析は一般的に多変量解析手法を用いて行われ、データの変換も含めて独特の情報処理技術を必要とする。複雑でわかりにくい部分が多いため十分に理解できていないところがあり、なかなか期待する結果が得られず困っておられる方がおられるのではないかと思います。メタボロミクスのデータ解析は目的とする成分の特定が重要なミッションであることから、今回の座談会のテーマである「キー成分の効率の良い見つけ方/データ解析の工夫」にはうってつけの題材ではないかと思います。

メタボロミクスは、大きく分けて代謝経路の構成メンバーなど目的成分を対象に絞ったターゲット解析と、観測される成分すべてを対象とするノンターゲットで解析になる。ノンターゲット解析の際には多成分の中からいかに目的成分を見つけ出すか、ターゲット分析の際には多成分の中からいかに効率的にかつ正確に成分の同定を行うか重要になる。いずれにせよ、質量分析計などの分析装置から得られる膨大な情報量のデータが多数のサンプル分積み上がった際に、その中からどのようにして目的とする成分をより分け見つけ出すことができるかが重要なポイントになる。したがって、データ解析の善し悪しによって結果が大きく異ってしまう。すなわち、得られたデータが宝の山となるか、ゴミの山となるかを左右すると言っても過言ではない。

メタボロミクスが提唱された10年前は標準化された技術はほとんどなかったため、我々のグループでは、これまでメタボロミクスのデータ解析に関して実用的な技術の構築に取り組んできた。ノンターゲット解析においては、ジーエルサイエンス社との共同研究により、二次的にピークリストを作成するのではなく、クロマトグラムのパターンを直接解析する手法の開発を行った¹⁾。クロマトグラムを構成する全てのデータポイントを使用するため、ピークリストを作成する手法に比べて圧倒的に変数が多く、微小なピークやコエリューションにも対応でき、また、ピーク同定を必要としないためスループットも高い。また、ターゲット解析においては、迅速簡便にGC/MSデータから代謝物同定が可能なシステムの構築を目的として、代謝物自動同定ソフトウェアの開発に取り組んだ。GC/MSデータから特徴検出によりピックアップされたピークについて、内生成分を疑似内部標準としてin-houseのライブラリとの較正を行った後に代謝物を精度良く同定するアルゴリズムを開発した²⁾。最近では専用のソフトやWebサイト、データベースが多数開発されており、また、各社の質量分析にも専用の解析ソフトが組み込まれるようになってきているが、それぞれの特性を理解して解析の目的に応じた使い分けが必要である。メタボロミクスのデータ解析においてはまだ解決すべき多くの課題を抱えており、今回の座談会において様々なご経験をお持ちの皆様にご意見を伺い出来れば幸いである。

1) E. Fukusaki, *et al.*, *Z Naturforsch [C]*, **61**, 267–272 (2006).

2) 馬場健史ら, *遺伝子医学MOOK 16*, 91–95 (2010).

キー成分の効率の良い見つけ方/データ解析の工夫

金をかけずにどれだけできるか？

((財)サントリー生物有機科学研究所)○小村 啓(こむらはじめ)

E-mail: hajime_komura@suntory.co.jp

近年、網羅的分析データを基に重要な成分を見つけ出す、「メタボロミクス」的なアプローチが広く用いられるようになり、多くの場合データ解析には多変量解析ソフトが用いられています。しかし、多変量解析は初心者にとっては「ブラック・ボックス」化しやすく、そこから得られた結果の解釈もままならないことも多々あるようです。また、新たに多変量解析ソフトを購入する場合比較的高価なため、導入を躊躇することも考えられます。そこで、お金をかけずにデータ解析がどこまで出来るか、汎用的に用いられているソフトで何とかならないかを考え、実際にエクセルの持つ関数の機能などを使って、官能評価結果と相関の高い成分を拾い出した例を紹介합니다。

通常、網羅的分析を行う場合、数多くあるサンプルをまず一斉に分析し、まとめてデータ処理をすることが多いと思いますが、そのような場合には個々のデータをじっくり眺める機会ほとんどありません。しかし、お金をかけずに仕事を進める場合には、「お金の代わりに汗をかきこと」と解釈して、データを自分の目でじっくり見ることが試みてください。特に、デジタルデータではなく、クロマトグラムそのものを比較することをお勧めします。まず、評点が満遍なくなるように4~5枚のクロマトグラムを選び、評点の順にクロマトグラムを並べて、ざっと眺めて「目力」で相関のありそうなピークを拾い出します。次に、更に2~3枚のクロマトグラムを追加し、候補ピークを絞り込んで、最終的に全てのクロマトグラムで自分が選んだ候補ピークが評点とよく相関していることを確認すれば良いわけです。もし候補ピークがいくつかある場合には、この段階でエクセルを用いてピーク強度と評点の相関係数を求め、同時に両者の散布図を作成して極端なはずれサンプルがないことを確認することも大切です。

ここで、皆さんもお気づきのように、クロマトグラムを眺める代わりに、最初からエクセルに全ピーク強度と評点を表としてまとめ、全てのピークについて、強度と評点の相関係数を求め、ピークを選択することも出来ます。しかし、どこかの段階で必ず散布図を作成して最終確認をしてください。そうすると、場合によってはデータポイントのばらつきから相補的になっているようなピークの存在にも気づくかも知れません。そんな場合には、双方のピークを足し合わせて、改めて相関を見ることも出来ます。

実は、これまでの話の中で一番重要で、手間のかかる部分はずして話を進めていました。それは、ピークのアラインメントです。GC/MS では似たような保持時間に違った成分に由来するピークがあっても、質量スペクトルを調べることで成分の異同が判別でき、ピークアラインメントをするのも比較的楽になります。しかし、お金をかけず GC/FID だけで分析を行い、保持時間だけを頼りにピークアラインメントを行う場合、結果的に間違ったアラインメントをしてしまって、後で悔しい思いをする危険性があります。こんな場合も、必要ならば着目する部分だけでもクロマトグラムを直接比較し、目でアラインメントを確認することで、問題を回避できる可能性があります。

お金をかけない場合には、やはりどこかで汗をかかないといけないわけですが、逆に言うと、汗をかき苦労をいとわなければ、お金をかけた以上の結果が得られる可能性もあるわけです。

キー成分の効率の良い見つけ方/データ解析の工夫

(ジーエルサイエンス) 秋本 智

演者のほかの方はそれぞれの研究においてデータ解析について工夫されている点をご披露頂けるとと思いますので、私はその前後の部分について少し紹介をしたいと思います。

・うまく行かなかった例

10年以上前のお話しですが、有るユーザーが熱分解クロマトグラムを利用した高分子の判定をソフトウェアで行っておりました。特定の成分の量を利用したデータベースを使用したものです。システムが古くなりまた既知成分のみでは判定精度も低くなっていたので更新したいが何か良い方法がありますか？との質問を受けましたが、その話しを聞いた人間が「クロマトグラムデータ全てを使った解析方法が出来ます」安易に引き受けてしまいました。

で、簡単に受けたのち当時担当では無かったのですが、私が導入(実際には導入後)に対応を行ったわけです。

目的はクロマトグラムデータを使用した分類/定量になる訳です。しかし各種サンプルの熱分解時データを持っている訳でも無く、また当時のPCのスペックでは多量なデータを読み込むことは出来ましたがそれ以上の事を行うと時間が掛かりすぎ、さらにその解析(+解析前の各クロマトグラムを一致させる方法も殆ど無い)についてのノウハウは無くユーザーへは出来ませんと平身低頭謝りに行きました。

・データ解析の工夫

普段の定性/定量を行う場合、各装置に付属・追加されているソフトウェアのみ利用するだけで十分と思われれます。しかし、題目のような方法を取る場合はメーカーのソフトウェアでは不十分な場合も見受けられ、それぞれの目的に即したソフトウェアが販売されております。

そのようなソフトウェアを利用する場合は、データを機器メーカーが想定していない使用方法をする為に利用者であるユーザーが苦勞するケースが多く見受けられます。解析には十分に時間をかける必要性はあると思いますが、このようところで時間をかけるのはもったいないと思います。装置やソフトウェアを選択する場合には互いのデータ移行がたやすく出来る事も機種選択のポイントとして検討した方が良いでしょう。

代謝物に関連するキー成分の効率のよい見つけ方

(味の素(株)ライフサイエンス研究所) ○宮野 博, 吉田寛郎

1) バイオマーカー探索

生体の内在性代謝物を直接 LC/MS で測定すると、未知化合物が数多く検出され、その評価や有用成分特定は容易ではない。これは、代謝物が膨大かつ多種多様な構造や物性を有するためであり、実用的な情報を速やかに確実に引き出すため、スペクトルやクロマトグラムを単純にする工夫が必要となる。

HPLC で発展した官能基特異的誘導体化は、その目的に非常に有用である。アミノ基、カルボキシル基、チオール基などをそれぞれ誘導体化することで、分析の選択性や感度を大幅に向上させることができる⁽¹⁾。特に、アミノ酸は代謝経路のハブであり、代謝変動の指標として有用である⁽²⁾。アミノ酸に絞り込んだあるいはアミノ酸をベースとするバイオマーカーの探索は、この分野の効率化に大いに役立つと考える。

2) 微量未知成分の同定

リン酸などの不揮発性の移動相と UV 検出器を組み合わせた HPLC 分析は簡便で、研究現場では、未だに数多く使われている。しかし、この系で微量な未知ピークが検出された場合、質量分析計等での構造解析のためには、手作業で、分取、濃縮、脱塩をおこなう必要があり、同定までに非常に時間がかかってきた。また時間がかかればかかるほど、対象となる物質の分解、散逸などの危険性が高くなる。このような分析系であっても、カラムスイッチングを駆使すれば、オンラインで、分取、濃縮、脱塩が可能で、不揮発性移動相を用いた HPLC で検出された化合物を、直接、質量解析に供することができる⁽³⁾。

誘導体化やカラムスイッチングのような HPLC のメリットを活かした前処理は、キー成分を効率よく見出し、同定するための強力な武器となる。

参考文献

- (1) Shimbo K et al., Multifunctional and highly sensitive precolumn reagents for amino acids in liquid chromatography/tandem mass spectrometry. *Anal. Chem.*, **81**, 5172-5179 (2009)
- (2) 安東敏彦 アミノインデックスー血中アミノ酸濃度による健康チェック法の開発 化学と工業 60,40-41 (2007)
- (3) Yoshida H et al., On-line desalting-mass spectrometry system for the structural determination of hydrophilic metabolites, using a column switching technique and a volatile ion-pairing reagent. *J. Chromatogr. A.*, **1119**, 315-321 (2006).