

氏 名 多田 一風太

学位(専攻分野) 博士(理学)

学位記番号 総研大甲第 2200 号

学位授与の日付 2020 年 9 月 28 日

学位授与の要件 生命科学研究科 遺伝学専攻  
学位規則第6条第1項該当

学位論文題目 Untargeted Metabolomics: Data Analysis Platform for All-Ion  
Fragmentation Mass Spectrometry

論文審査委員 主 査 教授 中村 保一  
教授 井ノ上 逸朗  
教授 黒川 顕  
准教授 川本 祥子  
教授 福崎 英一郎  
大阪大学 大学院工学研究科

(様式3)

## 博士論文の要旨

氏名 多田 一風太

論文題目 Untargeted Metabolomics: Data Analysis Platform for All-Ion Fragmentation Mass Spectrometry

New technologies make a leap forward in research fields by inspiring ideas and requiring new methods. As genetics has been advanced with DNA sequencers, metabolomics has gradually matured with mass spectrometry. Liquid chromatography-tandem mass spectrometry (LC-MS2) is a common technology for metabolomics studies. Major metabolomics experiments are roughly classified into two groups, targeted- and untargeted metabolomics. While targeted metabolomics aims to quantify pre-defined compounds with high accuracy, untargeted metabolomics tries to detect and identify as many compounds as possible for discovery studies. For reliable compound identification and estimation, MS2 spectrum is utilized because the fragmentation pattern of each molecule is consistent under the almost same experimental settings. In classical acquisition methods, a precursor ion is selected to acquire MS2 spectrum from co-eluting compounds. In contrast, All Ion Fragmentation (AIF) can generate all MS2 spectra by setting quite large  $m/z$  range (e.g. 40–1200 Da). Although AIF-MS is unbiased and reproducible, the acquired MS2 spectra are highly complex and difficult to interpret.

To solve the complex AIF MS2 spectra, I have developed a new Correlation-based Deconvolution (CorrDec) method. The CorrDec method utilizes intensity correlation between precursor ion and its fragment ions among samples. As a demonstration of CorrDec, it was applied to two datasets: dilution series of chemical standards and a 224-sample urinary metabolomics cohort. The serial dilution study showed that the peak intensities of fragment ions were highly correlated with their precursor ions. In the urine cohort study, 105 compounds were identified and CorrDec

could generate clean MS2 spectra for 85 compounds out of them (>80% MS2 match with reference). CorrDec can separate completely co-eluting compounds and work well in even low concentration compounds. Consequently, CorrDec enables more reliable compound annotations and identifications in multi-sample studies for untargeted metabolomics.

In order to confidently annotate and identify compounds in LC-MS2 data, reliable chemical standard libraries including three orthogonal properties—accurate mass (AM), retention time (RT), and MS2 spectrum—are required. In AIF projects, an MS2 spectrum is generated from a mixture of several adduct types and isoforms; therefore, MS2 spectral library measured by AIF mode can improve the quality of compound identification. However, AIF MS2 spectrum contains many noise peaks even in the measurement of a standard. I describe a workflow to confidently obtain AM, RT, and MS2 for a given compound using the AIF method and provide practical recommendations for library development. So far, 814 deconvoluted spectra by CorrDec and MS2Dec of 140 compounds were generated with manual curation as a chemical library. I illustrated how the library increases the confidence of compound identification in complex AIF data. The construction of high-quality, open-access libraries makes compound identifications more transparent, reliable, and transferable to the broader community.

I have proposed the AIF platform consisting of three metabolomics tools—MS-DIAL, MS-FINDER, and MS-LIMA. MS-DIAL and MS-FINDER were improved for AIF data, and MS-LIMA was newly developed. CorrDec was implemented into MS-DIAL, which is universal metabolomics software that supports various instruments. I have also improved MS-DIAL to adopt measurements of multiple collision energies and be fast and stable for large-scale study. MS-DIAL has grown as a modern user-friendly tool by my contributions. Second, MS-FINDER supports compound estimation by characterizing MS2 spectra, which is a key process in untargeted metabolomics.

MS-FINDER can annotate MS2 peaks as molecular formulas, chemical substructures, and types of adduct/isotopic ions. Lastly, MS-LIMA helps to properly manage MS2 spectra acquired from both biological samples and chemical standards. MS-LIMA is open-source software to curate, search, compare, and visualize MS2 spectra for stable and reliable management. The freely available AIF platform supports reliable data analysis, biological and technical insights, and reanalysis using public metabolomics raw data.

For further development in untargeted metabolomics, I believe that reusable data acquisition (such as AIF), reliable compound identification, and a universal and integrated data analysis platform are important. I hope that the AIF platform including CorrDec and the reliable library can improve the quality of compound identification, increase the number of annotated compounds, and help to exploit large-scale untargeted studies.

## 博士論文審査結果

Name in Full  
氏名 多田 一風太

論文題目 Untargeted Metabolomics: Data Analysis Platform for  
All-Ion Fragmentation Mass Spectrometry

液体クロマトグラフィー・タンデム質量分析法 (LC-MS2) は、メタボロミクスによる代謝産物の計測に広く用いられている技術である。All Ion Fragmentation (AIF-MS2) は可能な限り多くの化合物を検出・同定することを目的とするアンターゲットメタボロミクスに用いられる手法であり、液体クロマトグラフィーからのすべての前駆体イオンを断片化しスペクトルとして計測するためバイアスが少なく再現性が高い反面、得られる MS2 スペクトルは複雑でその解釈すなわちデコンボリューションを行うことが難しい。

多田さんは、この複雑な AIF-MS2 スペクトルの解釈法として、フラグメントのピーク強度はそれらの前駆体のピーク強度と相関するという発想に基づき、サンプル毎ではなく複数のサンプルを横断したピーク強度の情報を入力とし、前駆体とフラグメント強度の相関を利用した Correlation-based Deconvolution (CorrDec) 法を新たに開発した。この手法の有効性を示すために、多田さんは化学標準物質の段階的希釈サンプルと尿メタボロミクスコホートからの 224 サンプルのデータセットを解析することにより CorrDec 法は溶液中の化合物を完全に分離することができ、かつ低濃度の化合物に対しても有効であることを示した。この成果は多田さんが筆頭著者である論文として *Analytical Chemistry* 誌にアクセプト済である。

LC-MS2 のスペクトルデータに化合物を正確にアノテーションするためには、正確な質量、リテンションタイム、MS2 スペクトルを含む信頼性の高い化学標準ライブラリが必須となる。多田さんは 140 の化合物由来の 814 種のデコンボリューションスペクトルを元に、マニュアルキュレーションによって化合物データライブラリを生成し、このような高品質でオープンアクセス可能なライブラリにより、信頼性の高い化合物の同定が可能となることを示した。この成果は *Metabolites* 誌に多田さんを筆頭著者とする原著論文として発表されている。

また、多田さんは AIF 解析プラットフォームとして MS-DIAL, MS-FINDER, MS-LIMA の三種類のメタボロミクス解析ツールの開発と改良にも携わっており、MS-DIAL に前述した新規手法である CorrDec を実装するとともに、大規模メタボロミクス研究での高速かつ安定した測定を可能とする改善に貢献している。MS-FINDER はピークのアノテーションによりスペクトルからの化合物推定を支援するツールであり、MS-LIMA はスペクトルのキュレーション、検索、比較、可視化によりサンプルと標準物質から取得したスペクトルを管理できるオープンソースのソフトウェアである。このような自由に利用できる AIF 解析用のプラットフォームの提供により、すでに塩基配列で実現されているように、データベース化され公開されるメタボロミクスの生データを活用した信頼性の高い

データ解析、生物学的な考察、データの再利用をサポートすることができると多田さんは主張している。MS-DIAL の開発では共著者の一人として *Nature Biotechnology* 誌への論文がアクセプト済である。

アンターゲットメタボロミクスのさらなる発展のためには、AIF などの再利用可能なデータ蓄積、信頼性の高い化合物同定、データ解析プラットフォームの整備が重要であると申請者は考え、それらを実現するための技術とデータライブラリ、ソフトウェアの作成へのアプローチを行い、高い完成度でそれらを実現させている。得られた成果は将来のメタボローム解析の発展に寄与する学術上の優れた研究である。以上の理由により、審査委員会は、本論文が学位の授与に値すると判断した。