SHIMADZU

Eisai Co., Ltd.

ASMS2013
MP18-359

ms3d FIRST Program

Novel preprocessing method to align retention time of LC-MALDI and new implemented functions in Mass++ for differential analysis

Yuichiro Fujita[1], Natsumi Funakoshi[1], Yoshihiro Yamada[1], Satoshi Tanaka[1], Shin-ichirou Kawabata[1], Shinichi Iwamoto[1], Shinichi Utsunomiya[1], Shigeki Kajihara[1], Tsuyoshi Tabata[2], Ken Aoshima[2], Yoshiya Oda[2], Koichi Tanaka[1]
1 Koichi Tanaka Laboratory of Advanced Science and Technology, Shimadzu Corporation, Kyoto, Japan. 2 Eisai Product Creation systems Eisai Co., Ltd, Tsukuba, Japan

1. Introduction

Differential analysis plays an important role in biomarker discovery. Potential biomarker peaks can be discovered by comparing liquid chromatography (LC)-MS data of samples between healthy subjects and patients.

In differential analysis, it is necessary to align retention time (RT) utilizing similarity of total ion chromatogram (TIC) waveforms to match peak positions (RT and *m/z*) among multiple samples. However, conventional RT alignment is ineffective for LC-MALDI data because of its low TIC waveforms similarity due to spectral reproducibility issues.

Here, we developed and evaluated a novel preprocessing approach to calculate TIC in RT alignment for LC-MALDI data to improve differential analysis.

Our new preprocessing and the additional functions were implemented in freely available Mass++ software[1] (http://www.first-ms3d.jp/english/) (**Figure 1**).

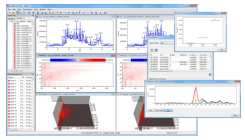

Figure 1 Overview of freely available Mass++ software.

2. Methods

2-1. Differential analysis

Differential analysis was performed by comparing peak intensities in the peak matrix.
The peak matrix is a list of positions of detected peaks in 2 dimensions (RT and *m/z*) from MS[1] and the peak height or areas in two samples.
If peaks exist in one sample group commonly at a certain position but do not exist in the other, peptides related to such unique peaks are considered biomarker candidates.

2-2. Flow

Differential analysis for LC-MALDI dataset was performed as follows,

(1) Normalization: all spectra were normalized using an internal standard.

(2) TIC with preprocessing: Spectral intensities were scaled by using variance and average of intensities in each spectrum to reduce the influences derived from intense peaks. TIC at RT = *t* was calculated after this preprocessing as follows (in this poster, we simply call this "recalculated TIC" as "TIC"),

$$\mathrm{TIC}(t) = \sum_{i=1}^{N} \mathrm{Int}_i \frac{a(t)}{v(t)}$$

N ; Peak number of spectrum at RT = t
i ; Peak index
Int_i ; Peak intensity of index i
$a(t)$; average of peak intensities at RT = t
$v(t)$; variance of peak intensities at RT = t

(3) RT alignment: RT alignment was performed using "AB3D" algorithm[2] for the TIC calculated in (2).

(4) Peak matrix: peak matrix was created from aligned spectra by existing Mass++ functions.

(5) Statistical analysis: Mann-Whitney's U test was performed to find peaks whose intensities were significantly different (p-value < 0.05) between sample groups.

(6) Identification: Mascot MS/MS Ion Search(MIS) was performed to identify the peptide for the peaks with significant difference.

(7) Multivariate analysis: PCA (Principal Component Analysis) and OPLS-DA (Orthogonal Partial Least Square – Discriminant Analysis) were performed to find biomarker candidates.

3. Result

3-1. Measurement

<Sample>
We evaluated our tools for differential analysis using two sample groups. "**4 mix**" group consists of 4 mixed proteins (i) ~ (iv) and "**5 mix**" group consists of 5 mixed proteins (i) ~ (v) shown in **Table 1**. Proteins (1 pmol/μL each) in **Table 1** were digested by trypsin. The first group and the second group were measured at five and four times respectively.
In this study, **BSA** is supposed to be detected as a biomarker candidate.

Table 1 Protein samples

No	Protein Name	Abbreviation
(i)	Alcohol Dehydrogenase	ADH
(ii)	Enolase	ENOL
(iii)	Bovine Hemoglobin	Hemo
(iv)	Phosphorylase b	PhosB
(v)	Bovine Serum Albumin	**BSA**

<Matrix>
CHCA 6.25 mg/mL(5 mg/well)
<Internal Standard>
Bradykinin (Brd)1-7, Adrenocorticotropic Hormone (ACTH)18-39 (125 fmol/mL(100 fmol/well) each) 50% acetonitrile (ACN)water
<instrument >
1D nanoflow LC(nanoLC)-AccuSpot-AXIMA system
・*Prominence nano-AccuSpot*™ (Shimadzu Corporation, Japan)
・*AXIMA Performance*™ (Shimadzu/Kratos, UK)

Figure 2 *AXIMA Performance*

3-2. RT alignment

Figure 3 shows a comparison of RT alignments. Upper figure was the result using TIC without preprocessing (conventional method) and lower figure was the result using TIC with preprocessing (proposed method).
In the proposed method, TIC waveform similarity was improved and the TIC peaks were aligned successfully.

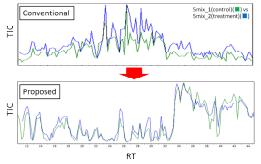

Figure 3 Comparison of RT alignments.

3-3. Statistical Analysis and Identification

Figure 4 shows an example of peak matrix (extracted). Peak positions (i.e. RT and m/z) were determined following the peak detection for all spectra after RT alignment. The 5th ~ 13th column of peak matrix are monoisotopic peak area of each spectrum. The 3rd and 4th column are results MIS and U test for each peak position respectively.
Note that "redundant peak" due to the margin error is appeared if RT alignment does not work well.

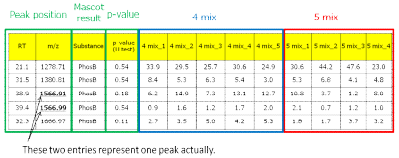

Figure 4 Example of peak matrix (extracted).

Table 2 shows the results of differential analysis "4 mix" and "5 mix" group. The number of differential peaks are shown and values enclosed in parentheses means the number of redundant peaks.
Using the peak matrix generated from the conventional method, not only BSA but also PhosB were found as the differential substance and many redundant peaks were included in BSA. In contrast, only BSA was identified as the differential compound using the peak matrix generated from our proposed method.

We therefore confirmed that our designed RT alignment preprocessing worked well and demonstrated the potential utility as a differential analysis tool in Mass++.

Table 2 Comparison of result of differential analysis using U test.

Identified protein	Number of differential peaks identified by MIS and p-value < 0.05 (U test).	
	Conventional	Proposed
ADH	2(1)	1(0)
ENOL	1(0)	1(0)
Hemo	0(0)	0(0)
PhosB	11(2)	2(1)
BSA	27(6)	21(1)

3-4. Multivariate analysis

The multivariate analysis was conducted using SIMCA ver.13 (Umetrics Inc.).

3-4-1. PCA

Figure 5 shows the comparison of PCA results between the conventional method and the proposed method. Sample groups were separated clearer using the proposed method than the conventional method.

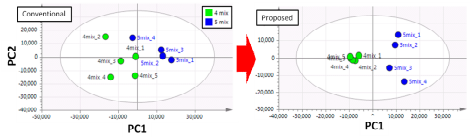

Figure 5 Comparison of PCA results.

3-5-2. S-plot

Figure 6 shows S-plot of OPLS-DA using the peak matrix generated by the proposed method. The horizontal axis means the contribution to group separation and the vertical axis means the reliability. Plotted marker type represents the significance in U test and the identified protein.
Even though many peaks have p-value < 0.05 (◆ ◆), most of them were plotted outside of biomarker candidate area. On the other hand, All peaks identified as BSA peptides (◆ ■) were plotted at biomarker candidate area.
We confirmed that it is important to use both statistical considerations and multivariate analysis. **We have developed a Mass++ plug-in which can display analytical results of SIMCA software.**

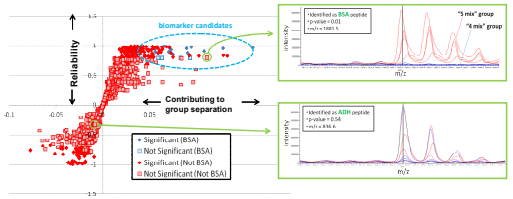

Figure 6 S-plot of OPLS-DA.

4. Conclusion

By using proposed method (see **2-2**), TIC waveform similarity of LC-MALDI was improved in RT alignment. As a result, the differential analysis was performed successfully in identification, statistical analysis, and multivariate analysis.

5. Reference

[1] S. Utsunomiya et. Al., 61th ASMS (2013), MP18-360
[2] K. Aoshima et. al., 56th ASMS (2008), TP-659

6. Acknowledgment

This research is granted by the Japan Society for the Promotion of Science (JSPS) through the "Funding Program for World-Leading Innovative R&D on Science and Technology (FIRST Program), "initiated by the Council for Science and Technology Policy (CSTP).