

Mass++ : ピーク情報検出技術の開発

Mass++: Development of the peak information detection techniques for mass spectrometry analysis

○宇都宮真一¹・姚 精文¹・藤田雄一郎¹・田中聡¹・山田賢志¹・梶原茂樹¹・川畑慎一郎¹・田畑剛²・高橋健太郎²・青島健²・小田吉哉²・田中耕一¹ 1 島津製作所 田中最先端研究所 2 エーザイ株式会社

○ S.Utsunomiya¹, J.W.Yao¹, Y.Fujita¹, S.Tanaka¹, S.Yamada¹, S.Kajihara¹, S.Kawabata¹, T.Tabata², K.Takahashi², K.Aoshima², Y.Oda², K.Tanaka¹ 1 Shimadzu Corp. Koichi Tanaka Laboratory of Advanced Science and Technology 2 Eisai Co., Ltd.

1. Introduction

生データからの雑音情報除去 (Signal de-noising)、ピーク検出 (Peak detection) から成る初段のデータ解析が、同定・定量等の後段解析の精度や信頼性に与える影響は大きい (Fig.1)。

今回、バイオマーカの探索や検定の局面で取得されるデータを想定して、これらのピーク情報検出技術の開発を行った。

なお、開発プラットフォームとして ms3d プロジェクトで開発・保守を行っている質量分析データ解析用ソフトウェア Mass++^{[1][2]} を用いた。今回開発した手法は同ソフトウェアのプラグインとして実装しており、定量解析の評価にも同ソフトウェアの Quantitation 機能、Batch 処理機能、3D-View 機能を用いている。

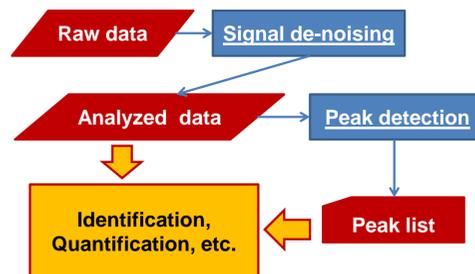


Fig.1 Peak information extraction

2. Method

2-1 Signal de-noising

(1) Noise reduction

Wavelet 解析は各種計測装置の画像処理・信号処理に応用されている^[3]。今回、過去のノウハウに質量分析データ用の工夫を加えたアルゴリズムを新たに構築した。

- ◆ m/z のピーク位置が重要となるので、基底関数には対称性に優れた直交基底である Coiflet (Fig.2) を採用
- ◆ 対象データの Wavelet 変換で得られる多重解像度成分に対し、指定 Parameter に忠実な雑音成分低減処理を施して再構成

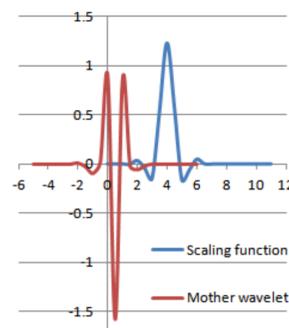


Fig.2 Coiflet (N = 6) Wavelet

An orthogonal wavelet base whose symmetric property is superior.

(2) Baseline subtraction

バイオマーカ探索フェーズで使用されるサンプルの種類・品質は多様で前処理やイオン化の条件も多岐に渡るため、ベースライン特性は大きく変わる。これらに簡便に正確に対応できるアルゴリズム BLT (Bottom Line Tracing) 法を構築した。

- ◆ 理解しやすいパラメータ (ピーク幅と個数) 設定により複雑な操作なしに多様なデータに対応

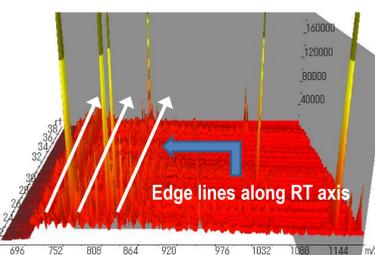


Fig.3 Contaminated peaks

Contaminating peaks are seen in LC-MS data as continuous edge lines along RT axis

(3) Contaminated peaks depression

LC-MS データにおいて RT 方向に顕著な夾雑ピークが見られる場合があり、自動分析における Precursor 選択にも悪影響を及ぼす。今回、これらを抑制する手法を構築した。

- ◆ RT 間の信号強度を正規化後、RT 方向の低周波の信号変動をベースラインとして除去

2-2 Peak detection

Peak list に含まれる雑音ピークが多いと同定信頼度が損なわれる。精度の高いピーク同定を目的としたピーク検出法 MWD (Multi Window Detection) 法^[4]を開発した。

- ◆ 領域ごとに Ion peak を Noise peak 中の孤立点と見なして検出することで、弱い Ion peak 検知しつつ Noise peak 混入を抑制
- ◆ 高分子測定時の MS スペクトル特性を反映したモデルにより、重なりのある Isotopic peak も正確に判定除去 (Fig.4)

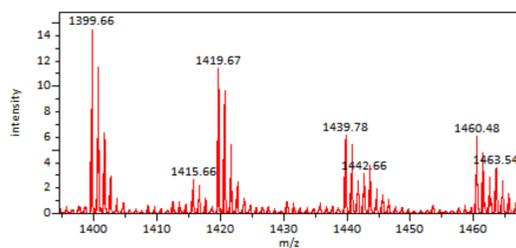


Fig.4 Mono-isotopic peaks detection

MWD can detect mono-isotopic peaks correctly as shown, even if some overlapped components exist.

3. Experimental Results

3-1 Effects on signal waveforms

(1) Noise reduction 効果 : ピーク波形を保てる範囲でノイズ低減パラメータ調整

⇒ ノイズ低減効果は Wavelet の方が高い。また、Wavelet の方が冗長度が高く Gaussian ほど丁寧な調整は不要。

Red : Wavelet (Width = 0.4, Rate = 1.5) Blue : Gaussian (Width = 0.08, Sigma = 3) Green : Raw data

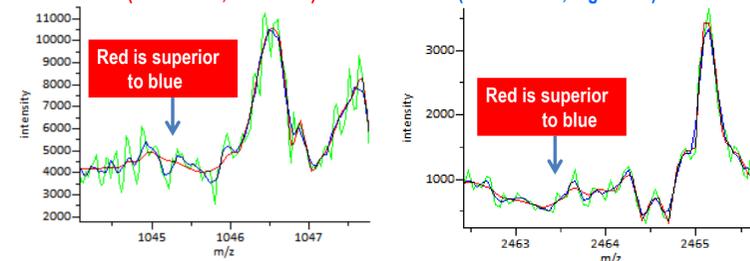


Fig.5 Effect of noise reduction

(2) Baseline subtraction 効果 : 大きなベースライン変動 (e.g. 核酸試料 MALDI 分析)

⇒ BLT 法では複雑な設定なしに適切なベースライン検出・除去が可能。

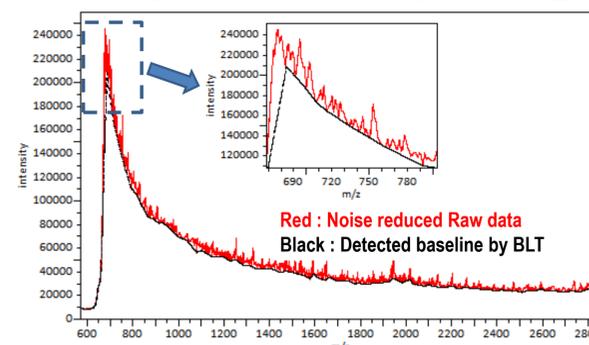


Fig.6 Effect of baseline subtraction

(3) Contaminated peak depression 効果 :

⇒ Fig.3 上に見られた RT 軸に沿う変動が低減できた。

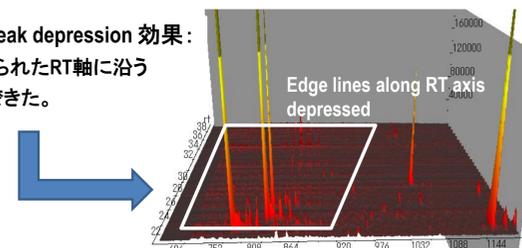


Fig.7 Effect of contaminated peaks depression

3-2 Effects on quantification

定量分析結果評価 : 既知タンパクの希釈系列データ (内標なし) に対して、熟練分析者による手動調整 (装置付属 SW 使用) と Mass++ の Default 値自動処理を比較した

⇒ Analyzed 波形はノイズ低減効果優位 (Fig.8(a)) で、波形劣化もなし (同(b))

⇒ 濃度直線性は両者波形劣化のない m/z = 1046.5 では同等 (Fig.9)

Red : Default operation with Mass++ (Wavelet and BLT) Green : Raw data Blue : Manual operation by a well-trained operator with the instrument attached software

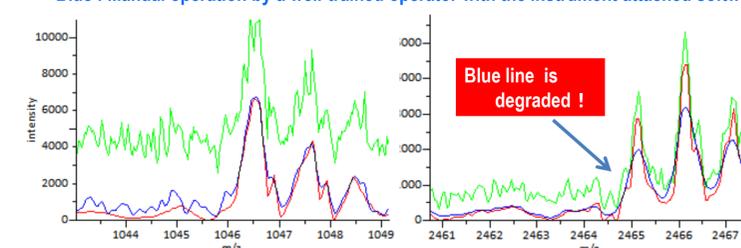


Fig.8 Analyzed data comparison

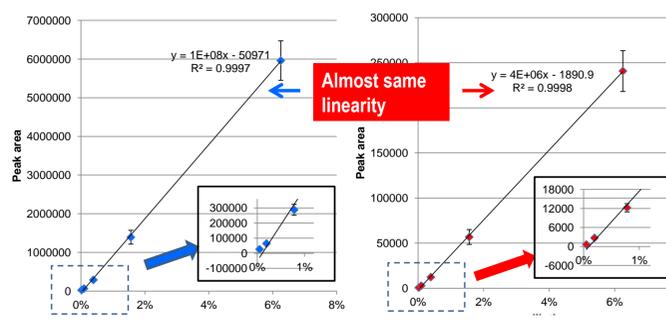


Fig.9 Linearity comparison : m/z = 1046.5

※ Peak area values are different between two software as their estimations are different.

3-3 Effects on identification

ピークリストに対する Mascot 検索の protein scores を市販ソフトウェア (SW-A) と比較 ⇒ MWD の方がほとんどの例で高いスコアを記録 (既知試料 MALDI-MS 計測例)

Table1 Protein scores of Mascot search

| | Samples | | | | | | | | | | | | | Average |
|------|---------|------|------|------|------|------|------|------|------|-------|-------|-------|-------|---------|
| | BSA1 | BSA2 | BSA3 | BSA4 | BSA5 | BSA6 | BSA7 | BSA8 | BSA9 | BSA10 | BSA11 | BSA12 | BSA13 | |
| MWD | 1072 | 955 | 890 | 701 | 440 | 650 | 823 | 1121 | 590 | 503 | 325 | 244 | 362 | 667 |
| SW-A | 951 | 689 | 473 | 338 | 234 | 378 | 438 | 1224 | 338 | 378 | 175 | 156 | 82 | 450 |

Conclusion

- ノイズ低減、ベースライン除去、夾雑ピーク除去、ピーク検出の新アルゴリズムを開発し、Mass++ プラグインとして実装した。
- ノイズ低減法とベースライン除去法はデータ特性変動に対して複雑な設定なしに対応可能で、定量解析試験で熟練操作者と同等の結果が得られた。
- ピーク検出法は同位体ピークを高精度に検出できて雑音ピークが少なく、同定試験において従来法 (他ソフト) より信頼度が高い結果が得られた。

ノイズ低減、ベースライン除去、ピーク検出の3つのプラグインは ms3d より一般公開している Mass++^[2] (V2.6.0以降) 上で利用できる。同ソフトウェアはマルチベンダー対応であるため、ユーザが使用中の装置データに対する効果確認も可能である。

References

- [1] S.Utsunomiya, S.Tanaka, et. al., "A platform for mass spectrometry to construct suitable software to achieve user's own purposes", ASMS(2013), MP18-360
- [2] Mass++ Download website : <http://www.first-ms3d.jp/achievement/software/mass2>
- [3] S.Utsunomiya, "Development of Signal and Image Processing Techniques Using Wavelet Transform", SHIMADZU REVIEW, Vol.57, 133-139 (2000)
- [4] J.Yao, et.al., "Peak Detection Method in Mass++", AOHUPO(2012)

Acknowledgement

本研究は、日本学術振興会の最先端研究開発支援プログラムにより、助成を受けたものである。This research is supported by the Japan Society for the Promotion of Science (JSPS) through its "Funding Program for World-Leading Innovative R&D on Science and Technology" (FIRST Program).