

京都大学	博士 (薬科学)	氏名	TAECHAWATTANANANT PASRAWIN
論文題目	Peak identification and quantification in proteomic mass spectrograms using non-negative matrix factorization		
<p>Shotgun proteomics using liquid chromatography coupled with mass spectrometry (LC/MS) has been used widely to identify and quantify digested peptides. Recent advances in LC/MS technologies have permitted detection of a wide range of signal parameters in a three-dimensional mass spectrogram, including m/z, retention time, and intensity. However, despite the high resolving power of modern LC/MS instruments, a large part of the mass spectrogram remains unresolved. A new approach based on mathematical and statistical methods is required to interpret mass spectrograms accurately and with higher sensitivity.</p> <p>Non-negative matrix factorization (NMF) is a well-used machine-learning technique that is applicable for source separation. NMF computation is purely additive, guarantees non-negativity, and generates a controllable sparse approximation with a sparsity constraint. Therefore, NMF should be suitable for analyzing combined signals of non-negative m/z, retention time, and intensity in sparse proteomic mass spectrograms. However, application of NMF to proteomics has barely been explored.</p> <p>We have developed a novel approach based on NMF to analyze three-dimensional mass spectrograms obtained from proteomic studies efficiently. Our proteomics NMF (pNMF) incorporates isotopic distribution, learns noise, exploits a protein-peptide hierarchical relationship by using a group sparsity constraint, and configures a reasonable initialization by using predicted retention times. The proposed update rule with the constraint of choice guarantees convergence to a local optimum, meaning that with appropriate initialization, pNMF guarantees convergence to the right solution.</p> <p>The effects of NMF modifications provided a significant improvement in extracting a high-quality chromatogram and annotating accurate peptide retention times, in comparison to the classical NMF. In benchmarking, pNMF identified and quantified 1319 peptide ions from the mass spectrogram obtained by LC/MS analysis of tryptic peptides of 48 standard proteins, and showed excellent agreement with the conventional identifications using Mascot and Skyline. For computational performance, despite the size of the computational tasks, which involve several million elements, the computation time and space required by pNMF are manageable on a common desktop computer. This is the first application of NMF to LC/MS-based proteome analysis. Since pre-processing is not required, this approach can maximize the efficiency of both protein and peptide identification and quantification.</p> <p>Further improvements for NMF algorithm-based approach are possible by incorporating other significant features such as product ions from proteomic mass spectrograms. We have added another NMF layer for analyzing LC/MS/MS spectra. The algorithm provided better performance for peptide identification. We also have been able to perform proteome analysis on a large scale while reducing running time.</p>			

(論文審査の結果の要旨)

液体クロマトグラフィータンデム質量分析 (LC/MS/MS) 法を用いるショットガンプロテオミクスでは、プロテオーム試料はあらかじめ消化酵素により断片化されてペプチド混合物となる。ペプチド混合物は、LCで時間分離されながらMS装置に送りこまれ、順次質量スペクトルを計測する。このとき、特定の成分のみが適宜抽出され、希ガス分子衝突によりさらなる分解が行われ、MS/MSスペクトルを計測する。この時系列スペクトル群を解析することで、試料中におけるタンパク質の存在比率=プロテオームプロファイル情報を取得する。オミクス技術の発展により、伝統的な生物学実験では不可能であった、複雑で微小な生命系を分子レベルで解析することができるようになった。TAECHAWATTANANANT PASRAWIN氏は、ショットガンプロテオミクスにおいて質量分析技術と統計的信号解析技術を融合させることにより、プロテオームの計測・解析技術を深化させ、次世代プロテオーム解析技術として、非負値行列因子分解 (NMF) に基づく新規手法を開発した。この手法では、従来のNMFにペプチドイオンの同位体分布プロファイル、ノイズ、およびタンパク質-ペプチド階層性を考慮したグルーピングを組み込んでおり、これにより分析の精度を大幅に向上させることに成功した。さらに、この新規手法を48種の混合タンパク質試料や大腸菌プロテオーム試料に適用し、従来法よりも優れた結果を得た。これは、LC/MSプロテオミクスにおけるNMFの最初の適用例であり、生データに対して前処理が必要ないため、このアプローチによりタンパク質とペプチドの同定・定量効率の大幅な向上が期待できる。

よって、本論文は博士 (薬科学) の学位論文として価値あるものと認める。また、令和2年4月3日、論文内容とそれに関連した事項について試問を行った結果、合格と認めた。