

戦略的創造研究推進事業 CREST  
研究領域「計測技術と高度情報処理の融合による  
インテリジェント計測・解析手法の開発と応用」  
研究課題「質量分析と統計解析の融合による  
メタプロテオミクス」

## 研究終了報告書

研究期間 2018年10月～2024年03月

研究代表者：石濱 泰  
(京都大学 大学院薬学研究科 教授)

## § 1 研究実施の概要

### (1)実施概要

現在、プロテオーム解析でもっともよく用いられているボトムアッププロテオミクスでは、タンパク質混合物試料をトリプシンで消化後、LC/ESI/MS/MS で計測し、配列データベース情報との照合からタンパク質を同定・定量する手法が用いられている。これに対し、以下の項目を達成することにより、従来法を超える解析法の開発に成功した。

#### ① 高速 LC/ESI/IMS/MS/MS 計測システムの開発(石濱 G)

一日 1000 検体の処理速度をもちながら、100 ng から 3000 ヒトタンパク質を同定定量可能な計測システムの開発に成功した。1試料あたりの分析時間は84秒であり、現時点で世界最速システムとなっている。

#### ② Proteomic Mass Spectrogram Decomposition (ProtMSD)による LC/ESI/MS/MS データ(Data-Independent Acquisition)の解析(石濱 G)

LC/ESI/MS データに対し、音響シグナル解析で用いられるデータの階層性に基づく非負値行列分解法(NMF)を適用し、従来法をこえる大腸菌プロテオーム解析結果を得ることに成功。Proteomic Mass Spectrogram Decomposition (ProtMSD)と命名。さらに、ヒト培養細胞株の DIA モードで取得した LC/ESI/MS/MS データに対し、ProtMSD を適用したところ、従来法を超える性能を確認した。

#### ③ 混合ガウスモデルによる Mass Spectrogram Clustering (LC/ESI/MS, LC/ESI/MS/MS) (石濱 G)

②とは異なる手法として混合ガウスモデルを LC/ESI/MS スペクトログラムに適用し、タンパク質ーペプチドーペプチドイオンの階層性に基づき、タンパク質の同定・定量が可能であることを確認した。さらに LC/ESI/MS/MS スペクトログラムに適用を拡大し、複雑な試料にも対応可能であることを確認した。

#### ④ LC/ESI/IMS/MS/MS 計測におけるタンパク質言語モデルの活用(イオンモビリティ予測、LC 保持時間予測)(田中 G)

タンパク質言語モデルを利用した pretrained protein language model-based network (PPLN)を開発し、LC/ESI/IMS/MS/MS で取得されるイオンモビリティの予測モデルを確立した。本予測値は ProtMSD の初期値としての利用や、実測値とのずれをペプチド同定の際のフィルタに応用した。

#### ⑤ 四重極におけるスパース推定(田中 G)

四重極マスアナライザにおけるイオンの透過特性を制御し、感度と質量分解能を向上させる手法を開発し、実測定データでその有効性を確認した。さらに DIA への応用を行った。

#### ⑥ 近縁種ゲノム情報を用いた未知生物種のプロテオーム解析(奥田 G)

既知生物種の進化系統樹における配列類似性に基づいて未知種の配列を生成する手法を開発し、ピロリ菌プロテオーム解析に適用した。

## (2) 顕著な成果

### < 優れた基礎研究としての成果 >

#### 1. マススペクトログラムに対する行列分解によるピークの同定定量法の開発

概要: 近年の LC/MS 技術の進歩により、プロテオーム解析の感度、分解能、速度が向上したことで、複雑なデータを解釈するためのより高度なアルゴリズムが求められている。ここでは、ペプチドとタンパク質を同時に同定・定量するための新しい統計的手法 ProtMSD (Proteomic Mass Spectrogram Decomposition) を提案した。ProtMSD は、プロテオミクスマススペクトログラムと、タンパク質に関連するすべてのペプチドイオンの参照マススペクトルを辞書として与え、従来の閾値処理やピークピッキングなどを行わずに、グループスパース性制約の下でそれらのペプチドイオンのクロマトグラムを推定するものである。LC/MS を用いたプロテオーム計測データにおいて行列分解を用いた初めての例として学術誌に成果を報告した。さらに、ヒトプロテオーム DIA データに対しても本手法は適用可能であった。

#### 2. 高速 LC/ESI/IMS/MS/MS 計測システムの開発

概要: 一日 1000 検体の処理速度をもちながら、ヒト培養細胞株抽出タンパク質 100 ng から 3000 タンパク質を同定定量可能な計測システムの開発に成功した。1 試料あたりの分析時間は 84 秒であり、現時点で世界最速システムとなっている。

#### 3. 未知生物種タンパク質のアミノ酸配列推定

概要: 従来のデータベース検索法の弱点である「未知生物種の未知アミノ酸配列をもつペプチドが同定できないこと」を克服するため、ゲノム既知の生物種におけるゲノム配列データベースから、進化系統樹をリファレンスに利用し、より確からしいアミノ酸配列にバイアスをかけながら MSMS スペクトルとの照合を行う手法を開発し、ピロリ菌プロテオームデータを用いて検証した。

### < 科学技術イノベーションに大きく寄与する成果 >

#### 1. 四重極マスアナライザーにおける帯観測モデルの適用による感度・分解能の向上

概要: 四重極マスアナライザーにおける透過特性を制御し、通常と比較して相対的に広い  $m/z$  範囲を透過させることで感度の高い低分解能観測信号を得るという観測手法を提案した。さらに、計測系に対して観測行列  $A$  を帯行列とする線形観測モデル  $y = Ax$  を仮定して圧縮センシングの発想を援用し、LASSO、すなわち L1 正則化つき最小 2 乗法によって高分解能質量スペクトルを推定するというアプローチを提案した。計算機上でのシミュレーションデータおよび実測データを用いて提案手法を評価したところ、既存手法と比較してマススペクトルの推定精度や感度の向上が確認された。

#### 2. タンパク質言語モデルを利用したペプチドのイオンモビリティ推定

概要: タンパク質言語モデルを利用した pretrained protein language model-based network (PPLN) を開発し、LC/ESI/IMS/MS/MS で取得されるイオンモビリティの予測モデルを確立した。

本予測値は ProtMSD の初期値としての利用や、実測値とのずれをペプチド同定の際のフィルタに応用可能であった。

#### <代表的な論文>

1. Taechawattananant P, Yoshii K, Ishihama Y., Peak Identification and Quantification by Proteomic Mass Spectrogram Decomposition, J Proteome Res. 2021, 20(5):2291-2298.

概要:LC/MS 測定データを用いた、ペプチドとタンパク質を同時に同定・定量するための新しい統計的手法 ProtMSD (Proteomic Mass Spectrogram Decomposition)を提案した。ProtMSD は、すべてのペプチドイオンの参照マスペクトルを辞書として与え、スペクトログラムの行列分解によりそれらのペプチドイオンのクロマトグラムを推定する手法である。また、タンパク質とペプチドの階層関係、同位体分布プロファイル、ペプチドイオンの参照保持時間、ノイズのマスペクトルを事前に学習しておくことで、本手法が大幅に改善されることを示した。本研究は、LC/MS を用いたプロテオーム計測データにおいて、同定・定量を目的として行列分解を用いた初めての試みである。

2. Ayana Tomioka, Ryota Tomioka, Issei Mori, Yosuke Isobe, Makoto Arita, Koshi Imami, Eisuke Kanao, Kosuke Ogata, Yasushi Ishihama, One thousand samples per day capillary-flow LC/MS/MS for high-speed, high-sensitivity and in-depth proteomics, bioRxiv 2023.06.05.543682.

概要:一日 1000 検体の処理速度をもちながら、ヒト培養細胞株抽出タンパク質 100 ng から 3000 タンパク質を同定定量可能な計測システムの開発に成功した。1 試料あたりの分析時間は 84 秒であり、現時点で世界最速システムとなっている。

3. Uemura, Kyoya; Obuchi, Tomoyuki; Tanaka, Toshiyuki (2024). Sparse Modeling for Spectrometer Based on Band Measurement. IEEE Transactions on Signal Processing, volume 72, pages 1728-1738. <https://doi.org/10.1109/TSP.2024.3381443>

概要:四重極マスアナライザーにおける透過特性を制御し、通常と比較して相対的に広い  $m/z$  範囲を透過させることで感度の高い低分解能観測信号を得るという観測手法を提案した。さらに、計測系に対して観測行列  $A$  を帯行列とする線形観測モデル  $y = Ax$  を仮定して圧縮センシングの発想を援用し、LASSO、すなわち L1 正則化つき最小 2 乗法によって高分解能質量スペクトルを推定するというアプローチを提案し、計算機上でのシミュレーションデータで精度や感度の向上が確認された。

## § 2 研究実施体制

### (1) 研究チームの体制について

### ① 石濱グループ

研究代表者:石濱泰(京都大学薬学研究科教授)

研究題目:質量分析と統計解析の融合によるメタプロテオミクス

研究実施項目:

研究項目 1:統計的信号解析に必要な実データ・標準データの大規模計測を行う。

研究項目 4: タンデム質量分析データを用いたタンパク質の同定 (DIA 解析、田中 G と共同)

研究項目 5:各階層・各次元において、独立した統計的情報処理を検討する。(奥田 G と共同)

研究項目 8:適応的制御による計測過程のデータ取得性能向上(田中 G と共同)

### ② 田中グループ

研究代表者: 田中利幸(京都大学情報学研究科教授)

研究題目: LC/MS/MS 統計情報解析と適応的制御計測

研究実施項目:

研究項目 2: 単一タンパク質からなる試料に対するペプチドの同定

研究項目 3: 複数タンパク質からなる試料に対するペプチドおよびタンパク質の同定

研究項目 4: タンデム質量分析データを用いたタンパク質の同定

研究項目 6: 試料中の未知タンパク質の同定

研究項目 8:適応的制御による計測過程のデータ取得性能向上(石濱 G と共同)

### ③ 奥田グループ

研究代表者: 奥田修二郎(新潟大学医歯学総合研究科教授)

研究題目:メタプロテオゲノム情報解析

研究実施項目:

研究項目 1:統計的信号解析に必要な実データ・標準データの大規模計測のうち、リファレンスライブラリの作成を行う。

研究項目 5:各階層・各次元において、独立した統計的情報処理を検討する。

研究項目 7:ゲノム情報を融合しプロテオーム検出能を向上させる。

### (2)国内外の研究者や産業界等との連携によるネットワーク形成の状況について

日本質量分析学会(石濱が現会長)、日本プロテオーム学会(奥田が現理事、石濱が前会長)、アジアオセアニア質量分析学会(AOMSC、石濱が理事)、アジアオセアニアプロテオーム機構(AOHUPO、石濱が副会長)、ヒトプロテオーム機構(石濱が前理事)などの学会主催の年大会などを通じて、本研究の成果発表を続けており、国内外の研究者から問い合わせを受けている。また、質量分析メーカーとも共同研究を通じて、本技術の実装について検討を行っている。