

研究開発課題 中間評価結果

➤ 課題情報

研究開発課題名 「次世代低分子マスペクトルデータベース シン・マスバンクの構築」

研究代表名 松田 史生（大阪大学 大学院情報科学研究科 教授）

➤ 中間評価結果：

マスペクトルデータベース MassBank に生体由来スペクトルを集積していくためのパイプラインである Shin-MassBank を構築する。生データリポジトリの MB-POST、ヒトおよびマウス由来のマスペクトルを収録した MassBank Human、予測マスペクトルを収録した MassBank in silico を新たに構築し、高品質な実測マスペクトルを MassBank に保存していく体制を構築する。ユーザー主導の開発体制で、生命科学研究者などのユーザーニーズを満たす次世代データベースを構築していく研究開発課題である。

MB-POST は 2024 年 12 月にβ版を構築後、関係者を中心に内部テストを実施し、2025 年 10 月 8 日に 100 データセット、10,830 ファイル、1.5TB を収録した正式版サイトの公開と、チーム外からのデータ登録の受入れを開始した。MB-POST の登録データは、12 月 7 日時点で 112 データセット、12,911 ファイル、1.7TB となっている。MassBank Human では、MB-POST のヒトおよびマウスで解析された 8 件の DDA データセットを再解析し、高品質な 21,923 スペクトルを MassBank フォーマットで収録して公開し、中間評価までの目標としていた 2 万件をクリアした。MassBank in silico では、保持時間予測・複数ソフト統合のパイプラインを構築し、これまでに 9.5 万の in silico スペクトルを作成して、現在、β版として公開し、中間評価までの目標としていた in silico スペクトル 5 万化合物を大きく上回った。MassBank のデータ拡充については、1,024 スペクトルを専門家レビューの承認を経て MassBank に登録し、こちらも中間評価までの目標としていたレビュー済みスペクトル 1,000 件以上を上回った。MassBank の普及については、学会等での講習会や展示等が多数実施された。ツール開発と RDF 化を含む MassBank Links については、マスペクトルデータの解釈作業用のソフトウェア MS-DIAL/MS-FINDER の MassBank フォーマットへの対応を行い、生成プログラムを GitHub で公開するとともに、作成した RDF を RDF ポータルに登録した。そのほか、サーバー検索速度を最大 1,000 倍に高速化し、MassBank Human/in silico へ実装したほか、合計で 160 種類の脂質クラスの in silico MS/MS スペクトルの生成を完了した。

今後は、MB-POST の国際リポジトリ連携参画を目指すとともに、MassBank Human の再解析パイプラインの自動化ならびに 10 万以上のスペクトルの公開、MassBank in silico の 20 万化合物の構造化を実施し、MassBank に 2,000 件以上の高精度スペクトルを登録することを目指す計画である。

上述の通り、進捗は順調で実施計画は着実に進められており、進捗状況と今後の成果見込みは優れている。MB-POST の迅速な立ち上げ、外部からのデータ登録が開始された点を評価する。生体由来スペクトルを高精度化して、これまで標準品を中心に拡張されてきた MassBank のデータを補完・拡張していくというアプローチは、利用者視点に立ったユニークでかつ実用性が高い取組みである。特に、MassBank Human と MassBank in silico で公開されている中間データ群は、実験で検出されたスペクトルの同定や解析を行おうとしているメタボロミクス研究者にとって価値が高い。それはβ版公開ながらアクセス数の急速な伸びにも繋がっているものと思われる。本プロジェクトの認知度はまだまだ高くはないが、産学・大型プロジェクトへの波及、応用例の創出が期待される。今後、論文発表や関連学会・出版社への働きかけにより、国際的にも認知され存在感を示すようになることを期待する。

以上