

2025 年度中間報告書

ライフサイエンスデータベース統合推進事業 統合化推進プログラム

【研究課題名】「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」

研究代表者：

松田 史生（大阪大学 大学院情報科学研究科 教授）

研究分担者：

奥田 修二郎（新潟大学 医学部 教授）

平山 明由（慶應義塾大学 先端生命科学研究所 准教授）

和泉 自泰（九州大学 生体防御医学研究所 准教授）

津川 裕司（東京農工大学 グローバルイノベーション研究院 教授）

1. 概要

(1) 研究および計画の概要

MassBank を刷新し、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、Shin-MassBank を構築する。生データリポジトリの MB-POST と予測マススペクトルを収載した MassBank in silico を新たに構築し、高品質な実測マススペクトルを MassBank に保存していく体制を作る。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する。

(2) 成果の概要

研究開始時に掲げた中間目標を達成した。開発項目 1 MB-POST の構築では、2024 年 12 月に MB-POST β 版サーバーを公開した。開発項目 2 MassBank Human の構築では、サーバーを構築し、ヒト、マウスの高品質スペクトルデータ 2 万件以上を作成し、MassBank Human サーバーから公開した。開発項目 3 MassBank in silico の構築では、in silico スペクトル 9.5 万化合物群を作成し、MassBank in silico サーバーから公開した。開発項目 4 MassBank のデータ拡充・普及・国際連携では、高品質スペクトルデータのレビュー作業を行い、レビュー済みスペクトル 1275 件を MassBank に登録した。開発項目 5 MassBank Links ポータルの開発では、MassBank データの RDF 化の指針を決定し、第一版を RDF ポータルに掲載した。生成プログラムも GitHub で公開した。

名称	概要
Shin-MassBank	本プロジェクトへの入り口となる WEB ポータル
MB-POST	生体サンプル由来の生マススペクトルデータレポジトリ
MassBank Human	ヒトを中心とした生体由来高品質マススペクトルライブラリ
MassBank in silico	低分子化合物構造から in silico 予測したマススペクトルのライブラリ
MassBank Links	MassBank、MB-POST と他 DB を連携する RDF およびツールの開発

2. 目的・目標の達成状況

(1) 中間評価時までの達成目標と達成状況

中間評価時までの達成目標	達成状況
【項目 1】 MB-POST の構築	
サーバー導入	サーバーを新潟大学に設置済
β版構築公開（メタ情報がプロテオミクス）	2024年12月24日にテスト環境を公開済
初期データアップロード	98プロジェクト1.5TBのデータをアップロード済
メタボロミクス用メタデータ策定	67項目のメタデータを策定済
公開版構築、公開	2025年10月8日にMB-POSTへのリンクを公開済
【項目 2】 MassBank Human の構築	
MassBank Human サーバー構築、公開	レンタルサーバー上に構築、公開済
スペクトル再解析法の開発：平均化	Python モジュール ddatoolbox として作成済
スペクトル再解析法の開発：組成式推定	ddatoolbox に機能追加済
スペクトルデータ再解析（ヒト、マウス）	78 DDA データセットの再解析済
他生物種への展開（腸内細菌、植物、環境、薬物など）	2025年の中間評価以降に実施する計画
【項目 3】 MassBank in silico の構築	
化合物表記項目の情報整理	24種の compound identifier を選定済
in silico MS/MS スペクトルライブラリの開発	4ソフトウェアを統合した手法を開発済
保持時間予測機能の開発	Unified-HILIC/AEX/MS の保持時間予測が完了
アノテーション法の開発	保持時間予測を統合した手法を開発済
β版 MassBank in silico サーバーの構築、公開	合計9.5万 in silico MS/MS スペクトルを登録済
公開版 MassBank in silico サーバーの構築、公開	in silico スペクトルの拡充に着手
MassBank および MassBank in silico サーバーの高速化	最大1000倍の高速化を達成し、MassBank Human および MassBank in silico に実装済
MS-DIAL/MS-FINDERのMassBankレコードフォーマット対応	対応済。対応したバージョンを公開済
【項目 4】 MassBank のデータ拡充・普及・国際連携	
MassBank データ登録	1024件のMassBankデータを登録済
ユーザー講習会	総計4回実施
国際連携（国際学会参加、国際シンポ）	国際学会発表を総計4回実施、MassBank EU チームとのミーティングをこれまで計6回実施
展示出展	メタボロームシンポジウムに3回展示を出展
アドバイザリ	これまでに4回実施
【項目 5】 MassBank Links ポータルの開発	
MassBank レコードの RDF 化の推進	RDF ポータルに登録済み
脂質 in silico MS/MS スペクトルの開発推進	33クラスを追加して公開済

MS-DIAL/MS-FINDERとMassBankを繋ぐプログラム開発	フラグメント注釈情報の出力機能を追加
MB-POSTの再解析基盤の構築	MS-DIALをLinuxに移植開始。

(2) 実施状況の詳細

【項目1】MB-POSTの構築

(項目1.1) サーバー導入

リポジトリとして機能するために必要な条件を満たしたストレージ部を持つサーバーを構築した。サーバー本体は新潟大学内の奥田研究室のサーバー室に設置した。

(項目1.2) β版構築公開（メタ情報がプロテオミクス）

MB-POST (<https://repository.massbank.jp>) として構築し、2024年12月24日にテスト環境として外部へのサービスを開始した。jPOSTリポジトリのプロテオミクス用メタ情報を基本として、メタボローム用メタデータに最適化した形で開発した。

(項目1.3) 初期データアップロード

アップロードの検証として、各グループが持つサンプルデータでアップロードのテストを実施した。2025年10月1日現在で、98プロジェクト10830ファイル1.5TBのデータをアップロードした。

(項目1.4) メタボロミクス用メタデータ策定

メタボローム用のメタデータおよびそのオントロジーについて、プロジェクト8件、ファイル2件など総計67件の入力項目を設定し、その中で計14種類のオントロジーと上記CVを用いた形で整備し、MB-POSTに実装した。

(項目1.5) 公開版構築、公開

システムのバグ修正等の調整期間として、プロジェクトメンバーを中心に関連する人に登録作業を進めてもらった結果、2025年10月8日に100プロジェクトの登録が完了した。概ね、改良すべき点の修正も完了していることから、これをもって公式サイト (<https://shin.massbank.jp/>) 上のMB-POSTのリンクを更新し、公式サイトでの公開とした(図1)。

The screenshot shows the MB-POST website interface. At the top, there is a navigation bar with the MB-POST logo and links for Data list, Submit, Mypage, and Help. On the right side of the navigation bar, there are buttons for Login and Signup. Below the navigation bar, there are two main sections: 'About MB-POST' and 'Statistics'. The 'About MB-POST' section contains a paragraph describing the repository. The 'Statistics' section displays two key metrics: '100 project are registered. 37 are opened.' and '11003 files amount to 1.5 TB.'

図1 公開したMB-POST (<https://repository.massbank.jp>)

【項目 2】 MassBank Human の構築

(項目 2.1) MassBank Human サーバー構築、公開

MassBank Human サーバー (<http://human.massbank.jp>)を構築、公開した。再解析した 8 データセット、合計 21,923 スペクトルを MassBank Human 及び GitHub (<https://github.com/Shin-MassBank/MassBank-Human>) にて公開した。

(項目 2.2) スペクトル再解析法の開発：平均化

スペクトル平均化作業用の Python スクリプトを拡張し、ddatoolbox という Python モジュールを作成した。また、類似のスペクトルをクラスター化するヒューリスティックな手法を実装し、100 データファイルからの平均化作業を可能とした。

(項目 2.3) スペクトル再解析法の開発：組成式推定

スペクトルのプリカーサーイオンのイオンフォーム ($[M+H]^+$, $[M+Na]^+$)を推定する機能、機能精密質量から化合物の組成式を推定する機能を ddatoolbox に追加した。再解析パイプライン要素技術の構築および実装を進め、図 2 に示す再解析パイプラインを構築した。

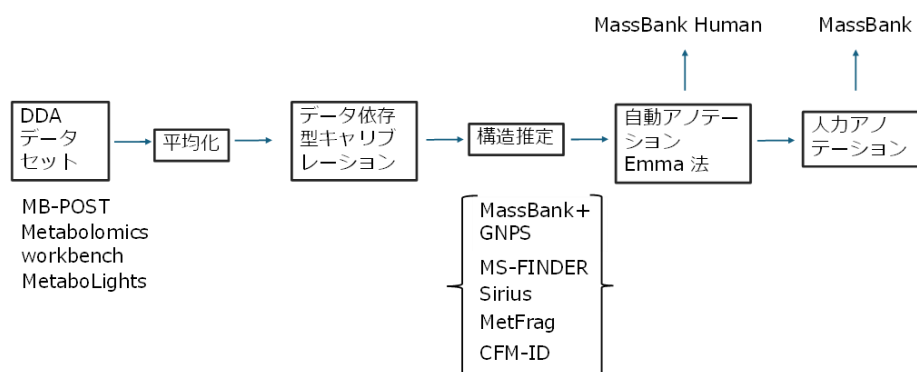


図 2 構築中の再解析パイプライン

(項目 2.5) スペクトルデータ再解析 (ヒト、マウス)

開発した再解析パイプラインを用いて、MB-POST および海外データリポジトリ (Metabolomics Workbench や MetaboLights)のヒト、マウス由来の 78 DDA データセットの再解析を行った。

(項目 2.6) 他生物種への展開 (腸内細菌、植物、環境、薬物など)

これまでのアドバイサリ委員会の意見や他プロジェクトの状況調査から、バイオものづくりプロセスの培地成分の解析にユーザーニーズがあることから、2025 年の中間評価以降に、NEDO プロジェクトと連携した展開を検討する。

【項目 3】 MassBank in silico の構築

(項目 3.1) 化合物表記項目の情報整理

19 種の既存の低分子化合物のデータベースに掲載されている compound identifier の項目調査を行い、分子化合物同定および他のデータベース連携により重要と考えられる 24 種の identifier を選定し

た。今後、登録する化合物に対してこれらの情報の記載・紐づけを実装する予定である。

(項目 3.2) in silico MS/MS スペクトルライブラリの開発

MetFrag, CFM-ID, MS-FINDER SIRIUS を統合した構造アノテーションパイプラインの開発を完了させ、MassBank Human の再解析パイプラインに実装した。

(項目 3.3) 保持時間予測機能の開発

親水性代謝物を包括的に測定可能な新規のメタボローム分析法である unified-HILIC/AEX/MS の開発に成功し、現在、松田 G や平山 G で汎用的に使用されている LC/MS 測定法における低分子化合物の保持時間予測モデルの構築を進めている。

(項目 3.4) アノテーション法の開発

in silico MS/MS スペクトルライブラリおよび unified-HILIC/AEX の保持時間予測を統合したアノテーション法の開発に成功した。

(項目 3.5) β版 MassBank in silico サーバーの構築、公開

レンタルサーバーを取得し、MassBank サーバースystemを改変することで MassBank in silico サーバー (<http://insilico.massbank.jp/>) のβ版を構築した。現在までに、3.2 で開発した 9.5 万種の親水代謝物の既存ルールに従った in silico MS/MS スペクトル情報を MassBank in silico サーバーに登録し、公開した。



Structure Annotation

About

Core Project

MassBank in silico is a library of MS/MS spectra predicted in silico from compound structures. This web tool has been developed as part of the Shi-MassBank project. Users can input exact (monoisotopic) mass and MS/MS spectra in text format, and compare them with in-silico predicted MS/MS spectra generated from human-related metabolites registered in HMDB and KEGG, enabling structure annotation. The tool is accessible via "Structure Annotation" and is freely available without registration or fees.

図3 公開したβ版 MassBank in silico サーバー(<http://insilico.massbank.jp/>)

(項目 3.6) 公開版 MassBank in silico サーバーの構築、公開

β版 MassBank in silico サーバーの開発を進めながら、2027 年度に in silico スペクトル 20 万化合物群を公開版 MassBank in silico サーバーから公開する予定。

(項目 3.7) MassBank および MassBank in silico サーバーの高速化

MassBank サーバーの高速化には、これまでの MassBank の形式を踏襲しながら最初からコードを書き直した方がスムーズであると結論付けた。ヨーロッパの MassBank グループから了承も頂き、開発を開始した。開発版の検索速度は、従来版と比べて、最大 1000 倍の高速化を達成し、当該機能を MassBank Human および MassBank in silico に実装した。

（項目 3.8）MS-DIAL/MS-FINDER の MassBank レコードフォーマット対応

マススペクトルデータの解釈作業用のスタンドアロンのソフトウェアである MS-DIAL/MS-FINDER に MassBank レコードフォーマットの書き出し機能を追加した。

【項目 4】 MassBank のデータ拡充・普及・国際連携

（項目 4.1） MassBank データ登録

MassBank Human に蓄積した高品質マススペクトルデータの中から、化合物構造とオントロジーが十分に付与された MassBank レコードを集めた。2025 年 5 月 19, 20 日に東京農工大にて代謝物アノテーション結果の専門家レビューを実施し、1275 件を承認し、MassBank に提供することとした。

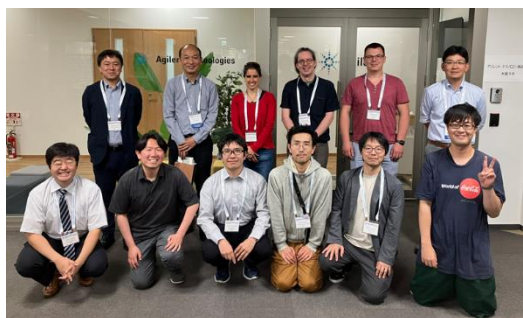
（項目 4.2） ユーザー講習会

日本質量分析学会および日本バイオインフォマティクス学会と連携してこれまでに 4 回実施した。

- 1) 松田史生、MassBank 入門中級編「レコードデータ作成」とシン・マスバンクの紹介、第 4 回日本質量分析学会スペクトルデータ部会シンポジウム、オンライン、2023 年 11 月 22 日（参加者 50 名程度）
- 2) 松田史生、講演「マススペクトルデータベース MassBank の使い方」、第 5 回スペクトルデータ部会シンポジウム、大阪大学 基礎工学研究科（ハイブリッド）、2024 年 5 月 15 日（参加者 90 名程度）
- 3) 松田史生、技術講習「マススペクトルデータ共有に向けた連携の最前線」、第 6 回日本質量分析学会スペクトルデータ部会、オンライン、2025 年 3 月 14 日（参加者 100 名程度）
- 4) 高橋 悠志、奥田修二郎、「シン・MassBank データリポジトリ利用講習会」、質量分析インフォマティクス研究会・第 10 回公開ワークショップ、東京農工大学、2025 年 8 月 1 日（参加者 46 名）

（項目 4.3） 国際連携（国際学会参加、国際シンポ）

海外の MassBank コミュニティーとの国際連携を進めるために国際学会発表を総計 4 回行った。また、ドイツ Leibniz Institute of Plant Biochemistry の Steffen Neumann をリーダーとする MassBank EU チームメンバーとのオンラインミーティングをこれまで計 6 回行い、マスバンクレコードフォーマットの拡張に関する意見交換を行った。うち一回は、Metabolomics 2024 に参加するため来日したメンバーらと対面での意見交換を実施できた（右写真 2024 年 6 月 18 日@大阪）。



（項目 4.4） 展示出展

最もコアなユーザーのコミュニティとして想定しているメタボロームシンポジウムにおいて 3 回展示を行った。大口のユーザーとして想定している下記研究者らにデモを行い必要な機能などについての情報を収集した。メタボロームデータの MB-POST への登録をテストケースとして実施するなどの成果に

つながっている。

第 17 回メタボロームシンポジウム 2023 年 10 月 18～20 日 川崎 Shimadzu Tokyo Innovation Plaza

第 18 回メタボロームシンポジウム 2024 年 10 月 23–25 日 鶴岡メタボロームキャンパス

第 19 回メタボロームシンポジウム 2025 年 10 月 15–18 日 神戸大学

(項目 4.5) アドバイザリ

これまでアドバイザリ委員会を計 4 回開催し、成果及び実施計画を説明した(4(2)有識者会議参照)。2025 年 1 月 30 日アドバイザリ委員会では 10 件以上にわたるコメントが挙がり、研究開発へ反映させた。

【項目 5】 MassBank Links ポータルの開発

(項目 5.1) MassBank レコードの RDF 化の推進

DBCLS が開催する Togothon および国内版バイオハッカソンに積極的に参加し、川島秀一研究員にアドバイスをもらいながら MassBank データの RDF 化を推進した。第一版として、RDF ポータルに掲載済みであり (<https://rdfportal.org/>)、生成プログラムも GitHub にて公開している (<https://github.com/systemsomicslab/MassBank-RDF>)。

(項目 5.2) 脂質 in silico MS/MS スペクトルの開発推進

33 種類の脂質クラスを追加し、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。in silico MS/MS は MS-DIAL に内蔵して公開した。和泉 G とも共有を完了した。

(項目 5.3) MS-DIAL/MS-FINDER と MassBank を繋ぐプログラム開発

項目 3.8 で開発した機能をもとに、MS-FINDER ではフラグメント注釈を付与したレコードの書き出しを可能とし。MS-FINDER のバージョンアップ時に機能追加した。MassBank-EU が提供する Validator による検証も行なった。

(項目 5.4) MB-POST の再解析基盤の構築

Windows OS 専用プログラムであった MS-DIAL を LINUX に移植し、GC/MS および LC-MS/MS データを解析可能とした。

(3) 主な成果論文等

1. Fumio Matsuda: Data Processing of Product Ion Spectra: Methods to Control False Discovery Rate in Compound Search Results for Untargeted Metabolomics. Mass Spectrometry 13, 2024, 1, A0155 <https://doi.org/10.5702/massspectrometry.A0155>

[概要] 本プロジェクトで開発しているデータ再解析パイプラインでは、専用ソフトウェアを用いて化合物構造推定(アノテーション)を行う。そこで、アノテーション結果中の False Discovery Rate を推定する手法を開発した。プロテオミクスのデコイ法を直接使えないため、入力データのデコイを用いる手法、検索結果の 2 番目ヒットを用いる FDR 推定法を開発した。開発した手法は再解析データの品質管理に利用している。

2. Yoshihiro Izumi et al.: Predicting retention time in unified-hydrophilic-interaction/anion-exchange liquid chromatography high-resolution tandem mass spectrometry (unified-HILIC/AEX/HRMS/MS) for comprehensive structural annotation of polar metabolome. Anal. Chem., 96(3), 1275-1283 (2024). <https://doi.org/10.1021/acs.analchem.3c04618>

【概要】 in silico MS/MS スペクトルライブラリおよび親水性代謝物を包括的に計測可能な unified-HILIC/AEX の保持時間予測情報を加味した構造アノテーション法の開発を行った。ヒト血漿メタボロミクスデータに対し、in silico MS/MS prediction と保持時間予測による構造アノテーションを行った結果、216 種類の代謝特性に対する構造アノテーションに成功した。特に、in silico MS/MS prediction によって割り当てた構造候補は、保持時間予測を加味することで約 50%が偽陽性として削減された。

(4) 主要なデータベースの利活用状況

MB-POST は 2024 年末に公開度、2025 年にはチーム内で公開し、仕様をすすめた。他の DB も総計でも訪問者数が 50%以上増加した。今後も 2025 年 10 月にメタボロミクスシンポジウム出展などでアピールに努め、アクセス数増加を図る。

(5) データベースを利用して得られた研究成果・産業応用の例

1. NEDO バイオものづくりプロジェクト（カーボンリサイクル実現を加速するバイオ由来製品生産技術の開発、https://www.nedo.go.jp/activities/ZZJP_100170.html）では、一気通貫で生産プロセスを検証し評価サンプルを創出できるバイオ生産システム基盤の構築とその周辺技術開発を行っている。その中で、質量分析法を用いたバイオ生産プロセス中の未知物質探索法の開発が進んでおり、MassBank silico を既知代謝物の構造同定に用いる研究が行われ、発酵阻害などの探索に成果を挙げている。
2. JST 革新的 GX 技術創出事業 (GteX) のバイオものづくりチームでは、バイオものづくりのボトルネックとなっている技術課題への対応や独創的な発想の下でゲームチェンジとなるような技術開発を行っている。その中で、バイオ生産微生物中の代謝ボトルネック同定のために、メタボローム解析を実施しており、そこで得た生データの公開に MB-POST を利用する方針を採用し、すでにデータの登録が開始されている。

3. 今後の計画および展望

実用化に向けた展開や社会還元的な展開活動として、NEDO バイオものづくりプロジェクトと共同して、バイオプロセス構築の課題解決に MassBank in silico を使い、その有用性を実証する。JST 革新的 GX 技術創出事業 (GteX) のバイオものづくりチームと連携して、MB-POST をメタボロームデータ公開のためのレポジトリとして活用する。

開発項目 1 MB-POST は、メタボローム版 PXC の傘下で、必要とされる機能を実装し、世界基準のメタボ

ロームレポジトリとして認められるよう活動する。また、論文や学会等での広報活動を進める、世界中のメタボローム研究者に必要とされるメタデータを実装した MB-POST サーバーを公開する。

開発項目 2 MassBank Human の構築では、再解析パイプラインの開発をさらに進め、パイプライン全体の自動化を進める。さらに、他生物種由来のスペクトルデータの再解析を行い、高品質スペクトルデータ 10 万件以上を作成し、MassBank Human サーバーから公開する。

開発項目 3 MassBank in silico の構築では、in silico MS/MS スペクトル予測法の開発をさらに進め、予測精度の向上を図る。サーバーの速度向上を行う。in silico スペクトル 20 万化合物群を作成し、MassBank in silico サーバーから公開する。サーバーを従来比 10 倍に高速化する。

開発項目 4 MassBank のデータ拡充・普及・国際連携では、MassBank Human に蓄積した高品質マススペクトルデータの中から、専門家レビューで承認されたものを MassBank に登録する作業を行う。レビュー済みスペクトルを 2000 件以上 MassBank へ登録する。また、普及・国際連携を進める。

開発項目 5 MassBank Links ポータルの開発では、質量分析の一次データと二次データの連携を RDF により実現し、MB-POST の CV とも連携させて二次データ抽出を加速させるプログラム開発も同時に行う。二次データ抽出においては、RDF および再解析パイプラインの有用性実証に向けた基礎検討を行う。

4. 計画・実施体制等の妥当性

(1) 各グループの担当項目

(1)-1. 松田グループ（大阪大学）

主に、開発項目 2 MassBank Human の構築を担当した。スペクトル再解析法の開発を開発し、19 個のメタボロームデータセットから 1275 件のアノテーション済データを作成し、MassBank に提供した。

(1)-2. 奥田グループ（新潟大学）

主に、開発項目 1 開発項目 1 MB-POST の構築を担当した。プロテオミクスデータレポジトリである jPOST をもとに、メタボロミクス用メタデータ策定、β 版構築公開を行った。

(1)-3. 平山グループ（慶應義塾大学）

主に、開発項目 2 MassBank Human の構築を担当した。公開メタボロームデータセットから、再解析パイプラインを用いて、データを生成し、MassBank Human サーバー β 版から公開した。

(1)-4. 和泉グループ（九州大学）

主に、開発項目 3 MassBank in silico の構築を担当した。特に、in silico MS/MS スペクトルライブラリ開発と β 版 MassBank in silico サーバー構築、公開を行った。

(1)-5. 津川グループ（東京農工大学）

主に、開発項目 5 MassBank Links ポータルの開発を担当した。特に、MassBank レコードの RDF 化を進め、方針決定および TTL ファイルの生成、RDF ポータルからの公開を行った。