

2023 年度 研究開発実施報告

概要

研究開発課題名	次世代低分子 マス スペクトルデータベース シン・マスバンクの構築
開発対象データベースの名称(URL)	Shin-MassBank (https://shin-massbank.jp)
研究代表者氏名	松田 史生 (50462734)
所属・役職	大阪大学 大学院情報科学研究科 教授(2024 年 3 月時点)



©2024 松田 史生(大阪大学) licensed under CC表示4.0国際

目次

概要	1
目次	2
§1. 研究実施体制	3
§2. 研究開発対象とするデータベース・ツール等	3
(1) データベース一覧	3
【主なデータベース】	3
【その他のデータベース】	3
(2) ツール等一覧	3
§3. 実施内容	4
(1) 本年度の研究開発計画と達成目標	4
開発項目1 MB-POST の構築	4
開発項目2 MassBank Human の構築	5
開発項目3 MassBank <i>in silico</i> の構築	5
開発項目4 MassBank のデータ拡充・普及・国際連携(松田、奥田、平山、和泉)	6
(2) 進捗状況	8
開発項目1 MB-POST の構築	8
開発項目2 MassBank Human の構築	9
開発項目3 MassBank <i>in silico</i> の構築	10
開発項目4 MassBank のデータ拡充・普及・国際連携(松田、奥田、平山、和泉)	16
§4. 成果発表等	19
(1) 原著論文発表	19
(2) その他の著作物(総説、書籍など)	19
(3) 国際学会および国内学会発表	20
(4) 知的財産権の出願 (国内の出願件数のみ公開)	22
(5) 受賞・報道等	22
§5. 主要なデータベースの利活用状況	23
(1) アクセス数	23
(2) データベースの利用状況を示すアクセス数以外の指標	23
(3) データベースの利活用により得られた研究成果(生命科学研究への波及効果)	23
(4) データベースの利活用によりもたらされた産業への波及効果や科学技術のイノベーション(産業や科学技術への波及効果)	23
§6. 研究開発期間中に主催した活動(ワークショップ等)	24
(1) 進捗ミーティング	24
(2) 主催したワークショップ、シンポジウム、アウトリーチ活動等	24

§1. 研究実施体制

グループ名	研究代表者または主たる共同研究者氏名	所属機関・役職名	研究題目
松田グループ	松田 史生	大阪大学・教授	MassBankのデータ拡充とスペクトルデータ高品質化法の開発
奥田グループ	奥田 修二郎	新潟大学・教授	メタボロームデータリポジトリの開発とデータ拡充
平山グループ	平山 明由	慶應義塾大学・准教授	MassBank Humanの構築とデータ拡充
和泉グループ	和泉 自泰	九州大学・准教授	MassBank <i>in silico</i> の構築とデータ拡充
津川グループ	津川 裕司	東京農工大学・准教授	MS-DIAL, MS-FINDERのMassBank対応

§2. 研究開発対象とするデータベース・ツール等

(1) データベース一覧

【主なデータベース】

No.	名称	別称・略称	URL
1	Shin-MassBank	シン・マスバンク	https://shin.massbank.jp

【その他のデータベース】

No.	名称	別称・略称	URL
1			

(2) ツール等一覧

No.	名称	別称・略称	URL
1	スペクトル二次解析ソフトウェア		https://github.com/fumiomatsuda/Data-processing-of-product-ion-spectra
2	インシリコスペクトル生成、構造推定ソフトウェア		現時点で未公開
3	MS-DIAL/MS-FINDER		https://github.com/systemsomicslab/MsdialWorkbench

§3. 実施内容

(1) 本年度の研究開発計画と達成目標

当初計画からの変更点:追加予算にて、3.8 MS-DIAL/MS-FINDER の MassBank レコードフォーマット対応(東京農工大津川)を追加実施した。

実施する研究開発項目	タスク
開発項目1 MB-POST の構築	・1.1 サーバー導入 ・1.2 8 版構築公開(メタ情報がプロテオミクス)
開発項目2 MassBank Human の構築	・2.1 MassBank Human サーバーの 8 版を構築 ・2.1 2.5 で作成した高品質スペクトルデータを登録してテスト ・2.2 スペクトル再解析法の開発(平均化法の拡張)と評価 ・2.5 スペクトルデータ再解析のテスト
開発項目3 MassBank <i>in silico</i> の構築	・3.1 化合物表記項目の情報整理 ・3.2 <i>in silico</i> MS/MS スペクトルライブラリの開発 ・3.5 8 版 MassBank <i>in silico</i> サーバーの構築 ・3.7 MassBank サーバーの高速化 ・3.8 MS-DIAL/MS-FINDER の MassBank レコードフォーマット対応(追加予算にて追加実施)
開発項目4 MassBank のデータ拡充・普及・国際連携	・4.1 MassBank データ登録テスト ・4.2 MassBank のユーザー講習会を実施 ・4.3 国際連携(Metabolomics 学会への参加) ・4.4 展示出展(メタボロミクスシンポジウム) ・4.5 アドバイザリ年 2 回(6 月、12 月)

マイルストーン

2023 年度	MassBank Human, MassBank <i>in silico</i> サーバーの構築
---------	---

開発項目1 MB-POST の構築

生体サンプル由来の生マススペクトルデータを収集し、それに必要なデータデポジット、閲覧、ダウンロード機能を持った物理サーバーの構築、メタデータ形式の設定を行う。現状の課題である、メタデータ、オントロジーの整備を進めるため、プロテオミクスデータレポジトリである jPOST が構築してきたシステムを活用する。メタボロミクスとプロテオミクスではデータ取得方法に 90%程度の共通点があるため、プロテオミクスと協調したメタデータ形式を採用して、将来的なマルチオミクスデータへの対応を可能にする。

1.1 サーバー導入(奥田 G)

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の構築を実施する。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する。jPOST リポジトリや GlycoPOST で実現しているシステムを導入し、稼働させる。

1.2 B 版構築公開(メタ情報がプロテオミクス)(奥田 G)

MB-POST のリポジトリとしての機能について、外部からアクセスして利用できる状態での運用を開始できる状態になるよう開発を実施する。当初の受入れメタ情報としては、jPOST リポジトリが採用しているプロテオミクスでのメタ情報を基本として設計する。

開発項目2 MassBank Human の構築

ヒト由来高品質マスマスペクトルライブラリの MassBank Human を構築する。MB-POST に集積した生体由来の代謝物のマスマスペクトルデータには、ノイズが多い、冗長性が高い、計測精度が低いなどの課題があり、検索時に擬陽性を生む原因となっている。そこで、ユーザーニーズの最も高いヒトに注目し、生マスマスペクトルデータの積算を行うことで、ノイズ・冗長性の除去、計測精度を向上した高品質スペクトルデータを生成する再解析法を構築する。作成した高品質スペクトルデータを公開し、ユーザーが取得したマスマスペクトルの過去の検出例を検索できる新たな機能を提供する。

2.1 MassBank Human サーバー構築、公開(平山、和泉 G)

レンタルサーバーを取得し、MassBank サーバーシステムを改変することで MassBank Human サーバーの B 版を構築する。2.5 スペクトルデータ再解析(ヒト、マウス)で作成したヒト、マウス由来の高品質スペクトルデータを登録してテストを行う。平山 G と和泉 G が連携して実施する。

2.2 スペクトル再解析法の開発:平均化(松田、平山 G)

これまで作成したスペクトル平均化作業用の Python スクリプトを拡張する。とくに類似のスペクトルをクラスター化する部分がボトルネックとなっているため、ヒューリスティックな手法を用いて 100 データファイルからの平均化作業を可能とする。また作業結果を検証するための支援ツールを作成する。松田 G がツールを開発し、平山 G が評価を行う。

2.5 スペクトルデータ再解析(ヒト、マウス)(平山、松田、和泉 G)

現状利用可能なスペクトル再解析法を用いて、ヒト、マウス由来のスペクトルデータの再解析を行う。初年度はまず、スペクトル再解析のテストを行うために、迅速に利用可能な海外のデータレポジトリから利用可能な小規模(〜30 ファイル程度)なデータセットや、和泉 G、平山 G ですでに取得したマウス由来のデータセットを用いる。また作成したマスマスペクトルの品質を評価する。松田 G が再解析を行い、平山 G と和泉 G が評価を行う。

開発項目3 MassBank *in silico* の構築

MassBank Human に蓄積した高品質マスマスペクトルデータに、構造情報を付与する解析パイプラインを構築する。まず化合物構造から *in silico* 予測したマスマスペクトルのライブラリである MassBank *in silico* を構築、公開する。これをもとにクエリのスペクトルに化合物構造を付与する手法と化合物クラス(オ

ントロジ)を付与する手法を開発し、ユーザーがマススペクトルに構造情報を付与する機能を提供する。また、MassBank Human に蓄積した高品質マススペクトルデータに、化合物構造とオントロジーの付与を行う。さらに、大量の *in silico* 予測スペクトルを取り扱うために MassBank サーバーの高速化を行う。

3.1 化合物表記項目の情報整理 (和泉 G)

MassBank *in silico* に掲載する化合物表記項目の情報整理を行う。具体的には、統一した SMILES 表記による化合物情報の管理と線形表記、化合物名の統一および表記のミスチェックなどを実施する。

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

HMDB や KEGG に登録されている代謝物を対象に、高精度の *in silico* MS/MS 予測スペクトルを算出する。今年度は、結合解離エネルギーと水素再配置規則に基づき、定性的なフラグメント情報を算出し、*in silico* MS/MS スペクトルライブラリへの登録を開始する。

3.5 8 版 MassBank *in silico* サーバーの構築、公開 (和泉 G)

レンタルサーバーを取得し、MassBank サーバーシステムを改変することで MassBank *in silico* サーバーの 8 版を構築する。

3.7 MassBank および MassBank *in silico* サーバーの高速化 (和泉 G)

既存の MassBank の高速化を図るために、GitHub のソースコードの一部書き換えを行う。

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

【2023 年度追加実施項目】

2023 年度 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として、高い性能を持つ MS-DIAL/MS-FINDER に MassBank レコードフォーマットの読み込み、書き出し機能を追加する(読み込みは MS-FINDER のみ)。

開発項目 4 MassBank のデータ拡充・普及・国際連携 (松田、奥田、平山、和泉)

MassBank のデータ拡充、普及、国際化を行う。MassBank Human に蓄積した高品質マススペクトルデータの中から、MassBank *in silico* で化合物構造とオントロジーが十分に付与されたものを集め、年に 1-2 回プロジェクト参加者による専門家レビューを行い、承認されたものを MassBank に登録する。これにより MassBank のデータを拡充させ、ライフサイエンスの基盤強化につなげる。また、本プロジェクトの普及のために国内学会での宣伝ブース出展(年 1 回程度)、国際学会での発表、ワークショップ提案(年 1-2 回程度)、ユーザーミーティング(年 1 回程度)を行う。さらに、産官学のユーザーコミュニティー

10 名程度のアドバイザリ会議を年 1-2 回実施し、ユーザーコミュニティのニーズの把握と、ユーザーコミュニティからの活発なデータ提供が可能な体制を構築する。

4.1 MassBank データ登録(6 月、12 月)

年 2 回(6 月と 12 月を想定)、本プロジェクトの研究者が対面で集まり、MassBank Human に蓄積した高品質マススペクトルデータの中から、MassBank *in silico* で化合物構造とオントロジーが十分に付与されたものを集め、年に専門家レビューを行い、承認されたものを MassBank に登録する作業を行う。初年度は 6 月のデータ登録はスキップし、12 月にテストデータを用いて作業プロトコルの作成を行う。

4.2 ユーザー講習会(年 1 回)

普及のためにユーザー講習会を年に 1 回以上実施する。日本質量分析学会スペクトルデータ部会と連携して実施する。初年度は MassBank のユーザー講習会を実施する。

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。そのために、国際学会参加、国際シンポ企画などを行う。初年度は Metabolomics 学会への参加を行う。

4.4 展示出展(年 1 件分生、マス学会、生化学)

普及のために学会への展示出典を行う。初年度はユーザーコミュニティであるメタボロミクスシンポジウムへの出展を行う。

4.5 アドバイザリ年 2 回(6 月、12 月)

10 名程度の有識者にアドバイザリ委員就任を依頼し、年 2 回(6 月と 12 月を想定)、オンライン・対面等でアドバイザリ委員会を開催する。そこでは研究開発の方向性、ユーザーコミュニティとの連携、他レポジトリとの連携について意見をいただく。

すでに、石濱泰(京都大学教授、日本質量分析学会会長、jPOST)、豊田岐聡(大阪大学教授、日本質量分析学会 Mass Spectrometry 編集長)、三枝大輔(帝京大学准教授、スペクトルデータ部会担当、東北メガバンク)、三浦大典(産総研 バイオメディカル研究部門)、岩橋福松(住友化学、創薬メタボロミクス)の 5 名からは内諾を得ている。今年度の初回のアドバイザリ委員会では他のアドバイザリ委員の選定を議論する。

(2) 進捗状況

まとめ:当初および追加計画通りの進捗があった。

実施する研究開発項目	タスク
開発項目1 MB-POST の構築	・1.1 【完了】サーバー導入 ・1.2 【完了】8 版構築公開(メタ情報がプロテオミクス)
開発項目2 MassBank Human の構築	・2.1 【完了】MassBank Human サーバーの 8 版を構築 ・2.1 【完了】2.5 で作成した高品質スペクトルデータを登録してテスト ・2.2 【完了】スペクトル再解析法の開発(平均化法の拡張)と評価 ・2.5 【完了】スペクトルデータ再解析のテスト
開発項目3 MassBank <i>in silico</i> の構築	・3.1 【完了】化合物表記項目の情報整理 ・3.2 【完了】<i>in silico</i> MS/MS スペクトルライブラリの開発 ・3.5 【完了】8 版 MassBank <i>in silico</i> サーバーの構築 ・3.7 【完了】MassBank サーバーの高速化 ・3.8 【完了】MS-DIAL/MS-FINDER の MassBank レコードフォーマット対応
開発項目4 MassBank のデータ拡充・普及・国際連携	・4.1 【完了】MassBank データ登録テスト ・4.2 【完了】MassBank のユーザー講習会を実施 ・4.3 【完了】国際連携(Metabolomics 学会への参加) ・4.4 【完了】展示出展(メタボロミクスシンポジウム) ・4.5 【完了】アドバイザリ年 2 回(6 月、12 月)

マイルストーン

2023 年度	【完了】MassBank Human, MassBank <i>in silico</i> サーバーの構築
---------	---

開発項目1 MB-POST の構築

1.1 サーバー導入(奥田 G)

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。jPOST リポジトリ及び GlycoPOST で実現しているリポジトリシステムを導入し、本サーバーで動くように調整している。現時点では図 1 のようなウェブインターフェイスがローカル環境として稼働している。

1.2 8 版構築公開(メタ情報がプロテオミクス)(奥田 G)

MB-POST のリポジトリとしての機能について外部からアクセスして利用できる状態での運用を開始できる状態になるよう開発中である。現時点の受入れメタ情報としては、jPOST リポジトリが採用しているプロテオミクスでのメタ情報を基本として実装した(図 2)。jPOST メタデータをメタボローム解析に適用させるために他のメタボロームデータベースを参考に改良案を作成している。

2.2 スペクトル再解析法の開発:平均化 (松田、平山 G)

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。また、精密質量から化合物の組成式を推定する機能を持ったソフトウェアを開発した。

これらをもとに、スペクトル平均化を担う DDAAveraging.py、作業結果を検証およびスペクトルの高精度化を担う DDAProcessing.py の2種類の Python スクリプトからなるデータ処理パイプラインを構築した。

2.5 スペクトルデータ再解析(ヒト、マウス)(平山、松田、和泉 G)

2.2 スペクトル再解析法の開発：平均化によって開発した手法を用いて、ヒト由来のスペクトルデータの再解析を行った。海外のメタボロームデータリポジトリ (Metabolomics Workbench や MetaboLights) および和泉 G で取得されたヒトサンプルの DDA 測定された 10 個のデータセットを対象に平均化を実施し、MassBank レコードを作成した。作成した MassBank レコードのうち 10 個は MassBank への追加リクエストを実施した。

開発項目3 MassBank *in silico* の構築

3.1 化合物表記項目の情報整理 (和泉 G)

MassBank *in silico* に掲載する化合物表記項目の情報整理を行った。具体的には、表 1 に示す 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 の項目調査を行った。その結果、低分子化合物同定および他のデータベース連携に重要な 57 種の表記が選定された。その中で、より重要と考えられる 24 種の identifier を選定した(表 2)。今後、MassBank *in silico* に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。

表 1 低分子化合物関連データベース情報

DataBase name	URL
Food Safety	https://ec.europa.eu/food/food-feed-portal/screen/food-additives/search
FOCUS	http://14.139.62.46:8888/focus/
ChemFOnt	https://www.chemfont.ca/unearth/q?query=%22Food+Additive%22&searcher=compounds&button=
HMDB	https://hmdb.ca/
HUMANCYC	https://humancyc.org/HUMAN/organism-summary
KEGG	https://www.kegg.jp/kegg/
MetaboLights	https://www.ebi.ac.uk/metabolights/search?organism.organismName=Homo%20sapiens
T3DB	http://www.t3db.ca/toxins
FOODB	https://foodb.ca/compounds
FooDisNET	http://140.113.120.248/FooDisNET/compound_list_json.php
VCF	https://www.vcf-online.nl/VcfCompounds.cfm
BitterDB	https://bitterdb.agri.huji.ac.il/dbbitter.php#compoundBrowse
KEGG DRUG Database	https://www.genome.jp/kegg/drug/drug_ja.html
DrugBank Online	https://go.drugbank.com/drugs
Drugs.com	https://www.drugs.com/drug_information.html
MassBank	https://massbank.jp/
GNPS	https://gnps.ucsd.edu/ProteoSAFe/static/gnps-splash.jsp
SIRIUS	https://bio.informatik.uni-jena.de/software/sirius/
MoNA	https://mona.fiehnlab.ucdavis.edu/

表 2 低分子化合物同定および他のデータベース連携に重要な表記法

Compound identifier etc.	19 種のデータベースでの頻度 (%)	コメント
Compound name (慣用名)	84.2	
Molecular weight	78.9	
SMILES	73.7	表記可能なものは Canonical SMILES と Isomeric SMILES を記載
Formular	68.4	
CAS	63.2	
ChEBI ID	57.9	
Synonyms	52.6	
PubChem CID	52.6	
InChikey	52.6	
IUPAC name	47.4	未知は SMILES⇒IUPAC 名変換、Chemaxon プラグイン購入する予定
InChi	47.4	
KEGG ID	42.1	
PDB ID	36.8	
LogP	36.8	
ChemSpider ID	31.6	標品の販売サイトとリンク
HMDB ID	26.3	
DrugBank ID	26.3	
KNAPSAcK ID	26.3	

Classy fire	26.3	代謝物クラス定義 https://cfb.fiehnlab.ucdavis.edu/
Phenol Explorer Compound ID	21.1	
FooDB ID	21.1	
BioCyc ID	21.1	
BiGG ID	21.1	
METLIN ID	21.1	
NIKKAJI ID	21.1	
Hydrogen Bond Donor count	21.1	
Hydrogen Bond Acceptor count	21.1	
LogS	21.1	
Water Solubility	21.1	
CSS value	21.1	
Marker DB ID	15.8	
PathBank ID	15.8	
Rotatable bonds	15.8	
Topological Polar Surface Area (TPSA)	15.8	
Physiological Charge	15.8	
Number of Rings	15.8	
Melting Point	15.8	
Boiling Point	15.8	
Biospecimen Locations	15.8	サンプル種の定義
Categories	10.5	
SPLASH Key	10.5	
3DMET ID	10.5	
LIPIDBANK ID	10.5	
LIPIDMAPS ID	10.5	
ZINC ID	10.5	
pKa	10.5	
Refractivity	10.5	
Polarizability	10.5	
HSDB ID	5.3	
RxNav ID	5.3	
LigandBox ID	5.3	
Binding DB ID	5.3	
MetaNetX ID	5.3	
MetaboLights ID	5.3	
Reactome ID	5.3	
RefMet ID	5.3	
ModelSEED ID	5.3	

*黄色ハイライトは重要、黄緑ハイライトは次点の候補。

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

SRM1950 ヒト血漿試料のノンターゲットメタボロームデータを用いて、標準品をベースに同定が可能な代謝物ピークに対して MS-FINDER でスコアを算出した。その結果、上位にくる代謝物もあれば、そうでないものも確認された。すなわち、MS/MS スペクトルによる部分構造情報から特に親水性代謝物の帰属精度を向上させるためには、既存の結合解離エネルギー規則と水素再配置規則に基づく *in silico* MS/MS 予測では必ずしも十分ではない(特に異性体の識別)。各種代謝物の構造情報から精度の高い

in silico MS/MS スペクトルが予想可能な方法論の構築に着手した(プロジェクト終了までの長期課題、図 4)。今年度は、上記と並行して、HMDB に登録されている約 8 万種の親水代謝物を対象に、従来手法に基づく *in silico* MS/MS スペクトルを算出した。

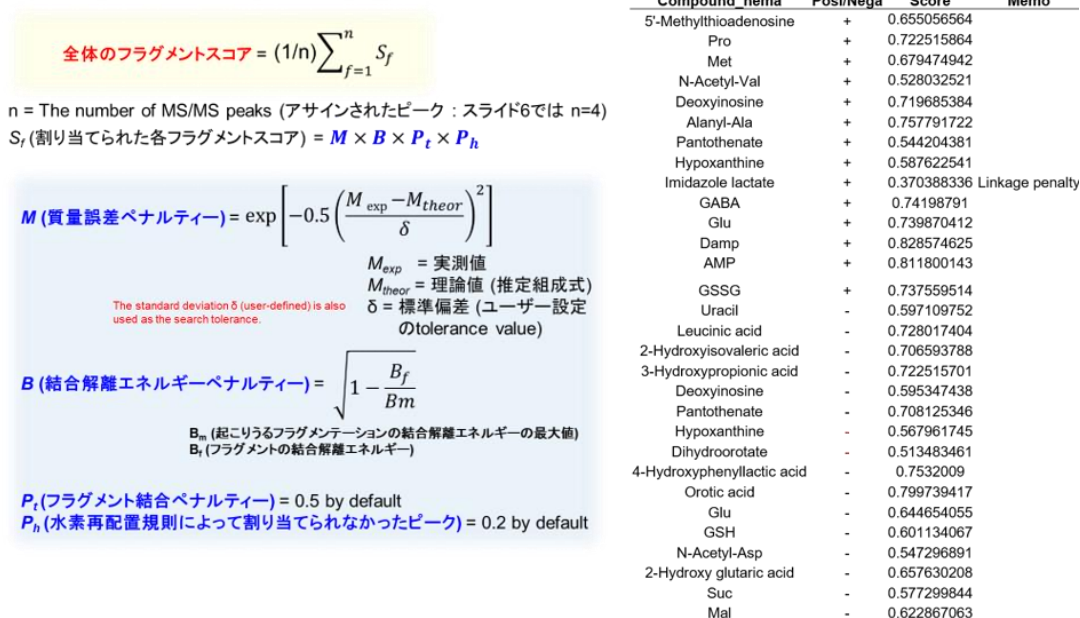


図 4 水素再配置規則に基づく *in silico* MS/MS 予測のスペック評価

3.3 保持時間予測機能の開発 (和泉 G)

昨年度、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/MS の開発に成功したため、今年度は、unified-HILIC/AEX/MS/MS の保持時間予測モデルの構築を前倒して実施した(図 5)。LC/HRMS/MS を用いたノンターゲットメタボローム解析においては、ブリカーサーイオンの精密質量と MS/MS スペクトルによる類似性により構造アノテーションが実施される。現行の構造アノテーション法では、構造異性体の関係にある偽陽性構造を含む場合が多々ある。未知代謝物ピークの実測保持時間と構造候補の予測保持時間の照合によって、代謝物の化学的性質に基づき偽陽性構造を削減することができる。しかし、昨年我々が開発した幅広い親水性代謝物を網羅的かつ一斉に分離検出できる unified-HILIC/AEX/HRMS/MS は、親水性相互作用 (HILIC) と陰イオン交換 (AEX) の分離モードを段階的かつ連続的に作用させる複雑な分離のため、本手法による代謝物ピークの保持時間予測には至っていなかった。そこで本研究では、unified-HILIC/AEX の保持時間予測モデルを構築した。トレーニングデータ (203 種類の極性代謝物) とその分子記述子 (12,420 種類) に対し、ランダムフォレストを用いて特徴量重要度をランク付けした。特徴量重要度に基づいて選出した 26 種類の分子記述子を用いて保持時間予測モデルを構築した。テストデータ (51 種類の極性代謝物) を用いて保持時間予測モデルの精度を評価した結果、 ΔRT (実測保持時間と予測保持時間の差) の 86.3% が ± 1.50 分以内、平均絶対誤差は 0.80 分であったことから、高い保持時間予測精度を示した。開発した保持時間予測モデルと *in silico* MS/MS 予測を用いて、グローバル品質管理 (global quality control, QC) 血漿として使用されている NIST SRM 1950 ヒト血漿サンプルから取得したノンターゲットメタボロームデータを解析した結果、標準物質に基づいて同定された 62 種類の既知代謝物に加え、216

種類の極性代謝物の構造アノテーションに成功した。本保持時間予測モデルは、メタボローム研究における重要な課題である未知の親水性代謝物の構造アノテーションを加速することに役立つと考えられる。

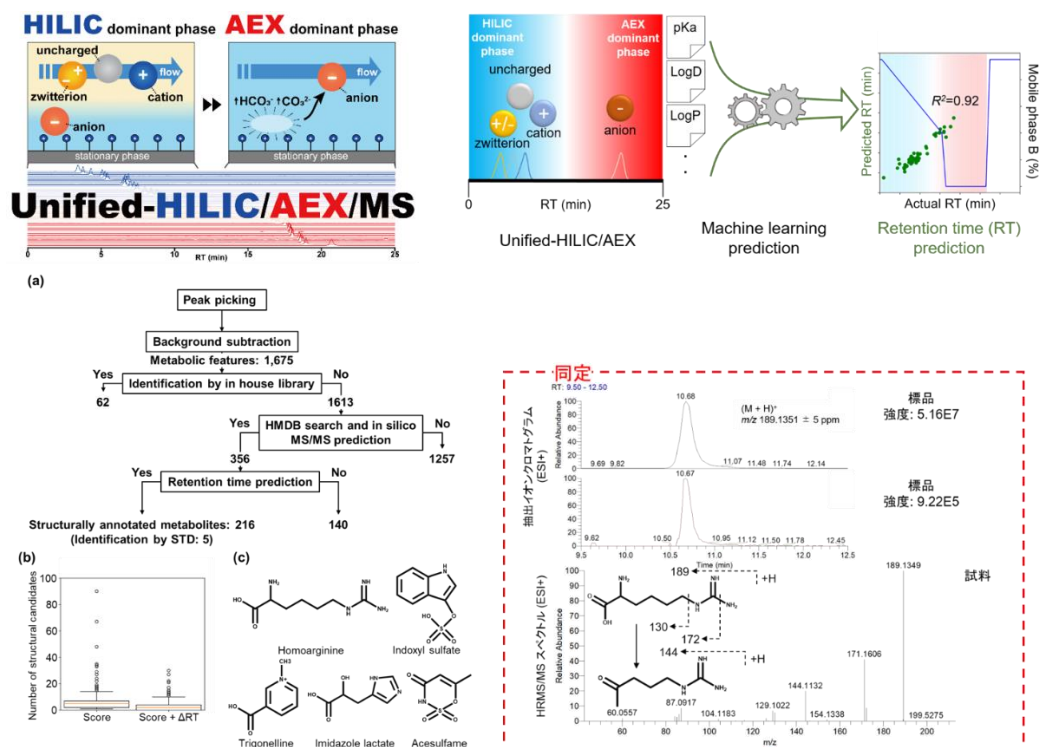


図 5 Unified-HILIC/AEX/MS/MS 保持時間モデルの開発

3.5 8 版 MassBank *in silico* サーバーの構築、公開 (和泉 G)

レンタルサーバーを取得し、MassBank *in silico* サーバーの α 版を構築した。具体的には、3.2 で開発した HMDB に登録されている約 8 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(図 6)を α 版の MassBank *in silico* サーバーに登録し、公開した(図 7)。

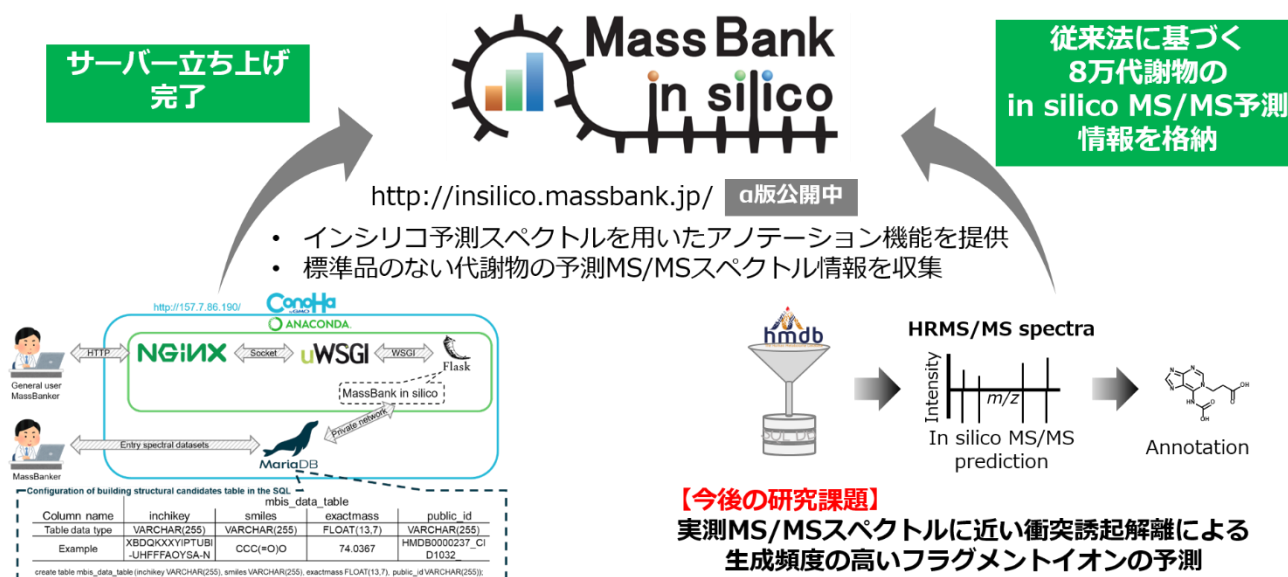


図6 α版 MassBank in silico サーバーの構築

http://insilico.massbank.jp/

1. ESI type, Exact mass, Exact mass tolerance (ppm), Fragment mass tolerance (ppm) を入力
* 現時点では
Chromatography type (IC or uni-H/A): 入力不要
Retention time tolerance (mins): 入力不要
2. Analysis start をクリックし、解析を開始する。

0: アノテーションスコア (0~1)
1: in silico フラグメントのマッチ率 (0~1)
2: 候補のSMILES
3: In silico fragment
4: in silico fragmentが割り当てられたプロダクトイオン
5: フラグメントスコア (0~1)
6: 水素再配置規則
7: In silico fragmentとproduct ionの質量誤差
8: 以降は、2~7の繰り返し

1. SMILES: 予測したい化合物のSMILESを入力する。
* 任意のSMILESはSMILES_generator等から取得可能) SMILES_generator URL:
https://www.cheminfo.org/flavor/malaria/Utilities/SMILES_generator_checker/index.html
2. Analysis start をクリックし、解析を開始する。

図7 α版 MassBank in silico サーバーの公開

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

MassBank の GitHub から取得したソースコードの問題を一部まとめた。結論としては、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ように考えられる。ヨーロッパの MassBank グループから了承も頂いているため、プロジェクト期間内で進めていく予定である。また、九大が保有する約 2,000 種の代謝物標準品の高品質のマススペクトルデータを MassBank に登録していくために、標準品の整理、オービトラップでのデータ取得(HCDを細かく設定)等を開始した。

3.8 MS-DIAL/MS-FINDER の MassBank レコードフォーマット対応(津川 G) <追加実施項目>

2023 年度追加実施項目として、マスペクトルデータ解析ソフトウェアとして、高い性能を持つ MS-DIAL/MS-FINDER に MassBank レコードフォーマットの読み込み、書き出し機能を追加した(読み込みは MS-FINDER のみ, 図 8)。MassBank Human プロジェクトを円滑に推進することを志向し、生体試料由来ピークのうち興味のあるスペクトルのみを MassBank フォーマットに変換できるようにするなど、なるべく分析化学者の視点に立って Utility を構築した。MS-DIAL は質量分析生データを読み込むため、MassBank フォーマットの対応は出力のみであるが、MS-FINDER では入力・出力両方をサポートした。MS-FINDER には、フラグメントイオンに対して組成式や部分構造情報を付与する機能が存在する。そこで、MS-DIAL で出力された MassBank フォーマットを MS-FINDER で読み込み、フラグメント注釈付けを行ったものを再度、MassBank フォーマットとして出力することで、情報量の多い MassBank レコードを作成することができるようになった。

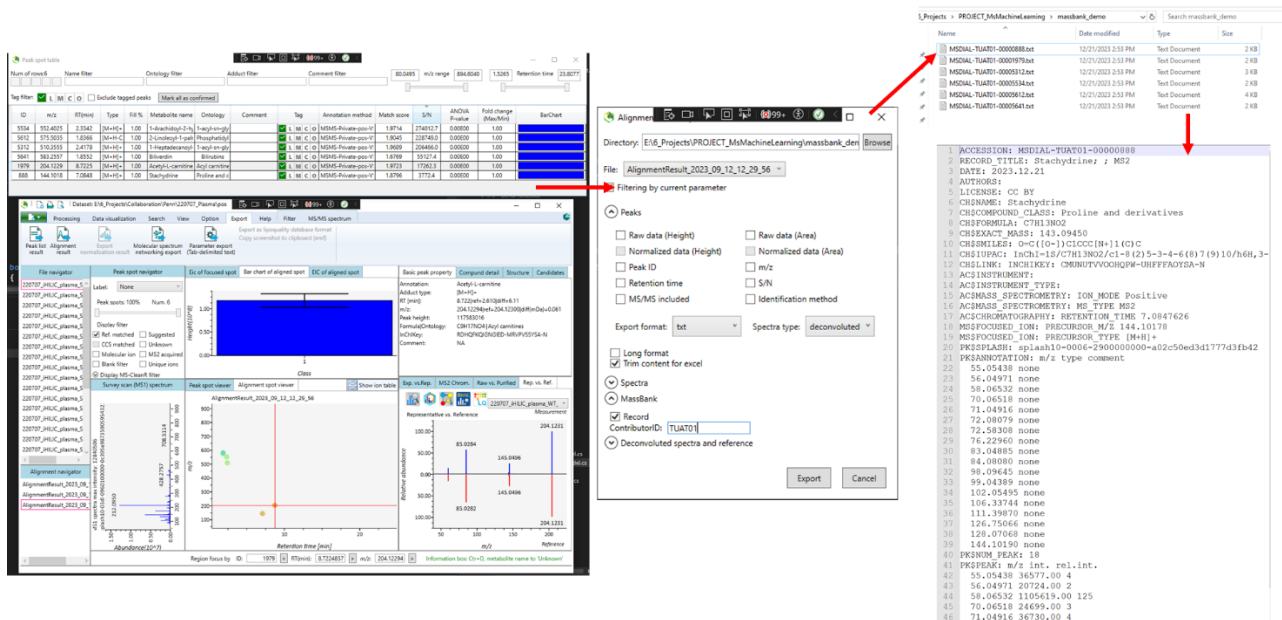


図 8 MS-DIAL から MassBank フォーマットへの出力の流れを示す

開発項目 4 MassBank のデータ拡充・普及・国際連携(松田、奥田、平山、和泉)

4.1 MassBank データ登録(6 月、12 月)

MassBank Human に蓄積した高品質マ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、専門家レビューを行い、10件を承認した。MassBank human の GitHub ページ上に、MassBank の MassBank-data レポジトリのブランチを作成して、10件をアップロードし、本家 MassBank-data レポジトリへの登録を依頼するプルリクエストを行った。これらの作業を通じて、MassBank レコードフォーマットの詳細、作業プロトコルの確認ができた。これらをもとに 2024 年度の MassBank データ登録を実施する。

4.2 ユーザー講習会(年1回)

日本質量分析学会スペクトルデータ部会と連携して下記の講習会を実施した。

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)

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、2023 年度は Metabolomics 2023(カナダ)に平山が参加し、情報収集を行った。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2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(図9、公開版では未掲載)。

日時	内容	場所	参加者数	目的
2023 年 7 月 11 日	MassBank EU メンバーとのミーティング(非公開)	オンライン	12 人	レコードフォーマット改定に向けたミーティング
2023 年 12 月 21 日	MassBank EU メンバーとのミーティング(非公開)	オンライン	12 人	レコードフォーマット改定に向けたミーティング

4.4 展示出展(年1件分生、マス学会、生化学)

シン・マスバンクのユーザーコミュニティである第17回メタボロームシンポジウム 2023 年 10 月 18 日(水)~20 日(金)川崎への出展を行った。メタボロームシンポジウムには 400 名の参加があり、150 名にチラシを配布できた。また、製薬企業研究者などから、マウスで見つかったバイオマーカー候補が人にもあるかすぐ確認できるのはうれしいといった、MassBank Human の機能に関するユーザーニーズを収集できた(図10)。



図 10 第 17 回メタボロームシンポジウム 2023 年 10 月 18 日(水)～20 日(金)出展

4.5 アドバイザリ年 2 回(6 月、12 月)

アドバイザリ委員として、石濱泰(京都大学教授、日本質量分析学会会長、jPOST)、豊田岐聡(大阪大学教授、日本質量分析学会 Mass Spectrometry 編集長)、三枝大輔(帝京大学准教授、スペクトルデータ部会担当、東北メガバンク)、三浦大典(産総研 バイオメディカル研究部門)、岩橋福松(住友化学、創薬メタボロミクス)の 5 名に嘱託した。これまでに小アドバイザリ委員会を 1 回、アドバイザリ委員会を 1 回開催し、成果及び実施計画を説明した。2024 年 1 月 23 日アドバイザリ委員会では 24 件以上にわたるコメントが挙がり、今後の研究開発に反映する計画である(図 11)

日時	内容	場所	参加者数	目的
2023 年 6 月 27 日	小アドバイザリ委員会 (非公開)	オンライン	12 人	実施計画へのコメント収集
2024 年 1 月 23 日	アドバイザリ委員会 (非公開)	大阪大学	16 人	成果、および実施計画への コメント収集

§4. 成果発表等

(1) 原著論文発表

① 論文数概要

種別	国内外	件数
発行済論文	国内(和文)	0 件
	国際(欧文)	6 件
未発行論文 (accepted, in press 等)	国内(和文)	0 件
	国際(欧文)	0 件

② 論文詳細情報

1. Ikeda K, Takahashi M, Bamba T, Izumi Y*. Comparison of amine-modified polymeric stationary phases for polar metabolomic analysis based on unified-hydrophilic interaction/anion exchange liquid chromatography/high-resolution mass spectrometry (unified-HILIC/AEX/HRMS), Mass Spectrometry, vol. 13, No. 1, 2024 (DOI: 10.5702/massspectrometry.A0143).
2. Taihei Torigoe, Masatomo Takahashi*, Omidreza Heravizadeh, Kazuki Ikeda, Kohta Nakatani, Takeshi Bamba, Yoshihiro Izumi*. 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, Analytical Chemistry, vol. 96, No. 3, 1275-1283, 2024 (DOI: 10.1021/acs.analchem.3c04618).
3. Noriyuki Tomiyasu, Masatomo Takahashi, Kenji Toyonaga, Sho Yamasaki, Takeshi Bamba, Yoshihiro Izumi*. Efficient lipidomic approach for the discovery of lipid ligands for immune receptors by combining LC-HRMS/MS analysis with fractionation and reporter cell assay., Analytical and Bioanalytical Chemistry, 2023 (DOI: 10.1007/s00216-023-05111-w).
4. Fumio Matsuda. Data Processing of Product Ion Spectra: Redundancy of Product Ion Spectra of Small Molecules in Data-Dependent Acquisition Dataset, Mass Spectrometry, vol. 12, No. 1, A0138-A0138, 2023 (DOI: 10.5702/massspectrometry.a0138).
5. Kanako Tokiyoshi, Yuki Matsuzawa, Mikiko Takahashi, Hiroaki Takeda, Mayu Hasegawa, Junki Miyamoto, Hiroshi Tsugawa. Using Data-Dependent and -Independent Hybrid Acquisitions for Fast Liquid Chromatography-Based Untargeted Lipidomics, Analytical Chemistry, vol. 96, No. 3, 991-996, 2024 (DOI: 10.1021/acs.analchem.3c04400).
6. Takaki Oka, Yuki Matsuzawa, Momoka Tsuneyoshi, Yoshitaka Nakamura, Ken Aoshima, Hiroshi Tsugawa. Multiomics analysis to explore blood metabolite biomarkers in an Alzheimer's Disease Neuroimaging Initiative cohort, Scientific Reports, vol. 14, 6797, 0 (DOI: <https://doi.org/10.1038/s41598-024-56837-1>).

(2) その他の著作物(総説、書籍など)

特になし

(3) 国際学会および国内学会発表

① 概要

種別	国内外	件数
招待講演	国内	11 件
	国際	5 件
口頭発表	国内	3 件
	国際	1 件
ポスター発表	国内	2 件
	国際	2 件

② 招待講演

〈国内〉

1. 松田史生、ユーザーコミュニティによる MassBank の発展、第 71 回質量分析総合討論会 2023、大阪、2023 年 5 月 15 日
2. 松田史生、使ってみよう MassBank ハンズオンセミナー、第49回 BMS コンファレンス、別府、2023 年 7 月 18 日
3. 津川裕司、質量分析データリポジトリを活用したメタボロミクス研究、第 71 回質量分析総合討論会 2023、大阪、2023 年 5 月 15 日
4. 津川裕司、メタボロミクスの網羅性と深度向上に向けた質量分析情報計測に関する研究、質量分析インフォマティクス研究会・第 8 回公開ワークショップ、横浜、2023 年 5 月
5. 津川裕司、生命の代謝を捉えるメタボロミクスとデータサイエンス、近未来への招待状～ナイスステップな研究者 2022 からのメッセージ～、オンライン、2023 年 7 月
6. 津川裕司、脂質多様性解明に資する質量分析情報計測、第49回 BMS コンファレンス、別府、2023 年 7 月 18 日
7. 津川裕司、マルチモーダル質量分析データの情報解析基盤開発、日本プロテオーム学会 2023 年大会、新潟、2023 年 7 月
8. 奥田修二郎、「質量分析データによるトランスオミクス統合」、第 12 回生命医薬情報学連合大会ワークショップ「データベース統合のフロンティア」、柏の葉、2023 年 9 月 8 日
9. 和泉自泰、ビデオを用いた実験現場紹介 プロテオミクス・メタボロミクス編、BINDS 発現機能解析インシリコ融合ユニット講習会、Web、2023 年 10 月 7 日
10. 和泉自泰、次世代質量分析オミクス計測技術による微小試料からの代謝解析、第 96 回 日本生化学会大会、福岡、2023 年 11 月 1 日
11. 和泉自泰、LC/MS によるメタボロミクス、第 51 回 質量分析講習会、大阪、2023 年 11 月 14 日

〈国際〉

1. Fumio Matsuda、The Shin-MassBank project: A collaboration between proteomics and metabolomics、HUPO-PSI Kyoto Symposium 2024 on Omics Repository, Database and Data Journal、Kyoto、2024 年 3 月 21 日
2. Hiroshi Tsugawa、Computational mass spectrometry to accelerate systems biology、Joint 11th AOHUPO and 7th AOAPC congress、Singapore、2023 年 5 月

3. Hiroshi Tsugawa, Multimodal mass spectrometry-based metabolomics environment in MS-DIAL 5, 3rd International BMS Symposium, Kyoto, 2023 年 10 月
4. Shujiro Okuda, Current status of the jPOST and JPDM projects and future approaches to multi-omics data integration, HUPO-PSI Kyoto Symposium 2024 on Omics Repository, Database and Data Journal, Kyoto, 2024 年 3 月 21 日
5. Yoshihiro Izumi, Development of single-cell multi-omics analytical system using nano-LC/MS, 2023 international joint meeting of the 23rd International Conference on Cytochrome P450 (ICCP450) and the 38th Annual Meeting of the Japanese Society for the Study of Xenobiotics (JSSX), Shizuoka, 2024 年 9 月 28 日

③ 口頭講演

〈国内〉

1. 鳥越大平、高橋政友、Heravizadeh Omidreza、中谷航太、池田和輝、馬場健史、和泉自泰、Unified-HILIC/AEX の保持時間予測と構造アノテーションの評価、第 71 回 質量分析総合討論会、大阪、2023 年 5 月 16 日
2. 中谷航太、池田和輝、高橋政友、馬場健史、和泉自泰、Unified-HILIC/AEX/MS/MS による親水性メタボロームとリピドームの同時分析、第 71 回 質量分析総合討論会、大阪、2023 年 5 月 16 日
3. 鳥越大平、高橋政友、中尾素直、Heravizadeh Omidreza、相馬悠希、池田和輝、中谷航太、馬場健史、和泉自泰、*in silico* エピメタボライトデータベース (IEMDB) を用いた代謝物の 包括的構造推定法の開発、第 49 回 BMS コンファレンス (BMS2023)、大分、2023 年 7 月 19 日

〈国際〉

1. Hiroshi Tsugawa, Reanalyzing mass spectrometry data to develop metabolome cartography database, Metabolomics 2023, Canada, 2023 年 6 月

④ ポスター発表

〈国内〉

1. 今戸優理、高橋政友、相馬悠希、油屋駿介、中谷航太、和泉自泰、馬場健史、血液メタボロームデータ統合に資する定量メタボロミクス解析システムの開発、第 49 回 BMS コンファレンス (BMS2023)、大分、2023 年 7 月 19 日
2. 今戸優理、高橋政友、相馬悠希、中谷航太、和泉自泰、馬場健史、SILIS を基盤とした血漿メタボロームデータ統合システムの開発、学術変革領域 A 自己指向性免疫学 若手ワークショップ 2023、大阪、2023 年 9 月 1 日

〈国際〉

1. Kohta Nakatani, Kazuki Ikeda, Masatomo Takahashi, Takeshi Bamba, Yoshihiro Izumi, Simultaneous analysis of polar metabolome and lipidome by unified-hydrophilic interaction/anion-exchange liquid chromatography tandem mass spectrometry (unified-HILIC/AEX/MS/MS), 71th ASMS Conference on Mass Spectrometry and Allied Topics (ASMS 2023), Houston, Texas, USA, 2023 年 6 月 6 日
2. Noriyuki Tomiyasu, Yoshihiro Izumi, Masatomo Takahashi, Naoya Nishimura, Kenji Toyonaga, Sho Yamasaki, Takeshi Bamba, A method for comprehensive investigation of lipid ligands using LC-FRC/HRMS/MS, 71th ASMS Conference on Mass Spectrometry and Allied Topics (ASMS 2023), Houston, Texas, USA, 2023 年 6 月 7 日

(4) 知的財産権の出願（国内の出願件数のみ公開）

出願件数

種別		件数
特許出願	国内	0 件

(5) 受賞・報道等

① 受賞

該当なし

② メディア報道

該当なし

③ その他の成果発表

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

§5. 主要なデータベースの利活用状況

(1) アクセス数

① 実績

表 1 研究開発対象の主要なデータベースの利用状況(月間平均)

DB 名	種別	2023 年度
Shin-MassBank	訪問者数	公開前
	訪問数	公開前
	閲覧ページ数	公開前
MassBank Human	訪問者数	公開前
	訪問数	公開前
	閲覧ページ数	公開前
MassBank <i>in silico</i>	訪問者数	公開前
	訪問数	公開前
	閲覧ページ数	公開前

注) 2024 年 4 月から AWStats による記録を開始した。

② 分析

現在プロジェクトの立ち上げ中であり、今後の対外宣伝活動を通じてユーザーを獲得していく。

(2) データベースの利用状況を示すアクセス数以外の指標

特になし。

(3) データベースの利活用により得られた研究成果(生命科学研究への波及効果)

特になし。

(4) データベースの利活用によりもたらされた産業への波及効果や科学技術のイノベーション(産業や科学技術への波及効果)

本プロジェクトの成果を NEDO バイオものづくりプロジェクト等に展開する検討を進めている。

§6. 研究開発期間中に主催した活動(ワークショップ等)

(1) 進捗ミーティング

年月日	名称	場所	参加人数	目的・概要
2023年 4月26日	チーム内ミーティング (非公開)	オンライン	8人	研究進捗報告のためのミーティング
2023年 5月19日	チーム内ミーティング (非公開)	オンライン	8人	研究進捗報告のためのミーティング
2023年 6月27日	チーム内ミーティング	大阪大学	8人	研究進捗報告のためのミーティング
2023年 6月27日	小アドバイサリ委員会 (非公開)	オンライン	12人	実施計画へのコメント収集
2023年 7月11日	MassBank EUメンバーと のミーティング(非公開)	オンライン	12人	レコードフォーマット改定に向けたミー ティング
2023年 7月21日	チーム内ミーティング (非公開)	オンライン	8人	研究進捗報告のためのミーティング
2023年 8月18日	チーム内ミーティング (非公開)	オンライン	9人	研究進捗報告のためのミーティング
2023年 9月15日	チーム内ミーティング (非公開)	オンライン	12人	研究進捗報告のためのミーティング
2023年 10月13日	チーム内ミーティング (非公開)	オンライン	12人	研究進捗報告のためのミーティング
2023年 11月22日	チーム内ミーティング (非公開)	オンライン	12人	研究進捗報告のためのミーティング
2023年 12月19日	チーム内ミーティング (非公開)	オンライン	12人	研究進捗報告のためのミーティング
2023年 12月21日	MassBank EUメンバーと のミーティング(非公開)	オンライン	12人	レコードフォーマット改定に向けたミー ティング
2024年 1月23日	チーム内ミーティング (非公開)	オンライン	12人	研究進捗報告のためのミーティング
2024年 1月23日	全体会議(非公開)	大阪大学	12人	次年度実施計画の作成
2024年 1月23日	アドバイサリ委員会 (非公開)	大阪大学	16人	成果、および実施計画へのコメント収 集
2024年 2月29日	チーム内ミーティング (非公開)	オンライン	11人	研究進捗報告のためのミーティング
2024年 3月26日	チーム内ミーティング (非公開)	オンライン	12人	研究進捗報告のためのミーティング

(2) 主催したワークショップ、シンポジウム、アウトリーチ活動等

年月日	名称	場所	参加人数	目的・概要
特になし				

以上

別紙1 既公開のデータベース・ウェブツール等

No.	正式名称	別称・略称	概要	URL	公開日	状態	分類	関連論文
1	Shin-massbank	シンマスバンク	生物サンプルから得たMS2スペクトルを蓄積するデータ処理パイプラインを開発する。MB-POST（ユーザーフレンドリーな生データリポジトリ）MassBank Human（）、MassBank in silico（in silicoスペクトルデータベース）、MassBank Links（RDF機能）の4つのデータベースから構成される。	https://shin.massbank.jp/	2023.9.12	新規	データベース等	準備中
2	MassBank Human		ヒトDDAデータセット由来のMS2スペクトルライブラリ	http://human.massbank.jp/	2024.3.31	新規	データベース等	準備中
3	MassBank in silico		in silicoスペクトルデータベース	http://insilico.massbank.jp/	2024.3.31	新規	ツール等	準備中
4	MassBank Links		マスバンクレコードと外部データベース・ツールとの連携	https://github.com/systemsomicslab	2023.10.01	維持・発展	ツール等	準備中