

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】「プロテオームデータベースの機能深化と連携基盤強化」

研究代表者：

石濱 泰（京都大学 大学院薬学研究科 教授）

研究分担者：

奥田 修二郎（新潟大学 医学部 教授）

河野 信（北里大学 未来工学部 教授）

1. 概要

(1) 研究および計画の概要

日本内外に散在している種々のプロテオーム情報を標準化・統合・一元管理し、多彩な生物種・翻訳後修飾・発現量も含めた横断的統合プロテオームデータベース jPOST (Japan ProteOme SStandard Repository/Database) を開発する。jPOST は、国際標準のデータリポジトリ部、高精度データ標準化機能を有する再解析部、および標準化データを蓄積・表示するデータベース部からなる。リポジトリ部では、プロテオミクス以外のオミクス MS データリポジトリとの連携を視野にいたした UniPOST 構想の基盤整備を行う。再解析部では、翻訳後修飾に関する国際プロジェクト PTMeXchange に参画し、独自の信頼性スコア UniScore の展開をはかるとともにデータジャーナル Journal of Proteome Data and Methods との連携を更に推進する。また、データ非依存性取得法やイオンモビリティ分離法とのハイブリッドなど、進化する計測系に対応したシステム整備を行う。データベース部では、様々なユーザーコミュニティとの連携を強化し、開発チームと連携ユーザーが一体となった双方向利活用環境 ” jPOST prime” を実現させ、プロテオームデータサイエンスリソースとしての有用性を向上させる。

(2) 成果の概要

リポジトリ部においては、10 月末日現在、登録プロジェクト数 3653 で、目標としていた 2025 年度登録数 3000 を達成。3 拠点でのサーバー体制の確立については、1 年目に新潟サーバーを設置し、現在は、年度末の遺伝研の旧サーバ (jq01) から新サーバ (jp03) への切り替えに伴い、リポジトリシステム全体の移植作業を行っている。2026 年の新サーバーでの運用開始に向けて順調に進行中。

再解析部においては、コミュニティ連携メタデータ収集システムが本格稼働し、Journal of Proteome Data and Methods(JPDM)との連携で、JPDM-empowered Sample Data Relationship Format (jeSDRF)を定義し、それに基づく再解析とメタデータ論文の投稿が開始された。さらに、jPOST 以外の ProteomeXchange リポジトリとも協力し、他リポジトリに登録されたメタデータの jeSDRF 化および JPDM 論文を開始した。また、Universal Spectrum Identifier ツールを利用した標準スペクトルライブラリや PTMeXchange は HUP0 Proteomics Standardization Initiative (PSI) との連携で順調に進行中である。さらに再解析の中心となる UniScore の論文化が終了し、本格的に再解析を開始するとともに、他のリポジトリ (KPOP) への展開について検討を開始した。

データベース部については、メタプロテオームデータの可視化ツールのプロトタイプを今年度中に公開するところまできた。また、非典型翻訳対応を含むタンパク質末端データベース EdgeProt の開発を進めており、再解析部と連動しながら大型国プロおよび HUP0—Human Proteome Project の国際協力のなかでデータを蓄積するシステムを確立した。

これらの成果により、jPOST ウェブ (リポジトリ、データベース) へのアクセス数は年々増加しており、2025 年度に入り、訪問者数は前期だけですでに 2024 年度実績の 1.5 倍(repository)、2.5 倍(database) となっている。

名称	概要
jPOST	国内外に散在する多様なプロテオームデータを標準化・一元管理する統合データベース。データリポジトリ、再解析によるデータ標準化、および標準化データを蓄積・表示するデータベース部からなる。国際連携・他データベースとの連携によって、生命科学データ統合化の中核となることを目指す。開発チームと連携ユーザーが一体となってプロテオームデータサイエンスリソースとしての双方向利活用環境“jPOST prime”を実現する。

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】 リポジトリ部の開発	
UniPOST 基盤の開発	他の DB とのコンセプト共有化（終了） 共通メタデータ入力システムの開発（進行中） 共通ピークピッキングツールの開発（進行中）
DIA データ対応	計画通り進行中
投稿データの品質保証	計画通り進行中
リポジトリ登録数の確保	3000project 登録達成
サーバー体制の確立	新潟サーバーの設置（終了） 3 拠点体制の確立（計画通り進行中）
再解析フローの半自動化	計画通り進行中
【項目 2】 再解析部の開発	
PTMeXchange 対応	一旦開発を中断後、Site localization に対する FDR 設定が進行中
Peptide-centric DIA 対応	既知法の評価（終了）
Isobaric tag 定量対応	バッチ間標準化対応（終了） プロジェクト間対応（進行中）
検索用リファレンスデータベース	単一生物種対応（進行中）
コミュニティ連携メタデータ収集	応募システムの確立（進行中）
再解析フローの半自動化	計画通り進行中
【項目 3】 データベース部	
複数生物種（メタプロテオームデータ）対応	複数生物種対応（進行中）
非典型翻訳対応データベースの開発	計画通り進行中
連携 prime ユーザー対応	大型国家プロジェクト対応（進行中）
再解析フローの半自動化	計画通り進行中

(2) 実施状況の詳細

【項目 1】リポジトリ部の開発

(項目 1-①) UniPOST 基盤の開発

他の DB とのコンセプト共有化を行うとともに理研のオミクスデータ受入用システムを開発中。共通メタデータ入力システムの開発は、今年度末のプロトタイプ完成に向けて進行中。さらに共通ピークピッキングツールの開発も計画通り進行中である。

(項目 1-②) DIA データ対応

MS 各社の DIA データに対応するべく、本年度より再解析フロー開発に取り組んでいる。装置開発スピードが目覚ましく、イオンモビリティ等を取り入れた新しい次世代 DIA が開発されつつあり、またそれに伴うスキャン手法の開発も続いているので、様々なデータ取得法に対応したメタデータ情報取得システムを開発中である。

(項目 1-③) 投稿データの品質保証

再解析クライテリアとしての生データやメタデータに対する各種パラメータの閾値の設定を目指している。本年度は JPDM と連動させた jeSDRF 収集を開始した。

(項目 1-④) リポジトリ登録数の確保

国内外の各種学会や研究者コミュニティを通じて jPOST の宣伝活動を継続するとともに開発内容の論文化を進め、高品質な再解析データをタイムリーに公開することにより、データ投稿数をコンスタントに増加させる試みを継続してきた。2025 年度の数値目標は累計 3000 projects を達成した。

(項目 1-⑤) 3 拠点でのサーバー体制の確立

初年度の新潟サーバーの設置は終了し、本年度末の遺伝研の旧サーバ (jq01) から新サーバ (jp03) への切り替えに伴い、リポジトリシステム全体の移植作業が進行中である。

(項目 1-⑥) 再解析フローの半自動化

再解析に必要なメタデータをオリジナル登録データ、JPDM 登録データおよびコミュニティ連携で取得する論文抽出データの統一フォーマット jeSDRF を確立した。

【項目 2】再解析部の開発概要

(項目 2-①) PTMeXchange 対応

HUPO-PSI における動向から、一旦開発を中断したが、再始動した。UniScore による信頼性評価法の開発が進行中である。

(項目 2-②) Peptide-centric DIA 対応

前年度に引き続き、DIA-NN 等の既知法の情報収集を継続するとともに、MS 装置の新モデル対応を行い

つつ、再解析フローのツール開発を進めている。

（項目 2-③）Isobaric tag 定量対応

定量法として今後も汎用的に使われるのは、DIA 法および isobaric tag 法であると考えられる。そこで Isobaric tag 定量において、大規模データ、多試料データを取得した際に問題となるバッチ間効果に対して行う補正法に対する標準化を行った。

（項目 2-④）検索用リファレンスデータベース

様々な大規模データが登録されているリポジトリからは、ペプチドの同定・定量のためのデータベース検索で用いることができるリファレンスデータベースを生成することが可能である。さらにこのリファレンスデータは DIA においてもライブラリとして使用が可能である。とくに複数生物種が混在するメタプロテオーム解析ではリファレンスデータライブラリが威力を発揮すると考えられる。現在は主たる単一生物種（ヒト、マウスなど）に対応したリファレンスデータベースの生成が進行中である。さらに、Microbiome Datahub の森チームと連携し、MAGs をメタプロテオミクス再解析用データベースとして用いるための基盤技術の開発を行っている。

（項目 2-⑤）コミュニティ連携メタデータ収集

jeSDRF をコアにしたコミュニティ連携メタデータ収集応募システムの確立に成功。本システムを使って、年間 100 報からのデータ抽出・収集ができるシステムの確立を 5 年目までに行う。

（項目 2-⑥）再解析フローの半自動化

再解析に必要なメタデータをオリジナル登録データ、JPDM 登録データおよびコミュニティ連携で取得する論文抽出データの統一フォーマット jeSDRF を確立した。

【項目 3】データベース部の開発概要

（項目 3-①）複数生物種（メタプロテオームデータ）対応

第 2 期からの積み残し課題のひとつとして、複数生物種データをどのようにデータベースに中に入れて可視化するか、がある。どのような機能が必要とされるか、情報収取から始め、DBCLS と協力し、技術的にはかなり高度になると思われるシステムの整備を行う。

（項目 3-②）非典型翻訳対応データベースの開発

大型国プロ（学術変革領域マルチファセットプロテインズ、AMED-CREST プロテオスタシス、学術変革領域タンパク質寿命学など）を中心にして、世界規模で様々な非典型翻訳情報が集積しつつある。これらのデータセットに対し、高品質再解析データを生成し、データサイエンスに資する専門データベース、すなわち、様々な翻訳開始点、終始点をトランスクリプトーム、トランスレイトーム、プロテオームのレベルで同時に可視化できるツールの開発を検討する。

（項目 3-③）連携 prime ユーザー対応

大型国家プロジェクトや HUP0-cHPP イニシアティブと連携し、PXC 全体の関連データを再解析する形

でのターミノームデータベース EdgeProt の作成を開始した。

(項目 3-④) 再解析フローの半自動化

再解析に必要なメタデータをオリジナル登録データ、JPDM 登録データおよびコミュニティ連携で取得する論文抽出データの統一フォーマット jeSDRF を確立した。

(3) 主な成果論文等

1. Okuda S, et al. “jPOST environment accelerates the reuse and reanalysis of public proteome mass spectrometry data”, *Nucleic Acids Res.*, 2025,53(D1), D462-D467 (DOI: 10.1093/nar/gkae1032).

[概要] jPOST リポジトリに登録されているメタデータは、再分析を行うには不十分で、この問題に対処するため、日本プロテオミクス学会は、再分析が可能なメタデータを詳細に記述したデータ記述子論文を受け入れるデータジャーナル「Journal of Proteome Data and Methods (JPDM)」を創刊した。jPOST 内では、JPDM データ記述記事に記載されたメタデータに基づき、UniScore を活用して生プロテオームデータを再解析する。再解析データは jPOSTdb に登録され、jPOST リポジトリには JPDM 記事へのリンクが追加される。jPOST 環境内でのこうした再解析の促進は、FAIR データ原則とオープンサイエンスを推進する。

2. Ishihama Y, et al. “UniScore, a unified and universal measure for peptide identification by multiple search engines”, *Mol Cell Proteomics*, 2025, 24(7), 101010 (DOI: 10.1016/j.mcpro.2025.101010).

[概要] UniScore を、LC/MS/MS ベースのボトムアッププロテオミクスにおけるデータ依存型取得データの解析において、複数の検索エンジンの出力を統合・標準化する指標として提案する。UniScore は、検索エンジンが提案した候補ペプチドのアミノ酸配列と生成イオンスペクトルを照合することにより、生成イオンのみに付随するアノテーション情報から算出される。UniScore アプローチは、大規模な全プロテオームおよびリン酸化プロテオームデータセットにおいて、従来の検索エンジンを上回る性能を示した。

(4) 主要なデータベースの利活用状況

年々訪問者数、ページ数ともに増加しているが、特に 2025 年度に入り、訪問者数は前期だけですでに 2024 年度実績の 1.5 倍(repository)、2.5 倍 (database) となっている。これは 2025 年 1 月および 6 月にそれぞれ *Nucleic Acids Res.* 誌および *Mol. Cell. Proteomics* 誌に関連論文が掲載されたことが影響したと思われる。

(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

基本的に、リポジトリに登録されたデータはその生物学的な成果を論文として発表する際のサポート

データとして用いられている。そういう意味では、2025/10/31 現在登録されている 3654 プロジェクトのうち、2734 プロジェクトが公開データとなっており、それに相当する数の論文が公開されていることになる。jPOST ではその論文情報を、それぞれのプロジェクトのフロントページに PubMed ID としてリンクできるようにしているが、プロジェクト数が必ずしも論文数と 1 : 1 対応しているわけではないので、その正確な数は把握していない。

一方、再解析後のデータベースに登録されたデータがどのように使われているかについては、我々は把握していない。

HUPO-PSI の活動の中で、jPOST repository に登録されている全マススペクトルは、Universal Spectrum Identifierによって自由に検索できるシステムが開発されている。それを利用して、mzSpecLib および mzPAF といった、スペクトルライブラリとフラグメントイオンピークのアノテーションのための標準フォーマットが開発された。

Klein J., Lam H., Mak TD., Bittremieux W., Perez-Riverol Y., Gabriels R., Shofstahl J., Hecht H., Binz PA., Kawano S, Van Den Bossche T., Carver J., Neely BA., Mendoza L., Suomi T., Claeys T., Payne T., Schulte D., Sun Z., Hoffmann N., Zhu Y., Neumann S., Jones AR., Bandeira N., Vizcaíno JA., Deutsch EW. The Proteomics Standards Initiative standardized formats for spectral libraries and fragment ion peak annotations: mzSpecLib and mzPAF. *Analytical Chemistry* 96, 18491-18501 (2024). doi: 10.1021/acs.analchem.4c04091

3. 今後の計画および展望

リポジトリ部は奥田グループが引き続き主担当となって開発を進める。UniPOST 基盤の開発についてはオミクス MS ストレージシステムの確立に向け、関係各所とのコンセプト共有を図るとともにその実現に向けて jPOST データストレージ体制の展開を図る。DIA データ対応として再解析フロー開発に取り組む。投稿データの品質保証としては、jeSDRF を積極的に利用しながら進める。リポジトリ登録数としては 5000 projects (5 年目) を目指す。

再解析部は石濱グループが引き続き主担当となって開発を進める。PTMeXchange 対応として、HUPO-PSI で動き始めたので、UniScore を組み合わせた信頼性評価を見据えて最終年度までに本格的な導入を開始する。DIA 対応としては UniScore の評価対象となっているプロダクトイオン情報がスコア化可能なはずなので、信頼性評価への適用を検討する。検索用リファレンスデータベースは HUPO-PSI の動向を見ながらメタプロテオーム解析に対応させる。コミュニティ連携メタデータ収集として JPDM と連携した jeSDRF を PXC レベルまで拡大し、年間 100 報からのデータ抽出・収集ができるシステムの確立を 5 年目までに行う。再解析フローの半自動化についてはリポジトリ部を担当する奥田グループと協力して、再解析に必要なメタデータをオリジナル登録データ、JPDM 登録データおよびコミュニティ連携で取得する論文抽出データの統一フォーマットによる半自動化エクスポートツールを開発するとともに、再解析データをデータベース部にエクスポート登録するにあたって河野グループと協力して半自動化ツールの開発を行う。

データベース部は河野グループが引き続き主担当となって開発を進める。メタプロテオームデータの可視化ツールの開発に注力する。非典型翻訳対応データベースとしてターミノーム DB の開発を国内国プロ、国際 HUP0-HPP と連携しながら引き続き進める。連携 prime ユーザー対応として広く研究者コミュニティ全体に対し、公募型の連携研究を募集し、第3期 jPOST プロジェクト終了後も継続が可能なシステムの構築も模索する。

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. 石濱グループ（京都大学）

PTMeXchange に準拠した修飾サイトの site localization に対する FDR の設定および UniScore を組み合わせた信頼性評価を行った。Peptide-centric DIA 法とはじめとして、現在の jPOST/PXC では対応できていない解析に対応した再解析法を確立しつつある。Isobaric tag 定量データに対し、バッチ間の標準化対応を行った。さらに検索用リファレンスデータベースの設定として、単一生物種対応を行った。メタデータ収集、それに基づく再解析および再解析データの Globe データベースへのエクスポートについて、奥田グループ、河野グループと連携して半自動化システムの検討を続けている。

(1)-2. 奥田グループ（新潟大学）

UniPOST 基盤の開発について、共通のメタデータ入力システム、ピークピッキングツール、およびオミクス MS リポジトリの開発を行ってきた。DIA データ対応として、MS 各社の DIA データに対応した。投稿データの品質保証として、再解析クライテリアとしての生データやメタデータに対する各種パラメータの閾値設定を行った。リポジトリ登録数として、3年目終了時に 3000 projects を達成した。3拠点でのサーバー体制の確立に向けて、新潟サーバーの整備を行った。メタデータ収集、それに基づく再解析および再解析データの Globe データベースへのエクスポートについて、石濱グループ、河野グループと連携して半自動化システムの検討を続けている。

(1)-3. 河野グループ（北里大学）

複数生物種（メタプロテオームデータ）対応として、システムの整備、および可視化ツール開発を行っている。また、非典型翻訳データベース対応として、可視化ツールの開発を行っている。さらに連携 prime ユーザー対応として、大型国家プロジェクトデータに対するデータサイエンス対応および公募型連携 prime ユーザーに対するデータサイエンスを遂行する。メタデータ収集、それに基づく再解析および再解析データの Globe データベースへのエクスポートについて、石濱グループ、奥田グループと連携して半自動化システムの確立を行う。