

## 研究開発課題別事後評価結果

### ▶ 課題情報

研究開発課題名	プロテオーム統合データベースの構築
研究代表名	石濱 泰

### ▶ 事後評価結果

#### 1.実施計画の達成度

jPOST (Japan ProteOme STandard Repository/Database) を主な研究開発対象とし、プロテオーム測定データのリポジトリを構築するとともに、プロテオーム測定データを再解析し、統合的な閲覧・比較なデータベースを構築するため、1)サーバー管理・運用、2)リポジトリ、3)キュレーション(再解析)、4)プロジェクト別 DB(Cube)構築、5)集積キューブ型 DB(Globe)構築、6)フォーカスト DB(Slice)構築を研究開発項目に掲げて実施された。

1)について、開発サーバーを京都大学に、運用サーバーをDBCLSに設置し、参加者の期中移籍があったものの安定的に運用された。

2)について、2016年5月にjPOSTrepoを公開した。同年8月にProteomeXchangeコンソーシアムとの連携を開始し、当該コンソーシアム加盟データベース間でデータセットごとに一意のIDを付与し、各メタデータを国際的に共有する体制に貢献することとなった。jPOSTrepoへは2018年3月末現在で298件が登録され、うち155件が公開された。公開データのなかには当初目標に掲げた20件のコアデータが含まれる。

3)について、リポジトリに登録されたデータに対してマルチピークピッキングツール、マルチ検索エンジンを使用して、生データからの情報抽出を最大化、不偏化し、またスペクトル品質を判断し偽陽性を低減させるためのjPOSTスコアを開発・導入することで検索エンジンやMS装置に依存しないユニバーサルなスペクトル評価スコアシステムを確立した。さらに偽陽性を除去するための新規クライテリア設定ワークフローも確立した。これにより、データの標準化、プロジェクト間の串刺し検索、横断的データベースの作製を可能にした。

4)について、様々な異なる測定プラットフォームデータ(具体的には、グローバル定量プロテオミクス、翻訳後修飾定量プロテオミクスおよびターゲット定量プロテオミクス)に適した情報の抽出・格納法を開発した。

5)および6)について、RDFスキーマ、オントロジー設計を行い、データ統合に向けた基本設計を行い、2018年3月にjPOSTDB公開を開始した。

上記から、実施計画は全て達成した。

#### 2.生命科学研究への波及効果

月間訪問ユニークIPアドレスは、公開初年度に平均195件、第二年次に平均320件であった。公開から2年弱であって利用が多いとはいいがたいが、リポジトリの開設および再解析データベースの構築を通じて、今後イノベーションに貢献していただくの素地が整えられたと考えられる。

#### 3.研究開発課題の

6つの研究グループで構成された実施体制について、頻繁なミーティング等を通じて適

運営	切に運営され、本研究開発課題の遂行に貢献した。
4.その他特記すべき事項	<p>プロテオームとグライコームとの連携に向けて糖鎖構造の質量分析データリポジトリを構築するために統合化推進プログラムの成松チーム、木下チームと、プロテオームとゲノムとをつなげるための検討を進めるため菅野チームと協議を開始した。DBCLS とは、SPARQLthon や BioHackathon に参加する他、合同で質量分析ハッカソンを開催するなどを通じて研究交流を深めた。また、有識者会議を設置し、研究開発に係る助言を受けるほか、研究者コミュニティでの認知度向上に向けて学会出展、ワークショップ開催などの取り組みを積極的に行った。</p> <p>人材育成の観点でも一定の貢献を果たした。研究開発に参加した多くの若手研究者がチーム内のミーティングに積極的に参加して研究開発に貢献し、またトレーニングコース、チュートリアルにおける講師や責任者を務める等の経験を積むこと等を通じて多くが本課題の期間中に昇任した。</p>
5.総合評価	<p>プロテオーム情報は基礎生物学の根幹を成すのみならず、医療や創薬に直接的に関わる最も重要な情報である。近年の急速な技術革新により、膨大な量のプロテオームデータが産生されているが、これらのデータの共有・維持・管理を行うためのインフラ整備は充分に進んでいなかった。本研究開発課題は、国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・絶対発現量も含めた世界初の横断的統合プロテオームデータベースを開発し、広く国内外に公開することを目的として実施された。研究開発に当たっては、国際標準になりつつあった ProteomeXchange と連携することでデータベース設計などの面で効率化を図った。当初計画より公開が遅れたものの企図した通りにデータベースの開発がなされた点で評価できる。総合的に再解析する人員が育成された点についても意義は高い。</p> <p>ただし、公開から日が浅いことから、現時点ではデータベースの利用者はプロテオーム研究者に限られているように思われる。具体的なユースケースを見据え、今後とも各研究機関、研究室と連携しながら、多様な研究分野から利用されるようなデータベースへと発展することを期待する。</p>