

プロテオーム・データベースの構築
Construction of proteome database

京都大学化学研究所 バイオインフォマティクスセンター 化学生命科学研究領域
吉沢 明康

研究成果概要

本研究は、JST NBDC(科学技術振興機構・バイオサイエンスデータベースセンター)統合化推進プログラム「プロテオーム統合データベースの構築」jPOST 計画のための研究及び開発である。jPOST 計画は以下の 3 部分から構成される:

- 1) 論文発表と同時に公表が義務づけられた、(質量分析法による)プロテオーム測定データ(いわゆる“生データ”)の deposit site である「jPOST Repository」
- 2) Repository で公開されたデータを、独自に解析するための「再解析プロトコル」
- 3) 再解析プロトコルによって解析した結果を格納する「データベース」

サーバはライフサイエンス統合データベースセンター(DBCLS)の三島ブランチ(遺伝学研究所 DDBJ センター内)及び京都大学化学研究所に設置され、データ処理に化学研究所スーパーコンピュータシステムを利用している。

本年度は上記 1) jPOST Repository (jPOSTrepo) の開発と改良、及び 2) 再解析プロトコルの開発を主に進めた。jPOST Repository は新潟大学及び DBCLS との協力により、2016 年 5 月 2 日に一般公開され、更にプロテオーム生データ・レポジトリの国際標準である ProteomeXchange consortium にも同 7 月 6 日から加盟した(アジア・オセアニア地区唯一)。

また再解析プロトコルの開発ためには、大別して

- (a) 偽陰性(False Negative)の減少
- (b) 質量分析データの特徴を利用して、信頼性の向上(偽陽性(False Positive)の減少)
- (c) データベースに新規データを順次追加することに対応できる統計的手法の開発

の 3 つの課題・目標が存在したが、主に本学薬学研究科との協力により、これら全てへの対応策と、更にそれを矛盾なく組み合わせたプロトコルの構築が完了した。現在、主に薬学研究科の共同研究者がこのプロトコルを実装中であり、今後來年度前半までの間にプロトコルの有効性を実証し、プロトコルの論文発表とデータベースの構築に移行する予定である。

発表論文(謝辞あり)

Okuda, S., *et al.* jPOSTrepo: an international standard data repository for proteomes. *Nucleic Acids Research*, **45** (D1, Database issue): D1107-D1111 (2017).

吉沢 明康. どのデータベースを使うか ～データベース検索と配列解析・誤解と難題～. *Proteome Letters*, **1** (2):63-80 (2016).