

## オープンソースの質量分析データ用ツールの開発

森澤拓<sup>†</sup>、岩本真知子<sup>†</sup>、津元裕樹<sup>†</sup>、廣田三佳子<sup>††</sup>、三浦ゆり<sup>†</sup>、戸田年総<sup>†††</sup>東京都健康長寿医療センター研究所 老化機構研究チーム<sup>†</sup>、帝京平成大学 薬学部<sup>††</sup>、横浜市立大学 先端医科学研究センター<sup>†††</sup>

生命科学研究において、質量分析装置の開発が進展し、検出精度とスループットの向上に従って、詳細かつ網羅的な解析が実現しつつある。そのため、膨大な生体サンプルに質量分析を行うことで発生する大量のデータを、効率よく解析するためのソフトウェアツールの開発が求められている。ヒトプロテオーム機構 (HUPO) は 2009 年に生体サンプルの質量分析データの交換と蓄積のために標準データ形式 mzML バージョン 1.10 を発表した。そこで、多くの質量分析機器メーカーが出力形式として mzML 形式をサポートすることになり、複数の装置メーカーや機種別の質量分析データの閲覧、解析、蓄積などのためのソフトウェアツールを開発することが現実的になった。筆者らは、プロテオーム用の研究情報管理システム Laboratory Information Management System (LIMS) の開発・改良を進める中で、質量分析データの標準形式である mzXML そして mzML への対応を進める必要性が生じた。そこで標準形式データの活用ライブラリを開発するためにオープンソースの質量分析データビューワの開発にとりかかった。質量分析装置から出力されるデータの主要部分は、縦軸が信号強度、横軸が  $m/z$  値 (質量電荷比) の質量スペクトルである。そしてその波形データから蛋白質の同定、さらには翻訳後修飾等の確認を行うためには各質量分析装置の手法に適したマススペクトルのデータベースと比較等の作業が必要となっている。筆者らはデータビューワの開発をとおしてスペクトルデータの表示機能、基礎的な解析機能を開発するとともに、研究連携を促進するためにオープンソースの質量分析データリポジトリの開発を行い、総合的に質量分析データ用ツールの開発を行ったので報告する。

## 材料と方法

質量分析データビューワ Basic MS Data Viewer は、Delphi (R) 2007 により開発し、Windows 7 Professional で動作確認を行った。動作検証には、研究所で利用していた島津製作所製質量分析装置 Shimadzu AXIMA TOF2 からエクスポートした mzXML 形式のデータと HUPO のホームページ上で配布されていた mzML 形式のサンプルデータを用いた。AXIMA TOF2 のデータは、装置付属アプリケーションと波形データの比較等を行った。質量分析データリポジトリの開発は、Fedora version 19 上で PostgreSQL-PHP-Apache システムとして開発を行った。Postgresql 9.2.4、PHP 5.5.3、Apache 2.4.6、JavaScript を用いて開発した。Web 形式のユーザビリティを実現し、ブラウザでのアップロード、データ閲覧をサポートした。動作確認は、Windows 7 Professional の Internet explorer で行った。動作検証には、質量分析装置 AXIMA-TOF2 から出力された mzXML 形式のデータ以外に研究所で利用している ABSCIEX 社の ABSCIEX TOF/TOF5800 からエクスポートした mzML 形式のデータを用いた。またリポジトリのアノテーション機能の開発として、HUPO が推し進めている Chromosome-Centric Human Proteome Project (C-HPP) の成果である Swiss Institute of Bioinformatics の nextProt における xml データ (2013 年 1 月) をダウンロードし、システムに取り込んで活用できるように開発した。

## 結果

質量分析データビューワ、質量分析データリポジトリ共に質量スペクトルデータの任意の  $m/z$  領域の表示 (拡大表示) をマウスドラッグで行えるよう開発した。2 つの mzML 形式ファイルを同時に取り込み、質量スペクトルデータを並べて表示して比較できるよう開発した。この機能により、同じ機種だけでなく、異なる機種の質量分析計のデータを比較できるようになった。基礎的な解析アプリケーションとして質量スペクトル上の任意の位置の  $m/z$  値、 $m/z$  値の差分、信号強度のピークエリアの計算が行えるよう開発した。(図 1)。ピークエリアはバックグラウ

Development of the open source repository and data viewer for mass spectrometry data

<sup>†</sup>Hiraku Morisawa, Iwamoto Machiko, Tsumoto Hiroki, Miura Yuri. Tokyo Metropolitan Institute of Gerontology

<sup>††</sup>Hirota Mikako. Teikyo Heisei University,

<sup>†††</sup>Toda Tosifusa. Yokohama City University

ンドを設定して計算が行えるよう開発した。質量分析データ用リポジトリについては、さらに、自動的に質量スペクトルのピーク抽出を行う機能を開発した。ピーク抽出する方法は、レベル補正を行うとともに、任意の  $m/z$  値領域の最大のシグナル強度の何分の1以上をピークとするかパラメーターを用意した(図2)。また、質量分析データと nextProt の蛋白質の遺伝子位置情報、参照文献情報、参照 URL 情報等をアノテーション情報としてキーワード検索でリンクできるよう開発を行った。(図3)

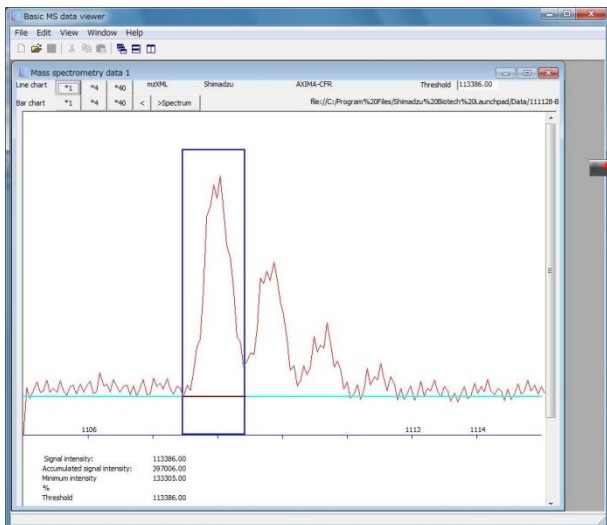


図1 質量分析データビューアにおいて、質量スペクトルの表示を行い、拡大表示を行ってマウスで領域を指定し、ピークエリアを計算表示した画面。

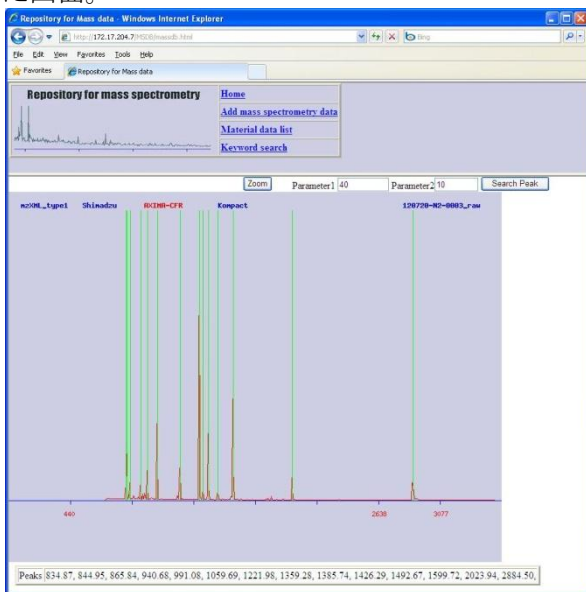


図2 質量分析データリポジトリにおいて、質量スペクトルの表示を行い、自動的にピーク抽出を行った画面

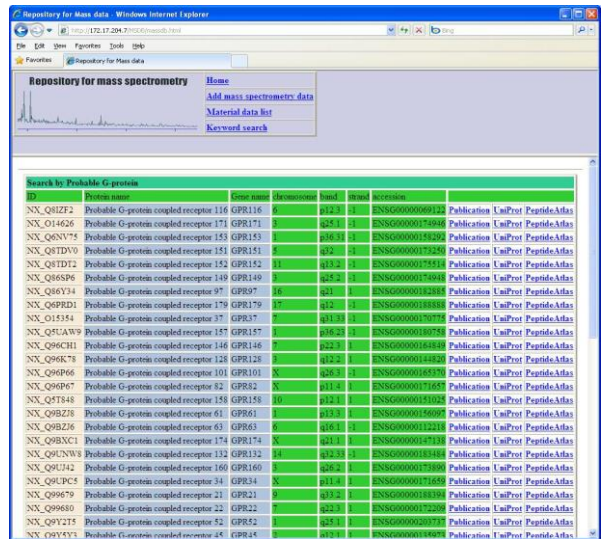


図3 質量分析データリポジトリにおいて、アノテーション情報を表示した画面

考察

本開発で用いた質量分析データの標準形式は、Seattle Proteome Center が作成した mzXML 形式はもちろん、mzML 形式においても装置メーカーや機種によりデータ形式にばらつきが許容されている。そのため、機種を判断しながらソフトウェアが動作するライブラリになってしまい、ソフトウェアが冗長化してしまった。今後については、開発したライブラリは、開発を行っているオープンソースの研究情報管理システムにも追加し、利便性を高めていきたい。オープンソースで質量分析のためのソフトウェアを開発することは、機種やメーカーに制限されずにデータ利用ができ、大きな役割を果たせると考えている。尚、本開発で作成したソフトウェアは、sourceforge.net でソースファイルの公開を予定しており、幅広い利用を期待している。

参考文献

[1] Sturm M, Bertsch A, Gröpl C, Hildebrandt A, Hussong R, Lange E, Pfeifer N, Schulz-Trieglaff O, Zerck A, Reinert K, Kohlbacher O. : OpenMS - an open-source software framework for mass spectrometry. BMC Bioinformatics. 2008 Mar 26;9:163.  
 [2] Gaudet P, Argoud-Puy G, Cusin I, Duek P, Evalet O, Gateau A, Gleizes A, Pereira M, Zahn-Zabal M, Zwahlen C, Bairoch A, Lane L. : neXtProt: organizing protein knowledge in the context of human proteome projects. J Proteome Res. 2013 Jan 4;12(1):293-8.