

# ProViewer: Rapid interactive visualisation of mzIdentML file content via the web

Suresh Hewapathirana, Ryan Smith, Jun Fan, Conrad Bessant\*

School of Biological and Chemical Sciences, Queen Mary University of London, E1 4NS. UK.

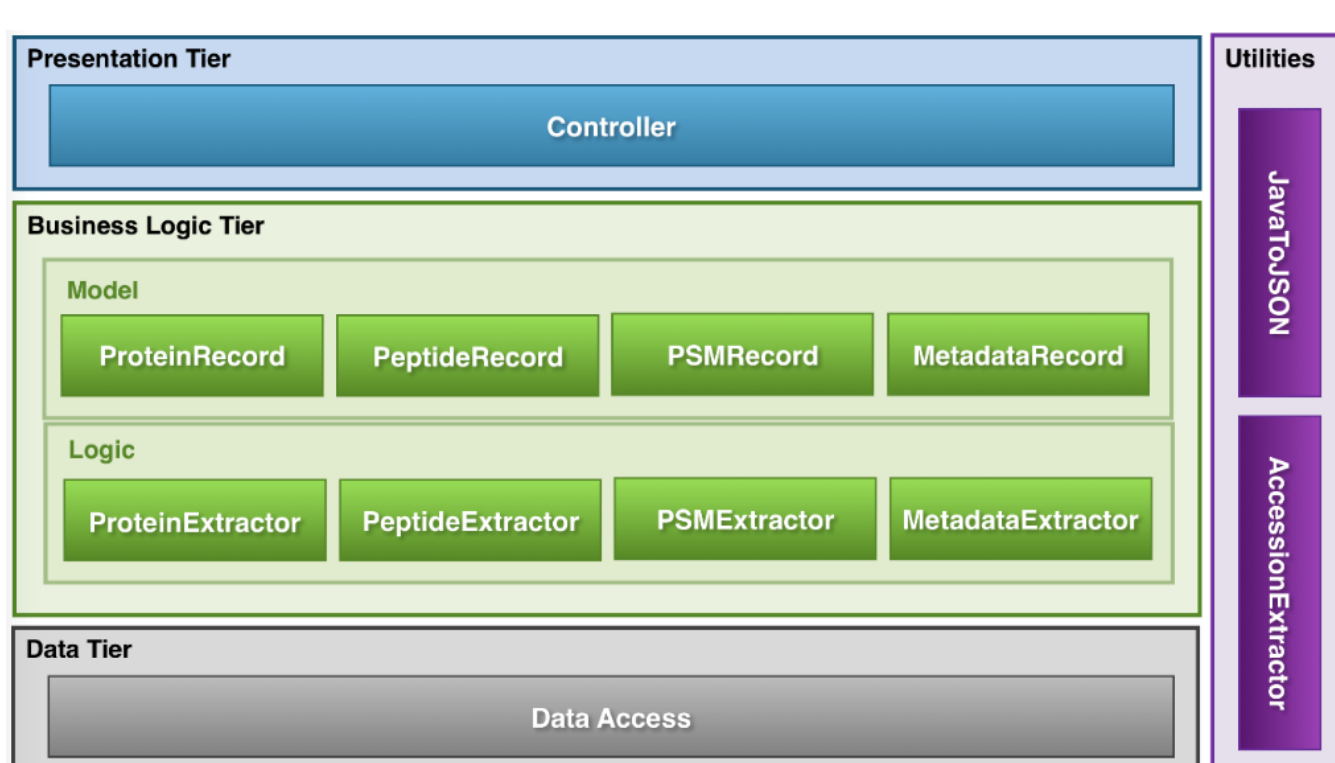
\*c.bessant@qmul.ac.uk

## Introduction

Uptake of the mzIdentML [1] format for peptide and protein identifications has increased substantially in recent years. Most peptide and protein identification software now produces results in this format, either natively or via converters, and mzIdentML is increasingly used for sharing results via databases and as supplementary materials. Viewers for mzIdentML have been developed, allowing access to the peptide and protein identification within in a way that is easy for biological researchers to comprehend. However, these viewers often require software to be downloaded and installed, and become uncomfortably slow when handling large files. This is partly due to the extensive internal cross-referencing that mzIdentML uses to minimise repetition – this leads to a large number of interrelationships among the data which are difficult to unravel rapidly in response to user requests. Recently, interest has been growing in using the popular Galaxy environment [2] for proteomic data analysis, as this provides a simple web interface for high performance large scale data analysis. While customised proteomics-centric versions of Galaxy such as GalaxyP [3] and GIO [4] have proven popular, inspecting results produced in mzIdentML format requires files to be converted to overly simple tabular formats, or downloaded for visualisation in a locally installed viewer. To overcome these limitations, we have developed a Java library for rapid mzIdentML data extraction and visualisation via the web, that can be used within Galaxy as a visualisation plugin. A novel Java library called ProExtractor was developed using the jmzIdentML library, to extract different proteomics data simultaneously from mzIdentML on the server-side and prepare extracted data to be transferred to the client-side in a lightweight JSON format. AJAX, JQuery, and other modern web technologies are then used to render the JSON data quickly to the browser on demand.

## Architecture and implementation

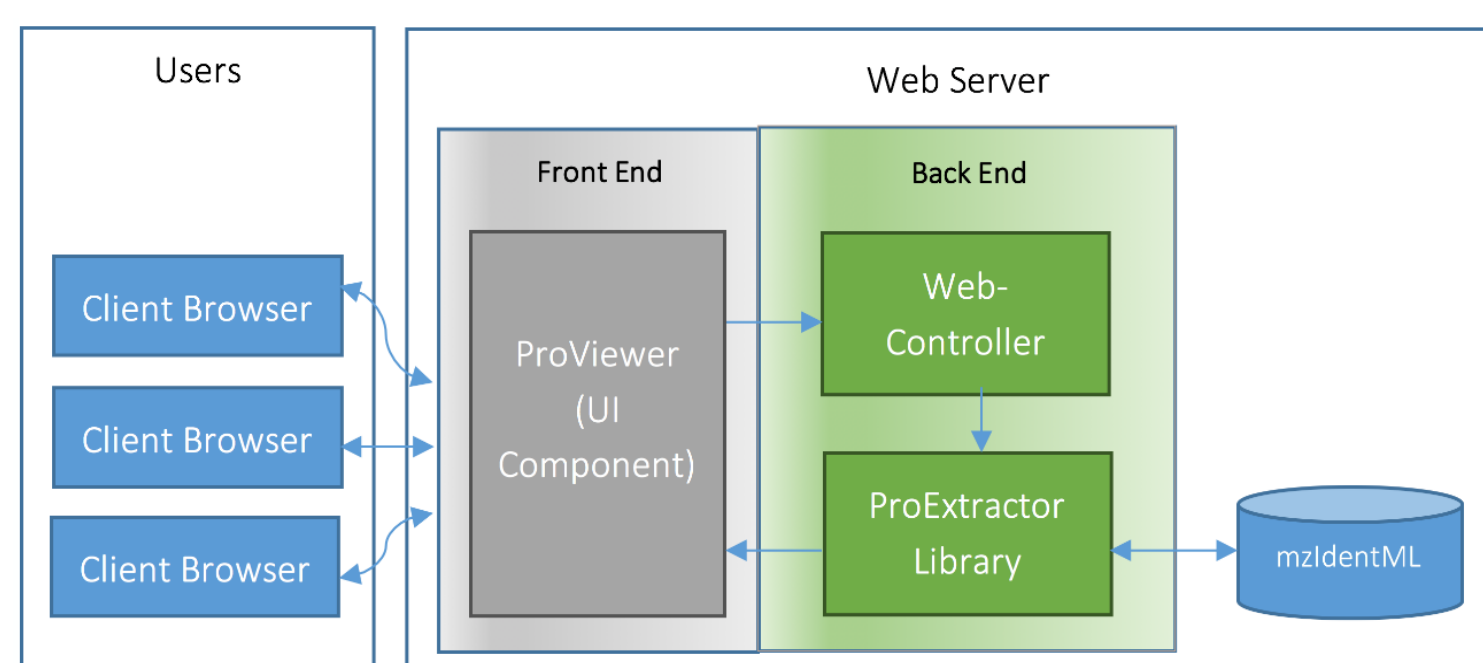
The main component of ProViewer is ProExtractor – a high performance Java library designed to extract pertinent information from an mzIdentML file and convert it to JSON format ready for web visualisation. This library (Figure 1) comprises a data tier to extract data from a mzIdentML file using the jmzIdentML API [5], a business logic tier that simultaneously processes extracted data into separate JSON files for each entity type (protein, peptide, PSM and metadata), and a presentation tier to control the entire process. These tiers are supported by two additional utilities providing common data extraction and conversation functionalities.



**Figure 1: Architecture of the ProExtractor Java library.** The data tier handles data extraction of the mzIdentML file. Extracted data are further processed according to the logic written in the business layer to produce JSON output files. The presentation tier is the main controlling unit, which decides the execution flow of the programme.

A standalone tool called MzIdentMLToJSON, using ProExtractor, was created to allow offline conversion of mzIdentML files for JSON, but the main purpose of ProExtractor is the on-demand preparation of information from mzIdentML files for presentation via the web, as shown in Figure 2.

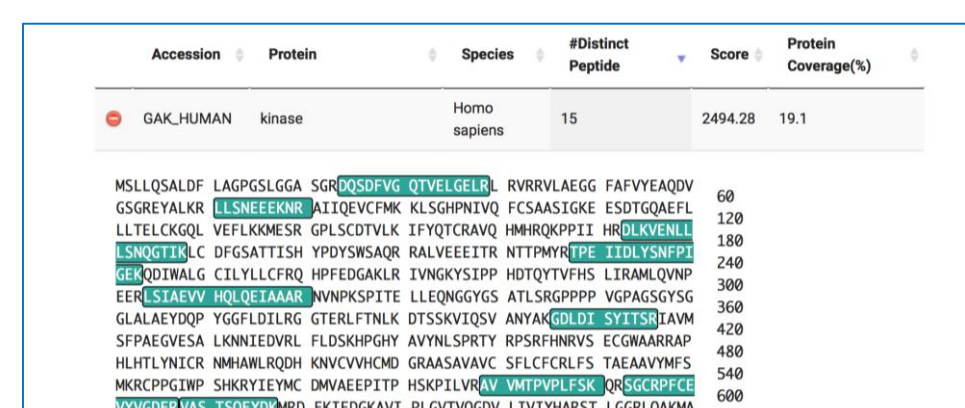
**Figure 2: Overall architecture of ProViewer.** Users request a view of a mzIdentML file via their web browser. This request is handled by the user interface component of ProViewer which calls ProExtractor on the web server to prepare the data.



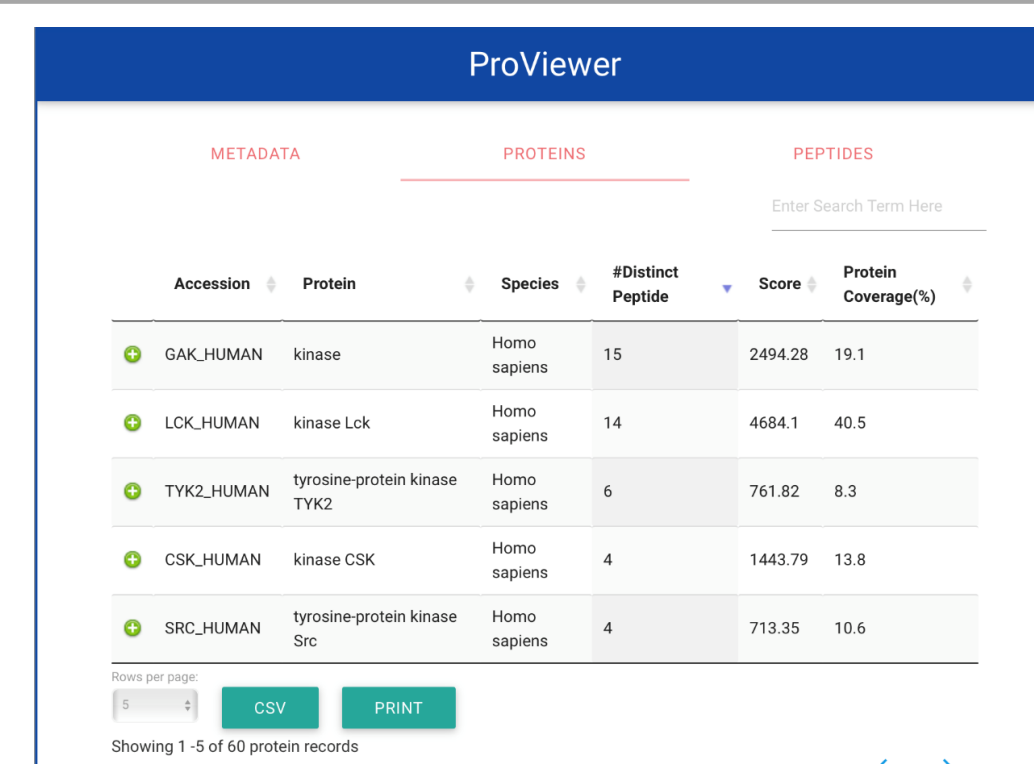
## ProViewer interface

ProViewer interactively visualises inferred proteins and identified peptides from JSON format files using the popular JQuery DataTable library. This highly optimised JavaScript library provides interactive searching, sorting and pagination functionalities, providing a user experience previously only expected from locally installed software.

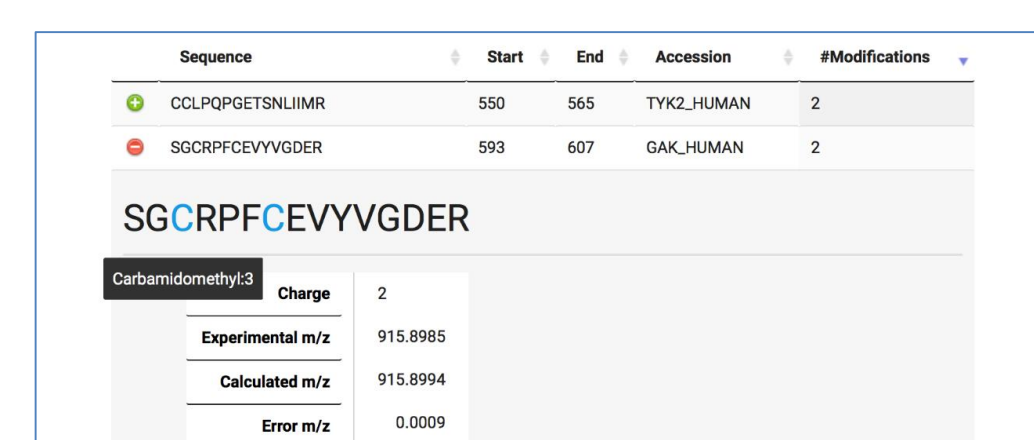
The file visualisation is arranged in three tabs – metadata, proteins and peptides – with buttons below allowing printing of the current table or download in CSV format. Clicking items in the view take the user to further supporting information.



**Figure 4: Protein coverage detail.** If the mzIdentML file contains full protein sequences, peptide coverage can be viewed by clicking the expand (+) button next to the protein name.



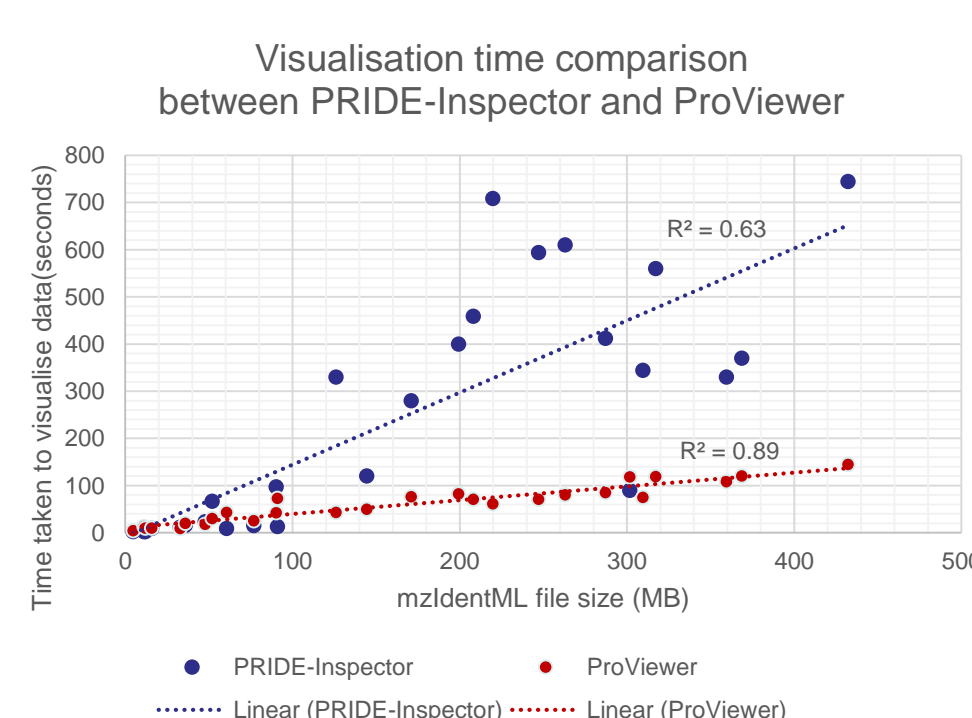
**Figure 3: Protein view.** The protein view (default view) provides an interactive list of proteins in the mzIdentML file, together with basic information about peptide evidence. This is one of three tabbed pages – the others show sample metadata and detailed peptide evidence.



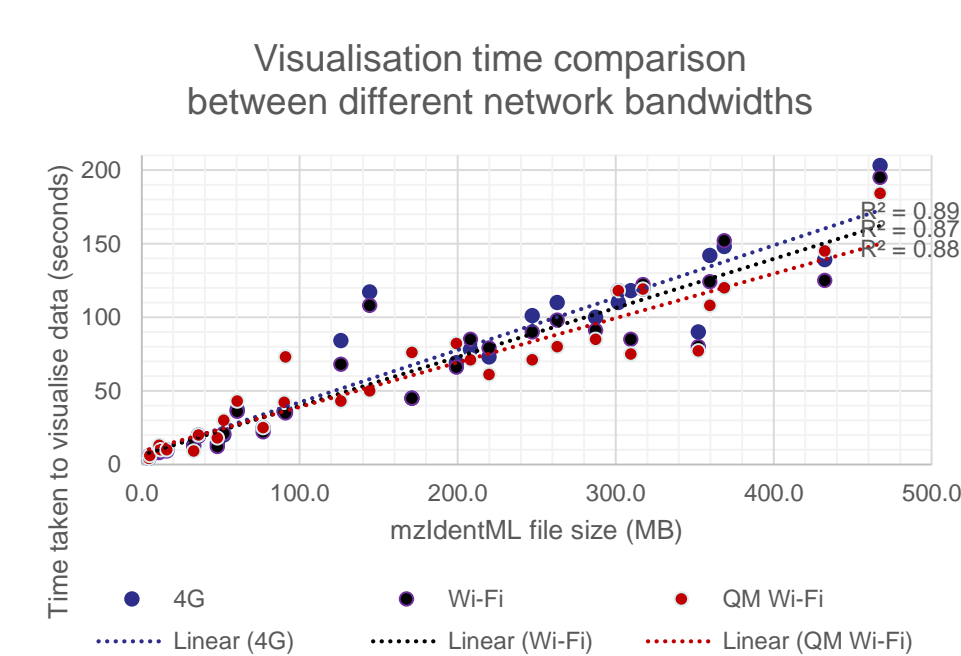
**Figure 5: Peptide modification detail.**

## Testing and performance

ProViewer and the underlying ProExtractor library were extensively tested on a wide range of mzIdentML files downloaded from PRIDE [6] to ensure compatibility with files of different sizes, generated by different software with different parameters, and on all major web browsers. Time from file selection to visualisation of a protein list increases linearly with file size (Figure 6) and is generally several times faster than other viewers. A server-side JSON caching mechanism provides a further speedup, rendering the protein list 50x faster if the file has been visualised before.



**Figure 6: Time from file selection to protein list view for a range of file sizes is generally much lower for ProViewer than for PRIDE-Inspector.**



**Figure 7: Network bandwidth has very little effect on visualisation time as most processing is performed server-side.**

## Conclusions and future development

Development of ProViewer has shown that fast, interactive visualisation of mzIdentML files in a web browser is possible. This is a major boon for proteomic analysis in web frameworks such as Galaxy, and could also be useful in other contexts, e.g. proteomic databases. An obvious future development is extending ProViewer support to mzQuantML and mzTab, for quantitative proteomics.

## References

1. www.psdev.info/mzidentml
2. usegalaxy.org
3. usegalaxy.org
4. gio.sbcs.qmul.ac.uk
5. Reisinger F, Krishna R, Ghali F, Ríos D, Hermjakob H, Vizcaino JA, Jones AR. *Proteomics* 2012, 12(6):790-4.
6. www.ebi.ac.uk/pride