Development of a Web-based Top-Down Data Analysis Tool

Christian Heckendorf, Roger Theberge, Catherine E. Costello, Mark E. McComb.
Cardiovascular Proteomics Center, Boston University School of Medicine, Boston, MA.

BOSTON UNIVERSITY

Overview

- Top-down proteomics has emerged as a technique that preserves labile post-translational modifications and offers full protein sequence coverage.
- One of the major problems facing topdown MS/MS is the assignment of peaks due to the possibility of a large number of fragments from an intact protein.
- Now in development is a web-based version of BUPID-Top-Down (Boston University Protein Identifier Top-Down) used to assign ions in top-down MS/MS spectra.
- Theoretical and experimental ion masses are compared and scored using a log-likelihood ratio driven algorithm to find matches.
- This program will facilitate the penetration of top-down techniques into a greater number of mass spectrometry laboratories.

Introduction to Top Down MS/MS

Top down proteomics involves introducing intact proteins into the mass spectrometer and fragmenting them using methods such as CID, ECD, ETD, etc. This has the potential for complete protein sequence and PTM identification without having to spend time digesting the protein. Making use of top-down data is very computationally taxing and the availability of software that can do this effectively is limited.

There are few tools available for top-down proteomics data analysis; these include:

- Mascot Top-Down (http://www.matrixscience.com): commercial, license required.
- ProsightPTM (http://prosightptm2.northwestern.edu; Nucleic Acids Research 2007; doi: 10.1093/nar/gkm371): free, web based, limited PTM searching capabilities.
- BUPID-Top-Down Desktop Version (BUPID-top-down: Database Search and Assignment of Top-Down MS/MS Data, in: Proceedings of the 57th American Society Conference on Mass Spectrometry and Allied Topics, Philadelphia, PA, 2009).

Here, we describe the development of a web-based continuation of BUPID-Top-Down redesigned for enhanced performance and with improved features which take advantage of the client-server approach.

Challenges

The original version of BUPID-Top-Down was developed as a desktop application to be run on laboratory PCs. This model is not ideal for the laboratory environment for several reasons:

- The speed of the search is limited by the resources available on the desktop minus the resources tied up by the other processes running on the system.
- In order to provide a familiar interface for the users, the original program was written in C# using the .NET framework. This, in part, caused the core code to become entangled with the user interface, preventing it from expanding to support new requirements.
- The PC platform restricted users to PC architecture and necessitated frequent updates to code, protein databases and PTM databases.

Algorithm for Pattern Matching

- BUPID-Top-Down begins by generating a list of theoretical ions that can be derived from the original sequence according to the properties associated with each fragmentation method. Primary fragmentation, internal fragments and post-translational modifications are included in the search.
- Score of the match is defined as the log-likelihood that the peak is generated by this ion over that the peak is generated by random background. If an ion matches with more than one peak, only the match with the highest score is kept for that ion.
- PTM searching is accomplished by shifting the original mass of each theoretical ion that meets the modification requirements before matching it with peaks. If a PTM is known to be on a specific residue, a fixed modification can be added to the search. Adding fixed modifications to a residue is equivalent to using a different mass for that residue in the subsequent search.
- If a modification is suspected to be present in the sample, a variable modification can be applied to the search. When searching for variable modifications, BUPID-Top-Down will search for additional ions using all possible combinations of modifications that can be applied to each ion.
- Definitions of variations and modifications can be loaded from the Unimod database. In addition, users may create custom definitions prior to searching.

ExD/CID Mechanisms for Top Down MS/MS

Collision Induced Dissociation (CID)

- Precursor ions smack into neutral gas and break into fragments
- Kinetic energy is converted to internal energy resulting in bond breakage
- The most common MS/MS technique
- Loss of labile PTM due to introduction of internal energy
- Complementary structure information to ETD and ECD

Electron Capture Dissociation (ECD) / Electron Transfer Dissociation (ETD)

$$[M+nH]^{n+} + e^- \rightarrow \left[[M+nH]^{(n-1)+}\right]^* \rightarrow fragments$$

- Multiply charged protein ions capture a free electron and the radical rearrangement on the backbone carbonyls occurs to break the N-Cα bond
- Outstanding protein sequence coverage
- Preservation of labile PTMs

Web Interface of BUPID-Top-Down

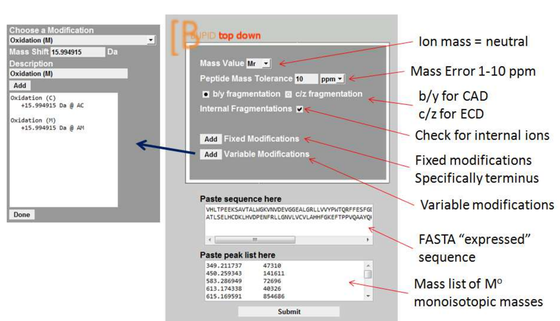


Figure 1, Example search parameters in BUPID-Top-Down.

Results

Example of Peak Assignment Using BUPID-Top-Down

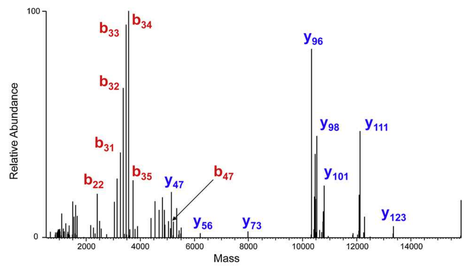
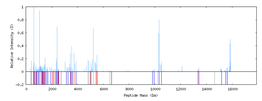


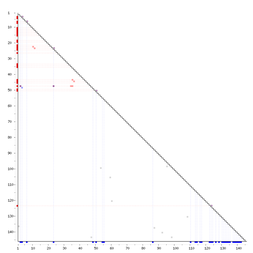
Figure 2, Source CID MS/MS of the beta chain of human hemoglobin: b and y ions shown. (R. Théberge, G. Infusini, W. Tong, M. E. McComb, C. E. Costello, Int. J. Mass Spectrom. 300, 2-3 (2010); doi:10.1016/j.ijms.2010.08.012)

Analysis of Hbβ by Top Down MS/MS

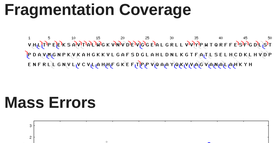
Matched Ions



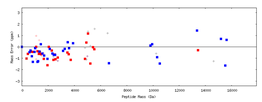
Matched Ions Matrix



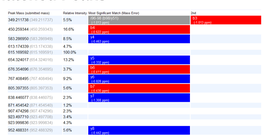
Fragmentation Coverage



Mass Errors



Matched Peaks



Matched Ions Table

Figure 3, BUPID-Top-Down analysis of MS/MS of Hbβ.

Comparison with ProsightPTM

Searches for unmodified ions on both BUPID-Top-Down and ProsightPTM produced comparable results. In addition to the b and y ions, BUPID-Top-Down was able to identify internal ions and variable modifications which were not included in the ProsightPTM search results.

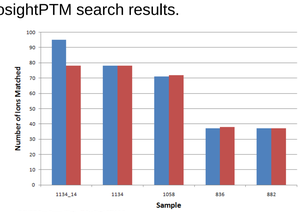


Figure 4, Comparison of matched ions. Samples were processed using equivalent parameters in both BUPID-Top-Down and ProsightPTM.

Performance Improvements

We compared the data processing speed of the original BUPID-Top-Down with our improved program. We found that the improved version is able to process more complex datasets in linear time.

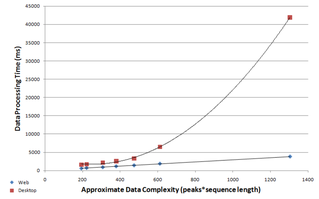


Figure 4, Comparison of BUPID-Top-Down Web vs. Desktop versions. We set up two virtual machines in VirtualBox 4.06, each consisting of 4GB RAM and 4 cores of an i7-950 CPU. The original program ran under Windows 7, while the improved program ran in FreeBSD 8.1. Each of the seven samples were processed four times. This chart shows the average run time for each sample.

Discussion

- By designing the program with a client-server model in mind, we were able to alleviate many of the problems facing the desktop version. At the core is a standalone program that will accept the user data in the form of standard text files and command line options. This allows us to easily deploy the program with a variety of front ends without having to alter the underlying code.
- By writing the code in C and designing it from the ground up with a multithreaded ideology, we were able to achieve greater performance than was possible in the previous design.
- We designed a web interface to attach to the core program. Our interface allows users to easily submit searches as well as review the results in a variety of formats independent of platform.
- The PTM searching capabilities were extended to include all possible combinations of modifications rather than limiting the search to one modification for each ion, thus improving search results.

Conclusions

- Faster processing than BUPID-Top-Down Desktop version.
- Comparable results to ProsightPTM for unmodified searches.
- Assignment of PTMs and internal ions.
- Improved usability of search results.
- Use of BUPID-Top-Down will enable facile and accurate data analysis thus expanding this approach to many users in the field.

Future Work

- Inclusion of a database search to identify unknown proteins.
- Improved ion matching accuracy.

Acknowledgements

This project was funded by NIH/NHLBI N01 HV-28178, NO1 HV-00239, NIH/NCRR P41 RR10888, S10 RR020946 and S10 RR025082.
Special thanks to the members of the CBMS laboratories for input.