

## **Development of a Web-based Top-Down Data Interpretation Tool**

Christian Heckendorf; Roger Theberge; Catherine E. Costello; Mark E. McComb

*Boston University School of Medicine, Boston, MA*

### **Novel Aspect**

Open access web-based software tool with a novel algorithm to interpret top-down MS/MS data obtained from CID, ETD and ECD.

### **Introduction**

Top-down proteomics has emerged as a technique that preserves labile post-translational modifications and offers full protein sequence coverage. However, the progress of the technique is hampered by the lack of availability of readily accessible data interpretation tools. Here we describe the development of a web-based open access version of BUPID Top-Down (Boston University Protein Identifier Top-Down) to assign product ions in top-down MS/MS spectra. The software can be used to analyze spectra obtained with various top-down fragmentation methods including CID, ECD and ETD. The development of an open access top-down data interpretation tool via a web interface will facilitate the penetration of top down techniques in a greater number of mass spectrometry laboratories.

### **Methods**

BUPID-Top-Down compares the masses of theoretical ions with experimental masses. The approach analyses every pair of experimental and theoretical masses to find matches. The score of a match is calculated as the log-likelihood ratio of the probability that the peak is generated by an ion versus the random background, which is derived from the distribution of masses. The algorithm from the desktop version was used as a template and fully implemented in a standalone program written in C to be used in conjunction with a web front-end. The web access version was designed for Linux/Unix servers to be accessible through standard http web browsing.

### **Preliminary Data**

In designing the web-based open access version of BUPID Top-Down emphasis was placed on increasing flexibility and interactiveness to the user and performance enhancement by making use of multithreading and large memory space. For example, the internal ion assignment option can be modulated to provide quicker data analysis in a preliminary assignment mode. BUPID Top-Down's performance was further improved by redesigning database searching to only store protein categorization information of matches rather than all proteins in the database. Currently, top-down benefits from this by requiring less memory for each ion search generated. This in turn results in reducing the "hardware" requirements and increasing accessibility to users on smaller server class hardware. Example results were obtained using BUPID Top-Down by analyzing the CID spectrum of a beta chain of hemoglobin. These were found to be comparable with Prosight (N. Kelleher) and Mascot Top-Down (Matrix Science). An overview of the algorithm, representative results and a comparison to other tools will be presented. The ultimate goal of this work is to develop a web-based Top-Down data interpretation tool with database searching capabilities which may be freely distributed to the MS community.

Acknowledgements: This project was funded by NIH/NHLBI N01 HV-28178, N01 HV-00239, NIH/NCRR P41 RR10888 and S10 RR020946.