Computational Analysis of Mass Spectrometry Isotopic Data

Donald L. Helseth Jr.¹, Don Wolfgeher², Kolbrun Kristjánsdóttir²; Stephen Kron² and David Sigfredo Angulo¹*

¹School of Computer Science, Telecommunications and Information Systems, DePaul University, Chicago, IL 60604

²Department of Molecular Genetics and Cell Biology, University of Chicago, Chicago, Illinois 60637

lhelseth@cti.depaul.edu

Abstract

Analyzing spectra from mass spectrometry (MS) instruments in proteomics is challenging enough under normal circumstances; however, researchers often need to quantify the amount of the protein, which is extremely difficult using MS. Still, MS is the instrument of choice, as it is the most accurate instrument for identifying proteins and is the most sensitive. In order to quantify proteins using MS, researchers frequently utilize the technique of introducing a molecular tag, usually via a stable isotope of naturally occurring biologically relevant atoms (i.e., carbon, nitrogen, oxygen, hydrogen), in order to detect subtle experimental changes in complex mixtures of samples. Existing computational tools for analyzing data were cumbersome and unsuitable for our use so we have developed two specific tools for our use in different types of MS experimental protocols. The first of these experimental protocols, SILAC (Stable Isotope Labeling with Amino Acids in Cell Culture), allows investigators to analyze an experimental condition and experimental control simultaneously (Ong, et al., 2002). This internal control helps remove instrument variation between runs and sample handling errors, as well as prevents inherent limitation on reproducibility from interfering with experimental interpretation, since this measurement variation can be 20% of the signal intensity or greater. IBG-SILAC Quant (the Illinois Bio-Grid tool for SILAC Quantitation) is an application developed by the authors to determine and report on the significant results of a SILAC experiment. The application displays the significant results of their SILAC experiment and indicates the location and intensity of those ions which demonstrated significant increase or decrease in abundance compared with the control.

The second IBG computational tool to assist in quantifying proteins identifies isotopic "fingerprints" in complex samples generated during cleavage of proteins with the proteolytic enzyme, trypsin. Trypsin cleaves peptide bonds at two specific amino acids (Lysine and Arginine), and can be used to introduce $^{18}$O at the newly generated carboxyl terminus by performing the digestion in the presence of $H_2^{18}O$ instead of natural abundance $H_2^{16}O$ (Yao, et al., 2001). Peptides generated in the presence of $^{18}O$ exhibit a characteristic fingerprint of five to eight peaks separated by approximately one Dalton. We have developed an algorithm and software (the IBG $^{18}$O Quantifier) to identify these $^{18}$O fingerprints in the presence of numerous non-labeled peaks in linear time. This algorithm identifies the primary monoisotopic peak and the related isotopic distribution along with the intensity of each of these ions.

This work is funded in part by the National Institutes of Health through the NIH Roadmap for Medical Research, Grant 5R01HG003864

*To whom correspondence should be addressed at dangulo@cti.depaul.edu