Sebastian Ortiz

CHEM 2332 – Lab for Bioanalytical Chemistry

Lab Report 7: Proteomics using HPLC-MS/MS

Lab Partner: Nicole Brossi

March 11, 2020

**Introduction**

This week, there were many objectives to lab. Proteomics, the study of the entire cohort of endogenous proteins in a cell, tissue, or organism, was segmented into two sections: “discovery-driven” (D-D) and “hypothesis-driven” (H-D). The former is an experimental design where the maximum number of peptides and polypeptides in a given matrix are analyzed and identified by LC-MS/MS and compared across samples with changes in biology to determine the changes in abundance of specific proteins. This produces candidate peptides that are changing in abundance, which can then be more precisely determine the significance of the changes—this is H-D. Specifically, this lab introduced both types of proteomics by showing how to analyze a sample with D-D proteomics to identify proteins and peptides in a sample, target a few candidate peptides using H-D proteomics, and analyze the peak area across many replicates and calculate the average peak are and %CV. This ultra-specific, sensitive, and accurate analysis of a sample is made possible by the powerful LC-MS/MS technique, which allows for the separation/isolation, quantification (of concentration), and composition determination of analytes such as peptides and the proteins they come from in a specific matrix. This extremely versatile and potent tool can be used in various industries where specific compounds must be isolated, quantified, and identified, such as the food or pharmaceutical industry.

This experiment was multi-faceted and more centered on the analysis of pre-generated data from prior-made samples. In the D-D phase, LC-MS/MS was used to identify peptides and proteins in a sample. Effectively, a pre-made sample of alpha and beta casein proteins (cleaved by trypsin) was run through LC-MS/MS to visualize peptide ion fragmentation. MS/MS spectrum data would then be used to perform de novo peptide sequencing by calculating the differences in m/z between adjacent fragment ions, and using the Amino Acid appendix in the lab handout to identify the amino acid sequence (*Proteomics Experiments using HPLC-MS/MS* 11). Lastly, the Mascot search engine would be used to perform the same peptide sequence analysis as before (with a given file containing the data), but much faster. This database’s information would then be used to answer some questions regarding the proteins and peptides identified, and the confidence in the identifications.

The H-D portion of the experiment mainly focused on analysis of mass spectrometer (MS) data from a triple quadropole MS to detect and quantitate peptides digested in a HeLa lysate sample. The MS/MS settings were experimentally changed to optimize fragmentation for MS1 and MS2 data; the most effective parameters were recorded. Then, the Skyline (proteomics data analysis) software would be used to analyze given pre-generated MS data and thus report average peak area and reproducibility (%CV – coefficient of variation).

**Materials and Methods**

The experiment was more theoretical than most previous, but nonetheless performed as described in the handout for this laboratory (*Proteomics Experiments using HPLC-MS/MS).* All samples used to generate data were pre-made by other personnel (likely the TA).

The main software used to analyze data were the MS data presentation and analysis software associated with each respective MS machine (Orbitrap and quadropole), Mascot MS/MS ions search, Skyline, and UCSF’s Protein Prospector.

The lab handout included many questions, tables, and figures to be completed; these will be presented in the most adequate sections and in order from D-D to H-D relevance. The questions will be presented in the discussion/conclusions section, while tables and figures will be presented in the results section, with further verbal analysis to follow in the discussion/conclusion section.

**Results**

The sample used for D-D detection of peptides via Orbitrap MS contained digested alpha- and beta-casein proteins. The sample was prepared as 56 µL of Mobile Phase (0.1% formic acid in water [v/v]) 4 µL of the 4 pmol/µL stock of the digested casein proteins. 25 µL of this was aliquoted into 2 autosampler vials for the Orbitrap and QqQ MS, respectively. Although not personally performed in lab, the necessary quantities of components needed was part of the first question, as shown and answered in Figure 1 below.

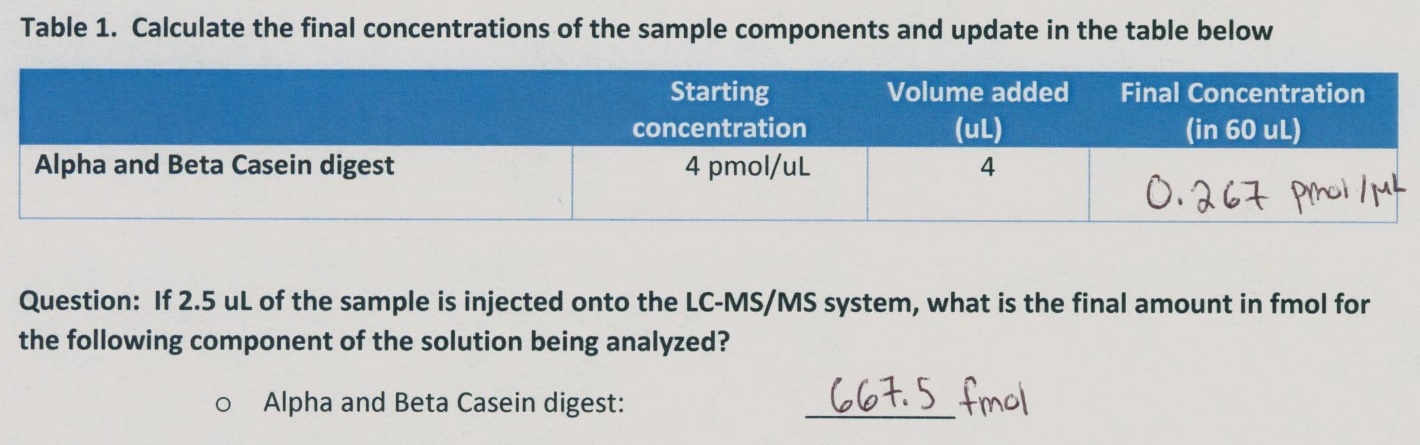
****

Figure 1 - Sample Preparation and Necessary Quantities for Analysis

Once the sample is run and the MS/MS spectrum data is generated by the MS machine, it is possible to perform de novo peptide sequencing by calculating the difference in m/z between adjacent fragments ions of the same type (b- or y-ions) and determining if this difference equates to the mass of an amino acid; if so, then that amino acid is part of the sequence of the fragmented polypeptide. This works because as the polypeptide is fragmented into peptides based on m/z, meaning that each subsequent fragment on the spectrum (as mass decreases) denotes the ‘loss’ of an amino acid, which was fragmented from the polypeptide. Another assignment of this lab was to label the fragment ion peaks of various MS/MS spectra with the proper fragment identity (i.e. b2+, y6+, y8++, etc.). The USCF prospector software was permissibly used to help identify fragment ions’ identities. However, this was only useful if the peptide sequence was known, so for the last (bonus) spectra, with an unknown sequence, it had to be done by hand as outlined earlier. These completed spectra are presented in Figures 2-5 below.

**A close up of text on a white background

Description automatically generated**

Figure 2 - Completed (identified) MS/MS Spectrum 1

**A close up of text on a white background

Description automatically generated**

Figure 3 - Completed (identified) MS/MS Spectrum 2

**A close up of text on a white background

Description automatically generated**

Figure 4 - Completed (identified) MS/MS Spectrum 3

**A close up of a map

Description automatically generated**

Figure 5 - Completed (identified) MS/MS Spectrum 4

That concluded the D-D analysis portion. The H-D analysis began with some peptides identified in D-D being analyzed via a triple quadropole MS. CID fragmentation optimization of peptides by infusion on the LTQ Orbitrap MS was troubleshooted; sensitivity was improve for detection on the Triple Quadropole MS by infusing a concentrated stock solution and optimizing the collision energy. This was performed as directed in the lab handout. A pre-generated MSQC data file was then analyzed using the Skyline software, and the average peak and % CV were determined and reported in Figure 6 below.

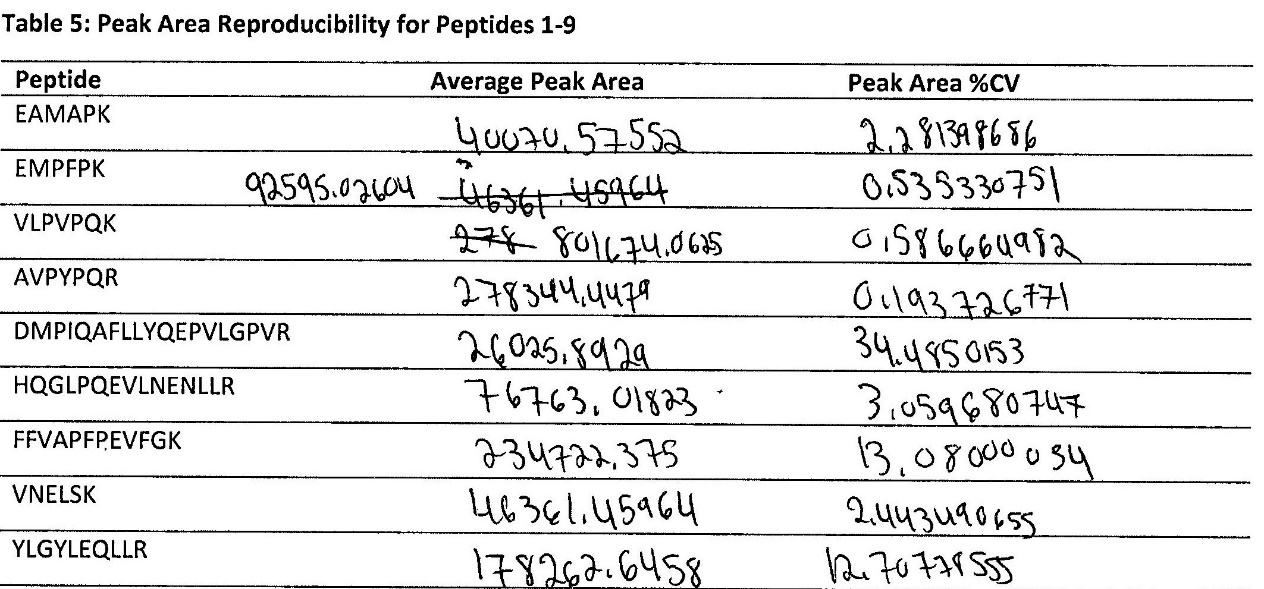


Figure 6 - Reported Analysis from Template MSQC Data via Skyline

**Discussion/Conclusions**

This lab being more theoretical than experimental, there not much data or significant results to discuss. However, that void is filled by various questions (found throughout the lab handout) that related to the D-D and H-D procedures and the theory behind the methods used. Those questions (not already answered in the results section) will be addressed now.

One page 6 of the lab handout, there was an investigation of LC-MS/MS data from the LTQ Orbitrap XL, where various parameters were tinkered with to elucidate the effects of said variables on the data, and thus reveal the differences in MS1 and MS2 scans. The answers to the questions on page 6 are included here. The MS1 and MS2 both separate ions by their m/z ratio, but the way in which they do so, and what results they yield are different. MS1 ionizes the molecules in the sample and separates them by their m/z. The ions produced here are sometimes referred to as precursor ions because some of these ions will be selected and fragmented into smaller ions by collision-induced dissociate (CID) and then introduced into the MS2. In the MS2, the smaller ions are separate by m/z and detected. This fragments in MS2 are separated ions of closely similar m/z, which is hard to visualize in MS1; in other words, MS2 allows for identification and separation of similar m/z ratio ions.

To identify the peptides in a sample used in the D-D LC-MS/MS procedure, the Mascot database was used. The procedure essentially consisted of uploading the LC-MS/MS data file (MGF) into the Mascot Proteome Discoverer, with specific parameters that would facilitate the prediction of peptides generated during trypsin digestion, and are outlined in the lab handout. The questions on page 7 of the lab handout were answered using the data from the database software, and detailed below. The proteins (and respective sequence coverage) identified in the file that we uploaded were: CASA1 (72%), CASA2 (50%), CASB (50%), CASK (16%), Trypsin (16%), 0002134.1 (3%), LACB (6%), 00701268.1 (2%), IPI00406377.3 (1%), and 00715287.2 (2%). The number of peptide matches identified (with the same order as the above proteins, meaning the first value presented here will relate to the first protein outlined above, and so on subsequently) were: 51, 28, 25, 4, 3, 1, 1, 1, 1, 1. In terms of confidence of identification, it is known that Trypsin was in the sample, thus, it can be hypothesized that any protein(s) above Trypsin (in terms of sequence coverage and peptide matches ) is likely present in the sample (denoted by red text color above).

Lastly, as was mentioned above, the parameters for MS/MS data was messed with to see how different things affect the data. When there is a higher charge (z) of 4, then there is a higher number of smaller peaks, and the same is true for more collision energy in the system. This makes sense because more energy being introduced would cause more fragmentation of larger peptides into smaller peptides, even single amino acids. When the parameters were changed to a lower z (1), but a similar m/z (1591 and 1521 m/z) there will be various small/medium fragments of a larger m/z peptide. This leads to a more even distribution of fragments across the m/z axis in comparison to a smaller fragments of lower m/z peptides. The peptides of lower m/z have fragments concentrated at the lower scale of m/z side of the axis, due to the fragmentation of a smaller peptide into even smaller pieces.

That concludes the general questions in the lab handout. Now, to finish, the final questions from the lab handout will be addressed to conclude this report.

1. The D-D is the whole fragmentation of the polypeptide, but the H-D shows select ionized peptides.
2. Because the analytes of interest are known, thus, specialized assays and methods for analyzing those specific analytes can be used, expediting the experiments.

**References**

Department of Chemistry. (Spring 2020). CHEM 2332 Lab Manual, *Proteomics Experiments using HPLC-MS/MS*. Boston, Massachusetts: Northeastern University, 2020.

MacLean, B., Tomazela, D. M., Shulman, N., Chambers, M., Finney, G. L., Frewen, B., Kern, R., Tabb, D. L., Liebler, D. C., & MacCoss, M. J. (2010). Skyline: an open source document editor for creating and analyzing targeted proteomics experiments. Bioinformatics (Oxford, England), 26(7), 966–968. <https://doi.org/10.1093/bioinformatics/btq054>

Mascot Discovery Proteomics Database. (n.d.). Retrieved from <http://www.matrixscience.com/cgi/master_results_2.pl?file=../data/20191107/FTemSGamO.dat;sessionID#7:pr.eh=(1-5)p,1-5;pr.show=proteins>

MS-Product Peptide Sequence Database. (n.d.). Retrieved from http://prospector.ucsf.edu/prospector/cgi-bin/msform.cgi?form=msproduct