Collaborative learning exercises for Teaching Protein Mass Spectrometry [post-print]

Michelle Kovarik
Trinity College, Hartford Connecticut, michelle.kovarik@trincoll.edu

Jill K. Robinson
Collaborative Learning Exercises for Teaching Protein Mass Spectrometry

Michelle L. Kovarik,* Jill K. Robinson**

*Department of Chemistry, Trinity College, 300 Summit St., Hartford, Connecticut, 06106
**Department of Chemistry, Indiana University, 800 E. Kirkwood, Bloomington, Indiana, 47405

ABSTRACT

A collaborative learning module for teaching protein mass spectrometry has been developed to overcome common obstacles to incorporating the modern topic of biological mass spectrometry into the undergraduate chemistry curriculum. Protein mass spectrometry data is provided to eliminate the need for expensive instrumentation, and a detailed instructor’s manual gives necessary details for those unfamiliar with the topic. The first section provides background information on proteins and the field of proteomics. The second section describes the use of electrospray ionization to determine the molecular weight of a protein. The third section shows how to identify a protein using peptide mass mapping, and the fourth section describes tandem MS experiments for de novo peptide sequencing. Each section also includes lessons on the analytical instrumentation used to make mass measurements including electrospray ionization, matrix assisted laser desorption ionization, and time of flight mass spectrometry. The module includes pre-class reading assignments and small group problem solving exercises to be used during class sessions. The module was implemented over several semesters at both a small liberal arts college and a large research university. Assessment data from both institutions suggest that the module is effective in helping students to learn about mass spectrometry-based proteomics. This freely available resource will assist instructors in introducing these topics to the undergraduate curriculum.
INTRODUCTION

The invention of electrospray ionization (ESI) and matrix assisted laser desorption ionization (MALDI) extended mass spectrometry measurements to large molecules, including biological macromolecules such as proteins. In the past 40 years, improvements in mass spectrometric methods for analyzing proteins have led to advances the field of proteomics. Figure 1 shows the dramatic increase in the number of publications found in the Web of Science database with the search terms protein, proteomics, and mass spectrometry since the 1990s. Many protein analysis methods based on mass spectrometry are now well-established and used routinely, while new methods continue to be developed to improve quantitation, decrease detection limits, identify post-translational modifications, and elucidate protein structures and interactions. Despite the rapid maturation of these technologies, undergraduate analytical chemistry curricula have mostly remained focused on mass spectrometry of small molecules, emphasizing interpretation of electron impact (EI) spectra to solve molecular structures and applications of GC-MS in the laboratory. While these skills remain relevant, exposure to newer techniques (e.g., ESI, MALDI, HPLC-MS, and tandem MS) would substantially enhance students’ preparation for post-graduate study or work in the biomedical and pharmaceutical industries.
Several obstacles exist to incorporating biological mass spectrometry in the undergraduate curriculum. Many current instructors completed their training before the advent of biomolecule applications of MS and may not feel fully confident teaching these topics to their students. Furthermore, commonly used text resources for teaching analytical chemistry and biochemistry have limited coverage of protein mass spectrometry. Commonly used textbooks in the analytical chemistry curriculum such as *Principles of Instrumental Analysis*, *Instrumental Analysis*, and *Analytical Chemistry* include descriptions of MALDI and ESI, but have little to no coverage of mass spectrometry of proteins.\textsuperscript{5–7} The books *Quantitative Chemical Analysis*, *Undergraduate Instrumental Analysis*, and *Biochemistry* both have sections on protein mass spectrometry, but few end-of-chapter problems and no use of bioinformatics tools.\textsuperscript{8–10} In contrast, a new analytical chemistry book, *Analytical Chemistry: A Chemist and Laboratory Technician’s Toolkit*, has an entire chapter on the mass spectral analysis of proteins.\textsuperscript{11}
The laboratory is an ideal venue for introducing biological mass spectrometry, and several papers have been published in this journal describing the implementation of proteomics experiments, such as determining the molecular weight of a protein, de novo peptide sequencing, and protein identification by peptide mass fingerprinting. However, the high cost of MS instrumentation in general and HPLC-ESI-MS and MALDI-TOF in particular means that few undergraduate programs have access to advanced instrumentation for student use. MS instrumentation is not required by the Committee on Professional Training for ACS certification of undergraduate programs, but it is encouraged since departments must have instrumentation from four of the five major instrumentation categories. Additionally, some protein separation methods (particularly 2D gels) include lengthy wait times that can be difficult to accommodate in a laboratory schedule, and some technical experience is generally required to obtain good results. Perhaps as a result, a 2005 survey of 64 instrumental analysis instructors found that 58% introduce students to LC-MS in lecture, but only 11% introduce LC-MS in the laboratory component of the course. However, the same survey found that coverage of MS beyond GC-MS has increased since previous surveys in 1992 and 1998. More recent survey data is unavailable; however, based on the increasing application of MS techniques to biological problems (Figure 1), it is likely that this trend has continued in the past 14 years. To address the problem of low coverage in textbooks and expensive MS instrumentation, we developed a module for teaching protein mass spectrometry which focuses on the fundamentals of instrumentation and the interpretation of protein and peptide mass spectra.

**PROTEIN MASS SPECTROMETRY MODULE**

The protein mass spectrometry module uses a collaborative pedagogy that is supported by research on how students learn. Class activities have been developed so students spend a significant portion of the period working in small groups to discuss and solve problems. This gives them opportunities to apply their knowledge and receive immediate and relevant feedback from their peers and/or the instructor. The instructor spends less time “lecturing” and transmitting information and more time helping students deepen their understanding. The module can be found on the Analytical Sciences Digital Library (ASDL) website in the active learning section under the In-class
Activities category. Materials on the ASDL site are intended for flexible and customizable use and are provided in an editable format that can be easily adapted for an instructor’s specific use. An instructor’s manual with teaching tips and solutions is available to make the topic more accessible for those who do not have experience with biological mass spectrometry, and data sets are provided so expensive MS instrumentation is not needed to implement module.

**Learning Objectives and Timeline**

The module is composed of four sections. Table 1 lists the topics, learning objectives, and approximate number of 50-75 min class periods needed for each section. (Class period estimates are approximate as the time required will vary with instructor, students, duration of period, and institution.) Section 1 is optional for courses in which students have already taken biochemistry as this section sets the foundation for learning protein mass spectrometry by providing background on the primary structure of proteins and computation of monoisotopic and average molecular weights of peptides. Sections 2-4 each cover a specific, MS-based proteomics experiment, and these sections could be adopted independently of one another. Section 2 describes the method for calculating the molecular weight of a protein from its ESI spectrum and the operation of a quadrupole mass analyzer. Section 3 covers the method of peptide mass mapping for protein identification and the principle of operation of MALDI-TOF mass spectrometry. Section 4 provides instruction on *de novo* sequencing of peptides using tandem MS data.

### Table 1. Topics and Learning Objectives for Protein Mass Spectrometry Module

<table>
<thead>
<tr>
<th>Topic</th>
<th>Number of Class Periods</th>
<th>Learning Objectives</th>
</tr>
</thead>
<tbody>
<tr>
<td>Section 1. Proteins and Proteomics</td>
<td>2</td>
<td>• Draw the chemical structure of an amino acid and small peptide.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Describe the difference between free and residue amino acid mass.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Calculate the monoisotopic and average mass of a peptide.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Define proteomics and interpret a scientific paper on the application of proteomics in cancer research.</td>
</tr>
<tr>
<td>Section 2. Calculating the MW of a Protein from its ESI-MS Spectrum</td>
<td>1</td>
<td>• Calculate the charge state and mass of a protein from its ESI-MS spectrum.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Describe the principle of operation of a quadrupole ion trap mass analyzer.</td>
</tr>
<tr>
<td>Section 3. Peptide Mass Mapping for Protein Identification</td>
<td>4</td>
<td>• Describe the technique of matrix assisted laser desorption ionization (MALDI) for creating gas phase ions of peptides and proteins.</td>
</tr>
</tbody>
</table>
Section 1 Proteins and Proteomics

The first section of the module provides the foundation for the study of proteomics and begins with the amino acid composition and primary structure of proteins. Once students are able to draw structures of peptides, they learn to calculate the average and monoisotopic mass of a peptide and discuss which mass will be observed in a mass spectrum. Figure 2 shows a small mass-to-charge region in the MALDI-TOF mass spectrum of a peptide. The peptide has a monoisotopic mass of 1898.7, and peaks with higher m/z values are due to the presence of higher mass isotopes of carbon, nitrogen, and oxygen such as $^{13}$C, $^{15}$N, or $^{18}$O. The $[\text{peptide} + \text{H} + 1]^+$ peak has a greater intensity than the $[\text{peptide} + \text{H}]^+$ peak because in a molecule with more than 100 carbon atoms, the chance of having at least one carbon-13 atom is greater than having all carbon-12 atoms. Many mass spectrometers have sufficient resolution to distinguish the monoisotopic peak from the isotope peaks, but in a lower resolution mass spectrometer the average molecular mass may be obtained. The field of proteomics and its utility are then introduced by having students read and identify the main goals of a comparative proteomics paper investigating proteins associated with oral cancer cell lines. Students also interpret images of two-dimensional gels and summarize the results.

Figure 2. The MALDI mass spectrum of a peptide with a monoisotopic molecular weight of 1898.7. Peaks at higher m/z are due to the presence of higher mass isotopes of carbon, nitrogen, and oxygen such as $^{13}$C, $^{15}$N, or $^{18}$O.
Section 2 Calculating the MW of a Protein from its ESI-MS Spectrum

The second section of the module introduces students to electrospray ionization (ESI) of proteins. The development of ESI advanced the field of proteomics in several ways. ESI ionizes charged species directly from a condensed, liquid phase making it an ideal ionization source to couple liquid chromatography (LC) separations to MS. Most proteomics samples are aqueous solutions containing a complex mixture of small molecules, peptides, and proteins which require a separation step such as LC in sample preparation. Additionally, ESI is a soft ionization method that produces multiply charged ions for intact proteins. This is important in the measurement of protein molecular weight, which may be outside the mass range of many mass analyzers (such as quadrupoles) when the protein is singly charged. Using the module, students learn the theory behind the operation of ESI and quadrupole mass analyzers, as well as the interpretation of ESI spectra. Because most students’ first introduction to MS data is through GC-MS in organic chemistry, they may not have previously seen mass spectra where \( z > 1 \). The exercises in this section help students recognize the characteristic charge envelope of a protein’s ESI spectrum (Figure 3), then perform the calculations to obtain the charge states and molecular masses of ubiquitin and cytochrome C.

![Figure 3. A representative ESI-MS spectrum of cytochrome C (12 kDa).](image)
Section 3 Peptide Mass Mapping for Protein Identification

Peptide mass mapping is a method used to identify a protein using mass spectrometry and bioinformatics. Although peptide mass mapping is not as widely used as more current LC-MS/MS methods, it is well suited to teaching the foundations of protein identification due to its simplicity. This section begins with a description of relevant mass spectrometry techniques including matrix assisted laser desorption ionization (MALDI) and time-of-flight (TOF) mass spectrometry. In the peptide mass mapping workflow, liquid chromatography or gel electrophoresis is first used to separate components in a complex mixture of proteins. Once the protein has been isolated, it is digested with the enzyme trypsin, which cleaves the protein after lysine and arginine residues. The purpose of digesting the protein is to produce many peptide masses, which leads to multiple matches in the identification process. In a similar manner to small molecule analysis by mass spectrometry, the molecular mass alone does not provide sufficient information to identify a substance. However, the masses of many specific fragment ions can be used to deduce structure and/or identify the substance when compared with a known or predicted spectrum.

Once the masses of the peptides are measured by mass spectrometry, their values are entered into a bioinformatics database such as Mascot, which matches experimental mass data with theoretical mass data generated from genome information. From a sequenced genome, a computer can generate the amino acid sequence of all the proteins for that organism, and then compute the predicted masses of all the peptides generated by digestion with trypsin. The list of peptide masses the computer generates upon digestion of a protein with trypsin is called an in silico digest. Figure 4 outlines the steps in the process of peptide mass mapping.
Figure 4. Steps in the process of peptide mass mapping for protein identification.

An example of a MALDI-TOF mass spectrum of a protein spot cut from a two-dimensional gel and digested with trypsin is shown in Figure 5. The module provides mass spectrometry data for several proteins along with recommended parameters for the Mascot database search. Some proteins do not produce statistically significant results using default search parameters and examples were chosen that require manipulation of search parameters. The lessons guide the students through the statistical thinking and experimentation needed to alter the parameters and achieve relevant results.

Figure 5. MALDI-TOF spectrum of human serum albumin digested with trypsin. The peaks correspond to the masses of different peptides fragments formed after the 66 kDa protein bovine serum albumin was digested with trypsin.

Section 4 Sequencing Peptides from Tandem MS Data

The fourth section of the module addresses tandem MS methods and de novo peptide sequencing. In tandem mass spectrometry, a mass analyzer is used to select the m/z value of a precursor ion,
which is then fragmented to product ions, which are measured in the MS-MS spectrum. (In trapping mass analyzers, MS^n experiments can be conducted, in which product ions are further fragmented.) Tandem MS experiments are useful for obtaining structural information, particularly in the case of isomers. For example, peptides composed of the same amino acid residues have the same molecular mass regardless of the amino acid sequence. MS-MS data can reveal the amino acid sequence based on fragmentation patterns alone, without the need for a genome database, in a process called de novo sequencing.

In Section 4, students are first introduced to a common method of fragmenting precursor ions – collision induced dissociation (CID) – and the operation of a triple quad instrument through a video (produced by Agilent and freely available online). A background reading assignment provides information on peptide fragmentation in CID and how the spectra are interpreted to determine the amino acid sequence of a peptide. During CID, peptides typically fragment along the backbone at the peptide (amide) bond. Depending on whether the charge stays with the C-terminus or N-terminus of the peptide, b or y ions are produced (Figure 6). Consequently, the amino acid sequence can be read off the mass spectrum because the change in mass between adjacent peaks often corresponds to the loss of a specific amino acid residue. In-class discussion questions guide students through an analysis of the MS-MS spectrum of the neuropeptide leu-enkephalin and a final challenge problem asks students to determine the sequence of an unknown peptide based on its MS-MS spectrum.
RESULTS AND DISCUSSION

The authors have used these materials in two distinct educational settings. These materials were implemented at a liberal arts college in the 300-level instrumental analysis course required for all chemistry majors. In this venue, class size ranged from 6-9 junior and senior students over the course of 3 semesters of implementation. Students completed portions of Sections 1-3 as homework and in class. For Section 4, they completed the module as an in-class activity along with an accompanying laboratory exercise, collecting their own data for de novo sequencing using direct infusion to an ESI-ion trap MS (Finnigan MAT LCQ). These materials have also been used in a bioanalytical class at a large public university. The class size ranged from 50-70 sophomore to senior-level students majoring in biochemistry and biotechnology over the course of 2 semesters of implementation. Students completed questions from the module as in-class small group activities. In both settings, the class periods consisted of small group work on module questions along with instructor facilitated whole class discussion.

We constructed assessment questions aligned with the learning objectives of the module (Table 1) at varying levels of Bloom’s taxonomy for each section. Assessment questions ranged from levels 1 and...
2 involving recall and comprehension to higher level questions involving critical thinking and interpretation. Level 3 requires old concepts to be applied to a new problem; level 4 requires analyzing patterns and organizing ideas such as interpretation of data; level 5 involves the creation new ideas such as the prediction of experimental results; and level 6 involves comparisons, judgements and recommendations.\textsuperscript{35} Table 2 shows the results of some representative assessment questions that students answered as part of routine semester exams during the most recent implementation of the module at two institutions. Student work was scored by the instructor, and scores represent the percentage of total points for a given question that students received (rather than the percent of student who received full credit.) We found that the student outcome data were non-normally distributed with many responses receiving full credit and small numbers of responses receiving little to no credit. For this reason, we have reported the median rather than the average and the interquartile range rather than the standard deviation to show the distribution of scores for each assessment item.

Students received high marks for most assessment questions with median scores of 100\% for over half of all questions (Table 2), including many challenging questions. For example, the learning objective “Identify \textit{b} and \textit{y} ions and use them to determine the sequence of a short peptide by \textit{de novo} sequencing” was assessed by the two questions given for Section 4. These questions required students to apply their knowledge of \textit{de novo} sequencing principles to a new peptide that they had not seen previously. Median scores were >85\% for both these questions. Similar results were obtained for high level questions based on other sections. For example, students earned about three-fourths of the available points for evaluating the utility of ESI-MS for antibody characterization (Section 2) and predicting changes to a MALDI-TOF spectrum with and without a reflectron (Section 3), even though they had never encountered these questions previously. Although students had lower scores on some assessment questions, averages on the exams covering the proteomics sections of courses were comparable to scores for exam questions on other topics. This outcome is particularly striking because the course textbooks included substantial coverage of other topics, but minimal to no coverage of proteomics. Since undergraduate students do not generally have prior knowledge of mass spectrometry or proteomics at the start of an analytical chemistry course, these student outcomes suggest that the module is an effective tool for introducing students to protein mass spectrometry.
Additionally, although no formal assessment of student attitudes toward the module was conducted, informal feedback from students suggested that they enjoyed and valued these activities.

**CONCLUSION**

We have described a module for teaching mass spectrometric methods for protein analysis to undergraduate students. Biological mass spectrometry, especially in its application to proteins, has become a major research area in analytical in the last several decades, and this module is a resource to faculty introducing these topics to their curriculum. We expect that the module will be particularly useful because it is freely available, assumes no background knowledge in proteomics on the part of instructor or students, and does not require access to MS instrumentation. The module has been constructed with “best practices” of active learning in mind, and initial assessment data suggest that it is effective in meeting the desired learning objectives.
### Table 2. Comparative Student Outcome Data Pooled from Two Institutions.

<table>
<thead>
<tr>
<th>Topic by Section</th>
<th>Assessment Items</th>
<th>Bloom’s Level(^a)</th>
<th>Interquartile Range, (N=59) %</th>
<th>Median, %</th>
</tr>
</thead>
<tbody>
<tr>
<td>Section 1. Proteins and Proteomics</td>
<td>• Calculate the monoisotopic mass of the intact FAVE peptide.</td>
<td>3</td>
<td>80-100</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td>Section 2. Calculating the MW of a Protein from its ESI-MS Spectrum</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>• In your own words explain how electrospray ionization generates charged, gas phase proteins for analysis by mass spectrometry.</td>
<td>2</td>
<td>50-100</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td>• On the mass spectrum below, label each peak with its charge state and calculate the molecular weight of the protein. Base your calculations on the peaks with (m/z) values of 1060.44 and 1131.05.</td>
<td>3</td>
<td>80-100</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td>• Compare and contrast electron impact ionization with electrospray ionization. Both methods result in mass spectra with multiple peaks for a single analyte. What factors result in multiple peaks for each ion source? Are they the same factors for both sources?</td>
<td>4</td>
<td>50-100</td>
<td>67</td>
</tr>
<tr>
<td></td>
<td>• You are using ESI-MS to characterize an antibody-based drug for a pharmaceutical company. An antibody is a large, Y-shaped protein. You determine the molecular weight by ESI-MS. Is ESI-MS a good technique for this application? Is ESI-MS characterization sufficient to ensure the drug molecule is in the correct molecular form? Justify your responses.</td>
<td>5-6</td>
<td>50-100</td>
<td>75</td>
</tr>
<tr>
<td>Section 3. Peptide Mass Mapping for Protein Identification</td>
<td>• Human serum albumin is the most abundant protein in human blood. The molecular weight of human serum albumin is 66.5 kDa. Sketch the expected appearance of a MALDI-TOF mass spectrum of a pure sample of human serum albumin protein.</td>
<td>3</td>
<td>24-100</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td>• The mass spectrum of two ions with similar masses in a linear TOF instrument is shown below. Sketch how the mass spectrum would change if a reflectron TOF were used to acquire the mass spectrum instead of a linear TOF instrument.</td>
<td>5</td>
<td>50-100</td>
<td>75</td>
</tr>
<tr>
<td></td>
<td>• In the original search, the mass tolerance was set to 1.0 Da. Predict how the probability score will change if the mass tolerance was set to 0.3 Da and the same 6 peptides were matched? Justify your reasoning.</td>
<td>4-5</td>
<td>50-100</td>
<td>100</td>
</tr>
<tr>
<td>Section 4. Sequencing Peptides from Tandem MS Data.</td>
<td>• Consider the structure below for the peptide FAVE. Mark the bonds that are cleaved to form the (b) and (y) ion series, then label each potential (b) and (y) ion by number.</td>
<td>2</td>
<td>77-100</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td>• The spectrum below shows tandem MS data used to sequence a peptide. Analyze the data to determine a sequence of three consecutive amino acids in the peptide. Indicate which end of the sequence is the C-terminus and which is the N-terminus.</td>
<td>4</td>
<td>75-88</td>
<td>88</td>
</tr>
</tbody>
</table>

\(^a\)Values range from 1–6, with lower levels (e.g., 1 and 2) involving recall and comprehension and higher levels (e.g., 5 and 6) involving critical thinking and interpretation. See ref 35.
Range from the 25\textsuperscript{th} to the 75\textsuperscript{th} percentiles of scores.

**AUTHOR INFORMATION**

**Corresponding Authors**

*E-mail: michelle.kovarik@trincoll.edu

**E-mail: jirobins@indiana.edu

**ACKNOWLEDGMENTS**

The authors thank Dr. Jon Karty for collection of mass spectrometry data and Prof. Tom Wenzel for helpful discussions. We also thank the National Science Foundation for supporting this work through a Transforming Undergraduate Education in STEM Award (DUE-1118600).

**REFERENCES**


