JOURNAL OF CHEMICAL EDUCATION

Laboratory Experiment

pubs.acs.org/jchemeduc

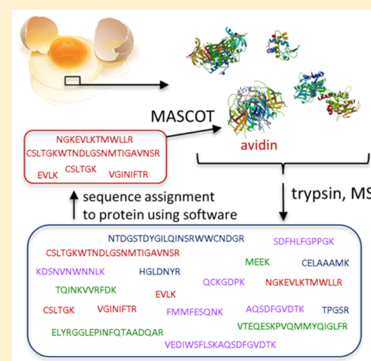
Peptide Mass Fingerprinting of Egg White Proteins

Lisa T. Alty* and Frederick J. LaRiviere

Department of Chemistry and Biochemistry, Washington and Lee University, Lexington, Virginia 24450, United States

Supporting Information

ABSTRACT: Use of advanced mass spectrometry techniques in the undergraduate setting has burgeoned in the past decade. However, relatively few undergraduate experiments examine the proteomics tools of protein digestion, peptide accurate mass determination, and database searching, also known as peptide mass fingerprinting. In this experiment, biochemistry students digest a protein mixture from egg white using the enzyme trypsin; liquid chromatography electrospray ionization time-of-flight mass spectrometry (LC–ESI-TOF-MS) separates the resulting peptides and determines their accurate masses. Instrument software is used to match these peptides to the sequences of known egg white proteins, obtained from an online source. Students then use online protein database search software to match the peptides to the protein and score the results.



KEYWORDS: *Upper-Division Undergraduate, Biochemistry, Laboratory Instruction, Hands-On Learning/Manipulatives, Internet/Web-Based Learning, Proteins/Peptides, Mass Spectrometry, Bioanalytical Chemistry*

INTRODUCTION

Use of advanced mass spectrometry techniques in the undergraduate setting has burgeoned in the past decade, in particular using electrospray ionization time-of-flight mass spectrometry (ESI-TOF-MS) or matrix-assisted laser desorption ionization mass spectrometry (MALDI-MS) to determine the accurate mass of a pure protein.[1−6] A few experiments in this *Journal*[6−8] and another biochemistry education journal[9−13] have used advanced MS techniques to determine the identity of a pure protein using peptide digest data and database searching, but none have started with a mixture of proteins as is done here. Peptide mass fingerprinting is an analytical technique for protein identification. First, the protein is digested by a protease into smaller peptides whose accurate masses are determined using MALDI-MS or ESI-TOF-MS. The peptide masses are then compared to a database containing known protein sequences and the resulting matches are statistically analyzed. Software can translate genomic data into a protein sequence, "digest" the protein sequence into peptides (given a protease), and calculate the masses of the peptides produced.[14]

In this experiment, students use trypsin to digest the proteins in egg white, determine the accurate mass of the resulting peptides using liquid chromatography ESI-TOF-MS (LC–ESI-TOF-MS), and use instrument software to match the peptide masses to peptide sequences in individual proteins in egg white. Students then submit a collection of peptides associated with one protein to an external database search engine to determine the "hits" from the database, as is done in a typical peptide fingerprinting experiment for a pure protein (Figure 1).

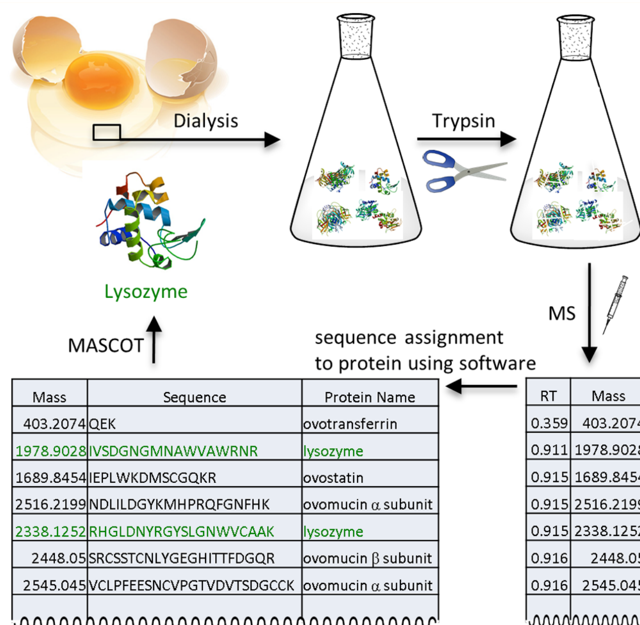


Figure 1. Workflow for peptide mass fingerprinting experiment.

Our upper-division undergraduate Biochemistry I laboratory course focuses primarily on the techniques of protein purification and characterization. Students initiate the multiweek process of purifying and characterizing lysozyme from hen egg white by dialyzing egg white. This dialyzed egg white is the source of the protein mixture used in this experiment.

The goals for this experiment are

Received: July 30, 2015
Revised: January 15, 2016

ACS Publications © XXXX American Chemical Society and Division of Chemical Education, Inc.

DOI: 10.1021/acs.jchemed.5b00625
J. Chem. Educ. XXXX, XXX, XXX−XXX

Table 1. Mass Spectral Data for Riboflavin Binding Protein

Riboflavin Binding Protein Peptides					
RT (min)	Mass (amu)	Volume	Sequence	Sequence Location	Hits
0.917	2712.1166	3538	HLLSESSEESSSMSSSEEHACQKK	A(182−205)	12
0.918	1276.5094	4922	DESGENHCKSK	A(127−137)	3
0.919	2430.105	2491	VSESSCLCLQMNKKDMVAIK	A(162−181)	13
0.922	1350.5563	2837	FEALQQEEGEE	A(209−219)	1
0.925	2814.1171	5051	HLLSESSEESSSMSSSEEHACQKK	A(182−205)	4
0.926	4065.5968	4027	DESGENHCKSKCVPYSEMYANGTDMCQSMWGESFK	A(127−161)	18
0.929	2680.119	2871	HLLSESSEESSSMSSSEEHACQKK	A(182−205)	9
1.664	871.3323	7088	SCEDFTK	A(63−69)	2
1.762	1656.7591	5502	VSNSYWNRCGQLSK	A(49−62)	1
11.283	1499.6351	59179	VSESSCLCLQMNK	A(162−174)	2
11.725	1594.7075	26054	IECFYRCSPHAAR	A(71−83)	13
11.92	2356.0826	21071	VSESSCLCLQMNKKDMVAIK	A(162−181)	10
12.027	1442.6141	9363	VSESSCLCLQMNK	A(162−174)	2
12.037	1744.7351	22360	VSNSYWNRCGQLSK	A(49−62)	4
12.072	1458.6062	13337	VSESSCLCLQMNK	A(162−174)	2
12.246	1663.7859	6853	LLKFEALQQEEGEE	A(206−219)	1
12.252	1615.7434	13155	CGQLSKSCEDFTKK	A(57−70)	8
12.288	1695.7831	3379	KIECFYRCSPHAAR	A(70−83)	3
12.489	1484.6343	76089	VSESSCLCLQMNK	A(162−174)	4
12.603	846.4572	7771	KDMVAIK	A(175−181)	3
12.744	1026.4421	57421	VSNSYWNR	A(49−56)	3
12.977	914.3399	8447	SCEDFTK	A(63−69)	1
13.808	783.3511	16027	CSPHAAR	A(77−83)	1
13.883	1394.5803	8235	EQYGCLEGDTHK	A(1−12)	1
14.596	1573.6449	36769	CGQLSKSCEDFTK	A(57−69)	1
14.596	1611.7472	5089	VSESSCLCLQMNKK	A(162−175)	3
14.81	1728.7604	18487	VSNSYWNRCGQLSK	A(49−62)	9
15.029	1918.9432	9707	KLLKFEALQQEEGEE	A(205−219)	4
15.245	3075.2573	6226	DDSICAHNWLTDWERDESGENHCKSK	A(112−137)	2
15.747	948.4791	4937	KDMVAIK	A(175−181)	1

1. Students enzymatically digest a protein mixture and understand the mechanisms of denaturation and digestion used.
2. Students learn that LC−ESI-TOF-MS is a powerful tool for analyzing peptides and proteins, and understand its role in proteomics.
3. Students use the National Center for Biotechnology Information database (NCBI)[15] to find protein sequences, use instrument software to "digest" proteins, and match those peptides to the sequences of proteins in the experimental mixture.
4. Students use online database peptide matching software to match the peptides from the experimental data to proteins in the database. They learn how manipulating the search parameters, such as amino acid modifications, can produce very different search scores.

The achievement of the four goals is assessed by prelab assignments and laboratory reports.

EXPERIMENTAL OVERVIEW

The experiment has two parts: (1) denaturation and digestion of the egg white proteins, and (2) analysis of the LC−ESI-MS-TOF data. The experiment is completed in two laboratory periods: one for the digestion (1.5 h, prep time plus an overnight digestion), another for data analysis (1−2 h). Or, using one lab period, students can prepare the sample and receive instruction on data analysis, and subsequently analyze data outside of the laboratory period. In our lab, students prepare the sample and generate data in pairs, then individually analyze their results.

Egg White Protein Digestion

Previously prepared dialyzed egg white (30 μL; dialysis buffer, 50 mM Tris hydroxymethyl aminomethane and 50 mM NaCl at pH = 8.2; dialysis tubing, Spectra/Por SP1 8K 40 mm) is transferred to a 1.5 mL microcentrifuge tube. In a chemical fume hood, denaturing solution (125 μL; 6 M guanidine HCl, 50 mM Tris, 2 mM 2-mercaptoethanol) is added. Sample is mixed using the vortex mixer and then placed in a 90 °C water bath for 25 min. Sample is cooled to room temperature and then briefly spun down in a microcentrifuge. Tris/$CaCl_2$ solution (875 μL; 50 mM Tris, 1 mM $CaCl_2$) is added and the sample is mixed. Trypsin solution (300 μL; 0.5 mg/mL trypsin in 50 mM Tris buffer) is added, sample is mixed, and then placed in a 37 °C water bath overnight. The next day, sample is briefly spun down in a microcentrifuge and 0.75 mL is transferred to a 1.5 mL microcentrifuge tube containing HPLC grade methanol (0.75 mL) and acetic acid (6 μL). The sample is mixed and transferred to a glass autosampler vial for LC−ESI-TOF-MS analysis.

LC−MS Parameters

LC−MS: Agilent 1200 series LC interfaced with an Agilent MS TOF, Model # G6224A. For the separation, an Agilent ZORBAX Eclipse XDB-C18 column (1.8 μm, 4.6 mm i.d. × 50 mm, thermostated at 40 °C) is used. Solvent A: 95% H_2O, 5% acetonitrile, 0.10% formic acid. Solvent B: 5% H_2O, 95% acetonitrile, 0.10% formic acid. LC flow, 0.5 mL/min; injection

DOI: 10.1021/acs.jchemed.5b00625
J. Chem. Educ. XXXX, XXX, XXX−XXX

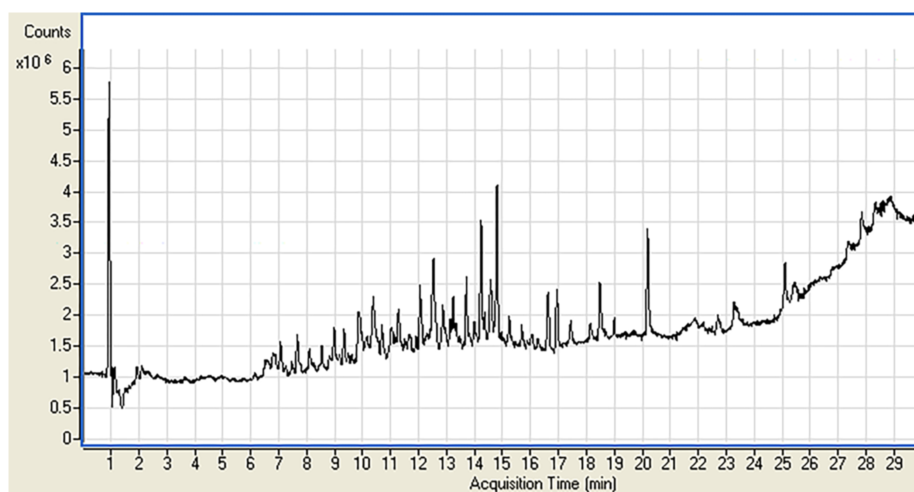


Figure 2. Chromatogram of egg white peptides, student data.

MASCOT Peptide Mass Fingerprint

Your name: Lisa Alty
Email: altyl@wlu.edu
Search title: riboflavin binding protein?
Database(s): SwissProt, NCBInr, contaminants, cRAP
Enzyme: Trypsin
Allow up to 2 missed cleavages
Taxonomy: . . . . bony vertebrates
Fixed modifications: --- none selected ---
Acetyl (K), Acetyl (N-term), Acetyl (Protein N-term), Amidated (C-term), Amidated (Protein C-term), Ammonia-loss (N-term C), Biotin (K), Biotin (N-term), Carbamidomethyl (C), Carboxymethyl (C), Cation:Na (C-term)
Display all modifications
Variable modifications: Carbamyl (N-term), Carbamyl (K), Deamidated (NQ), Oxidation (M)
Protein mass: kDa
Peptide tol. ± 10 ppm
Mass values: MH+ M_r M-H-
Monoisotopic Average
Data file
Query:
783.351100
846.457200
871.332300
914.339900
948.479100
1026.442100
Decoy
Report top 10 hits
Start Search ... Reset Form

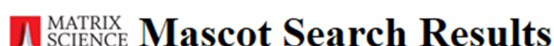
Figure 3. MASCOT screen with search parameters for riboflavin binding protein peptides. Image reproduced with permission from Matrix Science.

volume, 5 μL. Chromatographic conditions: initial solvent ratio B is 2%, held for 7 min, programmed to increase to 15% at 7 min, 45% at 20 min, 80% at 25 min; stop time is 30 min. The MS ion source is Multimode, detector mode is positive ion, scan parameters are 100−3000 amu. Complete details of the LC−ESI-TOF-MS run parameters are provided in the Supporting Information.

MS Data Analysis

The MS data process method created for this analysis contains sequence data for several of the proteins in egg white. Students use the instrument software (Agilent MassHunter BioConfirm)[16] to add two more proteins to the process method. Students extract the data from the MS run, obtaining an Excel spreadsheet with many columns, including RT, Mass, Volume, Sequence, Protein Name, Sequence Location, Hits, Predicted Modifications, among others (see Table 1; complete data is in the Supporting Information). Students sort the data by Protein Name, which groups peptides from the same protein. Then, students select one protein's set of peptides and use the external database (we use MASCOT,[17] but others are available[14,18]) to search for a match. They perform a number of searches, varying

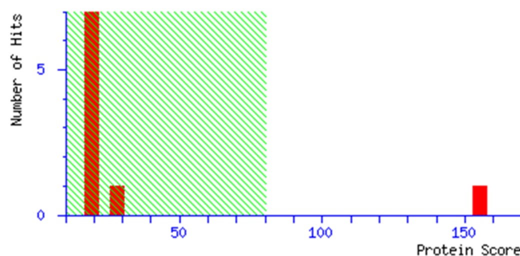
DOI: 10.1021/acs.jchemed.5b00625
J. Chem. Educ. XXXX, XXX, XXX−XXX




```
User            : Lisa Alty
Email           : altyl@wlu.edu
Search title    : riboflavin binding protein?
Database 1      : SwissProt 2015_04x (548208 sequences; 195282524 residues)
Taxonomy 1      : bony vertebrates (83761 sequences)
Database 2      : NCBInr 20150408 (64057457 sequences; 22884858569 residues)
Taxonomy 2      : bony vertebrates (5637050 sequences)
Timestamp       : 20 Apr 2015 at 18:35:11 GMT
Top Score       : 155 for 2::gi|352173, protein,riboflavin binding [Gallus gallus]
```

Mascot Score Histogram

Protein score is -10*Log(P), where P is the probability that the observed match is a random event.
Protein scores greater than 80 are significant (p<0.05).



Concise Protein Summary Report

Format As | Concise Protein Summary | Help
Significance threshold p< 0.05 | Max. number of hits 10
Preferred taxonomy | All entries

Re-Search All | Search Unmatched

```
1.   2::gi|352173    Mass: 25041   Score: 155   Expect: 1.8e-09   Matches: 26
     protein,riboflavin binding [Gallus gallus]
     2::gi|353188    Mass: 23752   Score: 116   Expect: 1.4e-05   Matches: 23
     protein,riboflavin binding [Gallus gallus]
     1::RBP_CHICK    Mass: 27193   Score: 108   Expect: 9.1e-05   Matches: 23
     Riboflavin-binding protein OS=Gallus gallus PE=1 SV=2
     2::gi|132070    Mass: 27193   Score: 108   Expect: 9.1e-05   Matches: 23
     RecName: Full=Riboflavin-binding protein; Short=RBP; Contains: RecName: Full=
     2::gi|45382617    Mass: 27207   Score: 108   Expect: 9.1e-05   Matches: 23
```

Figure 4. MASCOT search results for riboflavin binding protein peptides, showing only matches with scores >100. MASCOT reports scores as $-10*\text{Log}(P)$, where P is the probability that the observed match is a random event. A probability of 10^{-20} thus becomes a score of 200. Note that a score above 80 is considered significant ($p < 0.05$). Image reproduced with permission from Matrix Science.

the modifications and other search parameters in a systematic way, noting the match score each time.

This experiment has been performed for the last five years by 105 junior and senior biochemistry students. Full details are available in the Supporting Information.

HAZARDS

Appropriate personal protection (safety glasses, gloves, lab coat) should be worn at all times. All reactions should be carried out in a fume hood. Raw eggs may contain *Salmonella*; therefore, one should wash hands with soap after handling raw egg. The aqueous buffers present no significant hazards. Guanidine or 2-mercaptoethanol ingestion may cause gastrointestinal irritation, nausea, vomiting and diarrhea. 2-Mercaptoethanol may cause severe eye and skin irritation and may be fatal if absorbed through the skin, causing muscle paralysis and respiratory failure. Methanol is a flammable liquid and vapor that is irritating to the eyes and skin; it may be fatal or cause blindness, if swallowed. Acetic acid is a flammable liquid and causes severe skin burns and eye damage. Hazards for chemicals used in eluents for LC–MS analysis are detailed in the Supporting Information; they are not handled by students.

RESULTS AND DISCUSSION

A representative student chromatogram of the egg white peptides is shown in Figure 2. Many of the peaks represent multiple peptides eluting together; however, because the masses of these peptides are different, the data for individual peptides can be extracted without baseline separation of individual peptides being achieved. Table 1 shows mass spectral data output for one protein and illustrates the point that peptides with similar retention times have different masses.

The data for riboflavin binding protein, a 219 amino acid protein, will be used to illustrate the data processing workflow. Table 1 shows the MS data for the sequences assigned to riboflavin binding protein by the instrument software. These

DOI: 10.1021/acs.jchemed.5b00625
J. Chem. Educ. XXXX, XXX, XXX−XXX

peptides cover 155 of the 219 amino acids in the sequence, or 71% coverage, with no data for amino acids 13−48 and 84−111. Students select and copy all of the masses in the Mass column, and paste them into the MASCOT Query box, setting the parameters as shown in Figure 3. The variable modifications added to this search were those suggested by the MassHunter BioConfirm software (data in Supporting Information). The highest scoring results of the MASCOT search are shown in Figure 4.

Once students see how well MASCOT works when supplied with optimal data, we ask students to run the search several more times using different combinations of variable modifications. This process reveals that knowing the likely modifications is vital in finding a good match. Table 2 shows the results for riboflavin binding protein peptides, given different combinations of the variable modifications used in the previous search.

Table 2. MASCOT Scores for Riboflavin Binding Protein with Different Combinations of Variable Modifications

Variable Modifications Added to Search	Correct Hits with Genus Species	Highest MASCOT Score	Number of Matches to Riboflavin Binding Protein, Different Genus Species
None	0		0
Carbamyl (K)	6	29	3
Carbamyl (N-term)	7	56	30
Deamidated (NQ)	1	26	0
Oxidation (M)	0		0
Carbamyl (K), Carbamyl (N-term)	8	66	32
Carbamyl (K), Carbamyl (N-term), Deamidated (NQ)	8	144	33
Carbamyl (K), Carbamyl (N-term), Deamidated (NQ), Oxidation (M)	7	155	28

The quality of student data, student scores on the prelab assignment (AVG = 9.7/10, SD = 0.6, n = 24) and laboratory report (AVG = 92/100, SD = 9.3, n = 24) demonstrate that the goals of the experiment were met. The prelab assignment is designed to educate students about the technique of peptide and protein MS, including determining accurate mass of a larger protein. All students successfully digest the egg white proteins. Students understand the purpose for the methods used for denaturation as evidenced by correct answers to questions on lab reports and they see the specificity of trypsin in cleaving peptide bonds in the data. Laboratory reports also show that students successfully use the instrument software to add proteins to the analysis method, extract the MS data to get the masses and sequences of proteins, and use an external database search program to convert a list of peptide masses into a scored match to the protein's identity.

CONCLUSIONS

A peptide mass fingerprinting experiment teaches students how to digest a mixture of proteins, use LC−ESI-TOF-MS to separate and determine the masses of the peptides generated in the digestion, and use software to assign the peptides to an individual protein, as well as peptide database mining. In addition, students learn that database mining is most fruitful when the possible post-translational modifications, or modifications due to the experimental conditions, are known. Finally, students comprehend the wealth of data produced using an LC−ESI-TOF-MS and its usefulness in a proteomics experiment.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available on the ACS Publications website at DOI: 10.1021/acs.jchemed.5b00625.

Instructions for students, additional notes for the instructor, and representative student data (PDF, DOCX)

2014 EWP peptide report extract (XLS)

AUTHOR INFORMATION

Corresponding Author

*E-mail: altyl@wlu.edu.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

We gratefully acknowledge funding support for the purchase of the electrospray ionization time-of-flight mass spectrometer from the National Science Foundation (NSF-MRI Grant CHE-0922599). Additional financial support was provided by the Department of Chemistry and Biochemistry at Washington and Lee University. Thank you also to the 2014 Biochemistry I laboratory students who generated this data.

REFERENCES

(1) Stynes, H. C.; Layo, A.; Smith, R. W. LC−MS of Metmyoglobin at pH = 2. Separation and Characterization of Apomyoglobin and Heme by ESI−MS and UV−Vis. *J. Chem. Educ.* **2004**, *81* (2), 266−269.

(2) Weinecke, A.; Ryzhov, V. Fundamentals of Biomolecule Analysis by Electrospray Ionization Mass Spectrometry. An Instrumental Analysis Laboratory Experiment. *J. Chem. Educ.* **2005**, *82* (1), 99−102.

(3) Dopke, N. C.; Lovett, T. N. Illustrating the Concepts of Isotopes and Mass Spectrometry in Introductory Courses: A MALDI-TOF Mass Spectrometry Laboratory Experiment. *J. Chem. Educ.* **2007**, *84* (12), 1968−1970.

(4) Arnquist, I. J.; Beussman, D. J. Incorporating Biological Mass Spectrometry Into Undergraduate Teaching Labs, Part 1: Identifying Proteins Based on Molecular Mass. *J. Chem. Educ.* **2007**, *84* (12), 1971−1973.

(5) Arnquist, I. J.; Beussman, D. J. Incorporating Biological Mass Spectrometry into Undergraduate Teaching Labs, Part 2: Peptide Identification via Molecular Mass Determination. *J. Chem. Educ.* **2009**, *86* (3), 382−384.

(6) Rosado, D. A., Jr.; Masterson, T. S.; Masterson, D. S. Using the Mini-Session Course Format To Train Students in the Practical Aspects of Modern Mass Spectrometry. *J. Chem. Educ.* **2011**, *88* (2), 178−183.

(7) Counterman, A. E.; Thompson, M. S.; Clemmer, D. E. Identifying a Protein by MALDI-TOF Mass Spectrometry: An Experiment for the Undergraduate Laboratory. *J. Chem. Educ.* **2003**, *80* (2), 177−180.

(8) Cohen, A.; Reimann, C. T.; Mie, A.; Nilsson, C. Introduction to Biological Mass Spectrometry: Determining Identity and Species of Origin of Two Proteins. *J. Chem. Educ.* **2005**, *82* (8), 1215−1218.

(9) Eberhardt, E. S.; Hansen, J.; Riservato, L.; Cole, M.; Smaglo, B.; Szaniawski, P. Preparing undergraduates to participate in the post-genome era: A capstone laboratory experience in proteomics. *Biochem. Mol. Biol. Educ.* **2003**, *31* (6), 402−409.

(10) Jez, J. M.; Schachtman, D. P. R.; Berg, H.; Taylor, C. G.; Chen, S.; Hicks, L. M.; Jaworski, J. G.; Smith, T. J.; Nielsen, E.; Pikaard, C. S. Developing a new interdisciplinary lab course for undergraduate and graduate students: Plant cells and proteins. *Biochem. Mol. Biol. Educ.* **2007**, *35* (6), 410−415.

DOI: 10.1021/acs.jchemed.5b00625
J. Chem. Educ. XXXX, XXX, XXX−XXX

(11) Albright, J. C.; Dassenko, D. J.; Mohamed, E. A.; Beussman, D. J. Identifying gel-separated proteins using in-gel digestion, mass spectrometry, and database searching. *Biochem. Mol. Biol. Educ.* **2009**, *37* (1), 49−55.

(12) Short, M.; Short, A.; Vankempen, R.; Seymour, M.; Burnatowska-Hledin, M. Using HPLC-mass spectrometry to teach proteomics concepts with problem-based techniques. *Biochem. Mol. Biol. Educ.* **2010**, *38* (4), 242−246.

(13) Wilson, K. A.; Tan-Wilson, A. Seed storage proteins as a system for teaching protein identification by mass spectrometry in biochemistry laboratory. *Biochem. Mol. Biol. Educ.* **2013**, *41* (2), 79−86.

(14) Swiss Institute of Bioinformatics. ExPASy Bioinformatics Resource Portal. http://www.expasy.org/proteomics/ (accessed February 4, 2016).

(15) National Center for Biotechnology Information. Protein. http://www.ncbi.nlm.nih.gov/protein (accessed July 9, 2015).

(16) *MassHunter BioConfirm Software*, v.B.04.00, Agilent Technologies.

(17) Matrix Science. MASCOT Peptide Mass Fingerprint. http://www.matrixscience.com/cgi/search_form.pl?FORMVER=2&SEARCH=PMF (accessed July 9, 2015).

(18) *ProteinProspector*, v 5.14.1., University of California, San Francisco, CA. http://prospector.ucsf.edu/prospector/cgi-bin/msform.cgi?form=msdigest (accessed July 9, 2015).

DOI: 10.1021/acs.jchemed.5b00625
J. Chem. Educ. XXXX, XXX, XXX−XXX