Section 4

Sequencing peptides from tandem MS (MS-MS) data

Learning Objectives

At the end of this assignment, you should be able to:

1. Describe the operations of a triple quad mass spectrometer for tandem MS experiments.
2. Identify b and y ions and use them to determine the sequence of a short peptide by de novo sequencing.

Section 4A. Tandem MS

While enzymes, such as trypsin, can be used to cleave proteins and peptides at specific amino acid linkages, we can also fragment peptides inside of a mass spectrometer to obtain additional information. These types of experiments are called tandem MS or MS-MS experiments. These experiments are particularly helpful for "shotgun" proteomics or bottom-up proteomics. In these experiments, protein mixtures are first digested with enzymes (such as trypsin), then separated by one or more chromatography steps, and then electrosprayed into a mass spectrometer. A mass analyzer is then used to select a precursor ion with a specific *m*/*z* value for fragmentation. Fragmentation requires that some energy be added to the system. The most common method of fragmentation in MS-MS experiments is collision-induced dissociation (CID). In CID, the precursor ion is accelerated into an interaction cell that contains a collision gas, such as helium or nitrogen. When the precursor ion collides with the collision gas, the ion can fragment into two fragments, an ion and a neutral. The fragment ions are then analyzed to produce the MS/MS spectrum.

To better understand the operation of a commonly used triple quadruple or "triple quad" MS/MS, watch an animation, then answer the questions below.

Video Questions

1. What is the purpose of the skimmer in the instrument?

2. The animation uses color to indicate difference *m*/*z* value ions. Which "color" of ion is selected as the precursor ion?

3. Although this style of instrument is commonly called a "triple quad," the collision cell is not actually a quadrupole. What is it?

4. The last quadrupole selected fragment ions to be sent to the detector. Neutrals also pass through this quadrupole. Why don't they produce a signal at the detector?

MS/MS experiments are useful because the fragment m/z values give information about the analyte's molecular structure. Low energy CID often produces small neutral losses, such as H_2O, CH_3OH, CO, CO_2, NH_3, and CN. Higher energy collisions can lead to retrosynthetic reactions, in which characteristic bonds in the precursor ion are broken. For example, CID of peptides often results in cleavage of peptide bonds along the backbone.

Reading Question

1. Complete the table below to summarize the expected mass differences from common neutral losses. Round expected masses to the nearest amu.

Table 1. Common neutral losses produced by CID of peptide ions.

Neutral Loss	Mass (amu)
NH_3	
H_2O	
CN	
CO	
CH_3OH	
CO_2	

The identity of the precursor ion can be scanned through the mass range so that a mass spectrum of each precursor ion's corresponding fragments is obtained. These experiments produced large quantities of data extremely rapidly. Interpreting these large data sets usually involves specialized software programs that identify peptides from their fragments and then identify proteins from their peptides; however, for simple mixtures, the data may be interpreted manually since peptide fragmentation in tandem MS experiments is well-characterized.

Section 4B. CID of Peptides and *De Novo Sequencing*

Peptide fragments produced in tandem MS experiments are named using a letter-number scheme that identifies which bond was broken and which side of the peptide (the N-terminus or the C-terminus) became the charged fragment (Figure 1). As noted above, CID typically results in cleavage of the peptide bond and therefore produces *b* and *y* ions. When the N-terminal side of the peptide is the charged fragment and the C-terminus is a neutral loss, the result is a *b* ion. When the C-terminal side is the charged fragment, the result is a *y* ion. In either case, the ions are numbered sequentially from the charged terminus, meaning that the first N-terminal residue gives the b_1 fragment, and the first C-terminal residue gives the y_1 ion. Other fragmentation methods, such as photodissociation, electron capture, and electron transfer dissociation can cleave other bonds within the peptide, giving rise to *a*, *x*, *c*, and *z* ions. In some cases, CID may also produce *a* ions, which differ from *b* ions by the absence of a carbonyl, making them 28 amu lighter than the corresponding *b* ion.

For appropriate CID conditions, it is often possible to obtain a series of *b* and *y* ions for a peptide (Figure 1). The mass differences between *b* ions (or between *y* ions) are characteristic of the amino acid residues which have been lost, and careful examination of the mass spectrum can yield the peptide sequence from the MS-MS data. For example, Figure 3 shows model data for fragmentation of the MRFA peptide. The mass difference between peaks from right to left are 71, 147, 156 amu, corresponding in mass to sequential losses of A, F, and R amino acid residues (Table 2), indicating that these amino acids form a series in the peptide backbone of the analyte.

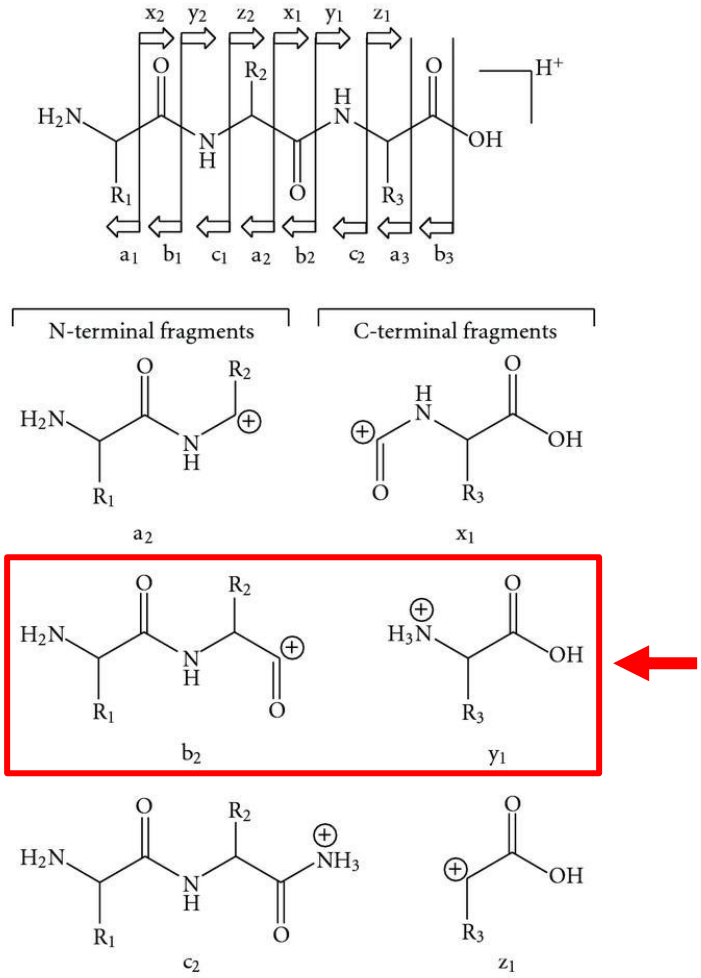

Figure 1. Fragment ions are named based on the bond cleaved and the location of the charge on the resulting fragments. CID most commonly produces *b* and *y* ions, which result from cleavage of the peptide bond with the charged fragment occurring on the N- or C-terminal fragment, respectively. Figure is adapted with permission from S. Banerjee and S. Mazumdar, *Int. J. Anal. Chem.*, **2012**, *2012*, 282574 under a Creative Commons Attribution License.

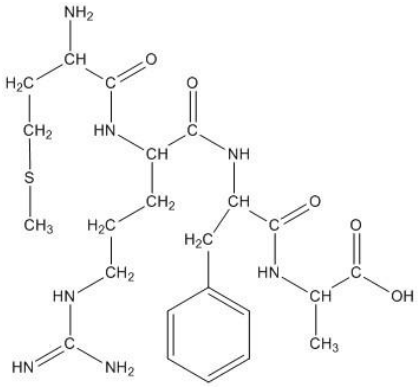

Figure 2. Structure of the tetrapeptide MRFA.

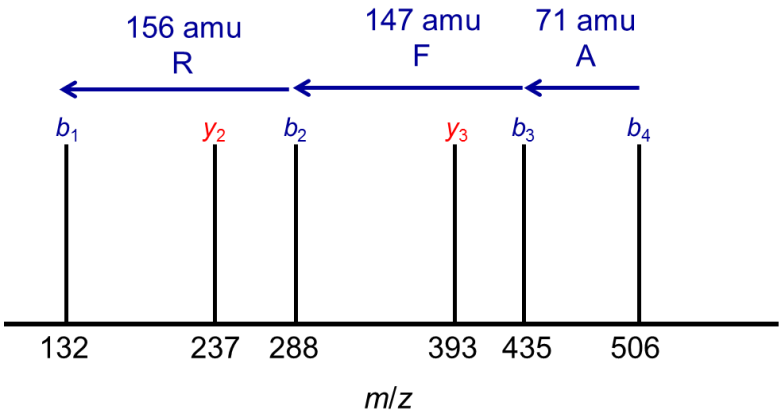

Figure 3. A model mass spectrum for MS-MS of the peptide, MRFA, which has a precursor *m*/*z* of 524.27.

Table 2. Molecular weight information for all 20 naturally occurring amino acids.

Amino Acid	Single-Letter Code	Residue MW (amu)	Amino Acid MW (amu)
glycine	G	57.02	75.03
alanine	A	71.04	89.05
serine	S	87.03	105.04
proline	P	97.05	115.06
valine	V	99.07	117.08
threonine	T	101.05	119.06
cysteine	C	103.01	121.02
isoleucine	I	113.08	131.09
leucine	L	113.08	131.09
asparagine	N	114.04	132.05
aspartic acid	D	115.03	133.04
glutamine	Q	128.06	146.07
lysine	K	128.09	146.11
glutamic acid	E	129.04	147.05
methionine	M	131.04	149.05
histidine	H	137.06	155.07
phenylalanine	F	147.07	165.08
arginine	R	156.10	174.11
tyrosine	Y	163.06	181.07
tryptophan	W	186.08	204.09

Reading Questions

2. Sketch the a_2, b_2, and y_2 ions for the tetrapeptide MRFA shown in Figure 2. What are the expected masses of these fragments?

3. What is the mass difference between the y_3 and the y_2 ion in Figure 3?

4. What amino acid residue corresponds to this mass difference? Does this make sense given the sequences of these two ions?

In practice, a complete *b* and *y* ion series may not be obtained, but it is often possible to deduce the peptide sequence from MS-MS data without referring to external databases or genomic data. This method of determining peptide structure is called *de novo* sequencing since the sequence is determined without reference to outside data. Because *de novo* sequencing does not rely on an external database, peptides can be identified even if they have unexpected post-translational modifications or arise from organisms with unsequenced genomes. Several research groups have developed algorithms to automate *de novo* sequencing from MS-MS data. These algorithms are designed to account for missing *b* and *y* ions, identify post-translational modifications, and address other challenges, including identification of oxidation and other chemical changes to proteins and peptides. Database searching is an alternative to *de novo* sequencing for longer peptides and proteins. Similarly to database searching for peptide mass fingerprinting, MS-MS database searching relies on genomic data to predict the expected spectra for MS-MS analysis. The experimental spectra are then matched to the predicted spectra to make peptide and protein assignments.

Discussion Questions

1. Consider the data in Table 2. By what value do the residue MW and the amino acid MW differ?

2. If needed, review the module introduction to peptides and proteins. Why is the MW of an amino acid residue in a peptide chain different from the mass of the full amino acid? Sketch a reaction to support your answer.

3. Are there any amino acids that could not be distinguished from one another using a mass analyzer with unit resolution (e.g., a quadrupole ion trap)?

4. Leu-enkephalin is a pentapeptide with the sequence YGGFL that is involved in neurotransmission. Tabulate the amino acid sequence and the expected monoisotopic MWs, and fragments of leu-enkephalin. Work in groups, dividing the required calculations among members.

Table 3. Expected precursor and fragment ions for CID MS-MS of leu-enkephalin.

Ion	Peptide Sequence	Expected *m/z*
$[M+H]^+$		
b_1		
b_2		
b_3		
b_4		
b_5		
y_1		
y_2		
y_3		
y_4		

5. Using your completed Table 3 above, identify as many of the peaks in the leu-enkephalin MS-MS spectrum shown in Figure 4 below as possible, remembering that CID can result in the small neutral losses that you noted in Table 1 in addition to retrosynthetic fragmentation.

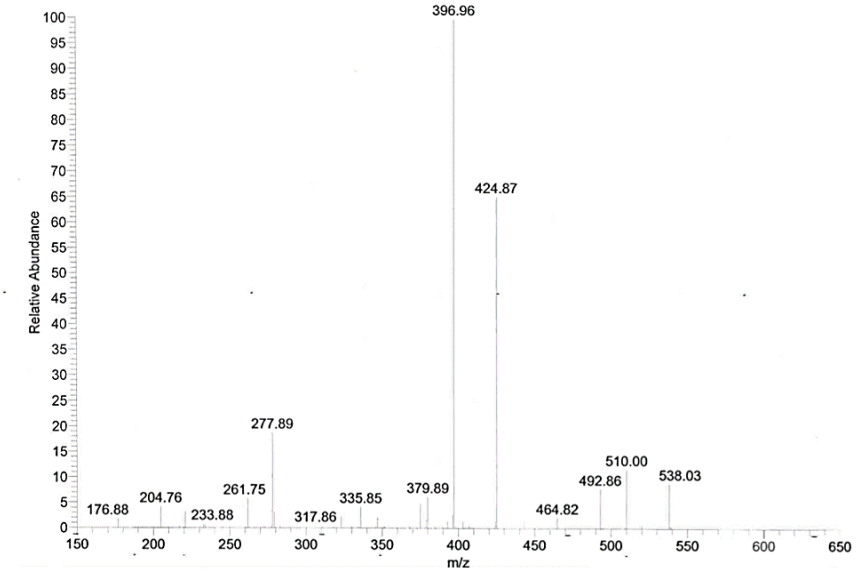

Figure 4. ESI-MS-MS spectrum of leu-enkephalin. Data obtained at Trinity College.

6. Imagine researchers take a sample of blood serum from a patient and perform a shotgun proteomics experiment with a trypsin digestion and tandem MS detection. Would you recommend a separation step in between the tryptic digest of the serum proteins and the mass spectrometric detection? Why or why not?

Challenge Question

To the best of your ability, use the MS-MS spectrum in Figure 5 below to *de novo* sequence an unknown tetrapeptide. To support your proposed sequence, make a table showing the expected mass of each fragment and the actual mass you observe for the peak.

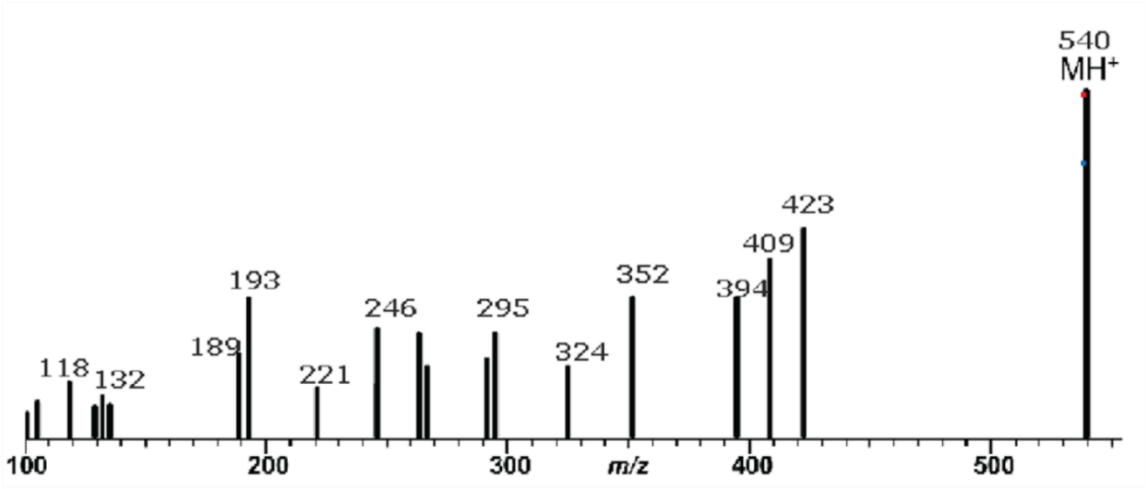

Figure 5. MS-MS mass spectrum of an unknown peptide. Data from F. Klink, *Separation Science.*

References and Resources

R. Aebersold and M. Mann, "Mass spectrometry-based proteomics," *Nature*, **2003**, *422*, 198-207.

The Broad Institute, "MRM (Multiple Reaction Monitoring)," https://www.broadinstitute.org/scientific-community/science/platforms/proteomics/mrm-multiple-reaction-monitoring

F. Klink, "MS Solutions #9: Peptide Sequencing with Electrospray LC/MS Part 1: Ion Types and Nomenclature," *Separation Science*, http://www.sepscience.com/Information/Archive/MS-Solutions/244-/MS-Solutions-9-Peptide-Sequencing-with-Electrospray-LCMS-Part-1-Ion-Types-and-Nomenclature

F. Klink, "MS Solutions #10: Peptide Sequencing with Electrospray LC/MS Part 2: Interpretation of a Simple Spectrum," *Separation Science*, http://www.sepscience.com/Information/Archive/All-Articles/246-/MS-Solutions-10-Peptide-Sequencing-with-Electrospray-LCMS-Part-2-Interpretation-of-a-Simple-Spectrum

Ion Source, "*De Novo* Peptide Sequencing Tutorial," **2012**, http://ionsource.com/tutorial/DeNovo/DeNovoTOC.htm