De Novo Peptide Sequencing Tutorial

Table of Contents

Introduction

Peptide Fragmentation Nomenclature

b and y Ions

The Rules (16 simple rules)

The Protocol

Example MS/MS Spectrum

- low mass
- mid mass
- high mass

First De Novo Exercise (Ion Trap Data)

- **Answer Page**

Second De Novo Exercise (q-TOF Data)

- **Answer page**

- Tutorial References and Additional Suggested Reading

  - De Novo Tables - AA Residue Mass, Mass Conflict, Di-peptide Mass Residue Combinations

http://ionsource.com/tutorial/DeNovo/introduction.htm

Introduction

De novo is Latin for, "over again", or "anew". A popular definition for "de novo peptide sequencing" is, peptide sequencing performed without prior knowledge of the amino acid sequence. Usually this rule is imposed by Edman degradation practitioners who perform de novo sequencing day in and day out, and perhaps feel a little bit threatened by that half million dollar mass spectrometer sitting down the hall, that can supposedly sequence peptides in a matter of seconds, and not days. Actually, any research project should be started with as much information as possible, there should never be a need to restrict your starting knowledge, unless of course you are performing a clinical trial or some other highly controlled experiment. Mass spectrometers do have the advantage when it comes to generating sequence data for peptides in low femtomole quantities. However, Edman degradation will always enjoy the advantage when pmol quantities of a peptide are available. At higher pmol quantities (2-10 pmol), Edman will often provide the exact amino acid sequence without ambiguity for a limited run of amino acids, 6-30 amino acids, usually taking 30-50 min per cycle of the sequencer. However, at lower quantities, gaps and uncertainties are often encountered, even with Edman sequencing. MS/MS enjoys sensitivity, and speed, and does not require an external standard for each amino acid or amino acid variant. MS/MS sequencing does have difficulty with isobaric or near isobaric masses, for example telling K from Q on low resolution, low mass accuracy mass spectrometers. Another advantage is that MS/MS sequencing is never stopped by a blocked amino terminus, as is the case for Edman degradation.

Edman practitioners will often blast MS/MS sequencing on it deficiencies. We do need to approach de novo sequencing with our eyes wide open to all of its challenges and also to all of it's advantages. As scientists we need to have faith in the derived de novo sequence without knowing the sequence ahead of time, I guess this is at the heart of de novo. Especially when software is involved, we need to be confident enough to point to the top ranked output sequence and say, "Yes, this is the most correct sequence!" It is appropriate to test your skills or the skills of a software package with blinded but know sequences. Throughout the tutorial we will look at some known and some blinded sequences to demonstrate some of the de novo sequencing principles and also to test your newly learned de novo sequencing skills.

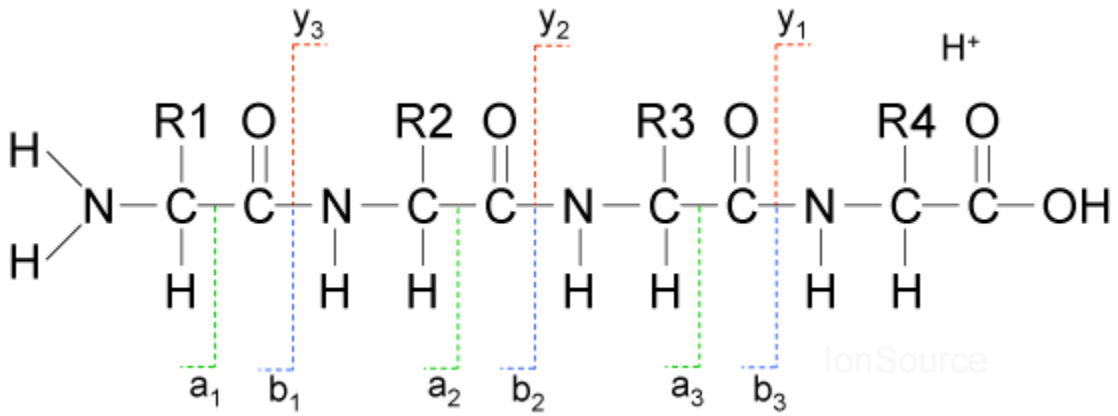
This tutorial leans heavily on a de novo sequencing course that was presented in 1992 at the University of Virginia, taught by Professor Donald F. Hunt. Dr. Hunt and his colleagues have generously taught this course for many years, educating generations of mass spectroscopists. It is impossible to calculate the enormity of the contribution that Dr. Hunt and his teaching efforts have made to countless research projects, both influencing basic, and medical, and drug research. One of the most notable early applications was the sequencing of peptides bound to MHC molecules. This was truly ground breaking work by the Hunt lab at the University of Virginia.

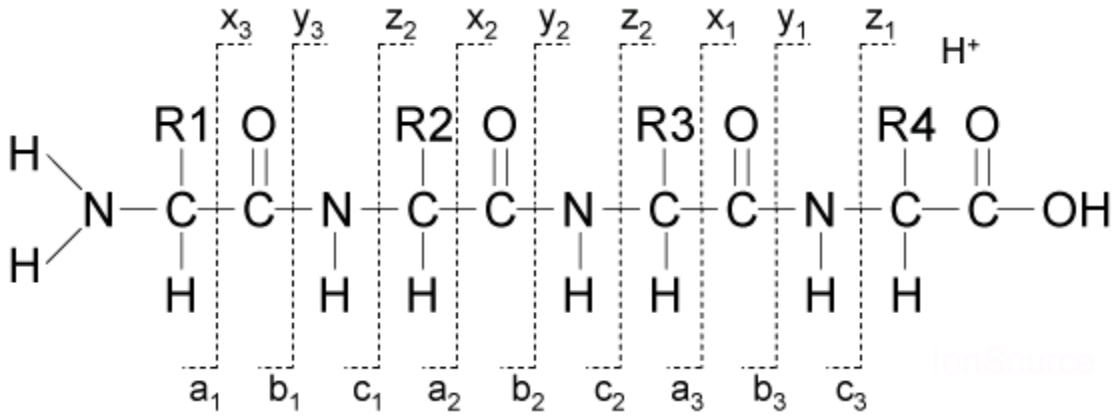
Let's Start:

A common question when one begins to talk about peptide fragmentation is, "What are **b** and **y** ions?" First we will look at the classical nomenclature, and then we will look at our first example peptide

Peptide Fragmentation Nomenclature

b, y and a ions





References:

1. Roepstorff P, Fohlman J. Proposal for a common nomenclature for sequence ions in mass spectra of peptides. Biomed Mass Spectrom. 1984 Nov;11(11):601.
2. K. Biemann In: J.A. McCloskey, Editor, Methods in Enzymology 193, Academic, San Diego (1990), pp. 886–887.
3. Biemann K. Contributions of mass spectrometry to peptide and protein structure. Biomed Environ Mass Spectrom. 1988 Oct;16(1-12):99-111.

The most common peptide fragments observed in low energy collisions are **a, b** and **y** ions, as described in the figure above. The **b** ions appear to extend from the amino terminus, sometimes called the N-terminus, and **y** ions appear to extend from the carboxyl terminus, or C-terminus. While readily observed and diagnostic for **b** ions, **a** ions occur at a lower frequency and abundance in relation to **b** ions. The **a** ions are often used as a diagnostic for **b** ions, such that **a-b** pairs are often observed in fragment spectra. The **a-b** pairs are separated by 28u, the mass for the carbonyl, C=O.

The fragment types listed above are the most common fragments observed with ion trap, triple quadrupole, and q-TOF mass spectrometers. Follow the link to see the fully annotated

fragmentation nomenclature as proposed by Biemann. An important note: an earlier nomenclature was proposed by Roepstorff and Fohlman and later modified by Biemann. The Biemann adaptation has been widely accepted.

More on b and y ions

Peptides do not fragment sequentially, that is to say, the first fragmentation event does not start at the amino terminus and proceed sequentially one residue at a time down the amino acid chain. The fragmentation events are somewhat random and definitely not sequential. In addition, some fragmentations are preferred over others as noted by the variation in the abundance of observed peaks in the spectrum below. Most of us can recognize a peptide fragment spectrum just by glancing at it. The peaks will appear to differ by the approximate mass of an amino acid residue as shown below.

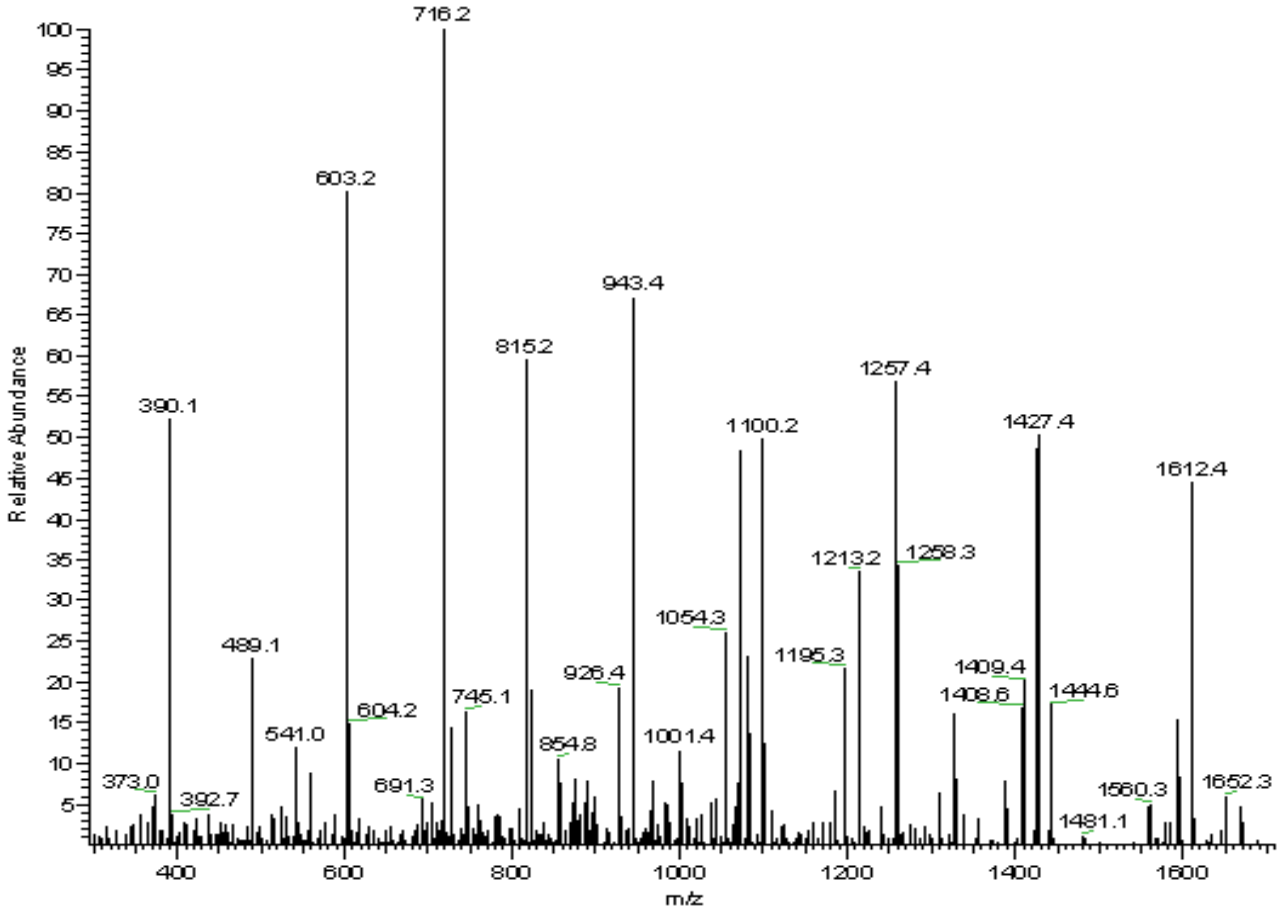


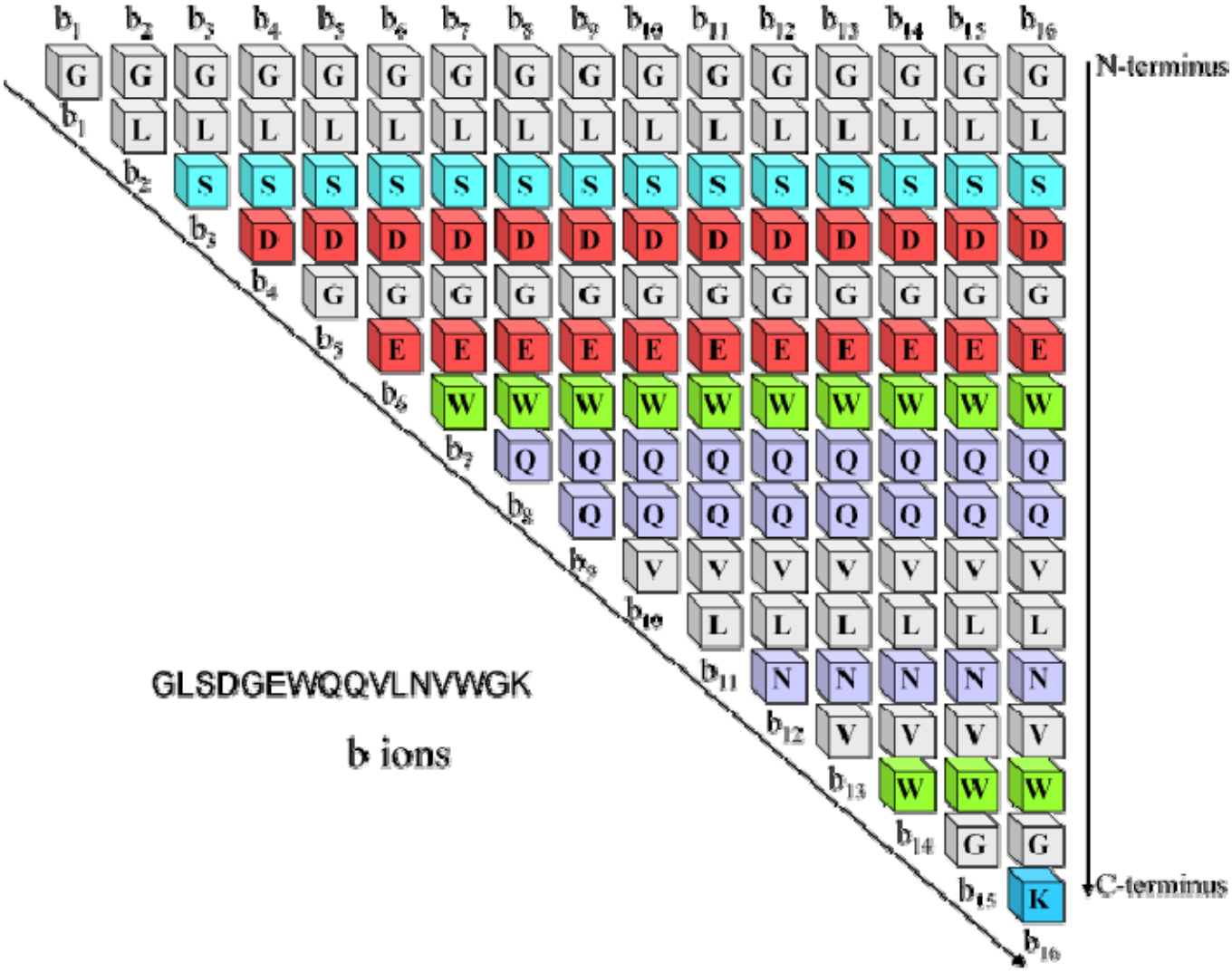
Figure 1. This is an MS/MS spectrum of the tryptic peptide GLSDGEWQQVLNVWGK. This data was collected on an ion trap mass spectrometer. This spectrum will be the subject of our first unblinded de novo sequencing example.

The mess of peaks normally observed in a fragment spectrum are a reflection of the population of fragment ions produced in the collision cell of a mass spectrometer. The sequence of the peptide is determined by the mass difference between these peaks. To complicate matters there will be **y** and **b** ions intermixed that may allow you to to establish a sequence, both forward and backward.

Those fragment peaks that appear to extend from the amino terminus are termed "**b** ions". Figure 2. below demonstrates the ladder or family of "**b** ions" that may be observed in the fragment mass

spectrum for this tryptic peptide. The **b** fragment peaks are labeled from the amino to the carboxyl terminus. The fragment containing only the amino terminal amino acid is termed **b**1. The fragment containing the first two amino terminal amino acids is termed the **b**2 ion, and so forth. The nomenclature is very simple to follow.

Figure 2.



Below is a closer look at the generic structure of the first six amino terminal b ions. You can calculate the mass of any **b** ion, basically it is the mass of the shortened peptide (M)-17 (OH) = **b** ion m/z or or more simply M-17 = **b** ion m/z. To keep it simple this is the calculation for a singly charged **b** ion.

http://ionsource.com/tutorial/DeNovo/introduction.htm

Figure 3.

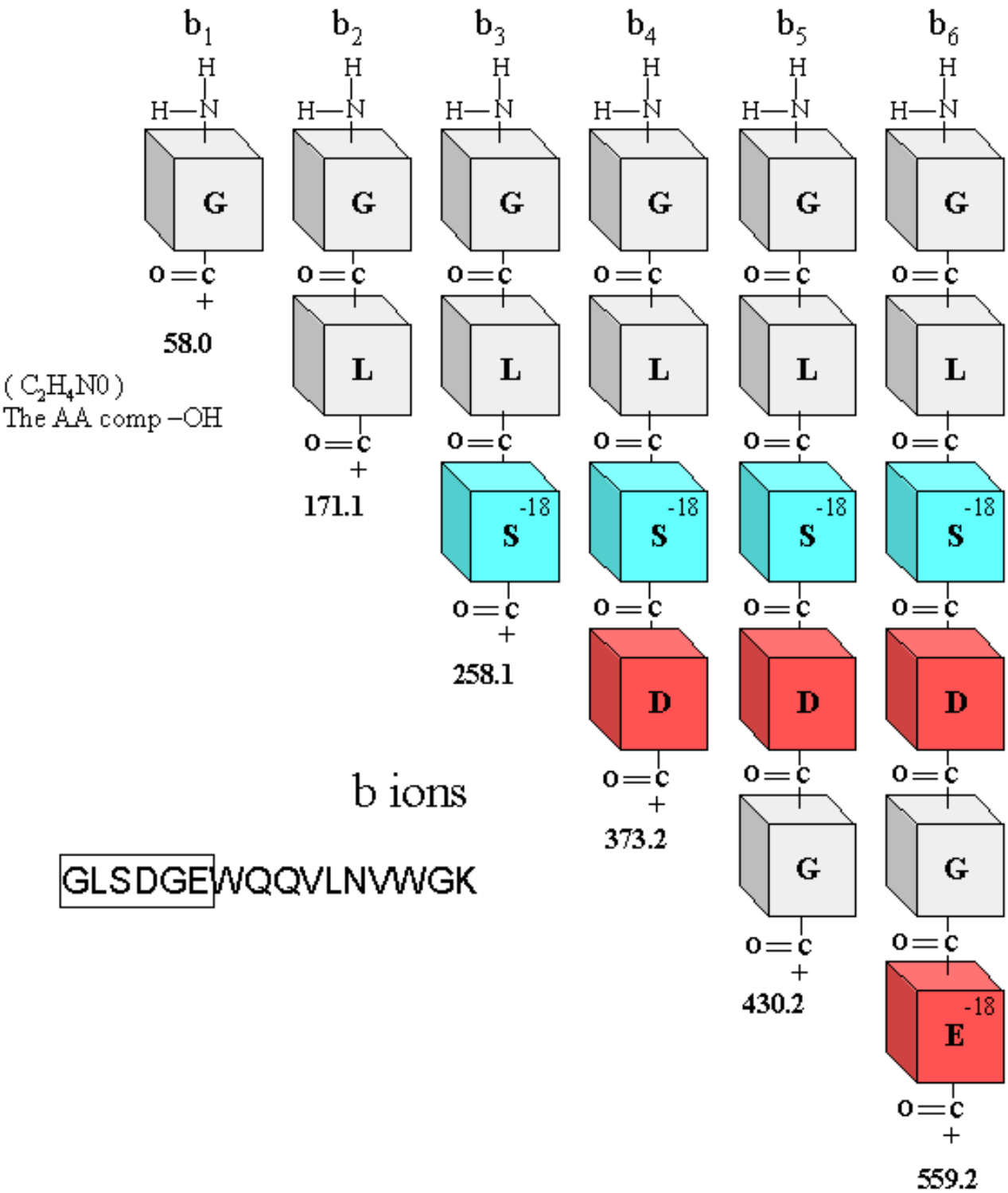
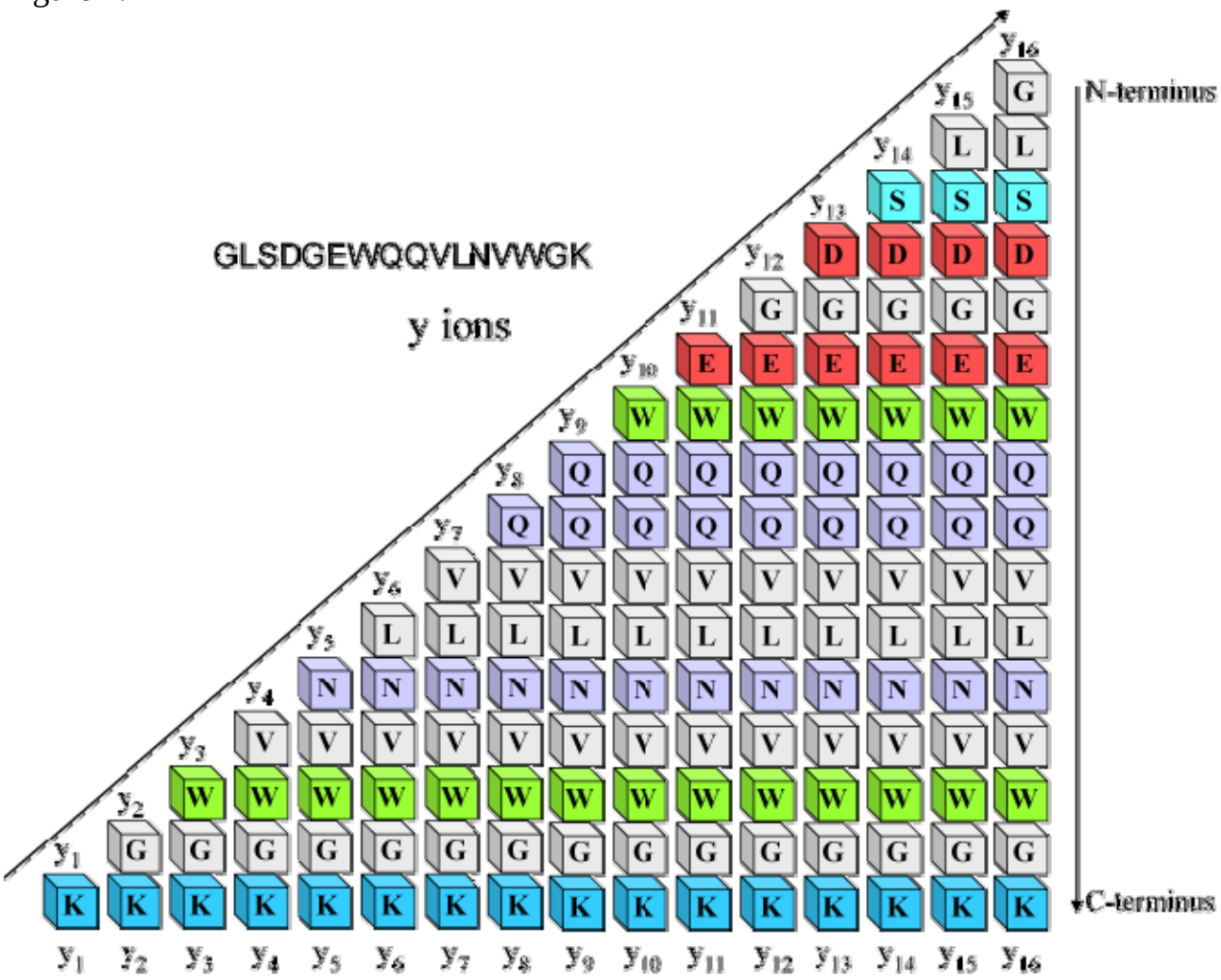


Figure 3. Shows the first six **b** ions in a little bit more detail. The **b** ion m/z value is basically the mass of the peptide minus OH, or -17u.

Similarly, groups of peptide fragment ions appear to extend from the C-terminus, these peaks are termed, "**y** ions". The **y** ion series for our example peptide GLSDGEWQQVLNVWGK is illustrated below in Figure 4.

http://ionsource.com/tutorial/DeNovo/introduction.htm

Figure 4.



http://ionsource.com/tutorial/DeNovo/introduction.htm

Below in Figure 5 the first six **y** ions are illustrated in some detail. To calculate the m/z value for the **y** ions just calculate the (M+H)1+ for the shortened peptide.

Figure 5

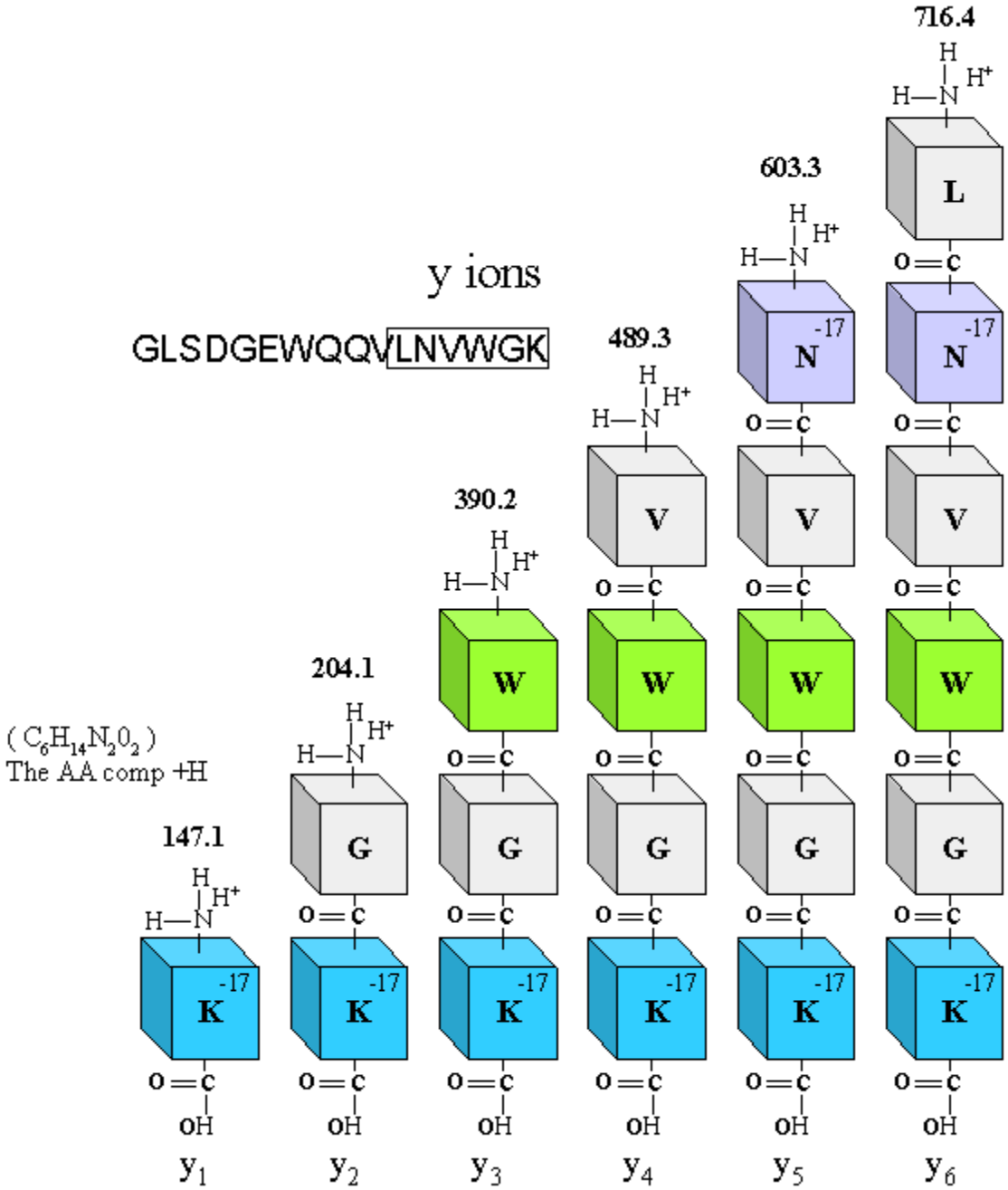


Figure 5. The first six **y** ions are illustrated. The calculated masses are shown above each **y** ion in bold numbers.

http://ionsource.com/tutorial/DeNovo/introduction.htm

y and b ion mnemonic:
To remember which are **y** ions and which are **b** ions you can remember that **b** ions are the series that extend from the amino terminus, or the front of the peptide. To us, it would make more sense if the **b** ions extended from the **b**ack or C-terminus, but just the opposite is true, **b** ions extend from the front of the peptide, the amino terminus.

The screen shot below is the output of a free on-line calculator provided by the Institute of Systems Biology. All you need to do is paste in the sequence of your peptide and it will output the expected y and b ions. The URL for this resource is
http://db.systemsbiology.net:8080/proteomicsToolkit/index.html

You can use these masses to casually match up the masses to the peaks in Figure 1, at the top of the page.

Figure 6.

Fragment Ion Calculator Results

Sequence: **GLSDGEWQQVLNVWGK**, pI: **4.37029**

Fragment Ion Table, monoisotopic masses

Seq	#	B	Y	# (+1)
G	1	58.02933	1815.90301	16
L	2	171.11340	1758.88155	15
S	3	258.14543	1645.79749	14
D	4	373.17237	1558.76546	13
G	5	430.19383	1443.73851	12
E	6	559.23642	1386.71705	11
W	7	745.31574	1257.67446	10
Q	8	873.37431	1071.59515	9
Q	9	1001.43289	943.53657	8
V	10	1100.50131	815.47799	7
L	11	1213.58537	716.40958	6
N	12	1327.62830	603.32551	5
V	13	1426.69671	489.28259	4
W	14	1612.77602	390.21417	3
G	15	1669.79749	204.13486	2
K	16	1797.89245	147.11340	1

De Novo sequencing seems petty simple right? Well, before we proceed onto our first example we should learn some of the rules and observations that scientists have previously made that will help us when we start looking at real data.

The Rules

(the observations)

Before we go though our first MS/MS example we should take a look at some of the rules that are generally applied to de novo sequencing. These rules or observations were adapted from a 1991 de novo sequencing course taught at the University of Virginia by Professor Donald F. Hunt and Dr. Jeffrey Shabanowitz. Here are a few of the rules and observations that were introduced in that course.

The Rules

Loss of Ammonia and Water

1. **y** and **b** ion fragments containing the amino acid residues R, K, Q, and N may appear to lose ammonia, -17.

2. **y** and **b** ion fragments containing the amino acid residues S, T, and E may appear to lose water, -18. In the case of glutamic acid, E must be at the N-terminus of the fragment for this observation to be made.

Spectral Intensity Rules

1. **b ion** intensity will drop when the next residue is P, G or also H, K, and R.

2. Internal cleavages can occur at P and H residues. An internal cleavage fragment is a fragment that appears to be a shortened peptide with P and or H at its amino terminus, for example the peptide EFGLPGLQNK may display the **b ions** PGLQNK, PGLQN, PGLQ, etc. These are the result of a double cleavage event. The **y ion** intensity will often be the most prominent peak in the spectrum.

3. It is common for **b** and **y** ions or **y** and **b** ions to swap intensity when a P is encountered in a sequence. This can also be true when the basic residues H, K, or R are encountered in the sequence.

4. When a cleavage appears before or after R, the -17 (loss of ammonia) peak can be more prominent than the corresponding **y** or **b** ion.

5. When encountering aspartic acid in a sequence, the ion series can die out.

Amino Acid Composition

1. It is possible to observe immonium ions at the low end of the spectrum that can give a clue to the amino acid composition of a peptide. One caveat is that if you do not see an immonium ion for a particular amino acid, this does not mean that that amino acid is absent from the sequence. You can follow this link to learn more about immonium ions.

Isobaric Mass

1. Leucine and Isolucine have isobaric masses and cannot be differentiated in a low energy collision. When we see this mass difference in a spectrum we will label it X or Lxx, adopting the Hunt nomenclature.

2. Lysine and Glutamine have near isobaric masses, **128**.09496 and **128**.05858 respectively. The delta mass is 0.03638 this difference can be used to differentiate K from Q on a mass spectroneter capable of higher mass accuracy and resolution, such as a q-TOF mass spectrometer. Usually triple quadrupole or ion trap mass spectrometers are incapable of this feat. On a lower mass accuracy mass spectrometers an acetylation can be performed to shift the mass of lysine by 42u. If you like to live dangerously, and we do not, one can assume that a 128 mass shift internally on a tryptic peptide is a glutamine unless followed by a proline or sometimes aspartic acid. Other instances of internal lysines left standing after a tryptic digest (this is our personal observation) is when double lysines occur in a sequence, so be careful.

3. There are instances where two residues will nearly equal the mass of a single residue, or a modified residue will nearly equal the mass of another amino acid. For more examples, see the following table.

More Rules

1. When starting a de novo sequencing project, start at the high mass end of the spectrum; the lower number of peaks at this end often makes it easier to start sequencing.

2. The region 60 u below the parent mass can be confounded by multiple water and ammonia losses, be careful. Realize that glycine may be your first amino acid and may fall in this region.

3. Do you want to know if your tryptic peptide ends in a K or an R? Look for the diagnostic **y**1 ions at the low end of the spectrum, you may observe 147 for K or 175 for R.

4. The **b1** fragment is seldom observed making it difficult to determine the order of the first two N-terminal amino acids in a peptide sequence. Solutions for this problem can include a one step Edman degradation or an acetylation.

5. Once you know the mass of a **b** or **y** ion the corresponding **y** or **b** ion can be calculated using the following formulas.

   $y = (M+H)^{1+} - b + 1$

   $b = (M+H)^{1+} - y + 1$

   Once you observe a **y** or **b** ion, calculate the mass of the corresponding **b** or **y** ion and go look for it in the spectrum!

You have learned the rules or at least know where to look for them, so now you can take a look at the basic protocol we use for de novo, when we sequence by hand.

The Protocol

Following the b Ion Series

1. If you are working with a tryptic peptide, look for the arginine or lysine **y**1 ion at the low end of the spectrum: 147 indicates a lysine, and a 175 indicates an arginine. This may give you a clue to the C-terminal residue of your tryptic peptide. Use this information to calculate the **b** ion that is the result of losing the C-terminal amino acid.

   Use this formula for Arginine $(M+H)^{1+}$ - 18 -156 = penultimate b ion

   Use this formula for Lysine $(M+H)^{1+}$ - 18 -128 = penultimate b ion

   (protonated peptide - 18 - AA residue mass = penultimate b ion)

   or you could use the standard formula for calculating the corresponding **b** ion once the **y** ion is known.

   $(M+H)^{1+} - y_1+1$ = penultimate b ion

   If you are using an ion trap you may not be able to observe the low end of the spectrum, and in this case, you will need to do both of these calculations

   Go back to the high end of the spectrum and look for this **b** ion that you just calculated.

2. Whenever you identify a **b** ion look for an **a** ion at -28u. This gives some assurance that your assignment is correct. Also look for ammonia and water losses, -17 and -18u respectively. Whenever you identify a **b** ion do the math to find the corresponding **y** ion.

   y = $(M+H)^{1+}$ - **b** +1

   Go and look for this calculated **y** ion. All of this data should fit together to help firm up your assignments.

3. Look for the next **b** ion residue in the series. Use the amino acid residue masses to look for the next peak, see table. Soon you will have the residue masses memorized. Take the smallest amino acid jump possible to search for the next **b** ion. It is important to make the smallest jump because some residue combinations equal the mass of a single residue, for example GG = N, see our conflicting masses table. Label the **b** ion that you find and look for the related **a** ion, and calculate the corresponding **y** ion. You will not always be able to find an **a** ion, however, sometimes you can, and it is an assurance that you are on the right track.

4. Continue to follow the **b** ions down to the low end of the spectrum. Once you reach the low end and cannot go any further construct the **b** ion series. Since you will not see a **b**1 ion you will often need to calculate the mass residue combinations that compose the gap at the end, in which case you will not be able to determine the order of these two amino terminal amino acids.

5. Since you have calculated all of the corresponding **y** ions go ahead and work up the **y** ion series that you have observed in the spectrum. It may be possible that you can determine the order of the amino terminal residues as you work the sequence back towards the high mass end. Hind sight being 20:20, even though you have calculated and observed the **y** ion series, it is always best to try to call the corresponding **y** or **b** ion sequence "de novo", in this case the **y** ion series. This may save you from calling a GG as an N.

Following the y Ion Series It may not always be easy to follow the **b** ion series when dealing with tryptic peptides. Tryptic peptides tend to be more basic at the C-terminus and may have a more prominent y ion series, this is definitely the case for the q-TOF data we will be looking at. If this is the case go to the high end of the spectrum and look for the next to last **y** ion. Here is the method for finding and following a **y** ion series from the high end of the spectrum.

1. Use the formula below to calculate the **y** ion formed by the loss of the amino terminal amino acid. You will need to plug in each of the common amino acid masses, then look for the peak in the mass spectrum.

   $(M+H)^{1+}$ - AA = penultimate **y** ion

   Or you could look at the spectrum and find a prominent peak, then do the math to see if it corresponds to one of the common amino acids. This **y** ion should be found between the smallest and the largest amino acid residue mass, between 57-186 u.

   $(M+H)^{1+}$ - observed ion = AA

2. Once a **y** ion is found, calculate the corresponding **b** ion and look for it, and label it in the spectrum.

   b = $(M+H)^{1+}$ - y +1

3. As outlined above, in the **b** ion series protocol, continue to follow the **y** ion series down to the low end of the mass spectrum.

4. Once you hit the end of the sequence, construct a **y** ion series from your observations. Again from your calculated and observed **b** ion series construct a **b** ion sequence.

Conclusion: It is pretty rewarding calling a peptide sequence using de novo sequencing techniques. It is hard, and you do need to be careful because there are pitfalls, like isobaric residues and residue combinations. Still you can get to be good at it. It is a puzzle and can be fun. Continue on to the examples and exercises and try to complete them without looking at the answers until you are done, you will be impressed at your newfound ability. Even if you will be using de novo sequencing software, it is good to know how it all works.

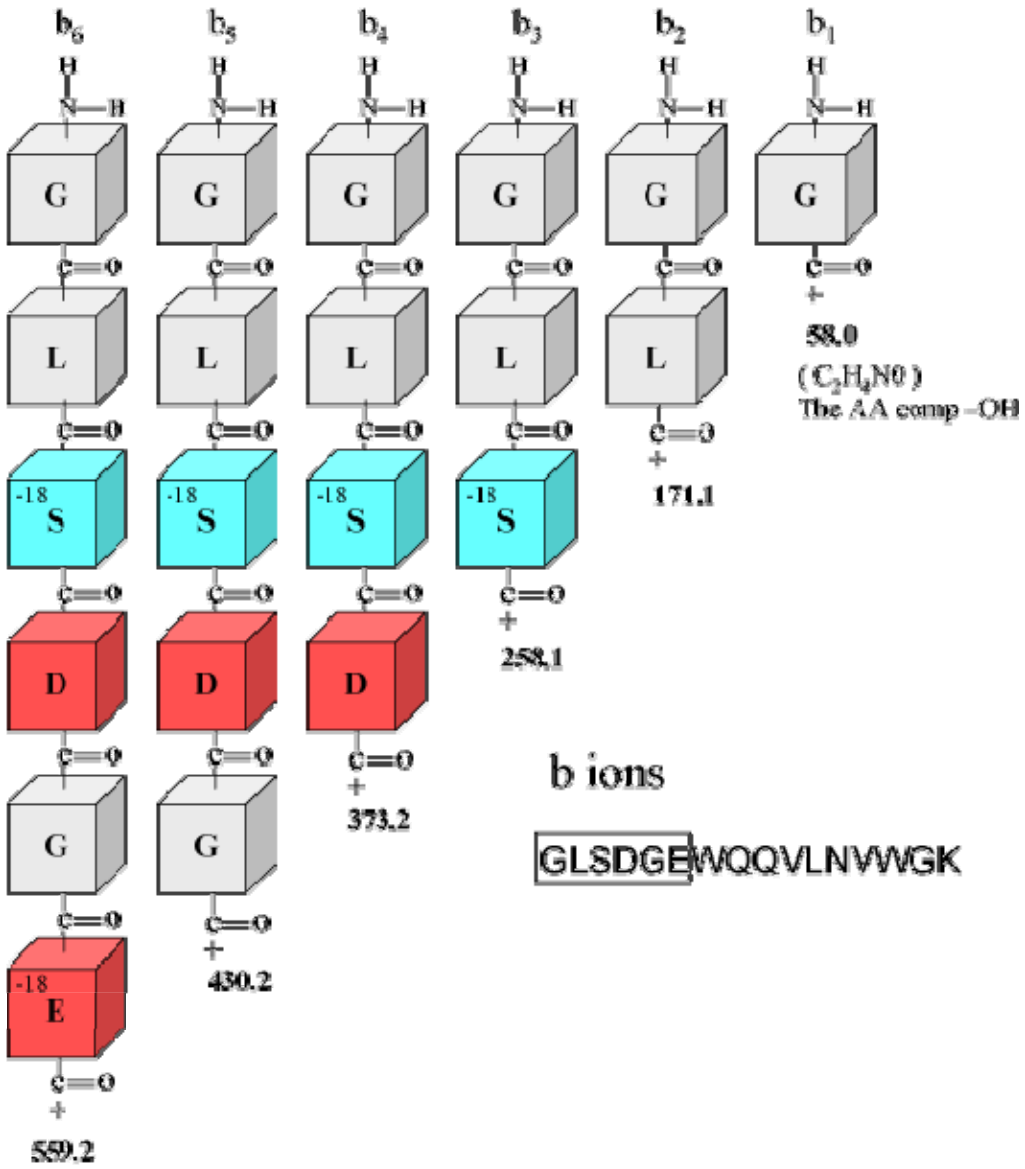
http://ionsource.com/tutorial/DeNovo/introduction.htm

Example Spectrum part one, low mass

For this example we will start at the low end of the spectrum and work our way up to the high end. This is just an example spectrum used for illustrative purposes. We will basically know the ions we are looking for and go look for them, and also we will learn what we can learn. Later on with the exercises we will follow the rules of our protocol and start at the high end of the spectrum. Below, in Figure 1, are six of the **b** ions described previously for the tryptic peptide GLSDGEWQQVLNVWGK. You may notice that we have annotated the serine and glutamic acid blocks with the notation -18, this is to remind us that these residues may be associated in the spectrum with water losses. In Figure 2 we have made a screen shot of the low end of the MS/MS spectrum to look for some of the predicted **b** ions and their associated fragment ions.

b Ion Details

Figure 1.

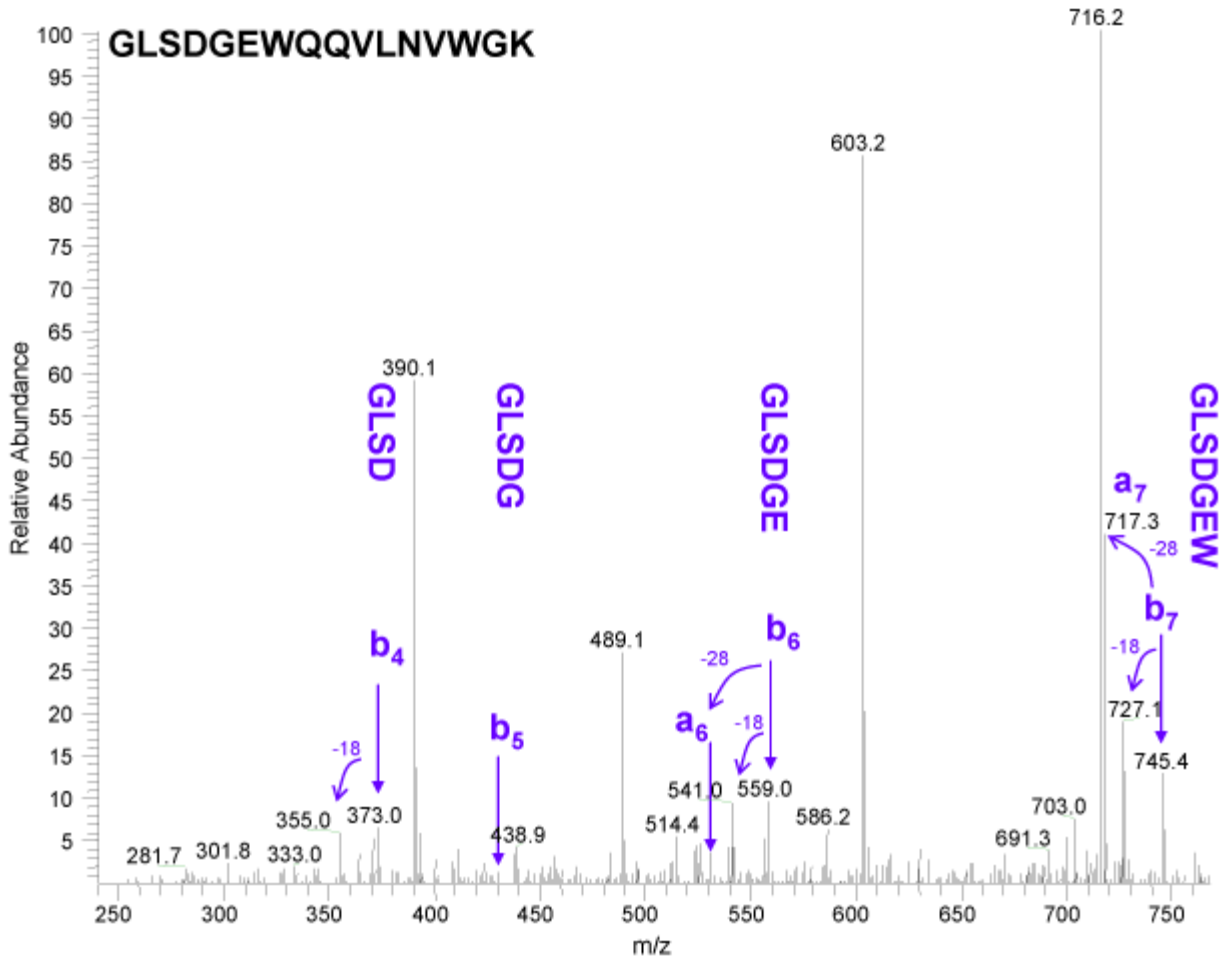


http://ionsource.com/tutorial/DeNovo/introduction.htm

Figure 1. We have taken the first 6 amino acids from the peptide GLSDGEWQQVLNVWGK to show in detail, **b** ions 1-6. The calculated b ion masses are noted in bold numbers below the block figures.

In Figure 2 below, it is easy to see the that the **b** ion series is not the predominant series. Data was not acquired below m/z 240 because this is ion trap data, and traps typically have a low mass cutoff, sometimes called the 1/3 rule. Due to this fact we have lost our diagnostic low mass immonium ions and several other important diagnostic ions. We were easily able to observe **b** ions 4, 6 and 7. Ion **b**5 was too low to call, most likely due to to the overall low abundance of the **b** ion series at the lower end of this spectrum. Calling a true de novo sequence with this **b** ion data would have been a challenge with these low ion intensities. There is an interesting feature in this spectrum that confirms that we are looking at a **b** ion series, which is that we observe **a** ions associated with **b**6 and **b**7. The loss of C=O or 28 u is the characteristic mass difference for the **a** / **b** pair. One can also observe water losses for **b** ions 4, 6 and 7 as predicted by the presence of serine and glutamic acid in this sequence.

Figure 2.



http://ionsource.com/tutorial/DeNovo/introduction.htm

Why did the **b**5 ion drop out in Figure 2? It looks like there was a significant water loss with a peak at m/z 412 which could have diminished the already small peak at m/z 430? Or, it could be the de novo sequencing rule, "**b** ion intensity may drop when the next residue is P, **G** or also H, K, and R", pretty cool, huh!

Figure 3.

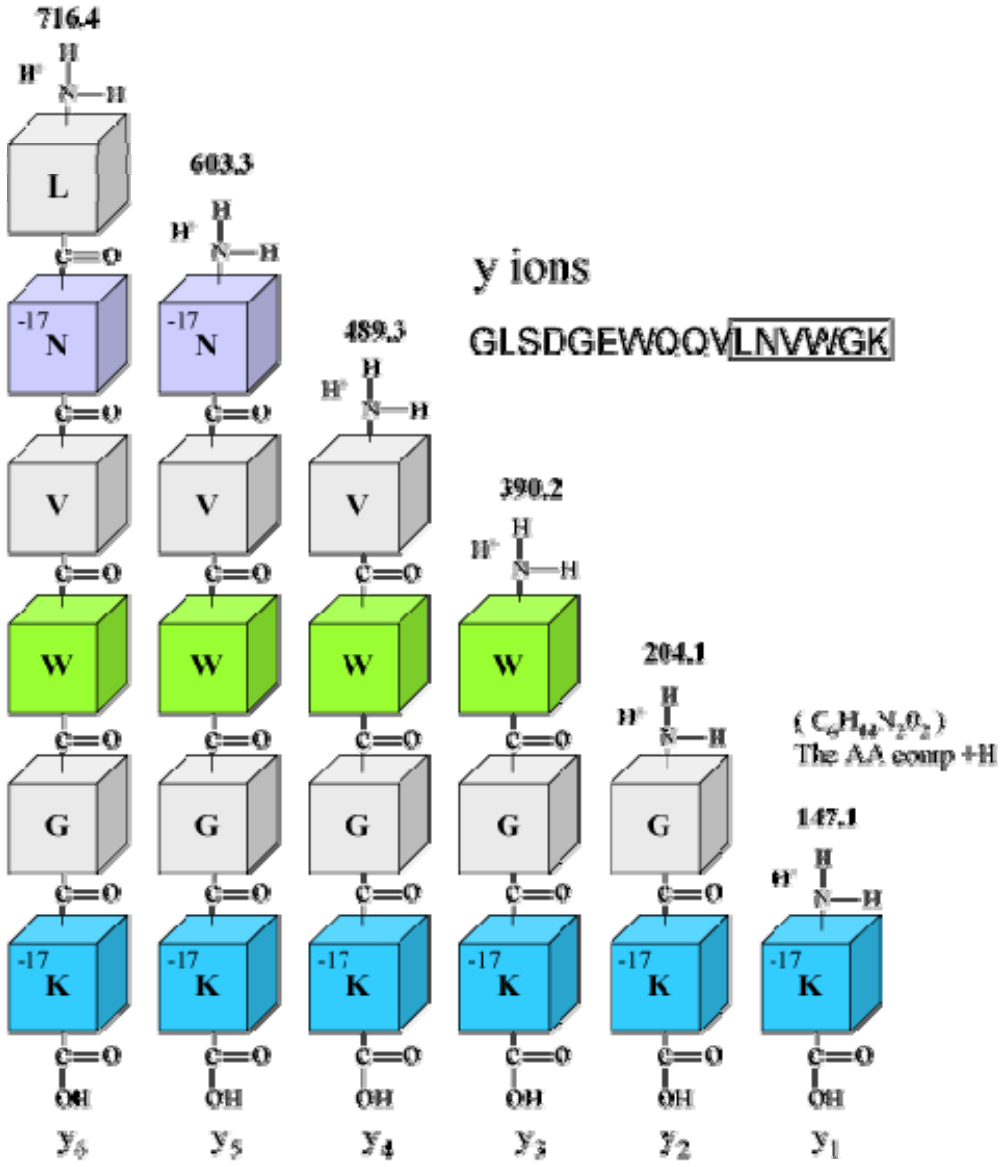
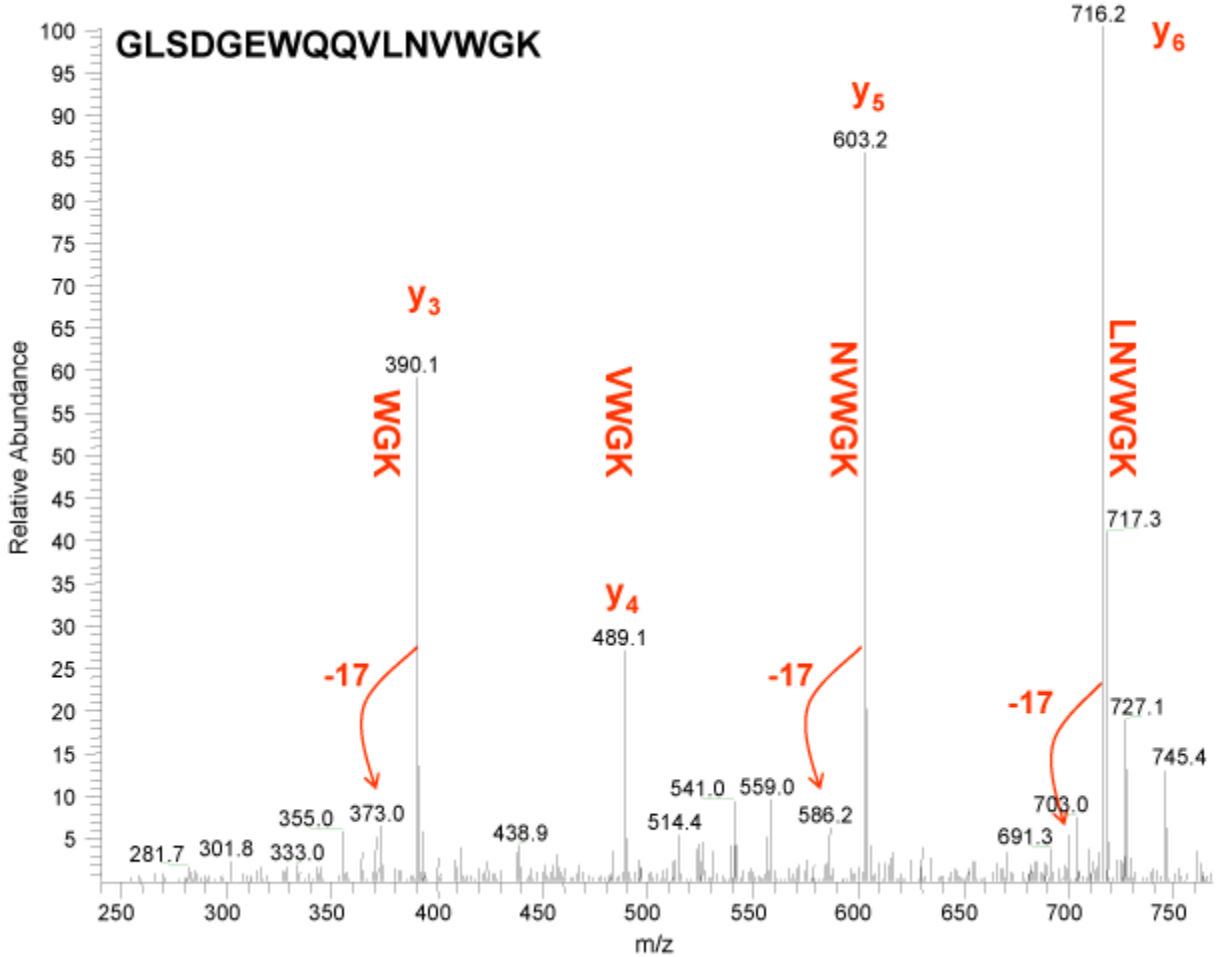


Figure 3 above illustrates the y ion series. Again note that we have annotated the lysine and asparagine blocks with -17. This is to remind us that these residues appear to lose ammonia, and for the majority of **y** ions observed below in Figure 4, we also observe a peak that is 17 u lower. The apparent loss of ammonia is diagnostic for these residues.

Figure 4.

http://ionsource.com/tutorial/DeNovo/introduction.htm



So far we have discussed **b** and **y** ions, and for simplicity we have relegated ourselves to the terminal 6 or 7 amino acids. We will detail the entire spectrum in the following slides. It is apparent that the amino terminal sequence may be difficult to determine from the **b** ion series, one because of the relative abundance, and two because of the limitations of the ion trap. At the end of this tutorial we will provide spectra obtained on a q-TOF mass spectrometer for you to practice with, which does not share the same ion trap low mass limitation.

If you are interested in following this sequence proceed onto the next page. It is getting kind of interesting, isn't it? I wonder what will happen next?

Reference Figures

---

b and y Ion Table

Fragment Ion Calculator Results

Sequence: **GLSDGEWQQVLNVWGK**, pI: **4.37029**

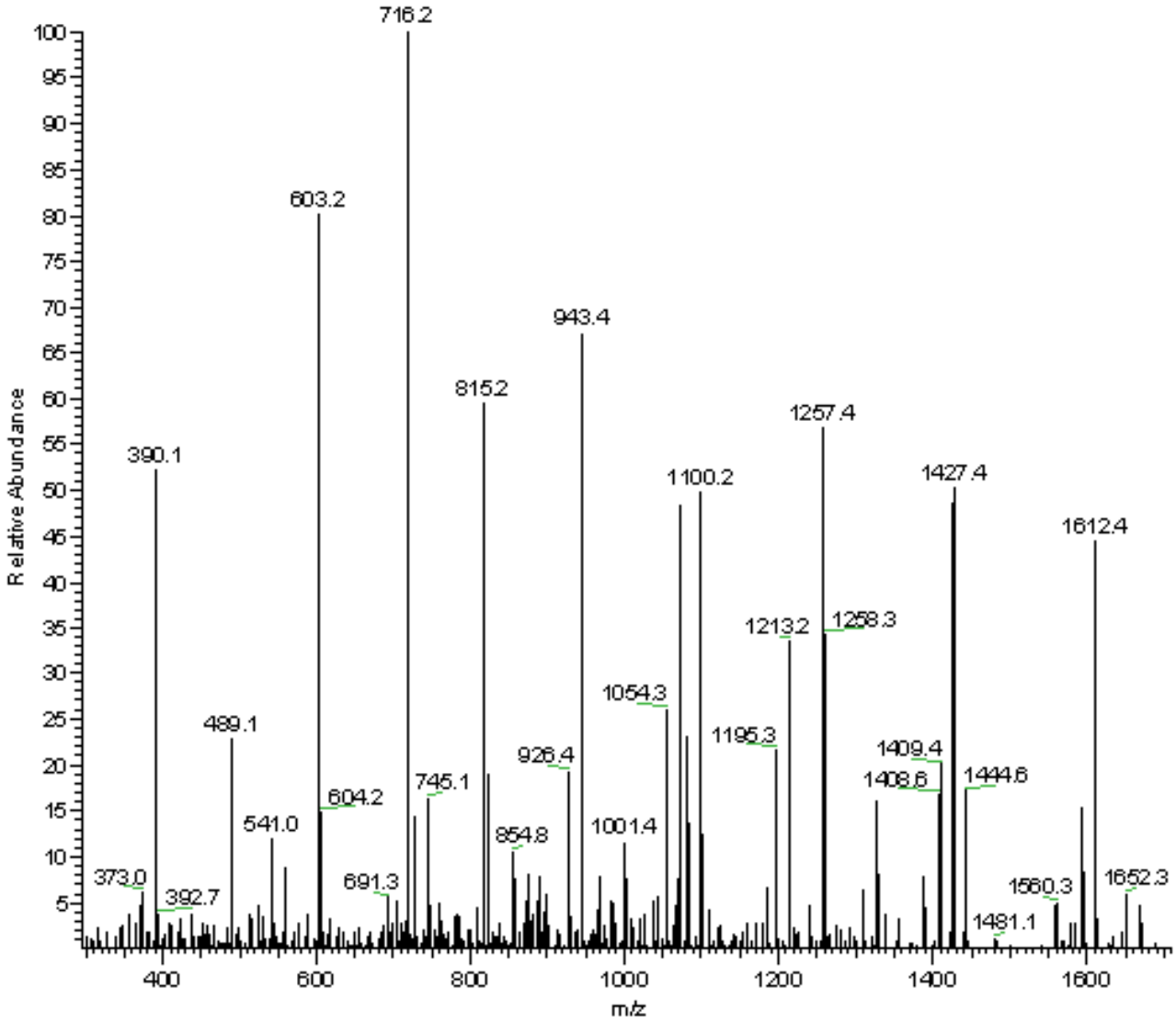
Fragment Ion Table, monoisotopic masses

Seq	#	B	Y	# (+1)
G	1	58.02933	1815.90301	16
L	2	171.11340	1758.88155	15
S	3	258.14543	1645.79749	14
D	4	373.17237	1558.76546	13
G	5	430.19383	1443.73851	12
E	6	559.23642	1386.71705	11
W	7	745.31574	1257.67446	10
Q	8	873.37431	1071.59515	9
Q	9	1001.43289	943.53657	8
V	10	1100.50131	815.47799	7
L	11	1213.58537	716.40958	6
N	12	1327.62830	603.32551	5
V	13	1426.69671	489.28259	4
W	14	1612.77602	390.21417	3
G	15	1669.79749	204.13486	2
K	16	1797.89245	147.11340	1

---

http://ionsource.com/tutorial/DeNovo/introduction.htm

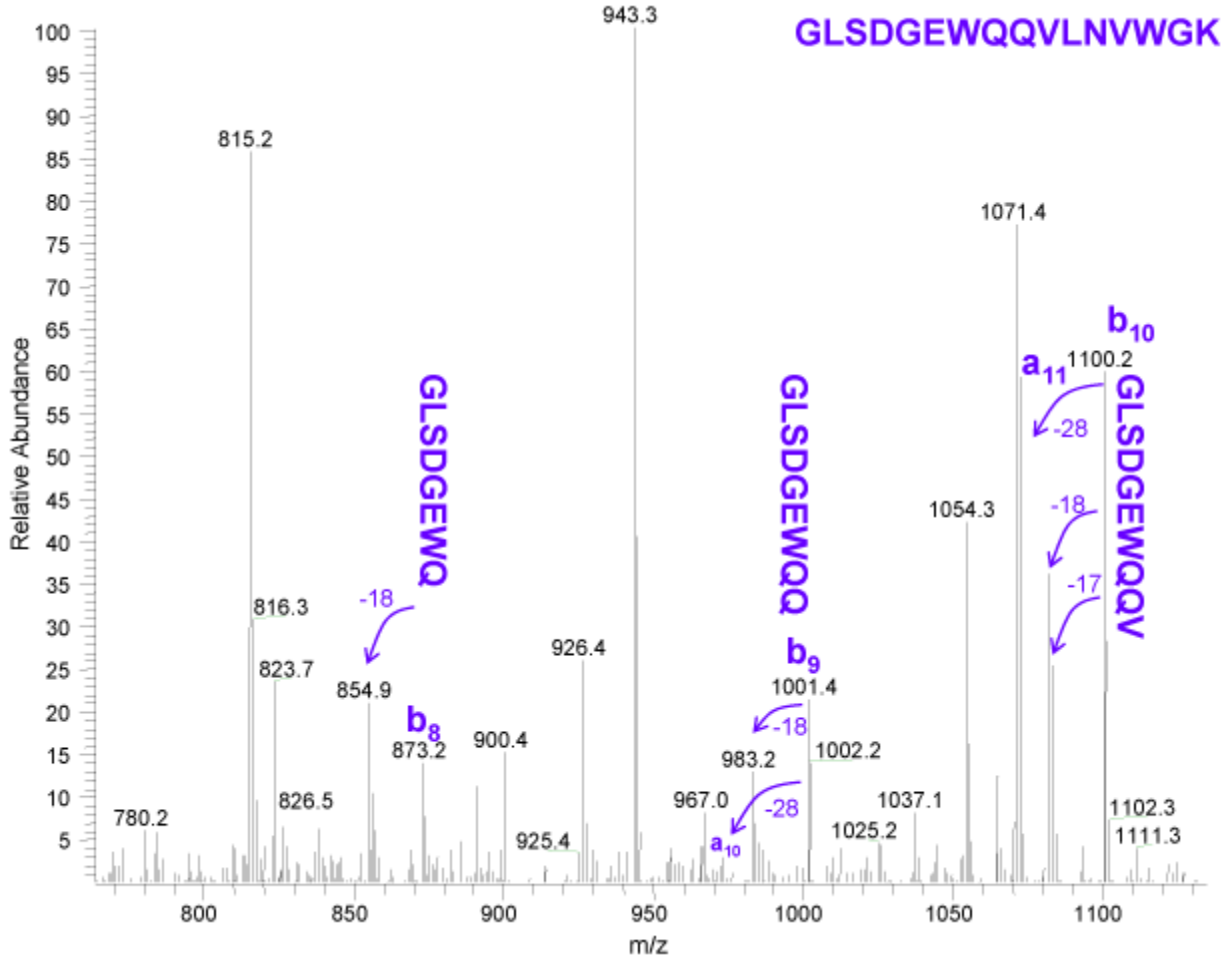
Full Fragment Spectrum



Example Spectrum part two, mid mass

As we move into the mid spectrum range the intensity of the **b** ions increase, see Figure 1 below. The **b** ions in Figure 1 are losing water and ammonia which is consistent with S (-18) and Q (-17) residues. Peaks consistent with **a** ions are also observed at -28u from two of the **b** ions.

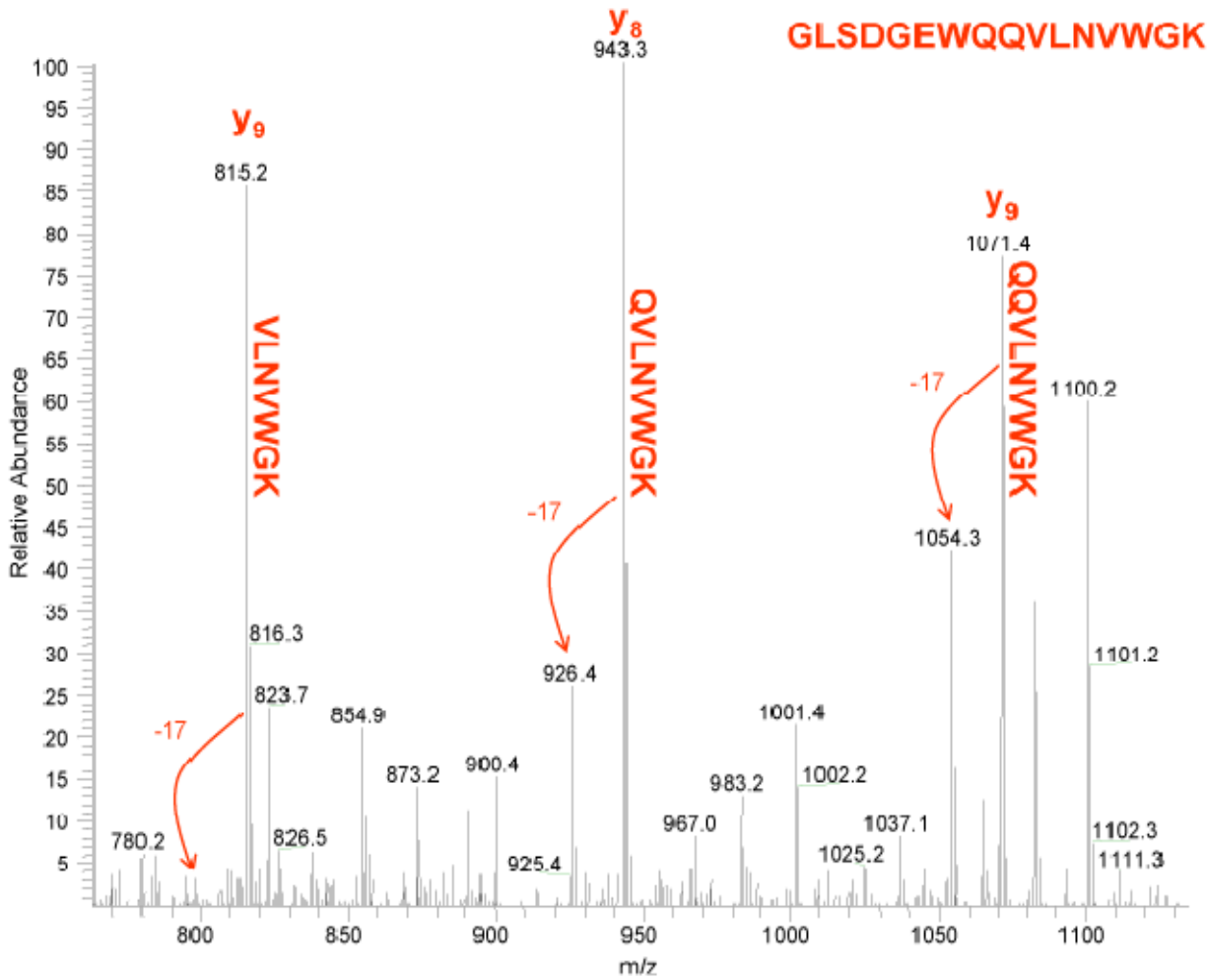
Figure 1.



The predominant ion in the mid range is the **y** ion. Notice that only ammonia loss (-17) is noted for the **y** ions in Figure 2. The -17u loss is consistent for fragments containing the residues K, Q and N. Also note that the greatest ammonia loss is observed for fragment $\mathbf{y}_9$ which contains the greatest number of these residues.

Figure 2.

http://ionsource.com/tutorial/DeNovo/introduction.htm



Shall we proceed onto the high mass range of the spectrum? Let's.

http://ionsource.com/tutorial/DeNovo/introduction.htm

Reference Figures

---

b and y Ion Table

Fragment Ion Calculator Results

Sequence: **GLSDGEWQQVLNVWGK**, pI: **4.37029**

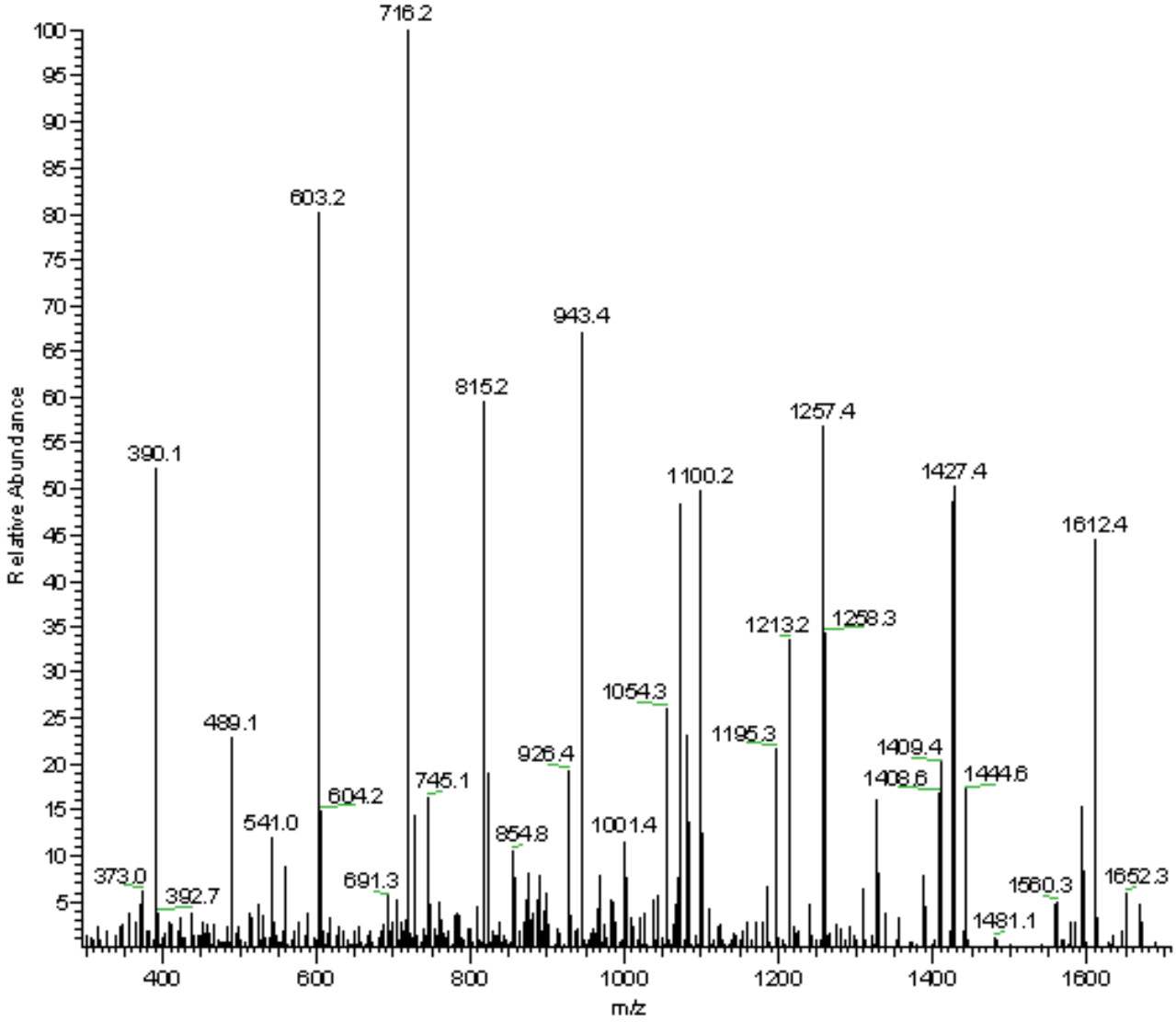
Fragment Ion Table, monoisotopic masses

Seq	#	B	Y	# (+1)
G	1	58.02933	1815.90301	16
L	2	171.11340	1758.88155	15
S	3	258.14543	1645.79749	14
D	4	373.17237	1558.76546	13
G	5	430.19383	1443.73851	12
E	6	559.23642	1386.71705	11
W	7	745.31574	1257.67446	10
Q	8	873.37431	1071.59515	9
Q	9	1001.43289	943.53657	8
V	10	1100.50131	815.47799	7
L	11	1213.58537	716.40958	6
N	12	1327.62830	603.32551	5
V	13	1426.69671	489.28259	4
W	14	1612.77602	390.21417	3
G	15	1669.79749	204.13486	2
K	16	1797.89245	147.11340	1

---

http://ionsource.com/tutorial/DeNovo/introduction.htm

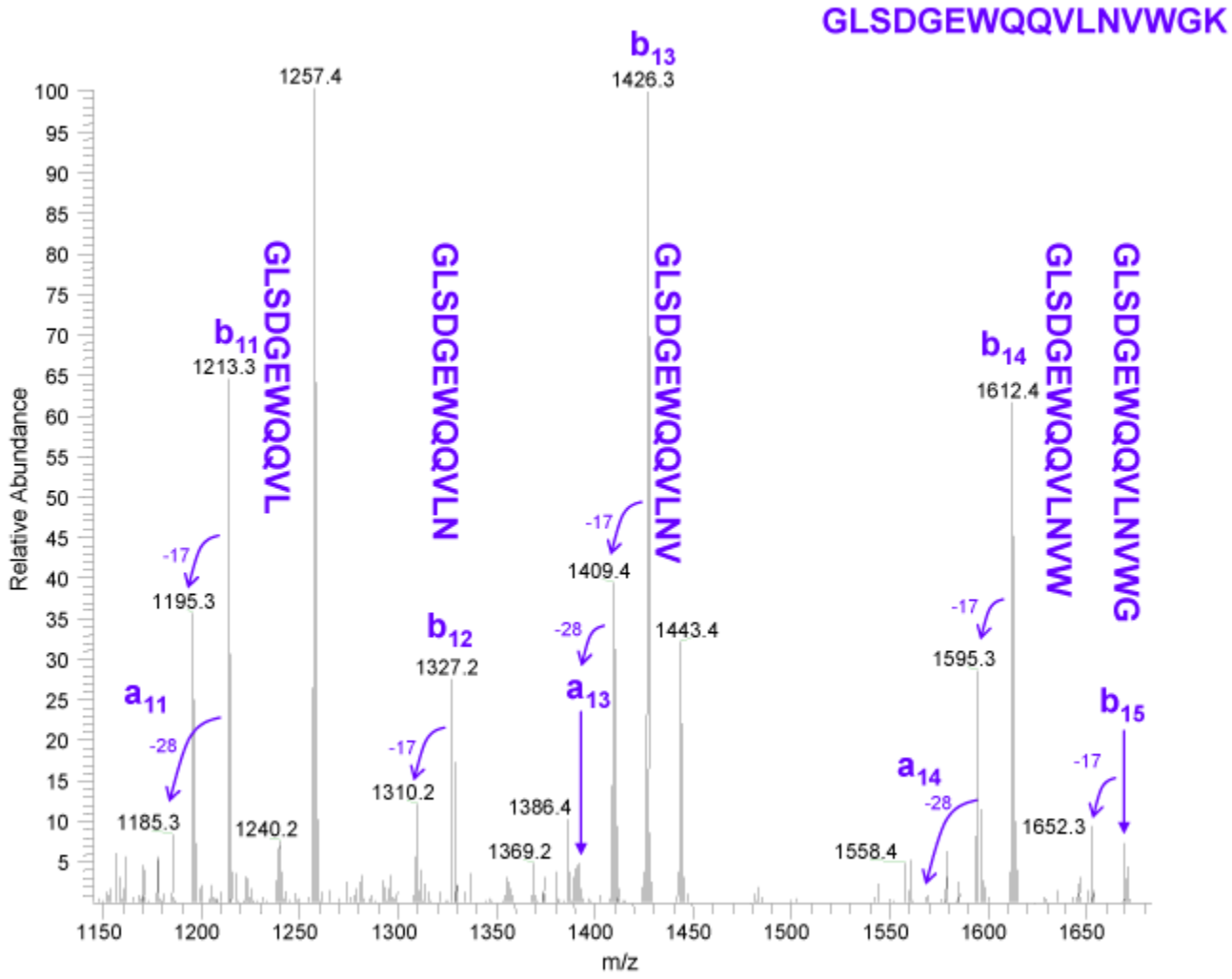
Full Fragment Spectrum



Example Spectrum part three, high mass

At the high end of the spectrum, the **b** ions increase dramatically in intensity as compared to the **b** ion intensity at the low end of the spectrum. In addition, **y** ion intensity has greatly decreased. It is as if the **y** and **b** ion intensity have swapped.

Figure 1.



http://ionsource.com/tutorial/DeNovo/introduction.htm

Figure 2.

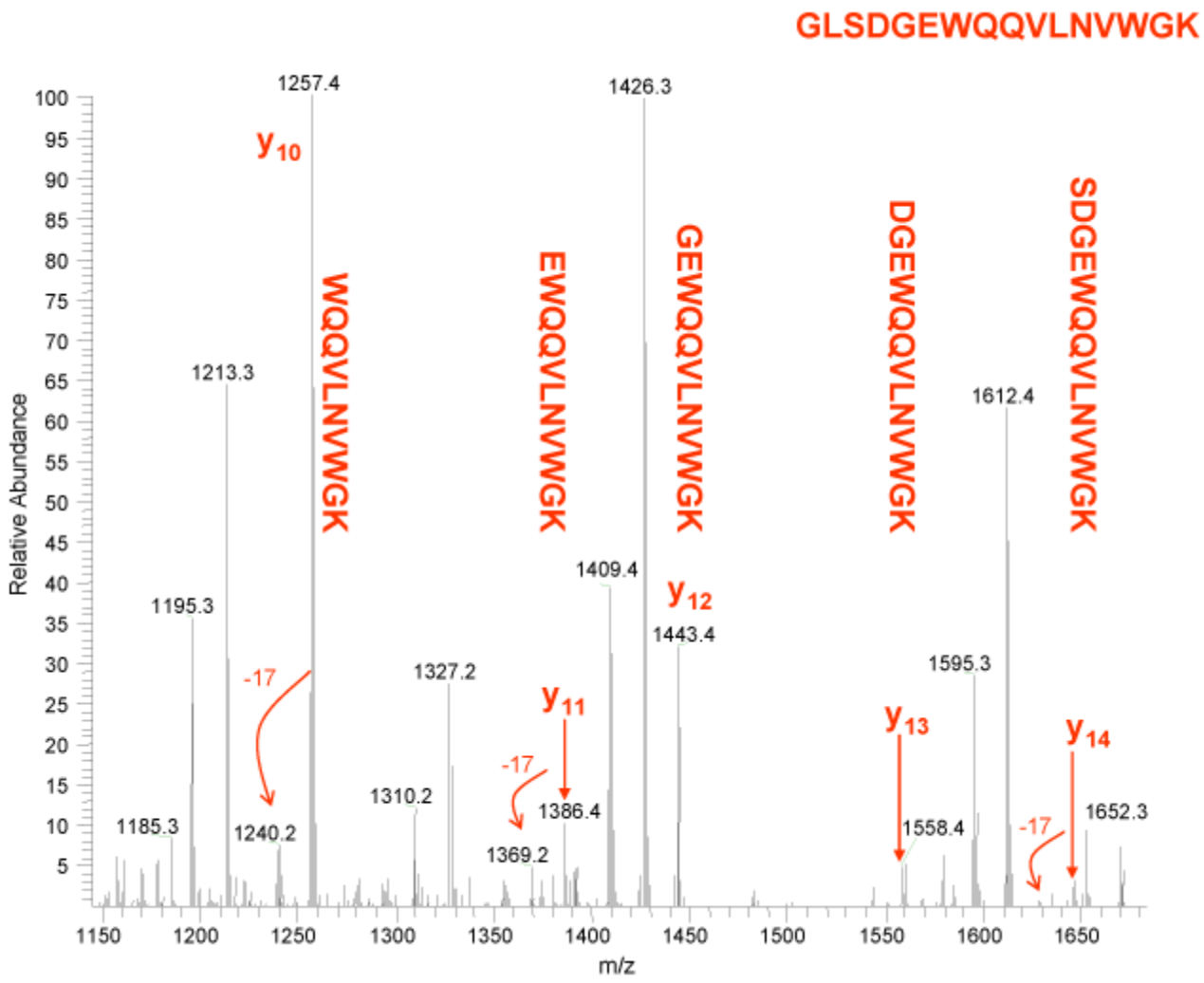


Figure 2 shows the greatly diminished **y** ion intensity at the high end of the spectrum. This occurs as we encounter an E. The reduction in peak intensity is pretty dramatic. Out of curiosity we plotted the pIs of the **y** and **b** ions, take a look. Amazing, between the **y** and **b** ion coverage we have accounted for almost all of the major peaks.

Feeling pretty bold? Proceed on to the next page for a true *de novo* example.

Reference Figures

---

b and y Ion Table

Fragment Ion Calculator Results

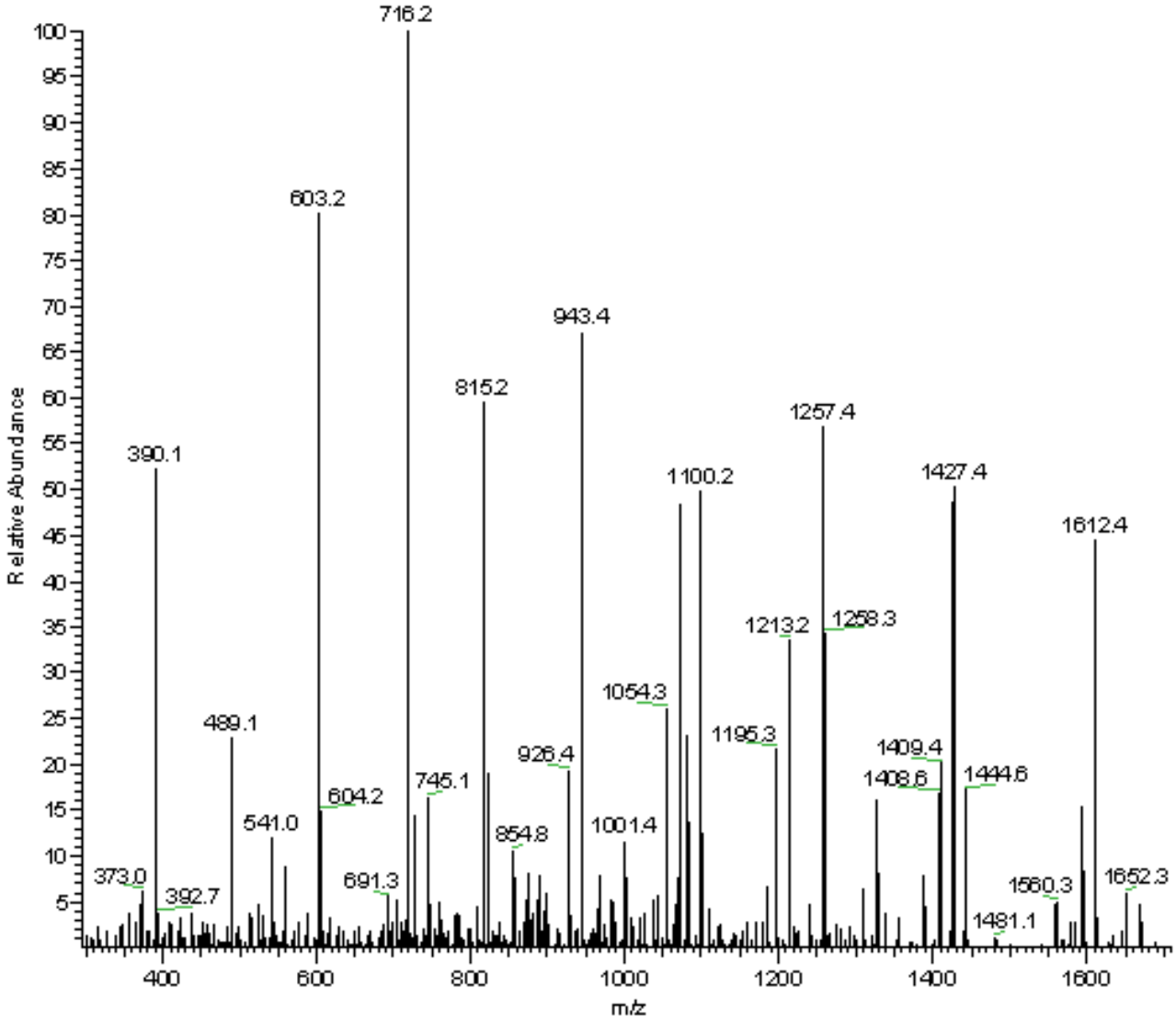
Sequence: **GLSDGEWQQVLNVWGK**, pI: **4.37029**

Fragment Ion Table, monoisotopic masses

Seq	#	B	Y	# (+1)
G	1	58.02933	1815.90301	16
L	2	171.11340	1758.88155	15
S	3	258.14543	1645.79749	14
D	4	373.17237	1558.76546	13
G	5	430.19383	1443.73851	12
E	6	559.23642	1386.71705	11
W	7	745.31574	1257.67446	10
Q	8	873.37431	1071.59515	9
Q	9	1001.43289	943.53657	8
V	10	1100.50131	815.47799	7
L	11	1213.58537	716.40958	6
N	12	1327.62830	603.32551	5
V	13	1426.69671	489.28259	4
W	14	1612.77602	390.21417	3
G	15	1669.79749	204.13486	2
K	16	1797.89245	147.11340	1

http://ionsource.com/tutorial/DeNovo/introduction.htm

Full Fragment Spectrum



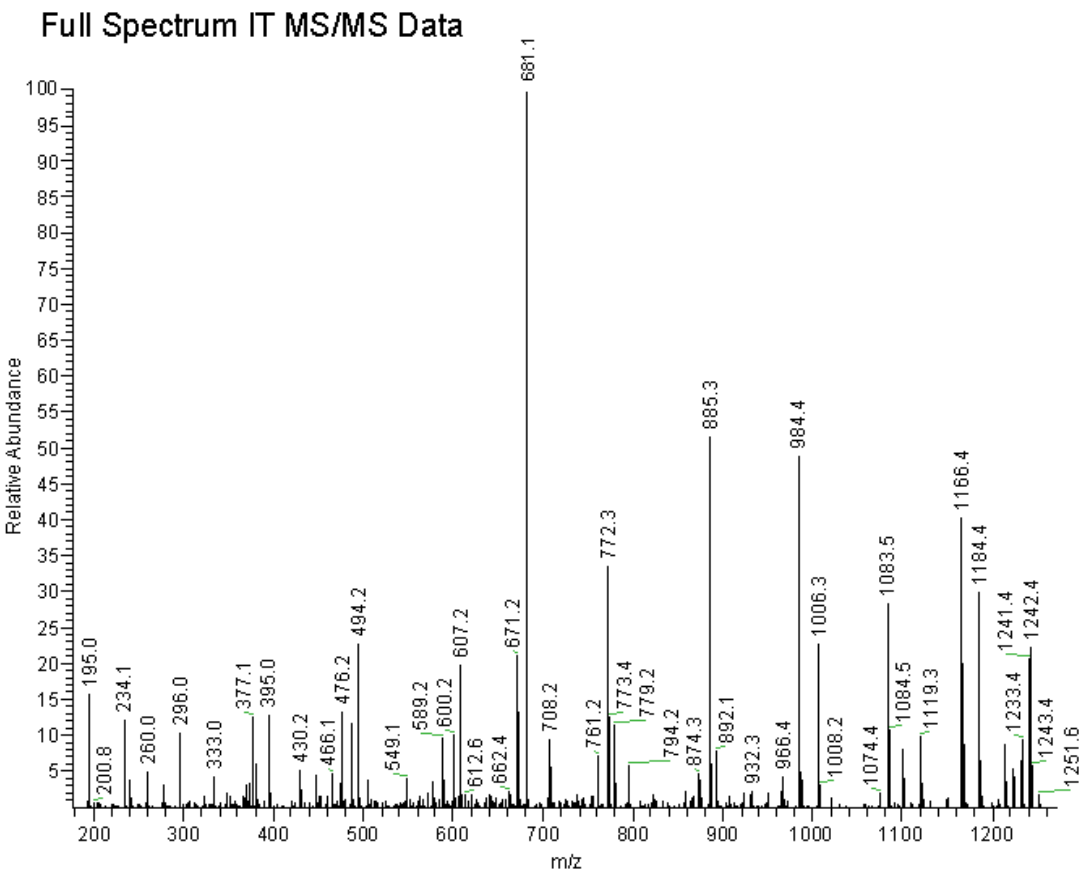
1st De Novo Exercise

Here is an MS/MS spectrum for you to test out your newly acquired de novo skills. Below we have an MS/MS spectrum acquired on an ion trap mass spectrometer. The first figure is the full fragment spectrum; the spectrum is then divided into three sections to make the details easier for you to see. Try not to be intimidated, just follow the simple protocol that we have described previously. Start with the high end of the spectrum, see Figure 2. We have made a pdf file for you so that you can download the spectrum and print them out so that you can get a pencil and a calculator and start sequencing. To download the spectra, right click on the link and save the pdf document to your computer.

Here are the hints:

1. This is a tryptic peptide ending in K or R.
2. The Parent [M+H]1+ mass is 1379.4
3. Since this is ion trap data, mass accuracy should be within 0.8u

Figure 1.



Here is a zoom in on the high end of the mass spectrum.

http://ionsource.com/tutorial/DeNovo/introduction.htm

Figure 2.

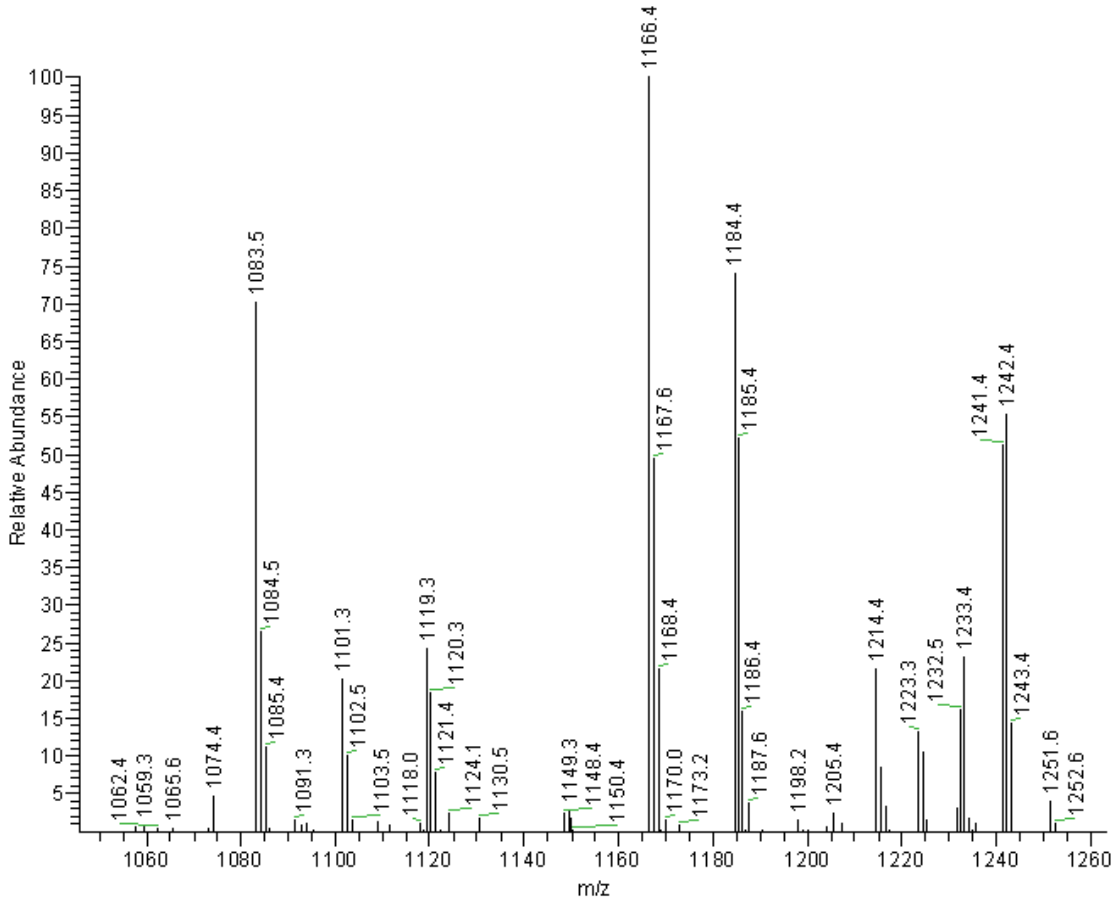
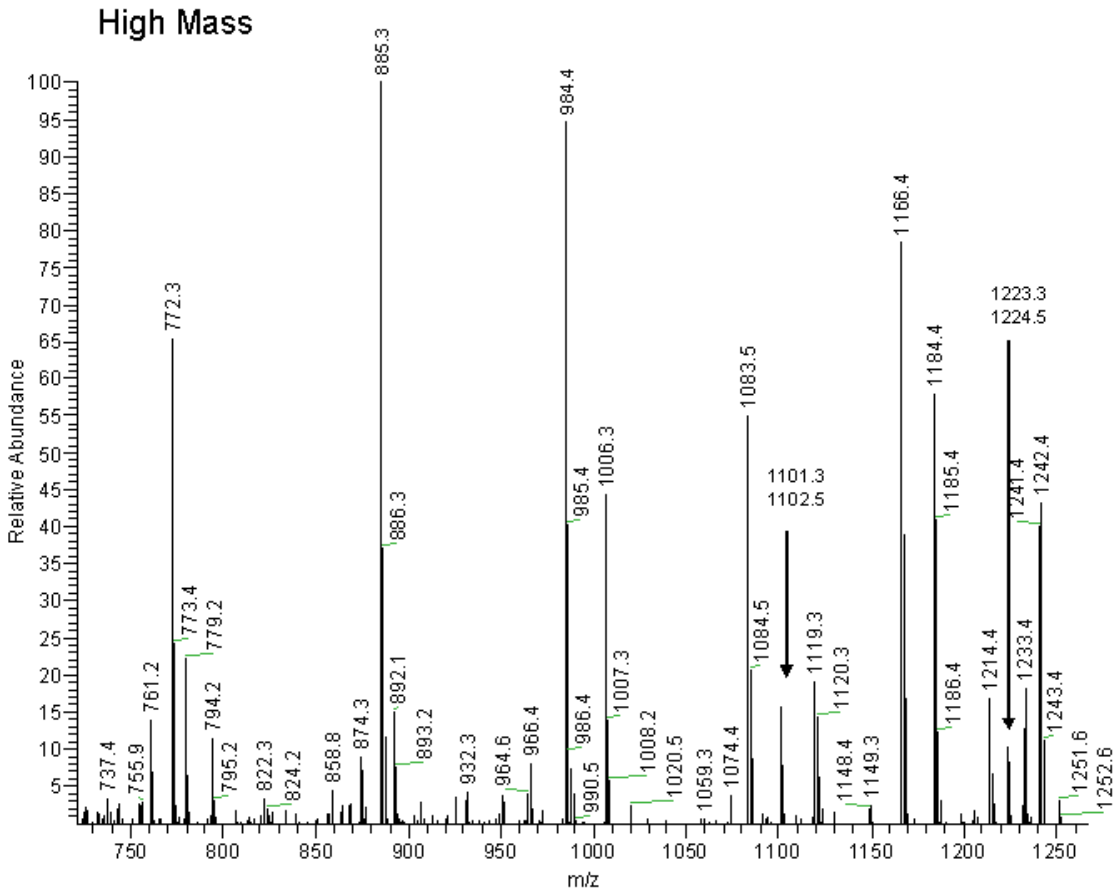
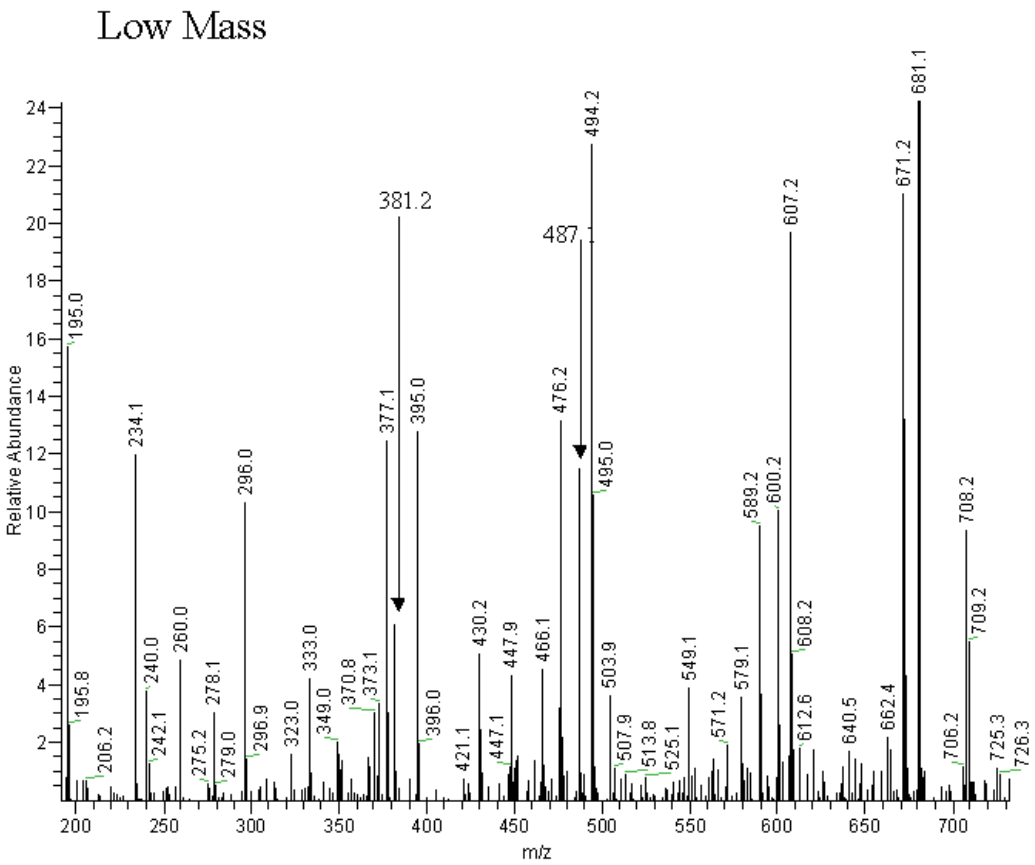


Figure 3.



http://ionsource.com/tutorial/DeNovo/introduction.htm

Figure 4.



Once you have done your best you can proceed onto the answer page to check out the known sequence and see some of the issues we ran into with this spectrum.

http://ionsource.com/tutorial/DeNovo/introduction.htm

1st De Novo Exercise the answer page

The correct sequence for this peptide is,

HGTVVLTALGGILK

or more accurately the best that low energy CID can do

HGTVVXTAXGGXXK

The first thing that we did was to do the math to determine the penultimate **b** ion.

If we could determine the C-terminal residue from the observed immonium ions we could calculate the penultimate **b** ion using the formulas below.

1379.4 - 18 -156 = 1205.4 for Arginine

1379.4 - 18 -128 = 1235.4 for Lysine

Since we could not observe the low end of the ion trap spectrum we had to do the math for both likely tryptic peptide C-terminal residues. We found that the C-terminal residue matched for Lysine. The **b** ion series extended all of the way into the low end of the spectrum. The **a** ion type fragments did accompany some of the **b** ions down the spectrum giving us some confidence that we were following the correct sequence. For the **b** ion series, once we got down to the low end of the spectrum we had to use mass and amino acid residue combinations to determine the last two amino terminal residues. From the observed **b** ion series we could not determine the order of the remaining two amino terminal amino acids. Once the **b** ions were determined we calculated the expected **y** ion series using the formula $[M+H]^+$ - observed **b** ion + 1 = corresponding **y** ion. We then went through and redid the delta masses on all of the calculated and observed **y** ions, obtaining the **y** ion sequence. The **y** ions were more abundant and let us determine the order of the two amino terminal amino acids as well as removing the N-X ambiguity that we observed in the **b** ion series. We had this N-X ambiguity because the **b** ion series was much less abundant that the **y** ion series, and a more accurate mass was easier to determine for the **y** ion series With the **b** ion series, we still missed the G-G sequence and substituted an asparagine, a rookie mistake.

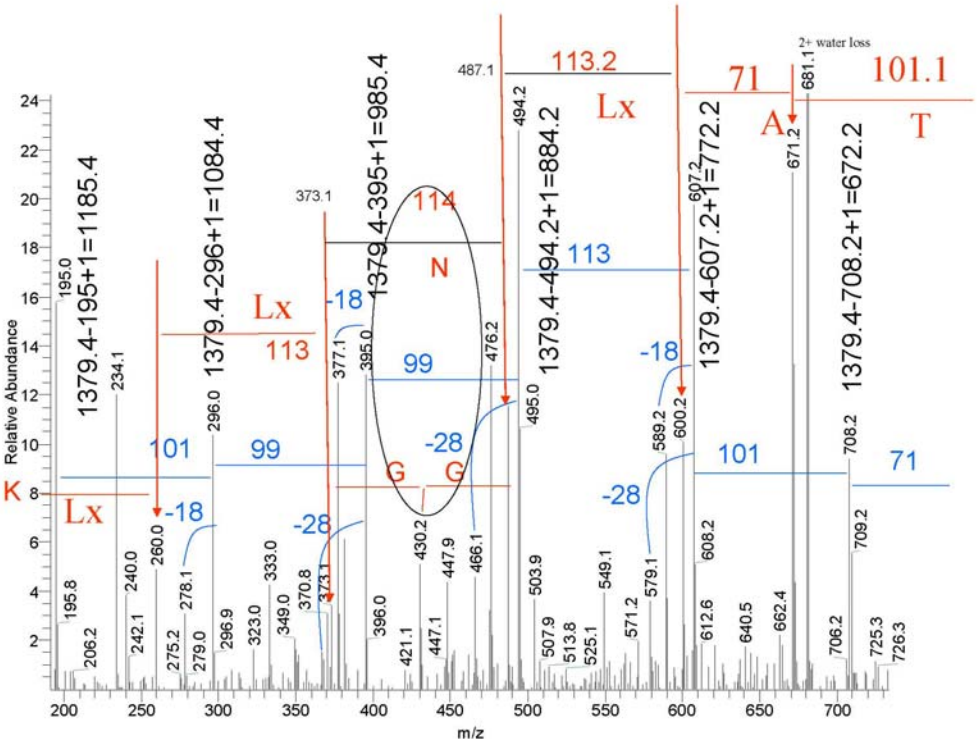
Here is the sequence that we determined (called) from both the **b** and **y** ion series.

X = leucine or isoleucine

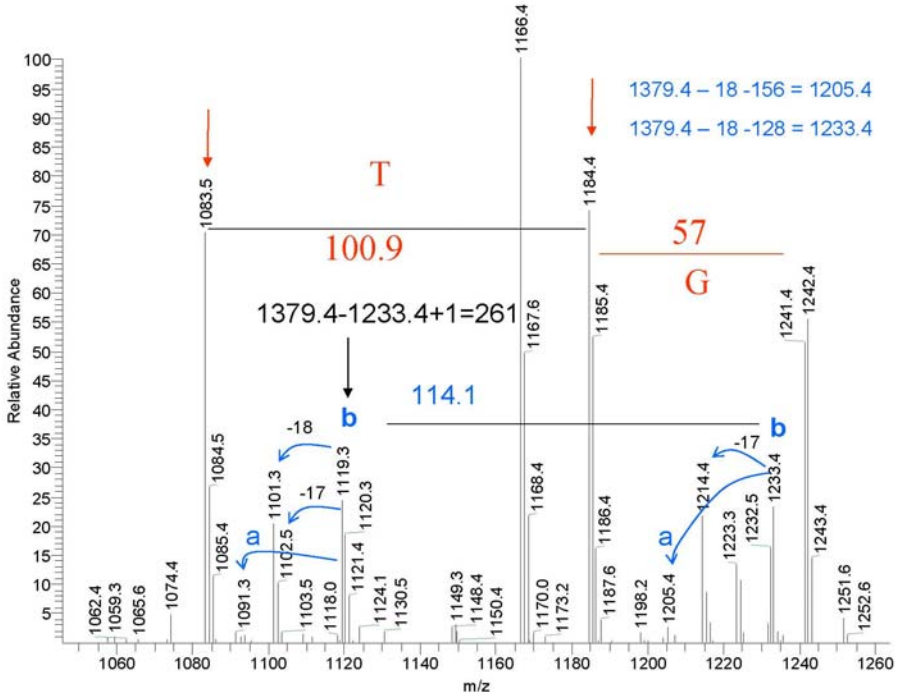
http://ionsource.com/tutorial/DeNovo/introduction.htm

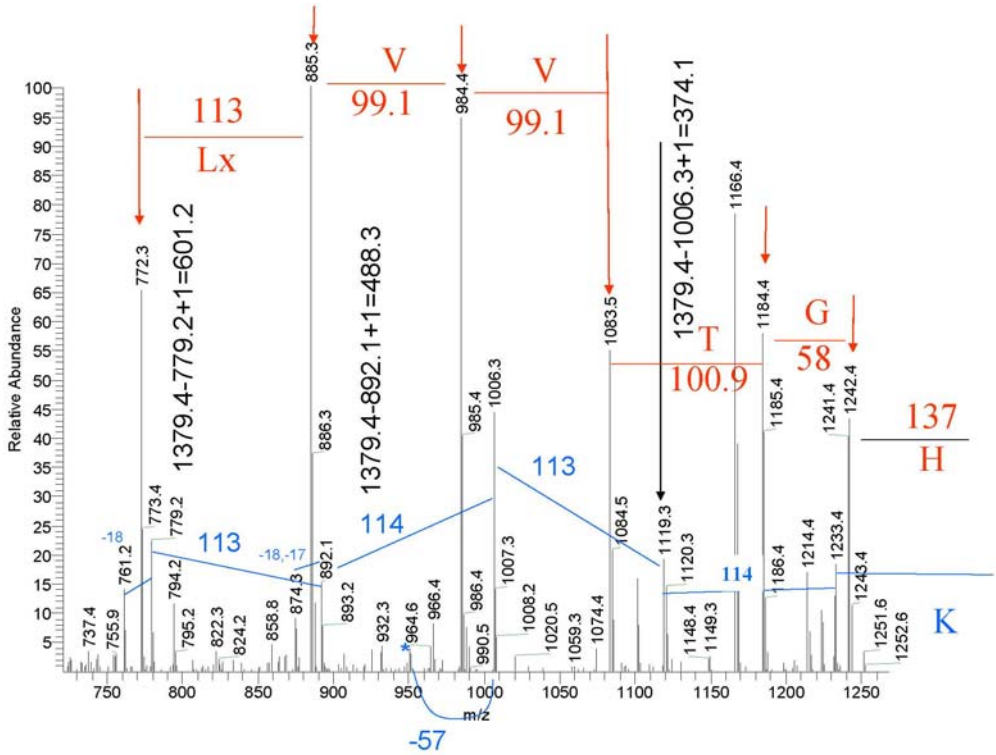
All considered, we feel that we did pretty good. We did mistake the double glycine for an asparagine. The peak that would have indicated the glycine in the **b** ion series was very small and we fell into the N = GG trap. Once this mistake was made we missed the glycine peak in the **y** ion series even though the peak was fairly observable. Even though one can calculate the corresponding **y** or **b** ion once a **b** or **y** ion is observed we would suggest calling both the **y** and **b** series independently to remove any bias. This would have saved us from calling GG as an N. Perhaps we could do better with q-TOF data, let's see as we proceed onto the next exercise. Still, we did pretty good; hope you did too. Don't let anyone tell you that trap data is bad for de novo, you just need to try a little harder and be a little wary.

If you would like to see the scribbling of a mad man, as I marked my spectrum up in this exercise:



http://ionsource.com/tutorial/DeNovo/introduction.htm





2nd De Novo Exercise q-TOF Data

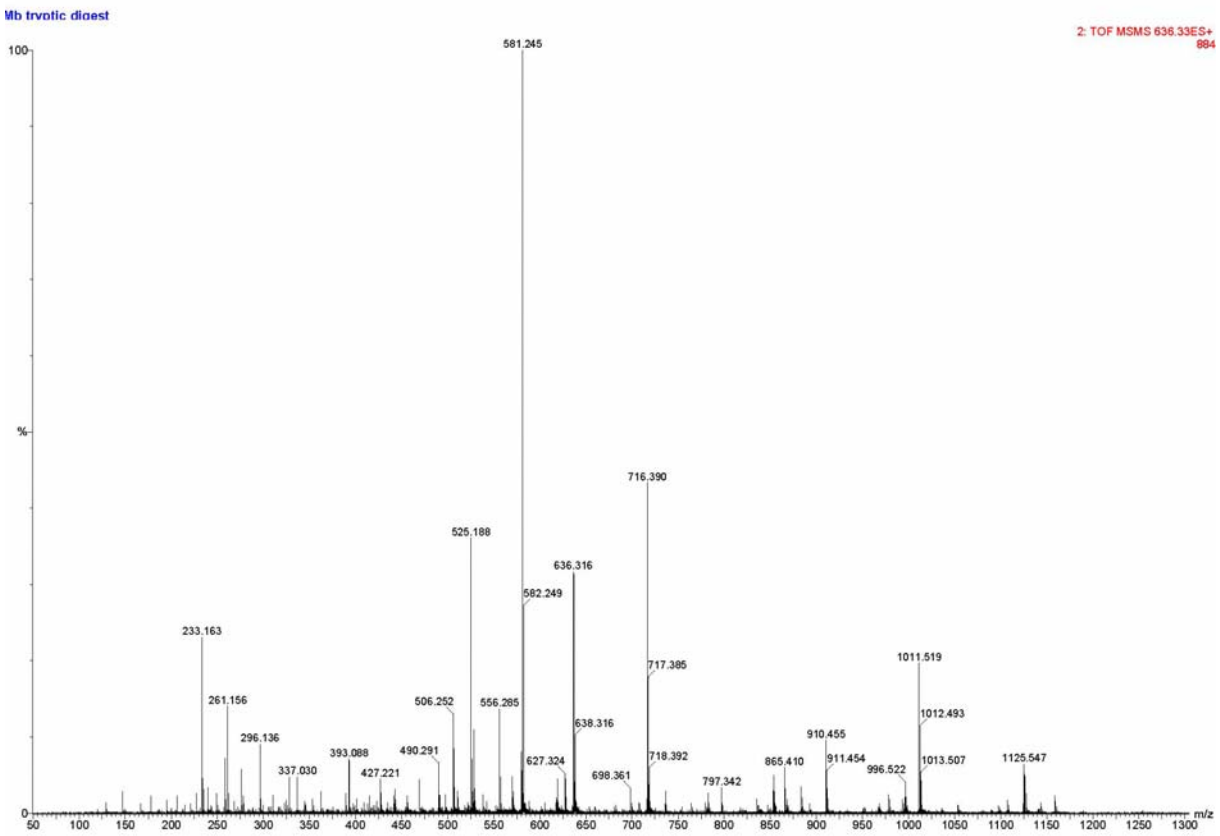
Here is an MS/MS q-TOF spectrum for you to test your de novo skills on. The first figure is the full fragment spectrum, the spectrum is then divided into four sections to make the details easier for you to see. Start with the high end of the spectrum, Figure 2. We have made a pdf file for you so that you can download the spectra and print them out so that you can get a pencil and a calculator and start sequencing. To download the spectra, right click on the link and save the pdf document to your computer.

Once you have done your best you can proceed onto the answer page to check out the known sequence and see some of the issues we ran into with this spectrum.

Here are several hints

1. The parent ion is m/z 636.33 which is a doubly charged ion, $[M+2H]^{2+}$.
2. This is a tryptic peptide ending in K or R
3. Since this is q-TOF data start by following the y ion series, try to start at the high mass end.
4. Another note is, **y** ions dominate this set of q-TOF data.
5. Mass accuracy will be on your side, mass accuracy should be better than 30 mmu, that's 0.030u.

Figure 1.



http://ionsource.com/tutorial/DeNovo/introduction.htm

Figure 2.

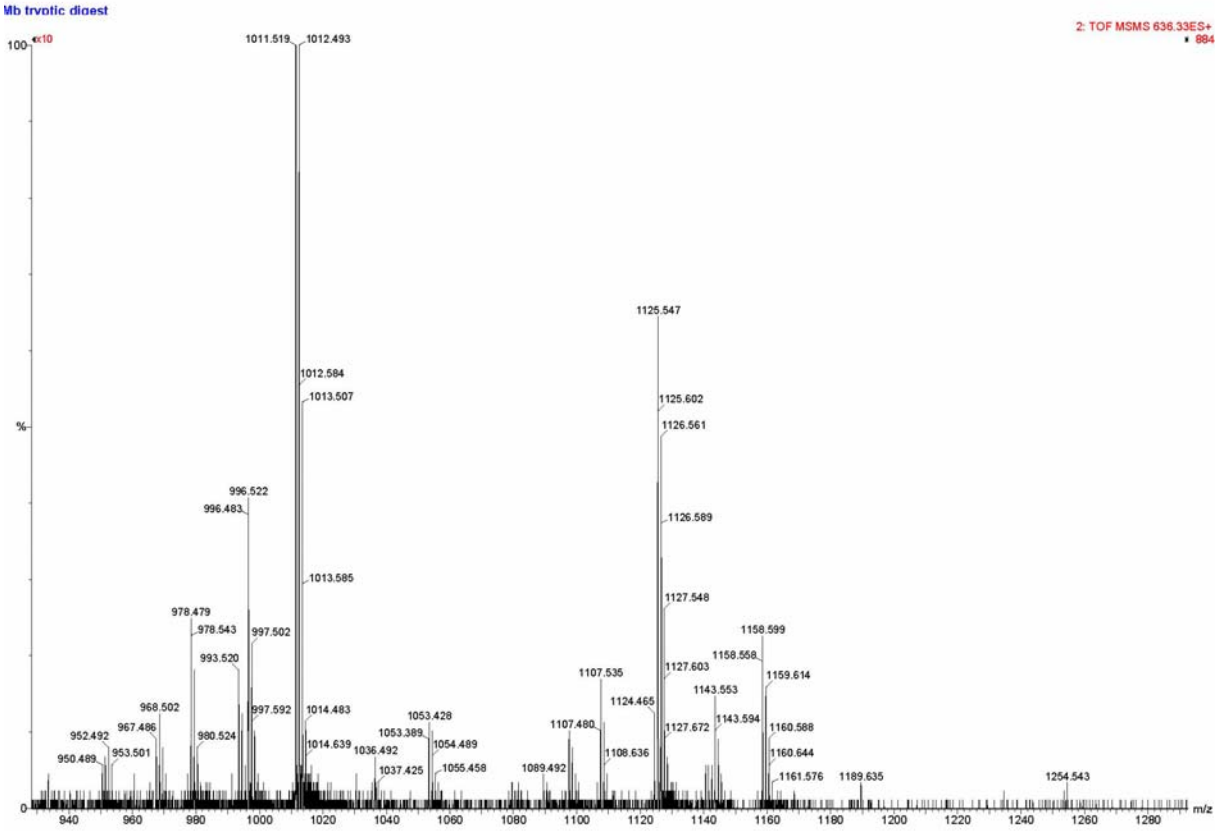
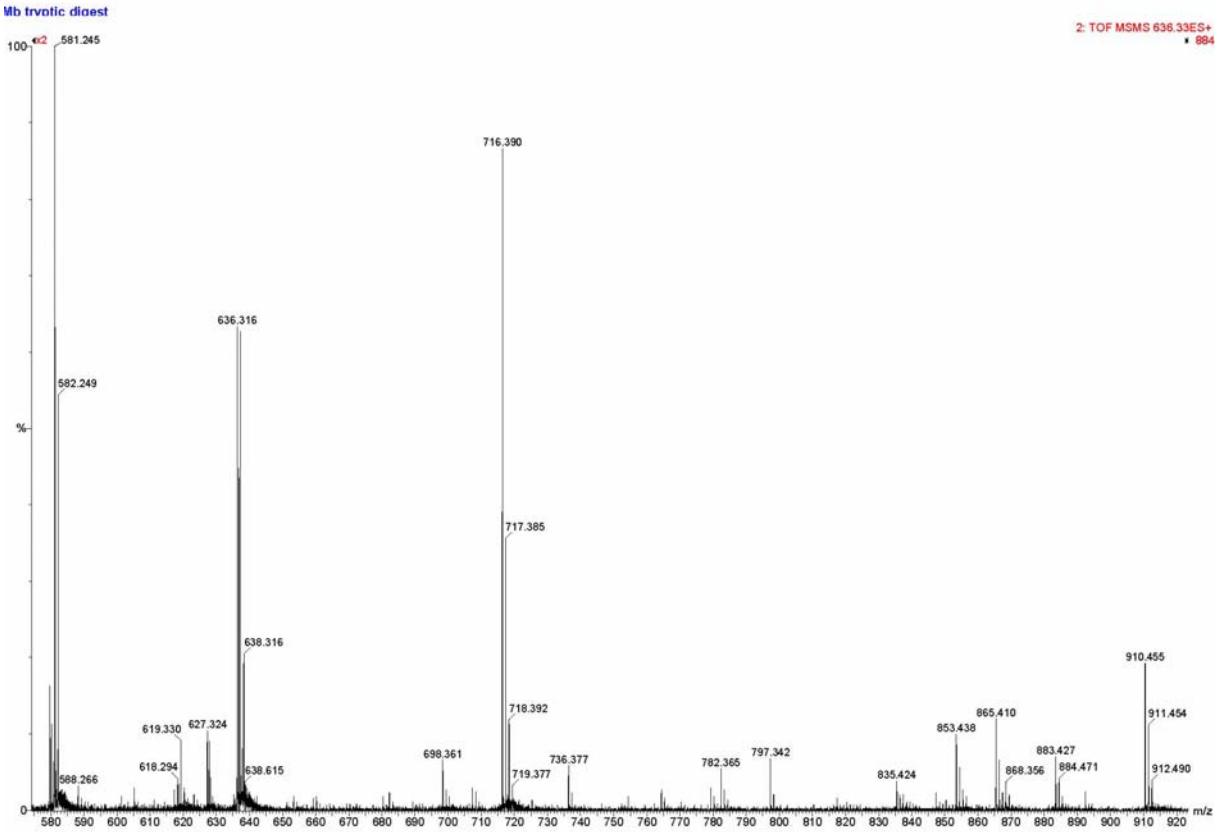


Figure 3.

.



http://ionsource.com/tutorial/DeNovo/introduction.htm

Figure 4.

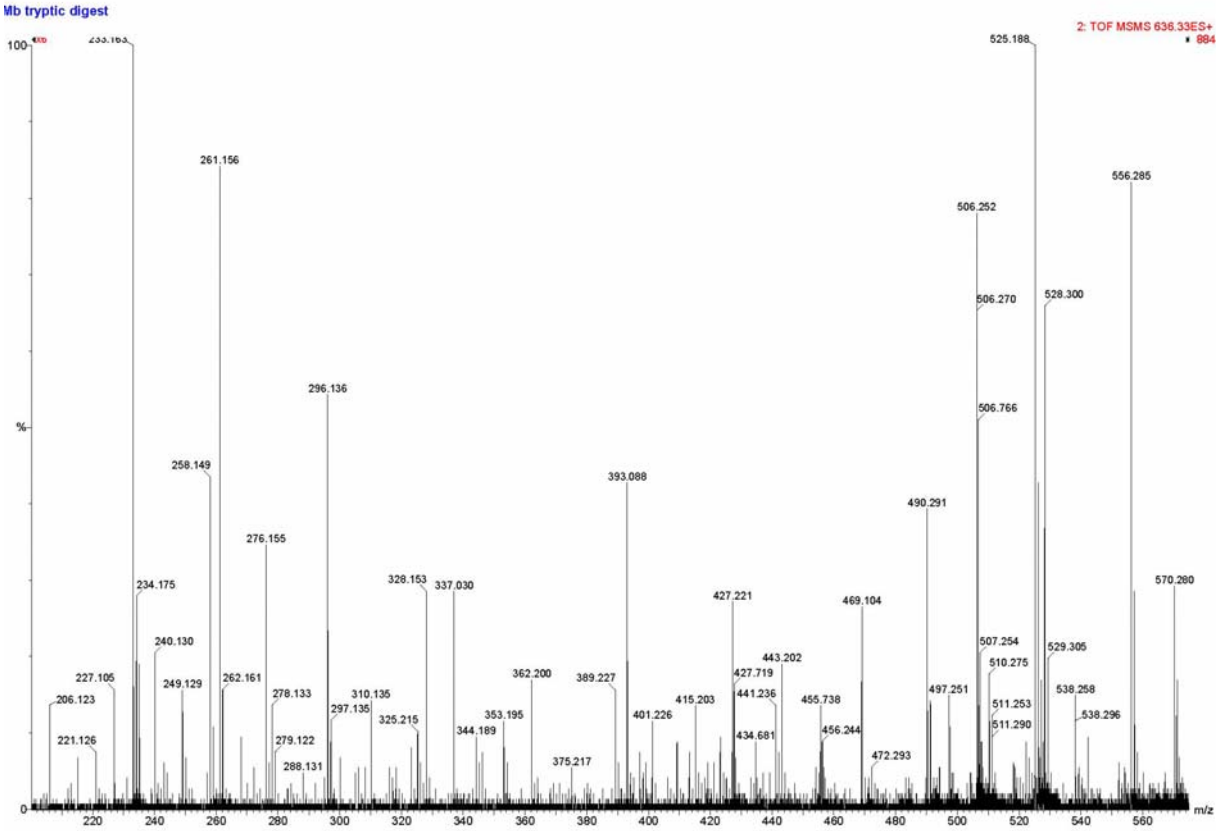
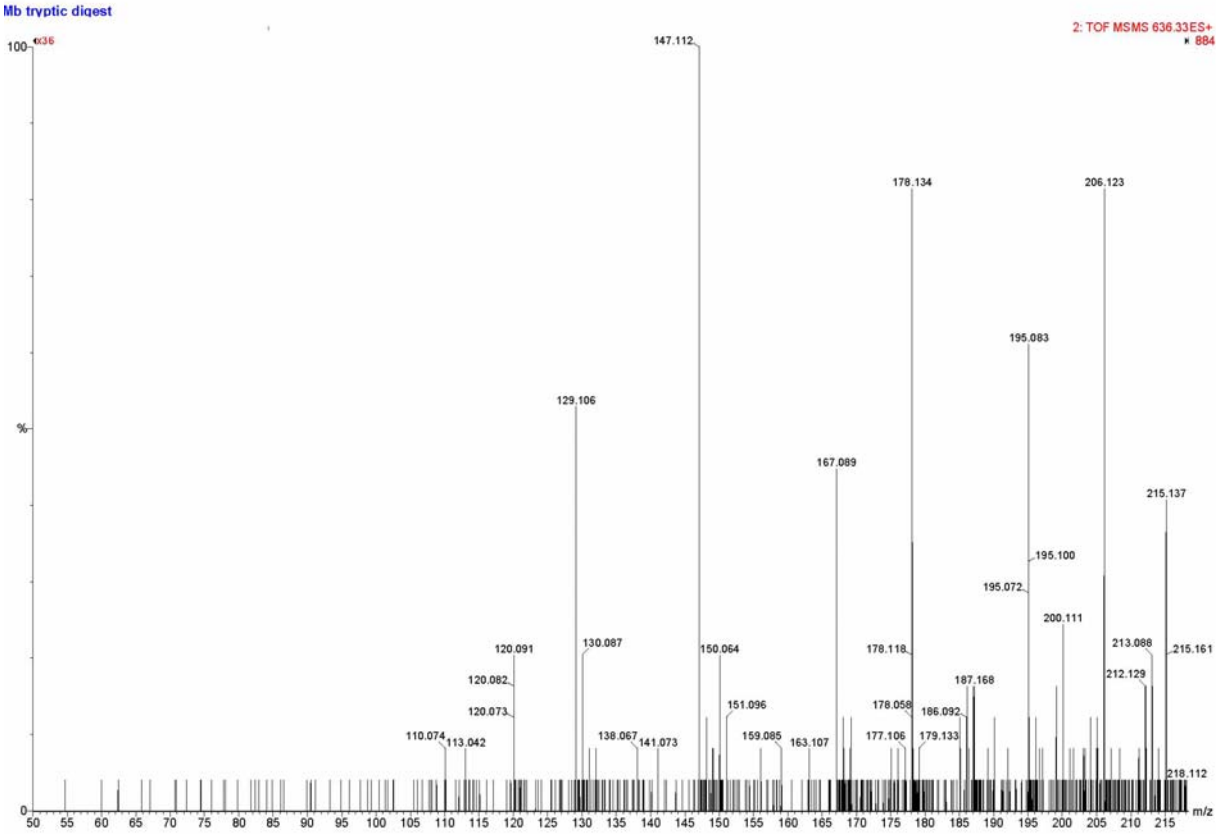


Figure 5.



2nd De Novo Exercise the q-TOF answer page

The correct sequence for this peptide is, LFTGHPETLEK

The first thing that we did was to look at the low end of the spectrum for low mass clues. We saw the immonium ions for H, F, and K, which let us know that we may encounter these amino acids in the sequence.

We did the math to determine the next to last **y** ion.

1271.66 - AA residue mass = penultimate **y** ion

We then proceeded to look for the next amino acid residue and labeled the peaks. We then did the calculation to determine the corresponding **b** ion and looked for **a** ions to confirm. With this data set, as long as you start with the **y** ion series the going is pretty straightforward. The mass accuracy of the q-TOF helps because you know that any amino acid call should be within 30 mmu.

If you tried to follow the **b** ion series your going would be tougher, because there are gaps that would force you to make several double amino acid jumps, and that's pretty hard, but still easier than with ion trap or triple quad data.. We don't have a lot of experience with q-TOF data, but it appears that the **y** ion series is dominant. Luckily we were able to determine the **b1** ion which was a good surprise.

Here is the correct sequence

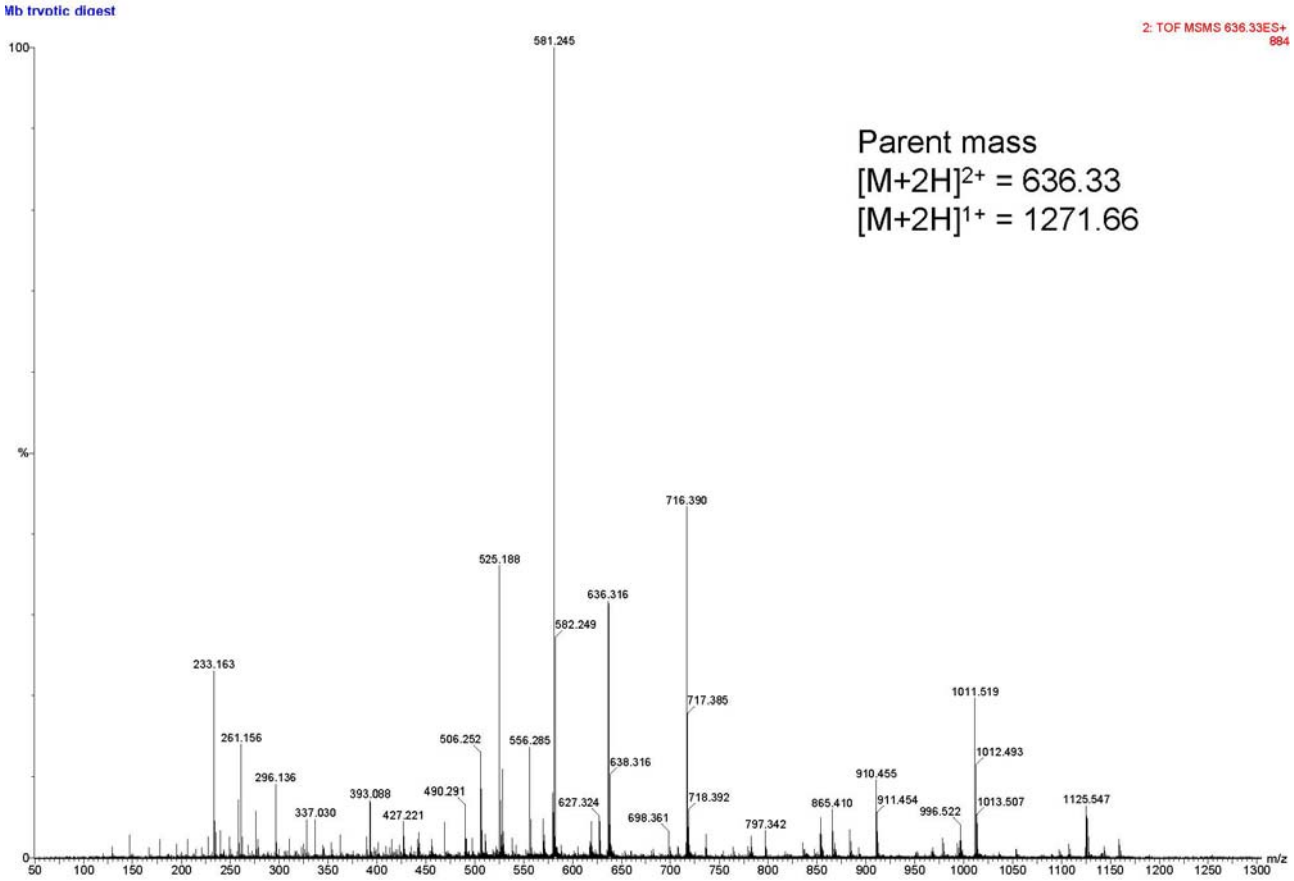
LFTGHPETLEK

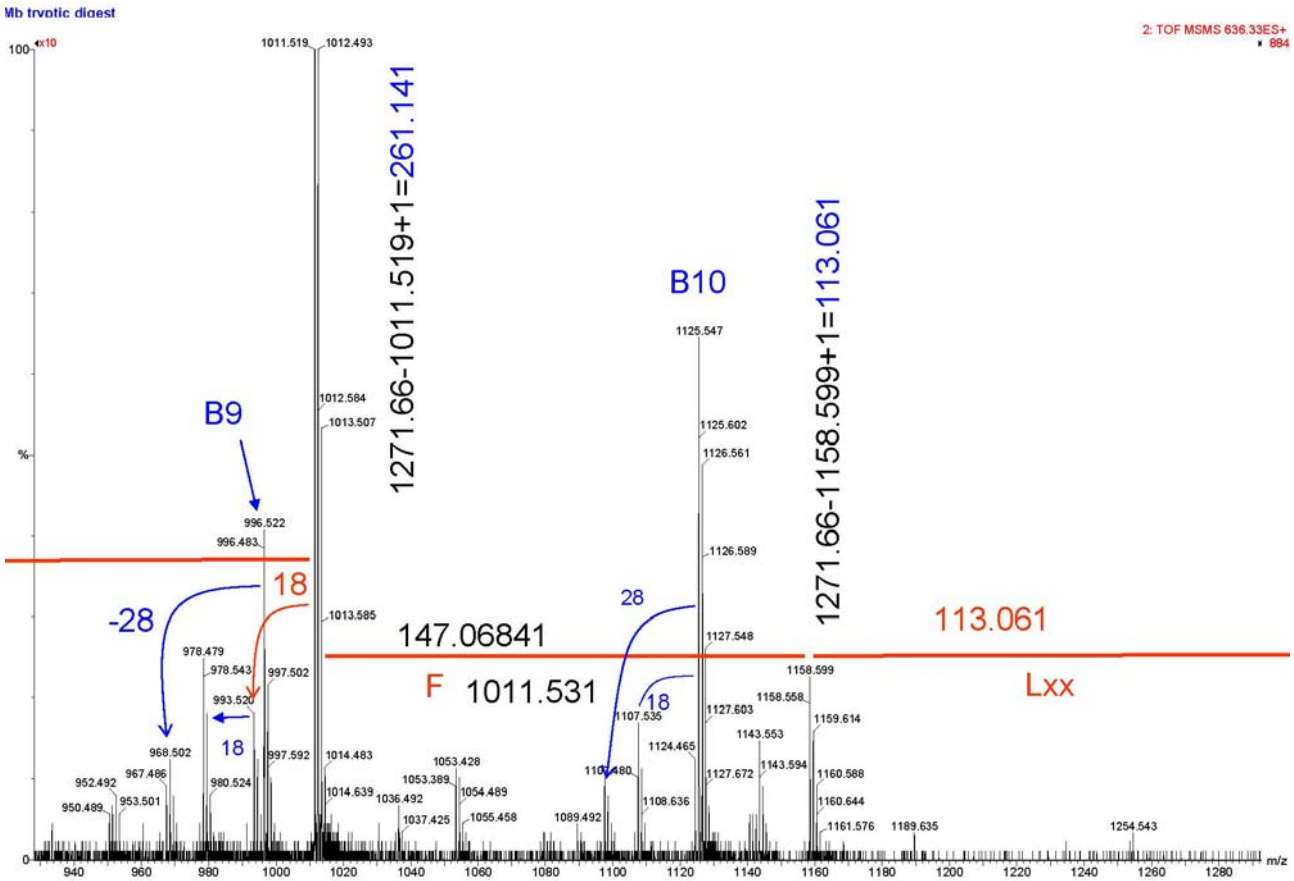
Here is the sequence that we determined from the **y** ion series

XFTGHPETXEK

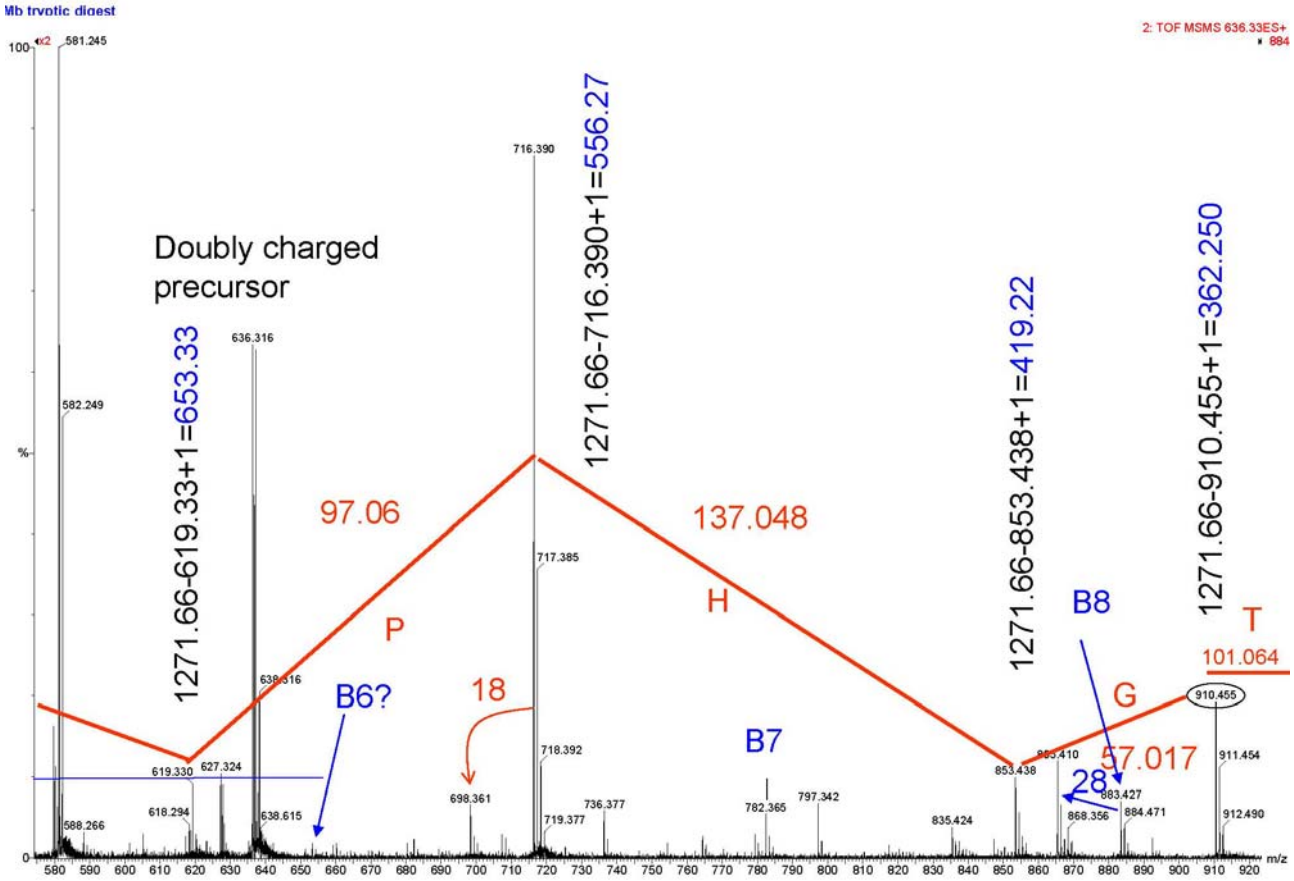
We determined the entire sequence, yippee! Hopefully you got it right too!. If you would like to look at our scribblings,

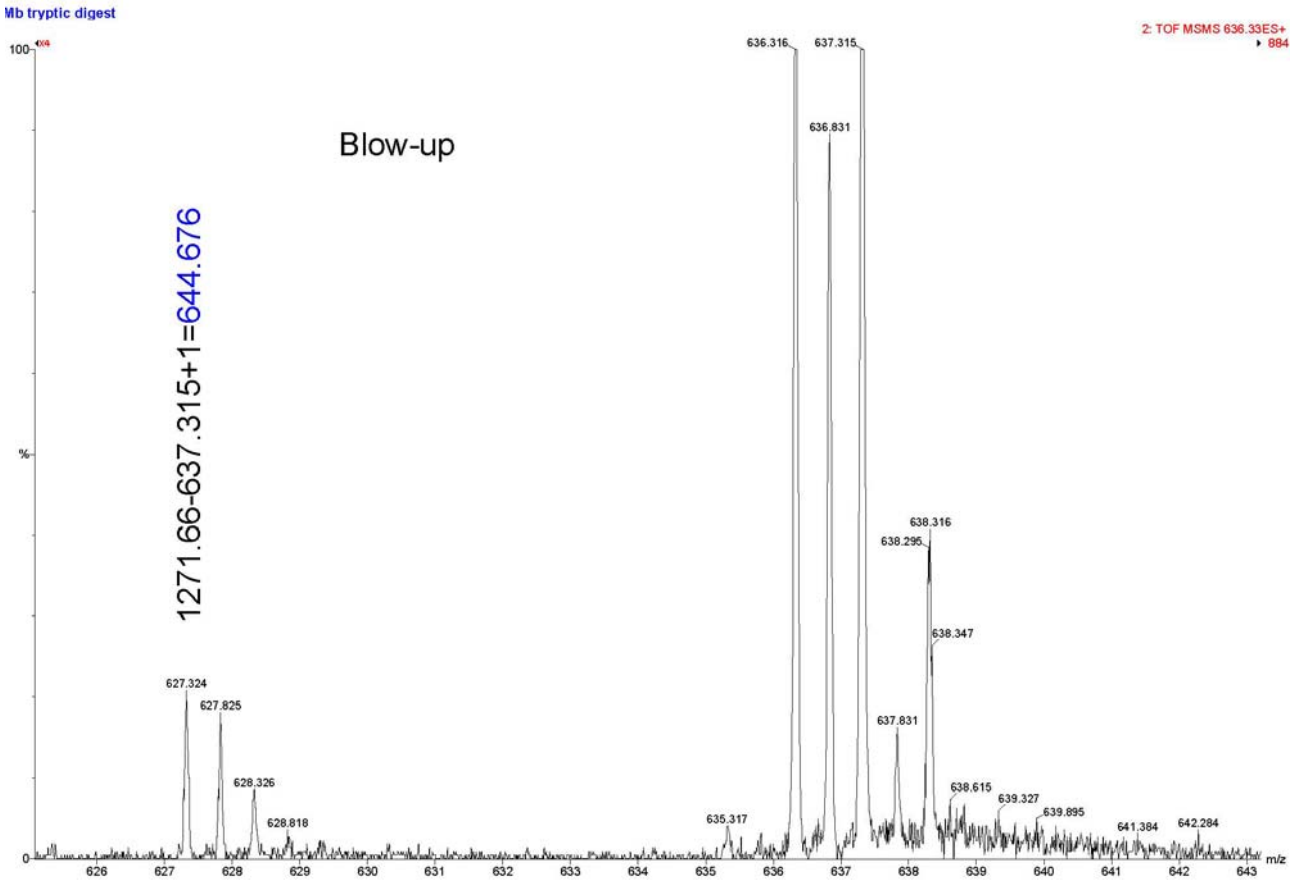
http://ionsource.com/tutorial/DeNovo/introduction.htm



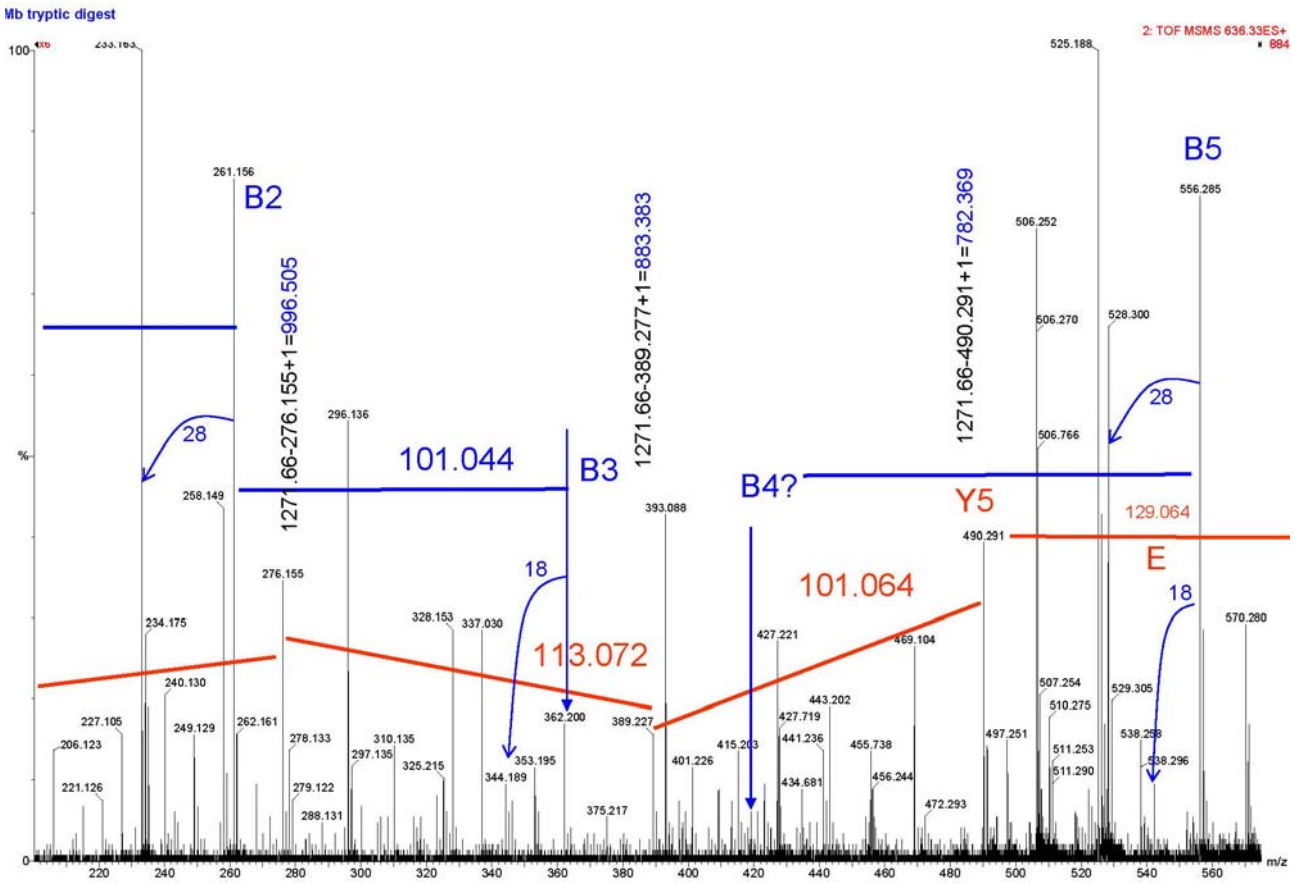


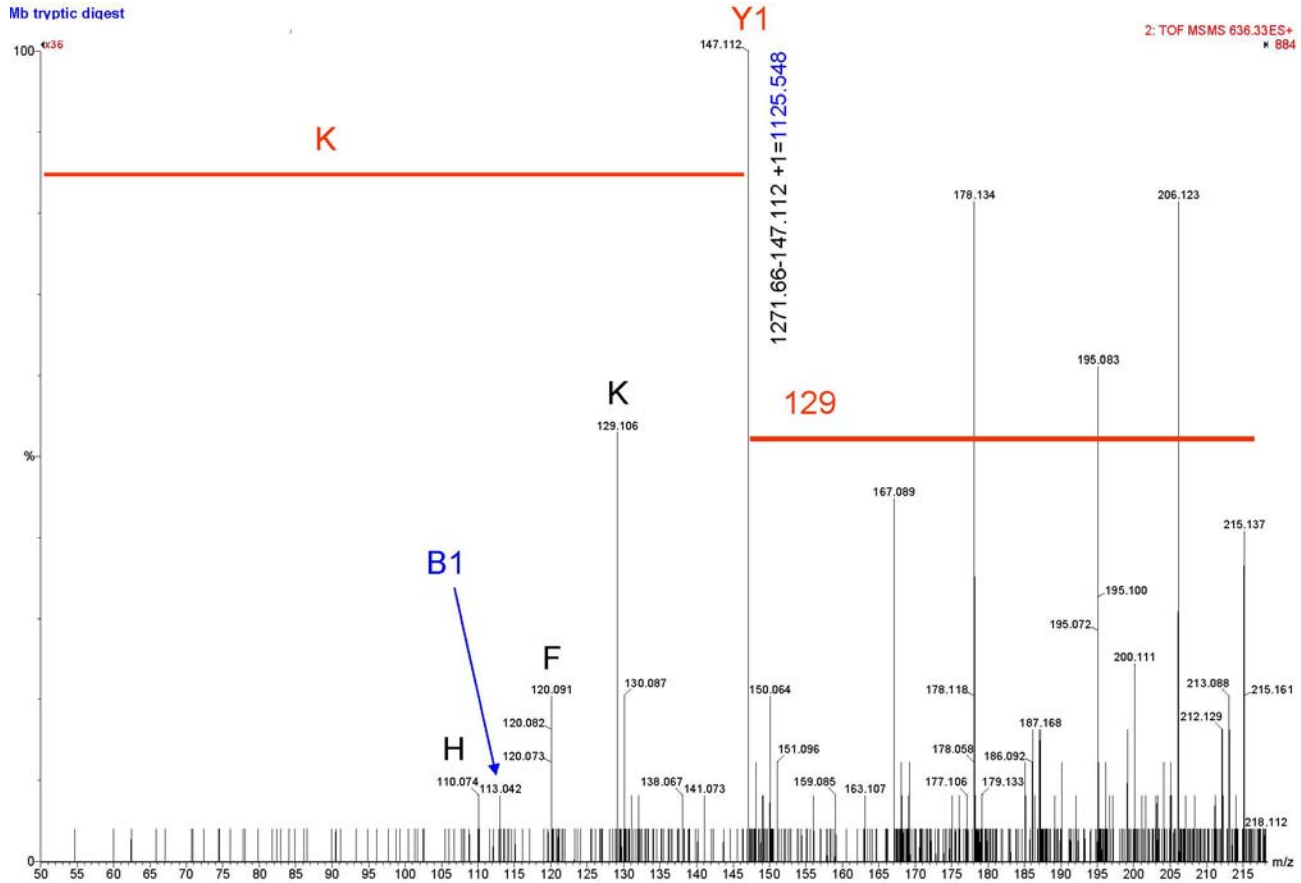
http://ionsource.com/tutorial/DeNovo/introduction.htm





http://ionsource.com/tutorial/DeNovo/introduction.htm





De Novo Tables

Common Amino Acids

AA Codes		Mono.	AA Codes		Mono.
Gly	G	**57**.021464	Asp	D	**115**.02694
Ala	A	**71**.037114	Gln	Q	**128**.05858
Ser	S	**87**.032029	Lys	K	**128**.09496
Pro	P	**97**.052764	Glu	E	**129**.04259
Val	V	**99**.068414	Met	M	**131**.04048
Thr	T	**101**.04768	His	H	**137**.05891
Cys	C	**103**.00919	Phe	F	**147**.06841
Leu	L	**113**.08406	Arg	R	**156**.10111
Ile	I	**113**.08406	CMC		**161**.01467
Asn	N	**114**.04293	Tyr	Y	**163**.06333
IonSource sm mass spectrometry educational resource			Trp	W	**186**.07931

Conflicting Masses

Amino Acids	Mono		Mono	delta mass
Val	**99**.068414	acetyl-Gly	**99**.032034	0.03638
Leu	**113**.08406	Ile	**113**.08406	0
Lxx	**113**.08406	acetyl-Ala	**113**.047684	0.036376
Asn	**114**.04293	Gly-Gly	**114**.04298	0.000002
Gln	**128**.05858	Lys	**128**.09496	0.03638
Gln	**128**.05858	Gly-Ala	**128**.058578	0.000002
Lys	**128**.09496	Gly-Ala	**128**.058578	0.036382
Glu	**129**.04259	acetyl-Ser	**129**.042599	0.000009
Phe	**147**.06841	Met Sulfoxide	**147**.0354	0.033
Arg	**156**.10111	Val-Gly	**156**.089878	0.011232
Arg	**156**.10111	acetyl-Asn	**156**.0535	0.04761
Tyr	**163**.06333	Met Sulfone	**163**.0303	0.033
Trp	**186**.07931	Ala-Asp	**186**.064054	0.015256
Trp	**186**.07931	Ser-Val	**186**.100443	0.021133
Trp	**186**.07931	Gly-Glu	**186**.064054	0.015256

http://ionsource.com/tutorial/DeNovo/introduction.htm

Di-peptide Table

Note: Where di-peptides conflict with single amino acid the masses are noted in blue.
Masses in red are notations where di-peptides conflict with other di-peptides.

		G	A	S	P	V	T	C	L	I	N	D	Q	K	E	M	H	F	R	CMC	Y	W
		57	**71**	**87**	**97**	**99**	**101**	**103**	**113**	**113**	**114**	**115**	**128**	**128**	**129**	**131**	**137**	**147**	**156**	**161**	**163**	**186**
G	**57**	**114**	128	144	154	156	158	160	170	170	171	172	185	185	186	188	194	204	213	218	220	243
A	**71**	**128**	142	158	168	170	172	174	184	184	185	186	199	199	200	202	208	218	227	232	234	257
S	**87**	144	158	174	184	186	188	190	200	200	201	202	215	215	216	218	224	234	243	248	250	273
P	**97**	154	168	184	194	196	198	200	210	210	211	212	225	225	226	228	234	244	253	258	260	283
V	**99**	**156**	170	**186**	196	198	200	202	212	212	213	214	227	227	228	230	236	246	255	260	262	285
T	**101**	158	172	188	198	200	202	204	214	214	215	216	229	229	230	232	238	248	257	262	264	287
C	**103**	160	174	190	200	202	204	206	216	216	217	218	231	231	232	234	240	250	259	264	266	289
L	**113**	170	184	200	210	212	214	216	226	226	227	228	241	241	242	244	250	260	269	274	276	299
I	**113**	170	184	200	210	212	214	216	226	226	227	228	241	241	242	244	250	260	269	274	276	299
N	**114**	171	185	201	211	213	215	217	227	227	228	229	242	242	243	245	251	261	270	275	277	300
D	**115**	172	**186**	202	212	214	216	218	228	228	229	230	243	243	244	246	252	262	271	276	278	301
Q	**128**	185	199	215	225	227	229	231	241	241	242	243	256	256	257	259	265	275	284	289	291	314
K	**128**	185	199	215	225	227	229	231	241	241	242	243	256	256	257	259	265	275	284	289	291	314
E	**129**	**186**	200	216	226	228	230	232	242	242	243	244	257	257	258	260	266	276	285	290	292	315
M	**131**	188	202	218	228	230	232	234	244	244	245	246	259	259	260	262	268	278	287	292	294	317
H	**137**	194	208	224	234	236	238	240	250	250	251	252	265	265	266	268	274	284	293	298	300	323
F	**147**	204	218	234	244	246	248	250	260	260	261	262	275	275	276	278	284	294	303	308	310	333
R	**156**	213	227	243	253	255	257	259	269	269	270	271	284	284	285	287	293	303	312	317	319	342
CMC	**161**	218	232	248	258	260	262	264	274	274	275	276	289	289	290	292	298	308	317	322	324	347
Y	**163**	220	234	250	260	262	264	266	276	276	277	278	291	291	292	294	300	310	319	324	326	349
W	**186**	243	257	273	283	285	287	289	299	299	300	301	314	314	315	317	323	333	342	347	349	372

http://ionsource.com/tutorial/DeNovo/introduction.htm

Di-peptide Table, Out to Four Decimal Places

		G	A	S	P	V	T	C	L	I	N	D	Q	K	E	M	H	F	R	CMC	Y	W
		57.0215	71.0371	87.0320	97.0528	99.0684	101.0477	103.0092	113.0841	113.0841	114.0429	115.0269	128.0586	128.0950	129.0426	131.0405	137.0589	147.0684	156.1011	161.0147	163.0633	186.0793
G	57.0215	114.0429	128.0586	144.0535	154.0742	156.0899	158.0692	160.0307	170.1055	170.1055	171.0644	172.0484	185.0800	185.1164	186.0641	188.0619	194.0804	204.0899	213.1226	218.0361	220.0848	243.1008
A	71.0371	128.0586	142.0742	158.0691	168.0899	170.1055	172.0848	174.0463	184.1212	184.1212	185.0800	186.0641	199.0957	199.1321	200.0797	202.0776	208.0960	218.1055	227.1382	232.0518	234.1004	257.1164
S	87.0320	144.0535	158.0691	174.0641	184.0848	186.1004	188.0797	190.0412	200.1161	200.1161	201.0750	202.0590	215.0906	215.1270	216.0746	218.0725	224.0909	234.1004	243.1331	248.0467	250.0954	273.1113
P	97.0528	154.0742	168.0899	184.0848	194.1055	196.1212	198.1005	200.0620	210.1368	210.1368	211.0957	212.0797	225.1113	225.1477	226.0954	228.0932	234.1117	244.1212	253.1539	258.0674	260.1161	283.1321
V	99.0684	156.0899	170.1055	186.1004	196.1212	198.1368	200.1161	202.0776	212.1525	212.1525	213.1113	214.0954	227.1270	227.1634	228.1110	230.1089	236.1273	246.1368	255.1695	260.0831	262.1317	285.1477
T	101.0477	158.0691	172.0848	188.0797	198.1004	200.1161	202.0954	204.0569	214.1317	214.1317	215.0906	216.0746	229.1063	229.1426	230.0903	232.0882	238.1066	248.1161	257.1488	262.0624	264.1110	287.1270
C	103.0092	160.0307	174.0463	190.0412	200.0620	202.0776	204.0569	206.0184	216.0933	216.0933	217.0521	218.0361	231.0678	231.1042	232.0518	234.0497	240.0681	250.0776	259.1103	264.0239	266.0725	289.0885
L	113.0841	170.1055	184.1212	200.1161	210.1368	212.1525	214.1318	216.0933	226.1681	226.1681	227.1270	228.1110	241.1426	241.1790	242.1267	244.1245	250.1430	260.1525	269.1852	274.0987	276.1474	299.1634
I	113.0841	170.1055	184.1212	200.1161	210.1368	212.1525	214.1318	216.0933	226.1681	226.1681	227.1270	228.1110	241.1426	241.1790	242.1267	244.1245	250.1430	260.1525	269.1852	274.0987	276.1474	299.1634
N	114.0429	171.0644	185.0800	201.0750	211.0957	213.1113	215.0906	217.0521	227.1270	227.1270	228.0859	229.0699	242.1015	242.1379	243.0855	245.0834	251.1018	261.1113	270.1440	275.0576	277.1063	300.1222
D	115.0269	172.0484	186.0641	202.0590	212.0797	214.0954	216.0746	218.0361	228.1110	228.1110	229.0699	230.0539	243.0855	243.1219	244.0695	246.0674	252.0859	262.0954	271.1281	276.0416	278.0903	301.1063
Q	128.0586	185.0800	199.0957	215.0906	225.1113	227.1270	229.1063	231.0678	241.1426	241.1426	242.1015	243.0855	256.1172	256.1535	257.1012	259.0991	265.1175	275.1270	284.1597	289.0733	291.1219	314.1379
K	128.0950	185.1164	199.1321	215.1270	225.1477	227.1634	229.1427	231.1042	241.1790	241.1790	242.1379	243.1219	256.1535	256.1899	257.1376	259.1354	265.1539	275.1634	284.1961	289.1096	291.1583	314.1743
E	129.0426	186.0641	200.0797	216.0746	226.0954	228.1110	230.0903	232.0518	242.1267	242.1267	243.0855	244.0695	257.1012	257.1376	258.0852	260.0831	266.1015	276.1110	285.1437	290.0573	292.1059	315.1219
M	131.0405	188.0619	202.0776	218.0725	228.0932	230.1089	232.0882	234.0497	244.1245	244.1245	245.0834	246.0674	259.0991	259.1354	260.0831	262.0810	268.0994	278.1089	287.1416	292.0552	294.1038	317.1198
H	137.0589	194.0804	208.0960	224.0909	234.1117	236.1273	238.1066	240.0681	250.1430	250.1430	251.1018	252.0859	265.1175	265.1539	266.1015	268.0994	274.1178	284.1273	293.1600	298.0736	300.1222	323.1382
F	147.0684	204.0899	218.1055	234.1004	244.1212	246.1368	248.1161	250.0776	260.1525	260.1525	261.1113	262.0954	275.1270	275.1634	276.1110	278.1089	284.1273	294.1368	303.1695	308.0831	310.1317	333.1477
R	156.1011	213.1226	227.1382	243.1331	253.1539	255.1695	257.1488	259.1103	269.1852	269.1852	270.1440	271.1281	284.1597	284.1961	285.1437	287.1416	293.1600	303.1695	312.2022	317.1158	319.1644	342.1804
CMC	161.0147	218.0361	232.0518	248.0467	258.0674	260.0831	262.0624	264.0239	274.0987	274.0987	275.0576	276.0416	289.0733	289.1096	290.0573	292.0552	298.0736	308.0831	317.1158	322.0293	324.0780	347.0940
Y	163.0633	220.0848	234.1004	250.0954	260.1161	262.1317	264.1110	266.0725	276.1474	276.1474	277.1063	278.0903	291.1219	291.1583	292.1059	294.1038	300.1222	310.1317	319.1644	324.0780	326.1267	349.1426
W	186.0793	243.1008	257.1164	273.1113	283.1321	285.1477	287.1270	289.0885	299.1634	299.1634	300.1222	301.1063	314.1379	314.1743	315.1219	317.1198	323.1382	333.1477	342.1804	347.0940	349.1426	372.1586

References and Additional Reading

Denovo Analysis:

1. Hunt DF, Yates JR 3rd, Shabanowitz J, Winston S, Hauer CR. **Protein sequencing by tandem mass spectrometry**. *Proc Natl Acad Sci*. 1986 Sep;83(17):6233-7.
2. Papayannopoulos, IA, **The interpretation of collision-induced dissociation tandem mass spectra of peptides.** *Mass Spectrom. Rev.*, 14(1) 49-73 (1995).
3. Martin SE, Shabanowitz J, Hunt DF, Marto JA. **Subfemtomole MS and MS/MS peptide sequence analysis using nano-HPLC micro-ESI fourier transform ion cyclotron resonance mass spectrometry.** *Anal Chem.* 2000 Sep 15;72(18):4266-74.
4. Schroeder MJ, Shabanowitz J, Schwartz JC, Hunt DF, Coon JJ. **A neutral loss activation method for improved phosphopeptide sequence analysis by quadrupole ion trap mass spectrometry.** *Anal Chem.* 2004 Jul 1;76(13):3590-8.
5. Syka JE, Coon JJ, Schroeder MJ, Shabanowitz J, Hunt DF.**Peptide and protein sequence analysis by electron transfer dissociation mass spectrometry.** *Proc Natl Acad Sci* U S A. 2004 Jun 29;101(26):9528-33. Epub 2004 Jun 21.
6. Deutzmann R. S**tructural characterization of proteins and peptides.** *Methods Mol Med.* 2004;94:269-97.
7. Medzihradszky KF **Peptide sequence analysis.** *Methods Enzymol.* 2005;402:209-44.
8. Coon JJ, Syka JE, Shabanowitz J, Hunt DF. **Tandem mass spectrometry for peptide and protein sequence analysis.** *Biotechniques.* 2005 Apr;38(4):519, 521, 523.

Sequence Tag:

1. Mann M, Wilm M., **Error-tolerant identification of peptides in sequence databases by peptide sequence tags.** *Anal Chem.* 1994 Dec 15;66(24):4390-9.
2. Mortz E, O'Connor PB, Roepstorff P, Kelleher NL, Wood TD, McLafferty FW, Mann M. **Sequence tag identification of intact proteins by matching tanden mass spectral data against sequence data bases.** *Proc Natl Acad Sci* U S A. 1996 Aug 6;93(16):8264-7
3. Sunyaev S, Liska AJ, Golod A, Shevchenko A, Shevchenko A. **MultiTag: multiple error-tolerant sequence tag search for the sequence-similarity identification of proteins by mass spectrometry.** *Anal Chem.* 2003 Mar 15;75(6):1307-15.

Software:

1. Horn DM, Zubarev RA, McLafferty FW.**Automated de novo sequencing of proteins by tandem high-resolution mass spectrometry.** *Proc Natl Acad Sci* 2000 Sep 12;97(19):10313-7.
2. Bruni R, Gianfranceschi G, Koch G. **On peptide de novo sequencing: a new approach.** *J Pept Sci.* 2005 Apr;11(4):225-34.
3. Han Y, Ma B, Zhang K. **SPIDER: software for protein identification from sequence tags with de novo sequencing error.** *J Bioinform Comput Biol.* 2005 Jun;3(3):697-716.
4. Savitski MM, Nielsen ML, Kjeldsen F, Zubarev RA. **Proteomics-grade de novo sequencing approach.** *J Proteome Res.* 2005 Nov-Dec;4(6):2348-54.
5. Zhong H, Li L. **An algorithm for interpretation of low-energy collision-induced dissociation product ion spectra for de novo sequencing of peptides.** *Rapid Commun Mass Spectrom.* 2005;19(8):1084-96.

6. Zhong H, Li L. **An algorithm for interpretation of low-energy collision-induced dissociation product ion spectra for de novo sequencing of peptides.** *Rapid Commun Mass Spectrom.* 2005;19(8):1084-96.
7. Searle BC, Dasari S, Wilmarth PA, Turner M, Reddy AP, David LL, Nagalla SR. **Identification of protein modifications using MS/MS de novo sequencing and the OpenSea alignment algorithm.** *J Proteome Res.* 2005 Mar-Apr;4(2):546-54.

Peptide Fragmentation Details:

1. Roepstorff P, Fohlman J. **Proposal for a common nomenclature for sequence ions in mass spectra of peptides.** *Biomed Mass Spectrom.* 1984 Nov;11(11):601.
2. Johnson RS, Martin SA, Biemann K, Stults JT, Watson JT. **Novel fragmentation process of peptides by collision-induced decomposition in a tandem mass spectrometer: differentiation of leucine and isoleucine.** *Anal Chem.* 1987 Nov 1;59(21):2621-5.
3. Biemann K. **Contributions of mass spectrometry to peptide and protein structure.** *Biomed Environ Mass Spectrom.* 1988 Oct;16(1-12):99-111.
4. K. Biemann In: J.A. McCloskey, Editor, *Methods in Enzymology* 193, 1990, pp. 886–887.
5. Falick, AM, Hines, WM, Medzihradszky, KF, Baldwin, MA and Gibson, BW, **Low-mass ions produced from peptides by high-energy collision-induced dissociation in tandem mass spectrometry.** *J. Am. Soc. Mass Spectrom.*, 4(11) 882-93 (1993).
6. Tang XJ, Thibault P, Boyd RK. **Fragmentation reactions of multiply-protonated peptides and implications for sequencing by tandem mass spectrometry with low-energy collision-induced dissociation.** *Anal Chem.* 1993 Oct 15;65(20):2824-34.
7. Patterson SD, Katta V. **Prompt fragmentation of disulfide-linked peptides during matrix-assisted laser desorption ionization mass spectrometry.** *Anal Chem.* 1994 Nov 1;66(21):3727-32.
8. Cramer R, Corless S. **The nature of collision-induced dissociation processes of doubly protonated peptides: comparative study for the future use of matrix-assisted laser desorption/ionization on a hybrid quadrupole time-of-flight mass spectrometer in proteomics.** *Rapid Commun Mass Spectrom.* 2001;15(22):2058-66.
9. Sadagopan N, Watson JT. **Mass spectrometric evidence for mechanisms of fragmentation of charge-derivatized peptides.** *J Am Soc Mass Spectrom.* 2001 Apr;12(4):399-409.
10. Samyn B, Debyser G, Sergeant K, Devreese B, Van Beeumen J. **A case study of de novo sequence analysis of N-sulfonated peptides by MALDI TOF/TOF mass spectrometry.** *J Am Soc Mass Spectrom.* 2004 Dec;15(12):1838-52.

De novo Approaches:

1. Schnolzer M, Jedrzejewski P, Lehmann WD. **Protease-catalyzed incorporation of 18O into peptide fragments and its application for protein sequencing by electrospray and matrix-assisted laser desorption/ionization mass spectrometry.** *Electrophoresis* 1996 May;17(5):945-53.
2. Keough T, Lacey MP, Youngquist RS. **Derivatization procedures to facilitate de novo sequencing of lysine-terminated tryptic peptides using postsource decay matrix-assisted laser desorption/ionization mass spectrometry.** *Rapid Commun Mass Spectrom.* 2000;14(24):2348-56.
3. Yergey AL, Coorssen JR, Backlund PS Jr, Blank PS, Humphrey GA, Zimmerberg J, Campbell JM, Vestal ML.**De novo sequencing of peptides using MALDI/TOF-TOF.** J *Am Soc Mass Spectrom.* 2002 Jul;13(7):784-91.
4. Heller M, Mattou H, Menzel C, Yao X. **Trypsin catalyzed 16O-to-18O exchange for comparative proteomics: tandem mass spectrometry comparison using MALDI-TOF,**

ESI-QTOF, and ESI-ion trap mass spectrometers. *J Am Soc Mass Spectrom.* 2003 Jul;14(7):704-18.

5. Dancik V, Addona TA, Clauser KR, Vath JE, Pevzner PA. **De novo peptide sequencing via tandem mass spectrometry.** *J Comput Biol.* 1999 Fall-Winter;6(3-4):327-42.
6. Heredia-Langner A, Cannon WR, Jarman KD, Jarman KH. **Sequence optimization as an alternative to de novo analysis of tandem mass spectrometry data.** *Bioinformatics* 2004 Sep 22;20(14):2296-304. Epub 2004 Apr 15.
7. Olsen JV, Mann M. **Improved peptide identification in proteomics by two consecutive stages of mass spectrometric fragmentation.** *Proc Natl Acad Sci* 2004 Sep 14;101(37):13417-22. Epub 2004 Sep 3.
8. Zhang Z. **De novo peptide sequencing based on a divide-and-conquer algorithm and peptide tandem spectrum simulation.** *Anal Chem.* 2004 Nov 1;76(21):6374-83.
9. Spengler B. **De novo sequencing, peptide composition analysis, and composition-based sequencing: a new strategy employing accurate mass determination by fourier transform ion cyclotron resonance mass spectrometry.** *J Am Soc Mass Spectrom.* 2004 May;15(5):703-14.

De Novo Combined with Database Searching

1. Sheng QH, Xie T, Ding DF. **De Novo Interpretation of MS/MS Spectra and Protein Identification via Database Searching.** Sheng Wu Hua Xue Yu Sheng Wu Wu Li Xue Bao (Shanghai). 2000;32(6):595-600.
2. Johnson RS, Taylor JA. **Searching sequence databases via de novo peptide sequencing by tandem mass spectrometry.** *Mol Biotechnol.* 2002 Nov;22(3):301-15.
3. Savitski MM, Nielsen ML, Zubarev RA. **New data base-independent, sequence tag-based scoring of peptide MS/MS data validates Mowse scores, recovers below threshold data, singles out modified peptides, and assesses the quality of MS/MS techniques.** *Mol Cell Proteomics* 2005 Aug;4(8):1180-8. Epub 2005 May 22.
4. Halligan BD, Ruotti V, Twigger SN, Greene AS. **DeNovoID: a web-based tool for identifying peptides from sequence and mass tags deduced from de novo peptide sequencing by mass spectroscopy.** *Nucleic Acids Res.* 2005 Jul 1;33(Web Server issue):W376-81.