FROM FOSSILS TO PHYLOGENIES PART 1: MASS SPECTROMETRY

[Written by: Baylee Goodwin, Dane Besser, Stephen Ramsey]

Dinosaur Vocabulary Words	**Chemistry Vocabulary Words**	**Biology Vocabulary Words**
Fossilization	Mass spectrometry	Central Dogma of Biology
T. rex	Mass-to-charge ratio	Proteins
Mastodon	Ion	Peptide
Hadrosaur	Relative abundance	Protein
		Collagen

Background:

Imagine it is time for your lunch break; you take your sandwich outside and you sit down to enjoy your lunch with a beautiful view of Montana's Rocky Mountains. As you look up, you see what appears to be a bone sticking out of the side of a rock wall. That bone just so happens to be part of one of the best preserved *Tyrannosaurus rex* fossils ever found. If you are Bob Harmon, a field crew chief of the Museum of the Rockies, this is exactly what happened. In the year 2000 Bob Harmon discovered a 68 million year old fossil, which is now named "B-Rex" after him.

Tyrannosaurus rex lived 65 to 70 million years ago, in what is now the western parts of the United States. They were the last of the large dinosaurs who lived during the Mesozoic era. After their extinction, the bones were trapped in the Earth for roughly 70 million years and preserved until the present day, through a process called fossilization. Much of what we know about dinosaurs comes from the scientific study of the shape, appearance, composition, and location of these fossil specimens. Dinosaurs' bodies were made up of the same general types of biological building blocks seen in present-day animals, such as tissues, cells, and proteins. However, since fossilization involves the replacement of dinosaur bone tissues with minerals over millions of years, the bone's biological material has long since degraded. Therefore, fossils usually do not give any molecular information about dinosaur proteins (i.e., they don't equip us to answer questions like "what *kinds* of proteins are in this fossil"). However, in the last decade scientists were able to isolate dinosaur

Figure 1: Tyrannosaurus rex fossil

proteins from some remarkably well-preserved dinosaur fossils. These discoveries open the door to the new era in paleontology, in which dinosaurs can be studied at the molecular level.

You and your team members are being called in to work with paleontologist Dr. Mary Schweitzer, in order to extract protein material from the “B-rex” fossil. You must determine what type of proteins it contains, and use it to learn more about how dinosaurs and present-day animals fit together in the evolutionary tree of life. It is your job obtain a protein sequence from the B-rex fossil to compare to the protein sequences from other present-day animals using bioinformatics tools, which you will learn about more about later.

To analyze the fossil sample, you will use mass spectrometry (MS), a standard technique in analytical chemistry. Protein mass spectrometry (MS) (Fig. 2) is a technique that ionizes chemical molecules and sorts the ions based on their mass-to-charge ratio (Fig 2). Using MS and specialized computer programs, scientists can take a protein mixture of unknown composition and identify the types of proteins in it. The whole process is analogous to how fingerprints can identify individuals: When a crime lab is provided with a fingerprint from a crime scene, they run it through a large computer database of fingerprints from known individuals, in order to find a matching result. Analogously, in MS, once you have the spectrum of an unknown protein you can use it to search a database of spectra of *known proteins* in order to identify the unknown protein. In today’s activity you will learn a bit about how this process works, by identifying protein sequences from the fossilized bones of a *T. rex*, a mastodon, and a Hadrosaur.

Figure 2: This is a picture of a Mass Spectrometer. It takes small samples and determines the molecular composition.

In the first step where MS is utilized you are working on the molecular level, understanding ion charges and masses. In the next step you will move up to the cellular level where proteins are made and used.

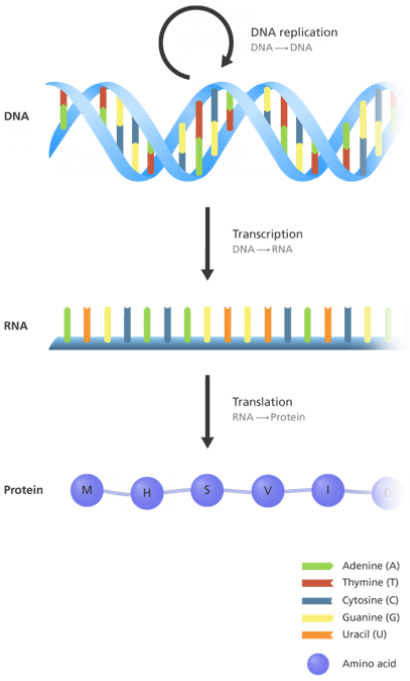


Figure 3. The Central Dogma: the first step is transcription of DNA into RNA. The second step is translating the RNA into proteins.

Proteins and where they come from: The process by which proteins are made in cells is known as the Central Dogma of biology (Fig 3). It is a two-step process involving, DNA, RNA, and amino acids (the building blocks of proteins). DNA carries all of the genetic information for an organism. In order for the DNA to be decoded and utilized in the cell, it must be transcribed into RNA, and then translated into an amino acid chain (sometimes referred to as a peptide)**.** Once the amino acid chain folds into its final shape (not shown in the figure), it is called a protein. Since you will be analyzing the protein content of a bone fossil, it is most likely that you will identify collagen proteins. Collagen proteins are sturdy and flexible in order to support our bones, and they make up 90-95% of the organic matter in bones.

Comparing DNA sequences across species is a powerful technique scientists use to learn more about the evolution of organisms. Sequences from specific DNA regions can be lined up with the same sequence from other organisms, in order to determine where mutations have occurred over time (Fig 4). This can be used to learn which animals have the same mutations, and how they evolved from each other. Although the B-rex fossil did not yield any DNA, it did yield protein. Since a protein is made up of a chain of amino acids (which has a corresponding letter sequence), it can be compared to other species' protein sequences in the same way as DNA. In the activity below, you will draw and analyze the mass spectrum of an unknown protein fragment from the B-rex fossil. In effect, you will be doing the work of the mass spectrometer and then replicating by hand the exact procedure that is today performed (much more efficiently!) by computers. This allows you to analyze protein MS spectra to identify proteins in a biological sample.

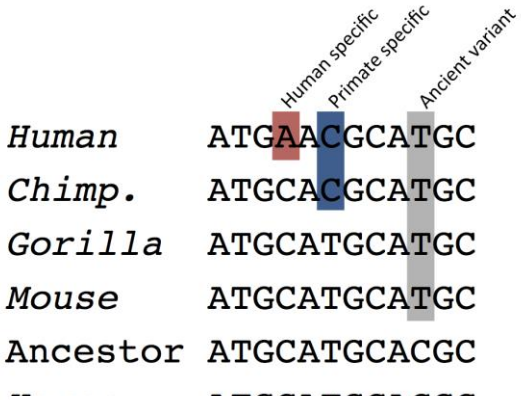


Figure 4. This is an example of DNA sequences from multiple species lined up together. Species who share mutations that others do not have are more closely related. This is how molecular biology can help determine evolutionary relationships.

Learning Objectives

After completing this activity, you should be able to:

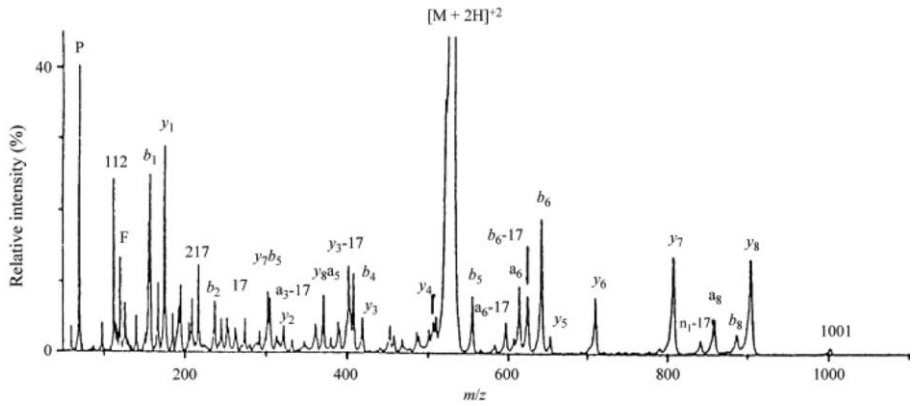
- Describe what protein mass spectrometry is
- Read and analyze protein mass spectra
- Describe a biological application of identifying and sequencing proteins from sample of unknown composition

Materials

- A bag with Legos clusters (**Please do NOT disassemble any of the Legos!**)
- Transparent paper with a blank spectra
- 12 known peptide spectra results

Procedure

You and Dr. Mary Schweitzer have collected a sample from the femur bone of the "B-rex" fossil. To understand how the MS takes a sample and determines its molecular composition, you can read the supplementary document, "Spectrometry in a Suitcase". However, it is not necessary to understand for this activity. What is necessary, however, is understanding the results after it has analyzed the sample.



The graph above is what your results will look like; this is a spectra of a peptide. The main parts of this graph that you need to understand are: the relative intensities, the m/z ratio, and what the peaks represent.

An ion is a molecule that has lost or gained an electron, changing its charge. Each peak corresponds to a different ion; the taller the peak, the more of that ion is found in the sample. Therefore, the y- axis (relative intensity, also referred to as relative abundance) determines **how much** of each ion is present in the sample. For example, the tallest peak on the graph is $[M + 2H]^{+2}$, which means that $[M+2H]^{+2}$ is the most abundant ion present. The mass-to-charge ratio (m/z) of an ion determines its **position** on the x-axis of the graph. To summarize:

- The peaks are specific ions
- The height of the peak (y-axis) is **how much** of that ion is present
- The location on the x-axis is unique to each ion since they all have a different **mass to charge ratio**.

You will be creating the spectra of the MS results from the 'B-rex" sample using Legos. Provided to you is a bag with 15 different clusters of Legos, when you remove them from the bag please **do not disassemble the Legos.** If you disassemble them the rest of the activity will not work. Each cluster of Legos represents one individual peak on the spectra from the "B-rex" sample. In order to draw the appropriate peak on your spectra, it is your job to decode what the Legos represent using the following rules,:

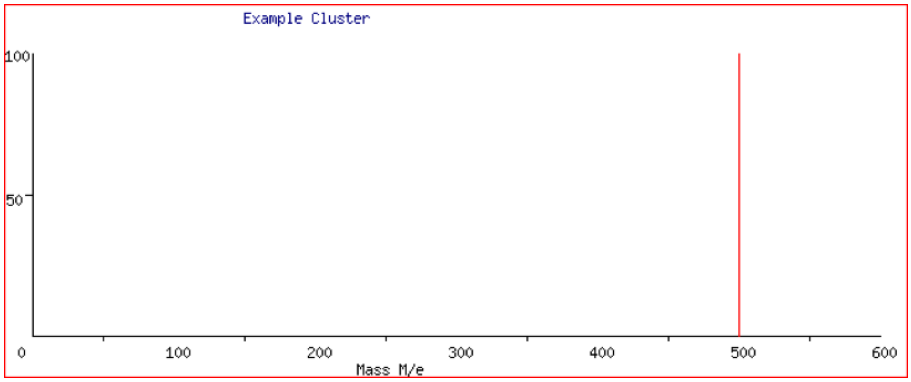
Lego Properties	**Corresponding Mass Spectrometry Properties**
The height (how many Legos **tall**) the cluster is	The locations on the x-axis; the mass-to-charge ratio (m/Z) where 1 Lego=100 m/Z
The width (size of the Lego) of the cluster	The height of the peak (y-axis); that represents relative abundance ● 2x1 Lego: <15 ● 2x2 Lego: 16 - 35 ● 2x3 Lego: 36-60 ● 2x4 Lego: >61
The additional Lego attached to each cluster	The precise location on the x-axis (only used when two or more clusters have the same height). Add the two numbers to the end of the mass-to-charge ratio determined from the height. ● Black: 00 ● Grey: 30 ● Green 50 ● Brown: 80

On the next page you will find examples of what a peak on your spectrum would look like based on the properties of a Lego cluster.

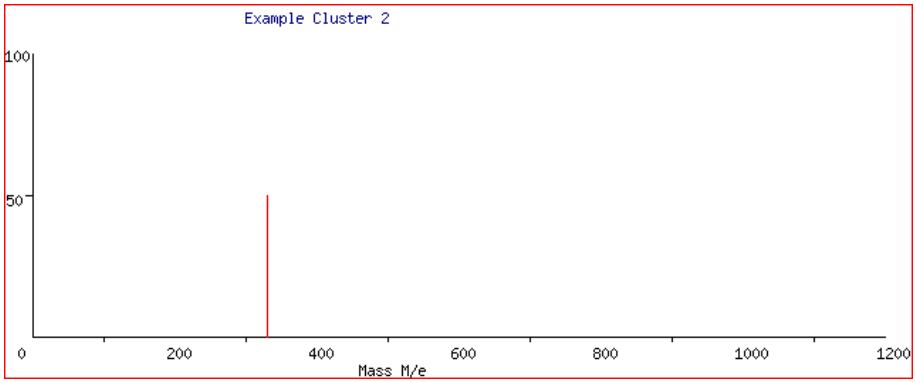
The final result from the mass spectrometer is the completed spectrum that you draw. This spectrum represents the ions in a peptide from the T. rex fossilized bone. In reality, this unknown spectrum would be used to search through a large database of mass spectra of known peptides, in order to find the closest match to a known peptide. To mimic this process, you are provided with 12 known spectra that are already identified to a specific protein or peptide. Take your drawn spectra on the transparency and line up the axes with each known spectra provided. The closest match to your spectra will tell you the exact sequence of amino acids from the peptide. In the next activity, you will learn how to use bioinformatics tools to identify what peptide this sequence belongs to.

Examples of Lego Clusters:

- A cluster that is 5 Legos tall, all size 2x4, and an additional black Lego attached on top it.
    - Since it is 5 Legos tall with a black additional Lego added, its mass-to-charge (x-axis) is 500, with a relative abundance over 60. This peak would look like this:

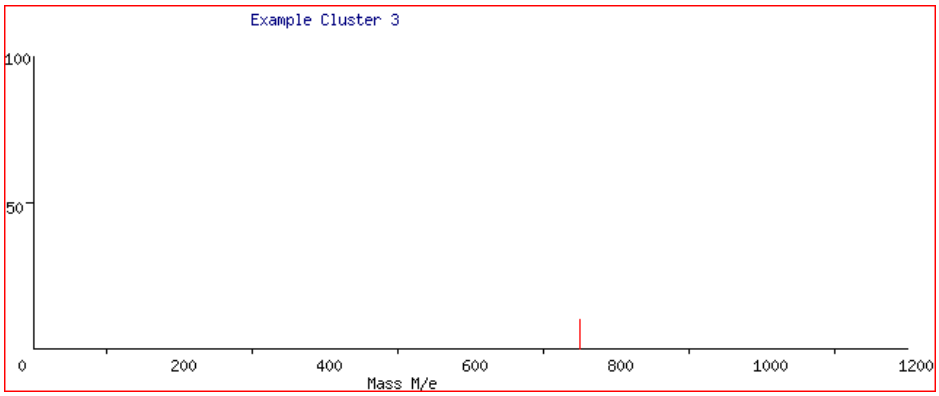


- A cluster that is 3 Legos tall, all size 2x3, and an additional grey Lego on top.
    - Since it is 3 Legos tall with a grey additional Lego, its mass-to-charge (x-axis) is 330, with a relative abundance between 35-60. This peak would look like this:



- A cluster that is 7 Legos tall, all size 2x1, and an additional green Lego on top.

- Since it is 7 Legos tall with a green additional Lego, its mass-to-charge (x-axis) is 750, with a relative abundance less than 15. This peak would look like this:



Analyzing Results

1. What is the sequence of the peptide according to your resulting spectra?

2. What is the purpose of using mass spectrometry in this activity? Would this be a good method for determining DNA/protein sequences in live animals?

3. What is the difference between a protein and a peptide? Explain why your results came up as one rather than the other.

Evaluating Results

1. What are any other real world applications of mass spectrometry that you can think of?

2. Why is it beneficial to know the exact order of the amino acids in a sequence, rather than what protein the amino acids make up? (*Figure 4)*

3. If you had to compare your spectrum against a stack of 200,000 spectra, how long do you think that process would take? Is there a better method for comparing spectra?