

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, that 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.



Figure : *Tyrannosaurus rex* fossil

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 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 other present-day animals using bioinformatics tools, which you will about more about later.

To analyze the fossil sample, you will use liquid chromatography mass spectrometry (LC-MS), a standard technique in analytical chemistry. Liquid chromatography (LC) is a "divide and conquer" technique that enables the separation of a complex sample (a sample containing multiple types of protein) into specific "fractions". The key point is that each fraction contains a much smaller (as in less diverse) set of proteins, and is thus easier to analyze. 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 LC-MS and specialized computer programs, scientists can take a protein mixture of unknown composition and identify the types of proteins in it. The



Comment [1]: This is actually how the fossil the students will be working with was found. If you or the students want to know more about the fossil and the research done with it, follow the "Dinosaur Shocker" link in the supplementary materials.

Comment [2]: To learn more about the theories behind the extinction of the dinosaurs, follow the "Dinosaur Extinction Information" link in the supplementary materials.

Comment [3]: When we mention bioinformatics here, it is to foreshadow that we will need bioinformatics tools in order to understand and make use of our results. This activity only focuses on the results themselves, but if you wish to know more about bioinformatics follow the links in activity 3's supplementary material.

Comment [4]: Mass Spectrometry is a confusing topic to teach. If the students need a visualization, or more information in order to understand this activity, check out "Mass Spectrometry in a Suitcase" or a youtube video from Bozeman science called "Mass Spectrometry".

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

Figure 2. This is a mass spectrometer, it is an analytical chemistry tool that can take very small samples, like fossilized bone, and determine its precise molecular composition.

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.

In the first step where LC-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.

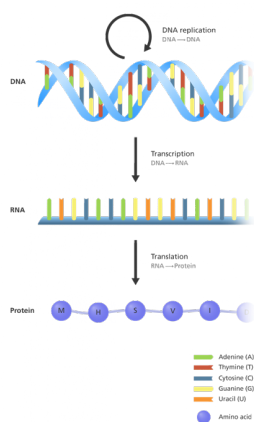


Figure 3. The central dogma, the first step is transcription of DNA into RNA. The second step is translating the RNA into proteins.

A bit about 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.

Comment [5]: Throughout all three activities the main extinct animal of focus is the T. rex. However, in activity 3 the students have the option to analyze peptide sequences from a Mastodon and a Hadrosaur, which is why they are mentioned here. Activity one only focuses on the T. rex.

Comment [6]: If students have not taken chemistry or need to refresh their knowledge on the subject they can check out the video "Chemistry 101" in the supplementary materials

Comment [7]: The central dogma is a very complex process in cellular biology. For the sake of this activity it is not necessary to understand all of the details behind how it works. If the students are interested in learning more, follow the link "Central Dogma" in the supplementary materials. This will take them to a khan academy page that has videos on each sub-topic of the central dogma.

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 B-rex fossil did not yield any DNA, it did yield protein. Since 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, and

compare it to mass spectra from a variety of known proteins in order to identify the specific protein that is present in the T. 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 LC-MS spectra to identify proteins in a biological sample.

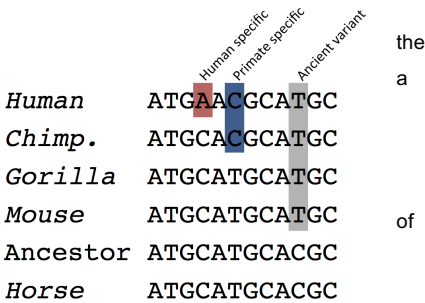


Figure 4. This is an example of DNA sequences from multiple species lined up together. Species who share mutations that others do not have are more closely related. This is how molecular biology can help determine evolutionary relationships.

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

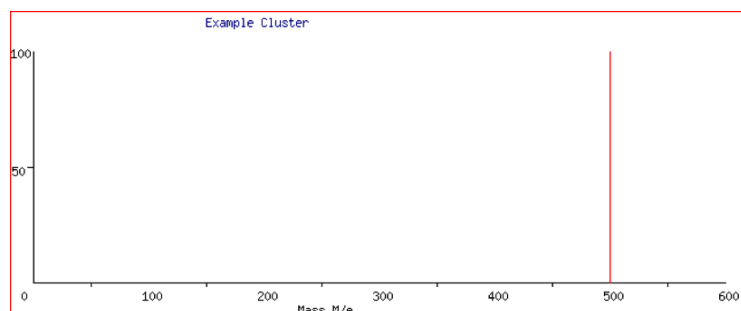
Procedure

You and Dr. Mary Schweitzer have collected a sample from the femur bone of the "B-rex" fossil. Protein fragments (which we call "peptides") have been extracted from the sample and divided into fractions using a technique called liquid chromatography; you will be analyzing a spectrum from one of these protein fractions. To understand how the MS takes a protein fragment 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 **color** of the Lego clusters determines the **type of the ion**. If the Lego cluster is white or yellow, it is a "y" ion. If the cluster is blue or red, it is a "b" ion.
 - Some clusters may have an additional Lego attached to it that is a different color from the rest of the cluster, ignore this Lego when determining the type of ion.
- The **number** of Legos in each cluster determines its **position on the x-axis**, this is its mass to charge ratio (m/z).
 - Each Lego is equivalent to 100 m/z , if you have 4 Legos then the mass to charge ratio is between 400-499 m/z . (Again, ignore the additional Lego that are attached when counting the number of Legos in each cluster.)

- The **color of the additional Lego** indicates the **precise order** the peaks fall on the **x-axis**. This is used when two or more clusters have the same number of Legos, this way you can determine which comes first. A pink Lego means it comes first, purple is second, green in third, and brown in fourth.
 - o For example; if you have 2 cluster with 3 Legos each, one will have a pink Lego and one will have a purple Lego. Having three Legos per cluster means the peak falls between 300-399 m/z, the cluster with the pink Lego will be drawn closer to 300, and the cluster with the purple Lego will be drawn closer to 399.
- The **size of the Legos in the cluster** determines the **relative abundance (y-axis)**. The 2x1 Legos indicates a relative abundance > 15. The 2x2 indicates a relative abundance between 15 and 35. The 2x3 indicates a relative abundance between 35 and 60. And the 2x4 indicates a relative abundance < 60.
 - o The precise height of each peak will vary slightly from the actual results, keep this in mind when comparing your final spectra to the ones provided for you.

For example: if a cluster has 5 white Legos all size 2x4, with an additional pink Lego attached on top it would be: a y ion, the first peak in the 500 m/z range, with a relative abundance over 60 (when you draw the peaks on your graph, label whether it is a y or b ion by simply writing the letter above the peak) this peak would



look like this:

The final result from the mass spectrometer is the completed spectrum that you draw. This spectrum represents the MS analysis of the fragment ions from a peptide from the fossilized bone specimen. In reality, this unknown spectrum would be used as a search key 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 axis' with each known spectra provided.** The closest match to your spectra will tell you exactly what peptide you have found, and thus, what dinosaur protein it likely came from!

Analyzing Results

Comment [8]: In order to create this activity using Legos I had to create ranges that the Legos represent. Size 2x2 Legos represent a y-axis between 15 and 35, their spectra won't be a 100% accurate match. Some of the lines will be slightly different heights, and the exact position on the x-axis will also slightly vary. If the students ask why theirs doesn't exactly match up tell them that is expected, and they didn't do anything wrong. They should still come to the same conclusion.

1. What protein did your peptide come from?

Collagen, more specifically alpha 2 type 1 collagen

2. Does the resulting peptide make sense as a result? (i.e. is your peptide something that could reasonably come from a bone sample?)

Yes, their result should be a collagen peptide. Collagen is the main protein found in all bones, therefore it make sense that the result is a collagen peptide. If their spectra matched the keratin or elastin proteins, this logically wouldn't make sense. Keratin is found in hair and nails, while elastin is found in the skin and organs.

3. 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?

The purpose of using mass spectrometry is to extract genetic material that still exists in fossilized bone.

Animals that have been extinct for over a thousand years have no remaining DNA that can be utilized in science. The only way to discover anything useful from a fossilized bone is to use LC-MS. LC-MS would not be an appropriate method for extant animals since they have DNA that can be extracted using much simpler methods.

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

As amino acids begin forming a chain during translation, they begin forming a peptide. Peptides do not have a function, and their shape is a simple string of amino acids. Once multiple peptides come together to fold, or a single peptide folds into a shape, it becomes a protein. A protein is always folded into a specific shape, this shape gives it its function. When the students get their results back they will have a collagen peptide, not protein. Mass spectrometry only works with very small samples, which is why the results are a peptide, a small sample of the collagen protein.

Evaluating Results

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

Students may provide a variety of answers, such as, forensics, environmental applications, quality control, medical applications, research applications, drug testing, and geology.

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? (hint: think back to y and b ions, and figure 4)

It is beneficial to know the order of amino acids in order to align them with the same protein from other animals in order to learn how they differ, which tells you how and when groups of animals evolved from one another. If you don't have the order of amino acids you can't determine evolutionary relationships.

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?

A very very long time. A computer program can do the same search in seconds! (this question is supposed to foreshadow how using computers makes these processes much quicker, which is why bioinformatics tools are necessary in science.)

