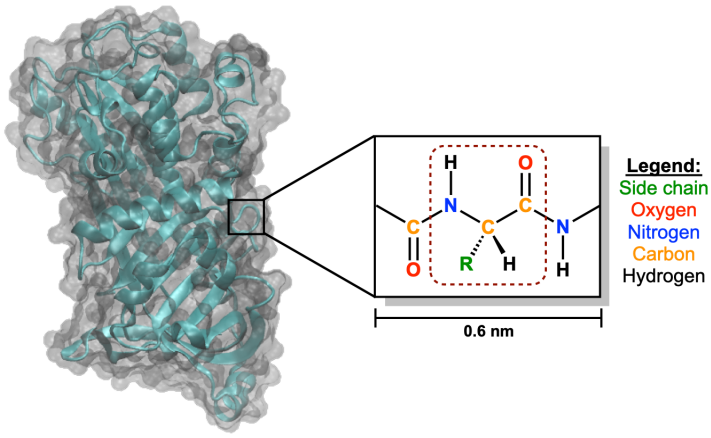
Protein Structure

Created with: Release 2021-3
Prerequisites: working knowledge of Maestro
Files Supplied: Protein_structure_worksheet
Categories: high school biology, AP/college general biology

About this Lesson

Proteins are among the most abundant organic molecules in living systems and come in many different shapes and sizes. Many different types of chemical bonds play an important role in maintaining a protein's structure. This lesson will focus on the primary, secondary, tertiary and quaternary structure of proteins.

Using Maestro, students will learn how to use the Sequence Viewer to look at individual amino acid residues, create a peptide sequence, visualize alpha helices and beta sheets, as well as generate electrostatic potential surfaces of a protein.



Learning Objectives

- Understand primary structure of proteins using the Sequence Viewer and the Build Biopolymer from Sequence panel
- Visualize secondary structure motifs such as alpha helices and beta sheets from the Maestro graphical user interface
- Generate electrostatic potential surfaces to view a protein's overall shape

Standards

- *ACS Guidelines*
  - Biological macromolecules (Section 5.1)
- *ETS Chemistry GRE*
  - Organic Chemistry – Amino acids, Peptides (3F)
- *AAMC MCAT*
  - Structure, function, and reactivity of biologically-relevant molecules (5D)

Assessments

The following types of formative assessments are embedded in this lesson:

- Assessment of student understanding through discussion of warm-up questions and filling in any knowledge gaps about structure-based virtual screen steps
- Visual assessment of student-generated docking scores from their own set of ligands

Warm-Up Questions: To be done on their own or at the beginning of class

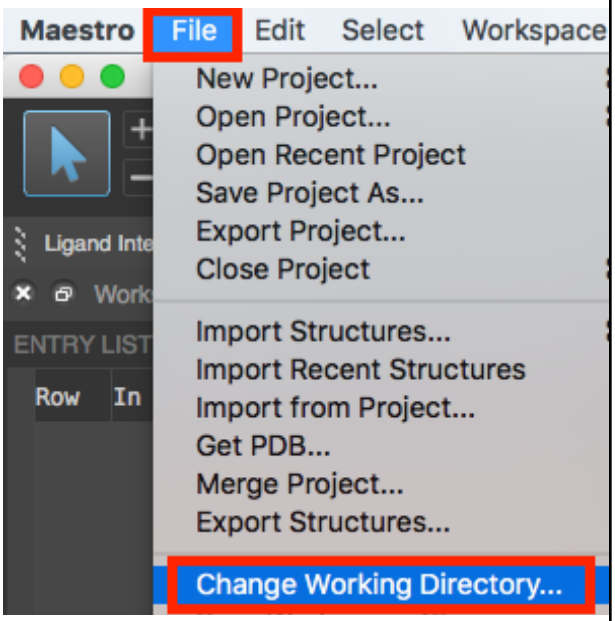
Read the Khan Academy article on Introduction to Proteins and Amino Acids and answer the following questions:

1) What are the 4 main categories of amino acids according to the article?

2) Define what protein denaturation is.

Lesson Outline

1. What you will need for this lesson

	1. Go to the 'Data' folder and open your Class Folder found on the virtual cluster's desktop. 2. Right-click on the folder called "Protein_Structure" and copy folder to Desktop ● Here, you will find the lesson plan, worksheet, and any additional resources
 *Figure 1-1. Open Maestro.*	3. Open **Maestro** a. See Starting Maestro if you need help
 *Figure 1-2. Change Working Directory option.*	4. Go to **File > Change Working Directory** 5. Find your "Protein_Structure" folder that you duplicated to your Desktop, and click **Choose**

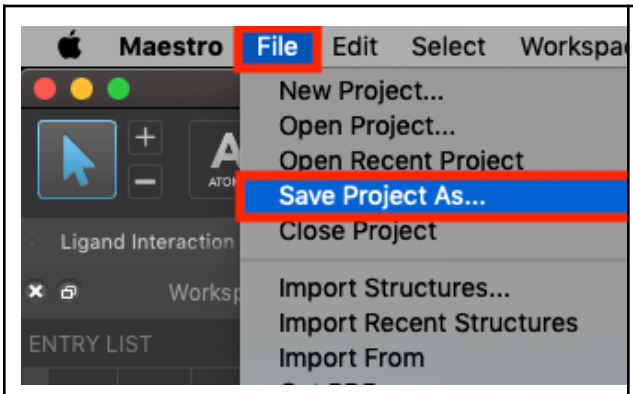


Figure 1-3. Save Project panel.

6. Next, go to **File > Save Project As**
7. Type "**Protein_structure_tutorial**" and click **Save**
    a. The project will be titled **Protein_structure_tutorial.prj**

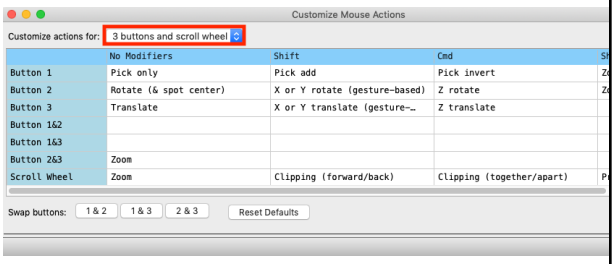


Figure 1-4. Choose the best mouse option for your set up.

8. Finally, check your **Mouse Actions**
    a. PC : **Edit > Customize Mouse Actions**
    b. Mac : **Workspace > Customize Mouse Actions**
9. Make sure you have **the best option chosen for your set up**. This lesson was written with a three-button mouse with a scroll wheel, meaning the scroll wheel is a button as well as a wheel. If you do not have a mouse, choose **Trackpad**.

10. If you are on the Teaching with Schrödinger virtual cluster, open the **Protein_Structure.prjzip** file in the Protein_Structure folder

11. If you are accessing this lesson from your personal computer, download the project file here

2. Introduction to Protein Structure

Let's start with what most people think of when they think of proteins – food! For example, eggs are a great source of protein. When we boil eggs, they go irreversibly from a liquid state to a solid (i.e. hard-boiled). To understand why that happens, we have to zoom in to the nanometer scale which is the scale at which proteins thrive. And what we will find is that heating eggs causes them to form solids because of a rather innocuous protein called ovalbumin pictured in **Figure 2.1**.

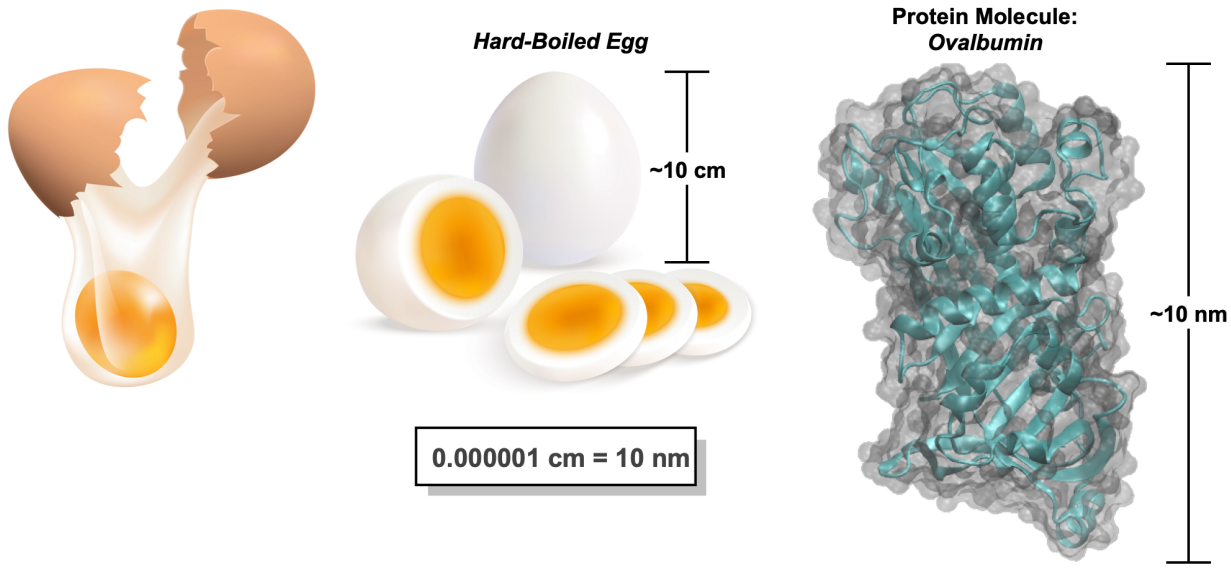


Figure 2.1. Comparing the size of an egg to a protein called ovalbumin.

The cyan ribbons are showing what is known as the backbone of the protein as it loops through space. These ribbons indicate the secondary structure of the protein while the grey area highlights the electron density surface of the protein, or the three-dimensional shape. This protein stands at 7.1 nanometers, and when we heat this protein up as we cook eggs, these ribbons unfold into long strings.

These long chains of unwound protein are kind of like spaghetti noodles. Have you ever cooked spaghetti so much that you accidently boiled off the water and left the cooked spaghetti in the sauce boil? What happens? All the noodles stick together into a massive clump that is unfortunately inedible. This is what is happening to ovalbumin when you heat up your eggs every morning, but on the nanometer scale. As the ovalbumin proteins in our eggs unwind they begin to interact with other unwound ovalbumin irreversibly. What are proteins made up of that cause them to behave this

way? Let's zoom in to the atomic level at a 0.1 nm scale and see what's going on in **Figure 2.2**.

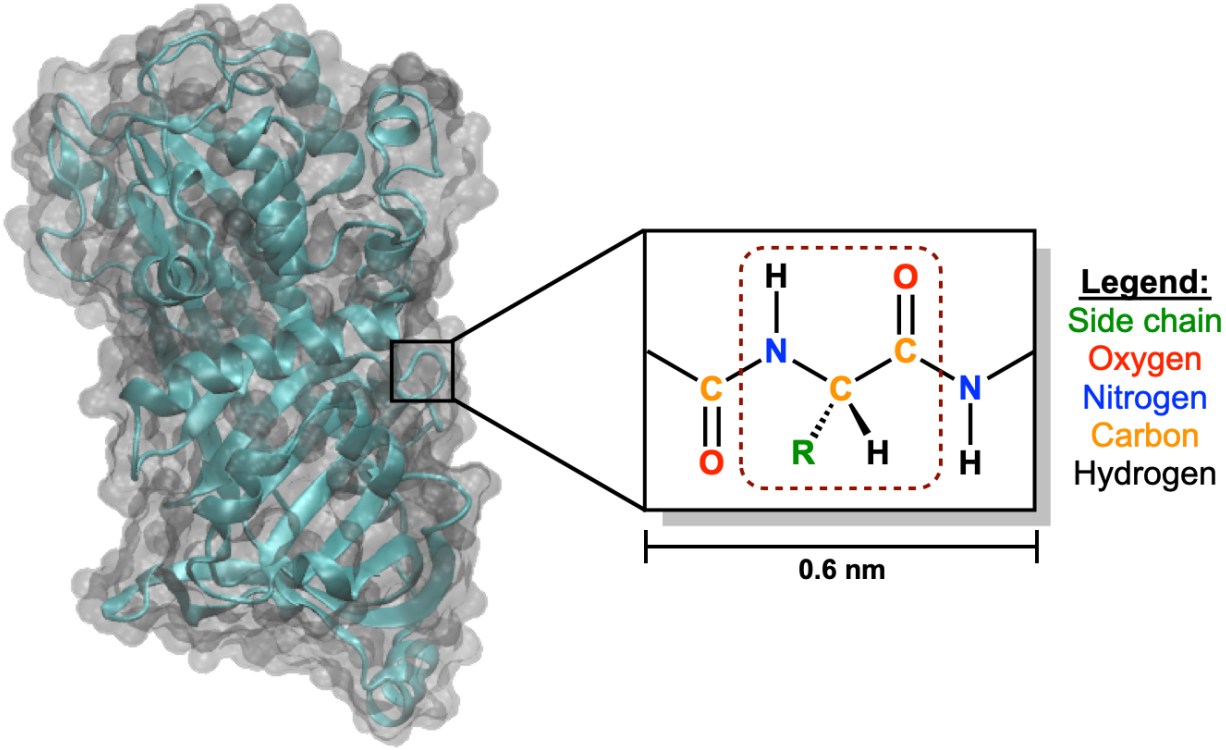


Figure 2.2. Zooming into an amino acid residue in the protein ovalbumin.

Proteins are made up of several main atoms: sulfur, oxygen, nitrogen, carbon, hydrogen, and the occasional phosphorus. And these atoms are arranged into building blocks called amino acids. In **Figure 2.2**, we are focusing on part of a protein with an amino acid highlighted within the dotted line box. There is an amino group shown with nitrogen and hydrogen, and then two carbons along the backbone until you get to the next amino acid which starts at the following blue nitrogen. The center carbon is attached to other atoms which are indicated here with the green R. This R is meant to represent one of twenty or more different side chains that a protein can have.

The twenty amino acids found naturally within our cells are the following divided up based on their chemistry. Hydrophobic amino acids include glycine, alanine, leucine, isoleucine, proline, and valine. The aromatic amino acids include phenylalanine, tryptophan, and tyrosine. Negatively charged amino acids are aspartic acid and glutamic acid. Positively charged amino acids are arginine, histidine, and lysine, depending upon their chemical environments. Hydroxylic amino acids include serine and threonine, and sulfur-containing amino acids include cysteine and Methionine.

Finally, the amidic amino acids are asparagine and glutamine. Each of the amino acids has their own one letter and three letter code associated with it. You can find a table that illustrates all of the amino acids here.

Proteins are polymers that are made up of long chains of these amino acids. Amino acids are linked into a long polymer through the formation of amide bonds via hydrolysis. Let's take the example in **Figure 2.3** of a short polymer made up of leucine, phenylalanine and alanine. Protein chains are built up from the N-Terminus with Leucine, which has a terminal amino group to the C-terminus, here with Alanine, which has a terminal carboxyl group. The amino acids in a protein fit together like a puzzle, except, unlike a typical puzzle which is two dimensions with similar sizes for each puzzle piece, the puzzle pieces of a protein are connected by a string (the backbone of the protein) and form a three-dimensional shape. Each individual puzzle piece is an amino acid, and the string in this analogy is the backbone of all the amino acids. Not only is the string mobile, but the amino acid side chains themselves can rotate and fit together in different ways over time.

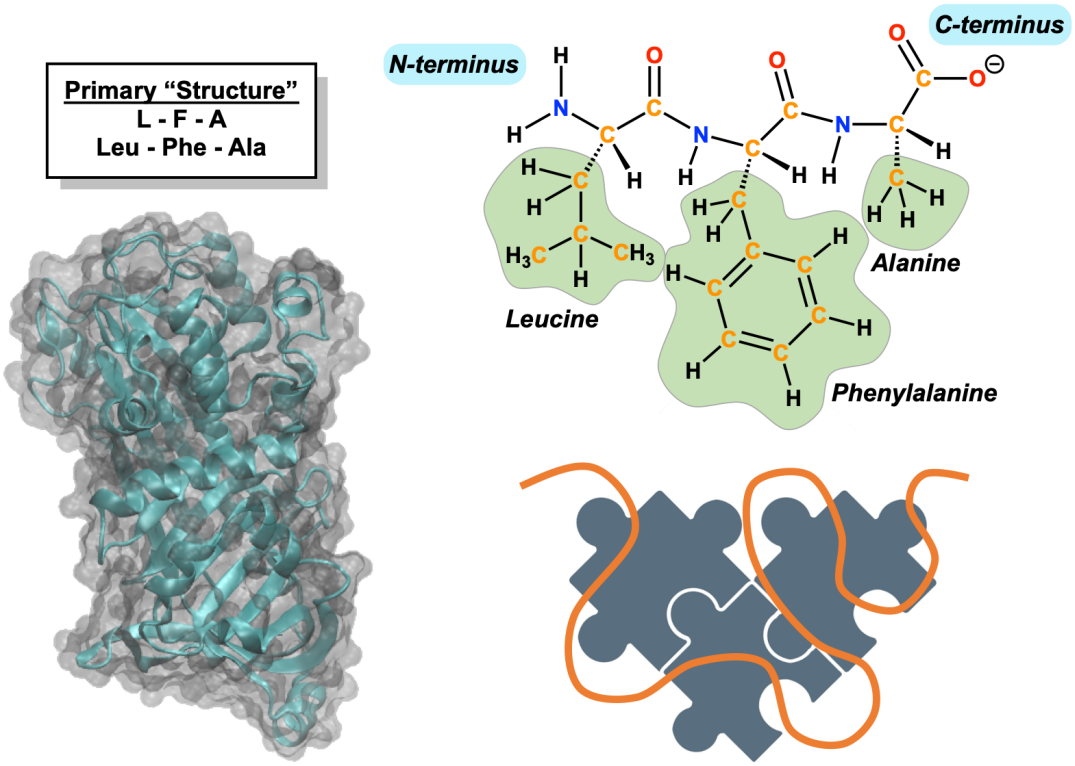


Figure 2.3. Amino acids in a protein fit together like a puzzle.

3. Primary Structure

3.1 Using the Sequence Viewer

The simplest level of protein structure is the primary structure which is a sequence of amino acids in a polypeptide chain. The sequence of a protein is determined by the DNA of the gene that encodes the protein, or a portion of a protein for multi-subunit proteins. A change in the gene's DNA sequence may lead to a change in the amino acid sequence of a protein. A single change (or mutation) of one amino acid in a protein's sequence may have drastic effects on a protein's overall structure and function. Let's use Maestro's Sequence Viewer to get a sense of the primary structure of the protein **1Y5M_prepared**.

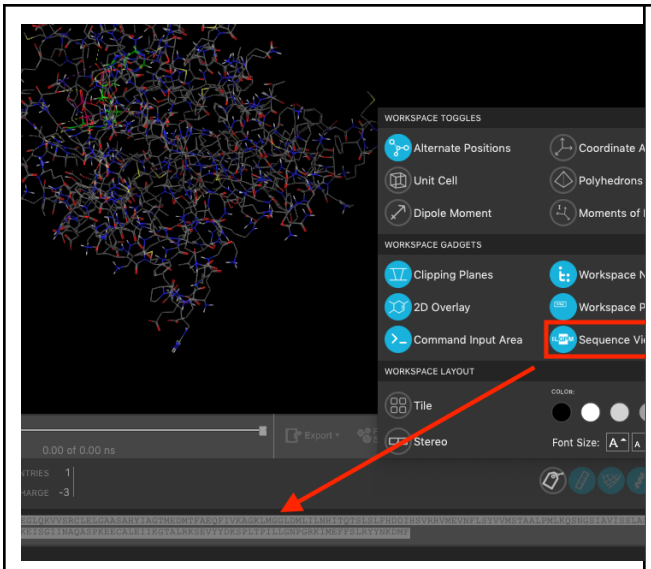


Figure 3-1. Turn on the Sequence Viewer.

1. In the bottom right, click on the ➕
2. Click on the **Sequence Viewer** to turn it on
    - The Sequence Viewer is now visible at the bottom of your Maestro window
    - This shows the amino acid sequence of the protein in your Workspace

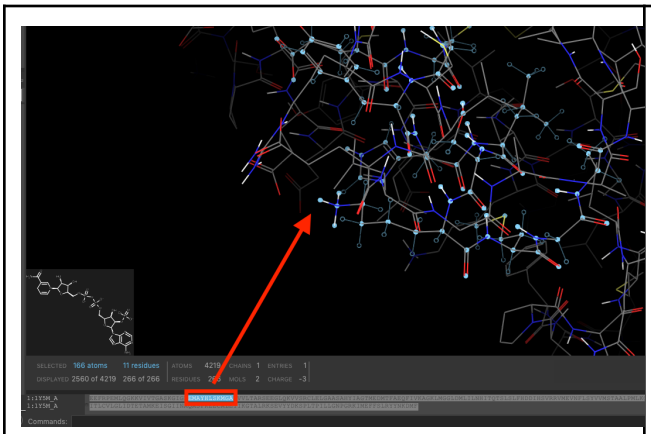


Figure 3-2. The part of the sequence is selected in the Workspace.

3. To see a specific part of the sequence, **Left-Click and drag**
    - That part of the protein will be highlighted in blue
4. Type **Z** on your keyboard,
    - Your Workspace is zoomed to that part of the protein

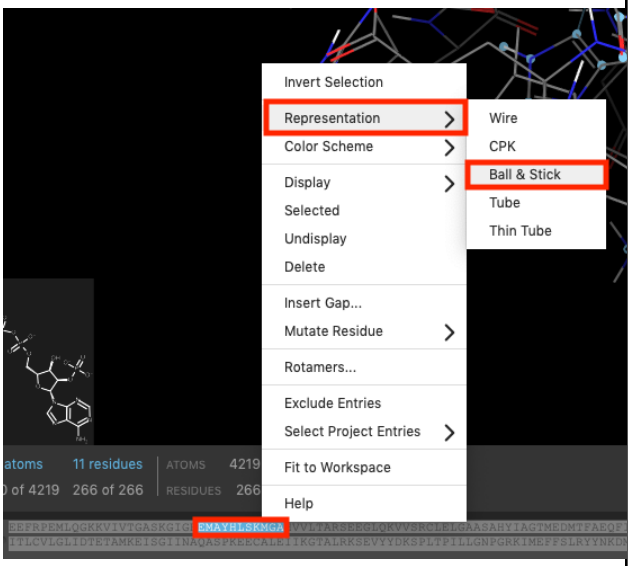


Figure 3-3. Change the representation of the sequence selection.

5. Right-click on the **selected sequence**
6. Go to **Representation > Ball & Stick**
    - Your sequence selection is now shown in Ball & Stick representation

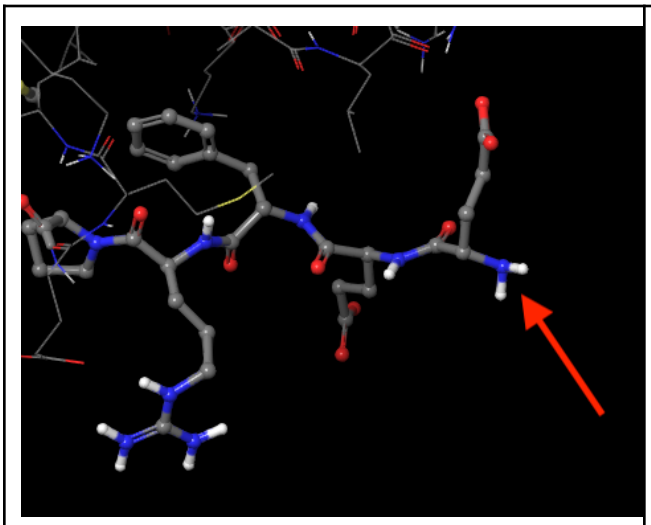


Figure 3-4. Highlight the N-terminus.

7. Now **select the first 5 residues** of the sequence and **represent** them as **Ball & Stick**
    - This first residue is the N terminus
    - The N-terminus is the start of a protein or polypeptide. The term comes from the free amine group located at the end of a polypeptide.
8. Type **Z** on your keyboard
    - Your Workspace is zoomed to the selected N-terminus

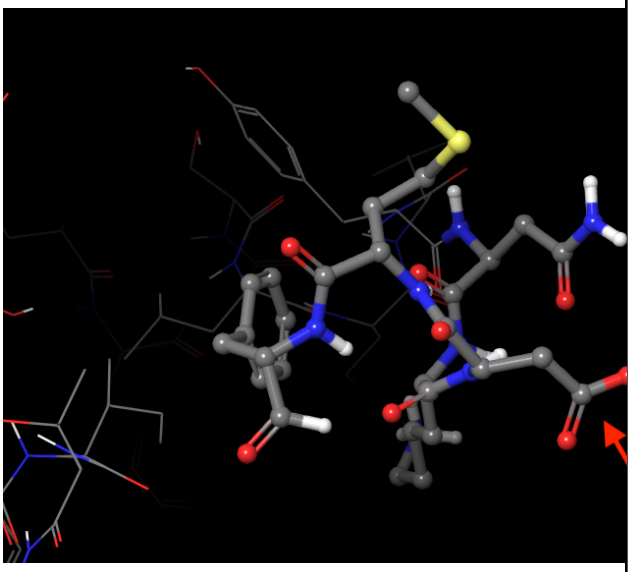


Figure 3-5. Highlight the C-terminus.

9. Now **select the last 5 residues** of the sequence and **represent** them as **Ball & Stick**
    - This last residue is the C terminus
    - The C-terminus is the end of an amino acid chain, terminated by a free carboxyl group (O-C=O)
10. Type **Z** on your keyboard
    - Your Workspace is zoomed to the selected C-terminus

Learn More: When the protein is translated from messenger RNA, it is created from N-terminus to C-terminus. The convention for writing peptide sequences is to write the sequence from N- to C-terminus.

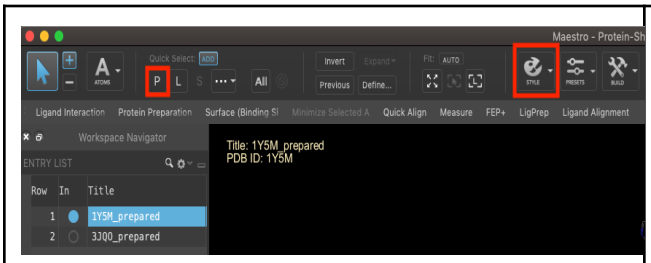


Figure 3-6.

11. Click on an **empty part** of the Workspace
    - The selection is cleared
12. Type **Z** on your keyboard
    - Your Workspace is zoomed to the full structure
13. In the top left of Maestro, under Quick Select, click **P**
    - The whole protein is selected in the Workspace
14. In the center of the screen, click **Style**
    - The Style toolbox opens

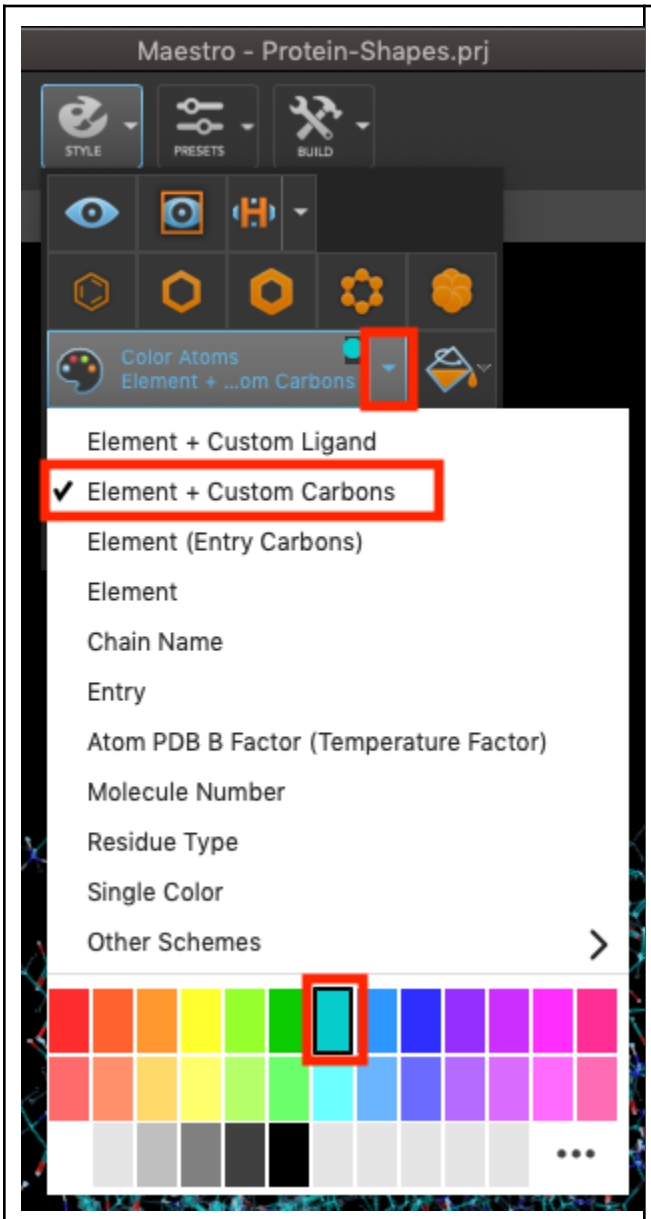


Figure 3.7.

15. Next to Color Atoms, click the **dropdown**
16. Choose **Element + Custom Carbons** and pick any color
    - The carbon atoms will be colored based on the color you chose
    - The sequence viewer color is updated with this color

3.2 Creating a peptide

A protein is made up of one or more polypeptide chains. Each polypeptide chain is made up of amino acids linked together in a specific order. A polypeptide chain is like a long word that is “spelled out” in amino acid letters. The structure and function of a polypeptide is determined by the chemical properties and order of the amino acids.

Now you will create your own peptide in Maestro using the **Build Biopolymer from Sequence** panel.

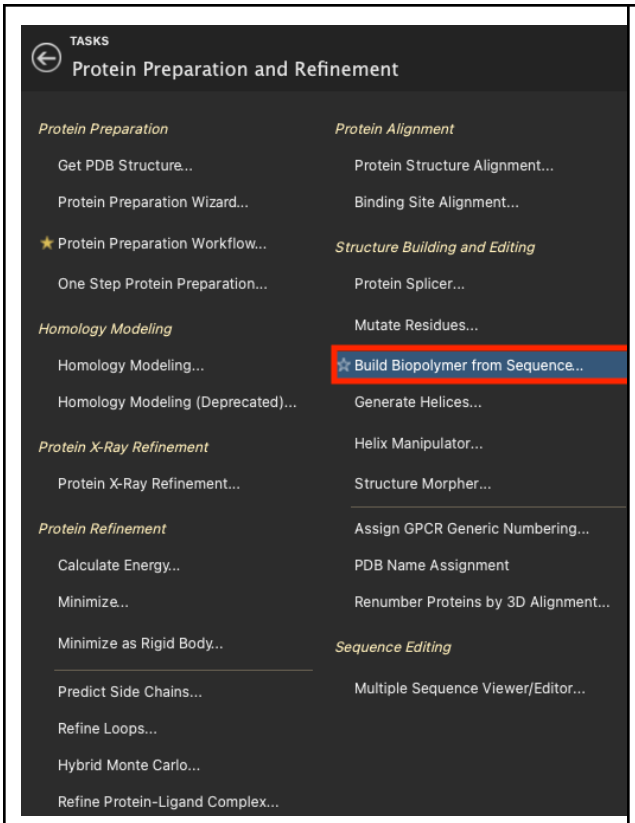


Figure 3-8. Open the Build Peptide from the Sequence panel.

1. Go to **Tasks > Browse > Protein Preparation and Refinement > Build Biopolymer from Sequence**
    - The Build Peptide from the Sequence panel opens.

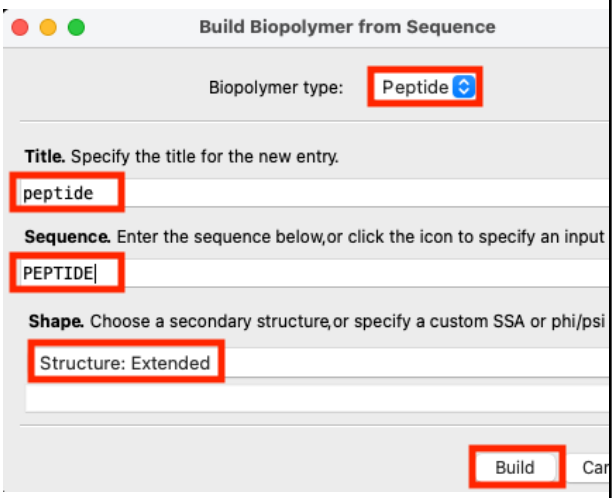


Figure 3-9. Build an extended peptide.

2. Under Title, type **Peptide**
3. Under Sequence, type **PEPTIDE**

Note: Capital letters must be used in the Sequence field

4. Under Shape, choose **Structure: Extended**
5. Click **Build**
    - A new entry titled Peptide is in the Entry List

<table>
<tr>
<td></td>
<td>

The peptide is shown in the Workspace

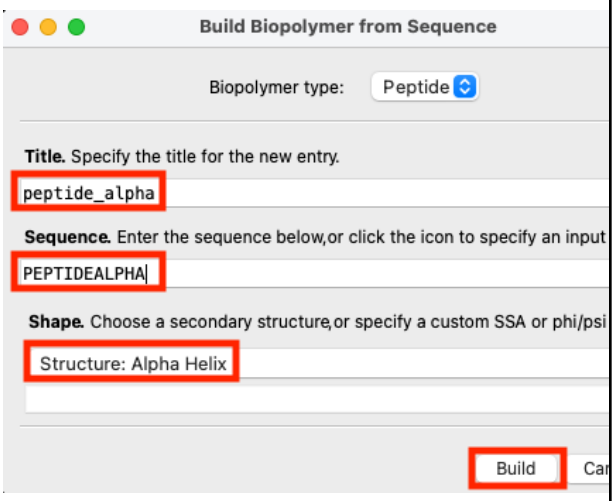
Note: Type Z if you cannot see the whole peptide.
</td>
</tr>
<tr>
<td>


Figure 3-10. Build an alpha helix peptide.
</td>
<td>
<ol start="6">
Under Title, type Peptide_alpha
Under Shape, choose Structure: Alpha Helix
Click Build

A new entry titled Peptide_alpha is in the Entry List
The peptide is shown in the Workspace

Activity: Build a peptide out of the letters in your name. Note that the letters B, J, O, U, X, and Z are not valid amino acids in this activity.
</td>
</tr>
</table>

Question #1: Build a peptide out of the letters in your name and take a screenshot of your peptide. Note that the letters B, J, O, U, X, and Z are not valid amino acids in this activity. Feel free to use any other letter to replace them.

4. Secondary Structure

In proteins, the primary structure will fold into 3D shapes based on the interactions that the various amino acids can make with each other. These shapes are known as secondary structures and there are three main types - alpha helix, beta sheet, and loops.

4.1 Visualizing Alpha Helices

Let's begin by taking a look at the alpha helices in our protein. In an alpha helix, the backbone of the polymer of amino acids forms a helical shape like an old-school telephone cord. The carbonyl of one amino acid forms a hydrogen bond with the amide hydrogen of an amino acid that is four places down in the amino acid polymer sequence. Oftentimes in biology and biochemistry courses, the hydrogen bonds are credited with dictating these folds. In reality, however, these secondary structure elements are driven to fold by hydrophobic forces that also govern tertiary folding. Let's visualize an alpha helix in Maestro.

Figure 4-1. The Sequence Viewer is colored by secondary structure.

1. In the Entry List, include **1Y5M** in your Workspace
    - This protein has already been prepared for you.

Note: To prepare a protein, use the Protein Preparation Workflow which helps address many of the issues inherent to a PDB structure, such as missing hydrogen atoms, missing side chains, missing loops, alternate positions for side chains. See the lesson on Protein Ligand Docking for more information.

2. In Quick Select, click **P** to select the whole protein

	○ The whole protein is highlighted in the Sequence Viewer 3. Left-click the **protein sequence** in the Sequence Viewer, and drag the mouse over the sequence to highlight a protein sequence.

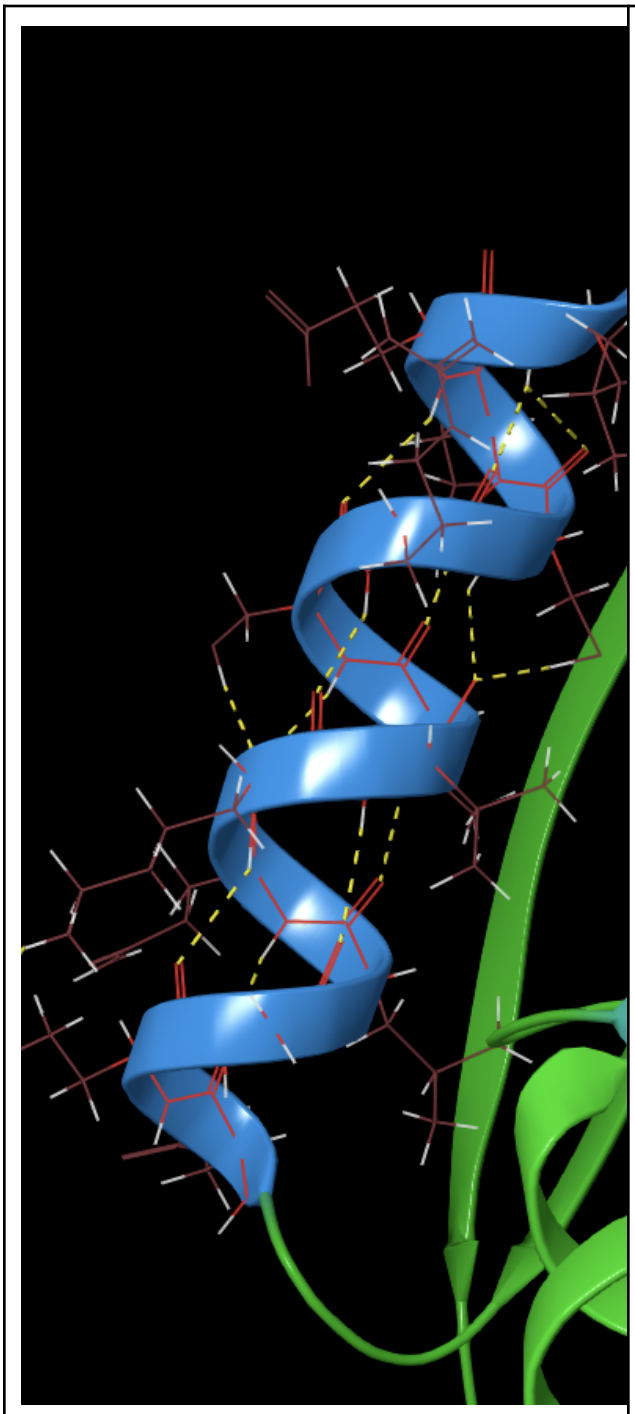


Figure 4-2. Visualizing hydrogen bonds in an alpha helix.

4. Then you can select certain alpha helices and they may look like this.
    - After the portion you want highlighted is highlighted you can click style and change the color of the highlighted portion, so that it stands out from the rest.

Note: Interactions in the bottom right corner of the screen must be turned on.

Question #2: In the alpha helix that you are visualizing, how many hydrogen bonds can you see? (Note that hydrogen bonds are represented by dotted yellow lines). What type of change do you think the hydrogen bonds are implementing on the structure? Are the hydrogen bonds there for the stabilization of the peptide?

4.2 Visualizing Beta Sheets

Now we will highlight the beta sheets. Beta sheets (also Beta-pleated sheets) are a common motif of a secondary structure in proteins. Beta sheets consist of beta strands (also B-strand) connected laterally by at least two or three backbone hydrogen bonds, forming a generally twisted, pleated sheet. In beta-sheets, the backbone of the polymer of amino acids forms an S curve, and the carbonyl carbon of one amino acid forms a hydrogen bond through space with the amide hydrogen of a different amino acid that is further down the polymer chain. In molecular modeling, a beta sheet is shown with arrows pointing in the direction where the carbonyl end of the amino acids are pointing. Let's visualize a beta sheet in Maestro.

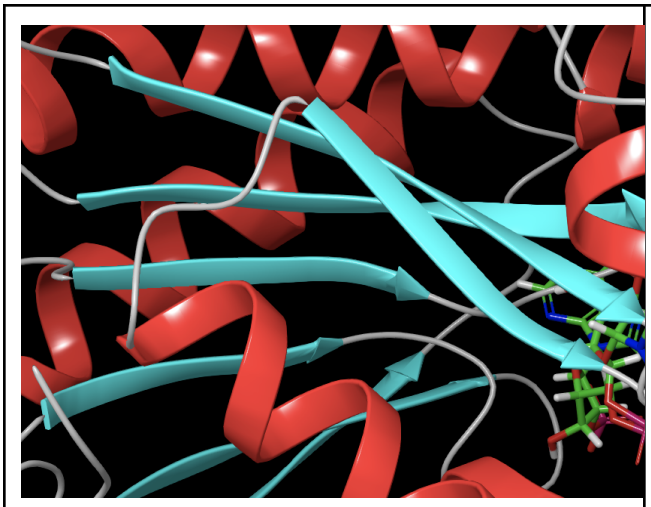


Figure 4-3.

1. Locate a Beta Sheet in a protein structure that looks similar to the image on the left.

Question #3: Secondary structures in proteins are all about interactions of the actual backbone. Go to **3JQO_prepared** in your entry list. Take screenshots of an alpha helix and a beta sheet. List what types of interactions are shown in each.

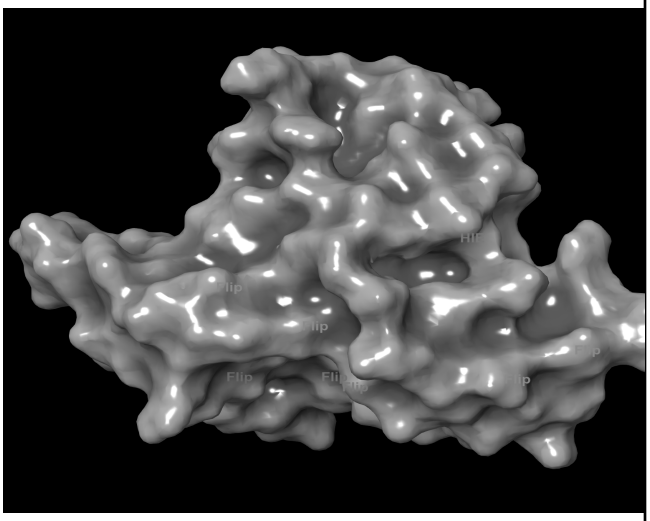
5. Tertiary Structure

In this section, we will be able to visualize the tertiary structure of a protein. Tertiary structure is the arrangement of the secondary structures into a final 3-dimensional shape. The secondary structure motifs fold into an overall arrangement that gives rise to a unique protein type with a unique shape and function. The chemical properties of the various R-groups or side chains of the amino acids within the protein influences such folds. Knowing the tertiary structure of a protein is often crucial to understanding how it functions and how to target it for drug therapy or other medical uses. Let's use Maestro to visualize an overall protein shape by generating an electrostatic potential surface.

5.1 Visualizing Protein Surfaces

Figure 5-1. Opening the Style panel to add a

1. To put a surface on a protein, click **Style > Surface** and it will appear in the workspace
    - If this action has been completed, you will see a grey surface form around the protein

Surface.	
 *Figure 5-2. Visualizing a protein surface.*	2. To edit the surface, click the S button to the right of the entry name in the Workspace Navigator. Right click the S and press **Display Options** 3. Change the **Style** to **solid** and the **Color Scheme** to **electrostatic potential**

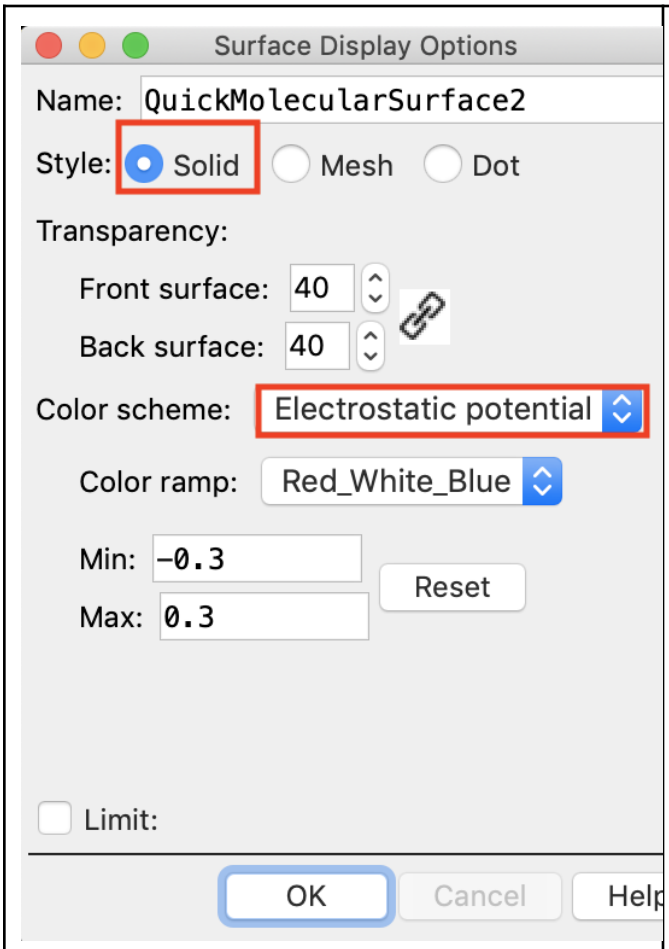


Figure 5-3. Changing the surface color scheme to electrostatic potential.

5.2 Visualizing Binding Sites

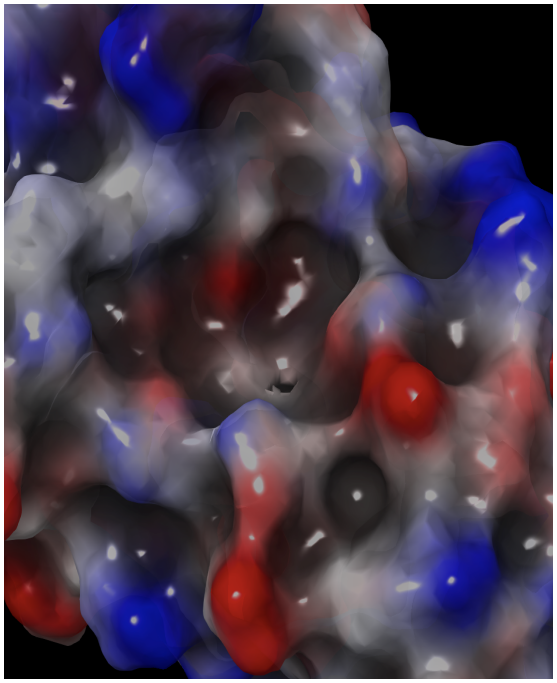


Figure 5-4. Visualizing the binding pocket of a protein.

4. The binding sites are the dips and holes in the structure where a receptor molecule can bind to it.

Analysis:

- Can you see the differences in the binding sites?
- Is it easier to understand why not every molecule can fit into a certain receptor?
- And why is shape important?

Question #3: Electrostatic surface potentials are colored red for negatively charged residues and blue for positively charged residues, while white/gray represents neutral residues. Look at the active site region where the ligand is bound. Can you see the differences in the binding sites? Is it easier to understand why not every molecule can fit into a certain receptor? And why is shape important?

6. Quaternary Structure

Some proteins are made up of more than one amino acid chain, giving them a quaternary structure. These multi-chain proteins are held together with the same forces as the tertiary structure of individual protein chains (hydrophobic, hydrophilic, positive/negative charges, etc.). Sometimes the protein chains in a complex are identical, making it a dimer, or may vary. Some examples of protein complexes with quaternary structure are potassium channels, antibodies, and G-proteins.

In this last section, we will be looking at the protein **3JQO_prepared**. It is a 42 chain structure with sequence from E. coli. Using what we learned from the previous sections, answer the following question:

Question #4:

Using the Sequence Aligner, list the first 5 amino acid residues in chain A (**3JQO_A)**. Zoom into these residues in your workspace and take a screenshot.

Question #5:

Find an alpha helix and a beta sheet in this protein complex and take a screenshot of each.

7. Summary, Additional Resources, and References

In this lesson, students learned how to use the Sequence Viewer to look at individual amino acid residues, create a peptide sequence, visualize alpha helices and beta sheets, as well as generate electrostatic potential surfaces of a protein. Using Maestro, students were able to interact with proteins by using the graphical user interface to zoom in/out, translate, and move around various protein structures.

For further information, please see:

[1] https://en.wikipedia.org/wiki/Alpha_helix

[2] http://proteopedia.org/wiki/index.php/Proteopedia:Table_of_Contents

[3] https://www.rcsb.org/search

[4] http://proteopedia.org/wiki/index.php/Introduction_to_protein_structure

Entry List - a simplified view of the Project Table that allows you to perform basic operations such as selection and inclusion

Included - the entry is represented in the Workspace, the circle in the In column is blue

Project Table - displays the contents of a project and is also an interface for performing operations on selected entries, viewing properties, and organizing structures and data

Scratch Project - a temporary project in which work is not saved, closing a scratch project removes all current work and begins a new scratch project

Selected - (1) the atoms are chosen in the Workspace. These atoms are referred to as "the selection" or "the atom selection". Workspace operations are performed on the selected atoms. (2) The entry is chosen in the Entry List (and Project Table) and the row for the entry is highlighted. Project operations are performed on all selected entries

Working Directory - the location that files are saved

Workspace - the 3D display area in the center of the main window, where molecular structures are displayed

Alpha Helices -

Peptide - The purpose of a peptide is to react with/ perform biological functions

Beta sheets - A secondary structure that can occur in proteins

Monomer - A molecule that binds to other identical molecules to form a polymer

Hydrophobic - an area that repels water, typically made of nonpolar atoms