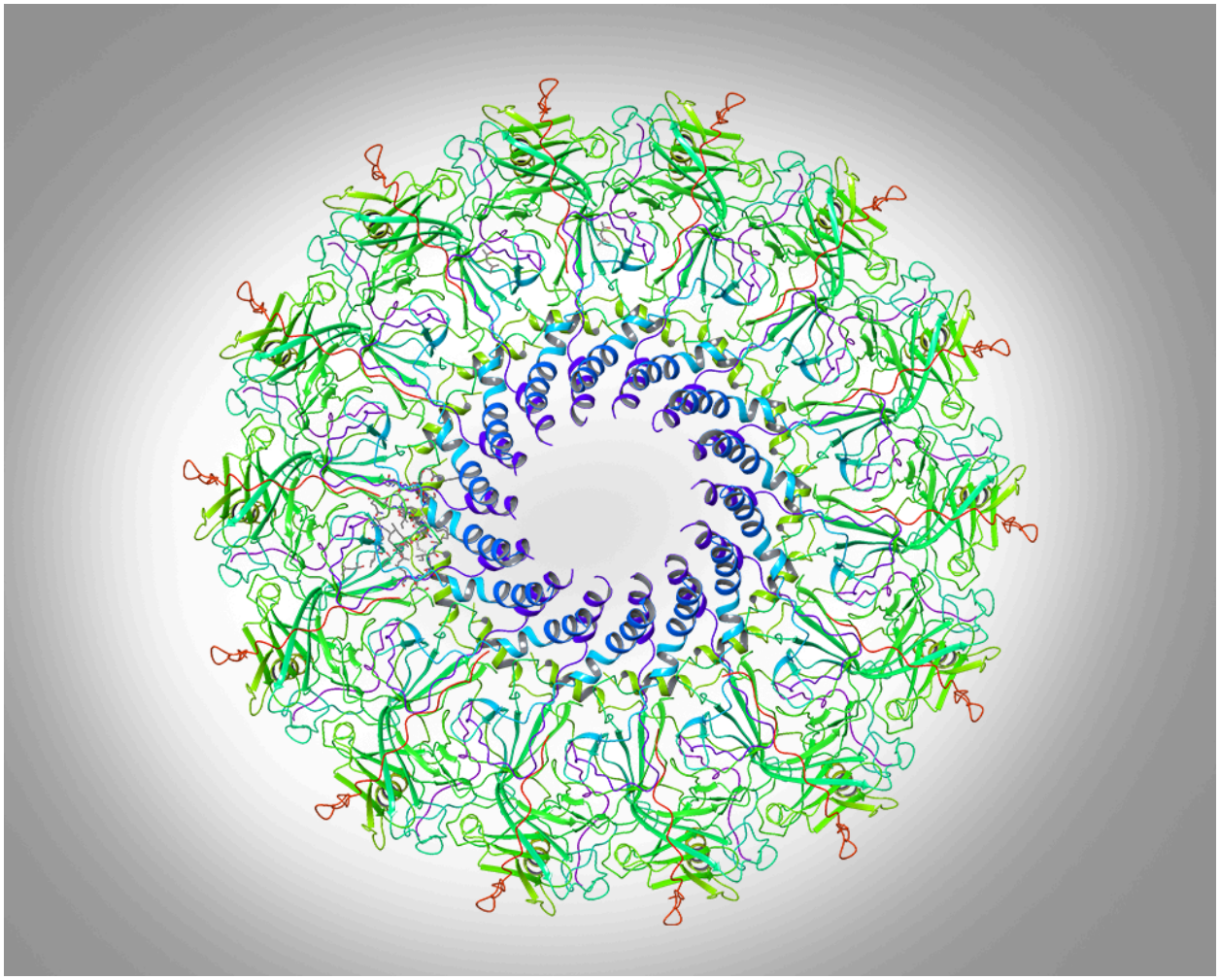
Protein Structure



Protein Structure

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:

- Identify primary, secondary, tertiary, and quaternary structure of proteins
- 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

Lesson Contents:

Standards Alignment:

- Connections to AP
  - Chemistry of Life (Unit 1)
- AAMC MCAT
  - Biological and Biochemical Foundations of Living Systems (1A)
- IB Diploma Programme
  - Enzymes (Core Topic 2.5)
- ACS Guidelines:
  - Biological Reactions - (Conceptual Topics)

Assessments for Understanding:

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 the four levels of protein structure
- Visual assessment of student-generated peptide chains, helices, and protein surfaces

Associated Documentation Pages: Building a Carbohydrate Polymer

Warm-Up Questions:

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.

Need help? Contact us at teaching@schrodinger.com

1. Setting Up the Maestro Session

At the start of the Maestro session, it is essential to 1) check your mouse actions, 2) change the file path to the Working Directory for this lesson, and 3) save your project file. The working directory indicated in this section contains the files necessary to complete this lesson. If you do not set the appropriate working directory, you will be unable to run any calculations.

1. Launch the Virtual Cluster

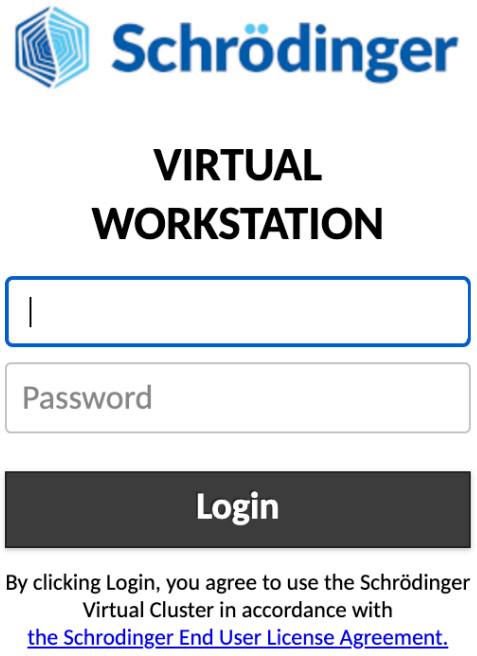


Figure 1-1. Virtual workstation login page.

2. Double-click the **course-data** folder on the desktop

Figure 1-2. Course-data folder on the desktop.

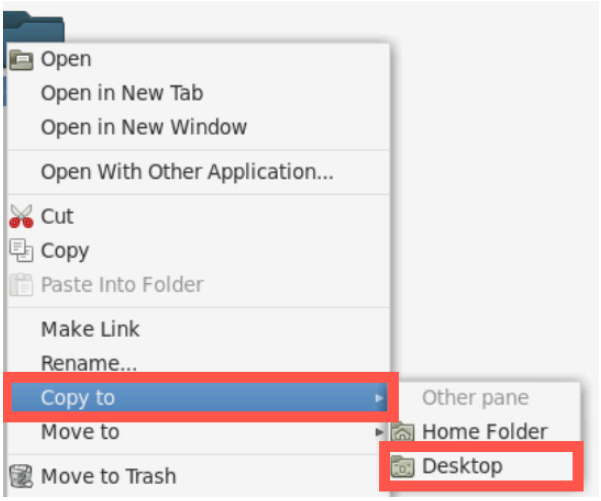


Figure 1-3. Copy the lesson folder to the Desktop.

3. Right-click the `Protein_Structure` folder and select **Copy to > Desktop**

4. Double-click the Maestro icon on the desktop

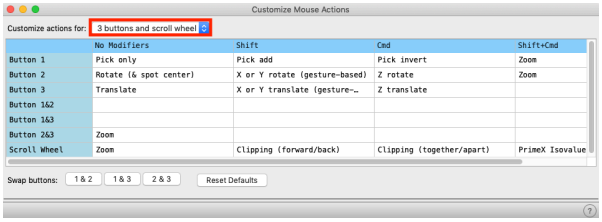


Figure 1-4. Change the mouse actions.

5. Check your mouse actions.
    - Go to **Workspace > Customize Mouse Actions**
    - *Note:* This lesson was made with a three-button mouse with a scroll wheel, but a trackpad can still be used
    - ***Trackpad keys:***
        - **Up/Down trackpad** = Zoom In/Out
        - **Option** = Rotate
        - **Control** = Translate

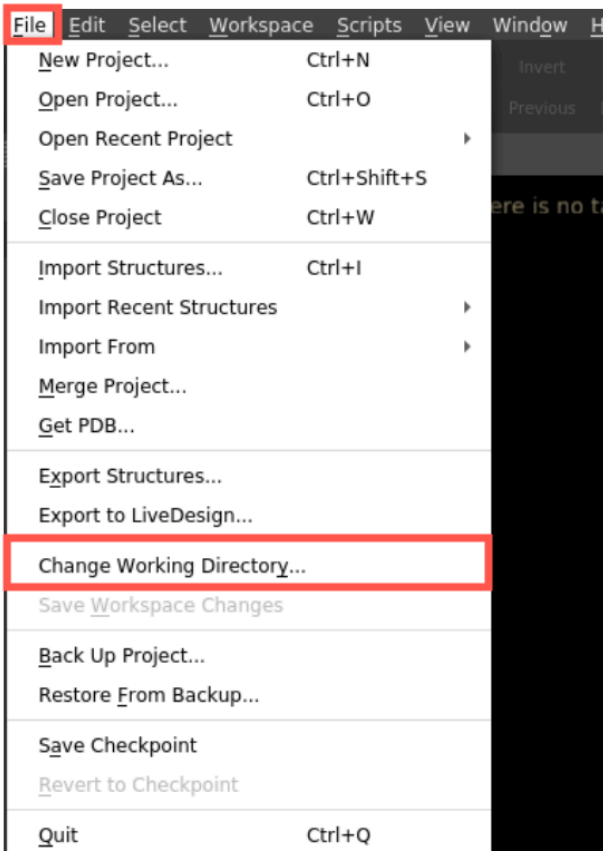


Figure 1-5. Change Working Directory option.

6. Go to File > Change Working Directory

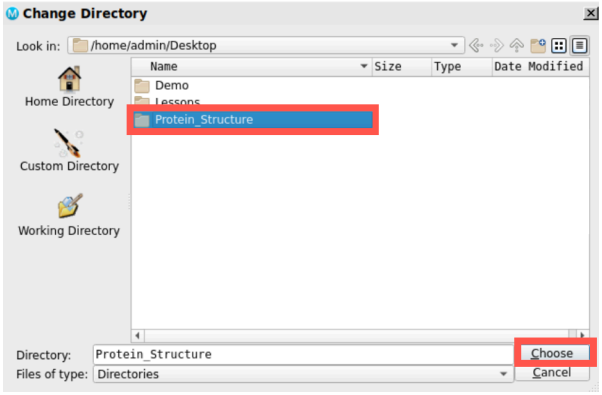


Figure 1-6. Change Working Directory panel.

7. Navigate to Desktop > `Protein_Structure` folder and click **Choose**

Pre-generated input and results files are included for running jobs or examining output

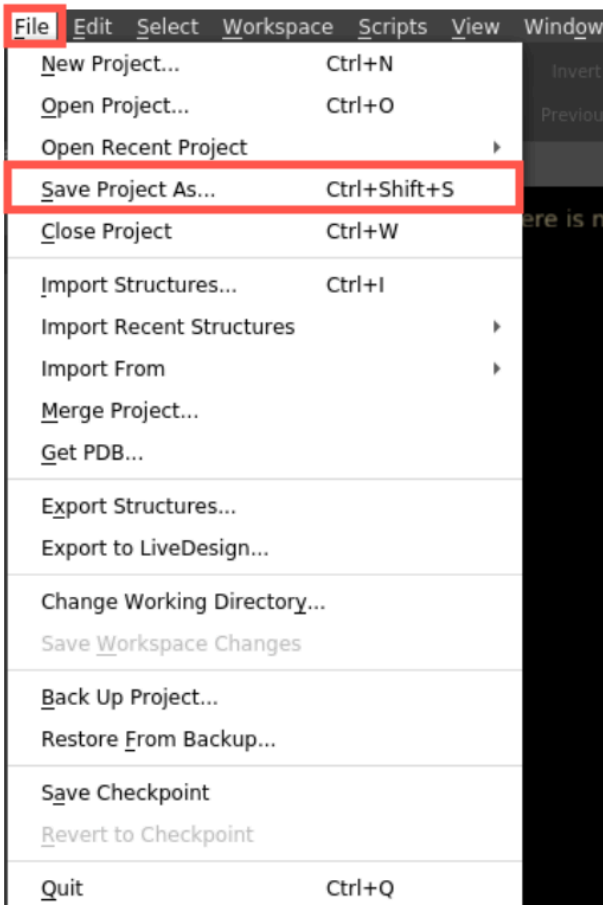


Figure 1-7. Save Project option.

8. Go to File > Save Project As

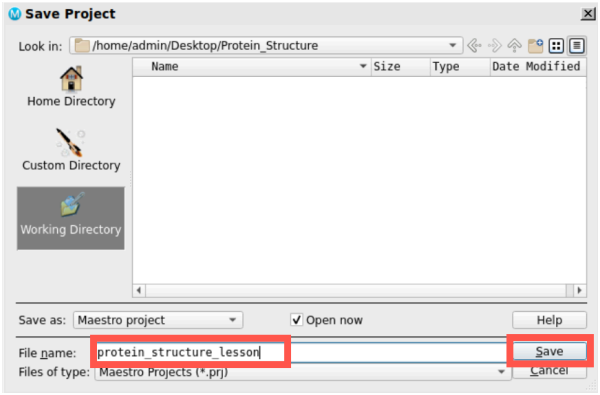


Figure 1-8. Save Project panel.

9. Change the *File name* to `protein_structure_lesson`, click Save
    - The project is now named protein_structure_lesson.prj

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**.

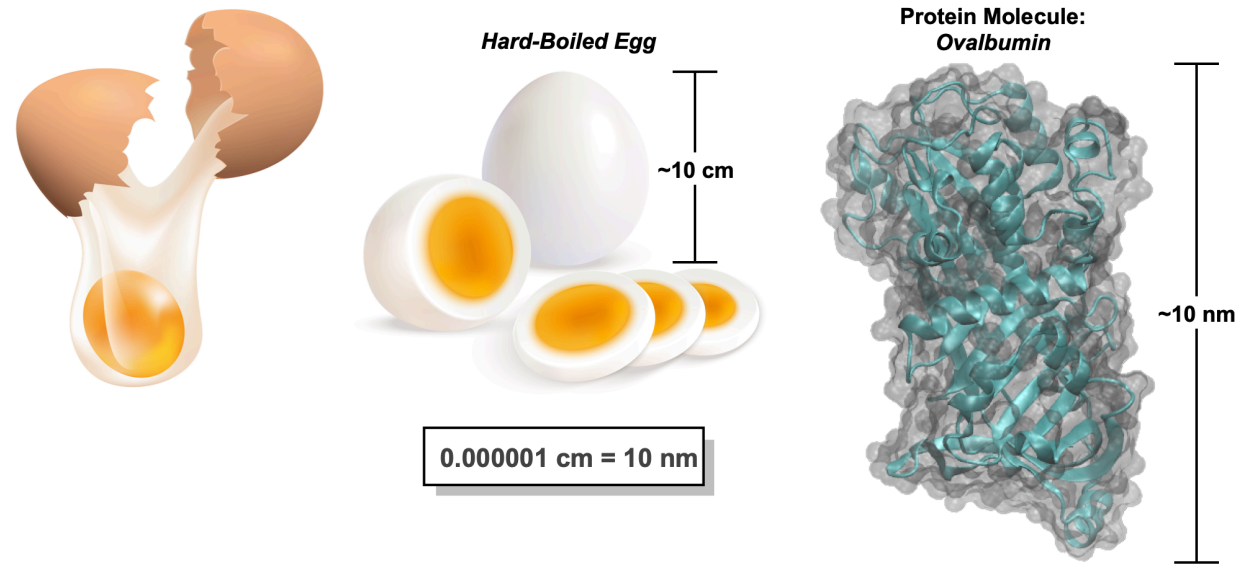


Figure 2-1. Comparing the size of an egg to a protein called ovalbumin.

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**.

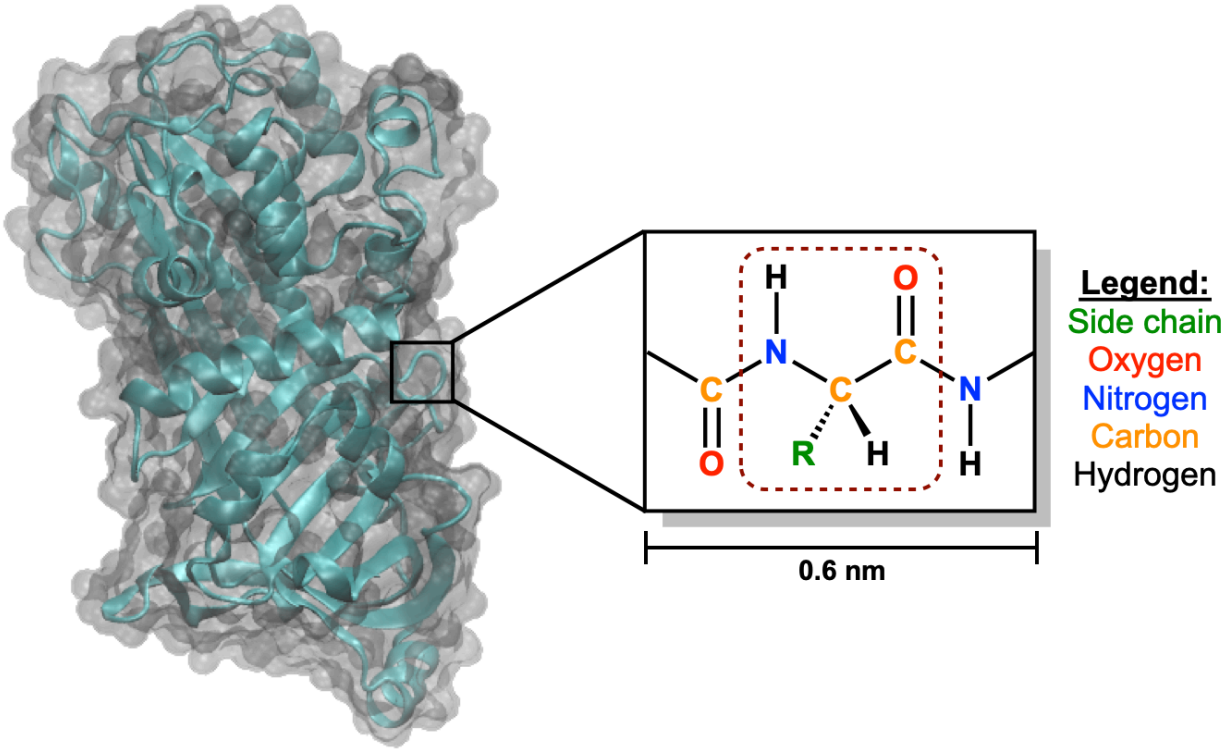


Figure 2-2. Zooming into an amino acid residue in the protein ovalbumin.

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.

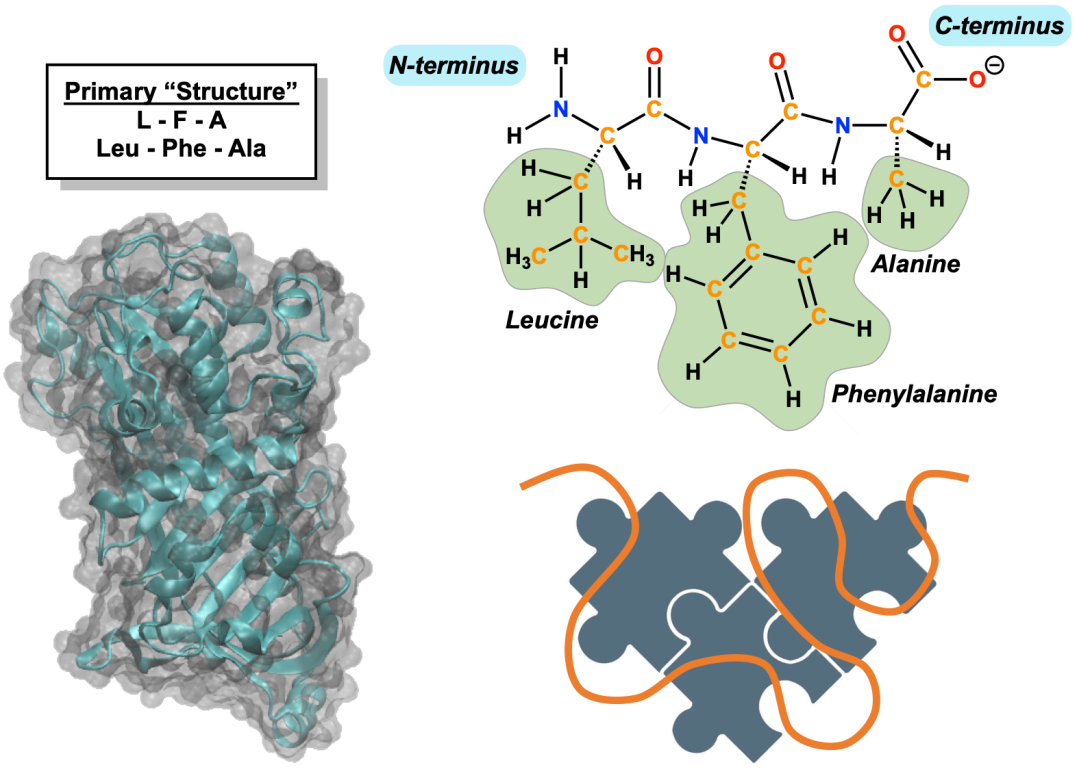


Figure 2-3. Zooming into an amino acid residue in the protein ovalbumin.

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**.

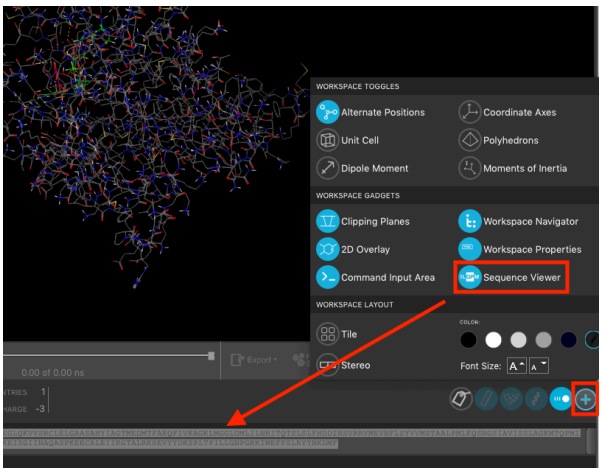


Figure 3-1. Turn on the Sequence Viewer.

1. In the bottom right, click on the ➕
2. Click on the **Sequence Viewe**r 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

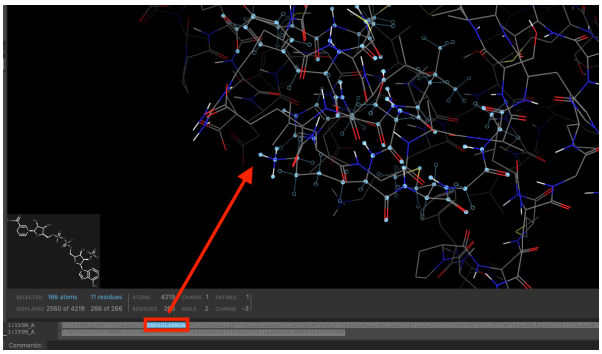


Figure 3-2. The part of the sequence is selected in the Workspace.

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

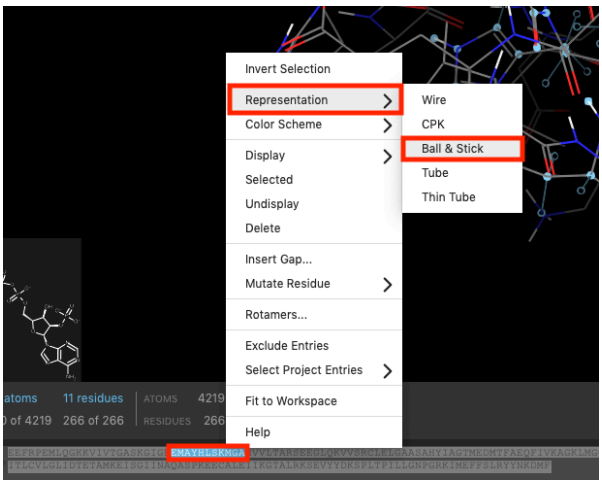


Figure 3-3. Change the representation of the sequence selection.

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

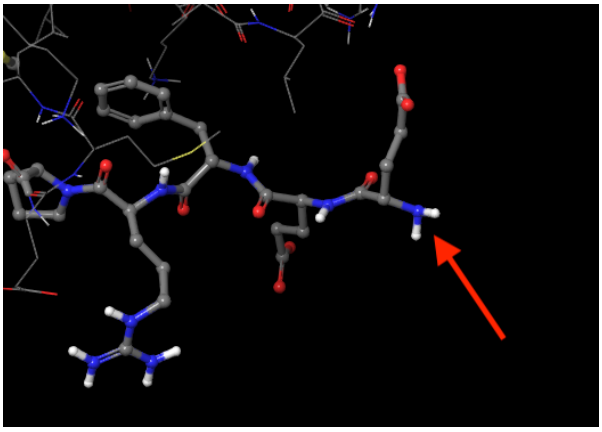


Figure 3-4. Highlight the N-terminus.

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

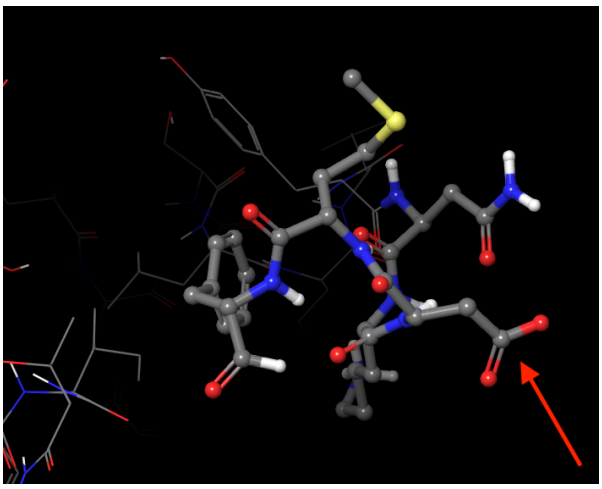


Figure 3-5. Highlight the C-terminus.

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.

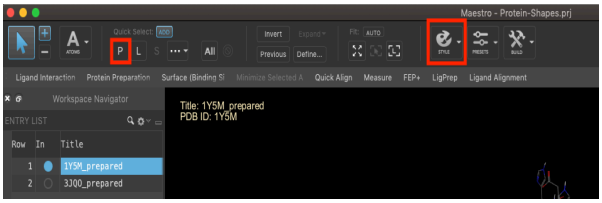


Figure 3-6. Open the styling tools.

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

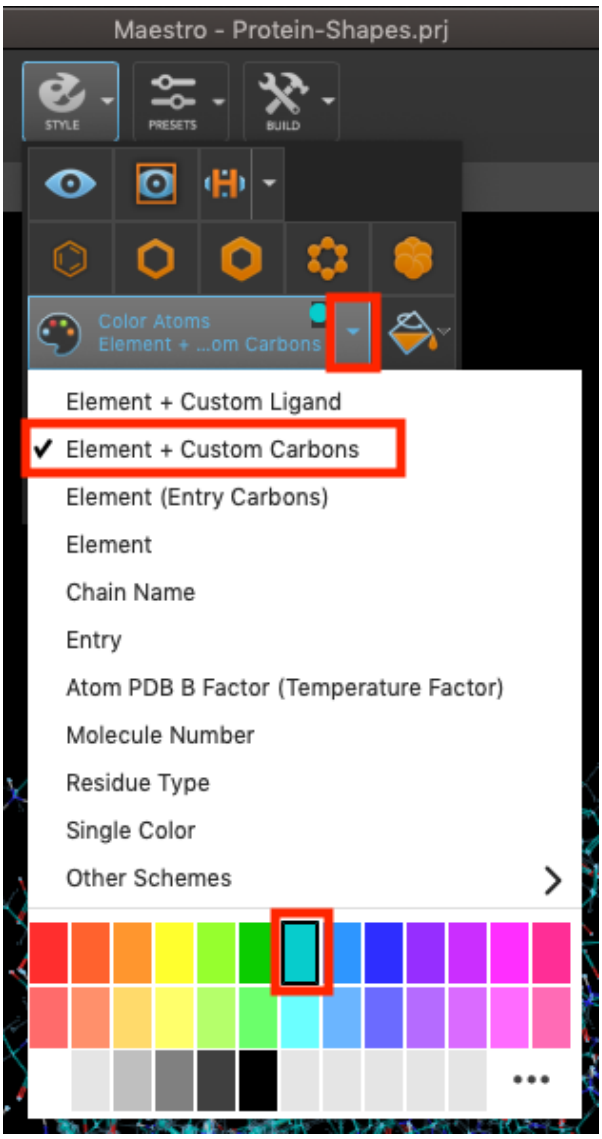


Figure 3-7. Coloring atoms in the Style toolbar.

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.

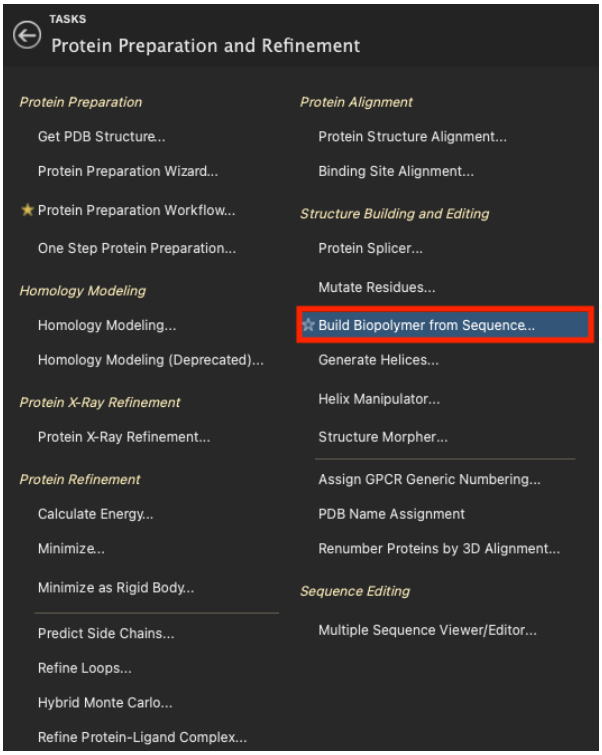


Figure 3-8. Open the Build Peptide from the Sequence panel.

1. Go to **Tasks** and type **Build Biopolymer from Sequence** in the search bar
    - The Build Peptide from the Sequence panel opens.

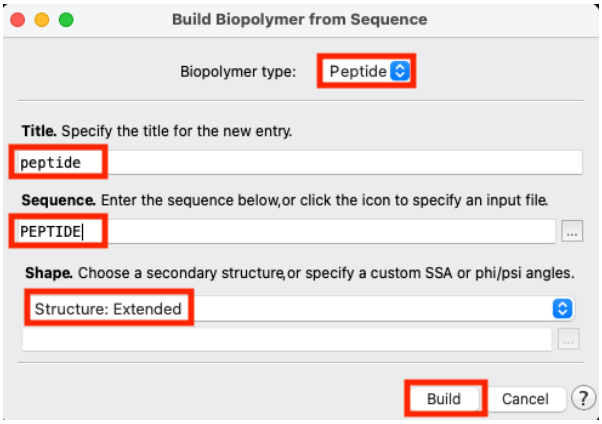


Figure 3-9. Build an extended peptide.

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
    - The peptide is shown in the Workspace
    - *Note:* Type **Z** if you cannot see the whole peptide.

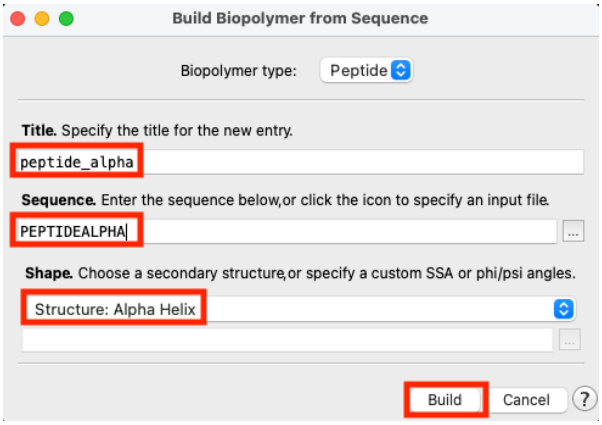


Figure 3-10. Build an alpha helix peptide.

6. Under Title, type **Peptide_alpha**
7. Under Shape, choose Structure: **Alpha Helix**
8. Click **Build**
    - A new entry titled Peptide_alpha is in the Entry List
    - The peptide is shown in the Workspace

Question #1: 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.

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. Go to **File > Import Structures** and import **1Y5M_prepared-out.maegz** into your workspace
2. In the Entry List, include **1Y5M - prepared** in your Workspace
    - This protein has already been prepared for you.
3. Left-click the **protein sequence** in the Sequence Viewer, and drag the mouse over the sequence to highlight a protein sequence.

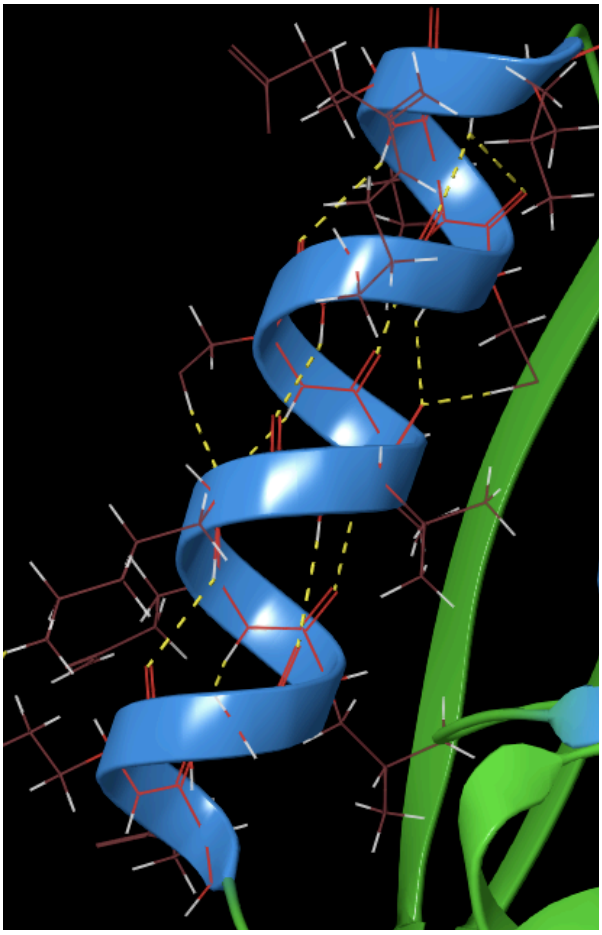


Figure 4-2. Visualizing hydrogen bonds in an alpha helix.

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 then click on the eyeball icon to visualize the residues and bonds in the highlighted sequence
    - *Note:* Interactions in the bottom right corner of the screen must be turned on.

Question: 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.

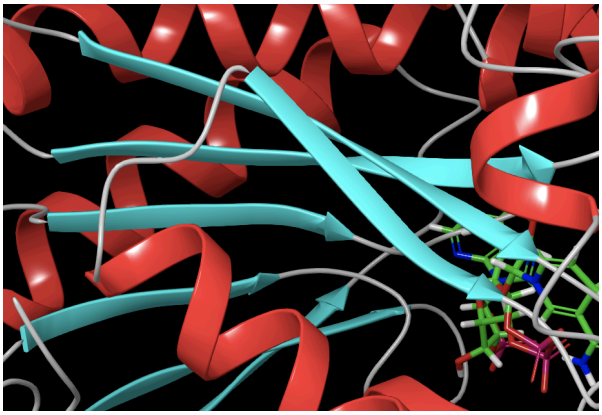


Figure 4-2. Visualizing a beta sheet.

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

Question: 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.

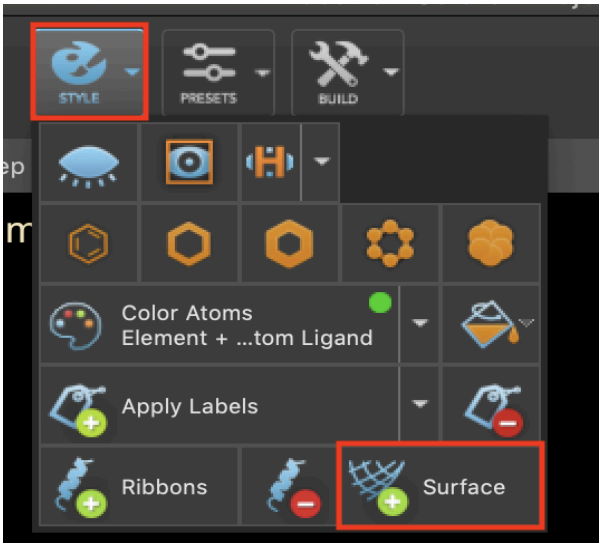


Figure 5-1. Opening the Style panel to add a Surface.

Part 1. Visualizing Protein Surfaces

1. Include **1Y5M - prepared** in the Entry List
2. 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

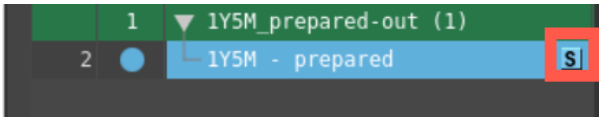


Figure 5-2. Visualizing a protein surface.

3. To edit the surface, click the **S** button to the right of the entry name in the Entry List.
4. Select **Manage** and a new window will open

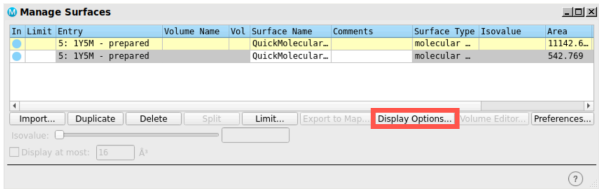


Figure 5-3. Opening the Manage Surface window.

5. Select **Display Options**

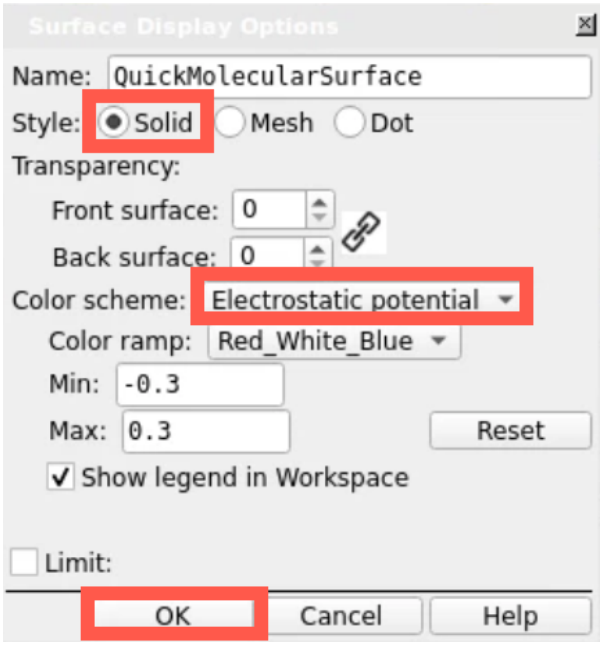


Figure 5-4. Adjusting Surface Display Options.

6. Change the **Style** to **solid**
7. Change the **Color Scheme** to **electrostatic potential**

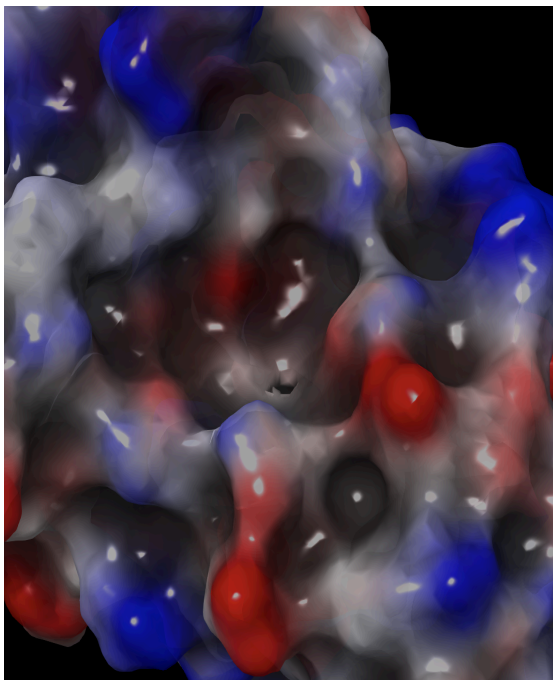


Figure 5-4. Visualizing the binding pocket of a protein.

Part 2. Visualizing Binding Sites

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

Color Analysis:

- **Red** = Negatively-charged residues
- **Blue** = Positively-charged residues
- **White/Gray** = Neutral residues

Question: 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 questions:

Question: 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: 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 learning:

- Alpha helices
- Protopedia
- RCSB
- Introduction to Protein Structure
- Enzymes

8. Glossary of Terms

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

Recent actions - This is a list of your recent actions, which you can use to reopen a panel, displayed below the Browse row. (Right-click to delete.)

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

Biological macromolecule - a large, organic molecule such as carbohydrates, lipids, proteins, and nucleic acids

Monosaccharide - a carbohydrate composed of one type of sugar

Disaccharide - a carbohydrate formed when two monosaccharides are joined together via a condensation reaction

Polysaccharide - a carbohydrate composed of a large number of monosaccharides

Glycosidic linkage - a type of covalent bond that joins a carbohydrate molecule to another group, which may or may not be another carbohydrate