Structure Analysis and Comparison Tutorial

This tutorial includes binding site analysis and comparison of related structures by superposition and morphing. Internet connectivity is required to fetch the structures **3w7f** and **2zco**.

- Background and Setup
- Distances, H-Bonds, Contacts
- Angles, Rotamers, Clashes
- Surfaces and Attributes
- Superposition and Morphing

← Background and Setup

The pathogenic organism *Staphylococcus aureus* makes a pigment called staphyloxanthin. The pigment imparts a golden color (hence *aureus*), but more importantly, contributes to virulence by protecting the bacteria from being killed by the host immune system. The *S. aureus* enzyme CrtM may be a good drug target because it catalyzes a key step in staphyloxanthin synthesis:

> A cholesterol biosynthesis inhibitor blocks *Staphylococcus aureus* virulence. Liu CI, Liu GY, Song Y, Yin F, Hensler ME, Jeng WY, Nizet V, Wang AH, Oldfield E. *Science*. 2008 Mar 7;319(5868):1391-4.

We will view and compare different structures of this enzyme.

Start Chimera by clicking or doubleclicking the **Chimera** icon ℚ (depending on its location). Typically, this icon will be present on the desktop. The Chimera executable can also be run from its installation location (details...).

A splash screen will appear, to be replaced in a few seconds by the main Chimera graphics window or **Rapid Access** interface (it does not matter which, the following instructions will work with either). If you like, resize the Chimera window by dragging its lower right corner.

Show the **Command Line** (**Tools... General Controls... Command Line**). Choose **Favorites... Add to Favorites/Toolbar** to place some icons on the toolbar. This opens the **Preferences**, set to **Category**: **Tools**. In the **On Toolbar** column, check the boxes for:

- **Side View** (under Viewing Controls)
- **FindHBond** (Structure Analysis)
- **Find Clashes/Contacts** (Structure Analysis)
- **Rotamers** (Structure Editing)
- **Reply Log** (Utilities, the last section)
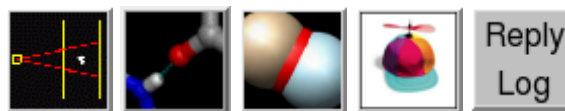
If you want these icons to appear in later uses of Chimera, click **Save** before closing the preferences.

Fetch a structure from the Protein Data Bank and use the ribbons preset:

Command: **open** **3w7f**
Menu: **Presets... Interactive 1 (ribbons)**

This preset displays the two protein chains (A and B) as red and blue ribbons; bound molecules and nearby sidechains are shown as sticks. The two chains are two copies of the enzyme. Delete chain B:

Command: **delete** **:.b**

Move and scale structures using the mouse and **Side View** as desired throughout the tutorial.

active site

The enzyme combines two 15-carbon molecules of farnesyl pyrophosphate to form a 30-carbon lipid. This structure contains farnesyl *thio*pyrophosphate, which differs from the substrate by having a sulfur in the place of one oxygen. Sulfur atoms are shown in yellow, phosphorus orange, oxygen red, and nitrogen blue. Label the ligand residues:
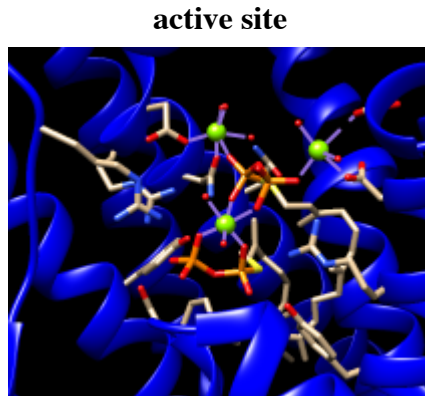
Command: **rlabel** **ligand**

In this structure, the farnesyl thiopyrophosphate molecules are named FPS. Three magnesium ions help to offset the negative charges on the phosphates. The ions are shown as greenish spheres; clicking into the Chimera window and hovering the mouse cursor over each shows information in a pop-up balloon. Metal coordination bonds from FPS, water, and the protein are shown with dashed purple lines. Lines drawn to indicate interactions other than covalent bonds are called *pseudobonds*. Hovering the cursor over a pseudobond or bond shows balloon information about its end atoms and length.

Delete the water, label residues with displayed sidechains, and place the labels near α-carbons instead of residue centroids:

Command: **del** **solvent**
Command: **rlab** **@/display**
Command: **setattr** **m residueLabelPos 2**

Water could be included in the various analyses, but is removed here to simplify the tutorial.

← Distances, H-Bonds, Contacts

It looks like several sidechains could be donating hydrogen bonds to phosphate oxygens. (Although the structure does not include hydrogens, we know they are there!)

One of the displayed residues is Ser 21. To measure a distance:

1. Ctrl-click to pick the sidechain oxygen of Ser 21
2. Shift-Ctrl-doubleclick on the nearest phosphate oxygen
3. click **Show Distance** in the resulting context menu

Similarly, measure the distance between the sidechain oxygen of Tyr 248 and the same phosphate

oxygen.

The distances seem consistent with hydrogen bonds. However, rather than measuring many distances and trying to remember the appropriate hydrogen-bonding distances for different types of atoms, just use **FindHBond**. We will limit the search to H-bonds involving the FPS residues:

1. Select the FPS residues (for example, with **Select... Residue... FPS** in the menu).
2. Start **FindHBond** by clicking its icon:
3. In that dialog, turn on the options:
   - **Only find H-bonds with at least one end selected**
   - **Write information to reply log**
4. Set **Line width** to 3.
5. Click **OK**.

The H-bonds are shown as pseudobonds of the specified color and line width. Both measured distances were found to be consistent with hydrogen bonding. Details can be viewed in the **Reply Log**, which can be opened by clicking its icon. (In your own work, you may prefer to write the information to a file instead.) Remove the "hydrogen bond" pseudobonds:

> *Command*: **~hbond**

Find Clashes/Contacts is similar to **FindHBond** but can also identify nonpolar interactions:

- *clashes* - unfavorable interactions where atoms are too close together; close contacts
- *contacts* - all kinds of direct interactions: polar and nonpolar, favorable and unfavorable (including clashes)

We will identify contacts of the FPS residues:

1. Start **Find Clashes/Contacts** by clicking its icon:
2. With the FPS residues still selected, click **Designate**. Now, 48 atoms should be designated for checking against **all other atoms**.
3. Set the **Clash/Contact Parameters** to the default **contact** criteria (by clicking the button marked **contact**).
4. Set the **Treatment of Clash/Contact Atoms** to:
   - **Select**
   - **If endpoint atom hidden, show endpoint residue**
   - **Write information to reply log**
   - (turn off any other treatment options)
5. Click **Apply**.

In the **Reply Log**, the atom-atom contacts are listed in order of decreasing VDW *overlap*, the amount of intersection of atomic van der Waals spheres: 0 if just touching, positive if interpenetrating, negative if separated by space. Slightly separated spheres are still considered contacts by default. The distances between atomic centers are also given.

One might simply want a list of the interacting residues rather than the details of each atomic contact. A list of the selected residues can be saved. First, deselect the ligand residues and ions, leaving the protein

residues selected:

> *Command*: **~sel** **ligand**
> *Command*: **~sel** **ions**

Open the **Selection Inspector** by clicking the magnifying glass icon 🔍 near the bottom right corner of the main window. It reports that 26 residues are selected. Click its **Write List...** button (or choose **Actions... Write List** from the main menu). In the resulting dialog, indicate that **selected residues** should be written. Click **Log** to write the list to the **Reply Log** instead of to a file.

← Angles, Rotamers, Clashes

Amino acid torsion angle values can be viewed in the **Selection Inspector**. Focus on Tyr 248:

> *Command*: **focus** **:248**

Ctrl-click to pick any atom or bond in Tyr 248. This will replace (clear) any pre-existing selection. Open the **Selection Inspector** again if it is not already open. Inspecting **Residue** (instead of **Atom**) shows that the sidechain angles chi1 and chi2 in this residue are approximately –107° and –142°, respectively. The backbone angles phi and psi are also given. **Close** the **Selection Inspector**.

We will set up clash checking and rotate the sidechain interactively.

clash checking

1. Display residues within 4 Å of Tyr 248:

   > *Command*: **disp** **:248 z<4**

2. Click into the main graphics window, then press the up arrow key on the keyboard to promote the atom or bond selection within Tyr 248 to the whole residue.

3. Start **Find Clashes/Contacts** by clicking its icon:

4. In that dialog, click **Designate**. The 12 atoms of Tyr 248 should be designated for checking against **all other atoms**.

5. Set the **Clash/Contact Parameters** to the default **clash** criteria (by clicking the button marked **clash**).

6. Set the **Treatment of Clash/Contact Atoms** to **Draw pseudobonds**; turn off any other treatment options.

7. Set the **Frequency of Checking** to **after relative motions**.
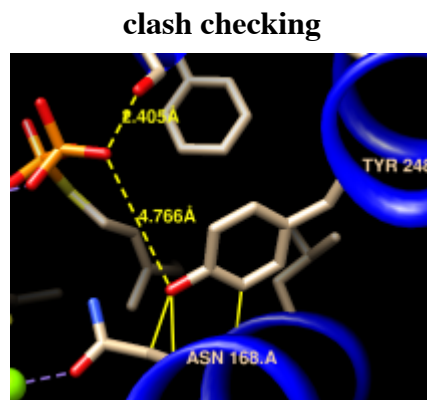
No clashes are found in the current position. To rotate the sidechain of Tyr 248 interactively:

1. Ctrl-doubleclick its CA-CB bond (the stick segment right next to the ribbon)
2. click **Rotate Bond** in the resulting menu

The rotatable bond will be listed in the **Adjust Torsions** dialog. In that dialog, if the **Near** atom is set to **N**, the value reported is the chi1 angle. There are several ways to rotate a bond, including dragging the pointer on the dial, clicking the black arrowheads, and editing the angle value directly. If the dialog

becomes obscured by other windows, it can be raised again from the menu (**Tools... Structure Editing... Adjust Torsions**).

Use whichever method you prefer to rotate the bond. As the sidechain moves, yellow pseudobonds show any clashes, and the distance monitor updates automatically. The sidechain is fairly constrained; only chi1 angles of approximately –(100–120)° avoid clashes if only that bond is rotated. The sidechain can be frozen in a new position, but for tutorial purposes, simply restore it to its original position: in the **Adjust Torsions** dialog, click the entry under **Bond** to show a menu. In that menu, choose **Revert** (move back to original position) and then **Deactivate** (make the bond no longer rotatable). **Close** the torsion dialog.

Click the **Find Clashes/Contacts** icon to bring the dialog to the front and **Close** it to halt clash checking.

Next, we will compare the conformation of Tyr 248 in the structure to tyrosine rotamers from a library. This will indicate whether the sidechain is in a frequently observed conformation and whether other conformations might also fit into the structure.

With part of Tyr 248 still selected (if not, Ctrl-click to to select any of its atoms or bonds), start **Rotamers** by clicking its icon . Click **OK** to show **TYR** rotamers from the **Dunbrack** backbone-dependent library. The rotamers are shown in the wire representation and listed in another dialog.

Clicking a line in the rotamer dialog displays just that rotamer. Clicking the lines one by one shows that none of the rotamers match the conformation in the structure.

The rotamer dialog reports chi1 and chi2 angles and probabilities from the library based on the residue's backbone conformation. Other than the backbone angles, the probabilities do not take into account any interactions specific to this structure. However, the rotamer dialog is integrated with clash and H-bond detection. Choose **Columns... Add... Clashes** and click **OK** to use the default settings; repeat with **Columns... Add... H-Bonds**.

The new columns show that each rotamer forms several clashes but no hydrogen bonds. We showed above that Tyr 248 H-bonds with the ligand and avoids clashes, which may compensate for its nonrotameric (presumably strained) conformation.

A sidechain can be replaced with a chosen rotamer. With a single rotamer shown, either click **OK** in the rotamer dialog to replace the sidechain with that rotamer and remove the others, or simply **Close** the dialog to remove the rotamers without replacing the sidechain. Although not done in this case, rotamers can be shown for a different amino acid type than in the structure, allowing for virtual mutations.

Unlike Tyr 248, Tyr 41 closely resembles the highest-probability tyrosine rotamer given its backbone angles. If you like, focus on Tyr 41, select one of its atoms or bonds, and use **Rotamers** to show and evaluate rotamers of tyrosine or some other amino acid at that position.

Clear the selection, unlabel the residues, zoom back out, and remove the distance pseudobonds:

> *Command*: **~sel**
> *Command*: **~rlab**
> *Command*: **focus**

Command: **~dist**

← Surfaces and Attributes

FPS and the natural substrate both have a highly polar/charged "head" and a long hydrocarbon "tail." The enzyme pocket is very large and deep, as needed to accommodate two of these molecules.

A surface representation is best for showing the shape of the pocket. Choose **Presets... Interactive 3 (hydrophobicity surface)** from the menu to display a molecular surface color-coded by amino acid hydrophobicity. This preset colors the surface ranging from dodger blue for the most polar residues to orange red for the most hydrophobic, with white in between.

As one might expect, the mouth of the pocket near the phosphates and ions is mostly polar, while the rest of the pocket is largely hydrophobic. However, the pocket is so deep that it is hard to see when the whole surface is shown. Restricting the surface display to the pocket region allows it to be viewed from the outside, as shown in the figure.

Hide the ribbon and the protein atoms, then show only surface near the ligand:

pocket surface
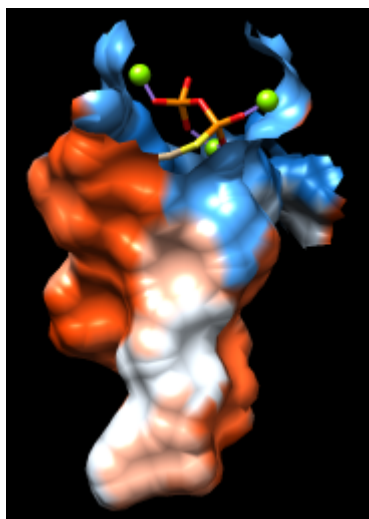
Command: **~ribbon**
Command: **~disp** protein
Command: **sop** zone #0 ligand 5.0 max 1

The **sop** ("surface operations") command limits the display of surface #0 to a 5.0-Å zone around the ligand and shows only the largest resulting fragment. Without **max 1**, a few additional disconnected shards of surface would be shown. If you like, try different cutoff values; the figure shows results for a cutoff of 4.7 Å.

Amino acid hydrophobicity is one example of an *attribute*. Another attribute commonly used in structure analysis is atomic B-factor. B-factor values indicate which parts of the structure are more or less flexible or disordered.

Display all atoms by choosing **Presets... Interactive 2 (all atoms)** from the menu. Now the ligand atoms are shown as spheres and the protein as wires. Show the protein as sticks:

Command: **repr stick protein**

The **Render by Attribute** tool shows attribute values with colors and other visual cues. Choose **Tools... Depiction... Render by Attribute** from the menu to start this tool. The dialog is set to **atoms** when it appears. Choose **bfactor** from the attribute menu. The B-factor values are shown in a histogram, and the vertical bars on the histogram define a color mapping. These bars or *thresholds* can be dragged along the histogram with the mouse, and their colors can be changed. Clicking the square color well allows changing the **Color** of the threshold that was most recently clicked or moved. The colors can also be changed collectively by choosing a different **Palette** or clicking the button to **Reverse threshold colors**. (The figure shows the dialog after a user has added more thresholds, chosen the **Rainbow** palette, and then reversed the colors.) Adjust the color mapping as desired and **Apply** it to the

structure. The lowest B-factors are in the protein core, the highest in a loop over the active site and the C-terminus on the opposite side. If any sidechain has been replaced with a rotamer, the new atoms will not have B-factor values, but they can be assigned a **No-value color**.

In the **Render** dialog, switch from **atoms** to **residues** and show the histogram for the attribute **kdHydrophobicity**, the Kyte-Doolittle hydrophobicity of amino acids. This is the same attribute as used by the surface preset, but different coloring schemes can be applied. Larger positive values correspond to more hydrophobic residues, negative values to hydrophilic residues. The **No-value color** pertains to residues without Kyte-Doolittle hydrophobicities, namely any that are not amino acids (such as the ligands in this structure). The colors can be changed as described above. Apply coloring by hydrophobicity if you like, then **Close** the dialog.

← Superposition and Morphing

Comparing different structures of a protein is another way to evaluate flexibility. We have been viewing a structure bound to substrate analogs, **3w7f**. A structure of the same enzyme without ligands is also available. Fetch the "empty" structure and apply the ribbons preset:

> *Command*: **open** **2zco**
> *Command*: **preset** **apply int 1**

Now the first structure is tan and the new structure is sky blue, as shown in the **Model Panel** (**Tools... General Controls... Model Panel**).

The structures are in completely different positions, so the next step is to superimpose them. This is easily accomplished with **MatchMaker** or its command equivalent:

> *Command*: **mm** **#0 #1 show true**

MatchMaker first generates a sequence alignment using residue types and secondary structure (tries to align helix with helix and strand with strand), then fits the sequence-aligned residues in 3D by pairing their CA atoms. The **show true** option displays the sequence alignment. By default, the fit is iterated so that far-apart pairs are not included in the final match. This allows the most similar parts to superimpose more tightly and the dissimilar parts to stand out. Final match statistics are reported in the **Reply Log**, and the pairs used in the final fit are shown with light orange boxes on the sequence alignment. The RMSD histogram above the sequences shows spatial variability, calculated as the root-mean-square deviation among the CA atoms of residues associated with each column. In this case (only two structures), the RMSD value is the same as the CA-CA distance.

The structures are highly similar except for the loop at approximately residues 52-56; you can see
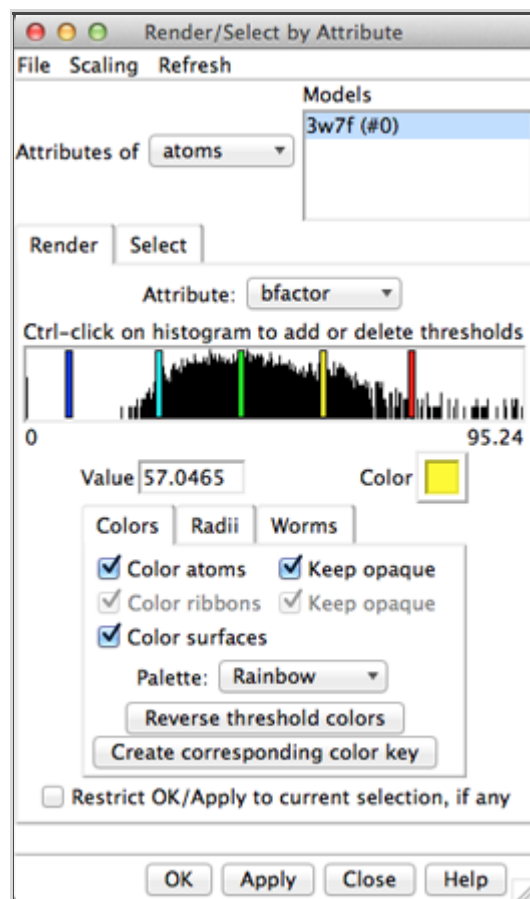
residue numbers by clicking into the sequence alignment window and hovering the cursor over the corresponding one-letter codes. Click any part of the light orange boxes to select all the residues used in the final fit; these form the common core, the most conformationally similar parts of the structures. The selection can be inverted if you want to do something to the less similar parts, for example:

different loop conformations

> *Command*: **sel invert**
> *Command*: **disp sel & protein**
> *Command*: **~sel**
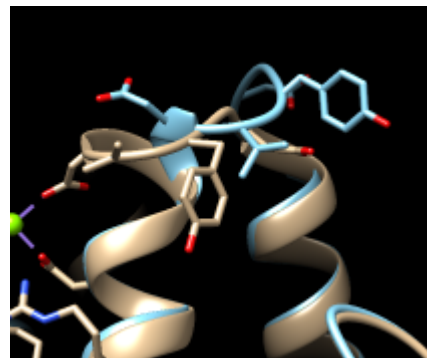
Morphing involves calculating a series of intermediate, interpolated structures between the original input structures. In Chimera, the series of structures is treated as a trajectory that can be replayed, saved to a coordinate file, or saved as a movie using **MD Movie**.

Start the morphing tool:

> *Menu*: **Tools... Structure Comparison... Morph Conformations**

Click **Add...** and in the resulting list of models, doubleclick to choose #0, #1, and #0 again, corresponding to a morph trajectory from the ligand-bound structure to the empty structure and back. **Close** the model list. In the main **Morph Conformations** dialog, set the **Action on Create** to **hide Conformations**, and then click **Create**.

The progress of the calculation is reported in the status line. When all the intermediate structures have been calculated, the input structures are hidden, the trajectory is opened as a third model (#2), and the **MD Movie** tool appears. The trajectory can be played continuously or one step at a time by clicking buttons on the tool, or a frame number can be entered directly. If the **MD Movie** playback dialog becomes obscured by other windows, it can be raised using its instance in the **Tools** menu (near the bottom of the menu, below the horizontal line).

The initial display of the morph trajectory is as ribbons rather than all atoms, but all atoms that are present in both input structures are also present in the morph. Their display can be controlled just as in any other structure, for example:

> *Command*: **disp #2:52-56**

Ligand residues were only present in one of the structures, so they are not included in the morph trajectory. However, part or all of the original structures can be displayed along with the trajectory. Display of the other models can be re-enabled by checking the **S**(hown) boxes in the **Model Panel** (**Tools... General Controls... Model Panel**) or with a command, for example:

> *Command*: **modeldisp #0**

Then, to show only the ligand residues and ions from that model:

> *Command*: **~ribbon #0**
> *Command*: **show :fps,mg**

In this case, the secondary structure does not change much between the input structures. When conformational differences are larger and ribbons will be displayed, one may want to re-evaluate the secondary structure at each step of the trajectory. That can be done by putting the command **ksdssp** in a per-frame Chimera script in **MD Movie**. Such scripts also allow performing actions (such as showing/hiding a ligand) at specific steps in the trajectory.

The **Morph Conformations** page lists a few additional example systems.

When finished enjoying the morph trajectory, choose **File... Quit** from the menu to exit from Chimera.