1. Protein preparation and homology modeling

BSR3101: Computer Aided Drug Design

The goal of this tutorial is to build homology model of GP171 (the "target" structure), a G-protein coupled receptor, from a homologous P2Y12 receptor crystal structure (the "template").

The tutorial consists of the following steps:

1. Open maestro and save a new project
2. Import structure of template from PDB
3. Change structure representations
4. Inspect the PDB file and delete superfluous components
5. Prepare the protein-ligand complex
6. Inspect protein ligand interactions
7. Homology modeling

What will you learn:

- How to visualize proteins and ligands (color, label, measure, display) in the Maestro 11 interface
- Protein structure modeling, refinement, and analysis.
- Binding site detection and characterization
- How to build a homology model

Review before you start:

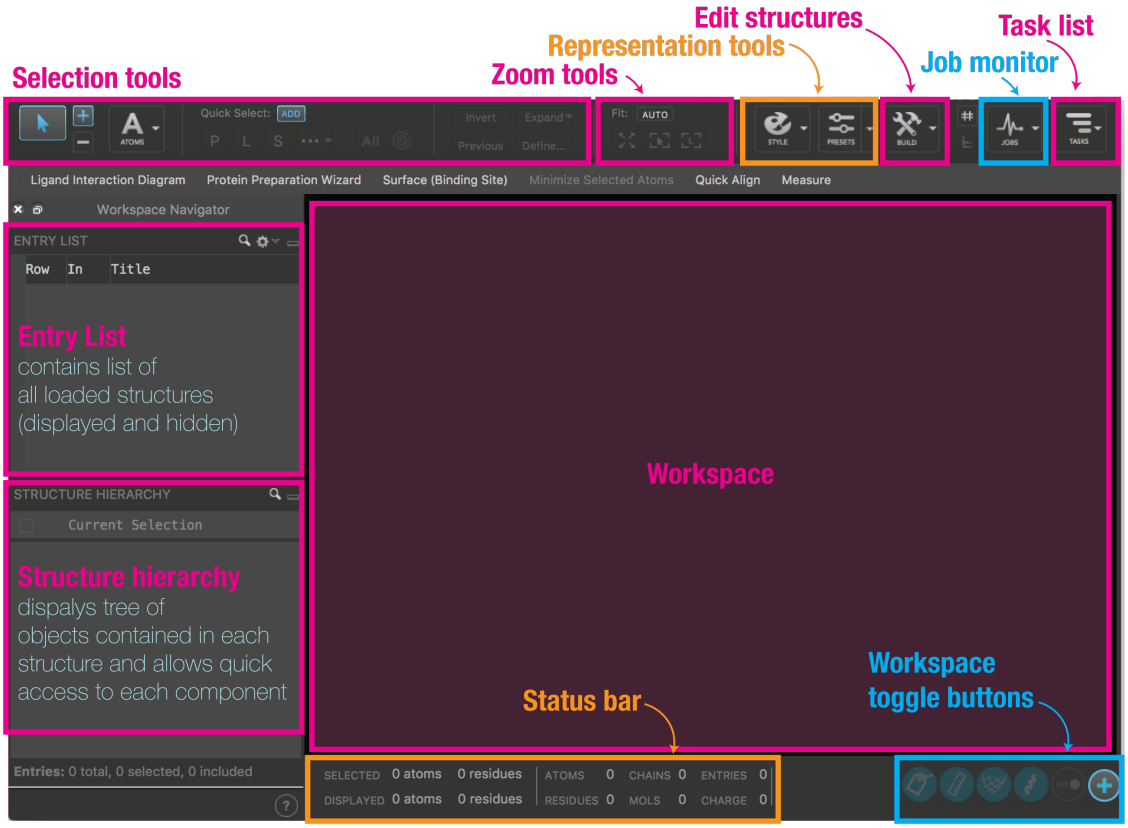
- Typical ligand-receptor interactions
- Protein Data Bank and Uniprot repositories
- Sequence alignment algorithms
- Homology modeling

Approximate time required to complete this tutorial:
30 minutes.

1. Open maestro and save a new project

First, download the tutorial folder **`CADD_tutorial_01`** from the link https://goo.gl/h3djH6, move the downloaded archive to the Desktop and extract it. This folder contains example output files that can be used to compare the results of the exercises executed in this tutorial.

Next, open Maestro by double clicking the Maestro icon on your desktop. The Maestro main window has many different components and their titles used throughout the tutorial are highlighted below.



Set the working directory to the tutorial folder by clicking on the menu bar File > Change Working Directory… navigate to **`Desktop/CADD_tutorial_01/`** and click the Choose button. All the files generated will be saved in this folder.
Since every newly opened Maestro session is considered a scratch project by default, click the File > Save Project As… common and save your project: type **`protein_prep`** in the file name field and click the Save button (or press Enter).

2. Import structure from PDB

The P2Y12 protein structure, which was solved in 2014, will serve as a template to build the model of GP171, for which no structure has been solved yet. We will start by importing the P2Y12 protein structure from Protein Data Bank (PDB), where it has been deposited and is identified by the PDB ID 4PXZ.

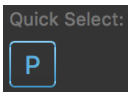
An easy way to proceed is using the File > Get PDB command, typing 4PXZ in the PDB ID field, and clicking the Download button (or press Enter).

After the 4PXZ structure file has been loaded, an entry named 4PXZ will appear in your entry list, and the protein structure will be shown in the workspace on the right. The experimental structure contains, apart from the receptor structure, several other molecules: a ligand bound to the receptor (2-methylthio-adenosine-5'-diphosphate), lipid molecules that were identified in the diffraction data, additional protein that was used to aid the crystallization process.

3. Visualize protein structures with different representations

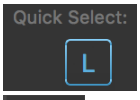
Changing representations. The default representation is the "wire". We will now select different representations so we can better visualize the P2Y12 structure. Specifically, we will represent the protein in "Ribbon" and the non-protein atoms in the structure in the "Tube" representation.

- Click on the protein quick select "P" button to select the protein. Open the "Style" panel, and click on the "Show ribbons for selected residues" button to turn on the ribbon representation for the protein backbone.



- Make sure the protein is still selected (for instance, have a look at the structure hierarchy panel), and click on the "Undisplay selected atoms" in the Style panel, to remove the wire representation of the protein.

- Now select the ligands, clicking on the "L" quick select button.

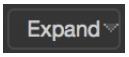


- In the style panel, click the Display only icon so that only the selected atoms (the ligand atoms in our case) are shown (besides the ribbon representation).

- In the style panel, click the Tube icon to the left of the entry list, to render the selected atoms in tube representation.

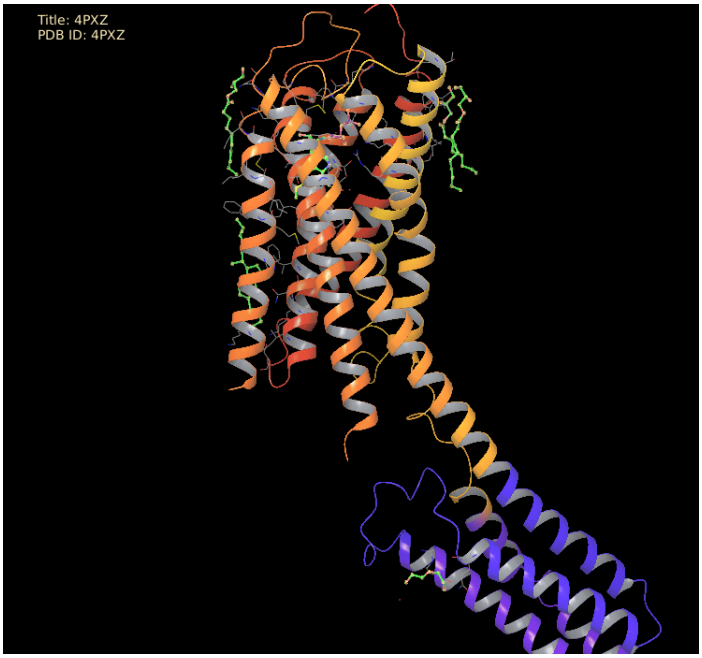
- Click the Center icon (or press the Z key) to center the selected atoms in the workspace.

- Make sure the ligands are selected (press again the "L" quick select, or click in the structure hierarchy), then slick on the "Expand selection…" menu in the selection toolbar, and click on +4 Å to select residues within 4 Å of the ligands.



- Once selected, represent them as "tubes" using the corresponding button in the Style panel. Remember to show the selected atoms by clicking on the "Display" button as well.

Navigating the structure. You can rotate the structure (keep the mouse scroll-wheel pressed and move the mouse), translate it (keep the mouse right-button pressed and move the mouse), or zoom (rotate the scroll wheel). You can automatically zoom on the whole structure pressing the Z key, and focus on each ligand pressing repeatedly the L key.
You workspace should now look similar to the picture below.



Additional exercises. Take a moment to navigate the structure and familiarize with its topology. Can you identify where the membrane would interact with the protein? Where is the main ligand-binding pocket?

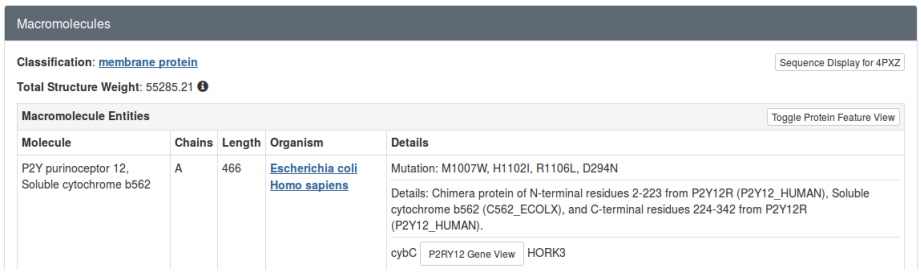
Using what you have learned about selection and representation, show the oxygen atoms of crystal waters as spheres (Select > Waters, then double click on the CPK icon in the Representation toolbar).

We can also inspect the protein for potential structural issues. Click on Tasks and type "Protein Analsysis" in the search bar: then click on Protein Analsysis … > Protein Reports to analyze the structure. Select "Bond Length Deviation" first and "Bond Angle Deviation" subsequently in the Display field. Click on each entry to zoom into the corresponding area in the structure. Do the same with the "Missing Atom" entry.
We will solve these issues before modeling the target receptor.

4. Inspect the PDB file and delete superfluous components

We will now identify and delete superfluous molecules (lipids and proteins). In your browser, open the PDB website http://www.rcsb.org/pdb/ and type the PDB ID of our protein (4PXZ) in Search by PDB ID field (top right). The resulting page reports several information about the proteins present in the file, as well as additional molecular entities.

- In the Macromolecules section, you can identify that the construct used includes an insertion of the cytochrome b562 protein, added to facilitate the formation of crystals



- In the Small Molecules section, we can see that the structure contains 4 non protein molecules: the bound ligand (identified by the 6AD residue name), and three lipid/solvent molecules (CLR, OLC and PEG)

To remove the superfluous atoms, we need to identify the atoms numbers related to the residues of the cytochrome and the small-molecules.
Display the raw PDB file by clicking on the Display files > PDB file in the top right area of the Structure summary. Scroll down the PDB file and write down the chain ID and the residue number of these components for deletion

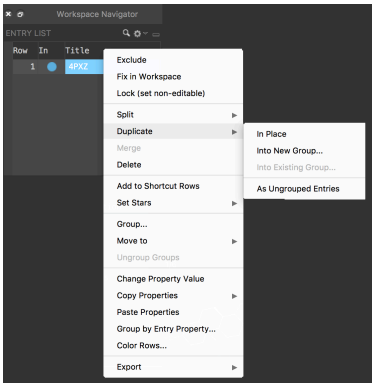
- The **DBREF** tags show that residues in chain A from residue 1001 to 1106 correspond to the cytochrome (while residues 2-223 and 224-342 correspond to the receptor).

```
REMARK 900 RELATED ID: 4NTJ   RELATED DB: PDB
REMARK 900 RELATED ID: GPCR-87   RELATED DB: TARGETTRACK
REMARK 900 RELATED ID: 4PY0   RELATED DB: PDB
DBREF  4PXZ A    2   223  UNP    Q9H244   P2Y12_HUMAN      2    223
DBREF  4PXZ A 1001  1106  UNP    P0ABE7   C562_ECOLX      23    128
DBREF  4PXZ A  224   342  UNP    Q9H244   P2Y12_HUMAN    224    342
SEQADV 4PXZ ASP A   -9  UNP  Q9H244              EXPRESSION TAG
SEQADV 4PXZ TYR A   -8  UNP  Q9H244              EXPRESSION TAG
SEQADV 4PXZ LYS A   -7  UNP  Q9H244              EXPRESSION TAG
```

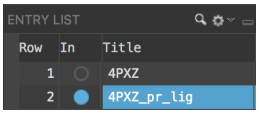
- A few lines below, the **HET** tags show that the residue in chain A with 1201 corresponds to the bound ligand (6AD), while residues 1202 to 1206 correspond to the crystallized lipids.

```
SEQRES  36 A  466  PRO MET GLY ARG PRO LEU GLU VAL LEU PHE GLN
HET    6AD  A1201      29
HET    CLR  A1202      28
HET    OLC  A1203      17
HET    OLC  A1204      17
HET    OLC  A1205      15
HET    PEG  A1206       7
HETNAM     6AD 2-(METHYLSULFANYL)ADENOSINE 5'-(TRIHYDROGEN
HETNAM   2 6AD  DIPHOSPHATE)
```

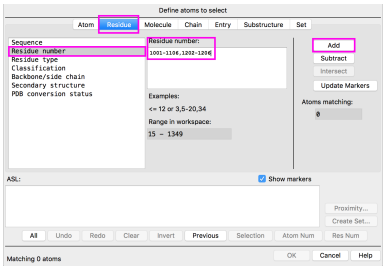
- Before deleting these atoms from our structure, let's duplicate the entry so as to keep a record of the original PDB file. In the entry list, right click on entry 1 (4PXZ), and select Duplicate > In Place. You would have the newly created entry 2, with the same name. Rename it by double clicking on the name, and typing **`4PXZ_pr_lig`**, followed by Enter.



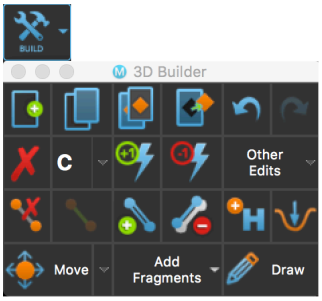
- Make sure the new entry (named **`4PXZ_pr_lig`**) is selected (the entry has blue background) and displayed in the workspace (filled circle in the **In** column in the entry list).



- Select residues corresponding to the lysozyme and lipids **by** clicking on the "Define…" button in the select panel. Go to the "Residue" tab, make sure the Residue Number entry is selected in the list, and type **`1001-1106,1202-1206`** in the Residue Number window. Click Add and the OK to select the residues.



- Open the "3D Builder" Panel and click on the "Delete" button.



The cytochrome and the lipids should have disappeared from the structure.

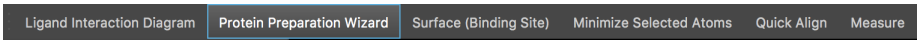
This is a good point to review the representations. Using what you have learned in section 3 ("Visualize protein structures with different representations"), make sure only the bound ligand is displayed in Tube, and the residues within 4 Å from it are shown.

We can also inspect more closely the interactions formed by the ligand and the protein. Use the L key to zoom in on the ligand. Hover with the mouse over the protein residues close to the ligand to see their number and type in the Status bar. How many interactions can you identify?

5. Prepare the protein—ligand complex

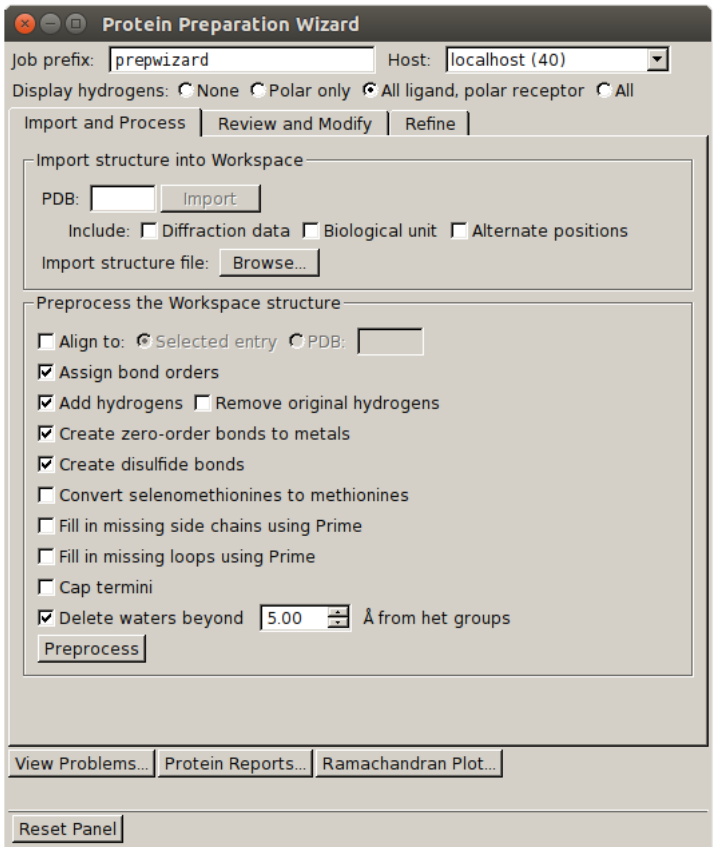
We will now prepare the protein structure by adding the hydrogen atoms (that are not resolved in x-ray crystallography and are therefore missing from the protein coordinates), and adding missing atoms that, because of poor resolution, were not resolved in the crystallographic reconstruction.

Make sure only the **`4PXZ_pr_lig`** entry is included in the workspace, open the Task panel, search for "protein preparation" keyword, and click on the Protein Preparation Wizard entry. (This can also be accessed directly on the toolbar above the workspace).



We will use the default settings (as shown below). Specifically, make sure that

- the Assign bond orders checkbox is selected
- the Add Hydrogens checkbox is selected
- the Create disulfide bonds checkbox is selected



Then click the Preprocess button. After a short processing, you should get warnings indicating that some atoms are missing in the crystal structure.

Click the Add Missing Side Chains button reconstruct the positions of the missing atoms. When the job finishes, all the warnings about missing atoms should disappear. Now click the OK button and close the Protein Preparation Wizard window.

Warning: this step could take up to 3 minutes on slow workstations.

You would now have 4 entries, with the last (and most recent) entry included in the workspace and selected. The previous 3 entries are the original crystal structure, the structure with unwanted components deleted, and the prepared structure before adding missing atoms.

6. Inspect protein ligand interactions

Once again, change representation of the ligand and residues surrounding it so that the ligand is shown as tubes and the surroundings are shown as lines.

- Hide non-polar hydrogens by selecting all objects, opening the style menu, and clicking on “Show Polar Hydrogens” button.
- Click on the Interactions toggle (bottom right) to show the hydrogen bonds between the ligand and the receptor. You can display and hide specific interaction types by opening the Interaction panel clicking on the “…” button on top of the Interaction toggle.

- Hover above the structure in the workspace to see residue information displayed in the status bar, and find residue N191, which is interacting with the ligand amine group. We would use this residue as an example for labeling residues we want to check more carefully.

- Make sure atom selection mode is active by clicking. Now you should see “A” (atom) instead of “R” (residue) shown on the Select icon.
- Click an atom of N191 (preferably one of the hydrogen atoms on its side chain amide nitrogen), then open the style panel and click on Apply Labels > Residue Information. Now the residue is labeled as “`ASN 191`”

7. Homology modeling

We will now build a homology model for GP171 based on the template we just prepared. We'll retrieve the aminoacid sequence of murine GP171 from the http://www.uniprot.org website, as a fasta format file.

- In your browser, open the http://www.uniprot.org website, and search for "**`GP171 mouse`**", then click on the entry with access code **`Q8BG55`** to open it.
- Scroll down to the Sequence section, and download a fasta file containing the sequence by clicking on the Fasta button, then Save As… and save the file with name **`Q8BG55.fasta`**

We are now ready to go back to Maestro and start the homology modeling task.

- In the Task panel, search for the keyword "homology modeling", and click on the Homology modeling > Structure Prediction Wizard > Open to open the structural prediction wizard.

The process is composed of a series of steps

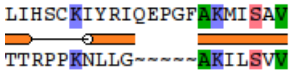
a. Input sequence
b. Find homologs
c. Edit Alignment
d. Build Structure

- First, we will select an **Input sequence**. Identify the Get sequence from area in the middle of the panel and click on the File… button to open the fasta sequence you saved in above as **`Q8BG55.fasta`**.
- Make sure the loaded sequence is highlighted in the Sequences text box, and click Next to proceed to the Find Homolog step.
- Make sure the prepared P2Y12 structure is selected in the Maestro window entry list. In the **Find Homolog** step, Import > From Project Table to load the selected entry in the project table (i.e. the sequence of the prepared P2Y12 structure) as sequence homolog.
- The sequence, as it is loaded, is approximately aligned to the query GP171 sequence. Click Next.
- If needed, click the Wrap Sequences icon to wrap the sequences (so you won't need to scroll from left to right to visualize the alignment).

- We can display secondary structure information from the template on the alignment, using the View SSA icon at the top of the Structure Prediction window.

- We can also mirror on the 3D structure in the main workspace the color-coding of conserved and non-conserved residues (between query and the template sequence), using the View Template Structure icon.

- In **Edit Alignment** step, we edit the existing, approximate alignment we got from last step, according to a more accurate alignment we obtained from multiple-sequence alignment. Specifically, we need to move 4PXZ sequence $_{135}$PKNLLG$_{140}$

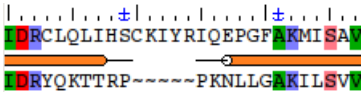


segment to the right, across the gap introduced by the automatic alignment.

- We do so by first anchoring the P_{129} and A_{141} residues in `4PXZ` sequence (you can hover on the aminoacids of the sequence to reveal their residue number in the form of **`(Xxx)AaaYyy`** where **`Xxx`** is the residue number in query sequence, **Aaa** is the residue name in the sequence your mouse is over, and **`Yyy`** is the residue number in the sequence your mouse is over.
- Click on the Anchor icon, then on P_{129} (a blue anchor sign indicates that the position is now anchored). Then click on A_{141} as well.

- Now we will push the $_{135}$PKNLLG$_{140}$ segment to the right by clicking on the Side Freely icon first, then click and hold on P_{135} and move the segment right until it closes the gap.

- The final alignment should look as indicated here. Click Next.



In **Build Structure** step, using the default setting (make sure the Knowledge based radio-button is selected, and that Return is set to 1 model), click Build Model.

Warning: this step could take up to 5 minutes on slow workstations. Be patient!

When the job finishes, decline the request for structure refinement (we can always do this step later by going to menu bar Task > Protein refinement). A new entry should have been added in the Maestro main workspace, and shown as ribbons. Blue color indicates residues for which all coordinates were taken from the template, cyan indicates residues with backbone coordinates taken from template, and red indicates residues that were rebuild ab-initio.

Change the representation of the newly created entry, and compare it with the 4PXZ crystal structure.