Tutorial: Virtual Ligand Screening and Protein-Ligand Binding

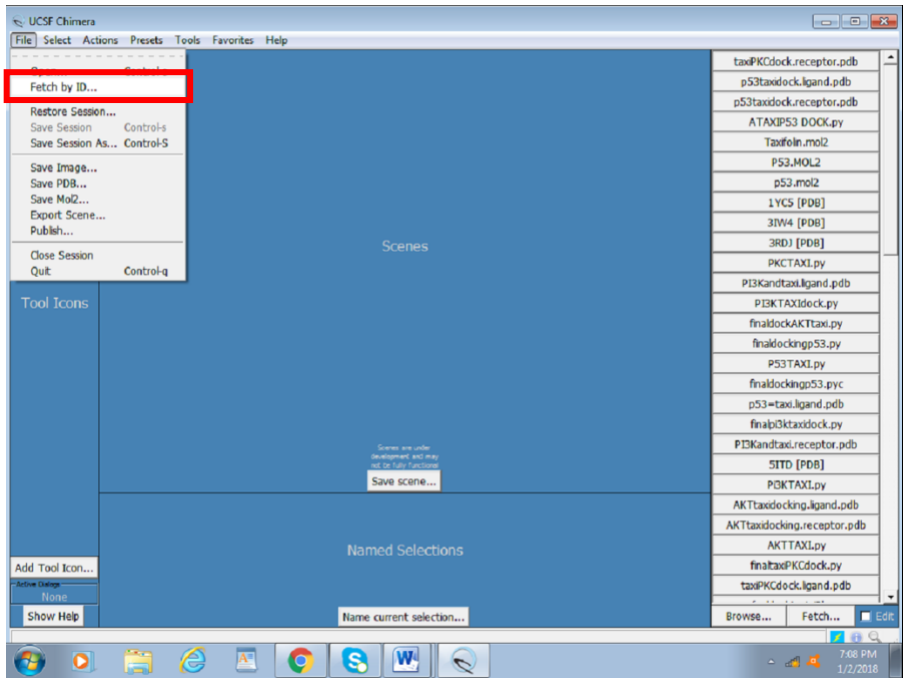
This tutorial will step-by-step guide a new user through the process of setting up a protein-ligand docking system using UCSF chimera containing a protein kinase B, also known as Akt, and a ligand fisetin and compute the protein-ligand binding affinity using Auto-Dock Vina.

Step 1. Create a working directory for the docking project that is convenient to access, such as *Users/Desktop/Docking/*.

Step 2. We must download the protein structure file we will be working with. For this tutorial, we will utilize the protein kinase (PDB code: 3QKK). Go to the RCSB website and download the PDB text for the crystal structure. Start saving all your prepared files in the working directory, for example, save 3QKK as Akt.pdb.

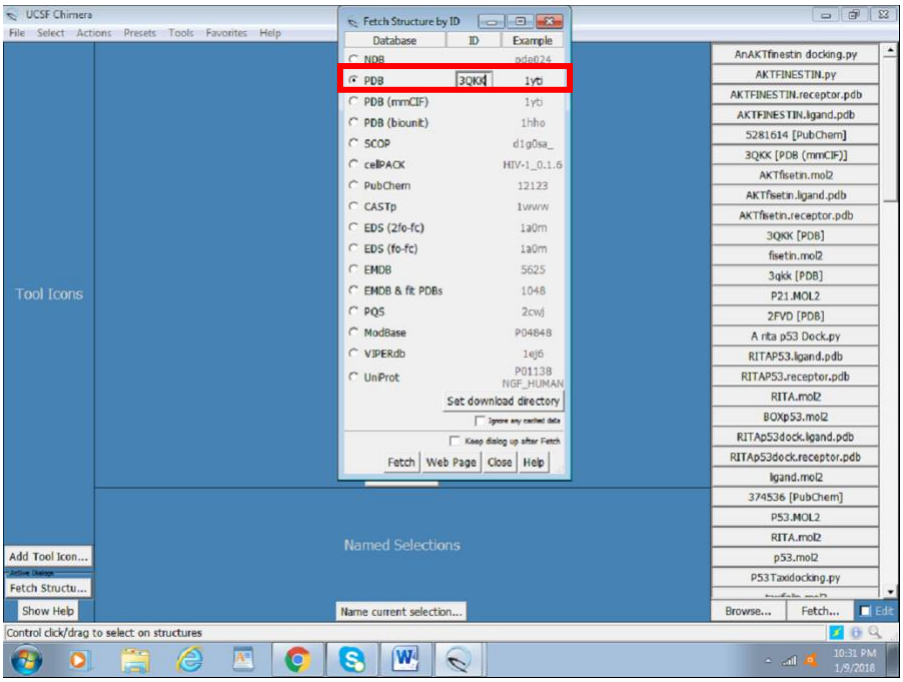
Step 3. Once you are ready and with your protein and ligand you want to bind, open UCSF Chimera.

Step 4. Click on the file and fetch by ID, as shown in the snapshot (outlined by a red box):



Step 5. Input the PDB ID of the protein (PDB ID: 3QKK).

The figure below presents a screenshot of how to obtain the protein structure through PDB ID in UCSF Chimera. Any protein can be fetched by inserting the PDB ID of the protein. Also, When the protein is fetched, its structure is downloaded through the website; hence, a working internet connection is required, or the PDB file can be downloaded beforehand and simply be opened thorough **File >open**.

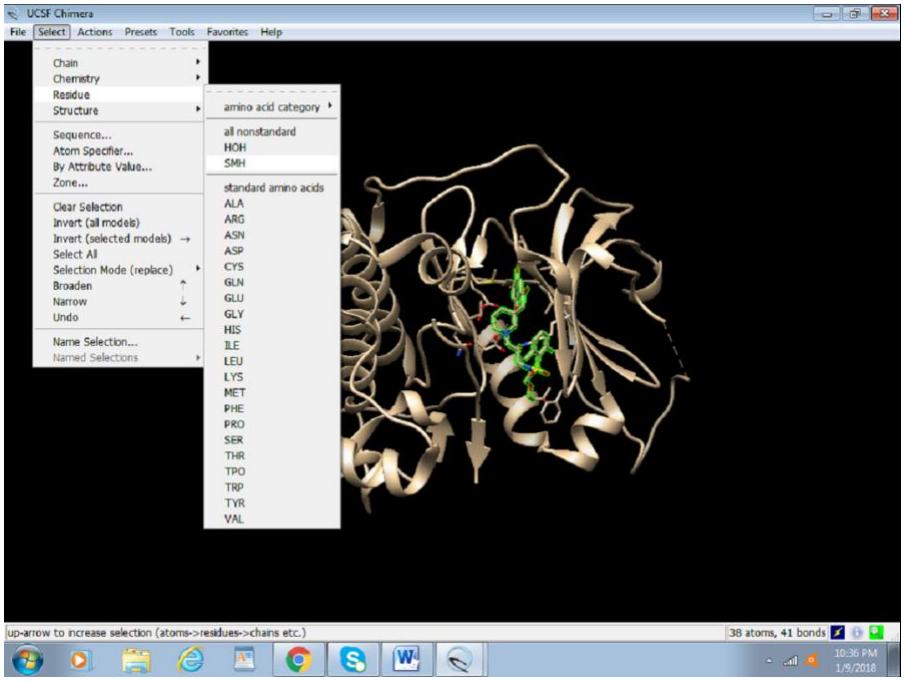


The following image displays the protein structure retrieved in UCSF Chimera. Once you have downloaded the structure, you can also visualize it using other a viewing program such as VMD, PyMOL, OVITO, etc.

Step 6. Once you have visualized, we must prepare the protein structure to be ready for the protein – ligand docking.

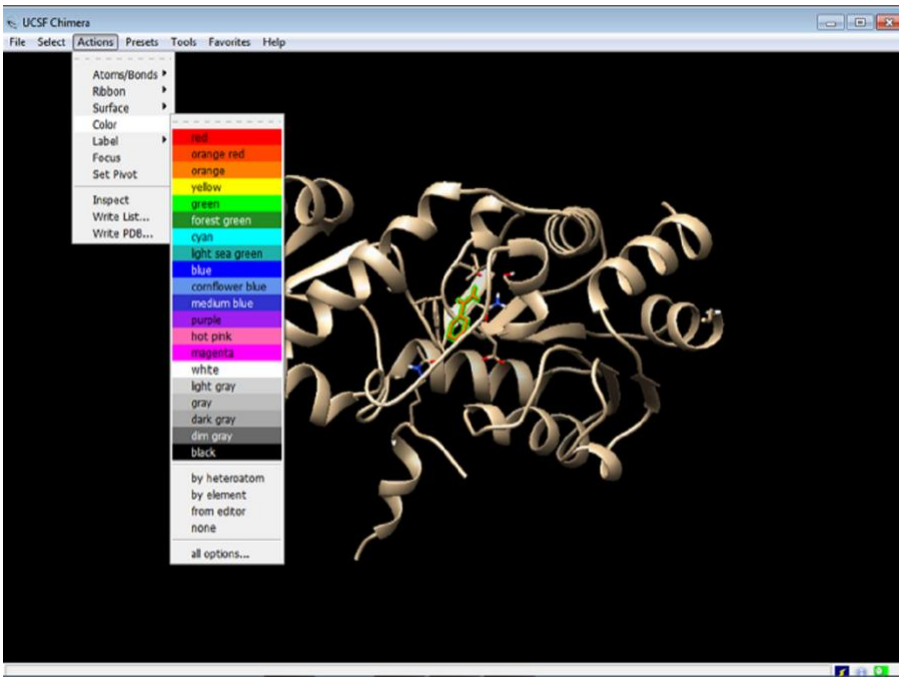
6.1. Protein preparation for docking:

1. To easily define the active site, the already present inhibitor needs to be identified. To do so, select the inhibitor by click on ***Select > Residue > SMH*** (nonstandard residue), as shown in the figure.
    - SMH typically denotes a nonstandard residue or an inhibitor within the protein structure. By selecting this, we can highlight and manipulate it separately from the standard amino acid residues.
    - Identifying and selecting the inhibitor helps in clearly defining the active site of the protein. Knowing the exact location and properties of the inhibitor (nonstandard residue) is crucial for studying its interaction with the protein and for accurate docking simulations.

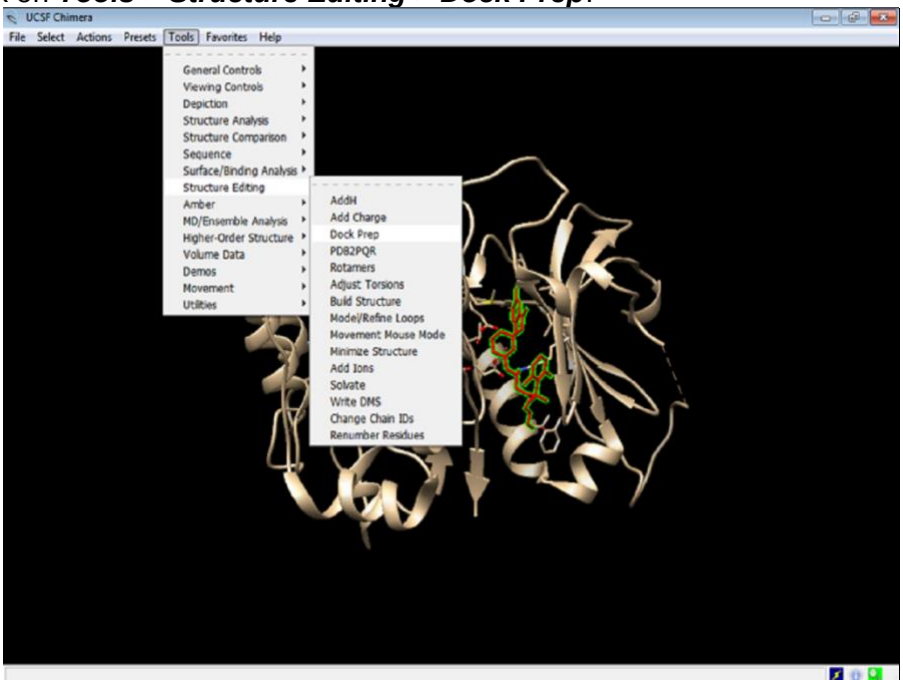


In this screenshot, protein bears an **HOH** group and SMH residues as nonstandard residues. Due to the selection, SMH appears to be highlighted in green.

2. After selecting the nonstandard (inhibitor) residues, the residues must be accorded a color. To distinguish the chosen residue from the rest of the protein. Change the color by clicking on ***Actions > Color > red*** (any color of your choice).
    - Visualizing the inhibitor separately ensures that it is correctly identified and can be distinguished from other parts of the protein. This step confirms the selection and helps in subsequent analysis and modifications.
    - Any residue of protein or nucleic acid that is not included in the list of Standard Residues is considered Non-Standard. The atomic coordinates for atoms in non-standard residues are given in records of type HETATM in the PDB file format.

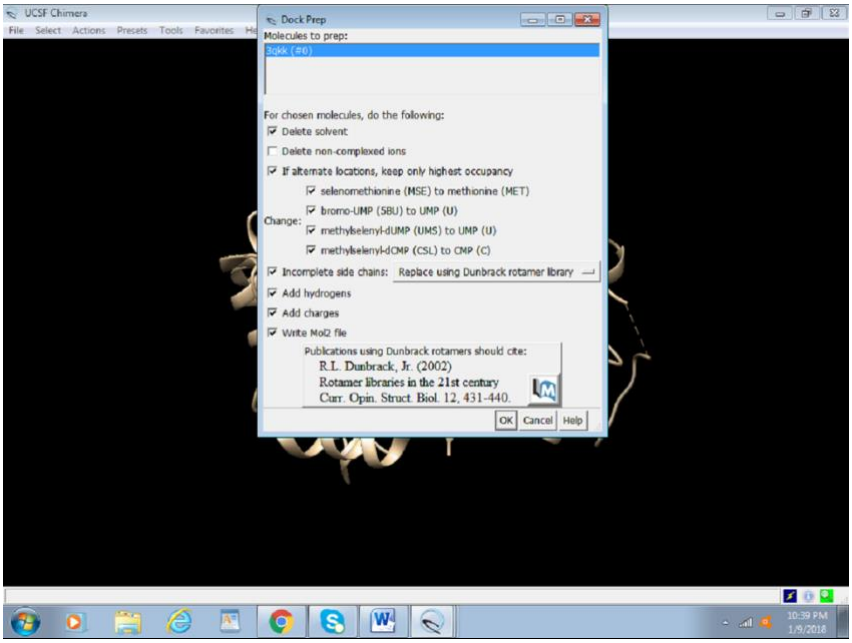


3. Once the coloring of the nonstandard residues is done, the protein needs to be optimized for docking. Click on ***Tools > Structure Editing > Dock Prep***.



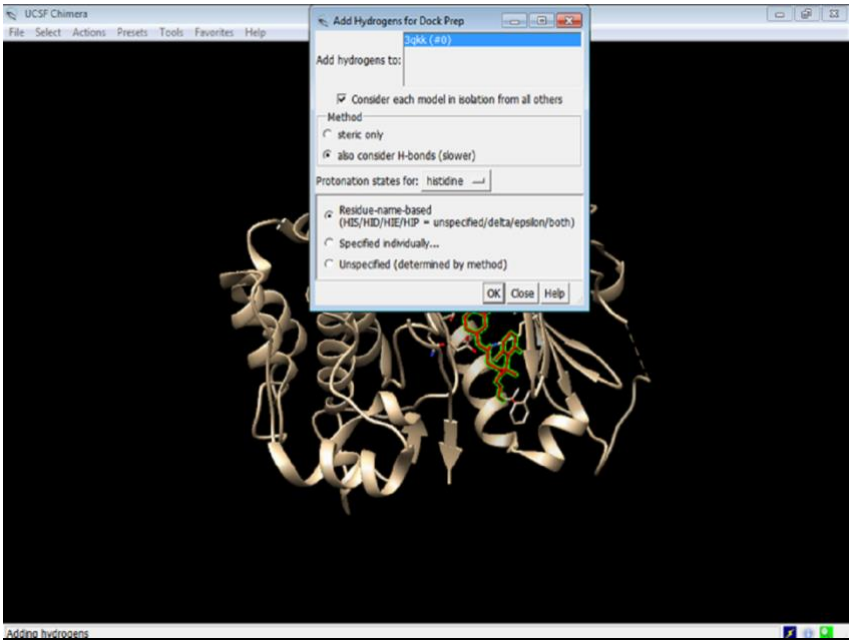
- The Dock Prep toolset in UCSF Chimera provides essential functionalities for preparing a protein structure for docking. This includes steps like adding hydrogens, assigning charges, and removing unnecessary elements that could interfere with docking simulations.
- Each option in Dock Prep ensures the protein structure is cleaned and optimized. For instance, retaining non-complexed ions might be necessary as they could play a role in the protein's stability or function, which could be crucial for accurate docking results.

4. In the dock prep box:
    - select all options except “Delete non-complexed ions” and click OK



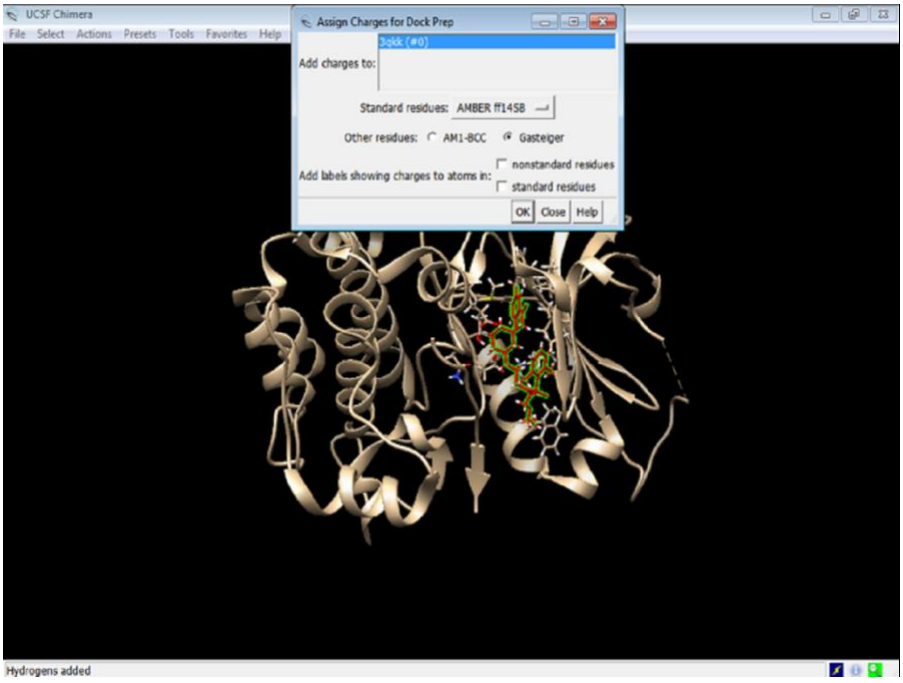
- **Add hydrogen** to the proteins by selecting the appropriate following options and click OK.
    - Adding hydrogen atoms is essential because X-ray crystallography data often lacks hydrogen positions. These atoms are critical for accurate modeling of hydrogen bonds and interactions during docking. Once the hydrogen bonds are added we must assign charges to the protein.

We allow the program to make the best choice according to the model by selecting the abovementioned options.

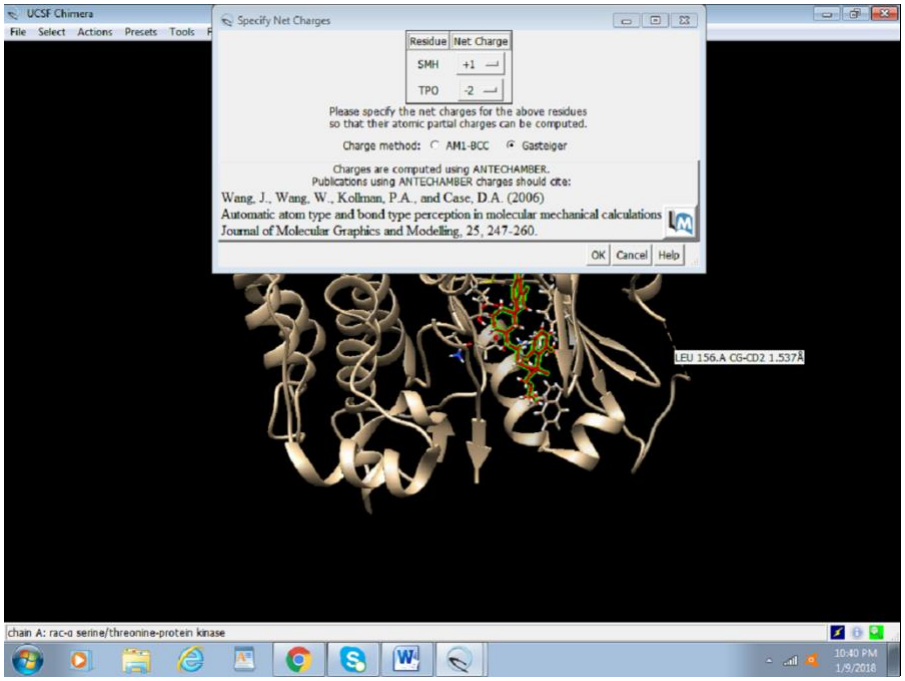


5. **Assign charges** to the protein by clicking on the Gasteiger charges and click OK.
    - Proper charge assignment is crucial for representing the electrostatic potential of the protein accurately. There are many charge models like Gasteiger, AM1-BCC or RESP charges that we can get from the Gaussian calculations for the Amber force fields, which eventually affect how the ligand interacts with the protein during docking. However, for the purpose of this tutorial we will assign Gasteiger charges, as they are an efficient method to approximate partial atomic charges.

    - Amber ff14SB is a widely used protein force field in the Amber suite of molecular dynamics (MD) simulations. It is designed to provide accurate modeling of protein structures, dynamics, and interactions. The ff14SB force field is an improvement over its predecessors, offering better accuracy in the representation of protein backbone and side-chain conformations.
    - ff14SB includes refined parameters for the protein backbone dihedral angles, which are critical for accurately modeling secondary structure elements like α-helices and β-sheets.
    - The ff14SB force field includes parameters that accurately describe hydrogen bonding interactions, which are crucial for maintaining protein structure and function.



6. ***Select the net charges*** and click OK.
    - Setting the net charge ensures the overall electrostatic environment of the protein is correctly represented. This is important as it affects the binding affinity and orientation of the ligand in the docking process.
    - The overall charge on the protein depends on its sequence composition.
    - Remember, typically R and L residues are positively charged (with +1 charge), and D and E residues are negatively charged (with -1 charge).



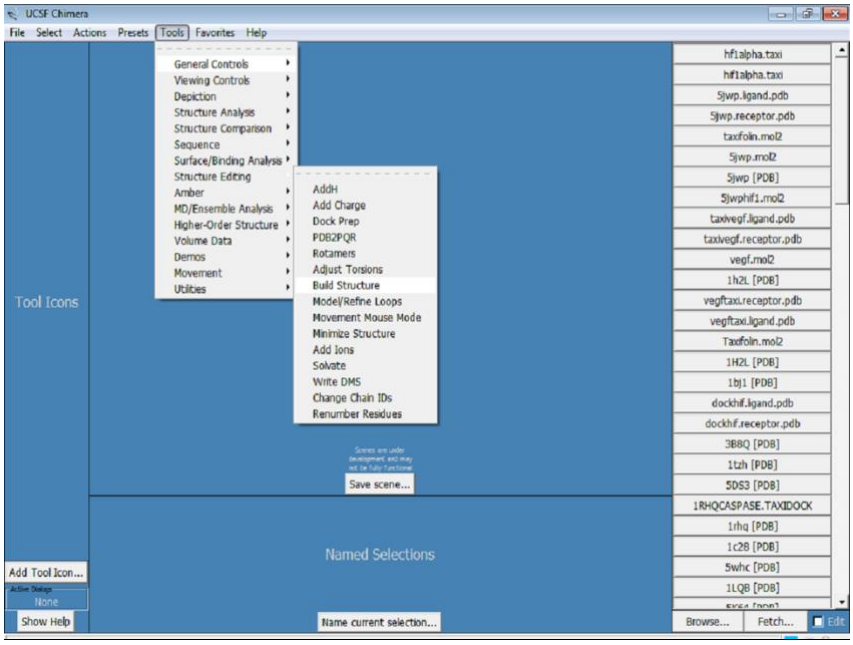
7. Save this file again as preped_Akt.PDB. or cleaned.PDB

6.2. Preparing the Ligand for Docking:

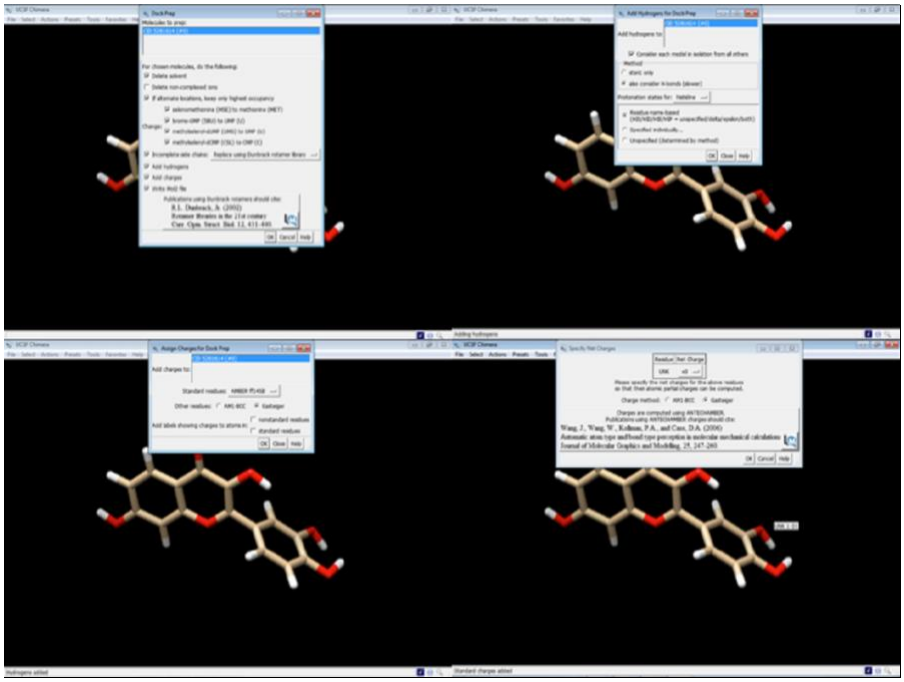
Similar to the process of obtaining the protein, drugs with Pub-chem compound ID (CID) can be fetched through the software with a working internet connection.

1. Click on ***Structure Editing > Build Structure > PubChem CID*** or you can even insert the simplified molecular-input line-entry system (SMILES) of the novel compound being used.

    - The image below shows how to fetch ligands from PubChem using its ID.
    - Enter the PubChem CID and click apply.

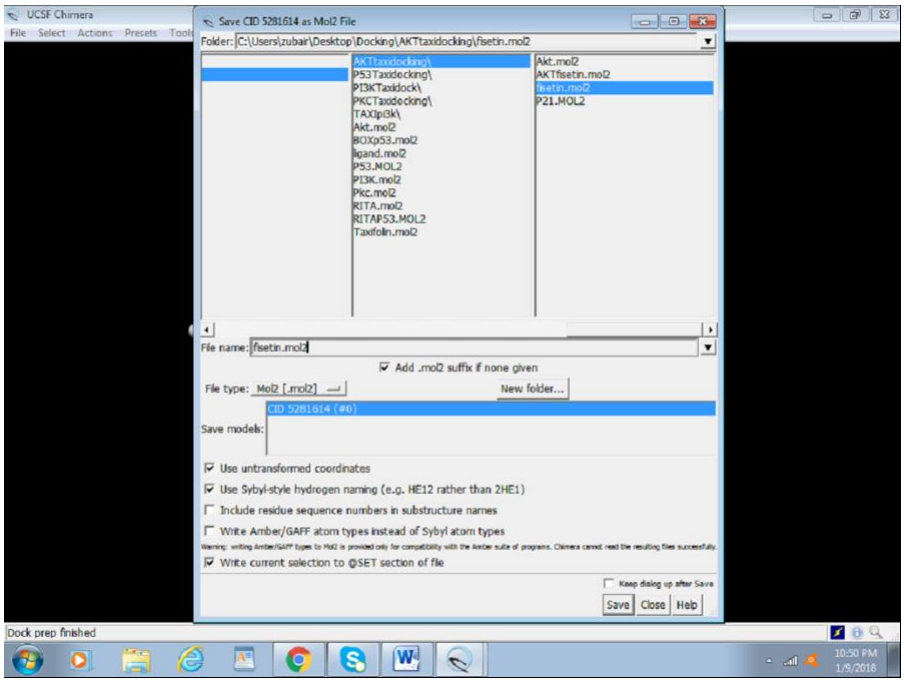


2. The ligand needs to be optimized as the protein was optimized. Click on ***Tools > Structure Editing > Dock Prep***, and repeat the same steps followed for preparing the protein. These steps include removing solvents, adding hydrogens, and determining the charge.



3. The ligand Fisetin is saved as prep_fisetin.mol2 file in the working directory earlier created.

    - Similar to PDB, mol2 is another format to save the coordinates.

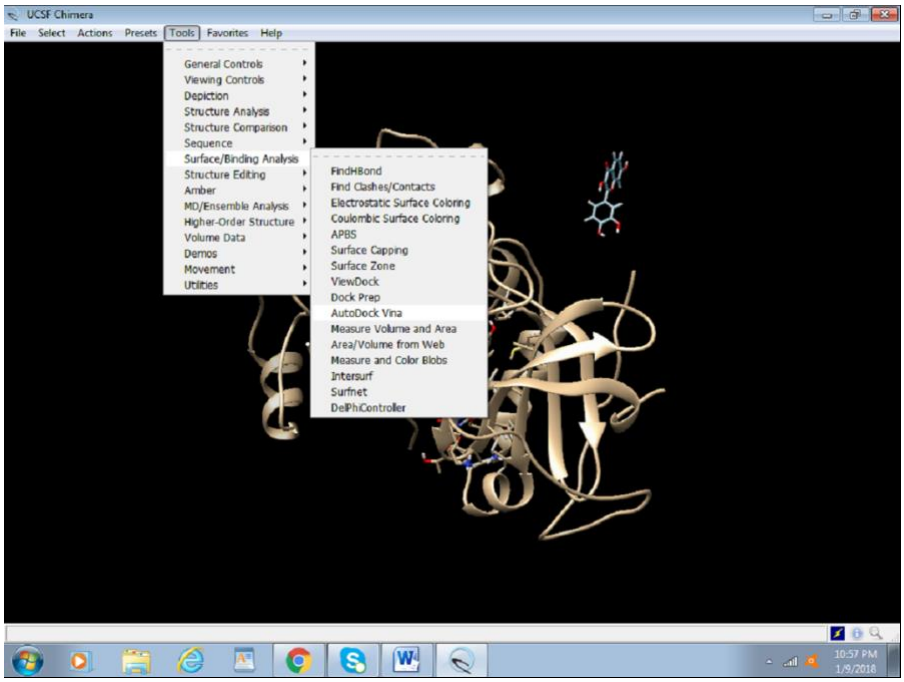


Step 7. Once the protein and ligand are ready the next step is we have to do the protein-ligand docking and calculate the binding affinity using Auto-Dock Vina.

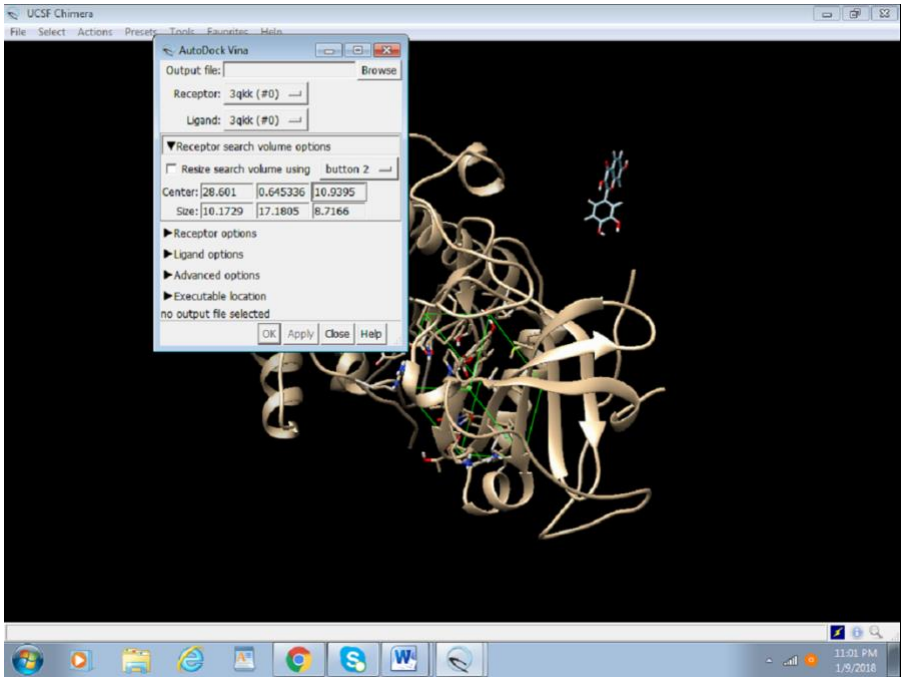
7.1. Docking

The following steps outline the process for docking:

1. Click on Tools > Surface or Binding Analysis > Auto dock Vina

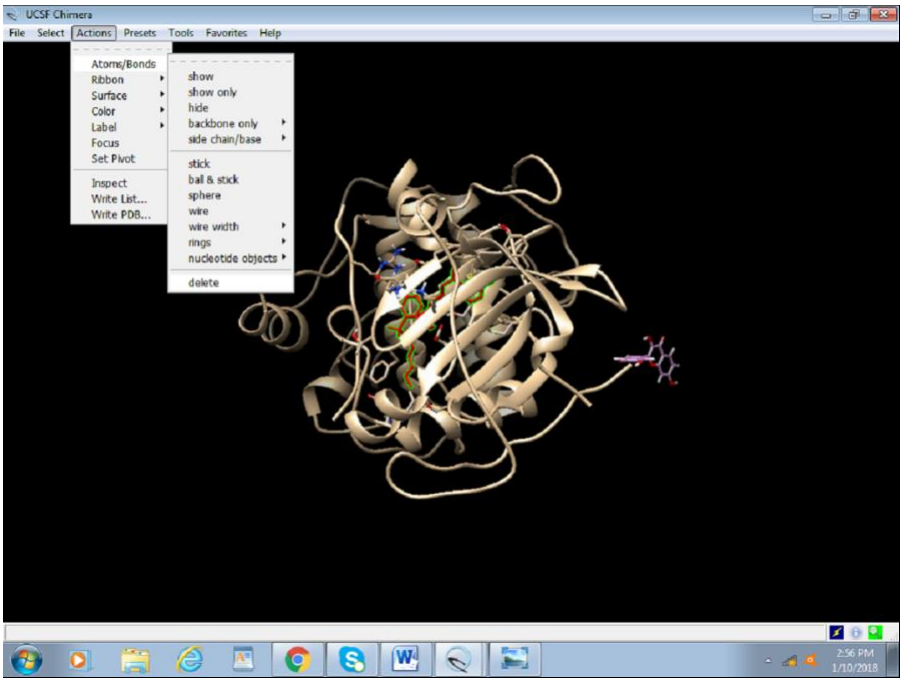


2. We will **set up the grid box values on the active site**; this is usually where the previous inhibitor was present. In case an inhibitor is absent, or the active site is relatively unknown, the size of the box and the location of the amino acids are determined by reading the literature.
    - The grid box defines a specific region of the protein where the docking simulation will take place. It restricts the search space to a smaller, more manageable area, usually around the active site or binding pocket where the ligand is expected to bind.
    - It ensures that the docking algorithm only considers the most relevant part of the protein, which increases the likelihood of finding meaningful binding interactions.

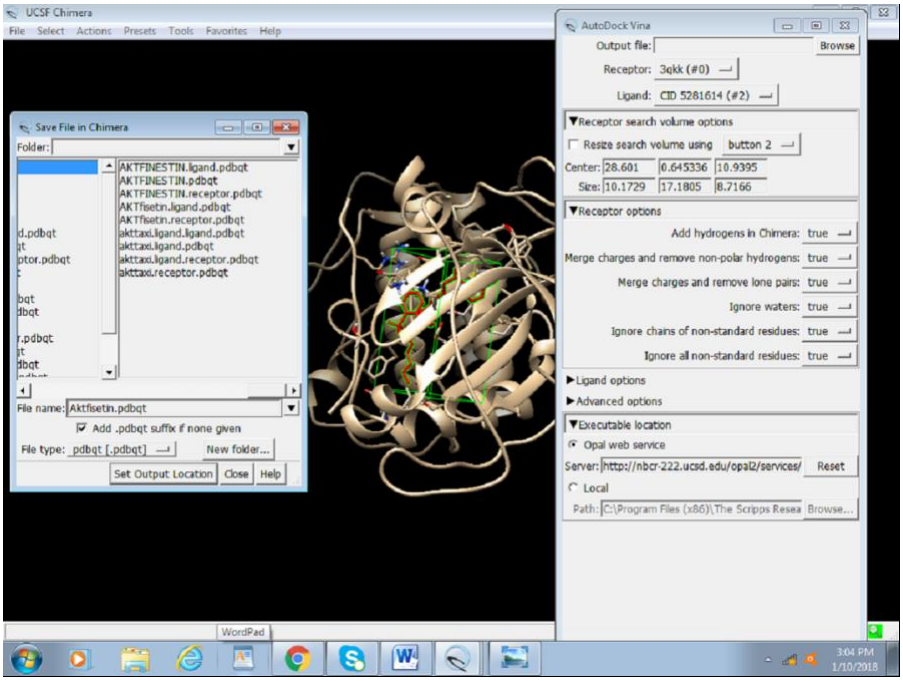


3. Browse the output file and save as Akt Fisetin.pdbqt in the same directory.
    - **Notice the extension of the file!**
    - A PDBQT file is a specialized format used by Auto Dock Vina for both protein and ligand structures. It extends the standard PDB (Protein Data Bank) format by including additional information necessary for docking simulations.
        - Components:
            - PDB: The basic structure of the molecule, including atomic coordinates.
            - Q: Partial charges on each atom.
            - T: Torsional degrees of freedom (rotatable bonds).
    - Partial Charges: The inclusion of partial charges (Q) in the PDBQT file is crucial for accurately modeling electrostatic interactions during the docking process. These charges influence how the ligand interacts with the protein, affecting binding affinity and orientation.
    - Torsional Information: The torsional degrees of freedom (T) are necessary to define rotatable bonds in the ligand. This information allows Auto Dock Vina to explore different conformations of the ligand during docking, which is essential for finding the optimal binding pose.

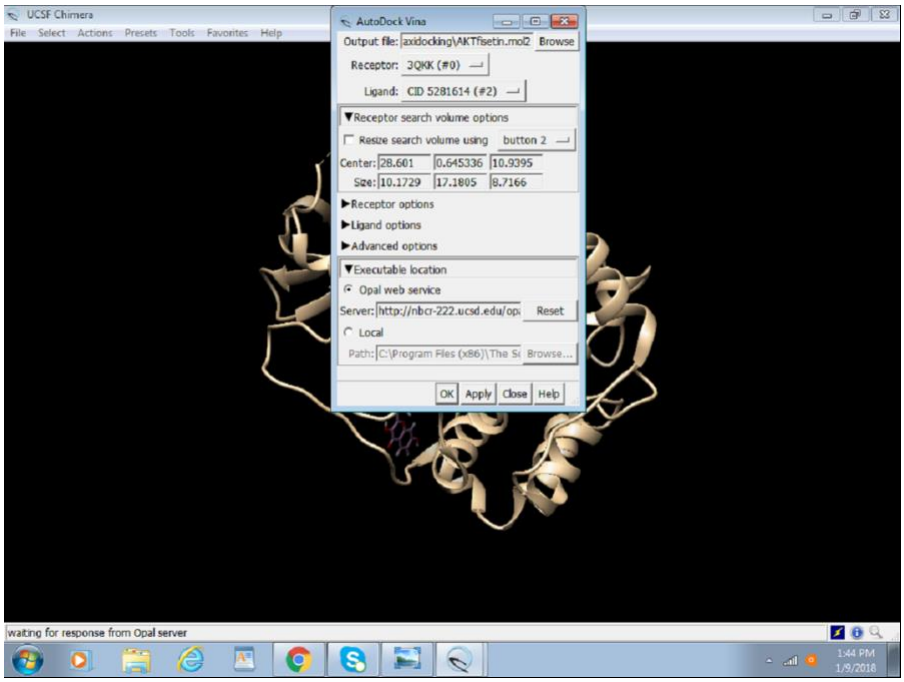
4. Delete the inhibitor molecule attached to the original 3D structure. Thereafter, ***select Actions > Atoms and Bonds > Delete***



5. Choose the receptor as the protein (cleaned_protein.pdb) from the drop-down menu and the ligand as prep_fisetin. mol2. It is important to set the right receptor and ligand. In the receptor and ligand options, change everything to TRUE.

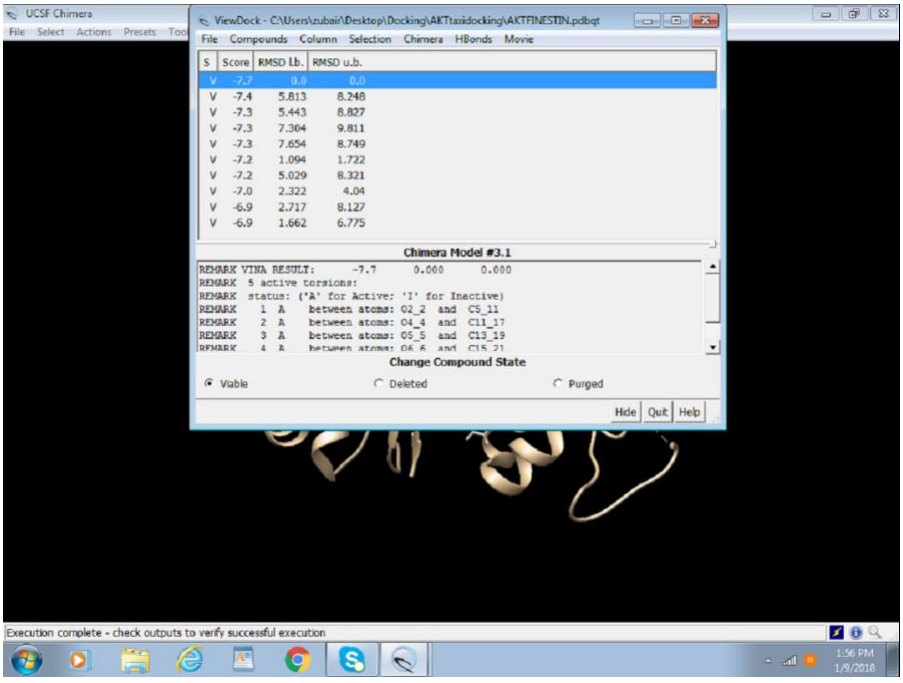


6. **Enter the local path where the installed version of Auto dock Vina is placed** and click on OK and Apply to run the docking calculations.



Step 8. Analyzing results/outcomes of docking:

After the successful run of Auto dock Vina, the following dialogue box will appear with the solution.

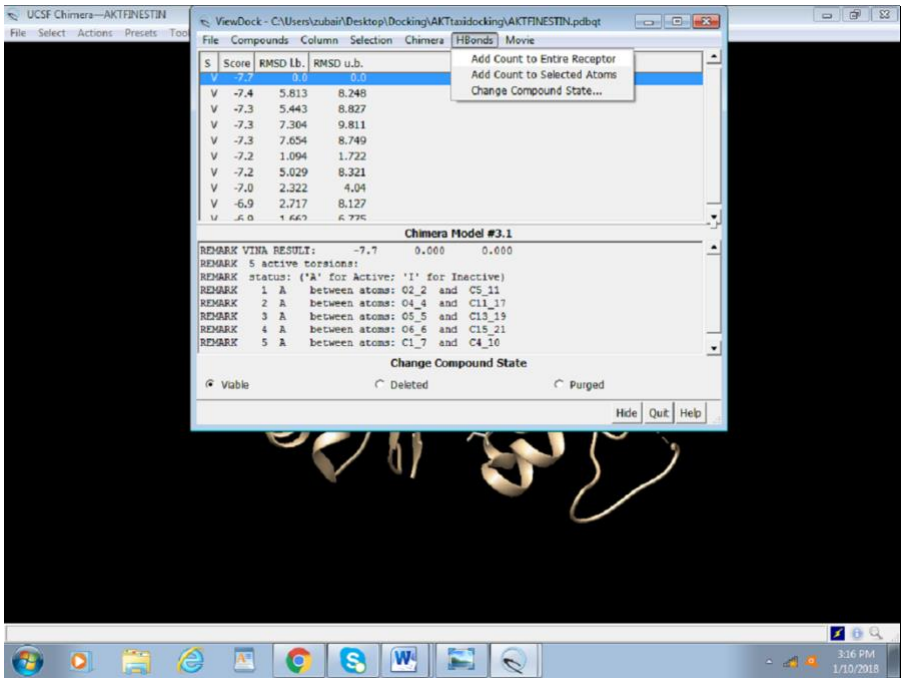


- The above figure portrays the final step of Docking, that is, outcome/results of docking, which are score, root-mean-square deviation (RMSD) lower bound, and RMSD upper bound.
    - Score: Indicates the predicted binding affinity between the ligand and the protein. Lower scores suggest stronger binding.
    - RMSD Lower Bound: Measures the minimum deviation of the predicted pose from a reference, providing insight into the best alignment and similarity.
    - RMSD Upper Bound: Measures the maximum deviation, indicating the variability and potential flexibility of the ligand's binding pose.

Visualization of docking can be done as follows:

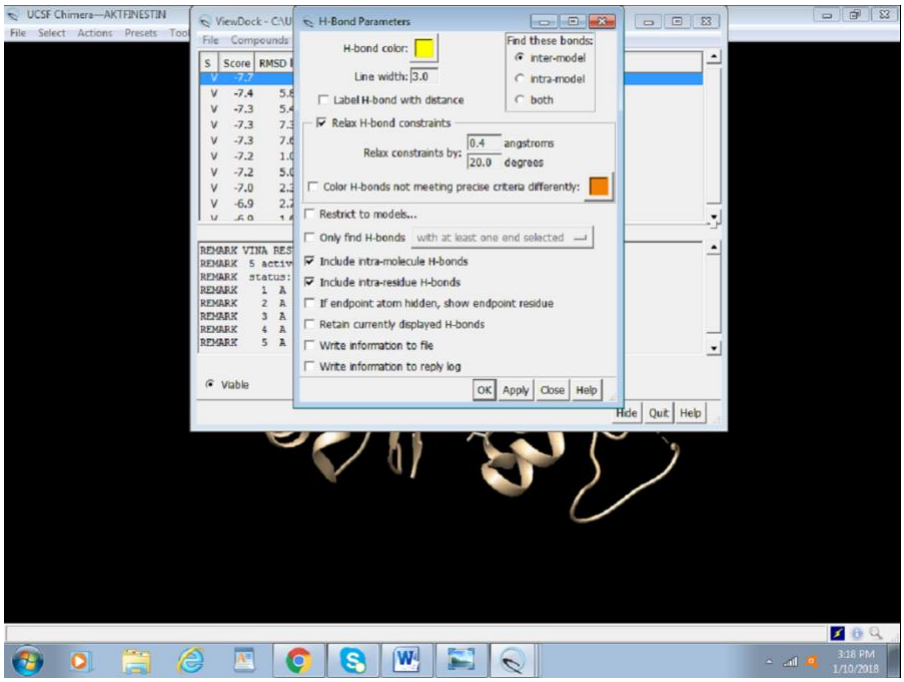
1. **Hydrogen bonding :** It plays a crucial role in the binding interactions between a ligand and its receptor. Here's why it is important to visualize these interactions using the result dialogue box in molecular docking studies. Hydrogen bonds contribute significantly to the binding affinity and specificity of the ligand-receptor interaction. By visualizing these bonds, researchers can understand how well a ligand fits into the receptor's binding site and which specific interactions stabilize this binding.
    - Strength of Interaction: Hydrogen bonds are relatively strong non-covalent interactions that can greatly influence the binding strength. Understanding these interactions helps in predicting the binding affinity of the ligand.
    - Selective Binding: Specific hydrogen bonds can explain why a ligand selectively binds to one receptor over another, providing insights into specificity.

To see the hydrogen bonding between the receptor and the ligands using the result dialogue box:

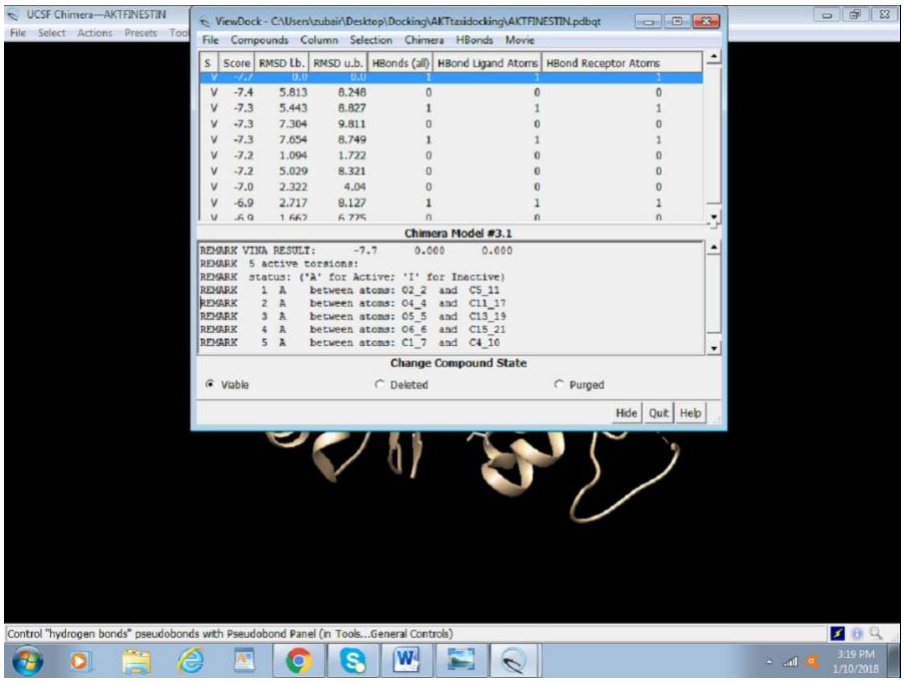


- ***Select H Bonds > Add Count*** to the Entire Receptor. This opens an H-bond parameter dialogue box.

- ***Select Intermodel*** to visualize bonding between receptor and ligand. Different parameters can be adjusted to better picture the bonding. The table showing all the information on hydrogen bonds and RMSD is presented at the end of the docking session.



Lastly, to be able to retrieve the docking session later at any stage, it can be saved by selecting File > Save Session as > An Akt Fisetin Docking (name the session).



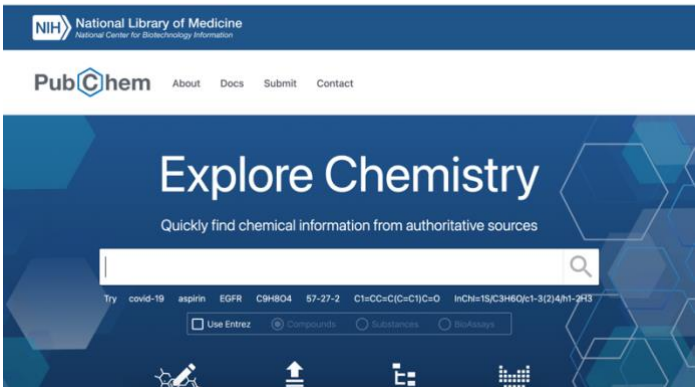
Multiple Ligand Binding using Auto Dock Vina

- module load miniconda3/23.11.0 – creation of the conda environment
- conda create -n openbabel-env -c conda-forge openbabel -y (This command downloads the openbabel which is useful for converting the ligands .sdf file to pdbqt conversion for the docking process)
- conda activate openbabel-env (Use this command to activate the conda environment)

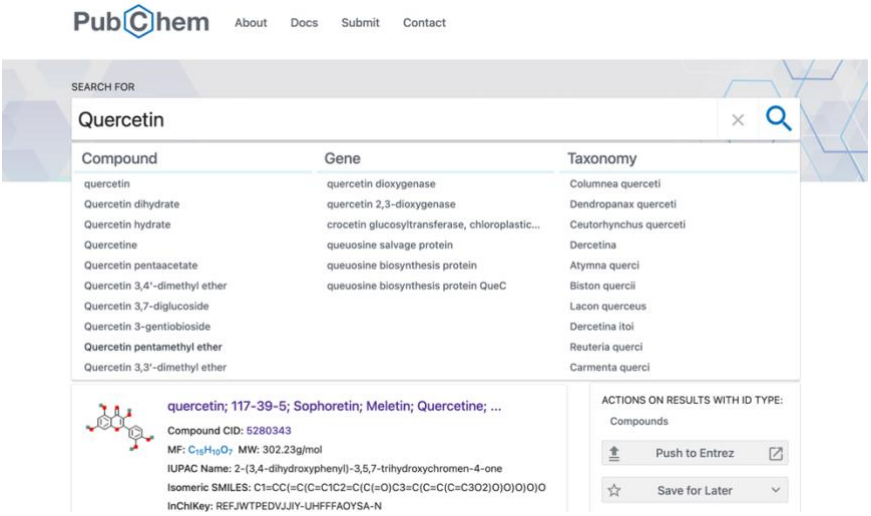
- Once it is done take the docking_receptor. pdbqt file from the single protein-ligand docking using chimera and auto dock vina and create a receptor. pdbqt file in your terminal.
- Similarly, create the config.txt file in the terminal. Just copy the lines from configuration file from the single protein – ligand docking.
- Then it's time to fetch the ligands you want to do docking from PubChem.
- Firstly, create a fetch_ligands.sh file using vim fetch_ligands.sh.
- Write the script mentioned below:

```
# Create a file named fetch_ligands.sh

#!/bin/bash

# List of PubChem CIDs
CIDS= (439533 5281607 221336) # Example CIDs, replace with your own

# Fetch SDF files for each CID
for CID in "${CIDS[@]}"; do
    wget "https://pubchem.ncbi.nlm.nih.gov/rest/pug/compound/CID/$CID/SDF" -O "ligand_$CID.sdf"
done
```

- In this script mention the PubChem ID's you want to fetch. In this file I took the PubChem ID's 439533 5281607 221336
- To get the PubChem ID - Follow these steps mentioned below
- Open the link: https://pubchem.ncbi.nlm.nih.gov/. You will open this kind of page.



- Type the ligand you want to dock. For example: If you want to do Quercetin just type that in the PubChem.



In the above picture see the Compound CID: 16224. This is the CID ID. Mention this ID in the fetch_ligands.sh file.

- Once all the CIDs are written in the fetch_ligands.sh file type these commands:
- chmod +x fetch_ligands.sh
- ./fetch_ligands.sh – (Once you run this command in your terminal it will download your files from PubChem)
- Now prepare_ligands.py file to do the energy minimization and convert those .sdf files to ligand. pdbqt files.

- In the prepare_ligands.py write this python script and then run then command **python prepare_ligands.py**

```python
from openbabel import pybel
import glob

# List of ligand files
ligand_files = glob.glob("ligand_*.sdf")

for ligand_file in ligand_files:
    # Load the molecule
    mol = next (pybel.readfile("sdf", ligand_file))

    # Add hydrogens
    mol.addh()

    # Energy minimizations
    mol.localopt()

    # Convert to PDBQT
    pdbqt_file = ligand_file.replace(".sdf", ".pdbqt")
    mol.write("pdbqt", pdbqt_file)
    print(f"Converted {ligand_file} to {pdbqt_file}")
```

- By running the above command python prepare_ligands.py it will add hydrogens, do the energy minimization and then convert those files to. pdbqt.
- After this we have the receptor.pdbqt and multiple ligand .pdbqt files. Now load the Auto dock vina.
- module use /project/biocompworkshop/ukapoor/codes/packages
- module load autodock_vina/1.1.2
- Now create the run_docking.sh file to do the Multiple ligands docking using auto dock vina.

```
 1.  # Create a file named run_docking.sh
 2.
 3. #! /bin/bash
 4.
 5. # List of ligand PDBQT files
 6. LIGANDS= ($(ls ligand_*.pdbqt))  # Automatically gather all ligand PDBQT files
 7.
 8. # Loop over each ligand and run AutoDock Vina
 9. for LIGAND in "${LIGANDS[@]}"; do
10.     vina --config config.txt --ligand $LIGAND --out "${LIGAND%.pdbqt}_out.pdbqt" --log
"${LIGAND%.pdbqt}_out.log"
11. done
```

- chmod +x run_docking.sh
- ./run_docking.sh
- Once you run this command Auto Dock Vina will perform the protein-ligand docking calculations and calculate the binding affinity.
- The below image is the snapshot where Auto dock vina performs the calculations and stores the output in the log file.

```
(openbabel-env) [hbalantr@blog2 protein_ligand_docking]$ chmod +x run_docking.sh
(openbabel-env) [hbalantr@blog2 protein_ligand_docking]$ ./run_docking.sh
###############################################################
# If you used AutoDock Vina in your work, please cite:          #
#                                                               #
# O. Trott, A. J. Olson,                                        #
# AutoDock Vina: improving the speed and accuracy of docking    #
# with a new scoring function, efficient optimization and       #
# multithreading, Journal of Computational Chemistry 31 (2010)  #
# 455-461                                                       #
#                                                               #
# DOI 10.1002/jcc.21334                                         #
#                                                               #
# Please see http://vina.scripps.edu for more information.      #
###############################################################

Detected 96 CPUs
WARNING: at low exhaustiveness, it may be impossible to utilize all CPUs
Reading input ... done.
Setting up the scoring function ... done.
Analyzing the binding site ... done.
Using random seed: 988421136
Performing search ...
0%   10   20   30   40   50   60   70   80   90   100%
|----|----|----|----|----|----|----|----|----|----|
***************************************************
done.
Refining results ... done.

mode |   affinity | dist from best mode
     | (kcal/mol) | rmsd l.b.| rmsd u.b.
-----+------------+----------+----------
   1        -7.7      0.000      0.000
   2        -7.6      2.099      2.960
   3        -7.2      1.780      7.426
   4        -6.9      1.907      3.752
   5        -6.7      1.748      3.400
```

- tail -n11 *.log>results.txt (To prepare the result table)

- Prepare protein ligand complex in chimera, for this open naked protein in Chimera, then open the output pdbqt file of ligand of your wish/need from Tools>Structure/Binding Analysis > ViewDock

- Save the appropriate pose using File > Save PDB > input file name, choose protein and pose of choice by pressing ctrl select> choose "a single file" under “Save multiple models in" and lastly press Save.

You can visualize protein-multiple ligands docking with the software of your choice. Below is the docked structure of the three ligands with the protein which is a complex PDB having both protein and ligands where the blue color, pink, and the dark blue are the three ligands we performed for the protein-ligand docking.

This will complete the Multiple Ligand docking with Protein calculations using the Auto Dock Vina.