Molecular Docking Flowchart

Molecular docking is a computational technique to predict the preferred orientation of one molecule (the ligand) when it binds to another molecule (the receptor), typically a protein, to form a stable complex. This method helps understand the molecules' interaction and is widely used in drug discovery and design.

Tools and software required for docking

1. **PyRx –** It is a popular open-source software tool for virtual screening and molecular docking

   Role: Ligand and protein preprocessing (Converting PDB and CIF into PDBQT format), Docking, Virtual screening.

   Download: PyRx

2. **PyMol –** It is a widely used molecular visualisation system that enables users to create high-quality 3D images of small molecules, proteins, nucleic acids, and other biological macromolecules.

   Role: Molecular visualisation, Molecular Interaction, Protein cleaning, Ligand processing, Saving molecules into PDB and CIF format.

   Download: PyMol 3.0

3. **Discovery Studio Client –** It is a comprehensive suite of software tools for computational chemistry, molecular modelling, and simulation. It is widely used in drug discovery, material science, and other fields that require molecular modelling and analysis.

   Role: Protein cleaning, Molecular visualization, Receptor-ligand Interaction, 2D plot

   Download: Discovery Studio Client (Windows 64-bit)

Steps in Molecular Docking

1) Data acquisition – Downloading Receptor and Ligand 3D structures
2) Data Preprocessing – Cleaning Receptor and Ligand molecules and saving them in PDB and CIF format respectively
3) Importing molecules into PyRx

Contact: Mail | LinkedIn | GitHub

4) Energy minimisation and converting molecules into autodock format (PDBQT)
5) Initiating the docking process
6) Save the results

A) Data Acquisition

Data acquisition in molecular docking involves gathering and preparing the necessary data for the docking simulation. This includes obtaining the molecular structures of the target (receptor) and the compounds (ligands) to be docked, as well as relevant biological and chemical information.

Receptor (Target Protein):

- **Protein Data Bank (PDB)**: The primary source for obtaining 3D structures of proteins and nucleic acids. Structures are typically obtained in PDB format.

  Website: Protein Data Bank

- **Alternative Databases**: Other databases such as UniProt, AlphaFold, and SwissModel provide protein structures and related data.

Ligands (Small Molecules):

- **Chemical Databases**: Databases like PubChem, ZINC, ChEMBL, and DrugBank provide 3D structures of small molecules in various formats (SDF, MOL2, etc.).
  - **PubChem**: PubChem
  - **ZINC**: ZINC Database
  - **ChEMBL**: ChEMBL
  - **DrugBank**: DrugBank

B) Data Preprocessing

Data preprocessing in molecular docking is a crucial step that ensures the quality and accuracy of the input data before performing docking simulations. This involves preparing both the receptor (target protein) and the ligands (small molecules) to be in the optimal state for docking.

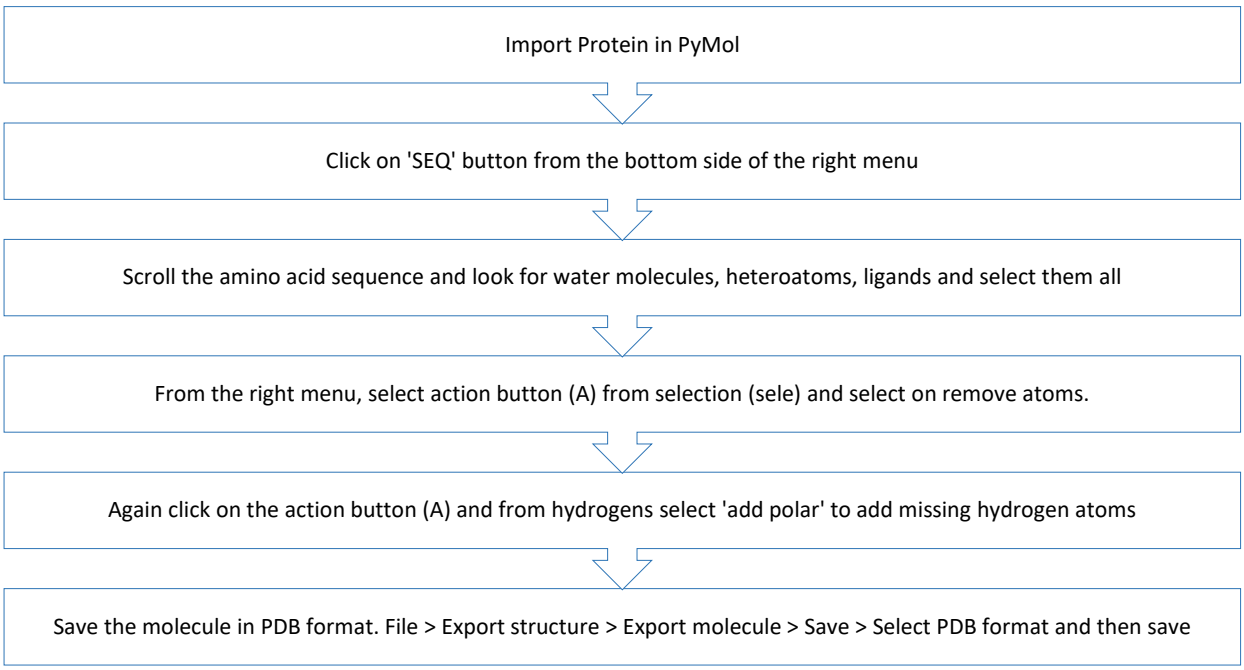
1. Receptor Preparation

a. Cleaning the Structure:

- **Remove Water Molecules**: Water molecules can interfere with docking simulations, so they are usually removed unless they are known to be critical for binding.
- **Remove Ligands and Cofactors**: Any bound ligands, cofactors, or ions that are not part of the binding process should be removed.

b. Adding Hydrogen Atoms:

- **Protonation States**: Add hydrogen atoms to the protein structure to ensure correct protonation states of ionizable groups. Tools like PROPKA can help predict the protonation states at physiological pH.



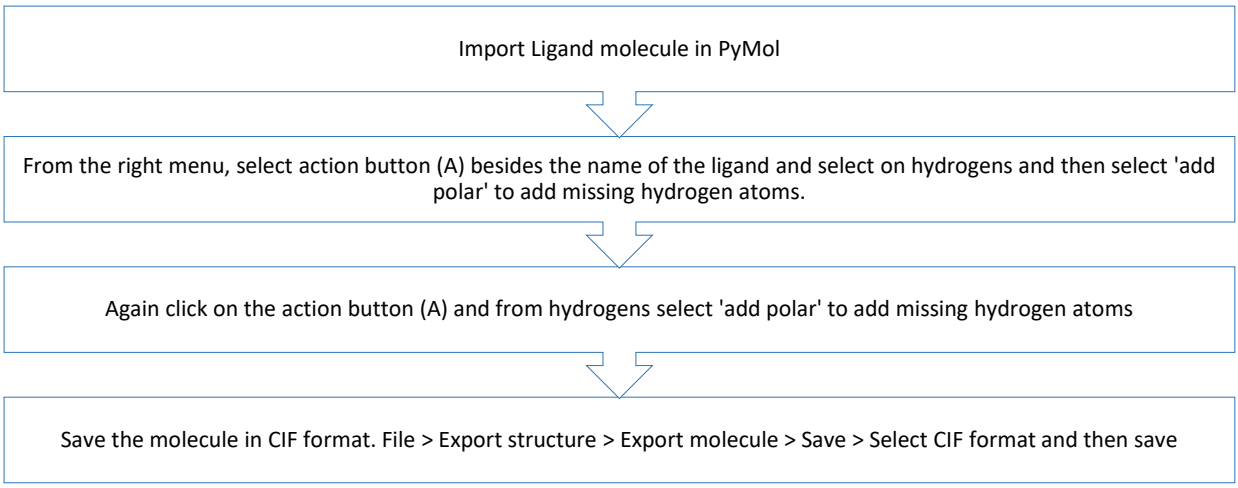
2. Ligand Preparation

a. Structure Acquisition:

- **2D to 3D Conversion**: Convert 2D structures of ligands into 3D. Tools like Open Babel or RDKit can be used for this purpose.
- **Database Retrieval**: Retrieve ligand structures from databases like PubChem, ZINC, or ChEMBL.

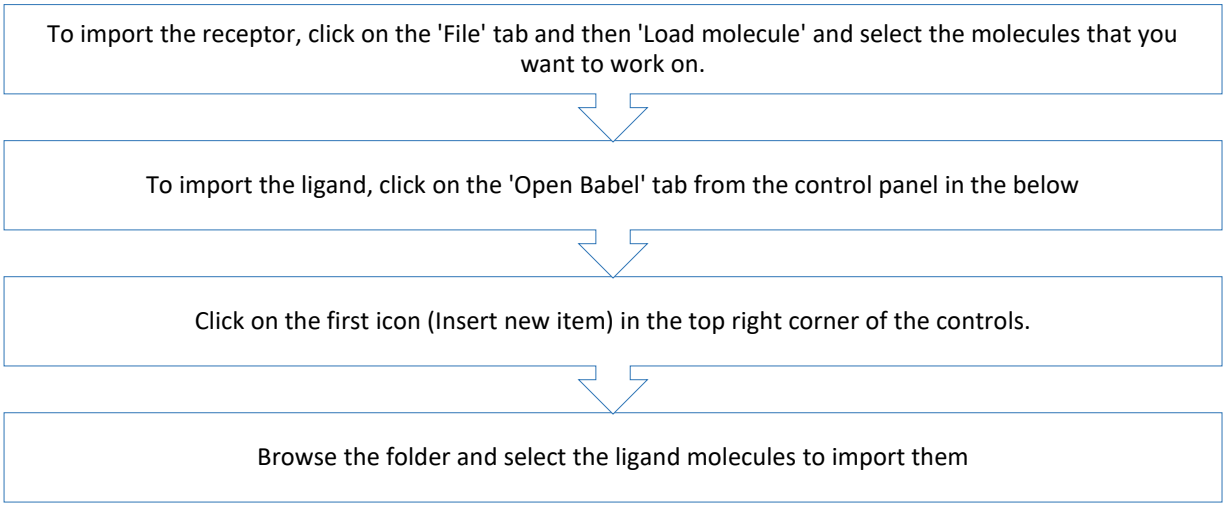
b. Adding Hydrogen Atoms:

- **Protonation States**: Add hydrogen atoms considering the correct protonation states at physiological pH.



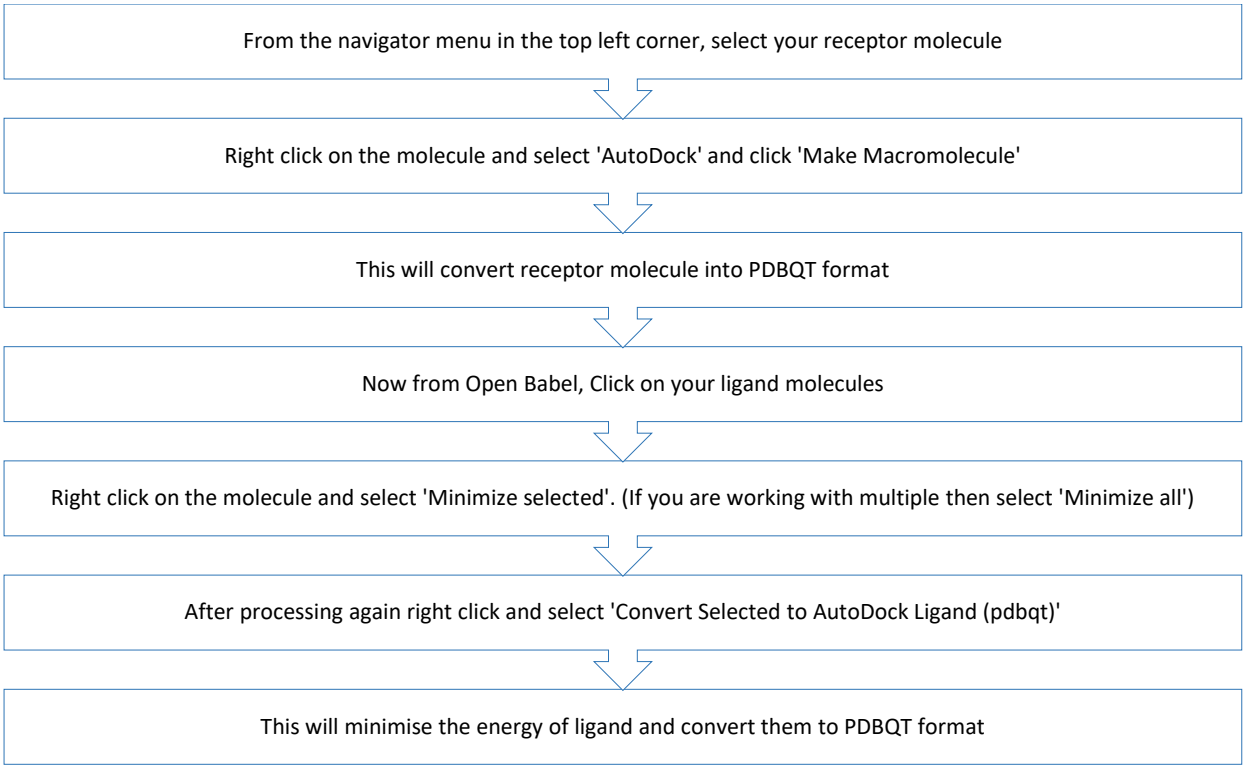
C) Importing molecules into PyRx

Before importing the molecule, we've to create a workspace where all the processes will be done and the molecules will be saved. For that click on the 'Edit' tab then select 'preferences' and browse the folder from the Workspace option where you want to save your files. After that close your PyRx and reopen it to apply the changes.



D) Energy minimisation and converting molecules into autodock format (PDBQT)

Energy minimisation and converting molecules into AutoDock format (PDBQT) are essential steps in preparing molecular structures for docking simulations using AutoDock or AutoDock Vina. Energy minimisation helps relieve any steric clashes or strains in the molecular structures and ensures they are in a low-energy conformation.



E) Initiating the docking process

To initiate the docking process, we have to specify the grid (space) that defines the region of the receptor where the docking simulation will take place. Usually, we specifically select the binding site where the reaction takes place. To know the binding site of the protein there are some webservers where we import the molecule and that program shows the region.

Tools for Binding Site Prediction:

- **CASTp**: Identifies surface pockets and internal cavities.

Website: CASTp

- **SiteMap (Schrödinger)**: A tool for predicting binding sites.

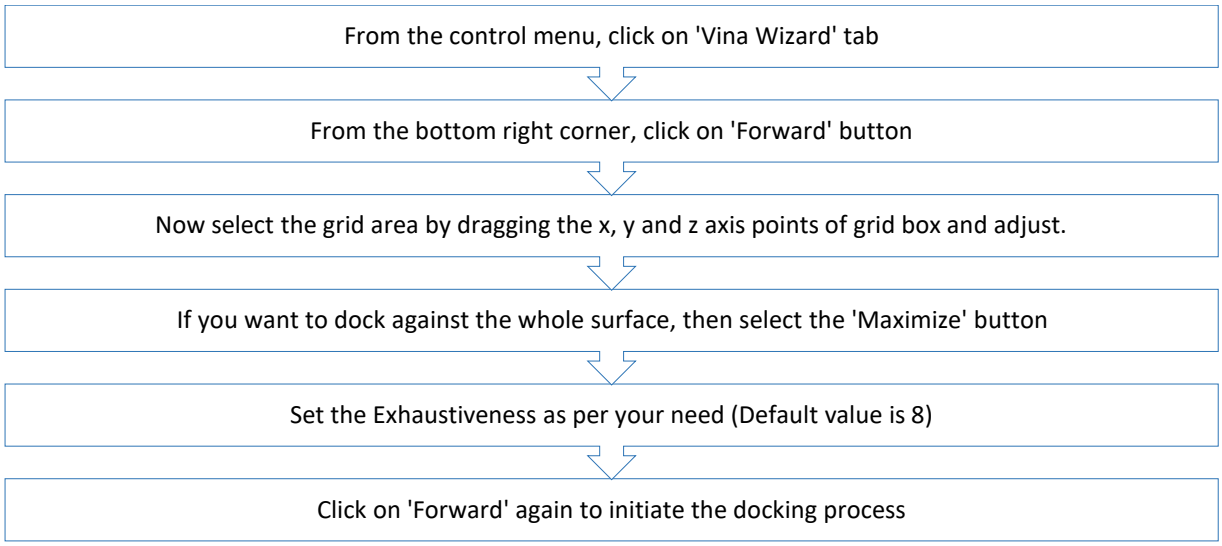
  Website: Schrödinger SiteMap

While initiating the docking process, you have set the exhaustiveness value. But what is Exhaustiveness?

Exhaustiveness in molecular docking refers to the thoroughness or extent of the search for optimal binding poses of the ligand within the receptor's binding site. It determines how extensively the conformational and positional space is sampled during the docking process.

So if you want to understand in simpler terms then imagine you're searching for a hidden treasure in a large field. The treasure represents the best way a small molecule (ligand) can fit into a larger molecule (receptor) in drug discovery.

- **Low Exhaustiveness**: This is like doing a quick scan of the field. You might find some interesting spots, but you could miss the best treasure. It's faster but less thorough.
- **High Exhaustiveness**: This is like doing a detailed search, checking every corner of the field. It takes more time and effort, but you have a better chance of finding the best treasure.



F) Saving the result

To save the results, click on the 4th icon (Save as CSV) located at the top right corner of the control panel. Select the folder where you want to save and click 'Ok' to save the file.

Understanding the Results

Understanding the docking results from a CSV file involves interpreting the binding affinities, identifying the best binding poses, and analyzing the molecular interactions. Before proceeding further, let's understand the basic terms and concepts.

a) **Binding Affinity (ΔG)**:

- Indicates the strength of binding between the ligand and the receptor.
- More negative values suggest stronger binding interactions.
- Units are typically in kcal/mol.

b) **RMSD (Root-Mean-Square Deviation)**:

- Measures the average distance between atoms of superimposed proteins.
- Used to compare different poses to a reference pose or the best pose.
- Lower RMSD values indicate more similar poses, suggesting less variability.

c) **Interaction Types**:

- Indicates specific interactions between the ligand and receptor, such as hydrogen bonds, hydrophobic interactions, electrostatic interactions, etc.
- Useful for understanding the nature of binding and identifying key residues involved.

A typical docking output CSV file may include the following columns:

- **Ligand Name**: The name or ID of the ligand.
- **Binding Affinity (ΔG)**: The binding free energy in kcal/mol.

- **RMSD/ub**: RMSD/ub represents the upper bound RMSD value. It measures the maximum deviation of atomic positions between the docked pose and a reference pose, considering all possible alignments.

  Interpretation: A higher RMSD/ub value indicates greater variability or deviation from the reference pose. It gives a sense of the worst-case deviation between the docked pose and the reference.

- **RMSD/lb:** RMSD/lb represents the lower bound RMSD value. It measures the minimum deviation of atomic positions between the docked pose and a reference pose, considering all possible alignments.

  Interpretation: A lower RMSD/lb value indicates that there are regions where the docked pose closely aligns with the reference pose. It provides a sense of the best-case scenario for how closely the docked pose matches the reference.

Analyzing the Data

1. Sorting and Filtering Data

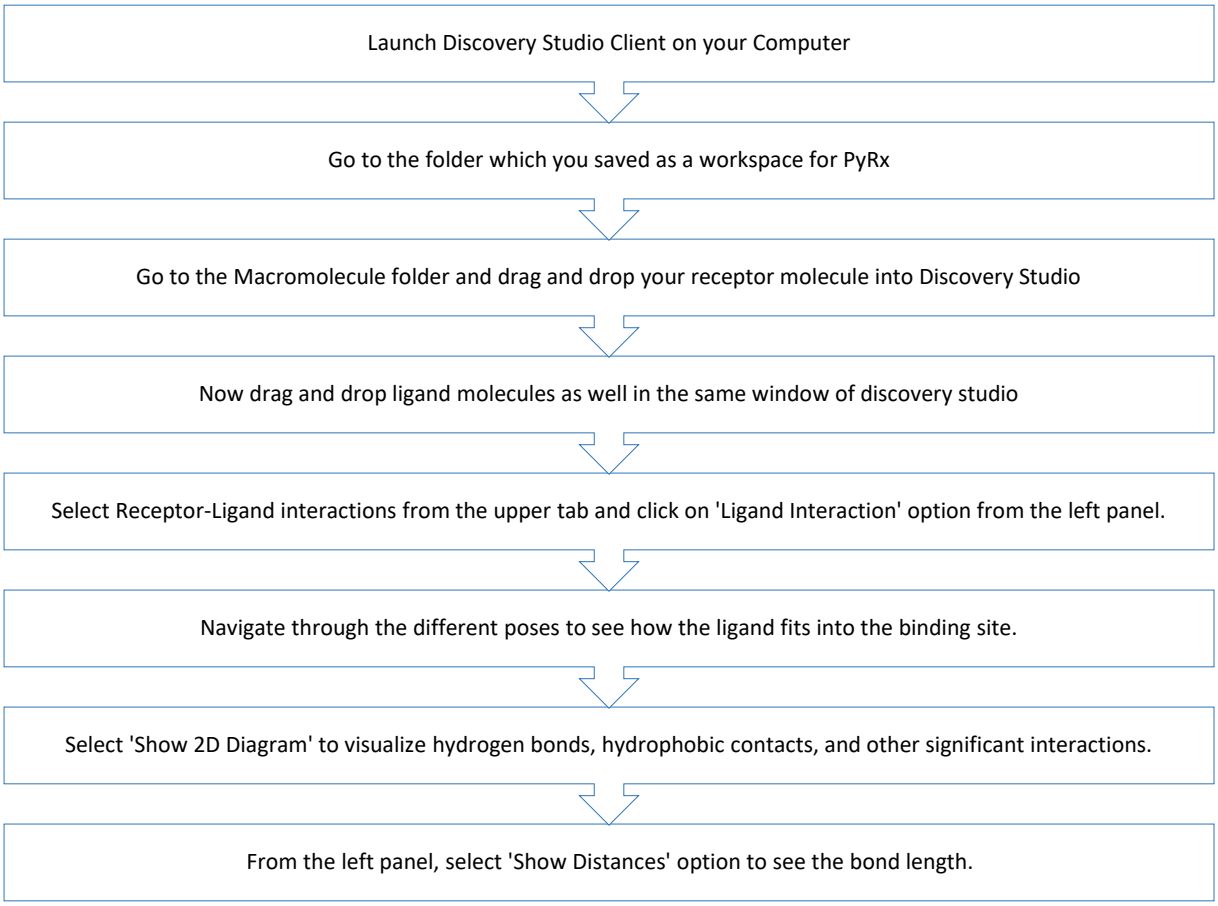
- **Sort by Binding Affinity:**
  - Sort the poses by the binding affinity column to identify the poses with the lowest (most negative) ΔG values.
  - These poses are considered the most favourable binding conformations.

2. Compare RMSD Values

- **Identify Consistent Poses**:
  - Look for poses with low RMSD/ub and RMSD/lb values. These poses are likely to be more consistent and reliable.

Visualizing the result in Discovery Studio Client

Analyzing docking results in Discovery Studio Client involves a series of steps to visualize and interpret the docking poses and interactions between the ligand and the receptor. Discovery Studio offers a comprehensive set of tools for this purpose.



Contact: Mail | LinkedIn | GitHub