# Documentation for Autodock Vina calculations

\Analysis software and scripts\Vina scripts this folder contains all the scripts needed to run a vina job on cluster.

## \ vina_PdbCode_SubstrateCode_minimised.sh

Change the file name accordingly by replacing ProteinDockfile to the name of protein. I used pdb code and substrate code as shown in file Name but any name could be used. Although if you want to use my matlab scripts use the file name as I used. Use notepad++ for all the editing,and change EOL in edit to unix if needed.

*#BSUB -q hp12* (assign queue on cluster)

*#BSUB -n 2* (number of processors for each process don't use more then 2)

*#BSUB -o /home/nshukla/Outputlogs/log_4wor.txt* (output log file path)

*#BSUB -R "span[hosts=1]" (only fire one node in hp12 queue)*

*#BSUB -J Nimesh_Dockings* (optional job name)

In order to run a vina job this file needs to be edited for ligand names in ligset_list, protein and ligand pdbqt file paths,grid box size and coordinate in vina command. Make 1 such file for each protein type and put it in the home folder on cluster. Exhaustiveness can be change for more through search,vina recommend 8 I used 100 by default in this script.

While changing the output pdbqt file name keep the string "Resultss" intact,as it is used to run.sh to modify by the number of docking run.

## \run.sh

this script runs the docking job on cluster.This scrpts need to be modified for number of runs and docking file name as explains in example below. Use notepad++ for all the editing,and change EOL in edit to unix in needed.

*for x in {1..20}; do sed "s/Resultss/Results_$x/"  vina_THP_minimised.sh > vina_THP_minimised_$x.sh ; done*

change x to number of docking run needed, default is 20 and in each run vina give 20 confirmations so total 400 conf. can achieved using the default runs.

Original dock file name *= vina_THP_minimised.sh*

New file name for each run*= vina_THP_minimised_$x.sh*

*for x in {1..20}; do bsub < vina_THP_minimised_$x.sh ;* (Change new dock file name accordingly)

## Output dock files

Running run.sh on cluster should output 20 confirmation for each run with their binding energies in .pdbqt format. Collect all output file in one folder on any windows computer for analysis.

Output file should be of the format : ***ProteinPdbCode_ Results_VinaRunNumber_LigandName.pdbqt***

For example:- ***1HEW_Results_1_sucralose.pdbqt***

\Analysis software and scripts contains all the scripts needed to analyze this data. Copy all these scripts to the folder containing vina result pdbqt files.

### \copy_original_files_to_folders.bat

This script copy all ligand files to their respective protein folders. Edit it accordingly for protein name and ligand name then double click on this batch file. Cut and paste all the newly created folder to a different location.

### \split.bat

Its not very important to split vina results but just in case you want to, this batch script split vina result pdbqt files into their individual confirmations. It will also create individual folders for each protein type and its ligand type. Edit the script accordingly for protein and ligand name. For running this batch file just double click on it.\

Vina_split.exe is required for this batch file to run.

# Data Analysis

### read_alldata.m

This matlab script will read values of binding energies from result pdbqt files and write them in a (**n+1)X20** matrix. Here n is number of Vina runs. n+1th raw contains values of max and min binding energies.

Copy this scripts in raw data folder and edit it accordingly for protein and ligand file names.

After running read_alldata.m then run **Parsepdb.m**

**Parsepdb.m**

This script will pullup coordinate of center of mass and create a matrix containing COM and binding energies for each confirmation. Edit script for file names accordingly. Then run **FindDistancesFromClusterCentroid.m**

**FindDistancesFromClusterCentroid.m**

This script will take centroid data obtained from pymol and generate and a python script to plot number and mean binding energy for each cluster. Edit accordingly.

**summaryBE.m**

This script will pull up max and mean values of BE of ligands.

# Pymol Commands

**set mouse_selection_mode=4**
It will set mouse mode to select a single atom.
**set all_state=off**
It will set all state a confirmation on or off. Use it to select any single atom of ligand as approximate position of centroid manually. Find the state which seems to have all clusters visible.

**print cmd.get_coords('sele',-1)**
**iterate_state -1,sele, print x,y,z**
Both these commands will print coordinate of centroid. Copy them to script *FindDistancesFromClusterCentroid.m*

**png C:\Users\nshukla\Desktop\PSE files\png\1BH6_substrate , width=1200, height=1200, dpi=2400**
This command will print a publication quality image in png format.

**rotate x, 45**
It will rotate view camera around x axis by the specified angle.

**set label_position, [x,y,z], objectA**
This command can be used to adjust position of label.
**delete sele**
This command will delete selection named sele.