

pMHCPred

pMHCPred is an initial step towards the structural based prediction of peptide binding to MHC class II molecule.

Prerequisites for MHC2model:

Data

1. PDB of MHC class II molecule
2. Template of the peptide
3. Peptide sequences and activity

Tools to run MHC2model

1. DatasetDivision
Download: <https://sourceforge.net/projects/qsardatasetdivisiongui/>
2. FoldX
Download and manual: <http://foldxsuite.crg.eu/documentation#manual>
3. ZDOCK
Download and manual: <http://zdock.umassmed.edu/software/>
4. PyPLIF-HIPPOS
Download: <https://github.com/radifar/PyPLIF-HIPPOS>
5. Weka v3.8
Download: <https://www.cs.waikato.ac.nz/ml/weka/downloading.html>

Refer the manual of the tools for the ease of the downloading and installation procedure and keep all the tools and scripts in one folder.

STEP-I: dataset division

- To divide the sequences in the test set and training set add sequence based descriptors with the activity as the input excel file in DatasetDivision.
- Set training and test set ratio 70:30 and run. It will divide the sequences in the training set and test set.

STEP-II: generation of peptide structures

- Make a template text file of the one letter code, chain name, residue no and amino acid code of the peptide.pdb file to generate the structure of the peptide sequence.
- The template and peptide sequence length has to be same.
- Run the perl script foldx_1.pl it will generate the structure. Give the command as follow:

```
perl foldx_1.pl [sequencelist.txt] [peptidefile]
```

STEP-III: docking

- To prepare the receptor follow the steps:

- Open the terminal and run the command to modify receptor:
`./mark_sur [proteinfile] [outputfile]`
- Make list text file of residue numbers which has to block to specify the binding site according to respective chains. Now run the command in terminal:
`perl block.pl [list.txt] [modified.pdb] > [Prot_receptor.pdb]`
- Execute the `zdock_2.sh` script to run the docking so it will generate a `zdock.out` file which the score and a `zdock.out.pdb` file which is docked complex.

STEP-IV: to generate structural fingerprints

- Give the residue index of the peptide interacting residues in the config file as shown in the example.
- Run the commands to activate PyPLIFHippos and script in terminal:
`conda activate base`
`conda activate hippos`
`bash hippos_3.sh`
- it will create cvs file of structural fingerprints.

STEP-V: affinity prediction using Weka

- To generate the model first prepare the data as shown in the example, add the descriptors in the first then fingerprints and at the end activity of the peptide.
- Open the weka 3.8. Click on “explorer” load the test and training set csv file and one by one select all attributes, save as the file in .arff format.
- Now load the training .arff file, go to classify tab, select **RandomForest** and give the output file format and name select true for all the options it will generate the model of the training set save the model, to get the better result run the autoweke.
- Load the training set model, select the “supplied test set” and set. Open the test.arff file and Class – (Num) Activity. After following all the steps right click on the model to select “Reevaluate on current test set”.
- The results will be displayed in the white space of “Classifier output”. We can see the binding affinity predictions for the MHC-II and peptide complexes under the “predicted” and “actual” column also correlation coefficient.

References:

1. Istyastono, E., Radifar, M., Yuniarti, N., Prasasty, V. and Mungkasi, S., 2020. PyPLIF HIPPOS: A Molecular Interaction Fingerprinting Tool for Docking Results of AutoDock Vina and PLANTS. *Journal of Chemical Information and Modeling*, 60(8), pp.3697-3702.
2. Chen, R., & Weng, Z. (2003). A novel shape complementarity scoring function for protein protein docking. *Proteins: Structure, Function, and Bioinformatics*, 51(3), 397-408.
3. Martin, T. M., Harten, P., Young, D. M., Muratov, E. N., Golbraikh, A., Zhu, H., & Tropsha, A. (2012). Does rational selection of training and test sets improve the outcome of QSAR modeling?. *Journal of chemical information and modeling*, 52(10), 2570-2578.
4. Van Durme, Joost, Javier Delgado, Francois Stricher, Luis Serrano, Joost Schymkowitz, and Frederic Rousseau. "A graphical interface for the FoldX forcefield." *Bioinformatics* 27, no. 12 (2011): 1711-1712.
5. Attwal, K. P. S., & Dhiman, A. S. (2020). Exploring data mining tool-Weka and using Weka to build and evaluate predictive models. *Advances and Applications in Mathematical Sciences*, 19(6), 451-469.