Read me for DeepBSRPred

We have developed DeepBSRPred, a binding site residue prediction method using protein sequence and predicted structures from AlphaFold2. DeepBSRPred is a deep neural network. It uses sequence-based features such as position-specific scoring matrix (PSSM), solvent accessibility, conservation score, and amino acid properties from the AAindex database, as well as structural features including residue depth, obtained from the predicted 3D structure of a protein using AlphaFold2. Our method performs better with an average F1 score of 0.73 and MCC of 0.22, which is better than existing methods.

To use codes, the user needs to install the following packages.
Kindly note that these codes will run on Linux operating system only.

Dependencies are:
1. python3.7 or above
2. argparse
3. pandas
4. TensorFlow
5. Numpy
6. Sklearn
7. Biopython
8. Requests

Software needs to be installed by the user, and an appropriate license should be taken from the author:

1. **DSSP:** The DSSP program defines the secondary structure of each amino acid residue, geometrical features, and solvent exposure using atomic coordinates. We employed DSSP to get the secondary structure of each residue (Kabsch & Sander, 1983).
2. **Naccess:** Naccess is a stand-alone program that calculates the accessible area of a molecule from PDB (Hubbard et al.,1993).
3. **HBPLUS:** HBPLUS is a hydrogen bond calculation program that provides hydrogen bond donors and bond acceptors (McDonald et al., 1994).
4. **AACon:** AACon is a Java implementation of 18 methods of scoring amino acid residue conservation in multiple sequence alignments (Agnieszka et al., 2018). The majority of the methods are described in Valdar's (2002), as well as the more complex SMERFS algorithm for predicting protein functional sites.
5. **PSI-Blast** (Altschul et al., 1997) to compute PSSM
6. **AlphaFold2** (Varadi et al., 2021) for predicting the 3D structure of a protein.

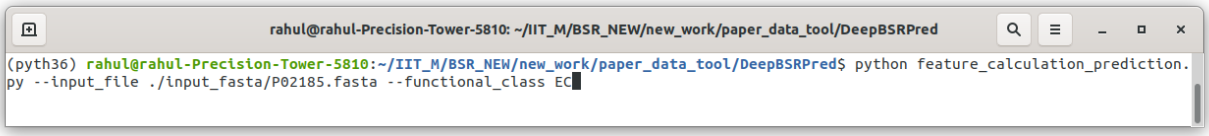
To use the standalone version, download and extract the codes from https://web.iitm.ac.in/bioinfo2/deepbsrpred/DeepBSRPred.zip and go to the DeepBSRPred directory. In this directory following codes are kept:

1. **feature_calculation_prediction.py**: this program will predict the binding site residues from a protein sequence. Results will be saved in the "**input_file_name_result.csv"** file.

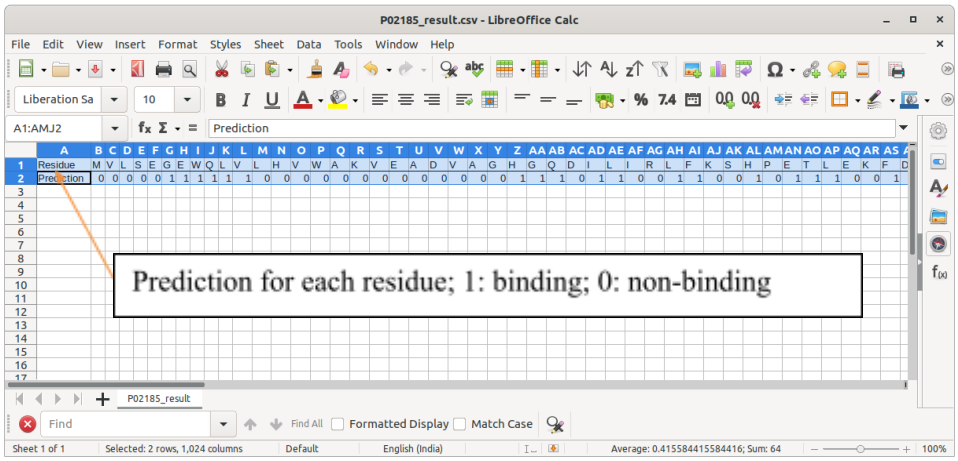
[Note: Download the standalone version extract the data, install all the tools in the DeepBSRPRed directory, and change the respective paths in the **feature_calculation_prediction.py** file to run the program].

Example:

To Run the program, go to the DeepBSRPred directory and run the following command: **python feature_calculation_prediction.py --input_file ./input_fasta/P02185.fasta –functional_class EC**



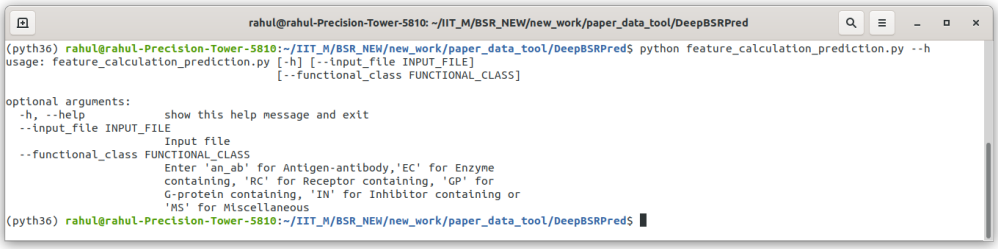
If everything works well, the result will be saved in the '**P02185_result.csv**' file in this case.



To know more about input arguments, run the following command:

Python feature_calculation_prediction.py –h

It will provide information regarding input arguments for the program.



Kindly contact us if you face any difficulty in running the program.
Contact details are available at
https://web.iitm.ac.in/bioinfo2/deepbsrpred/contact-1.html