ProtFLEXpreD

Laura Ciaran Alfano, Neus Pou Amengual

MSc in Bioinformatics for Health Sciences

Project of SBI - PYT

2021 - 2022

Introduction

ProtFLEXpreD is a python package to predict *b-factors* of each protein aminoacid, which are values used to predict the protein flexibility.

The package works as follows. Firstly, if you do not have the PDB database in your computer, it will be downloaded in `DB_pdb` directory. If you have it, please check that your directory is called `DB_pdb` and your files are called `PDB.phr`, `PDB.psq` and `PDB.pin`. Secondly, **BlastP** compares the protein query to the *PDB* database to gather its top 10 homologous proteins. Thirdly, *b-factors* of the homologues are obtained and normalized from the **PDB** PDB files. Fourthly, *Multiple Sequence Alignment* is performed with **ClustalW**. Fifthly, *b-factors* from homologous proteins are assigned to the aligned aminoacids by computing the mean between them, and the *b-factors* obtained according to their neighbours by computation are assigned to the non-aligned aminoacids. Finally, a text file containing the *b-factor* of each aminoacid of the query sequence and three plots are created to represent the flexibility results.

Initializating the program

Download the `ProtFLEXpreD-1.0.tar.gz` compressed file and execute the following command to install all required packages. Make sure that you are in the correct directory. It may be possible that you have to execute it as root.

```
pip3 install ProtFLEXpreD-1.0.tar.gz
```

Make sure that you have successfully downloaded the package and continue.

Running the program

Execute `ProtFLEXpreD` with the following command to run the program.

```
python3 -m ProtFLEXpreD.ProtFLEXpreD -i XXX
```

Where you have to write your **input -i** argument replacing `XXX` by the **Uniprot ID** or the **fasta** file of your ptotein query. You can also add the following arguments:

- `-o` : output file. It is saved as default in `/Results/predicted_bfactors.txt`.
- `-f`: output figure. It is saved as default in `/Results/flexibility_plots.png`.
- `-v`: verbose. It is `False` as default.

If you want to know more, write the following command-line.

```
python3 -m ProtFLEXpreD.ProtFLEXpreD -h
```

ProtFLEXpreD Results

After running the program successfully, you will find three new folders in your directory:

- `Downloads` : where you can find all downloaded files from **PDB** or **Uniprot**.

- `Intermediary` : where you can find **BlastP** results, **ClustalW** alignment files and the predicted flexibility calculated by the *biopython* package.

- `Results` : where you can find a file with the predicted *b-factors* of your query and representation plots of your results.

Examples

You will find below some examples to better understand how execute the program.

- Executing the program printing the progression log. You will find an uniprot ID as inout. Output and figure arguments are not added.

```
python3 -m ProtFLEXpreD.ProtFLEXpreD -i P65206 -v
```

- Executing the program without printing the progression log. You will find a fasta file uniprot ID as input and a text file as output. Figure argument is not added.

```
python3 -m ProtFLEXpreD.ProtFLEXpreD -i P65206.fasta -o ./Results/P65206.txt
```

- Executing the program printing the progression log. All possible argument will be added.

```
python3 -m ProtFLEXpreD.ProtFLEXpreD -i P65206 -o ./Results/P65206.txt -f ./Results/P65206.png -v
```

If you have any doubt, do not hesitate to contact with us, laura.ciaran01@estudiant.upf.edu and neus.pou01@ estudiant.upf.edu.