PPTStab User Manual

Introduction

PPTStab is a tool designed to predict and design thermostable proteins based on their melting temperature. It uses an ensemble regressor model combining ANN and MLP. The standalone version and web-server are available. For detailed information, including the underlying algorithms, visit PPTStab Web Server https://webs.iiitd.edu.in/raghava/pptstab.

Installation

To use pptstab, follow these steps to set up your environment:

- **Install using environment.yml**

  1. **Create a Conda Environment:**

     ```
     conda env create -f environment.yml
     ```

  2. **Activate the Environment:**

     ```
     conda activate pptstab
     ```

- **Standalone Version**

  The standalone version of **PPTStab** requires the following Python libraries:
  - scikit-learn==1.0.2
  - transformers==4.44.2
  - tensorflow==2.13.0

- pandas==2.0.3
- numpy==1.22.4
- torch==2.4.1

- **Minimum Usage**

3. To view available options, run:

> **python3 pptstab.py -h**

4. To run the example, use:

> **python3 pptstab.py -i example.fasta -f 1**

This command predicts whether the submitted sequences are thermostable and saves the results in outfile.csv.

- **Full Usage**

Command-Line Arguments:

usage: pptstab.py [-h]
        [-i INPUT]
        [-o OUTPUT]
        [-j {1,2}]
        [-d {1,2}]
        [-f {0,1}]
        [-m {EMB,AAC,SER}]

- **Optional Arguments:**

- **-h** , --help

    Show help message and exit.

- **-i** INPUT, --input INPUT
    Input: Protein or peptide sequence(s) in FASTA format or single sequence per line in single letter code.

- **-o** OUTPUT, --output OUTPUT

  Output: File for saving results (default: outfile.csv).
- **-j** {1,2}, --job {1,2}

  Job Type: 1 for Prediction, 2 for Design (default: 1).

- **-f** FLAG, --flag FLAG {0,1}

  Cell Flag: Value between 0 or 1 (default: 1).

- **-d** {1,2}, --display {1,2}

  Display: 1 for Thermophilic proteins only, 2 for All peptides (default: 1).

- **-m** {EMB,AAC,SER}, --method {EMB,AAC,SER}

  Method: EMB for embedding model (ProtBert), SER for Shannon entropy, AAC for amino acid composition (default: SER).

- **PPTStab Package Files**

INSTALLATION: Installation instructions.

LICENSE: License information.

Manual: Provides information about this package.

pptstab.py: Main Python program.

example.fasta: Example file containing peptide sequences in FASTA format.

example_predict_output.csv: Example output file for the prediction module.

example_design_output.csv: Example output file for the design module.