CRAFT_cmd

User Manual

Cavity Recognition Assisted by Flow Transfer algorithm (CRAFT) is designed to identify diverse set of protein cavities such as void, channel, pocket, and surface pocket. It also characterizes the identified cavities. Its online version is also available at **http://pitools.niper.ac.in/CRAFT.**

CRAFT_cmd, a command-line version of the CRAFT tool that is designed for cavity prediction and the estimation of associated physicochemical properties. It facilitates user to process single and multiple .pdb files. The user manual gives a structured document that guides users on how to utilize the script, its functionalities, accepted inputs, and expected outputs.

Installation:

The user manual provides step-by-step instructions for installing Python, installing the SciPy library, downloading and unzipping the CRAFT_cmd file, downloading PDB files into the PDB folder, and running the algorithm on the command prompt. The algorithm provides identified cavities in descending order of cavity volume, along with information about cavity residue, residue number, atom number, atom type, and chain.

a) Install Python:

Visit the official Python website at https://www.python.org/downloads/

Choose the appropriate version of Python for your operating system and click on the download link.

Run the installer once the download is complete.

During installation, make sure to select the option to add Python to your PATH environment variable. This allows you to run Python from the command prompt.

b) Pip install SciPy:

Open the command prompt on your computer.

Type "pip install scipy" and press enter.

The installation process will begin and may take a few minutes to complete.

Once the installation is finished, you can start using the SciPy library in your Python projects.

c) Download the CRAFT_cmd file (CRAFT_cmd.zip) from the webpage:

Visit the **https://github.com/PGlab-NIPER/CRAFT/** where the **CRAFT_cmd.zip** file is available for download.

Click on the download link to download the **CRAFT_cmd.zip** file to your computer.

d) Unzip the downloaded file CRAFT_cmd.zip:

Locate the CRAFT_cmd.zip file on your computer and right-click on it.

Choose the option to "Extract" or "Extract all" to unzip the file.

Choose a destination folder to extract the files to.

e) Download PDB files into a PDB folder present inside the downloaded CRAFT_cmd folder:

Open the extracted CRAFT_cmd folder.

Locate the PDB folder inside it.

Download the PDB files that you want to analyze and save them to this PDB folder.

f) Open the command prompt inside the CRAFT_cmd folder:

Open the extracted CRAFT_cmd folder.

Hold down the Shift key and right-click inside the folder.

Choose the option to "Open command window here" or "Open PowerShell window here".

g) Use the command "Python main.py":

Type "python main.py" in the command prompt and press enter.

The algorithm will begin and take user input to start the cavity scan inside the given protein.

h) Follow the user instructions available on the command prompt:

Once the algorithm starts running, it will prompt you for the name of the PDB file that you want to analyze.

Enter the name of the PDB file and press enter.

The CRAFT_cmd provides information about the identified cavities in the PDB file, along with information about the cavity residue, residue number, atom number, atom type, and chain. The cavities are listed in descending order of cavity volume.

CRAFT_cmd structure:

It contains the **craft** package and the main.py file to access the functionality. The **craft** package includes the following components:

- __init__.py
- Base.py
- ReadPdb.py
- Tetrahedral.py
- Properties.py
- CraftVariables.py

docs: holds text files utilized by the **craft** package.

PDB Directory: where users must place their input .pdb files for processing.

Results: It store results from a single run of main.py file. It contains:

- **Cavity_files:** Stores cavity surface information for each input file in a .cav (text) format.
- **Property files:** Contains cavity properties identified during processing, saved in .pro (CSV) format.
- **Non_standard PDB.csv:** Records details of non-standard input files when processing multiple files.

Usage:

The CRAFT_cmd can be utilize for single file and multiple .pdb files. The details are:

Single .pdb File Processing:

Input:

- The system will prompt if you want to proceed with default CRAFT parameters. Answer 'Yes' or 'No'.
- If 'Yes,' default parameters are utilized. For 'No,' individual parameters are requested, with acceptable ranges provided and validated.

File Selection:

- Users will be asked to provide the .pdb file to process. e.g. Abc3.pdb
- Display of chains within the file will occur next, allowing users to select among them.
- If applicable, the system displays available ligands in the selected chains for user selection. Advised to avoid selecting if unnecessary.
- Information on alternate positions of residues within selected chains will be presented.

Processing: Data submission to the Flow Transfer algorithm occurs for further processing.

Output: Details will be shown on-screen and stored in the Results folder for future reference.

Multiple .pdb Files:

Input: Similar prompt regarding default parameters, allowing 'Yes' or 'No' for parameter selection and validation if 'No' is chosen.

File Selection:

- Users input the names of .pdb files to process, separated by commas (e.g., abc, xyz, pqr, mno). Ensure these files exist in the PDB folder within CRAFT_cmd.
- Automated selection of all chains and exclusion of ligands present in the files.
- Provision to modify the main file if any specific changes are required.

Results: Processed results are saved in the Results folder for convenient access and review.

Instructions for Customizing Parameters:

When prompted for parameters, adhere to the specified ranges and guidelines for optimal processing. Default parameters can be used by opting 'Yes' when prompted or configure individually for precise analysis. It can process single file as well as multiple file. Users will can opts for any one of them from the prompted instruction on the command prompt.

User is provided with following parameters:

- **Maximum Circle Radius (MCR):** MCR defines cavity scanning criteria for the Flow transfer algorithm. Lower values imply denser cavity surface atoms; higher values imply porous ones. Default signifies a water molecule.
- **Peeling off radius:** It defines delimiter tetrahedra in the Flow Transfer algorithm. Lower value strict boundary condition and vice versa.
- **Susceptible radius:** It determines prior cavity locations in protein for the Flow Transfer Algorithm, favoring intense scanning at lower values, gentle at higher ones

- **Susceptible volume:** It determines prior cavity locations in protein for the Flow Transfer Algorithm, favouring intense scanning at lower values, gentle at higher ones
- **Extend Boundary:** It extends cavity entrance (default: false). For boundary cavities it can be set True i.e. 1.
- **Extend Forbidden Face:** It densifies the cavity surface atoms by incorporating atoms with limited involvement into it.
- **Cavity atoms:** Minimum number of atoms present in a cavity