**https://github.com/githubdas4/RAAPI**

*RAAPI* (Reduced amino acid pattern identification)is an executable program written in python script. It is developed to identify all possible non-overlapping conserved patterns among protein sequences. Generating and identifying pattern, it utilizes eight groups of reduced amino acid alphabets (See supplementary file: Table S1).

*How to use RAAPI: Reduced amino acid pattern identification?*

**1) Software and libraries (dependencies)**

**Python version: Python 3.6 or higher**

**Python libraires:**

biopython==1.79

easygui==0.98.2

numpy==1.22.3

**2) Operating system compatibility:**

Recommended: Windows 10 version 17763.0 or higher(64 bit). For Windows 7 or 8 users, one must check for the version of python and python libraries as listed above.

**3) Usages**

**Download executable file:**

There is a single executable file (***RAAPI***) written in python script. This can be downloaded to run any supported python installed personal computer (point 1 & 2).

**Run:** Execute the python script through command line or can be executed by double click on the executable file (***RAAPI)***. Once executed, a window will display, one needs to provide five inputs variables, which are described below.

1. **Input parameters:**
2. Targeted group protein sequences in fasta format (say, *Tg.fasta*) [pattern will identify for *Tg* group]
3. Remaining group protein sequences in fasta format (say, *Rg.fasta*) [identified patterns from *Tg* group will check their occurrences in *Rg* group]
4. Searching upper pattern length (*pu*), say, *25*
5. Searching lower pattern length (*pl*), say, *3*
6. Effectiveness score threshold (*Et*), say *20* [For an identified pattern in *Tg* group, Effectiveness threshold (Et) is the percentage of protein sequences in *Rg* group having the same identified pattern. The patterns with Ef score below this provided threshold will be discarded.

Graphical user interface, application

Description automatically generated

Once provided all the input variables, click ok to go next step to run the algorithm.

**Important points to be noted: (input parametrs)**

1. Keep both the ***Tg*** and ***Rg*** fast file in the same directory of executable file (***RAAPI***)***.***
2. If one needs to identify pattern only for a specific targeted group (***Tg***), provide the same for ***Rg*** file too. In that case no *Ef* score will be provided in the output file but must give any value >=0 for ***Et.*** (See below)
3. Must be ***pu>pl***

Graphical user interface, application

Description automatically generated

1. **Output files:**

If algorithm is successfully executing in every step, it will provide console output as intermediate results status (quantitative), and finally will produce/save the following 3 output files.

1. Only common patterns list (identified in ***Tg*** group) with pattern length (saved as ***Tg\_result1***)
2. Reduced amino acid patterns (pattern as numerical numbers: 1-8, as per grouping of amino acid used) with Ef score (<=Et), sequences ids (saved as ***Tg\_result2***) [both for *Tg* and *Rg* (if any)]
3. Numerical patterns with Ef score (<=Et), position, and amino acid sequence for each sequence (saved as ***Tg\_result3***) [both for *Tg* and *Rg* (if any)]

**C) Exestuation time:** [checked with windows 10 64-bit, 2.21 GHz, 16GB RAM]

Program usually takes less than 45s to generate, identify and save common pattern list for around 100 sequences (25 targeted group protein sequences and 75 remaining group protein sequences) with average length of each protein sequence about 250-300 amino acids.