

IPA: Probabilistic Annotations of Protein Sequences for Intrinsically Disordered Features

User's Guide

By: Nawar Malhis

This IDR Probabilistic Annotation (IPA) platform predicts 'Linker' regions and 'nucleic', 'protein', and 'all' (protein or nucleic) IDR binding sites within protein amino acid sequences.

Installing IPA:

- Download: `wget orca.msl.ubc.ca/nmshare/ipa.tar.gz`
- Unpack: `tar -zxvf ipa.tar.gz`
- Change directory: `cd ipa/`
- Create a conda environment:
 - To create a Cuda environment:
 - `conda env create -f ipa_cuda.yml` and then:
`conda activate ipa_cuda`
 - To create a CPU environment:
 - `conda env create -f ipa_cpu.yml` and then:
`conda activate ipa_cpu`

Input data:

All input sequences can be in a single fasta file, so each sequence has a fasta header line with a unique accession and a sequence line. Example of a sequence in a fasta format:

```
>P07766
MQSGTHWRVLGLCLLSVGWVGQDGNEEMGGITQTPYKVSISGTTVILTCPQYPGSEILWQHNDKNIGG...
>P05067
MLPGLALLLLAAWTARALEVPTDGNAGLLAEPQIAMFCGRLNMHMNVQNGKWSDPSGKTCTIDTKEG...
>P38398
MDLSALRVEEVQNVINAMQKILECPICLELIKEPVSTKCDHIFCKFCMLKLLNQKKGPSQCPLCKNDI...
```

To run IPA:

```
python3 ipa.py -p PATH -in INPUT_FILE [-af2] [-cpu]
```

-af2 is used to run IPA-AF2

-cpu can be used to force running on a CPU in a Cuda environment.

INPUT_FILE: is the input fasta file. This input file can be located in any accessible directory.

PATH: is the path (directory) of the INPUT_FILE

IPA will create a directory "results" inside PATH to store all IPA predictions.

Example:

```
python3 ipa.py -p test_data/ -in input.fasta
python3 ipa.py -p test_data/ -in input.fasta -af2
```