MrParse

Release 0.2.5

Adam Simpkin, Jens Thomas, Ronan Keegan & Daniel Rigden

November 11, 2021

CONTENTS

ABOUT

MrParse is a CCP4 program takes a protein amino acid sequence file and searches for homologs using PHMMER (default) or HHSEARCH. If supplied with a reflection data file (currently in MTZ format), it can then use PHASER to calculate the Expected Log Likelihood Gains (eLLG) values for the homologs. It also searches the EBI AlphaFold database for related models.

It also attempts to classify the sequence according to its secondary structure, and whether any regions are expected to be Coiled-Coil or Transmembrane.

Results are currently displayed in a simple HTML webpage that is rendered using VUE. The sequence graphics are created using the PFAM graphics library, a copy of which is distributed with this code.

INSTALLATION

MrParse is distributed with CCP4, although optional software can be installed to get the most out of MrParse. Full details are provided here

THREE

SIMPLE COMMAND LINE

```
mrparse --seqin <PATH TO SEQUENCE FILE>
```

To provide a reflection file and classify the sequence we can provide the following optional flags:

```
--hklin <PATH TO MTZ FILE>
--do_classify
```

FOUR

SEARCH MODEL FINDER

The search model finder by default uses PHMMER (distributed with CCP4) to search for homologs. If installed you can also use HHSEARCH. Examples of how to use MrParse are provided here

FIVE

EBI ALPHAFOLD DATABASE SEARCH

The search model finder currently uses PHMMER (distributed with CCP4) to search the EBI Alphafold database. This will be replaced with the 3Dbeacons API when it becomes available.

CLASSIFIERS

- secondary structure classification is currently carried by submitting jobs to the JPRED server.

- If installed, coiled-Coil classification is carried out with Deepcoil.

- If installed, transmembrane classification is carried out with TMHMM.