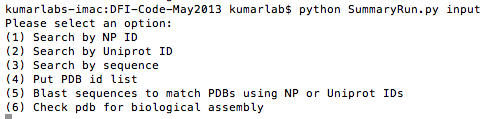
README

To run it:

python SummaryRun.py input



p.s. To get DSSP results, the DSSPPath defined in SummaryFunctions.py should be corrected. It changes accordingly to machine

**SummaryRun.py**: This code contains the main interface

**Input**: input file that includes minimum sequence identity and minimum query coverage you want to search in BLAST.

In the current version, there are 6 options:

When you choose:

Option (1), it asks

“Please enter the name of a .dat file containing NP IDs:”

This file should include the refseq id of protein you are interested.

Option (2), it is based on Uniprot ID of protein you are interested.

Option (3), it does BLAST using the sequence information

Option (4), if you want to skip BLAST and to run pdb IDs, you need to choose this option

Option (5), You provide a list of NP and Uniprot IDs and it automatically do BLAST and find the mapping structures

Option (6), If you want to check if pdb has biological assembly or not.

**SummaryFunctions.py** includes all necessary function used.

There may be some redundancies because each module were modified or developed for their unique purpose.

Functions: (Naming should be obvious for their purposes)

**read\_fasta**

**getPDBFiles**

**createCAfile**

**replaceBValues**

**npBLAST**

**seqBLAST**

**parseBlastFile**

**unBLAST**

**matchPDB** (Finds matching PDB structure)

**matchPDBHighThroughput** (This is modified for option (5))

**runPDB** (To run prs code)

**runPDBListforBA** (If the structure has BA, it automatically uses this function)

**runPDBList**

**readPDBFileforNP** (This fetches the pdb structure and build fortran code when you use NP id-Option(1))

**readPDBFile**(This fetches the pdb structure and build fortran code when you use NP id-Option(4))

**ObtainMonomerNMR** (It obtains the first model from NMR models)

**splitNMR**

**ObtainMonomer** (It obtains monomer chain from X-ray structure)

**ObtainMonomerFromComplex** (If the structure is complex, it obtains monomer from the complex one)

**RunFortranFile**

**GetData** (This contains mainly analyzing the obtained data)

**GetDataForNMRStructures** (This contains mainly analyzing the obtained data for the NMR structure)

**GetDataForComplexStructures** (This contains mainly analyzing the obtained data for the complex structure)

**writemas** (writes out all data to a single file)

**runDSSP**

**pdbsum** (This module works but not implemented into the main part yet)

**msv3d** (This module works alone to get data from MSV3D website)