Session I. Introduction to protein structure manipulation in python

Objective

To acquire the basic skills to load/save a protein structure in PDB format, identify their components, obtain basic information, and perform geometric measures on the structure

Software and libraries	• Python >= 3.6 • Biopython module (>= 1.72) • Molecular viewer (pymol, chimera) • biobb_structure_checking python module
Code Examples	https://github.com/jlgelpi/Biophysics
Biopython reference	• Biopython tutorial • Bio.PDB tutorial • Biopython reference

Conda installation (recommended)

Includes biopython, and python 3.6. Requires Anaconda or miniconda installed.

1. Create new environment
   ```
   conda create -n env_name
   ```
2. Install module Biobb_structure_checking
   ```
   conda install -c bioconda biobb_structure_checking
   ```
3. To add bioconda channel permanently
   ```
   conda config --add channels bioconda
   ```

Examples

https://github.com/jlgelpi/Biophysics/tree/master/Examples

- ex_cmd_line.py: Simple command line
- ex_distances.py: Search for contacts
- ex_distances_2res.py: Print distances between atoms
- ex_list_res.py: Print atoms and coordinates for ARG residues
- ex_list_res2.py: Print residue atoms of a residue number
- ex_chains.py: Remove a list of chains and save the remaining in a PDB file

Exercises

Prepare scripts for all exercises with the appropriate command line management (using argparse).
Output lists should be sorted (by residue o atom number) when appropriate and formatted for an easier read.

Upload a tar.gz file with codes (or github link) and examples of output for each exercise.

1. Determine the list of residues whose CA atoms are closer than a given distance
   Parameters: PDB file name, distance
1. Generate a list of all atoms for a given residue number
   Parameters: PDB file name, Residue number (Including Chain if applicable)
2. Determine all possible hydrogen bonds (Polar atoms at less than 3.5 Å).
   Parameters: PDB file name. Optional: cufoff distance (defaults to 3.5)
3. Generate a list of all CA atoms of given residue type with coordinates
   Parameters: PDB file name, residue type.
   Optional: accept residue codes in one-letter format
4. Generate a list of backbone connectivity (i.e. which residues are linked by ordinary peptide bonds).
   Parameters: PDB file name
5. Print all distances between the atoms of two given residues
   Parameters: PDB file name, Residue 1, Residue 2
6. Obtain a PDB file including only a list of chains from the original structure
   Parameters: PDB file name, list of chain ids