Project report: Secondary structure assignment

Myriam HANNA

September 18, 2019

Introduction

DSSP is a secondary structure assignment method. This algorithm is widely used in protein structure predictions. It is based on computing the energy of hydrogen bonds between atoms. Elementary hydrogen bonds pattern include turns and bridges. When those structures are assembled into specific patterns, they form helices and β sheets.
The purpose of this project is to implement this algorithm in Python.

Materials and methods

The program requires a file in the PDB format as input. First, it reads the file line by line. it then retrieves the information present in the header. The script retrieves next the coordinates of the atoms implicated in the hydrogen bonds and stores them in a dictionary so that the data could be easily accessible. The next step is to compute the energy of the hydrogen bonds. The function computes the distances between the atoms of the H bonds and returns the energy following this formula:

$$E = q1 * q2 * (\frac{1}{rON} + \frac{1}{rCH} - \frac{1}{rOH} - \frac{1}{rCN}) * 332$$

With r as the distance between the two following atoms, q1 = 0.42e and q2 = 0.20e, e being unit electron charge.
A hydrogen bond is found if this energy is < 0.5 kcal/mol.
The next two functions search for the basic structures that defines helices and sheets. Those structures are n-turns and bridges. The most important functions are listed following:

- def energy_Hbonds(self)
- def find_nturns(self)
- def find_bridges(self)
- def find_helices(self)
- def find_ladders(self)
- def find_sheets(self)

Results

First, n-turn pattern were tested. The residues that were in 3,4,5 turn were classed in a dictionary. Then, bridges were evaluated on every residues. Two types of bridges were found, anti-parallel and parallel. The residues were again classed into a dictionary. Furthermore, helices were computed based on the turns and ladders based on bridges. Finally, β sheets were obtained. Each residue is stored as turn, bridge, ladder or sheet.
The file used to test this program was: "1est.pdb" with an output file "1est.dssp". Comparable results were obtained.

Conclusion

In conclusion, the program written assigns elementary structures such as turns and bridges and cooperative H-bond pattern such as helices, β ladders, β sheets to the amino acid of a pdb file.

References

Kabsch W, Sander C, (1983) . Dictionary of protein secondary structure: pattern recognition of hydrogen bonded and geometrical features.

Annexe

I encountered some difficulties at first with the Hydrogen atoms that were not present in all PDB files. This problem was solved by using Pymol, which add the missing hydrogen atoms.
The more advanced structures and geometry, such as bends, chirality, SS bonds, chain breaks, were not evaluated due to the lack of time.