Project n° : ASSIGNMENT AND DETECTION OF TRANSMEMBRANE SEGMENTS OF A PROTEIN

Emma SOUFIR

M2 BI

# Introduction

Transmembrane proteins are a crucial class of molecules which essential roles in various cellular processes. Understanding the structures and functions of these proteins is central importance in biology and has many implications for drug discovery or disease understanding. However, the identification and classification of transmembrane proteins is a difficult process because the membrane is not resolved when the structure is.

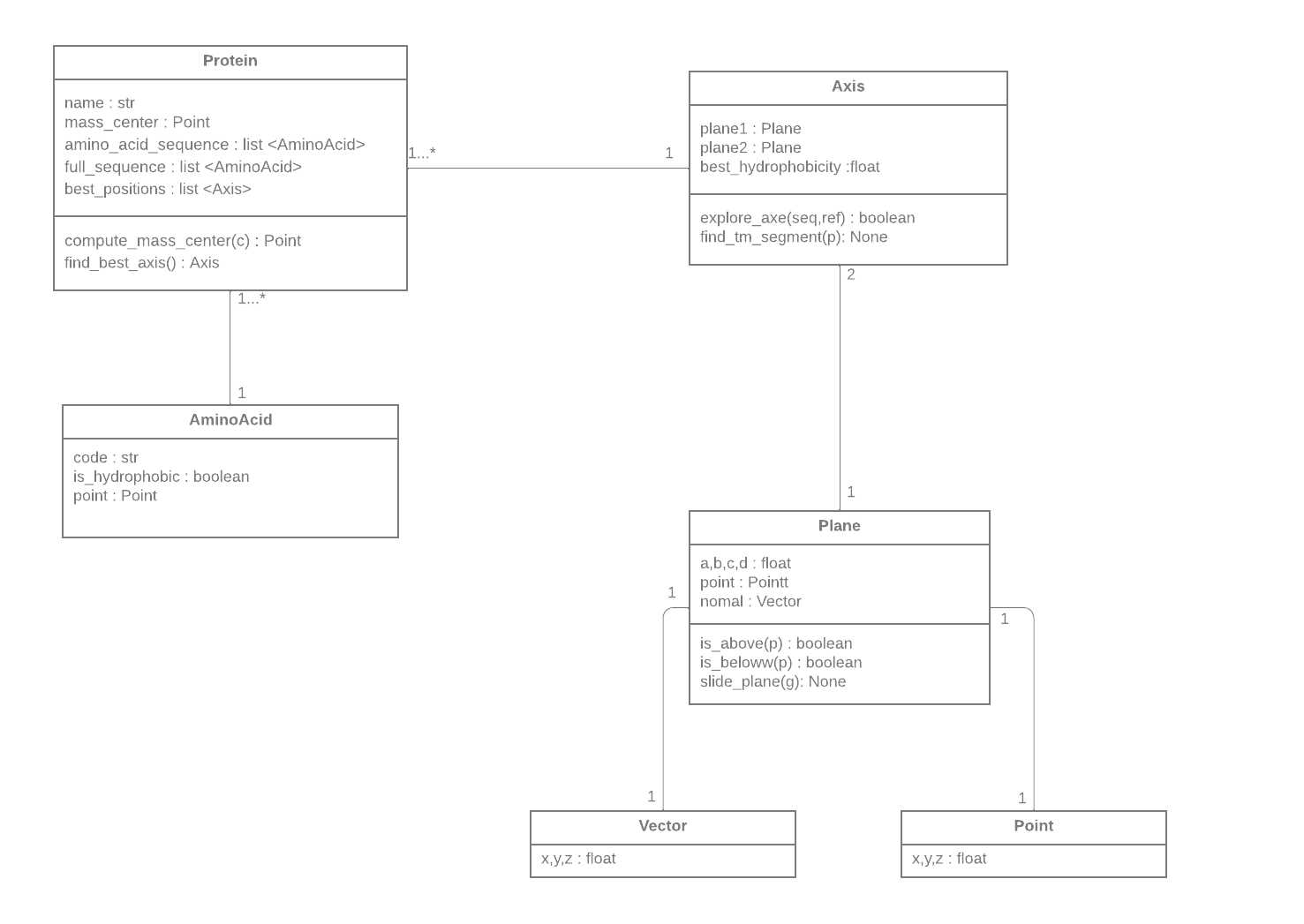
Here, we created a tool to identify membrane limits of transmembrane proteins inspired from the method used in [1,2].

# Material and Methods

## Algorithm

The main goal of this algorithm is to find the best position of a membrane, represented by two planes. The best position is defined by the positions of the two planes which maximizes the relative hydrophobicity. The purpose is to maximize the number of hydrophobic exposed-to-solvent residues between the two planes. First, the correct orientation of the membrane is identified, than the membrane’s width is adjusted (Fig ?)

To model this problem, we used OOP paradigm, which is quite suitable in order to encapsulate many information together on a particular object. Fig ?



Firstly, the PDB file is checked and parsed to create and initialize the Protein object, using AminoAcid objects to encapsulate information on residues. The solvent accessibility of each residue is computed using the DSSP [3] algorithm. A residue is considered accessible if its DSSP solvent accessibility is above 0.3. The coordinates of the Cα are also registered in each AminoAcid object. Finally, the hydrophobicity of the residue is registered. The residues considered as hydrophobic were the same as the article [1,2].

The center of mass of the chain A of the protein is calculated. From this center, N points equally distributed on a demi sphere using the Staff and Kuijlaars method [4] are generated. Therefore, there are N vectors generated between the center of mass and each point. For each of those vectors, two orthogonal planes are generated, separated by a certain gap (14 °A by default). Those information are stocked in an Axis object.

For each of those axis, the two planes are slide firstly above then below. After each slide (1 A by default), the relative hydrophobicity of the residues located between the two planes is calculated, using the formula :

If the relative hydrophobicity is higher than the best match yet, this axis is more advantageous. It is stocked in an object. This step is repeated again until there are no more atoms in between the two planes, updating the value of the more advantageous axis if necessary. This sliding is also performed in the other direction of the axis. At the end, for one axis, the best position for the planes is stocked with the associated relative hydrophobicity. This step is repeated for each point on the half-sphere. Then the axis with the best hydrophobicity ever is conserved.

Finally, the width of the membrane is adjusted by exploring the best axis step by step above and below for each plane.

## Data set used for evaluation

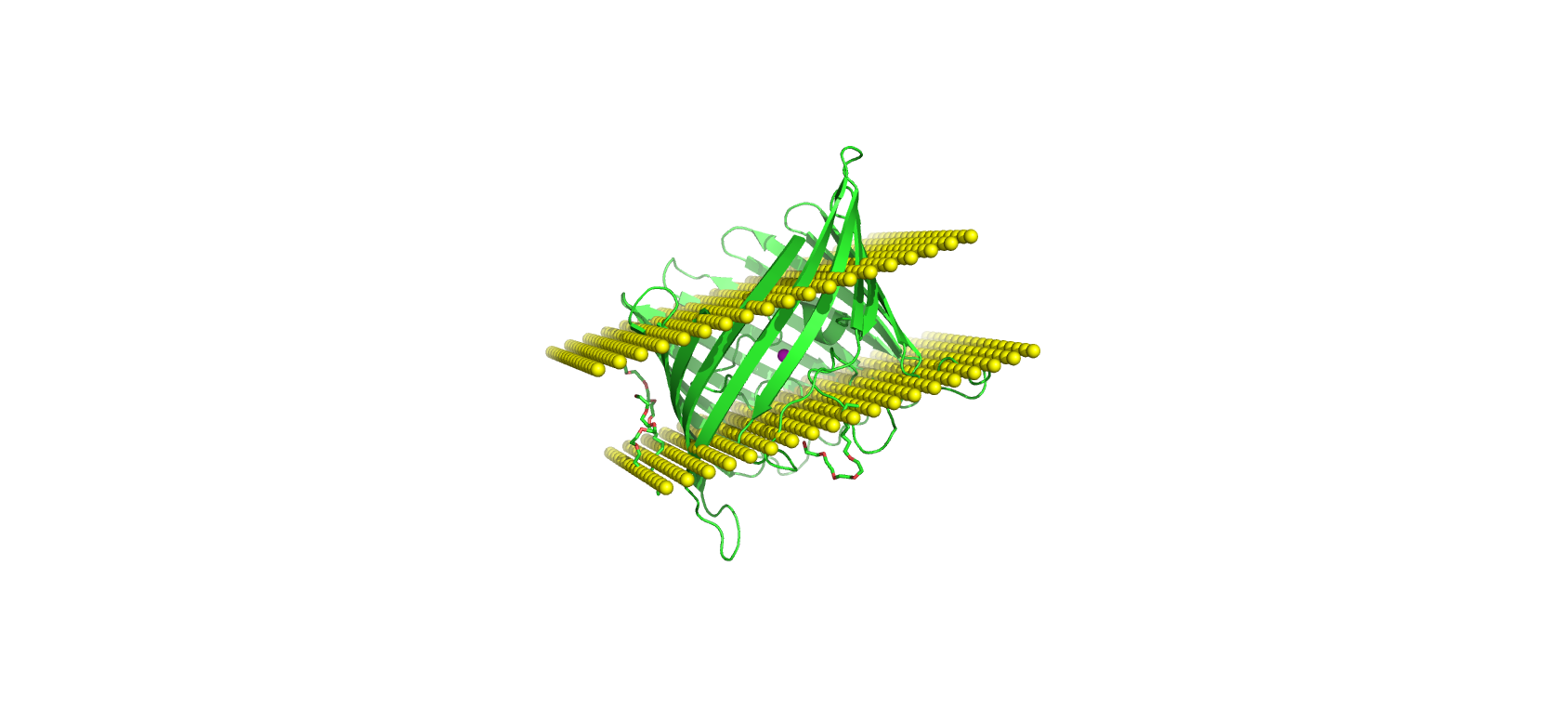
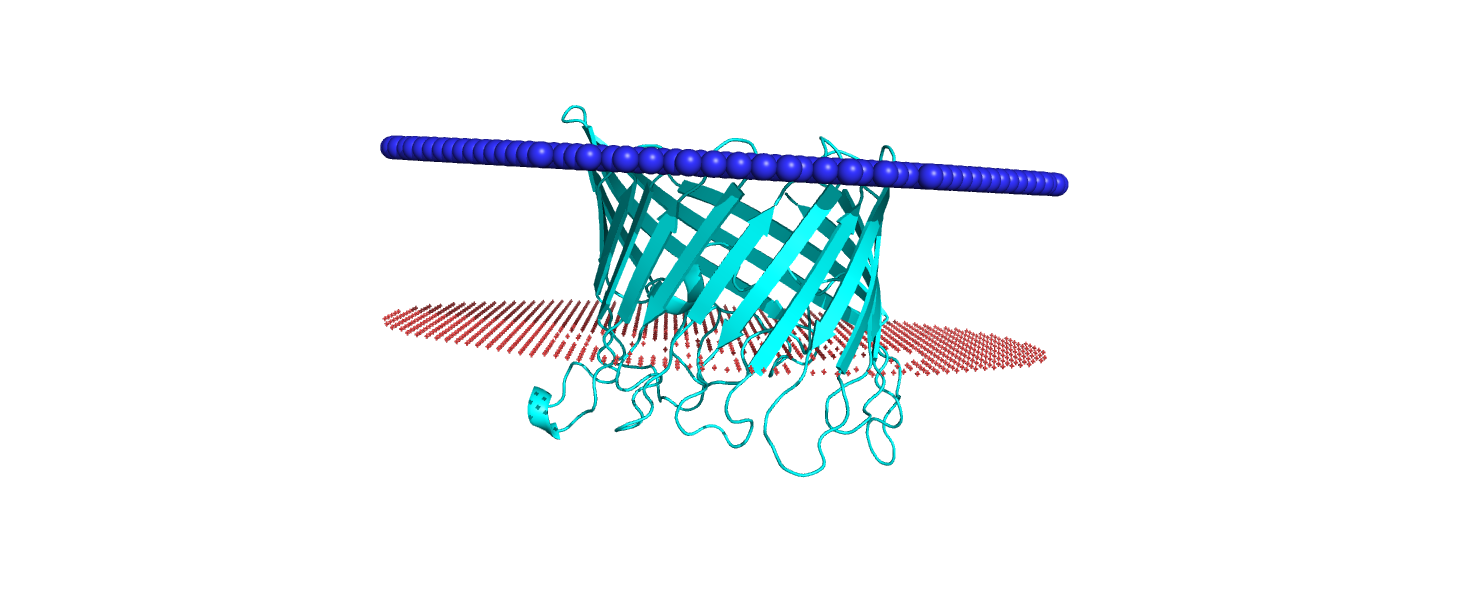
Since a protein is simplified here as a unique chain, we used only small and simple transmembrane proteins to evaluate our tool.

## Outputs

The program opens a PyMol [5] GUI with the protein and the predicted planes as heteroatom objects.

Another output file containing the transmembrane segments is generated in the ./results folder.

# Results and Discussion

Comparison between the output of our program (left) and the PDB file found in OPM (right). The following command was used to generate the output : python TM\_detect.py ../data/1prn.pdb -n 30.

In order to compare our results, we used the PDB files provided by the OPM data bank [6]. The orientation of the planes seems coherent between the two programs. We can see that the width of the membrane is smaller in our results than the one proposed by OPM. However the position of it seems coherent with what is given by databank.

In order to extend this tool several leads could be investigated. Firstly, we could consider adding the possibility of detecting the membrane on a poly-chain protein to extend the possibilities of the tool. We should also consider adapting the way of calculating the relative hydrophobicity because the tool has issues when it comes to deal with small transmembrane areas with large non transmembrane areas. Finally, as mentioned above, we could improve the optimization of the membrane’s width by being more tolerant in finding the best axis with the best planes. To solve this problem, we could also change the algorithm by starting with a quite large membrane and reduce hit until the hydrophobicity is significantly lower. ??? Essaye tu peux le faire !!!

### References

[1] Tusnády GE, Dosztányi Z, Simon I. Transmembrane proteins in the Protein Data Bank: identification and classification. *Bioinformatics*. 2004;20(17):2964-2972. doi:10.1093/bioinformatics/bth340

[2] Tusnády GE, Dosztányi Z, Simon I. TMDET: web server for detecting transmembrane regions of proteins by using their 3D coordinates. Bioinformatics. 2005;21(7):1276-1277. doi:10.1093/bioinformatics/bti121

[3] DSSP : <https://swift.cmbi.umcn.nl/gv/dssp/>

[4] Saff B, Kuijlaar A.B.J. , Distributing Many Points on a Sphere, 9 Springer Verlag New York, Volume 19. Number 1, 1997

[5] PyMol Software : <https://pymol.org/2/>

[6] OPM : <https://opm.phar.umich.edu/>