Development of an intra and inter protein calculator - CIP

Bertrand BOUVAREL - M2BI
September 2019

Introduction

Interactions, also called non-covalent bonds, play a major role in establishing the tertiary structure of a protein. Intra and inter protein interactions also allow the stabilization of protein conformation and thus complex formation. It's therefore easy to understand the importance of knowledge and calculation of protein interactions in structural biology issues, especially for questions about the relationship between the sequence and the structure of a protein, protein modeling or the structure prediction by sequence homology. So we had a goal with the CIP program, inspired by the article by K. G. Tina et al. [1], to perform the calculation of interactions such as hydrophobic, ionic, aromatic-aromatic, aromatic-sulfur, cation-pi, disulfide bridges or hydrogen bonds using one or more proteins from a PDB file.

Material and methods

1 PDB file

The protein interaction calculation of the CIP program is based on PDB format files. Indeed, this format facilitates the search for atoms / residues of the protein meeting the criteria of the different interactions, the characteristics such as their positions, their chain, their coordinates, etc, always at the same position in a line. For this report, we will use the 2C35 protein which corresponds to the Rpb4 and Rpb7 subunits of human RNA polymerase II, which has the advantage for this report of being a protein having 2 chains, which allow to test the functionality of calculation of inter-protein interactions. However, other PDB files of different size, number of subunits, and secondary structures are also available on the github deposit. Finally, the program takes into account the determination of protein by X-ray crystallography, in the case of an import of a protein not present in the deposit and determined by NMR, it will be necessary to retain only one of the models in the PDB file.

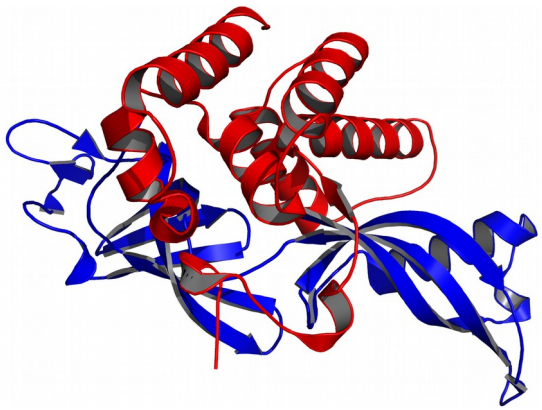


Figure 1: Visualization of the protein 2C35, Rpb4 in red and Rpb7 in blue

2 Interaction criteria

Default criteria for each interaction (specified in article [1]):

a. HYDROPHOBIC INTERACTIONS: pairs of ALA, VAL, LEU, ILE, MET, PHE, TRP, PRO, TYR within 5Å.

b. DISULFIDE BRIDGES: Pairs of cysteines (sulphur atoms) within 2.2Å.

c. HYDROGEN BONDS: distance between donor and acceptor is less than 3.50Å (oxygen and nitrogen), or 4Å (sulfur). [2]

d. INTERACTIONS IONIQUES: Ionic residue (ARG, LYS, HIS, ASP, GLU) pairs falling with 6Å.

e. AROMATIC-AROMATIC INTERACTIONS: Pairs of phenyl ring centroids that are seperated by a preferential distance of between 4.5 to 7Å. [3]

f. AROMATIC-SULFUR INTERACTIONS: Interactions between the sulphur atoms of cysteine and methionine and the aromatic rings of phenylalanine, tyrosine and tryptophan within 5.3 Å. [4]

g. CATION-PI INTERACTIONS: Cationic side chain (LYS, ARG) is within 6Å of an aromatic side chain (PHE, TRP, TYR). [5]

The CIP program also allows the user to modify the threshold distance values for hydrophobic, ionic, aromatic-aromatic, aromatic-sulfur and cation-pi interactions.

3 CIP program

The entire program and its modules were coded using python 3.7. From a PDB file and a list of arguments to inform the search on its intra or inter protein character, or on the interactions desired by the user, the program will display the interactions found in the bash and create a result file with the name of the protein in the results folder. For a more detailed explanation of its structure or for examples of program use, refer to the appendices and program README. The source code of the program, several PDB files, and the article of which it is inspired, are available on the deposit: https://github.com/Bbouvarel/CIP .

Results

We have therefore initiated the calculation of interactions of the 2C35 protein (as well as other proteins not mentioned in this report), both intraprotein and inter-protein, on our CIP program and on the PIC site in order to be able to compare the results obtained and thus verify that our program is working properly.

We thus noted two categories of interaction calculations. Indeed, for the hydrophobic, disulfide, aromatic-aromatic, aromatic-sulfur and cation-pi interactions, he two programs find identical results, allowing us to conclude the effectiveness of our CIP program for the determination of these interactions within a protein and between proteins (between chains). On the other hand, we can notice that for the ionic interactions and the hydrogen bonds, the PIC server identifies a greater number of interactions than the CIP program.

1 Differences of hydrogen bonds

With regard to hydrogen bonds, several theories can explain this difference in results. An element to be taken into account in advance is the fact that the CIP program, especially for the hydrogen bonds (main-side and side-side) shows only one interaction in the case where a hydrogen bond can be done with the two hydrogens of the same atom, unlike the PIC server. Then there is the choice of hydrogen donors and acceptors that can vary depending on the method. Unfortunately, the list of donors and acceptors used by the PIC server is not communicated to us, so we will not be able to compare it to ours (Table 1).

Residue	Donors	Acceptors
ARG	NE, NH1, NH2	-
ASN	ND2	OD1
ASP	-	OD1, OD2
CYS	SG	-
GLN	NE2	OE1
GLU	-	OE1, OE2
HIS	ND1, NE2	ND1, NE2
LYS	NZ	-
MET	-	SD
SER	OG	OG
THR	OG1	OG1
TRP	NE1	-
TYR	OH	OH

Table 1: List of donors and acceptors used for the calculation of hydrogen bonds by CIP

DONOR				ACCEPTOR							PARAMETERS	
POS	CHAIN	RES	ATOM	POS	CHAIN	RES	ATOM	MO	Dd-a	Dh-a	A(d-H-N)	A(a-O=C)
42	A	GLU	OE1	38	A	HIS	NE2	1	3.20	2.62	113.48	999.99

Figure 2: Output of PIC for the calculation of hydrogen bonds

Finally, by looking more closely at the results of PIC, we can see some strange data. Indeed, as we can see in Figure 2, the PIC server indicates that the OE1 of a glutamate is able to give a hydrogen atom whereas, as we can see in Figure 3, OE1 corresponds to the double bond O of the side chain of glutamate. We have seen this kind of choice of implementation for other residues (ASN, ASP, GLN, ...) and have decided not to implement this method in the CIP program which also explains the difference in output between the two programs .

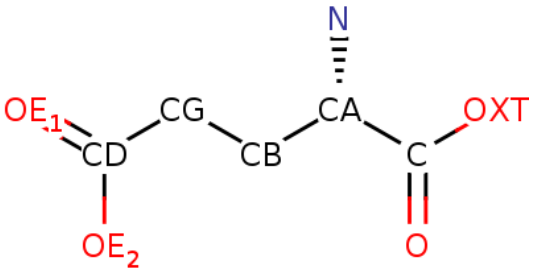


Figure 3: PDB format of glutamate residue

2 Differences in ionic interactions

For ionic interactions, after performing some tests, we realized that once again, the PIC server was doing something strange. Indeed, while we take into account, for our program, only the atom that can win or lose a hydrogen rendering positive or negative a residue, we realized that PIC considered two position by residues. To take the example of arginine (Figure 4), while we consider only NH2 (NH3 + position), PIC considers both NH2 and NH1, while NH1 is

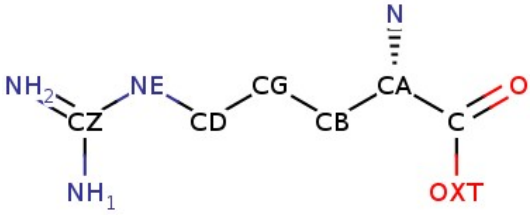


Figure 4: PDB format of arginine residue

not involved in the ionization of the residue. Again, we decided not to implement this method in CIP, which explains the differences between the two programs in this interaction.

Conclusion

By observing similar results at the level of hydrophobic, disulfide, aromatic-aromatic, aromatic-sulfur and cation-pi interactions, and by observing that the differences in ionic interactions and hydrogen bonds are explained by singular methods used by PIC that we have not reproduced, we can conclude from the effectiveness of the CIP program for the calculation of intra and inter protein interactions. A possible improvement of our program could be the implementation of new features such as depth calculation or access to the solvent.

References

[1] Tina, K. G., Bhadra, R., & Srinivasan, N. (2007). PIC: Protein Interactions Calculator. Nucleic Acids Research, 35(Web Server), W473–W476.

[2] J. Overington, et al. Proc. Roy. Soc. Biol Sci. 1990, pg.132145

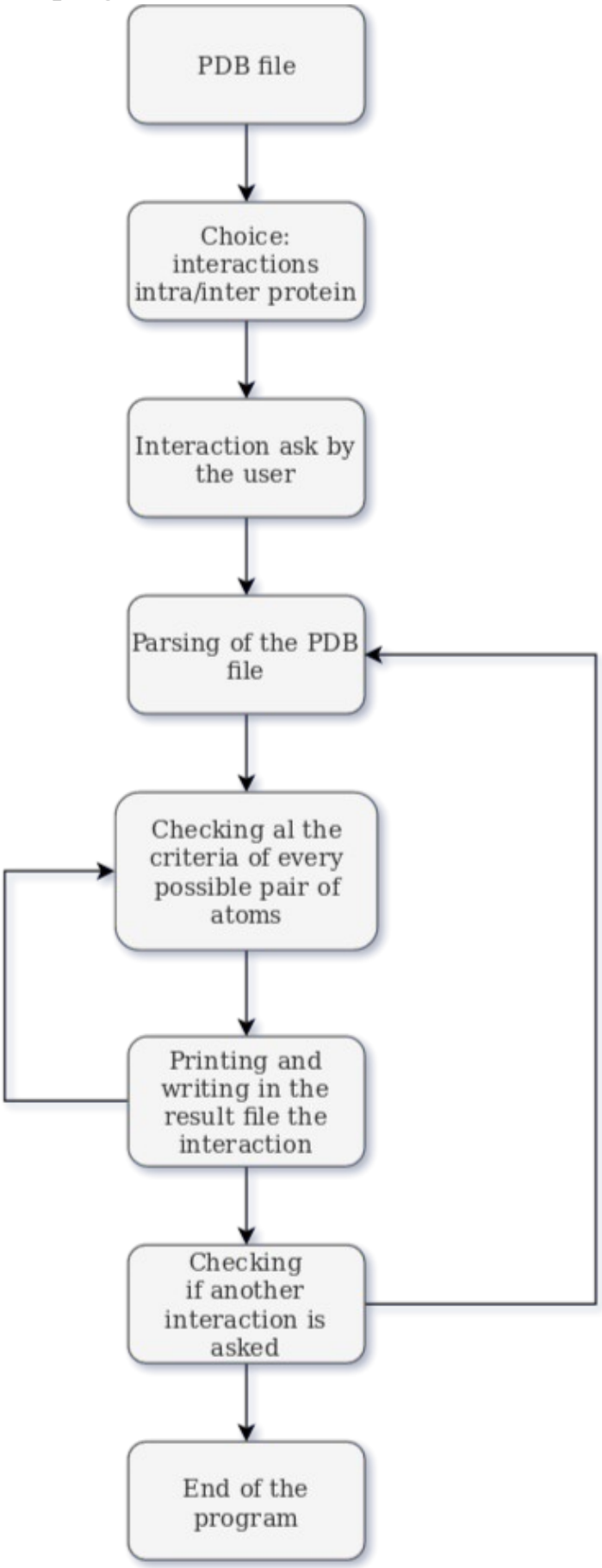
[3] S.K.Burley, G.A.Petsko, Science, 1985, Vol 299, pg.2328.

[4] K.S.C Reid, P.F.Lindley and J.M. Thornton, FEBS letter 1985, Vol.190, pg.209213.

[5] R.Satyapriya and Saraswathi Vishveshwara, Nucleic Acid Research , 2004, Vol 32, pg.4109-4118.

Appendices

1 Workflow of the program



2 Examples of use

Here are some commands to start the program and observe results easily. For this to work properly, it is necessary to clone the repository folder tree (https://github.com/Bbouvarel/CIP) and start the script from the bin directory. It is also possible to launch only certain interaction calculations, all the interaction arguments are not indispensable to launch the program.

Help of the program :
> python3 cip.py --help

Report calculations :
> python3 cip.py ../data/2c35.pdb --intra --inic --disu --arar --arsu --capi --hphb --mmhb --mshb --sshb

> python3 cip.py ../data/2c35.pdb --inter --inic --disu --arar --arsu --capi --hphb --mmhb --mshb --sshb

With modification of the default distance values :
> python3 cip.py ../data/2c35.pdb --intra --inic6.5 --disu --arar4/6.5 --arsu5.5 --capi7 --hphb4.5 --mmhb --mshb --sshb

To observe disulfide bridges (none in 2C35) :
> python3 cip.py ../data/2eti.pdb --intra --disu

More generally, we have to enter a command like this :
> python3 cip.py ../data/file.pdb --intra --arg1 --arg2 --argN

The results will be displayed in the bash and write in a file named as the PDB in the results directory.