

PyPIC: Python Protein Interaction Calculator

HAMRAOUI Ali

Master Bioinformatique : Biologie - Informatique

Université de Paris

2021 - 2022

INTRODUCTION :

Les interactions intra et inter protéines jouent un rôle crucial dans le maintien de la structure et la fonction des protéines, ainsi, elles jouent un rôle majeur dans les processus biologiques normaux et dans la régulation des réactions cellulaires qui affectent les gènes. L'analyse de ces interactions contribue à notre compréhension des bases structurales de la stabilité des protéines et de leurs évolutions. L'étude des interactions entre les chaînes latérales des résidus sont couramment utilisées dans la conception de méthodes de modélisation comparative des protéines et dans l'identification de stratégies pour les comparaisons structurales et la détection d'homologie entre les protéines (Eswar et al., 2006).

Différents types d'interactions tels que les interactions entre les résidus apolaires, les liaisons hydrogène, aromatique-aromatique, aromatique-soufre, ionique et cation- π ont été étudiés et leurs rôles et importance dans la stabilisation des structures secondaires, tertiaires et les assemblages de protéines sont devenus très bien connus (Gromiha and Selvaraj, 2004).

Plusieurs logiciels informatiques ont été développés pour analyser ces interactions, PIC (Protein Interactions Calculator: <http://pic.mbu.iisc.ernet.in>) est un serveur web particulièrement intéressant qui permet le calcul de liaisons disulfure, les interactions hydrophobes, les interactions ioniques, les liaisons hydrogènes, les interactions aromatiques-aromatiques, les interactions aromatiques-soufre et les interactions cation- π (Tina et al., 2007).

Dans ce projet nous proposons une implémentation du serveur web PIC écrite en python qui génère des sorties similaires et utilisable localement.

MATERIEL ET METHODE:

Les fichiers PDB:

Plus de 30 structures représentatives ont été téléchargées directement à partir de la base de données PDB (protein data bank). Ces structures ont été utilisées pour tester notre implémentation PyPIC et pour comparer nos résultats avec ceux obtenus avec le serveur web PIC.

Les outils informatiques :

Python 3.9.1 : est le principal langage de programmation utilisé dans ce projet.

PIC Protein Interactions Calculator version ...: le serveur web qui permet le calcul d'interactions protéiques (Tina et al., 2007).

PLIP version 1.1.0 fully automated protein–ligand interaction profiler: un serveur web permettant le calcul d'interactions protéine-protéine et protéine-ligand (Salentin et al., 2015).
Reduce version 3.16: l'outil utilisé pour ajouter les hydrogènes aux structures PDB.
BioPython version 1.77 : la bibliothèque Python utilisée pour la manipulation de fichiers PDB dans notre programme.

Identification et caractérisation des interactions:

Les critères suivants ont été fixés pour définir les différents types d'interaction selon la documentation de PIC:

Interactions hydrophobes :

Les résidus suivants sont considérés comme participant aux interactions hydrophobes s'ils se situent dans une distance de 5 Å : ALA, VAL, LEU, ILE, MET, PHE, TRP, PRO, TYR.

Ponts disulfures:

Les paires de cystéines (atomes de soufre) à moins de 2,2 Å sont considérées comme des ponts disulfure.

Liaisons hydrogènes:

Seuil de distance donneur-accepteur (oxygène et azote) = 3,5

Interactions ioniques:

Les paires de résidus ioniques (ARG, LYS, HIS, ASP, GLU) à moins de 6 Å contribuent aux interactions ioniques.

Interactions aromatiques aromatiques:

Les paires de centroïdes de cycle phényle avec une distance comprise entre 4,5 et 7 Å constituent une interactions aromatiques.

Interactions aromatiques soufres:

Les interactions entre les atomes de soufre de la cystéine et de la méthionine et les cycles aromatiques de la phénylalanine, de la tyrosine et du tryptophane dans une distance de 5,3 sont considérées comme une interaction aromatique-soufre.

Interaction Cation Pi:

Une chaîne latérale cationique (Lys ou Arg) et une chaîne latérale aromatique (Phe, Tyr ou Trp) à moins de 6 Å de distance, représentent une interaction cationique.

RÉSULTATS ET DISCUSSION :

Notre programme à été

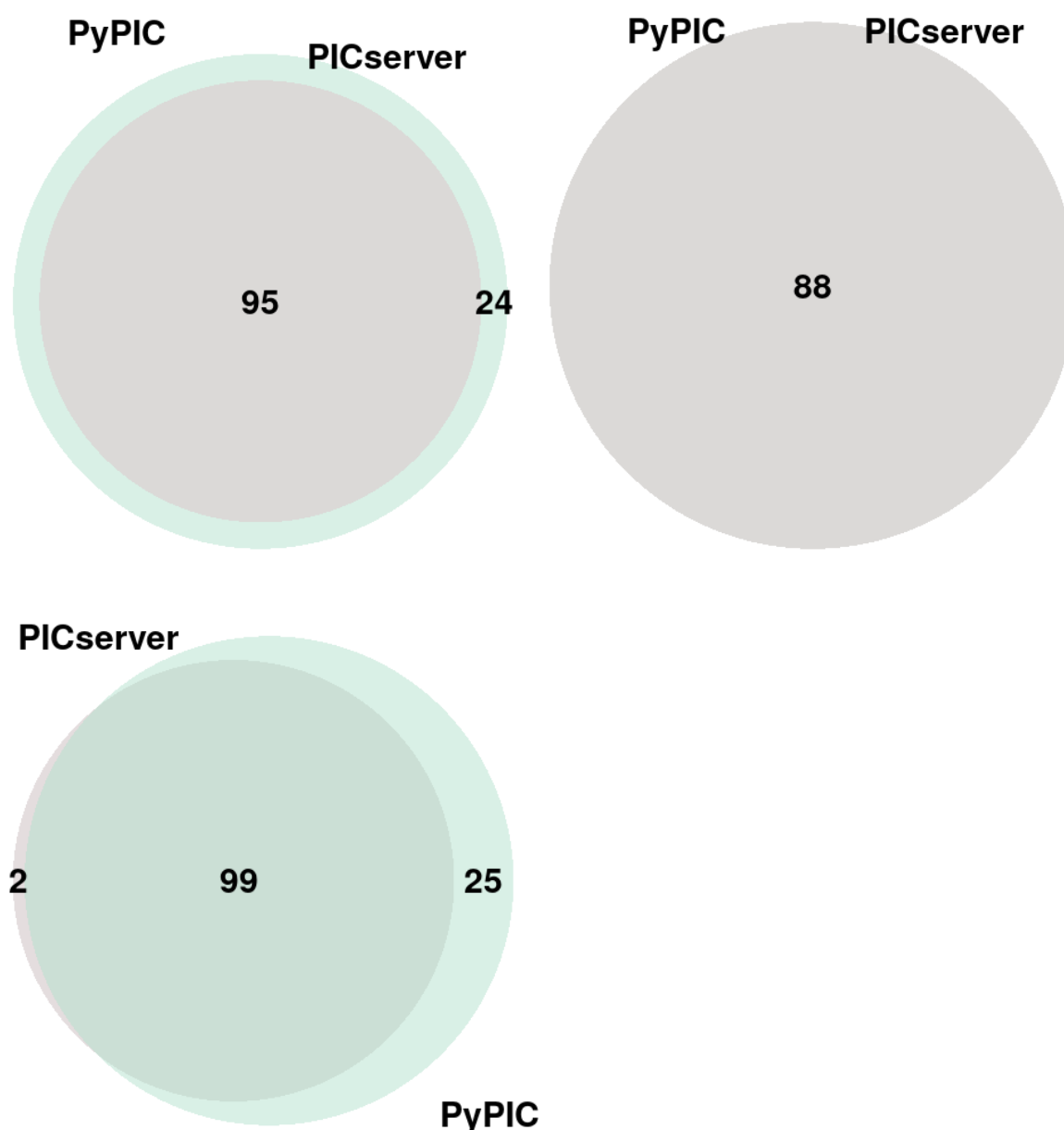


Figure 1 : Venn diagramme de comparaison des résultats obtenus avec notre programme (PyPIC) avec les sorties du serveur web PIC. La comparaison est basée sur le nombre d'interactions obtenues et la distance entre les deux atomes dans chaque interaction. En haut à droite : Liaisons hydrogènes (Main chain - Main chain); En haut à gauche : Liaisons hydrogènes (Main chain - Side chain); En bas à droite : Liaisons hydrogènes (Side chain - Side chain).

Les résultats obtenus avec notre implémentation (PyPIC) sont très corrélés avec la sortie de PIC, pour les liaisons hydrogènes entre les chaînes principales (Main chain - Main chain), les sorties obtenues sont parfaitement identiques avec les sorties de PIC. En revanche, des petites différences sont obtenues pour les liaisons hydrogènes impliquant les chaînes latérales, en effet, les distances entre l'atome d'hydrogène et l'atome accepteur d'hydrogène rapporté par

les deux programmes sont légèrement différentes, ce qui suggère que la méthode utilisée pour ajouter les hydrogènes dans la structure influence le résultat. Notre programme utilise le programme Reduce, alors que la méthode d'hydrogénation utilisée par PIC n'est pas renseignée dans la publication d'origine. Nous avons vérifié la correspondance de nos valeurs de distance et d'angle avec d'autres logiciels, en l'occurrence, le serveur PLIP v1.1.0 qui est destiné à l'analyse des interactions protéine-ligand et nous avons trouvé une parfaite corrélation. La méthode utilisée par PIC pour calculer les angles et la division dans les interactions aromatiques et Cation Pi n'ont pas été indiquées dans la publication ou la documentation du programme, ce qui impose une difficulté dans la reproductibilité des résultats.

Les résultats obtenus pour les interactions hydrophobes, ponts disulfures, aromatiques soufres et cation P sont très similaires aux sorties de PIC.

CONCLUSION :

Durant ce projet, nous avons développé une implémentation locale du serveur PIC qui génère des sorties similaires. En observons les différents résultats obtenus au niveau des interactions hydrophobes, ponts disulfures, aromatiques soufres, cation P et les liaison Hydrogènes (entres les chaînes principales) qui sont très similaire, et les interaction hydrogènes impliquant les chaînes latérale qui sont expliqué par les méthode utilisé par PIC que nous n'avons pas arrivé à reproduire, mais qui sont confirmé par PLIP, nous pouvons finalement conclure sur l'efficacité de notre programme PyPIC dans l'analyse d'interaction protéiques.

Références :

- Eswar, N., Webb, B., Marti-Renom, M.A., Madhusudhan, M.S., Eramian, D., Shen, M., Pieper, U., and Sali, A. (2006). Comparative Protein Structure Modeling Using Modeller. Curr. Protoc. Bioinforma. Ed. Board Andreas Baxevanis Al 0 5, Unit-5.6.
- Gromiha, M.M., and Selvaraj, S. (2004). Inter-residue interactions in protein folding and stability. Prog. Biophys. Mol. Biol. 86, 235–277.
- Salentin, S., Schreiber, S., Haupt, V.J., Adasme, M.F., and Schroeder, M. (2015). PLIP: fully automated protein–ligand interaction profiler. Nucleic Acids Res. 43, W443–W447.
- Tina, K.G., Bhadra, R., and Srinivasan, N. (2007). PIC: Protein Interactions Calculator. Nucleic Acids Res. 35, W473–W476.