



Master 2 Biologie Informatique

Projet Court : calcule de la surface accessible au solvant (ASA) d'une protéine

Auteur :

Taroumta Grace Djiraïbe

[https ://github.com/GraceDjiraibe/Calcul-de-la-surface-accessible-au-solvant-d-une-prot-ine-](https://github.com/GraceDjiraibe/Calcul-de-la-surface-accessible-au-solvant-d-une-prot-ine-)

Encadrants :

Jean-Christophe Gelly

Tatiana Galochkina

14 septembre 2021

1 Introduction

Dans l'étude du repliement protéique, connaître l'accessibilité est un paramètre non négligeable. L'accessibilité au solvant ou encore appelé "surface accessible au solvant" (SASA) [1] correspond à la surface de la molécule qui est en contact avec le solvant. Nous allons au cours de ce projet essayer d'implémenter l'algorithme de Shrake et Rupley [2]. De nos jours il existe des outils informatiques capable de fournir le paramètre SASA tel que NACCESS[3] ou freeSASA[4]. Nous les utiliserons plus tard afin de faire une comparaison avec les résultats expérimentales.

2 Objectif

Mon objectif dans ce projet est de calculer la surface accessible d'une protéine.

Pour se faire on va modéliser nos protéines comme étant constituées de sphères d'atomes et l'on essaiera de savoir si ces sphères peuvent ou non être en contact avec des molécules de solvant elles aussi approximées comme étant des sphères d'atomes.

On ne modélisera pas tous les points de la sphère mais une quantité N de points sur les sphères des atomes des protéines considérées. Chaque sphère possède un rayon égale au rayon de van der Waals de l'atome dont il est question.

Le projet prendra en compte la simulation du roulement d'une molécule d'eau (dont le rayon de van der Waals est de 1,4 Ångström) autour des sphères de vdW de chaque atome constituant la protéine. Tout ce projet a été codé en python (python 3.9) et a ensuite été compilé dans un environnement Linux. Les bibliothèques suivantes sont nécessaires pour le fonctionnement du programme :

- Argparse
- Matplotlib
- Numpy
- Pandas
- Scipy

2.0.1 Ressources

Les données brutes utilisées dans ce projet proviennent de la base de données Protein Data Bank (PDB), il y en a 2 fichiers .pdb qui sont :

- 1UBQ.pdb = Structure de l'ubiquitine humaine (protéine chromosomale)[5]
- 1BOQ.pdb = Melanotropine (hormone/facteur de croissance)[2]

2.0.2 FreeSASA et NACCESS

FreeSASA [4] est un outil possédant une ligne de commande de bibliothèques C/C++ qui est capable de calculer la surface accessible au solvant, avec différents paramètres d'analyses qui peuvent être personnalisés par l'utilisateur.

NACCESS [3] est un logiciel qui permet également d'obtenir la surface accessible au solvant à partir du fichier PDB.

2.1 Méthodes

Dans un premier il nous faut extraire les coordonnées de chaque atome (**fonction atom-coord()**) du fichier .pdb, l'extraction se fait sous la forme d'une data frame. A partir de cette data Frame obtenue on construit la matrice de distance de tous les atomes entre eux (**fonction dist-matrix()**).

Il faut maintenant déterminer quels sont les atomes qui sont spatialement voisins. Tous les atomes étant à une distance supérieure au diamètre d'une molécule d'eau + 2 fois le rayon de l'atome ayant le plus grand rayon de van der Waals à savoir le soufre ne sont pas considérés comme des voisins. (**fonction neighbors()**).

Une sphère à N points est ensuite construite autour de chaque atome voisins (trouvés précédemment), une fois que les sphères sont créées nous pouvons maintenant calculer la distance aux sphères voisines (**fonctions sphere() et dist-sphere-atome()**). Le point est considéré comme exposé au solvant si la distance entre un point de sa surface et le centre de l'autre sphère est plus grand que le diamètre d'une molécule d'eau + le rayon de vdw de l'atome.

Pour finir il faudra faire le calcul de la surface accessible au solvant de chaque atome puis sommer ces surfaces d'atome pour obtenir la surface accessible du résidu et enfin de la protéine (**fonction prop(), surface() et final-surface()**). Le calcul de surface accessible doit être pondéré par la proportion de points exposés.

3 Résultats & Discussion

Nous pouvons voir sur la **Figure1 et Table1** que les résultats expérimentaux obtenus pour la IUBQ ne sont pas vraiment proches des résultats obtenus avec NACCESS. La courbe des résultats naccess présente de grandes variations entre les faibles et grandes surfaces accessibles tandis que les valeurs obtenues expérimentalement sont plutôt constantes en comparaison. En revanche pour la 1bOq (**Figure2 et Table2**) il semble beaucoup plus proche des valeurs de NACCESS cela est sûrement dû à la taille car nous nous situons sur un peptide plus petit, les valeurs d'accessibilité par résidus sont également beaucoup plus proches. La différence des résultats obtenus entre NACCESS et le programme expérimental s'expliquent également par le fait que Naccess (ou fastsasa) prennent en compte beaucoup plus de paramètres. En effet ils tiennent compte de critères comme l'hydrophobicité des résidus, leur caractère polaire ou non et bien d'autres encore qui n'ont pas été implémentés dans ce projet court. Il est à noter également que lors de ce projet j'ai rencontré des problèmes d'optimisation, le programme a été testé avec plusieurs protéines de différentes tailles et plus la protéine était grande plus le temps de calcul était long et moins les résultats étaient bons. Les 2 résultats présentés reflètent les limites que j'ai rencontrées lors de l'implémentation.

Résidu	Implémentation du projet	Implémentation Naccess
MET	64,68	35,3
GLY	76,0	74,17
ILE	69,25	0
PHE	93,86	55,21
VAL	56,88	0,01
LYS	69,17	97,35
THR	53,28	21,32
LEU	71,30	136,63
THR	60,09	124,47
GLY	32,28	67,25
...
Total	5025,9	4782,2

TABLE 1 – Valeur de la surface accessible au solvant pour chaque atome de la protéine IUBQ pour les 10 premiers résidus

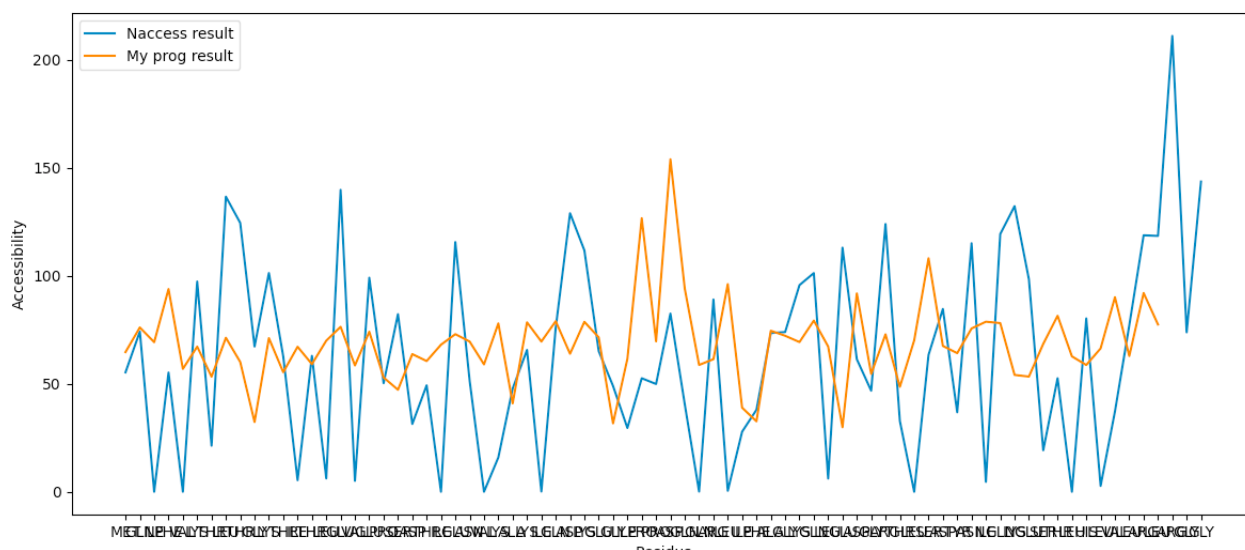


FIGURE 1 – Valeurs d'accessibilité de la IUBQ en fonction des résidus, la courbe bleu correspond aux resultats obtenus avec Naccess et la courbe orange les résultats que obtenus expermimentalement lors du projet

Résidu	Implémentation du projet	Implémentation Naccess
CYS	91,81	111,99
GLU	123,81	151,35
HIS	135,5	230,92
ARG	190,1	190,4
TRP	204,7	205,6
CYS	72,33	98,43
LYS	172,3	127,48
PRO	119,33	111,69
VAL	115,3	213,9
Total	1225,16	1442,9

TABLE 2 – Valeur de la surface accessible au solvant pour chaque atome de la protéine 1b0g

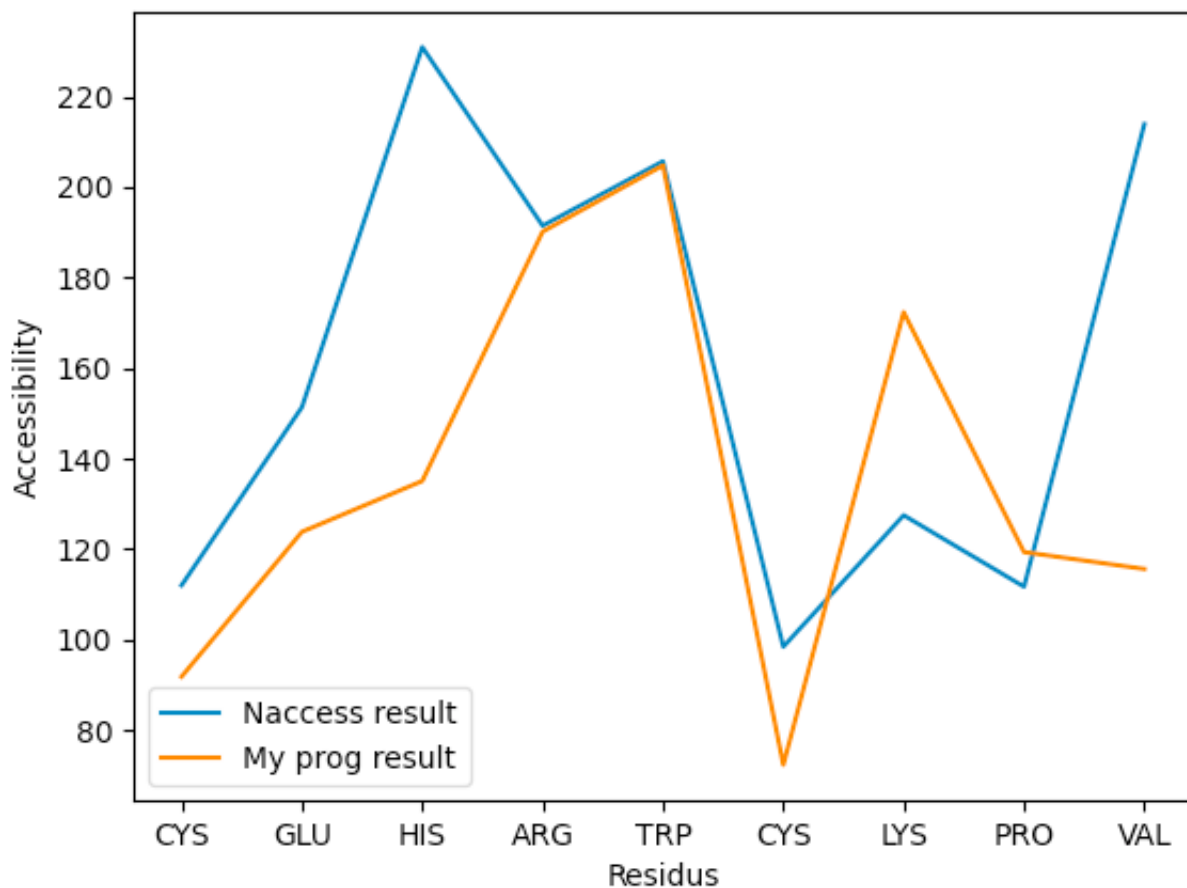


FIGURE 2 – Valeurs d'accessibilité de la 1b0q en fonction des résidus, la courbe bleu correspond aux résultats obtenus avec Naccess et la courbe orange les résultats que obtenus expérimentalement lors du projet

4 Conclusion & Perspective

En conclusion, nous pouvons dire que les résultats obtenus avec l'implémentation expérimentale sont assez différents de ceux obtenus grâce à NACCESS cependant sur la totalité de la protéine il reste tout de même relativement proche si l'on prends en compte le fait que naccess et fastsasa assimilent beaucoup plus de paramètres que le programme qui a été mis en place pour ce projet court. Pour obtenir des résultats plus satisfaisant il aurait été intéressant de se pencher sur les autres paramètres pouvant impacter sur l'accessibilité au solvant tel que caractère hydrophobe, hydrophiles etc.

Références

- [1] “Sasa.” https://en.wikipedia.org/wiki/Accessible_surface_area.
- [2] “1ubq.” <https://www.rcsb.org/structure/1UBQ>.
- [3] “naccess.” <http://www.bioinf.manchester.ac.uk/naccess/>.
- [4] “freesasa.” <https://github.com/mittinatten/freesasa>.
- [5] “1b0q.” <https://www.rcsb.org/structure/1B0Q>.