

# Projet court :

## Calcul de la surface accessible au solvant d'une protéine

Auteur : Aurélien Desvilles

M2 BI

Septembre 2023

Lien github : [https://github.com/amphioxusdu94/Projet\\_court\\_M2BI](https://github.com/amphioxusdu94/Projet_court_M2BI)

### Introduction

Les protéines sont constituées d'une ou plusieurs longues chaînes d'acides aminés (ou résidus), une chaîne pouvant se replier sur elle-même en fonction des caractéristiques (taille, polarité ou charge) des acides aminés qui la constituent. Elles sont généralement au moins en partie en contact physique avec des liquides physiologiques (sang, lymphe, LCR) principalement composés d'eau. Du fait du repliement protéique et de la disposition des atomes entre eux, la surface exposée est généralement différente de la surface. Le calcul de l'exposition au solvant est un sujet d'intérêt pharmacologique. Ce travail portera sur l'élaboration d'un script Python calculant la surface accessible d'une protéine.

### Matériel et méthodes

#### A) Algorithme Shrake-Rupley

L'algorithme Shrake-Rupley est un procédé numérique qui dessine un maillage de points équidistants de chaque atome de la molécule, et utilise le nombre de ces points qui sont accessibles au solvant, afin de déterminer l'aire de surface.[1] Les points sont attirés par une plage approximative d'une molécule eau sur au sein de van der Waals, qui est en fait similaire à rouler une balle le long de la surface. La région accessible est calculé en multipliant le nombre de points disponibles pour une partie de la zone de surface que chaque point représente. Le choix du rayon de la « sonde » affecte la surface observée, puisque l'utilisation d'un plus petit rayon de la sonde détecte plus de détails sur la surface, puis retourne une plus grande surface. Une valeur typique est de 1,52 Å, ce qui correspond approximativement au rayon d'une molécule d'eau. Un autre facteur qui influe sur les résultats est la définition de van der Waals des atomes dans la molécule étudiée. Par exemple, souvent la molécule peut ne pas avoir d'atomes d'hydrogène qui sont implicites dans la structure. Les atomes d'hydrogène peuvent être implicitement inclus dans le

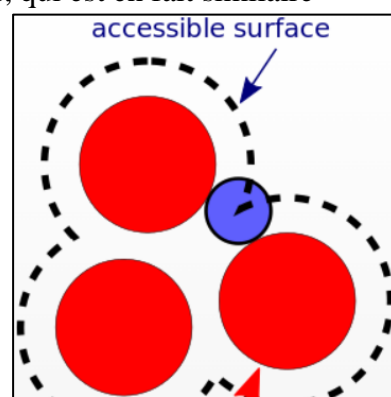


Figure 1: Représentation graphique de la surface disponible.

rayon atomique des atomes « lourdes », avec une mesure qui représente le rayon d'un groupe. De plus, le nombre de points créé sur van der Waals de surface de chaque atome détermine un autre aspect de discrétisation, où plus de points fournissent un plus grand niveau de détail. [2]

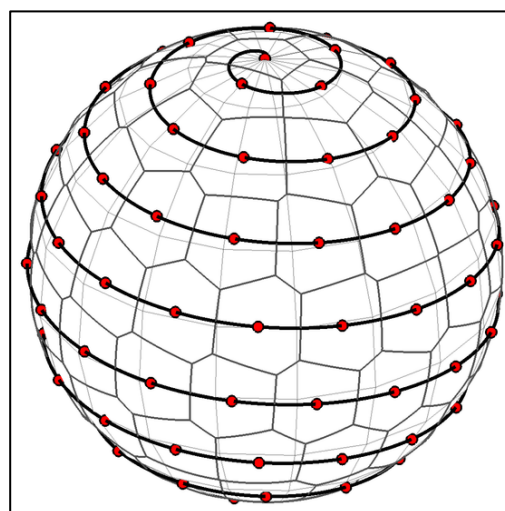
## B) Ecriture du code

Le code prend en entrée un fichier .pdb décrivant les coordonnées cartésiennes de chaque atome d'une protéine et effectue les étapes suivante :

- 1) Extraction des coordonnées de chaque atome ainsi que le numéro et la nature du résidu auquel il appartient. Stocker ces données dans un data frame.
- 2) A partir de chaque atome contenu, création d'un nuage de points uniformément sur la surface d'une sphère centrée sur l'atome. La sphère a pour valeur de rayon le rayon de Van der Waals de l'atome + rayon d'un atome d'oxygène (sonde). Stockage ces données dans le data frame.
- 3) Calcul de la surface de chaque atome grâce au rayon de Van der Waals.
- 4) A partir de chaque point de chaque sphère recherche des points pour lesquels aucun autre point d'une autre sphère n'est à une distance inférieure au rayon de la sonde.
- 5) Calcul de la fraction des points sans contact et en déduire la surface exposée relative (%) et absolue ( $\text{\AA}^2$ ) pour chaque atome.
- 6) Calcul des surfaces exposée relative et absolue de la protéine en faisant respectivement la somme des surfaces relative divisée par le nombre d'atomes et la somme des surfaces absolues.
- 7) Création d'un premier fichier contenant les surfaces absolues et relatives pour chaque atome.
- 8) Création d'un deuxième fichier contenant les surfaces absolues et relatives pour chaque acide aminé.

Les atomes d'hydrogène ne sont pas pris en compte pour la protéine et pour la molécule d'eau qui est réduite à un atome d'oxygène de 1,52 Å. Les valeurs des rayons de Van der Waals sont les mêmes que ceux utilisés dans l'article de Shrake et Rupley [1]. Les sphères générées sont des sphères dites de Fibonacci dont les points sont répartis de manière quasi uniforme.[3] Le nombre de points par sphère est fixé à 20 par défaut et ne dépasse pas 100 pour réduire le temps de calcul.

## C) Protéine



**Figure 2:** Points générés selon un algorithme de Fibonacci

La protéine testée est l'insuline humaine (code PDB : 1h59) [4], qui a l'avantage de n'être constituée que d'une seule chaîne et d'avoir une structure tridimensionnelle bien connue expérimentalement.

## D) Programme de référence

Les résultats ont été comparés avec le programme GetArea de l'Université du Texas. [5]. Ce programme aussi ne prend pas en compte les atomes d'hydrogène.

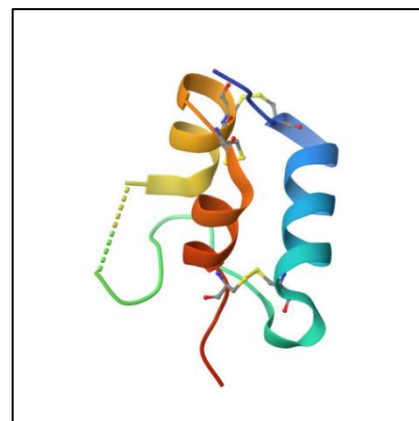


Figure 3: Insuline humaine

## Résultats

Le **Tableau 1** montre les résultats trouvés pour différents nombres de points par sphère. On y observe que la surface trouvée diminue avec la résolution (soit le nombre de points). GetArea, quand à lui, trouve une surface de 6294.63 Å<sup>2</sup> pour une sonde de même taille. Il semble y avoir une sous-estimation de la surface exposée qui diminue lorsque l'on augmente le nombre de point. Le **Tableau 2** montre que cette sous-estimation se retrouve pour chaque acide aminé.

	5 points	10 points	20 points	100 points
Surface absolue (Å <sup>2</sup> )	4359.86	4359.86	2488.18	1685.30
Surface relative (%)	17.15	17.15	9.95	6.84

**Tableau 1** : Surface trouvée pour différents nombres de points par sphère

<b>A</b>							
Residue Number	Type	Relative surface exposed (%)		Absolute surface exposed (Å <sup>2</sup> )			
2	PRO	15.0		38.13			
3	GLU	0.55		1.81			
4	THR	5.71		13.07			
5	LEU	0.0		0.0			
6	CYS	3.33		8.14			
<b>B</b>							
Residue		Total	Apolar	Backbone	Sidechain	Ratio(%)	In/Out
PRO	2	160.39	133.97	49.43	110.96	100.0	o
GLU	3	15.04	11.40	7.37	7.67	5.4	i
THR	4	48.22	14.14	6.39	41.83	39.4	
LEU	5	2.97	2.97	2.09	0.88	0.6	i
CYS	6	54.99	30.15	6.72	48.27	47.2	

**Tableau 2** : Surface exposée pour les six premier acides aminés calculé avec le programme pour 20 points par sphère (A) et avec GetArea (B).

## Discussion

Le programme a été modifié pour prendre en compte les atomes d'hydrogènes mais ces derniers ne semblent pas intervenir dans la surface exposée au solvant car étant systématiquement trop proche d'un autre atome.

Une des limites du code est qu'il ne prend pas en compte le mouvement des atomes dans l'espace qui augmenterait certainement la surface d'exposition au solvant.

## Conclusion

Le programme trouve une surface cohérente mais sous-évaluée. Cela est sans doute dû à des paramètres non pris en compte.

## Sources :

[1] Shrake, A; Rupley, JA. (1973). "Environment and exposure to solvent of protein atoms. Lysozyme and insulin". *J Mol Biol* 79 (2): 351–71. doi: 10.1016/0022-2836(73)90011-9

[2] [https://boowiki.info/art/chimie-computationnelle/aire-de-surface-accessible.html#Algorithme\\_di\\_Shrake-Rupley](https://boowiki.info/art/chimie-computationnelle/aire-de-surface-accessible.html#Algorithme_di_Shrake-Rupley)

[3] <https://extremelearning.com.au/how-to-evenly-distribute-points-on-a-sphere-more-effectively-than-the-canonical-fibonacci-lattice/>

[4] <https://doi.org/10.2210/pdb1H59/pdb>

[5] <https://curie.utmb.edu/getarea.html>