**Projets courts : Sujet n° 2 : Calcul de la surface accessible au solvant**



|  |  |
| --- | --- |
| **Trinôme** | Hocine Annaba  Hager Elharty  Sleheddine Kastalli |
| **Année universitaire** | 2020/2021 |

**Introduction**

La minimisation de la surface accessible aux solvants, appliquée à l'interaction entre les surfaces des protéines constituent une contrainte puissante pour la prédiction de la conformation de ces assemblages macromoléculaires (Chothia & Janin, 1975 ; Wodak & Janin, 1978 ; Richmond & Richards, 1978 ; Cohen et al., 1979,1982). Le site de la surface d'une molécule de soluté accessible aux solvants est la surface la plus proche des centres des molécules de solvant où le soluté et le solvant sont représentés par des sphères dures. Calculée, cette surface est définie comme la surface de van der Waals enveloppe de la molécule de soluté dilatée par le rayon de la sphère de solvant sur chaque centre d'atome soluté (Lee & Richards, 1971 ; Richards, 1977).

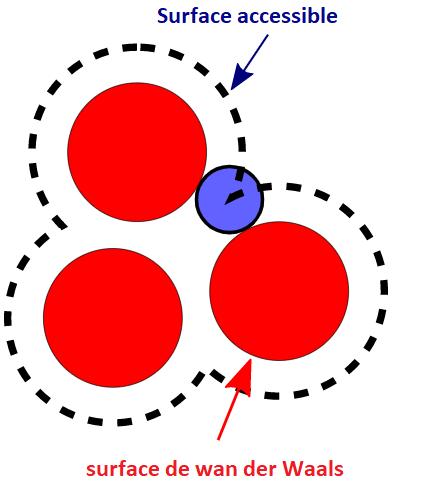
Le modèle des sphères de van der Waals est largement utilisé pour l'interprétation des structures des protéines et pour la corrélation avec les paramètres physico-chimique. L'algorithme le plus simple est probablement la méthode des incréments : Un incrément de volume est associé à chacun des n atomes, et le volume total est la somme des n contributions.

**Principe**

Rouler une molécule d’eau, représentée par une sphère, autour des atomes, également représentée par une sphère dont le rayon est celui du rayon de Wan Der Walls de chaque atome, de la protéine.

Le centre de la molécule d’eau est à une distance du centre de chaque atome égale au rayon de Wan Der Walls de la molécule d’eau avec le rayon de l’atome.

Le trajet décrit par le centre de la molécule d’eau délimite la surface accessible des atomes de la protéine.



La surface de wan der Waals tel que donné par les rayons atomique est représenté en rouge. La surface accessible est tracée en traits interrompus et est créé en suivant le centre de la sphère de la sonde (en bleu) qui roule le long de la surface de van der Waals. Notez que le rayon de sonde représenté ici est d'échelle plus petite que la 1.4a typique

**Objectif**

Calcul de la surface accessible au solvant qui est définie comme le lieu des centres d’une sphère modélisant une molécule d’eau (de rayon 1, 4 Å) lorsque celle-ci parcourt l’ensemble de la surface de la protéine.

**Matériel et Méthodes**

**Matériel :**

**Python** :

Ce programme a été condé en Python3 et compilé sous environnement Linux.

Python est un langage de programmation open source**créé par le programmeur Guido van Rossum en 1991.**

**Librairies Python :**

Pour exécuter ce programme nous avons utilisé des librairies non-standards pour Python :

-Pandas :

-Numpy :

-Argparse :

-Matplotlib :

**Données brutes :**

Les données brutes utilisées pour ce programme ont été extraites de la banque de données sur les protéines ou BDP du Research Collaboratory for Structural Bioinformatics via le site web [https://www.rcsb.org/.](https://www.rcsb.org/)

Il s’agit ici de deux fichiers PDB :

* 2om0.pdb : structure of human insulin in presence of urea at ph = 6.5.
* 7lzm.pdb : comparaison of crystal structure of bacteriophage T4 at lysozym 2 low, medium, and hight ionic strengh.

**Conda**

Conda est un puissant gestionnaire de packages et d'environnement utilisé avec les commandes de ligne de commande à l'invite Anaconda pour Windows, ou dans un terminal pour macOS ou Linux.

**Git et Github**

Git est un logiciel de gestion de versions décentralisé.

GitHub est un service web d'hébergement et de gestion de développement de logiciels.

Git et Github ont été utilisés dans ce projet pour faciliter le travail en groupe.

Tous les codes, ainsi que la documentation de ce sujet ont été déposés dans le dépôt Github ci-dessous :

<https://github.com/hocinebib/Projet_Court_H-H/tree/master>

**Méthodes :**

Le programme va commencer par lire un fichier .pdb et extraire les coordonnées de chaque atome dans un Data Frame grâce à la fonction **atom\_coor ()**.

La seconde étape consiste à générer une sphère pour chaque atome.

Ensuite, nous calculerons les distances entre les points pour 2 Data Frames, nous construirons une matrice avec toutes les distances entre les atomes et nous utiliserons la fonction **threshold\_dict ()** pour créer un dictionnaire qui a pour clé l’atome et pour valeur tous ces voisins.

**Références**

**1**. F. M. Richards, Annu. Rev. Biophys. Bioeng. 6,151 (1977).

**2**. Chothia. C. & Janin, J. (1975). Xutiuw (London). 256. 705 708

**3.** Cohen. F. E:.. Stern berg. M. J. E. & Taylor, W. R. J.Mol. Bio. 156. 821-862.