

Représenter le profil d'hydrophobicité d'une protéine à partir d'un fichier PDB

I. Introduction

Une protéine correspond à une succession d'acides aminés, dont les résidus possèdent des propriétés physico-chimiques. Ces propriétés, influencées par les interactions entre résidus et les conditions cellulaires, déterminent la conformation de la protéine, autrement dit sa structure tertiaire. Or, la structure d'une protéine lui permet d'assurer sa fonction biologique et ses interactions. Parmi les propriétés physico-chimiques influençant la structure des protéines, l'hydrophobicité joue un rôle majeur, il est donc essentiel de comprendre le profil d'hydrophobicité.

L'hydrophobicité désigne la capacité d'un acide aminé à repousser l'eau. En fonction de cette propriété, les acides aminés sont classés en deux grandes catégories : les hydrophobes et les hydrophiles. Le profil d'hydrophobicité d'une protéine est une représentation de la distribution des acides aminés hydrophobes et hydrophiles le long de sa séquence. En effet, les acides aminés hydrophobes vont créer des zones hydrophobes dans la protéines. La conformation spatiale des protéines est influencée par cette hydrophobicité car ces zones vont généralement se rassembler pour pouvoir s'adapter à l'environnement aqueux. Dans une protéine membranaire, les domaines hydrophobes sont souvent transmembranaires (ce souvent alors des hélices alpha chez les Eucaryotes). Ainsi, le profil d'hydrophobicité permet de comprendre comment une protéine se replie et interagit avec son environnement, de prédire sa structure tridimensionnelle ou, dans le cas des protéines membranaires, d'identifier des domaines transmembranaires, c'est-à-dire des segments hydrophobes qui traversent la membrane lipidique, et de comprendre les interactions protéine-membrane, caractérisées par des interactions avec les membranes cellulaires via des domaines hydrophobes.

Pour quantifier cette hydrophobicité

Pour quantifier cette hydrophobicité, Kyte et Doolittle² ont proposé en 1982 un indice d'hydropathie, un outil permettant d'évaluer l'hydrophobicité ou l'hydrophilie des chaînes latérales des acides aminés. Il s'agit d'une valeur numérique attribuée à chaque acide aminé, reflétant son affinité avec l'eau :

- Valeur positive » Acide aminé hydrophobe
- Valeur négative » Acide aminé hydrophile

Il est important de noter que, outre l'indice de Kyte et Doolittle, plusieurs autres échelles d'hydrophobicité ont été développées pour évaluer cette propriété des acides aminés.

Objectif

Notre objectif est donc de représenter le profil d'hydrophobicité d'une protéine à partir d'un fichier PDB, qui est un fichier texte qui décrit la position des atomes d'une molécule dans un espace à trois dimensions. Le format PDB¹ est structuré en lignes de texte où chaque ligne commence par un mot-clé indiquant le type d'information qu'elle contient, suivi de données spécifiques alignées selon des colonnes prédéfinies. Ces fichiers sont obtenus grâce à des techniques expérimentales. (Le fichier PDB permet de faire le lien entre les données obtenues expérimentalement et les logiciels de visualisation moléculaire notamment) .

Stratégie

Pour atteindre notre objectif, nous adopterons l'approche suivante :

1. Il faut préalablement trouver une protéine d'intérêt pertinente, c'est-à-dire une protéine dont les données structurales sont disponibles, qui est suffisamment bien documentée et bien annotée puis, pour des raisons pratiques, éventuellement une protéine de taille modérée. Dans notre contexte, les protéines membranaires sont idéales car leur interaction avec les membranes cellulaires est fortement influencée par leur hydrophobicité donc l'étude de leur profil d'hydrophobicité peut être informative.
 2. Récupérer le fichier PDB sur <https://www.rcsb.org/>
 3. Extraire la séquence d'acide aminé de la protéine
 4. Choisir une échelle d'hydrophobicité
 5. Calcul du profil d'hydrophobicité avec une moyenne glissante.
- » La **moyenne glissante** est une méthode mathématique utilisée pour lisser des données en calculant la moyenne sur une fenêtre mobile de taille fixe. Elle permet de réduire les fluctuations locales et de mieux identifier les tendances générales.
6. On détecte les zones transmembranaires
 7. On affiche graphique (usage de matplotlib) (hydrophobicité en fonction de la position) en mettant évidence les zones transmembranaires.

Exemple : Aquaporine⁷

Pourquoi cette protéine est-elle pertinente ?

- La structure d'AQP5 contient six hélices transmembranaires TM₁ à TM₆ et deux boucles hydrophobes insérées dans la membrane, la boucle E et B

- Sa structure est bien documentée et disponible en PDB.

II. Exemple

Détails sur la protéine

La protéine membranaire d'intérêt joue un rôle très important dans le transport de l'eau, à travers la membrane plasmique et donc elle va se retrouver, notamment dans les glandes salivaires, les poumons, les reins.

AQP 5 est un monomère de **269 acides aminés** sa structure se caractérise par **6 hélices transmembranaires (TM1 à TM6)** constitué majoritairement de résidus apolaire et de 2 **boucle internes B et E** qui contiennent notamment des motifs Asparagine , Proline , Alanine qui permettent de lier les molécules d'eau pour un passage sélectif et efficace de l'eau. La forme biologique native de la protéine est l'assemblage de 4 monomères identiques (chaîne A, B , C, D) chacun de ~269 acides aminés, donc dans notre contexte nous traiterons le tétramère AQP5.

Résultats

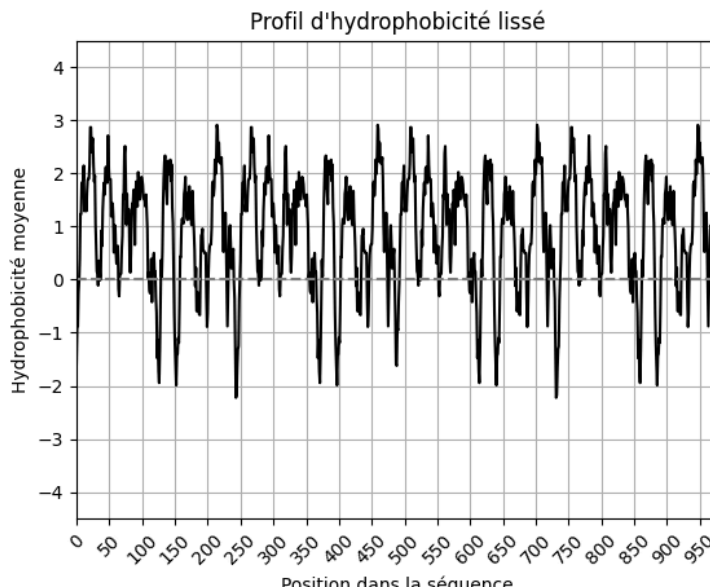


Figure 1.

Profil d'hydrophobicité concaténé d'un tétramère d'AQP5 lissée avec une fenêtre de 9.

L'axe des abscisses correspond à la position la séquence et l'axe des ordonnées à l'hydrophobicité moyenne. Les valeurs positives indiquent des segments hydrophobes ainsi que les négatives des segments hydrophiles.

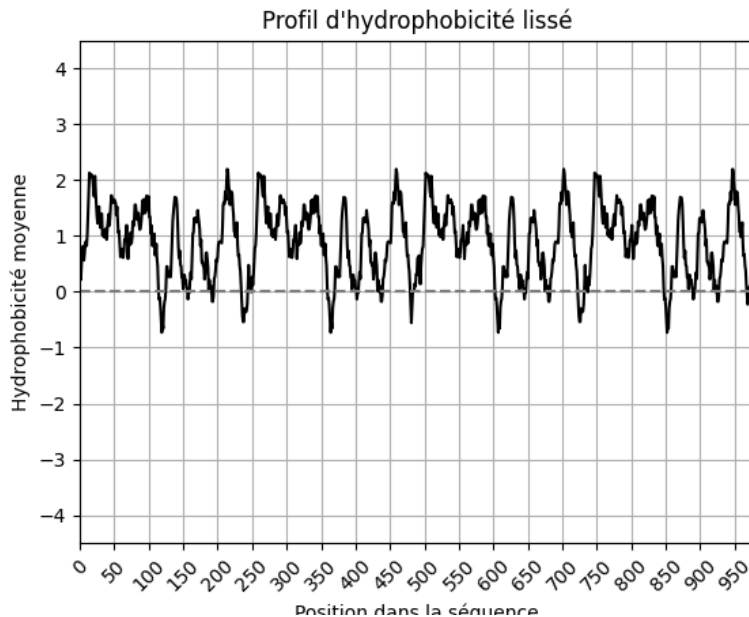


Figure 2.

Profil d'hydrophobicité concaténé d'un tétramère d'AQP5 lissée avec une fenêtre de 20 .

L'axe des abscisses correspond à la position la séquence et l'axe des ordonnées à l'hydrophobicité moyenne. Les valeurs positives indiquent des segments hydrophobes ainsi que les négatives des segments hydrophiles.

Discussion

Le profil d'hydrophobicité lissé obtenu pour la protéine d'intérêt, présente une alternance entre des pics représentant les segments hydrophobes et des creux hydrophiles.

En figure 1 et 2, nous pouvons identifier 24 pics majeurs (respectivement proches d'une hydrophobicité moyenne de 3 et 2) . Nous pouvons en déduire que ces pics représentatifs de segments hydrophobes peuvent correspondre aux 6 domaines transmembranaires (de TM1 à TM6^o) de chacune des chaînes (A,B, C, D). Ces domaines sont essentiels pour l'insertion de la protéine dans la bicouche lipidique et soutiennent l'hypothèse que la protéine étudiée est une protéine membranaire intégrale.

Sur la figure 2, on va mettre en évidence les régions hydrophiles et on peut en identifier une alternances 16 creux majeurs (en dessous de 0) qui sont des régions plus à même d'être

exposées à la surface de la protéines ou permettre le passage sélectif de l'eau, caractéristiques des aquaporines .⁸

Notre analyse basée sur l'échelle Kyte-Doolittle permet de confirmer la présence de segment transmembranaire et les portions les plus exposées à l'eau de la protéine.

III. Python

1. Lecture du fichier PDB : le module PDBParser lit la structure de la protéine AQP5 à partir du fichier 3d9s.pdb.
2. Extraction des chaînes polypeptidiques : Le module PPBuilder extrait les séquences d'acides aminés des différentes chaînes de la protéine.
3. Attribution d'un indice d'hydrophobicité : Chaque acide aminé est converti en une valeur numérique selon l'échelle de Kyte-Doolittle.
4. Calcul d'un profil lissé : Une moyenne glissante sur une fenêtre de 9 acides aminés est appliquée pour lisser les variations locales du profil d'hydrophobicité.
5. Tracé du profil : le profil est affiché à l'aide de Matplotlib, mettant en évidence les régions hydrophobes (lorsque les valeurs > 0) et hydrophiles (lorsque les valeurs < 0).

Bibliographie

1. Protein Data Bank. (s.d.). *Les fichiers PDB*. Consulté sur <https://www.rcsb.org>
2. Kyte, J., & Doolittle, R. F. (1982). *Journal of Molecular Biology*, 157(1), 105–132. [https://doi.org/10.1016/0022-2836\(82\)90515-0](https://doi.org/10.1016/0022-2836(82)90515-0)
3. Ministère de l'Éducation nationale. (s.d.). *BCPST1 SVT – Fiche technique TP SV D – Techniques d'étude des protéines*. Consulté sur <https://www.svt-tanguy-jean.com/uploads/1/2/0/4/120408978/bcpst1-tp-sv-d-proteines-techniques-complement.pdf>
4. [Jules Guittard](#) (2018). *Profil d'hydrophobicité des protéines*
5. Juke S. Lolkema , Dirk-Jan Slotboom(2005). Hydropathy profile alignment: a tool to search for structural homologues of membrane proteins. *FEMS Microbiology Reviews*, 29(4), 703–714. <https://academic.oup.com/femsre/article/22/4/305/593486>

6. Techno-Science.net. (s.d.). *Hydrophilie : définition et explications*.
7. Wikipédia. (s.d.). *Aquaporine*. <https://fr.wikipedia.org/wiki/Aquaporine>
8. ExPASy – ProtScale Tool (n.d). Retrieved April 27, 2025
<https://web.expasy.org/protscale/protscale-doc.html>