

Rapport de projet court  
Assignation et détection des parties transmembranaires d'une protéine  
Sujet n°4

## INTRODUCTION

Les protéines transmembranaires jouent un rôle important dans le fonctionnement des cellules du vivant. Bien qu'elles soient répertoriées en grand nombre dans les bases de données, leur parties transmembranaires sont peu identifiées, dû au manque de méthodes globales permettant leur identification.

Suivant le raisonnement de l'article *Transmembrane proteins in the Protein Data Bank: identification and classification* (Tusnády et al., 2004, *Bioinformatics*), nous avons cherché à établir une approche systématique définissant les zones transmembranaires de protéines, basée sur l'analyse de l'accessibilité au solvant et de l'hydrophobie de ses résidus.

Dans cette étude, nous mettons en place une telle approche que nous appliquons à une protéine transmembranaire en particulier.

### I. PREAMBULE THEORIQUE

La méthode mise en oeuvre ici se base sur l'hydrophobie des résidus protéiques et leur accessibilité au solvant.

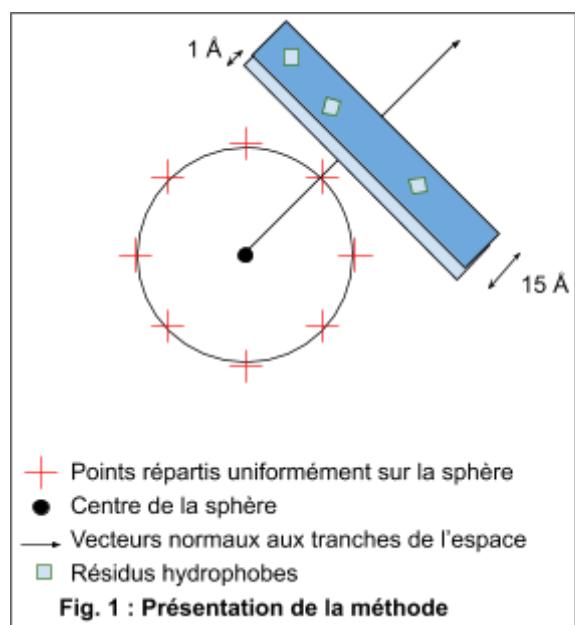
En effet, un moyen notable de déterminer de potentiels résidus transmembranaire est de regarder les résidus qui sont à la fois hydrophobes mais aussi accessibles au solvants; ces résidus sont en fait probablement en contact avec la membrane.

Ici, nous avons choisi d'étudier uniquement les résidus présentant une accessibilité relative au solvant supérieure à 30% de l'accessibilité maximale au solvant. Puis nous avons sélectionné les résidus hydrophobes par le biais d'une liste de résidus hydrophobes d'après l'article qui suggère qu'utiliser une échelle d'hydrophobie n'améliore pas les résultats.

Dans la seconde partie de l'étude, nous avons établi une sphère ayant pour centre le centre de masse de la protéine, et pour rayon 1Å. Sur cette sphère ont été placés N points répartis uniformément sur la surface, à l'aide de l'algorithme de la sphère de fibonacci qui utilise le nombre d'or.

Un vecteur a été défini entre chacun de ces points et le centre de la sphère. Puis des plans normaux à chaque vecteur ont été paramétrés. Ainsi, nous avons défini des "tranches", d'une épaisseur de 15Å, entre ces plans. Chaque tranche est obtenu en translatant la précédente d'1Å le long du vecteur normal (**Fig. 1**).

Dans chacune des tranches ainsi définies, nous avons comptabilisé les résidus hydrophobes, dans le but de déterminer les tranches ayant la plus grande hydrophobie, et étant vraisemblablement transmembranaires.



- Points répartis uniformément sur la sphère
- Centre de la sphère
- Vecteurs normaux aux tranches de l'espace
- Résidus hydrophobes

**Fig. 1 : Présentation de la méthode**

## II. MATERIEL ET METHODES

L'ensemble de la méthode ainsi élaborée a été implémentée en langage Python.

### Mise en forme des données PDB

Les données PDB sont obtenues de la base de données OPM qui est une base de données de protéines membranaires et transmembranaires. On extrait du fichier les coordonnées de chaque résidu et leur accessibilité au solvant, à l'aide des fonctions *recup\_coord\_carbones\_alpha()* et *recup\_acc\_solvant()*. Ce dernier utilise les modules *Biopython* et *dssp* pour calculer l'accessibilité au solvant.

Seuls les résidus hydrophobes sont conservés pour la suite de l'étude.

L'ensemble de ces données est organisé dans un DataFrame du module *Pandas*.

### Définition de la sphère étudiée

La moyenne des coordonnées des carbones alpha des résidus est calculée pour définir le centre de masse approximatif de la protéine avec la fonction *recup\_centre\_sphere()*. Ce centre de masse sert par la suite comme centre de la sphère présentée dans la **Fig. 1**.

La fonction *sphere\_fibo()* définit des points répartis uniformément sur la surface d'une sphère, que l'on translate ensuite pour que son centre corresponde au centre de masse de la protéine.

### Définition des plans et tranches

Pour chaque vecteur reliant le centre de la sphère et un point de sa surface, on définit des plans normaux. Les plans sont paramétrés par l'équation  $ax + by + cz + d = 0$ , avec  $x, y$  et  $z$  les coordonnées du vecteur normal au plan.

Cette équation permet de définir chaque plan par rapport au précédent, et de paramétriser des tranches de 15Å dans lesquelles nous pouvons chercher les résidus hydrophobes. Chaque plan est établi comme normal au vecteur reliant le centre et un point de la sphère, et passant par un point qui est l'extrémité d'un nouveau vecteur. En appliquant un coefficient  $k$  au vecteur d'origine, on obtient le nouveau vecteur et on définir ainsi le nouveau plan. Ces étapes utilisent les fonctions *EquationPlan()*, *DeterminerPlanSuivant()*, *PremierPlanProchaineTranche()*.

### Détermination des résidus inclus dans les tranches

Pour chaque tranche, on parcourt les coordonnées des résidus préalablement sélectionnés. On conserve pour chaque vecteur la/les tranche(s) ayant le plus grand nombre de résidus, et on stocke dans une liste les résidus figurant dans cette/ces tranche(s), à l'aide des fonctions *EstEntre2Plans()*, *ListeTranches()* et *ListeResiduesDansTranche()*.

On peut alors représenter dans l'espace les carbones alpha de la protéine en colorant en rouge les résidus obtenus en rassemblant les listes correspondant à chaque vecteur. Ces derniers dessinent les plans transmembranaires de la protéine.

## III. RESULTATS ET DISCUSSION

Cette étude réalisée principalement pour l'étude d'une protéine a été vérifiée comme étant applicable à d'autres exemples.

Nous observons ici que le programme mis en place fait apparaître plusieurs alignements de résidus transmembranaires, donnant un premier aperçu des zones transmembranaires de la protéine. L'étude pourra être étendue à de plus grands jeux de données, et les paramètres de l'épaisseur des tranches et de la répartition de la sphère peuvent être modifiés pour améliorer la précision du modèle.

## ANNEXES

le programme projet.py est notre programme principal il utilise le module argparse pour les argument en ligne de commande.

Il y a un argument obligatoire qui est le chemin d'accès au fichier pdb cible (-f) et 3 arguments facultatifs que sont :

le seuil d'accessibilité au solvant (-s) par défaut 0.3

le nombre de points de la sphère (-p) par défaut 50

la largeur de l'espace transmembranaire (-e) par défaut 15

Typiquement on lance le programme comme ceci:

`python3 projet_objet.py -f ..../data/1py6.pdb`

Ou bien si on veut changer les paramètres:

`python3 projet_objet.py -f ..../data/1py6.pdb -s 0.5 -p 100 -e 10`

On a deux fichiers de sorties un fichier.csv qui contient les résidus hydrophobes, accessibles au solvant qui ont permis de trouver notre membrane et un fichier.png qui montre ces résidus dans un espace 3d.