

# Etude et modélisation structurale de la protéine A6NFX1

## Introduction

Le but de cette étude est de réaliser un modèle de la protéine A6NFX1 grâce aux informations existantes dans les bases de données de protéine, aux outils en ligne de prédictions structurales et à la littérature scientifique. La méthode Blastp a été utilisée pour l'alignement de la séquence de A6NFX1, les outils en ligne Phyr2 et RaptorX ont été utilisés pour établir les modèles structuraux, ainsi que Mobyle HCA et PSIPRED pour la prédition des parties hydrophobes et de la structure secondaire. Enfin les modes normaux du modèle généré ont été calculés par le logiciel en ligne NOMAD-REF et les modèles ont été évalués par Verify3D.

La protéine A6NFX1 étant transmembranaire, son voisinage est très spécifique, ce qui rend difficile l'approche expérimentale [1]. Le manque de structure 3D disponible rend les outils automatiques de modélisation par homologie insuffisants, et c'est ce qui explique l'utilisation d'autant d'outils différents pour cette étude.

## Matériel et méthodes

### Matériel

Les informations sur la protéine A6NFX1 proviennent des bases de données Uniprot, pfam et de la littérature recensée sur PubMed.

A6NFX1 est une protéine transmembranaire présente chez l'humain, d'une taille de 497 paires de bases. C'est une protéine multi-pass, donc qui traverse la membrane plusieurs fois, ce qui induit la présence de plusieurs hélices alpha. Elle appartient à la super-famille intitulée : « Major facilitator Superfamily domain containing 2B »(MSFD2B) . Sa fonction est logiquement le transport membranaire.

SUPERFAMILY(supfam) associe directement à cette super-famille les processus GO suivants : [transmembrane transport](#), [anion transport](#), [response to drug](#), [carbohydrate derivative transport](#), [sodium ion transport](#), [response to ketone](#), [transporter activity](#), [anion transmembrane transporter activity](#).

La littérature scientifique précise que MSFD2B est une superfamille de transporteur orphelin impliqué dans le transport actif de la Sphingosine-1-phosphate(S1P), qui est un signal lipidique sécrété par les globules rouges et les plaquettes. MSFD2B partage avec une autre super-famille MSFD2A qui lui est très proche la particularité d'avoir 12 domaines transmembranaires en hélice alpha. Cette autre super-famille est aussi impliquée dans le transport de lipide et donc à leur métabolisme [2].

## Méthodes

### Alignement Blastp

Avec une base données de protéine (SwissProt) à disposition, la meilleure option pour chercher des protéines similaires est la recherche par homologie de la séquence. Le test d'homologie de séquences repose sur l'alignement de ces mêmes séquences et sur le score d'alignement. Un BLASTp permet l'alignement local (qui est idéal pour chercher des motifs communs) de la séquence avec chaque séquence protéique de la base de données ([www.ebi.ac.uk/Tools/sss.ncbiblast/](http://www.ebi.ac.uk/Tools/sss.ncbiblast/)). On estime qu'il y a homologie si l'on obtient plus de 30 % d'identité. L'hypothèse formulée ici est que les similitudes entre séquences induisent des fonctionnalités et des structures similaires.

### Threading par Phyr2

Si un modèle par homologie est impossible, il faut considérer un threading. Au lieu de se baser sur des séquences homologues, il utilise des calculs statistiques appliqués à la séquence donnée en entrée et les pdb déjà existants. Phyr2 est un outil en ligne utilisant le threading. La séquence et le template associé sont soumis à un PSI-Blast avant d'être formulés chacun de leur côté par une chaîne de Markov caché. Le template qui s'aligne le mieux avec la séquence sera le modèle final. (<http://www.sbg.bio.ic.ac.uk/phyre2/html/page.cgi?id=index>)

### Prédiction par hydrophobité avec Mobyle-HCA

Mobyle-HCA est un outil en ligne utilisant un graphique bidimensionnel appelé HCA (Hydrophobic Cluster Analysis). Il sépare les acides aminés hydrophobes de la séquence en clusters, et prédit statistiquement les structures secondaires selon les propriétés biologiques connues de ces dernières. Cette méthode sera utile pour notre étude pour vérifier l'information structurale donnée par notre modèle. (<http://mobyle.rpbs.univ-paris-diderot.fr/cgi-bin/portal.py#forms::HCA>)

### Prédiction des structures secondaires avec PSIPRED

PSIPRED est un algorithme de prédition utilisant une méthode de machine learning. Il utilise Psi-Blast pour trouver des séquences ressemblantes et un réseau neuronal artificiel (ensemble d'algorithmes chaînant leurs informations à la manière d'un réseau de neurone) pour traiter une matrice de score. Cette méthode donnera des résultats de prédictions de structure secondaires qui seront recoupés avec les autres. (<http://bioinf.cs.ucl.ac.uk/psipred/>)

### Modèle RaptorX

RaptorX est un outil en ligne de prédition structurale n'utilisant pas de template. Il utilise un modèle de deep learning DeepCNF (Deep Convolutional Neural Fields) pour prédire les

structures secondaires, l'accès au solvant et les régions désordonnées [3]. Les résultats consisteront en un modèle qui sera comparé avec celui de PHYR2 et aux différentes prédictions de structures secondaires. (<http://raptorx.uchicago.edu/>)

### Positionnement de la membrane avec PPM Server

PPM server est un calculateur de position de la membrane par rapport à une protéine transmembranaire qui utilise les coordonnées pdb. Il calcule des vecteurs rotationnels et translationnels pour le positionnement. Ces résultats permettront d'observer le positionnement de nos modèles dans une membrane. (<http://opm.phar.umich.edu/server.php>)

### Vérification de modèle avec Verify3D

Verify3D est un outil en ligne d'évaluation de la qualité d'un modèle. Il caractérise l'environnement de chaque position d'acide aminé (exposé, non-exposé, polaire, apolaire) pour construire un profil dont il va ensuite comparer la pertinence avec des structures 3D raffinées. Cette comparaison sera présenté sous la forme d'un score, qui permettra de voir les régions localement « bien » ou « mal » modélisées. Il servira à comparer nos modèles et à identifier les régions plus difficilement modélisables. ([http://services.mbi.ucla.edu/Verify\\_3D/](http://services.mbi.ucla.edu/Verify_3D/))

### Génération des modes normaux avec NOMAD-REF

NOMAD-Ref est un outil en ligne calculant les modes normaux d'un fichier pdb. Il utilise le modèle de réseau élastique. 60 modes ont été calculés avec les paramètres standards du logiciel. Les résultats sont dans notre étude une manière de voir les ouvertures du transporteur au ligand. (<http://lorentz.immstr.pasteur.fr/nma/submission.php>)

## Résultats

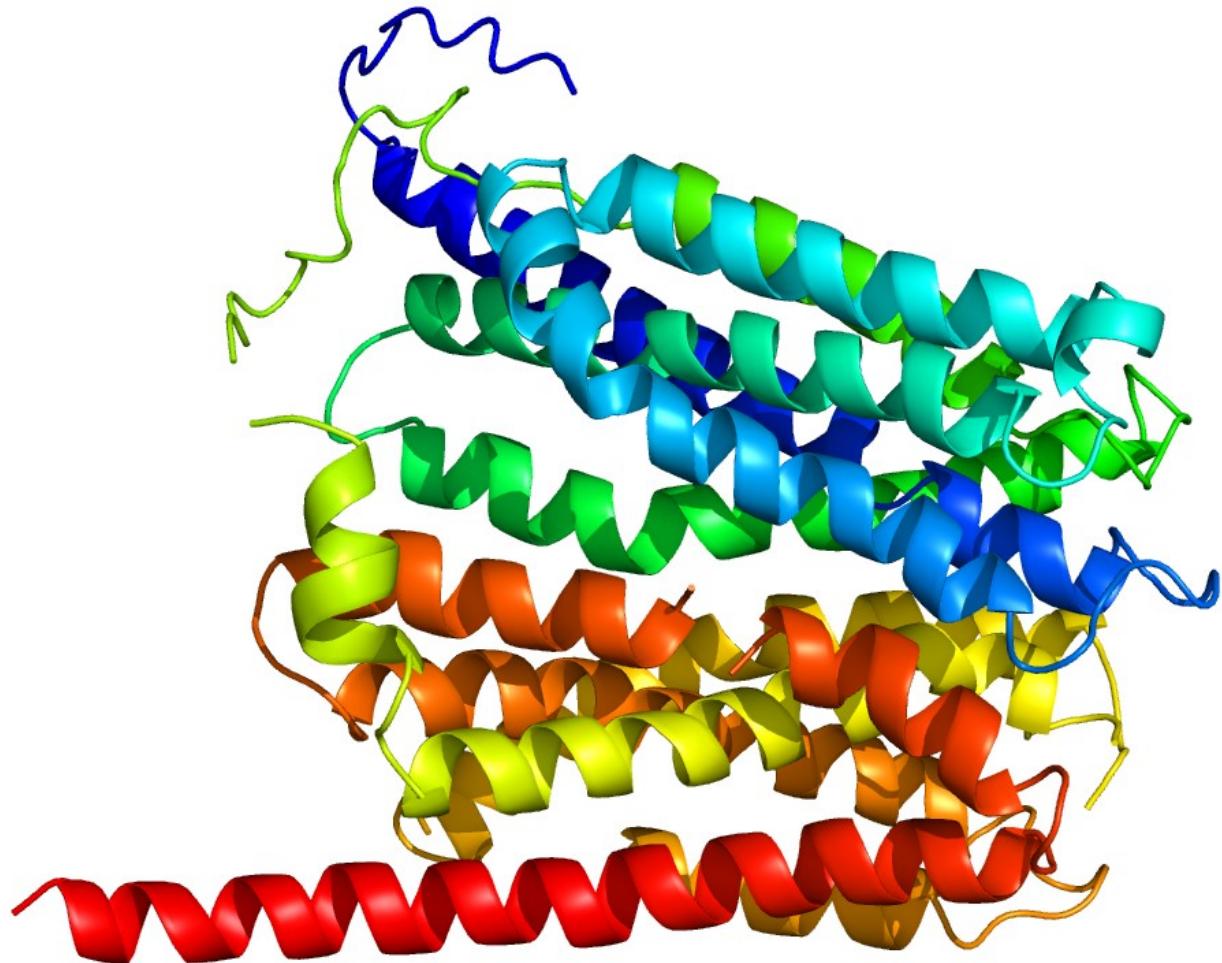
### Alignement Blastp

Les résultats ont été filtrés pour n'afficher que les protéines dont la structure 3D est connue. Le meilleur résultat présentait 20 % d'identité avec la séquence de A6NFX1, ce qui n'est estimé pas suffisant pour valider l'hypothèse d'homologie. Il est donc exclu de faire de la modélisation par homologie et l'approche initiale sera donc le threading.

Blastp nous montre tout de même que la signature de la super-famille MFSD2B correspond aux acides aminés des paires de bases 37 à 475.

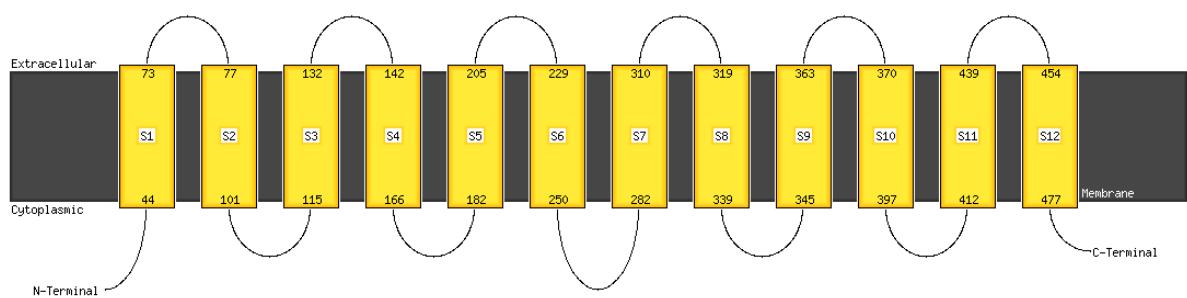
### Threading par PHYR2

Les résultats des figure 1 et 2 montent 12 domaines transmembranaires, ce qui correspond à la littérature. On remarque aussi que la signature de la super-famille (positions 37 à 475) correspond au début du premier domaine jusqu'à la fin du dernier domaine prédis dans le modèle. Les hélices alpha observées sont ordonnées et parallèles entre elles, comme on peut l'attendre d'une protéine transmembranaire multi-pass. La figure 3 ajoute que la majorité des hélices sont générées avec un haut niveau de confiance. On remarque par exemple(Figure 1) une anomalie prenant la forme d'une petite hélice alpha perpendiculaire à son propre domaine. Il pourrait en réalité s'agir d'une boucle.

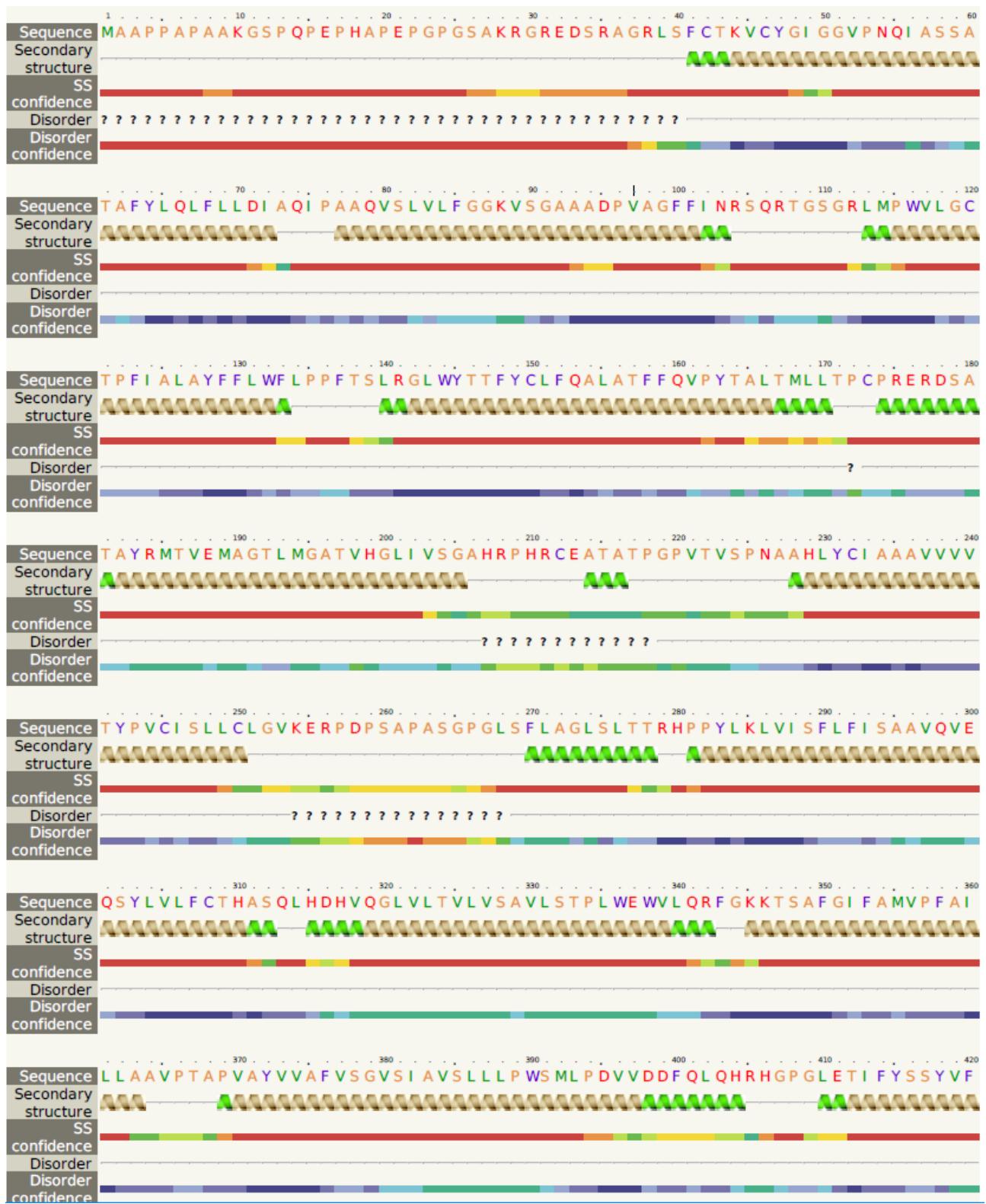


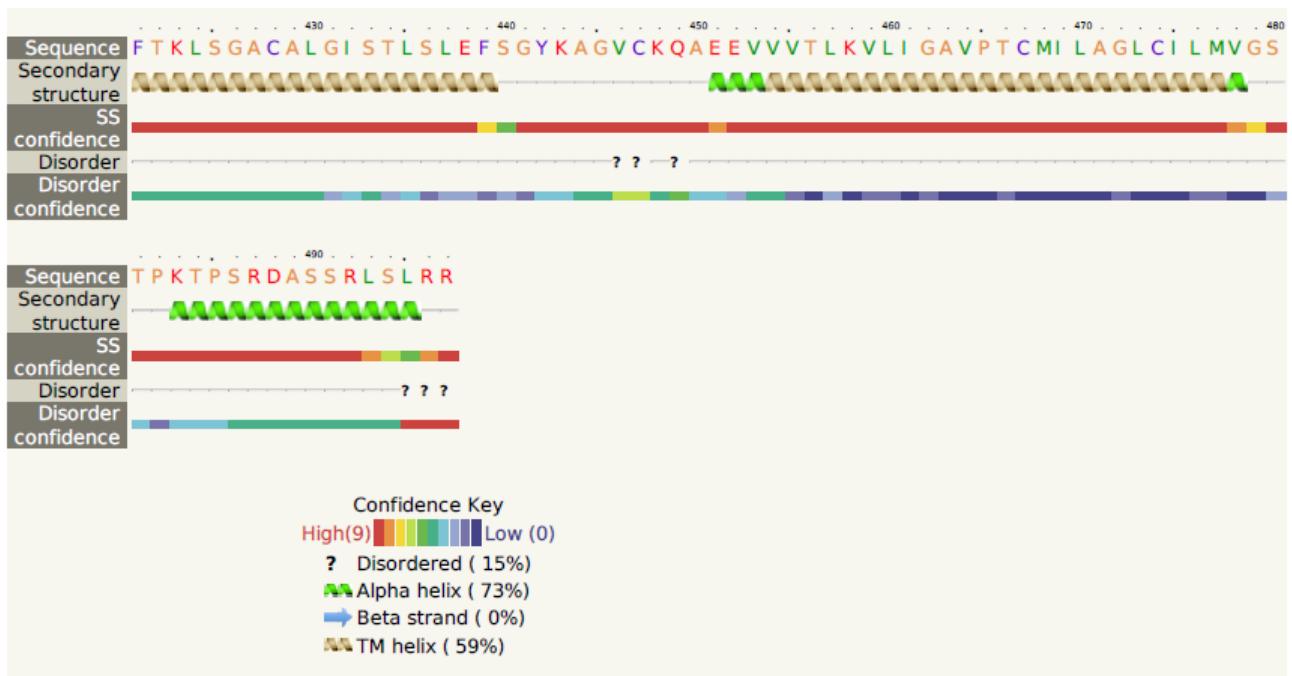
**Figure 1 :** modèle final généré par PHYR2, utilisant les meilleurs template.

query



**Figure 2 :** schéma des domaines transmembranaires du modèle. Les domaines transmembranaires sont positionnés au acides aminés : 44-73, 77-101, 115-132, 142-166, 182-205, 229-250, 282-310, 319-339, 345-363, 370-397, 412-439, 454-477.

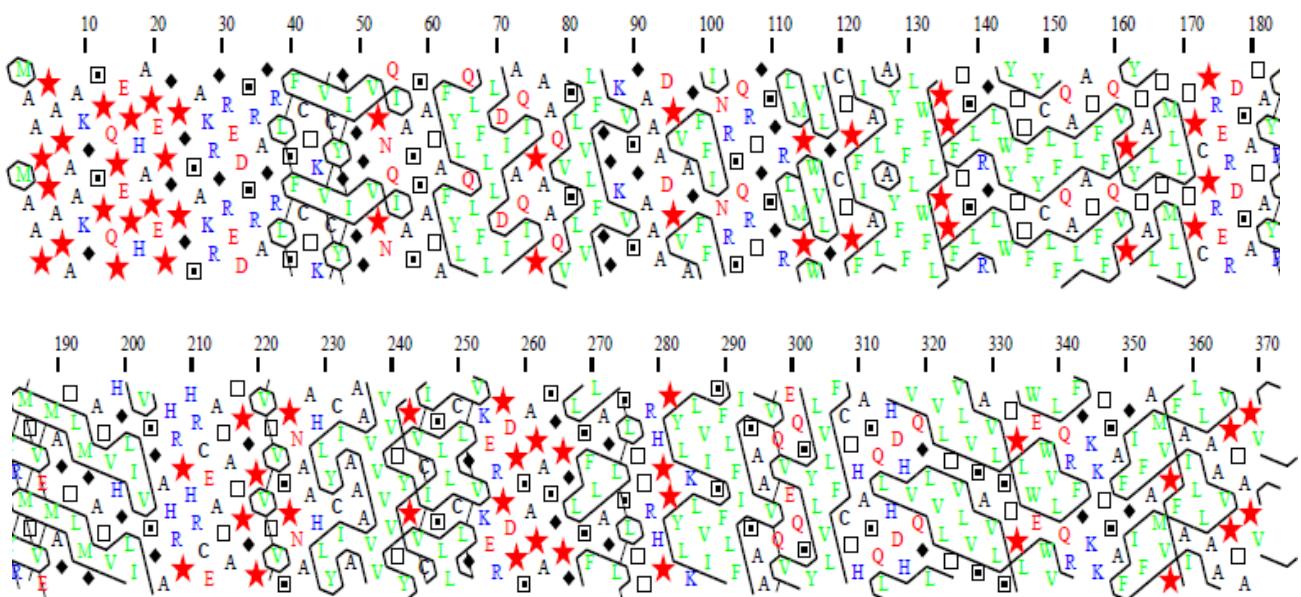


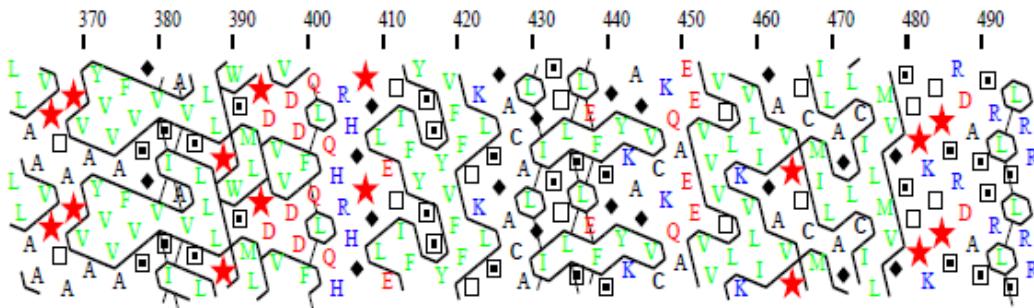


**Figure 3 :** Prédiction des structures secondaires de PHYR2 avec le taux de confiance et de désordre associé.

#### Prédiction de l'hydrophobicité avec Mobyl-HCA

La Figure 4 présente les résultats des prédictions des structures secondaire par hydrophobicité fait par une HCA. On remarque que les hélices alpha sont placés à : 140-160, 180-205, 230-255, 280-300, 350-390, 410-420, 430-440. Cela correspond en partie aux résultats de PHYR2 (7 hélices sur les 12), mais la HCA à tendance à ne pas repérer les hélices des 100 premières paires de bases. On note l'accumulation des prolines en N-terminal, ce qui laisse à penser une possible activité catalytique à cet endroit, mais il n'y a pas d'informations relatives à cette observation.





**Figure 4 :** Prédiction des structures secondaires par HCA. Les acides aminés verts sont les hydrophobes, les bleus les chargés – et les rouges les chargés +. Les formes horizontales des acides aminés hydrophobes représentent les hélices alpha et les formes verticales les feuillets bêta. Les étoiles sont le prolines, les losanges noirs les glycines, les carrés et les carrés pointillés sont les thréonines et les sérines.

Finalement cette méthode ne confirme que 7 hélices transmembranaires, correspondant aux domaines 4 à 11 du modèle PHYR2.

#### Prédiction des structures secondaires avec PsiPred

La Figure 5 montre les résultats de la prédiction des structures secondaires par PsiPred. On remarque que les hélices alpha ont été prédits avec un haut niveau de confiance. Les hélices prédites sont situés en : 44-72, 77-93, 123-131, 140-158, 175-204, 225-250, **270-280**, 282-301, **315-363**, 368-385, 410-440, 450-478, **485-495**. A part les valeurs mise en évidence en rouge(petites hélices et fusion de deux domaines) les 12 domaines transmembranaires de PHYR2 sont reconnaissables. Cette méthode semble plus sensible dans le repérage d'hélice alpha (plus de vrais positifs et de faux positifs par rapport au modèle) mais a aussi fourni des résultats plus proches que la HCA. Il est possible que cela vienne du fait que PHYR2 et PsiPred utilisent tous deux un psi-Blast.

#### Modèle RaptorX

La Figure 6 présente le modèle fait avec Raptor X. Un problème d'image de l'envoi des résultats de Raptor X empêche l'affichage d'une figure, mais les informations sur les bases individuelles sont bien arrivées avec l'image de la figure 6 et le fichier pdb. Les hélices alpha sont placés sur les positions 4-7, 77-10, 115-132, 142-168, 179-203, 230-250, 283-310, 316-341, 345-363, 369-392, 410-439, 451-477 et **489-494**. A part le dernier résultat qui n'est pas concordant, on a ici les mêmes domaines transmembranaires que ceux de PHYR2 (avec un maximum de 4 paires de bases de différence seulement).

Il est intéressant de constater que PHYR2 et RaptorX, deux méthodes différentes de prédictions de modèles, ont donné des résultats similaires. Les structures secondaires correspondent aussi bien à ceux de Psi-Pred renforçant l'idée de la justesse des modèles.

La différence de résultat de HCA laisse à penser que les parties latérales donc proches de N-terminal et C-terminal sont les moins bien prédits.



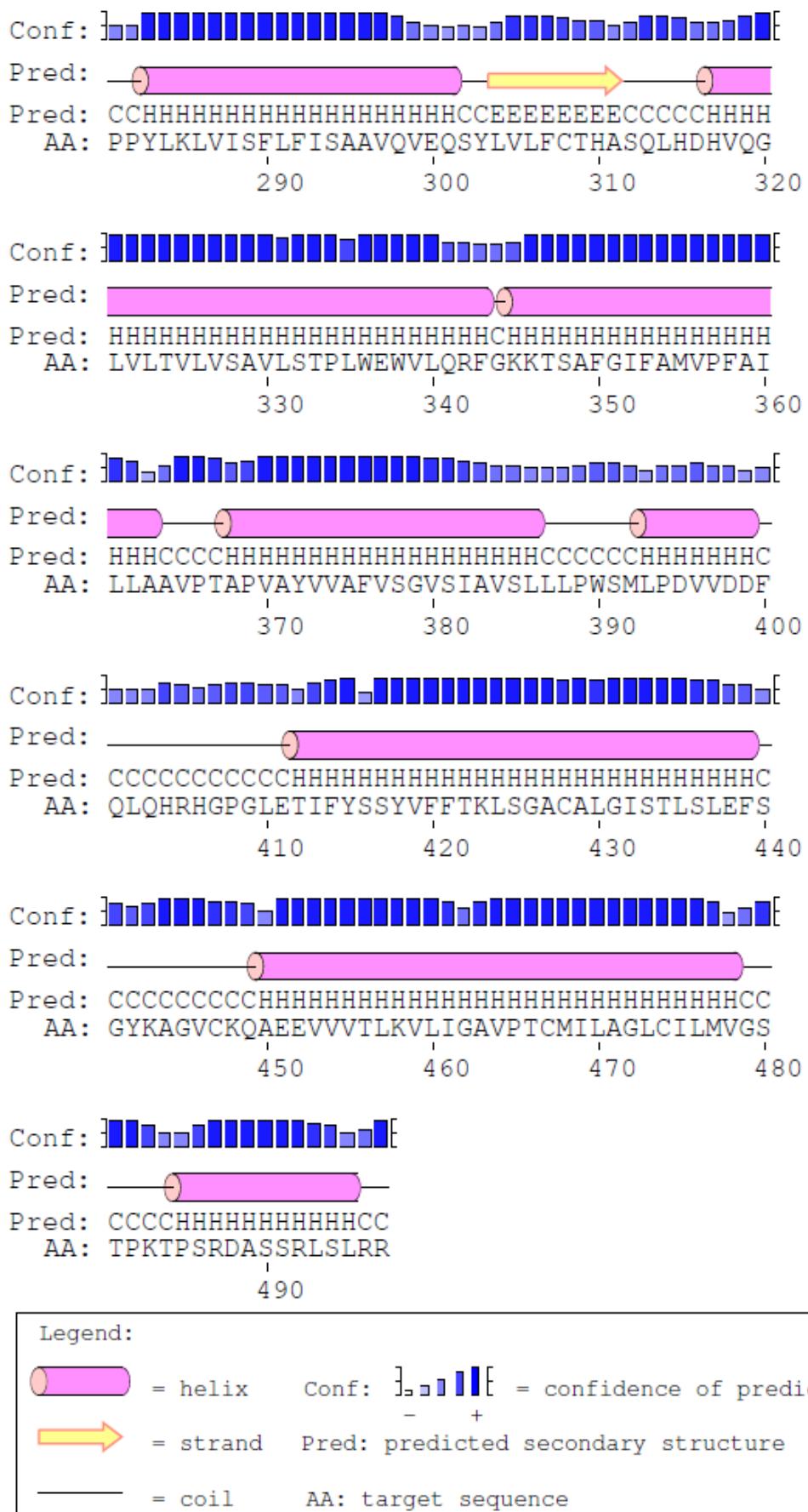
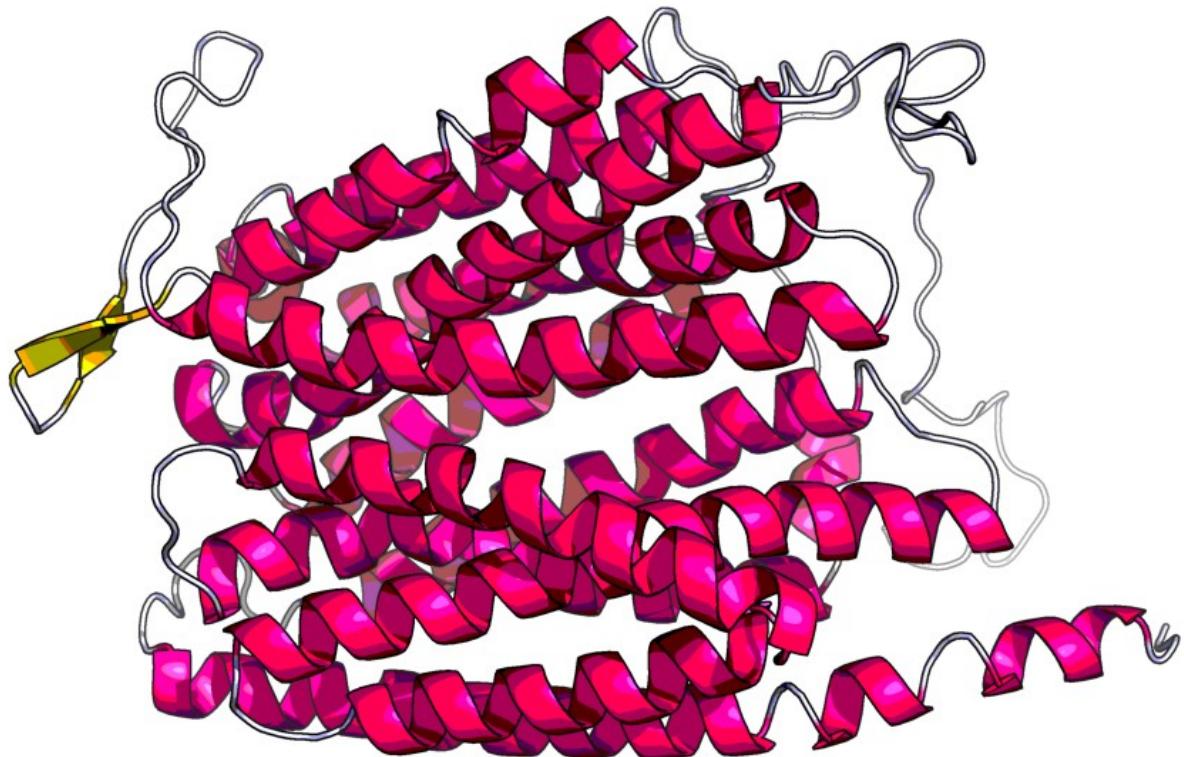


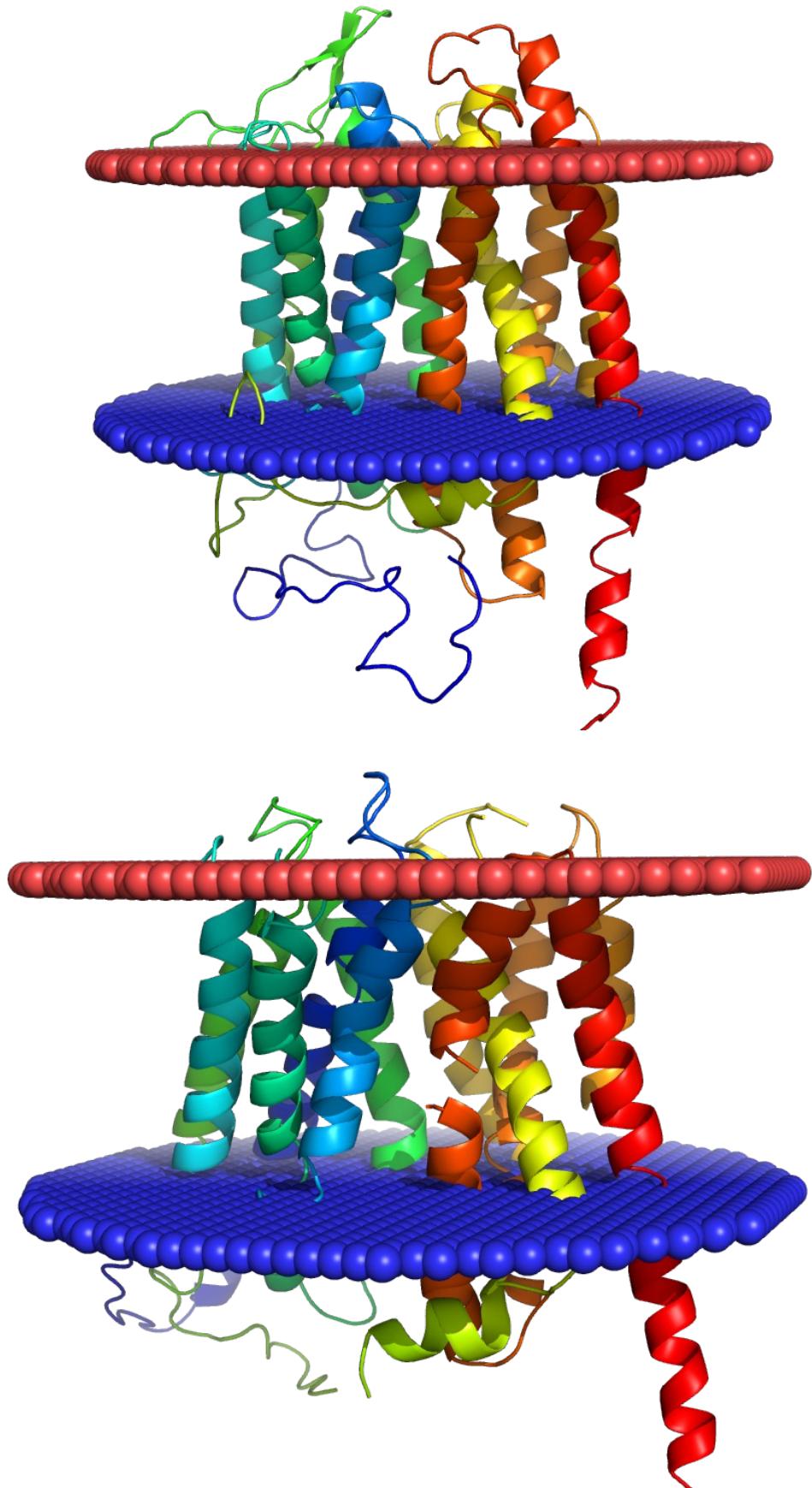
Figure 5 : Prédiction des structures secondaires selon Psi-Pred



**Figure 6:** Modèle produit par RaptorX. On note la ressemblance avec celui de PHYR2, ainsi que les 12 hélices transmembranaires. La présence des petites hélices en C-ter est questionable, mais Psi-pred les avaient aussi prédits à cette position.

#### Positionnement de la membrane avec PPM Server

La figure 7 montrent les modèles générés par PHYR2 et RaptorX dans une membrane. On remarque les 12 hélices hydrophobes dans la membrane et les boucles hydrophiles à l'extérieur entre les domaines. Les hélices alpha ont tendance à dépasser de la membrane, surtout le dernier domaine, dans les deux cas. Cela est du soit à un manque de précision des modèles soit au calcul de positionnement des membranes par PPM Server. Les modèles ont des caractéristiques équivalentes.



**Figure 7 :** Modèles des protéines transmembranaires dans une membrane calculée par PPM server(RaptorX, PHYR2) (bas = Cterm)

## Vérification de modèle avec Verify-3D

Résultats PHYR2 : 77,49 % des résidus ont un score moyen 3D-1D  $\geq 0.2$   
les points aberrants sont les positions 310 à 320.

Résultats RaptorX : 72,23 % des résultats ont un score moyen 3D-1D  $\geq 0.2$   
les points aberrants sont les positions 30 à 50, 20 à 220, 330 à 370.

Sachant qu'en dessous de 65 % il faut envisager une remise en question du modèle, les modèles vérifiés sont corrects en moyenne. Le modèle PHYR2 présente moins de points aberrants, donc il sera préféré au modèle RaptorX.

## Vérification de modèle par sources bibliographiques

La littérature scientifique sur la super-famille MFSD2B donne une piste indirecte pour le modellling de celle-ci. Il s'avère que cette super-famille est homologue par les segments introns et exons des gène qui codent pour ces protéines, ce qui ne peut être repéré par un blastp, à un transporteur bactérien. Il s'agit du bacterial Na(+)/melibiose symporters et son pdb est connu [4]. Il semblerait que son fonctionnement soit semblable aux transporteurs MSF et est responsable du transport du sucre melibiose à travers la membrane.

Une superposition des structures secondaires du modèle PHYR2 et de ce transporteur bactérien a été effectuée. L'image Figure 8 présente les résultats. Le RMS-fit calculé avec PyMOL est de 7.688 ce qui est élevé pour parler d'une homologie structurale. Mais on remarque que les domaines transmembranaires sont organisés de la même façon (les couleurs correspondent à la position de la structure dans la séquence, on peut donc identifier les domaines transmembranaires correspondants). Il s'agit d'un indice indiquant que le modèle PHYR2 a possiblement prédit une organisation structurale fonctionnelle, puisqu'elle correspond à une sensé être similaire trouvée dans un fichier pdb trouvé dans la base de données.

Ainsi on souhaite observer les modes normaux non seulement de PHYR2 mais aussi du transporteur bactérien.

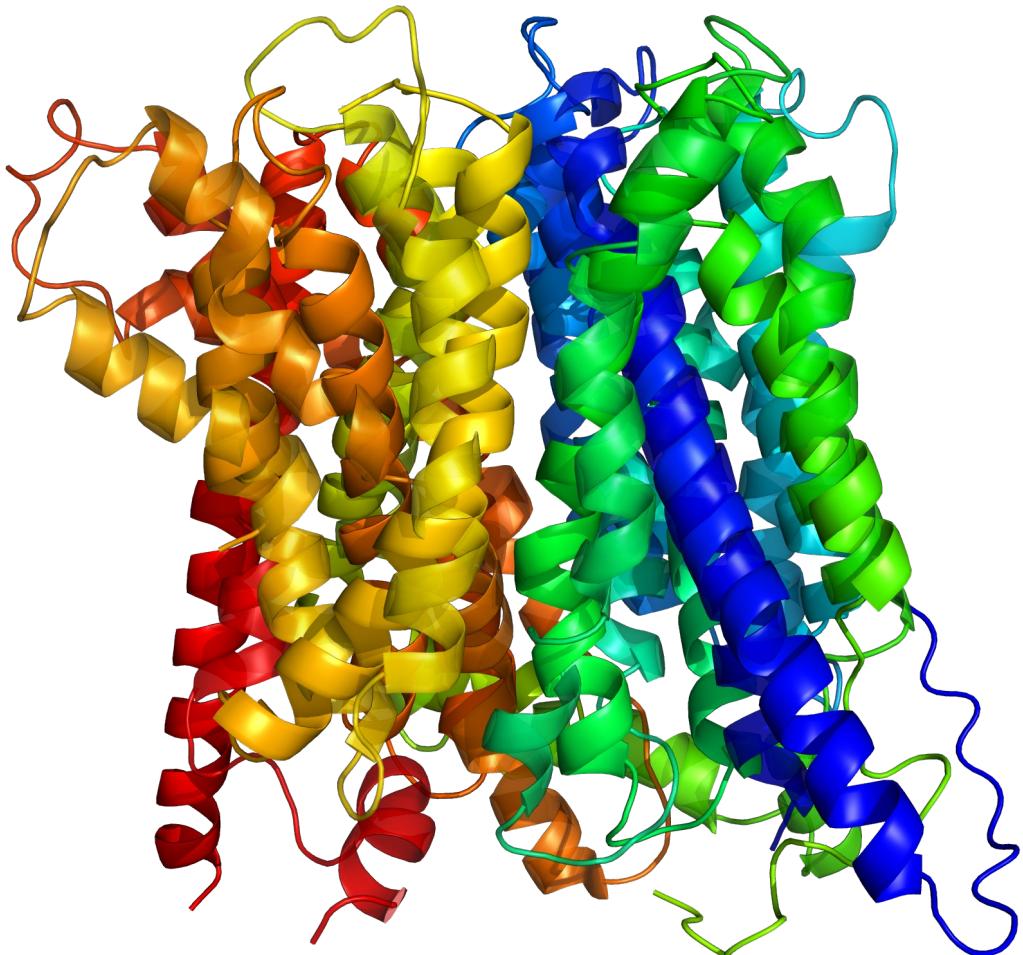


Figure 8 : Superposition du modèle PHYR2 avec le transporteur bactérien, ce dernier avec une transparence 0.3. Les couleurs différencient les domaines transmembranaires.

#### Génération des modes normaux avec NOMAD-REF

On cherche à étudier l'ouverture du récepteur à un éventuel ligand. Pour cela on a lancé le calcul des modes normaux (60) sur NOMAD-REF. Les modes sélectionnés sont ceux qui sont associés à des graphiques des cRMS présentant de nombreux pics, pour bien observer la molécule bouger.

La Figure 9 présente la position « ouvert e » et « fermée » du modèle de PHYR2, puisque il s'agit bien du mouvement observé en Cterm. Cela est encourageant pour la validité du modèle, puisque c'est le comportement attendu d'un transporteur. De plus, cela donne un indice pour où chercher une étude éventuelle de docking .

La Figure 10 représente le transporteur bactérien sous le même traitement. La similitude du comportement renforce nos précédentes hypothèses.

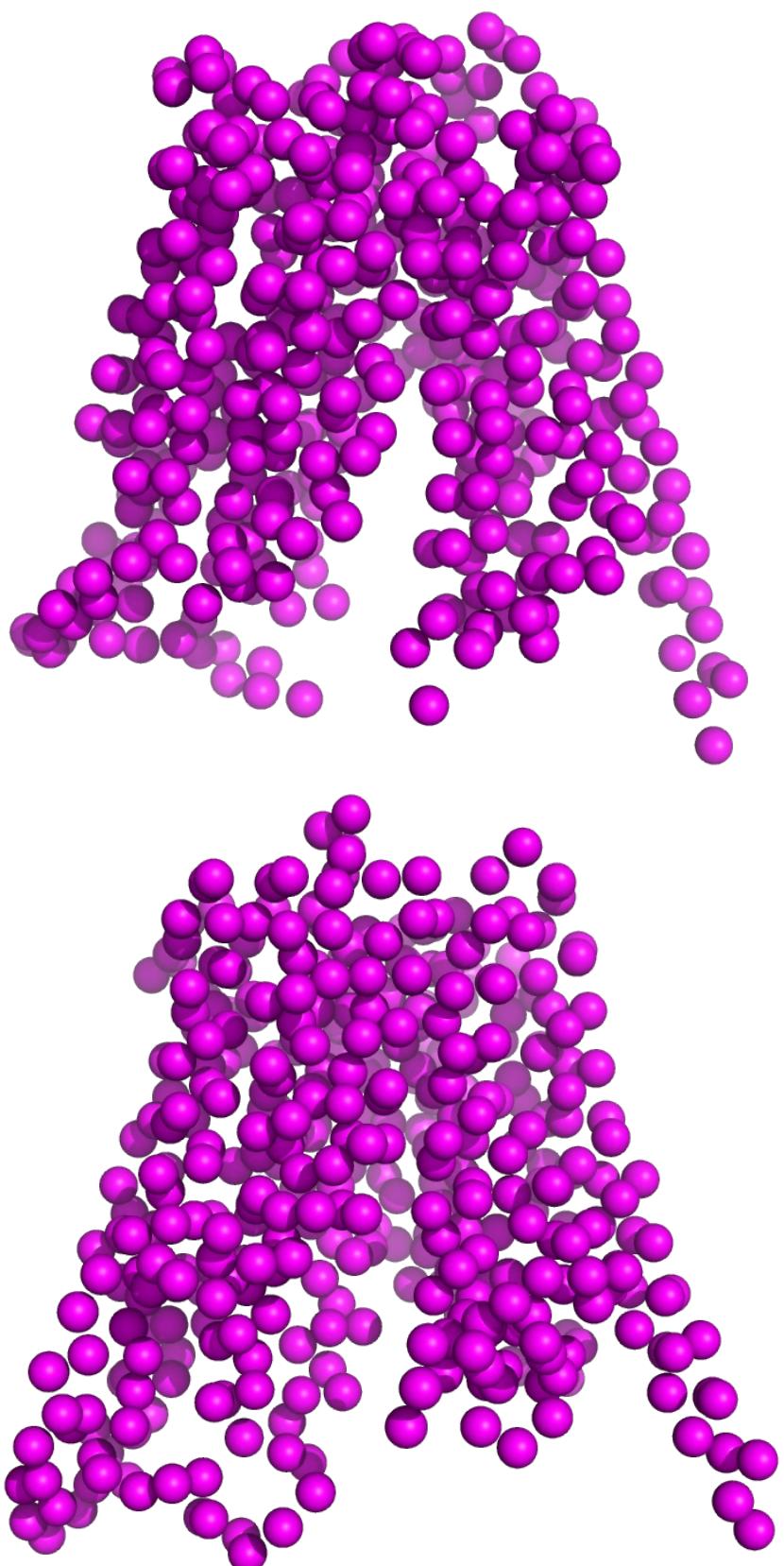


Figure 9: Arrêts sur image PyMOL en représentation sphère des modes normaux calculés avec NOMAD-REF du modèle généré avec PHYR2, successivement en position ouverte puis fermé.

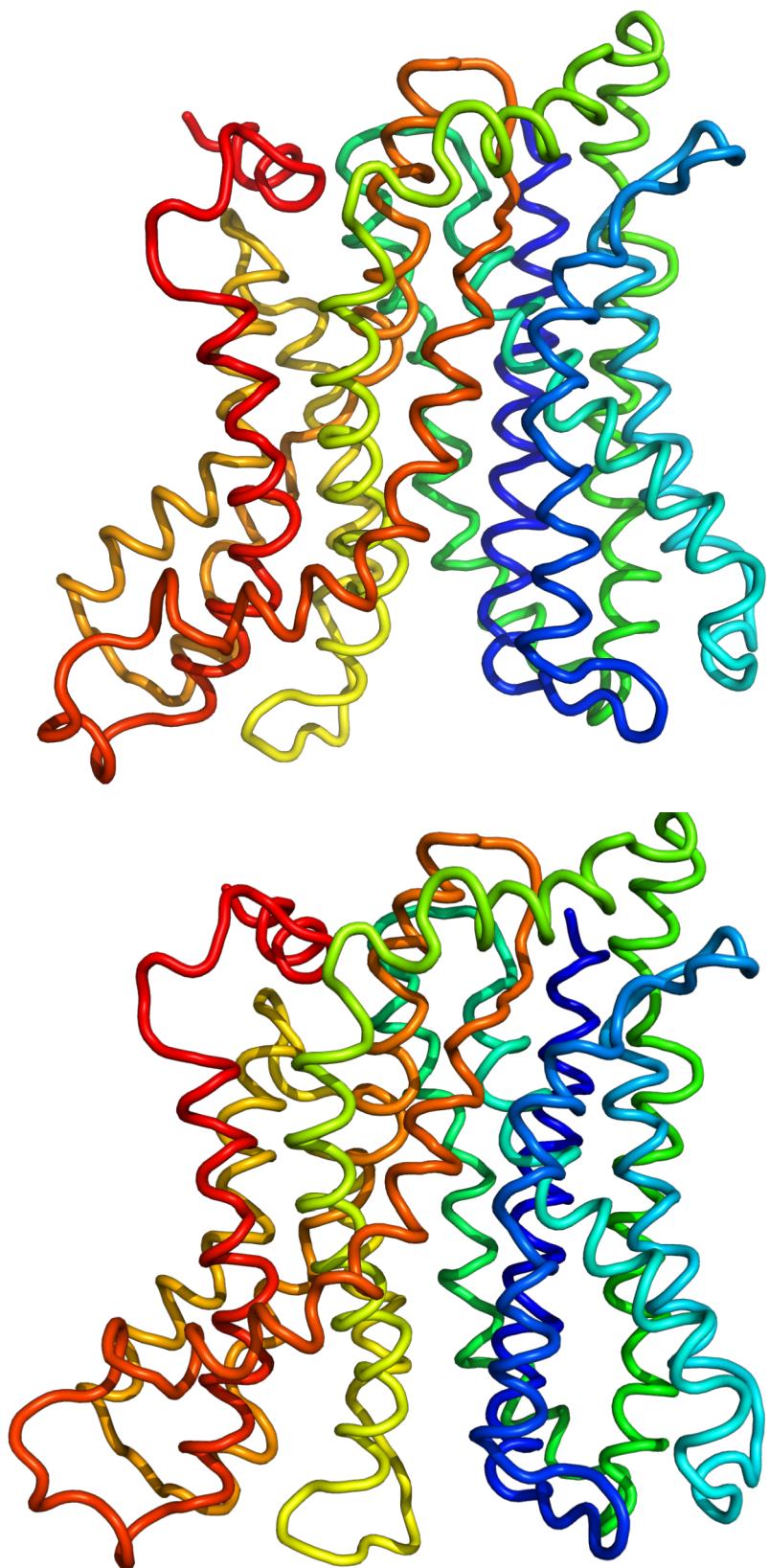


Figure 10 : Arrêts sur image PyMOL en représentation cartoon(les informations des structures secondaires ont été perdues, mais on reconnaît les hélices) des modes normaux calculés avec NOMAD-REF du transporteur bactérien, successivement en position ouverte puis fermé.

## Conclusion

Il est difficile de faire des modèles de protéines transmembranaires à cause du manque de structure cristallographiques expérimentales. Le modèle par homologie n'est pas toujours possible, ce qui complique les études. Les techniques comme le threading sont adaptés à ce type de problème, et le fait que la protéine soit transmembranaire demande un travail supplémentaire de prédiction de structure secondaire. Dans cette étude, il a été possible de générer 2 modèles avec une bonne prédiction de structures secondaires, qui sont ici les 12 hélices alpha transmembranaires. La comparaison avec le transporteur bactérien Na(+)/melibiose symporters selon la littérature renforce l'hypothèse selon laquelle notre modèle est fonctionnel. De plus, l'étude des modes normaux indique un emplacement intéressant pour étudier le docking d'un ligand sur cette protéine, cette information étant aussi donnée par les modes normaux du transporteur bactérien. Le docking est l'étape suivante de cette étude, ainsi que la recherche d'une ouverture en N-terminal qui n'a pas été mise en évidence.

## Bibliographie

- [1] Brevern, Alexandre G. de. « 3D Structural Models of Transmembrane Proteins ». *Methods in Molecular Biology (Clifton, N.J.)* 654 (2010): 387-401. [https://doi.org/10.1007/978-1-60761-762-4\\_20](https://doi.org/10.1007/978-1-60761-762-4_20).
- [2] Vu, Thiet M., Ayako-Nakamura Ishizu, Juat Chin Foo, Xiu Ru Toh, Fangyu Zhang, Ding Ming Whee, Federico Torta, et al. « Mfsd2b Is Essential for the Sphingosine-1-Phosphate Export in Erythrocytes and Platelets ». *Nature* 550, n° 7677 (26 octobre 2017): 524-28. <https://doi.org/10.1038/nature24053>.
- [3] Wang, Sheng, Wei Li, Shiwang Liu, et Jinbo Xu. « RaptorX-Property: A Web Server for Protein Structure Property Prediction ». *Nucleic Acids Research* 44, n° W1 (8 juillet 2016): W430-435. <https://doi.org/10.1093/nar/gkw306>.
- [4] Ethayathulla, Abdul S., Anowarul Amin, Gérard Leblanc, H. Ronald Kaback, Lan Guan, et Mohammad S. Yousef. « Structure-Based Mechanism for  $\text{Na}^+$ /Melibiose Symport by MelB ». *Nature Communications* 5 (6 janvier 2014): 3009. <https://doi.org/10.1038/ncomms4009>.