

Introduction

La modélisation moléculaire est un ensemble de techniques permettant la modélisation prédictive notamment de protéines. On utilise ces techniques le plus souvent en chimie et en biologie (pharmaceutique).

Faire un modèle, va permettre de simuler le comportement d'une protéine.

Le but de ce projet est de manipuler différents outils de modélisation et d'évaluation de structures de protéines et pouvoir comparer les résultats des uns par rapport aux autres. Pour cela nous nous sommes intéressés à une famille de protéines appartenant à la Major Facilitator Superfamily (MFS), la plus grande famille de transporteurs secondaires.

Matériel et Méthode

Au cours de ce projet, les outils suivants ont été utilisés:

NCBI protein blast afin de trouver des séquences homologues à celle de la protéine d'intérêt.

La Base de donnée PDB afin de trouver les fichiers structures des séquences homologues qui vont servir de support à la modélisation.

NCBI MUSCLE pour aligner les séquences homologues obtenus afin de retenir les régions conservées.

La Base de donnée Uniprot qui servira de point de départ pour cette étude.

PsiPred pour la prédiction de structure secondaire de la protéine.

Les outils de modélisation avec leurs paramètres :

- Swiss-Model

input: Séquence fasta de la protéine cible + protéine support

output: n modèles avec leur score de Qmean et leur score QMGE.

Cet outil après modélisation assigne un score et un certains nombres de modèles (les mieux) qu'il classe selon ces scores. Qmean pour Qualitative Model Energy Analysis, plus il est petit meilleur est le modèle. Et QMGE qui est une mesure de Z-score montrant la correspondance entre la séquence et la structure, plus il est grand mieux c'est.

- Modeller

Cet outil a été utilisé sous l'outil HHpred qui fait en même temps de la prédiction de structure secondaire et de la modélisation par homologie.

input: Séquence fasta de la protéine cible + protéine support

output: 1 modèle

- Memoir

input: fasta cible, code pdb du support, chaîne du support

output: 3 modèles.

Cet outil renvoie trois modèles avec ces caractéristiques : celui avec la plus grande précision, la plus couverture, la structure sans les boucles et la structure complète.

Ici je vais choisir la structure complète car les boucles contiennent aussi des sites clés pour cette protéine.

- Homer

input: alignement fasta cible - support, code pdb du support

output: 1 modèle

Les outils d'évaluation

- OREMpro(1)

input: modèle

output: score de MAIDEN, Nombre de segments transmembranaires et l'épaisseur d'hydrophobicité.

Cet outil permet également d'orienter et de placer la protéine dans la membrane

Le MAIDEN score est une pseudo fonction énergétique optimisée sur les structures de protéines membranaires, donc plus elle est petite, meilleure est la structure(1).

- Verify3D

input: modèle

output: le pourcentage de résidus ayant une valeur de supérieure à un seuil (ici 0.2).

- ProsA

input: modèle

output: Z score (plus il est petit meilleur est le modèle).

L'outil de raffinement GalaxyRefine

Séquence à modéliser

La protéine d'intérêt qui fera l'objet de cette étude a pour identifiant Uniprot Q9BYW1. C'est un transporteur de glucose de la superfamille majeure de facilitateurs (*major facilitator superfamily* (MFS)) de l'humain. Cette grande famille est divisée en 3 classes: classe I GLUT1-4 et GLUT14, la Class II (GLUT5, 7, 9, 11) et la Class III (GLUT6, 8, 10, 12 et 13). La protéine d'intérêt, elle est un GLUT11 de la Classe II. C'est donc une protéine membranaire et sa fonction est de faciliter le transport du glucose d'un côté à l'autre de la membrane d'où le nom de Glucose transporter type 11 (GLUT-11).

Cette protéine a une séquence de 496 acides aminés :

MRALRRLIQGRILLTICAAGIGGTFQFGYNLSIINAPTLHIQEFTNETWQARTGEPLP
DHLVLLMWSLIVSLYPLGGLFGALLAGPLAITLGRKKSLLVNNIFVVSAAILFGFSRKA
GSFEMIMLGRLLVGVNAGVSMNIQPMYLGESAPKELRGAVAMSSAIFTALGIVMGQ
VVGLRELLGGPQAWPLLLASCLVPGALQLASLPLLPESPRYLLIDCGDTEACLAALR
RLRGSGDLAGELEELEEERAACQGCRARRPWELFQHRALRRQVTSLVVLGSAME
LCGNDSVYAYASSVFRKAGVPEAKIQYAIIGTGSCCELLTAVVSCVVIERVGRRVLLIG
GYSLMTCWGSIFTVALCLQSSFPWTLYLAMACIFAFILSFGIGPAGVTGILATELFDQ
MARPAACMVCGALMWIMLILVGLGFPFIMEALSHFLYVPFLGVVCVCGAIYTGLFLPE
TKGKTFQEISKELHRLNFPARRAQGPTWRSLEVIQSTEL.

Les protéines de cette famille et de cette classe II ont 12 hélices transmembranaires (2) et sont sur 3 niveaux (cytoplasmique, extracellulaire et transmembranaire (constitué que d'hélices). avec les régions Nterminal et Cterminal du côté cytoplasmique.

Les régions importantes de cette famille sont(2) indiqué sur la figure

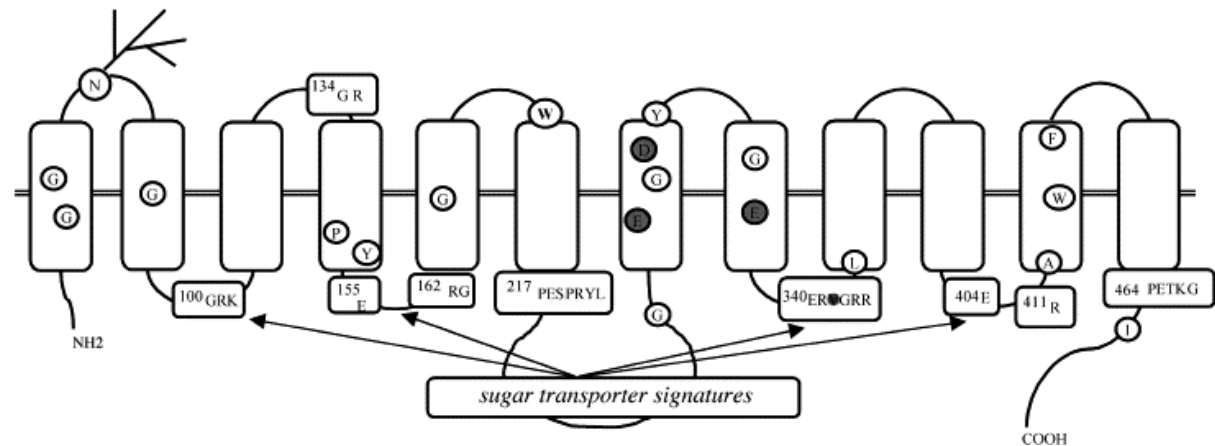


Figure1 : Organisation structurale des protéines de la famille des GLUT11

Cette Classe II dans laquelle les GLUT11 font partie a cette organisation structurale.

Démarche à suivre pour l'étude

Il existe différentes manières de modéliser (de novo, par threading, par homologie) et le choix de l'une ou l'autre des méthodes dépend des données que l'on a au départ. Dans cette étude, je ferai de la modélisation par homologie et j'explique le pourquoi de ce choix plus tard dans ce rapport. Ci-dessous la démarche suivie (figure2):

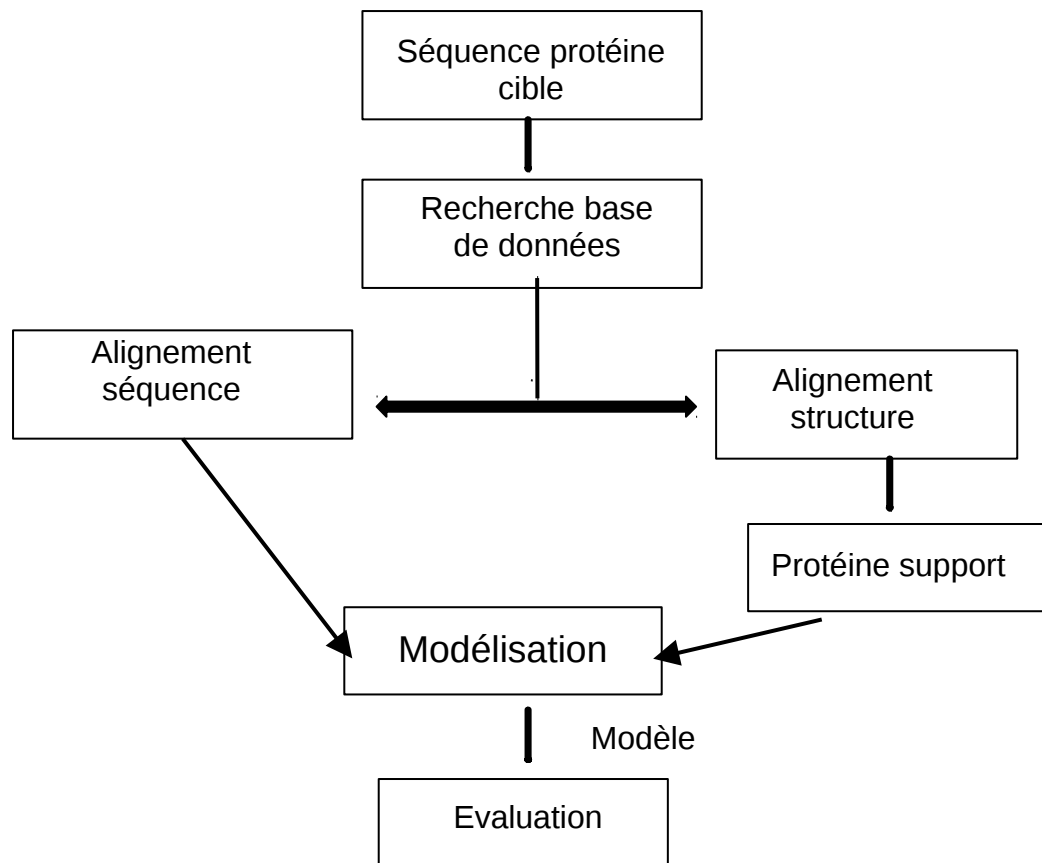


Figure2 : démarche à suivre lors d'une modélisation par homologie

Premier aperçu structural de ma protéine

Profil d'hydrophobicité

Avec l'outil protscale, je dresse le profil d'hydrophobicité de ma protéine cible. Cela me permet d'avoir un premier aperçu de la structure de ma protéine. Et je constate que la majeure partie de la protéine est hydrophobe, ce qui semble logique vu qu'on a à faire à une protéine membranaire.

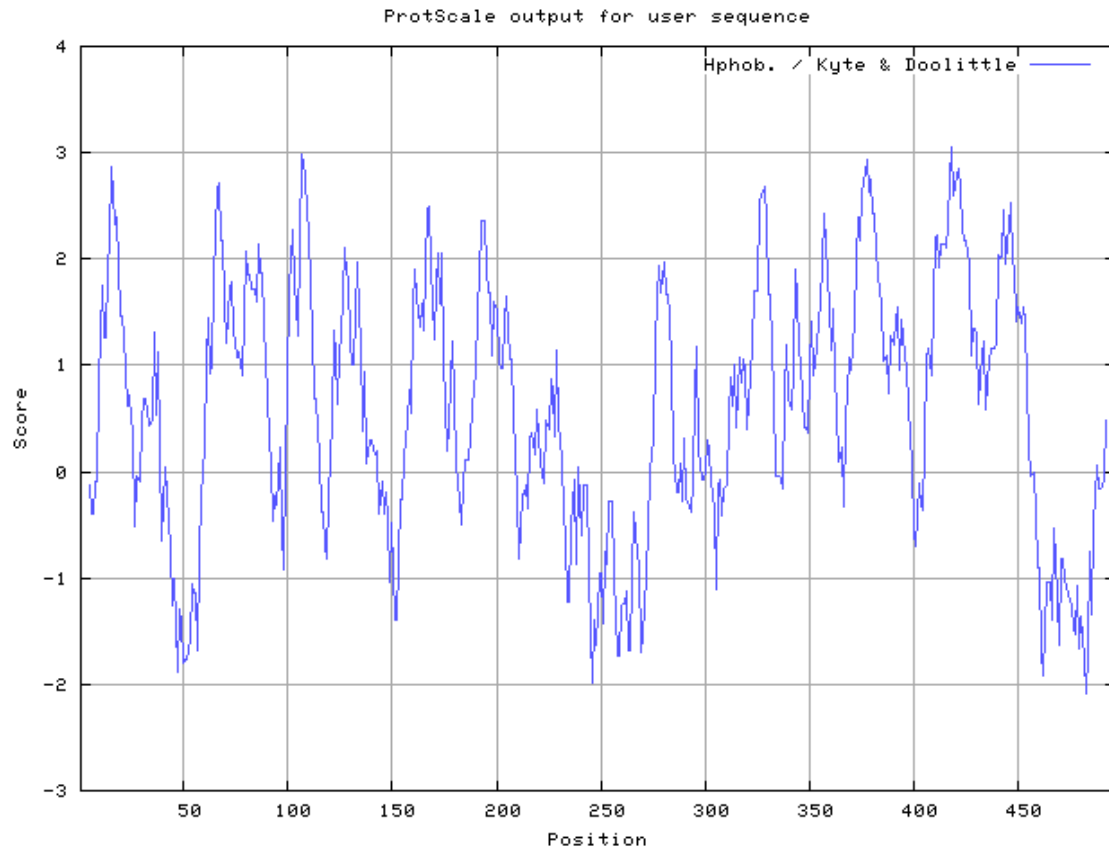
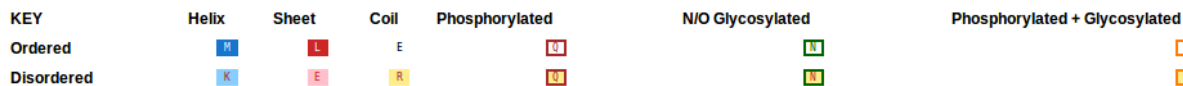
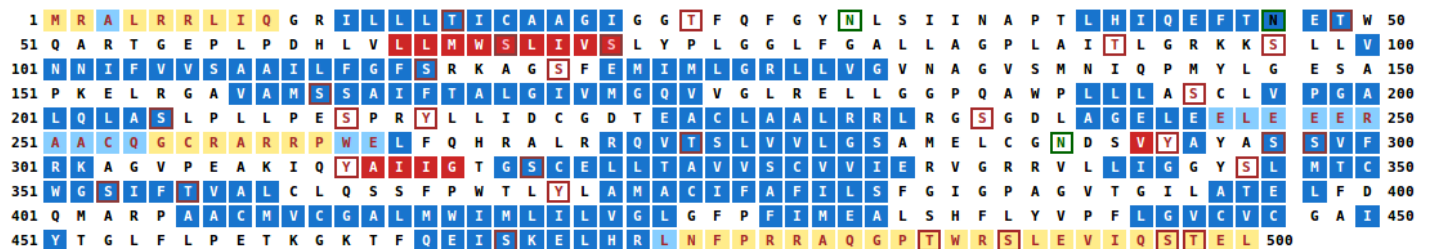


Figure3 : profil d'hydrophobicité de la protéine cible . LA position des acides aminés en abscisse, les valeurs négatives représentent les segments hydrophiles et les valeurs positives les segments hydrophobes.

Prédiction de structure secondaire



Predicted Transmembrane Topology

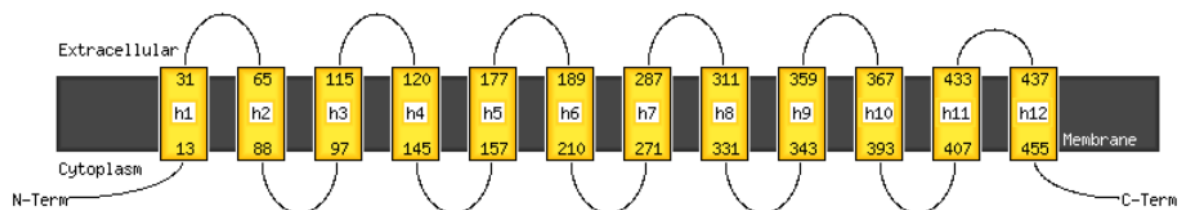


Figure4 : prédiction de la structure secondaire de la protéine d'intérêt (en haut). En bas, les positions des hélices transmembranaires.

Ensuite, je fais une prédiction de structure secondaire de ma protéine avec l'outil Psipred. Ceci me permet de voir la correspondance de ma structure avec les données de la littérature (figure4).

ET on constate qu'il ya d'autres hélices dans la structures mais qui ne sont pas dans la membrane.

Recherche de la séquence support

La 1ere étape est de chercher une séquence homologue à la nôtre en lançant un blastp contre la PDB pour chercher la structure d'une protéine homologue à la protéine d'intérêt, qui a un recouvrement maximum sur sa séquence avec un pourcentage d'identité acceptable et dont la structure résolue contient la zone d'alignement des deux séquences. On part de l'idée que deux séquences ayant des séquences proches pourraient avoir des structures proches.

Ce premier blast donne le résultat suivant comme sur la figure 5.

Selon le pourcentage d'identité obtenu, on choisit de faire du threading (<20%) ou de la modélisation par homologie (>20%-). Avec un pourcentage d'identité de 42.5%, je choisis donc de modéliser par homologie. La protéine d'identifiant P58353 est un bon candidat pour servir de support pour la modélisation.

La modélisation par homologie a pour principe de partir d'une séquence protéique connue et de rechercher des séquences homologues dont la structure tridimensionnelle a déjà été prédite.

<input type="checkbox"/>	Query: sp Q9BYW1 GTR11_HUMAN B20171124A7434721E10EE6586998A056CCD0537E0C405BW		
<input type="checkbox"/>	P22732	GTR5_HUMAN - Solute carrier family 2, facilitate... Homo sapiens (Human) - View alignment	E-value: 6.2e-129 Score: 998 Ident.: 42.9%
<input type="checkbox"/>	P58353	GTR5_BOVIN - Solute carrier family 2, facilitate... Bos taurus (Bovine) - View alignment	E-value: 1.4e-127 Score: 989 Ident.: 42.5%
<input type="checkbox"/>	Q6PXP3	GTR7_HUMAN - Solute carrier family 2, facilitate... Homo sapiens (Human) - View alignment	E-value: 8.5e-123 Score: 958 Ident.: 42.1%
<input type="checkbox"/>	P43427	GTR5_RAT - Solute carrier family 2, facilitate... Rattus norvegicu... - View alignment	E-value: 1.2e-119 Score: 936 Ident.: 40.3%
<input type="checkbox"/>	P11166	GTR1_HUMAN - Solute carrier family 2, facilitate... Homo sapiens (Human) - View alignment	E-value: 9.6e-102 Score: 815 Ident.: 36.5%
<input type="checkbox"/>	P11169	GTR3_HUMAN - Solute carrier family 2, facilitate... Homo sapiens (Human) - View alignment	E-value: 1.3e-91 Score: 747 Ident.: 35.2%
<input type="checkbox"/>	P22732-2	GTR5_HUMAN - Isoform 2 of Solute carrier family ... Homo sapiens (Human) - View alignment	E-value: 1.2e-51 Score: 453 Ident.: 51.5%
<input type="checkbox"/>	A0A0H2VG78	A0A0H2VG78_STAES - Bicyclomycin resistance protein Tca... Staphylococcus e... - View alignment	E-value: 7.6e-49 Score: 449 Ident.: 29.0%

Figure5 : résultat du blastp pour trouver la protéine support, avec les séquences rangées selon le score.

J'ai choisis la protéine d'identifiant P58353 dont la structure résolue va de la position 1-473. Même si la zone de l'alignement de la séquence cible avec cette séquence n'est pas complètement comprise dans cette zone de résolution (acides aminés), il m'a semblé judicieux de préférer cela car même si la séquence de certaines protéines avaient une meilleure couverture comme P11166, sa structure a été résolue en complexe, la conformation adoptée n'est peut être pas forcément celle prise de base.. Ce qui est aussi le cas pour les autres séquences même si la zone d'alignement est comprise dans la zone résolue. Pour cette protéine, sa structure résolue a pour code pdb 4YB9 à une résolution de 3.2Å. C'est une Structure cristalline du transporteur de fructose Bovine GLUT5 dans une

conformation ouverte vers l'intérieur. Les deux protéines appartiennent donc toutes les deux à la même classe II.

Résultats

Les outils d'évaluation ont donné les résultats suivants:

Pour les modèles de Swiss-Model qui a fourni 3 modèles

OREMpro:

Le meilleur modèle est le modèle3 de Swiss-Model, obtenu avec la protéine support que j'ai choisie.

Ce modèle a bien 12 hélices transmembranaires et il donne de bons score comparé aux autres.

•

Méthode de modélisation	Swiss-Model		
Modèles	01	02	03
Nombre d'hélices transmembranaires	11	10	12
Epaisseur de la membrane	-23.7	31.8	22
Score de MAIDEN	-41.81	-38	-40.3
Classement selon cet outil	2	3	1

•

Méthode de modélisation	Swiss-Model		
Modèles	01	02	03
	76%	54%	78%
Classement selon cet outil	2	3	1

•

Méthode de modélisation	Swiss-Model		
Modèles	01	02	03
Zscore	-6	-4.2	-5.8

Classement selon cet outil	1	3	2
----------------------------	---	---	---

Je choisis donc de prendre le modèle n°3 , pour l'outil Swiss-Model comme étant le meilleur modèle.

- Homer

Les modèles ici, donnent de mauvais scores avec les différents outils testés

verify3D	22%
prosA	-3.56
OREMpro	5 hélices transmembranaires
	20°A
	score de MAIDEN -8.87

- Modeller

verify3D	70%
prosA	-5.5
OREMpro	12 hélices transmembranaires
	22°A
	score de MAIDEN -41.27

Je choisis de prendre le modèle de Swiss-Model comme mon meilleur modèle. Car en faisant une comparaison des scores obtenus avec les différentes méthodes d'évaluation, ce sont ses modèles qui ont les meilleurs scores.

Après évaluation avec les différentes méthodes , swiss modèle semble être la meilleur méthode de modélisation car ça donne le meilleur score avec les différentes caractéristiques correspondant à ce qui était attendu.

J'ai ensuite effectué un raffinement et j'obtiens les valeurs suivantes, ceci n'améliore pas vraiment mais la structure obtenue est la suivante:

Après raffinement, ma structure finale sans (figure6) et avec(figure7) la membrane.

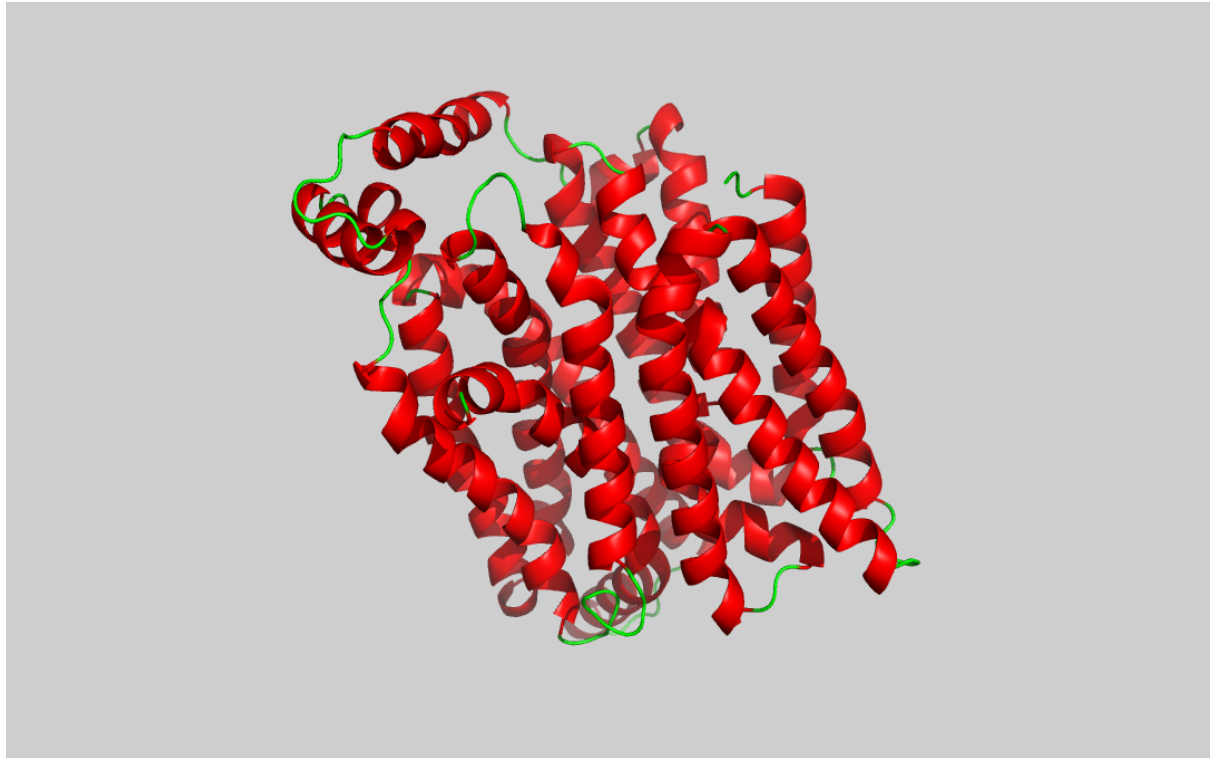


Figure6 : structure protéique de la protéine Q9BYW1 modélisé . Les hélices en rouge et les boucles en vert

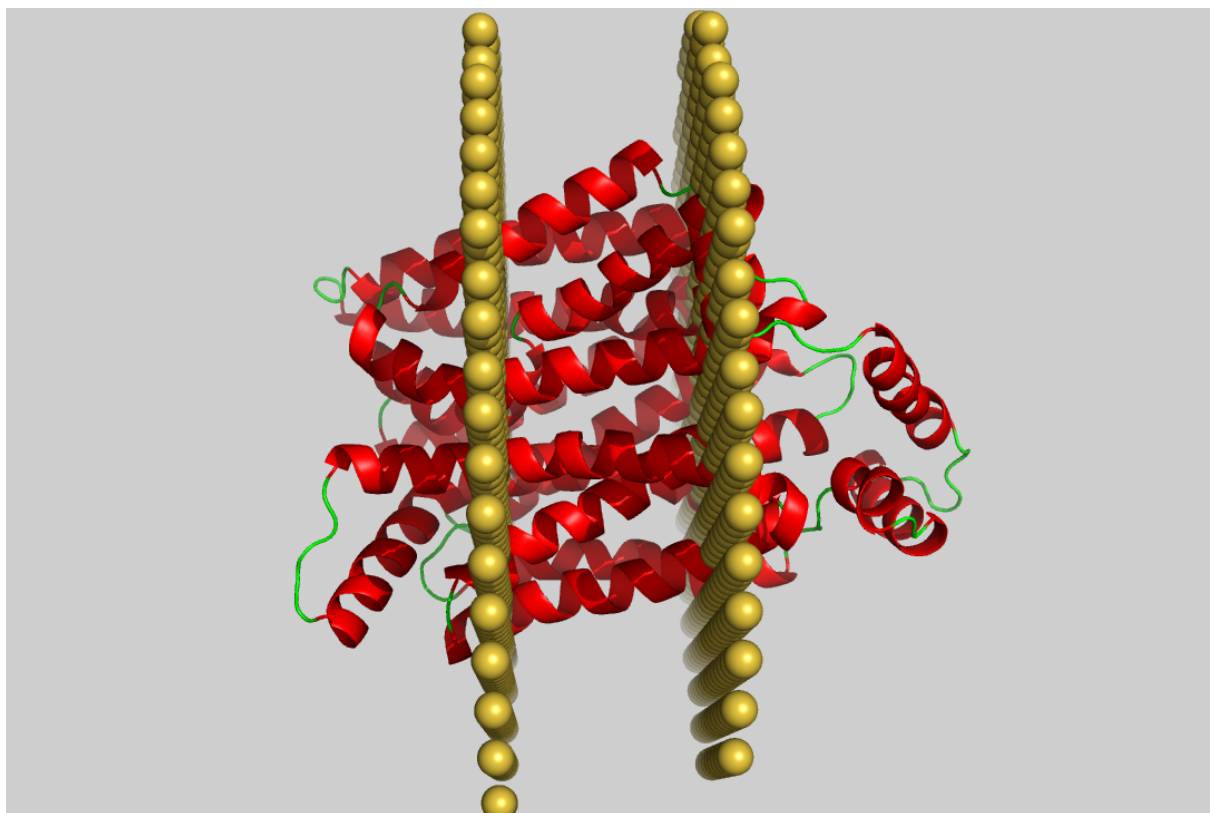


Figure 7: structure protéique de la protéine Q9BYW1 modélisée dans la membrane. Les hélices en rouge et les boucles en vert

Perspectives

Comme perspectives à ce projet, il serait bien pour et faire une analyse plus poussée des modes normaux et aussi de de lancer une dynamique moléculaire pour compléter cette analyse et savoir dans quel sens la protéine bouge. Cette étape pour aussi servir au raffinement de la structure protéique modélisée.

Bibliographie

1. Postic G, Ghouzam Y, Gelly JC. OREMPRO web server: Orientation and assessment of atomistic and coarse-grained structures of membrane proteins. *Bioinformatics*. 2016;32(16):2548–50.
2. Wu X, Li W, Sharma V, Godzik A, Freeze HH. Cloning and characterization of glucose transporter 11, a novel sugar transporter that is alternatively spliced in various tissues. *Mol Genet Metab*. 2002;76(1):37–45.