

Projet de Bioinformatique Structurale

Introduction:

La protéine GLUT-8 est une protéine transmembranaire de type transporteur de glucose qui un membre de la famille des facilitateurs du transport du sucre (Major Facilitator Superfamily ou MFS) [1]. La famille de protéines GLUT se compose de trois familles différentes qui peuvent être distinguées sur leurs homologues de séquence protéique: GLUT-8 appartiennent à la classe III [2].

GLUT8 est régulé par l'insuline et peut être inhibé de manière compétitive par le fructose, le galactose et la cytochalasine B. Ce transporteur est principalement exprimé dans les testicules, et faiblement exprimé dans le cerveau, la glande surrénale, le foie, la rate, le tissu adipeux brun et le poumon. La caractérisation fonctionnelle de GLUT8 dans les ovocytes de *Xenopus laevis* a immédiatement révélé la rétention intracellulaire du transporteur due à la présence d'un motif à base de dileucine dans son extrémité N-terminale. Lorsque les deux résidus de leucine sont changés en alanines, GLUT8 se localise à la membrane plasmique dans des cellules de mammifères qui permettent la détermination de la cinétique de transport de la protéine. GLUT-8 est constitué de 477 acides aminés, à la position 394, GLUT-8 possède un site de liaison qui est l'acide aminé tryptophane.

GLUT8 montre une forte affinité pour le glucose avec un K_m de 2 mM.

Cette protéine ne possède pas encore de structure 3D. L'objectif de ce projet est de construire un modèle 3D à partir de différents outils, puis évaluer ces modèles. Ensuite étudier la conformation des modèles générés et enfin tester les sites de fixation du ligand pour les différentes conformations en utilisant des outils de docking.

Matériel et Méthode:

I. Génération des modèles:

- 1) **Alignement multiple** : Recherche de support et identification des régions conservées:

Entry	Alignment overview	Info	Status
Query: sp Q9NY64 GTR8_HUMAN B20171121C2335653E4FA1B8AECF5153189FA788F366990N			
A0A0H2VG78	A0A0H2VG78_STAES - Bicyclomycin resistance protein Tca... - Staphylococcus e... - View alignment	E-value: 8.8e-63 Score: 546 Ident.: 30.6%	
P11169	GTR3_HUMAN - Solute carrier family 2, facilitate... Homo sapiens (Human) - View alignment	E-value: 4.5e-61 Score: 536 Ident.: 31.4%	
P0AGF4	XYLE_ECOLI - D-xylose-proton symporter - Escherichia coli... - View alignment	E-value: 5.4e-55 Score: 493 Ident.: 29.1%	
P11166	GTR1_HUMAN - Solute carrier family 2, facilitate... Homo sapiens (Human) - View alignment	E-value: 1.3e-51 Score: 469 Ident.: 28.8%	

Figure 1: alignement multiple basée sur la présence de structure 3D de GLUT-8 (code uniprot: Q9NY64)

Dans un premier temps, on effectue un alignement multiple dans le but d'identifier le support. On observe dans la figure 1 un bon recouvrement de l'alignement pour la protéine de résistance à la bicyclomycine TcaB (code uniprot A0A0H2VG78) mais en regardant les structure PDB, le recouvrement n'est pas total: (de la position 22 à 467). GLUT-1 a un bon recouvrement avec la protéine cible GLUT-8 même si il manque de l'information au niveau de l'extrémité N-terminale. De plus, elle possède plusieurs structure 3D. On choisit donc la protéine GLUT-1 comme support en prenant la structure cristallisée du transporteur de glucose humain GLUT1 (code PDB: 4PYP) résolue par X-ray à 3.17Å étant donné qu'il n'est pas complexé et recouvre bien la protéine cible. Le pourcentage d'identité est de 29%.

Ensuite on a effectué un autre alignement multiple (blast par défaut) dans le but de trouver les séquence homologues. On a sélectionné les 10 séquences homologues dans différents espèce dans le but d'identifier la conservation de certaines régions importante. Pour ce faire, on a utilisé l'outils MUSCLE [3][4] qui permet d'effectuer un alignement multiple.

CLUSTAL multiple sequence alignment by MUSCLE (3.8)

```

tr|B4YY04|B4YY04_CANLF      MTPENPEETQPLGGPPGGSAPLSRRVFLAAFAAASLGPLSGFGALGYSSPAIPSLRRAPPP
tr|A0A160CBN9|A0A160CBN9_BOSMU  HTPEDEETQPLRRPPGGSAPRGRVFLAAFAAALGPLSGFGALGYSSPAIPSLRRAPPP
tr|H0WGS2|H0WGS2_OTOGA      MTPDDDEEAQPLGGPPGGSAPRGRVFLAAFAAALGPLSGFGALGYSSPAIPSLRRAPPP
tr|A0A153FKJ6|A0A153FKJ6_DIPOR  MTPDDPEETQPLRRPPDHSVPRSRRLGFLAAFAAALGPLSGFGALGYSSPAIPSLRRAPPP
tr|F6SK66|F6SK66_CALJA      MTPEDPEETQPLLEPPGGSAPRGRVFLAAFAAALGPLSGFGALGYSSPAIPSLRRAPPP
tr|I2CUN0|I2CUN0_MACMU      MTPEDPEETQPLGGPPGGSAPRGRVFLAAFAAALGPLSGFGALGYSSPAIPSLRRAPPP
tr|H2PTF6|H2PTF6_PONAB      MTPEDPEETQPLGGPPGGSAPRGRVFLAAFAAALGPLSGFGALGYSSPAIPSLRRAPPP
tr|G156M4|G156M4_NOMLE      MTPEDPEETQPLGGPPGGSAPRGRVFLAAFAAALGPLSGFGALGYSSPAIPSLRRAPPP
sp|Q9NY64|GTR8_HUMAN        MTPEDPEETQPLGGPPGGSAPRGRVFLAAFAAALGPLSGFGALGYSSPAIPSLRRAPPP
tr|H2QXX2|H2QXX2_PANTR      * * * * *

tr|B4YY04|B4YY04_CANLF      ALRLDAAASWFGAIVTLGAAAGGVLGGWLYDRAGRKLSLLCSAPFVGGFVVITAAQNV
tr|A0A160CBN9|A0A160CBN9_BOSMU  APHLDEDAASWFGAIVTLGAAAGGVLGGWLYDRAGRKLSLVLCALPFFVAGFAVITAAQNL
tr|H0WGS2|H0WGS2_OTOGA      APRLNDSAAWFGAIVTLGAAAGGVLGGWLYDRAGRKLSLLCTVPFVVGFAVITAAQDV
tr|A0A153FKJ6|A0A153FKJ6_DIPOR  APRLDDEAASWFGAVVTLGAAAGGVLGGWLYDRAGRKLSLLCTVPFVVGFAVITAAQNV
tr|F6SK66|F6SK66_CALJA      TPRLDDEAASWFGAIVTLGAAAGGVLGGWLYDRAGRKLSLLCSVPFVVGFAVITAAQDV
tr|I2CUN0|I2CUN0_MACMU      APHLDAAASWFGAIVTLGAAAGGVLGGWLYDRAGRKLSLLCSVPFVVGFAVITAAQDV
tr|H2PTF6|H2PTF6_PONAB      APRLDAAASWFGAVVTLGAAAGGVLGGWLYDRAGRKLSLLCSVPFVVGFAVITAAQDV
tr|G156M4|G156M4_NOMLE      APRLDAAASWFGAVVTLGAAAGGVLGGWLYDRAGRKLSLLCSVPFVVGFAVITAAQDV
sp|Q9NY64|GTR8_HUMAN        APRLDAAASWFGAVVTLGAAAGGVLGGWLYDRAGRKLSLLCSVPFVVGFAVITAAQDV
tr|H2QXX2|H2QXX2_PANTR      : * : : * : : : * : : : * : : : * : : : * : : : * : : : * : : :

tr|B4YY04|B4YY04_CANLF      WMLLGGRLLTGLACGASLVAPVYIIEIAYPAVRGLLGSVCVLMVVGTILLAYLAGWVLE
tr|A0A160CBN9|A0A160CBN9_BOSMU  WMLLGGRLLTGLACGASLVAPVYIIEIAYPAVRGLLGSVCVLMVVGTILLAYLAGWVLE
tr|H0WGS2|H0WGS2_OTOGA      WMLLGGRLLTGLACGASLVAPVYIIEIAYPAVRGLLGSVCVLMVVGTILLAYLAGWVLE
tr|A0A153FKJ6|A0A153FKJ6_DIPOR  LMLLGGRLLTGLACGASLVAPVYIIEIAYPAVRGLLGSVCVLMVVGTILLAYLAGWVLE
tr|F6SK66|F6SK66_CALJA      WMLLGGRLLTGLACGASLVAPVYIIEIAYPAVRGLLGSVCVLMVVGTILLAYLAGWVLE
tr|I2CUN0|I2CUN0_MACMU      WMLLGGRLLTGLACGASLVAPVYIIEIAYPAVRGLLGSVCVLMVVGTILLAYLAGWVLE
tr|H2PTF6|H2PTF6_PONAB      WMLLGGRLLTGLACGASLVAPVYIIEIAYPAVRGLLGSVCVLMVVGTILLAYLAGWVLE
tr|G156M4|G156M4_NOMLE      WMLLGGRLLTGLACGASLVAPVYIIEIAYPAVRGLLGSVCVLMVVGTILLAYLAGWVLE
sp|Q9NY64|GTR8_HUMAN        WMLLGGRLLTGLACGASLVAPVYIIEIAYPAVRGLLGSVCVLMVVGTILLAYLAGWVLE
tr|H2QXX2|H2QXX2_PANTR      * * * * *

tr|B4YY04|B4YY04_CANLF      WRMLAVLGCVPASFMLLLMCFMPETPFLLTQHRQOEAMAAMQFLWGSQGWEEPPVGAQ
tr|A0A160CBN9|A0A160CBN9_BOSMU  WRMLAVLGCVPSPFMLLLMCFMPETPFLLTQHRQOEAMAAMQFLWGAQGWEEPPVGAQ
tr|H0WGS2|H0WGS2_OTOGA      WRMLAVLGCMPSPFMLLLMCFMPETPFLLTQHRQOEAMAAMQFLWGSQGWEEPPVGAQ
tr|A0A153FKJ6|A0A153FKJ6_DIPOR  WRMLAVLGCVPSPFMLLLMCFMPETPFLLTQHRQOEAMAAMQFLWGSQGWEEPPVGAQ
tr|F6SK66|F6SK66_CALJA      WRMLAVLGCVPSPFMLLLMCFMPETPFLLTQHRQOEAMAAMQFLWGSQGWEEPPVGAQ
tr|I2CUN0|I2CUN0_MACMU      WRMLAVLGCAPPSPFMLLLMCFMPETPFLLTQHRQOEAMAAMQFLWGSQGWEDPPIGAE
tr|H2PTF6|H2PTF6_PONAB      WRMLAVLGCVPSPFMLLLMCFMPETPFLLTQHRQOEAMAAMQFLWGSQGWEDPPIGAE
tr|G156M4|G156M4_NOMLE      WRMLAVLGCVPSPFMLLLMCFMPETPFLLTQHRQOEAMAAMQFLWGSQGWEDPPIGAE
sp|Q9NY64|GTR8_HUMAN        WRMLAVLGCVPSPFMLLLMCFMPETPFLLTQHRQOEAMAAMQFLWGSQGWEDPPIGAE
tr|H2QXX2|H2QXX2_PANTR      * * * * *

tr|B4YY04|B4YY04_CANLF      HOGFQLAQLRHPGVYKPFIIIGILLNAFQQLSGINAVMFYAETIFEEAKFKDSSLASVIVG
tr|A0A160CBN9|A0A160CBN9_BOSMU  HOGFHMAQLRRPGVYKPFIIIGILLNAFQQLSGINAVMFYAETIFEEAKFKDSSLASVIVG
tr|H0WGS2|H0WGS2_OTOGA      -OGFHLALLAQPGIYKPFIIIGVSLMAFQQLSGVNAIMFYAETIFEEAKFKDSSLASVIVG
tr|A0A153FKJ6|A0A153FKJ6_DIPOR  HOGFQALLRRPGIYKPFIIIGVSLMAFQQLSGVNAIMFYAETIFEEAKFKDSSLASVIVG
tr|F6SK66|F6SK66_CALJA      -OGFHLALLWQPGIYKPFIIIGVSLMAFQQLSGVNAIMFYAETIFEEAKFKDSSLASVIVG
tr|I2CUN0|I2CUN0_MACMU      -QSFHALLRQPGIYKPFIIIGVSLMAFQQLSGVNAIMFYAETIFEEAKFKDSSLASVIVG
tr|H2PTF6|H2PTF6_PONAB      -QSFHALLRQPSIYKPFIIIGVSLMAFQQLSGVNAIMFYAETIFEEAKFKDSSLASVIVG
tr|G156M4|G156M4_NOMLE      -QSFHALLRQPGIYKPFIIIGVSLMAFQQLSGVNAIMFYAETIFEEAKFKDSSLASVIVG
sp|Q9NY64|GTR8_HUMAN        -QSFHALLRQPGIYKPFIIIGVSLMAFQQLSGVNAIMFYAETIFEEAKFKDSSLASVIVG
tr|H2QXX2|H2QXX2_PANTR      * * * * *

tr|B4YY04|B4YY04_CANLF      IIQVLFAMAALIMDRAGRRLLLTSGVVMVFSTSAFGAYFKLTQGGPNSSHVALLTPI
tr|A0A160CBN9|A0A160CBN9_BOSMU  VIQVLFATAALIMDRAGRRLLLTSGVVMVFSTSAFGAYFKLTQGGPNSSHVALLTPI
tr|H0WGS2|H0WGS2_OTOGA      VIQVLFATAALIMDRAGRRLLLTSGVVMVFSTSAFGAYFKLTQGGPNSSHVALLTPI
tr|A0A153FKJ6|A0A153FKJ6_DIPOR  VIQVLFATAALIMDRAGRRLLLTSGVVMVFSTSAFGAYFKLTQGGPNSSHVALLTPI
tr|F6SK66|F6SK66_CALJA      VIQVLFATAALIMDRAGRRLLLTSGVVMVFSTSAFGAYFKLTQGGPNSSHVALLTPI
tr|I2CUN0|I2CUN0_MACMU      VIQVLFATAALIMDRAGRRLLLTSGVVMVFSTSAFGAYFKLTQGGPNSSHVALLTPI
tr|H2PTF6|H2PTF6_PONAB      VIQVLFATAALIMDRAGRRLLLTSGVVMVFSTSAFGAYFKLTQGGPNSSHVALLTPI
tr|G156M4|G156M4_NOMLE      VIQVLFATAALIMDRAGRRLLLTSGVVMVFSTSAFGAYFKLTQGGPNSSHVALLTPI
sp|Q9NY64|GTR8_HUMAN        VIQVLFATAALIMDRAGRRLLLTSGVVMVFSTSAFGAYFKLTQGGPNSSHVALLTPI
tr|H2QXX2|H2QXX2_PANTR      * * * * *

tr|B4YY04|B4YY04_CANLF      SMEPPSASVGLAWLAVGSMCLFIAGFVAGWGPWPWLLMSEIFPLHVKGATGVCVLTNWL
tr|A0A160CBN9|A0A160CBN9_BOSMU  SMEPADTNVGLAWLAVGSMCLFIAGFVAGWGPWPWLLMSEIFPLHVKGATGVCVLTNWL
tr|H0WGS2|H0WGS2_OTOGA      STEPTDTSVGLAWLAVGSMCLFIAGFVAGWGPWPWLLMSEIFPLHVKGATGVCVLTNWL
tr|A0A153FKJ6|A0A153FKJ6_DIPOR  SAQPVDAVGLAWLAVGSMCLFIAGFVAGWGPWPWLLMSEIFPLQVKGATGVCVLTNWL
tr|F6SK66|F6SK66_CALJA      STEPVDANVGLAWLAVGSMCLFIAGFVAGWGPWPWLLMSEIFPLHVKGATGVCVLTNWL
tr|I2CUN0|I2CUN0_MACMU      SAEPVDASVGLAWLAVGSMCLFIAGFVAGWGPWPWLLMSEIFPLHVKGATGVCVLTNWL
tr|H2PTF6|H2PTF6_PONAB      SAQPVDAVGLAWLAVGSMCLFIAGFVAGWGPWPWLLMSEIFPLHVKGATGVCVLTNWL
tr|G156M4|G156M4_NOMLE      SAEPVDASVGLAWLAVGSMCLFIAGFVAGWGPWPWLLMSEIFPLHVKGATGVCVLTNWL
sp|Q9NY64|GTR8_HUMAN        SAQPVDAVGLAWLAVGSMCLFIAGFVAGWGPWPWLLMSEIFPLHVKGATGVCVLTNWL
tr|H2QXX2|H2QXX2_PANTR      * * * * *

tr|B4YY04|B4YY04_CANLF      MAFLVTKFSSVMEVLRPYGAFWLASAFICFVSVLFTLFCVPETKGTLEQITAHFEGR
tr|A0A160CBN9|A0A160CBN9_BOSMU  MAFLVTKFSSVMEVLRPYGAFWLASAFICFVSVLFTLFCVPETKGTLEQITAHFEGR
tr|H0WGS2|H0WGS2_OTOGA      MAFLVTKFSSVMEVLRPYGAFWLASAFICFVSVLFTLFCVPETKGTLEQITAHFEGR
tr|A0A153FKJ6|A0A153FKJ6_DIPOR  MAFLVTKFSSVMEVLRPYGAFWLASAFICFVSVLFTLFCVPETKGTLEQITAHFEGR
tr|F6SK66|F6SK66_CALJA      MAFLVTKFSSVMEVLRPYGAFWLASAFICFVSVLFTLFCVPETKGTLEQITAHFEGR
tr|I2CUN0|I2CUN0_MACMU      MAFLVTKFSSVMEVLRPYGAFWLASAFICFVSVLFTLFCVPETKGTLEQITAHFEGR
tr|H2PTF6|H2PTF6_PONAB      MAFLVTKFSSVMEVLRPYGAFWLASAFICFVSVLFTLFCVPETKGTLEQITAHFEGR
tr|G156M4|G156M4_NOMLE      MAFLVTKFSSVMEVLRPYGAFWLASAFICFVSVLFTLFCVPETKGTLEQITAHFEGR
sp|Q9NY64|GTR8_HUMAN        MAFLVTKFSSVMEVLRPYGAFWLASAFICFVSVLFTLFCVPETKGTLEQITAHFEGR
tr|H2QXX2|H2QXX2_PANTR      * * * * *

```

Figure 2: Alignement multiple avec l'outil MUSCLE pour visualiser les régions conservées

Dans la figure 2, on observe bien la conservation du motif dileucine en N-terminale ainsi que le site de liaison tryptophane. Mise à part les motifs, plusieurs régions sont aussi bien conservées. Avec différents algorithmes d'alignements (TM-coffee, ClustalW), on observe les mêmes alignements que ce qu'on observe avec MUSCLE.

2) Prédiction de structure secondaire :

A partir de la séquence cible GLUT-8, on va prédire la structure secondaire pour identifier les segments transmembranaires, et pour avoir une vue d'ensemble de ce qui est attendue lors de la construction du modèle. Pour ce faire on a utilisé l'outil PSIPRED[5]. C'est une méthode de prédiction de structure secondaire simple et précise, incorporant deux réseaux de neurones à action directe qui effectuent une analyse sur la sortie obtenue à partir de PSI-BLAST (Position Specific Iterated - BLAST). PSIPRED inclut MEMSAT qui est une méthode de prédiction des protéines membranaires hélicoïdales. La méthode a été testée sur un ensemble de test de protéines transmembranaires de topologie connue.

a486715f-d72a-4196-8e68-22ef452b18ca.seq.job

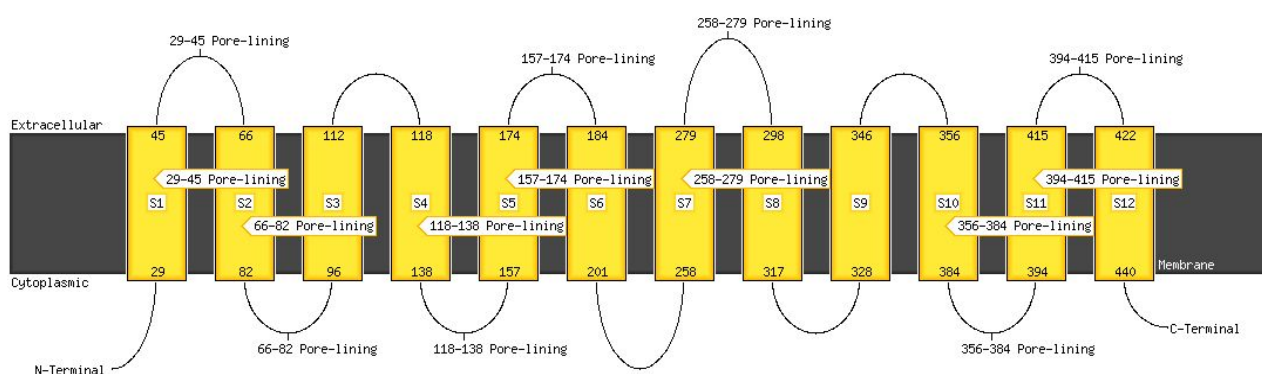


Figure 3: Les segments transmembranaires issus de la prédiction de structure secondaire du support

PSIPRED prédit 19 hélices avec des boucles pour la structure secondaire support GLUT-1, 12 des 19 hélices sont transmembranaires (figure 3). PSIPRED prédit aussi 19 hélices pour GLUT-8 mais on ne connaît pas pour le moment combien parmi les 19 sont transmembranaires.

3) Modélisation par homologie:

a) Membrane protein modelling pipeline(MEMOIR):

MEMOIR est un algorithme de modélisation d'homologie conçu pour les protéines membranaires. Les entrées sont la séquence qui doit être modélisée et la structure 3D du support. MEMOIR intègre des logiciels pour les protéines membranaires: iMembrane (annotation de la membrane), MP-T (alignement), Medeller (génération de coordonnées), Completionist (modélisation des boucles).

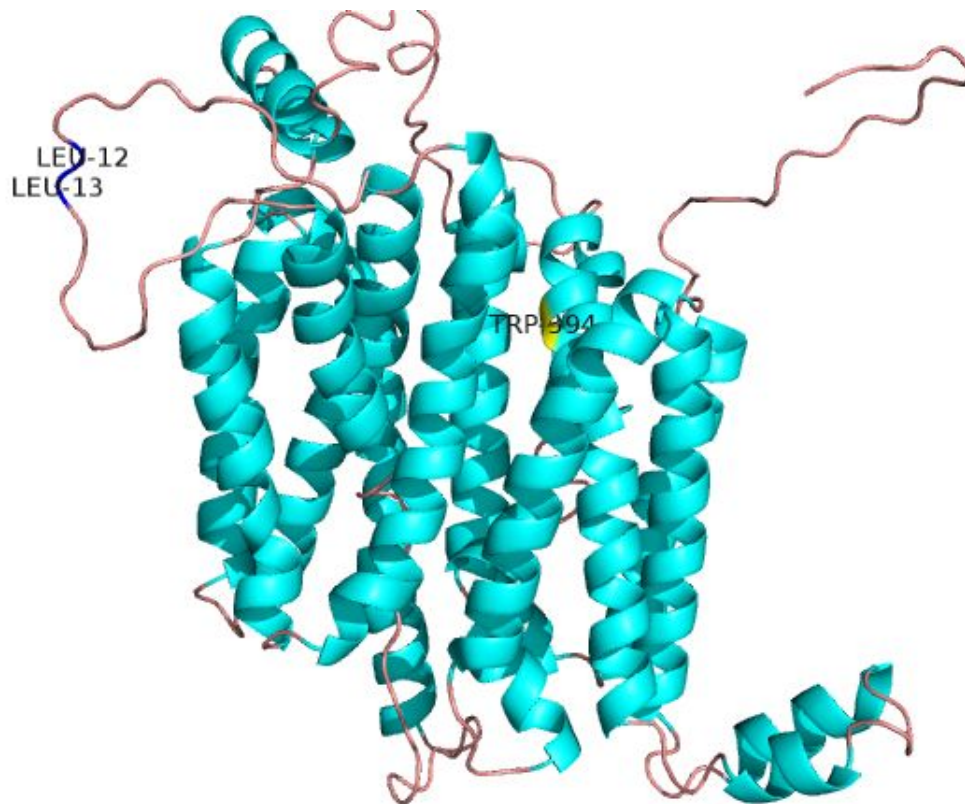


Figure 4: Modèle complet de GLUT-8 modélisé par l'outil MEMOIR, en bleu au niveau de la boucle les résidus dileucine en position 12 et 14, en jaune au niveau de l'hélice le tryptophane en position 394

On observe sur la figure 4 le modèle généré par MEMOIR, on identifie bien les résidus importants (dileucine en N-terminale et Tryptophane en position 394). Les boucles de la partie N-terminale et la partie C-terminale nécessitent un raffinement.

b) MODELLER:

Ensuite on a généré, un modèle avec l'outil MODELLER[6][7] une fois l'alignements avec HHpred effectué.

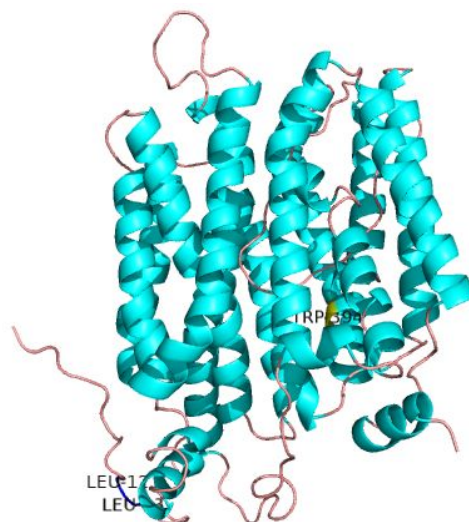


Figure 5: Modèle de GLUT-8 généré par MODELLER

On observe sur la figure 5 le modèle généré par MODELLER, on identifie bien les résidus importants (dileucine en N-terminale et Tryptophane en position 394). La boucle de la partie N-terminale nécessitent un raffinement.

c) SWISS-MODEL:

On a aussi généré un modèle avec l'outil SWISS-MODEL qui a donné trois modèle. Mais les structures ont été prédit qu'à partir de l'acide aminé 24, ce qui signifie qu'on perd certaines informations en N-terminale notamment la Dileucine en position 12 et 13. De plus le qmean est très faible, sans même passé l'étape d'évaluation de modèle on ne prend donc pas en compte des modèles générés par SWISS-MODEL.

Étant donné que le pourcentage d'identité est faible (29%), même en ayant un bon recouvrement, on va effectuer une modélisation par threading.

4) Modélisation par threading:

a) PHYRE2:

Comme le threading traite la séquence à modéliser dans l'alignement comme une structure, PHYRE2 nous donne un modèle en commençant par l'acide aminé 21, ce qui signifie qu'une partie de l'information de N-terminale est manquante. On ne prend pas en compte non plus donc de ce modèle dans l'évaluation.

b) Raptor-X:

RaptorX[8] est un serveur de prédiction de structures 3D pour des séquences de protéines sans homologues proches dans la PDB. Étant donné une séquence d'entrée, RaptorX prédit ses structures secondaires et tertiaires, ses contacts, l'accessibilité au solvant, les régions désordonnées et les sites de liaison. RaptorX attribue également des scores de confiance pour indiquer la qualité d'un modèle 3D prédit.

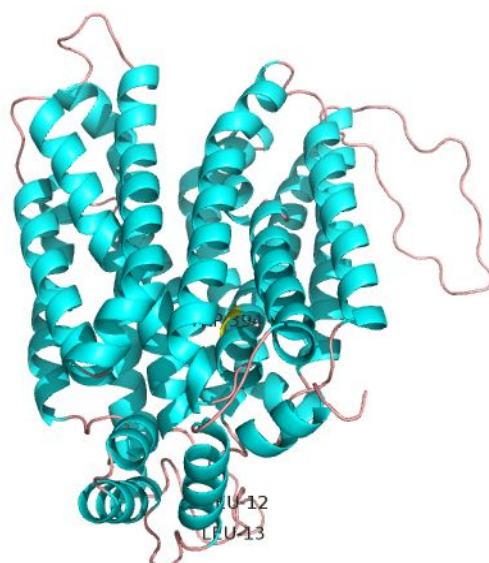


Figure 6: Modèle prédit par Raptor-X, identification des résidus importants (Leucine-12, Leucine-13, Tryptophane -394)

5) Raffinement des boucles:

Parmi les modèles présélectionnés(MODELLER,MEDELLER,Raptor-X), on va raffiner les boucles pour améliorer les modèles. Pour ce faire, on va utiliser l'outil Galaxy.

a) Raffinement des boucles du modèle généré par MODELLER:

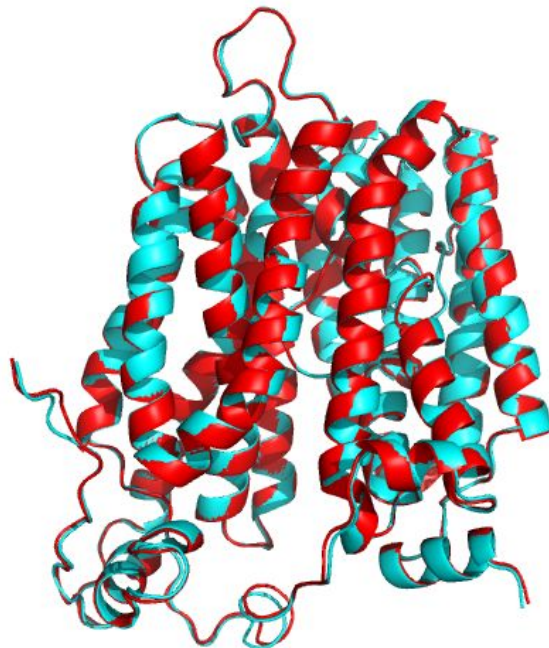


Figure 7: raffinement des boucles du modèle généré par MODELLER, en rouge le modèle raffiné en bleu le modèle initial.

Structure Information

Model	GDT-HA	RMSD	MolProbity	Clash score	Poor rotamers	Rama favored
Initial	1.0000	0.000	3.687	187.7	4.1	91.8
MODEL 1	0.9544	0.433	2.234	34.4	0.3	96.6
MODEL 2	0.9544	0.439	2.278	33.1	0.8	96.0
MODEL 3	0.9523	0.435	2.535	34.2	1.9	95.6
MODEL 4	0.9549	0.417	2.306	33.3	1.1	96.0
MODEL 5	0.9586	0.427	2.271	32.6	0.8	96.0

Tableau 1: Score évaluant le modèle raffiné par Galaxy

On observe dans la figure 7, qu'il n'y a pas de grand changement par rapport au raffinement. Pourtant quand on regarde les scores évalués par Galaxy, le modèle est raffiné, les clash stériques diminuent énormément, on passe de 187.7 à 34.4 pour le modèle 1, de même le score de Ramachandran augmente(Tableau 1).

b) Raffinement du modèle généré par MEDELLER:

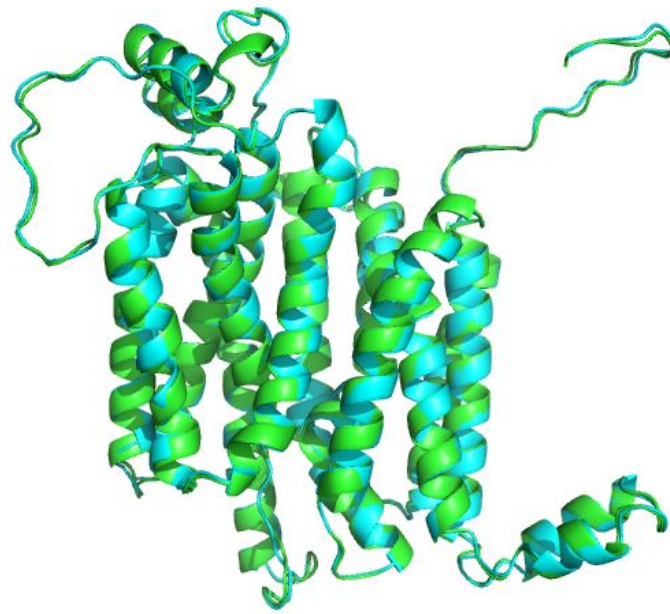


Figure 8: raffinement des boucles du modèle généré par MEDELLER, en vert le modèle raffiné en bleu le modèle initial.

Structure Information

Model	GDT-HA	RMSD	MolProbity	Clash score	Poor rotamers	Rama favored
Initial	1.0000	0.000	3.437	124.2	2.7	89.9
MODEL 1	0.9528	0.412	1.741	15.3	1.1	97.9
MODEL 2	0.9534	0.411	2.010	17.0	2.2	97.9
MODEL 3	0.9460	0.413	1.910	15.9	1.4	97.5
MODEL 4	0.9596	0.394	1.962	17.1	1.6	97.7
MODEL 5	0.9513	0.417	1.887	16.9	1.6	98.1

Tableau 2: Score évaluant le modèle raffiné par Galaxy

On observe dans la figure 8, qu'il n'y a pas de grand changement par rapport au raffinement. Pourtant quand on regarde les scores évalués par galaxy, le modèle est raffiné, les clash stériques diminuent énormément, on passe de 124.2 à 15.3 pour le modèle 1, de même le score de Ramachandran augmente mais pas beaucoup (Tableau 2).

c) Raffinement du modèle généré par Raptor-X:

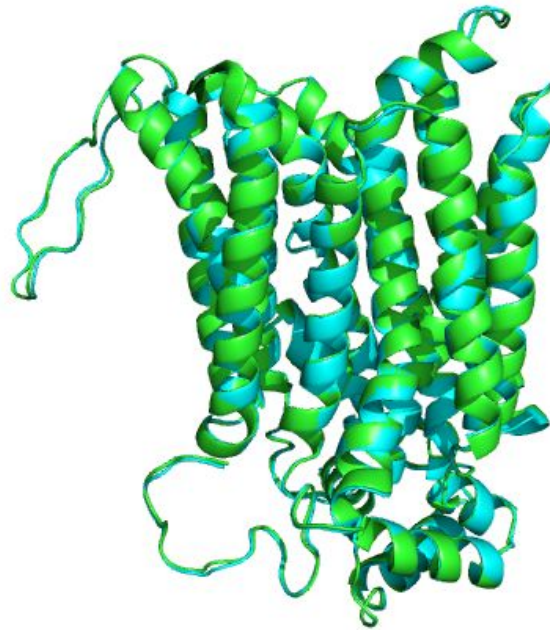


Figure 9: raffinement des boucles du modèle généré par Raptor-X, en vert le modèle raffiné en bleu le modèle initial.

Structure Information

Model	GDT-HA	RMSD	MolProbity	Clash score	Poor rotamers	Rama favored
Initial	1.0000	0.000	3.514	118.7	4.9	92.8
MODEL 1	0.9722	0.350	1.985	15.5	0.8	95.8
MODEL 2	0.9722	0.362	2.015	16.3	1.1	96.0
MODEL 3	0.9706	0.353	2.048	16.3	1.1	95.6
MODEL 4	0.9675	0.368	2.034	15.8	1.1	95.6
MODEL 5	0.9722	0.364	2.005	16.3	0.5	95.8

Tableau 3: Score évaluant le modèle raffiné par Galaxy

On observe dans la figure 9, qu'il n'y a pas de grand changement par rapport au raffinement. Pourtant quand on regarde les scores évalués par galaxy, le modèle est raffiné, les clash stériques diminuent énormément aussi comme les autres modèles, on passe de 118.7 à 15.5 pour le modèle 1, de même le score de Ramachandran augmente mais pas beaucoup (Tableau 3).

6) Évaluation des modèles:

Pour l'évaluation des modèles on a utilisé verify3D[9]. Cet outil détermine la compatibilité d'un modèle atomique (3D) avec sa propre séquence d'acides aminés (1D) en assignant une classe structurale basée sur son emplacement et son environnement (alpha, bêta, boucle, polaire, non polaire, etc.) et compare les résultats aux bonnes structures.

On va aussi évaluer les modèles avec l'outil PROSA. Cet outil est largement utilisé pour vérifier les modèles 3D de structures protéiques pour des erreurs potentielles sur la reconnaissance d'erreurs dans des structures déterminées expérimentalement et des modèles théoriques. Le score de qualité global calculé par ProSA pour une structure d'entrée spécifique est affiché dans un graphique qui montre les scores de toutes les chaînes de protéines déterminées expérimentalement actuellement disponibles dans la PDB. Cette caractéristique relie le score d'un modèle spécifique aux scores calculés à partir de toutes les structures expérimentales déposées dans PDB. Les parties problématiques d'un modèle sont identifiées par un graphique de scores de qualité locale et les mêmes scores sont mappés sur un affichage de la structure 3D à l'aide de codes de couleurs. Étant donné que notre protéine cible est une protéine membranaire, on a aussi évalué nos modèles par OREMPRO. Cet outil combine l'algorithme ANVIL pour positionner les structures protéiques dans la bicouche lipidique et le potentiel statistique MAIDEN pour l'évaluation structurale des domaines transmembranaires protéiques. Les deux méthodes sont basées sur les coordonnées uniques des alpha-carbones, ce qui permet à OREMPRO de traiter des modèles à haute et basse résolution.

a) Modèle généré par MODELLER:

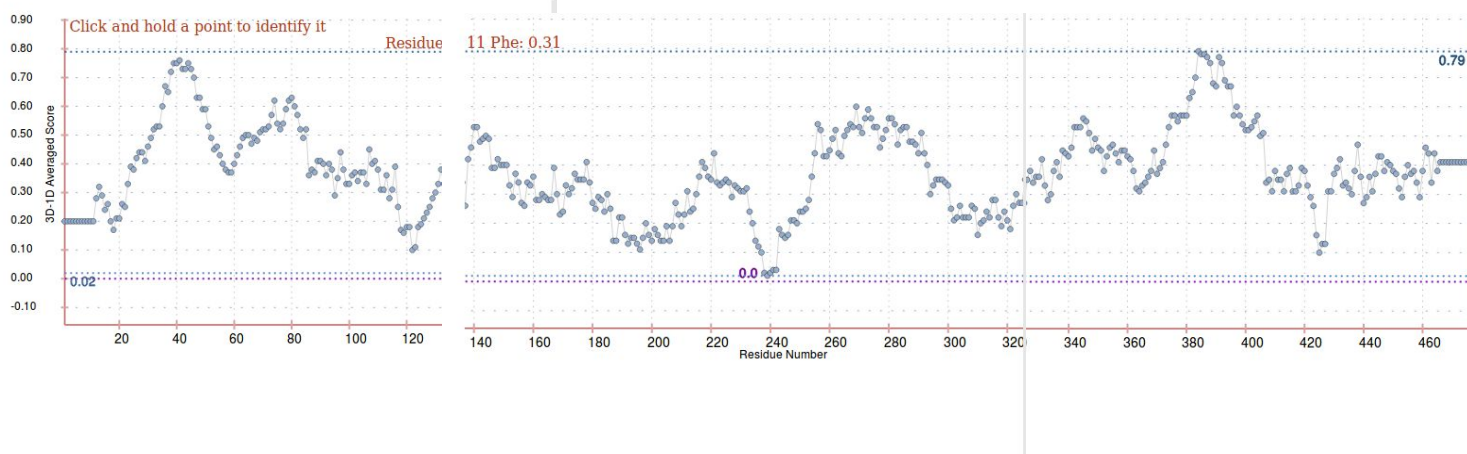


Figure 10: Évaluation du modèle généré par MODELLER par verify3D

90,15% des résidus avaient un score 3D-1D moyen $\geq 0,2$. toutes les résidus ont un score supérieur à 0,2. On peut dire donc qu'on a un bon modèle selon l'évaluation de verify3D. On aussi évaluer ce même modèle avec PROSA.

Overall model quality

[Help](#)

Z-Score: **-6.7**

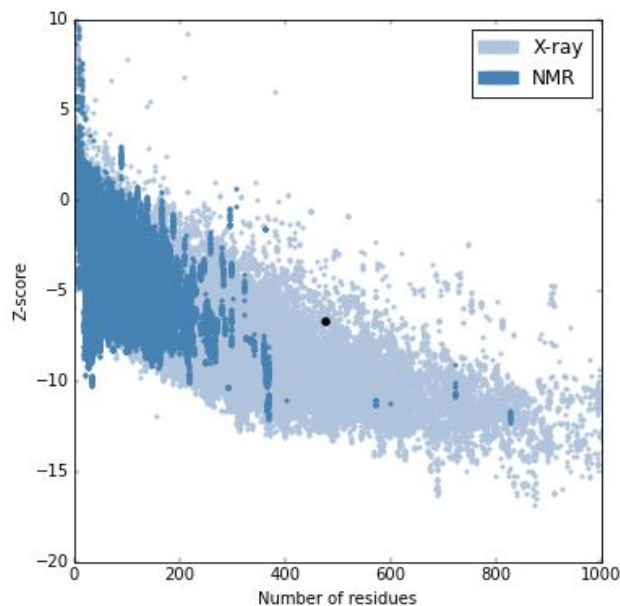


Figure 11: Évaluation du modèle généré par MODELLER avec l'outil PROSA

PROSA indique la qualité du PDB généré, ici le modèle(indiqué par le point noir sur la figure 11) est résolu par X-ray avec un Z-score de -6.7. Plus le Z-score est faible plus le modèle est meilleur. On peut en déduire que c'est un modèle ce qui est généré par MODELLER.

Input: model_1.pdb

Membrane assignment:

Hydrophobic thickness = 31.5 Å

Output PDB file [Download](#)

Number of TM segments = 12

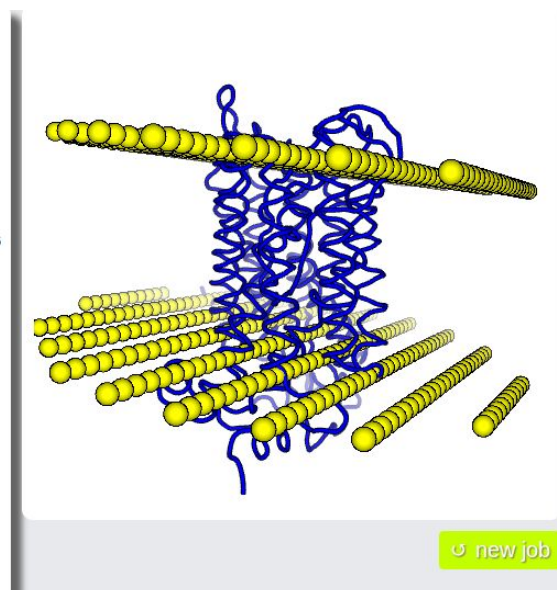
Sequence positions:
: 27-51|67-87|99-117|119-142|156-177|181-201|258-280|292-313|316-337|375-396|411-432|439-459

Tilt angles:

- Segment: 27-51 - Tilt: 34.0°
- Segment: 67-87 - Tilt: 21.2°
- Segment: 99-117 - Tilt: 12.8°
- Segment: 119-142 - Tilt: 34.5°
- Segment: 156-177 - Tilt: 16.1°
- Segment: 181-201 - Tilt: 22.3°
- Segment: 258-280 - Tilt: 34.1°
- Segment: 292-313 - Tilt: 24.8°
- Segment: 316-337 - Tilt: 25.6°
- Segment: 375-396 - Tilt: 16.0°
- Segment: 411-432 - Tilt: 15.6°
- Segment: 439-459 - Tilt: 10.8°

Structural assessment:

MAIDEN score = **137.7016**



[new job](#)

L'évaluation par OREMPRO nous indique que GLUT-8 possède 12 segments transmembranaires avec un bon score de MAIDEN de -137.

b) Modèle généré par MEDELLER:

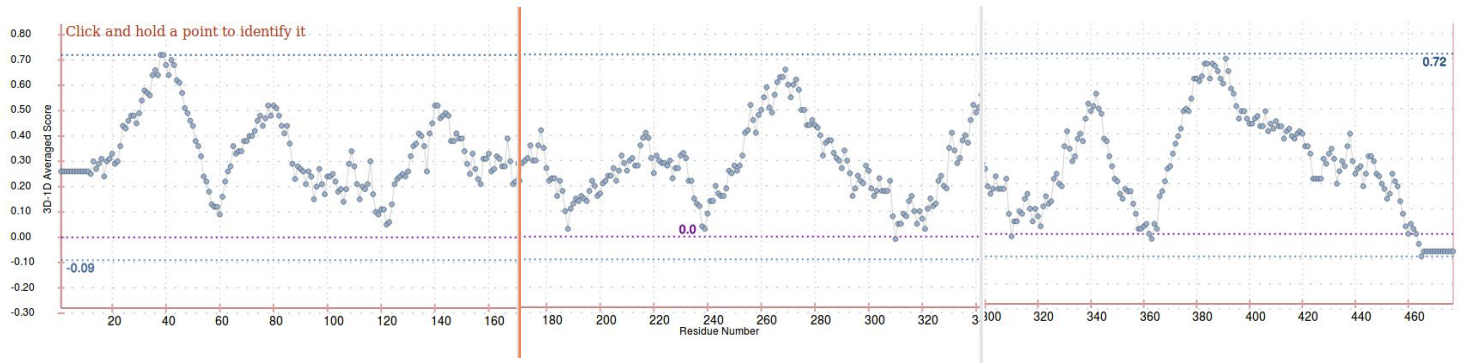


Figure 12: Évaluation du modèle généré par MEDELLER par verify3D

75.47% des résidus avaient un score 3D-1D moyen $\geq 0,2$. toutes les résidus ont un score supérieur à 0,2. On peut dire que le modèle généré par MEDELLER est meilleur mais pas autant que MODELLER selon l'évaluation de verify3D.

Z-Score: **-6.06**

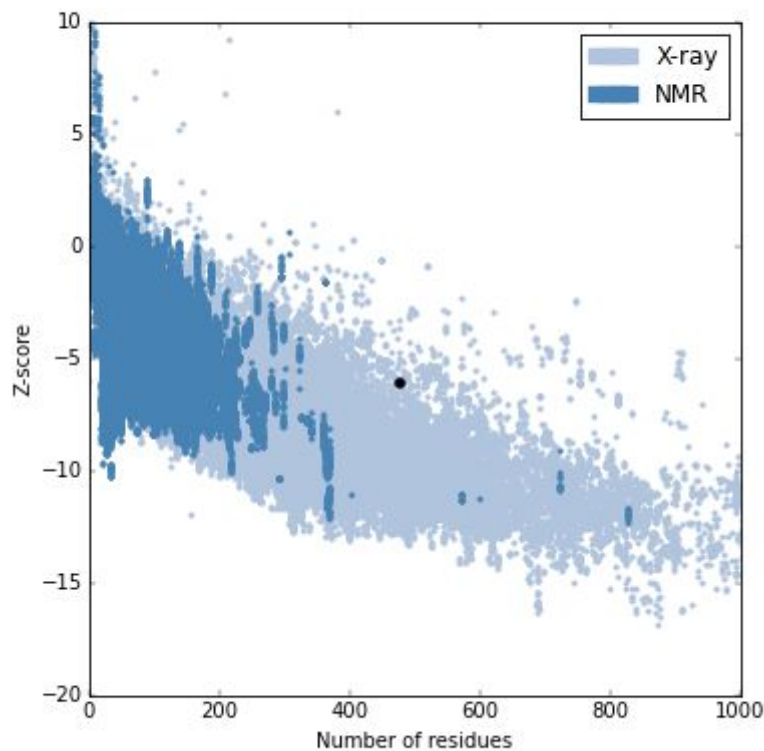


Figure 13: Evaluation du modèle généré par MEDELLER avec l'outil PROSA

Le modèle (indiqué par le point noir sur la figure 13) est résolu par X-ray avec un Z-score de -6.06. Plus le Z-score est faible plus le modèle est meilleur. On peut en déduire que c'est un modèle ce qui est généré par MEDELLER selon l'évaluation de PROSA.

Input: model_1.pdb

Membrane assignment:

Hydrophobic thickness = 33.9 Å

Output PDB file [Download](#)

Number of TM segments = 12

Sequence positions:

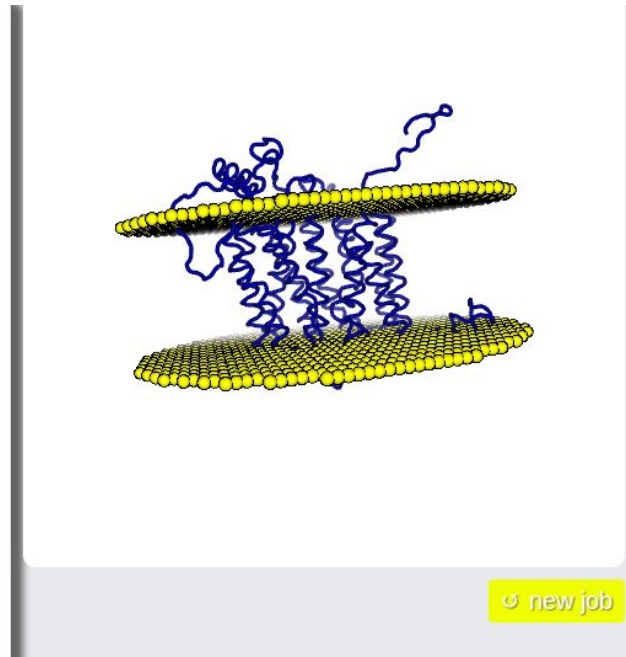
A: 27-47|67-91|96-117|120-145|156-179|181-202|257-283|292-313|318-340|372-396|410-432|434-459

Tilt angles:

A - Segment: 27-47 - Tilt: 32.8°
A - Segment: 67-91 - Tilt: 28.2°
A - Segment: 96-117 - Tilt: 10.4°
A - Segment: 120-145 - Tilt: 29.7°
A - Segment: 156-179 - Tilt: 17.9°
A - Segment: 181-202 - Tilt: 21.9°
A - Segment: 257-283 - Tilt: 28.4°
A - Segment: 292-313 - Tilt: 15.6°
A - Segment: 318-340 - Tilt: 9.5°
A - Segment: 372-396 - Tilt: 25.0°
A - Segment: 410-432 - Tilt: 20.4°
A - Segment: 434-459 - Tilt: 13.7°

Structural assessment:

MAIDEN score = [133.3553](#) ?



L'évaluation par OREMPRO nous indique que GLUT-8 possède 12 segments transmembranaires avec un bon score de MAIDEN de -133 pour le modèle généré par MEDELLER.

c) Modèle généré par Raptor-X:

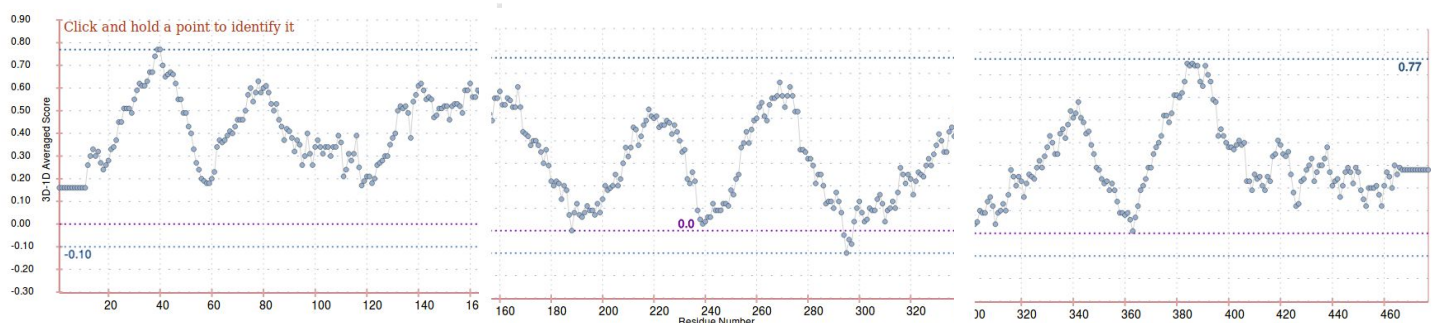


Figure 14: Évaluation du modèle généré par Raptor-X par verify3D

78.41% des résidus avaient un score 3D-1D moyen $\geq 0,2$. toutes les résidus ont un score supérieur à 0,2. On peut dire que le modèle généré par Raptor-X est meilleur mais pas autant que MODELLER selon l'évaluation de verify3D.

Z-Score: **-6.02**

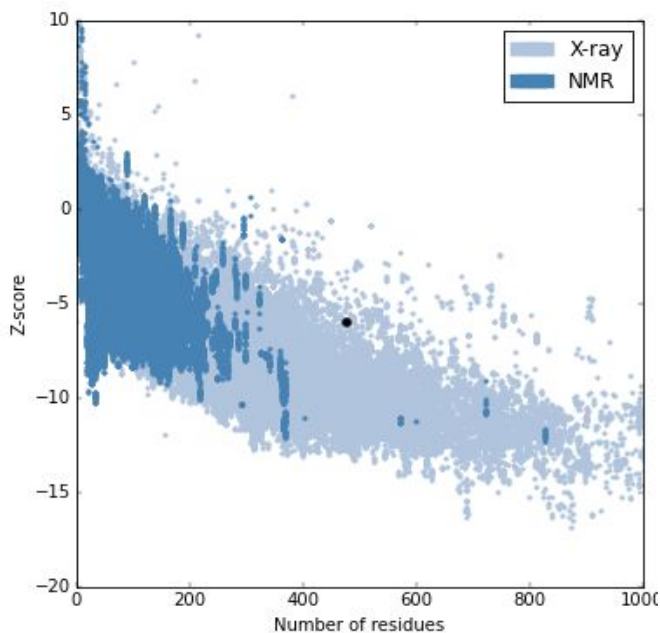


Figure 15: Evaluation du modèle généré par raptor-X avec l'outil PROSA

Le modèle (indiqué par le point noir sur la figure 13) est résolu par X-ray avec un Z-score de -6.06. Plus le Z-score est faible plus le modèle est meilleur. On peut en déduire que c'est un modèle ce qui est généré par Raptor-X selon l'évaluation de PROSA.

Input: model_1.pdb

Membrane assignment:

Hydrophobic thickness = 31.8 Å

Output PDB file [Download](#)

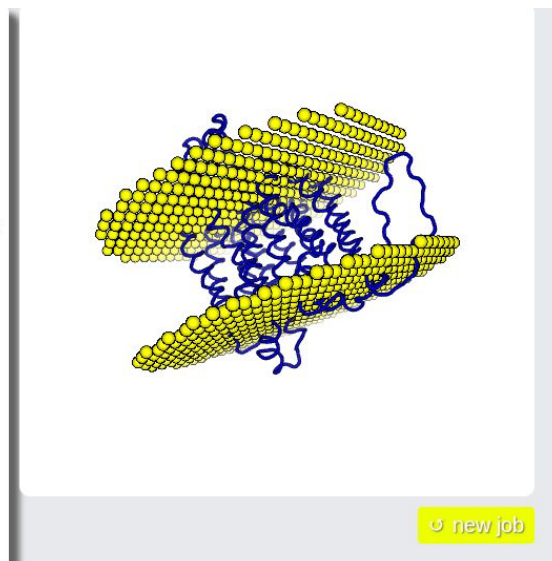
Number of TM segments = 12

Sequence positions:
: 30-50|70-91|96-116|121-143|158-178|181-200|257-280|290-312|318-340|370-397
|407-429|438-459

Tilt angles:
- Segment: 30-50 - Tilt: 3.4°
- Segment: 70-91 - Tilt: 18.0°
- Segment: 96-116 - Tilt: 17.4°
- Segment: 121-143 - Tilt: 22.4°
- Segment: 158-178 - Tilt: 26.2°
- Segment: 181-200 - Tilt: 16.6°
- Segment: 257-280 - Tilt: 38.8°
- Segment: 290-312 - Tilt: 9.4°
- Segment: 318-340 - Tilt: 16.1°
- Segment: 370-397 - Tilt: 43.8°
- Segment: 407-429 - Tilt: 25.8°
- Segment: 438-459 - Tilt: 16.8°

Structural assessment:

MAIDEN score = **-140.4114** ?



L'évaluation par OREMPRO nous indique que GLUT-8 possède 12 segments transmembranaires avec un bon score de MAIDEN de -140 pour le modèle généré par Raptor-X.

d) Représentation du modèle généré par modeller dans la membrane:

Pour ce faire on a utilisé l'outil PPM-server[11] qui calcule les positions de rotation et de translation des protéines transmembranaires et périphériques dans les membranes en utilisant leur structure 3D (fichier de coordonnées PDB) comme entrée.

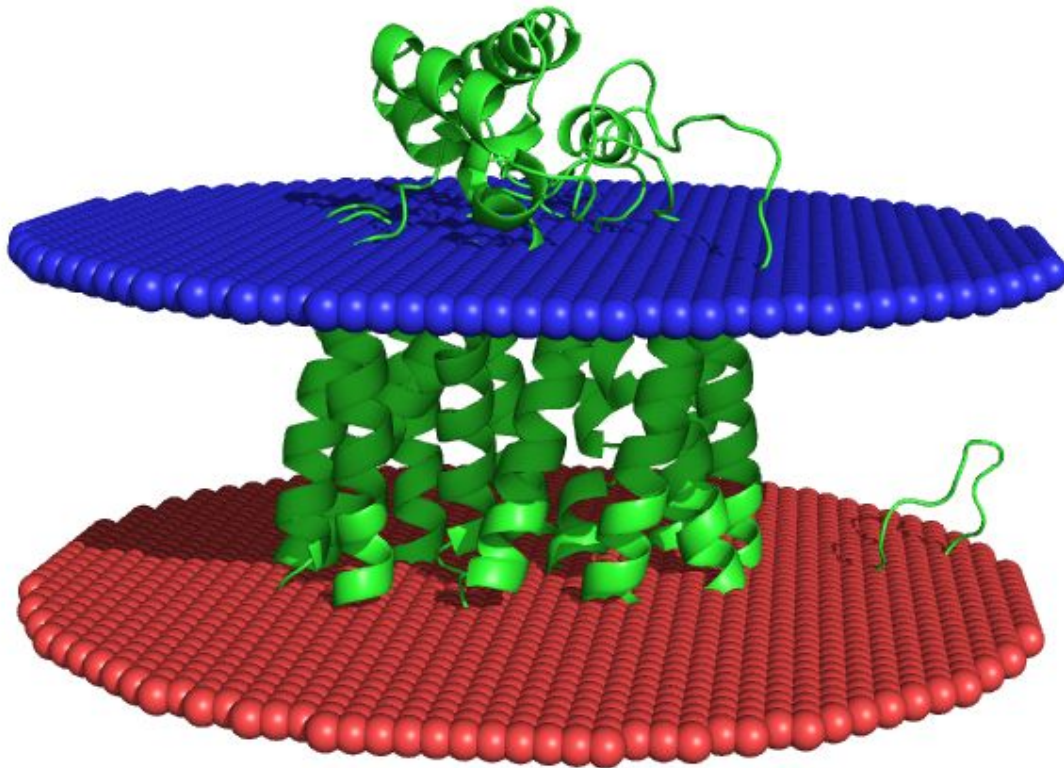


Figure 16: Représentation du modèle généré par modeller dans la membrane avec l'outil PPM server.

II. Étude de la conformation

Les modes normaux[10]:

Les protéines sont des objets dynamiques; leurs fluctuations structurales sont souvent la clé de leur fonction et une attention particulière doit donc leur être accordée. Les simulations de dynamique moléculaire (MD) sont une technique largement utilisée pour étudier la dynamique des protéines et des acides nucléiques. Cependant, la simulation de machines moléculaires contenant plusieurs milliers d'acides aminés, sur des échelles de temps suffisamment longues pour observer une déformation structurale pertinente, reste difficile. L'analyse en mode normal (NMA) est mieux adaptée pour étudier la dynamique lente des protéines.

L'analyse en mode normal fournit des informations sur les modes d'équilibre accessibles à un système, en supposant que le système est stabilisé par des potentiels harmoniques. Il a été utilisé pendant plusieurs décennies dans les phénomènes physiques classiques tels que les spectres vibrationnels atomiques et le transport à l'état solide. Cependant, seulement dans la dernière décennie est devenu un outil largement utilisé pour explorer les

mouvements fonctionnels. Une raison majeure derrière son utilisation générale est l'observation que les modes globaux élucidés par NMA portent une signification fonctionnelle. Cette caractéristique est devenue plus évidente avec l'utilisation de modèles simplifiés à gros grains (CG).

D'un point de vue physique, les modes globaux représentent simplement des reconfigurations le long des directions (axes principaux) les plus facilement accessibles (sur le paysage énergétique multidimensionnel). La fréquence de mode (au carré) fournit une mesure de la courbure (ou de la rigidité) du paysage énergétique avec la direction du mode donné, les modes de fréquence inférieure étant des mouvements plus doux. Une caractéristique de NMA est la robustesse observée des modes globaux de détail dans les interactions atomiques ou interatomiques spécifiques. Leur insensibilité aux interactions locales ou aux fonctions et paramètres énergétiques spécifiques qui définissent le champ de force résulte vraisemblablement de leur nature systémique.

On a étudié les modes normaux de nos modèles ce qui nous a permis de visualiser les mouvements de la protéine GLUT-8 ainsi que l'ouverture et la fermeture des canaux.

Conclusion et Perspective:

La protéine GLUT-8 est une protéine transmembranaire de type transporteur de glucose qui est un membre de la famille des facilitateurs du transport du sucre qui ne possède pas encore de structure PDB. Ce projet a permis de générer des modèles de structure 3D pour cette protéine en utilisant différents outils de modélisation par homologie ainsi que de la modélisation par threading. Tous les modèles sont en général bons mais il existe une différence au niveau des scores selon les méthodes. On identifie bien les résidus importants tels que la dileucine 12-13, ainsi que le tryptophane 394.

Les 12 protéines transmembranaires prédites dans la structure secondaire sont bien dans les modèles.

Les modes normaux ont permis de visualiser les mouvements de la protéine ce qui peut influencer sa fonction selon les mouvements. On a aussi pu voir l'ouverture et la fermeture des canaux. Cependant une simulation de la dynamique moléculaire (même en gros grain) des modèles est nécessaire pour mieux étudier la conformation de la protéine GLUT-8. Un étape de docking est aussi intéressante à faire pour voir si la fixation d'un ligand peut altérer la structure de la protéine.

Bibliographie:

GLUT8: [1] :GLUT8, a Novel Member of the Sugar Transport Facilitator Family with Glucose Transport Activity, Holger Doege, and al., February 2000.

GLUT8: [2] Mammalian Sugar Transporters Robert Augustin and Eric Mayoux, June 18, 2014

MUSCLE[3]: multiple sequence alignment with high accuracy and high throughput, Edgar, R.C. *Nucleic Acids Res.* 2004

MUSCLE[4]: a multiple sequence alignment method with reduced time and space complexity *BMC Bioinformatics*

PSIPRED[5] : Buchan, D.W. , Jones, D.T. (2017) EigenTHREADER: Analogous protein fold recognition by efficient contact map threading.

MODELLER[6] :The MPI bioinformatics Toolkit as an integrative platform for advanced protein sequence and structure analysis. Alva V, Nam SZ, Söding J, Lupas AN. *Nucleic Acids Res.* 2016 Jul 8;44(W1):W410-5

MODELLER[7]: Comparative Protein Structure Modeling Using MODELLER.

Webb B, Sali A. *Curr Protoc Protein Sci.* 2016 Nov 1;86:2.9.1-2.9.37.

RAPTOR-X[8]: Morten Källberg, Haipeng Wang, Sheng Wang, Jian Peng, Zhiyong Wang, Hui Lu & Jinbo Xu. **Template-based protein structure modeling using the RaptorX web server.** *Nature Protocols* 7, 1511–1522, 2012.

Verify3D[9]: A method to identify protein sequences that fold into a known three-dimensional structure. Bowie JU¹, Lüthy R, Eisenberg D. *Science* 12 Jul 1991

Modes normaux[10]: Normal Mode Analysis of Biomolecular Structures: Functional Mechanisms of Membrane Proteins *Chem Rev.* 2010 Mar 10; 110(3): 1463–1497. Ivet Bahar,^{*} Timothy R. Lezon,[†] Ahmet Bakan,[†] and Indira H. Shrivastava

PPM server[11]: Lomize M.A., Pogozheva I.D, Joo H., Mosberg H.I., Lomize A.L. OPM database and PPM web server: resources for positioning of proteins in membranes. *Nucleic Acids Res.*, 2012, 40(Database issue):D370-6