

PROJET DE BIO-INFORMATIQUE STRUCTURALE

I. INTRODUCTION

Le transport de métabolites à travers la membrane plasmique des cellules est une fonction cruciale dans la croissance et le maintien de l'homéostasie. La membrane étant imperméable aux molécules hydrophiles, des protéines transmembranaires, appelées transporteurs, sont présentes dans la bicouche lipidique afin d'assurer l'entrée et la sortie de ces molécules et ions.

La *Major Facilitator Superfamily* (MFS) est un des plus grands groupes de transporteur secondaire actif présent dans tout le règne du vivant. Les MFS permettent le transport d'une grande variété de substrat comme des ions, des carbohydrates, des lipides, des acides aminés et peptides, ainsi que d'autre molécules. [1]

La famille MFS contient en particulier les protéines de la famille SLC22 qui sont des transporteurs de cations organiques (OCT), des transporteurs zwitterion / cation (OCTN) ou des transporteurs d'anions organiques (OAT). Ces transporteurs contiennent 12 domaines transmembranaires en hélices alpha prédits (aucune structure 3D connues) et une grande boucle extracellulaire entre les hélices 1 et 2 et entre les hélice 6 et 7. Ils participent à l'absorption ou à l'excrétion de médicaments, de xénobiotiques et de composés endogènes dans l'intestin, le foie ou les reins, et remplissent des fonctions homéostatiques dans le cerveau et le cœur. [2]

L'objectif de ce projet est d'utiliser un ensemble d'outils et de méthodes bio-informatique et de modélisation moléculaire afin d'étudier la structure du transporteur transmembranaire SLC22A18 (*Solute Carrier family 22 member 18*), appartenant à la famille des Organic Cation Transporter (ORCT).

La protéine SLC22A18 humaine (code Uniprot Q96BI1) de 424 acides aminés est localisée principalement à la surface apicale des tubules proximales des cellules rénales.

Elle pourrait agir comme un transporteur de cations organiques basés sur un mécanisme antiport à efflux de proton. Elle joue également un rôle dans l'excrétion de chloroquine et de composés apparentés à la quinidine dans les reins. [3]

Le gène codant pour cette protéine pourrait être impliqué dans certaine maladie comme le cancer du poumon ou la Rhabdomyosarcoma embryonnaire.

La première étape consistera à générer des modèles de structure 3D pour cette protéine en utilisant plusieurs outils de threading, et de choisir le meilleur modèle après évaluation des différents modèles, ainsi qu'à partir des informations obtenues dans la bibliographie et par des outils de prédiction de structures secondaires. Enfin, les modes normaux seront étudiés afin d'observer les mouvements d'ouverture et fermeture du transporteur.

II. GENERATION DES MODELES

II.1. Prédiction des structures secondaires.

Une prédiction préalable des structures secondaire et de la topologie de la protéine permet d'avoir une idée globale du repliement 3D attendu. Trois méthodes différentes ont été utilisées afin de prédire les structures secondaires et notamment les hélices transmembranaires. Psi-Pred v3.3 basé sur un Psi-blast et une méthode de machine learning avec un réseau de neurone artificiel. TMHMM est une méthode de prédiction de la topologie des protéines membranaires basée sur un modèle de Markov caché et enfin MEMSAT-SVM qui est aussi un programme prédisant les structures secondaires et la topologie des protéines membranaires basé sur la méthode de support vector machines. Les régions en hélices alpha transmembranaires obtenues avec les 3 méthodes sont présentées dans le tableau 1.

Hélice TM	Psi-pred	TMHMM	MEMSAT-SVM
1	26-55	24-46	29-48
2	59-76	61-83	60-82
3	90-110	90-112	91-109
4	128-142	153-175	116-135
5	183-202	182-204	155-174
6	243-271	243-265	182-201
7	277-299	277-296	240-262
8	306-356	306-325	277-296
9	363-392	332-354	307-322
10	397-414	390-412	326-349
11	-	-	371-388
12	-	-	392-412

Tableau 1 : Prédiction des régions transmembranaire en hélice alpha obtenues avec les outils Psi-pred, TMHMM et MEMSAT-SVM.

Les 3 outils permettent d'obtenir globalement les mêmes résultats. 10 hélices alpha transmembranaires sont prédites avec Psi-Pred et TMHMM et 12 avec MEMSAT-SVM. La prédiction de MEMSAT-SVM est en accord avec la littérature qui indique que les protéines de la famille SLC22 possèdent 12 hélices transmembranaires [2]. Ces résultats permettront d'apporter des informations supplémentaires lors du choix du modèle et seront comparés aux hélices transmembranaires observées à partir de ce modèle.

II.2. Recherche de protéines homologues.

L'exécution d'un BLAST à partir de la séquence Fasta de la protéine SLC22A18 permet la recherche de protéines homologues pouvant servir de support pour la construction d'un modèle 3D par la méthode de modélisation comparative par homologie.

Le BAST sur UniprotKB_PDB avec structure 3D connues est effectué avec les paramètres suivants : blastp, matrice blosum62, seuil de e-value à 10, avec gap (ouverture : 10, extension : 1).

11 hits sont obtenus dont 2 sont des transporteurs d'E. coli avec un pourcentage d'identité de 22.5% et une e-value de 3.9e-6 : (1) C8U2B1_ECO10 - Predicted transporter, (2) A0A140NCX4_ECOBD - Major facilitator superfamily MFS_1.

Le pourcentage d'identité de seulement 22.5% et le recouvrement peu correct ne permettent pas d'utiliser une de ces 2 séquences comme support afin de réaliser un modèle par homologie.

II.3. Construction d'un modèle par la méthode du threading.

Les méthodes de threading sont des méthodes de reconnaissance de repliement basée sur le fait que les structures 3D sont mieux conservées que la séquence primaire d'une protéine. Ainsi, la structure est un bon moyen de caractériser des protéine homologue distante ayant un faible pourcentage d'identité.

Les méthodes de threading cherchent donc à mesurer l'adéquation entre une séquence donnée dont la structure 3D est inconnue et une autre structure connue. La séquence est « enfilée » sur chaque structure d'une base de donnée non redondante basée sur la PDB et la compatibilité entre la séquence et chaque structure est calculée. La séquence est optimisée par des insertions/délétions dans les boucles. Enfin, les repliements sont classés par ordre d'énergie, de la plus probable à la moins probable.

Trois méthodes de threading ont été utilisées, HHpred, RaptorX et MUSTER. Les modèles construits à partir des meilleurs templates sont réalisés en ligne avec Modeller directement par les outils. Les meilleurs modèles obtenus par chaque méthode seront raffinés puis évalués.

II.3.a HHpred

HHpred est exécuté avec les paramètres par défaut. Le meilleur modèle est obtenu avec la structure PDB 1PW4 (e-value 1.2E-31) qui est une structure de cristallographie aux rayons-X à 3.3 Å de la protéine transmembranaire Glycerol-3-Phosphate Transporter de E.Coli et de longueur 451 correspondant à une conformation ouverte vers l'espace intracellulaire.

Le modèle généré possède 12 hélices alpha transmembranaires et une grande boucle intracellulaire en H6 et H7.

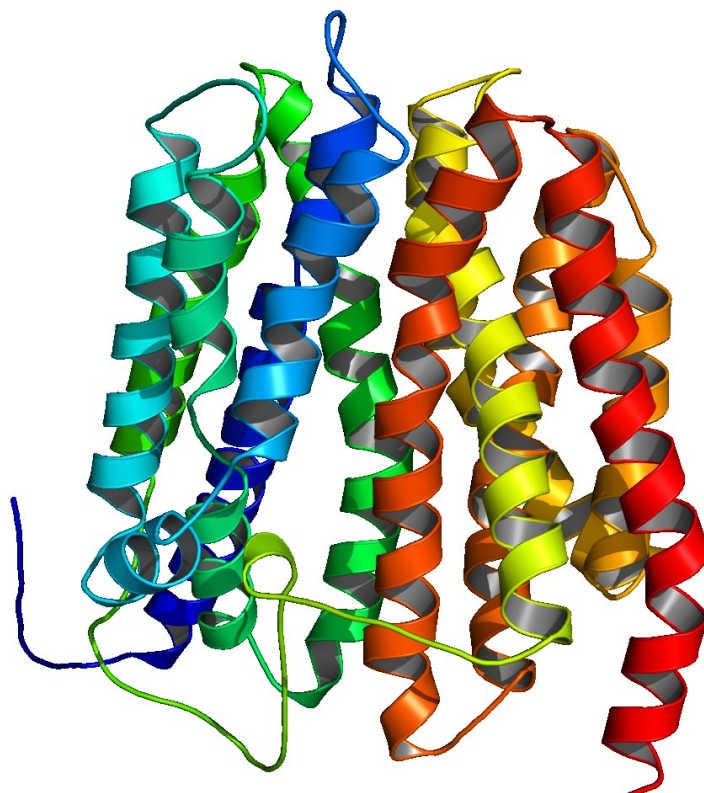


Figure 1 : Modèle de prédiction de repliement par threading du transporteur SLC22A18 obtenu avec l'outil HHpred et le template 1PWA après raffinement des boucles et des chaînes latérales avec GalaxyLoop et GalaxyRefine.

II.3.b MUSTER

MUSTER est exécuté avec les paramètres par défaut. Le meilleur modèle est obtenu avec la structure PDB 3WDO qui est une structure de cristallographie aux rayons-X à 3.15Å du transporteur transmembranaire YajR de E.Coli dans la conformation *outward* et de longueur 453. Cette structure appartient à la protéine obtenue en deuxième hit lors du BLAST et possédant un pourcentage d'identité de 22.5%. Il aurait donc été possible d'effectuer un modèle par homologie avec cette structure comme support.

Le modèle généré possède 12 hélices alpha transmembranaires et une grande boucle intracellulaire en H6 et H7.

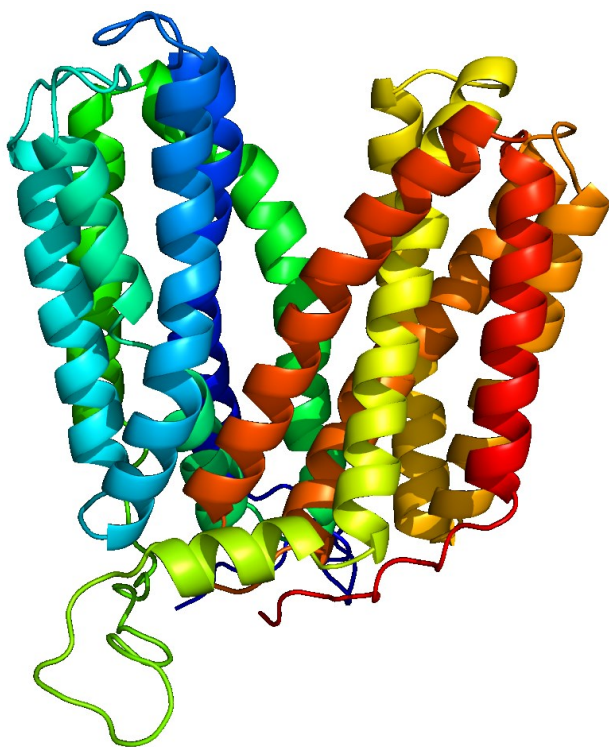


Figure 2 : Modèle de prédiction de repliement par threading du transporteur SLC22A18 obtenu avec l'outil MUSTER et le template 3WDO après raffinement des boucles et des chaînes latérales avec GalaxyLoop et GalaxyRefine.

II.3.b RaptorX

RaptorX est exécuté avec les paramètres par défaut. Le meilleur modèle est obtenu, comme pour MUSTER, avec la structure PDB 3WDO. La topologie est similaire, à la différence que pour ce modèle, la conformation ne semble pas ouverte vers l'espace extracellulaire comme pour le modèle généré par MUSTER bien que le même template soit utilisé.

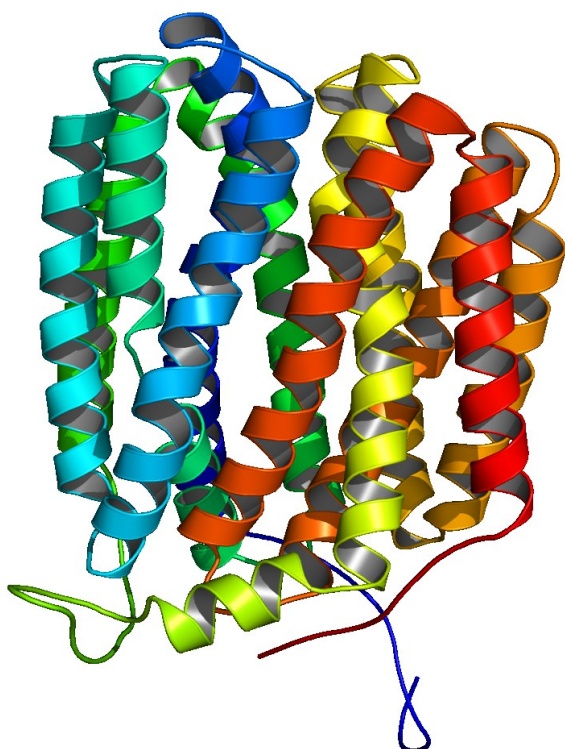


Figure 3 : Modèle de prédiction de repliement par threading du transporteur SLC22A18 obtenu avec l'outil RaptorX et le template 3WDO après raffinement des boucles et des chaînes latérales avec GalaxyLoop et GalaxyRefine.

II.3.c Raffinement des boucles et des chaînes latérales.

Pour chacun des 3 modèles, les régions de boucles à raffiner ont été déterminées avec l'outil d'évaluation de la qualité local Verify3D. Les régions ayant un mauvais score de qualité (inférieur à 0) correspondent à des acides aminés dans des boucles ou aux extrémités des hélices alpha. Le raffinement est effectué sur le serveur web GalaxyLoop. Seule une section de moins de 20 acides aminés peuvent être raffinées. Le meilleur modèle de boucles est choisi selon une évaluation avec Verify3D. Ensuite, les chaînes latérales sont raffinées en utilisant Galaxy Refine. Le modèle minimisant les clashes stériques sera sélectionné.

II.3.d Evaluation des modèles.

Les 3 modèles finaux obtenus (figure 1,2,3) sont comparés avec l'outil d'alignement structuraux TM-align et les TM-score et RMSD (tableau 2) montre que les repliements globaux sont très similaires, en particulier pour les modèles RaptorX et MUSTER qui utilisent tous les 2 le template 3WDO. Les différences observables se situent au niveau des régions C-ter et N-ter et de la grande boucle entre H6 et H7. La topologie des modèles correspond à celle prédite par MEMSAT-SVM (figure 4) avec 12 hélices alpha transmembranaires.

	MUSTER	RaptorX
HHpred	0.61 / 5.75	0.7 / 4.51
MUSTER	-	0.82 / 3.46

Tableau 2 : TM-score / RMSD (Å) entre les modèles obtenus selon les 3 méthodes calculées avec TM-align.

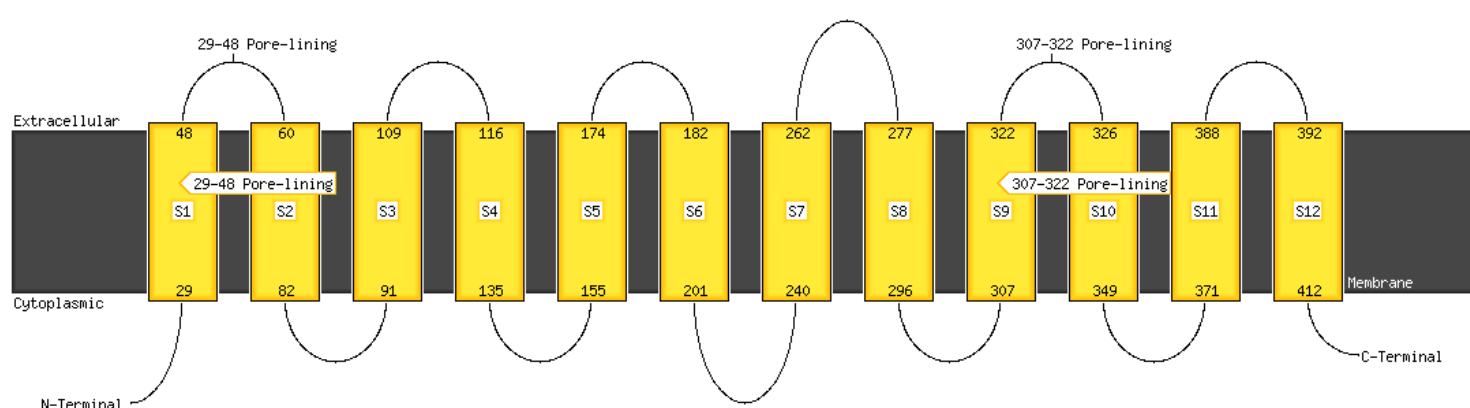


Figure 4 : Topologie du transporteur SLC22A18 prédite avec MEMSAT-SVM montrant 12 hélices alpha transmembranaires.

Il est indiqué dans la littérature la présence de 2 grandes boucles, une extracellulaire entre H1 et H2 et une intracellulaire entre H6 et H7. La première n'est pas observée sur les 3 modèles, due au choix du tamplate (3WDO et 1PWA) qui ne possèdent pas cette

boucle [4][5]. En revanche la région en coil N-ter est de longueur importante (de l'ordre de 20 acides aminés), ce qui peut montrer une mauvaise modélisation de cette région et un décalage de l'hélice H1 au niveau de la séquence qui devrait peut-être se situer plus tôt dans la structure.

L'évaluation de la qualité local et global est réalisée avec : (1) Verify3D, qui détermine la compatibilité d'un modèle atomique (3D) avec sa propre séquence d'acides aminés (1D) en assignant une classe structurale basée sur sa localisation et son environnement (alpha, bêta, boucle, polaire, non polaire, etc.) et compare les résultats aux bonnes structures, (2) QmeanBrane, qui est une version de QMEAN développée pour évaluer la qualité locale des modèles de protéines membranaires et (3) Procheck, vérifie la qualité stéréochimique d'une structure protéique en analysant sa géométrie globale et résidu par résidu.

L'évaluation de la qualité locale avec Verify3D (tableau 3) montre que les structures locales sont de meilleure qualité dans le modèle généré par RaptorX, néanmoins les régions en plus de l'extrémité C-ter sont peu nombreuses, de quelques acides aminés et correspondent à des extrémités d'hélices alpha et à des boucles.

Modele	% de score 3D-1D > 0.2	Régions de mauvaise qualité
HHpred	70.28	183-191; 291-297; 398-424
MUSTER	73.82	302-308; 404-424
RaptorX	79.48	403-424

Tableau 3 : Evaluation local de qualité des modèles avec Verify3D.

Le calcul de l'énergie d'insertion de la membrane avec Qmean Brane ainsi que les Z-score calculer par rapport au Z-score attendus pour des structures de cristallographie de longueur équivalente montrent que les 3 modèles sont de bonnes qualités (figure 5).

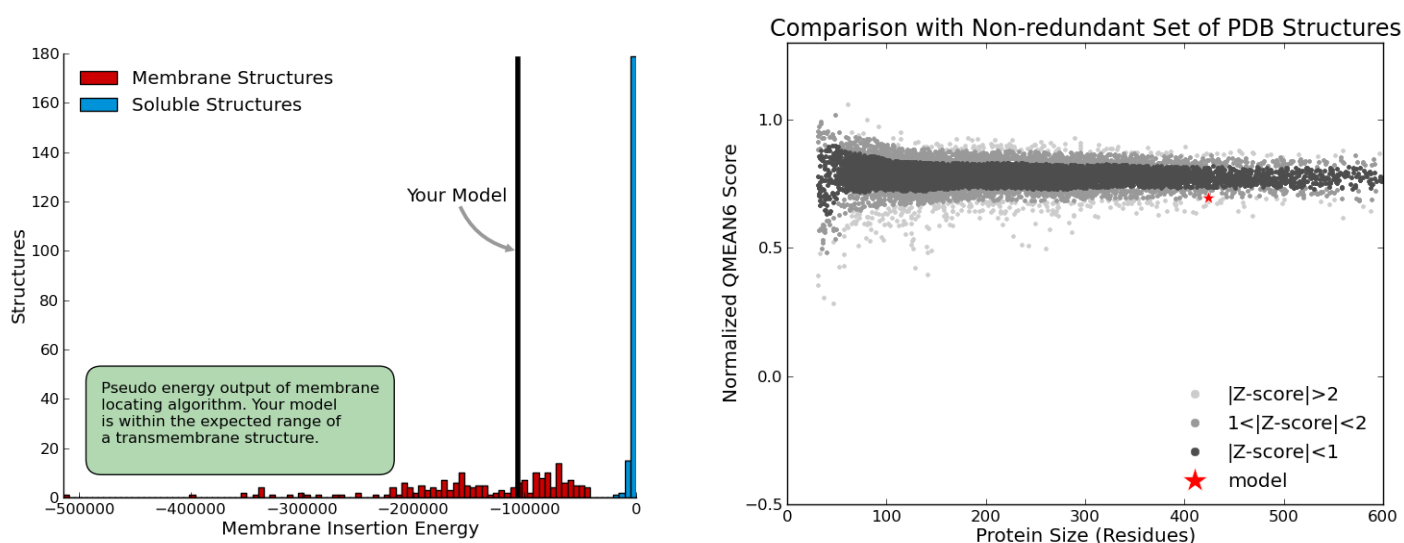


Figure 5 : Exemple de la mesure de la qualité du modèle générer par RaptorX avec QmeanBrane. A gauche l'énergie d'insertion de la membrane et à droite le Z-score.

III. ETUDES DES MODES NORMAUX

L'analyse en mode normal est une technique classique pour générer des mouvements fonctionnels dans les macromolécules biologiques. Les mouvements de basses fréquences des protéines, tels qu'on peut les calculer via l'approximation harmonique, ressemblent souvent à des mouvements fonctionnels.

Les modes normaux sont obtenus avec le server Nomad-ref avec les paramètres par défaut : tout atomes, 16 modes, paramètre de poids de distance pour la constante élastique 5 Å, Cutoff ENM à utiliser pour le calcul du mode 10 Å et RMSD moyen pour les trajectoire de sortie 1 Å.

Avec le modèle de RaptorX, correspondant à une conformation fermée, les mouvements globaux de pivots et de rotations sont observés pour les 6 premiers modes, puis les mouvements locaux de la région N-ter en coil et de la boucle entre H6 et H7. Il n'y a pas de mouvement d'ouverture du port observé.

Avec le modèle de MUSTER, les résultats sont similaires, mais une ouverture plus importante du port est observée avec les mode 14 à 16. Le modèle est une forme ouverte vers l'extracellulaire, il n'y a pas de mouvement de fermeture ou d'ouverture du coté intracellulaire observé.

Avec le modèle de HHpred, une très petite fermeture du port est observée avec les modes 7 à 9.

L'études des modes normaux à partir de ces modèles ne permettent pas d'observer un changement de conformation (ouverture ou fermeture) du transporteur.

IV. CONCLUSION

Les protéines de la famille des SLC sont des transporteurs membranaires dont la structure 3D est peu connue. La méthode de threading avec MUSTER, HHpred et RaptorX ont permis de générer 3 modèles de bonnes qualités au vue de la littérature et de la prédiction des structures secondaires avec MEMSAT-SVM. Les 12 hélices transmembranaire présente dans les protéine de cette famille sont bien retrouvées ainsi que la grande boucle entre les hélice 6 et 7. Néanmoins, il n'a pas été possible d'observer les changement de conformation avec les modes normaux.

Références

- [1] Structural advances for the major facilitator superfamily (MFS) transporters, Nieng Yan et al., Cell Press, Volume 38, Issue 3, p151–159, 2013
- [2] The SLC22 drug transporter family, Hermann Koepsell et al., Eur J Physiol, 447:666–676, 2004
- [3] Functional characterization of ORCTL2- an organic cation transporter expressed in the renal proximal tubules, Mark Reece et al., FEBS Letters 433 p245-250, 1998
- [4] Structure and Mechanism of the Glycerol-3-Phosphate Transporter from Escherichia coli Yafei Huang, Science, vol. 301, No. 5633 pp. 616-620, 2003
- [5] Structure of the YajR transporter suggests a transport mechanism based on the conserved motif, A Jiang, D et al. 2013