

Projet de **bioinformatique structurale:** **la ferroportine**

Sarah Kaddah
M2 BI

Année 2017-2018

Sommaire

1. Introduction
2. Matériel et méthodes
3.
 1. Génération du modèle
 2. Étude de la conformation
 3. Ligand
4. Conclusion

1. Introduction

Les protéines *Major Facilitors Superfamily* (MFS) est la plus grande famille de transporteurs secondaires, conservées de la bactérie à l'homme. Elles jouent un rôle primordial dans plusieurs processus physiologiques, le métabolisme et dans l'homéostasie énergétique. Plus de 10 000 membres sont séquencés et il existe environ 76 sous-familles. Aujourd'hui, seules sept MFS provenant de 6 familles différentes ont des structures connues. Ces structures ont été élucidées lors de conformations différentes prises par la protéine au cours du transport de son ligand. Ces protéines effectuent un transport sélectif au travers de la membrane cellulaire. En effet, chaque transporteur possède des substrats spécifiques (des ions, des glucides) qui peuvent alors traverser la membrane lipidique. Les MFS utilisent un mécanisme de couplage. Certaines molécules usent de la diffusion facilitée, c'est-à-dire que la diffusion se fait grâce au gradient de concentration. Les transporteurs actifs secondaires exploitent le potentiel électrochimique des ions ou du soluté pour faire avancer le substrat contre le gradient. Dans cette catégorie, les symporteurs ou antiporteurs catalysent le mouvement des TMs de deux ou plusieurs substrats dans un sens unique ou contraire, respectivement. Un changement de fonction, d'expression ou de localisation de ce type de protéine est associé à de nombreuses maladies humaines.

Cette étude sur les MFS se concentre sur la ferroportine-1 aussi appelée en anglais *Solute carrier family member 1*, appartenant à la sous-famille des ferroportines. Son code Uniprot étant Q9NP59, elle est codée par le gène SLC40A1. Elle se trouve chez l'*Homo Sapiens*. Cette protéine est localisée dans la membrane basolatérale des cellules épithéliales polarisées. Elle est détectée dans les érythrocytes et elle est exprimée dans le placenta, les intestins, les muscles et dans la rate. Elle est impliquée dans l'export du fer des cellules épithéliales duodénales et dans le transfert du fer entre la circulation foeto-maternelle. Elle sert de médiateur en présence de ferroxidase, une enzyme impliquée dans le métabolisme du fer. La ferroportine-1 est donc impliquée dans le transport et l'absorption du fer.

Lorsqu'elle subit des mutations, la ferroportine-1 cause l'Hémochromatose 4 (HFE-4). La protéine présente une baisse de sa capacité à transporter le fer, ayant pour conséquence une surcharge en fer dans l'organisme. Cet excès aboutit à une défaillance des organes et apparaissent alors d'autres maladies telles la cirrhose, des hépatomes, le diabète, la cardiomyopathie, l'arthrose et l'hypogonadisme hypogonadotrope. Les effets sévères n'apparaissent qu'au bout de décennies, la surcharge en fer étant progressive. Par ailleurs, la ferroportine est un bon indicateur dans le pronostic du cancer du sein.

Les MFS seraient constituées de 12 hélices transmembranaires (TMs) au moins d'après les prédictions bioinformatiques. Elles sont divisées en deux domaines N et C de 6 TMs consécutifs. Elles partagent un repliement de structure, bien que la similarité de séquence soit faible, les spécificités de substrats différentes et un mécanisme de couplage différent. Les structures disponibles proviennent de bactéries, surtout celles qui sont significativement impliquée dans un processus physiologique ou pathologique. La ferroportine-1 est formée de 571 acides aminés, avec une histidine conservée en position 32. Elle est composée de 10 hélices membranaires a priori. Sa structure n'a pas encore été résolue. L'objectif de cette étude est de prédire un modèle de la structure, puis d'étudier sa conformation et enfin d'étudier ses ligands.

2. Matériel et méthodes

Blast (*Basic Local Alignment Search Tool*) Cette méthode heuristique permet de trouver des régions similaires entre deux ou plusieurs séquences d'acides nucléiques ou d'acides aminés (dans notre cas). Un alignement est fait entre régions homologues à travers des bases de données. Dans cette étude, BLAST est utilisé pour repérer des protéines, dont on connaît la structure, susceptibles de modéliser la ferroportine-1.

PSIPRED (*PSI-blast based secondary structure PREDiction*) Cette méthode permet de prédire les structures secondaires d'une protéine. L'algorithme utilisé est l'apprentissage par réseau de neurones. Les structures secondaires prédites sont les hélices alpha, les feuillets bêta et les boucles à partir de la séquence primaire.

PyMOL Ce logiciel permet de visualiser les structures chimiques en 3D. Il sera également utilisé pour analyser les mouvements de la protéine durant l'étude des modes normaux.

VMD (*Visual molecular dynamics*) Ce logiciel permet de visualiser et de modélisation moléculaire.

I-TASSER (*Iterative Threading ASSEmbly Refinement*) Cette méthode permet de prédire des modèles de structure tridimensionnelle à partir de la séquence primaire d'acides aminés. Elle utilise des modèles de la PDB et l'algorithme de Monte-Carlo. Elle est une des meilleures prédiction de structure de protéine dans les expériences CASP. Ce logiciel procède par 6 étapes successives. Il prédit les structures secondaires à travers PSSpred, un logiciel de prédiction de structures secondaires. Ensuite il utilise des modèles détectés par LOMETS. Ce méta-serveur de threading prédit de façon automatisée et rapide les structures 3D de protéine. Puis les fragments de structures sont assemblés avec la simulation de Monte-Carlo. Après, les modèles sont sélectionnés par clustering des structures grâce à SPICKER. Cette stratégie permet d'identifier des repliements natifs en regroupant les structures de protéines. Un affinement de la structure à l'échelle atomique est effectué par FG-MD (*fragment-guided molecular dynamics*) ou ModRefiner, un algorithme à l'échelle atomique utilisé pour un affinement de haute résolution. Pour finir, les fonctions biologiques basées sur la structure sont annotées par COACH, par reconnaissance de protéine-ligand à partir de comparaison de séquence et d'alignements de structures.

MUSTER Cet algorithme de Threading est basé sur la programmation dynamique et l'alignement de séquences. Il combine aussi de multiples ressources de structures pour assister l'alignement de séquence.

PHYRE2 Ce serveur de Threading combine HHsearch, un logiciel qui recherche des séquences protéiques, et des modèles multiples de modélisation. La première étape de ce logiciel est de rassembler des séquences homologues. Deuxièmement, le repliement est pris en charge à partir de

bases de données. Troisièmement, les boucles sont modélisées au mieux. Cette étape est prise en charge par Poing si le modèle est pris en charge de façon intensive. Pour finir, les chaînes latérales sont placées.

HHpred Ce serveur qui prédit les structures de protéines et leur fonction. Il utilise HHsearch, un logiciel pour la détection d'homologies basé sur une comparaison par paire de modèles cachés de Markov.

ModLoop Ce serveur permet de modéliser les boucles dans la structure d'une protéine. Le procédé est lié à la modélisation de boucle de MODELLER. La prédiction est faite à partir de contraintes spatiales et non à partir de bases de données. Pas plus de 20 acides aminés sont pris en charge à la fois.

Galaxy Refinement Ce serveur permet d'affiner les boucles d'une structure protéique.

SidePro et Scwrl4.0 Ces logiciels permettent de compléter les chaînes latérales dans une structure protéique.

Orempro Ce serveur combine l'algorithme ANVIL pour positionner la structure des protéines dans une bicouche lipidique. Il utilise aussi le potentiel statistique MAIDEN pour évaluer les domaines transmembranaires. Ces deux méthodes sont basées sur les coordonnées des carbones alpha.

TMHMM Cet outil utilise les chaînes cachées de Markov pour prédire les hélices transmembranaires des protéines membranaires.

Nomad-ref L'analyse de modes normaux est un moyen efficace d'étudier le mouvement global de biomolécules. Elle est plus rapide que des simulations moléculaires. Ce serveur fournit des outils en ligne pour calculer les modes normaux de grandes molécules, conserve la représentation complète avec tous les atomes, et accède à un plusieurs programmes qui utilisent les déformations et raffinement des structures. La stéréochimie est correcte, mais avec de larges amplitudes de mouvements et modélise une flexibilité ligand/récepteur.

Verify 3D Cette méthode évalue la structure d'une protéine en utilisant des profils tridimensionnels. Ce programme analyse la compatibilité d'un modèle atomique à trois dimensions avec sa séquence en acides aminés à une dimension. Chaque résidu a une structure assignée, basé sur la localisation et l'environnement. Ensuite, une base de donnée générée à partir de bonnes structures est utilisée pour obtenir un score pour chacun des acides aminés dans cette classe structurale. L'axe vertical du graphe obtenu représente la moyenne du profil 1D-3D pour chaque résidu. Le score le plus mauvais est -1 et le meilleur score est 1.

ProQM Ce logiciel évalue la structure 3D d'une protéine. Il prédit le S-score pour chaque résidu. Le score est une transformation du RMSD normal pour chaque résidu basé sur la formule $S_i = (1/\sqrt{1 + \text{RMSD}_i^2/5})$. off between getting a high sum of S-score and the length of the structural alignment).

3.1.PARTIE GÉNÉRATION DES MODÈLES

A) Construction du modèle

Étude de la séquence protéique (Q9NP59)

Dans un premier temps, les acides aminés conservés sont repérés à partir de la séquence protéique de la ferroportine-1 et des autres protéines MFS. Une fois leur séquence regroupées, un alignement multiple est effectué. Les acides aminés les plus conservés sont P231, F320, G330 et G347.

Recherche de protéines homologues

Dans un premier temps, une recherche permettant de trouver des protéines homologues est faite. La recherche est faite pour des protéines homologues dont la structure est connue. Cependant, la protéine la plus proche, étant aussi une MFS (*Solute Carrier Family 39*), a seulement 23,7% d'identité. N'étant pas suffisant pour faire un modèle par homologie de séquence, la prochaine étape est le *Threading*.

Prédiction de structures secondaires

D'après la prédiction faite, la ferroportine-1 est constituée de 3 petits feuillets, de boucles et majoritairement d'hélices alpha. Pour délimiter le nombres d'hélices, il faut un modèle de la protéine.

Threading

Le threading ou méthode d'enfilage permet la prédiction des structures 3D. Elle s'effectue lorsque la structure 3D d'une protéine n'est pas connue. En général, le pourcentage de similitude entre la séquence requête et la séquence cible est entre 20% et 40%. Les bases de données de séquences de structure connues sont interrogées à partir de la séquence requête. Cette méthode de modélisation est comparative. C'est une méthode de reconnaissance du repliement 3D-1D. Le repliement obtenu est correct d'un point de vue stérique et énergétique. Une fonction de score évalue la compatibilité entre la séquence requête (1D) et la structure de la protéine identifiée dans la base de données (3D). Cette fonction de score permet de déterminer le choix du support structural. Le résultat est une liste de protéines associé à un score. Les méthodes de repliement inverse permettent d'évaluer la compatibilité d'une structure donnée avec un jeu de séquences issues d'une base de données, c'est-à-dire prédire quelles séquences sont susceptibles de prédire tel type de repliement.

Threading avec MUSTER

Ce logiciel utilise MODELLER. Le meilleur modèle trouvé par MUSTER est effectué à partir du fichier PDB 5AYM. Cette protéine est le cristal d'un homologue bactérien transporteur de fer (ferroportine). Elle est dans une conformation tournée vers l'extérieur. La structure est obtenue par diffraction au rayon X, avec une résolution de 3.0 Å. Cependant, les boucles et les extrémités s'éloignent de la protéine. Le modèle contient les 571 protéines.

Threading avec PHYRE2

Le fichier PDB utilisé correspond à la protéine 5AYN. Cette protéine est un homologue bactérien transporteur de fer (ferroportine), soit une solute carrier family 39. La méthode utilisée la diffraction par rayon X avec une résolution de 2.2 Å. Sa conformation est tournée vers l'extérieur. Ce logiciel ne modélise pas les boucles, mais seulement les hélices alpha. Seul 70% du modèle (402 résidus) est représenté mais avec une forte confiance de modélisation(100%). 91% de la structure peut être modélisée mais de façon intensive. En effet, la structure obtenue avec modélisation intensive contient les 571 résidus et la boucle est dans le même cadre que la structure.

Ce logiciel prédit les extrémités N-ter et C-ter d'une part, et les segments transmembranaires d'autre part (voir figure 1). Le premier modèle obtenu présente des hélices alpha modélisées sans les boucles.

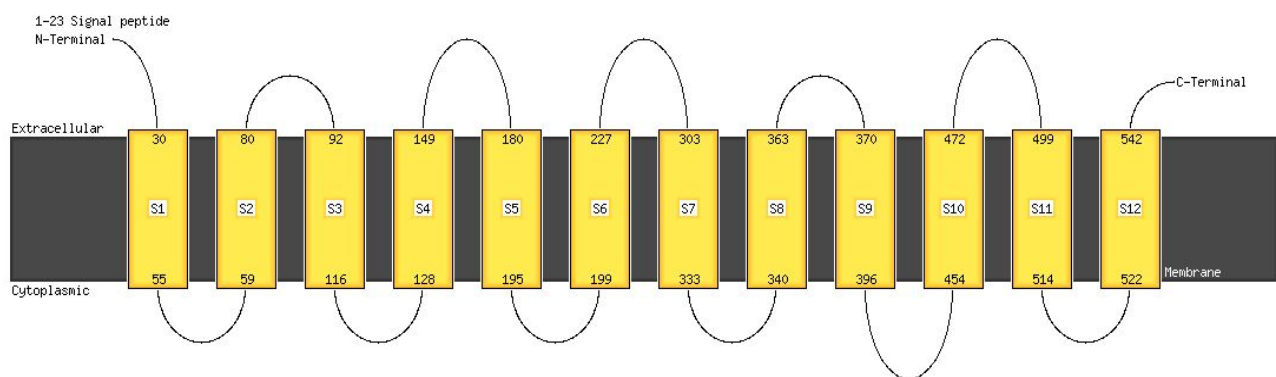


Figure 1: Prédiction des segments transmembranaires effectués par PHYRE 2.

Threading avec HHPRED

Le PDB utilisé pour le meilleur modèle est 5AYO. Cette protéine est un homologue bactérien transporteur de fer (ferroportine) dans un état tourné vers l'intérieur. Cette structure est obtenue par une diffraction au rayon X avec une résolution de 3.3 Å. Seuls 541 résidus sur 571 sont modélisés. Il manque 41 résidus. Une évaluation de la qualité du modèle est proposée (ANOLEA, SOLVX, VERIFY 3D). globalement, le modèle est acceptable, cependant certaines zones sont très mauvaises.

Threading avec I-TASSER

Le fichier PDB utilisé pour le meilleur modèle est 5AYM comme le logiciel MUSTER. Le modèle est complet et contient les 571 acides aminés.

Comparaison des différentes méthodes de threading:

Tous les modèles de la ferroportine-1 sont effectués à partir de la même protéine: un homologue bactérien transporteur de fer (ferroportine) ayant différents fichiers PDB. I-TASSER et MUSTER utilisent exactement le même fichier PDB. Malgré les différentes résolutions plus ou moins bonnes, les modèles les plus complets sont ceux produits par MUSTER, PHYRE2 et I-TASSER. En effet ils contiennent les 571 résidus qui composent la ferroportine-1. Cependant le modèle MUSTER nécessite un affinement des boucles, tandis que les deux autres modèles ont une structure regroupée.

B) Affinement du modèle

Affinement des petites boucles

Le modèle MUSTER est à affiner, surtout au niveau des grandes boucles qui dérivent loin des hélices. Les boucles sont réparties en deux catégories: les petites et les grandes boucles. Les outils Galaxy Refinement et ModLoop ne prennent pas plus de 20 acides aminés à affiner. Cela pose problème pour les grandes boucles. Les petites boucles peuvent être affinées. Avec ModLoop, les petites boucles sont affinées en plusieurs fois. Le logiciel Galaxy Refinement a un temps de calcul considérable et n'a pas amélioré la position des boucles.

| Boucles | Positions |
|-------------------|---|
| Petites (< 20 aa) | 1-23; 119-126;163-169; 195-204; 319-325 |
| Grandes (> 20 aa) | 233-305; 392-453; 544-571 |

Tableau 1: Position des petites et des grandes boucles à affiner.

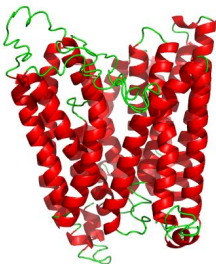
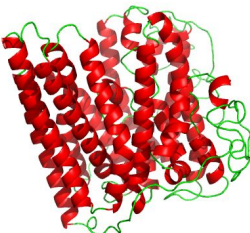

| | | |
|---|---|---|
|  |  |  |
| I-TASSER | PHYRE 2 | MUSTER |

Figure 2: Structure de la ferroportine-1 en fonction des différents logiciels de threading, et après affinement des petites boucles pour MUSTER.

Les grandes boucles sont plus difficiles à affiner. Une recherche sur la spécificité de ces boucles est faite dans le but de savoir si elle possède des propriétés ou une certaine fonction par exemple.

Choix du modèle après affinement des boucles

I-TASSER et PHYRE 2 ont des modèles déjà affinés car cette étape est comprise dans leur pipeline. Leur conformation diffère, probablement à cause du template utilisé pour chacun. Après avoir essayé d'améliorer les grandes boucles du modèle MUSTER, le résultat n'étant pas satisfaisant, le travail qui suit est effectué sur les modèles I-TASSER et PHYRE2.

Complétion des chaînes latérales

Le modèle PHYRE 2 a déjà ses chaînes latérales complétées. Les chaînes latérales sont donc complétées pour le modèle I-TASSER. Avec le logiciel Pro-side, le squelette de la protéine a légèrement dévié pendant la complétion des chaînes latérales. Au contraire, avec Scwrl4, la protéine n'a pas dévié et seules les chaînes latérales ont été complétées. Par conséquent, je conserve le modèle complété par Scwrl4.

Localisation de la membrane et des hélices transmembranaires

Le logiciel Orempred permet de visualiser la membrane (figure 3), repérer les hélices transmembranaires et donne un score au modèle. Plus le MAIDEN score est petit, meilleur est le modèle. Un code couleur est associé, du pire au meilleur score: rouge>jaune>vert>bleu. Le score associé au modèle I-TASSER amélioré est de -44,80 avec une couleur verte. Ce modèle est plutôt bon pour une première évaluation. La membrane a une épaisseur de 24,3 Å et 10 hélices alpha transmembranaires. Le score associé au modèle PHYRE 2 est de -27,90.

Une comparaison des positions des hélices alpha transmembranaires, avec logiciel TMHMM, permet d'avoir une autre approche. Certaines hélices trouvées sont semblables, mais d'autres hélices à différentes positions sont trouvées. Il faut noter que Orempro travaille à partir d'un fichier PDB et TMHMM à partir d'un fichier fasta. Par conséquent, je choisis de prendre en considération les hélices transmembranaires prédites par Orempro (figure 3).

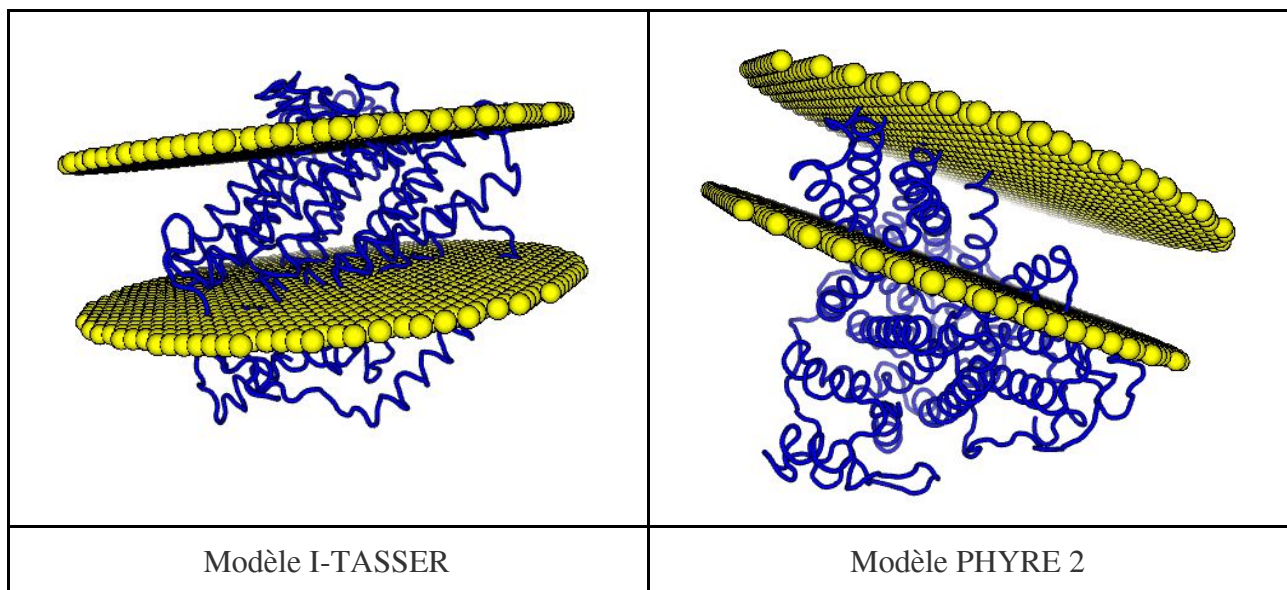


Figure 3: Représentation de la membrane et des hélices transmembranaires avec le logiciel Orempro.

C) Évaluation du modèle

Le modèle PHYRE 2 a 46.94% des résidus et moins de 65% des acides aminés ont une moyenne du score 3D-1D score supérieure ou égale à 0,2. Le modèle I-TASSER a 60.07% et moins de 65% des acides aminés ont une moyenne du score 3D-1D score supérieure ou égale à 0,2. Le modèle I-Tasser est meilleur que le modèle PHYRE 2 et sera utilisé pour le travail qui suit. En effet, ce modèle n'est pas excellent, mais plutôt acceptable. Le logiciel ProQM ne donne pas de différence significative au niveau du S-score pour des valeurs tournants autour de 0,525 en moyenne. Par contre cela confirme que le modèle est acceptable mais moyen.

D) Générer différents états conformationnels

Plusieurs tentatives de générer des états conformationnels différents avec Crankite et Open Babel, mais sans succès.

3.2. ÉTUDE DE LA CONFORMATION

1) Les modes normaux

La structure des protéines MFS a été résolue grâce aux différents états conformationnels: outward-facing, occluded, et inward facing. Ces mouvements en lieu lors du transport du substrat: lorsqu'il se lie au récepteur, puis se trouve à l'intérieur du transporteur et lorsqu'il est relâché dans le milieu intracellulaire. Ces différentes phases ont été observées pour les protéines GLUT.

Les modes normaux sont étudiés sur la base d'un réseau élastique. Chaque mode représente un mouvement. Plus la fréquence est basse, plus les mouvements sont globaux. Plus la fréquence est élevée, et plus les mouvements sont locaux.

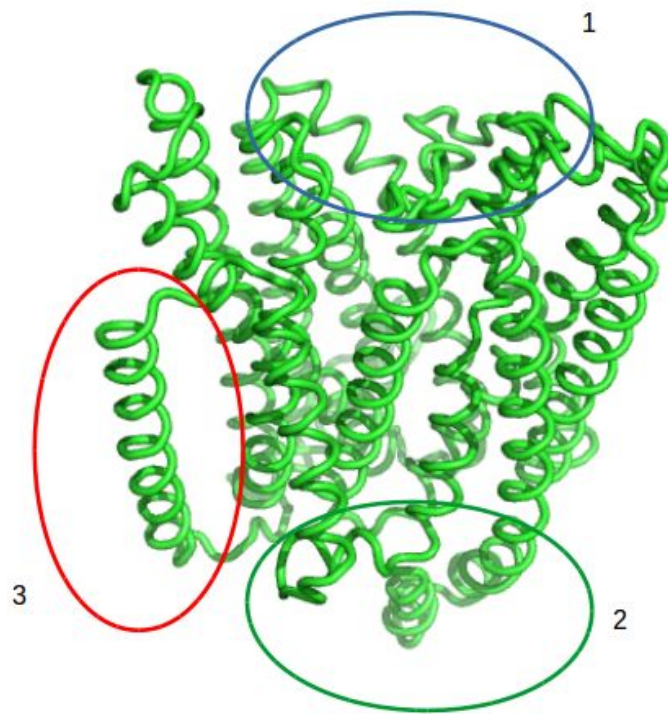


Figure 5: Description des mouvements de la ferroportine-1 avec le mode 14; 1 et 2 montrent des mouvements de clapets sous forme d'ouverture/fermeture antagonistes; 3 montre un mouvement de va et vient coordonné avec le mouvement d'ouverture/fermeture des parties 1 et 2.

Des recherches récentes ont montré que des séquences sont conservées. Leurs boucles sont courtes. Elles empêchent un mouvement trop ample des hélices reliées dans le cytoplasme. (TMs 2 et 3, 8 et 9). Au contraire, un linker connectant les deux domaines est long (30 à 100 acides aminés) se trouvant sous forme de boucles non structurées ou d'hélices alpha. Ces caractéristiques structurales peuvent supporter un large degré de mouvement durant les cycles alternatifs.

Le mode 14 permet de mieux visualiser le mouvement de la protéine et du transport du substrat. Les extrémités fonctionnent de façon antagoniste et une hélice guide le mouvement de fermeture/ouverture. L'axe de symétrie de ce pore couplé au mouvement sépare cette protéine en deux domaines distincts.

4. Conclusion

Déterminer les structures des MFS eucaryotes est un challenge. La plupart des structures obtenues proviennent des bactéries et les détergents utilisés ont probablement interféré avec la fonction des récepteurs membranaires. Les modèles sont obtenus par homologie, ce qui limite la qualité de l'information des structures obtenues. Par conséquent, ces modèles ne sont pas suffisamment fiables pour du drug design ou pour l'étude de la dynamique moléculaire. Les états conformationnels capturés ne sont pas suffisants pour décrire le mouvement détaillé du transporteur.

Références

This is about the in and the out, Peter J F Henderson & Stephen A Baldwin

Structural advances for the major facilitator superfamily (MFS) transporters, Nieng Yan, State Key Laboratory of Bio-membrane and Membrane Biotechnology, Center for Structural Biology, School of Medicine, Tsinghua-Peking Center for Life Sciences, Tsinghua University, Beijing 100084, China

The SLC2 (GLUT) family of membrane transporters, Mike Mueckler