

Caractérisation conformationnelle et mécanistique de la protéine Q8TDB8

Introduction

L'objectif de cette étude est de caractériser la conformation et la mécanistique de la protéine Q8TDB8, connu sous le nom de GLUT14 appartenant à la deuxième plus grande famille de transporteurs les *major facilitator superfamily* (MFS) qui sont des transporteurs membranaires qui facilitent la circulation de molécules dissoutes (sucres, métabolites, médicaments, oligosaccharides, acides aminés et oxyanions) à travers les membranes cellulaires. Les MFS sont exprimées chez tous les êtres vivants. Elles possèdent un mode de repliement homologue entre elles qui repose sur douze hélices α transmembranaires, séparé en deux groupes de six hélices constituées par les domaines *N*-terminal et *C*-terminal, relié entre elles par une longue boucle cytoplasmique. Une des familles de MFS la plus connue est celle des transporteurs de glucose (GLUT).

La famille des transporteurs de glucose est composé de 14 isoformes qui partagent des caractéristiques structurelles communes telles que 12 domaines transmembranaires, des extrémités *N*-terminal et *C*-terminal cytoplasmique et un site de N-glycosylation situé soit dans la première ou la cinquième boucle extracellulaire. Trois classes de GLUT peuvent être distinguées à partir de leur homologie de séquence (14-63% d'identité):

- Classe I : transporteurs de glucose « classiques » GLUT1-4 et GLUT14
- Classe II : GLUT5, 7, 9, 11
- Classe III : GLUT6, 8, 10, 12 et le transporteur de myo-inositol HMIT (ou GLUT13)

Malgré leurs similitudes structurelles, chaque isoforme est caractérisé par des caractéristiques distinctes comme sa localisation cellulaire et son substrat :

- Transporteurs de glucose : GLUT1-4, 8, 14
- Transporteurs de fructose : GLUT5, 7, 11
- Transporteur de polyol : GLUT12
- Transporteur de myo-inositol : GLUT13
- Transporteur d'urée : GLUT9

GLUT14 a été cloné et identifié en 2002 [3], il est associé au gène SLC2A14 situé sur le chromosome 12 et est situé en amont de GLUT3 avec qui il partage une grande similarité. Il est composé de 11 exons et possède une organisation génomique très similaire à celle de GLUT3, une des principales hypothèses émises et que GLUT14 résulte d'une duplication GLUT3. Il existe deux formes de GLUT14 lié à l'épissage alternatif, une forme courte GLUT14-S composée de 10 exons et produisant une protéine de 497 acides aminés avec un pourcentage de similarité à GLUT3 de 94,5%. Et une forme longue GLUT14-L qui possède un exon supplémentaire et qui code pour une protéine avec 520 acides aminés. Cette dernière diffère de GLUT14-S uniquement au niveau l'extrémité *N*-terminale. Comme les autres GLUT, GLUT14 contient 12 hélices couvrant la membrane ainsi que des motifs de signature de transporteur de sucre. Le site de glycosylation de GLUT 14 est présent dans la boucle 1. Contrairement, à GLUT3 dont l'expression est retrouvée dans de nombreux tissus

(cerveau, placenta, embryon, cancer) les deux isoformes de GLUT14 sont spécifiquement exprimées dans les testicules, GLUT14 est le transporteur prédominant dans les testicules et il n'est pas retrouvé chez la souris. 29 acides aminés sont retrouvés dans l'extrémité N-terminal de GLUT14-L alors que 6 -7 acides aminés sont retrouvés dans cette région dans le cas de GLUT3 et GLUT14-S. Uniquement 3 résidus sont différents dans le reste de la séquence entre GLUT3.

Nous allons donc nous concentrer sur la caractérisation conformationnelle et mécanistique de GLUT14-L par une méthode de *comparative modelling*.

Matériels et Méthodes

1) Recherche des partenaires et des interactions de la protéine

Les partenaires d'interaction et le type d'interaction associés à la protéine ont été recherchés sur la version 3.4.154 de BioGRID (*Biological General Repository for Interaction Datasets*). BioGRID est une base de données contenant des informations biologiques sur les interactions protéine-protéine, les interactions génétiques, les interactions chimiques et les modifications post-traductionnelles.

2) Recherche du meilleur alignement et étude de la conservation des résidus, des domaines et des motifs

Pour retrouver les résidus conservés un BLAST (paramètres par défaut de Uniprot) de notre séquence GLUT14 avec la banque de données de Uniprot a été réalisé. Les séquences provenant d'organismes différents de l'homo sapiens et partageant la même fonction que notre cible sont sélectionnées lorsque leur pourcentage d'identité se trouve entre 40 et 50%. Les séquences homologues ont ensuite été alignés selon différentes méthodes d'alignements. La version 8.93 de T-COFFEE, la version 7.215 de MAFFT (Multiple Alignment using Fast Fourier Transform), la version 3.8.31 de MUSCLE (MULTiple Sequence Comparison by Log- Expectation), la version 1.2.4 de clustalo (Clustal Omega) et TM-coffee ont été utilisés. La conservation des motifs et des domaines à ensuite été analysé par l'outil InterPro.

3) Caractérisation de la structure de la protéine

Pour caractériser la structure de la protéine c'est-à-dire identifier les structures secondaires et les régions hydrophobes les outils : PSIPRED v3.3 (outil de prédiction de structures secondaires), DomPred (outil de prédiction de domaines), MEMSAT3 & MEMSAT-SVM (outils de prédiction d'hélice membranaire) et HCA (outils de prédiction des régions hydrophobes) ont été utilisés.

4) Recherche d'une structure support "template" pour la modélisation

Pour retrouver une structure support ou *template* pour reconstruire la protéine, un BLASTp a été réalisé contre le banque de données PDB à partir de la protéine Q8TDB8. Les structures ayant une couverture maximale et un pourcentage d'identité comprise entre 20 et 40% ont été sélectionnées. Les structures associés aux clés Q6PXP3 (E-value: 1e-114 Score: 906 Ident.: 38.7%) et P22732 (E-value: 8.4e-114 Score: 899 Ident.: 39.7%) structures ont ainsi pu être sélectionnées. Mais les structures tridimensionnelles associées aux clés

sélectionnées ont été obtenue par homologie. Un BLASTp a donc été relancé sur ses deux clés par rapport à UniprotKB . Les mêmes filtres sont appliqués sur les résultats. Aucune structure 3D n'est retrouvé à partir de P22732, alors que deux clés ressortent pour Q6PXP3. Ces deux clés sont P11166 (une GLUT-1 avec une E-value de 1.1e-125, un score de 977 et un pourcentage d'identité 40.0%) et P11169 (une GLUT-3 avec une E-value de 6e-120, un score de 939 et un pourcentage d'identité 39.1%). Les structures associés au différentes clefs ont ensuite été observé. Les structures associés à P11169 ont été déterminées le plus adapté pour servir de support d'un point de vue résolutif et du point de vue de l'élément cristallisé dans notre cas GLUT3 qui partage un pourcentage de similarité de 95% avec notre cible. De plus nous retrouvons les conformations fermées outward open et occluded résolue par cristallographie avec une résolution inférieur à 3 Å.

Une recherche de template par PSI-BLAST, Swiss-model et HHpred a ensuite été réalisé à titre comparatif pour vérifier la pertinence des structures support sélectionnées sélectionnées.

5) Construction et évaluation des modèles

Les modèles ont ensuite été construit par homologie à partir des *templates* sélectionnées. Plusieurs outils ont été testés Swiss-model et HHpred, des serveurs de construction de modèle par homologie, Memoir et MEDELLER des serveurs de construction de protéine membranaire par homologie, et MUSTER un serveur qui crée des modèles par la méthode threading. Il servira de contrôle négatif, en effet les modèles générées par homologie sont de meilleures qualités que ceux générés par *threading*.

Les modèles générés seront ensuite évalués par différentes méthodes seuls les modèles les plus pertinents seront raffinés. Les outils d'évaluation sont ProQ, ProQM, ProSA-web, RAMPAGE, PPM server et OREMPRO. ProQ est un outils qui évalue la qualité d'un modèle à partir de caractéristique structurale. ProQM est la version de ProQ adapté à l'évaluation de protéine membranaire. ProSA-web est un serveur web qui évalue les erreurs dans les structures tridimensionnelle. RAMPAGE qui permet l'analyse du squelette peptidique des protéines par un diagramme de ramachandran. Et enfin, PPM server et OREMPRO qui sont des outils qui permettent de replacer la protéine dans une membrane.

6) Raffinement des modèles

Les meilleurs modèles, généré par Memoir sont ensuite raffiner entièrement par GalaxyRefine. Après le raffinement différents scores sont associés aux modèles raffinés. Le score MolProbity qui est basé uniquement sur les propriétés du modèle prédit, plus ils sont petits meilleurs sont les modèles. Le test de distance globale (GDT), qui est une mesure de la similitude de deux structures, avec compositions d'acides aminés similaires mais différentes structures tertiaires. Et le clash score qui est comprise entre 0 et 2,5 Å pour structures de haute résolution et entre 2,5 et 3,5 Å pour les structure de basse résolution.

7) Etude de la mécanistique du modèle par les modes normaux

Afin d'étudier le mouvement des trois modèles générés, ces derniers ont été simulés par des modes normaux par NOMAD-ref les paramètres par défaut on été conservé pour les simulations.

Résultats

1) Recherche des partenaires et des interactions de la protéine

Les partenaires et le type d'interaction associé à GLUT14 ont été recherché dans BioGRID.

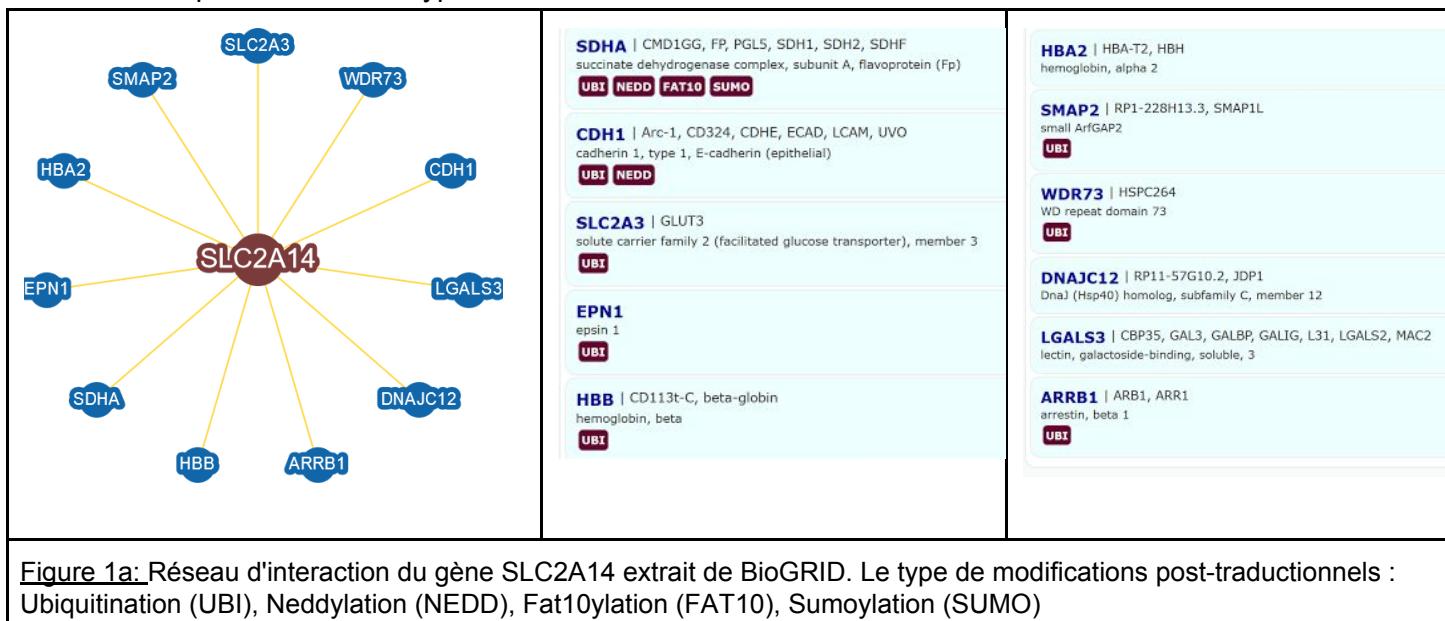


Figure 1a: Réseau d'interaction du gène SLC2A14 extrait de BioGRID. Le type de modifications post-traductionnels : Ubiquitination (UBI), Neddylate (NEDD), Fat10ylation (FAT10), Sumoylation (SUMO)

La protéine d'intérêt n'est relié à aucun GO Process et GO function mais possède une GO component qui indique que c'est une protéine nucléaire, cette annotation GO a également été retrouvé dans la fiche unprot de la protéine. Elle semble interagir avec 11 partenaires sur lesquels elle ajoute des modifications post-traductionnelles (fig 1a) telle que l'ubiquitination, la neddylation, la fat10ylation ou la sumoylation. Les interactions entre GLUT14 et les partenaires ARRB1, DNAJC12, EPN1, HBA2, HBB, LGALS3, SMAP2 et WDR73 ont été retrouvé par spectroscopie de masse ou par des anticorps polynomiaux. L'interaction entre GLUT14 et l'enzyme CDH1 a été déduite par spectroscopie de masse. La protéine SDHA a été retrouvé dans la même fractions purifié que GLUT14 (fig 1b).

| Switch View: Interactors (11) Interactions (11) Network PTM Sites (1) | | | | | | | | |
|-----------------------------------------------------------------------|------|------------|----------------------------|---------------------------|------------|--------|------------|-------|
| Displaying 11 total unique interactions | | | | | | | | |
| Interactor | Role | Organism | Experimental Evidence Code | Dataset | Throughput | Score | Curated By | Notes |
| ARRB1 | HIT | H. sapiens | Affinity Capture-MS | Huttlin EL (2014/pre-pub) | High | 0.7824 | BioGRID | |
| CDH1 | BAIT | H. sapiens | Proximity Label-MS | Guo Z (2014) | High | - | BioGRID | - |
| DNAJC12 | HIT | H. sapiens | Affinity Capture-MS | Huttlin EL (2014/pre-pub) | High | 0.7592 | BioGRID | |
| EPN1 | HIT | H. sapiens | Affinity Capture-MS | Huttlin EL (2017) | High | 0.998 | BioGRID | |
| HBA2 | HIT | H. sapiens | Affinity Capture-MS | Huttlin EL (2017) | High | 0.9613 | BioGRID | |
| HBB | HIT | H. sapiens | Affinity Capture-MS | Huttlin EL (2017) | High | 0.9968 | BioGRID | |
| LGALS3 | HIT | H. sapiens | Affinity Capture-MS | Huttlin EL (2014/pre-pub) | High | 0.7785 | BioGRID | |
| SDHA | BAIT | H. sapiens | Co-fractionation | Havugimana PC (2012) | High | 0.766 | BioGRID | |
| SLC2A3 | HIT | H. sapiens | Affinity Capture-MS | Huttlin EL (2017) | High | 1 | BioGRID | |
| SMAP2 | HIT | H. sapiens | Affinity Capture-MS | Huttlin EL (2017) | High | 0.8943 | BioGRID | |
| WDR73 | HIT | H. sapiens | Affinity Capture-MS | Huttlin EL (2017) | High | 0.7999 | BioGRID | |

Figure 1b : Tableau des interactions retrouvés dans BioGRID pour la GLUT14

2) Recherche du meilleur alignement et étude de la conservation des résidus, des domaines et des motifs

A la fin du BLAST, cinq séquences correspondant à nos critères ont été sélectionnées (Table I).

| Identifiant Uniprot | Espèces | Informations | Reviewed |
|------------------------|-------------------|----------------------------------------------------|----------|
| P12336 | Rattus norvegicus | E-value: 9.2e-158 Score: 1,215 Ident.: 47.7% | Oui |
| Q90592 | Gallus gallus | E-value: 1.9e-157 Score: 1,214 Ident.: 48.7% | Oui |
| P14246 | Mus Musculus | E-value: 7.6e-157 Score: 1,209 Ident.: 48.0% | Oui |
| P58351 | Bos taurus | E-value: 5.4e-145 Score: 1,129 Ident.: 48.2% | Oui |
| Q8WMN1 | Ovis aries | E-value: 3.4e-119 Score: 956 Ident.: 42.3% | Oui |

Table I : Sequence sélectionnées après le BLAST de Q8TDB8 sur la banque de données Uniprot

T-COFFEE, Version_11.00.d625267 (2016-01-11 15:25:41 - Revision d625267 - Build 507)
Cedric Notredame

| IN | HEL | OUT | |
|----------------------|-------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------|-------------------------|
| sp Q8TDB8 GTR14 | M E H F N G H G V S G I G G F L V S L T S R M K P H T | L A V T P A L I F A I T V A T I G - S P O F G Y N T G V I N A P E T I | |
| sp P12336 GTR2- | - - - - - S E | D K I T G L A F T F V T A V I G L - S P O F G Y D I V G I N A P E T I | |
| sp Q90592 GTR2- | M D G K S K M D A E | - - - - - K H L T G L V S P T A V I G L - F R O Y G S L G V I N A P O K V I | |
| sp P14246 GTR2- | M - - - - - S E | - - - - - D K I T G L A F T F V T A V I S L S - S P O F G Y D I V G I N A P E T I | |
| sp P58351 GTR2- | M - - - - - T E | - - - - - D K I T G L A F T F V T A V I S L S - S P O F G Y D I V G I N A P O K V I | |
| sp Q8WMN1 GTR5_- | M E P Q D P V K R E | - - - - - G R L T P V I V L A T I A A F G S S P O F G Y N T G V A T I S P E F M | |
| sp Q8TDB8 GTR14 | K E F I N K T L T D K A | - - - - - | N A P P S E V L L T N |
| sp P12336 GTR2- | I S H Y R H V L G P V L D D R R A T I N Y D I N D G T D P I | - - - - - I T P V A H T T P D A M E | E E T E G S A H T I |
| sp Q90592 GTR2- | E A H Y R M G L A P V L D D R R A T I N Y D I N D G T D P I | - - - - - E E T E G S A H T I | E D P T V S P H I L T M |
| sp P14246 GTR2- | I S H Y R H V L G P V L D D R R A T I N Y D I N D G T D P I | - - - - - I T P V A H T T P D A M E | D E P T V S P H I L T M |
| sp P58351 GTR2- | I T H Y R H V L G P V L D D R R A T I N Y D I N D G T D P I | - - - - - I T P V A H T T P D A M E | E E T E G S A H T I |
| sp Q8WMN1 GTR5_- | K D F Y N T Y Y D R V | - - - - - I S L G D P T P V S H A E E T I R S A S I L T M | G E Y N H E F T L |
| sp Q8TDB8 GTR14 | L W S L S V A I F S V G G M I G S F V G L F V N R F G R R N S M I L V N L L A T G G C L M G L C K I A E S V E M I L I G R | | |
| sp P12336 GTR2- | L W S L S V S F A V G G M V A S F F G G W L G D K L G R I K A M L A N S L T G A L M G C S K P G P H A L I A G R | | |
| sp Q90592 GTR2- | L W S L S V S F A V G G M V A S F F G G W L G D K L G R I K A M L A N S L T G A L M G C S K P G P H A L I A G R | | |
| sp P14246 GTR2- | L W S L S V S F A V G G M V A S F F G G W L G D K L G R I K A M L A N S L T G A L M G C S K P G P H A L I A G R | | |
| sp P58351 GTR2- | L W S L S V S F A V G G M V A S F F G G W L G D K L G R I K A M L A N S L T G A L M G C S K P G P H A L I A G R | | |
| sp Q8WMN1 GTR5_- | L W S L S V S F A V G G M V A S F F G G F G L S L M V G P L V N N L G R K G T L L F N N I S I V P A L L M G F S D L A K S F E M I I V A R | | |
| sp Q8TDB8 GTR14 | L V I G L F C G L C T G F V P M Y I G E I S P T A L R G A F G T L N L Q G L I V I G I L V A Q I F G L E L I L G S E E L W P V L | | |
| sp P12336 GTR2- | S V S G L Y C G L I S G L V P M Y I G E I A P T T L R G A L G T H L O L A L V T G I L I S O I A G L S F I L G N D O Y W H M L | | |
| sp Q90592 GTR2- | A I T G L Y C G L S S G L V P M Y I G E I A P T T L R G A L G T H L O L A I V T G I L I S O I A G L S F I L G N D E L W P L L | | |
| sp P14246 GTR2- | S V S G L Y C G L I S G L V P M Y I G E I A P T T L R G A L G T H L O L A L V T G I L I S O I A G L S F I L G N D O H W D M L | | |
| sp P58351 GTR2- | G I S G L Y C G L S S G L V P M Y I G E I A P T T L R G A L G T H L O L A I V T G I L I S O I A G L S F I L G N D E L W P L L | | |
| sp Q8WMN1 GTR5_- | V L V G I A C G L S S N V V P M Y I G E L G A P K N R G R A G V P Q V P L I T I G L V A Q I F G R L S L A N E G W P L | | |
| sp Q8TDB8 GTR14 | L G F T I L P A L I O S A A P C C P E S P R F L L I N R K K E E N A T R I L O R L W G T Q D V S Q D I Q E M K D E S A R M S | | |
| sp P12336 GTR2- | L G L S A V P A L L O C C L L L F C P E S P R Y L I N L E E V R A K K S L R K R G T E D I T D K I N E M K E K E A E S | | |
| sp Q90592 GTR2- | L G L S A V P A L L O C C L L L F C P E S P R Y L I K L G K V E A K K S L R K R G T E D I T D K I N E M K E K E A E S | | |
| sp P14246 GTR2- | L G L S A V P A L L O C C L L L F C P E S P R Y L I K L G K V E A K K S L R K R G T E D I T D K I N E M K E K E A E S | | |
| sp P58351 GTR2- | L G L S A V P A T O C L L L F C P E S P R Y L I K L D E E A K A K S L R K R G T E D D I T D K I T D M E K K E R E A E S | | |
| sp Q8WMN1 GTR5_- | L G L T G I P A V L O L L F P P F E S P R Y L I Q K Q D E E A K A R L R R L R G W H D V D A I E E I L E E D E A R E A E S | | |
| sp Q8TDB8 GTR14 | Q E K O V T I F L R V F S Y R O P I I T I S T W L Q S Q L G S I N A V F Y Y S T G I F K D A V Q O - - P Y A T I S A | | |
| sp P12336 GTR2- | T E O K V S V I O L F T D P N P R O I V V A M L H L A O F G S I N G I F Y Y S T I S F O T A G I S O - - P Y A T I G V | | |
| sp Q90592 GTR2- | S E K R V S T I G O L F P S S K S Y K R O A V T V A M V O T S Q I S F N A I F Y Y S T I N F O T A G V G O - - P Y A T I G V | | |
| sp P14246 GTR2- | T E O K V S V I O L F T D O N A R Y O P I V A M L H L A O F G S I N G I F Y Y S T I S F O T A G I S O - - P Y A T I G V | | |
| sp P58351 GTR2- | N E K K V S T I F T N A S Y R O P I V A M L H L A O F G S I N G I F Y Y S T I S F O T A G I S O - - P Y A T I G V | | |
| sp Q8WMN1 GTR5_- | A A G F I S V L K F K N R S L R O W V I S I V L M A Q O S G V I N A I Y Y A D O I Y L S G A V K E D D V O Y V T A G | | |
| sp Q8TDB8 GTR14 | G V V N T I F T L L S L F L V E A G R R T L H M I G L G M A F C S T L M T V S L L L K N H Y N G M S F V C I G A I L F V | | |
| sp P12336 GTR2- | G A I M N I T F A V S V L V E K A G R R T L F L A G M I G M F C A V F M S L G V L L V D F K T W S M Y S V H T A I F L V | | |
| sp Q90592 GTR2- | G V V N T V F T V S V L V E K A G R R S T L F L A G M I G M L T S A V A M T V G L V L L S F O F A M Y S V W H T A I F L V | | |
| sp P14246 GTR2- | G A I M N I T F A V S V L V E K A G R R T L F L T G M I G M F C T I I M P G V L V L L D F K T W S M Y S V H T A I F L V | | |
| sp P58351 GTR2- | G V V N T V F T V S V L V E K A G R R S T L F L T G M I G M F C V A T I M P G V L V L L S F O F A M Y S V W H T A I F L V | | |
| sp Q8WMN1 GTR5_- | G A V N V I L I T C A I F V E L M G R R F L L L G S F V C T A C V L T G A L A N D O V I S W M P Y Y S I A C V I S Y | | |
| sp Q8TDB8 GTR14 | A C F E I G P G I P W F I V A E L F S O G P R P A A M A V G C S N W T S N F L V G L L F P S A A Y Y L G A Y V I I F T G | | |
| sp P12336 GTR2- | S F F E I G P G I P W F I V A E F F S O G P R P T A L A L A F S N W V C N F I I A L C F O Y I A D F L G P V V F F L V | | |
| sp Q90592 GTR2- | I F F E V G P G I P W F I V A E F F S O G P R P T A A I A V G C S N W C A N C I F V G M C F O Y I A D F L G P V V F F V A F | | |
| sp P14246 GTR2- | S F F E I G P G I P W F I V A E F F S O G P R P T A L A L A F S N W V C N F I I A L C F O Y I A D F L G P V V F F L V | | |
| sp P58351 GTR2- | S F F E I G P G I P W F I V A E F F S O G P R P T A A L A I A F S N W T G N F I I A L C F O Y I A D F L G P V V F F L V | | |
| sp Q8WMN1 GTR5_- | I G H A L G P S P I A L L V T E I F L Q S P R A A Y M V A G H W L M S T F V G L V P F E T Q V G L S F V A Y F I V A F | | |
| sp Q8TDB8 GTR14 | F L I T F L A F T F F K V P E T R G R T E D I T A R F E G O A H G D R S G K D G V M G M S I E P A K E T T N - V | | |
| sp P12336 GTR2- | V L V L F V T L F F K V P E T K G S K F D T E I A R F R K K S G A P P R K - - A T V O M E F L G S S - - E T - V | | |
| sp Q90592 GTR2- | L L L V F F L A F T F F K V P E T K G S K F D T E I A R F R K K L P K A P S M - - T E L D R G G - - E E - A | | |
| sp P14246 GTR2- | V L V L F V T L F F K V P E T K G S K F D T E I A R F R K K S G A P P R K - - A A V Q E M F L A S S - - E S - V | | |
| sp P58351 GTR2- | W S P L F C S H K L F O K P K E N P L R K S O O S D R E R G R V G O L K R O K - - L | | |
| sp Q8WMN1 GTR5_- | I C F L T V V I F L I P E T K S T F T E T N O I F T K M N K V P G V H P - - E K E E L K E F P P S - - T A R Q | | |

Figure 2a : Alignement multiple de la protéine Q8TDB8 avec ses homologues (50 et 40% identité)

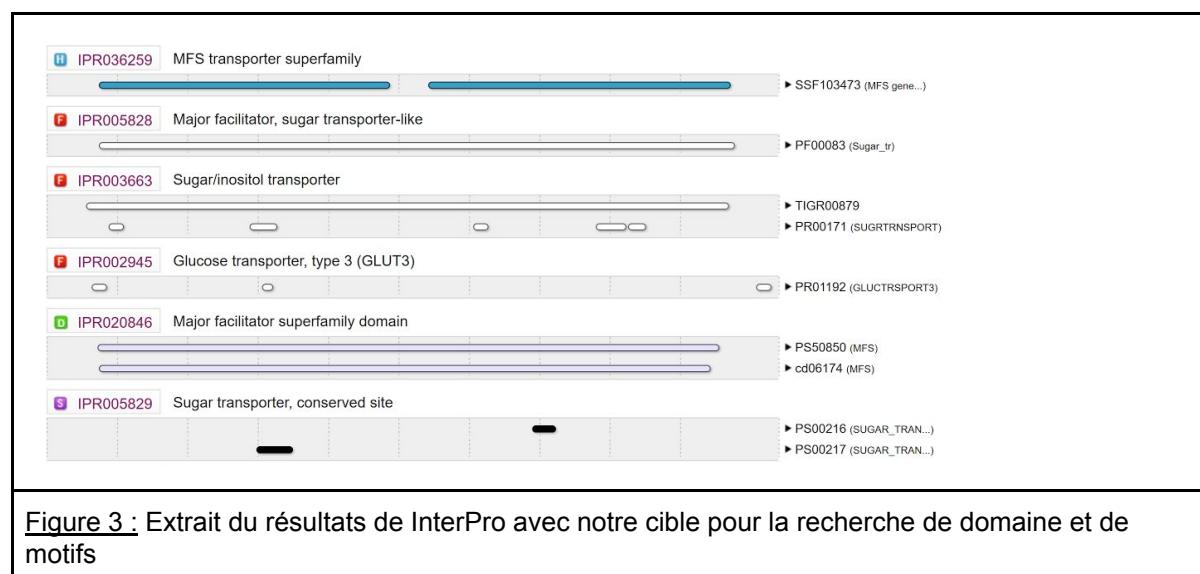
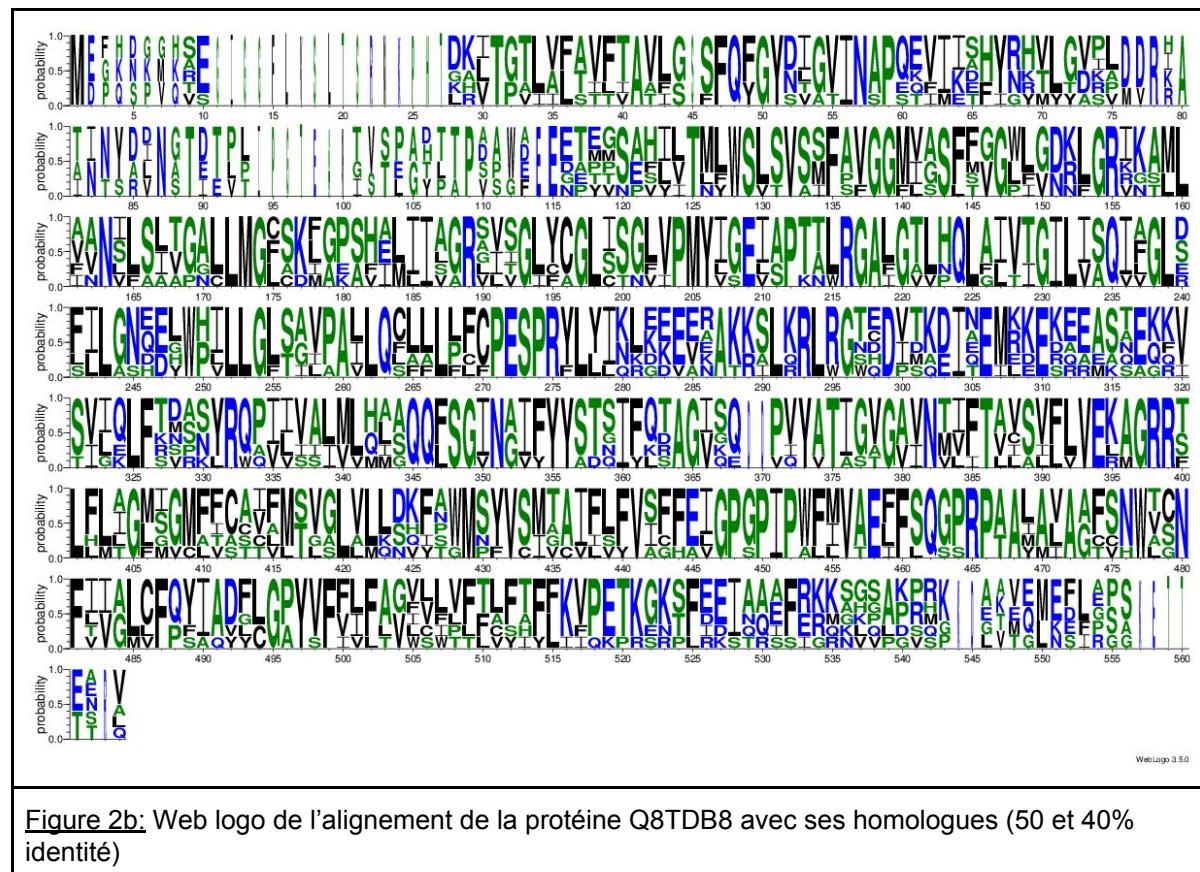
sont associés à des positions très conservé. Une di-leucine, une di-glutamine, une di-tyrosine et une di-arginine sont retrouvé à des positions très conversés.

Les résultats associées à la recherche de domaine et de motifs dans InterPro avec notre séquence cible (fig 3), indique que cette dernière possède des domaines conservé associés à la famille des MFS et qu'elle ressemble à un facilitateur du transport de glucose qui possède des sites associé au transporteur de glucose GLUT3 c'est informations sont également retrouvées dans la fiche uniport de notre protéine.

Les résultats de l'alignement sont globalement les mêmes pour les cinq méthodes d'alignement sauf sur la région N-terminal. Certains alignement pénalise moins les gap c'est le cas de T-coffee et TM-coffee contrairement à Clustalo et MUSCLE. Le pourcentage d'identité de la méthode TM-coffee 21,12% est meilleur que pour les quatre autres méthodes. Cela n'est pas étonnant puisque TM-coffee est une méthode d'alignement adapté aux protéines membranaire. Cet alignement sera donc sélectionnées pour étudier la conservation des résidus au sein de notre famille de protéine (fig 2a). Pour cela, le web logo de l'alignement a été généré (fig 2b).

Douze hélices transmembranaires, et des régions *N-terminal* et *C-terminal* cytoplasmique sont retrouvées après l'alignement ceci n'est pas étonnant puisque c'est une des caractéristique de la famille des MFS (fig 2a), cela conforte le choix de l'alignement.

D'après le web logo, les régions *N-terminal* et *C-terminal* ne sont pas conservées. La famille MFS est très riche en tryptophane, proline, glycine, tyrosine et glutamine qui



3) Caractérisation de la structure de la protéine

Douze régions hydrophobes sont retrouvés dans la sortie de HCA (fig 4a) pour notre protéine. Ces régions hydrophobes pourraient correspondre aux 12 hélices transmembranaires prédictes par MENSAT (fig 4b) ou MENSAT-SVM (fig 4c). La position des hélices transmembranaires prédictes varient selon la méthode, nous ne pouvons pas discriminer les bonnes positions car les structures secondaires prédictes par PSIpred (fig 4d) ne correspondent pas. En effet dans la prédition de structure secondaires on retrouve 24 hélices et 3 feuillets par PSIpred.

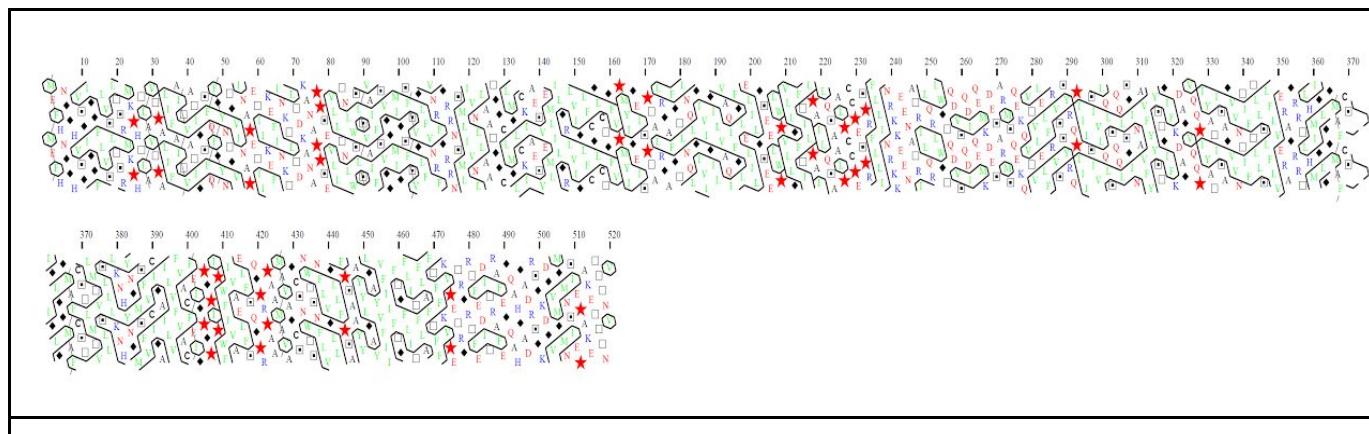


Figure 4a: Répartition des régions hydrophobes dans notre protéine cible GLUT14. Les acides aminés colorés en verts sont situés dans une région hydrophobe.

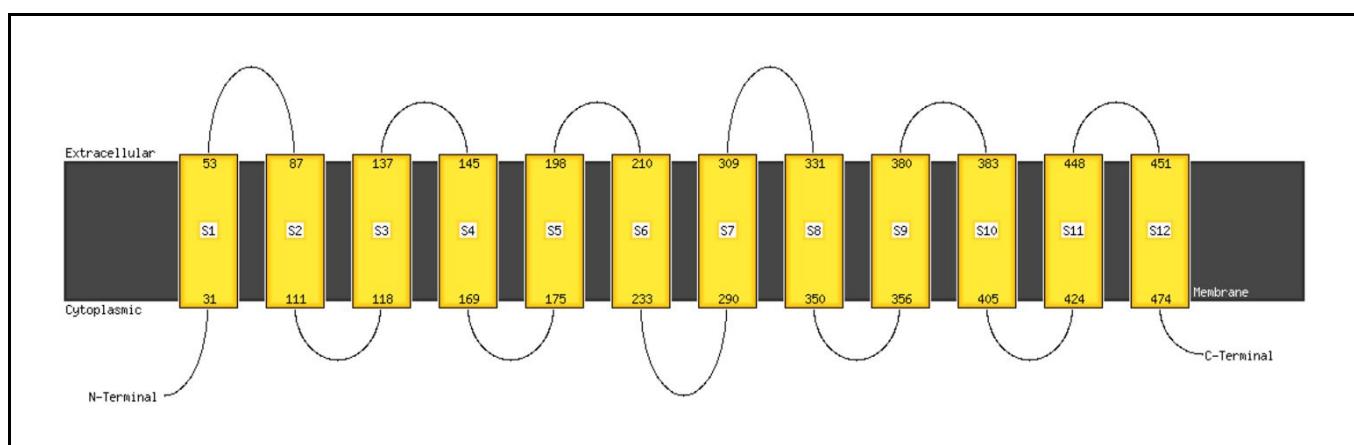
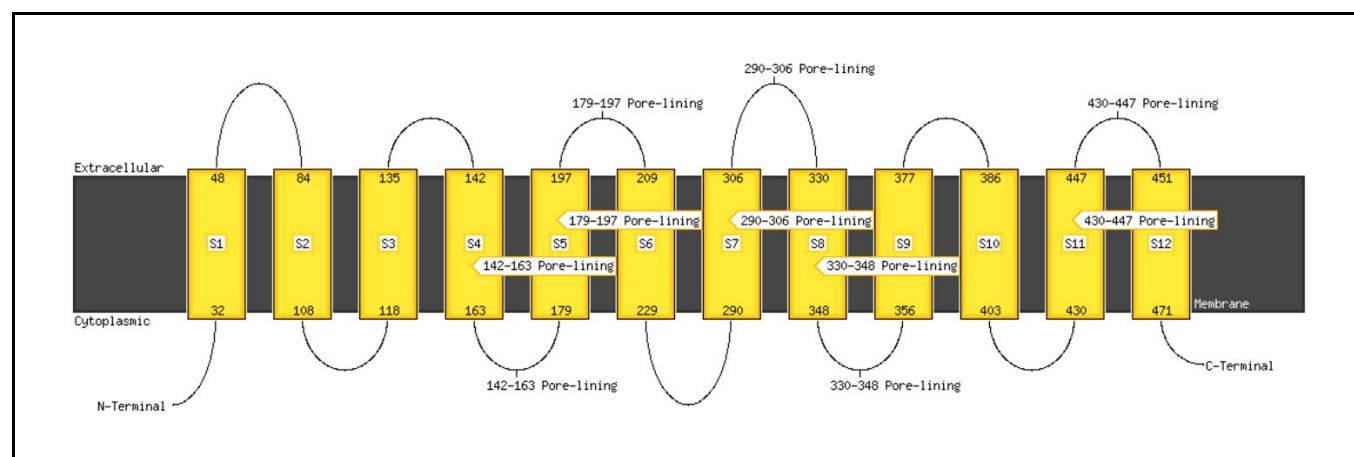


Figure 4b : Structuration des potentielles hélices transmembranaires de GLUT14 selon MEMSAT



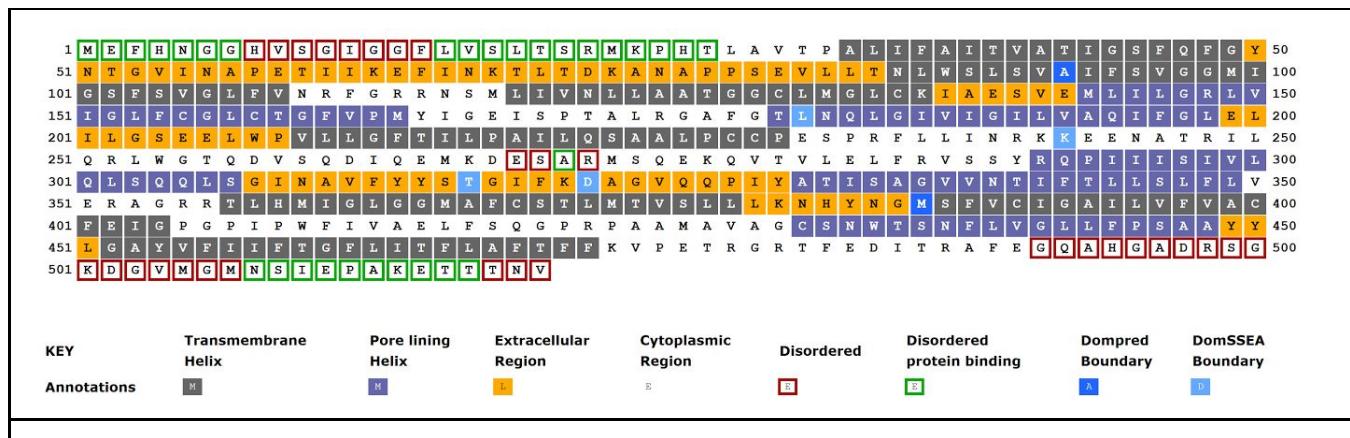


Figure 4c: Structuration des potentielles hélices transmembranaires de GLUT14 selon MEMSAT - SVM

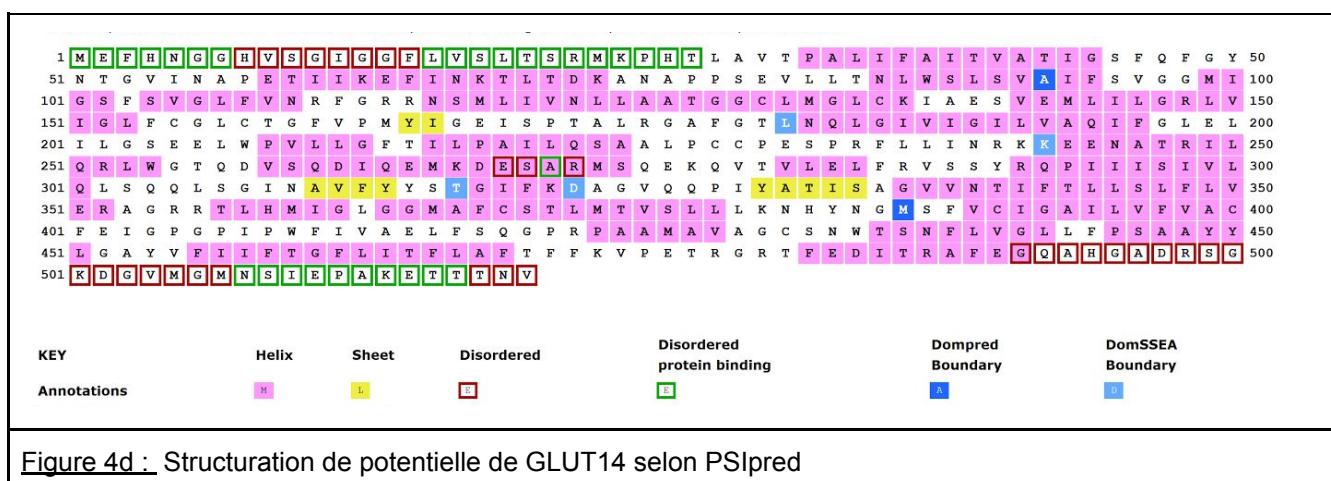


Figure 4d : Structuration de potentielle de GLUT14 selon PSIPred

4) Recherche d'une structure support “template” pour la modélisation

Trois *template* on donc été sélectionnés à partir de P11169 : 4zw9, 4zwc et 5c65.

| PDB | 4zw9 | 4zwc | 5c65 |
|--------------------------|------------------------|--------------------|-------|
| Résolution | 1.5 | 2.6 | 2.65 |
| Structure - conformation | GLUT3 outward-occluded | GLUT3 outward-open | GLUT3 |

Pour vérifier les pertinences de *templates* sélectionnés une recherche automatique de *template* a été réalisé sur HHpred, Swiss-Model et par PSI-BLAST. Les *templates* 4zw9 et 5c65 sont retrouvés dans le top 5 des templates proposés par HHpred (Figure 5a). Les trois *templates* sélectionnées sont retrouvés parmi les six meilleurs *templates* pour Swiss-Model (Figure 5b). Et enfin, par PSI-BLAST les trois *templates* sont retrouvés dans le top 10 des meilleurs *templates* (Figure 5c).

| Nr | Hit | Name | Align. | DB-ID | Source | Length | Score |
|----|--------|------------------------------------------------------------------------------------------------------------------------------|--------|------------|------------------------------------------------------------------------------------------|--------|-------|
| 1 | 4ZW9_A | Solute carrier family 2, facilitated; transporter, TRANSPORT PROTEIN; HET: BGC, GLC, OLC; 1.502A {Homo sapiens} | ✓ 1 | PDB-4ZWC_B | mol protein length:518 Solute carrier family 2, facilitated glucose transporter member 3 | 518 | 922.0 |
| 2 | 5EQI_A | Solute carrier family 2, facilitated; MFS transporter, glucos transporter, TRANSPORT; HET: 5RH; 3.002A {Homo sapiens} | ✓ 2 | PDB-4ZWC_A | mol protein length:516 Solute carrier family 2, facilitated glucose transporter member 3 | 516 | 922.0 |
| 3 | 5EQG_A | Solute carrier family 2, facilitated; MFS transporter, glucose transporter, TRANSPORT; HET: SRE; 2.9A {Homo sapiens} | ✓ 3 | PDB-4ZWB_A | mol protein length:518 Solute carrier family 2, facilitated glucose transporter member 3 | 518 | 922.0 |
| 4 | 5C65_A | Solute carrier family 2, facilitated; MEMBRANE TRANSPORTER, SUGAR TRANSPORT, Structural; HET: Y01, 37X; 2.65A {Homo sapiens} | ✓ 4 | PDB-4ZWB_A | mol protein length:518 Solute carrier family 2, facilitated glucose transporter member 3 | 518 | 922.0 |
| 5 | 4YB9_D | Bovine Fructose Transporter GLUT5; GLUT5 Transporter, Fructose, Bovine MFS; 3.2A {Bos taurus} | ✓ 5 | PDB-5C65_B | mol protein length:481 Solute carrier family 2, facilitated glucose transporter member 3 | 481 | 882.0 |
| | | | ✓ 6 | PDB-5C65_A | mol protein length:481 Solute carrier family 2, facilitated glucose transporter member 3 | 481 | 882.0 |
| | | | ✓ 7 | PDB-5EQI_A | mol protein length:492 Solute carrier family 2, facilitated glucose transporter member 1 | 492 | 618.0 |
| | | | ✓ 8 | PDB-5EQH_A | mol protein length:492 Solute carrier family 2, facilitated glucose transporter member 1 | 492 | 618.0 |
| | | | ✓ 9 | PDB-5EQG_A | mol protein length:492 Solute carrier family 2, facilitated glucose transporter member 1 | 492 | 618.0 |
| | | | ✓ 10 | PDB-4PYP_A | mol protein length:504 Solute carrier family 2, facilitated glucose transporter member 1 | 504 | 614.0 |

Figure 5a : Top 5 des *templates* associés à la protéine Q8TDB8 d'après HHpred

Figure 5b : Top 10 des *templates* associés à la protéine Q8TDB8 d'après Swiss Model

Figure 5c : Top 10 des *templates* associés à la protéine Q8TDB8 d'après PSI-BLAST vs banque de données de structures

5) Construction et évaluation des modèles

Les différents modèles ont ensuite été généré par les différentes méthodes expliqué dans la partie matériels et méthodes.

Aucun modèle n'a pu être généré par MEDELLER le programme a été interrompu peu importe le template et les différents alignements proposés .

MUSTER quant à lui a permis la génération de 10 modèles, comme le premier modèle a été construit à partir d'un des *templates* sélectionnées, il a été choisi pour être le contrôle négatif (fig 6a).

| Rank | Template | Align_length | Coverage | Zscore | Seq_id | Type | Target-template alignments | 3-D models from threading alignments | Full-length models by MODELLER |
|------|----------|--------------|----------|--------|--------|------|------------------------------|--------------------------------------|--------------------------------|
| 1 | 5c65A | 457 | 0.878 | 25.813 | 0.950 | Good | alignment_1 | threading_1 | model_1 |
| 2 | 4pypA | 447 | 0.859 | 25.271 | 0.655 | Good | alignment_2 | threading_2 | model_2 |
| 3 | 4ybqA | 450 | 0.865 | 23.761 | 0.411 | Good | alignment_3 | threading_3 | model_3 |
| 4 | 4ldsA | 417 | 0.801 | 21.721 | 0.290 | Good | alignment_4 | threading_4 | model_4 |
| 5 | 4gbyA | 444 | 0.853 | 21.235 | 0.297 | Good | alignment_5 | threading_5 | model_5 |
| 6 | 4j05A | 405 | 0.778 | 12.907 | 0.153 | Good | alignment_6 | threading_6 | model_6 |
| 7 | 5c65A1 | 248 | 0.476 | 12.470 | 0.935 | Good | alignment_7 | threading_7 | model_7 |
| 8 | 4pypA1 | 245 | 0.471 | 12.309 | 0.624 | Good | alignment_8 | threading_8 | model_8 |
| 9 | 4ybqA2 | 218 | 0.419 | 10.911 | 0.399 | Good | alignment_9 | threading_9 | model_9 |
| 10 | 5c65A2 | 209 | 0.401 | 10.702 | 0.967 | Good | alignment_10 | threading_10 | model_10 |

Figure 6a: Liste des modèles produit par MUSTER

Pour chaque *template* Swiss Model génère un seul modèle, les trois modèles générés seront donc tous évaluer (Table II).

| Swiss Model | | Quality model (Z-score) | LGscore | MaxSub | Q-means |
|---------------------------------------------------------------------------------------|----------|----------------------------|---------|--------|---------|
| Template 1 (4zw9) | Modèle 1 | -6.01 | 5.920 | 0.357 | -3.57 |
| Template 2 (4zwc) | Modèle 1 | -5.89 | 5.952 | 0.359 | -3.74 |
| Template 3 (5c65) | Modèle 1 | -6.07 | 5.789 | 0.347 | -3.98 |
| <u>Table II:</u> Evaluation des modèles générés par Swiss Model par ProSa-web et ProQ | | | | | |

| Memoir | | Quality model (Z-score) | LGscore | MaxSub |
|-----------------------------------------------------------------------------------|---------------------|----------------------------|---------|--------|
| Template 1 (4zw9) | Modèle 1 (complete) | -5.93 | 5.111 | 0.320 |
| | Modèle 2 (core) | -5.96 | 6.199 | 0.386 |
| | Modèle 3 (hiacc) | -5.96 | 6.199 | 0.386 |
| | Modèle 4 (hicov) | -5.96 | 6.199 | 0.386 |
| Template 2 (4zwc) | Modèle 1 | -5.73 | 4.969 | 0.335 |
| | Modèle 2 | -5.96 | 6.131 | 0.375 |
| | Modèle 3 | -5.96 | 6.131 | 0.375 |
| | Modèle 4 | -5.96 | 6.131 | 0.375 |
| Template 3 (5c65) | Modèle 1 | -6.16 | 4.376 | 0.310 |
| | Modèle 2 | -5.22 | 5.724 | 0.350 |
| | Modèle 3 | -5.22 | 5.724 | 0.350 |
| | Modèle 4 | -6.33 | 5.605 | 0.345 |
| <u>Table III:</u> Evaluation des modèles générés par Memoir par ProSa-web et ProQ | | | | |

En ce qui concerne Memoir, les modèles ont été générée à partir des trois templates sélectionnées puis évalué par différents outils (Table III). Pour chaque template le meilleur modèle est le modèle 4 car il possède à chaque fois le meilleur compromis entre les trois scores et il a été réalisé pour avoir la plus grande couverture.

| | MUSTER | Swiss Model | | | Mémoir | | |
|--------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------|
| | Modèle 1 | Modèle 1.1 | Modèle 2.1 | Modèle 3.1 | Modèle 1.4 | Modèle 2.4 | Modèle 3.4 |
| OREMPRO | | | | | | | |
| Hélice TM | 12 | 12 | 12 | 12 | 12 | 12 | 12 |
| Position TM | 39-58 88-108 117-138 140-163 177-198 207-225 291-316 329-350 353-374 389-413 424-44 455-474 | 39-59 86-109 117-138 140-163 179-204 206-226 294-320 329-350 355-379 385-413 424-448 451-474 | 38-59 87-109 115-138 140-163 179-204 206-226 294-320 329-355-375 389-413 424-448 451-474 | 38-57 88-108 117-138 140-164 176-204 207-222 291-317 328-350 353-378 386-414 423-448 450-474 | 39-59 86-109 114-138 140-166 176-204 207-222 291-317 328-350 353-378 386-414 423-448 450-474 | 35-59 86-109 117-138 140-163 179-204 206-226 294-320 329-353-378 386-414 424-448 450-473 | 39-59 86-108 117-138 140-163 179-204 206-226 294-320 329-356-380 385-413 424-448 450-473 |
| MAIDEN score | error | error | error | error | -125.4173 | -78.0709 | -109.8143 |
| Hydrophobic thickness | 31.2 Å | 34.8 Å | 35.0 Å | 31.5 Å | 34.5 Å | 34.0 Å | 34.0 Å |
| PPM server | | | | | | | |
| Hélice TM | error | 10 | 10 | 10 | 12 | 12 | 12 |
| Position TM | | 1(231-251),2(257-273),3(281-305),4(316-332),5(343-361),6(468-487),7(492-513),8(536-559),9(569-590),10(599-616) | 1(231-251),2(257-273),3(281-305),4(316-332),5(343-361),6(468-487),7(492-513),8(536-559),9(569-590),10(599-616) | 1(231-251),2(257-273),3(281-305),4(316-332),5(343-361),6(468-487),7(492-513),8(536-559),9(569-590),10(599-616) | 1(38-58),2(87-109),3(116-135),4(140-164),5(179-199),6(208-227),7(294-320),8(328-350),9(355-379),10(385-413),11(423-448),12(453-472) | 1(38-58),2(87-109),3(116-135),4(140-164),5(179-199),6(208-227),7(294-320),8(328-350),9(355-379),10(385-413),11(423-448),12(453-472) | 1(38-58),2(87-109),3(116-135),4(140-164),5(179-199),6(208-227),7(294-320),8(328-350),9(355-379),10(385-413),11(423-448),12(453-472) |
| Depth/Hydrophobic Thickness | | 29.2 ± 1.1 Å | 29.2 ± 1.1 Å | 29.2 ± 1.1 Å | 32.2 ± 1.4 Å | 32.2 ± 1.4 Å | 32.2 ± 1.4 Å |
| ΔG _{transfer} | | -67.8 kcal/mol | -67.8 kcal/mol | -67.8 kcal/mol | -90.2 kcal/mol | -90.2 kcal/mol | -90.2 kcal/mol |
| Tilt Angle | | 3.± 2.° | 3.± 2.° | 3.± 2.° | 7.± 3.° | 7.± 3.° | 7.± 3.° |
| ProQM | | | | | | | |
| Global Quality | 0.686 | 0.773 | 0.755 | 0.762 | 0.784 | 0.772 | 0.771 |
| <u>Table IV:</u> Table d'évaluation des modèles générés par les différentes méthodes | | | | | | | |

Pour l'évaluation des modèles générées par les différentes méthodes des nouveaux outils d'évaluations on été rajouté ces outils sont spécifiques aux protéines membranaires. Notre principale critère d'évaluation est le nombre d'hélice transmembranaire. Avec OREMPRO tous nos modèles possèdent 12 hélices transmembranaires, un autre outils a donc été utilisé PPM server. Avec PPM server seul les modèles générés par Mémoir possèdent 12 hélices transmembranaires. De plus ils possèdent des scores MAIDEN, une énergie favorable et un score de qualité meilleurs que les autres modèles. Ces modèles sont donc nos meilleurs modèles et seront raffinés.

6) Raffinement des modèles

Après raffinement du Modèle 1.4 les modèles raffinés sont moins pertinents que les résultats non raffinés, en effet dans tous les modèles les MolProbity score est plus élevé alors que les score les plus faibles sont associés au meilleur modèle. En ce qui concerne le clash score il est plus élevé dans les modèles raffiner, or un clash score élevé est associé à une structure de faible résolution. Le meilleur modèle est donc celui que l'on possède au début sans raffinement. Les mêmes observations sont retrouvées au cours de l'analyse des modèles obtenues après le raffinement des modèles Modèle 2.4 et Modèle 3.4 (Table V). Les modèles non raffinés sont donc les meilleurs modèles obtenus au cours de notre étude. Ils seront analysés par mode normaux.

Modèle 1.4

| Model | GDT-HA | RMSD | MolProbity | Clash score | Poor rotamers | Rama favored |
|---------|--------|-------|------------|-------------|---------------|--------------|
| Initial | 1.0000 | 0.000 | 1.280 | 5.2 | 0.3 | 98.5 |
| MODEL 1 | 0.9936 | 0.253 | 1.837 | 13.6 | 1.8 | 98.7 |
| MODEL 2 | 0.9957 | 0.243 | 1.798 | 14.0 | 1.6 | 98.9 |
| MODEL 3 | 0.9941 | 0.250 | 1.672 | 14.7 | 0.5 | 98.9 |
| MODEL 4 | 0.9893 | 0.266 | 1.809 | 14.4 | 1.6 | 99.1 |
| MODEL 5 | 0.9909 | 0.259 | 1.827 | 15.1 | 1.6 | 98.5 |

Modèle 2.4

| Model | GDT-HA | RMSD | MolProbity | Clash score | Poor rotamers | Rama favored |
|---------|--------|-------|------------|-------------|---------------|--------------|
| Initial | 1.0000 | 0.000 | 1.604 | 7.8 | 0.6 | 97.0 |
| MODEL 1 | 0.9903 | 0.253 | 1.577 | 11.6 | 0.5 | 99.1 |
| MODEL 2 | 0.9930 | 0.250 | 1.549 | 10.7 | 0.5 | 98.9 |
| MODEL 3 | 0.9930 | 0.253 | 1.533 | 10.3 | 0.5 | 99.1 |
| MODEL 4 | 0.9946 | 0.250 | 1.600 | 12.2 | 0.8 | 98.9 |
| MODEL 5 | 0.9941 | 0.252 | 1.549 | 10.7 | 0.8 | 99.1 |

Modèle 3.4

| Model | GDT-HA | RMSD | MolProbity | Clash score | Poor rotamers | Rama favored |
|---------|--------|-------|------------|-------------|---------------|--------------|
| Initial | 1.0000 | 0.000 | 1.892 | 9.8 | 2.2 | 97.4 |
| MODEL 1 | 0.9898 | 0.273 | 1.590 | 11.9 | 0.8 | 99.4 |
| MODEL 2 | 0.9919 | 0.277 | 1.650 | 13.4 | 1.0 | 98.9 |
| MODEL 3 | 0.9892 | 0.292 | 1.664 | 14.4 | 0.8 | 99.1 |
| MODEL 4 | 0.9865 | 0.289 | 1.660 | 14.3 | 0.5 | 98.7 |
| MODEL 5 | 0.9860 | 0.300 | 1.608 | 12.5 | 0.5 | 99.1 |

Table V : Evaluation des différents modèles raffinés par GalaxyRefine.

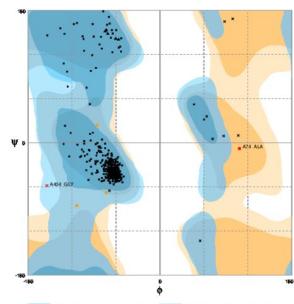
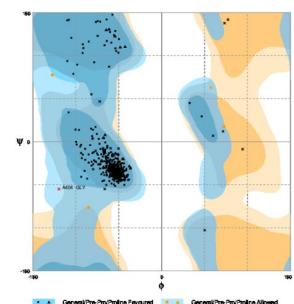
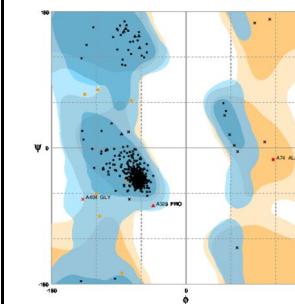
| | Modèle 1.4 | Modèle 2.4 | Modèle 3.4 |
|---------------------------------------------------------|------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------|
| Diagramme de Ramachandran |  |  |  |
| Number of residues in favoured region (~98.0% expected) | 461 (98.9%) | 455 (98.5%) | 451 (97.8%) |
| Number of residues in allowed region (~2.0% expected) | 3 (0.6%) | 6 (1.3%) | 7 (1.5%) |
| Number of residues in outlier region | 2 (0.4%) A404 GLY A74 ALA | 1 (0.2%) A404 GLY | 3 (0.7%) A404 GLY A74 ALA A328 PRO |

Table VI : Evaluation des trois meilleurs modèles par un diagramme de Ramachandran

Les diagrammes de Ramachandran des trois meilleurs modèles (Table VI) a ensuite été généré pour vérifier la robustesse des modèles. Tous les modèles ont quasiment le bon nombre résidus favorables et ont moins que le seuil fixé pour les résidus non favorables. Les mêmes outliers sont retrouvés chez les trois modèles.

Les trois modèles ont ensuite été visualisé sur PyMOL au sein de la membrane, pour les trois modèles le site N glycosylé et retrouvé dans une boucle entre la première et la

deuxième hélices, ce qui est concordant avec la littérature puisque c'est une des caractéristique de la classe 1 des GLUT (fig 6a, 6b, 6c).

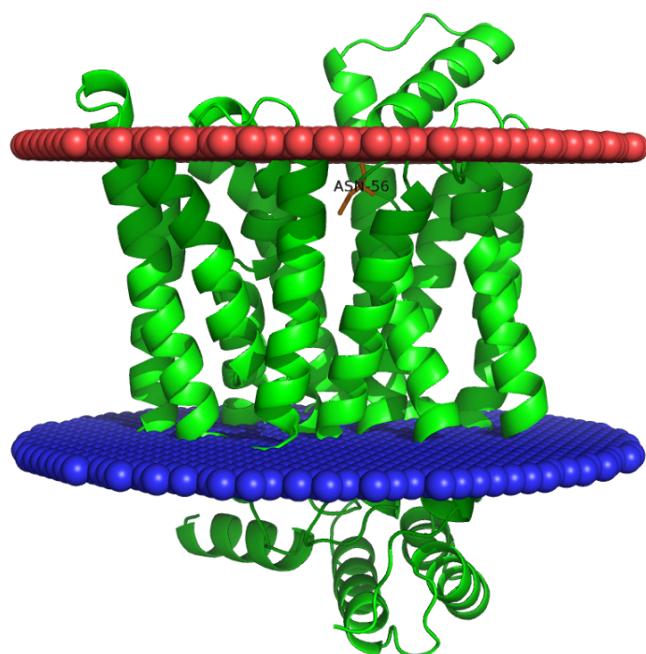


Figure 6a : Représentation de modèle 1.4 (outward-occluded) dans sa membrane. En bleu la membrane interne (cytosolique) et en rouge la membrane externe.

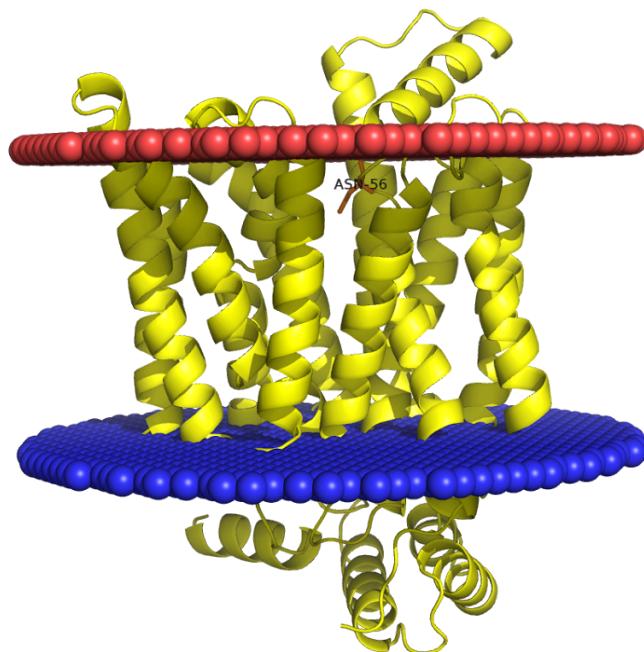


Figure 6b : Représentation de modèle 2.4 (outward-open) dans sa membrane. En bleu la membrane interne (cytosolique) et en rouge la membrane externe.

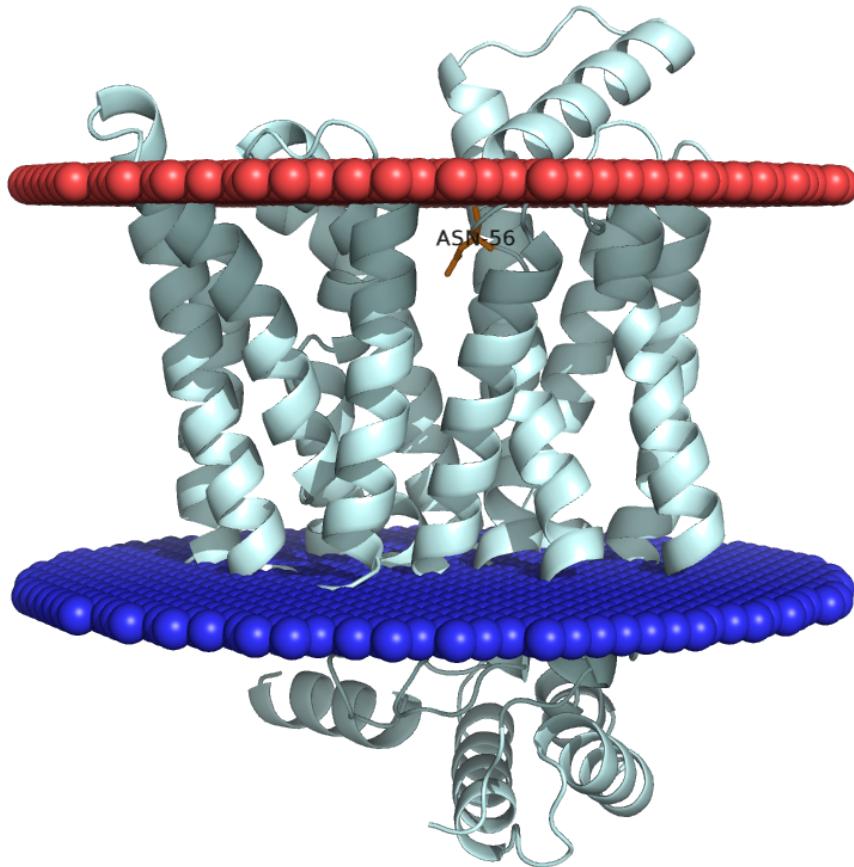


Figure 6c : Représentation de modèle 3.4 (glut3) dans sa membrane. En bleu la membrane interne (cytosolique) et en rouge la membrane externe.

7) Etude de la mécanistique du modèle par les modes normaux

Les modes pertinents pour l'analyse de l'ouverture et la fermeture du canal sont le 14, 10, 12, 11, 15, 13 et 16. Il permettent de visualiser le mouvement de la troisième boucle extracellulaire qui peut interférer dans l'entrée d'un substrat ainsi que la contraction des hélices 11 et 4 permettant l'ouverture ou la fermeture du canal.

Conclusions et Perspectives

Pour conclure, les conformations normal, outward-occluded et outward-open de ont été prédites et semble pertinente d'un point de vue biologique. Les modèles sont composés de 12 hélices transmembranaires avec le N-terminal et le C-terminal orienté vers le cytoplasme donc appartiennent bien à la famille des MFS. Le N glycosylé présenté dans la première boucles extracellulaire caractéristique des GLUT de classe 1 est retrouvé.

Cependant les modèles pourraient être raffiné par dynamique moléculaire ainsi que le mouvement globales de GLUT14.

Références bibliographiques

- Structural advances for the major facilitator superfamily (MFS) transporters. [Yan N¹. Trends Biochem Sci.](#) 2013
- Mammalian Sugar Transporters, Robert Augustin and Eric Mayoux, Chapitre 1
- The SLC2A14 gene, encoding the novel glucose/dehydroascorbate transporter GLUT14, is associated with inflammatory bowel disease [Amir Shaghahi M¹, Zhouyao H¹, Tu H², El-Gabalawy H³, Crow GH⁴, Levine M², Bernstein CN^{3,5}, Eck P⁶.](#), 2017
- Structure validation by Calpha geometry: phi,psi and Cbeta deviation. S.C. Lovell, I.W. Davis, W.B. Arendall III, P.I.W. de Bakker, J.M. Word, M.G. Prisant, J.S. Richardson and D.C. Richardson (2002) *Proteins: Structure, Function & Genetics*. **50**: 437-450.
- Molecular basis of ligand recognition and transport by glucose transporters, Yan, Nature