



Projet de Bioinformatique Structurale :
Prédiction de la structure 3D d'une protéine de la
famille des Major Facilitator Superfamily (MFS)

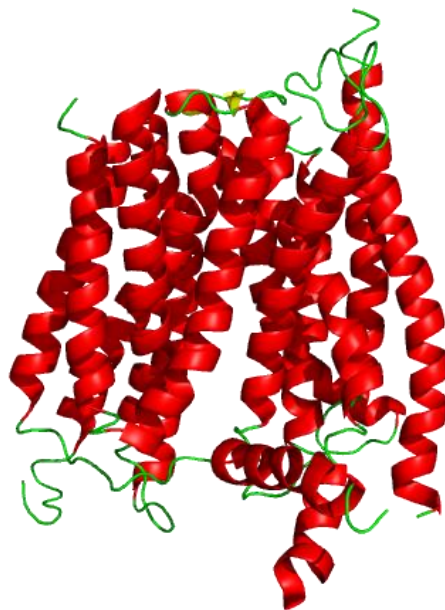


Table des matières

Introduction	3
Matériels et Méthodes	3
Etude de SLC15A5.....	3
BLAST	3
PSI/T-Coffee	3
TMHMM.....	3
Génération des modèles	4
HHpred et MODELLER	4
SWISS-MODEL.....	4
Affinement des modèles.....	4
Evaluation des modèles sélectionnés	4
VERIFY3D.....	4
ProQM.....	4
OREMPRO.....	4
Résultats et discussion.....	4
Etude de SLC15A5.....	4
Sélection des modèles	6
.....	9
Evaluation des modèles sélectionnés	11
Conclusion.....	12
Références	13

Introduction

Les protéines de la famille des Major Facilitaty Superfamily (MFS) appartiennent au groupe le plus large de transporteurs secondaires. [1] Ces transports incluent des uniports, (déplacement de 2 molécules différentes dans le même sens), symports (déplacement de 2 molécules dans le sens contraire) ou antiports (déplacement d'une seule molécule ou ion dans un sens). Les MFS transportent sélectivement un grand nombre de molécules à travers la membrane cellulaire et jouent ainsi un rôle primordial dans de nombreux processus biologiques. Cependant, peu de structures tridimensionnelles de cette famille ont été résolues.

Le projet consiste donc à étudier et prédire la structure 3D d'une protéine appartenant à la famille des MFS.

La protéine choisie pour cette étude est **Solute carrier family 15 member 5**, dont le numéro d'accès sur Uniprot est **A6NIM6**. Elle fait donc partie du groupe des solute carrier (SLC), elle-même faisant partie de la famille des MFS. C'est un transporteur de type symport de proton et d'oligopeptide. Au cours de ce projet, j'ai généré des modèles à partir de la séquence de cette protéine grâce à divers outils de prédiction de structure 3D. Par la suite, j'ai évalué la pertinence de ces modèles grâce à d'autres outils pour sélectionner le meilleur modèle.

Matériels et Méthodes

Etude de SLC15A5

BLAST

BLAST (*Basic Local Alignment Search Tool*) [2] est une méthode qui permet de trouver des régions similaires entre une ou plusieurs séquences. Dans le cadre du projet, j'ai réalisé des blasts entre des séquences protéiques (blastp).

PSI/T-Coffee

PSI/TM-Coffee [3] est serveur web qui effectue des alignements multiples de protéines. Ce serveur est adapté pour les protéines transmembranaires.

TMHMM

TMHMM [4] est une méthode qui permet de prédire la localisation et l'orientation des hélices α dans la membrane. Cette méthode est basée sur les modèles de Markov cachés.

Génération des modèles

HHpred et MODELLER

HHpred [5] est un serveur pour la prédiction de structure. Il utilise la comparaison deux à deux de profils de modèles de Markov cachés. HHpred peut ainsi produire des modèles 3D calculés par le programme MODELLER [6], qui permet de calculer des modèles comparatives.

SWISS-MODEL

SWISS-MODEL [7] est un serveur pour la modélisation comparative de structure 3D de protéines. Il permet de créer un modèle à partir d'un *template* en entrée.

Affinement des modèles

GalaxyRefine [8] est un serveur web qui permet l'affinement de protéines. Il prend en entrée la structure PDB.

Evaluation des modèles sélectionnés

VERIFY3D

VERIFY3D [9] est un outil en ligne qui permet d'évaluer des modèles de protéines. Cette méthode mesure la compatibilité de chaque acide aminé à son environnement local (hydrophobe ou hydrophile). La qualité du modèle 3D est alors mesurée par un profil 3D par la compatibilité du modèle à sa propre séquence en acides aminés.

ProQM

ProQM [10] est aussi un outil en ligne pour évaluer des modèles de protéines. L'avantage de cette méthode c'est qu'elle est adaptée aux protéines membranaires.

OREMPRO

L'orientation et l'évaluation du modèle au sein de la membrane sera effectué à l'aide du serveur web OREMPRO [11]. La qualité de la protéine est quantifiée par un MAIDEN score. Plus ce score est faible, plus le modèle est de bonne qualité.

Résultats et discussion

Etude de SLC15A5

Premièrement, la fiche Uniprot du numéro d'accès A6NIM6 (SLC15A5) nous donne un certain nombre d'informations concernant cette protéine. Sa séquence possède 579 résidus. Elle a été

étudiée chez l'Homme. Elle fait partie de la famille des MFS. Néanmoins, Uniprot nous indique 11 hélices transmembranaires alors que la littérature en indique 12 [12]. Par ailleurs, la littérature nous renseigne sur une longue boucle extracellulaire.

Tout d'abord, j'ai effectué un BLAST dans la base de données Uniprot en ne gardant que les séquences qui possèdent une structure connue dans la Protein Data Bank (PDB).

Parmi les alignements obtenus, j'ai cherché celui dont la couverture est la plus grande et dont le pourcentage d'identité est le plus élevé. Cet alignement est montré sur la Fig. 1.

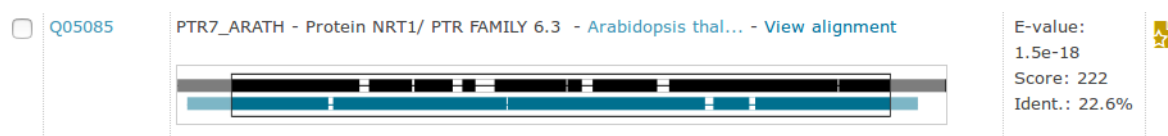


Fig. 1 Alignement obtenu à l'aide d'un BLAST et qui correspond au recouvrement le plus grand et le pourcentage d'identité le plus élevé.

On voit bien que la séquence protéique de Q05085 recouvre quasiment toute la séquence étudiée, hormis les extrémités N-terminale et C-terminale. Avec plus de 20 % d'identité de séquence, on peut donc considérer que ces 2 séquences dérivent d'une ancêtre commun. Par conséquent, j'ai donc effectué une homologie comparative.

Par la suite, j'ai étudié plus en détail la protéine Q05085. Cette structure correspond à une protéine NTR1, un transporteur du nitrate proton-dépendant de la famille des MFS, dans l'organisme *Arabidopsis thaliana*. Dans la PDB, ce code UniprotKB correspond à 3 structures PDB (4OH3, 5A2N et 5A2O). Ces 3 structures sont présentes sous forme de dimère. 4OH3 et

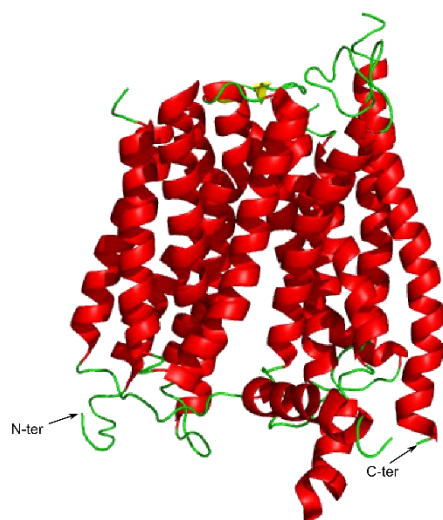


Fig. 2 Représentation de la protéine 5A2N à l'aide de PyMOL (Schrödinger, LLC). En vert : les boucles ; En rouge : les hélices α . Les extrémités N-terminal et C-terminal sont indiqués par des flèches.

5A2O sont en présence de ligands. J'ai choisi d'utiliser 5A2N (Fig. 2) comme *template* pour créer un modèle avec une conformation sans le ligand.

Ensuite, cette protéine étant de la famille des MFS, j'ai étudié plus en détail les zones conservées avec les autres membres de cette famille. Pour cela, à partir d'un BLAST sur la base de données Uniprot, j'ai sélectionné quelques séquences qui ont entre 30 et 40 % d'identité avec ma séquence d'intérêt. Puis j'ai effectué un alignement multiple de ces séquences choisies à l'aide de PSI/TM-Coffee.

Comme il s'agit d'une protéine transmembranaire, j'ai utilisé l'outil TMHMM pour détecter les régions transmembranaires. TMHMM nous propose 11 hélices transmembranaires (Fig. 3), tout comme Uniprot, alors que la littérature sur les SLCA15 nous en indique 12.

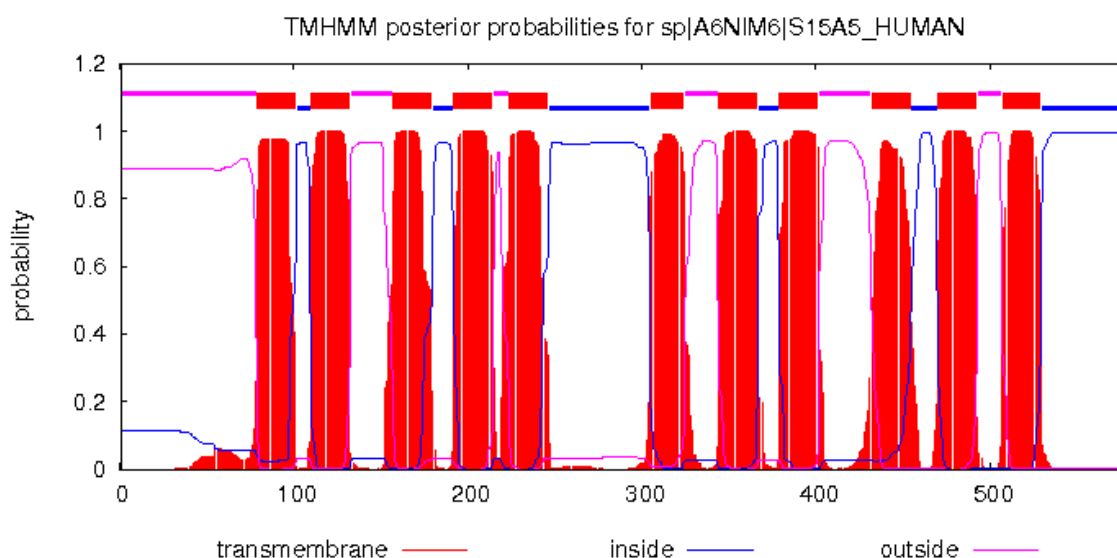


Fig. 3 Représentation de la probabilité d'être à l'extérieur, transmembranaire ou à l'intérieur pour chaque résidu. Figure obtenue à l'aide de TMHMM.

Sélection des modèles

Les 2 modèles obtenus par MODELLER (Fig. 4) et SWISS-MODEL (Fig. 5) nous prédisent bien 12 hélices transmembranaires, avec les extrémités C-terminale et N-terminale dans l'espace intracellulaire. Cependant, on n'observe pas de longue boucle au niveau extracellulaire comme indiqué dans la littérature.

Par ailleurs, avec SWISS-MODEL, on a une très bonne couverture de la protéine.

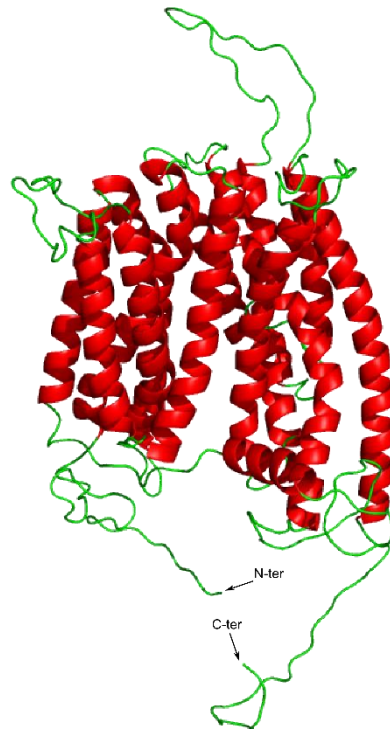


Fig. 4 Représentation du modèle obtenu par MODELLER à l'aide de PyMOL (Schrödinger, LLC). En vert : les boucles ; En rouge : les hélices α . Les extrémités N-terminal et C-terminal sont indiqués par des flèches.

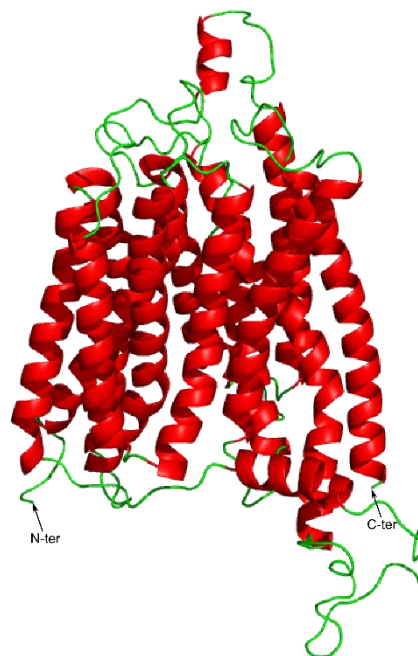


Fig. 5 Représentation du modèle obtenu par SWISS-MODEL à l'aide de PyMOL (Schrödinger, LLC). En vert : les boucles ; En rouge : les hélices α . Les extrémités N-terminal et C-terminal sont indiqués par des flèches.

Sur la Fig. 6, on voit que la prédiction de la similarité locale n'est pas la même selon les résidus pour le modèle de SWISS-MODEL. On a des pics de valeurs. Les valeurs les plus basses correspondent probablement aux boucles entre les hélices. On peut penser que du fait de leur variabilité, ils ont du mal à être bien prédit.

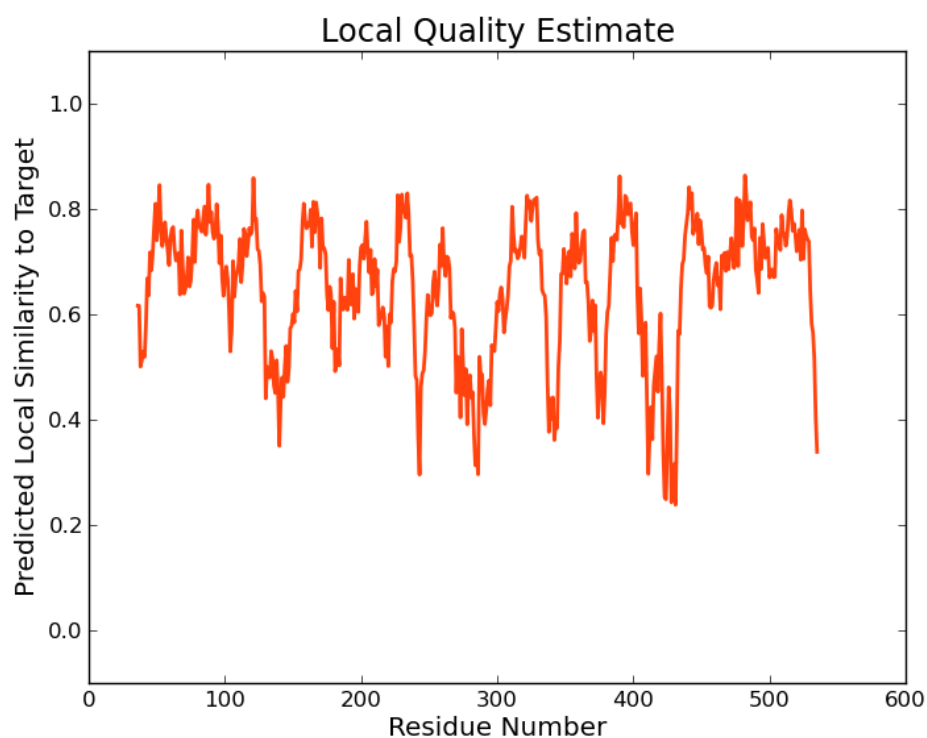


Fig. 6 Représentation de la prédiction de la similarité locale en fonction des résidus. Figure obtenu dans SWISS-MODEL.

Par la suite, j'ai affiné les modèles obtenus à l'aide de GalaxyRefine.

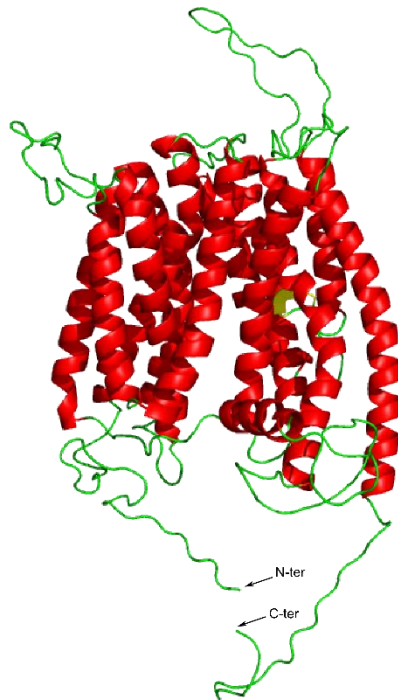


Fig. 7 Représentation du meilleur modèle obtenu par MODELLER et un affinement par GALAXYRefine à l'aide de PyMOL (Schrödinger, LLC). En vert : les boucles ; En rouge : les hélices α . Les extrémités N-terminal et C-terminal sont indiqués par des flèches.

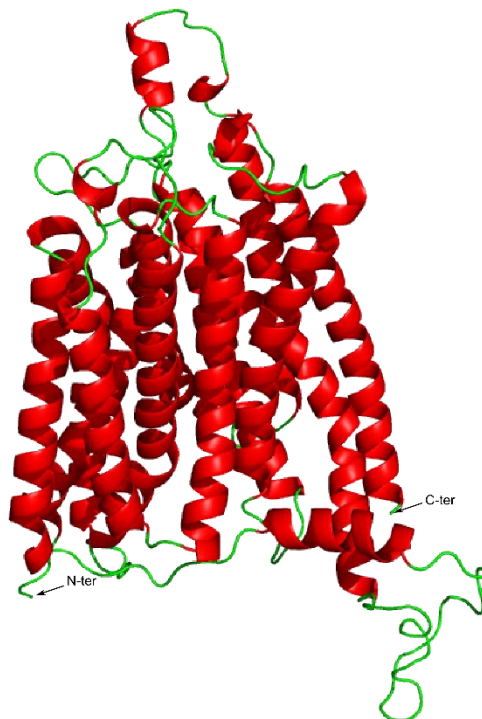


Fig. 8 Représentation du meilleur modèle obtenu par SWISS-MODEL et un affinement par GALAXYRefine à l'aide de PyMOL (Schrödinger, LLC). En vert : les boucles ; En rouge : les hélices α . Les extrémités N-terminal et C-terminal sont indiqués par des flèches.

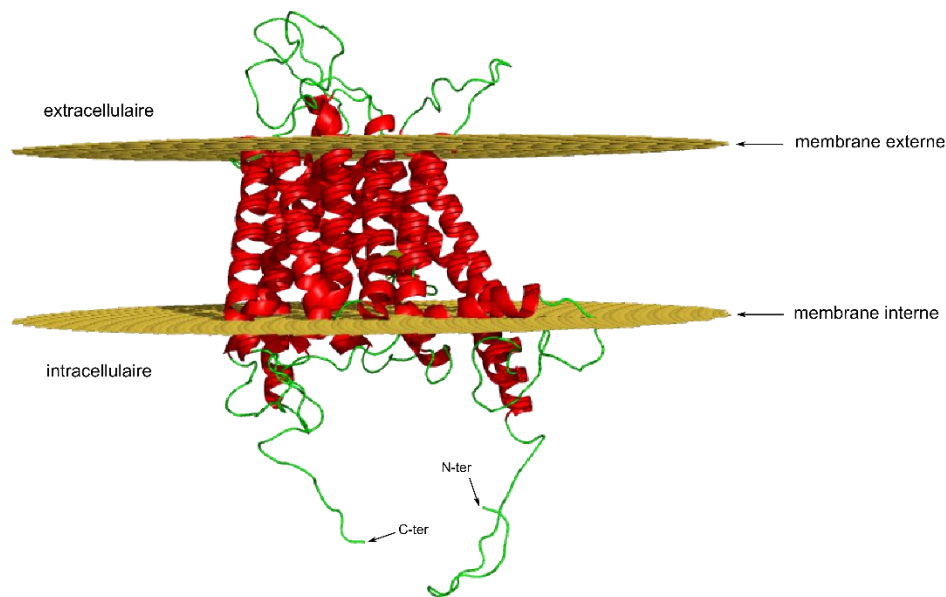


Fig. 9 Représentation du modèle obtenu par SWISS-MODEL avec la membrane plasmique à l'aide de PyMOL (Schrödinger, LLC). En vert : les boucles ; En rouge : les hélices α ; En orange : la membrane interne et externe. Les extrémités N-terminale et C-terminale sont indiquées par des flèches.

Cependant, lorsque l'on rajoute la membrane à l'aide d'OREMPRO, on remarque que la méthode prédit seulement 10 hélices transmembranaires et non 12. C'est pourquoi, malgré l'affinement, j'ai gardé le modèle de base de SWISS-MODEL.

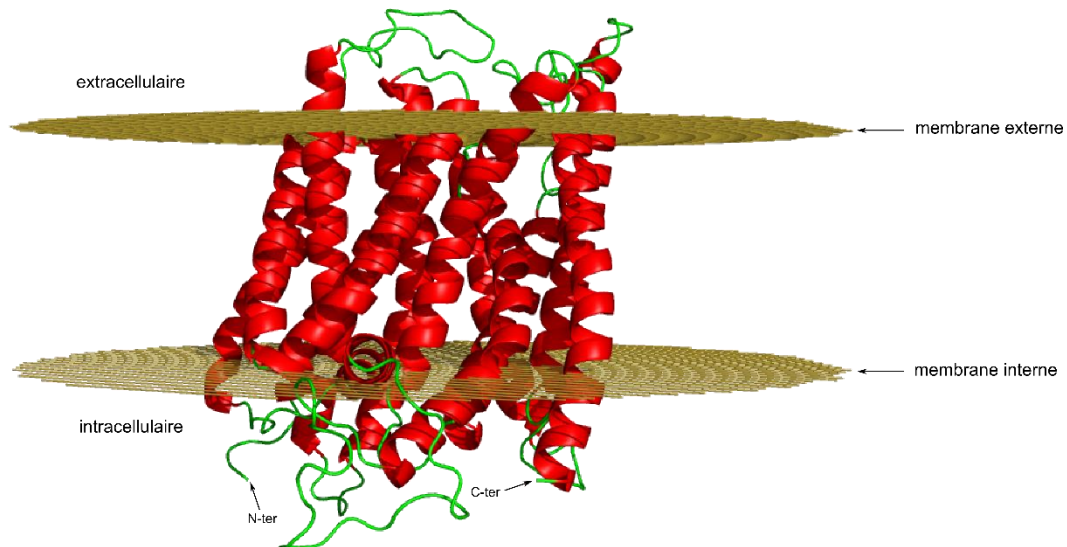


Fig. 10 Représentation du modèle obtenu par MODELLER avec la membrane plasmique à l'aide de PyMOL (Schrödinger, LLC). En vert : les boucles ; En rouge : les hélices α ; En orange : la membrane interne et externe. Les extrémités N-terminale et C-terminale sont indiquées par des flèches.

Pour le modèle de MODELLER, OREMPRO nous renvoie 12 hélices transmembranaires, l'épaisseur de la zone hydrophobe à 27.3 Å et un MAIDEN score de -7.5382.

Pour le modèle de SWISS-MODEL, OREMPRO nous renvoie 12 hélices transmembranaires, l'épaisseur de la zone hydrophobe à 28.5 Å et un MAIDEN score de -101.7232.

On voit que le MAIDEN score est très inférieur pour le modèle de SWISS-MODEL. Ce qui montre que le modèle SWISS-MODEL est de meilleures qualités.

Evaluation des modèles sélectionnés

Les modèles sont évalués à l'aide des programmes Verify-3D et ProQM.

Pour le modèle de Modeller, on a 54.41 % de résidus avec un 3D-1D score moyenne supérieur à 0.2 ; alors que pour le modèle de SWISS-MODEL, on a 58.22 % de résidus avec un 3D-1D score supérieur à 0.2. Même si le modèle de SWISS-MODEL semble meilleur que celui de Modeller, on remarque que ces valeurs sont plutôt proches. On ne peut donc pas vraiment trancher quant à la qualité de ces 2 modèles qui se valent.

De plus, on remarque aussi que ces valeurs de pourcentage de score ne sont pas très élevées, ce qui nous indique que ces modèles ne sont pas très bons.

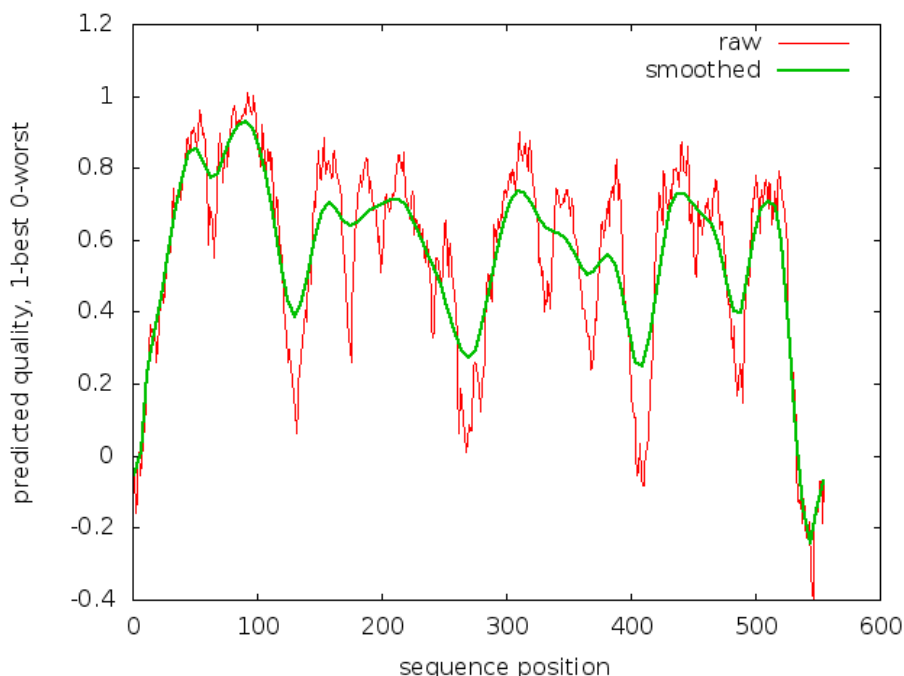


Fig. 11 Pour le modèle de Modeller, courbes de la qualité de la prédiction (1 pour le meilleur et 0 pour le plus mauvais) en fonction de la position de la séquence. Figure obtenue par ProQM.

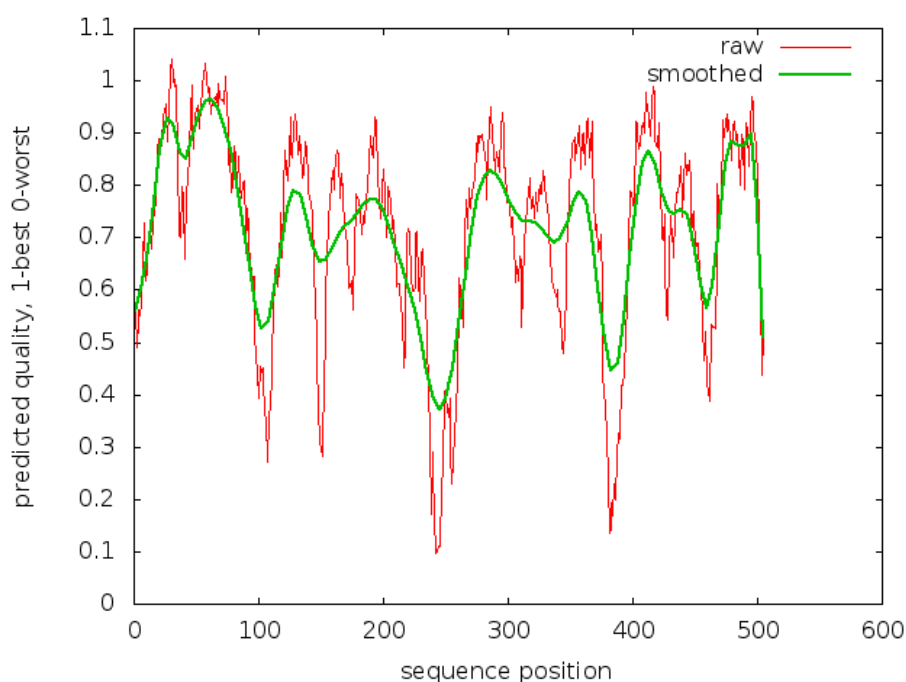


Fig. 12 Pour le modèle de SWISS-MODEL, courbes de la qualité de la prédiction (1 pour le meilleur et 0 pour le plus mauvais) en fonction de la position de la séquence. Figure obtenue par ProQM.

Sur les Fig. 11 et 12, on voit que sur l'ensemble de la séquence, la qualité prédite est globalement plus élevée pour le modèle SWISS-MODEL que le modèle MODELLER. On peut penser que les pics de valeurs faibles de la qualité prédites correspondent à des boucles.

Au vu des différents outils utilisés pour la prédiction et l'évaluation des modèles obtenus, j'ai donc gardé le modèle de SWISS-MODEL comme étant la meilleure prédiction de la structure 3D de SLC15A5.

Conclusion

En conclusion, au cours de ce projet, j'ai donc cherché à prédire la structure 3D de la protéine solute carrier family 15 member 5, de la famille des MFS.

On voit que cette prédiction n'est pas évidente compte tenu des modèles qui peuvent différer selon les outils de prédiction. Cela est notamment dû aux particularités des outils utilisés, en termes d'algorithmes et d'évaluations.

J'aurai pu aller plus loin dans la création du modèle, avec notamment l'utilisation d'autres outils pour confronter d'autres résultats. J'aurai aussi pu étudier la conformation de la protéine en présence de ligand en utilisant les 2 autres *templates*.

De plus, j'aurai pu aller plus loin dans l'analyse du modèle obtenu, avec étude des mouvements de la protéine à l'aide des modes normaux, mais aussi à l'aide d'une dynamique moléculaire à gros grains.

Références

1. Yan, Nieng. « Structural Advances for the Major Facilitator Superfamily (MFS) Transporters ». *Trends in Biochemical Sciences* 38, n° 3 (1 mars 2013): 151-59. <https://doi.org/10.1016/j.tibs.2013.01.003>.
2. Altschul, S. F., W. Gish, W. Miller, E. W. Myers, et D. J. Lipman. « Basic Local Alignment Search Tool ». *Journal of Molecular Biology* 215, n° 3 (5 octobre 1990): 403-10. [https://doi.org/10.1016/S0022-2836\(05\)80360-2](https://doi.org/10.1016/S0022-2836(05)80360-2).
3. Floden, Evan W., Paolo D. Tommaso, Maria Chatzou, Cedrik Magis, Cedric Notredame, et Jia-Ming Chang. « PSI/TM-Coffee: a web server for fast and accurate multiple sequence alignments of regular and transmembrane proteins using homology extension on reduced databases ». *Nucleic Acids Research* 44, n° Web Server issue (8 juillet 2016): W339-43. <https://doi.org/10.1093/nar/gkw300>.
4. Krogh, Anders, Björn Larsson, Gunnar von Heijne, et Erik L. L. Sonnhammer. « Predicting transmembrane protein topology with a hidden markov model: application to complete genomes » Edited by F. Cohen. *Journal of Molecular Biology* 305, n° 3 (19 janvier 2001): 567-80. <https://doi.org/10.1006/jmbi.2000.4315>.
5. Söding, Johannes, Andreas Biegert, et Andrei N. Lupas. « The HHpred interactive server for protein homology detection and structure prediction ». *Nucleic Acids Research* 33, n° Web Server issue (1 juillet 2005): W244-48. <https://doi.org/10.1093/nar/gki408>.
6. Eswar, Narayanan, Ben Webb, Marc A. Marti-Renom, M.S. Madhusudhan, David Eramian, Min-yi Shen, Ursula Pieper, et Andrej Sali. « Comparative Protein Structure Modeling Using Modeller ». *Current protocols in bioinformatics / editorial board, Andreas D. Baxevanis ... [et al.]* 0 5 (octobre 2006): Unit-5.6. <https://doi.org/10.1002/0471250953.bi0506s15>.
7. Schwede, Torsten, Jürgen Kopp, Nicolas Guex, et Manuel C. Peitsch. « SWISS-MODEL: an automated protein homology-modeling server ». *Nucleic Acids Research* 31, n° 13 (1 juillet 2003): 3381-85.
8. Heo, Lim, Hahnbeom Park, et Chaok Seok. « GalaxyRefine: protein structure refinement driven by side-chain repacking ». *Nucleic Acids Research* 41, n° Web Server issue (juillet 2013): W384-88. <https://doi.org/10.1093/nar/gkt458>.
9. Eisenberg, David, Roland Luthy, et James U. Bowie. « [20] VERIFY3D: Assessment of protein models with three-dimensional profiles ». In *Methods in Enzymology*, 277:396-404. Macromolecular Crystallography Part B. Academic Press, 1997. [https://doi.org/10.1016/S0076-6879\(97\)77022-8](https://doi.org/10.1016/S0076-6879(97)77022-8).
10. Ray, Arjun, Erik Lindahl, et Björn Wallner. « Model Quality Assessment for Membrane Proteins ». *Bioinformatics (Oxford, England)* 26, n° 24 (15 décembre 2010): 3067-74. <https://doi.org/10.1093/bioinformatics/btq581>.
11. Postic, Guillaume, Yassine Ghouzam, et Jean-Christophe Gelly. « OREMPRO web server: orientation and assessment of atomistic and coarse-grained structures of membrane proteins ». *Bioinformatics* 32, n° 16 (15 août 2016): 2548-50. <https://doi.org/10.1093/bioinformatics/btw208>.

12. « Catalog of 238 variations among six human genes encoding solute carriers (hSLCs) in the Japanese population | Journal of Human Genetics ». Consulté le 25 novembre 2017. <https://www.nature.com/articles/jhg200295>.