

Projet de Bioinformatique Structurale

Le but de ce projet est de se familiariser avec les outils de Bioinformatique et de Modélisation Moléculaire grâce à l'étude de Q9Y6L6, une protéine appartenant à MFS (Major Facilitator Superfamily), une grande famille de transporteurs. Après la présentation des Matériels & Méthodes, les résultats seront donnés dans l'ordre de la démarche utilisée pour modéliser la structure de la protéine.

Matériels & Méthodes

Recherche d'informations : Uniprot, PFAM, STRING, PDB RSCB, Pubmed

Prédiction de structures secondaires ou autres : DIANNA, PSIPRED, TMHMM, TMPRED, HCA

Visualisation : Pymol, WEBLOGO

Alignement multiple : T-COFFEE, MAFFT, MUSCLE

Recherche de template : BLAST, HHPRED, HHBLITS, MUSTER, SPARKS-X, PHYRE2, SWISSMODEL, GenTHREAD

Modélisation de structure 3D : MODELLER, MEMOIR, SWISSMODEL, PHYRE2

Test de la qualité d'un modèle : OREMPO, PROSA, QMEMBRANE, QMEAN

Affinement des boucles : GALAXYWEB

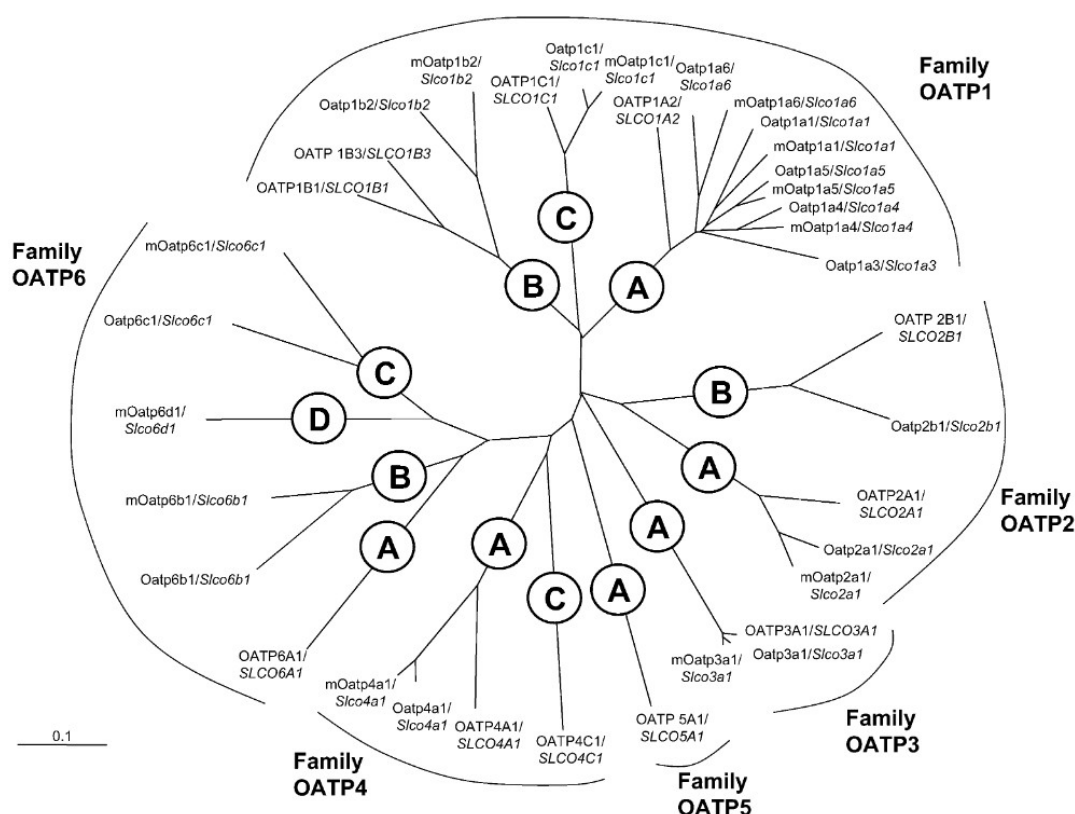


Figure 1 : Classification phylogénétique avec la nouvelle nomenclature des membres de la superfamille des transporteurs membranaires OATP/SLCO. Au sein de chaque famille, il y a au moins 40% d'identité de séquence protéique et 60% d'identité au sein de chaque sous-famille.

Résultats

Superfamille SLCO

La superfamille SLCO (Solute carrier organic) est composée des familles OATP (Organic anion transporter polypeptide), OAT (Organic anion transporter) et OCT (Organic cation transporter). Elle-même appartient à la superfamille des MFS, plus spécifiquement à SLC (Solute carrier). A l'origine, SLCO avait pour nom SLC21. Mais, étant donné qu'il y a une grande divergence entre la famille SLC21 et les autres SLC, notamment au niveau des substrats, cette famille a été mise un peu à part et est désormais connue sous le nom de SLCO [2].

Famille OATP

La famille OATP est composée de six membres (Figure 1). Elle regroupe des transporteurs membranaires de la plupart des anions organiques à travers la membrane cellulaire. La fonction de ces protéines est la sécrétion d'acides biliaires et d'hormones via la membrane basolatérale des hépatocytes majoritairement [7] (Figure 2).

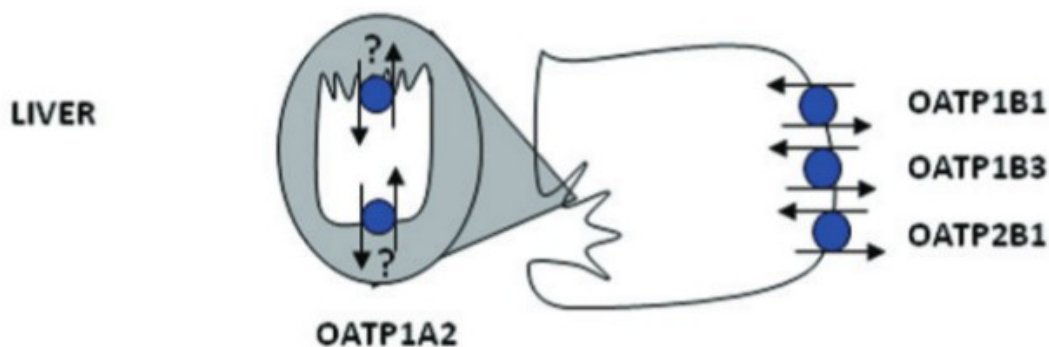


Figure 2 : Exemple de localisation de la protéine étudiée OATP1B1. Elle se trouve au niveau basolatérale de la membrane plasmique. Il y a également d'autres transporteurs de la même famille.

Les OATPs sont présents chez beaucoup d'animaux mais pas chez les bactéries. De plus, les homologues des OATPs humains proviennent vraisemblablement de duplication de gènes. Ces OATPs sont également retrouvés chez les insectes, ce qui montre une évolution tôt dans la formation du règne animal [4].

Généralement, les protéines de la famille OATP ont une séquence entre 643 et 724 acides aminés [6, 7], 12 domaines transmembranaires, les deux termini dans la partie intracellulaire [5, 6, 7] (Figure 3). Les boucles extracellulaires 2 et 5 contiennent plusieurs sites de N-glycosylation dont la localisation peut varier selon les protéines. Ces sites de glycosylation sont conservés chez la plupart des isoformes OATPs [5, 7]. La boucle extracellulaire 5, entre les segments transmembranaires 9 et 10, est très large avec 11 cystéines conservées impliquées dans les ponts disulfures [5, 7]. Entre la boucle extracellulaire 3 et le segment transmembranaire se trouve une signature de la famille OATP qui est : D-X-RW-(I,V)-GAWW-X-G-(F,L)-L [7]. Nous retrouvons cette signature dans la séquence de notre protéine (DSRWVGAWWLNFLVSGLF), ce qui prouve bien son appartenance à cette famille.

En pharmacologie, les protéines OATP1B1 et OATP1B3 ont un rôle important dans le transport des médicaments.

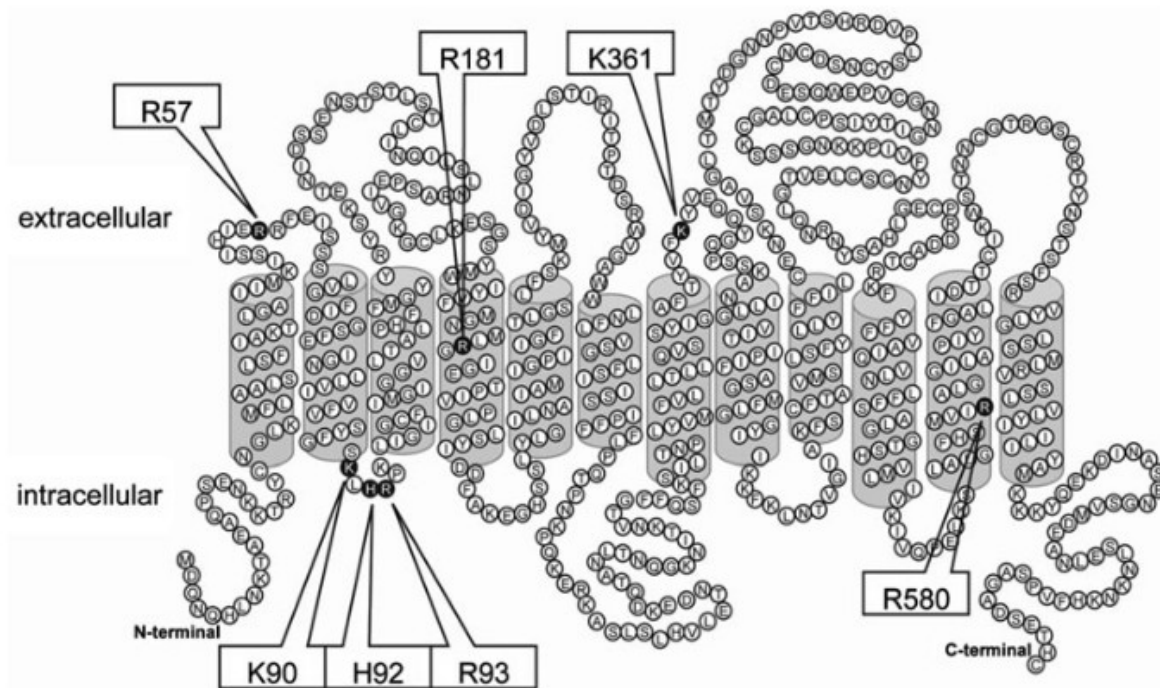


Figure 3 : Représentation de la protéine OATP1B1. Les résidus importants pour le site de fixation du substrat retrouvés dans notre protéine sont K361 et R580. les deux termini se trouvent dans la partie intracellulaire.

Protéine Q9Y6L6

La protéine qui nous concerne se nomme « Solute carrier organic anion transporter family member 1B1 », LST1, OATP2, OATP1B1, OATP-C ou encore SLC21A6. Elle est synthétisée à partir du gène *SLCO1B1* situé sur le chromosome 12. Exprimée au niveau de la membrane plasmique des hépatocytes, c'est un transporteur transmembranaire d'ions organiques (bile, sodium, hormone thyroïdienne). Cette protéine humaine est impliquée dans beaucoup de pathologies et est la cible d'un nombre important de médicaments. Bien qu'ayant une séquence polymorphe, *SLCO1B1* est composée de 691 acides aminés et de 12 hélices transmembranaires d'environ 20 acides aminés chacune [4]. Plusieurs de ses sites sont considérés comme important pour la glycosylation (résidus 130, 134, 432, 503, 516, 617), pour les ponts disulfures (459/485, 463/474, 465/489) ainsi que pour la phosphorylation des sérines (293, 295, 672, 682). Trois résidus sont considérés comme étant important pour le site de fixation du substrat (R51, K361, R580) et seulement les deux derniers se trouvent dans notre protéine (sans doute à cause du polymorphisme) [7, 8] (Figure 4). Avec PSIPRED, on trouve des désordres dans la protéine.

Selon la base de données PFAM, ce membre de la famille OATP contient :

- 1) un domaine Kazal-like 2, souvent en tandem. Il est constitué d'une hélice alpha centrale, d'un court feuillet bêta anti-parallèle composé de deux brins, ainsi que de plusieurs ponts disulfures (Figure 5).
- 2) un domaine transmembranaire hélice (une des signatures de la famille MFS).
- 3) un domaine OATP.

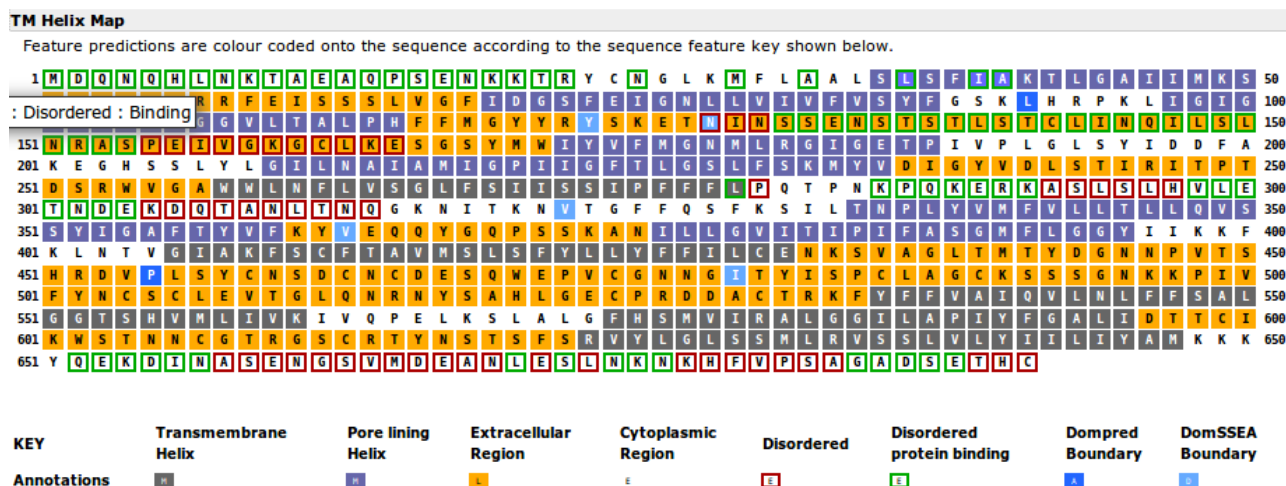


Figure 4 : Résultat de PSIPRED. Représentation de la séquence protéique. En bleu et gris, ce sont les résidus de l'hélice, les résidus jaunes montrent les régions extracellulaires, les encadrements des résidus en rouge et vert montrent les désordres et en bleu clair sont mis en évidence les domaines de liaison.

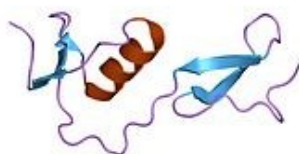


Figure 5 : Structure du domaine Kazal 2 (image provenant de la base de données PFAM)

Prediction des caractéristiques de la protéine (Tableau 1)

Predicted bond by DIANNA
24 - 485
101 - 504
142 - 463
162 - 412
430 - 607
459 - 524
465 - 489
474 - 506
530 - 613
599 – 691

Tableau 1 : Prédiction de ponts disulfures par le programme DIANNA.

Modélisations de l'OATP1B1

1) OATP1B1 n'a pas de structure 3D résolue, donc n'a pas de PDB. Cependant, la fiche Uniprot indique qu'un modèle a été proposé avec le template 1LDT, c'est-à-dire la protéine « Leech-derived tryptase inhibitor C » (Figure 6). Ce template, avec une résolution de 1.90 Angströms, est indiqué comme ayant 42 % d'identité de séquence avec notre protéine cible.

La fiche Uniprot de cette protéine indique qu'elle a 46 acides aminés et a été résolue deux fois par X-ray et quatre fois par NMR. Pratiquement la totalité de cette séquence est un domaine Kazal.

Le modèle réalisé sur les résidus 463 à 514 de l'OATP1B1 montre donc le domaine Kazal du transporteur. Il contient les trois ponts disulfures (459/485, 463/474 465/489) et la glycosylation sur le résidu 503 et, les caractéristiques du domaine (deux brins bêta et une petite hélice). Cependant, il reste la modélisation de la structure 3D de la plus grande partie de la protéine, avec les hélices transmembranaires.



Figure 6 : Modélisation de la structure 3D de Q9Y6L6 avec comme support 1LDT.
Seulement le domaine Kazal a pu être prédite.

2) Un modèle par homologie a également été généré avec Phyre2 et basé sur une protéine d'*Escherichia Coli* (glycerol3 phosphate transporter) afin de construire un modèle 3D putatif des OATPs [7] (Figure 7). Cette image montre la protéine OATP1B1 avec une vue du côté extracellulaire et une autre depuis l'intérieur de la bicouche lipidique (les domaines transmembranaires 2 et 4 sont enlevés). En rouge sont représentés les acides aminés importants de la protéine. Dans ce cas-là, nous avons bien la représentation de la partie transporteur de la protéine. Si nous observons attentivement le modèle de droite, nous pouvons voir qu'il y a des boucles « coupées ». De plus, il manque le domaine Kazal, identifiable par son feuillet bêta.

Le but de ce projet est d'essayer de faire la modélisation de la structure tridimensionnelle de la protéine Q9Y6L6.

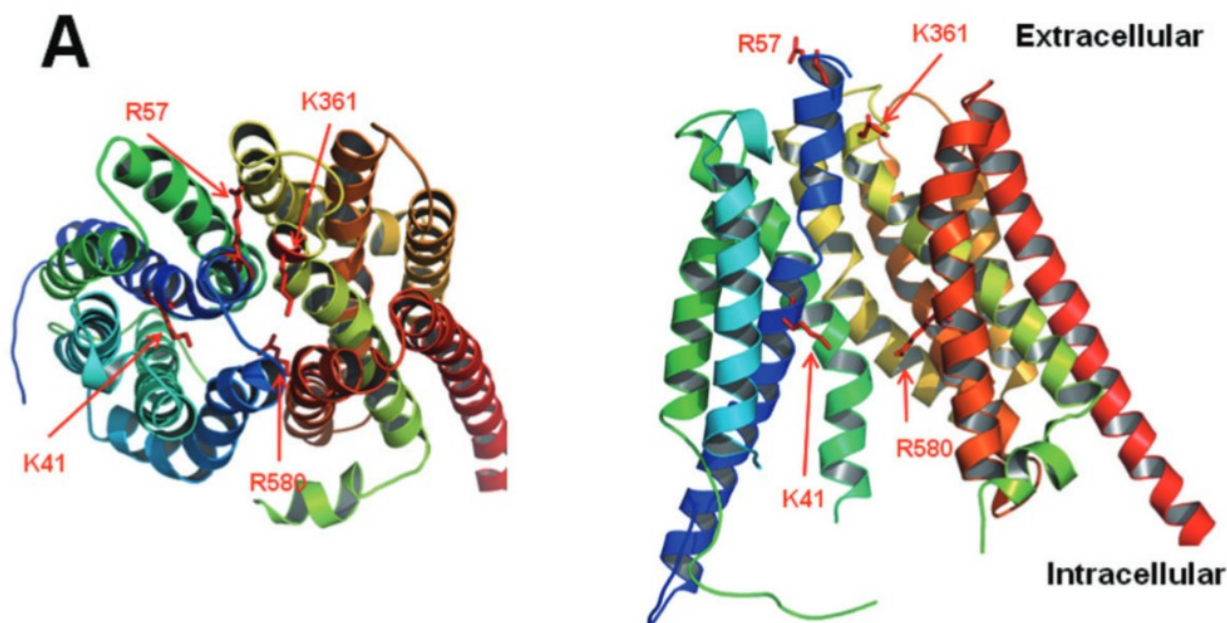


Figure 7 : Modèle par homologie d'un membre de la famille SLCO (OATP1B1), généré avec Phyre2 et basé sur le "Glycerol 3 Phosphate Transporter" de Escherichia Coli. A gauche, c'est une vue extracellulaire de la protéine et à droite, une vue depuis la bicouche lipidique en ayant enlevé les segments transmembranaires 2 et 4 pour plus de clarté.

Caractérisation des acides aminés importants pour la fonction par comparaison avec d'autres membres de la famille

Dans un premier temps, nous décidons de travailler avec les membres de la même sous-famille que OATP1B1. Pour cela, sur Uniprot, nous obtenons 7 résultats avec le mot-clé OATP1. Ensuite, un alignement multiple est réalisé avec notre protéine et les résultats en utilisant T-COFFEE, MUSCLE et MAFFT. Il est pertinent d'utiliser une présentation des séquences en Weblogo fin de visualiser les conservations d'acides aminés. Le but a été de trouver les acides aminés important pour le site de fixation du substrat. Avec ces trois alignements, les sites K361 et R580 sont retrouvés et bien conservés (Figure 8). Beaucoup d'autres sites sont conservés également mais ne nous intéresse pas spécialement à ce stade. Une remarque qui pourrait être faite est que les trois alignements donnent globalement le même résultat mas pas dans la précision. En effet, les deux acides aminés conservés de notre protéine ne sont pas à la même place selon l'alignement multiple considéré.

Dans un deuxième temps, nous avons décidé de prendre plus de protéines. Pour cela, la même méthode est réalisée avec les 40 protéines « reviewed » de la famille OATP. Avec le mot clé « OATP », Uniprot donne 4 190 résultats, mais la sélection des « reviewed » est plus sûre puisque c'est une annotation manuelle et vérifiée. Dans ce cas-là, seul R580 est retrouvé comme étant conservé dans la famille OATP.

L'alignement multiple avec TCOFFEE prend du temps quand le nombre de séquences augmente. Par contre, nous retrouvons bien les 12 hélices transmembranaires. De plus, il faudrait vérifier si le site R580 est réellement conservé en prenant l'ensemble des 4 190 résultats. S'il est conservé, il faudrait également voir cette question au niveau de l'ensemble de la superfamille SLCO.

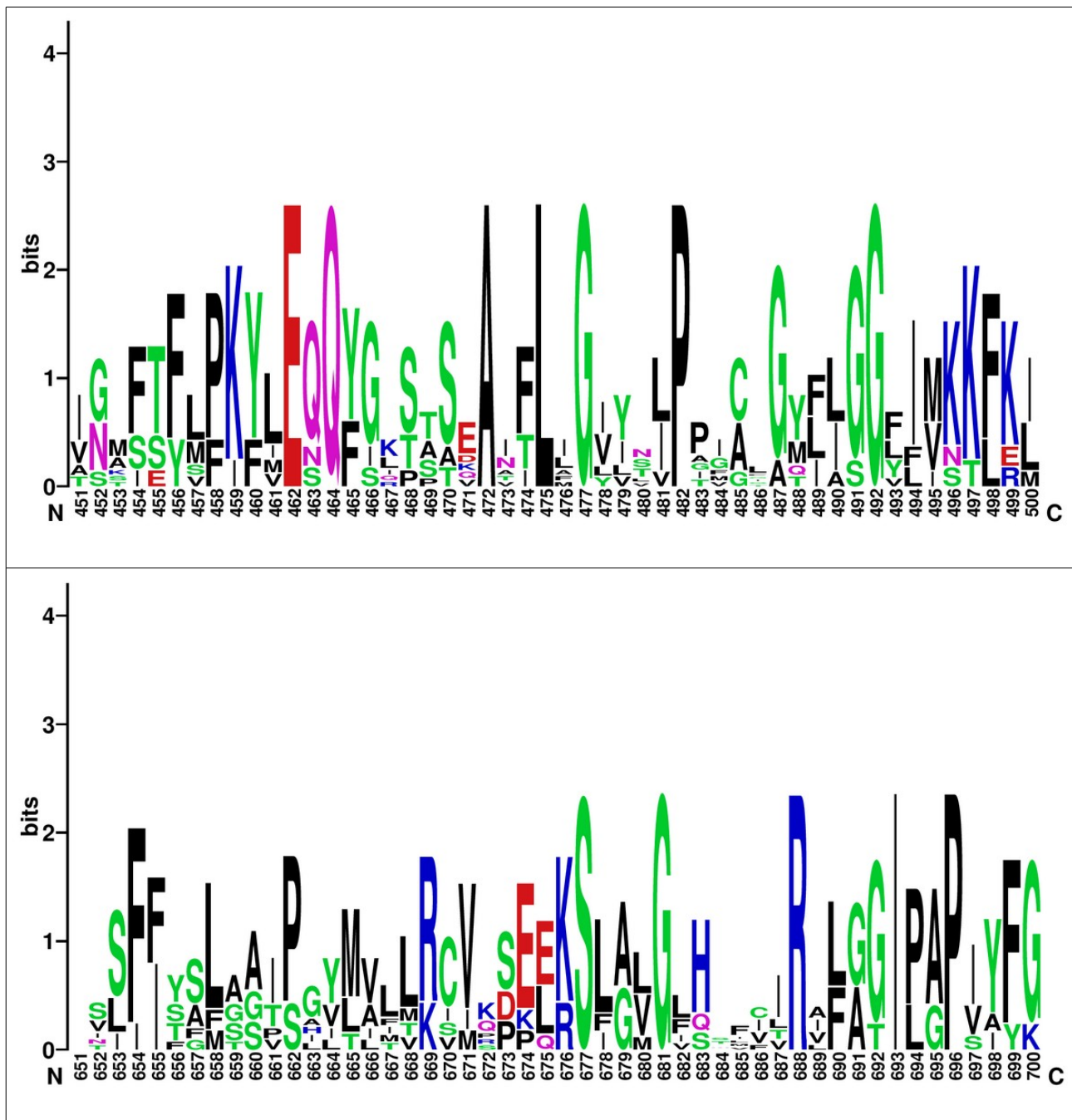


Figure 8 : Représentation en Weblogo de la conservation des résidus dans la famille des OATPs. Notamment, nous notons la conservation des résidus important dans le site de fixation du substrat : en haut, K en position 459 et en bas, R en position 688.

Détection des segments transmembranaires

Pour la détection des hélices transmembranaires, nous avons utilisé des méthodes de prédiction. **HCA** permet la visualisation des segments (Figure 9). Nous voyons bien qu'il y a une grande majorité d'hélices alpha (résidus représentés en vert), correspondant aux hélices transmembranaires. Avec cette méthode, nous avons à peu près leur localisation dans la séquence mais non précise. De plus, il est difficile de compter le nombre de séquences présentes.

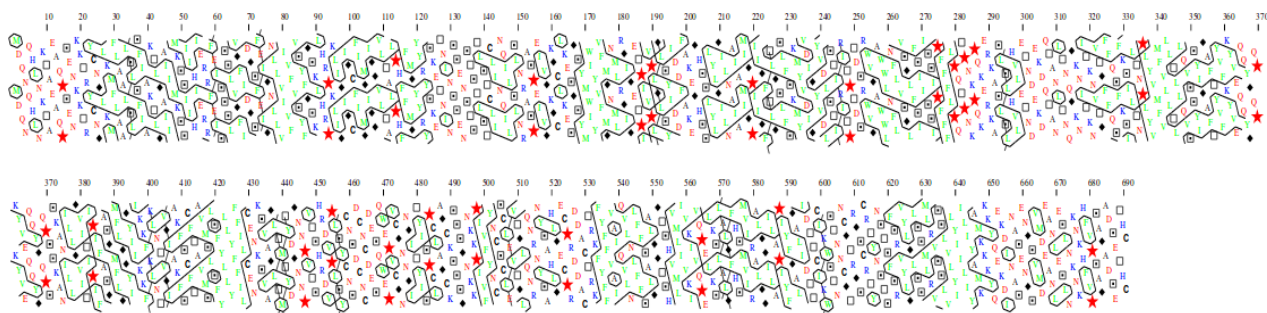


Figure 9 : représentation des parties transmembranaires de la protéine (résidus verts) par la méthode HCA.

TMHMM ne prédit que 11 segments transmembranaires (Tableau 2). De plus, la partie N-terminale est bien prédite au niveau intracellulaire, ce qui n'est pas le cas pour la partie C-terminale de la protéine.

TMPRED prédit bien les 12 segments transmembranaires (Tableau 2). Cependant, là aussi uniquement le N-terminal est prédite en intracellulaire. De plus, sachant que notre protéine a 691 résidus, il y a un problème au niveau du segment 12.

En comparant les trois méthodes, les positions prédites par TMHMM correspondent un peu à celles de HCA. En tout cas, ces méthodes nous prédisent plus de 10 segments transmembranaires.

Méthode	TMHMM	TMPRED
Segment 1	30-52	111-134
Segment 2	67-89	151-174
Segment 3	96-118	182-207
Segment 4	208-230	255-282
Segment 5	257-279	293-316
Segment 6	332-354	345-365
Segment 7	375-397	418-442
Segment 8	407-429	462-483
Segment 9	536-558	496-516
Segment 10	573-595	622-646
Segment 11	624-646	656-681
Segment 12	-	710-733

Tableau 2 : Prédiction des segments transmembranaires avec différentes méthodes.

Recherche de similitude de séquence par PDB

Cette partie permet de trouver des homologues - ayant déjà une structure - à OATP1B1, notre protéine afin de pouvoir les utiliser comme modèle (template) pour modéliser sa structure.

Alignement séquence-séquence

Nous commençons par ce type d'alignement qui se réalise avec BLAST en prenant comme base de données 3D structure et les paramètres par défaut. Le taux d'identité des séquences

protéiques avec notre séquence cible va de 27.3 % à 58.8 %. Ceci correspondrait à une modélisation par homologie. Mais, il n'y a pas de couverture maximale de notre séquence et les résultats proposés ne correspondent pas à des transporteurs, sauf pour Q92959 avec 33.4 % d'identité et couvrant une grande partie de notre cible. L'alignement est fait sur les 643 acides aminés de l'homologue.

En regardant de plus près sa fiche Uniprot, la protéine provient du gène SLCO2A1 (Solute carrier organic anion transporter family member 2A1) de l'humain. Sa structure a été résolue par la méthode X-ray à une résolution de 1.60 Angströms. Elle a un domaine Kazal-like et 12 hélices transmembranaires. Seulement la chaîne P de la structure cristallisée correspond à la protéine Q92959, ce qui correspond à 10 acides aminés. Le reste correspond à deux autres fragments de deux autres protéines des gènes HLA-A et B2M. Donc, Q92959 ne peut pas servir de template.

Alignement séquence-profil

Ceci permet de rendre possible la détection d'homologues plus distants. La recherche de séquences homologues pour notre cible se fait donc grâce à l'établissement de profils afin de réduire la distance évolutive. Nous utilisons HMMER pour obtenir finalement 9 hits au total allant de nombre de positions alignées de 38 à 318. La Evalue n'est pas assez importante ($2.4e-9$). Le meilleur résultat obtenu est 3o7p et 3o7q, deux structures provenant d'une même protéine, la L-fucose_proton symporter (3o7q étant le meilleur car il est plus complet _ une partie de la protéine n'est pas présente chez 3o7p) ([Figure 10](#)). C'est en fait une structure cristallisée d'un transporteur MFS d'une seule chaîne de 438 résidus, le FucP, dans une conformation ouverte (« outward ») de la bactérie Escherichia Coli. Un des problèmes est que la méthode Xray est faite avec une résolution de 3.14 Angström. Sa fiche Uniprot (P11551) montre que ce transporteur a 12 hélices transmembranaires et des domaines MFS et FucP. Nous pouvons le garder comme template mais nous préférons quand même voir ce que propose d'autres méthodes.

Alignement profil-profil

Nous utilisons donc HHBLITS qui nous donne 250 réponses. Les critères de sélection pris en compte sont une forte probabilité, une Evalue faible et une taille de séquence environ égale à celle de notre protéine. Le problème est qu'avec cette méthode nous avons obtenu des transporteurs pratiquement de la même superfamille que OATP1B1 mais sans structure. Cette méthode n'a en effet pas de paramètres pour sélectionner une base de données spécifique à notre recherche. Donc cette méthode ne nous donne pas de résultat.

Avec HHPRED, ayant comme base de données PDB_mmCIF70_28_Oct, nous obtenons 172 résultats. Il détecte 11 hélices transmembranaires dans notre protéine cible, ce qui a été prédit avec d'autres méthodes auparavant. Nous prenons les 10 premiers.

Avec SWISSMODEL, les meilleurs modèles ont été obtenus avec 3wdoA et 4zw9, ces templates (nous donnons à chaque fois l'identifiant PDB) ayant été trouvés avec la méthode HHBLITS.

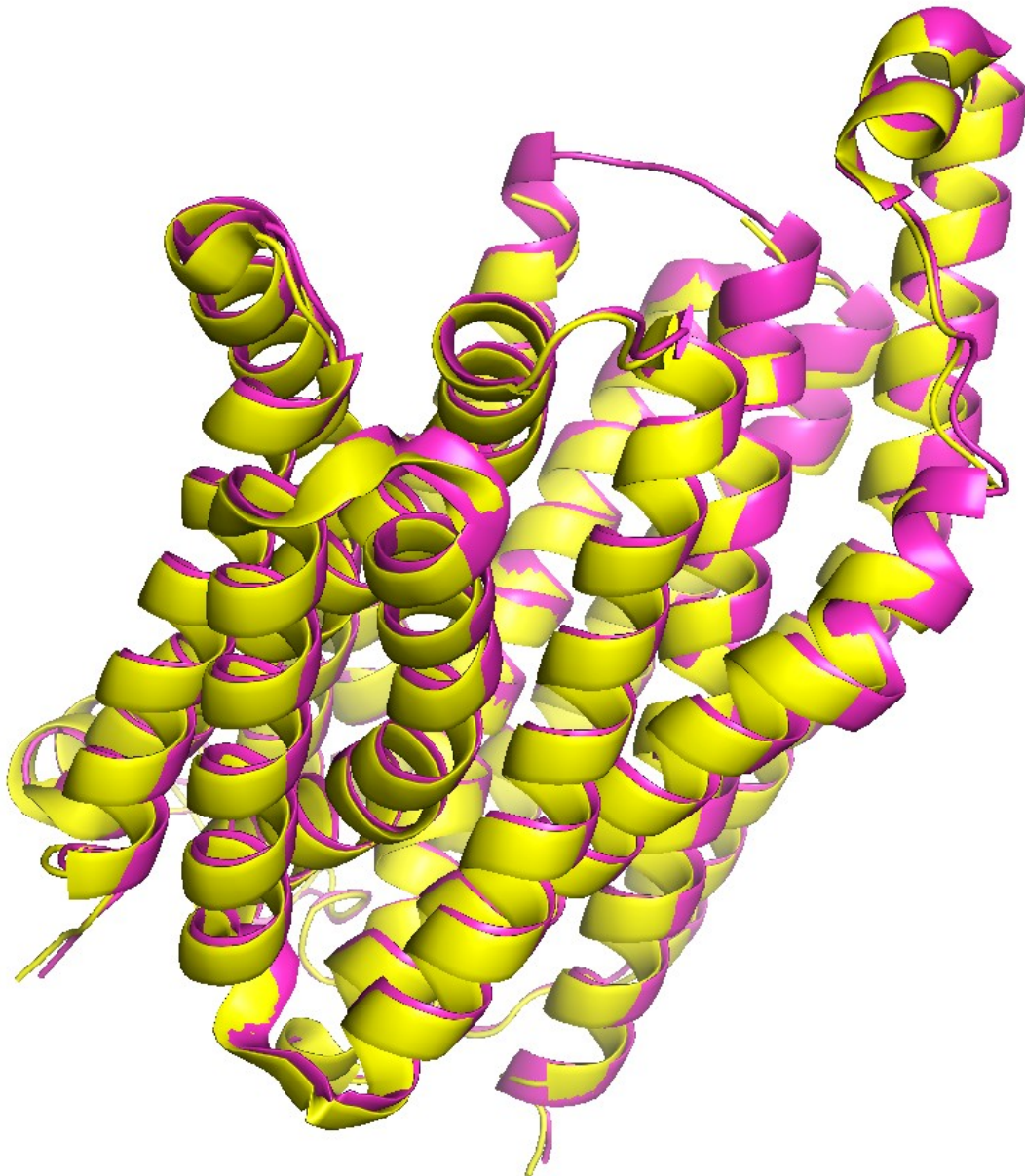


Figure 10 : Représentation sur Pymol des structures 3o7p (jaune) et 3o7q (violet). 3o7q paraît plus complet que 3o7p.

Threading

Nous utilisons également des méthodes de threading pour avoir un aperçu des top 10 des templates proposés (Tableau 2). Une méthode n'est pas indiquée dans le tableau et est Genthread qui propose comme meilleur solution 3wdoA, 1pw4A et 3o7qA (homologue trouvé par HMMER).

Méthode	MUSTER	Phyre2 (confidence, id)	SPARKS-X (zscore, seqID)	HHPRED
1	3wdoA	1pw4a (100, 16)	4fk7A (7.35, 6.6)	1pw4A
2	1pw4A	4ldsB (99.9, 16)	5ijoJ (6.96, 7.2)	4ldsA
3	4zowA	4zp0A (99.9, 17)	4lepA (6.07, 6.8)	4zw9A
4	4gbyA	3o7pA (99.9, 16)	4ifqA (5.93, 7.6)	4ikwA
5	4lepA	4cl5B (99.9, 11)	4ui9A (5.92, 5.7)	4ikvA
6	4jo5A	4ybqB (99.9, 12)	4av3A (5.88, 8.2)	4j05A
7	4ikvA	4ikyA (99.9, 14)	4k7hA (5.85, 5.8)	4tpgA
8	3wajA	3wdoA (99.9, 17)	2x1gF (5.82, 6.9)	4tphB
9	4iuwA	4apsB (99.9, 13)	5hb4B (5.77, 5.6)	5ay0A
10	4cl4A	4gbzA (99.9, 15)	2yevA (5.76, 6.5)	5aynA

Tableau 2 : Méthodes permettant de trouver des templates avec le threading (MUSTER, Phyre2, SPARKS-X) ou la comparaison par homologie (HHPRED). Pour MUSTER, tous les résultats sont indiqués comme étant mauvais. Pour Phyre2 sont indiqués le score de confiance et le pourcentage d'identité avec la séquence cible. Pour SPARKS-X, si le zscore est supérieur à 6, le template est considéré comme étant bon. Cette méthode indique également le pourcentage d'identité de séquence. HHPRED est mis dans ce tableau afin de pouvoir comparer les templates obtenus avec ceux de la méthode par threading.

Conclusion

En regardant toutes ces méthodes, deux templates reviennent souvent : 1pw4A et 3wdoA. Nous allons faire de la modélisation par homologie et décidons de choisir ces deux templates comme support pour modéliser la structure 3D de notre molécule. En plus d'être repéré par plusieurs méthodes, les deux protéines appartiennent à MFS. 1pw4A a une longueur de 451 acides aminés et 3wdoA une longueur de 434 acides aminés. Nous savons d'ores-et-déjà que ce ne sont pas de bons supports car ils ne pourront pas représenter le domaine de Kazal avec ses feuillets bêta.

Comparaison avec les conservations identifiées

Nous utilisons TCOFFEE pour faire un alignement multiple avec les trois séquences (protéine cible, 1pw4A et 3wdoA). Comme précédemment, nous réalisons un Weblogo qui nous permet de voir une absence des deux sites conservés (K361 et R580) de notre protéine (Figure 11).



Figure 11 : Résidus non conservés (à gauche, K361 et à droite, R580) suite à l'alignement des séquences protéiques Q9Y6L6, 1pw4A et 3wdoA.

Prédiction modèle 3D

Malgré cela, nous décidons de continuer afin de découvrir les autres outils de modélisation. Il est effet très difficile de trouver un support pour notre protéine comme nous l'avons pu voir. Il nous faudrait au moins une structure résolue d'une protéine de la famille des OATPs.

MODELLER

Les 10 templates trouvés avec HHPRED sont sélectionnés pour effectuer un alignement multiple et une modélisation par MODELLER. Toutes ces méthodes se font automatiquement les unes à la suite des autres. La modélisation n'est pas obtenue car un problème non identifié intervient.

SWISSMODEL

Ces deux templates sont aussi donnés par cette méthode. Ce modèle permet un ajustement des corps rigides sur la matrice puis de reconstruire les parties non conservées (boucles et chaînes latérales)(Figure 12). La remarque qui pourrait être faite ici est que des brins bêta sont présents dans le modèle, en plus des hélices. Le modèle obtenu avec 1pw4A prédit très mal notre protéine car il a moins de 10 hélices. Un autre souci est rencontré dans cette partie. En regardant la structure sur Pymol grâce au fichier PDB obtenu, nous apercevons une absence de brins bêta (Figure 12).

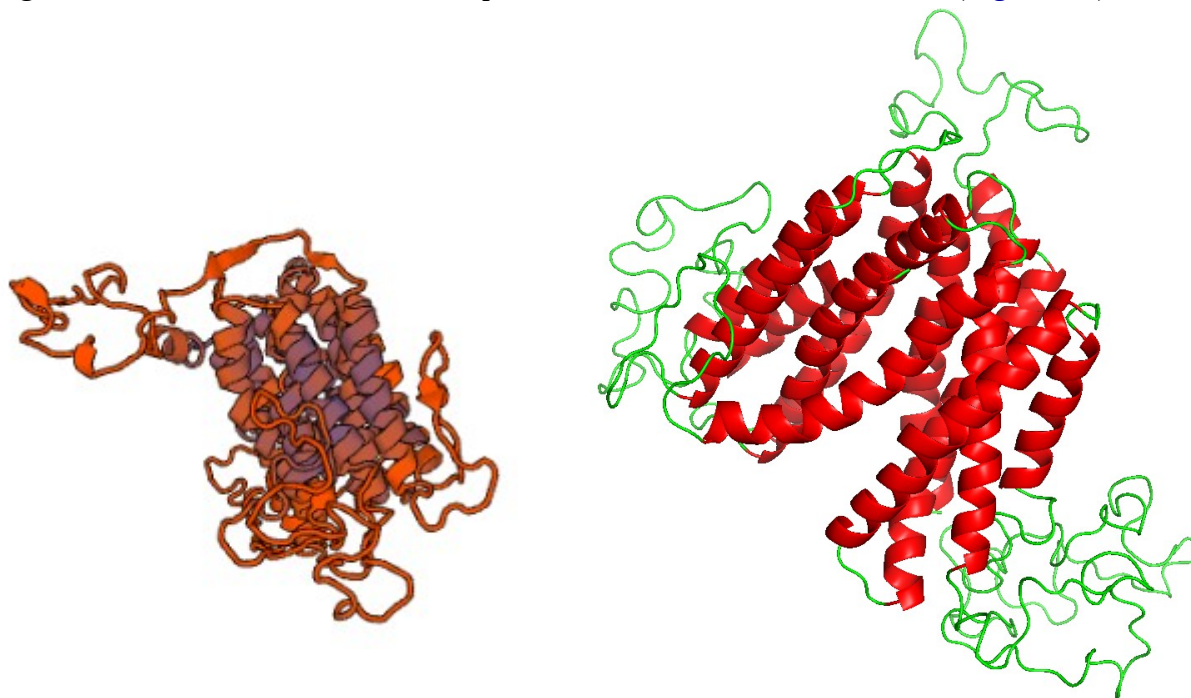


Figure 12 : Représentation du modèle de notre protéine par SWISSMODEL avec comme support 1wdoA. A droite, l'image est donnée par SWISSMODEL et à droite l'image est obtenue avec Pymol avec le même fichier PDB.

PHYRE2

Il est utilisé normalement utilisé pour une modélisation par threading car il fait à la fois la recherche de template(s) et donne également le modèle prédit de notre protéine. Mais nous l'avons quand même utilisé pour voir le genre de résultat qu'il proposait. En regardant son meilleur résultat,

c'est-à-dire avec 1pw4A, nous voyons que le modèle n'est en fait pas si parfait que cela (Figure 13). En repérant les résidus N (asparagine) sur la structure (résidus roses sur la Figure), le but est de trouver les sites de N-glycosylation de l'OATP1B1 afin de discerner les boucles 2 et 5. En faisant cela, nous avons pu remarquer que le modèle n'est pas complet car la protéine est coupée à plusieurs endroits.

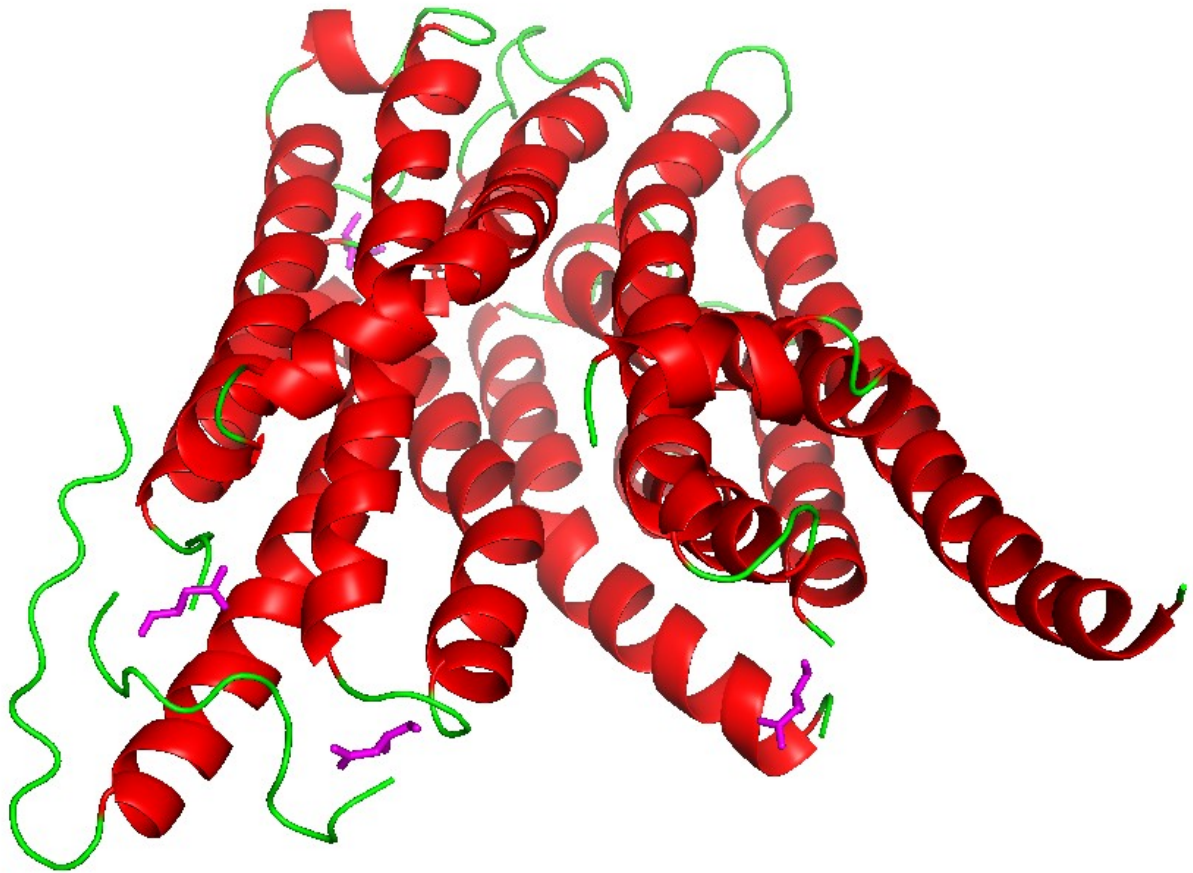


Figure 13 : Représentation sur Pymol du modèle de la protéine cible obtenue avec le template 1pw4A. En vert, ce sont les boucles, en rouge les hélices et en rose/violet les résidus N. La protéine est coupée à plusieurs endroits.

MEMOIR (MEmbrane protein Modelling pipeline)

Grâce à un alignement entre notre séquence cible et une séquence template, ainsi que l'identifiant PDB de ce template, ce serveur sort des modèles. Il est utilisé pour la prédiction de la structure 3D des protéines membranaires et utilise la méthode MEDELLER (MODELLER pour les protéines membranaires). PSIPRED ayant prédit du désordre dans notre protéine et que l'on connaît les positions de ces désordres, nous avons éliminé à la main ces résidus. Ceci nous permet d'avoir une protéine sans désordre. En effet, les molécules avec désordre sont très mal prédites généralement. Le template utilisé est 3wdoA car c'est celui qui a donné de meilleur résultat par rapport à 1pw4A. Pymol montre que certaines parties ont été omises. Cela pourrait s'expliquer par l'absence de parties dues à l'enlèvement des désordres. De plus, il n'y a pas de brins bêta. Ceci peut paraître normal parce qu'il n'y a pas non plus de brins dans le support. Mais comment SWISSMODEL a pu prédire donc des brins avec le même support ?

Test de la qualité du modèle.

Les deux premiers tests (OREMPRO et QMEMBRAN) sont utilisés pour des protéines membranaires. Le dernier (PROSA) est utilisé pour n'importe quelle protéine.

OREMPRO

Il indique notamment le nombre de segments transmembranaires trouvés et permet la visualisation du modèle dans une bicouche lipidique. Si pour le template 1pw4A il a bien prédit le nombre de segments, ce n'est pas le cas pour 3wdo (Figure 14). En effet, alors que sur Pymol nous pouvons compter 12 segments, OREMPRO nous en compte 3. Ce test n'est donc pas fiable.

Input: sm_3wdoA.pdb

Membrane assignment:

Hydrophobic thickness = 40.0 Å

Output PDB file [Download](#)

Number of TM segments = 3

Sequence positions:
A: 171-200|204-231|259-283

Tilt angles:

A - Segment: 171-200 - Tilt: 34.3°

A - Segment: 204-231 - Tilt: 11.8°

A - Segment: 259-283 - Tilt: 16.2°

Structural assessment:

MAIDEN score = 22.057 ?

Scores for all residue pairs [Download](#)

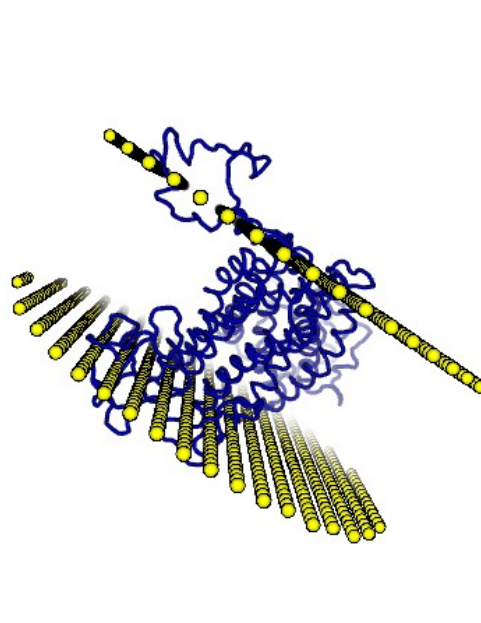


Figure 14 : Résultat donné par OREMPRO avec à gauche les informations et à droite l'illustration de la protéine dans une bicouche lipidique. Le nombre de segments proposés ne correspond pas du tout à ce qui a été observé sur Pymol.

QMEMBRANE

Ce test de SWISSMODEL est choisi car en plus de montrer le modèle prédit une bicouche lipidique, il nous donne sa séquence en colorant résidus selon un score de qualité locale. Plus la couleur est bleue, plus la qualité du modèle est bonne. Pour les modèles obtenus avec les deux templates, la qualité est mauvaise (Figure 15). Cette mauvaise qualité se trouve également au niveau du calcul de QMEAN. Plus sa valeur est petite (moins de -6), plus la qualité de la prédiction est mauvaise. Or là nous avons un QMEAN à -8.29 pour 3wdoA et -10.32 pour 1pw4A. Donc ce sont bien de mauvais modèle.

PROSA

La qualité du modèle est donnée par un zscore, et peut être visualisée par un graphe du zscore en fonction du nombre de résidus. Les modèles prédits sont de qualité moyenne.

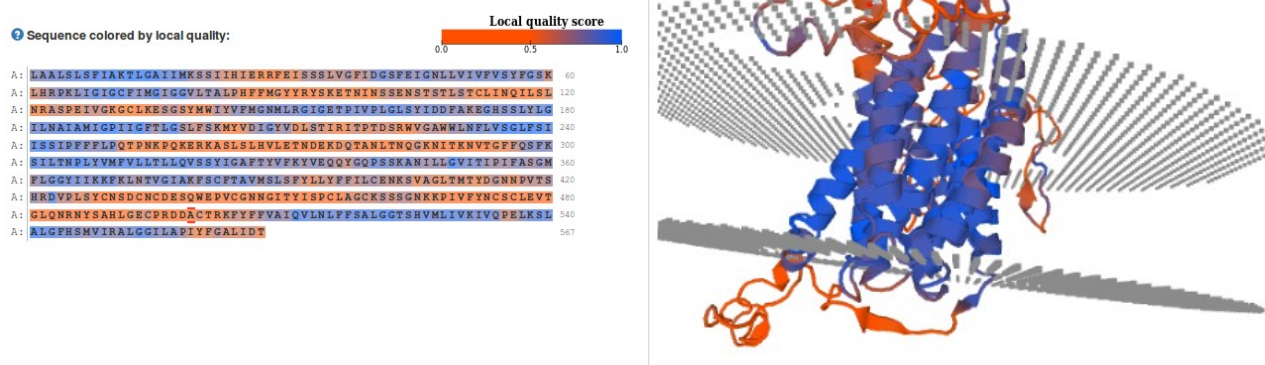


Figure 15 : Résultats de QMEMBRANE avec à gauche la qualité donnée sur la séquence en couleur et à droite la visualisation avec une bicouche lipidique

Affinement des boucles

GALAXYWEB

Cette étape a été réalisée avec le modèle prédit avec le support 3wdoA car il donnait de meilleur résultat que le second support. De plus, c'est le modèle obtenu avec SWISSMODEL. Après cette étape d'affinement, nous avons tester la qualité du nouveau modèle. Elle s'est améliorée sans pour autant donner une très bonne prédiction. Nous avons ensuite décidé de visualiser notre modèle, bien que pas très bonne (Figure 16). Il y a beaucoup plus de résidus N (en rose) au niveau des boucles, les deux résidus importants K361 et R580 sont en jaune, les termini (en cyan) ne se trouvent pas tous les deux du côté intracellulaire, comme ce qui avait été prédit et non pas comme ce qui avait été écrit dans un article [7].

Effet du désordre dans la prédiction de la structure 3D de Q9Y6L6

Avant de faire toute cette analyse, nous avons travaillé sur 4jO5A, un transporteur de phosphate. Ce template a été parmi les meilleurs avec HHBLITS et PHYRE2. Le modèle prédit a à la fois hélices et des brins. Le souci était qu'une boucle était vraiment très large, ce qui était bizarre. Nous n'avions donc pas poussé cette analyse plus loin car à ce moment nous ne connaissions pas assez notre protéine. Par contre, l'observation intéressante est que la présence des désordres dans la molécule influence énormément la modélisation de la structure (Figure 17).

Conclusion

Le choix d'un template s'avère être très difficile. Plusieurs critères rentrent en compte pour bien le choisir. Ensuite, l'étape importante est l'alignement qui permet ou non la conservation des résidus importants. Ces étapes ont pris beaucoup plus de temps que prévu. Ces modèles sont à revoir et ne sont pas de bonne qualité. Bien que le projet n'est pas terminé, il nous a permis de découvrir les différents outils de la Bioinformatique Structurale et les problèmes rencontrés pour la prédiction de la modélisation d'une protéine, qui plus est transmembranaire.

Bibliographie

- 1 [Analysis and update of the human solute carrier (SLC) gene superfamily (2008)]
- 2 [comparison of human solute carriers (2009)]
- 3 [The SLC2 (GLUT) family of membrane transporters]
- 4 [The multispecific organic anion transporter (OAT) family]
- 5 [Organic anion transporting polypeptides of the OATP/SLC21 family : phylogenetic classification as OATP/SLCO superfamily, new nomenclature and molecular / functional properties (2004)]
- 6 [Organic anion-transporting polypeptides (OATs/SLCOs)(2013)]
- 7 [OATPs, OATs and OCTs : the organic anion and cation transporters of the SLCO and SLC22A gene superfamilies (2005)]
- 8 [Several Conserved Positively Charged Amino Acids in OATP1B1 are Involved in Binding or Translocation of Different Substrates (2010)]

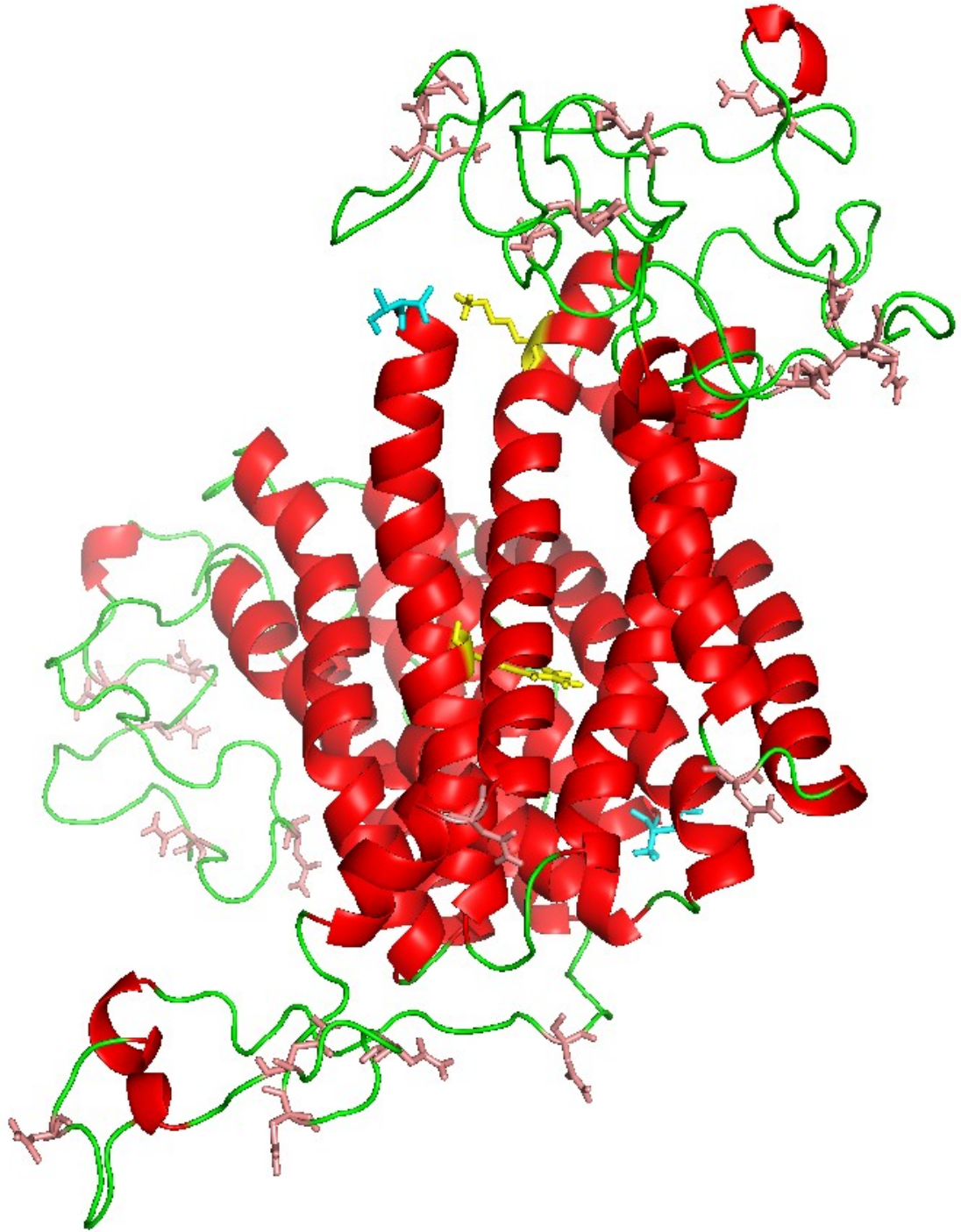


Figure 16 : Représentation sur Pymol du modèle avec le support 3wdoA, suite à l'affinage des boucles. En rouge sont représentées les hélices, en vert les boucles, en jaune les résidus importants pour la fixation du substrat, en cyan les termini et en rose les résidus N.

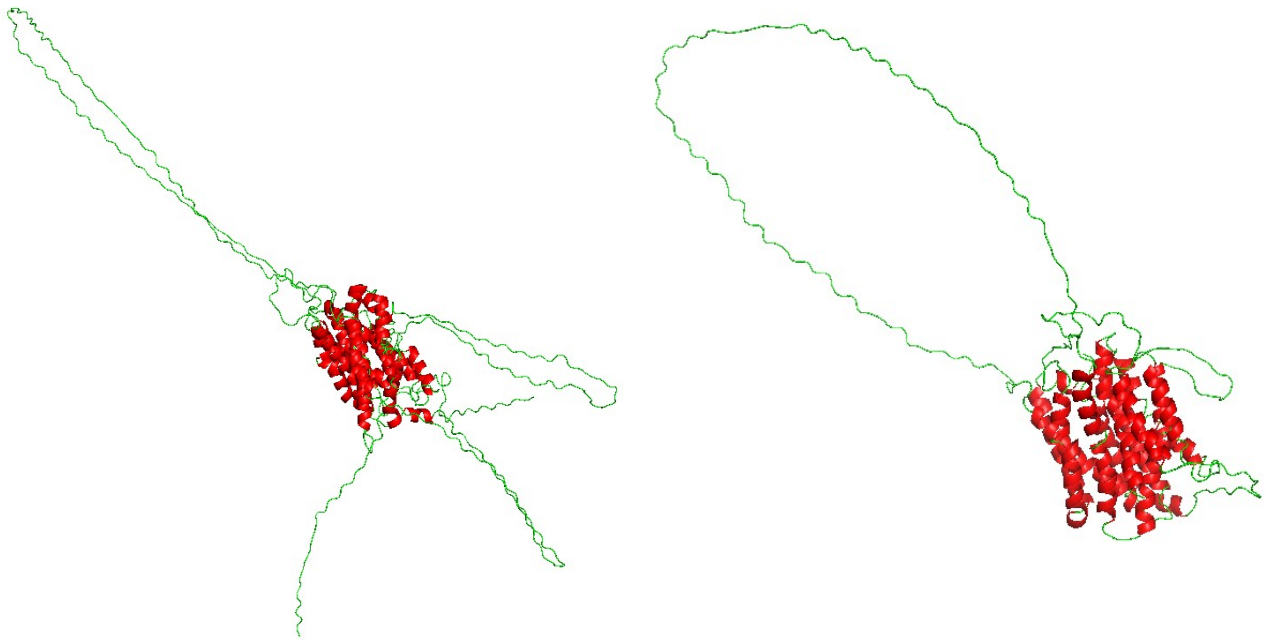


Figure : A gauche, le modèle est obtenu avec la séquence protéique complète de Q9Y6L6 et à droite, les désordres ont été enlevés. Nous remarquons notamment que le nombre de boucles larges diminue.