

## **Projet Bioinformatique structurale**

### **Protéine Q8IVW8**

#### **1. Introduction**

L'objectif de ce projet est de faire une modélisation de structure 3D pour une protéine transmembranaire qui n'est pas encore résolue. La protéine étudiée la Q8IVW8 (code Uniprot).

##### **1.1. fiche uniprot**

La protéine est une Homologue de la protéine spinster chez l'humain. Elle fait partie de la famille protéique des 'major facilitator superfamily' (MFS).

Sa fonction principale est : Transporteur de sphingolipide. Ces sphingolipides sont requis pour la migration des précurseurs du myocarde (muscle cardiaque). Elle Transporte plus spécifiquement la sphingosine 1-phosphate (S1P), un médiateur lipidique sécrété qui joue un rôle essentiel dans le développement et la fonction cardiovasculaire, immunologique et neurale.

La taille de sa séquence est de 549 pb.

On remarque d'après les information expérimentale il y a un conflit dans la séquence en position 74 par le changement d'acide aminé T en A. De plus la position 200 est très importante car lors de mutagènes, la mutation de R vers S provoque une perte de fonction.

##### **1.2. Informations des articles cités sur Uniprot**

Il a été montré qu'en plus des 12 domaines transmembranaires, il existe une similitude structurelle prédite entre la protéine spinster du poisson-zèbre(zSpns2) et le transporteur de glycérol-3-phosphate bactérien d'après Atsuo Kawahara and al (2009)

De plus, il a été proposé que les transporteurs ABC, y compris ABCC1 et ABCA1 sont nécessaires pour le transport S1P. Ces transporteur transportent le même sphingolipide que la protéine Q8IVW8, ils donc surement une structure similaire.

#### **2. Caractérisation de la protéine Q8IVW8**

##### **2.1. Séquence consensus**

###### **2.1.1. Séquence logo**

La recherche de séquence s'est fait via un blastp sur la base de donnée de séquence 'UniProtKB' sur le site Uniprot.org avec 1000 hit. L'alignement multiple a été réalisé avec Clustal Omega sur Uniprot.. Les séquences choisis sont d'autres espèces et avec des

séquences homologues suffisamment différente. Le pourcentage de similarité d'homologie a donc été pris autour de 55-40 %, pour éviter une forte redondance.

La séquence logo a été obtenu via WebLogo 3:



### 2.1.2. Séquence motif de famille

La fiche Uniprot de Q8IVW8, montre que dans la section Family&Domain, cette protéine fait partie d'une Famille de motif de protéines (Pfam), plus précisément du clan des MFS Major Facilitator Superfamily. Ce clan contient 24 membres:

[Acatn](#)  
[MFS\\_1](#)  
[Nodulin-like](#)  
[TLC](#)

[ATG22](#)  
[MFS\\_1 like](#)  
[Nuc H symport](#)  
[TRI12](#)

[BT1](#)  
[MFS\\_2](#)  
[Nucleoside tran](#)  
[UNC-93](#)

[Folate carrier](#)  
[MFS\\_3](#)  
[OATP](#)

[FPN1](#)  
[MFS\\_4](#)  
[PTR2](#)

[FTR1](#)  
[MFS\\_5](#)  
[PUCC](#)

[LacY symp](#)  
[MFS Mycoplasma](#)  
[Sugar tr](#)

## 3. Choix de la méthode de modélisation

### 3.1. Méthode par homologie:

#### 3.1.1. Méthode pas à pas

##### 3.1.1.1. Recherche de séquence

A partir de la séquence et l'exécution d'un *advanced* blast sur la base de donnée '3D structure PDB', on recherche s'il existe un *template*, un support qui possède une similarité de structure à Q8IVW8. Les résultats montrent des protéines entre 20 et 39 % de similarité (Fig. 1), ce qui suggère d'utiliser une méthode par homologie.

Entry	Alignment overview	Info	Status
Query: sp Q8IVW8 SPNS2_HUMAN B20171122A7434721E10EE6586998A056CCD0537E095398G			
P39352	YJHB_ECOLI - Putative metabolite transport prote... - Escherichia coli... - View alignment	E-value: 4.6e-8 Score: 134 Ident.: 22.8%	
Q3K554	EBNA1_EBVG - Epstein-Barr nuclear antigen 1 - Epstein-Barr vir... - View alignment	E-value: 2.6e-4 Score: 104 Ident.: 39.4%	
P03211	EBNA1_EBVB9 - Epstein-Barr nuclear antigen 1 - Epstein-Barr vir... - View alignment	E-value: 3.4e-4 Score: 103 Ident.: 33.7%	
P07663	PER_DROME - Period circadian protein - Drosophila melan... - View alignment	E-value: 3.9e-4 Score: 103 Ident.: 33.0%	
P07663-2	PER_DROME - Isoform PER-A-short of Period circa... - Drosophila melan... - View alignment	E-value: 3.9e-4 Score: 103 Ident.: 33.0%	
P07663-4	PER_DROME - Isoform PER-C of Period circadian p... - Drosophila melan... - View alignment	E-value: 3.9e-4 Score: 103 Ident.: 33.0%	
A0A0H2VG78	A0A0H2VG78_STAES - Bicyclomycin resistance protein Tca... - Staphylococcus e... - View alignment	E-value: 3e-4 Score: 103 Ident.: 22.3%	

Figure : Aperçue de la sortie du Blast de structure 3D sur Uniprot de Q8IVW8.

### 3.1.1.2. Choix du support

La protéine P39352 (Putative metabolite transport protein, E.Coli), ayant le meilleur recouvrement de la structure initiale ainsi qu'un bon score et un bon pourcentage de similarité d'identité, est sélectionnée pour servir de *template*. Cependant sa fiche Uniprot montre que sa structure n'est résolue que pour les positions 101-106, ce qui est trop faible pour servir de support.

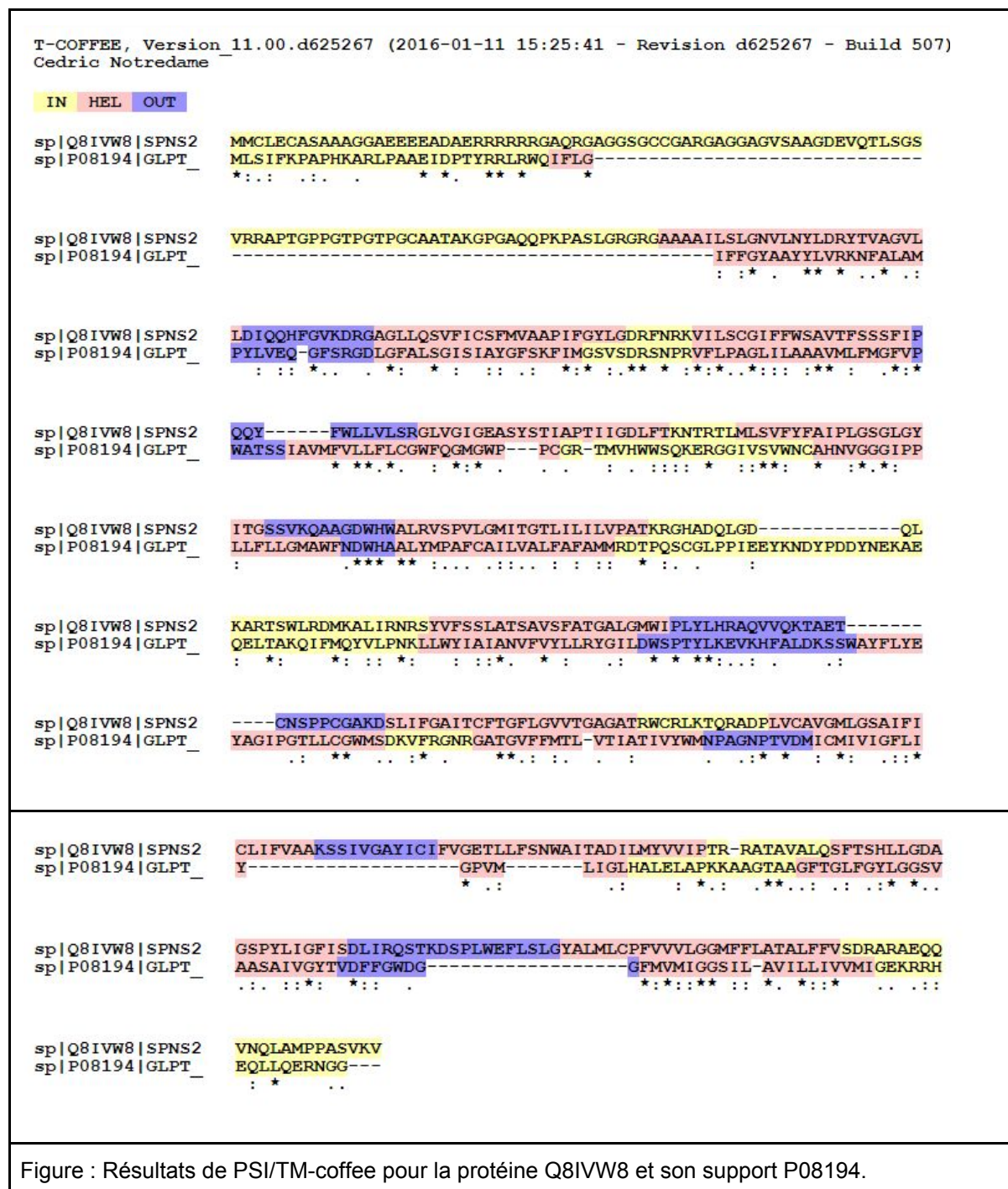
D'autres protéines ont un recouvrement moins étendu, mais possèdent une bonne e-value, un bon score et un bon pourcentage d'identité (Tab. 1). Les trois possèdent une structure 3D résolue, mais on constate un meilleur recouvrement avec la cible pour la protéine P08194. De plus, il s'agit d'un Glycerol-3-phosphate transporter qui est prédit comme ayant une similarité structurelle avec une Protein spinster homolog 2 (SPNS2). Le support choisit est la protéine P08194, avec pour PDB 1PW4.

Uniprot Entry	PDB	Recouvrement	E-value	Score	% identité
<a href="#">P0AGF4</a>	<a href="#">4JA3</a> , <a href="#">4JA4</a>	107-261	4e-1	77	23,4
<a href="#">P08194</a>	<a href="#">1PW4</a>	129-410	4,7e0	68	20,1
<a href="#">A0A0H2VG78</a>	<a href="#">4LDS</a>	109-272	3e-4	103	22,3

Tableau 1: Trois supports potentiels obtenus après un blast sur une base de donnée de structure 3D.

### 3.1.1.3. Alignement

L'alignement choisit est celui de T-coffee, car il possèdent un algorithme spécifique pour les protéines transmembranaire : PSI/TM-coffee. La particularité de cet algorithme est d'indiquer les parties qui sont à l'intérieur de la cellule en jaune, les parties transmembranaires en rouge et les parties de la protéine à l'extérieur de la cellule en violet (Fig. )



### 3.1.1.4. Modelisation : MEDELLER

Le programme crash et impossible de modeliser sur le site :

<http://opig.stats.ox.ac.uk/webapps/medeller/home.pl?app=MEDELLER>

### 3.1.2. Méthode automatique

#### 3.1.2.1. SwissModel

Après soumission de la séquence FASTA de Q8IVW8, 50 *template* ont été proposé, parmi eux on retrouve la 1PW4(P08194). Des protéines de transport comme les 'facilitated glucose transporter member 5', 'Di-tripeptide ABC transporter (Permease)', 'MFS Transporter' et des 'Proton:oligopeptide symporter POT family'.

Au total 4 modèles avec different *template* ont été effectué:

- ABC transporteur : 4IKXA
- MFS transporteur : 3WDOA
- Glycerol-3-phosphate transporte 1PW4A
- Transporteur facilité de glucose : 4YB9A

#### 3.1.2.2. HHpred - Modeller

Pour comparer les méthodes "automatiques" par homologies, la recherche de *template* et ma modélisation avec MODELLER a été faite sur le site <https://toolkit.tuebingen.mpg.de>.

On retrouve en tête des supports proposés 1PW4 (Glycerol -3- phosphate transporter), puis des ABC transporter, des perméases et des transporteurs multi-drug. Deux modèles ont été générés avec MODELLER (Tab. ).

PDB	Type protéine	
1PW4	Glycerol-3-phosphate transporter	
4IKW	ABC transporteur	
Tableau : <i>Template</i> sélectionnés via HHpred-Modeller.		

### 3.2. Méthode par threading

I-TASSER result: Your job id number is S364725

<http://zhanglab.ccmb.med.umich.edu/I-TASSER/output/S364725/>.

### 3.3. Analyse qualité des modèles avant affinement

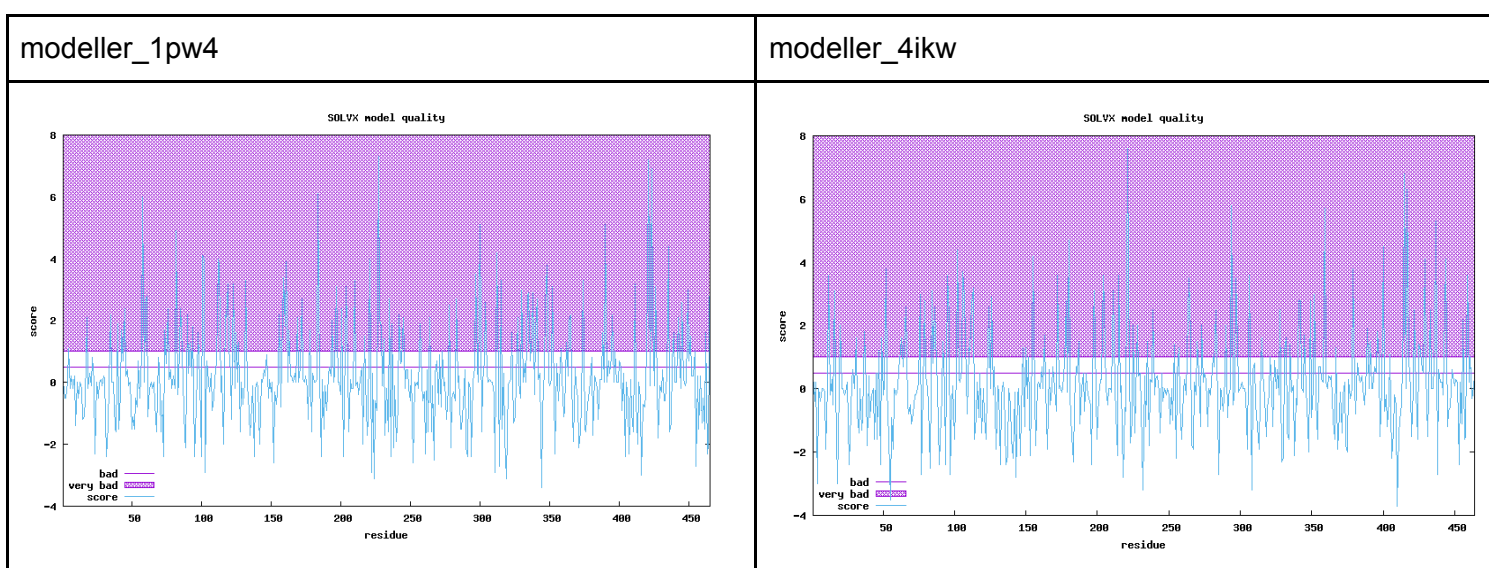
### 3.3.1. Modele SwissModel

L'utilisation de Q-mean comme critère d'évaluation des modèles n'est pas adapté pour des protéines membranaire, le seuil à ne pas dépasser est de -4, ici tous les modèles sont en dessous autour de -9 et sont donc classés comme étant mauvais. On remarque que parmi les modèles de SwissModel, les protéines de transporteur de glucose (4yb9) et les ABC transporteur (4ikx) sont tout de même meilleur avec un Q-mean de -6,6 et -6,8 respectivement. Les modèles de MFS transporteur et de transporteur de Glycerol-3-phosphate (G3P) sont très mauvais malgré que le transporteur de G3P était le meilleur support proposé d'après SwissModel et HHpred .

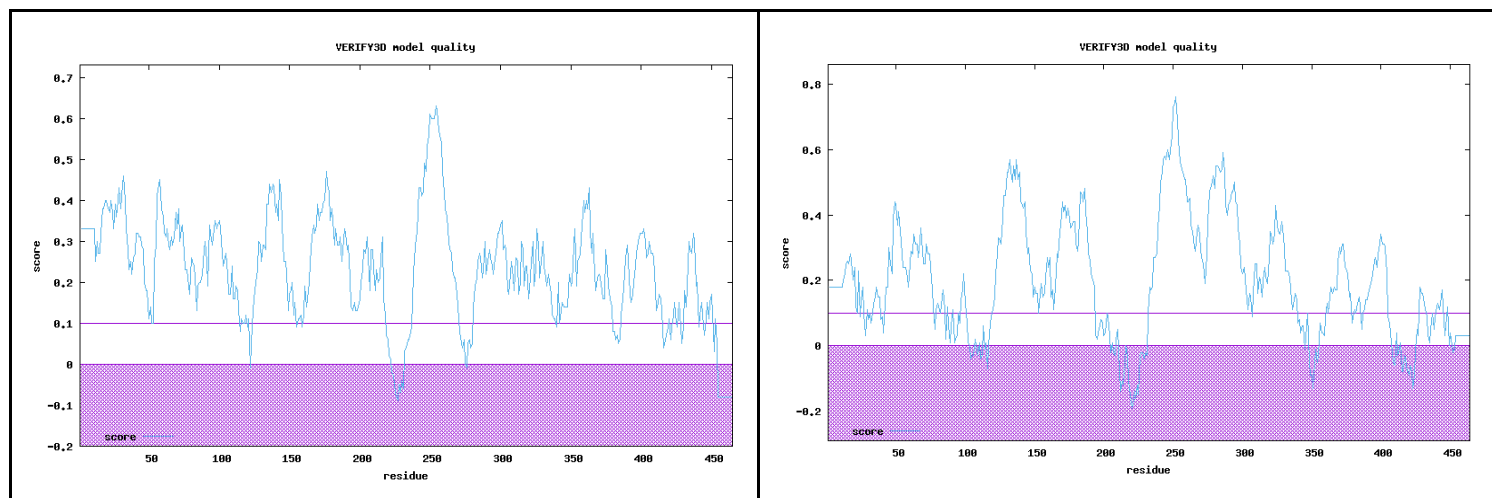
PDB	Type protéine	Q-mean	GMQE	Result
4YB9A	Transporteur facilité de glucose	-6,62	0,39	Mauvais
4IKXA	ABC transporteur	-6,79	0,39	Mauvais
3WDOA	MFS transporteur	-8,29	0,40	Très mauvais
1PW4A	Glycerol-3-phosphate transporter	-8,98	0,40	Très mauvais

### 3.3.2. Modele HHpred-Modeller

On constate que la qualité de solvation des deux modèles est très mauvaise pour beaucoup de résidus tout au long de la chaîne protéique. De plus, la qualité vérifiée du modèle de 1PW4 est mauvaise pour les residues proches des positions 120, 225 et 275. Pour la protéine 4IKW, on constate des zones de mauvaise qualité autour des positions 100-120, 200-225, 350-360 et 410-430.







### 3.4. Affinement des modèles

Les modèles ont été affinés via GalaxyRefine (<http://galaxy.seoklab.org/cgi-bin/submit.cgi?type=REFINE>) et seul le meilleur affinement proposé est sélectionné. Pour déterminer si l'affinement est nécessaire et apporte une plus value aux modèles générés, on prend en compte les critères de Clash et de MolProbity (validation de structure qui renseigne sur la qualité du modèle aux niveaux global et local pour les protéines et les acides nucléiques). On constate que les modèles MODELLER sont moins bons par rapport à ceux de SwissModel (Tab. ).

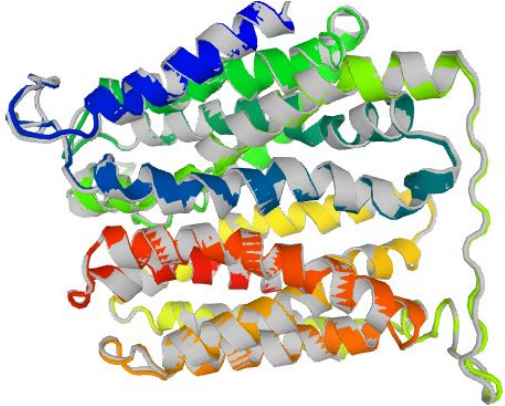
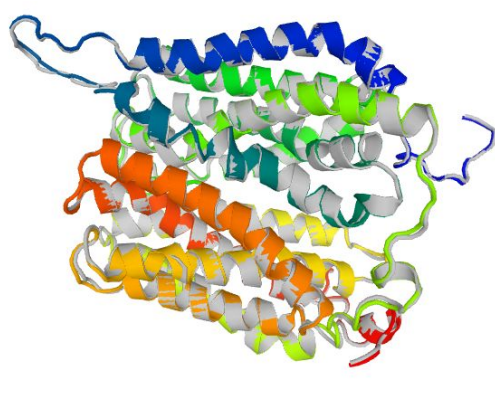
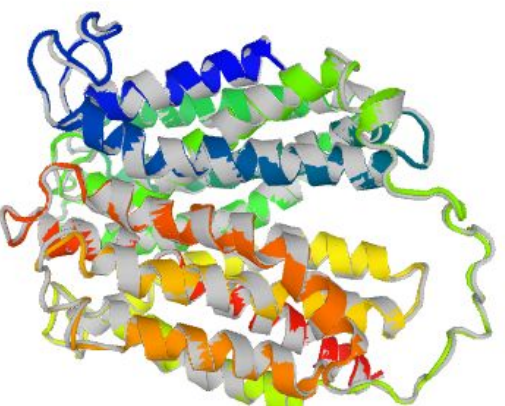
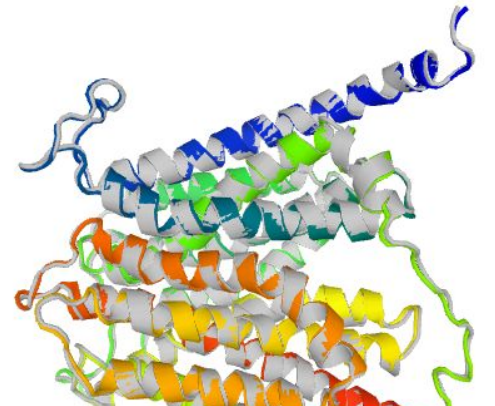
On remarque que les modèles affinés sont en générale les meilleurs (swiq\_1pw4, swiss\_3wdo, modeller\_4ikx, modeller\_1pw4) avec exception du modèle swiss\_4ikx qui possède le meilleur score de clash et de MolProbity avant affinement.

Il reste un cas plus difficile à déterminer pour swiss\_4yb9, le modèle initial possède un meilleur score de clash et un mauvais score de MolProbity, et à l'inverse le modèle affiné a un meilleur score de MolProbity et un mauvais score de clash.

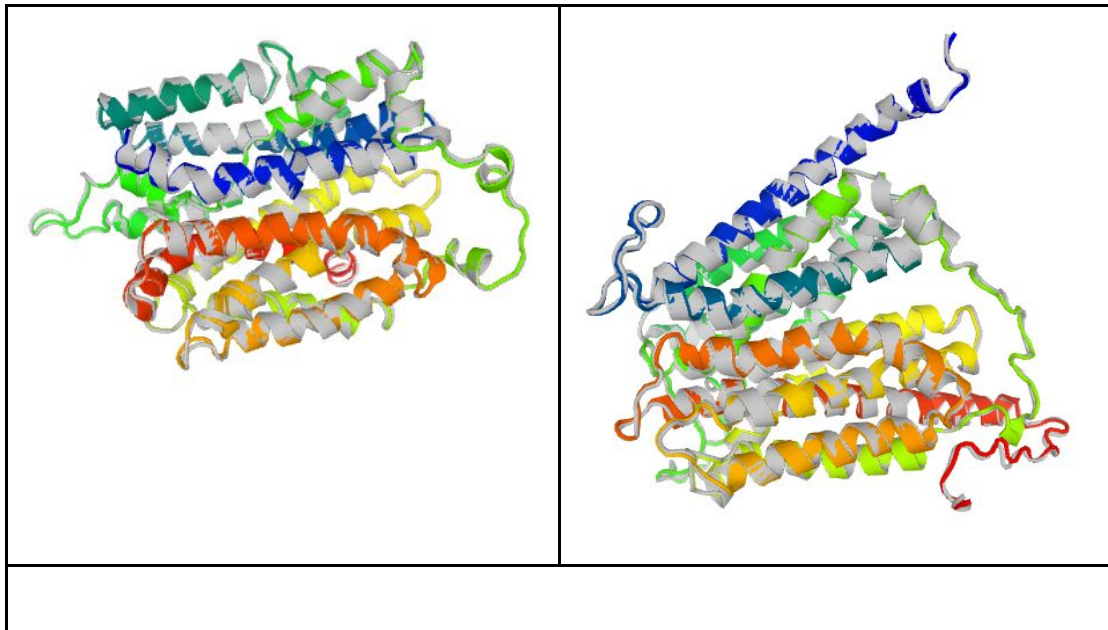
Nom model	Modèle	GDT-HA	RMSD	MolProbity	Clash score	Choix
swiss_1PW4	Initial	1.0000	0.000	2.584	19.5	Non
	MODEL 2	0.9496	0.414	<b>2.153</b>	<b>20.5</b>	Ok
swiss_4ikx	Initial	1.0000	0.000	<b>2.167</b>	<b>12.0</b>	Ok
	MODEL 1	0.9665	0.379	2.089	19.3	Non
swiss_4yb9	Initial	1.0000	0.000	2.310	<b>14.7</b>	
	MODEL 1	0.9557	0.413	<b>1.834</b>	20.8	
swiss_3wdo	Initial	1.0000	0.000	3.511	138.0	Non
	MODEL 4	0.9398	0.446	<b>2.252</b>	<b>18.2</b>	Ok

modeller_4ikw	Initial	1.0000	0.000	3.511	138.0	Non
	MODEL 1	0.9504	0.410	<b>2.394</b>	<b>29.1</b>	Ok
modeller_1pw4	Initial	1.0000	0.000	3.511	138.0	Non
	MODEL 1	0.9242	0.480	<b>2.248</b>	<b>22.4</b>	Ok

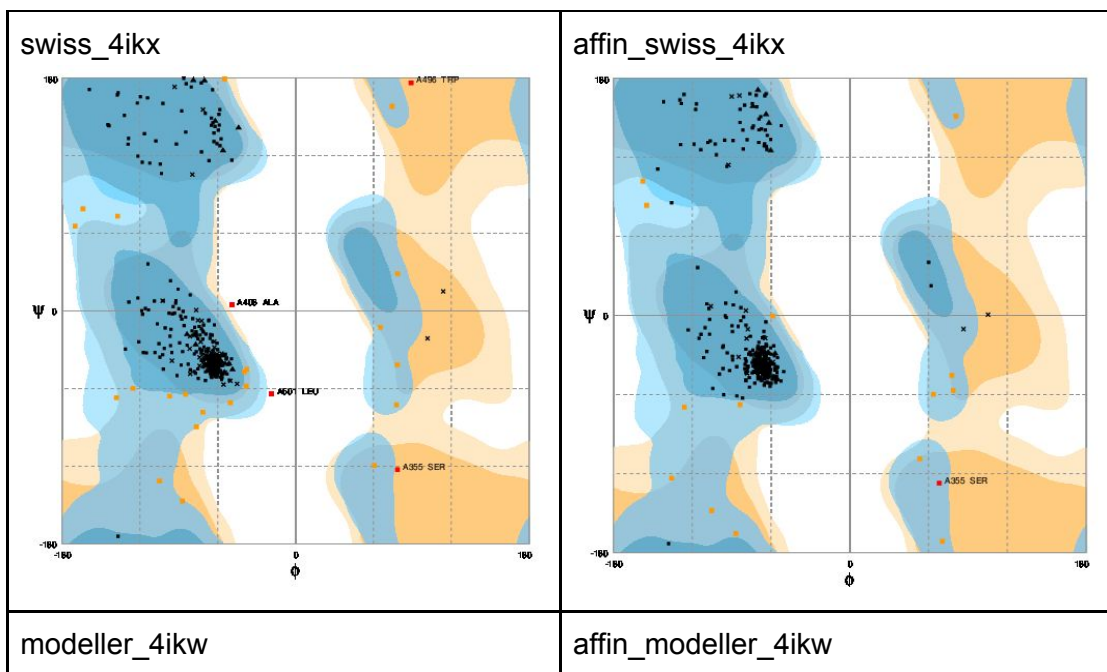
Tableau : Score des modèles SwissModel et Modeller avant et après affinement.

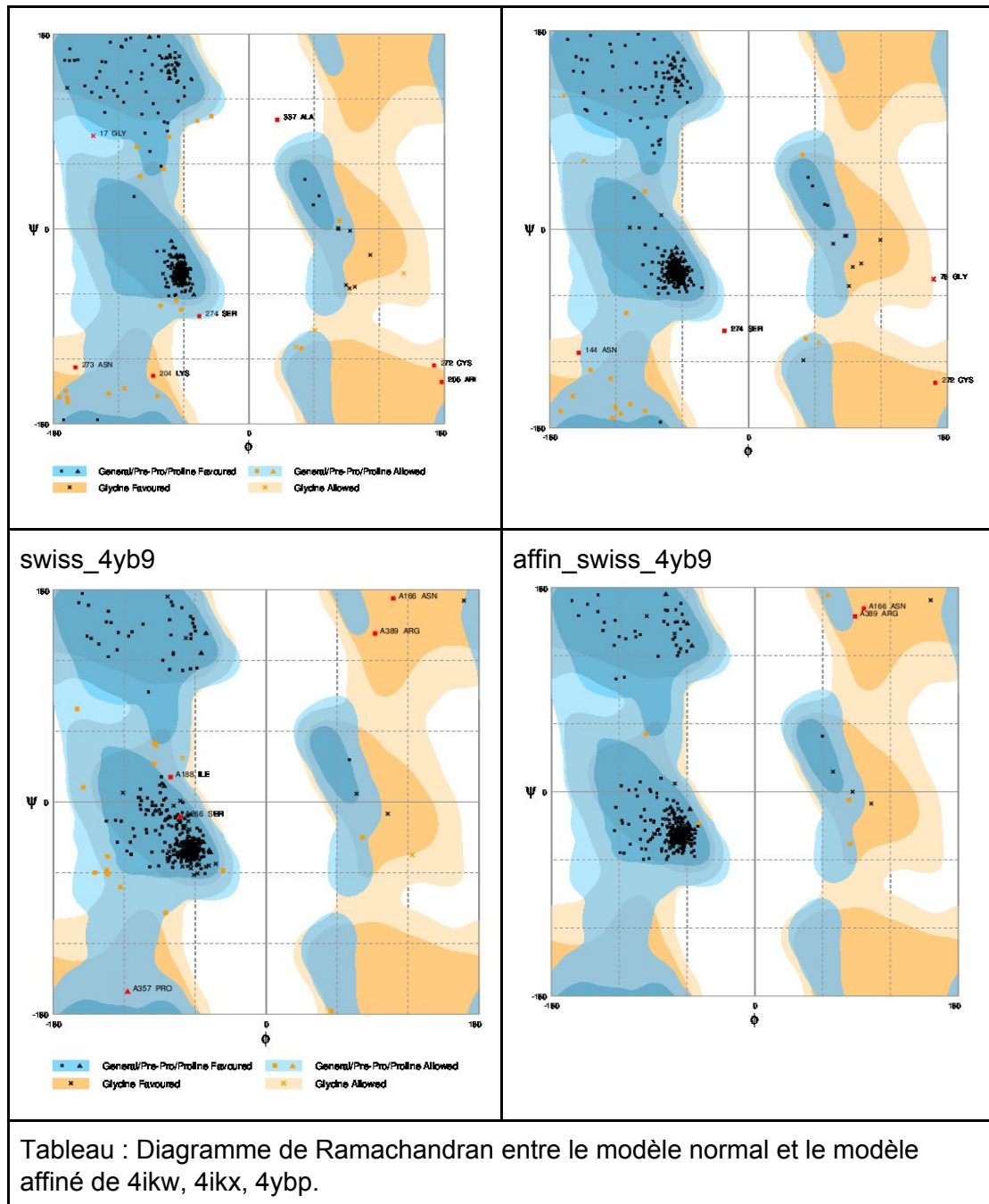
<p>swiss_4kx</p> 	<p>modeller_4ikw</p> 
<p>swiss_1pw4</p> 	<p>modeller_1pw4</p> 
<p>swiss_4yb9</p>	<p>swiss_3wdo</p>





3.5. Qualité des modèles : Comparaison des modèles simple aux modèles affinés.  
La qualité des modèles sont mesuré grâce au diagramme de Ramachandran sur le site RAMPAGE (<http://mordred.bioc.cam.ac.uk/~rapper/rampage.php>).





#### 4. Evaluation des parties transmembranaire: Orempro

#### 5. Bibliographie

"The sphingolipid transporter spns2 functions in migration of zebrafish myocardial precursors."

[Kawahara A.](#), [Nishi T.](#), [Hisano Y.](#), [Fukui H.](#), [Yamaguchi A.](#), [Mochizuki N.](#)  
Science 323:524-527(2009)

