STRUCTURAL BIOINFORMATICS REPORT

Introduction

In this study, we are interested in a family of proteins belonging to the Major Facilitator Superfamily (MFS)[1]. This Family is the largest family of secondary carriers.

More specifically, we are interested in the Mammalian Sugar Transporters.This study case we concentrate on the protein named Solute carrier family 2, facilitated glucose transporter member 4 (GLUT4).This protein (UniProt id : P14672) is coded by a gene named SLC2A4 from Homo sapiens (Human)[2].

This gene is part of the GLUT family protein which is composed of 14 isoforms [3].
The GLUT protein family is classified into three different families that can be distinguished based on their protein sequence homologies (Fig 1).

Class	1	2	3
GLUT	1,2,3,4,14	5,7,9,11	6,8,10,12,13

FIG 1. Repartition of the different GLUT protein into classes. a class groups proteins with strong homology

The protein GLUT 4 is part of the first family and it's an Insulin-regulated facilitative glucose transporter.

This protein is intensively studied among the glucose transporters because it has an important physiological role. It regulates the glucose uptake of skeletal and cardiac muscle, brown and white adipose tissue and is regulated by the insulin level .
Moreover, impaired GLUT4 translocation is bind to insulin resistance[4]. Indeed, default of the insulin-mediated translocation of GLUT4 to the plasma membrane is known to be a peripheral insulin resistance.

Because most of the blood glucose is absorbed by skeletal muscle in the presence of high insulin and the GLUT4 transport activity is limiting for this process, GLUT4 plays a vital role in the regulation of body glucose homeostasis.

The 3D structure of GLUT4 has yet to be resolved but some homological structure has already been resolved like GLUT1 (Uniprot id:P11166) and GLUT3 (Uniprot id : P11169).

Plus GLUT1, GLUT3, GLUT4 have a really similar sequence. So we proposed to use an homology approach to try to model the GLUT4 structure and if possible find two differents conformations of the receptor. After the modeling, we are going to use normal mode analysis to find the transition between the open and closed conformation. Because a receptor have different structure according to it conformation[5].

Materials and methods

METHODS

The whole process to obtain an accurate model need a lot of online or desktop software (Fig 2). Mostly it is more convenient to use online tools, it allows to speed up the process of modeling except for the main task which is the modeling itself that requires a lot of computing power.

steps	Software or server used
Structure and transmembrane prediction	T-Coffee[6], PsiPred[7], Blastp[8]
Modeling	Swiss-Model[9], Memoir[10], Medeller[11]
refining	Galaxyweb refine[12], sidepro2[13]
Quality check	ProQ[14], ProQM, Verify 3D, Orempro[15], Qmean[16]
After modeling	Bio3D(NMA)[17], Swiss-dock (protein ligand docking)[18]

Fig 2. Different online or desktop software use during this study

This study has required the use of Fasta files to select sequences that have a similarity between 40% and 60%. After this selection, we use one PDB for each Fasta This PDB is going to be the template used in order to model the secondary and tertiary structure of our reference Fasta.At the same time, we predict the secondary structure and the putative position of the transmembrane helix.

Each selected PDB is used as a template that is going to provide the secondary and the tertiary structure. Then the modeling can begin, using mostly Memoir. This software is going to provide a full modeling pipeline, using Medeller for the coordinate generation and Completionist and Galaxyweb refine for loop modeling. After each step, a quality check is performed with tools detailed in figure 2. The most Useful has been Orempro and Verify

3D which seems to be more compatible with our protein which is a transmembrane protein.

Then we select different conformation like open conformation and closed conformation. After theses steps, we use the Bio3D tools to analyses the normal modes and Swiss-Dock to find a potential binding site for potentials ligands like D-glucose.

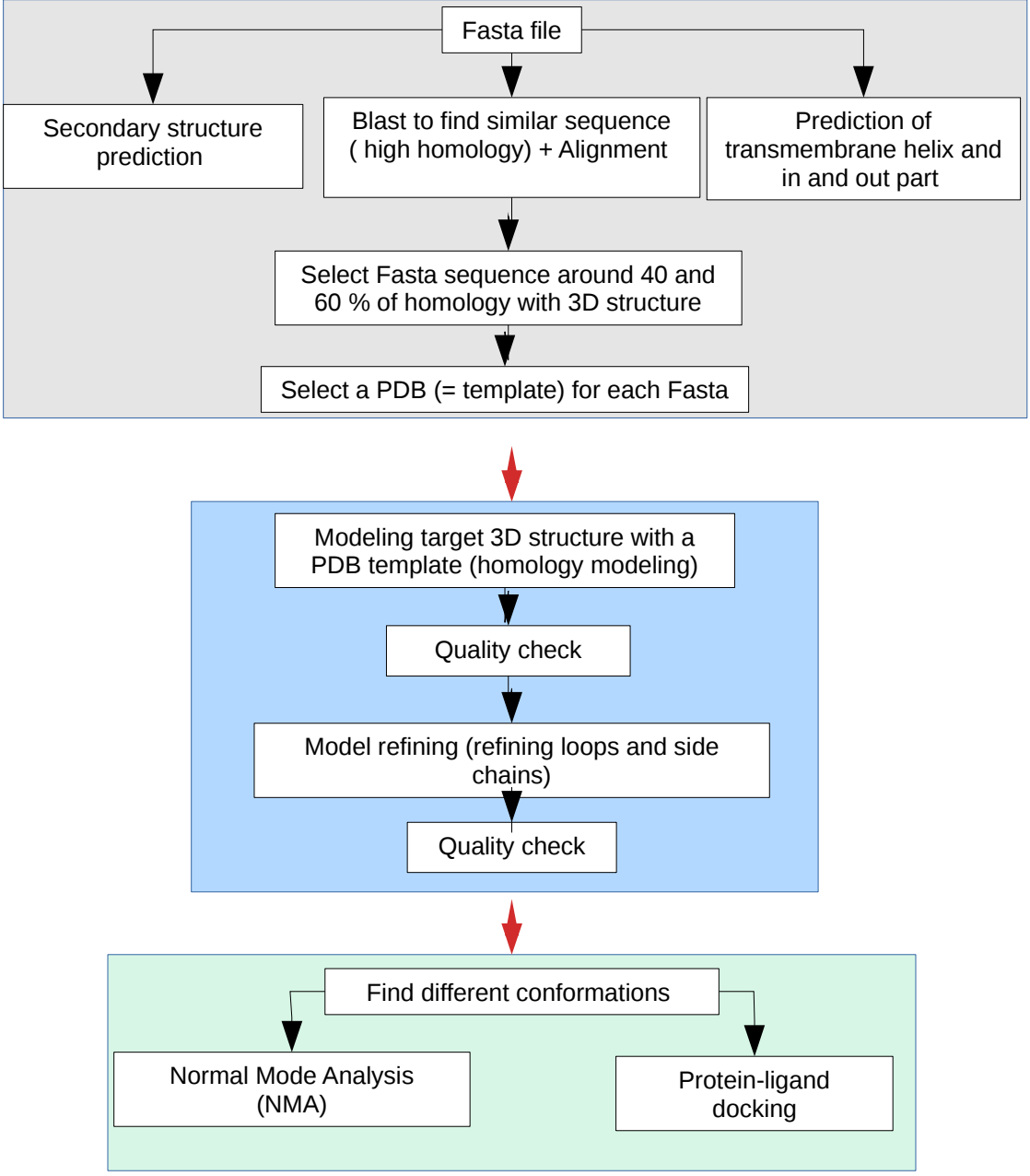

Figure 3. Workflow of steps for the protein modeling.

The steps before the modeling are in the grey rectangle the modeling is represented by the blue rectangle and the analysis step are on green

Results

Details of the fasta sequence to model

The details about the Fasta sequence are in Fig 5. But it important to insist on the fact that it is from the GLUT family

Features	Data
Uniprot id	P14672
name	Solute carrier family 2,facilitated glucose transporter member 4 (GLUT4)
Gene id	SLC2A4
function	Insulin-regulated facilitative glucose transporter.
Localization	Cellular membrane
length	519

Figure 4. Details of the sequence selected to be 3D models.

Secondary structure prediction

The Psipred analysis predcit 21 helix and some strand. Instead the Uniprot web site predict 12 transmembrane helix.T-coffe predict 12 helix and it's position in the membrane(Annexe 1).

Finally, It is important to find the domain related to the molecular function of P14672.Especially the domain related to the sugar and glucose transporter. On Uniprot the position 333 (N) and 404 (W) are highlighted as binding site for Monosaccharide. The figure 6 precise these position and the other great domain like th MFS positions on P14672.

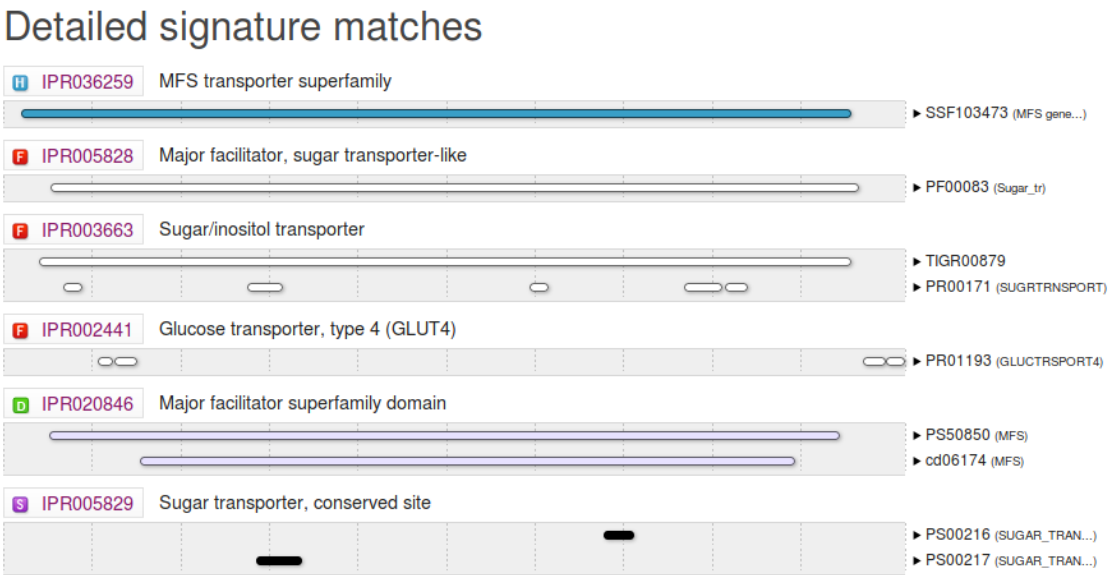

Figure 5. Principal domain found for the GLUT4

Selecting the Fasta sequence and alignment

According to the homology method, I choose 4 Fasta that have a high homology with P14672.They also have a similar sequence, a similar origin, and function. The first three sequence share the same group and the next two are from a different group.It is going to be interested to see if the modeling is going to be more difficult be the last two.

The figure 6 details the choose sequences. In annexe 2, the align provided by the T-Coffee server is quite good (annexe 2).

It show that there is a gap at the beginning between position 3 to 13.But after that one the sequence fits well. All the selected sequence provide a very large cover as seen in the figure 6.

Pdb number	Gene	function	% of Identity	3D structure	organism
P14672	SLC2A4	Insulin-regulated facilitative glucose transporter.	100,00 %	0	Homo sapiens
P11166	SLC2A1	Facilitative glucose transporter	65,9	4	Homo sapiens
P11169	SLC2A3	Facilitative glucose transporter	60,9	4	*Homo sapiens*
P43427	SLC2A5	fructose transporter	43.4	1	*Rattus norvegicus (Rat)*
P58353	SLC2A5	fructose transporter	41.8	1	*Bos taurus (Bovine)*

Figure 6. Fasta sequence with 3D sequence choose

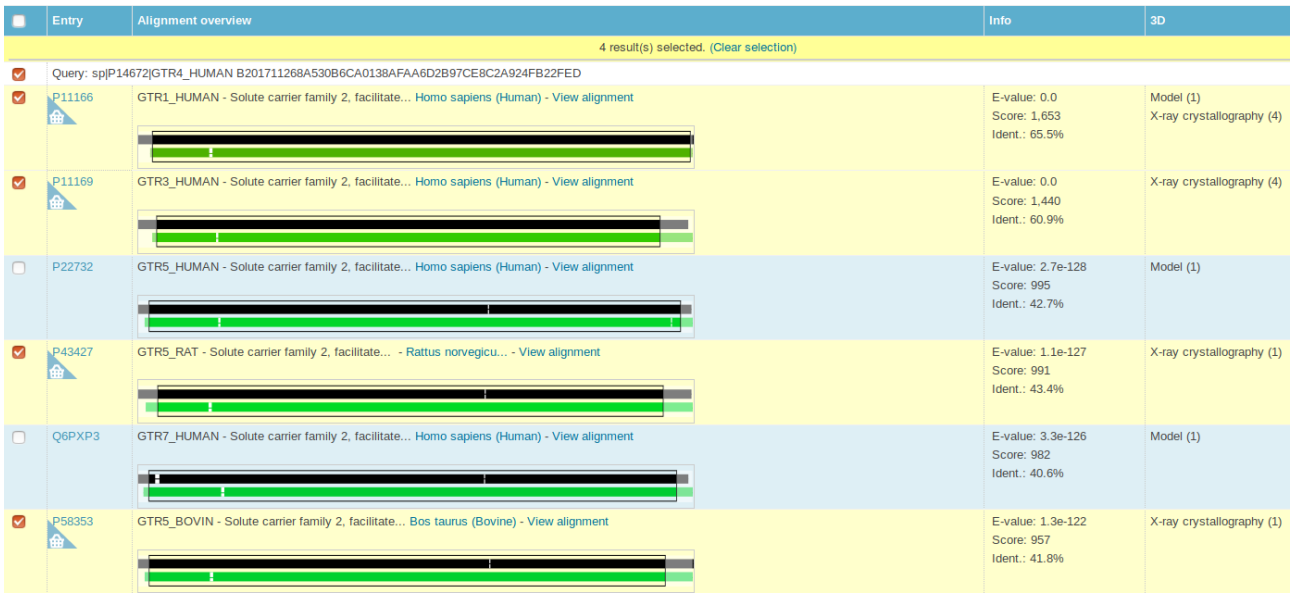

Figure 7. Covering provided by the Fasta sequence on the original sequence

Selecting the PDB template

For each Fasta sequence, a PDB has been chosen. I took four PDB, two PDB have an inward-open conformation and the other two have an outward-closed conformation. Each template was carefully selected. I align all the PBD and compute the RMSD in order to see if there is a big difference (annex 3). I only align the PDB that have the same conformation.

For P11166, all the conformation are inward open, with pymol the 3D structure are all very close (Fig 8) and the average RMSD is around 0.3 A (annex 3) which show that all the 3D structure are really similar.

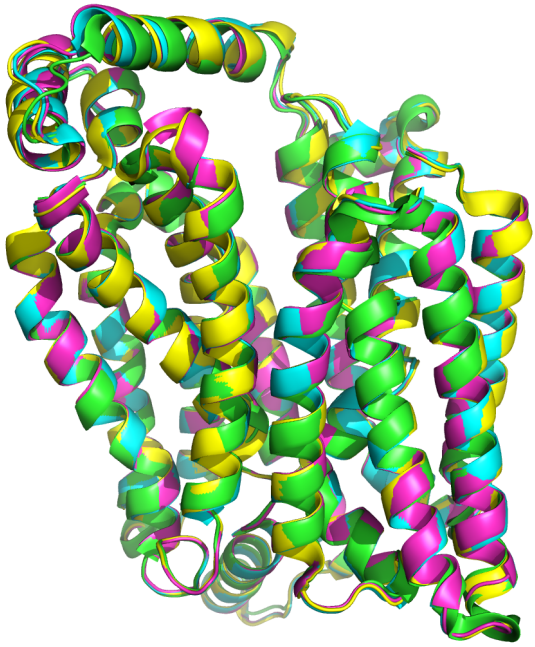

Figure 8. All the PDB of P11166 align with pymol to show their similarity

The same work has been made with the three PDB of P11169 and shows similar results. All the 3D structure are well aligned and the average RMSD is around 0.3 A.

For both P58353 and P43427, only one structure is available.

So we can conclude that only a PDB for each Fasta is necessary. Figure 9 regroups chosen PDB a provide more information about each structure

PDB entry	Method	Reso-lution (Å)	Chain	Positions	conformation
4PYP	X-ray	3.17	A	1-492	inward-open
4ZW9	X-ray	1.50	A	1-496	outward-occluded
4YB9	X-ray	3.20	D	1-473	open inward-facing
4YBQ	X-ray	3.27	A/B	1-502	outward-open

Figure 9. The four PDB chosen to be the template in order to model GLUT4's 3D structure

Modeling the PDB and quality check

As precise in the materials and methods, a huge part of all models have been realized with Memoir, a full membrane protein modeling pipeline. To compare the result I have built another model with Swiss-Model but the best model was built with the default option of memoir. Also memoir provide a tool that refine the model quality but to be sure of the best quality

Model	Template from	helix number	Predicted LGscore	Predicted MaxSub	PRO QM	Qmean score	Verify 3D(%)	Valid Model
1	4pyp	12	6.022	0.405	0.775	-6.67	85.46	OK
2	4zw9	10	6.128	0.377	0.785	-3.26 -2.22	90.21	NON
3	4ybq	15	4.469	0.277	0.762	-5.40	73.00	NON
4	4yb9	19	5.085	0.333	0.774	-5.40	63.72	NON

Figure 10. Details of each model and the quality score obtain

To check the validity of the models after the modeling and before refining I use a lot of tools to validate the quality.Some of them were supposed to be more suitable for membrane protein like PROQM and verify 3D.

The ProQM score was the most useless because the difference between the value is too little to be able to find the real good conformation.The Qmean score was also not very adapted.
Instead the Lgscore, MAXsub and verify 3D score were the best and the most suitable to determine the best model.
So we can see that just with the Lgscore, the MaxSub and the verifi3D score the last two model are not valid

After checking the number of the helix with Pymol and with Orempro I realize that only the first model was really a valid one.
So after that step, I use Galaxy-Web refine then SidePro2 to refine the model and the side chain. Galaxy-web gave me five model different models. So I use the two best tools to check the quality.
But the five models'quality was below the quality of the raw model obtain with memoir.Indeed, the models obtained are below 2 points than my previous model and the verify 3D score is also below (Figure 11).

Model	1	2	3	4	5
LGscore	4.247	4.114	4.430	4.425	4.378
MaxSub	0.217	0.226	0.244	0.264	0.240
verify 3D	70	68	76	82.10	75

Figure 11. Quality score obtain after the refining step

The best Model

At the beginning, I had planned to built two model one with a inward open conformation and another one with an closed outward conformation. The first model is supposed to be in inward open conformation.
If the second model had been valid, I would also had a model with a outward closed which should have been useful to use with the normal modes analysis and to perform some molecular dynamic.

The best model has an Hydrophobic thickness of 29.1 Å and 12 helix as predicted by Uniprot and T-coffee.

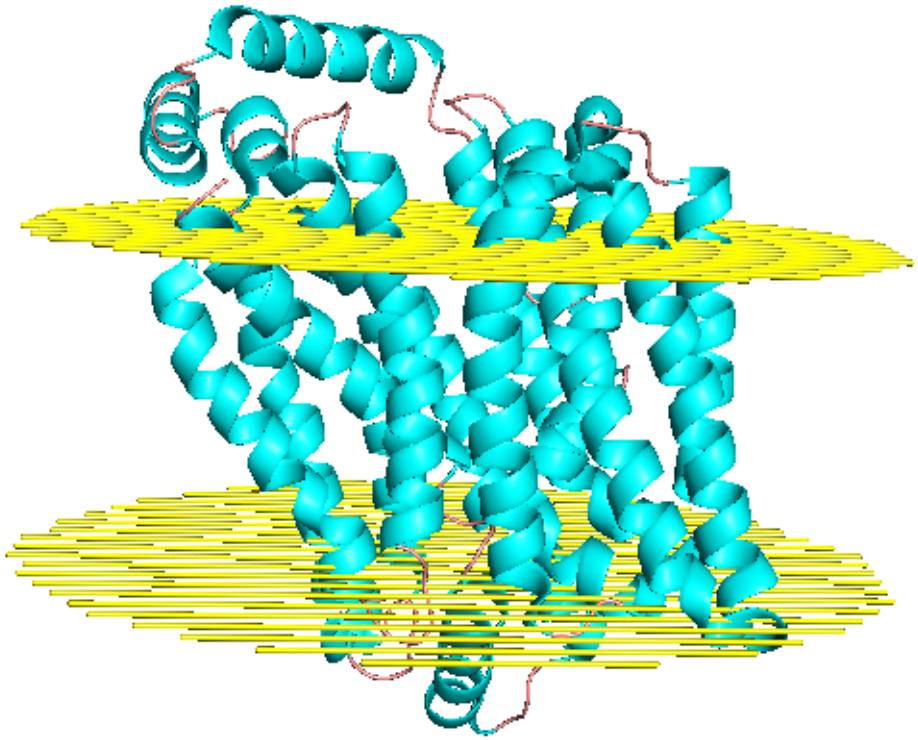

Figure 13. The best 3D model for GLUT4. the membrane is in yellow, the helix in blue and the loop in red

Normal Mode Analysis

This analysis shows that the parts that move the most are in the peripheric regions and most of the time the moving structure are loops (Fig 14). But some helix in the membrane has a translation move. The size of the barrel is changing in the same time as the entire structure move.

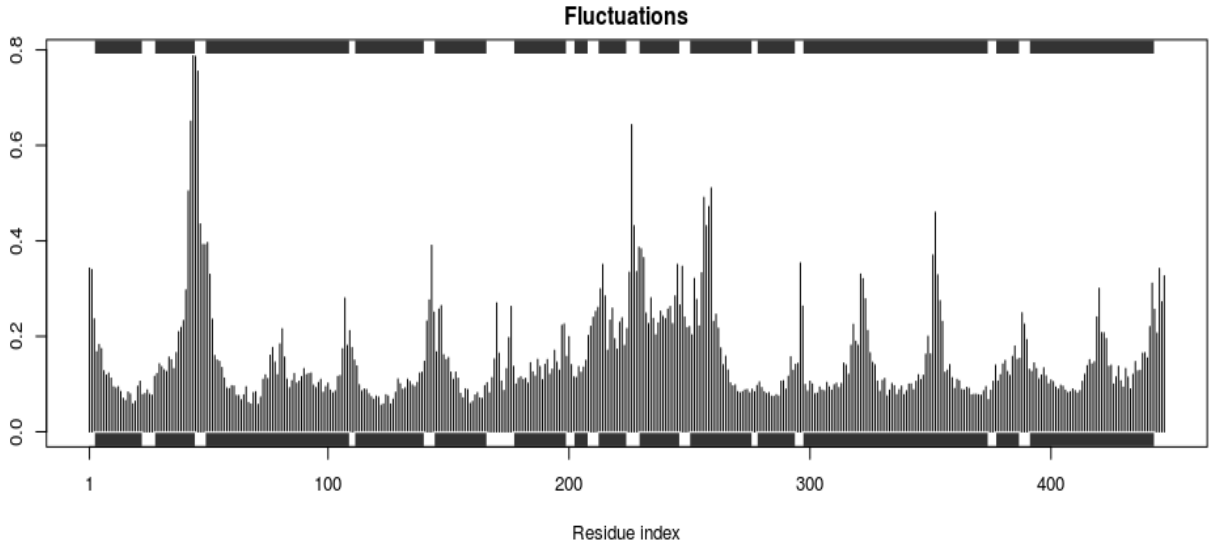

Figure14. Fluctuation 's analysis of the range of the secondary structure. Helix are represented in Black and loop in white.

Conclusion

In this study we manage to find one of the putative 3D structure of the GLUT4 family.The next step would be to realize a molecular dynamic to confirm the move of the whole protein and its stability.After these steps a docking can be the best way to confirm that the protein really binds it’s ligand.

ANNEX

ANNEX 1 .

MSA

The multiple sequence alignment result as produced by T-coffee.

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

 IN  HEL  OUT

sp|P14672|GTR4_    MPSGFQQIGSEDGEPPQQRVTGTLVLAVFSAVLGS-LQFGYNIGVINAPQKVIEQSYNETWLG
sp|P11166|GTR1_    MEP------------SSKKLTGRLMLAVGGAVLGS-LQFGYNTGVINAPQKVIEEFYNQTWVH
sp|P11169|GTR3_    MG--------------TQKVTPALIFAITVATIGS-FQFGYNTGVINAPEKIIKEFINKTLTD
sp|P43427|GTR5_    MEKE-------D-QEKTGKLTLVLALATFLAAFGSSFQYGYNVAAVNSPSEFMQQFYNDTYYD
sp|P58353|GTR5_    MEPQ-------DPVKREGRLTPVIVLATLIAAFGSSFQYGYNVAAINSPSEFMKDFYNYTYYD
                   *                  ::*  : :*   *.:** :*:*** ...*:*.:::::  * *


sp|P14672|GTR4_    RQGPEGPSSIPPGTLTTLWALSVAIFSVGGMISSFLIGIISQWLGRKRAMLVNNVLAVLGGSL
sp|P11166|GTR1_    RYGE----SILPTTLTTLWSLSVAIFSVGGMIGSFSVGLFVNRFGRRNSMLMMNLLAFVSAVL
sp|P11169|GTR3_    KGNA----PPSEVLLTSLWSLSVAIFSVGGMIGSFSVGLFVNRFGRRNSMLIVNLLAVTGGCF
sp|P43427|GTR5_    RNKE----NIESFTLTLLWSLTVSMFPFGGFIGSLMVGFLVNNLGRKGALLFNNIFSILPAIL
sp|P58353|GTR5_    RVGE----YMNEFYLTLLWSVTVSMFPFGGFLGSLMVGPLVNNLGRKGTLLFNNIFSIVPALL
                   :           ** **:::*::*..**::.*: :* : : :**: ::*. *:::.  . :


sp|P14672|GTR4_    MGLANAAASYEMLILGRFLIGAYSGLTSGLVPMYVGEIAPTHLRGALGTLNQLAIVIGILIAQ
sp|P11166|GTR1_    MGFSKLGKSFEMLILGRFIIGVYCGLTTGFVPMYVGEVSPTALRGALGTLHQLGIVVGILIAQ
sp|P11169|GTR3_    MGLCKVAKSVEMLILGRLVIGLFCGLCTGFVPMYIGEISPTALRGAFGTLNQLGIVVGILVAQ
sp|P43427|GTR5_    MGCSKIAKSFEIIIASRLLVGICAGISSNVVPMYLGELAPKNLRGALGVVPQLFITVGILVAQ
sp|P58353|GTR5_    MGFSELAKSFEMIIVARVLVGICAGLSSNVVPMYLGELAPKNWRGALGVVPQLFITIGILVAQ
                   ** .: . * *::* .*.::*  .*: :..****:**::*.  ***:*.: ** *.:***:**


sp|P14672|GTR4_    VLGLESLLGTASLWPLLLGLTVLPALLQLVLLPFCPESPRYLYIIQNLEGPARKSLKRLTGWA
sp|P11166|GTR1_    VFGLDSIMGNKDLWPLLLSIIFIPALLQCIVLPFCPESPRFLLINRNEENRAKSVLKKLRGTA
sp|P11169|GTR3_    IFGLEFILGSEELWPLLLGFTILPAILQSAALPFCPESPRFLLINRKEEENAKQILQRLWGTQ
sp|P43427|GTR5_    LFGLRSVLASEEGWPILLGLTGVPAGLQLLLLPFFPESPRYLLIQKKNESAAEKALQTLRGWK
sp|P58353|GTR5_    IFGLRSLLANEEGWPILLGLTGIPAVLQLLFLPFFPESPRYLLIQKKDEAAAKSALRRLRGWH
                   ::**  ::.. . **:**.:  :** **   *** *****:* * :: *  *.. *: * *


sp|P14672|GTR4_    DVSGVLAELKDEKRKLERERPLSLLQLLGSRTHRQPLIIAVVLQLSQQLSGINAVFYYSTSIF
sp|P11166|GTR1_    DVTHDLQEMKEESRQMMREKKVTILELFRSPAYRQPILIAVVLQLSQQLSGINAVFYYSTSIF
sp|P11169|GTR3_    DVSQDIQEMKDESARMSQEKQVTVLELFRVSSYRQPIIISIVLQLSQQLSGINAVFYYSTGIF
sp|P43427|GTR5_    DVDMEMEEIRKEDEAEKAAGFISVWKLFRMQSLRWQLISTIVLMAGQQLSGVNAIYYYADQIY
sp|P58353|GTR5_    DVDAEIEEILEEDRAEKAVGFISVLKLFKMRSLRWQVISIIVLMAGQQLSGVNAIYYYADQIY
                   **   : *: .*.        ::: :*:   : *  ::  :**  .*****:**::**:  *:


sp|P14672|GTR4_    ETAGVGQ--PAYATIGAGVVNTVFTLVSVLLVERAGRRTLHLLGLAGMCGCAILMTVALLLLE
sp|P11166|GTR1_    EKAGVQQ--PVYATIGSGIVNTAFTVVSLFVVERAGRRTLHLIGLAGMAGCAILMTIALALLE
sp|P11169|GTR3_    KDAGVQE--PIYATIGAGVVNTIFTVVSLFLVERAGRRTLHMIGLGGMAFCSTLMTVSLLLKD
sp|P43427|GTR5_    LSAGVKSNDVQYVTAGTGAVNVFMTMVTVFVVELWGRRNLLLIGFSTCLTACIVLTVALALQN
sp|P58353|GTR5_    LSAGVNEDDVQYVTAGTGAVNVLITVCAIFVVELMGRRFLLLLGFSVCFTACCVLTGALALQD
                     *** .    *.* *:* **. :*: ::::**  *** * ::*:.    .. ::* :* * :


sp|P14672|GTR4_    RVPAMSYVSIVAIFGFVAFFEIGPGPIPWFIVAELFSQGPRPAAMAVAGFSNWTSNFIIGMGF
sp|P11166|GTR1_    QLPWMSYLSIVAIFGFVAFFEVGPGPIPWFIVAELFSQGPRPAAIAVAGFSNWTSNFIVGMCF
sp|P11169|GTR3_    NYNGMSFVCIGAILVFVAFFEIGPGPIPWFIVAELFSQGPRPAAMAVAGCSNWTSNFLVGLLF
sp|P43427|GTR5_    TISWMPYVSIVCVIVYVIGHAVGPSPIPALFITEIFLQSSRPSAYMIGGSVHWLSNFIVGLIF
sp|P58353|GTR5_    VISWMPYVSIACVISYVIGHALGPSPIPALLVTEIFLQSSRPAAYMVAGTVHWLSNFTVGLVF
                      *.:::.* .:: :*  . :**.*** ::::*:* *..**:*  :.*  :* *** :*: *


sp|P14672|GTR4_    QYVAEAMGPYVFLLFAVLLLGFFIFTFLRVPETRGRTFDQISAAFHRTPSLLEQEVKPST-EL
sp|P11166|GTR1_    QYVEQLCGPYVFIIFTVLLVLFFIFTYFKVPETKGRTFDEIASGFRQGGASQSDKTPEEL--F
sp|P11169|GTR3_    PSAAHYLGAYVFIIFTGFLITFLAFTFFKVPETRGRTFEDITRAFEGQAHGADRSGKDGVMEM
sp|P43427|GTR5_    PFIQVGLGPYSFIIFAIICLLTTIYIFMVVPETKGRTFVEINQIFAKKNKVSDVYPEKEEKEL
sp|P58353|GTR5_    PFIQVGLGAYSFVIFAVICLLTTVYIFLIIPETKSKTFIEINRIFIKMNKVPGVHPEKE--EL
                         *.* *::*: : :    : :: :***:.:** :*   *                 :


sp|P14672|GTR4_    EYLGPDEN----D
sp|P11166|GTR1_    HPLGADSQ----V
sp|P11169|GTR3_    NSIEPAKETTTNV
sp|P43427|GTR5_    NDLPPATRE---Q
sp|P58353|GTR5_    KEFPPSTAR---Q
                   . : .
```

ANNEX 2

```
T-COFFEE, Version_11.00.d625267 (2016-01-11 15:25:41 - Revision d625267 - Build 507)
Cedric Notredame
CPU TIME:0 sec.
SCORE=96
*
 BAD  AVG  GOOD
*
sp|P14672|GTR4_     :  98
sp|P11166|GTR1_     :  97
sp|P11169|GTR3_     :  94
sp|P43427|GTR5_     :  93
sp|P58353|GTR5_     :  98
cons                :   9

sp|P14672|GTR4_    MPSGFQQIGSEDGEPPQQRVTGTLVLAVFSAVL-GSLQFGYNIGVINAPQKVIEQSYNETWLG
sp|P11166|GTR1_    ------------MEPSSKKLTGRLMLAVGGAVL-GSLQFGYNTGVINAPQKVIEEFYNQTWVH
sp|P11169|GTR3_    --------------MGTQKVTPALIFAITVATI-GSFQFGYNTGVINAPEKIIKEFINKTLTD
sp|P43427|GTR5_    --------MEKEDQEKTGKLTLVLALATFLAAFGSSFQYGYNVAAVNSPSE------------
sp|P58353|GTR5_    ------MEPQDPVKREGRLTPVIVLATLIAAFG-SSFQYGYNVAAINSPSEFMKDFYNYTYYD

cons                                .  :  :   *   .*:*:*** ..:*:*.:

sp|P14672|GTR4_    R------------QGPEGPSSIPPGTLTTLWALSVAIFSVGGMISSFLIGI-ISQWLGRKRAM
sp|P11166|GTR1_    R------------YG----ESILPTTLTTLWSLSVAIFSVGGMIGSFSVGLFVNRF-GRRNSM
sp|P11169|GTR3_    K------------GN----APPSEVLLTSLWSLSVAIFSVGGMIGSFSVGLFVNRF-GRRNSM
sp|P43427|GTR5_    FMQQFYNDTYYDRNK----ENIESFTLTLLWSLTVSMFPFGGFIGSLMVGFLVNNL-GRKGAL
sp|P58353|GTR5_    R------------VG----EYMNEFYLTLLWSVTVSMFPFGGFLGSLMVGPLVNNL-GRKGTL

cons                          ** **:::*::*..**::.*: :*  :..  **: ::

sp|P14672|GTR4_    LVNNVLAVLGGSLMGLANAAASYEMLILGRFLIGAYSGLTSGLVPMYVGEIAPTHLRGALGTL
sp|P11166|GTR1_    LMMNLLAFVSAVLMGFSKLGKSFEMLILGRFIIGVYCGLTTGFVPMYVGEVSPTALRGALGTL
sp|P11169|GTR3_    LIVNLLAVTGGCFMGLCKVAKSVEMLILGRLVIGLFCGLCTGFVPMYIGEISPTALRGAFGTL
sp|P43427|GTR5_    LFNNIFSILPAILMGCSKIAKSFEIIIASRLLVGICAGISSNVVPMYLGELAPKNLRGALGVV
sp|P58353|GTR5_    LFNNIFSIVPALLMGFSELAKSFEMIIVARVLVGICAGLSSNVVPMYLGELAPKNWRGALGVV

cons               *. *:::.  . :** .: . * *::* .*.::*  .*: :..****:**::*.  ***:*.:

sp|P14672|GTR4_    NQLAIVIGILIAQVLGLESLLGTASLWPLLLGLTVLPALLQLVLLPFCPESPRYLYIIQNLEG
sp|P11166|GTR1_    HQLGIVVGILIAQVFGLDSIMGNKDLWPLLLSIIFIPALLQCIVLPFCPESPRFLLINRNEEN
sp|P11169|GTR3_    NQLGIVVGILVAQIFGLEFILGSEELWPLLLGFTILPAILQSAALPFCPESPRFLLINRKEEE
sp|P43427|GTR5_    PQLFITVGILVAQLFGLRSVLASEEGWPILLGLTGVPAGLQLLLLPFFPESPRYLLIQKKNES
sp|P58353|GTR5_    PQLFITIGILVAQIFGLRSLLANEEGWPILLGLTGIPAVLQLLFLPFFPESPRYLLIQKKDEA

cons               ** *.:***:**::**  ::.. . **:**.:  :** **   *** *****:* * :: *

sp|P14672|GTR4_    PARKSLKRLTGWADVSGVLAELKDEKRKLERERPLSLLQLLGSRTHRQPLIIAVVLQLSQQLS
sp|P11166|GTR1_    RAKSVLKKLRGTADVTHDLQEMKEESRQMMREKKVTILELFRSPAYRQPILIAVVLQLSQQLS
sp|P11169|GTR3_    NAKQILQRLWGTQDVSQDIQEMKDESARMSQEKQVTVLELFRVSSYRQPIIISIVLQLSQQLS
sp|P43427|GTR5_    AAEKALQTLRGWKDVDMEMEEIRKEDEAEKAAGFISVWKLFRMQSLRWQLISTIVLMAGQQLS
sp|P58353|GTR5_    AAKSALRRLRGWHDVDAEIEEILEEDRAEKAVGFISVLKLFKMRSLRWQVISIIVLMAGQQLS

cons               *.. *: * *  **   : *: .*.        ::: :*:   : *  ::  :**  .****

sp|P14672|GTR4_    GINAVFYYSTSIFETAGVG--QPAYATIGAGVVNTVFTLVSVLLVERAGRRTLHLLGLAGMCG
sp|P11166|GTR1_    GINAVFYYSTSIFEKAGVQ--QPVYATIGSGIVNTAFTVVSLFVVERAGRRTLHLIGLAGMAG
sp|P11169|GTR3_    GINAVFYYSTGIFKDAGVQ--EPIYATIGAGVVNTIFTVVSLFLVERAGRRTLHMIGLGGMAF
sp|P43427|GTR5_    GVNAIYYYADQIYLSAGVKSNDVQYVTAGTGAVNVFMTMVTVFVVELWGRRNLLLIGFSTCLT
sp|P58353|GTR5_    GVNAIYYYADQIYLSAGVNEDDVQYVTAGTGAVNVLITVCAIFVVELMGRRFLLLLGFSVCFT

cons               *:**::**:  *:  *** .:  *.* *:* **. :*: ::::**  *** * ::*:.

sp|P14672|GTR4_    CAILMTVALLLLERVPAMSYVSIVAIFGFVAFFEIGPGPIPWFIVAELFSQGPRPAAMAVAGF
sp|P11166|GTR1_    CAILMTIALALLEQLPWMSYLSIVAIFGFVAFFEVGPGPIPWFIVAELFSQGPRPAAIAVAGF
sp|P11169|GTR3_    CSTLMTVSLLLKDNYNGMSFVCIGAILVFVAFFEIGPGPIPWFIVAELFSQGPRPAAMAVAGC
sp|P43427|GTR5_    ACIVLTVALALQNTISWMPYVSIVCVIVYVIGHAVGPSPIPALFITEIFLQSSRPSAYMIGGS
sp|P58353|GTR5_    ACCVLTGALALQDVISWMPYVSIACVISYVIGHALGPSPIPALLVTEIFLQSSRPAAYMVAGT

cons               .. ::* :* * :    *.::.* .:: :*  . :**.*** ::::*:* *..**:*  :.*

sp|P14672|GTR4_    SNWTSNFIIGMGFQYVAEAMGPYVFLLFAVLLLGFFIFTFLRVPETRGRTFDQISAAFHRTPS
sp|P11166|GTR1_    SNWTSNFIVGMCFQYVEQLCGPYVFIIFTVLLVLFFIFTYFKVPETKGRTFDEIASGFRQGGA
sp|P11169|GTR3_    SNWTSNFLVGLLFPSAAHYLGAYVFIIFTGFLITFLAFTFFKVPETRGRTFEDITRAFEGQAH
sp|P43427|GTR5_    VHWLSNFIVGLIFPFIQVGLGPYSFIIFAIICLLTTIYIFMVVPETKGRTFVEINQIFAKKNK
sp|P58353|GTR5_    VHWLSNFTVGLVFPFIQVGLGAYSFVIFAVICLLTTVYIFLIIPETKSKTFIEINRIFIKMNK

cons               :* *** :*: *       *.* *::*: : :    : :: :***:.:** :*   *

sp|P14672|GTR4_    LLEQEVKPSTELEYLGPDEND-----
sp|P11166|GTR1_    SQSDKTPEELFHPLGADSQV------
sp|P11169|GTR3_    GADRSGKDGVMEMNSIEPAKETTTNV
sp|P43427|GTR5_    VSDVYPEKEEKELNDLPPATREQ---
sp|P58353|GTR5_    VPGVHPEKEELKEFPPSTARQ-----

cons
```

Annexe 3

Fasta origin	PDB entry	Method	Resolution (Å)	Chain	Positions	conformation	RMSD (ref 5EQG)
P11166	4PYP	X-ray	3.17	A	1-492	inward-open	O.38
P11166	5EQG	X-ray	2.90	A	1-492	inward-open	0
P11166	5EQH	X-ray	2.99	A	1-492	inward-open	0 .22
P11166	5EQI	X-ray	3.00	A	1-492	inward-open	0.323

Fasta origin	PDB entry	Method	Resolution (Å)	Chain	Positions	conformation	RMSD (ref 5EQG)
P11169	4ZW9	X-ray	1.50	A	1-496	outward-occluded	0.261
P11169	4ZWB	X-ray	2.40	A	1-496	outward-occluded	0
P11169	5C65	X-ray	2.65	A/B	1-474	outward-occluded	0.503

Fasta origin	PDB entry	Method	Resolution (Å)	Chain	Positions	conformation
P58353	4YB9	X-ray	3.20	D	1-473	Open inward-facing
P43427	4YBQ	X-ray	3.27	A/B	1-502	outward-open

Bibliography

[1] R. Augustin et E. Mayoux, « Mammalian Sugar Transporters », 2014.

[2] « SLC2A4 - Solute carrier family 2, facilitated glucose transporter member 4 - Homo sapiens (Human) - SLC2A4 gene & protein ». [En ligne]. Disponible sur: http://www.uniprot.org/uniprot/P14672. [Consulté le: 24-nov-2017].

[3] M. Mueckler et B. Thorens, « The SLC2 (GLUT) Family of Membrane Transporters », *Mol Aspects Med*, vol. 34, n° 0, p. 121-138, 2013.

[4] J. Kusari, U. S. Verma, J. B. Buse, R. R. Henry, et J. M. Olefsky, « Analysis of the gene sequences of the insulin receptor and the insulin-sensitive glucose transporter (GLUT-4) in patients with common-type non-insulin-dependent diabetes mellitus. », *J Clin Invest*, vol. 88, n° 4, p. 1323-1330, oct. 1991.

[5] N. Yan, « Structural advances for the major facilitator superfamily (MFS) transporters », *Trends in Biochemical Sciences*, vol. 38, n° 3, p. 151-159, mars 2013.

[6] « T-Coffee Server ». [En ligne]. Disponible sur: http://tcoffee.crg.cat/apps/tcoffee/do:regular. [Consulté le: 25-nov-2017].

[7] « psipred : index ». [En ligne]. Disponible sur: http://bioinf.cs.ucl.ac.uk/psipred/. [Consulté le: 25-nov-2017].

[8] « Protein BLAST: search protein databases using a protein query ». [En ligne]. Disponible sur: https://blast.ncbi.nlm.nih.gov/Blast.cgi?PROGRAM=blastp&PAGE_TYPE=BlastSearch&LINK_LOC=blasthome. [Consulté le: 25-nov-2017].

[9] « SWISS-MODEL | Workspace ». [En ligne]. Disponible sur: https://swissmodel.expasy.org/interactive. [Consulté le: 25-nov-2017].

[10] J.-P. Ebejer, J. R. Hill, S. Kelm, J. Shi, et C. M. Deane, « Memoir: template-based structure prediction for membrane proteins », *Nucleic Acids Research*, 2013.

[11] S. Kelm, J. Shi, et C. M. Deane, « MEDELLER: homology-based coordinate generation for membrane proteins », *Bioinformatics*, vol. 26, n° 22, p. 2833-2840, nov. 2010.

[12] L. Heo, H. Park, et C. Seok, « GalaxyRefine: Protein structure refinement driven by side-chain repacking », *Nucleic Acids Res.*, vol. 41, n° Web Server issue, p. W384-388, juill. 2013.

[13] K. Nagata, A. Randall, et P. Baldi, « SIDEpro: a novel machine learning approach for the fast and accurate prediction of side-chain conformations », *Proteins*, vol. 80, n° 1, p. 142-153, janv. 2012.

[14] « ProQ - Protein quality prediction ». [En ligne]. Disponible sur: http://proq.bioinfo.se/cgi-bin/ProQ/ProQ.cgi. [Consulté le: 24-nov-2017].

[15] « OREMPRO: a web server for positioning and assessing membrane protein structures ». [En ligne]. Disponible sur: http://www.dsimb.inserm.fr/dsimb_tools/OREMPRO/en/index.php. [Consulté le: 25-nov-2017].

[16] « SWISS-MODEL ». [En ligne]. Disponible sur: https://swissmodel.expasy.org/qmean/. [Consulté le: 24-nov-2017].

[17]« Bio3D ». [En ligne]. Disponible sur: http://thegrantlab.org/bio3d/index.php. [Consulté le: 24-nov-2017].

[18]A. Grosdidier, V. Zoete, et O. Michielin, « SwissDock, a protein-small molecule docking web service based on EADock DSS », *Nucleic Acids Res.*, vol. 39, nº Web Server issue, p. W270-277, juill. 2011.