Report of the Structure Project

Masters second year, Bioinformatics

University Paris 7

Template-based modelling of NPT4

Authors:

Etienne JEAN

Adam BELLAICHE

Teacher:

Pr. Catherine Etchebest

Years 2018/2019

Introduction

The Sodium-dependent phosphate transporter protein 4, NPT4, is a human membrane transport protein which belongs to the Major Facilitator Superfamily (MFS). This family of active transporters gathers proteins conserved from bacteria to humans [1]. The proteins in this family allow the transport of small solutes in response to chemiosmotic gradients. The solutes can be sugars, drugs, metabolites, oligosaccharides, amino acids and oxyanions [1].

Although the amino acids sequences of the members of the MFS can differs significantly due to their evolutionary distance, their structures remains highly similar. They are typically composed of 12 (sometimes 14) transmembrane helices, with the N-ter and C-ter ends facing the cytoplasm [1]. Their structures can be viewed as two 6-helices domains symmetrical to each other by a 180° rotation around an axis perpendicular to the membrane plane [1].

The structure of the MFS proteins have been observed in six different conformations: inward open, outward open, substrate-bound outward facing partly occluded, ligand bound inward open, outward occluded, inward occluded. These conformations are dependent on the presence of a ligand [1].

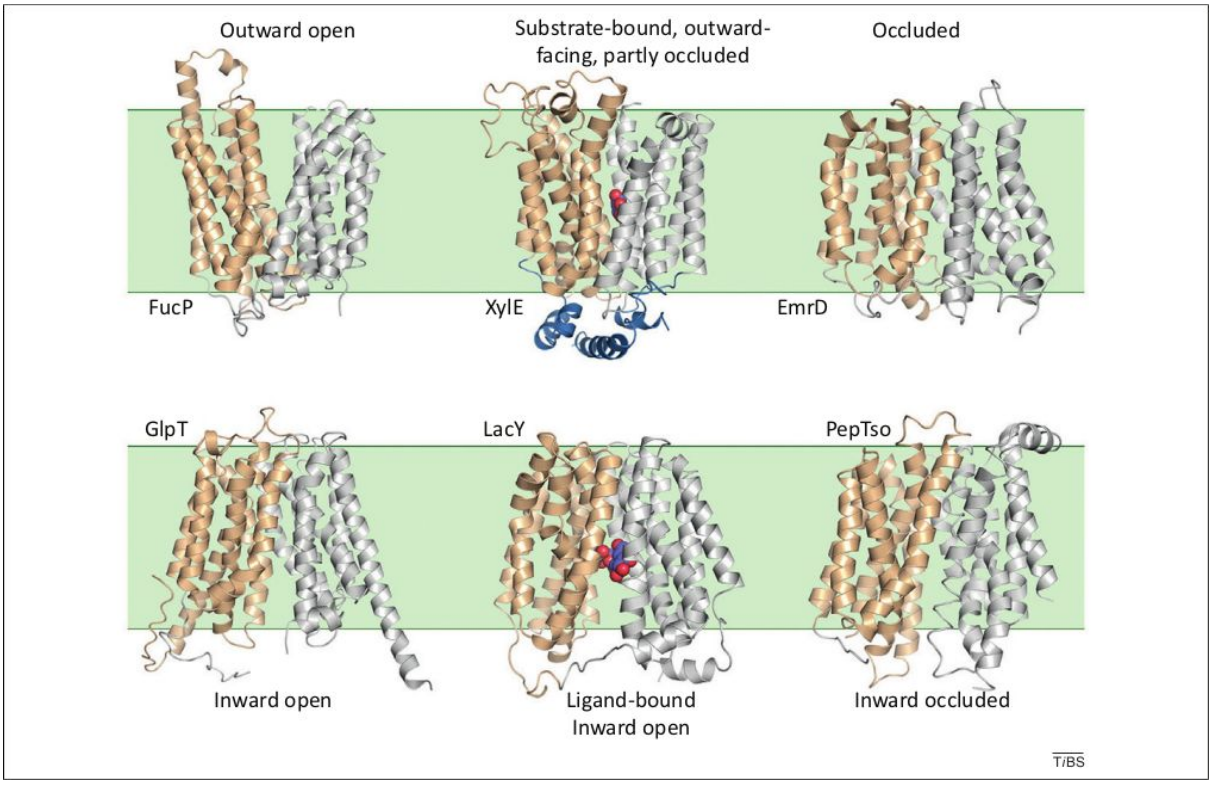


Image1: the differents conformations in the MFS.

To date, only a very few of the MFS proteins' structures have been resolved, and mainly in the bacteria kingdom. In this study, we try to generate two models of the structure of NPT4 by homology, corresponding to the inward-open and outward-open conformations. We also compare results between raw and refined models, inward-open and outward-open conformations, and single-template versus multiple-templates modelling.

Based on the last studies (2018), NPT4 is voltage dependent efflux transport for urate, anionic compounds and drugs in renal proximal tubule cells. But the community is not sure about that because somes results indicates others transports [3].

In this report, NPT4 is considered as a transporter of urate, anionic compound and drugs. That's why, it could be interesting to get the structure of NPT4. In fact, a disease called Gout is linked to the metabolism of urate.

This disease is linked to the increase of the urate in the blood, and causes a lot of suffering, so if NPT4 is a transporter of urate, it is a good target to understand this increase and find a solution to it, because an increase of the urate in blood could link to a bad transport of urate from blood to cell.

Material and methods

Transmembrane segments prediction

The prediction of transmembrane fragments was performed from the amino acids sequence using Phobius[8].

Secondary structure prediction

The prediction of secondary structure was made from the amino acids sequence using Jpred4[6].

Template structures search

A search for homolog proteins was made in order to find template structures. We performed a HMM-HMM search with HHblits [4] on the Protein Data Bank (PDB). Templates for both inward-open and outward-open structures were searched separately, and this criteria was inferred from the literature of each structure. Among

the results, the best 3 matches in terms of e-values were kept for inward-open and outward-open conformations.
Alongside, a PSI-BLAST [7] was launch to compare the obtained results.

Inputs

We use some inputs to do a modelling of NPT4:

- PDB from RCSB (found by HHBlits):

INWARD-OPEN	4J05 [PDB 1]	4LDS [PDB 2]	4IKV [PDB 3]
OUTWARD-OPEN	4GC0 [PDB 4]	4ZW9 [PDB 5]	5AYN [PDB 6]

Table 1: the PDB chosen from HHpred.

- Fasta from Uniprot:
    - NPT4 isoform 2

Single-template homology-based modelling

The single-template modelling was performed using Swiss-Model[9-14] with default parameters. 3 models for both inward-open conformations and outward-open conformations were generated, using the structures previously selected.

Multiple-template homology-based modelling

The multiple-template modelling was made using HHpred on the web site MPI Bioinformatics Toolkit [4]. The HHblits was performed on the PDB with default parameters. Then the obtained alignments are annotated with the prediction of secondary structure. On the Image 2, it is possible to observe an alignment in HHpred.

```
No 2
>4LDS_B Bicyclomycin resistance protein TcaB; alpha helical transmembrane protein, glucose; 3.2A {Staphylococcus epidermi
Probab=100.00  E-value=7.8e-34  Score=270.78  Aligned_cols=413  Identities=14%  Similarity=0.109  Sum_probs=0.0  Template

Q ss_pred             HHHHHHHHHHHHHHHHHHHHHHHHHhhhheeecccCCccccCCCCCcccccccCccccCCcCCCCcCcccCCChhHHHH
Q sp|O00476-2|NP   35 CSARYGIALVLHFCNFTTIAQNVIMNITMVAMVNSTSPQSQLNDSSEVLPVDSFGGLSKAPKSLPAKAPVYDWSPQIQGI  114 (498)
Q Consensus       35 ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~lp~~~~~~g~s~~~~g~  114 (498)
                        +..++.....+.+..++.................                                   ++|.|..+.|+
T Consensus        2 ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~-----------------------------------~~~~s~~~~g~   46 (446)
T 4LDS_B          2 KANKYLIFILGALGGLLYGYDNGVISGALLFIHK-----------------------------------DIPLNSTTEGI   46 (446)
T ss_dssp            -----CHHHHHHHHHHHHHHHHHHTTTHHHHGGG-----------------------------------GCCCCSHHHHH
T ss_pred             chHHHHHHHHHHHHHHHHHHHHHhhhhHHHHHHHHHh-----------------------------------cCCCCCcchHH
```

Image 2: example of what contains an alignement in HHpred.

Next, a HMM profile of the query is generated from the multiple alignment that includes the information about predicted secondary structure. It use to compare, with

the HHsearch software, query HMM profile to each HMM profile of the selected database (here uniclust30). At the end, there are lot of hits returned with scores which could be a potential template for the modelling. And on the MPI web site, HHpred go to search on PDB/mmCIF database, files matched with the returned hits by HHsearch/HHblits.

Then HHPred create a PIR file based on the choice of the templates to send it to MODELLER [5]. And MODELLER returns a PDB file.

We previously run our inputs on MEDELLER, a programs to template based modelling for transmembrane proteins, but the result was not very differents.

Models refining

The refining of the obtained models was made in 2 steps. First, the Galaxy Refine[10] tool was used to refine the whole molecules by doing two rounds of relaxation. Then, a more specific refining was performed on the outer-membrane and cytoplasmic loops.

Quality assessment of models

Swiss-Model's tool QMEANBrane scoring function[15] was used to check energy and evaluate the quality of the models.
Swiss-Model's tool Structure assessment was used to assess, knowledge-based, the model. ProSA/proQM[16-17] was used to complete the assessment.

Results and discussion

Choice of NPT4 isoform

NPT4 has two known isoforms produced by alternative splicing[2]. It has been showed by Jutabha P. and al.[2] that the canonical isoform 1 (420 amino acids long) does not show any transport activity and neither is located on the plasma membrane, unlike the isoform 2 (498 amino acids long).
Moreover, the prediction of transmembrane fragments with Phobius shows that the isoform 1 should have only 8 transmembrane fragments, which does not correspond to the characteristics of the MFS, whereas the isoform 2 has 12 (fig. 1).
Thus, the choice was made to use the sequence of the isoform 2 to model NPT4.

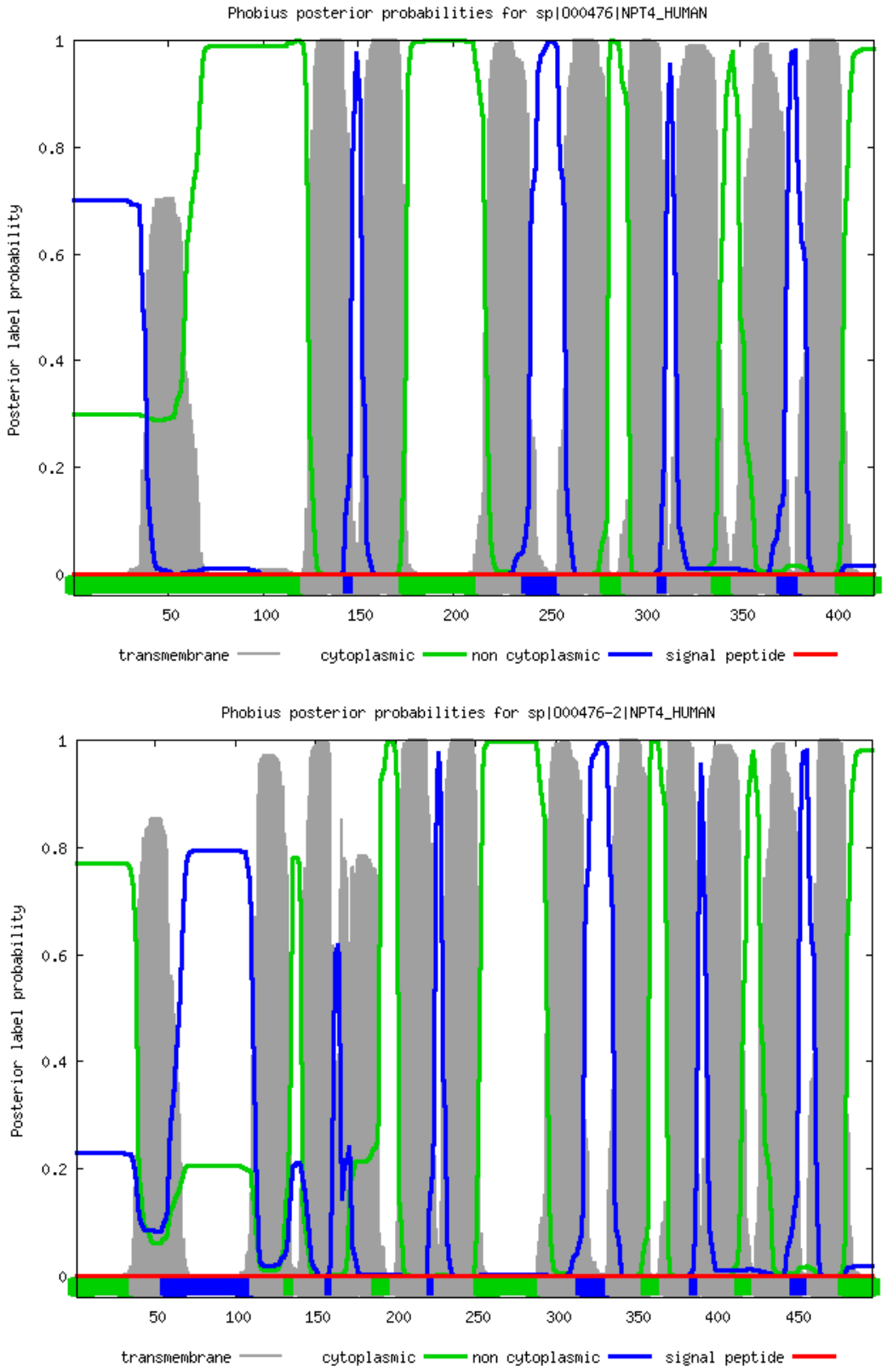


Fig 1 : Transmembrane segments prediction. The horizontal axis shows the sequence position. The vertical axis show the probability of a portion of the sequence to be either an outer-membrane (blue), cytoplasmic (green) or transmembrane (grey) portion. Upper plot = isoform 1. Lower plot = isoform 2.

Secondary structures correlates with transmembrane fragments.

The prediction of secondary structures with JPred4 shows 12 alpha helices which superimpose with the transmembrane segments previously described (fig. 2). This confirms the belonging of NPT4 to the MFS, and will later guide the choice of the template structure.

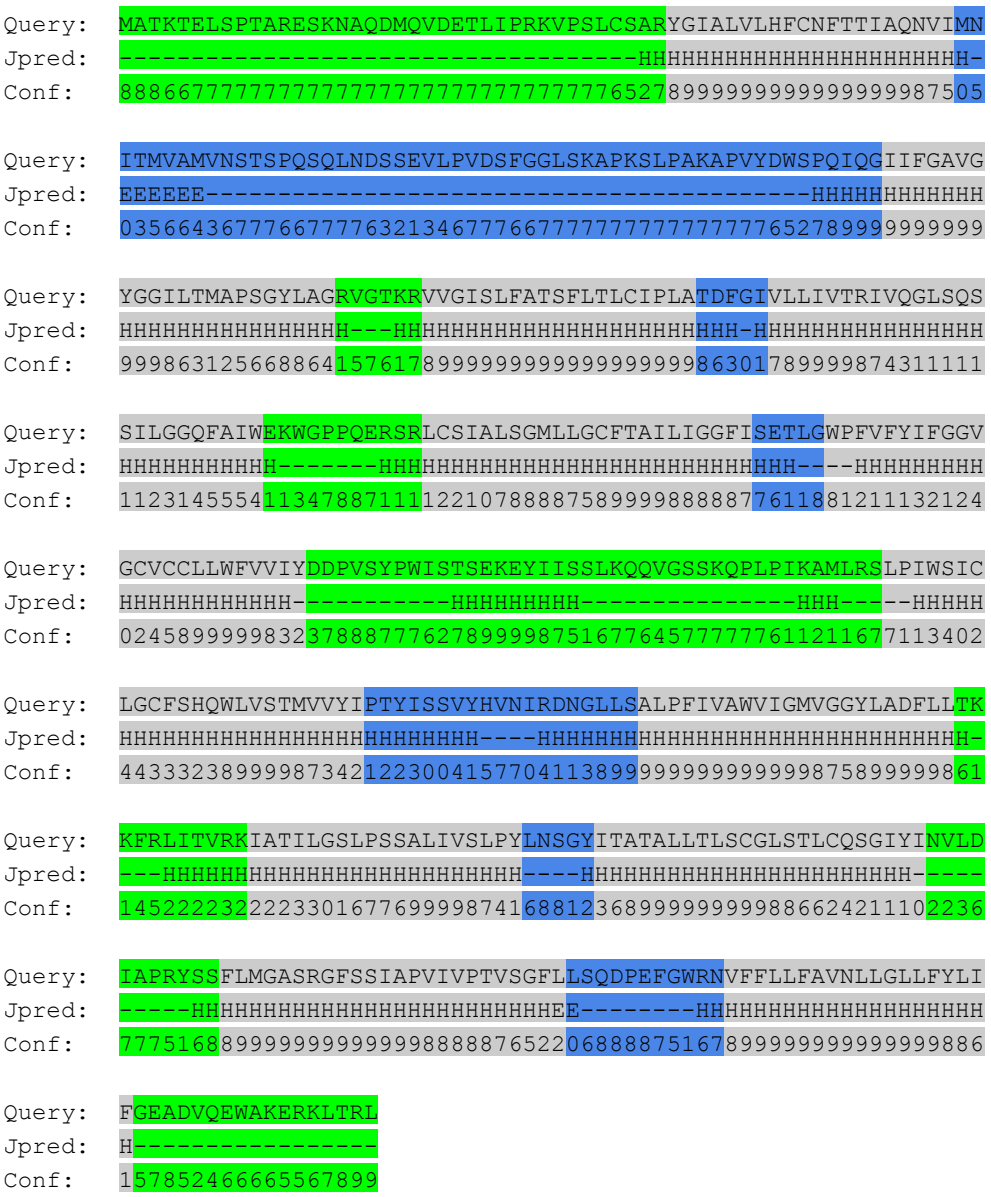


Fig 2 : Prediction of secondary structures with Jpred. H = Alpha helices, E = Beta sheets. The colors represent the prediction of transmembrane fragments made by Phobius. Grey = transmembrane segment, Green = cytoplasmic segment, Blue = outer-cell segment.

Comparison before and after model refinement

The models were generated using Swiss-prot and HHpred. Here, we describe the quality improvement brought by a detailed refinement of those models. As an example, we used the model generated with the single template structure of PDB accession code 4LDS.

Raw model quality assessment

The model was evaluated with both QMEAN and QMEANBrane tools from Swiss-model (fig. 3).
The global QMEAN value for the model was -7.79. But the quality of the model is heterogen, as the transmembrane alpha helices seem to be of good quality, whereas the loop are poorly predicted (fig. 3A & 3B). The same result is found with ProQM, which gives a poor score for regions from position 30 to 70 which correspond to a big loop in the outer-cell side, and from position 210 to 260 which is another big loop in the cytoplasmic side. This can probably be explained by the high conservation of the structure of the 12 transmembrane segments across the MFS. The loops however depend from one protein to another, and are not under high selective pressure, which makes their prediction a lot more random.

Model refining

The model was then refined in two steps using Galaxy web server. A first global refinement was performed, followed more precise but deeper refinement of the loops. This global refinement eliminated almost all poor rotamers and clashes in the structure, whereas the loop refinement aimed at improving the biggest loops structures (fig. 4).

Model final quality assessment

The same quality analysis was performed after the refining. The global qmean value was -5.90 which is slightly better. The overall quality of the model seemed improved, and so did the quality of the refined loops. Indeed, the QMEAN value has slightly increased in loops regions (fig. 3C & 3D).
The Galaxy web server allows the refinement of no more than 3 loops of 20 residues maximum in length. The biggest loops in the model are around 50 residues long, so they could have been better refined without such limitations.

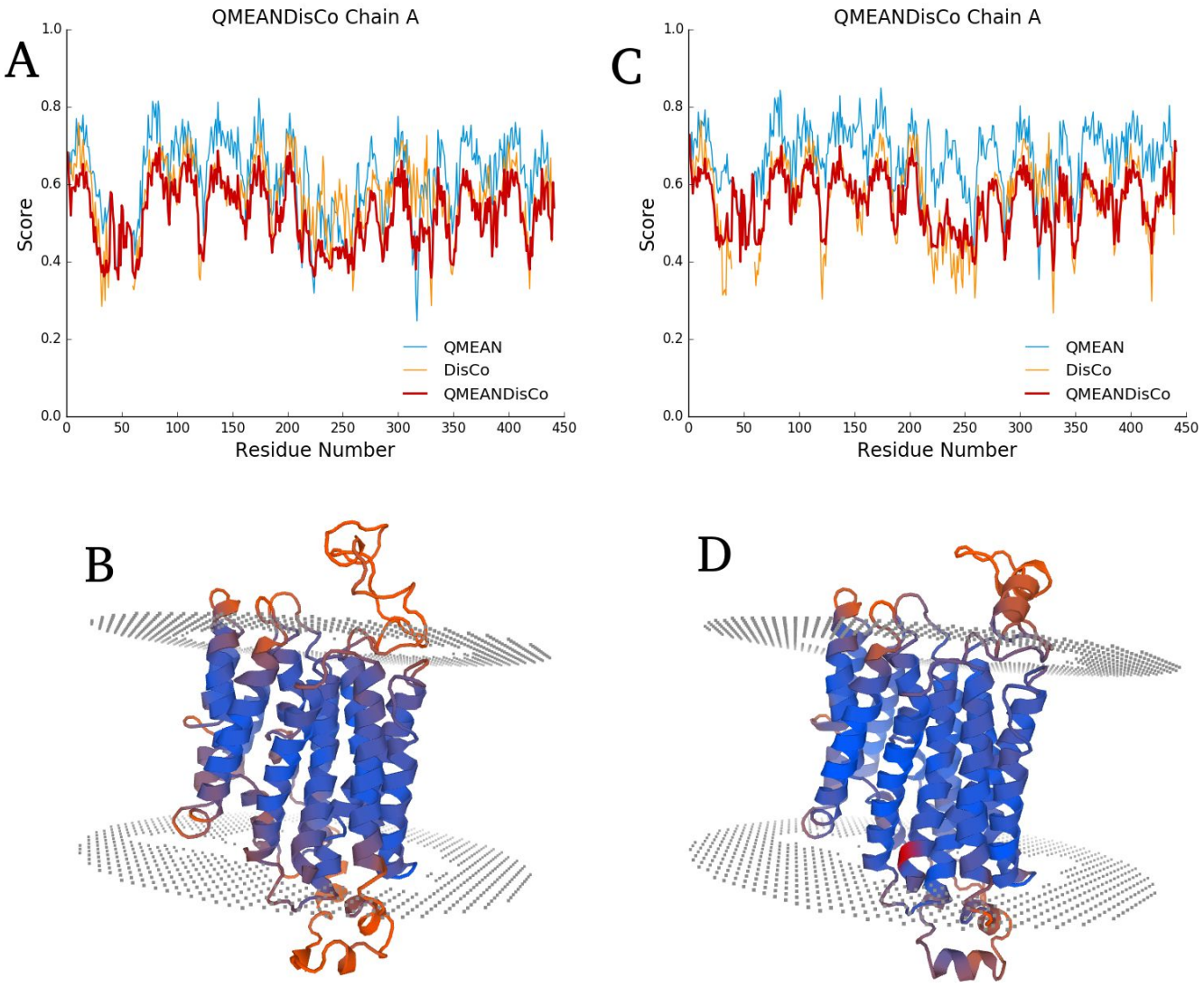


Figure 3 : Refinement of the model of NPT4. **A :** QMEAN curves as a function of the position in the sequence of the model of NPT4 before refining (from Swiss-model). The QMEAN curve (blue) shows two low quality sections, from amino acids 30 to 70 and 210 to 260 correspondings to the biggest loops in the model. **B :** Cartoon representation of the model of NPT4 before refinement. In blue are the good QMEAN values whereas in orange are the bad QMEAN values. The 12 transmembrane fragments are well predicted, unlike the loops on either sides of the membrane. **C :** QMEAN curves as a function of the position in the sequence of the model of NPT4 after refining (from Swiss-model). The QMEAN curve (blue) shows a better quality loops quality than in figure 3 (from amino acids 30 to 70 and 210 to 260 correspondings to the biggest loops in the model). **D :** Cartoon representation of the model of NPT4 after refinement.

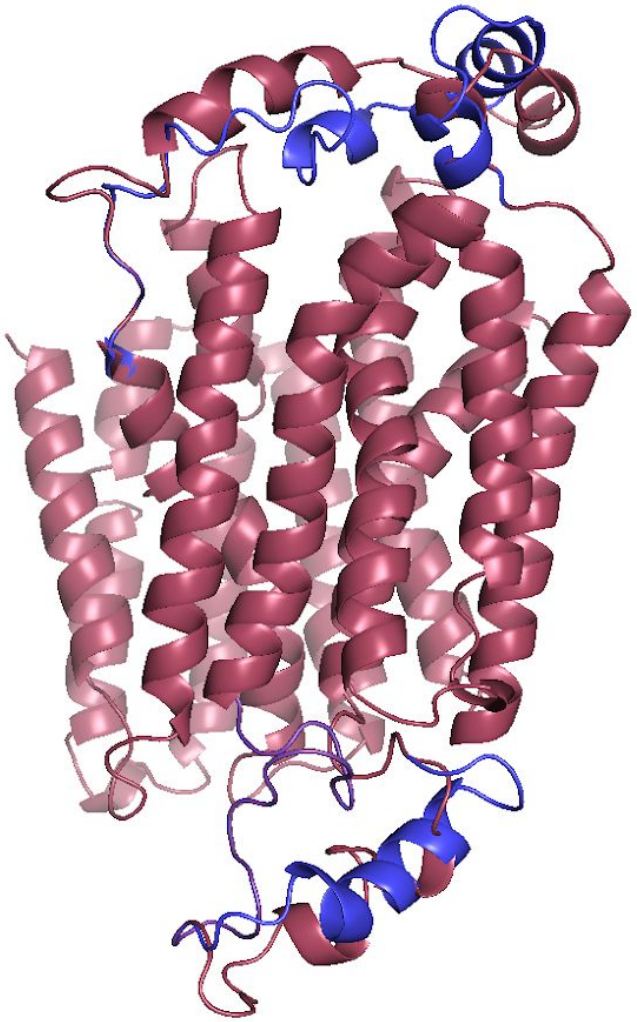
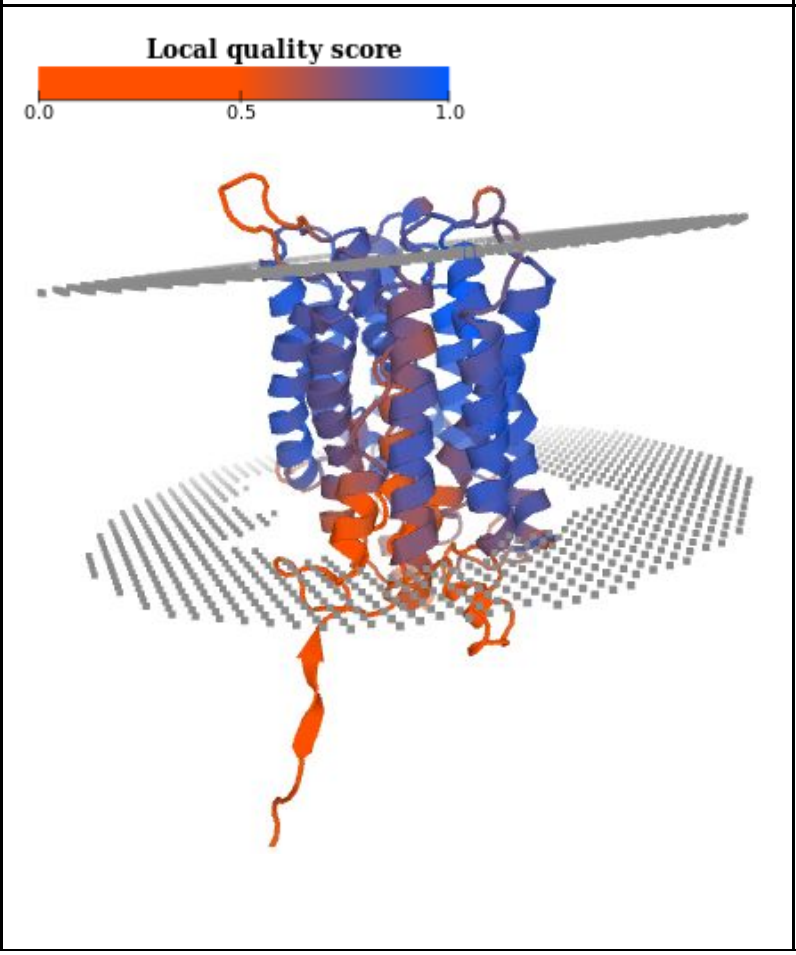
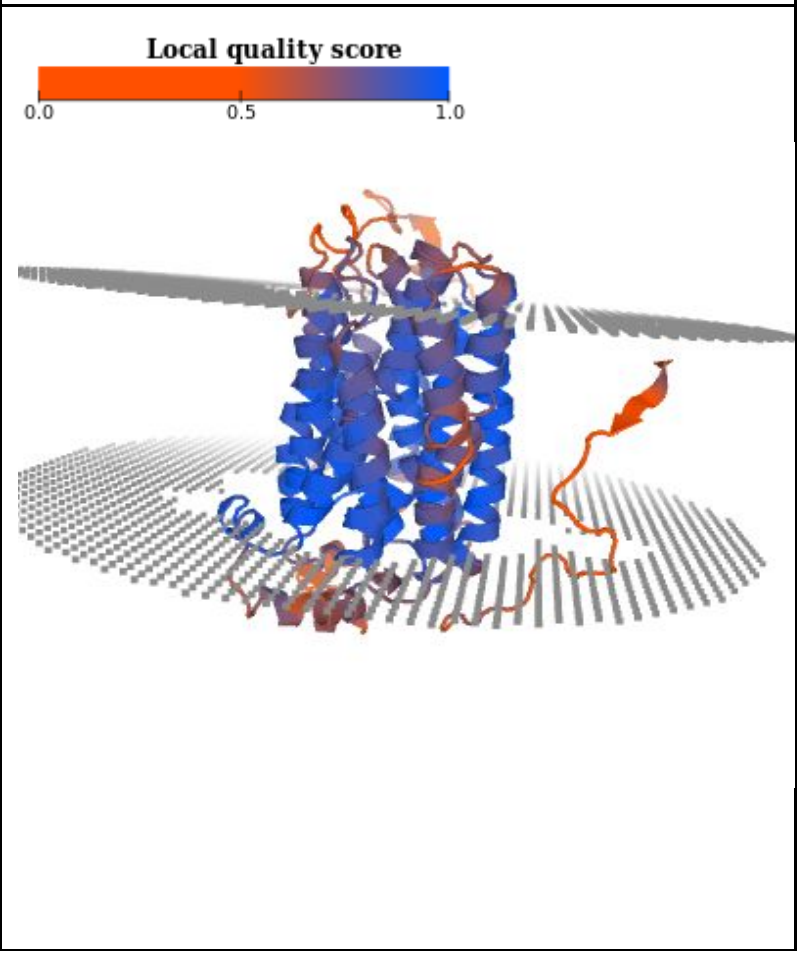
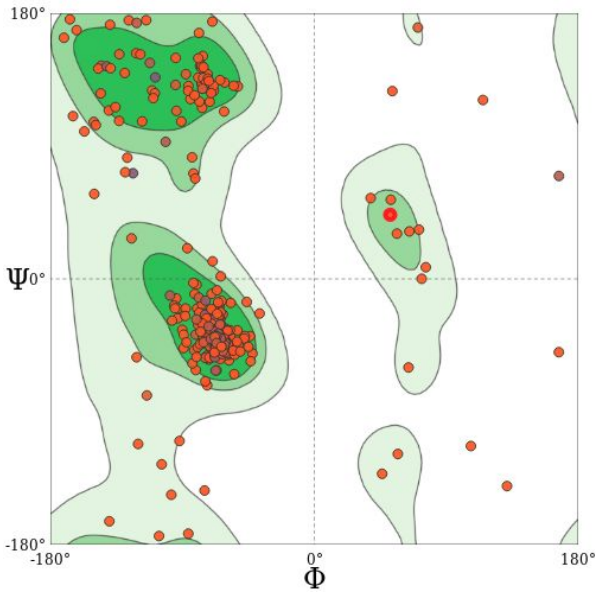


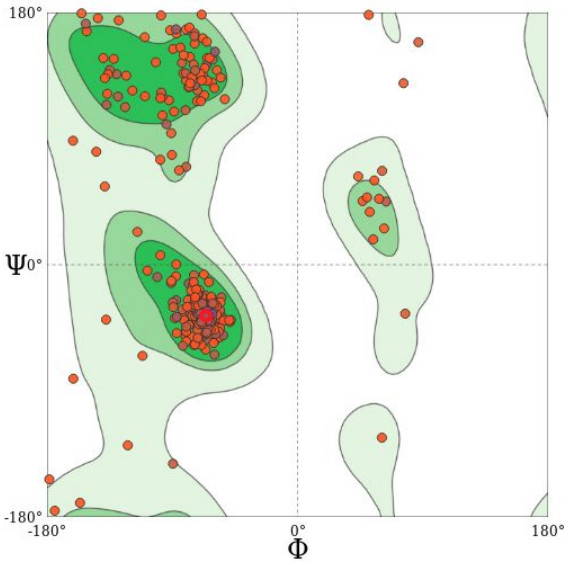
Figure 4 : Illustration of the movement of the loops after refining of the biggest loops in the model.

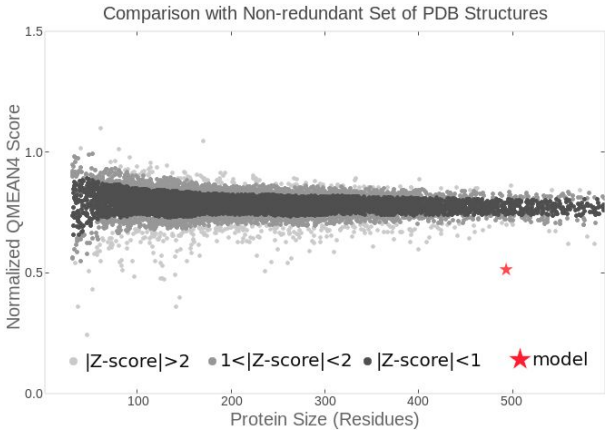
Comparison between inward-open and outward open models

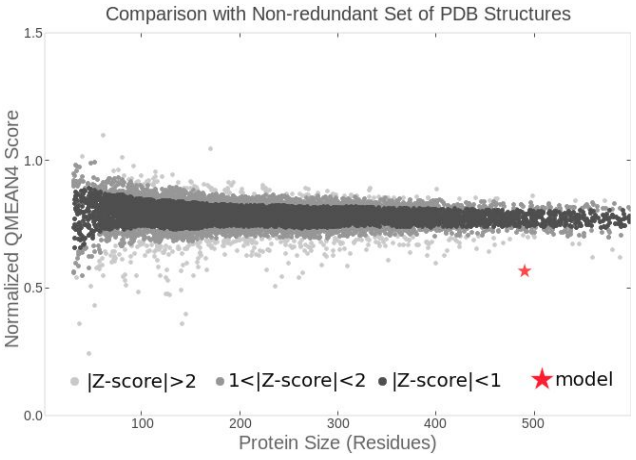
Here, the refined results, assessment and analysis of the models obtained by multiple-template homology based modelling for the inward-open and outward-open models (HHpred, Swiss-model tools, proSA/proQM).

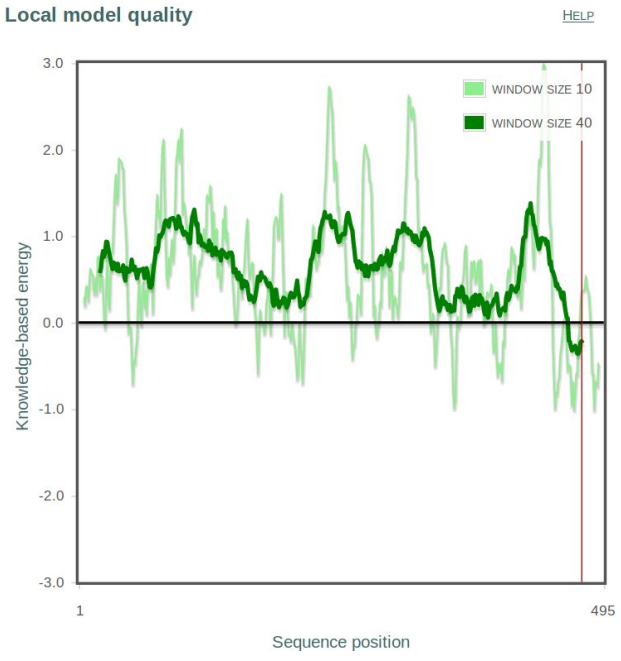
INWARD-OPEN	OUTWARD-OPEN
	

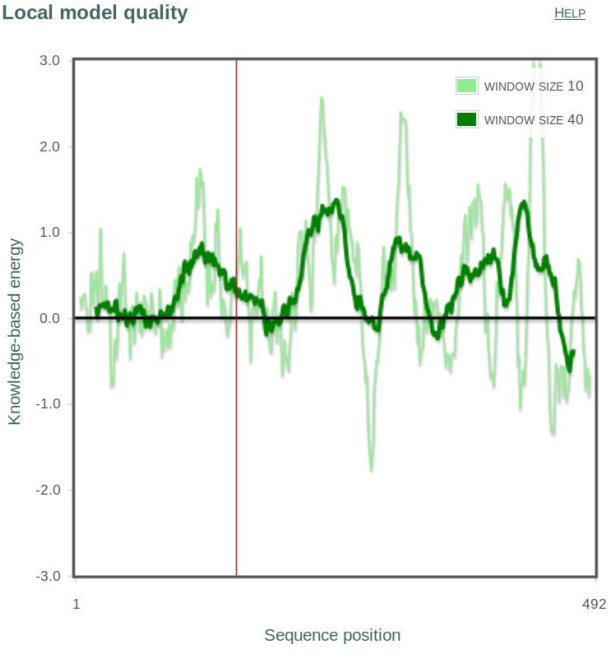

180°
Ψ 0°
-180°
-180°
0°
180°
Φ

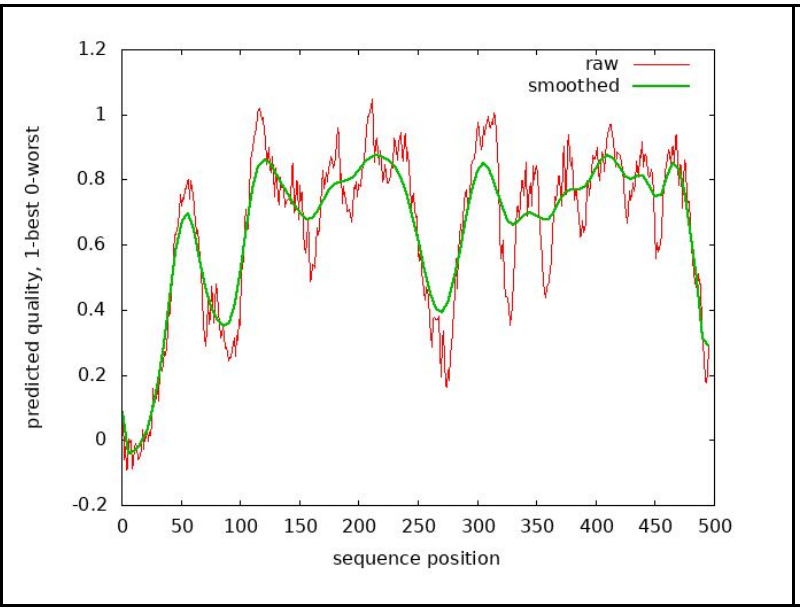

180°
Ψ 0°
-180°
-180°
0°
180°
Φ


Comparison with Non-redundant Set of PDB Structures
Normalized QMEAN4 Score
1.5
1.0
0.5
0.0
100
200
300
400
500
Protein Size (Residues)
|Z-score|>2
1<|Z-score|<2
|Z-score|<1
model


Comparison with Non-redundant Set of PDB Structures
Normalized QMEAN4 Score
1.5
1.0
0.5
0.0
100
200
300
400
500
Protein Size (Residues)
|Z-score|>2
1<|Z-score|<2
|Z-score|<1
model


Local model quality
HELP
WINDOW SIZE 10
WINDOW SIZE 40
Knowledge-based energy
3.0
2.0
1.0
0.0
-1.0
-2.0
-3.0
1
495
Sequence position


Local model quality
HELP
WINDOW SIZE 10
WINDOW SIZE 40
Knowledge-based energy
3.0
2.0
1.0
0.0
-1.0
-2.0
-3.0
1
492
Sequence position



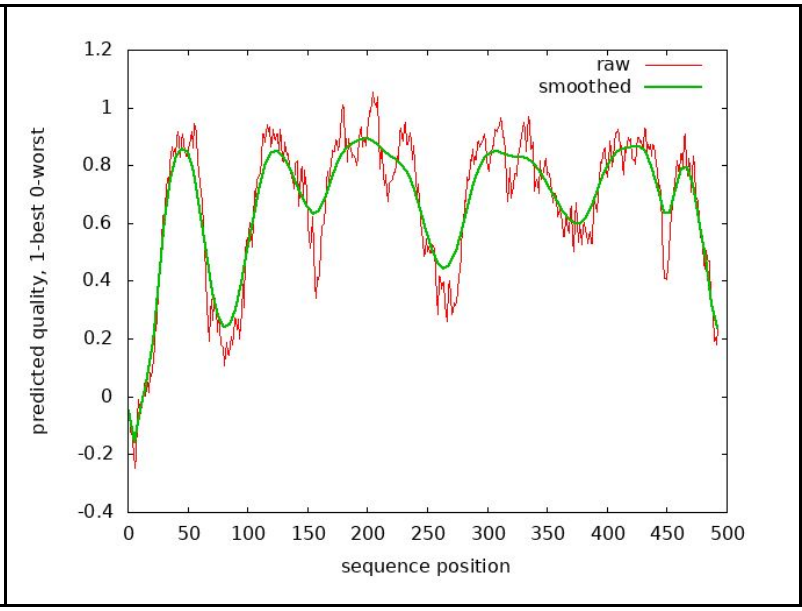


Table 2: the results of HHpred+MODELLER and the differents plots of assessments analysis. In the order: QMEANbrane, Swiss-model assessment (Ramachandran), Swiss-model assessment (QMEAN score), proSA-web server, proQM web server.

By observing the assessments of inward/outward models based on multi-templates, there are somes differences:

- **QMEANbrane**: the outward model seems to be better on the quality score of its helices. In fact, inward model has more red parts(low quality) than outward model.
- **Ramachandran diagrams**: the outward model has more residues on the favorable areas than inward model while inward model has more residues on the unfavorable areas than outward model.
- **Knowledge-based QMEAN score**: the outward model seems to bed better place than inward model, indeed, the QMEAN score of the inward model is near to 0.4 and the QMEAN score of the outward model is near to 0.3.
- **Knowledge-based energy**: According to the QMEAN score, the knowledge-based energy of the outward model seems to be better than inward model. In fact, at somes positions has a better superposition and follows better the knowledge-based curve. For instance, at the position between 200-300, the knowledge-based curve goes under zero, like the outward curve and unlike the inward curve.
- **Predicted quality**: Outward model seems to be better than inward model.

We can think that this difference of quality comes from two reasons:

- the choice of the templates: in fact, the only thing which differs between the inward modeling and the outward modelling is the templates. There are possibly one reason:

- PDB for outward conformation has a better quality than PDB for inward conformation.

This difference of quality can bring us the second reasons of the difference between outward/inward quality models:

- the natural conformation of NPT4. Indeed, the standard state/conformation of NPT4 is the outward facing [18]. Modeling outward model is more judicious than modeling inward model because of this “natural” conformation. In fact, the inward conformation is induced by a ligand and try to modelling this conformation without considering the ligand could bring low quality scores. For instance, the PDBs: 4LDS and 4IKV doesn't have a ligand unlike 4J05 which has a ligand. We should use 5OXO or 4ZP0.

So we think that the best choice to modeling NPT4 is to take outward templates, restart all the processus (HHpred, MODELLER, QMEANSbrane, Swiss-model assessment, proSA/proQM). Then, launch a MD with a complex composed by: predicted model outward in a membrane, urate to see how the NPT4-outward switches to inward conformation, and see the structure of the inward model from the MD. Or we can just restart the modelling of inward model and take only templates which has a ligand. But there are few PDBs with high quality and with a ligand. And it could be difficult to find two or more PDBs with ligand and the same ligand on all the PDBs. As we said, NPT4 can has severals ligands.

Comparison between single-template and multiple-template models

To compare the models that we generated, we looked at the QMEAN value of each of them with Swiss-model (tab. 3). Surprisingly, the multiple-template based models do not have the best QMEAN, as we would have expected. This can be due to the fact that the single-template models and multiple-template models use different algorithms. In particular, the single-template models were generated with Swiss-model, which especially tries to optimize the QMEAN value.
Another reason might be close similarity between our template structures. Multiple-template modelling is useful for combining several template structures that align on different portions of the query sequence, which was not our case.
On the other hand, the quality of the model is very dependent on the choice of the model, and the best template are not necessarily those with the best e-value. Thus, it is important to create many models from many templates, and choose the best by looking especially at structure of the more variable parts. In our case, the model quality mainly depends on the structure of the loops, which, unlike the transmembrane segments, are very variable from one template to another.

INWARD-OPEN conformation	QMEAN of the model
Multiple-template	-7.51
Single-template - Template 4LDS	-4.93
Single-template - Template 4IKV	-5.92
Single-template - Template 4J05	-6.77

OUTWARD-OPEN conformation	QMEAN of the model
Multiple-template	-6.11
Single-template - Template 4ZW9	-6.81
Single-template - Template 4GC0	-5.48
Single-template - Template 5AYN	-6.11

Table 3 : QMEAN comparison of the different models generated, and after refinement.

Conclusion

In this study, we generated several structural models for the human sodium-dependent phosphate transporter 4, in its two main conformations, inward-open and outward-open. The quality of our models is good for the transmembrane part, the more conserved part in the MFS. It is not as good for the variable parts however, and those would need both a good template structure and heavy refining in order to improve their conformation.
To follow the analysis of NPT4, and with two good inward-open and outward-open structures, it would be interesting to perform a normal mode analysis. With the right parameters, this could show the transition from inward-open to outward-open.
Secondly, it would be very interesting to perform a docking of several types of ligands. The ligands of NPT4 are thought to be phosphate molecules, but this assumption is still discussed and uncertain[3].
Finally, with a good outward model and a good docking, it would be interesting to perform a molecular dynamic. Starting from the outward-open conformation embedded into a membrane, and with a chemical docked inside it, we could try to observe the transition from outward-open to inward-open.

References

[1]: Yan N (2013) Structural advances for the major facilitator superfamily (MFS) transporters. Trends Biochem Sci 38:151–159. doi: 10.1016/j.tibs.2013.01.003

[2]: Jutabha P., Anzai N., Kitamura K., Taniguchi A., Kaneko S., Yan K., Yamada H., Shimada H., Kimura T., Katada T., Fukutomi T., Tomita K., Urano W., Yamanaka H., Seki G., Fujita T., Moriyama Y., Yamada A., Uchida S., Wempe M.F., Endou H., Sakurai H. Human sodium phosphate transporter 4 (hNPT4/SLC17A3) as a common renal secretory pathway for drugs and urate. J. Biol. Chem. 2010;285(45):35123–35132. doi: 10.1074/jbc.M110.121301.

[3]: Weifeng Z., Deng Y., Xiaodong Z. Multiple Membrane Transporters and Some Immune Regulatory Genes are Major Genetic Factors to Gout. Open Rheumatol J. 2018; 12: 94–113. doi: 10.2174/1874312901812010094

[4] : Zimmermann L, Stephens A, Nam SZ, Rau D, Kübler J, Lozajic M, Gabler F, Söding J, Lupas AN, Alva V. A Completely Reimplemented MPI Bioinformatics Toolkit with a New HHpred Server at its Core. J Mol Biol. 2018 Jul 20. S0022-2836(17)30587-9.

[5] : Webb B, Sali A. Comparative Protein Structure Modeling Using MODELLER. Curr Protoc Protein Sci. 2016 Nov 1;86:2.9.1-2.9.37.

[6]: Alexey Drozdetskiy, Christian Cole, James Procter, Geoffrey J. Barton; JPred4: a protein secondary structure prediction server, Nucleic Acids Research, Volume 43, Issue W1, 1 July 2015, Pages W389–W394, https://doi.org/10.1093/nar/gkv332

[7]: Altschul, S.F., Madden, T.L., Schäffer, A.A., Zhang, J., Zhang, Z., Miller, W. & Lipman, D.J. (1997) "Gapped BLAST and PSI-BLAST: a new generation of protein database search programs." Nucleic Acids Res. 25:3389-3402.

[8]: Lukas Käll, Anders Krogh and Erik L. L. Sonnhammer. A Combined Transmembrane Topology and Signal Peptide Prediction Method. Journal of Molecular Biology, 338(5):1027-1036, May 2004.

[9]: Waterhouse, A., Bertoni, M., Bienert, S., Studer, G., Tauriello, G., Gumienny, R., Heer, F.T., de Beer, T.A.P., Rempfer, C., Bordoli, L., Lepore, R., Schwede, T. SWISS-MODEL: homology modelling of protein structures and complexes. Nucleic Acids Res. 46(W1), W296-W303 (2018).

[10]: Bienert, S., Waterhouse, A., de Beer, T.A.P., Tauriello, G., Studer, G., Bordoli, L., Schwede, T. The SWISS-MODEL Repository - new features and functionality. Nucleic Acids Res. 45, D313-D319 (2017).

[11]: Guex, N., Peitsch, M.C., Schwede, T. Automated comparative protein structure modeling with SWISS-MODEL and Swiss-PdbViewer: A historical perspective. Electrophoresis 30, S162-S173 (2009).

[12]: Benkert, P., Biasini, M., Schwede, T. Toward the estimation of the absolute quality of individual protein structure models. Bioinformatics 27, 343-350 (2011).

[13]: Bertoni, M., Kiefer, F., Biasini, M., Bordoli, L., Schwede, T. Modeling protein quaternary structure of homo- and hetero-oligomers beyond binary interactions by homology. Scientific Reports 7 (2017).

[14]: Ko J, Park H, Heo L, Seok C. GalaxyWEB server for protein structure prediction and refinement. Nucleic acids research. 2012;40:W294–W297. doi: 10.1093/nar/gks493.

[15]: Studer, G., Biasini, M. and Schwede, T. (2014). "Assessing the local structural quality of transmembrane protein models using statistical potentials(QMEANBrane)." Bioinformatics, 30(17):i505-11.

[16]: Wiederstein & Sippl (2007); ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. Nucleic Acids Research 35, W407-W410.

[17]: Lars A. Bratholm, Jan H. Jensen. Protein structure refinement using a quantum mechanics-based chemical shielding predictor. Chemical science, Volume 8, Issue 3, 2017. doi: 10.1039/C6SC04344E

[18]: Yufeng L., Meng K., Haipeng G. Protonation of Glu135 Facilitates the Outward-to-Inward Structural Transition of Fucose Transporter. Cell Press, BioPhysical Journal, Volume 109, Issue 3, 4 August 2015, Pages 542-551. doi: 10.1016/j.bpj.2015.06.037

[PDB 1]: Bjørn P. Pedersen, Hemant Kumar, Andrew B. Waight, Aaron J. Risenmay, Zygy Roe-Zurz, Bryant H. Chau, Avner Schlessinger, Massimiliano Bonomi, William Harries, Andrej Sali, Atul K. Johri, Robert M. Stroud. Crystal structure of a eukaryotic phosphate transporter. Nature. 2013 Apr 25; 496(7446): 533–536. Published online 2013 Mar 31. doi: 10.1038/nature12042

[PDB 2]: Cristina V. Iancu, Jamillah Zamoon, Sang Bum Woo, Alexander Aleshin, Jun-yong Choe. Crystal structure of a glucose/H+ symporter and its mechanism of action. Proc Natl Acad Sci U S A. 2013 Oct 29; 110(44): 17862–17867. Published online 2013 Oct 14. doi: 10.1073/pnas.1311485110

[PDB 3]: Shintaro Doki, Hideaki E. Kato, Nicolae Solcan, Masayo Iwaki, Michio Koyama, Motoyuki Hattori, Norihiko Iwase, Tomoya Tsukazaki, Yuji Sugita, Hideki Kandori, Simon Newstead, Ryuichiro Ishitani, Osamu Nureki. Structural basis for dynamic mechanism of proton-coupled symport by the peptide transporter POT. Proc Natl Acad Sci U S A. 2013 Jul 9; 110(28): 11343–11348. Published online 2013 Jun 24. doi: 10.1073/pnas.1301079110

[PDB 4]: Linfeng S., Xin Z., Chuangye Y., Xiuyun S., Xinqi G., Yu R. & Nieng Y. Crystal structure of a bacterial homologue of glucose transporters GLUT1–4. Nature volume 490, pages 361–366 (18 October 2012). doi: 10.1038/nature11524

[PDB 5]: Dong D., Pengcheng S., Chuangye Y., Meng K., Xin J., Lei X., Wenlin R., Kunio H., Masaki Y., Shilong F. & Nieng Y. Molecular basis of ligand recognition and transport by glucose transporters. Nature volume 526, pages 391–396 (15 October 2015). doi: 10.1038/nature14655

[PDB 6]: Reiya Taniguchi, Hideaki E. Kato, Josep Font, Chandrika N. Deshpande, Miki Wada, Koichi Ito, Ryuichiro Ishitani, Mika Jormakka, Osamu Nureki. Outward- and inward-facing structures of a putative bacterial transition-metal transporter with homology to ferroportin. Nat Commun. 2015; 6: 8545. Published online 2015 Oct 13. doi: 10.1038/ncomms9545