A voltage dependent anion efflux transporter hNPT4 protein modelling: from a sequence to a structure

GRAVEY François

Master 2 Biologie informatique

Année Universitaire 20017 - 2018

1. Introduction:

In the last decade, with the reduction of DNA and RNA sequencing cost, the biological information has significantly increased. By now, the genomic organisation is better understood. However, if it is possible to learn from the DNA by it nucleic acids sequence, the sequence of amino-acids of a protein is less informative. In order to correctly characterise a protein, its structure is needed.

Nowadays, three major techniques allow to solve protein structure which are x-ray crystallography, nuclear magnetic resonance and cryo-mycroscopy Unfortunately, these methods are fastidious, time consuming and not easy to use. This means an enormous gap between protein sequences and structure resolved. At the end of 2017, 135201 proteins structure have been solved whereas the number of proteins in the Uniprot database is more than 80 millions.

One way to learn about protein structure without knowing the resolved one, is to apply the existing knowledge of protein structure on a protein of interest.

The *in-silico* method is called protein modelling. The technique is now widely employed and can be divided into two different processes. First the homology protein modelling is based on the sequences and the structure similarities between the protein of interest and some proteins for which the structure has been solved. During the modelling process, known structures are used as a model to define the structure of the study protein. The major limitation of this technique, is the necessity among related proteins to have a structure solved for some of them. Moreover, the distance between the proteins sequences have an important impact on the accuracy of the modelling. In fact, more the distance is important, less the structural model will be precise.

The second modelling method is employed when no relation could be established between the sequence of the study protein and the sequence of proteins for which the structure has been resolved. This modelling method is called threading. In this approach, the sequence of the protein will be decomposed in several fragments. For each part of the protein, a modelling template will be look for in databases. When all the pieces of the protein have a 3D conformation, the process

consist of assembly all the structures part by part. Nevertheless, modelling by threading is less accurate than homology modelling.

In this study, we proposed to predict the tertiary structure of an orphan transporter, belonging to the Major Facilitator Superfamily (MFS), the human sodium phosphate transporter 4 (hNPT4).

2. Materials and methods:

2.1 Protein structure and functional analysis:

First, we regrouped all the information based on the literature, the Uniprot data base (uniprot.org) and the Interpro web site(1). Having done these researches, provided a general point of view of hNPT4.

We used PSIPRED v3.3(2) to predict the secondary structure of hNPT4. Then we employed several algorithms to determine the transmembrane (TM) domains: Octopus(3), Spoctopus(4) and Phobius(5). Finally, we studied the domains of hNPT4 with Dompred(6) and determined disorder regions by Disopred v3(7).

2.2 Amino-acids sequence study:

With the information about secondary structure and the TM domains, we looked for related protein sequences in order to study the amino-acid which are conserved in all the linked proteins. We used the National Center for Biotechnology Information (NCBI) resources, in particular, the Basic Local Alignment Search Tool (BLAST): protein-protein BLAST (Blastp). The research was made using the database: UniprotKB/Swiss-Prot. Indeed, the proteins in this database are published after manual verification of the amino-acid sequence, allowing more accurate information.

2.3 Templates researches:

Protein modelling by similarities required protein with kwon structures which are related to the study protein in order to used them as a model. These proteins structures are called

templates. We looked for templates using NCBI resources, Blastp and Position-Specific Iterated BLAST (PSI-Blast). The research was made using the database: Protein Data Bank proteins with default parameters: expected threshold 10, word size 6 and 3 for blastp and psi-blast respectively.

We also used a program based on Hidden Markov Model (HMM), HHPred(8). The selected database was PDB_mmCIF70_28_Oct. The Multi-Sequences alignment was performed by HHblits(9) using different parameters. We browsed three different databases: uniprot20_2016_02 database, Uniclust30_2017_07 and nre70. All the researches were performed with a E-value threshold at 1e-3 and a minimum coverage of MSA hits to 10%.

2.4 Protein modelling:

We employed two different approaches in our study to propose hNPT4 models. First, we used homology modelling algorithms in several web-servers. Two of them were generalist methods, Swiss-model (SM)(10,11) and HHPred(8) whereas Membrane protein modeller (Memoir)(12) is a tool dedicated to the transmembrane protein modelling.

Using the SM and Memoir algorithms, we performed a single template homology modelling. After alignment between hNPT4 sequence and the template, these web servers proposed a protein model. We simulated five proteins using the top templates available with both algorithms.

Then, we did a multi-template modelling with HHPred. For this approach, we selected the top five templates available based on similarity and coverage. Then we performed a multi-sequence alignment using T-coffe aligner(13) (Expresso algorithm), converted the fatsa files to PIR files using formatseq, a format file converter included in the MPI Bioinformatics Tool Kit(8). Then we used Modeller(14) to propose protein models. As we browsed three different databases, we performed three models.

We also employed threading modelling algorithms: Iterative Threading ASSEmbly Refinement (I-TASSER)(15), Phyre^2(16), Muster(17) and Lomets(18). Lomets summarised

predictions from nine threading modelling servers including Muster. All the threading servers were used at default parameters.

2.5 Models evaluation:

In order to evaluate our models, we used different approaches. First, we estimated the overall quality of our models by defining the Z-Score with the ProSA-Web server(19). We also performed on every model a Ramachandra analysis to detect any anormal steric conformation of the model using the dihedral angles of the protein backbone. The Ramachandra plots were performed using the Rampage Server.

Then we used the Errat2 tool(20) on the Structure Analysis and Verification Server (SAVE web server) from UCLA. This approach is based on non-boned interactions in the protein. Indeed, these interactions are not randomly distributed through a native protein structure. The Errat2 scoring function analyse the non-bonded interactions in our proposed protein, testing if this particular distribution of interactions is conserved or not.

Finally, the top thirty remaining models according to the Z-score where tested using the specific ORientation and Evaluation of Membrane PROtein server (OREMPRO server). This algorithm inspects and score the protein insertion in the membrane using MAIDEN and ANVIL(21).

2.6 Structure refinement:

In order to improve our best model, we used the GalaxyWeb server(22). This tool allowed us to reduce the steric problems and increased the dihedral angles accuracy. Moreover, it was to possible to refined some specific regions such as the parts of the protein which are not included in the membrane, which are called loops(23).

2.7 Normal mode calculation and evaluation:

We performed a Normal mode calculation using the Nomad-Ref server. This calculation was performed using these following parameters: The process was performed in “all-atoms” model. Sixteen modes were calculated, the distance weight for elastic constant was 5 Å, the cut-

off for the Elastic Network Model (ENM) was established to 10 Å. For the output trajectories, the average Root Mean Square Deviation was 1 Å. Finally, the algorithm used the LAPACK diagonalizer. The results were visualized using the PyMol software.

3. Results:

3.1 Protein general presentation:

hNPT4 is a voltage dependent anion efflux transporter(24,25), only present in two different human organs, the liver and the kidney(24). This protein, based on the renal tubular cell, is involved in the secretion of urate and drugs from the blood to the urine. His function was suggested by the study of hyperuricemia cases, which reveal two missense mutations in *SLC17A3* gene based on 6p22.2(24). Two isoforms of these protein are found, hNPT4_L (isoform 1) and hNPT4_S (isoform 2). These isoforms are a result of different alternative splicing, exon 4 is present in hNPT4_L sequence and not in hNPT4_S. Moreover, *in vitro* studies revealed that, only the isoform 1 is found on the cell membrane whereas the isoform 2 is based on the cytoplasm. These results suggest that, amino-acid coded by the exon 4 are essentials to the adhesion of the protein to the cell membrane(24).

In addition, a reduce membrane incorporation was found during *in vitro* studies of hNPT4 from patients harbouring mutations in *SLC13A3* at nucleotide 1021, a phenylalanine substitution by a serine in the predicted sixth transmembrane domain(24).

3.2 Secondary and domains prediction:

According to Psipred, the secondary structure of hNPT4 protein is composed by coils and fourteen helices. Moreover, the Disopred analysis highlighted that, the N-Terminal region of hNPT4 as a disordered protein binding site.

The prediction of TM regions was quite confusing. Indeed, predictions which were proposed were divergent according to which method was employed. Table I summaries all the predictions obtained.

Among these nine different algorithms, six predicted nine TM segments (Topcons, Octopus, Scampi, Spoctopus, Mensat-SVM and Mensat-Cartoon), two predicted eight TM segments (Philius and TMHMM) and polyphobius predicted ten TM regions (Table I).

Moreover, among methods which predicted the same number of TM segments, the length of the helices and the starting amino-acid were divergent. As an example, according to the "nine TM segments" algorithms, the length of the helices was constant (20 amino-acids) for Topcons, Octopus, Scampi, Spoctopus and inconstant for Memsat-svm and Memsat cartoon (Table I).

Furthermore, in the MFS, the N- and C-terminal regions are known to be both in the cytoplasmic part of the cell. Any impair number of TM regions seems not possible to use in order to modelling hNPT4(26–28). This observation was support by the information contained in Uniprot and Interpro whereby hNPT4 should harbouring eight TM segments.

2.3 Templates research:

Many methods have been used in order to find the best template. The first ones were, Blastp and Psi-Blast. Unfortunately, no answers were obtained using Blastp and only five using Psi-Blast, all of them were out of the threshold and harbouring very low similarities and coverage (from 26% to 36% and from 10% to 39% respectively). As a consequence, none sequence found by Blast have been used for our study.

We next used, different methods using alternative algorithms to look for possible templates such as HHBlits. This method was used by several web servers such as Swiss-model and MPI Bio-informatics Tool kit. With this new approach, multiple sequences usable as template were found. Nevertheless, the similitude and the coverage between the sequence of hNPT4 and the possible templates were also very low.

Table I: Transmembrane regions according to several prediction algorithms:

Algorithms	TM segment 1	TM segment 2	TM segment 3	TM segment 4	TM segment 5	TM segment 6	TM segment 7	TM segment 8	TM segment 9	TM segment 10
TOPCONS	34-54	125-145	151-171	218-238	257-277	290-310	313-333	352-372	383-403	
OCTOPUS	34-54	126-146	150-170	218-238	257-277	290-310	313-333	350-370	384-404	
Philius		123-145	153-173	215-238	256-279	291-310	317-340	352-373	384-403	
Polyphobius	38-64	97-111	123-145	151-174	215-238	258-279	291-311	317-340	352-373	381-402
SCAMPI	33-53	125-145	156-176	215-235	259-279	291-311	313-333	353-373	383-403	
SPOCTOPUS	34-54	126-146	150-170	218-238	257-277	290-310	313-333	350-370	384-404	
TMHMM		124-146	151-173	221-243	258-280	292-314	319-341	353-375	385-404	
MEMSAT-SVM	35-63	125-149	155-173	211-240	257-279	289-309	316-335	351-370	385-403	
MEMSAT Cartoon	38-63	124-148	152-176	213-236	255-279	289-313	316-338	350-374	385-404	

Abbreviations: **TM** Transmembrane

The situation leads us to employed alternative protein modelling methods which were modelling by threading. We resort to I-tasser, Phyre2, Muster and Lomets.

The sequences which were used as a template by both strategies (homology and threading modelling) are summaries in Table II. All templates which has been selected belonged to Alpha Helical group, a MFS sub-group. All the structures were solved by X-ray diffraction at different resolutions, from 1,5 to 4,2 Angstrom (Å) (Table II). Once aligned to the hNPT4 sequence, the identity percentage of the templates was 8% to 27% with a percentage of coverage from 77% to 94% (Table II). Furthermore, when the coverage was correct, the identity was very low or the opposite. In addition, the resolution of certain structures was quite high, reducing the accuracy of our models.

Beside the quality problems, one template was find by several methods, a Glycerol-3-Phosphate Transporter, solved with the resolution of 3,3 Å, harboring 13% identity and a coverage of 94% (Table II).

Due to theses templates issues, we processed to a multi-sequences alignment in order to detect where homology models were lacking in our protein. One region was highlighted, the N-terminal extremity of hNPT4. In order to solve this problem, we first performed a HHBlits using a sequence harboring only the N-terminal amino-acids to find related proteins and then look for templates. Unfortunately, if related sequences were found, no new template sequences were recovered. Indeed, with a sixty amino-acid sequence, nineteen templates have been found but all of them were out of the E-value threshold and harbouring 10% of similarity and a coverage 80% for the “best of them”. We did not include these template in our modelling process.

Table II: Summary of the most frequent templates used by homology and threading methods

Name	Protein family	Group	Function[a]	Chain	Method	Resolution (Å)	Identity[b] (%)	Coverage[b] (%)	Methods
1PW4	MFS	A-H	Glycerol-3-Phosphate Transporter	A	X-Ray	3,3	13	94	Several[c]
4ZOW	MFS	A-H	Multidrug Transporter MdfA	A	X-Ray	2,45	11	85	Muster
3WDO	MFS	A-H	YajR Transporter	A	X-Ray	3,15	12	90	Muster
4J05	MFS	A-H	Phosphate Transporter	A	X-Ray	2,9	11	88	Muster
5A2O	MFS	A-H	Nitrate Transporter	B	X-Ray	3,71	11	*NA*	Phyre2
4ZP0	MFS	A-H	Multidrug Transporter MdfA	A	X-Ray	2	13	*NA*	Phyre2
4LDS	MFS	A-H	Glucose Transporter	B	X-Ray	3,2	13	*NA*	Phyre2
4IKY	MFS	A-H	Peptide Transporter POT	A	X-Ray	2,1	11	*NA*	Phyre2
4ZW9	MFS	A-H	Glucose Transporter	A	X-Ray	1,5	8	85	HHPread
5AYO	MFS	A-H	Iron Transporter Ferroportin	A	X-Ray	3,3	11	87	HHPread
5AYN	MFS	A-H	Iron Transporter Ferroportin	A	X-Ray	2,2	11	87	HHPread
4XNJ	MFS	A-H	PepTst2	A	X-Ray	2,3	11	87	HHPread
4IKW	MFS	A-H	Peptide Transporter POT	A	X-Ray	2	12	93	HHPread
2CFQ	MFS	A-H	Lactose Permase	A	X-Ray	2,95	14	81	HHPread
4IKV	MFS	A-H	Peptide Transporter POT	A	X-Ray	1,9	11	85	HHPread
4ZIR	MFS	A-H	Lactose Permase	A	X-Ray	3,31	12	77	Swiss-Model
4QIQ	MFS	A-H	D-Xylose Proton Symporter	A	X-Ray	3,51	12	85	Swiss-Model
4JA4	MFS	A-H	Xylose Transporter	A	X-Ray	4,2	27	85	Swiss-Model
4GBY	MFS	A-H	Xylose Transporter	A	X-Ray	2,81	26	84	Swiss-Model

Abbreviations: **MFS** Major Facilitator Supra-Family, **A-H** Alpha-Helicase, Å Angstrom, NA Not Available

Notes:

[a] According to the pdb database.

[b] Identity and coverage between the hNPT4 and the template sequences in a modelling or a threading usage.

[c] The exhaustive list of modelling methods which used the 1PW4 template is: I-Tasser, HHPread, Memoir, Muster, Phyre2, Swiss-Model, Sparks-X, SP3, FFAS-3D, Prospect2, HHsearch, FFAS03, HHsearch2, Neff-PPAS, pGEN Threader.

3.3 Protein modelling:

Despite the low similarity between hNPT4 and sequences available, we tried to produce models with the homology approach. First of all, using Swiss-model and Memoir, we produced "simple" homology models. Then we used the association of HHPred and Modeller to propose multi-template models. Three different models were performed by using the templates obtained browsing three different databases (Materials and methods). Finally, we obtained different results from "all in one" threading methods which were I-Tasser, Phyre2, Lomets and Muster.

More than 50 different models were produced. Then, we used several evaluation methods to sort our results.

3.4 Model Evaluations:

First, all the models were sorted using the Z-score. Indeed, this method allow to compare our models with native proteins structures. Then, we performed all the evaluation only on the top thirty models according to the Z-score. Results are summaries in table III.

Regardless of a specific model, "the performances" were low and inhomogeneous. Indeed, according to a specific test, there were good results but no model had good scores among all of tests.

Using OREMPRO server was very informative for several reasons. First of all, this algorithm predict how many TM segments are in the modelling protein. This functionality allowed us two different things. The first one was to exclude proteins for which the membrane insertion was unrealistic. The Phyre[2] models 3, 11, 15 and the Muster model 4, had very low TM domains predicted, (2, 4, 5 and 3 respectively). These "impossible" results for a MFS protein were the consequence of very low-quality prediction leading to an impossibility for the algorithm to properly simulate the protein model into a cell membrane. Secondly, as mentioned above, a protein which belong to the MSF family has to harbor a pair number of TM. As a consequence, all models for which the number of TM domains were impair were not considered as appropriate model.

Table III: Evaluation of the top thirty models according to the Z-Score evaluation:

Names	Z-Score[a]	ERRAT2[b]	Ramachandran server[c]			ProQ[d]		OREMPRO[e]		
			Residues in favoured region (%)	Residues in allowed region (%)	Residues in outlier region (%)	LG score	Max sub	MAIDEN	TM	Hydrophbic thickness (Å)
Swiss_model_1	-2,86	77	89,7	5,7	4,6	3,476	0,260	-14,2	9	22,0
I-tasser_3	-2,64	92	77,3	16,3	6,4	3,613	0,336	-27,0	10	34,0
Loomets_8MPI	-2,37	60	93,5	4,8	1,7	4,196	0,338	23,0	8	35,0
Loomets_7	-1,97	60	90,7	5,7	3,6	4,090	0,334	2,0	9	34,0
Phyre2_11	-1,80	53	90,8	6,4	2,8	2,878	0,248	-16,4	2	30,0
Loomets_9	-1,80	53	91,9	6,0	2,2	4,258	0,382	26,4	9	30,0
Loomets_5	-1,68	34	81,8	12,9	5,3	3,853	0,219	-15,0	9	29,0
Loomets_6	-1,67	53	93,1	5,3	1,7	4,359	0,271	27,0	11	33,0
Muster_3	-1,63	63	93,5	4,1	2,4	3,954	0,137	57,0	12	20,0
Memoir_hicov[f]	-1,59	88	89,1	7,5	3,4	2,681	0,212	19,0	10	36,0
Phyre2_3	-1,59	54	95,1	2,9	2,0	2,774	0,225	29,0	4	32,0
I-tasser_2	-1,57	81	80,6	13,4	6,0	3,660	0,237	-5,6	10	31,0
HHPred_1	-1,56	55	89,4	7,0	3,6	1,872	0,143	-0,3	10	27,3
I-tasser_1	-1,55	85	79,4	15,1	5,5	3,580	0,237	32,0	11	22,8
Loomets_1	-1,54	54	91,1	5,7	3,1	4,073	0,308	54,0	11	27,0
Muster_2	-1,52	58	95,7	2,9	1,4	4,162	0,188	1,4	11	27,6
Muster_1	-1,48	54	90,4	7,2	2,4	4,070	0,197	138,0	12	27,0
Loomets_3	-1,45	56	92,1	6,7	1,2	4,169	0,289	35,0	9	30,0

Loomets_4	-1,44	53	92,1	5,7	2,2	4,187	0,476	70,0	11	24,6
Swiss_model_3	-1,38	82	84,2	8,6	7,2	1,893	0,148	-15,0	8	24,9
Muster_10	-1,22	62	94,3	3,8	1,9	4,063	0,126	48,0	11	34,0
Muster_8	-1,07	67	92,6	6,4	1,0	3,915	0,149	106,0	11	34,5
Phyre2_1	-1,02	55	83,7	12,2	4,1	2,533	0,179	63,0	10	33,0
Loomets_10	-0,98	31	83,7	10,3	6,2	3,914	0,356	46,0	11	29,0
Modeller_78	-0,91	70	92,0	6,8	0,8	1,969	0,145	-22,6	11	36,0
Muster_4	-0,73	60	90,7	6,7	2,6	4,014	0,322	-18,0	3	29,4
Phyre2_15	-0,62	56	84,9	10,9	4,2	2,184	0,135	17,0	5	25,5
HHPred_2	-0,61	57	89,9	5,5	4,6	1,564	0,110	-57,8	8	37,5
Phyre2_18	-0,56	47	84,2	9,8	6,0	2,654	0,184	21,0	3	26,4

Notes:

[a] Z-score according to the ProSa web server.

[b] Overall quality factor predicted by the ERRAT2 algorithm. Analyses were performed on the Structure Analysis and Verification Server (SAVE web server) from UCLA.

[c] Ramachandran plots were obtained using the Rampage Server.

[d] ProQ evaluation was performed by the ProQ web server form the Stockholm Bioinformatics Center

[e] OREMPRO evaluation was made using the ORientation and Evaluation of Membrane PROtein server (OREMPRO server)

f Protein model performed with Memoir algorithm using the high coverage option.

Only ten models had pair number of TM segments, Muster models 1 and 3, I-tasser models 2 and 3, the “high coverage” Memoir model, the HHPred models 1 and 2, the SM model number 3 and the Loomets model 8 which was performed by HHsearch2.

We did not consider models with twelve TM as adequate models. Indeed, none prediction of twelve TM domains was find in our pre-modelling protein study.

Among the “ten or eight TM segments models”, only four had a negative MAIDEN score. HHPred model two (HHPred_2) had the lowest MAIDEN score and the lower outlier number residues in the Ramachandran plot. Nevertheless, it had the worse ERRAT2, LG and Max sub scores. Because our main goal of the study is to propose a transmembrane protein, we considered the HHPred_2 as our best model. Indeed, as already mentioned it had the lowest MAIDEN score and also, it harbored eight TM segments. The number of TM segments corroborated the predictions performed by Philus and TMHMM, and the information found in the Uniprot database and Interpro Website. Moreover, the evaluation of HHPred_2 was performed before the refinement. After this step, the different score obtained by HHPred_2 were: Z-score -1,2, ERRAT2 89, 93% of the residues were in a favored region and 1,8% of the residues were considered as outliers. Moreover, his MAIDEN score improved to -60. Figure 1(A - C) represented the HHPred_2 model obtained after refinement.

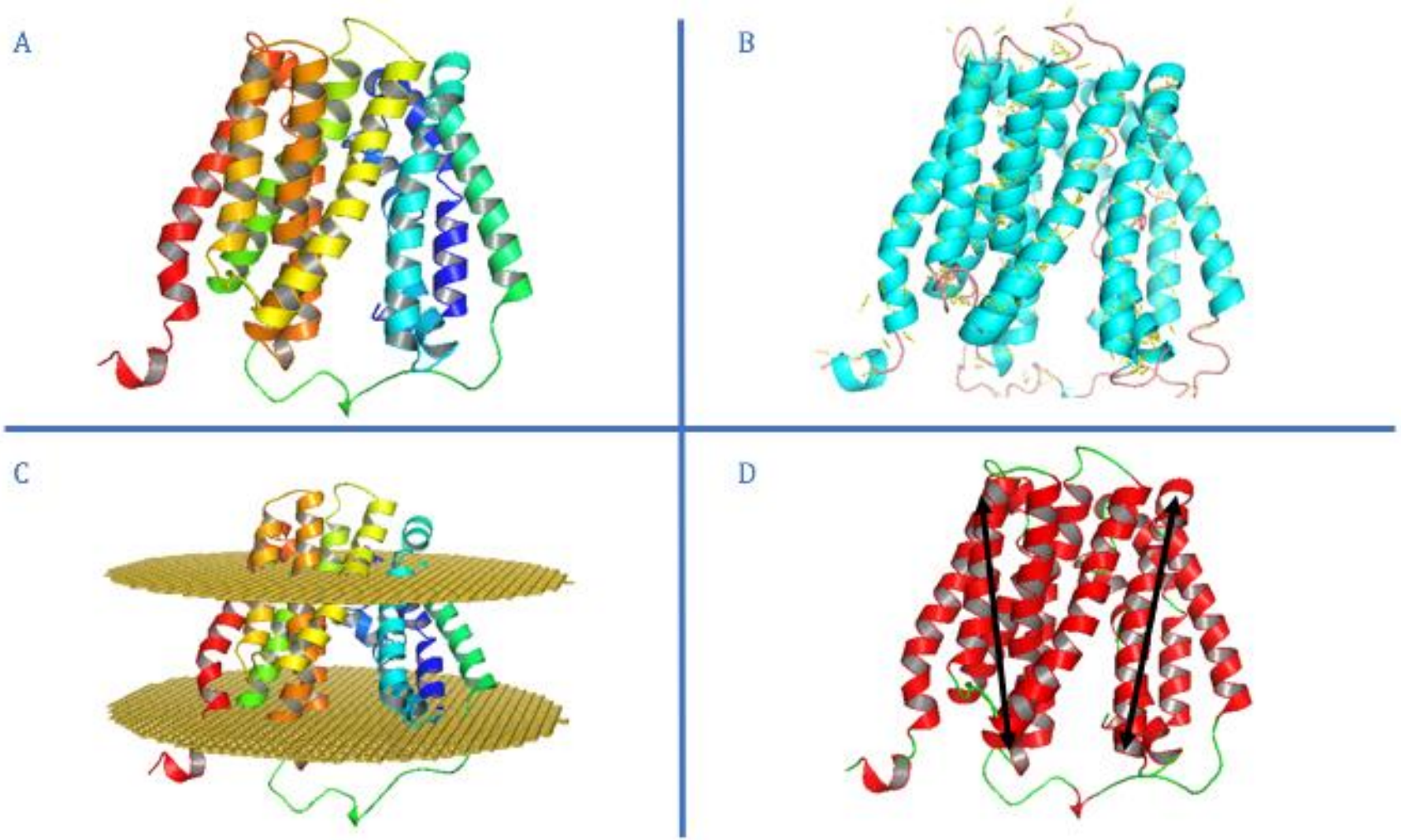

Figure 1: Best model of hNPT4 HHPred_2 represented by PylMol software
A - General organisation of the protein
B – Hydrogen bonds through the alpha-helixes
C – Insertion of hNPT4 in a cell membrane
D- Asymmetric movements observed from the normal modes analysis. During the simulations, we observed two sub-unities with antero-posterior displacements. Arrows indicates the axes of movements of the protein.

3.5 Normal mode analysis:

Using the Pymol software, we analysed all the 3D movements of our models. Table IV summaries all the movements found during the normal mode analyses.

Table IV: Normal mode analysis:

Names	Type of movement
Mode_1	Lateral movements
Mode_2	Antero-posterior movements
Mode_3	Antero-posterior movements
Mode_4	Lateral movements
Mode_5	Up and down movements
Mode_6	Lateral movements
Mode_7	Shearing movements between two parts of the protein
Mode_8	Lateral movements between two rigid parts of the protein
Mode_9	Shearing movements between two parts of the protein
Mode_10	Lateral movements between two rigid parts of the protein
Mode_11	Shearing and lateral movements of the two parts of the protein
Mode_12	Ondulation movements
Mode_13	Lateral movements
Mode_14	Lateral movements between two rigid parts of the protein
Mode_15	Lateral movements between two rigid parts of the protein
Mode_16	Up and down movements

Among all the modes obtained, Mode_7, Mode_9 and Mode_11 were very interesting. Indeed, these dynamics modes presented the protein moving as two different sub-unities (Figure 1D). Theses observation were in agreements with several data in the literature(26,29).

Unfortunately, we were unable to obtain distant modes. As a consequence, we could not simulate the two states of our protein: outward open and inward open. We 'only' were able to model an intermediate conformation between the two functional states.

4. Discussion:

This study suffered from several limitations. First of all, we were unable to rich our "ultimate" goal which was to propose a dynamic molecular model of hNPT4. Moreover, the quality and the similarity between hNPT4 sequence and the templates available were very low. As a

consequence, the accuracy of our model was poor. During our work, we did not use Rosetta. We used the Robetta server, but did not obtain answer before the end of our work.

Nevertheless, during this study, we learned how difficult it is to perform protein modelling. Despite the enormous number of methods available, we were not able to propose a correct model of hNPT4. It was very interesting to used two different approaches, the homology and the threading. As far as we are concerned, homology modelling is more interesting. Indeed, using the process, we have a lot of opportunities to decide how the algorithm will work. On the other side, threading modelling was more as a "black box".

Another very interesting (and very difficult point) was the evaluation of our models. We performed quite a lot of tests to our protein structure predictions and were always amazed by the discordances for a considered model between all the tests. The main explanation was because used tests did not evaluate the same "characteristics" of the protein. Another explanation could be, we used general evaluations. The only transmembrane protein test we used was the OREMPRO server.

5 Conclusion:

Protein modelling is a best way to learn about protein when his structure is unknown. During this study, we tried to propose the best model to a voltage dependent anion efflux transporter hNPT4. We learned how it is difficult to obtain an acceptable model. Nevertheless, this work introduced us to a gigantic scientific world.

6 – Bibliography:

1. Finn RD, Attwood TK, Babbitt PC, Bateman A, Bork P, Bridge AJ, et al. InterPro in 2017—beyond protein family and domain annotations. Nucleic Acids Res. 4 janv 2017;45(D1):D190-9.

2. Jones DT. Protein secondary structure prediction based on position-specific scoring matrices. J Mol Biol. 1999;292(2):195–202.

3. Viklund H, Elofsson A. OCTOPUS: improving topology prediction by two-track ANN-based preference scores and an extended topological grammar. Bioinformatics. 1 août 2008;24(15):1662-8.

4. Viklund H, Bernsel A, Skwark M, Elofsson A. SPOCTOPUS: a combined predictor of signal peptides and membrane protein topology. Bioinformatics. 15 déc 2008;24(24):2928-9.

5. Kall L, Krogh A, Sonnhammer ELL. Advantages of combined transmembrane topology and signal peptide prediction--the Phobius web server. Nucleic Acids Res. 8 mai 2007;35(Web Server):W429-32.

6. Marsden RL, McGuffin LJ, Jones DT. Rapid protein domain assignment from amino acid sequence using predicted secondary structure. Protein Sci. 13 avr 2009;11(12):2814-24.

7. Jones DT, Cozzetto D. DISOPRED3: precise disordered region predictions with annotated protein-binding activity. Bioinformatics. 15 mars 2015;31(6):857-63.

8. Alva V, Nam S-Z, Söding J, Lupas AN. The MPI bioinformatics Toolkit as an integrative platform for advanced protein sequence and structure analysis. Nucleic Acids Res. 8 juill 2016;44(W1):W410-5.

9. Remmert M, Biegert A, Hauser A, Söding J. HHblits: lightning-fast iterative protein sequence searching by HMM-HMM alignment. Nat Methods. 25 déc 2011;9(2):173-5.

10. Schwede T. SWISS-MODEL: an automated protein homology-modeling server. Nucleic Acids Res. 1 juill 2003;31(13):3381-5.

11. Bordoli L, Kiefer F, Arnold K, Benkert P, Battey J, Schwede T. Protein structure homology modeling using SWISS-MODEL workspace. Nat Protoc. déc 2008;4(1):1-13.

12. Ebejer J-P, Hill JR, Kelm S, Shi J, Deane CM. Memoir: template-based structure prediction for membrane proteins. Nucleic Acids Res. 1 juill 2013;41(W1):W379-83.

13. Notredame C, Higgins DG, Heringa J. T-coffee: a novel method for fast and accurate multiple sequence alignment 1 1Edited by J. Thornton. J Mol Biol. sept 2000;302(1):205-17.

14. Webb B, Sali A. Comparative Protein Structure Modeling Using MODELLER: Protein Structure Modeling Using MODELLER. In: Coligan JE, Dunn BM, Speicher DW, Wingfield PT, éditeurs. Current Protocols in Protein Science [Internet]. Hoboken, NJ, USA: John Wiley & Sons, Inc.; 2016 [cité 26 nov 2017]. p. 2.9.1-2.9.37. Disponible sur: http://doi.wiley.com/10.1002/cpps.20

15. Zhang Y. I-TASSER server for protein 3D structure prediction. BMC Bioinformatics. 2008;9(1):40.

16. Kelley LA, Mezulis S, Yates CM, Wass MN, Sternberg MJE. The Phyre2 web portal for protein modeling, prediction and analysis. Nat Protoc. 7 mai 2015;10(6):845-58.

17. Wu S, Zhang Y. MUSTER: Improving protein sequence profile-profile alignments by using multiple sources of structure information. Proteins Struct Funct Bioinforma. 4 févr 2008;72(2):547-56.

18. Wu S, Zhang Y. LOMETS: A local meta-threading-server for protein structure prediction. Nucleic Acids Res. mai 2007;35(10):3375-82.

19. Wiederstein M, Sippl MJ. ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. Nucleic Acids Res. 8 mai 2007;35(Web Server):W407-10.

20. Colovos C, Yeates TO. Verification of protein structures: patterns of nonbonded atomic interactions. Protein Sci. 1993;2(9):1511–1519.

21. Postic G, Ghouzam Y, Gelly J-C. OREMPRO web server: orientation and assessment of atomistic and coarse-grained structures of membrane proteins. Bioinformatics. 15 août 2016;32(16):2548-50.

22. Lindahl E, Azuara C, Koehl P, Delarue M. NOMAD-Ref: visualization, deformation and refinement of macromolecular structures based on all-atom normal mode analysis. Nucleic Acids Res. 1 juill 2006;34(Web Server):W52-6.

23. Ko J, Park H, Heo L, Seok C. GalaxyWEB server for protein structure prediction and refinement. Nucleic Acids Res. 1 juill 2012;40(W1):W294-7.

24. Jutabha P, Anzai N, Kitamura K, Taniguchi A, Kaneko S, Yan K, et al. Human Sodium Phosphate Transporter 4 (hNPT4/ *SLC17A3*) as a Common Renal Secretory Pathway for Drugs and Urate. J Biol Chem. 5 nov 2010;285(45):35123-32.

25. Jutabha P, Anzai N, Hayashi K, Domae M, Uchida K, Endou H, et al. A Novel Human Organic Anion Transporter NPT4 Mediates the Transport of Ochratoxin A. J Pharmacol Sci. 2011;116(4):392-6.

26. Yan N. Structural advances for the major facilitator superfamily (MFS) transporters. Trends Biochem Sci. mars 2013;38(3):151-9.

27. Augustin R, Mayoux E. Mammalian Sugar Transporters. In: Szablewski L, éditeur. Glucose Homeostasis [Internet]. InTech; 2014 [cité 26 nov 2017]. Disponible sur: http://www.intechopen.com/books/glucose-homeostasis/mammalian-sugar-transporters

28. Mueckler M, Thorens B. The SLC2 (GLUT) family of membrane transporters. Mol Aspects Med. avr 2013;34(2-3):121-38.

29. Henderson PJF, Baldwin SA. This is about the in and the out. Nat Struct Amp Mol Biol. 5 juin 2013;20:654.