Structural Bioinformatic – Q8N697

Introduction

Q8N697 is a proton oligopeptide cotransporter of the membrane, meaning it transports free histidine and certain di- and tripeptides through the membrane. It is a multi-pass membrane protein (= protein spanning the membrane more than once) belonging to the PTR2/POT transporter (TC 2.A.17) family. It goes by multiples names : “Solute carrier family 15 member 4”, “Peptide transporter 4”, “Peptide/histidine transporter 1” and “hPHT1”.

Gene SLC15A4, situated in chromosome 12, code for this protein. hPHT1 is highly expressed in skeletal muscle, moderately expressed in liver, kidney and heart, and weakly expressed in brain and colon. It is also expressed in low levels throughout the gastrointestinal tract. (1)

hPHT1's family, called peptide transport family (or PTR), are proteins composed by pairs of salt bridge interactions between transmembrane helices, working in together to alternate access transport within the PTR family. Mammalian members of this family are responsible for the uptake of many pharmaceutically important drug molecules, including antibiotics and antiviral medications. (2)
Studies made indicate that The generalized transport reaction catalyzed by the proteins of the PTR family is (3) :

$$\text{Substrate (out) } n\text{H (out)} \rightarrow \text{substrate (in) } n\text{H}^{+}\text{ (in)}$$

hPHT1's sequence has been resolved, as its 2D structure which is composed by helix α and coils alternating.

query

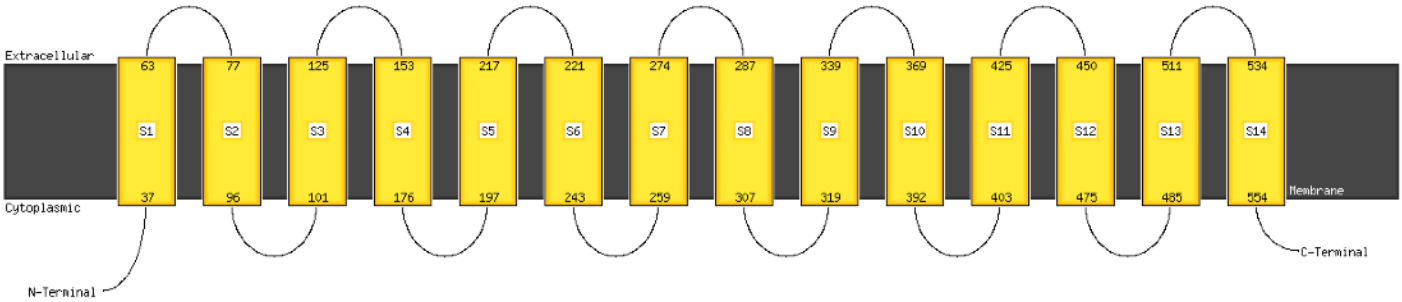


Fig 1 : Q8N697 2D structure in a membrane (yellow = helix α). Sources : Phyre2

Material & Methods

Uniprot describes 2 isoforms produced by alternative splicing. I choose isoform 1 (Q8N697-1) because it is made of 577 amino acids wherease isoform 2 contain only 240 amino acids. Moreover, isoform 1 has been chosen to be the 'canonical' sequence. (1)

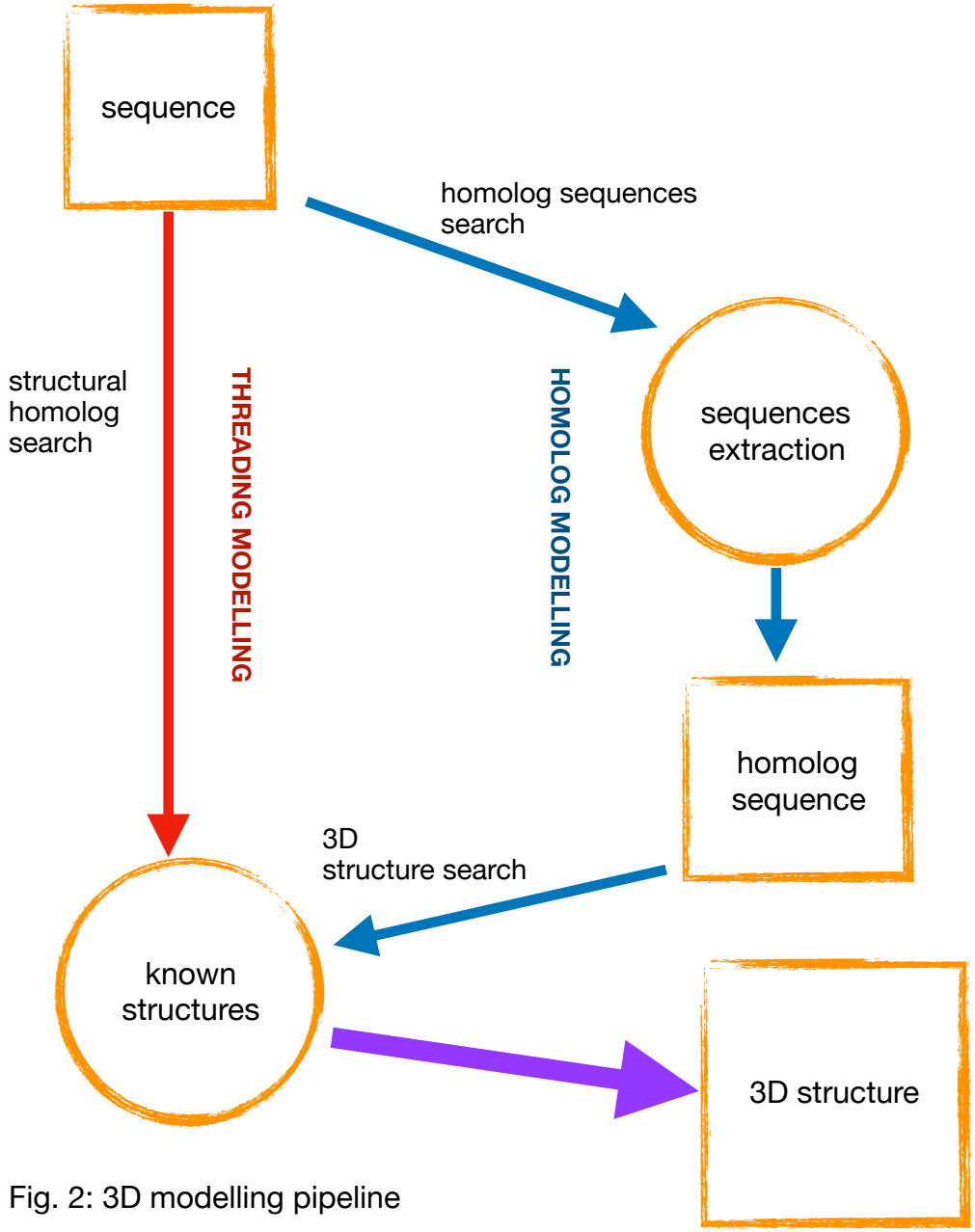


Fig. 2: 3D modelling pipeline

homolog sequences

To know if a homolog modelling could be made, a blast was made on every sequences with a 3D structure known of UniProt database with these parameters :
- an E-threshold (a statistical measure of the number of expected matches in a random database. The lower the e-value, the more likely the match is to be significant) of 10 because I wanted an overview of all matches and an alignement with an E-threshold higher than 10 is very unlikely to have a biological significance,
- a matrix (which assigns a probability score for each position in an alignment) set automatically depending on the query sequence length,

- no filtering,
- gaps allowed (to be introduced in the sequences when the comparison is done).

family domains

An alignment of all 124 reviewed sequences of PTR family was made to see if domain(s) were preserved, and may help to make a 3D model of hPHT1. Seeing the affiliated tree, only the 8 nearest sequences were used to test it (Q8IY34, Q8BPX9, Q924V4, Q91W98, O09014, A6QQL0, Q8N697, and Q68F72). A logo was created at the issue of the alignment, using WebLogo (5).
The alignment was made with default parameters : Gonnet default transition matrix, gap opening penalty = 6 bits, gap extension = 1 bit. HHalign algorithm was used by Clustal-Omega and its default settings were its core alignment engine. (6)

2D structure model

Two 2D structure models were made, using jpred4 (7), Quick2D of Hhpred (8) and Phyr2 (9), to corroborate UniProt informations and to be able to make comparisons between the protein 2D and 3D models.

JPred incorporates the Jnet algorithm. (10) In addition to protein secondary structure JPred also makes predictions on Solvent Accessibility and Coiled-coil regions. (11)

Quick2D gives an overview of secondary structure features like alpha-helices, extended beta-sheets, coiled coils, transmembrane helices and disorder regions. Its parameters are the e-value threshold and the database used. (8)

Phyre2 is a suite of tools available on the web to predict and analyze protein structure, function and mutations. It allows to predict 2D and 3D structures or any protein with threading methods. Intensive results will show a horizontal (possibly multiline) colored bar that represents the user sequence together with a color-coded confidence key.(12)

3D structure model

Protein threading is a protein modeling method used to model proteins with the same fold as proteins of known structures, but do not have homologous proteins with known structure. It actually is the case.
The prediction is made by "threading" (or aligning) each amino acid in the target sequence to a position in the template structure, and evaluating how well the target fits the template. After the best-fit template is selected, the structural model of the sequence is built based on the alignment with the chosen template. Protein threading is based on two basic observations : 90% of the new structures submitted to the PDB in the past three years have similar structural folds to ones already in the PDB and the number of different folds in nature is fairly small. (13) So If we place residues instead of those of this structure, the obtained folding remains correct from a steric and energy point of view.
The main idea is to detect the fold the most compatible with the query sequence. In practice we need to test different structural templates and rank them. This set two major drawbacks : it needs an exhaustive library of folds and an accurate tool for aligning a sequence on a 3D structure. (4)

I used multiple tools to be able to compare various models and make informed choice to choose the final model : Phyre2 (9), I-TASSER (14), LOMETS (15), and SWISS-MODEL (16).

Phyre2 made the model with 5 templates which were selected to model the protein, based on heuristics to maximise confidence, percentage identity and alignment coverage. 61 residues were modelled by ab initio.

I-TASSER is a hierarchical approach to protein structure and function prediction. It first identifies structural templates from the protein by multiple threading approach LOMETS (15), with full-length atomic models constructed by iterative template fragment assembly simulations. Function insights of the target are then derived by threading the 3D models through protein function database BioLiP (a semi-manually curated database for high-quality, biologically relevant ligand-protein binding interactions. The structure data are collected primarily from the Protein Data Bank, with biological insights mined from literature and other specific databases) (17).

LOMETS (Local Meta-Threading-Server) generates 3D models by collecting high-scoring target-to-template alignments from 9 locally-installed threading programs : FFAS-3D, HHsearch, MUSTER, pGenTHREADER, PPAS, PRC, PROSPECT2, SP3, and SPARKS-X.

SWISS-MODEL has an automatic pipeline which identifies suitable templates based on Blast (18) and HHblits (19). Template quality is then estimated from its properties (20). Priority is given to templates that maximize the expected quality of the models and the coverage of the target, and several models can be generated to reflect the structural diversity.

Models evaluation

Models evaluation were made with Verify3D (21) and ProSA (22).

Verify3D determines the compatibility of an atomic model with its own amino acid sequence by assigning a structural class based on its location and environment (alpha, beta, loop, polar, nonpolar etc) and comparing the results to good structures.

ProSA generates scores reflecting the quality of protein structures. They are derived from known protein structures, capturing the average properties of native globular proteins in terms of atom-pair, and protein-solvent interactions. The program also provides graphs which pinpoint problematic sections in a particular structure.

Results

homolog sequences

Of all alignments with a identity percentage higher 25%, none match more than 20% of the original sequence but one. Of all alignments with a identity percentage between 20 and 25%, only two matches more than 50% of the original sequence, but they have quite a lot of huge gaps.
As it is very few concrete informations, I decided to do a threading modelling.

family domains

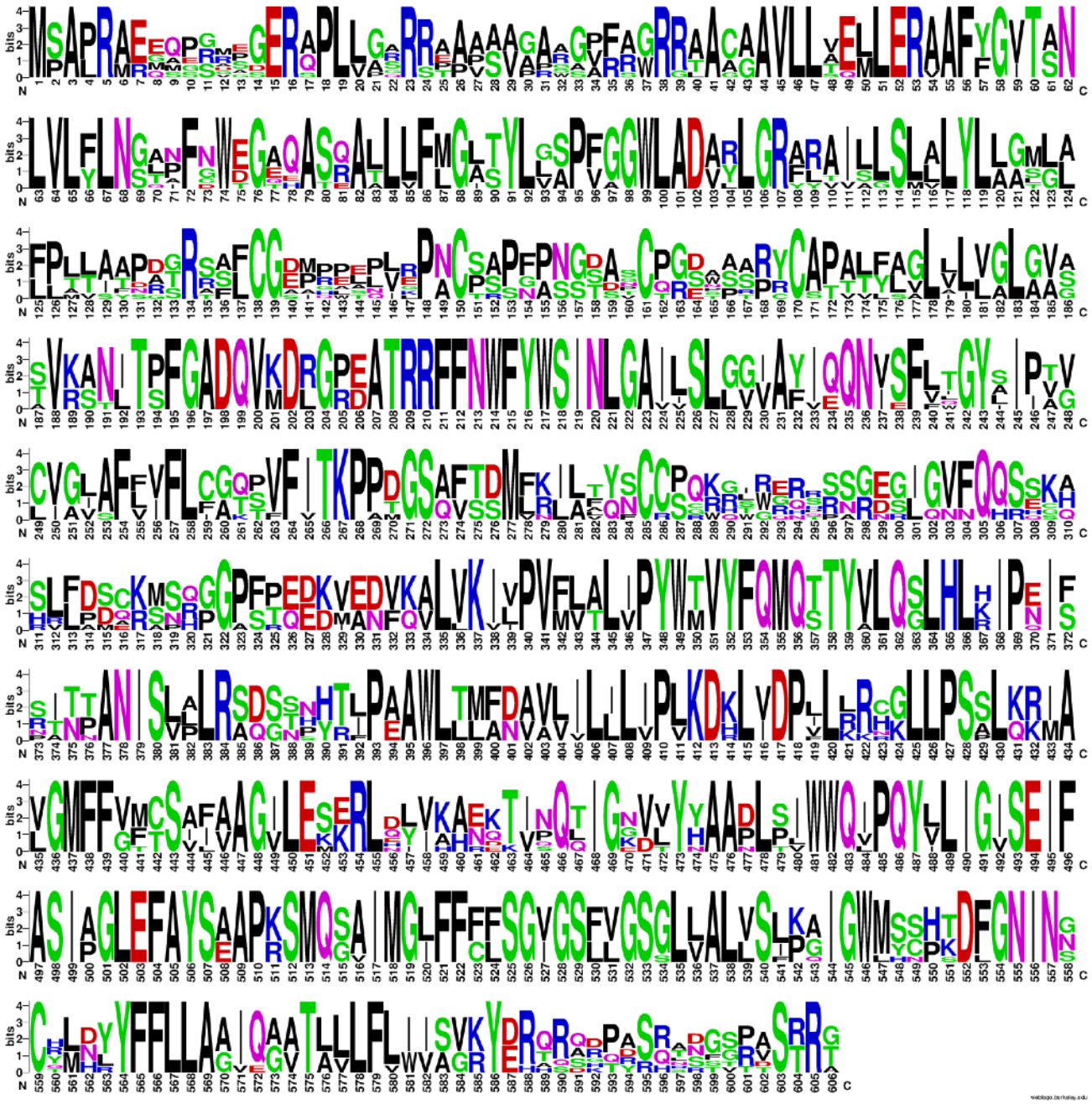


Fig 3.: logo of the 8 nearest sequences of hPHT1

We can see on fig 3 that quite some parts of the sequence are preserved. Unfortunately, none have a 3D structure known, and neither have their active domain known.

2D structure model

The results of the 3 tools (jpred4, Phyr2 and Quick2D) have given similar results and concur with Uniprot informations. hPHT1 is a transmembrane protein alternating helix alpha and coils. Helix seems to be integral part of the membrane (see fig. 1).

3D structure model

LOMETS (15), and SWISS-MODEL

Phyre2 has given the model below (Fig 4.). We can see all predicted helix, and it seems to corrobore the idea that hPHT1 is a membrane transporter.

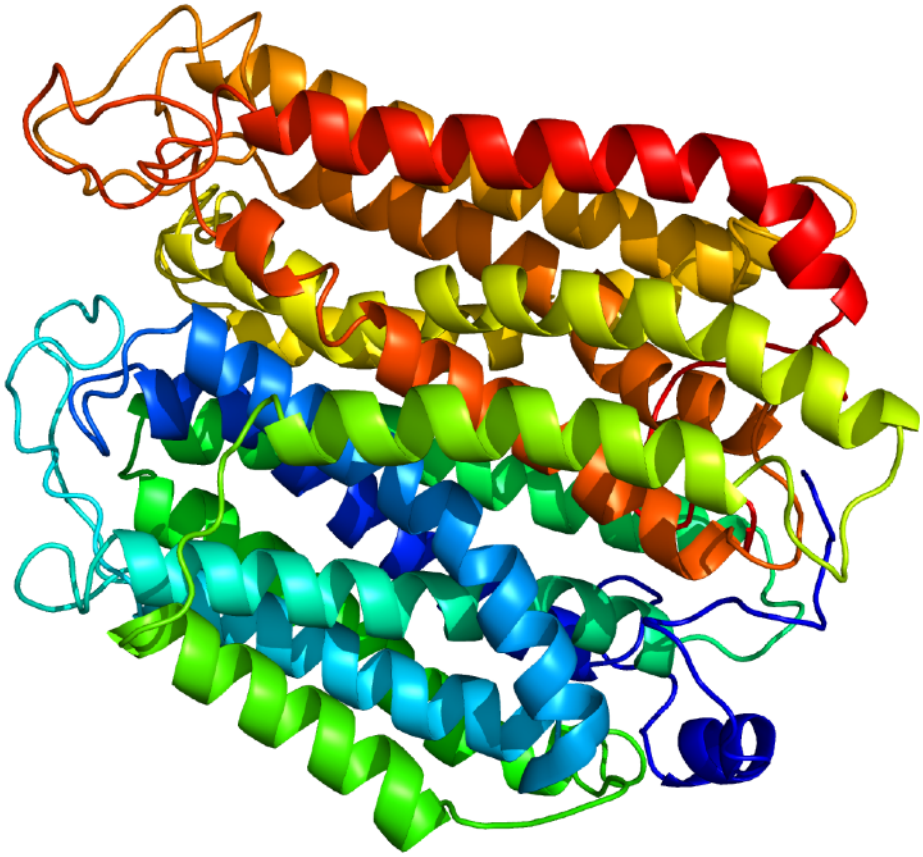


Fig 4. : 3D predicted structure of hPHT1 by Phyre2

I-TASSER has given 3 models and LOMETS 10 for each methods.

The results did not all seems bad.

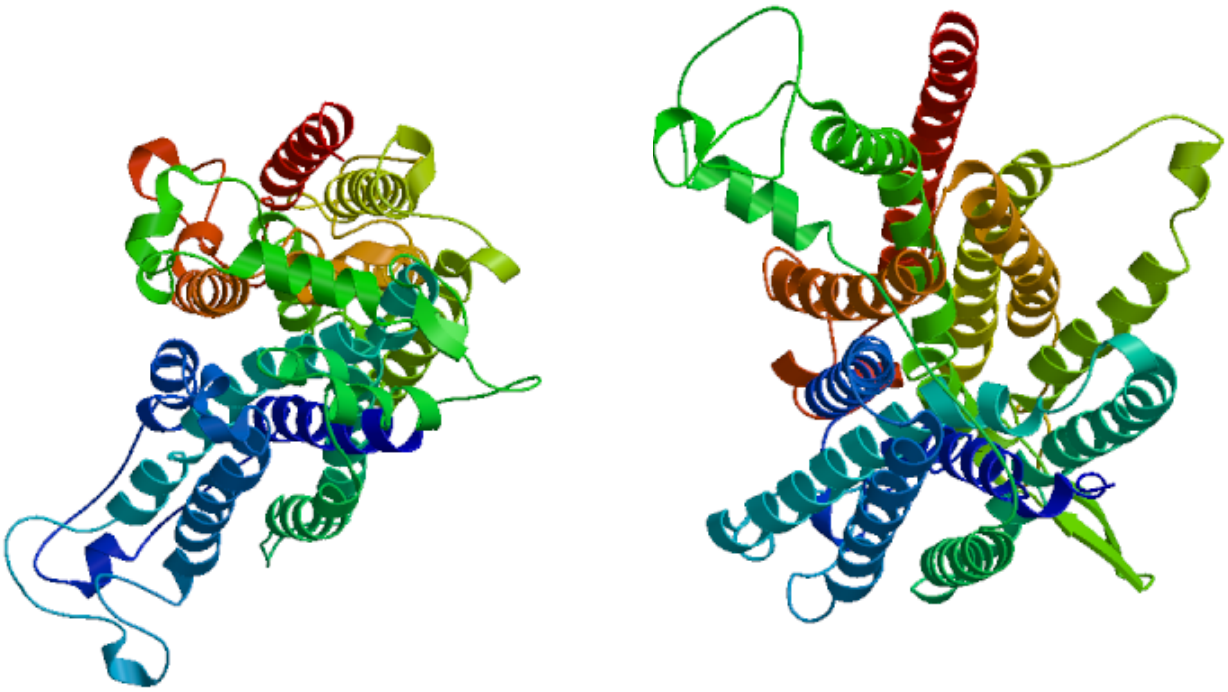


Fig 5. : 3D predicted structures of hPHT1 by SWISS-PROT

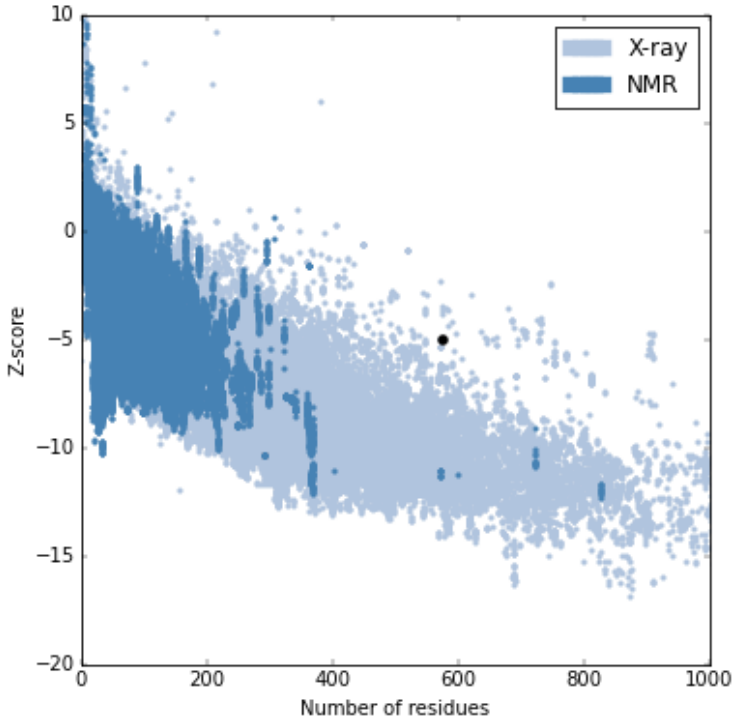
SWISS-MODEL gave two models (Fig 5.). We can see that they do not appear to be sensitively correct.

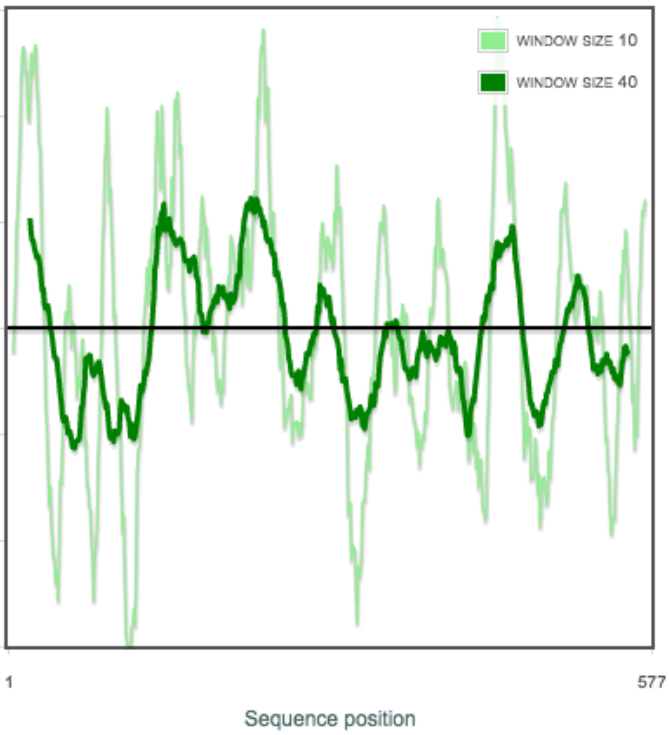
After looking to all pdb and removing the absurd ones (like the first model of SWISS-PROT), I evaluated these models.

Models evaluation

From I-TASSER, the 1st model has 73.48% of the residues had an averaged 3D-1D score >= 0.2 with Verify3D).

We can see that the model is not overall bad, is not perfect.





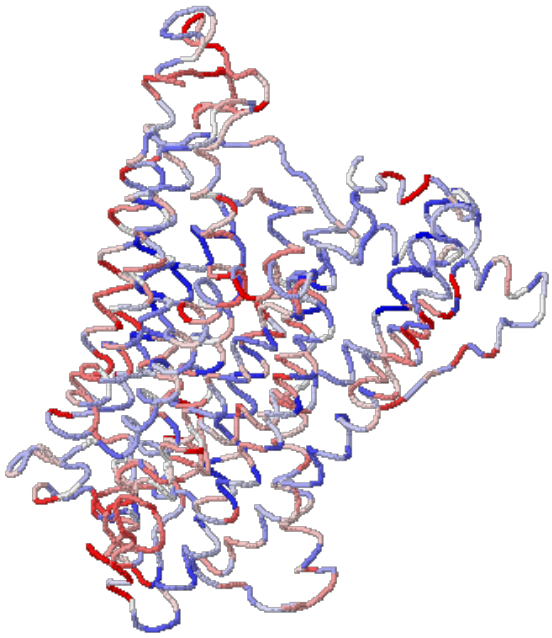


Fig. 6 : results of the first model with ProSA.

Analysing similarly all the other models, it appeared that models with LOMETS were best. They all have some problems of energy but are, in average, correct.

Conclusion

Models with different techniques were created, and some were better than others, but more work is needed to have an operational model of hPHT1.

Références

(1) http://www.uniprot.org/uniprot/Q8N697

(2) Molecular insights into proton coupled peptide transport in the PTR family of oligopeptide transporters, Newstead, Biochim Biophys Acta, 2015

(3) Determinants for Arabidopsis peptide transporter targeting to the tonoplast or plasma membrane, Traffic, Komarova *et al.*, 2012

(4) "3D Modelling: Homology, threading, & ab initio / de novo methods" by Catherine Etchebest

(5) http://weblogo.berkeley.edu

(6) "Protein homology detection by HMM–HMM comparison", Söding, J., Bioinformatics 21, 951-960., 2005

(7) http://www.compbio.dundee.ac.uk/jpred/

(8) https://toolkit.tuebingen.mpg.de/#/tools/quick2d

(9) http://www.sbg.bio.ic.ac.uk/~phyre2/html/page.cgi?id=index

(10)"Application of multiple sequence alignment profiles to improve protein secondary structure prediction", James A. Cuff et al., Proteins, 2000

(11)"Predicting coiled coils from protein sequences", Lupas A. et al., Science, 1991

(12)"The Phyre2 web portal for protein modeling, prediction and analysis", Lawrence A Kelley et al., Nature Protocols, 2015

(13)"A new approach to protein fold recognition", DT Jones et al., Nature, 1992

(14) https://zhanglab.ccmb.med.umich.edu/I-TASSER/

(15)https://zhanglab.ccmb.med.umich.edu/LOMETS/

(16)https://swissmodel.expasy.org/

(17)https://zhanglab.ccmb.med.umich.edu/BioLiP/

(18)"Gapped BLAST and PSI-BLAST: a new generation of protein database search programs", Altschul SF et al., Nucleic Acids Res, 1997

(19)"HHblits: lightning-fast iterative protein sequence searching by HMM-HMM alignment", Remmert M et al., Nat Methods, 2011

(20)"SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information", Marco Biasini, Nucleic Acids Research, 2014

(21)http://services.mbi.ucla.edu/Verify_3D/

(22)https://prosa.services.came.sbg.ac.at/prosa.php