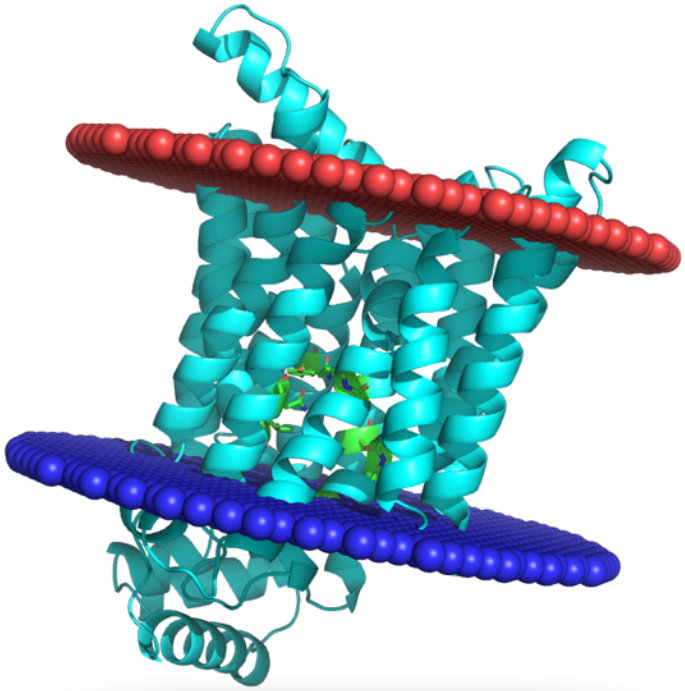

Structural Bioinformatics analysis of the facilitated glucose transporter member 9 of the Solute carrier family 2

Maxime Borry

November 2017

Structural Bioinformatics M2BI report
Paris Diderot University

Abstract

SLC2A9 is a member of the glucose transporter family. However, the structure of its protein GLUT9 has not been solved yet. Modelling by homology of GLUT9's structure based on other GLUT known structures allows to explore the mechanistic and functions of this transporter. Normal mode analysis allowed to show the GLUT9 change of conformation from inward open to outward open, as already demonstrated for other member of this GLUT family. Even though the transport of Glucose and Fructose by GLUT9 is not a settled debate, the docking of Urate on GLUT9 seems to confirm its transport by this member of the GLUT family.

Acronyms

GUI Graphical User Interface

NMA Normal Mode Analysis

MD Molecular Dynamic

MSA Multiple Sequence Alignment

PCA Principal Component Analysis

SLC SoLute Carrier

SS Secondary Structure

TM Trans Membrane

Contents

1. *Introduction*

1.1 SLC2 Glucose transporters

The Glucose Transporters are a protein family originally thought to be involved in facilitating the glucose transport across cells plasma membrane. Because glucose if very often used as an energy source in many different organisms, this family of protein is present is a very wide range of eukaryotic organisms [20]. In *Homo sapiens*, the glucose transporter family is composed of 14 members, regrouped under the SLC2 (**S**o**L**ute **C**arrier family 2) appellation. Since 1985, this family of facilitative transporters have been associated with the transport of many more substrates, including different hexose, myo-inositol, urate, glucosamine, ascorbate and probably many more yet to be discovered [21]. The 14 members of this SLC2 family share the following characteristics:

- ~ 500 amino-acid residues
- 12 transmembrane segments
- C-ter and N-ter positioned in Cytoplasm

At least one SLC2 family member should be expressed in almost every human cell, while 11 out of the 14 members can transport glucose. Mueckler et al. explain this redundancy by the critical importance of glucose as a fuel source for human cells [21].

The SLC2 family can be divided in three different classes, based on similarity of sequences. (table 1)

Class I	Class II	Class III
Glut 1	Glut 5	Glut 6
Glut 2	Glut 7	Glut 8
Glut 3	Glut 9	Glut 10
Glut 4	Glut 11	Glut 12
Glut 14		HMIT

Table 1: Classes of the SLC2 protein family

The function of transporting substrates in the GLUT family is thought to be associated with a conformational change giving a minima two distinct conformations [10] (fig 1):

- The **outward open** conformation where the 12 TM α-helix form a funnel open towards the outside of the cell.
- The **inward open** conformation where the 12 TM α-helix form a funnel open towards the inside of the cell.

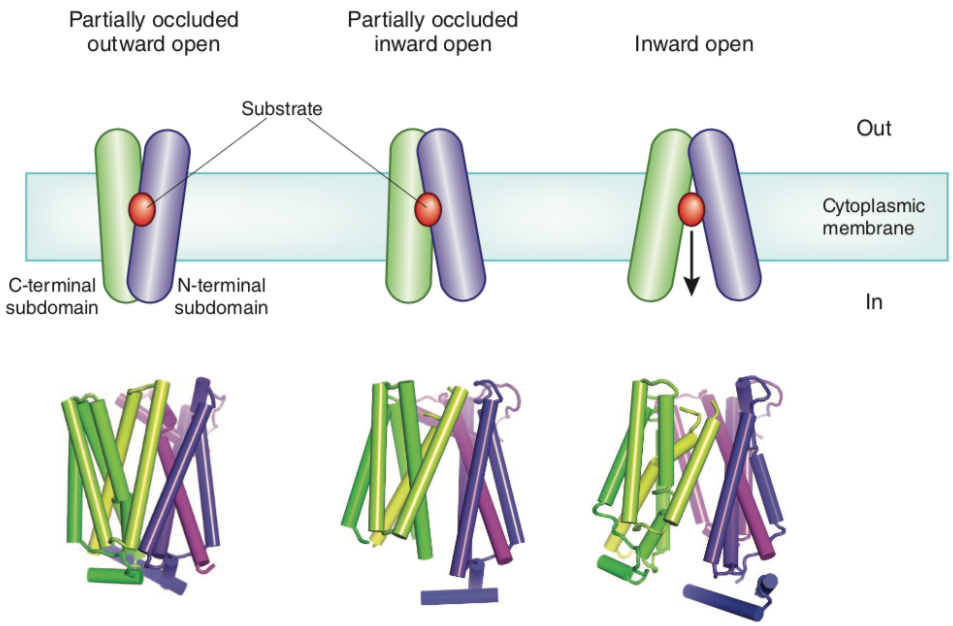

Figure 1: Conformational changes of Sugar transporters. Taken from Henderson et al. [10]

1.2 SLC2A9 - GLUT9

In this work, I chose to work on the protein `GLUT9` expressed by the `SLC2A9` gene. This protein, member of the class II of the SLC2 family is so far know to transport Glucose, Fructose, and Urate [14] (see substrates structure in fig 2). While the GLUT9 gene, SLC2A9 is located on the chromosome 4, the protein GLUT9 is found in the membrane of Kidney, liver, small intestine, placenta, lung and leukocytes. Because of alternative splicing, there are two splice variants of SLC2A9: SLC2A9a and SLC2A9b. The major difference between SLC2A9a and SLC2A9b is a shorter N-ter for SLC2A9b, otherwise the rest of the sequence is identical. Here I chose to focus on the "canonical" sequence, SLC2A9a, also known as SLC2A9-L, which encodes the protein GLUT9a. For the rest of this report, GLUT9a will be abbreviated as GLUT9 and SLC2A9a in SLC2A9.

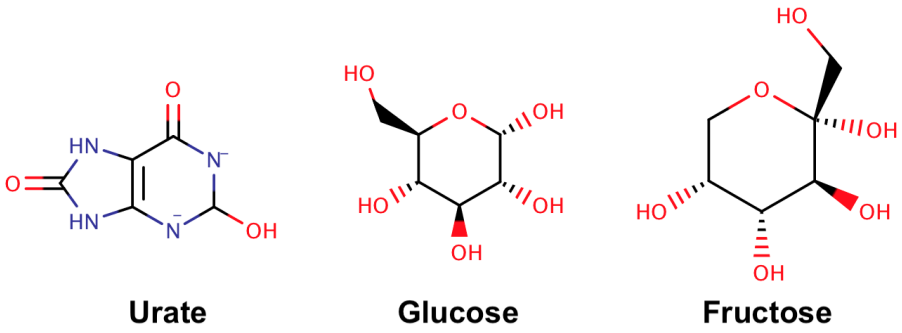

Figure 2: Structure of the substrates transported by GLUT9

Urate/uric acid is normally present in human blood, as a product of the digestion of purine containing food. While its concentration is normally regulated in the blood, an elevated concentration has been linked to different disorders, such as obesity, hyperlipidemia, atherosclerosis, and hyperinsulinemia [9] and GLUT9 has been directly associated with the regulation of urate [1].

Regarding Fructose, it has been shown in the past that a special motif `NXI/NXV` in the GLUT family was determinant in the ability of transporting this hexose: GLUT proteins exhibiting the NXI motif could transport fructose, while proteins exhibiting the NXV motif could not. In GLUT9,

this motif is located on $ASN_{333} - ALA_{334} - ILE_{335}$ which corresponds to the `NXI` motif, giving GLUT9 the ability to transport fructose [17].

However recent in-vitro analyses have brought new insights on the transport of fructose by GLUT9: according to Ebert et al. [7] GLUT9 does not transport Fructose, nor Glucose.

Based on all those information, I chose to orientate my work on the study of urate transporting by GLUT9.

2. *Material and Methods*

2.1 Secondary structure prediction

The prediction of the secondary structure was performed using PSIPRED v3.3 [18] on the amino-acid sequence of SLC2A9.

2.2 Detection of transmembrane segments

The detection of the transmembrane segments was performed using TOPCONS [2] on the amino-acid sequence of SLC2A9.

2.3 Gene selection

To select the protein for the reconstruction of GLUT9 structure, I first performed a PSI-Blast iterative alignment was performed using SLC2A9 sequence, with a word size of 2, a BLOSUM62 matrix, a gap existence cost of 11, and a gap extension cost of 1. Duplicates were then eliminated. Among the top hits, only the gene sequences of which the structure was solved were selected.

2.4 Protein Clustering and structural exploration

In order to proceed with homology modelling, the different available structures were clustered in the two different conformational stage outward open, and inward open. This process was performed both by checking the literature of those structures [5][22][4][11], verifying by structural alignments performed in PyMOL [3], and visualizing using the Bio3D R library [8].

2.5 Single-template homology modelling

Proteins with a solved structure whose sequence similarity were the highest when aligned to SLC2A9 were chosen. One protein was chosen for the inward open conformation, and another one for the outward open conformation. Homology modelling was then performed using Memoir, a tool designed specifically for membrane proteins.[6].

2.6 Multiple sequence alignment

The multiple sequence alignment was performed with MAFFT v7.363b using a BLOSUM62 matrix in the Aliview [13] MSA GUI [12]. Two MSA were realized separately for the amino-acid sequences of protein structures in the inward open conformation, and for the amino-acid sequences of protein structures in the outward open conformation version. MSA were then converted from the `fasta` format to the `pir` format using a custom Python script.

2.7 Multi-template homology modelling

Two different models were realized: one for inward open conformation structures, and another one for outward open conformation structures. The code for the generation of the different templates is available on GitHub. `https://git.io/vFjtD`. When no C-ter and/or N-ter structural information were available, the C-ter and N-ter residues modelled just form a long coil. To avoid this fraudulent

structural information to infer for NMA and Docking, those residues were deleted from the predicted structure.

2.8 Model Evaluatation

The different models generated were evaluated using ProQ [25] using the pdb of the model, and the secondary structure predicted by PSIPRED. The top 2 reliable models on LGScore and MaxSub criteria were re-evaluated using a membrane specific scoring tool: ProQM [23] Finally, the best model was chosen based on the highest ProQM global quality.

2.9 Model refinement

The best model was processed using ModRefiner[26] for side-chain and loop refinement using the structure in same conformation that had the highest sequence identity with SLC2A9.

2.10 Model embedding in Membrane

The best model was embedded in a cell membrane using PPM server [15] with both C-ter and N-ter in the cytoplasm.

2.11 Normal mode analysis

Normal mode analysis was performed using the Bio3D R library [8] on the best inward open model. The first 20 modes were explored by looking at the RMSD variation compared to the outward open model, and visually inspected for structural changes. The mode showing a movement bringing it close to the outward open conformation was investigated more thoroughly.

Protein Name	Organism	identity	E-Value	nb structure
GLUT5	Rattus norvegicus	45%	3e-148	1
GLUT5	Bos taurus	45%	1e-151	1
GLUT1	Homo sapiens	39%	6e-120	4
GLUT3	Homo sapiens	37%	1e-103	4

Table 2: Protein sequences having a resolved PDB structure found by PSI-BLAST of SLC2A9

PDB id	Protein	Organism	Conformation	Resolution (Å)
4PYP	GLUT1	Homo sapiens	inward open	3.17
5EQG	GLUT1	Homo sapiens	inward open	2.9
5EQH	GLUT1	Homo sapiens	inward open	2.99
5EQI	GLUT1	Homo sapiens	inward open	3
4ZW9	GLUT3	Homo sapiens	outward-occluded	1.5
4ZWB	GLUT3	Homo sapiens	outward-occluded	2.4
4ZWC	GLUT3	Homo sapiens	outward-open	2.6
5C65	GLUT3	Homo sapiens	outward-open	2.65
4YBQ	GLUT5	Bos taurus	outward-open	3.27
4YB9	GLUT5	Rattus norvegicus	inward-open	3.2

Table 3: PDB selected for homology modeling, and their conformation.

2.12 Docking of Urate on the best model

The urate molecule, was prepared using MarvinSketch and AutoDockTools-1.5.6 [19]. A first docking box was designed enclosing the whole protein (the best inward open model), then a second smaller box was designed to enclosing only the center channel of the protein to precise the docking location. The docking itself was performed using Vina [24] using an exhaustiveness of 8.

3. *Results*

3.1 Gene selection

The selected gene sequences after the PSI-Blast are described in table 2. The lowest identity score 37% while the highest is 45%. The structures associated with the selected sequences are described in table 3. All structures were obtained by X-ray diffraction.

3.2 GLUT proteins clustering and structure variation

The clustering of the GLUT family based on the RMSD between structures shows two different groups (fig 3): one at the top left representing the inward open structures, and one at the bottom right representing the outward open structures. The same observation can be made on a PCA plot (fig 4) where the first component allows to discriminate between inward and outward open conformations. It can also be noted that this distinction inward/outward carries most of the variance explained.

3.3 Secondary structure

PSIPRED predicted 16 α helix longer than 5 residues and 6 β strands (fig 5). Looking at the Ramachandran diagram (fig 6) of the best predicted inward open structure, we can observe a similar secondary structure distribution: mostly α helix, and a few β strands.

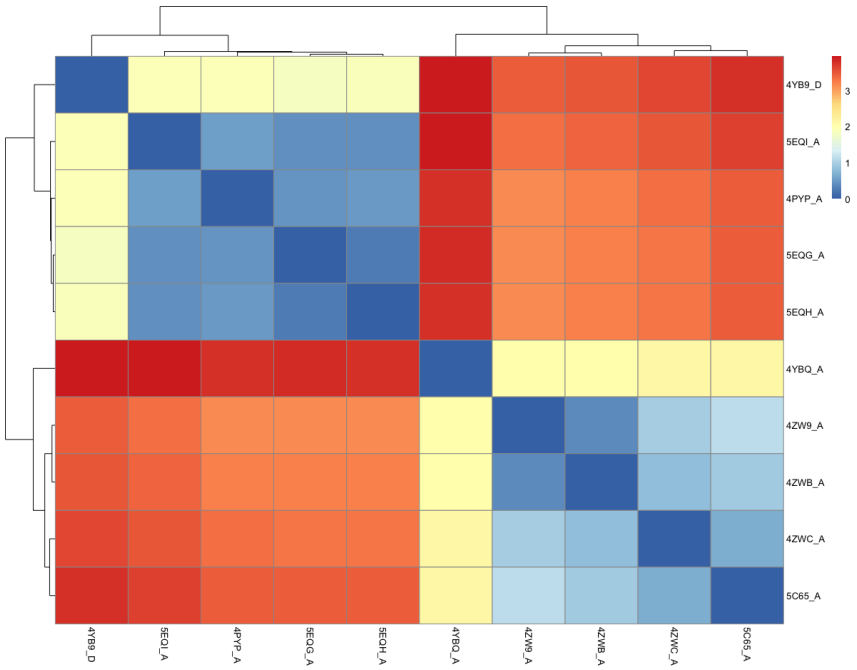

Figure 3: Clustering of the selected structures based on the RMSD. Scale in Angstrom

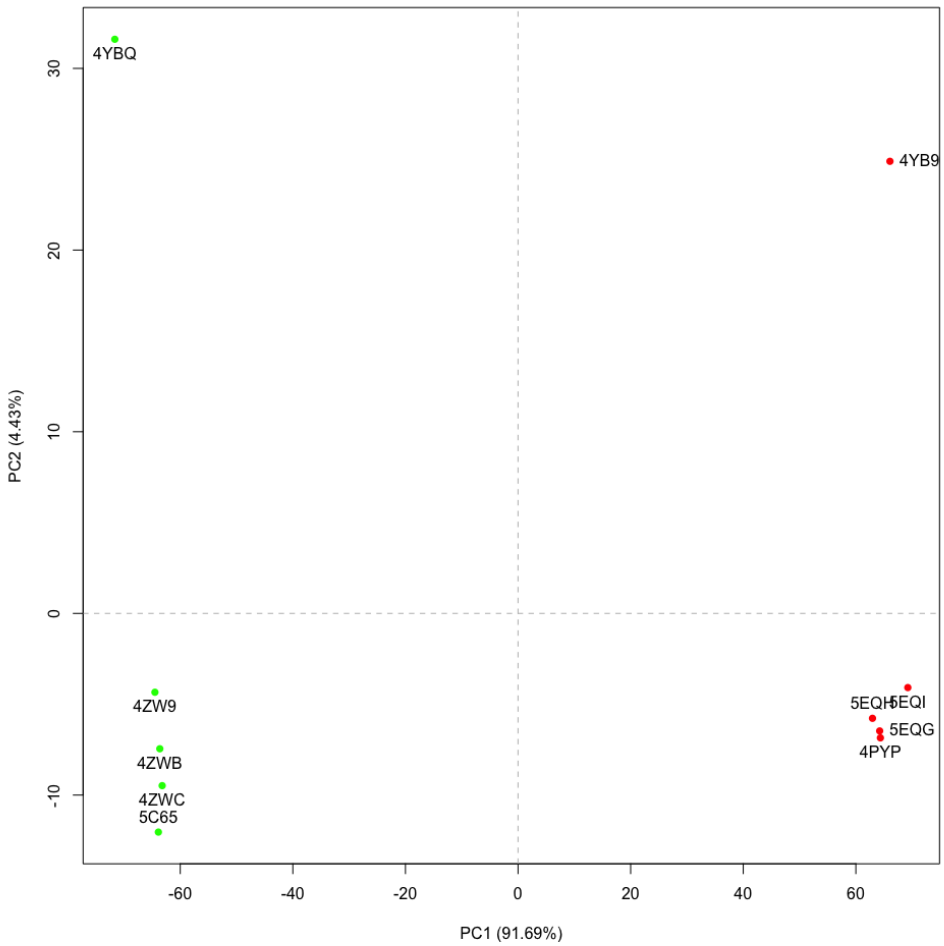

Figure 4: PCA of the selected structures based on atom positions. Inward open conformation are shown in red, outward open conformation are shown in green

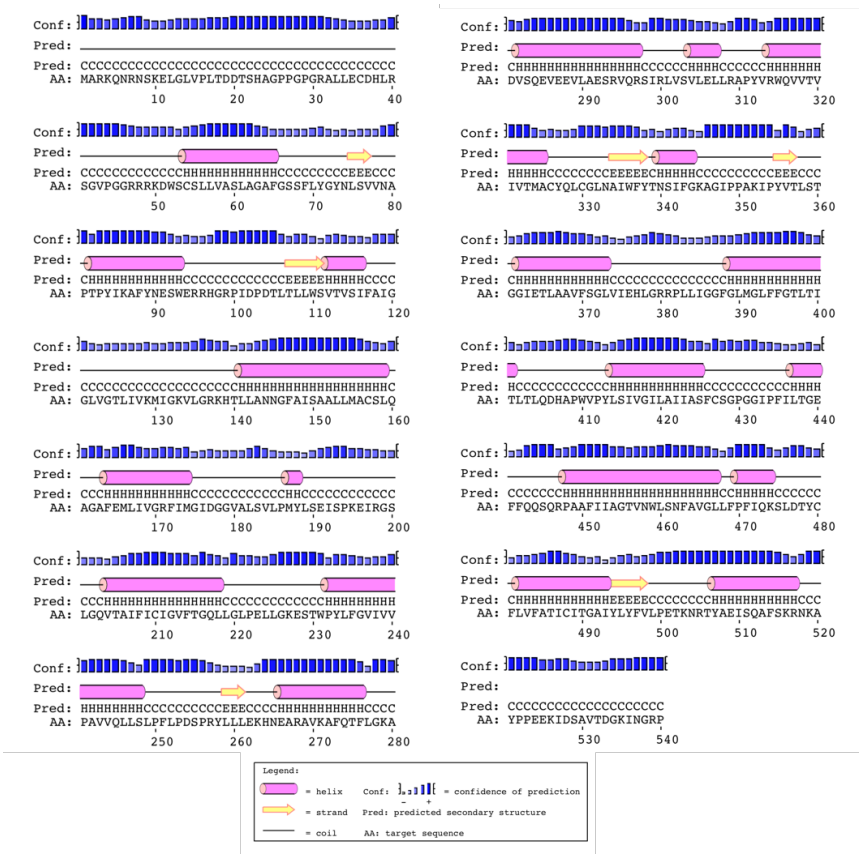

Figure 5: Prediction of secondary structures of SLC2A9 by PSIPRED.

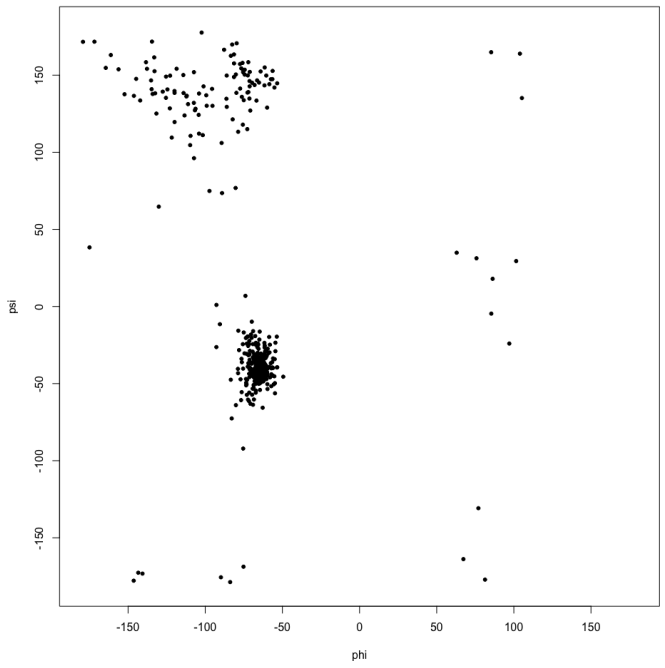

Figure 6: Ramachandran diagram of the inward open structure as predicted by Modeller.

3.4 Transmembrane segments

TOPCONS predicted 12 transmembrane α helices (fig 7) which is expected for the GLUT family. There are 6 helices having an inside -> outside orientation, and 6 helices having an outside -> inside orientation.

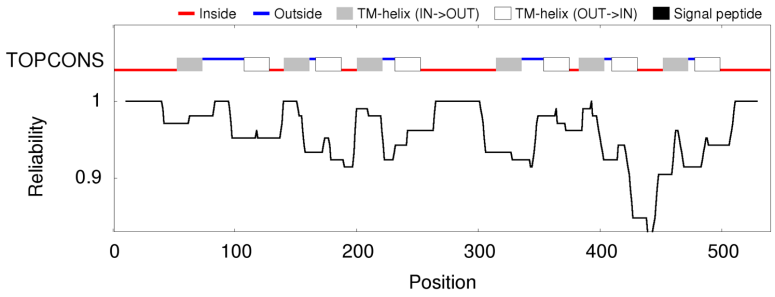

Figure 7: TOPCONS prediction of transmembrane segments

3.5 Choice of the best model

Based on the result of ProQ, and ProQM (table 4), two best models were selected: the modeller multi-template model for the inward-open conformation, and the modeller multi-template model for the outward-open conformation as they had the best combination of LGScore, MaxSub, and ProQM score. Therefore, for the rest of this work, inward open will refer to the inward open multi-template model built by modeller from inward open conformations structures, and outward will refer to the outward open multi-template model built by modeller from outward open conformations structures.

3.6 Model refinement

The refined model with ModRefiner showed a RMSD 3.738Å to the reference model (4YB9) and a RMSD of 2.830Å to the non-refined model. This very high structural difference seems a bit too artificial, and therefore, the model refined by ModRefiner was not used for the rest of this report, as

Model	Using SS	LGScore	MaxSub	ProQM
inward open (modeller)	yes	4.157	0.606	0.661
outward open (modeller)	yes	4.176	0.354	0.678
inward open (memoir complete)	yes	4.133	0.538	0.611
inward open (memoir core)	no	5.575	0.197	-
inward open (memoir hi-acc)	no	5.637	0.198	-
inward open (memoir hi-cov)	no	5.446	0.194	-
outward open (memoir complete)	yes	4.058	0.493	-
outward open (memoir core)	no	5.541	0.346	-
outward open (memoir hi-acc)	no	5.541	0.346	-
outward open (memoir hi-cov)	no	5.475	0.344	-

Table 4: Model selections and Metrics.

the predicted refined structure doesn't seem to be in accordance with the overall structure of GLUT family. (fig 8)

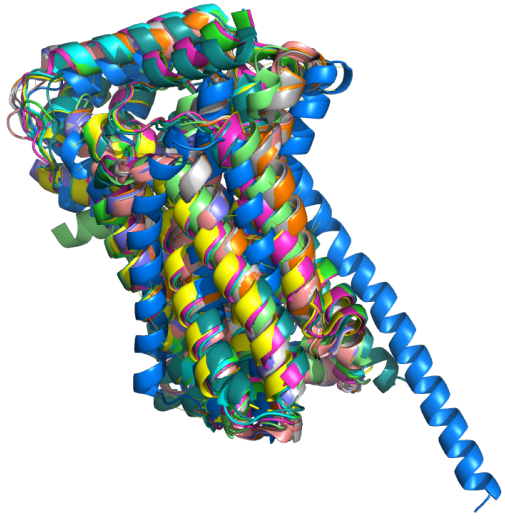

Figure 8: The whole GLUT Family (including the predicted inward open structure of GLUT9) aligned with the refined structure of GLUT9 inward open by ModRefiner (in blue)

3.7 Normal Modes

The analysis of the normal modes of the inward open structure reveals that 80% of the movement from the inward open conformation to the outward open conformation, is described by the first 10 modes (fig 9b). This can also be observed on the mode frequencies, where there is a break between the first 10 frequencies, and the 11th one (fig 9a). The more important movements happen in coil regions, even though there also movements in helices regions, especially in the region of the middle of the protein, around residue 250 (fig 9c). As the first 6 modes are trivial (frequency=0), which correspond to the rotation and translation of the rigid protein itself, [8] only the movement of modes 7 to 10 are displayed in figure 10.

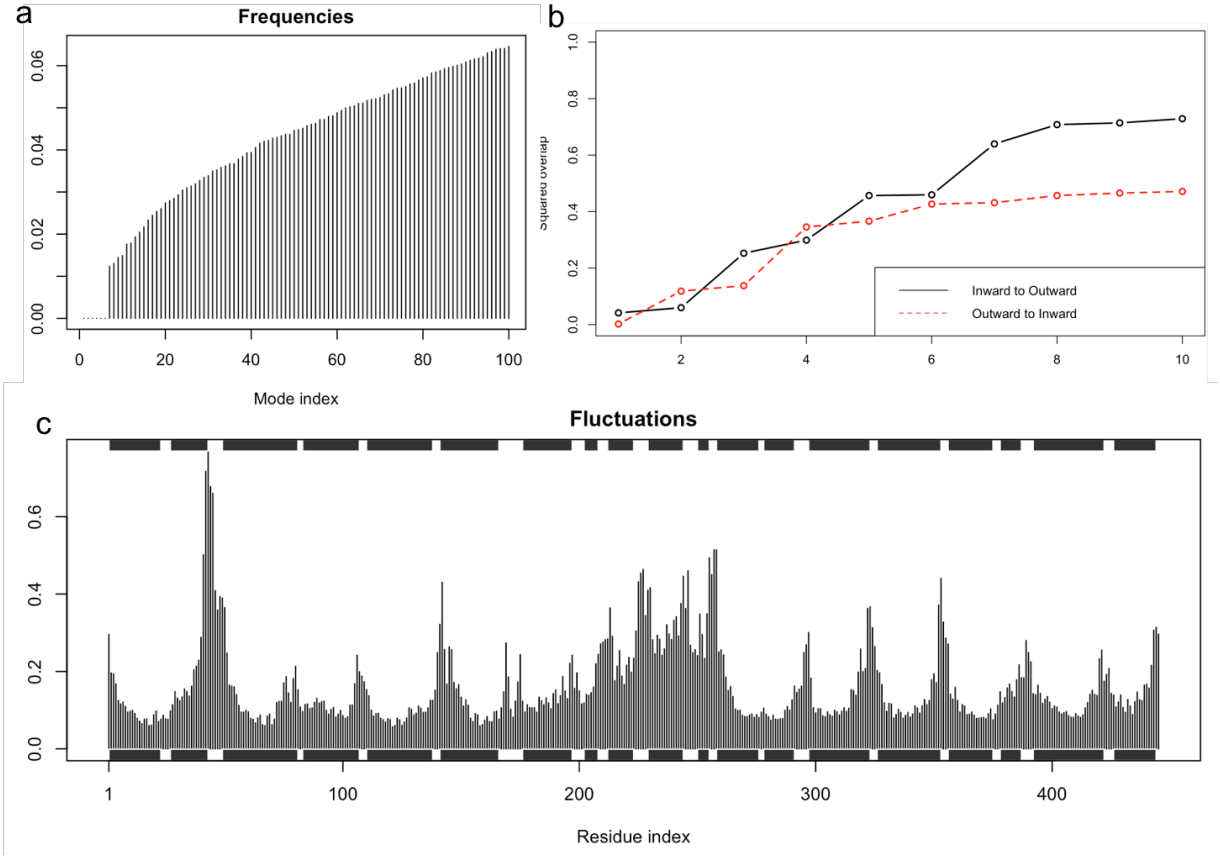

Figure 9: Normal mode analysis. a: Frequencies of the first 100 normal modes. b: cumulative coverage of the first 10 modes for inward open to outward open movement (black) and outward open to inward open (red). c: Atomic fluctuation plot in Å (helices in black and strands in gray)

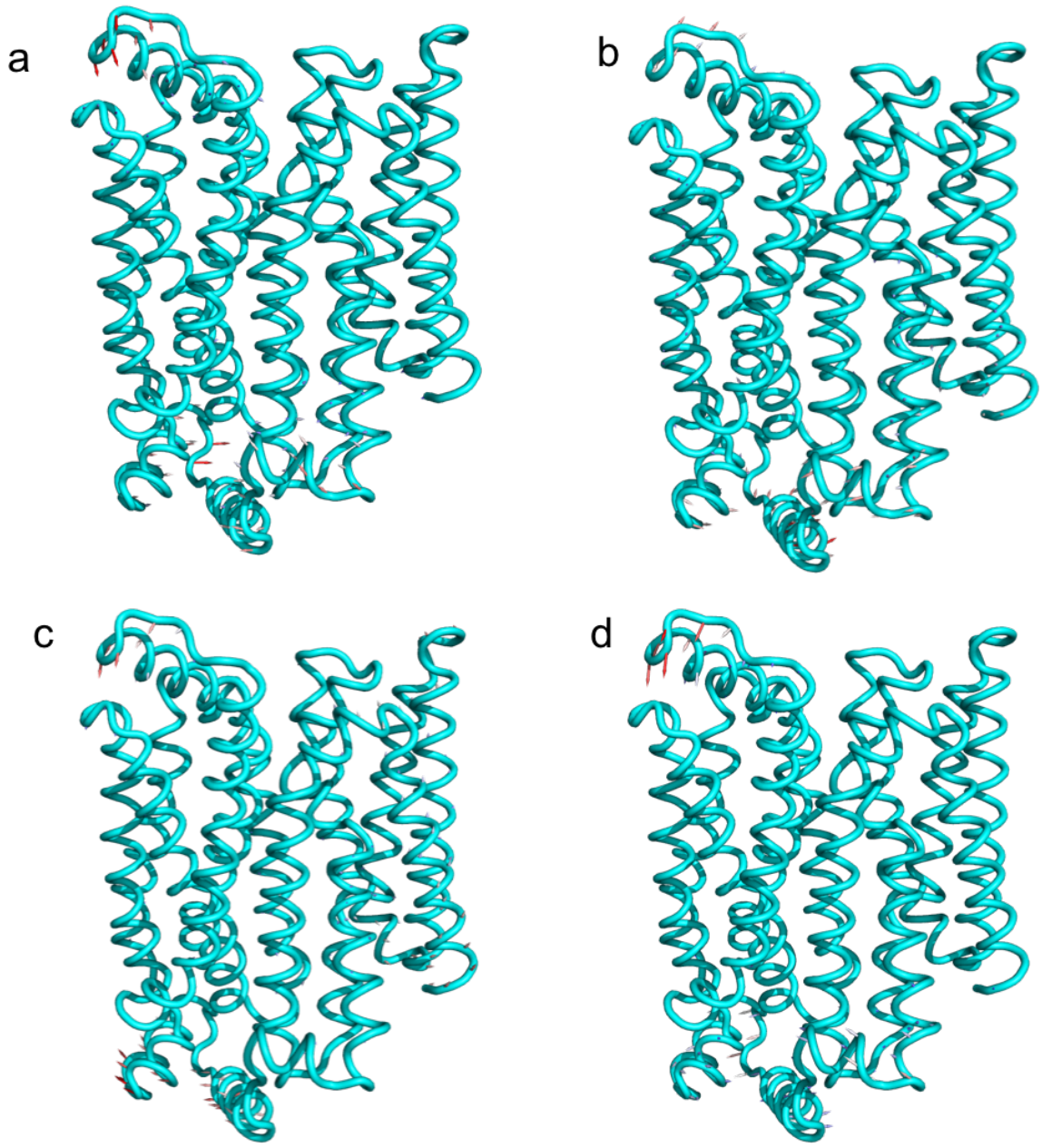

Figure 10: Normal mode movements. Arrows represent the movement described by a normal mode a: mode 7. b: mode 8. c: mode 9. d: mode 10

3.8 Docking

The docking site of urate as described by Long et al. is shown in figure 11. In this outward open conformation, the urate docking site is near the inner membrane, and is composed of 10 amino-acids [16]. In first docking performed by Vina with the grid enclosing the whole protein showed that only 6 of the 20 poses were in the docking site of urate. However, in a second docking where the grid was enclosing only to the whole inner channel of GLUT9 (12) 11 of the 20 poses were in the docking site predicted by Long et al. Two of the best poses are shown in figure (12). From all the poses, it seems to confirm the role of TYR 328 and TYR71 in the transfer of Urate across the cell membrane, first shown by Long et al.

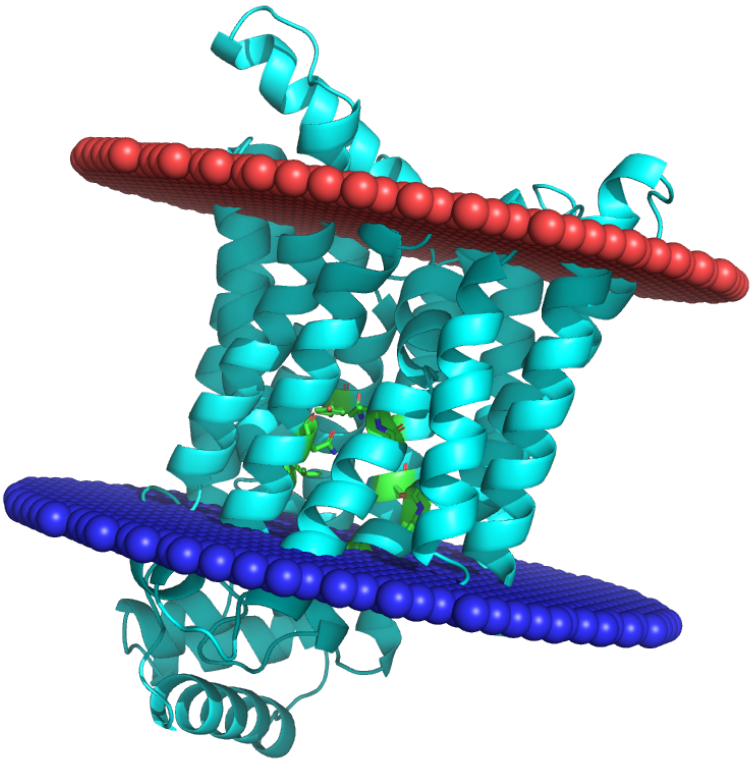

Figure 11: Urate docking pocket in the GLUT 9 inward open conformation embedded in the cell membrane by PPM server. The location of the pocket in GLUT9 is show by the green residues. The outer membrane is shown in red, the inner membrane is shown in blue.

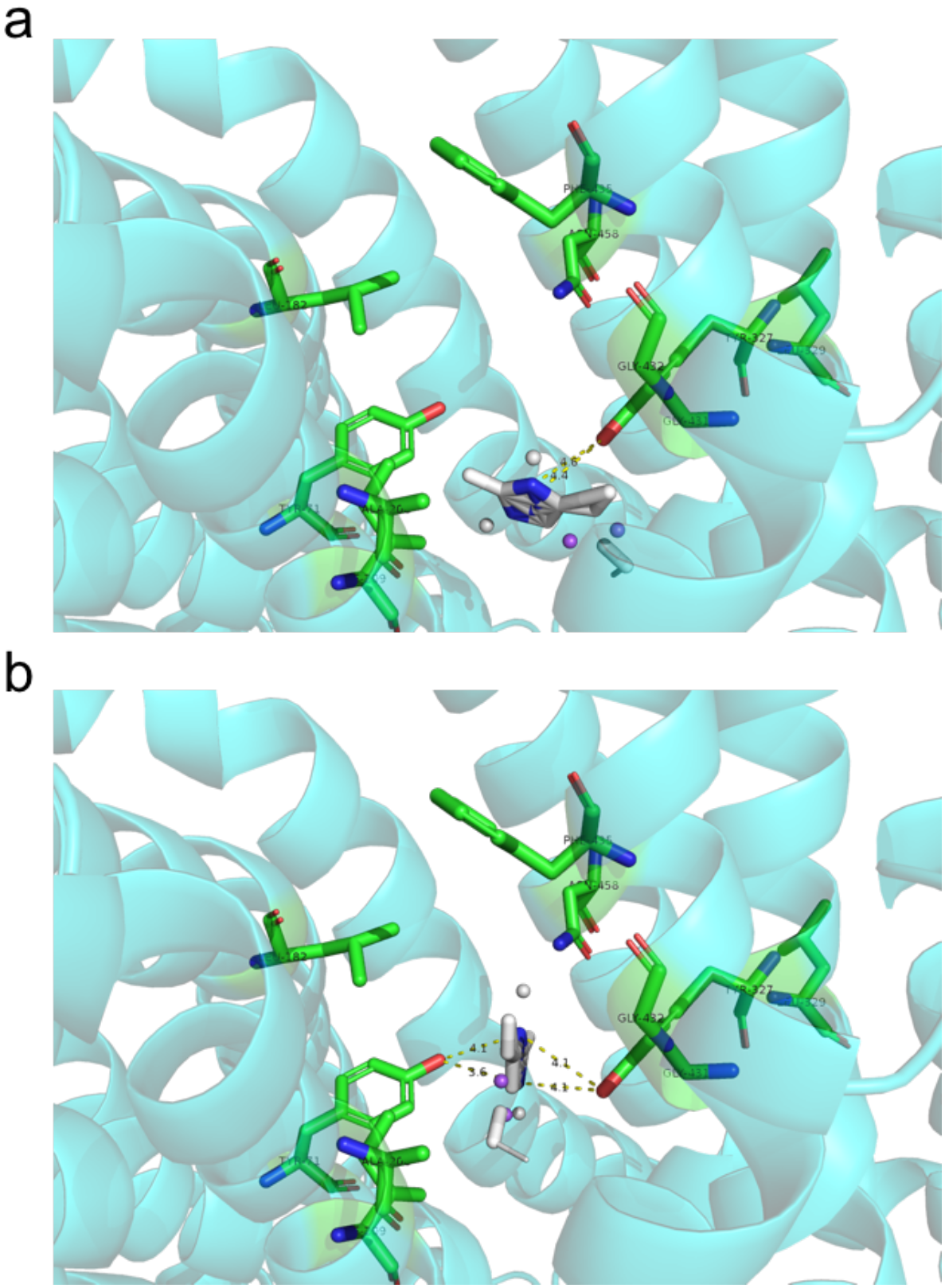

Figure 12: Docking of the urate. Oxygen atoms are shown in red, Nitrogen atoms are shown in blue, and carbon atoms are shown in green. The backbone is shown in cyan. Distances are shown with a dotted yellow line. a: Pose of an urate molecule interacting with TYR 227. b: Pose of an urate molecule interacting with TYR 327 and TYR 71

4. *Discussion and Conclusion*

The predicted secondary structure of GLUT9 by PSIPRED is in accordance with the Ramachandran diagram. However, it predicts too many different secondary structures compared to the expected model of a GLUT family member. It can be assumed that some of the SS predicted are errors, especially given their short size. Another explanation to this SS count discrepancy could be that in the GLUT family, some helices are not revolving along the same axis, but might have some flexibility which could be interpreted as coils by PSIPRED. Predicted $beta$ strands sizes are overall very short, which makes them likely to be artifacts. The canonical number of 12 TM helices is confirmed in GLUT9 by TOPCONS with a good reliability. The drop of reliability in TOPCONS prediction around 450 amino-acids can be explained by the absence of amino-acids residues at this position in many available solved structures. Overall, the best homology-modelling technique is unsurprisingly the one that borrows information from a maximum of available structures. Multi-template modelling should be preferred to single-template modelling when enough relevant structures are available. Still, dedicated membrane protein homology modelling software such as Memoir perform quite decently. The scoring score of model evaluation methods seem to agree on the best model. Refining the model might be a good idea, however, here, there are no long coils in the middle of the structure. Yet side chain refinement could improve the docking result, but the method employed here didn't deliver reliable results. Another side-chain refinement method should be preferred to ModRefiner for this model. The NMA showed that a combination of the first 10 modes provides a good description of the conformational change from inward open to outward open. Given the very short computational time of NMA compared to a MD, it is an option that shouldn't be neglected in a bioinformatic structural analysis. The docking of Urate on the inward open conformation modelled structure of GLUT9 tends to agree with its urate transporter function. However, only reproducible specie specific experimental confirmations of the transport of those molecules will bring a definitive conclusion to the subject.

Bibliography

[1] Anzai, N., Ichida, K., Jutabha, P., Kimura, T., Babu, E., Jin, C. J., Srivastava, S., Kitamura, K., Hisatome, I., Endou, H., et al. (2008). Plasma urate level is directly regulated by a voltage-driven urate efflux transporter uratv1 (slc2a9) in humans. *Journal of Biological Chemistry*, 283(40):26834–26838.

[2] Bernsel, A., Viklund, H., Hennerdal, A., and Elofsson, A. (2009). Topcons: consensus prediction of membrane protein topology. *Nucleic acids research*, 37(suppl_2):W465–W468.

[3] DeLano, W. L. (2002). The pymol molecular graphics system. *http://pymol. org*.

[4] Deng, D., Sun, P., Yan, C., Ke, M., Jiang, X., Xiong, L., Ren, W., Hirata, K., Yamamoto, M., Fan, S., et al. (2015). Molecular basis of ligand recognition and transport by glucose transporters. *Nature*, 526(7573):391–396.

[5] Deng, D., Xu, C., Sun, P., Wu, J., Yan, C., Hu, M., and Yan, N. (2014). Crystal structure of the human glucose transporter GLUT1. *Nature*, 510(7503):nature13306.

[6] Ebejer, J.-P., Hill, J. R., Kelm, S., Shi, J., and Deane, C. M. (2013). Memoir: template-based structure prediction for membrane proteins. *Nucleic Acids Research*, 41(Web Server issue):W379–W383.

[7] Ebert, K., Ludwig, M., Geillinger, K. E., Schoberth, G. C., Essenwanger, J., Stolz, J., Daniel, H., and Witt, H. (2017). Reassessment of GLUT7 and GLUT9 as Putative Fructose and Glucose Transporters. *The Journal of Membrane Biology*, 250(2):171–182.

[8] Grant, B. J., Rodrigues, A. P., ElSawy, K. M., McCammon, J. A., and Caves, L. S. (2006). Bio3d: an r package for the comparative analysis of protein structures. *Bioinformatics*, 22(21):2695–2696.

[9] Hayden, M. R. and Tyagi, S. C. (2004). Uric acid: A new look at an old risk marker for cardiovascular disease, metabolic syndrome, and type 2 diabetes mellitus: The urate redox shuttle. *Nutrition & metabolism*, 1(1):10.

[10] Henderson, P. J. and Baldwin, S. A. (2013). This is about the in and the out. *Nature structural & molecular biology*, 20(6):654–655.

[11] Kapoor, K., Finer-Moore, J. S., Pedersen, B. P., Caboni, L., Waight, A., Hillig, R. C., Bringmann, P., Heisler, I., Müller, T., Siebeneicher, H., and Stroud, R. M. (2016). Mechanism of inhibition of human glucose transporter GLUT1 is conserved between cytochalasin B and phenylalanine amides. *Proceedings of the National Academy of Sciences*, 113(17):4711–4716.

[12] Katoh, K. and Standley, D. M. (2013). Mafft multiple sequence alignment software version 7: improvements in performance and usability. *Molecular biology and evolution*, 30(4):772–780.

[13] Larsson, A. (2014). Aliview: a fast and lightweight alignment viewer and editor for large datasets. *Bioinformatics*, 30(22):3276–3278.

[14] Li, S., Sanna, S., Maschio, A., Busonero, F., Usala, G., Mulas, A., Lai, S., Dei, M., Orrù, M., Albai, G., et al. (2007). The glut9 gene is associated with serum uric acid levels in sardinia and chianti cohorts. *PLoS genetics*, 3(11):e194.

[15] Lomize, M. A., Pogozheva, I. D., Joo, H., Mosberg, H. I., and Lomize, A. L. (2012). Opm database and ppm web server: resources for positioning of proteins in membranes. *Nucleic acids research*, 40(D1):D370–D376.

[16] Long, W., Panigrahi, R., Panwar, P., Wong, K., ONeill, D., Chen, X.-Z., Lemieux, M. J., and Cheeseman, C. I. (2017). Identification of Key Residues for Urate Specific Transport in Human Glucose Transporter 9 (hSLC2a9). *Scientific Reports*, 7.

[17] Manolescu, A. R., Augustin, R., Moley, K., and Cheeseman, C. (2007). A highly conserved hydrophobic motif in the exofacial vestibule of fructose transporting SLC2a proteins acts as a critical determinant of their substrate selectivity. *Molecular Membrane Biology*, 24(5-6):455–463.

[18] McGuffin, L. J., Bryson, K., and Jones, D. T. (2000). The psipred protein structure prediction server. *Bioinformatics*, 16(4):404–405.

[19] Morris, G. M., Huey, R., Lindstrom, W., Sanner, M. F., Belew, R. K., Goodsell, D. S., and Olson, A. J. (2009). Autodock4 and autodocktools4: Automated docking with selective receptor flexibility. *Journal of computational chemistry*, 30(16):2785–2791.

[20] Mueckler, M., Caruso, C., Baldwin, S. A., Panico, M., Blench, I., Morris, H. R., Allard, W. J., Lienhard, G. E., and Lodish, H. F. (1985). Sequence and structure of a human glucose transporter. *Science*, 229:941–946.

[21] Mueckler, M. and Thorens, B. (2013). The SLC2 (GLUT) family of membrane transporters. *Molecular Aspects of Medicine*, 34(2-3):121–138.

[22] Nomura, N., Verdon, G., Kang, H. J., Shimamura, T., Nomura, Y., Sonoda, Y., Hussien, S. A., Qureshi, A. A., Coincon, M., Sato, Y., Abe, H., Nakada-Nakura, Y., Hino, T., Arakawa, T., Kusano-Arai, O., Iwanari, H., Murata, T., Kobayashi, T., Hamakubo, T., Kasahara, M., Iwata, S., and Drew, D. (2015). Structure and mechanism of the mammalian fructose transporter GLUT5. *Nature*, 526(7573):nature14909.

[23] Ray, A., Lindahl, E., and Wallner, B. (2010). Model quality assessment for membrane proteins. *Bioinformatics*, 26(24):3067–3074.

[24] Trott, O. and Olson, A. J. (2010). Autodock vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *Journal of computational chemistry*, 31(2):455–461.

[25] Wallner, B. and Elofsson, A. (2003). Can correct protein models be identified? *Protein science*, 12(5):1073–1086.

[26] Xu, D. and Zhang, Y. (2011). Improving the physical realism and structural accuracy of protein models by a two-step atomic-level energy minimization. *Biophysical journal*, 101(10):2525–2534.