RESEARCH ARTICLE

Proteins
DOI10.1002/prot.26453

In silico investigation on structure-function relationship of members belonging to the human SLC52 transporter family

Molecular modeling of SLC52 transporters

Omar Ben Mariem[1], Simona Saporiti[1], Uliano Guerrini[1], Tommaso Laurenzi[1], Luca Palazzolo[1], Cesare Indiveri[2], Maria Barile[3], Emma De Fabiani[1], Ivano Eberini[1,4,*]

[1] Dipartimento di Scienze Farmacologiche e Biomolecolari, Via Giuseppe Balzaretti 9, 20133 – Milan, Italy
[2] CNR Institute of Biomembranes, Bioenergetics and Molecular Biotechnologies (IBIOM) via Amendola 122/O, 70126 Bari, Italy
[3] Department of Biosciences, Biotechnology and Biopharmaceutics, University of Bari A.Moro, via Orabona 4 –, 70126 Bari, Italy
[4] DSRC, Università degli Studi di Milano, Milan, Italy
[*] Correspondence to ivano.eberini@unimi.it

ACKNOWLEDGMENTS

O.B., S.S., U.G, L.P. and I.E., were supported by grants from MIUR—Progetto Eccellenza.

This work was partially financed by
"Cure RTD Foundation" http://curertd.org/news/new/ to M.B.

Received: Jul 15, 2022;Revised: Nov 18, 2022;Accepted: Dec 08, 2022

ABSTRACT

Riboflavin is an essential water-soluble vitamin that needs to be provided through the diet because of the conversion into flavin adenine dinucleotide (FAD) and flavin mononucleotide (FMN), important cofactors in hundreds of flavoenzymes. The adsorption and distribution of riboflavin is mediated by transmembrane transporters of the SLC52 family, namely RFVT1-3, whose mutations are mainly associated with two diseases, MADD and the Brown-Vialetto-Van Laere syndrome. Interest in RFVTs as pharmacological targets has increased in the last few years due to their overexpression in several cancer cells, which can be exploited both by blocking the uptake of riboflavin into the cancerous cells, and by performing cancer targeted delivery of drugs with a high affinity for RFVTs. In this work we propose three-dimensional structural models for all three human riboflavin transporters obtained by state-of-the-art artificial intelligence-based methods, which were then further refined with molecular dynamics simulations. Furthermore, two of the most notable mutations concerning RFVT2 and RFVT3 (W31S and N21S respectively) were investigated studying the interactions between the wild-type and mutated transporters with riboflavin.

Keywords: riboflavin, SLC52, RFVT, artificial-intelligence, molecular dynamics, 3D modeling, Brown-Vialetto-Van Laere syndrome.

INTRODUCTION

Riboflavin (RF), more commonly known as vitamin B2, is an essential molecule for aerobic life, as its two main derivatives, flavin adenine dinucleotide (FAD) and flavin mononucleotide (FMN); these molecules are used as cofactors of hundreds of flavoenzymes, involved in redox reactions[1].

The human body is unable to synthesize riboflavin, so it must be introduced with the diet. In particular, it can be found in a wide variety of animal products and several vegetables [2–4]. There, it exists as FAD and FMN and prior to its absorption it is converted to riboflavin by the intestinal enzymes FAD pyrophosphatase and FMN phosphatase[2]. Additionally, it is synthesized by many species of bacteria present in the human gut microbiota and made available for bsorption in its free form[4].

The cellular uptake of the vitamin is mainly exerted by three solute carrier transporters (SLC), also named riboflavin transporters (RFVTs), belonging to the novel human SLC52 family. SLCs are ubiquitous transmembrane proteins which carry ions and small molecules across the membranes of cells and organelles[5]. SLCs are divided into several subfamilies according to their phylogenetic analysis. The transport itself can be uni- or bidirectional and it involves major conformational rearrangements of the transporters for bringing the structure from the “inward-open” to the “outward-open”, that are the end conformations. Three transport mechanisms have been identified: rocker-switch, gated-pore and elevator, each one involving different helices. In particular, RFVT1-3 are assumed to belong to the major facilitator superfamily (MFS)[6] which is correlated with a rocker-switch transport mechanism.

The three transporters share a relatively high degree of sequence identity (Table 1); they are characterized by 11 transmembrane helices, connected by intra- and extra-cellular loops, one of which, between TM6 and TM7, spans over 70 residues in all three transporters. RFVT3 also presents a large extracellular loop between TM5 and TM6.

To ensure the trans-epithelial transport of the vitamin into the blood, RFVTs are differently located in the polarized intestinal epithelium. RFVT3 is maximally expressed on the apical membrane of enterocytes, whereas RFVT1, and to a minor extent RFVT2, perform their activity in the baso-lateral membrane[7]. Differential tissue expression and functional specialization of each transporter allow the distribution of the vitamin to the liver and to all the other tissues which have differential requirements for flavocoenzyme [8]

Despite a comparable, and high affinity for riboflavin (K_m ranging from 0.26 to 1.38 μM) [9,10], they seem to exhibit different kinetic features[2,10]. The transport of riboflavin is ion-independent, but there is evidence suggesting that the sole RFVT3 may be sensitive to pH variation[11]. Certain compounds, such as methylene blue, could inhibit riboflavin transporters, with a higher selectivity for RFVT3 than for RFVT1 and RFVT2[12]. Other molecules have been tested as competitive inhibitors of RFVTs, such as lumiflavin and lumichrome, with both in vivo and in vitro studies via cell models[13,14]. Nevertheless, from the kinetical point of view there is still a lack of information: in particular, it is not completely clear if, although with a much lower affinity compared to riboflavin, FMN and FAD might also be transported by RFVTs as experiments have found that the presence of the two leads to a decrease in the transport of riboflavin in cell systems[13].

From a pathology perspective, the importance of these transporters resides in the severe effects caused by some mutations. Alterations of RFVT1 were firstly discovered as a cause of a transient neuro-muscular disorder named RR-MADD or riboflavin responsive Multiple Acyl CoA Dehydrogenase Deficiency (OMIM #615026).

The most notable disorder caused by defects in RFVT2 and RFVT3 is the Brown-Vialetto-Van Laere syndrome (RTD type 2 and RTD type 3; OMIM #614707 and #211530, respectively). This disease can lead to blindness, deafness, and severe nerve palsies, but it can be in some cases successfully treated by high doses of vitamin B2 supplements. The pathogenic mutations in *SLC52A2* and *SLC52A3* described to date are reported in Console et al.[15], Tolomeo et al.[8], CureRTD database[16].

In the last few years, the interest in studying these transporters has grown also because of their role in cancer[6,17,18], making them promising targets for antineoplastic therapy. In fact, RFVTs are overexpressed in many tumor types, such as colorectal cancer, squamous cell carcinoma, skin melanoma, and luminal A breast cancer[6,17,18], to accomplish the increased need for riboflavin and its derivatives due to their increased metabolism[6]. Therefore, the use of drugs that reduce the transport of riboflavin could be at the basis of a powerful strategy to exploit the increased expression of these transporters.

Accordingly, understanding the three-dimensional (3D) structure of these proteins is pivotal to performing drug discovery, as it would allow the identification of the key residues participating in the recognition and transport of solutes. However, the experimentally solved structures of these transporters are not available yet, and little is known about their structure-function relationships. Currently, most of the knowledge regarding RFVTs is derived by sequence analysis[19] that can provide information with a certain degree of confidence and from some homology structural models. However, homology models were difficult to set due to the lack of suitable templates, i.e., proteins with solved 3D structures with sequence identity with RFVTs transporters of more than 20%. The most refined models were recently obtained for RFVT2 starting from a multi-template approach [15] and using threading methods[20–2320].

[212223]The goal of this investigation is to obtain more reliable 3D models of RFVT1, RFVT2 and RFVT3, using *in silico* molecular modeling methods [2,10,15,24–27]. We propose here to improve the reliability of structural models by using innovative ab initio methods based on artificial intelligence (AI), including AlphaFold (AF)[28] and the newly released RoseTTaFold[29]. Although AF obtained impressive results in the CASP2020[30] we decided to use multiple methods and compare their results, given how little is known about the 3D structures of these transporters.

After the evaluation of the most suitable model via molecular dynamics (MD) simulations, an investigation of the important structural features of the transporter was carried out via molecular docking, as well as *in silico* mutagenesis of key residues for substrate recognition.

MATERIALS AND METHODS

Modeling of RFVT1, RFVT2, and RFVT3

The sequence of all three human riboflavin transporters was retrieved from the UniProt knowledgebase database (UniProtKB IDs: Q9NWF4 (RFVT1), Q9HAB3 (RFVT2), and Q9NQ40 (RFVT3)). Their secondary structure was predicted using TMHMM server[31], PsiPred[32], Jpred[33] and RaptorX[34] that use the protein primary structure. Since the protein BLAST searches run in the Protein Data Bank (PDB) did not identify any suitable homolog for modeling procedures for the three RFVTs, we decided to build the 3D models for these three proteins using state-of-the-art AI-based software. For each protein we generated three models: one obtained by RosettaTR[35], one by RoseTTAfold[29], and one taken from the AF database[28], for a total of 9 models. AF allows the user to download the 3D structure of the query protein directly, while the other two software need the primary structure as input. The default options were kept while using these two tools.

Equilibration and MD analysis

MD simulations to equilibrate the obtained models and frame clustering procedures were carried out with tools available with the Schrödinger Suite 2022-1, specifically Desmond and the analysis modules (D. E. Shaw Research, New York, NY; Schrödinger, New York, NY)[36].

The Desmond System Builder tool was used to place the apo-models of all RFVTs into a reference membrane bilayer made up by 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine (POPC). Protein orientation was set up according to the OPM server[37], which provides spatial arrangements of membrane proteins with respect to the hydrocarbon core of the lipid bilayer. The system was solvated with SPC water molecules in a sufficiently large box, to fit the whole protein plus a margin (buffer) to account for protein movements (see Supplementary Table 1 for more details on the box size and number of atoms). Chloride ions were added to neutralize the exceeding positive charge and sodium chloride to reach a 0.15 M concentration. The system was energy-minimized to relax the assembly and remove clashes among protein, membrane, and solvent in the new setup.

To produce equilibrated models, each system was submitted to two MD simulations of 1000 ns using the Desmond Molecular Dynamics tool. Periodic boundary conditions (PBCs) and the following parameters were set: 300K and Nose–Hoover thermostat for temperature coupling, 1 bar and Martyna–Tobias–Klein piston for pressure coupling, and 2 fs as the integration time step. The force field used for all MD simulations was OPLS4[38]. Coordinates and velocities of each atom were saved every 0.5 ps.

The Desmond Trajectory Frame Clustering tool was used to cluster the whole MD simulations to select the most representative frame (the medoid) for each cluster. Distances between trajectory frames were computed from the root mean square deviation (RMSD) matrix of alpha carbons.

The RMSD, the root-mean-square-fluctuation (RMSF), the solvent accessible surface area (SASA), and the secondary structure properties were calculated via Python scripts using the Schrödinger analysis API.
Cluster analysis based on RMSD of the alpha carbon atoms was performed via the Schrödinger analysis API in order to identify the most frequent conformational clusters, in particular using the gromos clustering algorithm[39]. When considering the whole models, the threshold chosen was 4Å, while when considering only the transmembrane helices, the threshold was 2Å.
The number of water molecules within 5Å of riboflavin and forming water bridges during the MD simulations were calculated via Python scripts provided by Schrödinger and the MDAnalysis Python library[40,41] respectively.
A solvent analysis was performed on the medoids of the MD simulations by the ‘‘Solvent analysis’’ tool included in MOE 2020.09. The goal of this analysis was to estimate the solvent contribution to the RF-transporter interaction. With this method, the distribution functions of the solvent moieties and the associated free energy are computed. The calculation was performed with the AMBER10:EHT force field and using the 3D reference interaction site model (3D-RISM)[42–44]. The ‘‘solute’’ mode was used and those residues forming the RF recognition site were selected considering a cutoff distance of 5 Å from RF. The average of water molecules with a negative solvation dG among the replicas was calculated as reported in Saporiti et al.[45]
A study of the models' conformations was performed via the analysis of the number of water molecule in the extracellular-facing side (“upper”) and intracellular-facing side (“lower”) using the Schrödinger analysis API. This was performed by considering water molecules within 5 Å of residues located towards the extracellular and intracellular part of the protein respectively.
Two 500 ns MD simulations of one molecule of riboflavin in solvent have been performed to evaluate if the obtained docking poses were comparable to its behavior in simulated physiological conditions.

Molecular Docking

The medoids of the most populated cluster for each of the resulting trajectories were used as receptors to perform molecular docking of the main transported solute, riboflavin. The grid generation and the molecular docking protocols offered by Glide, as available in the Schrödinger suite 2022-1[46] were used. The .sdf file of riboflavin was downloaded from the ChEBI database[47]. The molecule was prepared using the LigPrep collection available in Schrödinger 2022-1 (Schrödinger Release 2022-2: LigPrep, Schrödinger, LLC, New York, NY, 2021).
The receptor box was built using the translocation funnel residues as reference, in particular the ones on the upper part of the funnel. Multiple conformations of riboflavin were considered by keeping the ligand flexible in all docking procedures.
The “XP precision” docking procedure was used, and 10 poses were generated for each docking procedure. Prime MM-GBSA was used to estimate the binding free energies of the complexes.

The top-scoring resulting poses were submitted to three replicas of 200 ns molecular dynamics simulations in order to assess the binding interactions and stability of the complexes. The average interaction energies throughout the replicas were calculated using the Schrödinger 2022-1 analysis modules, considering both Coulomb and van der Waals interaction energies.

Mutagenesis

The Residue Scanning Calculation tool available in the Schrödinger 2022-1 suite was used to generate RFVT2 and RFVT3 mutants in complex with riboflavin. The mutations to be introduced were chosen among the previously mentioned mutations, which were obtained from the proteins' UniProtKB pages and their associated bibliography. Values of the change in the energies related to protein stability (ΔStability) and complex affinity (ΔAffinity) were calculated. Additionally, the same residues were mutated on the centroid of the equilibration MD simulations, to perform a molecular docking of riboflavin, following the same procedure as the wild-type transporters. Three replicas of MD simulations were run for each complex to observe differences caused by the introduced mutations. Each replica is 200 ns long. In Table 2 the mutations introduced are reported.

RESULTS and DISCUSSION

Model generation, equilibration, and analysis

As no experimentally solved 3D structures were available for any of the three transporters, the first attempt at generating a reliable model for these transporters was done *via* a classical homology modeling approach. The first step in the homology modeling procedure is the identification of a homologous template with a sufficiently high sequence identity/similarity level with the query protein. In Supplementary Figure 1 the results of a search using protein BLAST[48] are reported for each transporter. It became immediately obvious from these results that no single template would have allowed the building of a reliable model. In fact, even if the percentage of identity could seem high, the protein coverage is too low, as the alignment can cover at most ~100 residues out of ~500. Indeed, the homology model of RFVT2 was recently constructed using a three-template strategy[15].

To generate suitable models of RFVT1 and 3 and to improve the suitability of the 3D model structure of RFVT2, AI-based methodologies, which have shown significant improvements during the last few years[30], were chosen as the preferred methodologies for generating these models. In the attempt of choosing the most accurate out of the available tools, the three most reliable methods were used, and the results were compared. These methods are AF, RoseTTAFold and RosettaTR.

Some members of the SLC transporters superfamily are known to be functional as homodimers[21], homotrimers[22], or heterodimers[23]. Therefore, AlphaFold-Multimer[31] was used to obtain tentative homodimer and homotrimer complexes (Supplementary Figures 2-3). While the homodimeric models may appear realistic at a first look, the homotrimeric models lack any sort of reliability. In fact, some monomers are even generated with the intracellular side turned to the extracellular side of the other ones. However, as the oligomerization state of riboflavin transporters is still unknown, this study will focus on the generation of monomeric models and their interaction with riboflavin. Reliable modeling of oligomeric structures would require first direct evidence that RFVTs form oligomers in biological membranes in cellular contexts.

After a preliminary investigation of the 3D models generated, none resulted so critical to be immediately excluded. For this reason, we decided to place all models in a POPC membrane and to perform, for each model, two MD simulations in replicates in order to equilibrate the structures and obtain data that could help us choose the best models. After the simulations, trajectories were analyzed by calculating the RMSD, RMSF, performing a cluster analysis of the protein conformations and quantifying the amount of time each residue spent in an ordered helical secondary structure.

In Figure 1, the RMSD and RMSF profiles of the models throughout the simulations are reported. The structural superposition of the three medoids isolated from the most populated clusters for each model of each RFVT are superposed and colored according to RMSD in Figure 2.

For RFVT1, the prediction of transmembrane helices is quite comparable among the methods, giving also very similar results to the prediction provided by UniProtKB. The main differences are instead found in the prediction of extracellular and intracellular loops. This can be correlated to the relatively low mobility expected from the transmembrane domains, while the loops undergo higher variability and fluctuation. The evaluation of RMSD and RMSF leads to a similar conclusion (Figure 1A), as all models reach a *plateau* of the RMSD value, and the RMSF reaches the highest values around the residues belonging to the loops, while it is relatively low around the transmembrane helices. The high RMSF observed in the loops is due to the intrinsic unorganized nature of these regions, but it does not necessary correlate with a low reliability of the model. So, the differences observed in the prediction of the loops should not be considered relevant for this study but might be investigated in the future to account for interactions with other proteins.

Very similar results were obtained by the analysis of the trajectories of RFVT2 (Figure 1B). In fact, all the transmembrane domains of the models are essentially identical, while the loops show higher variability. However, all models show a *plateau* of RMSD and low RMSF in the transmembrane residues.

The models of RFVT3, on the other hand, showed some differences. While having an identical overall architecture, particularly concerning the transmembrane helices and the identification of the residues located in the intracellular and extracellular loops, there were slight differences in the conformation of the models, as well as the secondary structure arrangement of the two largest loops. In fact, unlike in the models RFVT1 and RFVT2, all three of the AI-based methods were able to recognize larger ordered regions, identifying stretches of alpha-helices and beta-strands, albeit with small differences (Figure 2C).

Because of this, no model could be immediately chosen as the best one for any of the transporters. However, the RosettaTR models are associated to an overall higher RMSF during dynamics, which might have been caused by a lower quality model. The SASA values were calculated considering the residues facing the translocation funnel, but, again, no significant differences could be observed among the models. Ultimately, the differences were not considered significant enough to justify choosing one model over the others, particularly regarding the transmembrane domains. Therefore, because the current study does not involve intracellular and extracellular loops, we decided to perform the following calculations on the medoids of the most populated cluster of the MDs of the models obtained by AF, supported by the impressive performance at CASP2020[30], demonstrating the ability of generating much better models than the competition.

The 3D models of RFVTs share a high structural similarity, as expected by the high sequence identity. The biggest differences can be observed in the large loops connecting the helices of RFVT3.

A cluster analysis based on RMSD was carried out on AF models' trajectories, after merging the replicas, in order to evaluate the systems equilibration according to the presence of one most populated cluster. As expected, considering the whole proteins did not allow to identify a unique conformational cluster among the replicas (see Supplementary Figure 4, left panels), because of the large intracellular and extracellular loops, which have many degrees of freedom. However, if the cluster analysis was performed only on the transmembrane helices, a single conformational cluster could be observed throughout the replicas (see Supplementary Figures 4, right panels, and Supplementary Figure 5).

Riboflavin recognition site analysis

The medoid conformations of the most populated clusters obtained by the equilibration via MDs were extracted, energy minimized and used as receptors for molecular docking of riboflavin using Glide. The docking score is calculated by an empirical scoring function, approximating the binding free energy. The top scoring poses are shown in Figure 4, while their docking scores are reported in Table 3.

The order of magnitude of the predicted affinity for the obtained poses can be correlated to known dissociation constants found in literature for SLC transporters[49–51]. The difference between RFVT3 and the other two transporters can be attributed to the conformation of the model in the specific frame of the MD simulation used for molecular docking. Therefore, to obtain more reliable data and to better characterize the stability of the interactions between riboflavin and the transporters, three replicas of MD simulations 200 ns long were performed using the top scoring poses of each of the docked complexes using Desmond.

A further investigation about the reliability of the docking poses was performed by simulating one molecule of riboflavin alone in SPC solvent and comparing the conformations assumed to the ones generated by the docking algorithm, as well as the ones assumed during the MD simulations. The results (Supplementary Figure 6) showed that the docking poses and the behavior of riboflavin inside of the transporter was consistent with behavior in physiological simulated conditions.

The most significant interactions between riboflavin and the transporters as well as the average interaction energies are reported in Figure 3A. In particular, interactions were considered significant if they were preserved for more than 80% of the total time throughout the replicas.

In the docking performed on the model of RFVT1, riboflavin interacts the most with N28 *via* a hydrogen bond. There are two other notable hydrogen bonds with Q301 and N324, although they are not preserved long enough to be considered relevant. Both the protein and the ligand show good stability, according to their energy profiles, that are around –66.6 kcal/mol with a standard error of 6.7, and constant throughout the MD runs.

Regarding the best pose obtained with the molecular docking on RFVT2, the most impacting interactions were between the solute and Q298, K390, and S387, in all three cases with the formation of hydrogen bonds stabilizing the complex. Hydrogen bonds were also formed with G317, N291, and N321, although below the chosen significancy threshold. The interaction energy (the sum between the Coulomb and VdW interaction energies) between the transporter and riboflavin averages –100.7 kcal/mol with a standard error of 3.7.

During the simulation of the RFVT3::riboflavin complex, the most significant interactions throughout the replicas were the ones between riboflavin and N21 and W17, both via hydrogen bonds. There were few other hydrogen bonds, such as the ones with K414 and N345, but, again, below the significancy threshold. In these complexes, the average interaction energy was –106.4 kcal/mol with a standard error of 2.5. All identified interactions and corresponding significancy are reported in Supplementary Table 2.

The complexes demonstrated good stability throughout the simulations, without major conformational changes in both the protein and the solute, aside from initial

rearrangements. Also, the solute remained around its initial position in all simulations, with only minor oscillations.

Effects of W31S::RFVT2 and N21S::RFVT3 on the interaction with riboflavin

After the identification of the most significant interactions between riboflavin and the transporters, key mutations known to affect riboflavin transport were introduced. The introduced mutations were W31S for RFVT2 [52], and N21S for RFVT3 [53]. These mutations are known to significantly decrease the transport rate of riboflavin, without affecting the expression of these transporters on the cell membrane[54]. One hypothesis is, therefore, that the mutations directly influence the riboflavin recognition.[54]

The mentioned mutations were initially introduced in the complexes obtained by molecular docking using Schrödinger's Biologics suite, in order to estimate Δstability and Δaffinity values between wild-type and mutated complexes.

Table 4 shows that the two studied mutations produced a decrease in stability and affinity of the complexes of W31S_RFVT2::riboflavin and N21S_RFVT3::riboflavin. In particular, the energies related to both the stability of the model and the affinity between the protein and riboflavin were higher, when compared to the wild-type complexes. This means that the mutants are less stable, and that the affinity of riboflavin is lower than in wild-type complexes. These results suggest that, as expected according to literature data[15,52–54], these mutations can destabilize the transporter::solute complex.

To verify the impact of the mutations on the transport mechanism of RFVT2 and RFVT3, we introduced the same changes in the medoids obtained from the equilibration MD of the 3D models. After introducing the mutations, the structures were energy minimized and a molecular docking of riboflavin was performed using the same site and the same procedure of the docking on the wild-type form. Also in this case, the top-ranking poses were submitted to three MD simulation replicas 200 ns long, to compare the interactions after the introduction of the mutations, like in the wild-type complexes simulations.

The interaction energy profiles of the complexes were then calculated. Overall, during the simulations of the W31S_RFVT2 transporter with riboflavin, the average interaction energy between the transporter and the transported solute was –84.9 kcal/mol with a standard error of 3.8, which is ~15 kcal/mol higher than during the simulations of the wt_RFVT2::riboflavin complex. Similarly, the N21S_RFVT3::riboflavin complexes had an average of –97.7 kcal/mol with a standard error of 3.1, ~10 kcal/mol higher than the wild type complex.

During the molecular dynamics of the W31S_RFVT2::riboflavin complexes, none of the interactions between the transporter and the solute were preserved for more than the predetermined threshold time value (80%). The most preserved interactions were a hydrogen bond with K390 (around 79% of the overall time), and another one with Q298 (68% of the time) while all other hydrogen bonds were maintained for much less time (see Supplementary Table 3).

Similarly, only one interaction is close to be considered significant throughout the simulations of the N21S_RFVT3::riboflavin complexes, namely a hydrogen bond between the solute and E145 that occurs for 76% of the overall time.

According to the obtained data, the introduced mutations have a significant role in the stabilization of the complex and of the interactions between the transporters and riboflavin. This is a possible explanation for the decreased riboflavin transport observed in the mutated proteins and suggests that the models generated and refined are reliable to perform further studies.

Analysis of water molecules during MD simulations

Water molecules are known to play a role in the recognition and stabilization of ligands by proteins[55]. Therefore, we analysed how water could influence the interactions between RFVTs and riboflavin. An analysis of the water-mediated hydrogen bonds (also called water bridging) was performed using MDAnalysis. However, no water bridges were detected. The number of molecules in the riboflavin recognition site was also calculated during MD simulations with a distance threshold of 5 Å from the residues identified as pivotal for the binding and reported in the Supplementary Table 2. The average number of water molecules was calculated for the replicas. Then, the difference between the number of water molecules identified in the protein::RF complexes and those in the systems with protein alone was computed (Figure 5A). 3D-RISM on the medoids of the MD simulations was also performed, and the results confirmed the ones obtained by the dynamic analysis of the number of water molecules (Figure 5B)

As a result, all three systems show a significant decrease in the number of molecules when proteins are in complex with RF, suggesting that water displacement by ligand may have a role in the stabilization of the complex.

Additionally, a solvent analysis was performed to discuss the inward/outward-facing conformations of the models. Indeed, as these models were not generated by homology modeling, their conformation was not known a priori. Interestingly, all models were in one of the many intermediate conformations between the extreme outward and inward facing ones, but the models of RFVT1 and RFVT3 displayed a larger number of water molecules in the intracellular side of the translocation funnel, suggesting a conformation closer to the inward facing, which was also observed in the replicas of the complexes with riboflavin (see Figure 6). On the other hand, the model of RFVT2 showed a much lower difference between the two openings and an overall lower number of water molecules, suggesting an intermediate closed conformation. The difference between the extracellular and intracellular openings of RFVT2 significantly increased during the MD simulations of the complex with riboflavin both in the wild-type and mutated models (Figures 7-8). Future studies will be focused on exploring more conformations of the transporters, including a simulation of the whole transport, using non-classical MD techniques.

CONCLUSIONS

In this work, multiple 3D models of RFVT1, RFVT2 and RFVT3 were generated thanks to the most recent advances in AI-based 3D structure predictions. The reliability of the models was then supported by the study of the interactions between riboflavin and the three SLC52 transporters. Molecular docking and the following MD simulations helped us to hypothesize the molecular recognition mechanism between each transporter and the solute. On the other hand, the mutated structures showed a significantly lower stability of the complexes, that is associated to a more positive interaction energy. This further corroborates the reliability of the models, as well as supports the hypothesis that the mutations W31S in RFVT2 and N21S in RFVT3 have a direct effect on the decrease riboflavin transport rate by impairing the riboflavin recognition. In future studies, the whole riboflavin transport could be simulated to identify the key residues along the substrate transport path. In particular, the knowledge of these important residues for the recognition and transport of the solutes will also help the development of therapeutics since these transporters are highly expressed in several types of cancer. According to this perspective, two strategies could be devised: either developing drugs that prevent the transport of riboflavin into the cancer cells, causing a depletion of FMN and FAD which would block cell growth and replication, or riboflavin-like anti-cancer drugs that exploit the RFVT transporters to specifically target the cancer cells overexpressing them.
In this context, the understanding of the 3D structure of these proteins is a key step to identify residues involved in recognition and transport of solutes, to improve the knowledge base about both their structure-function relationships and their biological role.

TABLES

Sequence Identity (%)	RFVT1	RFVT2	RFVT3
RFVT1	100	86	41
RFVT2	86	100	41
RFVT3	43	43	100

Table 1. Value of sequence identity between the members of SLC52 family. The matrix was obtained by pairwise sequence alignment performed by MOE 2020.09 software[56]. It is not symmetrical because the lengths of the primary structures of the three transporters are different.

Transporter	Mutation	Effects
SLC52A2/RFVT2	W31S	Km is increased and Vmax is almost unaffected.[15] Transport of riboflavin is almost completely abolished. [52]
SLC52A3/RFVT3	N21S	Drastic reduction in riboflavin transport.[53]

Table 2. Mutations introduced and related effects observed in literature.

Transporter	Docking score [kcal/mol]	Dissociation constant (K_d)
RFVT1	-10.8	1.34×10^{-8} M
RFVT2	-10.1	4.34×10^{-8} M
RFVT3	-6.4	2.16×10^{-5} M

Table 3. Docking scores relative to the top-scoring docking poses for each transporter.

Mutation-Transporter	Δstability [kcal/mol]	Δaffinity [kcal/mol]
W31S-RFVT2	+14.1	+2.4
N21S-RFVT3	+11.8	+2.7

Table 4. Changes in stability and affinity after the introduction of mutations in RFVT2 and RFVT3.

FIGURES

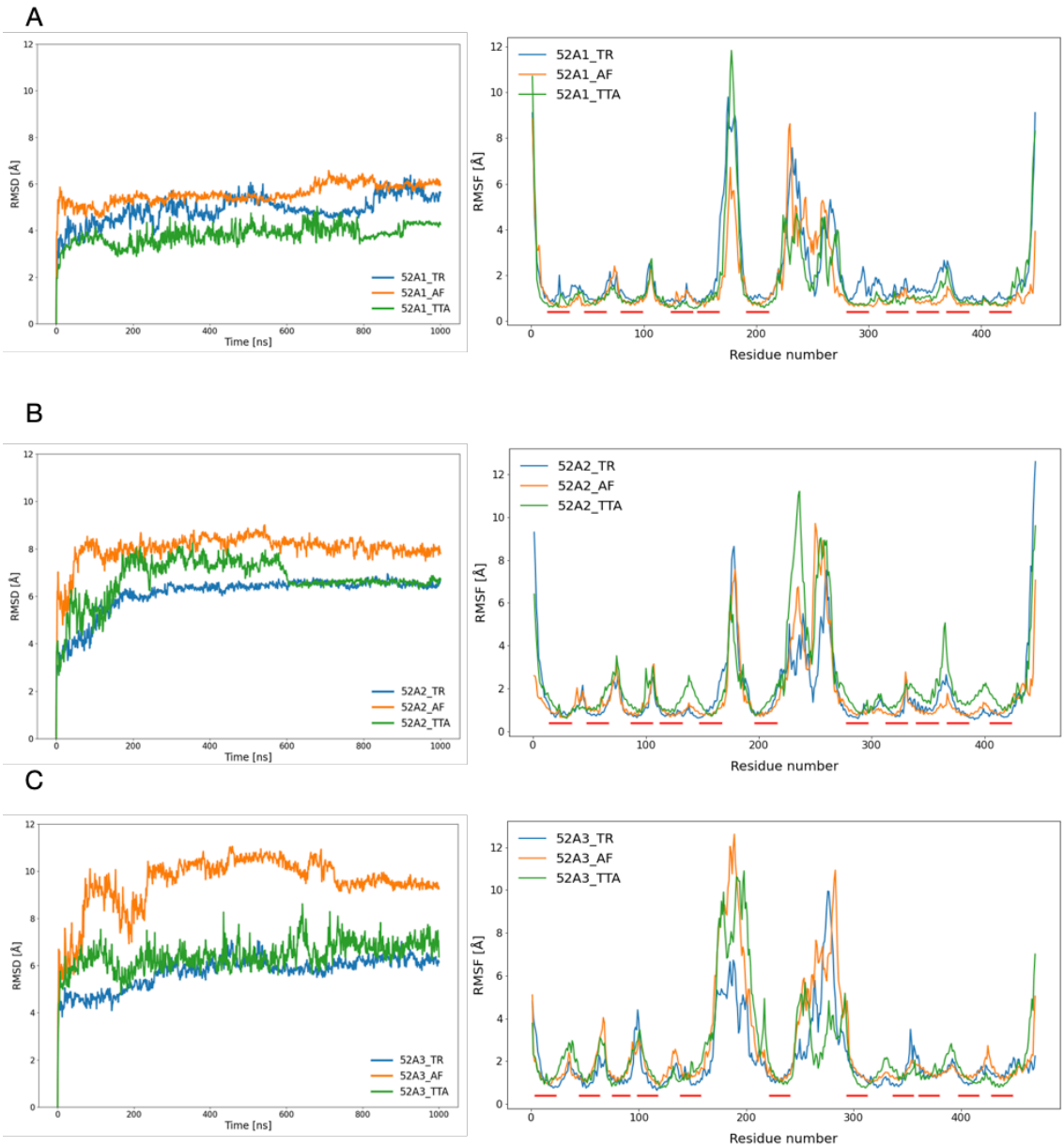


Figure 1. RMSD and RMSF profiles of each model of RFVTs during the MD simulations. for each of the three models generated for A) RFVT1 B) RFVT2, and C) RFVT3, respectively. In all graphs, the model generated by RosettaTR is in blue, the one generated by AlphaFold is in orange, and the one generated by RoseTTAFold is in green. The red lines in the RMSF plots identify the predicted transmembrane helices as found in UniProt. As expected from reliable models, the residues belonging to alpha helices are characterized by lower values of RMSF, while loops are much more flexible in all models of all proteins.

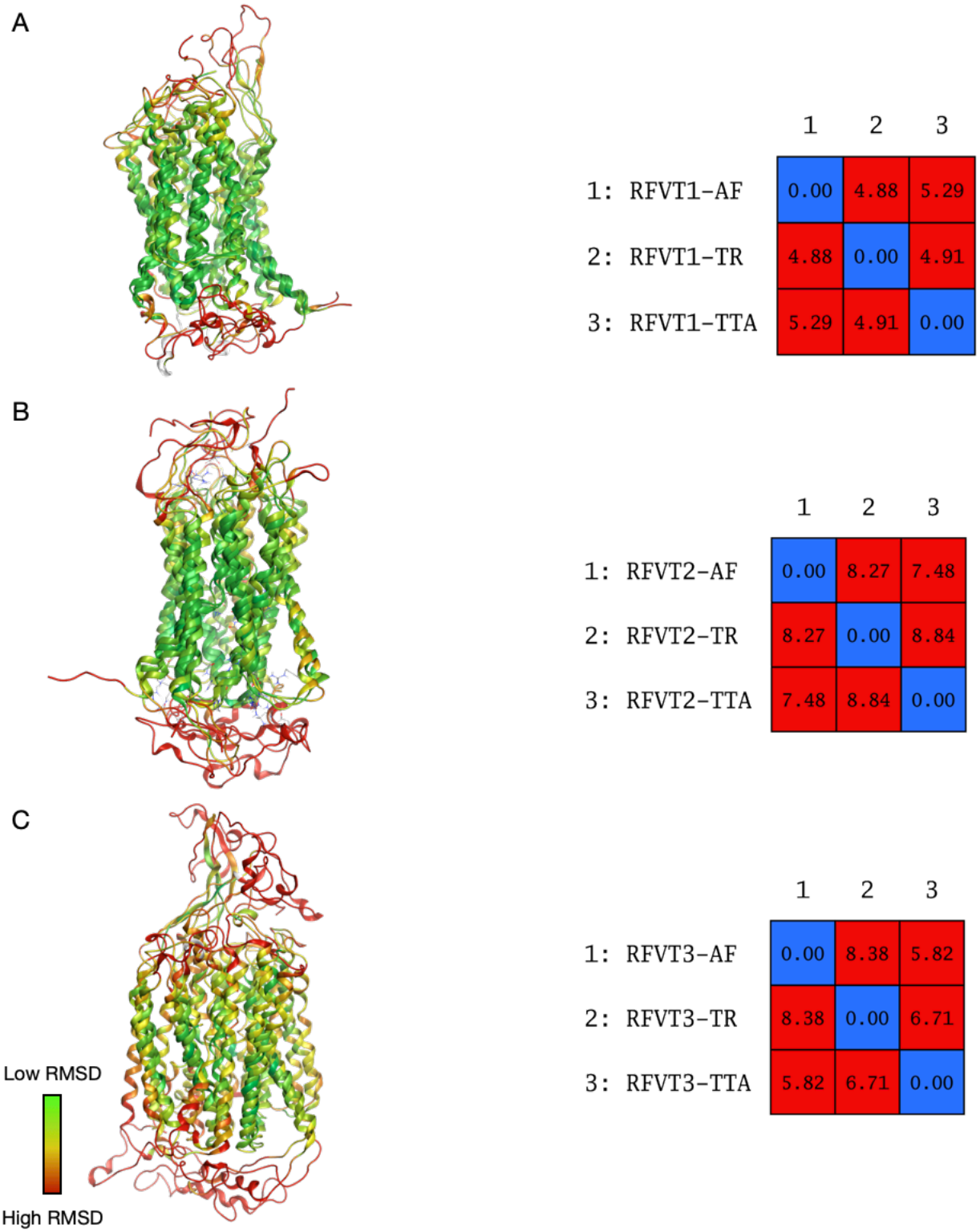


Figure 2: Structural superposition of medoids of the most populated cluster and corresponding RMSD matrix. A) RFVT1, B) RFVT2, and C) RFVT3. The superposed structures are colored by RMSD, green being the lowest values, red the highest. It is easy to see that the transmembrane helices are predicted very similarly by all three software for all three transporters, while the intracellular and extracellular loops show much more variance.

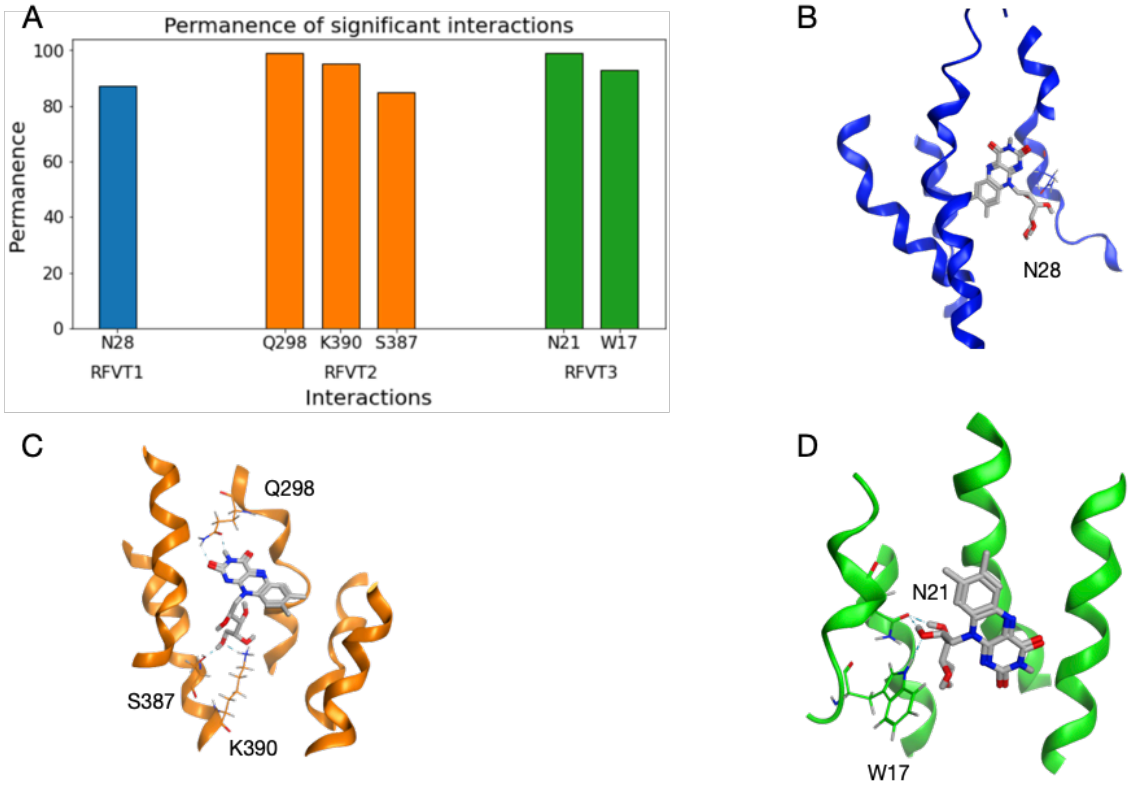


Figure 3. A) Bar graph of the most significant interactions for each transporter throughout the MDs replicas. The medoids of the most populated clusters for each transporter are reported in B) for RFVT1, C) for RFVT2, D) for RFVT3. Although several interactions are identified in all docking procedures, only few were considered relevant, in particular, one for RFVT1, three for RFVT2, and two for RFVT3. All relevant interactions were characterized by the formation of hydrogen bonds.

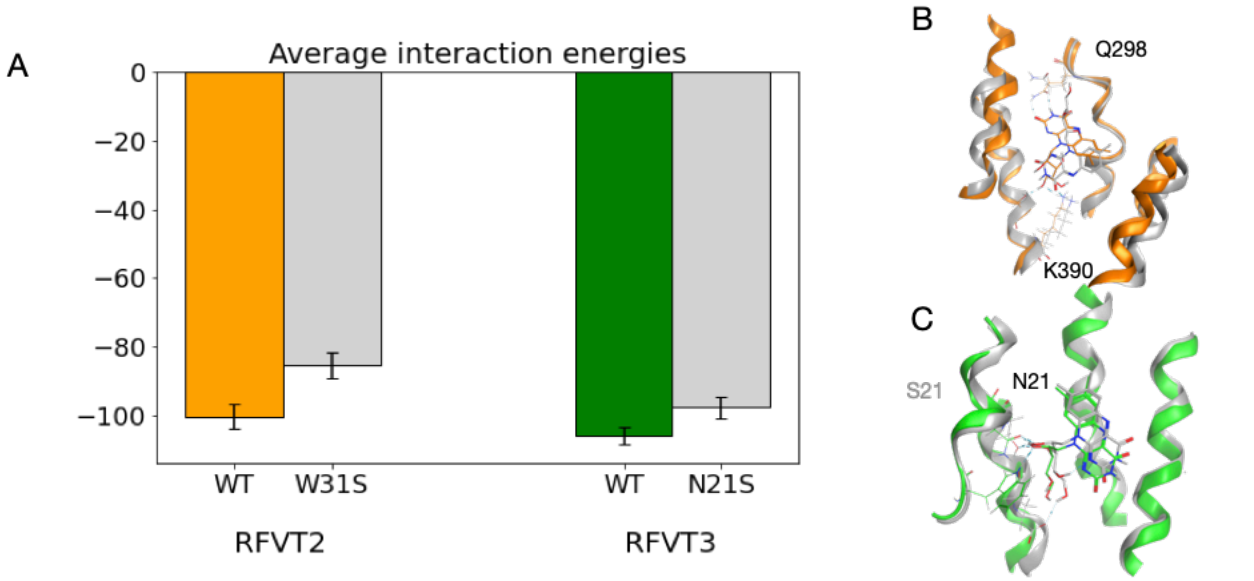
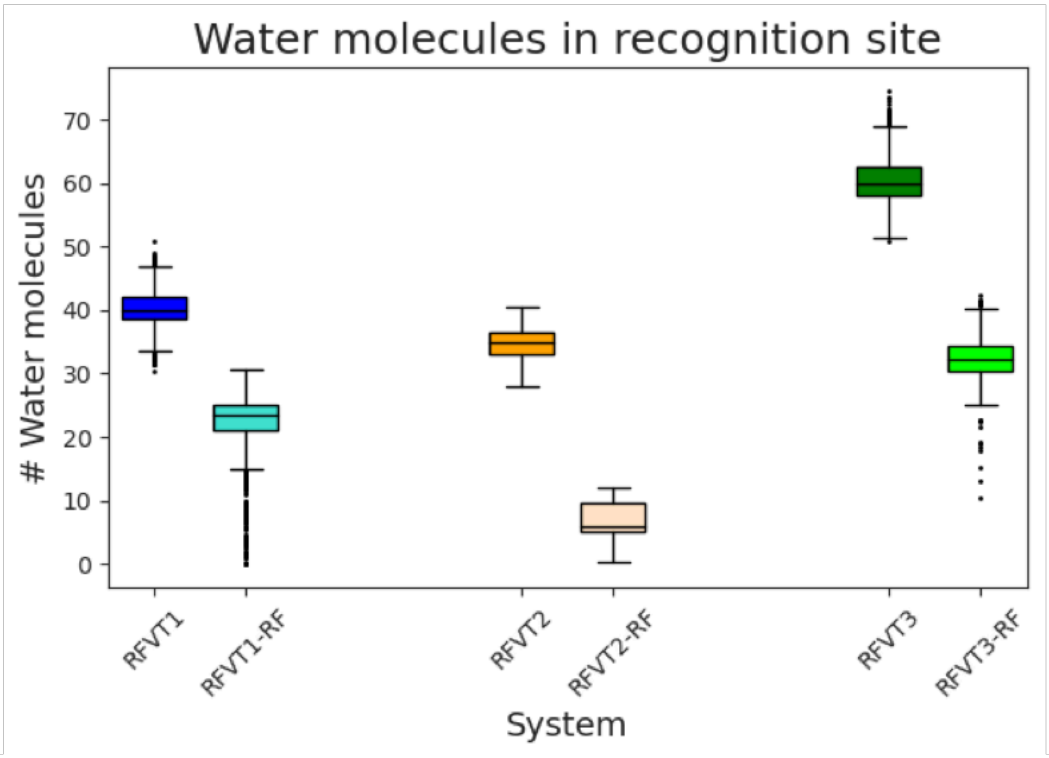


Figure 4. A) Bar graph reporting the average interaction energies between both wild type and mutated RFVT2 and RFVT3 with riboflavin. Superposition of the poses of the wild-type and mutated B) RFVT2 and C) RFVT3. Residue names in black are interacting residues found only in the wild-type model or in both the wild type and the mutated model, while in grey, the residues interacting only in the mutated protein MD. There is a clear increase in interaction energy due to the introduced mutations.

A



B

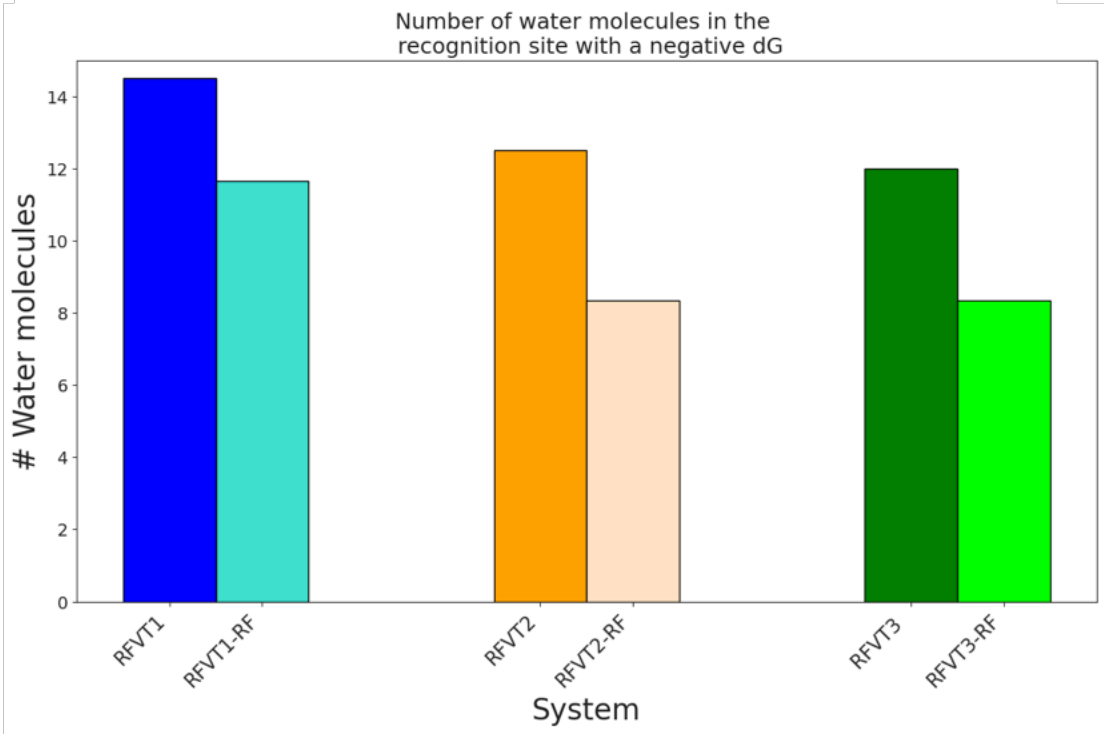


Figure 5. A) Difference between the number of molecules in the riboflavin recognition site with and without the solute. In all simulations, the number of water molecules is much lower in the complex than in the model alone. This result suggests that water displacement may have a role in the stabilization of the complex. B) Number of water molecules with a negative dG obtained with the 3D-RISM analysis. This result confirms the decrease in the number of water molecules in the recognition site if RF is present

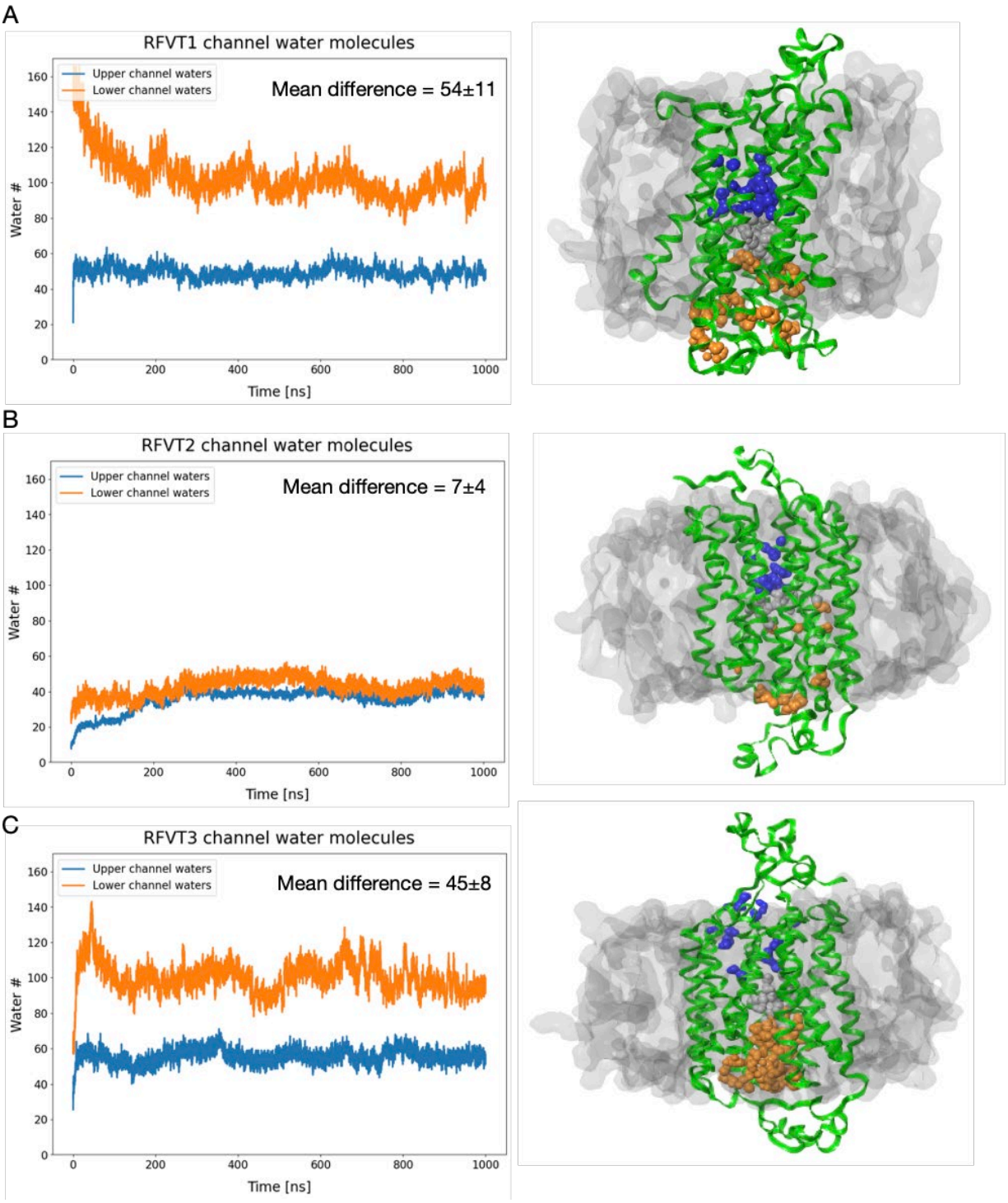


Figure 6. Comparison of extracellular and intracellular openings of the apo models of A) RFVT1, B) RFVT2, C) RFVT3. RFVT1 and RFVT3 exhibit a meaningful opening toward the intracellular side of the membrane (inward facing), while RFVT2 appears to be in a closed conformation, one of the many between the two extremes. On the right, water molecules considered to be towards the extracellular opening have been colored blue, the ones towards the intracellular opening orange, and the ones in the center grey.

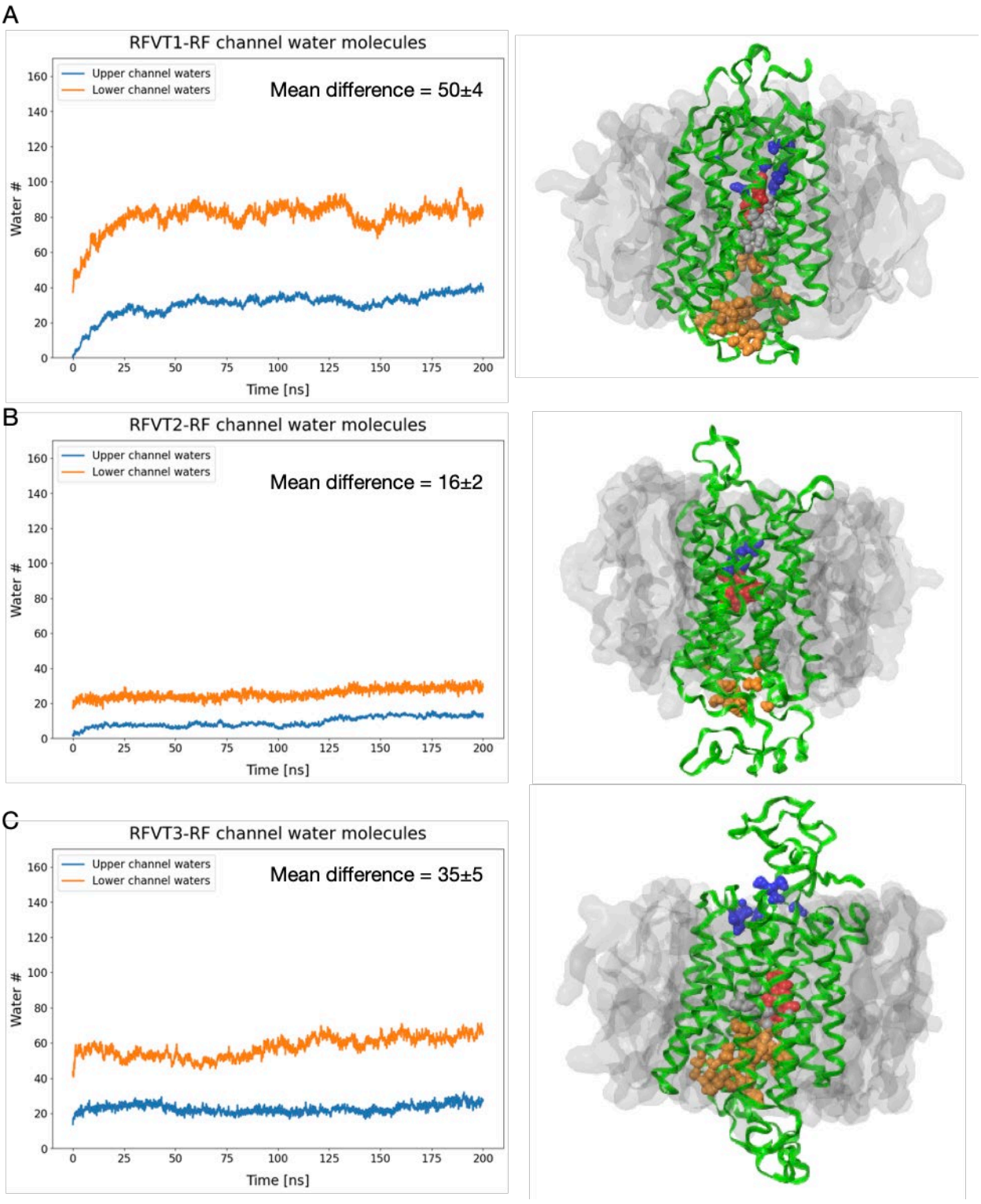


Figure 7. Comparison of extracellular and intracellular openings of the models of A) RFVT1, B) RFVT2, C) RFVT3 in complex with riboflavin. On the right, water molecules considered to be towards the extracellular opening have been colored blue, the ones towards the intracellular opening orange, and the ones in the center grey. Riboflavin has been colored in red.

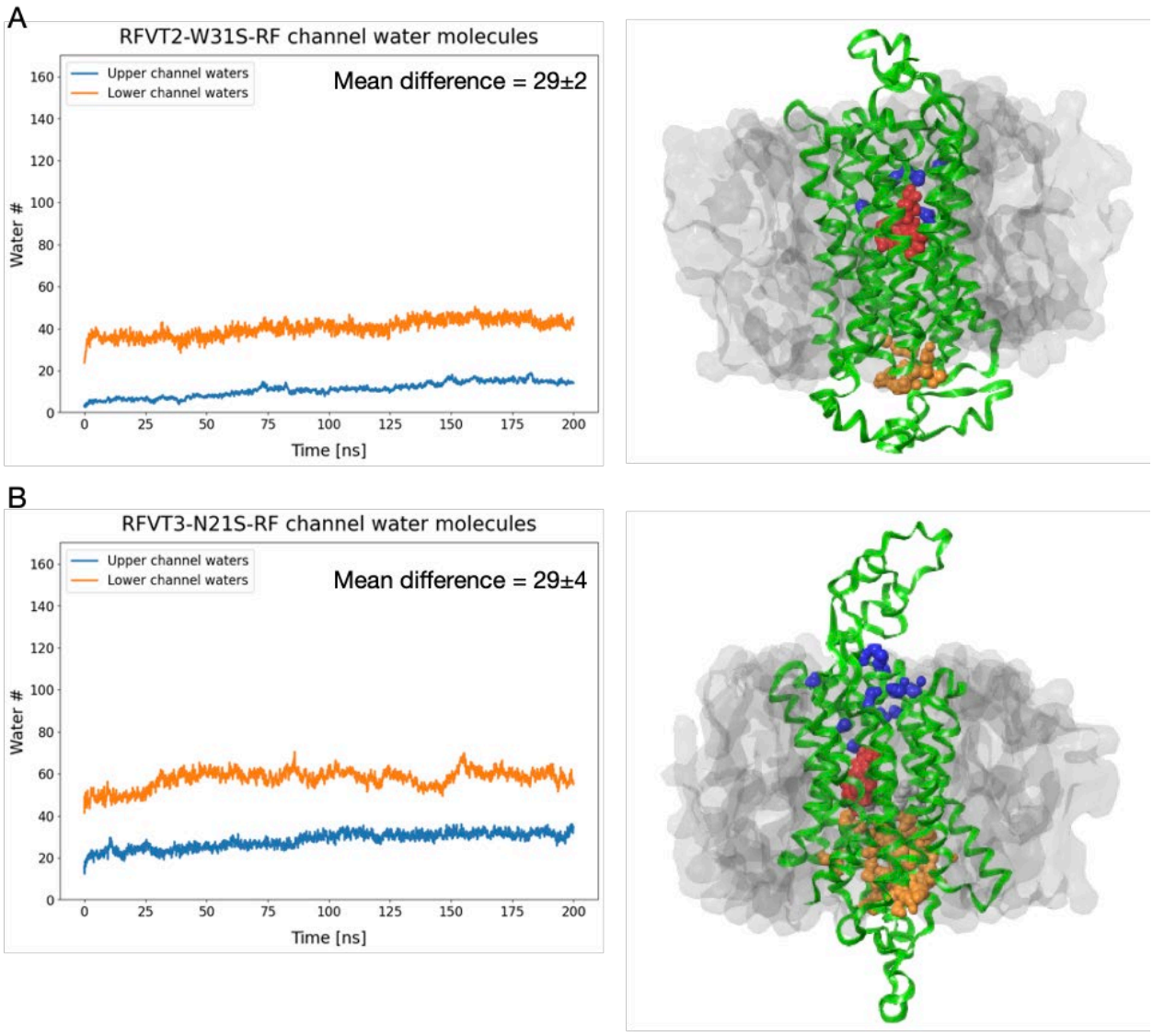


Figure 8. Comparison of extracellular and intracellular openings of the mutated models of A) RFVT2, B) RFVT3 in complex with riboflavin. On the right, water molecules considered to be towards the extracellular opening have been colored blue, the ones towards the intracellular opening orange, and the ones in the center grey. Riboflavin has been colored in red.

BIBLIOGRAPHY

1. Thakur K, Tomar SK, Singh AK, Mandal S, Arora S. Riboflavin and health: A review of recent human research. *Crit Rev Food Sci Nutr*. 2017;57(17):3650-3660. doi:10.1080/10408398.2016.1145104
2. Barile M, Giancaspero TA, Leone P, Galluccio M, Indiveri C. Riboflavin transport and metabolism in humans. *J Inherit Metab Dis*. 2016;39(4):545-557. doi:10.1007/s10545-016-9950-0
3. Pinto JT, Zempleni J. Riboflavin. *Advances in Nutrition*. 2016;7(5):973-975. doi:10.3945/an.116.012716
4. Yoshii K, Hosomi K, Sawane K, Kunisawa J. Metabolism of dietary and microbial vitamin b family in the regulation of host immunity. *Front Nutr*. 2019;6. doi:10.3389/fnut.2019.00048
5. Lin L, Yee SW, Kim RB, Giacomini KM. SLC transporters as therapeutic targets: Emerging opportunities. *Nat Rev Drug Discov*. 2015;14(8):543-560. doi:10.1038/nrd4626
6. Bartmann L, Schumacher D, von Stillfried S, et al. Evaluation of riboflavin transporters as targets for drug delivery and theranostics. *Front Pharmacol*. 2019;10(FEB). doi:10.3389/fphar.2019.00079
7. Subramanian VS, Subramanya SB, Rapp L, Marchant JS, Ma TY, Said HM. Differential expression of human riboflavin transporters -1, -2, and -3 in polarized epithelia: A key role for hRFT-2 in intestinal riboflavin uptake. *Biochim Biophys Acta Biomembr*. 2011;1808(12):3016-3021. doi:10.1016/j.bbamem.2011.08.004
8. Tolomeo M, Nisco A, Leone P, Barile M. Development of novel experimental models to study flavoproteome alterations in human neuromuscular diseases: The effect of rf therapy. *Int J Mol Sci*. 2020;21(15):1-32. doi:10.3390/ijms21155310
9. Yonezawa A, Inui KI. Novel riboflavin transporter family RFVT/SLC52: Identification, nomenclature, functional characterization and genetic diseases of RFVT/SLC52. *Mol Aspects Med*. 2013;34(2-3):693-701. doi:10.1016/j.mam.2012.07.014
10. Console L, Tolomeo M, Colella M, Barile M, Indiveri C. Reconstitution in proteoliposomes of the recombinant human riboflavin transporter 2 (SLC52A2) overexpressed in E. coli. *Int J Mol Sci*. 2019;20(18). doi:10.3390/ijms20184416
11. Yao Y, Yonezawa A, Yoshimatsu H, Masuda S, Katsura T, Inui KI. Identification and comparative functional characterization of a new human riboflavin transporter hRFT3 expressed in the brain. *Journal of Nutrition*. 2010;140(7):1220-1226. doi:10.3945/jn.110.122911
12. Yoshimatsu H, Yonezawa A, Yao Y, et al. CALL FOR PAPERS Innovative and Emerging Technologies in GI Physiology and Disease Functional involvement of RFVT3/SLC52A3 in intestinal riboflavin absorption. *Am J Physiol Gastroin-test Liver Physiol*. 2014;306:102-110. doi:10.1152/ajpgi.00349.2013.-Riboflavin
13. Kubo Y, Miki S, Akanuma S ichi, Hosoya K ichi. Riboflavin transport mediated by riboflavin transporters (RFVTs/SLC52A) at the rat outer blood-retinal barrier. *Drug Metab Pharmacokinet*. 2019;34(6):380-386. doi:10.1016/j.dmpk.2019.08.002

14. Kubo Y, Yahata S, Miki S, Akanuma S ichi, Hosoya K ichi. Blood-to-retina transport of riboflavin via RFVTs at the inner blood-retinal barrier. *Drug Metab Pharmacokinet*. 2017;32(1):92-99. doi:10.1016/j.dmpk.2016.09.006
15. Console L, Tolomeo M, Cosco J, Massey K, Barile M, Indiveri C. Impact of natural mutations on the riboflavin transporter 2 and their relevance to human riboflavin transporter deficiency 2. In: *IUBMB Life*. John Wiley and Sons Inc; 2021. doi:10.1002/iub.2541
16. SLC52 Variants Database - Cure RTD. Accessed June 23, 2022. http://curertd.org/research/slc52variants/
17. Kenyon J, Liu W, Dalgleish A. Report of objective clinical responses of cancer patients to pharmaceutical-grade synthetic cannabidiol. *Anticancer Res*. 2018;38(10):5831-5835. doi:10.21873/anticanres.12924
18. Darguzyte M, Drude N, Lammers T, Kiessling F. Riboflavin-targeted drug delivery. *Cancers (Basel)*. 2020;12(2). doi:10.3390/cancers12020295
19. Drew D, North RA, Nagarathinam K, Tanabe M. Structures and General Transport Mechanisms by the Major Facilitator Superfamily (MFS). *Chem Rev*. 2021;121(9):5289-5335. doi:10.1021/acs.chemrev.0c00983
20. Dilly S, Garnier M, Solé M, Bailly R, Taib N, Bestel I. In Silico Identification of a Key Residue for Substrate Recognition of the Riboflavin Membrane Transporter RFVT3. *J Chem Inf Model*. 2020;60(3):1368-1375. doi:10.1021/acs.jcim.9b01020
21. Yu X, Yang G, Yan C, et al. Dimeric structure of the uracil:proton symporter UraA provides mechanistic insights into the SLC4/23/26 transporters. *Cell Res*. 2017;27(8):1020-1033. doi:10.1038/cr.2017.83
22. Stecula A, Schlessinger A, Giacomini KM, Sali A. Human Concentrative Nucleoside Transporter 3 (hCNT3, SLC28A3) Forms a Cyclic Homotrimer. *Biochemistry*. 2017;56(27):3475-3483. doi:10.1021/acs.biochem.7b00339
23. Scalise M, Galluccio M, Console L, Pochini L, Indiveri C. The human SLC7A5 (LAT1): The intriguing histidine/large neutral amino acid transporter and its relevance to human health. *Front Chem*. 2018;6(JUN). doi:10.3389/fchem.2018.00243
24. Palazzolo L, Parravicini C, Laurenzi T, et al. In silico description of LAT1 transport mechanism at an atomistic level. *Front Chem*. 2018;6(AUG). doi:10.3389/fchem.2018.00350
25. Napolitano L, Galluccio M, Scalise M, et al. Novel insights into the transport mechanism of the human amino acid transporter LAT1 (SLC7A5). Probing critical residues for substrate translocation. *Biochim Biophys Acta Gen Subj*. 2017;1861(4):727-736. doi:10.1016/j.bbagen.2017.01.013
26. Palazzolo L, Paravicini C, Laurenzi T, et al. SLC6A14, a Pivotal Actor on Cancer Stage: When Function Meets Structure. *SLAS Discovery*. 2019;24(9):928-938. doi:10.1177/2472555219867317
27. Napolitano L, Scalise M, Koyioni M, et al. Potent inhibitors of human LAT1 (SLC7A5) transporter based on dithiazole and dithiazine compounds for development of anticancer drugs. *Biochem Pharmacol*. 2017;143:39-52. doi:10.1016/j.bcp.2017.07.006
28. Jumper J, Evans R, Pritzel A, et al. Highly accurate protein structure prediction with AlphaFold. *Nature*. 2021;596(7873):583-589. doi:10.1038/s41586-021-03819-2
29. Baek M, DiMaio F, Anishchenko I, et al. Accurate prediction of protein structures and interactions using a three-track neural network. *Science (1979)*. 2021;373(6557):871-876. doi:10.1126/science.abj8754

30. Kryshtafovych A, Schwede T, Topf M, Fidelis K, Moult J. Critical assessment of methods of protein structure prediction (CASP)—Round XIV. *Proteins: Structure, Function and Bioinformatics*. 2021;89(12):1607-1617. doi:10.1002/prot.26237
31. Krogh A, Larsson B, von Heijne G, Sonnhammer ELL. Predicting transmembrane protein topology with a hidden Markov model: Application to complete genomes. *J Mol Biol*. 2001;305(3):567-580. doi:10.1006/jmbi.2000.4315
32. Mcguffin LJ, Bryson K, Jones DT. *The PSIPRED Protein Structure Prediction Server*. Vol 16.; 2000. http://globin.bio.warwick.ac.uk/psipred/
33. Drozdetskiy A, Cole C, Procter J, Barton GJ. JPred4: A protein secondary structure prediction server. *Nucleic Acids Res*. 2015;43(W1):W389-W394. doi:10.1093/nar/gkv332
34. Wang S, Li W, Liu S, Xu J. RaptorX-Property: a web server for protein structure property prediction. *Nucleic Acids Res*. 2016;44(W1):W430-W435. doi:10.1093/nar/gkw306
35. Designed Research; J DBY. Improved protein structure prediction using predicted interresidue orientations. 2020;117(3):1496-1503. doi:10.1073/pnas.1914677117/-/DCSupplemental
36. Bowers KJ, Bowers KJ, Chow E, et al. Scalable algorithms for molecular dynamics simulations on commodity clusters. *IN SC '06: PROCEEDINGS OF THE 2006 ACM/IEEE CONFERENCE ON SUPERCOMPUTING*. Published online 2006. Accessed June 23, 2022. http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.98.2121
37. Lomize MA, Pogozheva ID, Joo H, Mosberg HI, Lomize AL. OPM database and PPM web server: Resources for positioning of proteins in membranes. *Nucleic Acids Res*. 2012;40(D1). doi:10.1093/nar/gkr703
38. Lu C, Wu C, Ghoreishi D, et al. OPLS4: Improving force field accuracy on challenging regimes of chemical space. *J Chem Theory Comput*. 2021;17(7):4291-4300. doi:10.1021/acs.jctc.1c00302
39. Daura X, Gademann K, Jaun B, et al. *Peptide Folding: When Simulation Meets Experiment*. Vol 31.; 1998.
40. Michaud-Agrawal N, Denning EJ, Woolf TB, Beckstein O. MDAnalysis: A toolkit for the analysis of molecular dynamics simulations. *J Comput Chem*. 2011;32(10):2319-2327. doi:10.1002/jcc.21787
41. Gowers RJ, Linke M, Barnoud J, et al. *MDAnalysis: A Python Package for the Rapid Analysis of Molecular Dynamics Simulations*.; 2016. http://mdanalysis.org
42. Saporiti S, Parravicini C, Pergola C, et al. IgG1 conformational behavior: elucidation of the N-glycosylation role via molecular dynamics. *Biophys J*. 2021;120(23):5355-5370. doi:10.1016/j.bpj.2021.10.026
43. Halgren TA, Murphy RB, Friesner RA, et al. Glide: A New Approach for Rapid, Accurate Docking and Scoring. 2. Enrichment Factors in Database Screening. *J Med Chem*. 2004;47(7):1750-1759. doi:10.1021/jm030644s
44. Hastings J, Owen G, Dekker A, et al. ChEBI in 2016: Improved services and an expanding collection of metabolites. *Nucleic Acids Res*. 2016;44(D1):D1214-D1219. doi:10.1093/nar/gkv1031
45. Altschup SF, Gish W, Miller W, Myers EW, Lipman DJ. *Basic Local Alignment Search Tool*. Vol 215.; 1990.
46. Taslimifar M, Oparija L, Verrey F, Kurtcuoglu V, Olgac U, Makrides V. Quantifying the relative contributions of different solute carriers to aggregate substrate transport. *Sci Rep*. 2017;7. doi:10.1038/srep40628

47. Burtscher V, Schicker K, Freissmuth M, Sandtner W. Kinetic models of secondary active transporters. *Int J Mol Sci*. 2019;20(21). doi:10.3390/ijms20215365
48. Schicker K, Bhat S, Farr C, et al. Descriptors of secondary active transporter function and how they relate to partial reactions in the transport cycle. *Membranes (Basel)*. 2021;11(3):1-16. doi:10.3390/membranes11030178
49. Foley AR, Menezes MP, Pandraud A, et al. Treatable childhood neuronopathy caused by mutations in riboflavin transporter RFVT2. *Brain*. 2014;137(1):44-56. doi:10.1093/brain/awt315
50. Udhayabanu T, Subramanian VS, Teafatiller T, et al. SLC52A2 [p.P141T] and SLC52A3 [p.N21S] causing Brown-Vialetto-Van Laere Syndrome in an Indian patient: First genetically proven case with mutations in two riboflavin transporters. *Clinica Chimica Acta*. 2016;462:210-214. doi:10.1016/j.cca.2016.09.022
51. O'Callaghan B, Bosch AM, Houlden H. An update on the genetics, clinical presentation, and pathomechanisms of human riboflavin transporter deficiency. *J Inherit Metab Dis*. 2019;42(4):598-607. doi:10.1002/jimd.12053
52. Zsidó BZ, Hetényi C. The role of water in ligand binding. *Curr Opin Struct Biol*. 2021;67:1-8. doi:10.1016/j.sbi.2020.08.002
53. Molecular Operating Environment (MOE). Accessed July 12, 2022. 2022.02 Chemical Computing Group ULC, 1010 Sherbooke St. West, Suite #910, Montreal, QC, Canada, H3A 2R7, 2022.

DATA AVAILABILITY STATEMENT

The data that supports the findings of this study are available in the supplementary material of this article.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflicts of interest.