# In Silico Identification of a Key Residue for Substrate Recognition of the Riboflavin Membrane Transporter RFVT3

Sébastien Dilly, Mélanie Garnier, Marion Solé, Rémi Bailly, Nada Taib,
Isabelle Bestel

▶ **To cite this version:**

## HAL Id: hal-03101452
## https://hal.science/hal-03101452

Submitted on 7 Jan 2021

# In Silico Identification of a Key Residue for Substrate Recognition of the Riboflavin Membrane Transporter RFVT3

Sébastien Dilly, Mélanie Garnier, Marion Sole, Remy Bailly, Nada Taib, and Isabelle Berque-Bestel

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

# *In Silico* Identification of a Key Residue for Substrate Recognition of the Riboflavin Membrane Transporter RFVT3

*Sébastien Dilly,* *Mélanie Garnier, Marion Solé, Rémi Bailly, Nada Taib, Isabelle Bestel*

L'Institut de Chimie et Biologie des Membranes et des Nano-Objets (CBMN), Unité Mixte de Recherche (UMR) 5248, Centre National de la Recherche (CNRS), University of Bordeaux, Pessac, France.

**ABSTRACT**

Due to its specific physicochemical properties (fluorescence, photosensitizing and redox reactions), the vitamin B2, also called riboflavin (RF), has been generating a lot of interest in nanotechnologies and bioengineering for the last decade. RF, by targeting its RFVT transporters overexpressed in some cancers, is particularly used to functionalize nanovectors for anti-cancer drug delivery. From a physiopathological point of view, a RF deficiency has been implicated in various pathologies, including mendelian diseases. RF deficiency is mainly due to natural variants of its RFVT transporters that make them inactive and therefore prevent RF transport. The lack of structural data about RFVT is a major drawback for a better understanding of the role of the mutations in the molecular mechanism of these transporters. In this context, this work was aimed at investigating the 3D structure of RFVT3 and its interactions with RF. For this purpose, we used an *in silico* procedure including protein threading, docking and molecular dynamics. Our results propose that the natural variant W17R, known to be responsible for BVVL syndrome, prevents the recognition of RF by RFVT3 and thus blocks its transport. This *in silico* procedure could be used for elucidating the impact of pathogenic mutations of other proteins.

Moreover, the identification of RF binding site will be useful for the design of RF-functionalized nanovectors.

## 1. INTRODUCTION

Riboflavin (RF) (Figure 1A), also known as vitamin B2, is a water-soluble vitamin essential for normal cellular metabolic activities (energetic process, lipid metabolism).[1] Mammals are not able to synthesize RF *in situ* and should obtain it from diet. In humans, RF internalization is mediated by three membrane transporters, recently identified and named RFVT1-3.[1,2] These proteins, also known as SLC52 transporters, are the last members of the solute carrier SLC transporter superfamily to be characterized.SLC transporters are secondary or indirect active transporters. They do not directly use ATP, but the energy produced by an electrochemical gradient produced by an primary active transporter (e.g. ATP Binding Cassette transporters).[3]RFVT transporters consist of about 450 amino acids and are expressed in the CNS, as well as in peripheral tissues.[1,2] From a therapeutic point of view, these transporters can be targeted for two pathologies: mendelian diseases and cancer.[1,4]

Mendelian diseases are monogenic and rare diseases due to the mutation of a single gene whose impact on the function of some proteins remains unknown, thus preventing the development of therapeutic solutions. Previous studies have shown that mutations in the RF transporter genes SLC52A1-3 (coding for RFVT1, RFVT2, RFVT3) make the transporters inactive (example of RFVT3 mutations in Figure 1B) and cause mendelian diseases such as Brown-Vialetto-Van Laere (BVVL) syndrome and glutaric aciduria type 2.[1] Despite these findings, RFVT transporters are still not completely characterized and the exact role of the pathogenic mutations in RF transport is not known due to the lack of structural data about RFVT transporters.
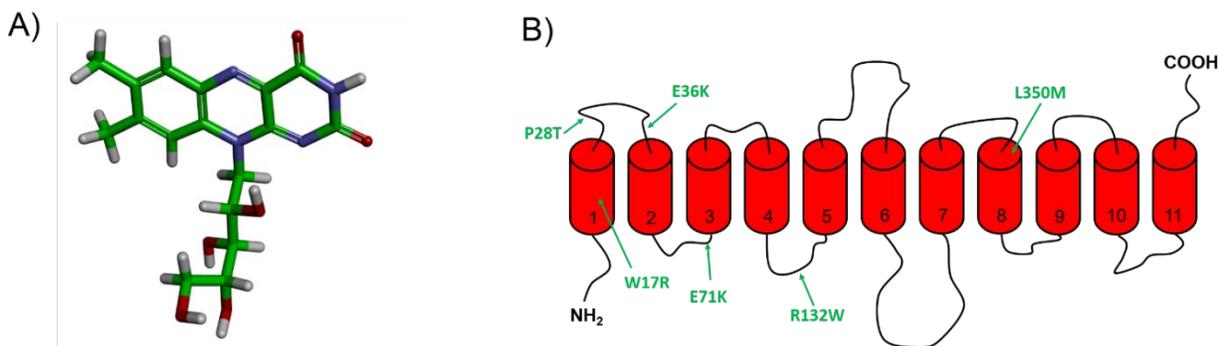
**Figure 1.** A) 3D structure of riboflavin (C, H, O et N atoms are coloured green, white, red and blue, respectively). B) Schematic representation of the RFVT3 transporter. Pathogenic mutations are indicated in green.

RFTV transporters can be also targeted for cancer therapy. The main problem with usual anticancer treatments is the absence of selectivity for tumor cells. In this context, the development of nanovectors carrying the drug towards cancer cells appears as an interesting approach to increase therapeutic efficacy and to reduce toxicity.[5] These nanovectors can be polymers, albumin-based particles or liposomes that are functionalized by chemical agents capable to recognize specific biological markers of the cancer cells.[6] RF-based nanovector formulations have been already developed and evaluated in *in vitro* and *in vivo* studies.[4,7] The results indicated a better internalization of functionalized nanovectors in comparison to controls, thus indicating the importance of the RF analogue and the presence of an active and RF-specific membrane transport.[4] According to the database of the Human Protein Atlas,[8] membrane transporters RFVT are over-expressed in breast and prostate cancer cell lines, suggesting that these proteins are involved in the recognition of RF-derived nanosystems.

In the context of the two pathologies described above, the availability of the 3D structure of RFVT transporters appears as a first and important step in the future development of original therapeutic solutions. In the absence of experimental data about the 3D structure of proteins, homology modeling is a classical method for generating protein models. However, in a case like ours, this approach is not appropriate due to the absence of homologous templates whose 3D structure has been resolved experimentally (i.e. with at least 30% of sequence identity with the target). Another modeling approach based on fold recognition, called threading, can overcome this drawback. Like homology modeling, threading is a template-based method. However,

whereas homology modeling treats the template in an alignment as a sequence, threading also considers the template as a structure, and both structure and sequence information extracted from the alignment are used for prediction. Threading relies on the observation that the 3D structures are better preserved than the sequences and that structures are thus a better support for the characterization of protein architectures.[9] Thus, through this method, it is possible to detect a distant homologous template with a similar fold for generating a robust target model. Threading was shown to be a more suitable method than homology modeling for the modeling of many proteins including membrane transporters.[10,11]

The availability of RFVT 3D models will be useful not only to better understand the impact of the mutations in the RF transport, but also to identify the molecular interactions between RFVT and RF for the design of novel nanovectors. In this work, we especially focused on the RFVT3 transporter given that the pathogenic mutations leading to mendelian diseases mainly affect this protein function (Figure 1B).[1,12-14]

## 2. MATERIALS AND METHODS

Protein modeling and docking experiments were performed on a DELL precision 5810 workstation, while molecular dynamics simulations were carried out using our own computer servers and the local supercomputer AVAKAS.[15]

### 2.1. Transporter Modeling

In the absence of experimental data (NMR or X-ray) about the 3D structure of the RFVT3 transporter, it was built by protein threading and according to the following procedure.

#### 2.1.1. Template selection

Templates are generally selected according to their sequence homology with the targeted protein. For this purpose, BLAST program[16] included in the Protein Data Bank (PDB)[17] was used. The human sequence of RFVT3 was extracted from the Universal Protein Resource[18](entry Q9NQ40). BLAST did not detect any known 3D structure with a sequence identity higher than 30%, which is the cutoff for detecting a reliable template.[19] Moreover, the sequencing coverage was very low with percentages lower than 50%. Therefore, another method based on protein folding appeared more suitable to our study: protein threading, also known as fold recognition. Protein threading is aimed at searching the adequacy between the sequence of an unknown structure and a known 3D structure. Indeed, the sequence is "threaded" on each folding of a library of non-redundant protein folds extracted from the PDB. The compatibility of the sequence with a particular folding is measured by statistical potentials or knowledge-based potentials.[20]Protein threading programs are numerous. Those approved by the scientific community and validated by the CASP ("Critical Assessment of Protein Structure Prediction") project[21] were used: I-Tasser,[22] Pcons[23] and RaptorX.[24]

### 2.1.2. Model generation

Following the template selection, the next step was the transfer of a set of constraints derived from template structures to the corresponding residues of the sequence to be modeled using the same programs previously described. Threading programs, such as comparative modeling programs, are often not effective to predict the 3D structure of the loops. Thus, they were refined using the Modeller's Loop Refinement protocol.[25,26]

Since each threading program generated several models, it was thus necessary to detect leads.

First, models showing a structural similarity were grouped in the same family or cluster. An in-house python program was used to automate the clustering analysis and its preliminary steps

(structural alignment, assessment of structural similarity, generation of distance matrix). The structural similarity was evaluated according to the TM-Score.[27] TM-score is aimed at fixing two main problems in traditional metrics such RMSD. Firstly, TM-score evaluates the global fold similarity and is less sensitive to local structural variations. Secondly, the value of TM-score for random structure pairs is independent from length. TM-score values are between 0 and 1 where 1 indicates a perfect match between two structures. Following statistics performed on structures extracted from the PDB, scores below 0.17 corresponds to randomly chosen unrelated proteins whereas scores higher than 0.5 indicate generally the same fold. For our study, we tightened similarity criteria by considering two structures with an identical fold for a score higher than 0.6.

Then, for each generated cluster, the closest model to other models in terms of TM-score was selected to be the lead.

### 2.1.3.  Lead Validation

The resulting leads were validated by checking the geometry and the topology of their 3D structures using the tools RAMPAGE[28] and TOPCONS.[29] Moreover, a special attention was paid to models presenting an outward-open conformation. Indeed, this is the preferred conformation to study RF binding according to the "rocker-switch" mechanism (See the 3.1 section for more details about this mechanism).

### 2.1.4.  Generation of the mutated transporter model W17R

The mutated transporter W17R was built by replacing the appropriate residue in the model previously generated, with the help of Discovery Studio 3.1.[30]

### 2.2. Binding modes of riboflavin

The study of the binding mode of riboflavin for native RFVT3 and mutated RFVT3 transporters required the definition of the binding site. The potential binding site was determined

by blind docking to the whole protein. For this purpose, two docking programs were used: Gold[31] and AutoDock Vina.[32] The docking space was defined visually in order to encompass the proteic model with a 2 Å margin at least, as a sphere for Gold, as a tetragonal box in the case of Vina. For each program, five replicas with the default scoring function (Goldscore[31] and Vina,[32] respectively) were performed to check the convergence of the experiment. Based on the obtained results, showing that all solutions were situated in the same cavity, a site-docking was then performed using the same programs to more precisely identify the RF binding mode. In this docking, five replicas were also applied but on a restricted space, encompassing the cavity where the ligand was blind docked.

### 2.3. Study of the stability of the RF-transporter complexes

For each complex, the stability of RF into its binding sites was studied by molecular dynamics simulations in an explicit membrane/solvent environment using the GROMACS version 5.1.2 package[33] and the GROMOS 53A6 force field.[34] The topology of RF was obtained from the online program Automated Topology Builder version 2.2.[35]

### 2.3.1. Membrane preparation

A DMPC bilayer of 512 lipids was constructed from a 128 DMPC lipid membrane generated by Peter Tieleman.[36] This 128 DMPC bilayer was placed in the presence of water molecules, in a cubic box of 60 Å side. For water, the simple point charge (SPC) model was employed.[37] Energy Minimizations by steepest descent and conjugate gradient were performed. After equilibration, the system has undergone a production of 10 ns. This initial membrane was then duplicated along the x and y axes of an orthonormal frame to obtain the 512 DMPC lipid bilayer. This square membrane was placed in the presence of water molecules in a cubic box of 120 Å of side and has been minimized with the two successive steepest descent and gradient conjugated

methods. The 512 DMPC bilayer was further equilibrated in the constant Number of particle, Volume and Temperature (NVT) ensemble for 4 ns. All production MD simulations were done in the constant Number of particle, Pressure and Temperature (NPT) ensemble with the v_rescale thermostat[38] at 300 K and the Parrinello-Rahman pressure coupling algorithm[39,40] with the pressure set to 1 bar in semi-isotropic conditions. All bonds were constrained to their equilibrium lengths with the linear constraint(LINCS) algorithm. Periodic Boundary conditions were applied. Lennard–Jones interactions and the real space part of electrostatic interactions were cut off at 12 Å and a time step of 2 fs. The particle-mesh Ewald (PME) method was used to compute electrostatic interactions.The resulting lipid bilayer has an area per lipid equal to 61.9 Å². The surface air of a DMPC is 60.6 Å² according the literature,[41] thus indicating this DMPC membrane model is robust enough.

### 2.3.2. Molecular dynamics simulations

The complexes were inserted in the 512 DMPC lipid bilayer previously prepared and centered in a cubic box of edge length 120 Å using the InflateGRO methodology.[42] As that insertion step required the removal of the solvent, the box was refilled with water using the SPC model[37] and the system was then made neutral using the Genion tool of the GROMACS package (salt concentration of 0.15 mol.L$^{-1}$). The resulting system was then subjected to an energy minimization (2000 steps of steepest descent followed 2000 steps of conjugate gradient with short-range electrostatic and VDW cutoff of 12 Å until the maximum force was smaller than 1000 kJ.mol$^{-1}$.nm$^{-1}$) to remove steric clashes and to correct inappropriate geometry. After, a preparation step was applied to equilibrate the molecular environment around the complexes that were restrained to their original position. Equilibration was carried out in two phases using periodic boundary conditions in all directions and a time step of 2 fs. The first was conducted in

a NVT ensemble to reach a temperature of 310 °K using a velocity rescaling thermostat.[38] The second phase was conducted in a NPT ensemble to equilibrate the system under a 1 atm pressure using the Nosé-Hoover thermostat[43,44]and the Parrinello-Rahman barostat.[39,40].All equilibration phases were carried out during 2 ns. Upon completion of the equilibration step, the position restraints of the complexes were released for a 500 ns production.

### 2.3.3. Data analysis

The gmx rms and gmx energy tools of the GROMACS version 5.1.2 package[33] were used to evaluate the displacement of RF into its binding site and the interaction energies between RF and the binding site residues, respectively. The results were visualized using the Pymol Molecular Graphics System version 2.0[45] and VMD.[46]

## 3. RESULTS AND DISCUSSION

### 3.1. Transporter Modeling

Three threading softwares, I-Tasser,[22] Pcons,[23] and RaptorX,[24] were used to build, respectively, 5, 10 and 5 models. The number of generated models for each software is sufficient to obtain a representative and robust panel of the different folds that a protein target can adopt and to avoid duplicate structures. Since models with close structures can be built by the different softwares, a clustering analysis was carried out and revealed 3 groups and 4 outliers (Figure 2).
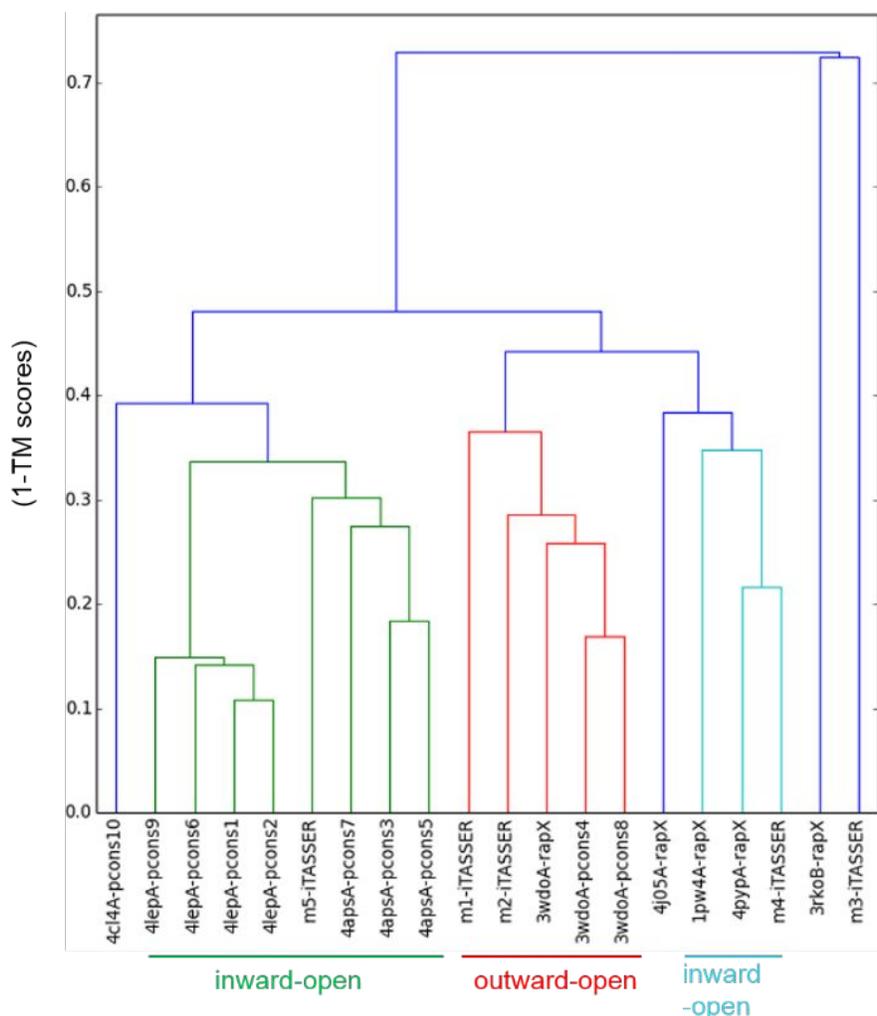
**Figure 2.** Dendrogram of the cluster analysis of RFVT3 models. For practical convenience, the y-axis indicates 1-TM scores. 3 groups (coloured green, red and cyan) and 4 outliers (blue) were revealed

The first group (in green) is composed of 8 models built by I-TASSER and Pcons programs in an inward-open conformation. The 5 models of the second group (in red) generated by the three programs presents an outward-open conformation. The last group (in cyan) possesses 3 models built by the I-TASSER and RaptorX programs in an inward-open conformation. For each group, we selected as the most representative model (lead) the one which is closer to others in terms of TM-score. These 3 models are presented in Figure 3.

| Software | Pcons | Pcons | RaptorX |
|---|---|---|---|
| Template | Proton dependent oligopeptide transporter from E.Coli (PDB code : 4LEP) | YaJR transporter from E.Coli (PDB code: 3WDO) | Human glucose transporter GLUT1 (PDB code: 4PYP) |
| Conformation | Inward-open | Outward-open | Inward-open |

**Figure 3.** Structures and modeling protocols of the 3 generated RFVT3 leads

They were generated by Pcons and RaptorX programs from templates belonging to the Major Facilitator Superfamily (MFS). This transporter family is one of the largest families of membrane transporters found in bacteria as well as in archaea and eukarya. From a structural point of view, most of these transporters are composed of between 10 and 12 transmembrane α helices (TMH) forming two symmetrical domains, called bundles, connected by extracellular or intracellular loops.[47] Two types of symmetrical folding were observed: the 6+6 configuration for most of them and the 5+5 configuration for the others such as the Mhp1 transporter.[48] Additional TMH can be found, but these extra domains are likely to be involved in oligomerization, trafficking or turnover, rather than transport activity.[47] Regarding their transport mechanism, they are characterized by an access of the binding site either from the extracellular side or from the cytoplasm, thus preventing the formation of a continuous pore across the membrane such as in the case of ion channels.[47] In this alternating-access mechanism, described as the "rocker-switch" mechanism, the substrate binds to a site localized at the interface of the two bundles and

approximately in the middle of the membrane, thus leading to a conformational change of the protein and allowing the transport of the substrate (Figure 4). Upon substrate release, the transporters recycle to the extracellular domain facing orientation. These transporters thus oscillate between an outward-open conformation and an inward-open conformation due to the symmetrical movement of the bundles around the binding site.[47,49]
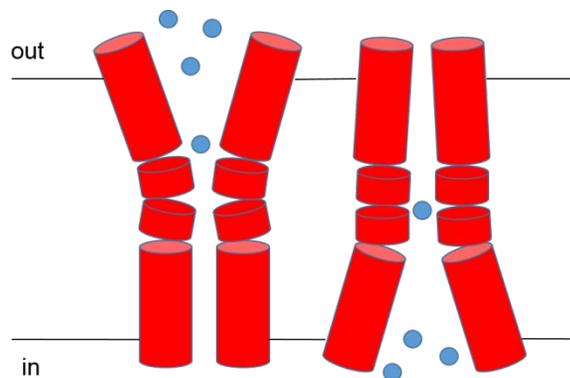


**Figure 4.** Schematic representation of the "rocker-switch" mechanism.

The first model is composed of 12 TMH arranged in a 6+6 configuration with extracellular Nter and Cter regions. The second model presents 11 TMH, suggesting a 5+5 configuration with an extra TMH. The Nter and Cter regions are in the intracellular and extracellular domains, respectively. As the second model, the third model is made of 11 TMH with an intracellular Nter region and an extracellular Cter region. This model is also characterized by the presence of 3 short intracellular helices and a large loop between TMH5 and TMH 6. The first and third models are in inward-open conformation whereas the second model is in an outward-open conformation.

Since an outward-open conformation is preferred when the substrate is recognized by the binding site, we selected the second lead model ("RFVT3-3WDO") generated by PCons to study the interactions between riboflavin and RFTV3. The topology of the model is consistent with

that predicted by the TOPCONS program,[29] the Universal Protein Resource[18] (entry Q9NQ40) and the work of Yamamoto and al.[2] which also predict 11 transmembrane domains with an intracellular Nter region and an extracellular Cter region. Moreover a fairly good superimposition of the TMH is observed between the generated model and the predictions of TOPCONS and Uniprot (Figure 5). Indeed, the localization and the length of the TMH are close, except for TMH6 which has a location further upstream of the sequence.
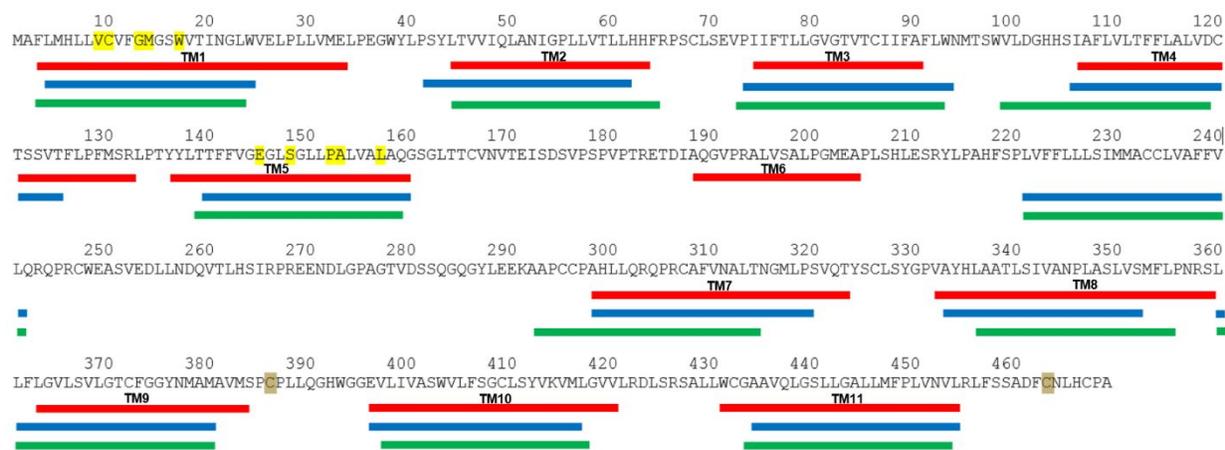


**Figure 5.** Sequence of the RFVT3 transporter indicating the localization of the TM domains observed in our model (red) and predicted by TOPCONS (blue) and Uniprot (green). The residues of the potential binding site are highlighted in yellow. A disulfide bridge is formed between C486 and C463 (highlighted in brown).

Then, the Ramachandran plot generated by the RAMPAGE program[28] indicates high quality stereochemical parameters with a very good distribution of φ and ψ angles (Figure 6). Indeed, more than 98% of the residues were located in the favoured and allowed regions.
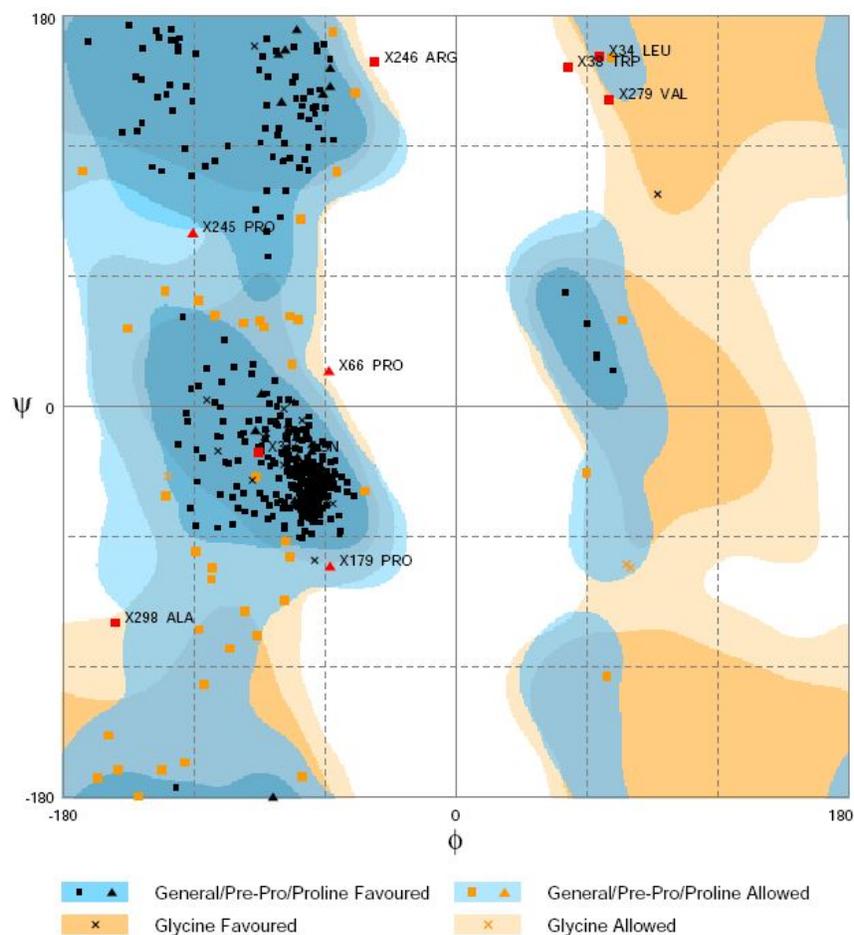
**Figure 6.** Ramachandran plot of the « RFVT3-3WDO » model

### 3.2. Docking Studies

The resulting RFVT3 model was used in a docking analysis for the identification of the binding mode of RF. The results of the blind docking using Gold led to a potential binding site located in the middle of the transporter (Figures 5 and 7A) between the transmembranedomains1, 4, 5, 7 and 11. This binding site was confirmed by Autodock Vina with a pose very close to that found by Gold (RMSD = 0.92 Å) (Figure 7C).

The site-docking analysis performed by Gold on the predicted binding site revealed a binding mode of RF involving: an aromatic stacking between the isoalloxazine ring of RF and the residue W17, hydrogen bonds between the ribitol moiety and residues W17 and E145, an hydrogen bond

between the isolalloxazine ring and the residue S148, and van der waals contact between the isoalloxazine ring and residues V9, C10, G13, M14, P152 and A153 (Figure 7C). The site-docking analysis realized with Autodock Vina confirmed these interactions with a slightly shifted pose.
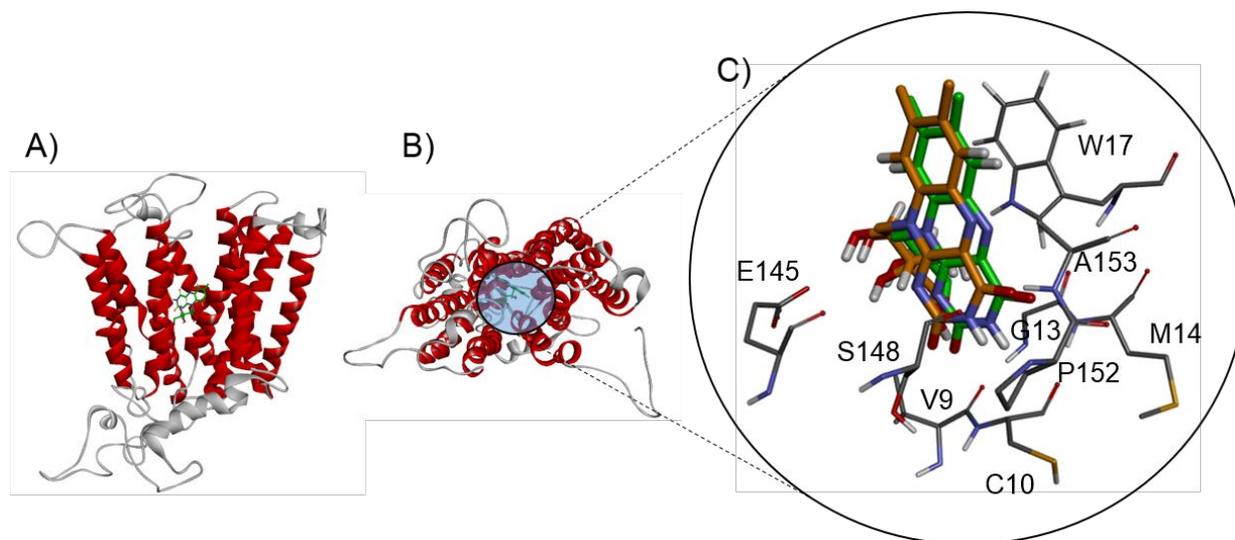


**Figure 7.** Docking of RF into RFVT3: **A)** Side view, **B)** Extracellular view and **C)** magnified view of the potential RF binding mode detected by GOLD  (RF is colored in green, red and blue) and Autodock Vina (RF is colored in orange, red and blue)

It is firstly interesting to note that the residues composing the binding site are conserved in other species such as rat, mouse, bovine and giant panda (Figure 8),thus indicating the importance of these residues. Then, in Monaco's work relative to the study of the interactions of RF with its plasma carrier protein, called RCP,[50] X-ray crystallography data (generously provided by Pr. H.L. Monaco) had also highlighted the involvement of aromatic residues, including a tryptophane, in the stabilization of RF in its binding site (Figure 9). Indeed, CH-π and π-π stacking interactions were found between the group isoalloxazine and residues F160, W156 and Y160. Moreover, our results are consistent with the work of Nabokina et al.[51]which

demonstrated the importance of the residue W17 in the transport of RF. Indeed, their results

showed an impairment in RF uptake by Caco-2 cells transiently expressing W17R, one of the

pathogenic mutations characterized in RFVT3. W17R mutant was expressed at the cell

membrane, thus suggesting that this mutation could directly affect the transport mechanism of

RF. The residue W17 was also previously found to be in the binding site of RF predicted by

Subramanian et al.,[52] supporting the relevance of our finding.

This body of evidence seems to show the key role of the residue W17 in the molecular

recognition of RF by RFVT3. In order to confirm this assumption, we performed molecular

dynamics (MD) simulations on RFVT3-RF and RFVT3W17R-RF complexes to compare their

stability over time. For this purpose, RF was also previously docked into the same region of
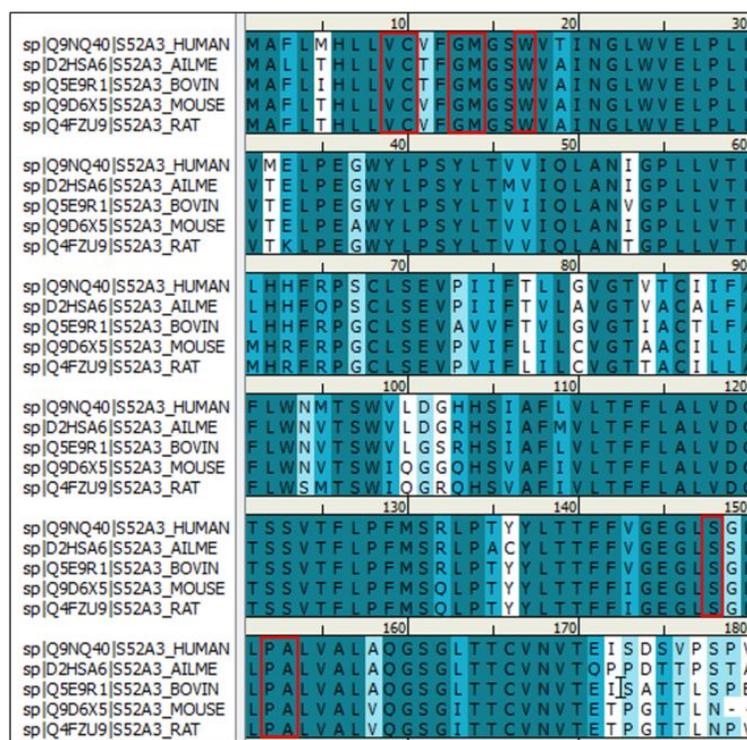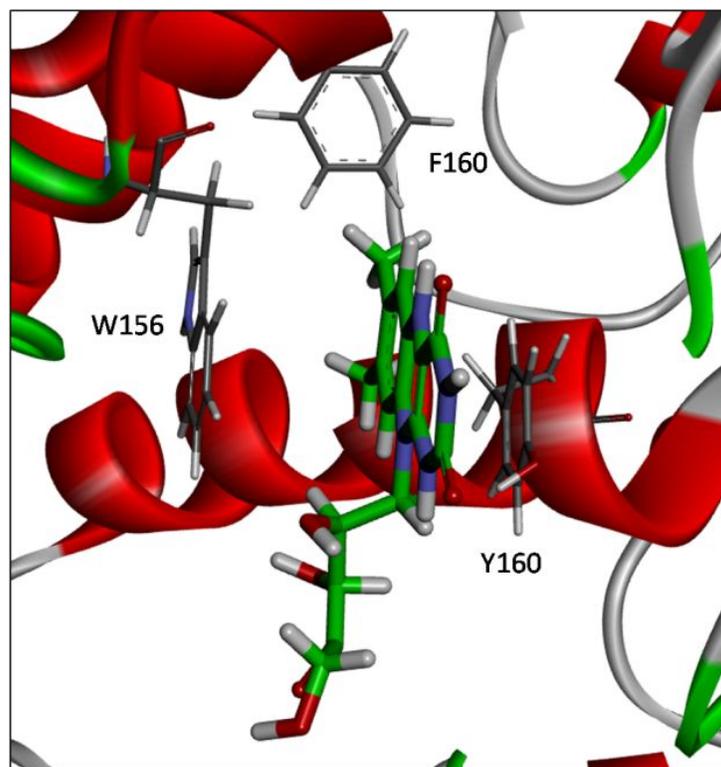
RFVT3W17R.



**Figure 8.** Alignment of RFVT3 sequences of different species. The residues of RF binding site

are framed in red. The shades of blue indicate the similarity between compared residues (from

dark blue to light blue for strong to weak similarity, respectively). The white colour indicated the

absence of similarity.



**Figure 9.** Binding mode of RF in the plasma carrier protein RCP (RX structure generously

provided by Pr. H.L. Monaco).

### 3.3. MD simulations

The MD simulation performed on the RFVT3-RF complex indicated that RF moves 4 Å from

its initial position (around 60 ns) before stabilizing at a position distant from 2 Å (Figure 10A).

These results indicated that the initial RF binding found by the docking analysis was not optimal

and needed to be refined. The comparison of the position of RF at 0 and 500 ns (Figure 10A)

showed that the interaction with the residue W17 was maintained. Moreover, a study of the

average interaction energy between RF and the residues of the binding site (based on

electrostatic and VDW interactions) indicated that RF binding was mainly stabilized by its

interactions with the residue W17 (Figure 10B).
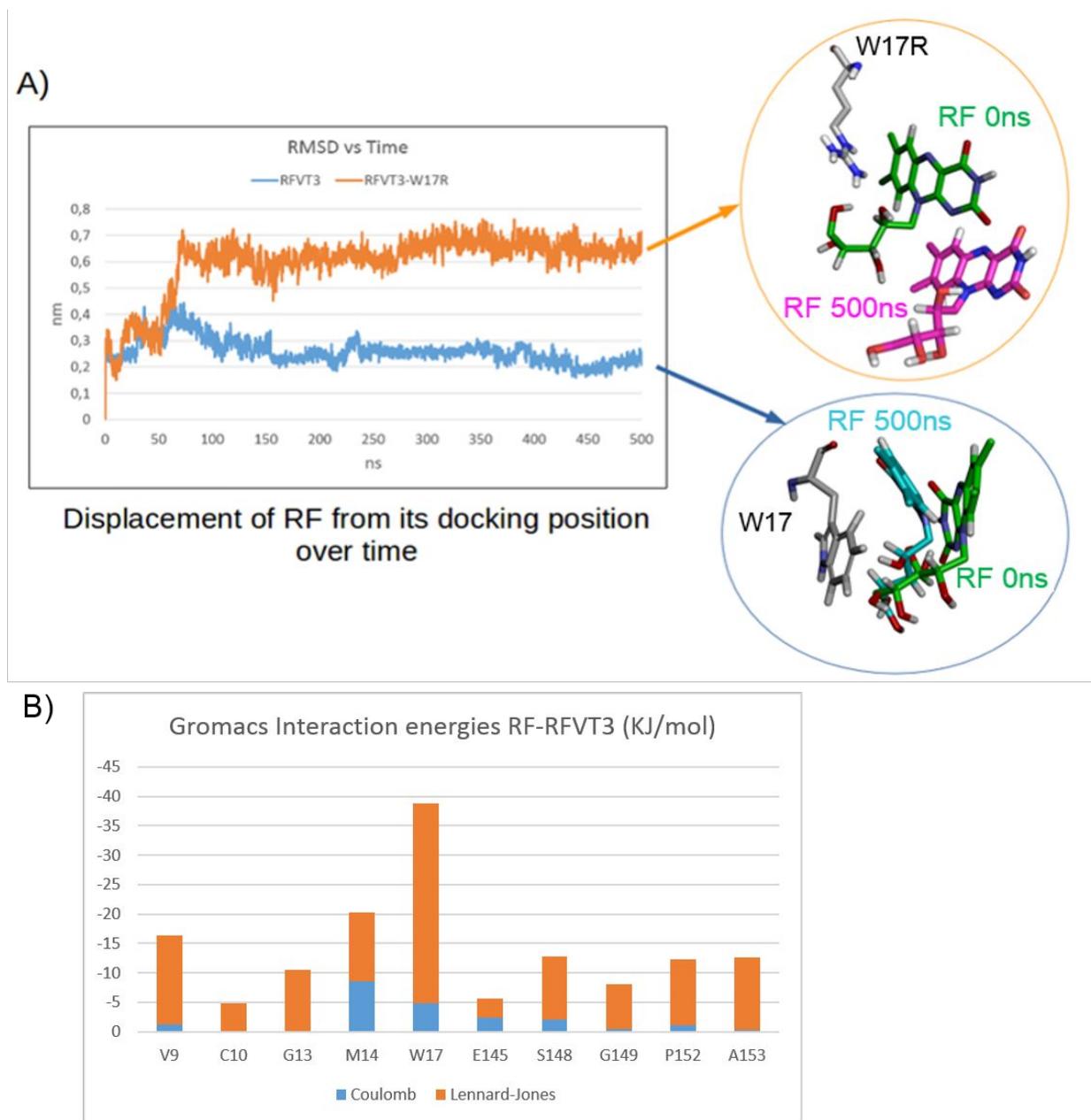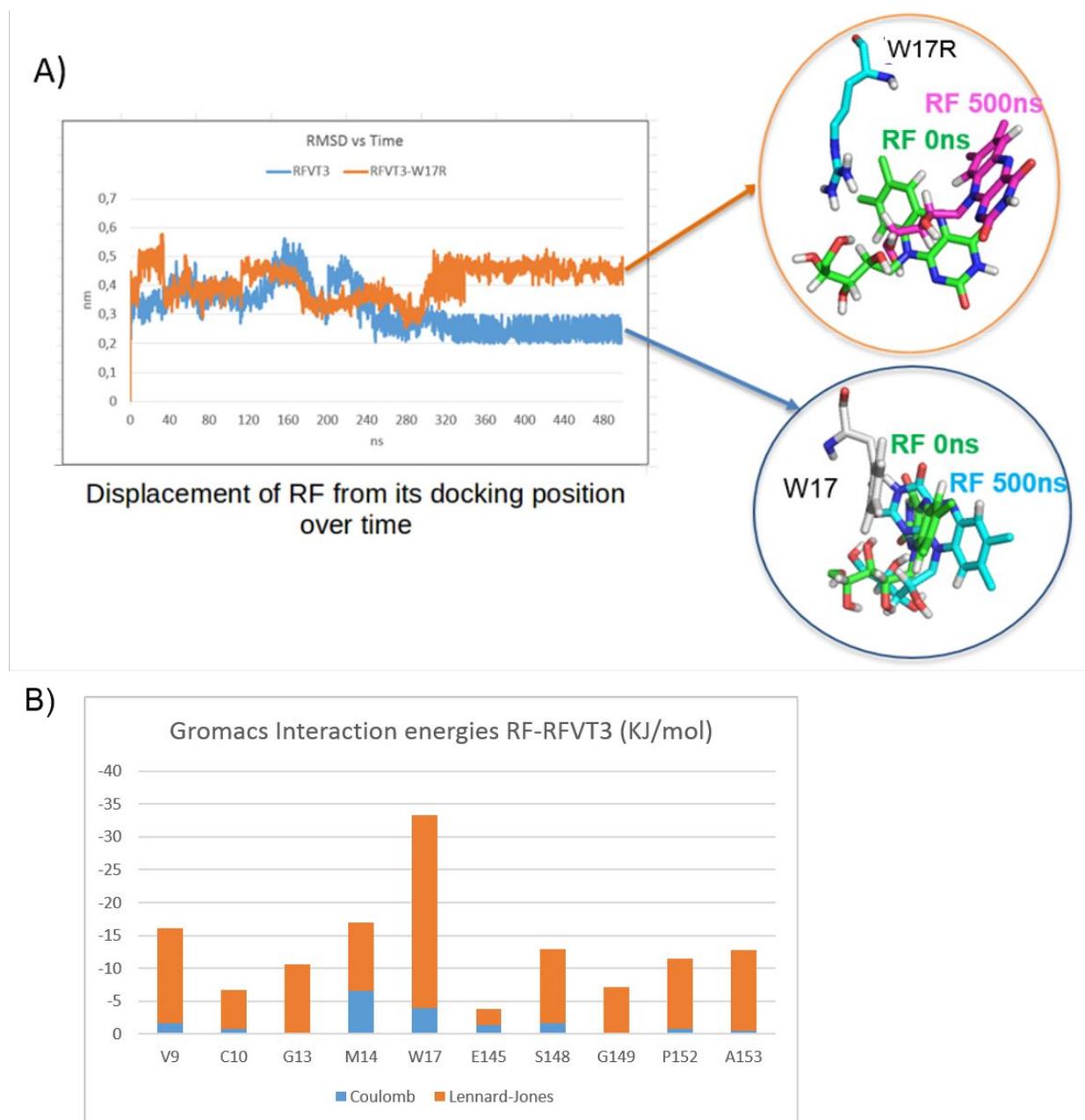


**Figure 10.** Molecular dynamics simulations on the complexes RFVT3-RF and RFVT3W17R-

RF. **A)** Study of the displacement of RF. **B)** Average interaction energy between RF and the

residues of the binding site. The contribution of the electrostatic (calculated by the Coulomb

Page 19 of 30                    Journal of Chemical Information and Modeling

potential) and Van der Waals (calculated by the Lennard-Jones Potential) are indicated by blue

and orange bars, respectively.

The position of RF into the binding site of RFVT3W17R was also investigated (Figure 10A).

By its nature, the guanidinium group of arginine can form a cation- $\pi$ interaction with pi system

such as the phenyl moiety of isoalloxazine group that could stabilize the binding of RF.

However, this interaction did not occur if we consider the distance measured between the

guanidinium group and the phenyl moiety (Dcat-$\pi$ = 6.67 A°). Indeed, for this type of

interaction, the distance Dcat-$\pi$ cutoff is 5A°.[53] In fact, the substrate was shown to move away

from its original position after 50 ns of simulation with a RMSD score higher than 6Å, thus

indicating that the replacement of the tryptophan by the arginine disturbs the binding of RF.

MD simulations are inherently subject to the problem of reproducibility. Indeed, even the same

input files, run on the same computer, may not produce the same trajectory. This is why running

a single trajectory may not mean much. At least, one replicate is almost always required to verify

the convergence of the MD results. Thus, one replicate was done. The same tendency was

observed. Firstly, the initial RF binding also needed to be refined before stabilizing at a position

close to 2 Å from its initial position, but after 240 ns of simulation (Figure 11A).Then, the

average interaction energy between RF and the residues of the binding site confirmed the main

contribution of the residue W17 in the stabilization of RF binding (Figure 11B). Lastly, the

W17R substitution was also shown to affect the stabilization of RF with a significant

displacement after 300 ns (Figure 11A). As well as in the first simulation, no cation-$\pi$ interaction

that could stabilize the binding of RF was found (Dcat-$\pi$ > 6A°).

ACS Paragon Plus Environment

19

**Figure 11.** Replicate of Molecular dynamics simulations on the complexes RFVT3-RF and RFVT3W17R-RF. **A)** Study of the displacement of RF. **B)** Average interaction energy between RF and the residues of the binding site. The contribution of the electrostatic (calculated by the Coulomb potential) and Van der Waals (calculated by the Lennard-Jones Potential) are indicated by blue and orange bars, respectively.

## 4. CONCLUSION

In this work, we propose an *in silico* procedure to explore the interactions between RF and one of its membrane transporter RFVT3 at the atomic level. The generation of a reliable model by a threading approach allowed us to predict the residues involved in the binding of RF. More precisely, our investigations indicated that the residue W17 plays a major role in the molecular recognition of RF by its transporter and its replacement by an arginine residue appears to disturb RF binding. That finding thus suggests that the pathogenic mutation W17R prevents the molecular recognition of RF by its binding site and therefore blocks RF transport. Moreover, this study provides a proof of concept of our computational protocol for the substrate recognition, not only of RFVTs, but also of other membrane transporters whose 3D structure is unknown.

## ASSOCIATED CONTENT

**Supporting Information**.

The sequence Sequence Alignment between RFVT3 and the YaJR transporter (PDB code: 3WDO) is provided.

The following files are available free of charge.

3D coordinates of the RFVT3 model (PDB)

3D coordinates of the RFVT3-RF complexe (PDB)

3D coordinates of the RFVT3-RF complexes after 500 ns of MD simulations (PDB)

3D coordinates of the mutated RFVT3W17R model (PDB)

3D coordinates of the mutated RFVT3W17R-RF complexe (PDB)

3D coordinates of the mutated RFVT3W17R-RF complexes after 500 ns of MD simulations

(PDB)

**Corresponding Author**

* E-mail: sebastien.dilly@gmail.com. Phone: +33 621320357

**Present Addresses**

† Institute Gustave Roussy, 114 Rue Edouard Vaillant, 94800 Villejuif, France

**Author Contributions**

The manuscript was written through contributions of all authors. All authors have given approval to the final version of the manuscript.

**Notes**

The authors declare no competing financial interest.

**ABBREVIATIONS**

RF, riboflavin; RFVT, riboflavin transporter; SLC solute carrier; DMPC, 1,2-Dimyristoyl-*sn*-glycero-3-phosphocholine

**Authors will release the atomic coordinates and experimental data upon article publication**

**REFERENCES**

1. Yonezawa, A.; Inui, K., Novel Riboflavin Transporter Family RFVT/SLC52: Identification, Nomenclature, Functional Characterization and Genetic Diseases of RFVT/SLC52. *Mol. Aspects Med.* **2013**, 34, 693-701.

2.      Yamamoto, S.; Inoue, K.; Ohta, K. Y.; Fukatsu, R.; Maeda, J. Y.; Yoshida, Y.; Yuasa, H., Identification and Functional Characterization of Rat Riboflavin Transporter 2. *J. Biochem.* **2009**, 145, 437-443.

3.      Colas, C.; Ung, P. M.; Schlessinger, A., SLC Transporters: Structure, Function, and Drug Discovery. *MedChemComm* **2016**, 7, 1069-1081.

4.      Beztsinna, N.; Tsvetkova, Y.; Bartneck, M.; Lammers, T.; Kiessling, F.; Bestel, I., Amphiphilic Phospholipid-Based Riboflavin Derivatives for Tumor Targeting Nanomedicines. *Bioconjugate Chem.* **2016**, 27, 2048-2061.

5.      Ferrari, M., Nanovector Therapeutics. *Curr. Opin. Chem. Biol.* **2005**, 9, 343-346.

6.      Hsueh, C. T.; Selim, J. H.; Tsai, J. Y.; Hsueh, C. T., Nanovectors for Anti-Cancer Drug Delivery in the Treatment of Advanced Pancreatic Adenocarcinoma. *World J. Gastroenterol.* **2016**, 22, 7080-7090.

7.      Beztsinna, N.; Sole, M.; Taib, N.; Bestel, I., Bioengineered Riboflavin in Nanotechnology. *Biomaterials* **2016**, 80, 121-133.

8.      Human Protein Atlas : http://www.proteinatlas.org/.

9.      Chothia, C.; Lesk, A. M., The Relation between the Divergence of Sequence and Structure in Proteins. *EMBO J.* **1986**, 5, 823-826.

10.    Dakal, T. C.; Kumar, R.; Ramotar, D., Structural Modeling of Human Organic Cation Transporters. *Comput. Biol. Chem.* **2017**, 68, 153-163.

11.    Shaikh, S. A.; Tajkhorshid, E., Modeling and Dynamics of the Inward-Facing State of a Na+/Cl- Dependent Neurotransmitter Transporter Homologue. *PLoS Comput. Biol.* **2010**, 6, 1-14.

12.    Bosch, A. M.; Abeling, N. G. G. M.; Ijlst, L.; Knoester, H.; van der Pol, W. L.; Stroomer, A. E. M.; Wanders, R. J.; Visser, G.; Wijburg, F. A.; Duran, M.; Waterham, H. R., Brown-Vialetto-Van Laere and Fazio Londe Syndrome is Associated with a Riboflavin Transporter Defect Mimicking Mild MADD: a New Inborn Error of Metabolism with Potential Treatment. *J. Inherited Metab. Dis.* **2011**, 34, 159-164.

13.    Green, P.; Wiseman, M.; Crow, Y. J.; Houlden, H.; Riphagen, S.; Lin, J.-P.; Raymond, F. L.; Childs, A.-M.; Sheridan, E.; Edwards, S.; Josifova, D. J., Brown-Vialetto-Van Laere Syndrome, a Ponto-Bulbar Palsy with Deafness, Is Caused by Mutations in C20orf54. *Am. J. of Hum. Genet.* **2010**, 86, 485-489.

14.    Johnson, J. O.; Gibbs, J. R.; Megarbane, A.; Urtizberea, J. A.; Hernandez, D. G.; Foley, A. R.; Arepalli, S.; Pandraud, A.; Simón-Sánchez, J.; Clayton, P.; Reilly, M. M.; Muntoni, F.; Abramzon, Y.; Houlden, H.; Singleton, A. B., Exome Sequencing Reveals Riboflavin Transporter Mutations as a Cause of Motor Neuron Disease. *Brain* **2012**, 135, 2875-2882.

15.    AVAKAS : https://redmine.mcia.univ-bordeaux.fr/projects/cluster-avakas/wiki.

16.    Altschul, S. F.; Gish, W.; Miller, W.; Myers, E. W.; Lipman, D. J., Basic Local Alignment Search Tool. *J. Mol. Biol.* **1990**, 215, 403-410.

17.    Protein Data Bank : http://www.rcsb.org/pdb/search/advSearch.do?st=SequenceQuery.

18.    UniProt: the Universal Protein Knowledgebase. *Nucleic Acids Res.* **2017**, 45, D158-D169.

19.    Xiang, Z., Advances in Homology Protein Structure Modeling. *Curr. Protein Pept. Sci.* **2006**, 7, 217-227.

20.    Melo, F.; Sanchez, R.; Sali, A., Statistical Potentials for Fold Assessment. *Protein Sci.* **2002**, 11, 430-448.

21.    Moult, J.; Fidelis, K.; Kryshtafovych, A.; Schwede, T.; Tramontano, A., Critical Assessment of Methods of Protein Structure Prediction (CASP) - Round x. *Proteins: Struct., Funct., and Bioinf.* **2014**, 82, 1-6.

22.    Zhang, Y., I-TASSER Server for Protein 3D Structure Prediction. *BMC Bioinf.* **2008**, 9, 40.

23.    Wallner, B.; Fang, H.; Elofsson, A., Automatic Consensus-Based Fold Recognition using Pcons, ProQ, and Pmodeller. *Proteins* **2003**, 53 Suppl 6, 534-541.

24.    Kallberg, M.; Wang, H.; Wang, S.; Peng, J.; Wang, Z.; Lu, H.; Xu, J., Template-Based Protein Structure Modeling using the RaptorX Web Server. *Nat. Protoc.* **2012**, 7, 1511-1522.

25.    Modeller : https://salilab.org/modeller/.

26.    Fiser, A.; Do, R. K. G.; Šali, A., Modeling of Loops in Protein Structures. *Protein Sci.* **2000**, 9, 1753-1773.

27.    Zhang, Y.; Skolnick, J., Scoring Function for Automated Assessment of Protein Structure Template Quality. *Proteins* **2004**, 57, 702-710.

28. Lovell, S. C.; Davis, I. W.; Arendall, W. B., 3rd; de Bakker, P. I.; Word, J. M.; Prisant, M. G.; Richardson, J. S.; Richardson, D. C., Structure Validation by C-alpha Geometry: phi,psi and C-beta Deviation. *Proteins* **2003**, 50, 437-450.

29. Tsirigos, K. D.; Peters, C.; Shu, N.; Kall, L.; Elofsson, A., The TOPCONS Web Server for Consensus Prediction of Membrane Protein Topology and Signal Peptides. *Nucleic acids Res.* **2015**, 43, W401-407.

30. Systèmes, S. D. D. *Dassault Systèmes BIOVIA, Discovery Studio Client*, 16.1.0; 2015.

31. Jones, G.; Willett, P.; Glen, R. C.; Leach, A. R.; Taylor, R., Development and Validation of a Genetic Algorithm for Flexible Docking. *J. Mol. Biol.* **1997**, 267, 727-748.

32. Trott, O.; Olson, A. J., AutoDock Vina: Improving the Speed and Accuracy of Docking with a New Scoring Function, Efficient Optimization, and Multithreading. *J. Comput. Chem.* **2010**, 31, 455-461.

33. Berendsen, H. J. C.; van der Spoel, D.; van Drunen, R., GROMACS: A Message-Passing Parallel Molecular Dynamics Implementation. *Comp. Phys. Commun.* **1995**, 91, 43-56.

34. Oostenbrink, C.; Villa, A.; Mark, A. E.; Van Gunsteren, W. F., A Biomolecular Force Field Based on the Free Enthalpy of Hydration and Solvation: The GROMOS Force-Field Parameter sets 53A5 and 53A6. *J. Comput. Chem.* **2004**, 25, 1656-1676.

35. Malde, A. K.; Zuo, L.; Breeze, M.; Stroet, M.; Poger, D.; Nair, P. C.; Oostenbrink, C.; Mark, A. E., An Automated Force Field Topology Builder (ATB) and Repository: Version 1.0. *J. Chem. Theory Comput.* **2011**, 7, 4026-4037.

36. University of Calgary website : http://wcm.ucalgary.ca/tieleman/downloads.

37.    Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; Hermans, J., Interaction Models for Water in Relation to Protein Hydration. *In: B. Pullman Ed., Intermolecular Forces, Dordrecht: D. Reidel Publishing Company, Boston, pp. 331-342.* **1981**.

38.    Woodcock, L. V., Isothermal Molecular Dynamics Calculations for Liquid Salts. *Chem. Phys. Lett.* **1971**, 10, 257-261.

39.    Parrinello, M.; Rahman, A., Crystal Structure and Pair Potentials: A Molecular-Dynamics Study. *Phys. Rev. Lett.* **1980**, 45, 1196-1199.

40.    Parrinello, M.; Rahman, A., Polymorphic Transitions in Single Crystals: A New Molecular Dynamics Method. *J. Appl. Phys.* **1981**, 52, 7182-7190.

41.    Wohlert, J.; Edholm, O., Dynamics in Atomistic Simulations of Phospholipid Membranes: Nuclear Magnetic Resonance Relaxation Rates and Lateral Diffusion. *J. Chem. Phys.* **2006**, 125, 204703.

42.    Kandt, C.; Ash, W. L.; Tieleman, D. P., Setting Up and Running Molecular Dynamics Simulations of Membrane Proteins. *Methods* **2007**, 41, 475-488.

43.    Hoover, W. G., Canonical Dynamics: Equilibrium Phase-Space Distributions. *Phys. Rev. A* **1985**, 31, 1695-1697.

44.    Nosé, S., A Unified Formulation of the Constant Temperature Molecular Dynamics Methods. *J. Chem. Phys.* **1984**, 81, 511-519.

45.    Delano, W. L. *The PyMOL Molecular Graphics System*, 2002.

46. Humphrey, W.; Dalke, A.; Schulten, K., VMD: Visual Molecular Dynamics. *J. Mol. Graphics* **1996**, 14, 33-38.

47. Diallinas, G., Understanding Transporter Specificity and the Discrete Appearance of Channel-like Gating Domains in Transporters. *Front. Pharmacol.* **2014**, 5, 207.

48. Weyand, S.; Shimamura, T.; Yajima, S.; Suzuki, S.; Mirza, O.; Krusong, K.; Carpenter, E. P.; Rutherford, N. G.; Hadden, J. M.; O'Reilly, J.; Ma, P.; Saidijam, M.; Patching, S. G.; Hope, R. J.; Norbertczak, H. T.; Roach, P. C.; Iwata, S.; Henderson, P. J.; Cameron, A. D., Structure and Molecular Mechanism of a Nucleobase-Cation-Symport-1 Family Transporter. *Science* **2008**, 322, 709-713.

49. Drew, D.; Boudker, O., Shared Molecular Mechanisms of Membrane Transporters. *Annu. Rev. Biochem.* **2016**, 85, 543-572.

50. Monaco, H. L., Crystal Structure of Chicken Riboflavin-Binding Protein. *EMBO J.* **1997**, 16, 1475-1483.

51. Nabokina, S. M.; Subramanian, V. S.; Said, H. M., Effect of Clinical Mutations on Functionality of the Human Riboflavin Transporter-2 (hRFT-2). *Mol. Genet. Metab.* **2012**, 105, 652-657.

52. Subramanian, V.S.; Sabui, S.; Teafatiller, T.; Bohl, J.A., Said H.M., Structure/functional Aspects of the Human Riboflavin Transporter-3 (*SLC52A3*): Role of the Predicted Glycosylation and Substrate-Interacting Sites. *Am. J. of Physiol.* **2017**, 313, C228-C338.

53. Khandelia H.; Kaznessis Y.N., Cation-pi Interactions Stabilize the Structure of the Antimicrobial Peptide Indolicidin near Membranes: Molecular Dynamics Simulations. *J. Phys. Chem. B.* **2007**, 111, 242-250.

"For Table of Contents Use Only"

*In Silico* Identification of a Key Residue for Substrate Recognition of the Riboflavin Membrane

Transporter RFVT3

*Sébastien Dilly,\*Mélanie Garnier, Marion Solé, Nada Taib, Rémi Bailly, Isabelle Bestel*