Versatile Dimerisation Process of Translocator Protein (TSPO) Revealed by an Extensive Sampling Based on a Coarse-Grained Dynamics Study

Rajas Rao, Julien Diharce, Bérénice Dugué, Mariano A. Ostuni, Frédéric Cadet, Catherine Etchebest

A VERSATILE DIMERISATION PROCESS OF TRANSLOCATOR PROTEIN (TSPO) REVEALED BY AN EXTENSIVE SAMPLING BASED ON A COARSE-GRAINED DYNAMICS STUDY

Rajas Rao[1,2,3]†, Julien Diharce[1,2,3], Bérénice Dugué[1,2,3], Mariano A. Ostuni[1,2], Fréderic Cadet[1,2,3,4], Catherine Etchebest[1,2,3]*

[1]Université de Paris, Biologie Intégrée du Globule Rouge, UMR_S1134, BIGR, INSERM, F-75015, Paris, France

[2]Laboratoire d'Excellence GR-Ex, Paris, France

[3]Université de la Réunion, Biologie Intégrée du Globule Rouge, UMR_S1134, BIGR, F-97715, Faculté des Sciences & Technologies Saint-Denis, France.

[4]PEACCEL, Artificial Intelligence Department, 6 Square Albin Cachot, box 42, 75013 Paris, France.

ABSTRACT: Translocator protein (TSPO), a mitochondrial membrane protein, has been extensively studied and its role is still debated and continues to be enigmatic. From a structural perspective, despite availability of atomic structures from different species, the possible oligomeric state and its 3D structure remains elusive. In the present study, we attempted to study dynamics of TSPO from the perspective of oligomerization. In this aim, we examined if and how TSPO monomers could assemble to form dimer. Accordingly, we performed several coarse-grained molecular dynamics simulations considering two different initial configurations, one with pair of TSPO monomers distantly placed in a model of bilayer composed of DMPC/cholesterol mixture and the other with preformed dimer models with different starting interactions. We identify stable TSPO dimers with diverse interfaces, some of which being consistent with earlier experimental observations on putative TSPO oligomer interfaces. For most of the stable ones, interactions between aromatic residues were significantly overrepresented in diverse oligomeric organisation. Interestingly, we identified different communication pathways that involve dimer interfaces. Additionally, we observed that a cholesterol molecule in close interaction with TSPO dimer was able to translocate through the bilayer. This phenomenon might be related to the putative mechanism of cholesterol transport, and could be increased and favoured by the dimer formation. Overall, our observations shed new light on TSPO oligomerization and bring new perspectives on its dynamics, as well its interactions with protein and ligand partners.

INTRODUCTION

Translocator protein (TSPO) is an 18 kDa transmembrane protein that was primarily found in outer mitochondrial membrane, though it has also been reported in other organelles, such as plasma membrane,[1] nuclear envelope[2] and endoplasmic reticulum.[3] It has a characteristic property of being enriched in tryptophan residues. While it is active in various tissues and is present in organisms from all the three domains of life, it was first observed in nervous system, as a putative receptor that binds to benzodiazepines.[4] It was named as peripheral benzodiazepine receptor (PBR), due to the fact that it appeared to be predominantly concentrated in peripheral tissues. However, it was later found to be involved in transport of cholesterol into the mitochondria.[5] Successive experiments later found that PBR has a major role in translocation of cholesterol into the mitochondria,[6-8] due to which it was renamed as TSPO, the TranSlocator PrOtein.[9,10] This was supported by another study in which incorporation of TSPO in *E. coli* cells resulted in increased cholesterol uptake. Since *E. coli* mitochondria does not have cholesterol, increased uptake of cholesterol in TSPO transfected cells further reinforced the idea that TSPO is indispensable for cholesterol translocation.[11-15] Further site-specific mutagenesis studies revealed the presence of a cholesterol-binding motif, called as Cholesterol Recognition And Consensus motif or 'CRAC' motif, to which cholesterol was found to bind at nanomolar affinities.[6] Located at the C-terminal of the protein, the motif is largely hydrophobic in nature, and consists of residues L/V-(X)1–5-Y-(X)1–5-R/K.[6,8] The presence of a positively charged residue at the end of the motif suggests that it may act as a tether that interacts with the hydroxyl group of cholesterol. This is also supported from the recent gene-deletion studies by means CRISPR/Cas9 method, where TSPO deletion resulted in reduced steroid transport into mitochondria.[16] Furthermore, an alternate cholesterol-binding motif adjacent to the CRAC motif was characterized recently.[17,18] This motif has a sequence pattern palindromic to CRAC motif, and thus was called 'CARC' motif. It has to be noted that the implication of TSPO in steroid biosynthesis has been questioned in several studies,[19-21] and no clear conclusions have not yet been made on this aspect nowadays.

Apart from cholesterol transport, TSPO has also been observed to play a role in various other processes, such as transport of porphyrin,[22] erythropoiesis,[23] reactive oxygen species (ROS)-mediated apoptosis,[24-27] neuroinflammation[28-30] to name a few. However, despite these significant advances in knowledge of TSPO and its implication in various biological processes, its exact function continues to remain a mystery.

Similarly, until recent past, the structural knowledge of TSPO at atomic level also remained unknown. Earlier molecular modelling studies on TSPO supported by NMR experiments suggested that it folds into a five-helices transmembrane domain (TM1 to 5),[9,31] which was confirmed from low-resolution cryo-electron microscopy studies.[32-34] The first atomistic structure was finally solved for the mouse sequence (mTSPO) by NMR spectroscopy[35] in DPC micelles. In this environment, the protein folds as a monomer. Meanwhile, X-ray crystal structures of bacterial TSPO were resolved and revealed the existence of different TSPO oligomers,[36,37] confirming earlier cryo-electron microscopy results[14,38] showing that TSPO may be found in oligomeric arrangement. For instance, *Bacillus cereus* TSPO was found as a homodimer[37] or a monomer depending on the crystallization conditions. *Rhodobacter sphaeroides* TSPO wild type[36] is a dimer in the crystal environment as its A139T mutant. Noteworthy, in this last case, a third chain makes significant contacts with a chain involved in the dimer, which brings additional stabilization forces on the assembly present in an asymmetric unit cell. Moreover, transmembrane helical segments involved in the dimerization may differ between the bacterial dimers. Interestingly, recent solid-state NMR experiments also revealed another picture for the mouse TSPO: TSPO was found to exist in a dynamic equilibrium between monomer and dimer, with the dimer interface[39] being also different from the bacterial forms. The change in stoichiometric orientation was found to be dependent on the interaction with cholesterol that would trigger the dissociation of dimer into monomer.

In summary, though the assembly of TSPO monomers into quaternary structures are relatively well-documented, there are significant differences in the nature of oligomeric interfaces and the prevalent stoichiometric arrangement, depending on the sequence and/or the experimental conditions, in particular, the cholesterol concentration. As a consequence, even though it is unequivocal that TSPO has the capability to oligomerize, the interface of the oligomer continues to be elusive.

Overall, all these results underline the difficulty to conduct experiments on TSPO not only because it is a transmembrane protein but also because it seems to be highly sensitive to the environment. Consequently, to gain further insights in this enigmatic protein, alternative approaches are needed. From several decades now, molecular simulations have demonstrated their usefulness to complete experiments and their capacity to help in understanding mechanisms at a resolution that is difficult to reach in experimental studies. Therefore, in the present study, we examine the dynamics of

TSPO dimerization using molecular dynamics simulation and address the following questions: Can TSPO monomers diffuse in a model membrane, assemble into a dimer that remains stable? What is the nature of the dimer interface? What are the stabilizing forces? What are the key residues involved in the dimerization process? And what about the behavior of cholesterol molecules near the dimers? Accordingly, we performed a large number of molecular dynamics simulations using a coarse-grained force-field and choosing different initial conformations of the system to better sample the conformational space. Our results bring support to some conclusions drawn from experiments but importantly, they revisit and enlarge the vision of the oligomeric state of TSPO. Interestingly, we have also discovered that distant residues from the interface participate in the dimerization through a network of long-range interactions. Finally, and importantly, we observed how TSPO can accompany cholesterol transfer, playing its reputed role of cholesterol transporter.

MATERIALS AND METHODS

Molecular dynamics simulations

Inspired by the experimental work of Jaipura et al.[39] who examined dimerization process of TSPO in DMPC/Cholesterol liposomes with solid-state NMR experiments, we performed simulations in the same lipid-cholesterol ratio. The starting monomer structure was the solution NMR model #1 of mouse TSPO (PDB ID: 2MGY). Coarse-grained models were described by Martini 22 force field.[42-45] Classical parameters have been chosen for ion and water representations, meaning a non-polarizable model. Secondary structure calculated with DSSP software was maintained. The choice of such force field can be questionable, because of its known weaknesses.[46] However, we have preliminary compared CG simulations of mouse (and bacterial TSPO monomers) with an all-atom force field (Charmm36 and amber ff99SB-ILDN) simulation. Results show that fluctuations between the two cases are in the same order of magnitude (maximum about 2 Å for TM regions). We even observed an overlap between the conformational space sampled in the simulations for each cases (Supp. Figure 1). Those observations brought some support to the validity of our choice to use the Martini force-field with he given protocol.

For the initial conformations, we considered two types of coarse-grained setups: i) TSPO monomers were placed at a large separation distance in a lipid bilayer and ii) TSPO monomers

were placed at a short distance as a preformed dimer in a lipid bilayer. Figures 1A and 1B represent both cases.

For the first case, the monomers were separated by a distance of 55-60Å and oriented with respect to the membrane using PPM web server[40,41].

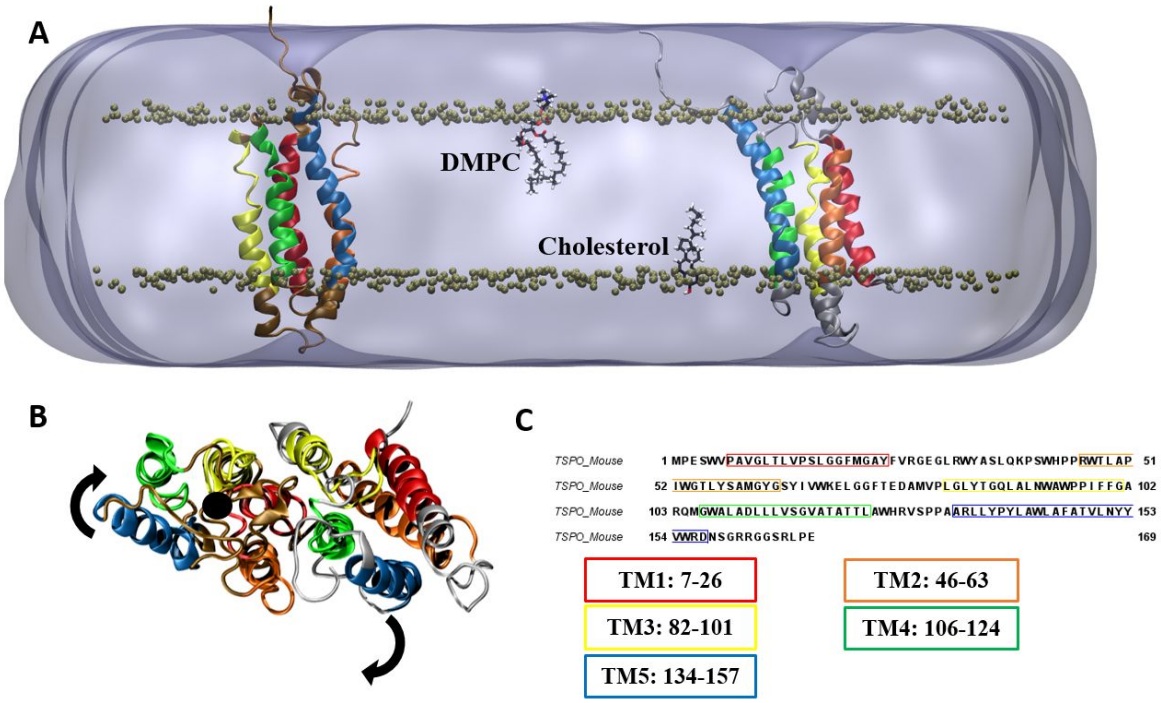

Figure 1: Presentation of the systems studied. TSPO is a transmembrane protein, composed by 5 TM helices. A) Representation of the "Sim_Dimer_X_rY" (X is the rotation angle and can be 90, 180 or 270 degrees and Y the replicate) simulations. Two TSPO monomers (loops of the first and second monomer in brown and silver, respectively) are separated with a distance of 60 Å and are embedded in a mixture of DMPC/cholesterol bilayer. Membrane is shown in surface and its limits have been showed by golden spheres corresponding to phosphate groups. B) Pre-formed TSPO dimer based on the dimer structure of Jaipiruia et al.[39], representing the several simulations named "Dimer_model_X" and “Dimer_model_X_Y” (where X and Y is the rotation angle for each monomer). The 0° of rotation for both monomers corresponds directly to the NMR model and is called “Dimer_model_ NMR”. Black arrows represent the two possibilities of monomers rotation considered in our several calculations. C) TSPO mouse protein sequence, with position of the 5 TMs, their definition and the colour code for each of them.

Replicates of starting system were constructed by rotating one of the monomer chains by 0°, 90°, 180° and 270° with respect to the other monomer. The protein chains were embedded in a bilayer consisting of 1,2-dimyristoyl-sn-glycero-3-phosphocholine (DMPC) lipid and cholesterol, in the molar ratio of 10:5 (about 400 lipid molecules and 200 cholesterol molecules), using the Charmm-Gui server . This ratio was used in order to reproduce the observations made from the solid-state NMR experiments.[39] For each system, two replicates have been carried on, with a change on the initial input velocities. The protocol is as follow: the system was subjected to energy minimization by steepest descent method for 50,000 steps. The energy-minimized system was then subjected to equilibration in NPT ensemble, with position restraints imposed on the backbone as well as side chain beads of the protein. The equilibration was performed for 1ns, with 0.01ps time step. The temperature was maintained at 300K, by means of Berendsen thermostat[47] and pressure was maintained at 1 atm by means of Parinello-Rahman barostat with semi-isotropic pressure coupling.[48] The equilibration was performed in 5 steps of 1ns each, for a total of 5ns, where each step involves gradual relaxation of protein by reduction of position restraints by 1,000, 500, 100, 10 and 0 kJ/mol.nm². After equilibration, production simulation was performed for 10µs, with time step of 0.01ps in NPT ensemble, using the V-rescale thermostat.[48] All the preparatory steps and the simulations were performed using GROMACS program (version 4.6.7 or later).[49,50] In global, 8 simulations, of 4 different orientations, with lipid-cholesterol ratio of 10:5 are realized. The total sampling was 80µs in 8 simulations. Those simulations are mentioned in the text "Sim_Dim_X_rY", where X is the rotation of the second chain and Y the number of replicas (1 or 2)

For the second system, a TSPO dimer model was constructed to reproduce the interface observed in the solid-state NMR experiments. The dimer model was constructed by HADDOCK protein-protein docking web server.[51,52] To ensure that the dimer interface in our model is similar to the interface described from the solid-state NMR experiments, the residues observed to contribute to the dimer interface, 83-GLYTG-87, were constrained as the putative dimer residues. The best-scoring complex was selected, and was found to have interface similar to the interface observed in NMR study.

Starting from these results, four more starting orientations were constructed, by rotating one of the monomer chain with respect to the main axis of the other chain by 60, 90, 180 and 270° respectively, perpendicular to the membrane plane. In this way, we postulate that those systems

can explore another part of the conformational space of the TSPO-TSPO dimer, and thus complete the sampling obtained with the simulations of separated monomer. The dimer models oriented with respect to membrane axis were embedded in a bilayer with the same composition as for systems 1, but using directly martinize.py and insane.py scripts, and choosing similar membrane composition as above, in terms of number of lipids, cholesterol and water. The system was prepared by means of energy minimization and equilibration using the same protocol as described above. Simulations were performed on all the dimer orientations for 10μs each resulting in the total sampling of 50μs. Those calculations are named in the text "Dimer_model_X" with X the rotation angle of second chain (and NMR for the initial NMR model[39]).

To provide a more exhaustive sampling, 14 additional simulations have been carried on. For four rotation of the second monomer (0, 90, 180 and 270°), we have considered four rotations of the first monomer at 0, 90, 180 and 270° around its own axis (as depicted in Figure 1B). Two systems on the 16 are equivalent to some systems already sampled, leaving 14 new systems. This pushes our study to 27 simulations, for a total duration of 270 μs. Those supplementary calculations are named "Dimer_model_X_Y", with X and Y the rotation angles for both monomers. For the sake of clarity, those simulations won't be detailed as the 13 others (for example, not for the structural protein network analyses) but their results are considered for every global analyses we present. A summary of all the simulations can be found in the Supplementary Table 1.

Analysis of all those simulations (RMSF, RMSD, distance calculation, etc.) are been realized with the GROMACS tools[49,50]

Representative conformations and calculation of interface areas

For the simulations where the dimerization was successful, we attempted to study in details the interface, starting from the surface area of contact. To answer this question, we extracted dimer conformations that are representative of the given simulation, and calculated interface area for the dimer representative. To do that, part of the considered trajectory where the monomer chains are in contact with each other, *i.e.* the minimum distance between monomers <5.2 Å, is extracted. All the following analyses, including Protein Structural Network (PSN) and Dynamic Cross-Correlation Matrix (DCCM), are made on this subset of simulations with the Bio3D tool.[53]

On the extracted conformations from each simulation, RMSD-based clustering was performed using g_cluster tool of GROMACS[49,50]), with RMSD threshold above which given pair of conformations are separated into distinct clusters is 3.5Å. For each simulation, the conformation

representative of the cluster with largest population of frames was extracted, and back mapped into all-atomic representation using 'backward' tool.[54] On the representative dimer structures, interface surface area was calculated using the following method:

$IA = (ASA_a + ASA_b) - ASA_d$ (1)

Where IA is the interface area, ASA_a is the solvent-accessible surface area for the monomer chain-A, ASA_b is the solvent-accessible surface area for the monomer chain-B, and ASA_d is the solvent accessible surface area for the dimer. The solvent accessible surface areas were calculated using Naccess program.[55] The interface area was similarly calculated for the experimental X-ray crystallographic structures of TSPO dimers form *R. sphaeroides* and *B. cereus*, and compared with the interface area of representative structures from MD simulations. The representative structures were visualized and the TM regions present at the interface were noted. TM regions were obtained from the OPM database[40] and are reported in Table 1.

Analysis of interface frequencies and interacting residue pairs

We also attempted to understand what was the frequency of interaction associated with a given pair of TM regions. To answer this question, we calculated distance between centre of mass of all the 5 TM regions (as detailed in Figure 1) with each other from the dimer ensemble, using g_dist tool of GROMACS of residues.

For an interaction involving given pair of TM regions, we calculated the frequency for every pair of TM regions. To determine which residues participate in the dimer interface, we calculated the frequency of residue contacts for the ensemble of dimer trajectories, using an in-house program. We used bead-bead distances with a distance threshold about 5.2 Å but different thresholds were used to support the main conclusions (and other criteria have been tested). Interacting TMs were deduced from the list of these residues using the TM definitions given in Figure 1.C.

The gmx mdmat tool of GROMACS,[49,50] combined with CONAN program.[56] was used create a representative trajectory. The top-5 residue pairs that have maximum contact time (in terms of % of trajectory time) were identified, and the nature of interaction was identified based on the side-chain interactions, using CONAN program.

Cholesterol Dynamics

The translocation mechanism of cholesterol was followed by computing the number of instances for which the ROH bead of cholesterol is in close proximity (<5.2 Å) to C3B bead of DMPC, since this bead is found in the center of the membrane. We have made the analysis only on the frames in which a TSPO dimer is formed. Presence of a polar ROH bead in hydrophobic environment of C3B bead indicates possible translocation of cholesterol across the bilayer. To differentiate between such instances occurring in proximity to TSPO and far from the TSPO protein, we performed the calculations accounting for both the conditions. The ratio of number of possible translocated cholesterol molecules over the total number of molecules (in all the system and near the TSPO dimers) has been calculated and analysed.

All-atom simulation

We have extracted a few representative structures from CG simulations and all-atom simulations have been carried on in order to verify the stability of the dimer formed in a more detailed resolution. The three structures considered, with 3 different interfaces (TM1-TM3/TM1-TM3-TM4, TM2-TM5/TM2-TM5, TM1-TM2/TM3-TM4) have been back-mapped from Martini to the Charmm36 force field. The entire system (dimer with DMPC membrane and cholesterol) has been simulated in all-atom resolution with GROMACS software. The protocol is as follows: an energy minimization of 10000 steps with steepest descent method has been employed, followed by a heating phase of 25 ps at 300k with Berendsen thermostat. Restraints have been put on the backbone and the sidechains of the protein. Equilibration phases begun, in 5 phases of 50 ps each in the NPT ensemble. In each phase, restraints have been gradually decreased from 2000 to 1000, 500, 200, 50 and 0 kJ/mol.nm², respectively. Finally, MD simulations have been performed for a total duration of 200 ns, with a time step of 2 fs.

RESULTS

TSPO monomers assemble to form dimers in most of the simulations

The assembly of TSPO monomers into dimers were first assessed by calculating the *minimum distance* between any atom of the two TSPO monomers as a function of time for every simulation. The dimer was considered as formed when this distance was smaller than 5.2 Å. Amongst the eight replicates of 10μs starting with separated monomers, TSPO monomers come in close contact (Supp. Fig 2), forming long-lifetime dimers in half of them. In the resting cases, dimers of short-

lifetime may also form but finally dissociate (Supp. Fig 2). This result clearly indicates that the monomers contacts may not occur incidentally but dimer formation needs specific interactions to be stabilized. This dissociation process also shows that the Martini CG force field does not necessarily artificially favour association at least in the case of this polytopic membrane protein.

When starting with pre-formed dimers with different relative orientations of the two monomers (19 replicates of 10 μs each), whatever the simulations, the two monomers remained close to one another but interacting regions vary along the simulation. Thus, despite stabilizing contacts, fluctuations remain rather large (Supp. Fig 3), which shows that the complex is highly dynamic. As expected, the regions with the largest fluctuations correspond to connecting loops but transmembrane domains are not deprived of motions. We can also observe that the dimer starting from the NMR is dynamical, with a minimal distance between the center of mass of the two chains about 27 Å (average distance: 30 Å). This system is even more dynamic, with higher fluctuations, than a dimer formed during the simulation of two initially separated monomers (minimum distance: 23 Å, average distance: 27 Å). Clearly, alternative stable conformations do exist beside the NMR model.

The minimum distance is a good criterion *per se* but it does not guarantee that the interface is large enough to stabilize the complex. Accordingly, we examined the number and nature of residues that participate in the interface, *i.e.* residues of one monomer in contact with residues of the other monomer. We focused on the transmembrane region TM and chose a 5.2 Å distance criterion between any beads to define a contact. The number of contacts covers a large range of values, extending from 15 to 35, the largest number of contacts being observed for the simulation mentioned above starting with a pre-formed dimer (Fig. 2 and Supp. Fig 4).

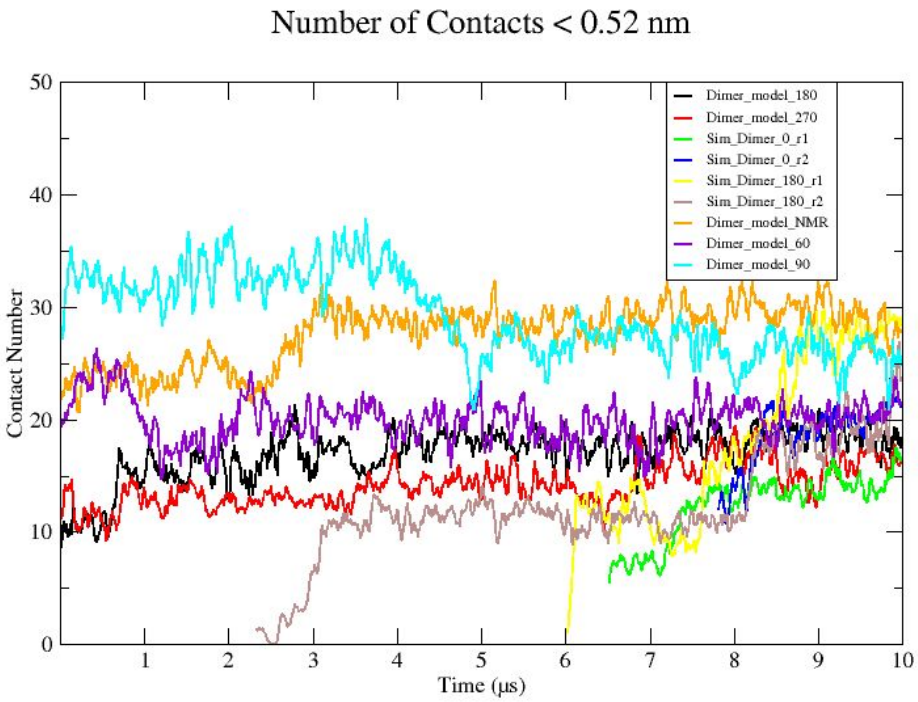

Figure 2: Contact number between the two TSPO chains during the course of the several simulations. Only 9 main simulations in which at least one contact has been observed are shown here.

Along the time course of the simulations, the number of contacts tends to decrease for the pre-formed dimers and to increase when starting from separated monomers. We have also analysed the distribution of the number of contacts in all simulations in association with the part of the structure involved. It appears that TM1, TM2, TM3 and TM5 are the most frequently involved, in terms of contact number, in the different interfaces sampled, as shown on figure 3A. Interestingly, there is a tendency of TM1 and TM3 in one hand, and TM2-TM5 on the other, to interact together within an interface, Since the threshold choice might be arbitrary, we checked its impact using a slightly increased value*, i.e.* 6 Å (Supp. Figure 3). Interestingly, using this value, the number of contacts tends to converge between the two sets of simulations, reaching on average 35 at the end of the simulations. The conclusions remain the same.

Dimerization can occur through large different interfaces but some TM pairings are preferred:

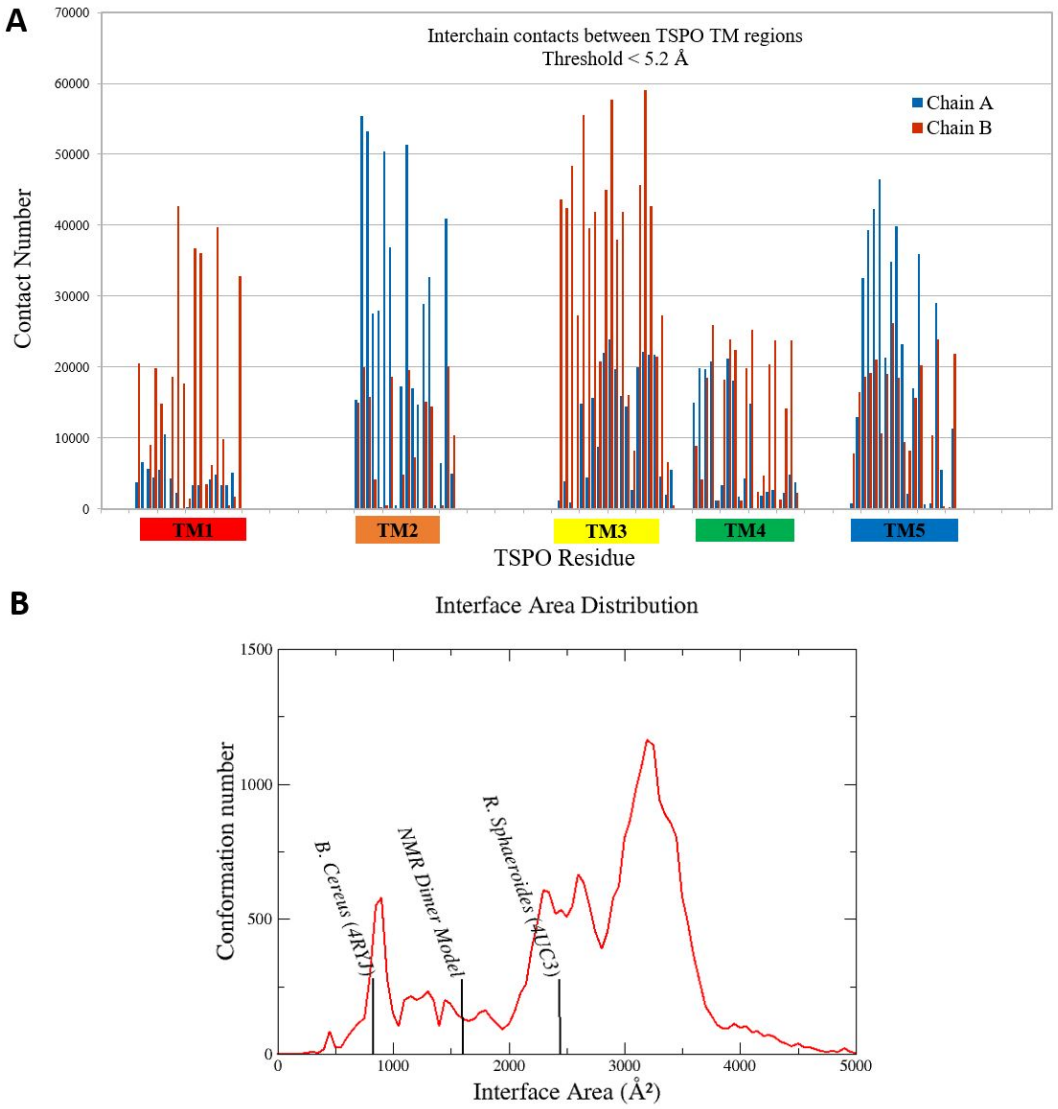

Figure 3. A) Total contact numbers for each residue calculated for all the frames saved from the whole set of trajectories. Contacts are defined on a distance-based threshold < 5.2 Å. B) Distribution of interface area of all the dimer formed during the 9 simulations. Black lines represent the position of some interface area of experimental dimers.

Besides these distance criteria, we also examined the size of the buried interface. We first rebuilt as all-atom structures of the most representative dimer CG-structures extracted from the different trajectories (see Material and Methods) and calculated the corresponding interface area. The distribution of dimer interface (Figure 3.B) is bimodal, the first one centered on [2000-2500 Å^2] and a second one on [3300-3500 Å^2]. We can observe that the interface area of the dimer we sampled is much higher than the one currently listing in the experimental structures, but we also

retrieve equivalent structures in term of interface area, especially near the *B. Cereus* dimer from 4RYJ PDB file.

We analysed the number and nature of transmembrane segments that participate in the dimer contacts using the same distance criterion as above, as detailed in Table 1 and figure 3A. They show that TM3 is the most involved in TM-TM interfaces, but its pairing partner varies, the preferred one being by far TM1 (24% of all the pairing sampled). The second most represented pairing is TM2-TM5, with 22% of the sampling.

Table 1: Heatmap representing the percentage of TM pair contacts during all simulations in which a dimer is formed or already formed.

	TM1	TM2	TM3	TM4	TM5
TM1	3%	2%	24%	14%	7%
TM2	2%	2%	5%	13%	22%
TM3	24%	5%	14%	10%	5%
TM4	14%	13%	10%	1%	7%
TM5	7%	22%	5%	7%	9%

The number of TM regions participating in the interface can significantly vary, e.g. 1 TM from each chain (1TM:1TM), 1TM from one chain with 2 TM from the other chain (1 TM:2 TM) and 2 TM from each chain (2 TM:2TM). Actually, the 2 TM:2TM association mode is the most frequent one, *i.e.* ~ 4 TMs are involved in the interface. This mode of association corresponds to the second largest peak in the interfacial area distribution (Figure 3B). Interestingly, even though the system is highly dynamical, in 55% of the dimer conformations, the interface involves a symmetrical pairing, *e.g.* TM4 and TM5 of monomer A interacts with TM5 and TM4 of monomer B, respectively. Note that symmetrical pairings would correspond to even oligomers while non symmetrical interface would lead to odd oligomers.

Table 2 shows the distribution of the nature of TMs involved in these symmetrical interfaces. The results show that three interfaces are overrepresented: TM5-TM5 (29,7%), TM3-TM3 (25,7%) and TM1-TM3 (16,1%). Results show clearly than TM1, TM3 and TM5 are the three most preferred TMs regions in those symmetrical interfaces. Details about the main interactions and residues involved for each main simulation are given in Supp. Table 2.

It has to be noted that very similar results were obtained with different thresholds and different measures tested, e.g., from distances between centers of mass of transmembrane segments.

Table 2: Distribution of TM associations in the interface involving symmetrical pairings. It results from the number of occurrences of this symmetrical pairing by the total number of symmetrical pairing.

Nature of the symmetrical interface	Percentage of existence
1 - 1	8.1%
1 - 3	16.1%
2 - 2	5.5%
2 - 5	5.4%
3 - 3	25.7%
4 - 5	5.7%
5 - 5	29.1%

Interactions between aromatic residues were the most dominant at the interface

To better understand the nature of interactions occurring between the residues of TSPO chains, we focused on the frequencies associated with residue contacts for the whole set of trajectories and for the representative set of conformations. As an example, the eight most abundant residue pairs observed in the whole set of simulations are given in Table 3. A most detailed table is available in Supp. Table 3. Depending on the threshold value, the ranking of preferred pairs of residues may slightly change, but we systematically found that Leu, Phe, Tyr and small residues like Gly, Ala and Thr were frequently involved. This result is expected since these last three types of amino acids are the most abundant ones among the accessible residues in the structure. Especially, Leu residues are the most accessible, and consequently the most frequently observed in contact. However, *considering their frequency of accessibility*, the percentage of observed contacts seems to be underrepresented, except for Thr (see Supp.Table 3). In contrast, despite a smaller number of accessible Phe in comparison with Leu, Phe-Phe pairing is extremely frequent and significantly overrepresented. Similarly, but to a lesser extent, Trp-Trp and Tyr-Tyr contacts are abundant and slightly more represented than would be expected from their accessibility.

Table 3: Distribution of most frequent (above 3%) type of pair of contacts. The observed contacts have been calculated by the ratio between the identified pair over the total number of contacts. Cut-off distance has been set to 5.2 Å. Score is defined by the ratio between the observed contacts over the combination of both accessibility frequencies (calculated with Naccess). A "+" value means that this contact is overrepresented in respect to pair accessibility, and a "-" value means that the contact is underrepresented, despite a high accessibility. "++" symbol means that the residues pair is highly overrepresented compared to the number of residues accessible. A more detailed version is given as Supp. Table 3.

Contact pair		% Observed Contacts	Score
LEU	LEU	10.9%	-
PHE	PHE	5.6%	++
PRO	PRO	5.3%	+
TYR	TYR	4.6%	+
THR	THR	4.3%	+
GLY	GLY	4.2%	-
ALA	ALA	3.62%	-
TRP	TRP	3.2%	+

Thus, despite the fact that TSPO has abundance of 'SmxxxSm' motifs, that are hypothesized to mediate dimerization in various TM proteins,[57-59] and besides the Leu-Leu most frequent contacting residues, aromatic residues seem also frequently involved to mediate dimerization. As an additional support to this observation, when we started with a model based on solid-state NMR data, which involves only the TM3 and a GxxxG dimerization motif (residues 83-87 of TM3, Fig. 4A),[39] this dimer interface is not retained along the simulation. We observe that the interface evolves to finally involve TM1 and TM3 for both monomers. Consequently, the interaction between GxxxG motifs, which was the only interaction in the starting model, is now enriched with other kind of interaction (aromatic interactions manly), and especially Phe-Phe pairing (see Fig 4.B).

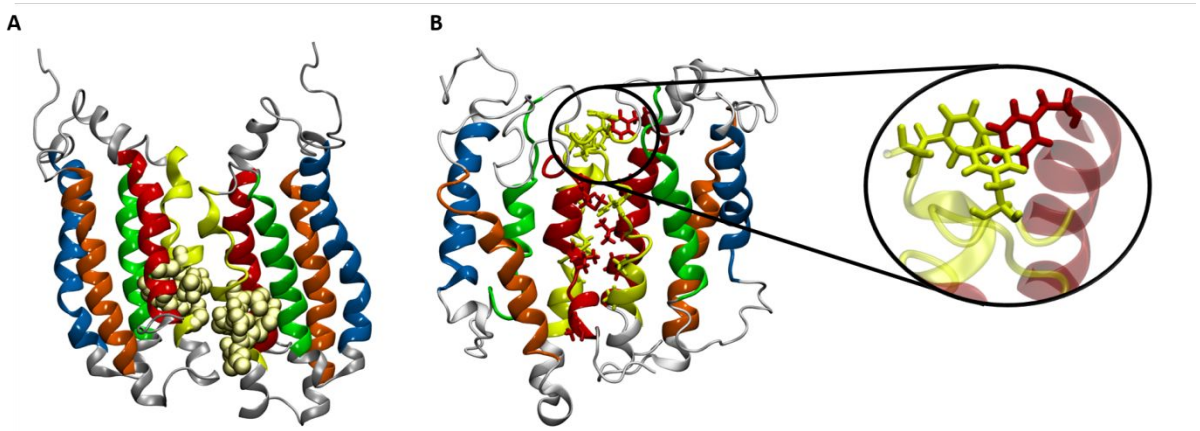

Figure 4. A: Starting structure of TSPO dimer model, with interface similar as described by Jaipuria et. al. Residues of GxxxG motif (83-GLYTG-87) are represented as yellow spheres. **B**: A representative conformation of TSPO dimer from the simulation with starting conformation described by Jaipuria et. al,[39] with a zoom on the newly formed aromatic interactions at the interface. Residues at the newly formed interface are shown by the *Licorice* representation and belong to TM3 and TM1. TM regions 1 to 5 are represented as red, orange, yellow, green and blue ribbons respectively.

Possible explanations for these observations might be: i) the sampling of the conformational space corresponding to these particular interfaces was not sufficient, ii) the Martini coarse-grained force field is less favorable for these interactions but more favorable for aromatic-aromatic interactions, iii) the absence of ligand in our simulations might have affected the monomer dynamics and *per se* might have modified the dimer formation and finally, iv) the interface size of the putative experimental dimer model is too small, (only about 5 residues are involved)[39] to maintain the stability of the complex. Consequently, initial interface of TM3 changed to a larger interface involving TM1 and TM3. Interestingly, the interface between TM1 & TM3 from our simulations is very similar to the dimer interface observed in *R. sphaeroides*.[36] In addition, it is important to note that our results are in agreement with recent molecular modelling studies, showing that an interface between TM1 & 3 is more stable than that of TM3 observed from NMR studies.[60] Note that in the present study, this result was reached without making any hypothesis on a putative similarity with the dimer interface observed *R. sphaeroides*, but was obtained consecutive to an exhaustive sampling.

As mentioned above, whatever the followed protocol and even when the dimers are formed, the complex remains highly flexible, as well at the level of the interfaces as for the monomers themselves. This is more clearly visible, when average contact maps for the ensemble of all the

dimer trajectories are calculated and compared with the average contact map for the 20 NMR models available for mouse TSPO. The average NMR topology is largely conserved in the ensemble of dimer trajectories (Fig. 5 A & C), as indicated by conservation of contacts between different TM regions. However, we also observed that while initial contacts are conserved, new contacts were also formed in the average ensemble of dimer trajectories, e.g TM regions 1 with 5 and 3 with 5 (Fig 5, B & D). The fact that TM regions initially located far from each other, become in contact suggests that the overall topology of the protein has become more compact. This may be attributed to the absence of ligand that was present in the experimental structure but not included in the present calculations.

In order to verify the stability of the main dimers observed in the CG simulations and assess the robustness of our observations with respect to the force field and protocol chosen, all-atom simulations have been realised for three representative interfaces: the two main ones in terms of contact number (TM1-TM3/TM1-TM3 and TM2-TM5/TM2-TM5) and a third one, less representative (TM1-TM2/TM3-TM4). On the 200 ns duration of those simulations, we observed that the three dimers remain stable all along the simulations. Interactions are mainly hydrophobic, involving σ-π and aromatic interactions (such as depicted in Figure 4B), but some H-bonds interactions are also present. The predominance of aromatic contacts was still maintained, which means that it seems not due to an artefact of the force field chosen, confirming the agreement with experimental results.

A putative cholesterol translocation mechanism was observed:

As TSPO is assumed to be a cholesterol transporter, we attempted to investigate the behavior of the cholesterol molecules with respect to the different TM regions and more specifically in relation with the dimer formation. First, we analyzed the location of cholesterol molecules with respect to the TM domain in the whole set of simulations considering a distance criterion. Whatever the threshold chosen, we did not notice any specific location with respect to the two

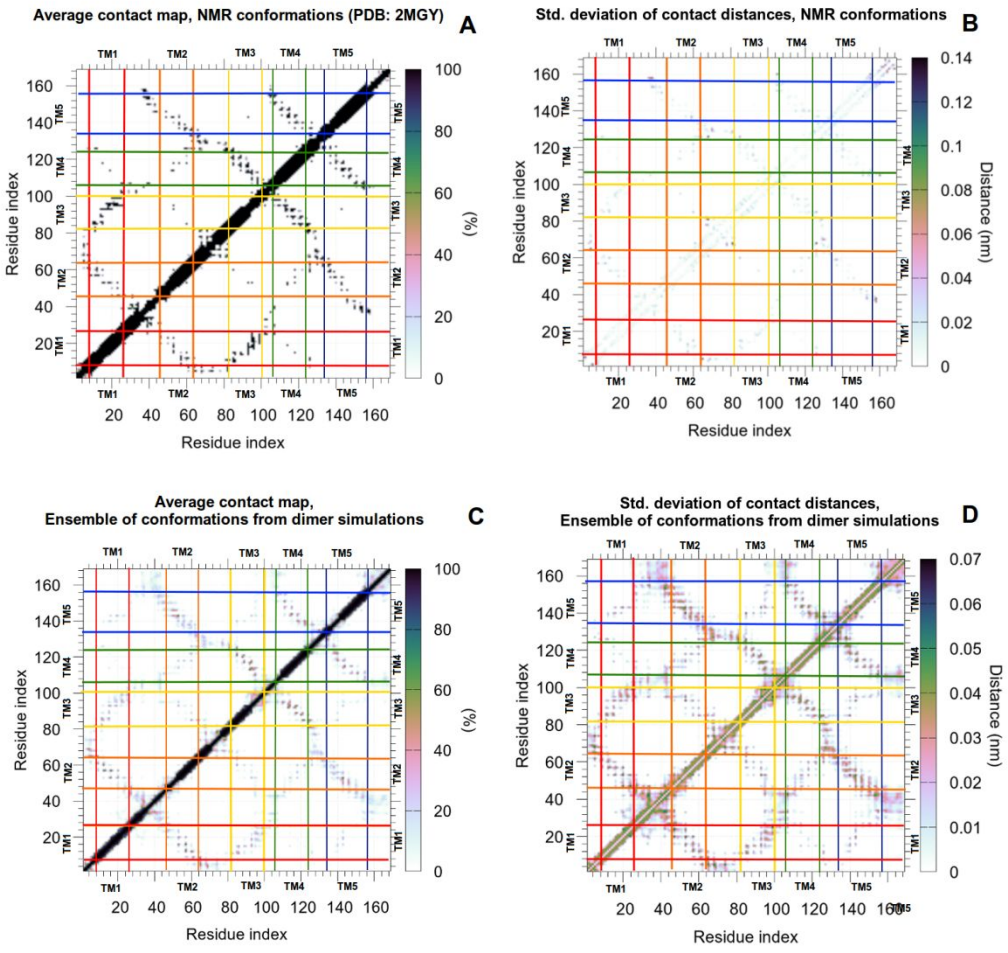

Figure 5 A: Plot showing average contact map of conformations from the NMR structure of mouse TSPO (PDB ID: 2MGY). **B:** Plot showing standard deviation of contact distances from the average contact map of conformations from the NMR structure. **C:** Plot showing average contact map of ensemble of conformations from all the simulations involving dimer state. **D:** Plot showing standard deviation of contact distances from the contact map.

well-known cholesterol-binding motifs, CRAC (residues 149-157) and its reverse counterpart CARC (residues 135-141), located in TM5, in comparison to the rest of accessible residues. Cholesterol molecules are spread homogeneously in the membrane despite the presence of dimers. Yet, looking thoroughly at the simulations, we found several cholesterol molecules with a long residence time, close to the putative functional motifs. While many cholesterol molecules that were close to CRAC motif diffused away along the *x*-*y* plane, those cholesterol molecules were translocating between the lower and upper leaflet, by means of a flipping mechanism involving several residues of the protein. Some of them interacts with A147 residue of TM5 (Figure 6.A&B). Flipping of cholesterol between the two leaflets has already been observed by means of molecular

dynamics simulations, in absence of any proteins and we actually observed such a flipping mechanism for many other cholesterol molecules, far away from the protein. Consequently, we have calculated the ratio of flipping cholesterol molecules in all the membrane over the total number of cholesterol and the ratio of flipping cholesterol molecules close to the dimer over the total number of cholesterol molecules near the protein, respectively. Averages on all the 19 CG simulations in which dimerization occurs show that the overall ratio within the membrane is about 1.24% of flipping cholesterol, and about 3.4% near the TSPO dimers.

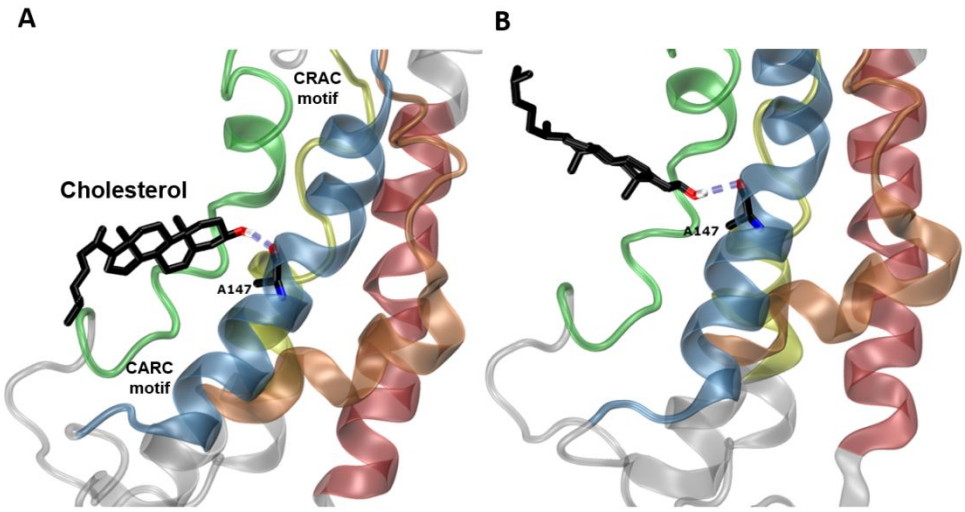

Figure 6 A: Position of cholesterol next to the TM5 (interacting with the A147 residue) in the lower leaflet during our simulations. B: Position of cholesterol molecule after the flipping mechanism, induced by the TM5, in the upper leaflet.

Results over all the simulations shown that the average ratio is higher when the cholesterol is near the protein compared to those spread over the full membrane (about three times higher in average). It means that TSPO could facilitate the translocation of cholesterol within the membrane. In the several simulations, including the all-atom ones, we do observe that some residues of TM3 and TM5 of both monomers are involved in the translocation mechanism, such as depicted in Figure 6. Thus, it suggests that TSPO would facilitate cholesterol translocation, using residues of the protein as a hook, compared to pure membrane. This observation requires further investigation to be definitely confirmed by all-atom simulations that would be more appropriate in this case.

Different long-range communication pathways were found:

The significant motions observed for the whole set of trajectories question on how they impact key functional regions. Accordingly, we calculated dynamic cross-correlations matrix (DCCM) for each trajectory and performed a protein structural network (PSN) analysis that has proved to be a useful approach to pinpoint long-range interaction pathways in protein dynamics. Depending on the simulations, the DCCM pattern varies but in any case, long-range interactions exist, either inside one monomer or between the two chains. For 4 simulations, we observe *intramonomer* long-range interactions that involve several parts of one or both monomer but with no communications between them (see Supp. Figure 5). In these cases, TM1, TM2 and TM5 are mostly involved.

For the 5 others simulations, intermonomer communication takes place through the dimer interface and could involve several parts of both chains. Especially, the parts of the two chains located in the lower leaflet are mainly involved in those intermonomer communications, and imply TM1, TM3 and TM5 for both monomers in majority (Figure 7) and TM2 in a lesser extent (cf Supp. Figure 5).

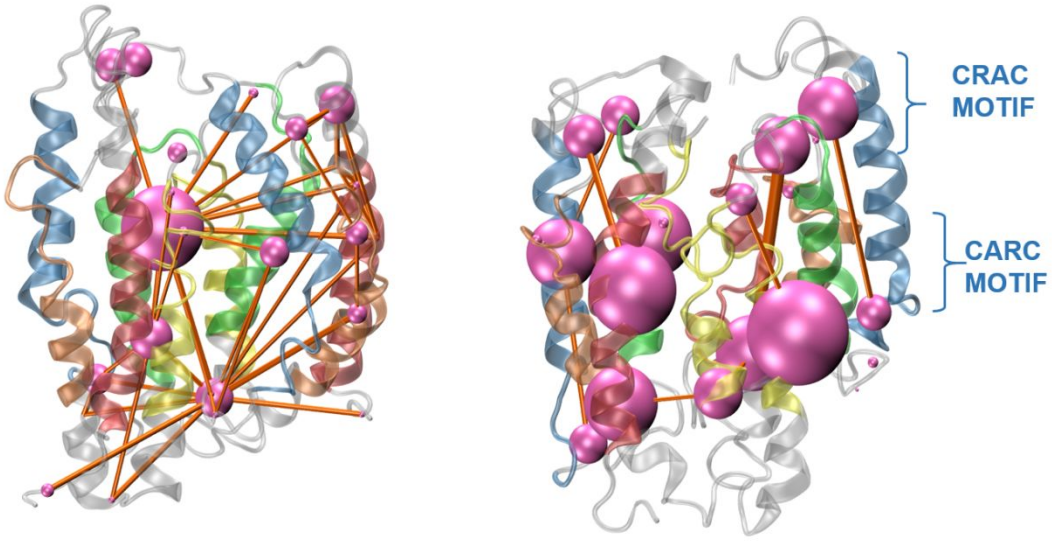

Figure 7: TSPO dimer structures indicating allosteric communication pathways in dimer orientations from different simulations in 33% cholesterol concentration, in both simulations with dimerization process (left) or already formed dimer (right). TM regions 1 to 5 are represented as red, orange, yellow, green and blue ribbons respectively. Communication nodes (pink dots) and allosteric communication pathways (orange cylinder) were traced from the matrix containing correlations of residue motions from different simulations, using Bio3d MD analysis package.[52] Their size are proportional to their importance. A cut-off of 0.7 is chosen for representing the most important connection of the pathway.

Intermonomer communication dominates in those simulations. It appears that these interfaces involve TM1 and TM3, and the predicted network shed light on a connection with those TMs and TM5. However, while the allosteric pathway described by Jaipuria et. al. involved the residues of dimer interface (83-87), residues of TM2, 4 and the CRAC motif (TM5), we found that residues of CARC motif are also involved in this pathway (residues 135 and 136 of TM5) [14,15]. Despite this slight difference, the observed connection between the TM1-TM3 interface and the TM5 can activate its own dynamics, and especially the CRAC and CARC motifs, allowing the cholesterol to translocate more easily.

Finally, it is worth to mention that connecting loops between the TMs are also involved sometimes in the intramonomer or intermonomer communications.

DISCUSSION

In the present work, the simulations have been performed starting with the NMR model. However, this structure is suspected not to correspond to the native folded structure, due to the DPC detergent used in the experiment. As it was recently argued by Xia et al., a homology model based on the *R. Sphaeroides* X-ray structure would better reflect the folded state while being compatible with NMR data.[61] However, the very long simulations that we performed in the presence of a DMPC bilayer, underline high dynamics properties in absence of ligand. To check whether this might be an artefact due to some induced-detergent defects of the initial structure, we compared in a CG resolution, the dynamics of the *R. Sphaeroides* monomer with the mouse monomer, both in apo form. Starting from PDB ID: 4UC1, which possess a better resolution, we build a model of *R. Sphaeroides* TSPO to reverse the A139T mutation to the native sequence. PCA analysis on the two simulations reveal that an overlap exists between both dynamics (overlap between both PC 2) and the conformations explored (unpublished results, PhD Thesis of Rajas M. Rao). This overlap shed lights on common dynamical and structural properties between both proteins. This fact allows us to propose that the experimental conditions on the NMR model by itself have not produced artefacts in terms of dynamics. Similarity might be due to CG representation but when we simulated *Bacillus Cereus* using the same protocol and equivalent sampling, the conformational landscape was different. In addition, on the structural point of view, it is important to remind that the biological environment of the bacterial is quite different from the mammalian form, because of the lack of cholesterol molecules in the bacterial mitochondria membrane. It is well established that membrane composition affects protein structure, dynamics and oligomerization. Thus, such

differences could imply structural modifications compared to the mammalian forms. On the sequence level, the shortened loop between TM1 and TM2 on the bacterial sequence can also have an impact on the stability of the tertiary structure. Consequently, a single structure, *i.e.* neither the NMR model nor the X-ray based-structure is able to reflect this diversity of conformational states, which seems to be the hallmark of the protein in absence of ligand. The presence of ligand in the experimental structures may have contributed to trap only one conformation amongst a huge number of possibilities. Accordingly, the large conformational sampling obtained in the present study thanks to the coarse-grained force field might better represent the states that can be adopted by the protein, which is also reflected in the different assemblies obtained experimentally.

Consequences of interfaces diversity: existence of higher order oligomers

In the context of dimerization, the large set of simulations, which differ in starting structures and velocities, shows that TSPO may form a large diversity of stable dimers by combining different TM regions as interfaces. Our study is the first work that conducts an exhaustive study of the TSPO dimerization process. It demonstrates that stable TSPO dimers can be formed. Interfaces not detected in any of the experimental studies were obtained, suggesting that our simulations have sampled a wider interaction space than the experiments. The presence of multiple complexes suggests that the interaction between the TSPO monomers may be opportunistic, or more interestingly, it may hint towards existence of TSPO higher-order oligomers. Indeed, the diversity of interfaces area, number and nature of TMs involved is quite large and compatible with high-order oligomer organisation. Notably, this diversity in oligomeric state was already observed in X-ray structures and NMR studies.[36,37] As an example, the asymmetric unit of *R. Sphaeroides* TSPO crystal, for which the authors also suggest the existence of higher-order TSPO oligomers, contains three chains that are involved in two pairs of dimers. A first large interface involves TM1-TM3:TM3-TM1, which is also one of the most preferred in our simulations. The second interface slightly smaller is formed by TM1-TM2:TM5.[36] This might be an artefact of crystallization conditions (as described by the authors themselves) but in any case, it shows that significant forces stabilize this assembly. Interestingly, the existence of a large number of TM2-TM5 contacts in our simulations (about 22% of the contacts sampled) strengthen the possibility of a real stable interface. In summary, we observed presence of both the interfaces observed for *R. Sphaeroides* in two separate simulations. In contrast, the interface of *B.cereus* TSPO dimer, involving TM2

from each monomer was indeed observed in our study, but with a low percent of our conformations sampled compared to the *R. Sphaeroides* interfaces.[37] Eventually, the dimer model of TSPO dimer observed from NMR studies,[39] which we attempted to reproduce, did not persist in our simulations in its native state but we found that a similar long-range mechanism exists between the dimer interface and the TM5 region, as suggested in NMR studies. Apart from these X-ray and NMR structures, larger assemblies have also been documented in previous experimental studies.[14,38]

Overall, our simulations are supported by experimental observations and show that dimer formation can be accomplished through a wider range of possibilities. AA simulations reveal a good stability of our dimers sampled by CG simulations and assess our results obtained with the Martini force field, despite its known weaknesses. To go further, metadynamics-like simulations combined with CG simulations could be employed to assess quantitatively the energy of the several dimers we sampled.[62]. Nevertheless, convergence of free energy calculations requires a huge sampling of all the degrees of freedom, which is difficult to attain considering the size of the system, even using biased-methods.

Importance of aromatic residues in a tryptophan-rich protein

Solid-state NMR experiments have proposed residues 83 to 87 to form a dimerization motif, the so-called 'GxxxG' motif.[39] A similar motif is also involved in the dimer interface in *B. cereus* TSPO.[37] GxxxG motifs have been observed to be part of dimer interface in a variety of TM proteins,[57,63-67] such that they have been often hypothesized as the primary drivers of interactions between TM helices.[68,69] This motif, where Glycine is the key element, has been extended to consider other small residues leading to the general motif referred to as 'SmxxxSm' motif,[59,70] which has been observed to be abundant in TM proteins. The interaction by such motifs involves hydrogen bonds between the backbone atoms of the small residues.[67,69,71] Aromatic residues have also been documented to mediate TM dimerization. In certain proteins, tryptophan residues have been found to be a primary driver of oligomerization, by means of planar π-π interactions.[72,73] Other aromatic residues such as phenylalanine and tyrosine have also been found to be part of oligomer interactions, where aromatic residues supplement the dimerization by SmxxxSm motifs.[74]

In case of TSPO, despite the significant presence of SmxxxSm motifs, they seem to be less preferred in the dimerization interface. In contrast, the interface was preferentially filled of large residues, first Leu-Leu and then mainly aromatics residues, which is not surprising since the

protein comprises a large amount of aromatics residue. Among the aromatic-aromatic interactions, Phe-Phe contacts are the most predominant considering all the simulations, but Tyr-Tyr is also greatly represented. In addition, Tyr138 and Tyr140, which are part of the 'CARC' binding motif[17,18] were also found to stabilize the interface in simulations involving the TM5 in the interface (see Supp. Table 2). Interestingly, in a previous experimental study,[38] TSPO was observed to form oligomers that were mediated by formation of covalent dityrosine interactions, upon treatment with ROS. Moreover, in the case of the symmetrical interfaces, the large existence of TM5-TM5 interfaces strengthen this hypothesis and it could be the first step to the creation of those dityrosine covalent interactions. Our simulations, in good agreement with experimental information, provide a structural and more detailed support to those experiments. However, accessibility of CRAC as well as CARC motifs is affected in such oligomeric arrangement. A possible explanation to reconciliate with putative functional properties, would be that the interface involving TM 1 and 3 may be involved in cholesterol transport, since CRAC as well as CARC motifs are more accessible in such arrangement whereas the dityrosine-mediated interface involving TM5 may be involved in response to ROS.

The fact that NMR interfaces are not fully be retained may be due to different reasons. Firstly, the interface comprising the GxxxG motif in our simulations has not disappeared, but simply evolved to a larger interface, involving now aromatic interactions between TM1 and TM3 and not only TM3. Moreover, our observations agree the work of Zeng et al.,[60] who built an alternative dimer model based on homology modelling with bacterial X-ray structures, to palliate putative deficiencies mentioned for the NMR structure. Their rigid model agrees NMR data. Our simulations retrieve a very similar interface to this model of dimer, without making any hypothesis on homology with bacterial structures. It elegantly completes the initial model proposed by Jaipuria et al.,[39] by underlining the importance of dynamics. In addition, AA simulations have also sampled the formation of those aromatic interactions at the interface, reducing the probability of a bias induced by the Martini Force field.

Another possible explanation may rely on the experimental and simulations conditions that are different. Although we took care to be as close as possible to the NMR conditions, the NMR protein concentration is probably higher than in the simulation. This may have impacted the results. In addition, our simulations have been performed in absence of any ligand. Since we have observed that the presence of a ligand affects the dynamics of a monomer (unpublished data, PhD.

Thesis of Rajas M. Rao), it will contribute to modify the conformational space of the two chains and in consequence, impacts the dimerization process. Simulations of dimerization in the presence of a ligand need to be run to confirm this hypothesis.

Putative translocation mechanism of cholesterol

Our studies have highlighted the existence of a putative translocation mechanism, where cholesterol molecules use TM regions as a tether to flip between the upper and lower lipid bilayer leaflets. This has been illustrated here by the A147 residue of TM5 (figure 6). Interestingly, this residue belongs to the "enhancement motif ", located between the CARC and CRAC motif. However, we also observe this kind of hook mechanism with other TMs residues, such as TM3.

Our statistical analyses tend to illustrate that presence and closeness of TSPO dimer seem to facilitate such process. While this is indeed an interesting and exciting observation about the possible TSPO biological function, existence of this mechanism needs to be corroborated rigorously in future studies. Firstly, it remains to be checked if this mechanism is unique to TSPO, or if it is simply due to change in membrane properties consecutive to the presence of any protein. Indeed, a general explanation would be that the presence of a membrane protein possibly reduces energy barrier associated with flipping between two leaflets. Secondly, the robustness of this observation should be insensitive to the force-field used. In AA simulations on three dimers, we have sampled also such phenomena, which by consequent validates our hypothesis. However, longer all-atomistic simulations should still be run to confirm the relevance of this translocation mechanism, especially in biological-relevant membrane, with physiological cholesterol concentration. Ultimately, the role of TSPO ligands also cannot be discounted. Since several experimental studies have demonstrated an effect of TSPO ligands on cholesterol translocation,[5,7,75] question of whether this flipping mechanism can be observed in ligand-bound environments could also be explored.

Long-range communications may be connected to the cholesterol translocation process

Despite the existence of unique pathways depending on the interface, a few common denominators across all the communication pathways can be extracted. Firstly, the pathway implying inter-chain communication involves transfer of information via the dimer interface. Secondly, TM5 region is involved in communications in all the simulations, regardless of the kind of intercommunication. Insofar as we have observed that CARC motif in TM5 helix might play a role in the cholesterol translocation, a relation between dimerization and cholesterol transfer might be inferred, which

corroborates with some conclusions drawn by Jaipuria et al.,[39] who propose a long-range regulation of the dimerization by cholesterol binding but in the CRAC motif. However, further simulations and analyses, in physiological cholesterol concentration and with presence of ligand, are still needed to make a clear conclusion of this aspect.

Though we observed inter-dimer communication in half of our simulations, intra-monomer communication is also present in certain cases, in particular for the models obtained by diffusion of the monomers. This communication may be initially required for the dimerization process to further occur.

In addition, we also observed that the communication pathways observed in monomer were also present in dimer, suggesting that such communication may take place irrespective of the stoichiometry. For example, TM helices involved in the interdimer communication are also implied in the monomer pathway.

Finally, since all the simulations were performed in absence of any ligand, the effect the binding of a ligand in such communication pathways would be of major interest. Similarly, the impact of cholesterol in these communications, for which preliminary data let us to suspect a significant role, needs to be confirmed.

The above questions could be addressed in future by performing multiple and longer molecular dynamics studies in different environments, *i.e.,* in absence of any cholesterol and in presence of a ligand bound to TSPO. More details should also be gained by performing longer all-atom simulations.

CONCLUSIONS

Despite its involvement in important biological processes, TSPO remains a mysterious protein. In the present study, we have addressed a first important question related to a putative oligomerization process. Here, we did not make any hypothesis on the dimer structure in comparison to what was done before. In contrast, we have conducted the most exhaustive study realized so far on this putative mechanism. By doing so, we have been able to observe the versatility of the dimerization and to bring valuable information on dimer interface and long-range communication pathways. The existence of several ways for the TMs to interact even in a non-symmetrical manner, questions about the possible existence of higher-order TSPO oligomers. We have also provided key elements about the nature of interactions that contribute to the formation of the dimers, in terms of TM nature and interaction type. TM1, TM3 and TM5 seems to be the most preferred for the

dimerization process, and lead to different interfaces. The most famous dimerization motif, *i.e.* GXXXG motif, was indeed observed but actually, aromatic residues seem also to play an important role. The results we obtained also confirm that TSPO has noteworthy dynamical properties that explains the difficulty obtained to get proper characteristics of the protein. Importantly, despite the dynamic nature of TSPO, we were able to capture a putative translocation mechanism of cholesterol between the two lipid leaflets, once the dimer is formed. The next step that will be rather exciting to examine, will consist in exploring in-depth, if the cholesterol translocation is *specifically* associated to TSPO or if it is a more general process linked to the presence of a protein in the membrane. In summary, our findings have brought to light important features of TSPO but have also open up a new set of questions on TSPO dynamics-function relationships, which could be explored in future studies.

Supporting Information. : Supplementary information_revision_Final.pdf

Supplementary Table 1: Summary and details of all the simulations made

Supplementary Table 2: Preferred residue pairings between each TSPO monomer

Supplementary Table 3: Distribution of most frequent type of pair of contacts.

Supplementary Figure 1: Comparison of mTSPO monomer simulations in AA and CG resolutions.
Supplementary Figure 2: Minimum distance between two TSPO chains for the thirteen simulations as a function of time
Supplementary Figure 3: Average RMSF for TSPO in monomer vs. dimer conditions, calculated over the 9 simulations in which we have the dimerization process
Supplementary Figure 4: Number of residues within a chosen distance as a function of time for a set of simulations.
Supplementary Figure 5: Protein Structural Network (PSN) for the 9 simulations in which a dimer is formed

Corresponding Author

***corresponding Author, mail : catherine.etchebest@u-paris.fr**

Present Addresses

†CNRS UMR-7369 Laboratoire Matrice Extracellulaire et Dynamique Cellulaire, Université de Reims Champagne-Ardenne

Author Contributions

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

F.C. is associated to Peaccel. The other authors declare no conflict of interest.

ACKNOWLEDGMENT

Authors are thankful to the GR-EX label of excellence, Région Réunion and European Union for financial support (grant: 210939) The authors also would like to thank CINES (Centre Informatique National de l'Enseignement Supérieur) for computational resources (project: A0040710480).

ABBREVIATIONS

TSPO: TranSlocator PrOtein
TM: Transmembrane helix region
PBR: Peripheral Benzodiazepine Receptor
CRAC: Cholesterol Recognition And Consensus motif
ROS: Reactive Oxygen Species
NMR: Nuclear Magnetic Resonance
CRISPR: Clustered Regularly Interspaced Short Palindromic Repeats
DPC: Dodecylphosphocholine
DMPC: 1,2-dimyristoyl-sn-glycero-3-phosphocholine
CG: Corse-Grained
AA: all-atoms
PSN: Protein Structural Network
DCCM: Dynamic Cross-Correlation Matrix
RMSD: Root-Mean Square Deviation
RMSF: Root-Mean Square Fluctuation
OPM: Orientation of Protein in Membrane
ASA: Accessible Surface Area

Authors will release the atomic coordinates upon article publication.

REFERENCES

[1] Bouyer G., Cueff A., Egée S., Kmiecik J., Maksimova Y., Glogowska E., et al., Erythrocyte peripheral type benzodiazepine receptor/voltage-dependent anion channels are upregulated by Plasmodium falciparum., *Blood*, **2011**, *118*, 2305–2312.

[2] Hardwick M., Fertikh D., Culty M., Li H., Vidic B., Papadopoulos V., Peripheral-type benzodiazepine receptor (PBR) in human breast cancer: correlation of breast cancer cell aggressive phenotype with PBR expression, nuclear localization, and PBR-mediated cell proliferation and nuclear transport of cholesterol., *Cancer Res*., **1999**, *59*, 831–842.

[3] Fan J., Rone M.B., Papadopoulos V., Translocator protein 2 is involved in cholesterol redistribution during erythropoiesis., *J. Biol. Chem*., **2009**, *284*, 30484–30497.

[4] Braestrup C., Squires R.F., Specific benzodiazepine receptors in rat brain characterized by high-affinity (3H)diazepam binding., *Proc. Natl. Acad. Sci. USA*., **1977,** *74*, 3805–3809.

[5] Krueger K.E., Papadopoulos V., Peripheral-type benzodiazepine receptors mediate

translocation of cholesterol from outer to inner mitochondrial membranes in adrenocortical cells., *J. Biol. Chem*., **1990**, *265*, 15015–15022.

[6] Papadopoulos V., Amri H., Li H., Boujrad N., Vidic B., Garnier M., Targeted disruption of the peripheral-type benzodiazepine receptor gene inhibits steroidogenesis in the R2C Leydig tumor cell line., *J. Biol. Chem*., **1997**, *272*, 32129–32135.

[7] Papadopoulos V., Nowzari F.B., Krueger K.E., Hormone-stimulated steroidogenesis is coupled to mitochondrial benzodiazepine receptors. Tropic hormone action on steroid biosynthesis is inhibited by flunitrazepam., *J. Biol. Chem*., **1991**, *266*, 3682–3687.

[8] Li H., Yao Z., Degenhardt B., Teper G., Papadopoulos V., Cholesterol binding at the cholesterol recognition/ interaction amino acid consensus (CRAC) of the peripheral-type benzodiazepine receptor and inhibition of steroidogenesis by an HIV TAT-CRAC peptide., *Proc. Natl. Acad. Sci. USA*., **2001**, *98,* 1267–1272.

[9] Papadopoulos V., Amri H., Boujrad N., Cascio C., Culty M., Garnier M., et al., Peripheral benzodiazepine receptor in cholesterol transport and steroidogenesis., *Steroids*, **1997**, *62*, 21–28.

[10] Papadopoulos V., Baraldi M., Guilarte T.R., Knudsen T.B., Lacapère J.-J., Lindemann P., et al., Translocator protein (18kDa): new nomenclature for the peripheral-type benzodiazepine receptor based on its structure and molecular function., *Trends Pharmacol. Sci.,* **2006**, *27,* 402–409.

[11] Papadopoulos V., Liu J., Culty M., Is there a mitochondrial signaling complex facilitating cholesterol import?, *Mol. Cell. Endocrinol*., **2007,** *265–266*, 59–64.

[12] Jefcoate C., High-flux mitochondrial cholesterol trafficking, a specialized function of the adrenal cortex., *J. Clin. Invest*., **2002**, *110*, 881–890.

[13] Mukhin A.G., Papadopoulos V., Costa E., Krueger K.E., Mitochondrial benzodiazepine receptors regulate steroid biosynthesis., *Proc. Natl. Acad. Sci. USA*., **1989**, *86,* 9813–9816.

[14] Papadopoulos V., Boujrad N., Ikonomovic M.D., Ferrara P., Vidic B., Topography of the Leydig cell mitochondrial peripheral-type benzodiazepine receptor., *Mol. Cell. Endocrinol*., **1994**, *104*, R5-9.

[15] Li H., Papadopoulos V., Peripheral-type benzodiazepine receptor function in cholesterol transport. Identification of a putative cholesterol recognition/interaction amino acid sequence and consensus pattern., *Endocrinology*., **1998,** *139,* 4991–4997.

[16] Fan J., Wang K., Zirkin B., Papadopoulos V., CRISPR/Cas9–Mediated Tspo Gene Mutations Lead to Reduced Mitochondrial Membrane Potential and Steroid Formation in MA-10 Mouse Tumor Leydig Cells., *Endocrinology*., **2018**, *159*, 1130–1146.

[17] Papadopoulos V., Fan J., Zirkin B., Translocator protein (18 kDa): an update on its function in steroidogenesis., *J. Neuroendocrinol*., **2018**, *30*.

[18] Fantini J., Di Scala C., Evans L.S., Williamson P.T.F., Barrantes F.J., A mirror code for protein-cholesterol interactions in the two leaflets of biological membranes., *Sci. Rep*., **2016**, *6*, 21907.

[19] Selvaraj V., Stocco D.M., The changing landscape in translocator protein (TSPO) function., *Trends Endocrinol. Metab*., **2015**, *26* , 341–348.

[20] Morohaku K., Pelton S.H., Daugherty D.J., Butler W.R., Deng W., Selvaraj V., Translocator protein/peripheral benzodiazepine receptor is not required for steroid hormone biosynthesis., *Endocrinology*, **2014**, *155,* 89–97.

[21] Tu L.N., Morohaku K., Manna P.R., Pelton S.H., Butler W.R., Stocco D.M., et al., Peripheral benzodiazepine receptor/translocator protein global knock-out mice are viable with no

effects on steroid hormone biosynthesis., *J. Biol. Chem*., **2014**, *289,* 27444–27454.
[22] Zeno S., Veenman L., Katz Y., Bode J., Gavish M., Zaaroor M., The 18 kDa mitochondrial translocator protein (TSPO) prevents accumulation of protoporphyrin IX. Involvement of reactive oxygen species (ROS)., *Curr. Mol. Med*., **2012**, *12,* 494–501.
[23] Rampon C., Bouzaffour M., Ostuni M.A., Dufourcq P., Girard C., Freyssinet J.M., et al., Translocator protein (18 kDa) is involved in primitive erythropoiesis in zebrafish., *FASEB J*., **2009**, *23* , 4181–4192.
[24] Kugler W., Veenman L., Shandalov Y., Leschiner S., Spanier I., Lakomek M., et al., Ligands of the mitochondrial 18 kDa translocator protein attenuate apoptosis of human glioblastoma cells exposed to erucylphosphohomocholine., *Cell. Oncol*., **2008**, *30*, 435–450.
[25] Maaser K., Höpfner M., Jansen A., Weisinger G., Gavish M., Kozikowski A.P., et al., Specific ligands of the peripheral benzodiazepine receptor induce apoptosis and cell cycle arrest in human colorectal cancer cells., *Br. J. Cancer*., **2011**, *85*, 1771–1780.
[26] Veenman L., Gavish M., The role of 18 kDa mitochondrial translocator protein (TSPO) in programmed cell death, and effects of steroids on TSPO expression., *Curr. Mol. Med*., **2012**, *12,* 398–412.
[27] Veenman L., Shandalov Y., Gavish M., VDAC activation by the 18 kDa translocator protein (TSPO), implications for apoptosis., *J. Bioenerg. Biomembr*., **2008**, *40*, 199–205.
[28] Liu G.-J., Middleton R.J., Hatty C.R., Kam W.W.-Y., Chan R., Pham T., et al., The 18 kDa translocator protein, microglia and neuroinflammation., *Brain Pathol*., **2014**, *24,* 631–653.
[29] Chechneva O.V., Deng W., Mitochondrial translocator protein (TSPO), astrocytes and neuroinflammation., *Neural Regen. Res*., **2016**, *11*, 1056–1057.
[30] Bonsack F., Alleyne C.H., Sukumari-Ramesh S., Augmented expression of TSPO after intracerebral hemorrhage: a role in inflammation?, *J. Neuroinflammation*., **2016**, *13*, 151.
[31] Bernassau J.M., Reversat J.L., Ferrara P., Caput D., Lefur G., A 3D model of the peripheral benzodiazepine receptor and its implication in intra mitochondrial cholesterol transport., *J. Mol. Graph*.,**1993**, *11,* 236–44.
[32] Korkhov V.M., Sachse C., Short J.M., Tate C.G., Three-dimensional structure of TspO by electron cryomicroscopy of helical crystals., *Structure*, **2010**, *18*, 677–687.
[33] Hinsen K., Vaitinadapoule A., Ostuni M.A., Etchebest C., Lacapere J.-J., Construction and validation of an atomic model for bacterial TSPO from electron microscopy density, evolutionary constraints, and biochemical and biophysical data., *Biochim. Biophys. Acta*. , 2015, *1848*, 568–580.
[34] Jaremko L., Jaremko M., Becker S., Zweckstetter M., Toward the functional oligomerization state of tryptophan-rich sensory proteins., *Protein Sci*., **2014**, *23*, 1154–1160.
[35] Jaremko L., Jaremko M., Giller K., Becker S., Zweckstetter M., Structure of the mitochondrial translocator protein in complex with a diagnostic ligand., *Science*, **2014**, *343*, 1363–1366.
[36] Li F., Liu J., Zheng Y., Garavito R.M., Ferguson-Miller S. Crystal structures of translocator protein (TSPO) and mutant mimic of a human polymorphism., *Science*, **2015**, *347,* 555–558.
[37] Guo Y., Kalathur R.C., Liu Q., Kloss B., Bruni R., Ginter C., et al. Structure and activity of tryptophan-rich TSPO proteins., *Science*, **2015**, *347*, 551–555.
[38] Lacapère J.-J., Delavoie F., Li H., Péranzi G., Maccario J., Papadopoulos V., et al., Structural and functional study of reconstituted peripheral benzodiazepine receptor., *Biochem. Biophys. Res. Commun*., **2001**, *284*, 536–541.
[39] Jaipuria G., Leonov A., Giller K., Vasa S.K., Jaremko Ł. , Jaremko M., et al., Cholesterol-

mediated allosteric regulation of the mitochondrial translocator protein structure., *Nat. Commun*., **2017**, *8* ,14893.
[40] Lomize M.A., Pogozheva I.D., Joo H., Mosberg H.I., Lomize A.L., OPM database and PPM web server: resources for positioning of proteins in membranes., *Nucleic Acids Res*., **2012**, *40,* D370-6.
[41] Lomize A.L., Pogozheva I.D., Mosberg H.I., Anisotropic solvent model of the lipid bilayer. 2. Energetics of insertion of small molecules, peptides, and proteins in membranes., *J. Chem. Inf. Model*., **2011**, *51*, 930–946.
[42] de Jong D.H., Singh G., Bennett W.F.D., Arnarez C., Wassenaar T.A., Schäfer L.V., et al., Improved parameters for the MARTINI coarse-grained protein force field. *J. Chem. Theory Comput*., **2013**, *9*, 687–697.
[43] Yesylevskyy S.O., Schäfer L.V., Sengupta D., Marrink S.J., Polarizable water model for the coarse-grained MARTINI force field., *PLoS Comput. Biol*., **2010**, *6*, e1000810.
[44] Monticelli L., Kandasamy S.K., Periole X., Larson R.G., Tieleman D.P., Marrink S.J., The MARTINI coarse-grained force field: Extension to proteins., *J. Chem. Theory Comput*., **2008**, *4*, 819–834.
[45] Marrink S.J., Risselada H.J., Yefimov S., Tieleman D.P., de Vries A.H., The MARTINI force field: coarse grained model for biomolecular simulations., *J. Phys. Chem. B*., **2007**, *111*, 7812–7824.
[46] Javanainen M.; Martinez-Seara H. and Vattulainen I. Excessive Aggregation of Membrane Proteins in the Martini Model. *PLoS One*, **2017**, *12* (11).
[47] Berendsen H.J.C., Postma J.P.M., van Gunsteren W.F., DiNola A., Haak J.R., Molecular dynamics with coupling to an external bath, *J. Chem. Phys*., **1984**, *81*, 3684.
[48] Parrinello M., Polymorphic transitions in single crystals: A new molecular dynamics method, *J. Appl. Phys*., **1981**, *52,* 7182.
[49] Berendsen H.J.C., van der Spoel D., van Drunen R., GROMACS: A message-passing parallel molecular dynamics implementation, *Comput. Phys. Commun*., **1995**, *91*, 43–56.
[50] Páll S., Abraham M.J., Kutzner C., Hess B., Lindahl E., Tackling Exascale Software Challenges in Molecular Dynamics Simulations with GROMACS, *in Solving Software Challenges for Exascale*, S. Markidis, E. Laure, Eds., Springer International Publishing, Cham, **2015**, pp. 3–27.
[51] Van Zundert G.C.P., Rodrigues J.P.G.L.M., Trellet M., Schmitz C., Kastritis P.L., Karaca E., et al., The HADDOCK2.2 Web Server: User-Friendly Integrative Modeling of Biomolecular Complexes., *J. Mol. Biol*., **2016**, *428*, 720–725.
[52] Dominguez C., Boelens R., Bonvin A.M.J.J., HADDOCK: a protein-protein docking approach based on biochemical or biophysical information., *J. Am. Chem. Soc*., **2003**, *125*, 1731–1737.
[53] Grant B.J., Rodrigues A.P.C., ElSawy K.M., McCammon J.A., Caves L.S.D., Bio3d: an R package for the comparative analysis of protein structures., *Bioinformatics*, **2006**, *22*, 2695–2696.
[54] Wassenaar T.A., Pluhackova K., Böckmann R.A., Marrink S.J., Tieleman D.P., Going Backward: A Flexible Geometric Approach to Reverse Transformation from Coarse Grained to Atomistic Models., *J. Chem. Theory Comput*., **2014**, *10*, 676–690.
[55] S.J. Hubbard, Thornton J.M. Naccess, Computer program., **1993,** Department of Biochemistry and Molecular Biology, University College, London.
[56] Mercadante D., Gräter F., Daday C., CONAN: A Tool to Decode Dynamical Information

from Molecular Interaction Maps., *Biophys. J.*, **2018**, *114*, 1267–1273.
[57] Zhou F.X., Merianos H.J., Brunger A.T., Engelman D.M., Polar residues drive association of polyleucine transmembrane helices., *Proc Natl Acad Sci USA*., **2001**, *98*, 2250–2255.
[58] Lemmon M.A., Flanagan J.M., Treutlein H.R., Zhang J., Engelman D.M., Sequence specificity in the dimerization of transmembrane alpha-helices., *Biochemistry*, **1992**, *31*, 12719–12725.
[59] Duneau J.-P., Vegh A.P., Sturgis J.N. A dimerization hierarchy in the transmembrane domains of the HER receptor family., *Biochemistry*, **2007**, *46*, 2010–2019.
[60] Zeng J., Guareschi R., Damre M., Cao R., Kless A., Neumaier B., et al., Structural prediction of the dimeric form of the mammalian translocator membrane protein TSPO: A key target for brain diagnostics., *Int. J. Mol. Sci*., **2018**, *19*.
[61] Xia Y., Ledwitch K., Kuenze G., Duran A., Li J., Sanders C.R., Manning C., Meiler J. A unified structural model of the mammalian translocator protein (TSPO). *J. Biomol. NMR,* **2019**, *73*, 347-364
[62] Lelimousin M., Limongelli V., Sansom M.S.P., Conformational Changes in the Epidermal Growth Factor Receptor: Role of Transmembrane Domain Investigated by Coarse-Grained Metadynamics Free Energy Calculations. *J. Am. Chem. Soc*., **2016**, *138*(33), 10611-10622.
[63] Jaremko M., Jaremko Ł., Jaipuria G., Becker S., Zweckstetter M., Structure of the mammalian TSPO/PBR protein, *Biochem. Soc. Trans*., **2015**, *43*, 566–571.
[64] Sulistijo E.S., Mackenzie K.R., Structural basis for dimerization of the BNIP3 transmembrane domain., *Biochemistry*, **2009**, *48*, 5106–5120. doi:10.1021/bi802245u.
[65] Lemmon M.A., Flanagan J.M., Hunt J.F., Adair B.D., Bormann B.J., Dempsey B.E., et al., Glycophorin A dimerization is driven by specific interactions between transmembrane alpha-helices., *J. Biol. Chem*., **1992**, *267* , 7683–7689.
[66] Pogoryelov D., Klyszejko A.L., Krasnoselska G.O., Heller E.-M., Leone V., Langer J.D., et al., Engineering rotor ring stoichiometries in the ATP synthase., *Proc Natl Acad Sci USA*, **2012**, *109*, E1599-608.
[67] Li E., Wimley W.C., Hristova K., Transmembrane helix dimerization: beyond the search for sequence motifs., *Biochim. Biophys. Acta*., **2012**, *1818*, 183–193.
[68] Kleiger G., Grothe R., Mallick P., Eisenberg D., GXXXG and AXXXA: common alpha-helical interaction motifs in proteins, particularly in extremophiles., *Biochemistry*, **2002**, *41*, 5990–5997.
[69] Mueller B.K., Subramaniam S., Senes A., A frequent, GxxxG-mediated, transmembrane association motif is optimized for the formation of interhelical Cα-H hydrogen bonds., *Proc Natl Acad Sci USA*, **2014**, *111*, E888-95.
[70] Eilers M., Patel A.B., Liu W., Smith S.O., Comparison of helix interactions in membrane and soluble alpha-bundle proteins., *Biophys. J*., **2002**, *82*, 2720–2736.
[71] Weber M., Schneider D. Six amino acids define a minimal dimerization sequence and stabilize a transmembrane helix dimer by close packing and hydrogen bonding., *FEBS Lett*., **2013**, *587*, 1592–1596.
[72] Zoued A., Duneau J.-P., Durand E., España A.P., Journet L., Guerlesquin F., et al., Tryptophan-mediated Dimerization of the TssL Transmembrane Anchor Is Required for Type VI Secretion System Activity., *J. Mol. Biol*., **2018**, *430*, 987–1003.
[73] Ridder A., Skupjen P., Unterreitmeier S., Langosch D., Tryptophan supports interaction of transmembrane helices., *J. Mol. Biol*., **2005**, *354*, 894–902.
[74] Unterreitmeier S., Fuchs A., Schäffler T., Heym R.G., Frishman D., Langosch D.,

Phenylalanine promotes interaction of transmembrane domains via GxxxG motifs., *J. Mol. Biol*., **2007**, *374*, 705–718.

[75] McCauley L.D., Park C.H., Lan N.C., Tomich J.M., Shively J.E., Gee K.W., Benzodiazepines and peptides stimulate pregnenolone synthesis in brain mitochondria., *Eur. J. Pharmacol*., **1995**, *276*, 145–153.