Construction of dimeric hTSPO protein model using homology modeling and molecular dynamics

Hien T T Lai and Toan T Nguyen∗
Key Laboratory for Multiscale Simulation of Complex Systems, VNU University of Science, Vietnam National University - Hanoi, 334 Nguyen Trai street, Thanh Xuan district, Hanoi, Vietnam
E-mail: toannt@hus.edu.vn, toannt@vnu.edu.vn

Abstract. A model for the dimeric form of the human TSPO (hTSPO) protein is constructed homologically using the RsTSPO dimer template. Then, Molecular dynamics simulation of 1µs is carried out on the model to investigate its stability as well as to study its various physico-chemical properties. Our model shows good stability, with various important conserved residues involving in monomer–monomer interactions, in good agreements with available experimental data. Several functioning motifs are well demonstrated. This model hence can serve as a good basis for further study of this system as well as for discoveries of candidate radioligands targeting TSPO protein as a biotracer for medical imaging of brain inflammation.

1. Introduction
The translocator protein TSPO 18kDa is an important membrane protein that is expressed at high levels in the outer mitochondrial membrane of steroidogenic cells of the nervous system [1, 2, 3]. It is known to show elevated expression in response to a variety of brain inflammations such as cancers, Alzheimer’s and Parkinson’s diseases, disorders such as depression and anxiety [4]. Therefore, it has been exploited as a biotracer for brain inflammation using highly specific synthetic ligands. In this aspect, the radioactive PK11195 ligand has been widely used for brain imaging. However, this ligand has a high lipophilicity and high non-specific binding which makes it difficult to quantify uptake, and efforts are being made to develop more sensitive radioligands for this purpose [5, 6, 7, 8].

Recently, molecular structures of TSPO protein of various species have been resolved experimentally leading to renewed research interests in understanding structural and dynamical properties of this protein [9, 10, 11]. Among them, one particular recent discovery is the role of cholesterol in regulating the structure and function of this membrane protein. It is demonstrated experimentally that cholesterol can shift the equilibrium of TSPO dimer toward monomer dissociation [12]. However, the molecular mechanism as well as physical understanding of this cholesterol effect is still very limited. In this work, as a first step toward better understanding of the mechanism, we homologically model the hTSPO dimer based on the Rhodobacter sphaeroides TSPO (RsTSPO) dimer template [9] and carry out molecular dynamics simulation to investigate its stability and various physico-chemical properties. Our model is stable during a long 1µs molecular dynamics simulation. The monomer–monomer interface shows various important interactions in good agreements with various experimental results. Hence, our model can serve as a basis for further investigations of these systems.
2. Materials and methods

2.1. Prepare the model and the simulation system

The amino acid sequence of hTSPO was exported from UniProtKB web-server [13] with ID P30536. It was aligned with sequences of various species: Rhodobacter sphaeroides TSPO (RsTSPO, PDB ID: 4UC1) [9], Bacillus cereus TSPO (BcTSPO, PDB ID: 4RYI) [10] and Mus musculus TSPO (MoTSPO, PDB ID: 2MGY) [11]. We set cut-off values of 95% sequence identity and 35% sequence identity for fully conserved residues and semi-conserved residues, respectively. The weight matrix BLOSUM [14] was chosen for sequences alignment. In this work, RsTSPO [9] was chosen as the template for our hTPSO dimer model. The 3D structure of hTPSO dimeric model based on this template was built using Python scripts and Modeller library [15, 16]. Out of the 20 models that was suggested, the best model that has lowest Discrete optimized protein energy (DOPE) score is chosen for further molecular dynamics simulation study [17].

For simulation system, a lipid bilayer membrane was constructed from a mixture of phosphatidylcholine (POPC), phosphatidylethanolamine (POPE) and cholesterol (CHL) with the number ratio of 3:3:1, respectively [18, 19]. Simulation box was generated for system with the dimensions of 13.2nm × 13.2nm × 20.4nm. Water molecules were filled to the simulation box outside the membrane. Lastly, the system was neutralized by randomly replace water molecules by ions of NaCl monovalent salt at the physiological 150mM concentration.

2.2. Molecular dynamics simulation

MD simulations are carried using the GROMACS 2018.6 software package [20]. The forcefields AMBER99SB-ILDN [21], Slipids [22] and TIP3P [23] are used to parametrize proteins and ions, lipids, and water molecules due to their compatibility. Particle Mesh Ewald method was used for treatment of long-range electrostatic interaction[24]. Semi-isotropic pressure coupling was chosen to uniformly scale the x−y axes, while independently scaling the z axis. Cutoff radius of 1.2nm is used for treating short−range electrostatic and van der Waals interactions.

The system is first optimized using steepest descent minimization method. After that, 20ns of simulated annealing is carried out, raising the temperature from 2K to 310K with 50 annealing points. A NVT-ensemble simulation of 50ns was then carried out at the reference temperature of 310K using the V-rescale thermostat [25]. This is followed by a NPT-ensemble simulation of 50ns at the same temperature and at the reference pressure of 1 atm using the Berendsen barostat. Finally, 1 μs of MD production run are conducted at the same temperature and pressure but using the equilibrium Parrinello-Rahman barostat and the Nosé-Hoover thermostat[25, 27].

For analyses, frames were collected every 100 ps. Various tools within the GROMACS package, the VMD program [2], and some in-house Python scripts were used for these tasks (and also for performing various tasks in setting up the system).

3. Results and discussion

3.1. The homologically constructed hTSPO dimeric model

Although the MoTSPO structure [11] has higher sequence identity (81%) with hTSPO model, it has been known that the experimental construction of this model has strong bias that affects the fidelity and stability of the model [28]. On the other hand, despite lower sequence identity (29%), important functional and structural regions of RsTSPO sequence [9] are conserved in hTSPO [29]. Therefore, in this work, we construct the hTSPO dimer model using the RsTSPO template (PDB ID: 4UC1) [9].

The resultant homologically constructed model of hTSPO protein dimer built using RsTSPO template [9] is shown in Fig. 1. The model composes of two sequence−identical monomeric TSPO proteins, called sub-unit A and sub-unit B. As shown in Fig. 1(a), both sub-units have five transmembrane (TM) helices arranged consecutively in clockwise direction as TM I, TM II,
Figure 1. The structural model of hTSPO dimeric protein. (a) Top view - hTSPO dimer contains two sequence-identical monomeric structures with five transmembrane helices arranged in the clockwise order as TM I – TM II – TM V – TM IV – TM III. (b) Side view - this model (orange and red cartoon representation) is built using RsTSPO template (black tube representation) with their interface (red colored) between sub-units. (c) The interface regions in each sub-unit: all TM III, the N-terminal part of TM II and the C-terminal parts of TM I and TM IV. (d) The interface surface of the hTSPO dimeric model.

TM V, TM IV, and TM III (the order in the name of the helices comes from their order in the primary sequence of the hTSPO protein).

The helices in the same monomer are connected by flexible loops. The LP I, LP III and LP V loops are in the cytoplasm, while the LP II and LP IV loops are in the cytosol. Fig. 1(b) shows that these helices and loops in our model (orange and red colors) matched similar helices and loops of its RsTSPO template (black curve) respectively, except for LP I and LP II loops. The conformations of LP I and LP II are different because the sequences of LP I and LP II in hTSPO are longer by three residues (G30 - R32) and five residues (G72 - K77), respectively, than those in RsTSPO sequence. However, these additional residues in the connecting loops of hTSPO
Figure 2. The stability of hTSPO dimer model in MD simulation: (a) the root–mean–square deviation of backbone $C_{\alpha}$ atoms as a function of time; (b) the root–mean–square fluctuation of the $C_{\alpha}$ atom of each residue; and (c) the number of hydrogen bonds of the monomer–monomer interface as a function of time.

sequence do not affect significantly the interface between the sub-unit A and the sub-unit B of the hTSPO dimer. Indeed, as Fig. 1(b) shows, the interface between sub-units of hTSPO model (red color) closely matches the same interface in its RsTSPO template (black curves).

The two monomers of the hTSPO dimer touch each other at the same region in each sub-unit, shown by the red color in Figs. 1(c) and 1(d). The structural model contains two parts, the C-terminus located in the upper part and in the outer membrane leaflet, and the N-terminus located in the lower part and in the inner membrane leaflet. As one can see from Fig. 1(c), the monomer–monomer interface of hTSPO model in each sub-unit composed of all outer residues of the large helix – TM III, and the upper part of TM I, and the lower part of TM II, similar to its RsTSPO template. Because of these two components (upper and lower parts) of the interface, the monomers in our model are kept together by both C- and N-termini of the proteins. This is to contrast with that of MoTSPO model [11] where the monomers are kept together only by the N-terminus; or to contrast with that of BcTSPO model [10] where the monomers are kept together only by the C-terminus. Therefore, RsTSPO structure [9] is a good candidate template for building hTSPO model [28], especially for oligomerization states that have not been known.

3.2. MD simulation analysis
Next, we investigate the stability of our model by carrying out an MD simulation of the system in a typical membrane for 1000ns. The result of root–mean–square deviation (RMSD) calculation for backbone $C_{\alpha}$ atoms of the proteins is shown in the Fig. 2(a). In this figure, the RMSD value reaches a plateau around 200ns indicating that equilibrium has been established. In later analyses, we drop the first 200ns from statistics, since this is the relaxation time of our protein. Note that the saturated RMSD value is only about 0.3nm for our TSPO model. This value is lower than RMSD values for typical proteins and is a good indicator that our model is well...
constructed and stable.

We also analyze flexibility of each residue by calculating the root–mean–square fluctuation (RMSF) of the backbone C\(_\alpha\) atoms. Fig. 2(b) shows RMSF values of C\(_\alpha\) atoms of sub-unit A (purple line) and of sub-unit B (green line). Despite the fact that the monomers are identical in sequence, and very similar in structure, the degree of flexibility of same residues in each sub-unit are not the same. In the sub-unit A, the RMSF values of TM I (4 - 23), TM II (45 - 70) and LP I (24 - 44) are larger than those in sub-unit B. Oppositely, the remaining TMs of sub-unit A are more stable. Such symmetry breaking among identical monomers in a dimer is not uncommon, and is usually contributed to the strong interaction among the monomers [30]. Some regions have rather large RMSF values of more than 1.5Å. These are expected because most of these residues are located in the flexible loops connecting the helices: LP II - III - IV with residue ranges 70 - 77, 103 - 107 and 128 - 134, respectively.

Interestingly, in sub-unit A the TM I is quite flexible, but not uniform along the helix. The first half (the lower, N–terminal part of TM I) shows high flexibility (RMSF higher than 1.5Å), while the last half (the upper, C–terminal part of TM I) shows only about 1Å in RMSF values. This phenomenon is also observed for TM II but in opposite order. The RMSF values decrease from 1.8Å in the upper, C–terminal part to 0.8Å in the lower, N–terminal part. The different and opposite behaviors in fluctuation values between C– and N–termini of TM I and TM II is due to residues in the upper part of TM I and in the lower part of TM II of this sub-unit interacts with these residues in the other sub-unit, a part of the monomer–monomer interface of our hTSPO model. These interactions reduce significantly fluctuations of the residues in the corresponding half of the helices.

As already mentioned, the monomer–monomer interface of our hTSPO model is composed of whole TM III helix and upper parts of TM I and TM IV, together with the lower part of TM II helices of both sub-units. This interface is stable at all time during our simulation thanks to the hydrogen bonds (H–bonds) among opposite residues (Fig. 2(c)). The interface has about 10 H–bonds initially. This number increases to and saturates at about 20 H–bonds after about 200ns. Therefore, the interface in our hTSPO model is very strong and stable.

Lastly, our model has two motifs, G83xxxG87 and W95xPxF99 that are commonly found at the monomer–monomer interface of some proteins, including eukaryotic and prokaryotic TSPOs [28, 31, 32, 33, 34, 35, 36]. In our simulation, the side chains in these motifs point out of the interface, toward and interact with residues of the other sub-unit during the whole simulation.

The strong interaction between the monomers with increasing hydrogen bonds and stable interacting motifs observed during simulation is also in good agreement with results of previous experiments and simulations of another mammalian TSPO, the MoTSPO [12, 28, 37]. They show that the functional form of MoTSPO is the dimeric form, and its NMR structure (PDB ID: 2MGY) is monomer because of detergent’s condition of the experiment effects its structure.

4. Conclusion

In this work, a model for the hTSPO dimer protein is built using the RsTSPO template. This model is not only structurally stable during a long 1µs simulation time, but also show strong interaction between sub-units, especially with regards to the number of structurally–important hydrogen bond interactions at the interface. Analyses of simulation results show that our model agrees well with various experimental data. We believe this is a good model for discoveries of structural – functional relationships of human TSPO protein, which will be our aims in the near future. Along this future direction, some preliminary investigations of our group on the effects of cholesterol binding also show very good agreement with experiment results (manuscript in preparation).

The authors acknowledge financial supports of the Vietnam National University - Hanoi under grant number TXTCN.20.03 and the World Bank and the Ministry of Science and Technology.
of Vietnam under grant number 13/FIRST/1.a/VNU1.

References
[1] Laurent B, Chavent M, Cragnolini T, Dahl A C E, Pasqualli S, Derreumaux P, Sansom M S and Baaden M 2015 *Bioinformatics* 31 1478–1480
[2] Humphrey W, Dalke A and Schulten K 1996 *J. Mol. Graph.* 14 33–38
[3] Ghahezeinampour M M, Arab S S, Aghazadeh S B, Zhang J and van der Spoel D 2014 *Bioinformatics* 30 439–441
[4] Rupprecht R, Papadopoulos V, Rammes G, Baghai T C, Fan J, Akula N, Groyer G, Adams D and Schumacher M 2010 *Nat. Rev. Drug Discov.* 9 971
[5] Myers R, Manjil L G, Cullen B M, Price G W, Frackowiak R S and Cremer J E 1991 *J. Cereb. Blood Flow Metab.* 11 314–322
[6] Chauveau F, Boutin H, Van Camp N, Dollé F and Tavitian B 2008 *Eur. J. Nucl. Med. Mol. Imaging* 35 2304–2319
[7] Werry E L, Bright P M, Pignet O, Ittner L M, Halliday G M, Hodges J R, Kiernan M C, Loy C T, Kril J J and Kassiou M 2019 *Int. J. Mol. Sci.* 20 3161
[8] Best L, Ghadery C, Pavese N, Tai Y F and Strafella A P 2019 *Curr. Neurol. Neurosci. Rep.* 19 24
[9] Li F, Liu J, Zheng Y, Garavito R M and Ferguson-Miller S 2015 *Science* 347 555–558
[10] Guo Y, Kalathur R C, Liu Q, Kloss B, Bruni R, Ginter C, Klopkmann E, Rost B and Hendrickson W A 2015 *Science* 347 551–555
[11] Jaremko Ł, Jaremko M, Giller K, Becker S and Zweckstetter M 2014 *Science* 343 1363–1366
[12] Jaipuria G, Leonov A, Giller K, Vasa S K, Jaremko Ł, Jaremko M, Linser R, Becker S and Zweckstetter M 2017 *Nat. Commun.* 8 14893
[13] Consortium U 2019 *Nucleic Acids Res.* 47 D506–D515
[14] Mount D W 2008 *Cold Spring Harb. Protoc.* 2008 pdb–top39
[15] Eramian D, Shen M y, Devos D, Melo F, Sali A and Marti-Renom M A 2006 *Protein Sci.* 15 1653–1666
[16] Lai H T, Ha D M, Nguyen D M and Nguyen T T 2020 *Int. J. Mol. Phys.* 31 2050036
[17] Karplus M and Petsko G A 1990 *Nature* 347 631
[18] Hallermayer G and Neupert W 1974 *Hoppe-Seyler’s Zeitschrift für physiologische Chemie* 279–288
[19] Martin B and Lily S 1976 *Chem. Phys. Lipids* 17 416–422
[20] Berendsen H J, van der Spoel D and van Drunen R 1995 *Comput. phys. commun.* 91 43–56
[21] Ponder J W and Case D A 2003 Force fields for protein simulations *Advances in protein chemistry* vol 66 (Elsevier) pp 27–85
[22] Jaembeck J P and Lyubartsev A P 2012 *J. Chem. Theory Comput.* 8 2938–2948
[23] Price D J and Brooks III C L 2004 *J. Chem. Phys.* 121 10096–10103
[24] Essmann U, Perera L, Berkowitz M L, Darden T, Lee H and Pedersen L G 1995 *J. Chem. Phys.* 103 8577–8593
[25] Hünenberger P H 2005 Thermostat algorithms for molecular dynamics simulations *Adv. Theory Simul.* (Springer) pp 105–149
[26] Parrinello M and Rahman A 1981 *J. Appl. Phys.* 52 7182–7190
[27] Martyna G J, Tobias D J and Klein M L 1994 *J. Chem. Phys.* 101 4177–4189
[28] Zeng J, Guareschi R, Damre M, Cao R, Kless A, Neumaier B, Bauer A, Giorgetti A, Carloni P and Rossetti G 2018 *Int. J. Mol. Sci.* 19 2588
[29] Lai H T, Giorgetti A, Rossetti G, Nguyen T T, Carloni P and Kranjc A 2021 *Molecules* 26 1250
[30] Nguyen L H, Tran T T, Truong L T N, Mai H H and Nguyen T T 2020 *Biochemistry* 59 1378–1390
[31] Russ W P and Engelman D M 2000 *J. Mol. Biol.* 296 911–919
[32] Senes A, Gerstein M and Engelman D M 2000 *J. Mol. Biol.* 296 921–936
[33] Senes A, Ubarretxena-Belandia I and Engelman D M 2001 *Proc. Natl. Acad. Sci. U.S.A* 98 9056–9061
[34] Doura A K and Fleming K G 2004 *J. Mol. Biol.* 343 1487–1497
[35] Brosig B and Langosch D 1998 *Protein Sci.* 7 1052–1056
[36] Li F, Xia Y, Meiler J and Ferguson-Miller S 2013 *Biochemistry* 52 5884–5899
[37] Li F, Liu J, Liu N, Kuhn L A, Garavito R M and Ferguson-Miller S 2016 *Biochemistry* 55 2821–2831