Doctorate thesis

Thèse de doctorat

STUDY OF STRUCTURAL AND FUNCTIONAL DYNAMICS OF TRANSLOCATOR PROTEIN, A NEW TARGET FOR TREATMENT OF MALARIA

Etude de la dynamique structurale et fonctionnelle de la protéine translocatrice, une nouvelle cible pour traiter le paludisme.

Thesis submitted to École doctorale de Sciences, Technologie et Santé

Thèse soumise à l'École doctorale de Sciences, Technologie et Santé

Discipline: Biologie Informatique

Laboratoire: Dynamique des Structures et Interactions des Macromolécules Biologiques

Université de la Réunion

by Rajas M Rao

Directed by Pr. Catherine ETCHEBEST & Pr. Fréderic CADET

Presented on 30th april 2019

JURY:

Pr. Patrick FUCHS	**President/ Président**
Pr. Manuel DAUCHEZ	**Reviewer/Rapporteur**
Pr. Isabelle BERQUE BESTEL	**Reviewer/Rapportrice**
Pr. Catherine ETCHEBEST	**Supervisor/Directrice**
Pr. Fréderic CADET	**Supervisor/Directeur**
Pr. Thérèse MALLIAVIN	**Jury member/Examinatrice**
Pr. Jean-Jacques LACAPÈRE	**Jury member/Examinateur**

Dedicated to the giants,

on whose shoulders I stand today.

ACKNOWLEDGEMENTS

There are many people I would like to thank, without whose guidance and support, this thesis work would not have seen the light of its day. I would like to thank my supervisors, Pr. Catherine Etchebest and Pr. Fréderic Cadet, without whose guidance and support, I would never have come this far. I would also like to express my sincere gratitude to the thesis reviewers, Pr. Isabelle Berque Bestel and Pr. Manuel Dauchez, and the reporters Pr. Jean-Jacques Lacapère, Pr. Thérèse Malliavin and Pr. Patrick Fuchs for their constructive and helpful critique of this work.

I would also like to thank the members of my thesis committee, Pr. Alexandre de Brevern, Pr. R. Sowdhamini, Pr. Bernard Offmann, Pr. Thérèse Malliavin, Pr. Mariano Ostuni and Pr. Patrick Fuchs for their helpful guidance and counsel during different stages of my doctoral studies, which were critical in steering my doctoral studies. I would specially like to thank Dr. Julien Diharce, who was my 'third supervisor', provided guidance, conducted long critical discussions that helped me better interpret my results, and provided essential helpful inputs on writing this thesis. I also wish to express special gratitude to Pr. R. Sowdhamini, who not only nurtured my nascent skills in bioinformatics before the start of my doctoral studies, but also encouraged me to apply for the doctoral program.

I would also like to express my heartfelt gratitude to our collaborators on TSPO, Pr. Mariano Ostuni, Pr. Jean-Jacques Lacapère, Dr. Sophie Lefevre, Pr. Luminita Duma, Dr. Leeyah Issop, Pr. Stéphanie Finet and Pr. Laurent Gouya for fruitful discussions on TSPO. Importantly, I would also like to thank the agencies and institutions who funded my doctoral studies, la Région Réunion, Centre Informatique National de l'Enseignement Supérieur for provision of computation time in their supercomputer 'Occigen', and most importantly, Inserm.

I would also like to thank the members of Inserm UMR-1134, Université de la Réunion and INTS. I would like to thank the members of my laboratory, Alex, JC, Sofiane, Jirika, Akhila, Tatiana, Charlotte, Aline and all the interns, as well as former members of the lab, including Hubert, Aurore, Mathieu, Tarun, Sylvain, Sonia, Agata and NIcolas. I would also like to thank former interns who worked with me, Bérénice, Alper, Amiel and Alexandre. They have contributed significantly to this work, helped me immensely with my pedagogical skills and made the overall experience of working with this project a great joy. I would also like to thank Pr. Fernando Barroso da Silva, with whom I had various fruitful discussions on bioinformatics in general, during his visits to Paris, and Dr. Pierre Poulain and Dr. Samuel Murail, whose courses on programming were very helpful for me during the course of the doctoral studies.

Lastly but most importantly, I want to express my heartfelt gratitude to my family and my friends who have been my pillar of strength and provided support throughout the course of my doctoral studies.

TABLE OF CONTENTS

List of figures 6
List of tables 9
Abbreviations 10
Abstract 12
1. Chapter-1: Introduction 14
1.1. Biology of TSPO 18
1.1.1. Steroidogenesis and cholesterol transport 18
1.1.2. Tetrapyrrole metabolism, heme biosynthesis and erythropoieisis 21
1.1.3. Apoptosis 24
1.1.4. Regulation of photosynthesis in green-photosynthetic bacteria 26
1.1.5. Response to stress in plants 27
1.1.6. Other functions associated with TSPO 27
1.1.7. TSPO2-A mysterious paralog of TSPO 28
1.2. Involvement of TSPO in different pathologies 30
1.2.1. Malaria 30
1.2.2. Neurodegenerative diseases 30
1.2.3. Cancers 31
1.3. Structures and structure-dynamics relationships of TSPO 32
1.3.1. Early studies on TSPO structure 32
1.3.2. Studies on atomic structures of TSPO 33
1.4. Interacting partners of TSPO 41
1.4.1. Endogenous ligands of TSPO 41
1.4.2. Exogenous ligands of TSPO 43
1.4.3. Interacting protein partners of TSPO 45
2. Chapter-2: Materials and methods 53
2.1. A brief introduction to molecular dynamics simulations 54
2.1.1. Basic principles 55
2.1.2. Force fields 56
2.1.3. Methods in molecular dynamics 59
2.1.4. Coarse-grained force fields 62
2.2. Analysis methods of molecular dynamics simulations 64
2.2.1. Selection of residues for analysis 64
2.2.2. Principal component analysis 66
2.2.3. Protein structural networks 67
2.2.4. Protein-protein docking 68
2.3. Methods employed for comparison of TSPO paralogs 71
2.3.1. Multiple sequence alignment 71
2.3.2. Phylogenetic tree construction 72
2.3.3. Normal mode analysis 73
3. Chapter-3: Impact of PK11195 on TSPO dynamics 76
3.1. Introduction 78
3.2. Results 89
3.2.1. The structure of mouse TSPO in bilayer environment is stable, but slightly differs from the NMR structure with DPC micelles 89

3.2.2. PK11195 ligand impacts TSPO secondary structure 92
3.2.3. PK11195 impacts global dynamics of TSPO 94
3.2.4. Long range correlated motions are activated in TSPO bound to PK11195 99
3.2.5. The long-range motions correspond to a possible allosteric pathway from the conserved to variable regions 101
3.2.6. PK11195 adopts a wide range of conformations when it is bound to TSPO 103
3.2.7. PK11195-interacting residues around the binding cavity have strong evolutionary conservation 108
3.3. Discussion 109
3.3.1. Common binding sites of PK11195 and PP-IX ligands 109
3.3.2. Dynamics of PK11195 ligand 110
3.3.3. Helical rotations of TSPO 110
3.3.4. Effect of PK11195 ligand on TSPO dynamics 111
3.3.5. Functional implications of allosteric communication 116
3.4. Conclusions and Future directions 117

4. Chapter-4: Dynamics of Translocator protein oligomerization 119
4.1. Introduction 121
4.2. Results 129
4.2.1. TSPO monomers assemble to form dimers in most of the simulations 129
4.2.2. Dimerization can occur through different interfaces but some TM pairings are preferred 131
4.2.3. Interactions between aromatic residues were the most dominant at the interface 136
4.2.4. A putative cholesterol translocation mechanism was observed 139
4.2.5. Different long-range communication pathways were found 141
4.3. Discussion 144
4.3.1. Asymmetric interfaces and existence of higher order oligomers 144
4.3.2. Importance of aromatic residues in a tryptophan-rich protein 145
4.3.3. Putative translocation mechanism of cholesterol 146
4.3.4. Prevalence of long-range communications in varied dimer interfaces 147
4.4. Conclusions 148

5. Chapter-5: Comparative dynamics of mammalian and bacterial TSPO 149
5.1. Introduction 150
5.2. Results 155
5.2.1. Bacterial and eukaryotic TSPO explored unique sets of conformations 155
5.2.2. Similar dynamics were observed between certain regions of bacterial and mouse TSPO despite difference in conformational exploration 159
5.2.3. PP-IX binding residues are conserved across bacteria as well as eukaryotes 162
5.3. Discussion 165
5.4. Conclusions and future directions 167

6. Chapter-6: Structural and evolutionary comparison of TSPO paralogs 169
6.1. Introduction 170
6.2. Results 171
6.2.1. TSPO2 diverged from TSPO1 during evolution of early vertebrates 171
6.2.2. Many PP-IX binding and cholesterol interacting residues are not conserved in TSPO2 175
6.2.3. Large differences were observed between TSPO1 and TSPO2 at the level of structures 178
6.2.4. Both differences and similarities were observed in the TSPO1 vs. TSPO2 dynamics 181
6.3. Discussion 184
6.4. Conclusions and future directions 189

7. Concluding remarks 190
8. References 197
9. Annexures 234

LIST OF FIGURES

1. Chapter-1

1.1. Life cycle of malarial parasite *Plasmodium falciparum(Klein 2013)*from(Klein 2013). The stages merozoite, trophozoite and schizont are labelled (A), (B) and (C) respectively.

1.2. Effect of treatment of TSPO-binding drugs on the growth of *P. falciparum* in *P. falciparum* infected erythrocytes (from Bouyer et. al, 2011).

1.3. Protein sequence of mouse TSPO (Uniprot ID:P50637), indicating the boundaries of TM regions and the cholesterol-binding motif.

1.4. Consequence of CRAC residue mutagenesis in cholesterol uptake in TSPO-transfected *E. coli* cells (Li and Papadopoulos 1998).

1.5. Change in absorbance peak reflecting catalytic degradation of PP-IX in *Bacillus cereus* TSPO (Guo et al. 2015).

1.6. Effect of TSPO ligand PK11195 treatment versus ethanol treatment (control) on hemoglobin content in zebrafish (Rampon et al. 2009).

1.7. Schema showing arrangement of TSPO, VDAC and their interacting partners in mitochondrial membranes. B: Activated TSPO-VDAC complex, indicating cascade of events leading to apoptosis.

1.8. Growth (A) and cell death (B) of *Physcomitrella patens* wild-type (labelled WT) and TSPO knockout lines.

1.9. Immunoblot showing expression of TSPO1, TSPO2 and VDAC in healthy and *P. falciparum*-infected red blood cell membranes.

1.10. Traverse (A) and cross-sections (B) of brain of Parkinson's disease patient and Traverse (C) and cross-sections (D) of brain of healthy control patient showing radiolabelled PK11195 binding.

1.11. Interpretation of TSPO oligomeric interfaces from the cryo-EM studies.

1.12. NMR structure of mouse TSPO, showing 20 lowest energy conformations.

1.13. Experimental structures of *R. sphaeroides* TSPO.

1.14. Experimental structures of TSPO.

1.15. Structure-based sequence alignment of TSPO sequences, based on structural superposition experimental structures.

1.16. Exogenous and endogenous ligands of TSPO.

1.17. Interacting partners of TSPO, with respect to their subcellular localizations and involved physiological processes.

2. Chapter-2

2.1. Illustration of different interactions calculated by a force field.

2.2. Illustration comparing all-atom structures of biomolecules with respective coarse-grained models.

3. Chapter-3

3.1. Orientation of TSPO in membrane.

3.2. TSPO in lipid bilayer undergoes helical rotations.

3.3. TSPO conformations have high RMSD with respect to experimental structures.

3.4. PK11195 binding stabilized TSPO secondary structure.

3.5. PK11195 influences TSPO dynamics.

3.6. TM regions of TSPO have different dynamics in holo- and apo- forms.
3.7. Long range correlated motions between conserved residues are activated in TSPO in complex with PK11195.
3.8. TSPO residues are part of a protein structural network in presence of PK11195 ligand, and an allosteric communication pathway was identified involving TSPO residues.
3.9. Plot showing comparison of betweenness centralities of protein structural network and protein energy network of TSPO trajectory ensemble in presence of PK11195.
3.10. PK11195 is dynamic in complex with TSPO.
3.11. TSPO-PK11195 interactions are hydrophobic in nature.
3.12. PK11195 transitions between E- and Z- rotamer forms.
3.13. Plot showing rotational energies of PK11195.
3.14. Certain residues have dominant interactions with PK11195.
3.15. PK11195 and PP-IX share a common binding cavity.
3.16. Plot showing RMSF of TSPO simulations in presence of PK11195.

4. Chapter-4

4.1. Plot showing minimum distance between two TSPO chains.
4.2. Properties of simulated dimer as a function of time.
4.3. A: Number of residues within a distance < 5.6 Å of residues of the other chain; B: Number of residues within a distance < 6 Å of residues of the other chain.
4.4. Total contact numbers for each residue calculated for all the frames saved from the whole set of trajectories.
4.5. Plot showing TSPO residues of each chain within 10Å distance of each other, for the replicates 1 (A) and 2 (B) of simulations with 90° initial orientation.
4.6. A: Chart showing dimer interface areas of representative TSPO dimer structures; B: Table showing representative TSPO dimer conformations and their interfaces obtained from each trajectory
4.7. A: Starting structure of TSPO dimer model, with interface similar as described by Jaipuria et. al.; B: A representative conformation of TSPO dimer from the simulation with starting conformation involving TM3 region, based on the interface described by Jaipuria et. al.
4.8. A: Plot showing number of cholesterol molecules interacting with CRAC motif in the simulation of TSPO model with interface involving TM3 region. B: Plot at (A) zoomed to show the decrease of cholesterol binding between time points 650ns-2,000ns.
4.9. A: Snapshot of TSPO and interacting cholesterol molecule at time point 650ns; B: Snapshot of TSPO and interacting cholesterol molecule at time point 712ns.
4.10. TSPO dimer structures indicating allosteric communication pathways in dimer orientations from different simulations, compared with an ensemble of monomer conformations.
4.11. Representation of long-range communication pathway between the dimer interface and residues of TM5 region, occuring between the time points 650ns-2000ns of TSPO dimer model-1, where a putative translocation of cholesterol was also observed.

5. Chapter-5

5.1. Consensus logo of sequences comprising TM domain, aligned with the TM boundaries of mouse TSPO sequence.
5.2. Structural superposition of experimental TSPO structures showing RMSD of TM regions.

5.3. A: RMSD of coarse-grained mouse TSPO (NMR structure as starting structure) trajectory replicate-1; B: Heatmap showing correlations between the top principal components representing TSPO coarse-grained dynamics and equivalent all-atom dynamics.

5.4. Plot showing projections of mammalian and bacterial TSPO conformations onto the conformational space of PC1 and PC2.

5.5. Heatmap showing correlations between PC 1-5 of mouse TSPO versus PC 1-5 of bacterial TSPO.

5.6. Plot showing RMSF from the trajectory representing dynamics of PC1 from the mouse ensemble and PC2 from bacterial TSPO ensemble.

5.7. Porcupine plot showing the dynamics of first principal components from *R. sphaeroides* (32% covariance) (A) and mouse (52% covariance) (B) TSPO simulations.

5.8. A: X-ray crystal structure of R. sphaeroides TSPO (PDB ID: 4UC1, Li et al. 2015) in complex with PP-IX ligand showing protein ligand contacts and corresponding residue conservation. B: Illustration of PP-IX interacting residues of X-ray crystal structure of *R. sphaeroides* TSPO (PDB ID: 4UC1), showing different types of interactions as interpreted by interatomic distances.

6. Chapter-6

6.1. A: Phylogenetic tree showing evolution of animal TSPO sequences; B: The same phylogenetic tree zoomed into the vertebrate clade showing divergence of TSPO1 and TSPO2 sequences.

6.2. Pairwise sequence alignment between the TSPO sequences of Chinese alligator and Coelacanth with that of mouse TSPO sequence.

6.3. Multiple sequence alignment of TSPO1 and 2 sequences with the most probable ancestral sequence from which divergence of TSPO1 and 2 occurred

6.4. Superposition of mouse TSPO1 and TSPO2 structures. A: Horizontal view B: Vertical view.

6.5. Alignment of TSPO1 sequences of mouse, *R. sphaeroides* and *B. cereus*, whose structures were used as template to construct homology model of TSPO2, along with the sequence of mouse TSPO2.

6.6. Heatmap showing correlation between the dynamics of non-trivial lowest-frequency normal modes of TSPO1 and TSPO2.

6.7. Porcupine plot of mouse TSPO1 and mouse TSPO2 coarse-grained models showing dynamics of normal modes having high collectivity (>0.5).

LIST OF TABLES

1. **Chapter-1:**
   - 1.1. List of commonly used drugs for treatment of malaria (From Vaitinadapoulé 2015).
   - 1.2. Summary of experimental TSPO structures.

2. **Chapter-2:**
   - 2.1. Table describing boundaries of TM regions, which were employed for all the analysis of TSPO simulations.
   - 2.2. Table representing boundaries of each layers, in terms of z-coordinates. The above represented layers are used to calculate frequencies of amino acid residues occurring at respective z-coordinates, which in turn are used to calculate membrane transfer energies for a given docked complex in Memdock web server (Hurwitz et al., 2016).

3. **Chapter-3:**
   - 3.1. Summary of TSPO simulations performed in complex with PK11195 ligand.
   - 3.2. Summary of TSPO simulations performed in absence of PK11195 ligand, with NMR structure(PDB: 2MGY) as the starting structure
   - 3.3. Summary of TSPO simulations performed in absence of PK11195 ligand, with snapshots from coarse-grained simulation as starting points.
   - 3.4. Definitions of TM regions used for PCA and protein structural network.
   - 3.5. Table showing residues having contact (<5.0Å distance of PK11195) with PK11195 ligand for >90% of total simulation time with PK11195 (1.23μs), and corresponding conservation scores.

4. **Chapter-4:**
   - 4.1. Summary of TSPO dimerization and TSPO dimer model simulations
   - 4.2. Heatmap representing the frequency of interactions between different TM regions of TSPO.
   - 4.3. Best-scoring TSPO homo-dimer complexes obtained from blind-docking of NMR structure of mouse TSPO.
   - 4.4. Table showing residue pairs from each TSPO monomer with highest interaction frequencies, as described by proportion of trajectory where given residue pairs are in <5.6Å distance of each other.

5. **Chapter-5:**
   - 5.1. Table showing summary of simulations performed along with preparatory steps.
   - 5.2. Table describing the boundaries of TM regions, which were employed for all the analyses of simulations of bacterial and mouse TSPO.
   - 5.3. Table showing residues of *R. sphaeroides* TSPO having strongest contacts (in terms of VdW radii overlap) with PP-IX ligand in the X-ray structure of A139T mutant (PDB: 4UC1, Li et al., 2015) and their respective residue conservation scores.

6. **Chapter-6:**
   - 6.1. Table showing RMSD between TM regions of mouse TSPO1 NMR structure and mouse TSPO2 homology model.
   - 6.2. Table describing degree of collectivity of residue dynamics of TSPO1 and TSPO2 normal modes.

ABBREVIATIONS

TSPO……Translocator protein

PBR……..Peripheral Benzodiazepine Receptor

CRAC……..Cholesterol Recognition And Consensus motif

PP-IX……..Protoporphyrin-IX

TM……..……Trans-Membrane

MPTP……..……Mitochondrial Permeability Transition Pore

VDAC……..……Voltage Dependent Anion Channel

ANT……..……Adenine Nucleotide Transporter

ErPC……..……Erucyl Phospho Choline

OMM……..……..……Outer Mitochondrial Membrane

IMM……..……..……Inner Mitochondrial Membrane

NPP……..……..……..……New Permeability Pathway

NMR……..……..……Nuclear Magnetic Resonance

SDS……..……..……Sodium Dodecyl Sulphate

Cryo-EM……..……..……Cryo-Electro Magnetic micrograph

DPC……..……..……..……..……Dodecyl Phospho Choline

RMSD……..……..……..……..……Root Mean Square Deviation

StAR……..……..……..……Steroidogenic Acute Regulatory protein

PRAX-I……..……..……..……..……PBR Associated Protein-I

PAP-7……..……..……..……..……..……PBR-Associated Protein-7

ACBD3……..……..……..……..……..……Acyl-CoA Binding Domain-3

LJ……..……..……..……..……..…Lennar-Jones potential

VdW……..……..……..……Van der Waals potential

MD……..……..……..……Molecular Dynamics simulations

QM……..……..……..……Quantum Mechanics simulation

CHARMM……..……..……Chemistry at Harvard using Molecular Mechanics

OPLS……..……..……Optimized Potentials for Liquid Simulations

AMBER……..……..……Assisted Model Building with Energy Refinement

DOPC……..……..……..……1,2-dioleoyl-sn-glycero-3-phosphocholine

GROMACS……..……..……..……GROningen MAchine for Chemical Simulations

NVT……..……..……..……..…Constant Number of particles, Volume and Temperature

NPT……..……..……..……..…..Constant Number of particles, Pressure and Temperature

PBC……..……..……..……..……..…Periodic Boundary Conditions

AA……..……....…..…..…All-Atomic molecular dynamics simulations

CG……..……....…..…Coarse-Grained molecular dynamics simulations

DSSP……..…….…….....……..……...……Define Secondary Structure of Proteins

PDB……...……....……...……...……...……...…Protein Data Bank

OPM……..……....……...…..…Organisation of Proteins in Membranes

PCA……..……....……..……...……...……Principal Component Analysis

PSN……..……....……..……...……...…….Protein Structural Networks

DCCM……..……....……...…..……Direct Cross Correlation Matrix

MSA……..……....…...……....Multiple Sequence Alignment

NMA……..……....…...……..…..…….Normal mode analysis

ANM……..……....…...……..…...…Anisotropic Network Model

RMSD……..……....…...……...…...…Root Mean Square Deviation

RMSF……..……..……....…...……...…..…Root Mean Square Fluctuation

DMPC……..……....…1,2-dimyristoyl-sn-glycero-3-phosphocholine

Abstract

Despite availability of different drugs, malaria remains a serious health challenge in tropical countries, and is rapidly making inroads in the countries with temperate climate. To compound the problem, resistance to the antimalarial drugs by the Plasmodial parasite has been documented in many instances, underscoring the need for novel therapeutic approaches for the disease. However, discovery of involvement of a transmembrane protein Translocator protein (TSPO) in the Plasmodium-infected erythrocytes showed the potential of this protein as a target for novel antimalarial drugs. TSPO is an 18 kDa transmembrane protein found in various organelles of cell and involved in variety of processes, including steroidogenesis and cholesterol transport, tetrapyrrole transport/catalysis, erythropoiesis. apoptosis, response to stress to name a few. As a result, it is also involved in various diseases apart from malaria, and is considered a potential drug target. Though recent studies on its atomic structure provided some interesting insights into the structure, several questions on the structure-dynamics-function relationship of TSPO remained unanswered.

In this thesis, we attempted to answer some of these questions using various bioinformatics and molecular modelling approaches. We describe how does the ligand PK11195 influence the dynamics of TSPO using all-atom molecular dynamics simulations, and how does PK11195 itself behave in complex with TSPO. We also studied different aspects of TSPO oligomerization using coarse-grained molecular dynamics simulations, and explored how do mammalian TSPO dynamics compare with those of bacterial homologs. Ultimately, we attempted to make a comprehensive comparison between the two paralogs of TSPO in mammals, TSPO1 and TSPO2, from sequence, structural as well as phylogenetic perspectives. Our results give interesting insights into the structure-dynamics-function relationships of TSPO, and pave way to some new hypotheses on the same, which could be exploited to develop new generation of antimalarial drugs that target TSPO.

Keywords:
Translocator protein, molecular dynamics, PK11195, oligomerization.

Résumé

Le paludisme reste un grave problème de santé, bien que différents médicaments sont disponibles pour le traitement. En outre, il se propage à l'une vitesse alarmante dans les pays avec climat tempéré. Pour aggraver le problème, la résistance aux antipaludiques du parasite Plasmodium a été documentée dans de nombreux cas, soulignant la nécessité de nouvelles approches thérapeutiques pour la maladie. Cependant, la découverte de l'implication d'une protéine transmembranaire, la protéine translocateur (TSPO), dans les érythrocytes infectés par Plasmodium a montré le potentiel de cette protéine en tant que cible pour de nouveaux médicaments antipaludiques. La TSPO est une protéine transmembranaire de 18 kDa présente dans divers organites de cellules et impliquée dans divers processus, notamment la stéroïdogenèse et le transport du cholestérol, le transport / catalysme du tétrapyrrole, l'érythropoïèse. apoptose, réponse au stress, et cetera. En conséquence, il est également impliqué dans diverses maladies en dehors du paludisme et est considéré comme une cible potentielle pour un médicament. Bien que des études récentes sur sa structure atomique aient fourni des informations intéressantes sur la structure, plusieurs questions sur la relation structure-dynamique-fonction de TSPO sont restées.

Dans cette thèse, nous avons tenté de répondre à certaines de ces questions en utilisant diverses approches de bioinformatique et de modélisation moléculaire. Nous décrivons comment le ligand PK11195 influence la dynamique de TSPO en utilisant des simulations de dynamique moléculaire tout atome et comment PK11195 se comporte-t-il en complexe avec TSPO. Nous avons également étudié les aspects différents de l'oligomérisation de TSPO à l'aide de simulations de dynamique moléculaire

à grains grossiers et exploré comment la dynamique de TSPO de mammifère se compare à celle d'homologues bactériens. Finalement, nous avons tenté de faire une comparaison complète entre les deux paralogues de TSPO chez les mammifères, TSPO1 et TSPO2, du point de vue de la séquence, tant structurel que phylogénétique. Nos résultats fournissent des informations intéressantes sur les relations structure-dynamique-fonction de TSPO et ouvrent la voie à de nouvelles hypothèses similaires, qui pourraient être exploitées pour développer une nouvelle génération d'antipaludiques ciblant TSPO.

Mots clés:
Protéine translocateur, dynamique moléculaire, PK11195, oligomérisation.

Chapter 1: Introduction

Table of contents

1.1. Biology of TSPO 18
1.1.1. Steroidogenesis and cholesterol transport 18
1.1.2. Tetrapyrrole metabolism, heme biosynthesis and erythropoieisis 21
1.1.3. Apoptosis 24
1.1.4. Regulation of photosynthesis in green-photosynthetic bacteria 26
1.1.5. Response to stress in plants 27
1.1.6. Other functions associated with TSPO 27
1.1.7. TSPO2-A mysterious paralog of TSPO 28
1.2. Involvement of TSPO in different pathologies 30
1.2.1. Malaria 30
1.2.2. Neurodegenerative diseases 30
1.2.3. Cancers 31
1.3. Structures and structure-dynamics relationships of TSPO 32
1.3.1. Early studies on TSPO structure 32
1.3.2. Studies on atomic structures of TSPO 33
1.4. Interacting partners of TSPO 41
1.4.1. Endogenous ligands of TSPO 41
1.4.2. Exogenous ligands of TSPO 43
1.4.3. Interacting protein partners of TSPO 45

Malaria is among the most persistent health issues of the world, causing an estimated 445,000 deaths globally in 2016(Health 2017). It is particularly rampant in tropical countries, though new cases are also being reported in countries with temperate climate, at an increasing frequencies. The disease is caused by a protozoan parasite *Plasmodium falciparum*. *P. falciparum* has a complex life cycle in two hosts: humans and female anopheles mosquitoes, which act as vectors for the parasite. In humans, *P. falciparum* proliferates by means of asexual reproduction, in liver and blood tissues. The merozoite, trophozoite and schizont stages of *P. falciparum* life cycle involves invasion of erythrocytes, multiplication inside the hosts and breaking-free from erythrocytes by killing them (Fig. 1.1). At these stages, the patient experiences symptoms such as fever, shivering, anemia, convulsions etc. If untreated, the infection ultimately results in coma, or death.

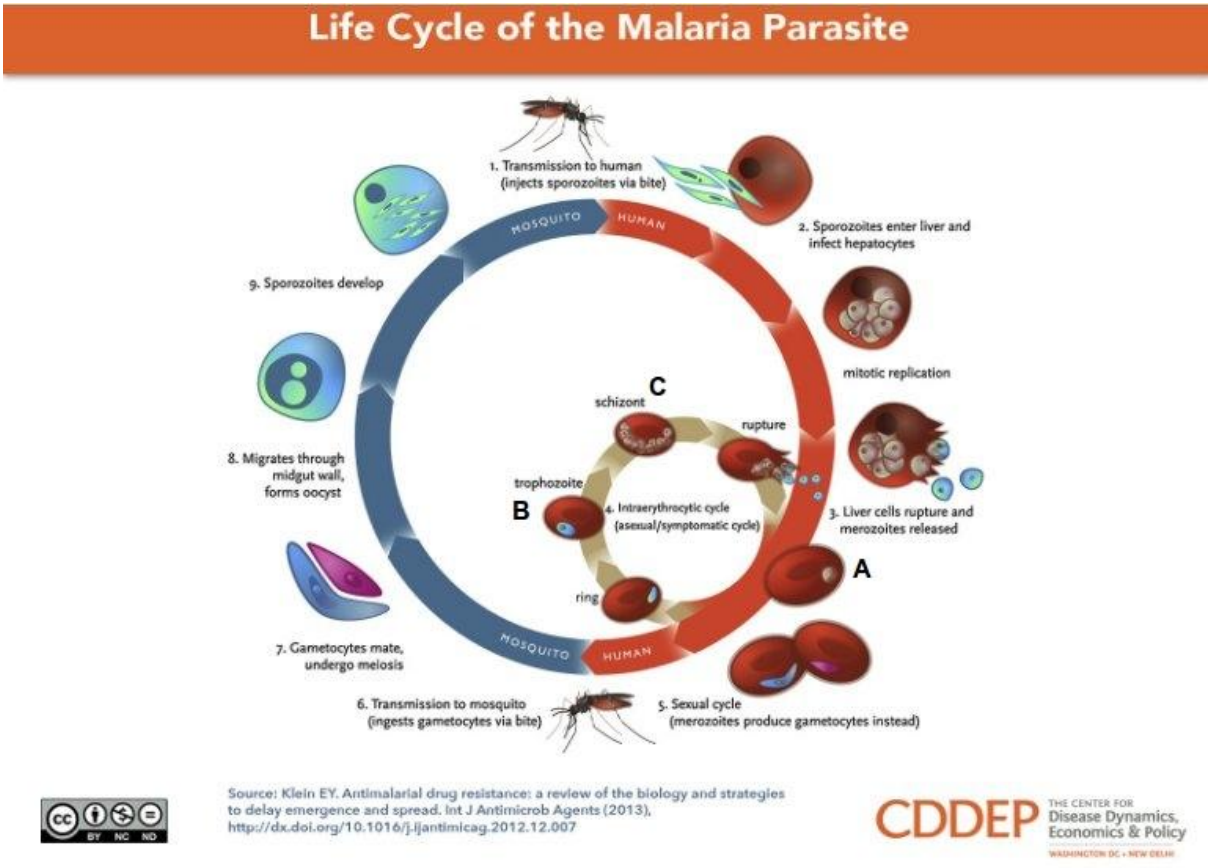


Figure 1.1. Life cycle of malarial parasite *Plasmodium falciparum(Klein 2013)*from(Klein 2013). The stages merozoite, trophozoite and schizont are labelled (A), (B) and (C) respectively.

Though drugs are available for the treatment of the disease (Table 1.1), there are problems associated with them, such as large number of side effects and resistance by *P. falciparum* to these drugs(Health 2017). A novel anti-malarial drug Artemisinin was developed, which has been found to be very effective in treatment of malarial infections(van Agtmael et al. 1999). However, Artemisinin resistant *P. falciparum* strains have been found in various parts of Asia, including India, Thailand, Cambodia, underscoring the need for alternative and more effective malarial treatment regimens (Table

1.1)(Amaratunga et al. 2012; Das et al. 2018). Indeed, the parasite is constantly mutating and gaining resistance against drugs. As a result, the search for new drugs and new possible targets still remains a real challenge to treat malaria infected patients**.**

Molecules	***Mechanism of action***	***Presence of resistance by P. falciparum***	***Side Effects***
Quinine/Quinidine	Blood schizonticide and breakdown of ability of the parasite to digest hemoglobin.	Yes	Haemolytic anemia, α-adrenergic blocking effect, headache, nausea, cinchonism, hypoglycaemia, hypotension, seizures, cardiotoxic effects
Chloroquine	Blood schizonticide; targets heme polymerization.	Yes (in addition to resistant strains of *P. vivax, P. ovale, P. malariae*)	Nausea, headache, generalized pruritus, retinal toxic effect, worsening of psoriasis and rare cases of seizures and psychosis, central nervous system dysfunction.
Atovaquone/Proguanil	Folate metabolism : blocks parasite proliferation; targets electron transport.	Yes	Rashes, fever, vomiting, diarrhea, abdominal pain and headache. Forbidden on infants, pregnant and breastfeeding women and patients with renal complications
Artemisinins and its derivatives	Inhibition of heme polymerization, targetting electron transport chain	Yes	Abdominal pain, anorexia, nausea, diarrhea, vomiting, rare cases of rash and pruritus
Mefloquine	Produce swelling of the *P. falciparum* food vacuoles; targets heme polymerization.	Yes	Cardiac depressant effect, anti fibrillatory activity, Nausea, seizures (rare) and psychosis.
Doxycycline	Slow acting blood schizontocidal agent.	Yes	Photosensitivity, gastrointestinal upset and oesophageal ulcerations (rare), black coloration of tongue, fongic infection of anus and vagina, Oral contraceptive failure (lacking evidence)

Table 1.1. List of commonly used drugs for treatment of malaria (From Vaitinadapoulé 2015).

However, a recent finding gave some hope for finding more effective approaches to malarial treatment. During *P. falciparum* infection of erythrocytes, a transmembrane protein called TSPO was found to be active, along with its other interacting protein partners(Bouyer et al. 2011). Interestingly, drugs that bind to TSPO in high affinity and specificity were found to be very effective in inhibition of both intraerythrocytic as well as intererythrocytic parasite growth in trophozoite and schizont stages (Fig. 1.1, B & C, Fig. 1.2). Understanding the structure, dynamics and interactions of this protein

would be a very valuable tool for the discovery of new generation of antimalarial drugs or therapeutic strategies, and will be the prime focus of this thesis.

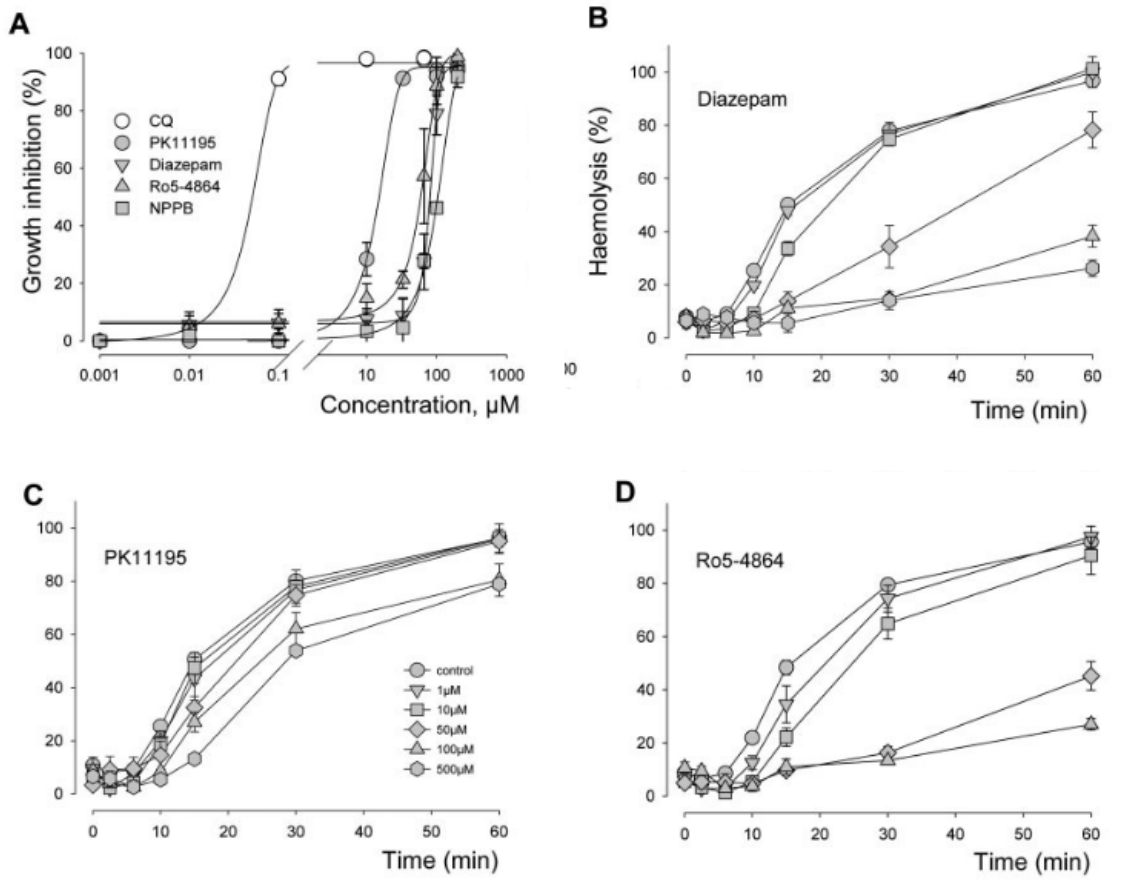


Figure 1.2. Effect of treatment of TSPO-binding drugs on the growth of *P. falciparum* in *P. falciparum* infected erythrocytes (from Bouyer et. al, 2011).
A: Growth inhibition of cultures of 3D7 *P. falciparum* strain in response to different TSPO-binding drugs.
B: Haemolysis of *P. falciparum* infectred erythrocytes in response to different concentrations of Diazepam.
C: Above in response to PK11195
D: Above in response to Ro5-4864.

Translocator Protein (TSPO, also referred to as Peripheral Benzodiazepine Receptor/PBR and TSPO1) is an 18kDa transmembrane protein, consisting of five membrane-spanning TransMembrane (TM) regions (Fig. 1.3). It is found in various tissues in humans, as well as in organisms from all the three domains of life. In certain cases, it is also referred to as tryptophan-rich sensory protein, due to the fact that this protein is abundant in tryptophan residues (Fig. 1.3). Among the earliest studies on TSPO described it as a receptor that binds to the drug benzodiazepine with high affinities(Braestrup and Squires 1977). Since the high-affinity diazepam binding was observed particularly in the peripheral nervous system, it was first named as peripheral benzodiazepine receptor (PBR). PBR was later found

to bind to numerous other drug molecules, such as Ro5-4864(Marangos et al. 1982), PK11195(Bénavidès et al. 1984; Inoue et al. 1985), Flunitrazepam(Chang and Snyder 1978) to name a few. Similarly, it was later found that porphyrins also bind to PBR with high affinities as an endogenous ligand(Verma et al. 1987; Snyder et al. 1987).these The drug-molecule binding was characterized to such an extent that the radio-labelled variants of drugs are used for diagnosis of various diseases, and PBR was proposed as a biomarker for diagnosis of these diseases(Olson et al. 1988). Though numerous drugs and an endogenous ligand interaction with PBR were described, its physiological function remained mysterious. One of the earliest studies into the PBR function proposed that it was involved in tracheal smooth muscle contraction, in association with calcium ion channels(Raeburn et al. 1988). However, another set of studies found it to be primarily localized in the outer mitochondrial membrane (Hirsch et al. 1989; Olson et al. 1992). TSPO was proposed to play an important role in regulation of mitochondrial function, that include modulation of respiration, mitochondrial potential etc.(Verma and Snyder 1989). These processes were proposed to be modulated by TSPO ligands such as PK11195, Ro5-4864, as well as certain porphyrins. In another study, TSPO was found to be localized in leydig cells, and was observed to have a role in steroidogenesis(Papadopoulos et al. 1990). Subsequently, a series of studies, (discussed in following pages) over a period of next two decades firmly established the role of TSPO in steroidogenesis. In light of these studies, PBR was renamed as translocator protein (TSPO), and this has been in use till date(Papadopoulos et al. 2006).

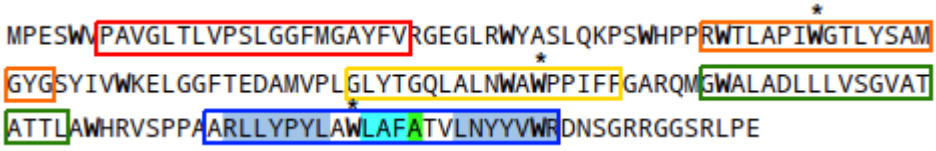


Figure 1.3. Protein sequence of mouse TSPO (Uniprot ID:P50637), indicating the boundaries of TM regions and the cholesterol-binding motif. The TM regions 1 to 5 are marked in red, orange, yellow, green and dark blue boxes respectively; The cholesterol binding motifs are highlighted in sky blue, cholesterol-binding enhancement motif is coloured cyan, and the residue Ala147 is highlighted light green. Tryptophan residues in the sequence are shown as bold, and tryptophan residues conserved between mouse and bacterial TSPO sequence (*Rhodobacter sphaeroides*) are marked '*' above the residue. The sequence was retrieved from Uniprot database.

1.1 Biology of TSPO

TSPO has been associated with various physiological processes, in various tissues and organisms. Below is an attempt to discuss all the physiological roles TSPO has been found to be associated with:

1.1.1. ***Steroidogenesis and Cholesterol transport***: Role of TSPO in steroidogenesis is one of the most extensively studied, and perhaps the most debated function. Earliest series of studies suggested that TSPO played a regulatory role in steroidogenesis, which was inferred from the observation that TSPO ligands significantly modulated steroid biosynthesis(F. Li et al. 2016; Papadopoulos et al. 1990). This

was also supported by subsequent studies, where PK11195 significantly accelerated the biosynthesis of steroid hormone pregnenolone(Krueger and Papadopoulos 1990). On the other hand, flunitrazepam, a diazepam derivative which binds to TSPO in nanomolar affinity, inhibited peptide-hormone regulated steroidogenesis(Papadopoulos, Amri, Boujrad, et al. 1997; Papadopoulos et al. 1991). Gene deletion/disruption studies of TSPO in different models supported this hypothesis. In Leydig cells, targeted disruption of TSPO gene resulted in significant inhibition of steroidogenesis(Papadopoulos, Amri, Li, et al. 1997). Similarly, mouse TSPO gene expressed recombinantly in *Escherichia coli* DE3 cells resulted in enhanced uptake of cholesterol, despite the fact that bacterial cells are unable to uptake cholesterol. In the same study, it was also observed that mutant mouse TSPO expressed recombinantly in *Escherichia coli* DE3 cells were unable to interact with the cholesterol (Lacapère et al. 2001). These observations reinforced the hypothesis that TSPO is involved in steroidogenesis, and it can indeed interact with cholesterol. The residues interacting with cholesterol were subsequently identified as a 'Cholesterol Recognition and Consensus motif' (CRAC motif). It consists of distinct signature composed of hydrophobic residues and positively charged residues (motif: –L/V-(X)1–5-Y-(X)1–5-R/K-; Fig. 1.3) that interacts with cholesterol (Li and Papadopoulos 1998; Jamin et al. 2005). The molecule binds with nanomolar affinities (Li, Yao, et al. 2001). Of these residues, the residues tyrosine (Tyr152) and arginine (Arg156) in mouse TSPO were found to be particularly important, since mutagenesis of these residues resulted in loss of cholesterol translocation (Fig. 1.4, Box 1.2; Li and Papadopoulos 1998). Furthermore, recent computational studies on TSPO indicated that there exists another cholesterol-binding motif adjacent to the CRAC motif, with a similar sequence but arranged in a reverse order(Fantini et al. 2016). Moreover, this motif in the mouse TSPO was found to have lower binding energy with the cholesterol, compared with the already identified CARC motif. It was named 'CARC' motif, due to the fact that the residue composition is palindromic with respect to CRAC motif (Fig. 1.3).

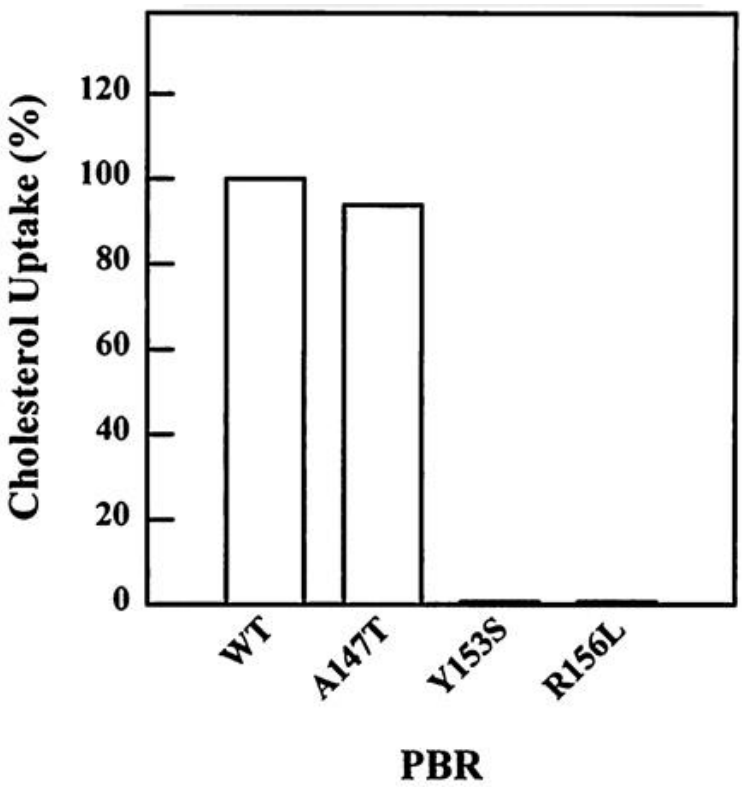


Figure 1.4. Consequence of CRAC residue mutagenesis in cholesterol uptake in TSPO-transfected *E. coli* cells (Li and Papadopoulos 1998).

In addition, another motif in between both the cholesterol-binding motifs was characterized recently. In bacteria, though CRAC motif is conserved in TSPO, cholesterol uptake does not take place. Though this is due to the fact that bacterial membranes do not have cholesterol, it was found that bacterial TSPO lacks an 'enhancement motif' that enables uptake of cholesterol. However, when the enhancement motif-containing TSPO gene was transfected into the bacteria, cholesterol uptake was found to increase significantly(Li, Liu, Valls, et al. 2015). These observations led to suggest that this motif also play a crucial role in enabling translocation of cholesterol by TSPO. The motif consists of three residues between both the cholesterol binding (CRAC & CARC) motifs (Motif: 144-LAF-146 in mouse TSPO, Fig. 1.3)(Li, Liu, Valls, et al. 2015). However, the structural mechanism by which this motif enables translocation remains unclear.

Despite growing body of evidence from large number of experiments that point towards the notion that TSPO is indispensable for steroidogenesis, the exact role of TSPO in this process is still enigmatic and poorly understood. There are three dominant hypotheses on the role of TSPO in steroidogenesis(F. Li et al. 2016; Selvaraj and Stocco 2015):

***i. TSPO is the cholesterol transporter*:** Since TSPO consists of a motif that binds cholesterol in nanomolar affinities, and binding of ligands such as PK11195, that binding with high specificity to TSPO augment the cholesterol translocation into mitochondria, it was initially suggested that TSPO

itself is the cholesterol transport apparatus(Krueger and Papadopoulos 1990). Computational modelling and simulation studies on the mouse and human TSPO revealed that TSPO can indeed accommodate cholesterol molecule inside the transmembrane pore and function as a channel(Papadopoulos et al. 1994), In addition, cholesterol binding motif is also well characterized, where mutagenesis of residues of the motif lead to an adverse impact on steroidogenesis (Fig. 1.4, Box-2,(Li and Papadopoulos 1998; Li, Yao, et al. 2001)). However, there exists no direct experimental evidence for cholesterol-transport activity of TSPO.

***ii. TSPO is a cholesterol transport regulator as a part of cholesterol import complex*:** Due to the absence of experimental evidence for the cholesterol transport activity of TSPO, it was suggested that TSPO by itself may not transport cholesterol into mitochondria(F. Li et al. 2016). An alternative role of TSPO in steroidogenesis was proposed: it plays a regulatory role in the import of cholesterol into mitochondria. TSPO was suggested to form a part of a larger complex that imports cholesterol into the intermembrane space(Liu et al. 2006). In support of this hypothesis, an 800 kDa complex on the junction of outer and inner membranes of mitochondria was found to bind to cholesterol (Rone et al. 2012). This complex consisted of TSPO, along with other proteins such as Steroidogenic Acute Regulatory Protein (STAR), which has been experimentally demonstrated to transport cholesterol, with VDAC, Cyclophilin D, etc.(Rone et al. 2012; Bose et al. 2008).

***iii. TSPO has no role in steroidogenesis & cholesterol transport*:** While above studies have supported the hypothesis that the TSPO plays a critical role in steroidogenesis, a series of recent studies refute this hypothesis. *In vivo* experiments performed on TSPO-conditional knockout mice revealed that deletion of TSPO had no impact on steroidogenesis, and as a result, physiological indicators such as testosterone production, gametogenesis were normal(Morohaku et al. 2014; Tu et al. 2014). Another study on a global TSPO knockout mice revealed that there was no effect of TSPO deletion on steroidogenesis as well as on fertility(Tu et al. 2014; Banati et al. 2014). The role of the ligand PK11195 on TSPO-mediated regulation of steroidogenesis was also questioned, as progesterone synthesis increased upon treatment with PK11195, both in wild-type as well as TSPO-knockout variants of mice(Tu et al. 2015). Though there have been a large number of experimental studies that supports as well as contradicts the notion that TSPO plays an important role in steroidogenesis and cholesterol transport, the role of TSPO in steroidogenesis still remains unclear and continues to be debated.

1.1.2. Tetrapyrrole metabolism, Heme biosynthesis and erythropoiesis: Several studies suggest an important role of TSPO in tetrapyrrole metabolism. Protoporphyrin-IX (PP-IX), a tetrapyrrole derivative, is one of the endogenous TSPO ligands, whose binding has also been characterized experimentally as well as structurally (Li, Liu, Zheng, et al. 2015a; Verma et al. 1987).

TSPO-tetrapyrrole interactions have been observed in context of different physiological processes in different organisms.

In mammalian cells, occurrence of disease porphyria has been associated with decreased PK11195 binding (Cantoni et al. 1992). Porphyria is a condition where porphyrin derivative get accumulated in the cells, arising from the failure to transport porphyrins across organelles. This observation suggested a possible role of TSPO in PP-IX transport (Cantoni et al. 1992). Later studies revealed a more complex role of TSPO-PP-IX interactions in regulation of apoptosis. Silencing of TSPO mRNA in glioblastoma cell lines revealed accumulation of PP-IX in cells. However, this accumulation also led to cell death (Zeno et al 2012). Indeed, free PP-IX reacts with oxygen to form Reactive Oxygen Species (ROS) in presence of light, and exposure of TSPO-knockdown glioblastoma cells to light led to cell death (Zeno et a, 2012, Batoko et al, 2015). Similar results have also been observed in *Arabidopsis thaliana* plant (Vanhee et al, 2011). In the bacteria *Bacillus cereus*, a TSPO homologue was found to catalyse degradation of PP-IX into an unidentified heme derivative (Fig. 1.5, Guo et al, 2015), suggesting that TSPO is involved in the catalytic degradation of PP-IX. In addition, tryptophan residues in TM2 and TM5 regions (residues Trp51 and Trp138 in *B. cereus*; equivalent to residues Trp53 and Trp143 in mouse TSPO) was found to be essential for this catalytic degradation (Ginter et al. 2013; Guo et al. 2015).

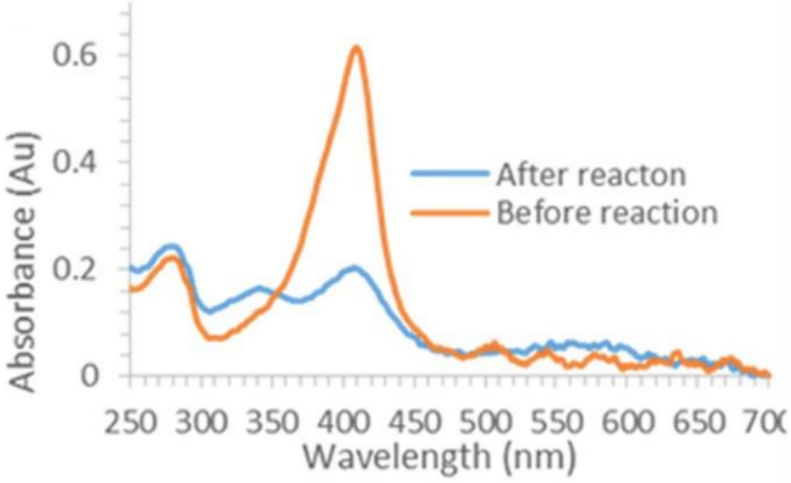


Figure 1.5. Change in absorbance peak reflecting catalytic degradation of PP-IX in *Bacillus cereus* TSPO (Guo et al. 2015).

Another hypothesis suggests that TSPO is involved in the trafficking and transport of porphyrin derivatives, as a part of heme metabolic pathway. Initial studies showed that binding of hemin significantly increased in monkey Cos-1 cells upon transfection with TSPO cDNA (Taketani et al 1995). Upon treatment with TSPO ligands (such as PK11195, Diazepam, Ro5-4864), conversion of coproporphyrinogen into protoporphyrinogen, which was performed by the enzyme coproporphyrinogen oxidase (located in the mitochondrial intermembrane space), was inhibited (Taketani et al. 1995). This suggested that the TSPO was involved in the transport of porphyrin

derivatives into the intermembrane space, which was inhibited by the TSPO ligands (Taketani et al 1995). In the moss *Physcomitrella patens*, knockout of TSPO resulted in accumulation of heme and PP-IX in the cytoplasm, in support of this hypothesis (Frank et al 2007). Some studies have also suggested that PP-IX transport by TSPO occur in association with other mitochondrial membrane proteins, such as voltage-dependent anion channel and adenine nucleotide transporter (Pastorino et al. 1994; Veenman et al. 2016).

TSPO has also been found to be involved in the process of erythropoiesis. During the development of Zebrafish embryos, TSPO has been found to be localized to the cells involved erythroid differentiation. Treatment of these cells with TSPO ligands resulted in inhibition of differentiation into blood cells (Fig. 1.6(Rampon et al. 2009). These observations led to suggest that TSPO is involved in erythropoiesis, by involvement in heme transport for synthesis of hemoglobin (Rampon et al, 2009). Similar observations have also been made in chickens (Nakazawa et al. 2009).

However, contradictory hypothesis also exists on the role of TSPO in heme metabolism and erythropoiesis. In TSPO knockout mice, no significant differences were observed in erythropoiesis and porphyrin metabolism, compared to the wild-type mice ((Tu et al. 2015; Zhao et al. 2016). Though many studies on the role of TSPO in porphyrin metabolism revealed some interesting insights (Veenman et al. 2016; Guo et al. 2015; Taketani et al. 1995; Ginter et al. 2013), it is still not clear whether the binding to porphyrins is part of transport or catalytic activity of TSPO, or both. However, some hypothesis put forth to clear this ambiguity suggests a dichotomy of both the roles in prokaryotes and eukaryotes, suggesting that the catalysis function by TSPO may have become redundant in eukaryotes (Li, Liu, Garavito, et al. 2015). Furthermore, the exact mechanism of porphyrin binding and transport involving TSPO is still unclear and remains to be studied.

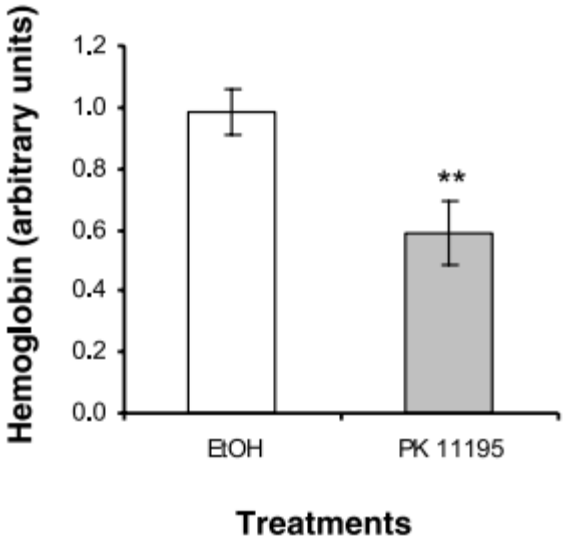


Figure 1.6. Effect of TSPO ligand PK11195 treatment versus ethanol treatment (control) on hemoglobin content in zebrafish (Rampon et al. 2009).

1.1.3. Apoptosis: Mitochondria has been found to play a critical role in the cell death. A complex on the mitochondrial membrane is proposed to initiate a cascade of events that ultimately leads to apoptosis. This multiprotein complex is termed as Mitochondrial Permeability Transition Pore (MPTP). MPTP constitutes of various proteins, including TSPO, Voltage Dependent Anion Channel (VDAC), Adenine Nucleotide Transporter (ANT), Cyclophilin D, Hexokinase, Bcl-2 to name a few(Veenman et al. 2008; Veenman et al. 2007; Gatliff et al. 2014). Activation of MPTP occurs upon receipt of specific cues, i.e., influx of Ca2+ ions, oxidative stress, depletion of adenine nucleotides and accumulation of ROS(Halestrap et al. 2002), In certain cases, it is also activated by treatment with anticancer agents, such as erucyl phosphocholine (ErPC).

MPTP exports the inner membrane protein cytochrome c into cytoplasm, which initiates a cascade of events involving activation of Caspase 3 and Caspase 9, ultimately leading to collapse in Mitochondrial Transmembrane Potential (MTP) and apoptosis. Initially, this complex was thought to be localized exclusively in the inner membrane. Subsequent studies revealed that proteins from outer membrane are also a part of MPTP (Szabó et al. 1993), and TSPO, among many other proteins were identified as components of MPTP (Kunduzova et al. 2004).

TSPO is shown to be involved in the process of cell death, along with other mitochondrial proteins that constitute MPTP. Involvement of TSPO in apoptosis has been demonstrated from two types of studies: Ligand binding and mutation/knockdown studies. Treatment of various TSPO ligands to cancerous cell lines resulted in both initiation, as well as inhibition of apoptosis, depending on various factors, including the type of ligand and dosage. It has been observed that treatment of TSPO ligands PK11195 and Ro5-4864 to cancerous cell lines induced apoptosis(Mendonça-Torres and Roberts 2013; Santidrián et al. 2007). On the other hand, treatment of ligands SSR180575, TRO40303 and 4-Chlorodiazepam inhibited MPTP formation and mitochondrial permeability, inhibiting the cell death(Morin et al. 2016; Obame et al. 2007; Hansson et al. 2015). In a study performed on C6 glioblastoma cell lines, survivability of cells decreased with increasing concentration of PK11195 treatment at micromolar quantities(Chelli et al. 2004). Similar observations have also been made from various other cancerous cell lines(Sutter et al. 2003; Sutter et al. 2004; Sutter et al. 2002; Tanimoto et al. 1999; Maaser et al. 2001; Mendonça-Torres and Roberts 2013; Santidrián et al. 2007). These observations have led to consensus that TSPO itself can be a novel target for cancer therapy, and TSPO ligands such as PK11195 and Ro5-4864 can be employed as anti-cancer drugs(Austin et al. 2013). However, PK11195 and Ro5-4864 also induced opposite effects, i.e., it inhibited apoptosis in another set of studies. Studies on glioblastoma cell lines indicated that upon treatment of PK11195 and Ro5-4864 with the anticancer agent Erucyl phosphohomocholine at micromolar quantities, the cell

proliferation increased, rather than initiating cell death(Kugler et al. 2008; Veenman and Gavish 2012). Similar observations was also made in human lymphoblastoid cell lines, where TNFα-induced apoptosis was inhibited(Bono et al. 1999).

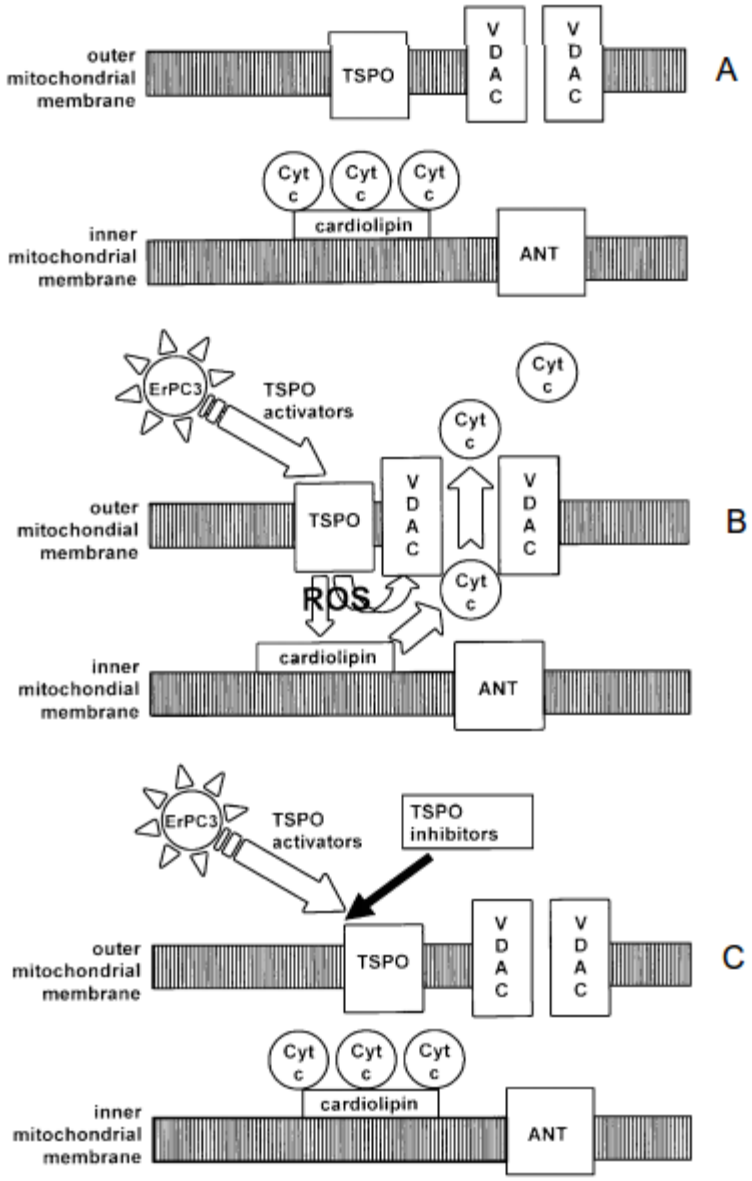


Figure 1.7. A: Schema showing arrangement of TSPO, VDAC and their interacting partners in mitochondrial membranes. B: Activated TSPO-VDAC complex, indicating cascade of events leading to apoptosis, including activation by binding of anticancer agent erucyl phosphocholine (ErPC). C: Inhibition of ErPC binding by TSPO ligands (Veenman et al. 2008).

Studies involving knockdown/ knockout of TSPO also indicate a role of TSPO in apoptosis. Crispr-Cas9 mediated TSPO mutation resulted in impaired MPTP formation(Fan et al. 2018), and antisense knockdown studies revealed that binding of TSPO ligands PK11195 & Ro5-4864 was significantly reduced, impairing MPTP-mediated apoptosis(Levin et al. 2005; Veenman and Gavish 2012; Veenman et al. 2008). On the other hand, overexpression of TSPO in rat glioma cells led to

increased motility and proliferation of the cells, suggesting a role of TSPO in the metastasis stage(Rechichi et al. 2008).

While it is evident that TSPO has a role in cell death, its exact function in this process remains unclear. Furthermore, TSPO ligands can elicit contrasting effects depending on the type of ligand and dose, which has also led to suggest that TSPO may play a protective role in pathologies pertaining to tissue injury and inflammation, where TSPO may prevent cell death by blocking the MPTP opening and apoptosis(Morin et al. 2016; Chechneva and Deng 2016; Daugherty et al. 2013).

A different hypothesis from recent set of studies suggest that TSPO has no role in the apoptosis, or the opening of MPTP. In TSPO deleted mitochondria and cardiac tissues, no difference was observed in functioning of MPTP, as well as in apoptosis. This led to propose that TSPO is not a component of MPTP, which must be regulated by other OMM proteins of unknown nature(Šileikytė et al. 2014). Furthermore, PK11195 was found to bind to ATP binding cassette transporter and prevent drug efflux in leukemia cells. This mechanism was found to be independent of expression of TSPO, which suggested that PK11195-mediated initiation of apoptosis did not involve opening of opening of MPTP via TSPO(Walter et al. 2005).

While TSPO has been extensively studied in context of maintaining mitochondrial potential and apoptosis, ambiguity persists on the exact role of TSPO, due to presence of contradictory narratives. Apart from the exact mechanism of activation of permeability transition, and the transition itself, the mechanism of interaction between the OMM and IMM proteins is still unclear, and the role played by TSPO ligands is still ambiguous.

1.1.4. Regulation of photosynthesis in green-photosynthetic bacteria: In *Rhodobacter sphaeroides*, a member of bacterial phylum alphaproteobacteria, a homologue of TSPO (36% identity with mouse TSPO) has been observed to regulate the photosynthesis. Bacterial TSPO were formerly termed as CrtK, but later named as TSPO, which was also an abbreviation for tryptophan rich sensory protein. Seminal studies indicated that TSPO, localized in outer membrane, may regulate photosynthesis, as null mutation of TSPO gene resulted in increased synthesis of photosynthetic pigments (Yeliseev and Kaplan 1995). It was then found that TSPO was involved in photosynthesis by negative regulation of coproporphyrinogen III oxidase. This regulation was proposed to occur by selective transport of tetrapyrrole derivatives across the membrane, which are further metabolized into photosynthetic pigments(Yeliseev and Kaplan 1999). The regulation process was found to be oxygen-dependent, where the expression of photosynthetic genes was repressed in presence of oxygen. This was also observed when rat TSPO was transfected to TSPO-deleted *R. sphaeroides* cells (Yeliseev et al. 1997).

1.1.5. Response to stress in plants: TSPO in plants was first characterized in *Arabidopsis thaliana*, as a homologue of mammalian TSPO(Lindemann et al. 2004). It had identical cholesterol and PP-IX-binding properties, suggesting strong homology with mammalian TSPO. Mutation studies on another TSPO homologue from the moss *Physcomitrella patens* suggested that TSPO is involved in response to the oxidative stress(Frank et al. 2007)Fig. 1.8(Frank et al. 2007). TSPO knockout cells were found to have high concentrations of H_2O_2 and cell death, hinting towards a role of TSPO in response to oxidative stress, similar to mammalian TSPO. Further studies proposed the role of TSPO in salt stress, osmotic stress(Guillaumot et al. 2009) and nutrient deprivation(Davey and de Bruijn 2000). Subsequently, heme binding to *A. thaliana* TSPO was characterized *in vitro*, and it was found that the residue His81 is essential for heme binding(Vanhee et al. 2011). Similar to TSPO in mammals, TSPO in plants are involved in response to stress through the means of interaction with tetrapyrroles. However, it remains unclear whether this interaction involves transport or catalytic degradation, and whether this response involves a multiprotein complex similar to MPTP.

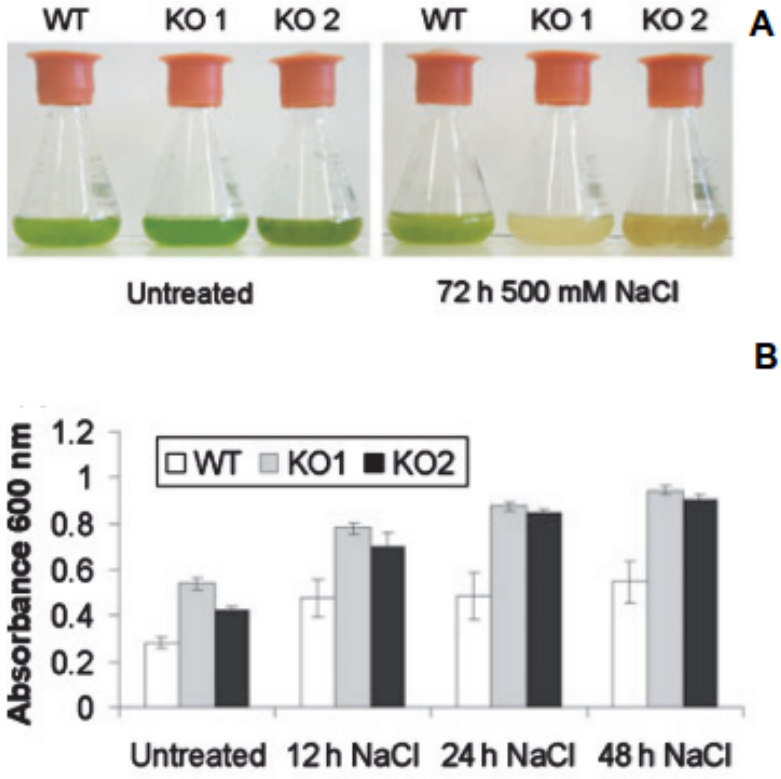


Figure 1.8. A: Growth of *Physcomitrella patens* wild-type (labelled WT) and TSPO knockout lines (labelled KO1 and KO2) under normal conditions (left) and under salt stress (after incubation with 500mM NaCl for 72 hours). B: Cell-death in wild-type cells and TSPO knockout lines (labelled KO1 and KO2) measured spectrophotometrically (Frank et al. 2007)

1.1.6. Other functions associated with TSPO:

i. Inflammation and Injury: Large number of studies suggest an active role of TSPO in conditions involving tissue injury and inflammation(Chechneva and Deng 2016; Morin et al. 2016; Lavisse et al. 2012; Chen and Guilarte 2008; Bonsack et al. 2016), and TSPO has been reported to have higher expression levels and activity in such tissues(Rey et al. 2000; Agnello et al. 2000; Raghavendra Rao et al. 2000; Bonsack et al. 2016). This prompted the proposal to use TSPO as a biomarker for diagnosis of such conditions by means of radioligands(Vowinckel et al. 1997; Lockhart et al. 2003; Rojas et al. 2007). Furthermore, treatment of popular TSPO ligands such as PK11195, Ro5-4864(Torres et al. 2000), Vinpocetine(Zhao et al. 2011), TRO40303(Schaller et al. 2010), 4'-Chlorodiazepam(Paradis et al. 2013) resulted in inhibition of collapse in MTP and cell death, which suggests possible protective effects of ligand binding to TSPO(Morin et al. 2016).

ii. Regulation of mitochondrial metabolism: While studies on the role of TSPO in regulation of metabolism are still in nascent stages, there is a considerable body of evidence that suggests a central role of TSPO in regulation of mitochondrial metabolism and homeostasis. In microglia cells, TSPO was found to be active during transition from dormant to proliferating stages(Liu et al. 2014; Gut 2015). Significant decrease in oxygen consumption and ATP production was observed in TSPO knockout cells(Banati et al. 2014). In agreement with the previous two studies, insertion of TSPO gene into Jurkat cells, which have low TSPO expression and activity, resulted in increase in transcription of genes involved in mitochondrial electron transport chain, and ultimately ATP production(Liu et al. 2017). Though much more needs to be studied on the exact role played by TSPO in regulation of metabolism, initial studies hint towards a central and critical role played by TSPO.

1.1.7. TSPO2-A mysterious paralog of TSPO: Recently, a paralog of TSPO, called TSPO2 was identified from search for remote TSPO homologs(Fan et al. 2009). TSPO2 was found to be distributed in mammals and birds along with TSPO1(Fan et al. 2009; Nakazawa et al. 2009), from which it was suggested that TSPO2 may have evolved from an ancient gene duplication event of TSPO gene. While TSPO2 retained cholesterol binding properties, evident from the conservation of CRAC motif, it lost the drug-binding properties, particularly, PK11195 binding properties(Fan et al. 2009). This was inferred from the fact that PK11195 did not bind to TSPO2 transfected to yeast cells, (Fan et al. 2009). Since TSPO2 was specifically observed in golgi bodies and endoplasmic reticulum of primitive erythrocytes, a niche role of TSPO2 was inferred, involving redistribution of free cholesterol. This step is essential for the final stage in the red blood cell maturation, involving the removal of nucleus to form enucleated erythrocytes. Because the involvement of TSPO1 in erythropoiesis was also observed in zebrafish(Rampon et al. 2009), a division of labour among TSPO1 and TSPO2 in differentiating erythrocytes was hypothesised(Fan et al. 2012). While TSPO1 in lower vertebrates is involved in erythropoiesis (Rampon et al. 2009), role of TSPO1 in higher

vertebrate erythropoiesis was hypothesised to be restricted to porphyrin transport and metabolism for biosynthesis of hemoglobins, whereas TSPO2 was proposed to be involved in the cholesterol redistribution(Fan et al. 2012; Fan et al. 2009). This division of labour was attributed to the fact that erythrocytes in lower vertebrates were nucleated, whereas erythrocytes in higher vertebrates were enucleated, which gave rise to the need for redistribution of cholesterol during the later stages of erythropoiesis.

However, recent studies revealed the presence of TSPO2 in matured erythrocytes as well (Marginedas-Freixa et al. 2016; Bouyer et al. 2011). In context of malarial infection in schizont stage, TSPO2, along with VDAC and ANT was found to be active in *P. falciparum* infected red blood cells (Bouyer et al. 2011). Interestingly, it was also found that treatment of TSPO ligands, such as PK11195 and Ro5-4864 resulted in inhibition of *P. falciparum* growth (Bouyer et al. 2011). While this is contrary to the earlier study on TSPO2 (Fan et al. 2009), it was observed that the ligands bound with lower affinity than TSPO1 (Marginedas-Freixa et al. 2016). Subsequently, expression of TSPO2 was also confirmed in healthy red blood cells as well (Marginedas-Freixa et al. 2016). TSPO2 was observed to exist as a dimer and tetramer, and also as a complex with Voltage-Dependant Anion Channel (VDAC). In erythrocytes infected with the malarial parasite *Plasmodium falciparum*, both the proteins were observed to be a part of a larger 200 kDa complex, which is hypothesized to facilitate import of nutrients for the parasite growth (Fig. 1.9). Furthermore, TSPO2 was observed to incorporate Zinc-Protoporphyrin-IX, in both normal and infected erythrocytes. Unlike TSPO1, where PP-IX uptake is inhibited upon treatment with ligands, the Zn-PP-IX uptake here was augmented upon treatment with ligands(Marginedas-Freixa et al. 2016).

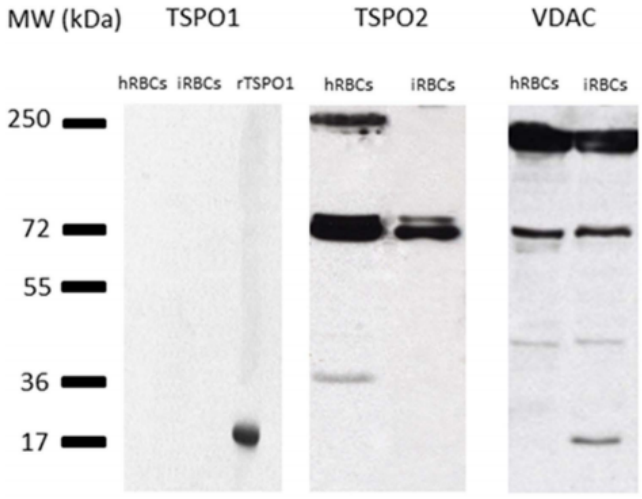


Figure 1.9. Immunoblot showing expression of TSPO1, TSPO2 and VDAC in healthy and *P. falciparum*-infected red blood cell membranes. Recombinant human TSPO1 (rTSPO1) was used as a positive control (Marginedas-Freixa et al. 2016)from(Marginedas-Freixa et al. 2016).

Though it is clear that TSPO2 appears to have a role relegated to red blood cell development, recent observations on presence in erythrocytes and ability to uptake Zn-PP-IX implies a broader role for

TSPO2. On the other hand, though comparison of TSPO1 and TSPO2 sequences suggested that TSPO2 lacks mitochondrial localization sequence (Fan et al. 2012), it is unclear what are the differences between TSPO1 and TSPO2 at the level of ligand binding residues, particularly PP-IX and cholesterol binding residues. The mystery is also compounded by the fact that there are currently no structures of TSPO2 available, which raises the question of differences in dynamics of TSPO1 and TSPO2, due to the fact that both the protein sequences have considerably low sequence identity (40%).

1.2. Involvement/association of TSPO with different pathologies:

***1.2.1. Malaria*:** During the 'blood stage', or erythrocytic stage (merozoite, schizont and trophozoite stages, Fig. 1.1) of malarial infection by *Plasmodium falciparum*, the membrane permeability of erythrocytes increases, to various molecules including carbohydrates, polyols, amino acids and anions(Ginsburg et al. 1983; Ginsburg et al. 1985). The increase in permeability was characterized by formation of a pore of 0.7nm diameter, which was termed as New Permeability Pathway (NPP). It was later observed that NPP consist of anion channels that transport the nutrients(Ginsburg and Stein 2005). Recently, it was observed that certain TSPO variants, (especially TSPO2), along with VDAC and ANT constitute the NPP, and are active in erythrocyte membrane during the erythrocytic invasion stage of malaria, though similar observations have also been made with TSPO2 ((Bouyer et al. 2011; Marginedas-Freixa et al. 2016)). Furthermore, TSPO ligands PK11195 and Ro5-4864 were able to inhibit the growth of *P. falciparum,* suggesting involvement of TSPO in NPP(Bouyer et al. 2011; Marginedas-Freixa et al. 2016).

1.2.2. Neurodegenerative diseases: Involvement of TSPO has been suggested in various age-related neurodegenerative diseases, such as Alzheimer's and Parkinson's disease to name a few, due to the fact that there is increased uptake of TSPO ligands in these conditions (Fig. 1.10; Gerhard et al. 2006). It has been also suggested that TSPO has a role in the pathogenesis, particularly in the microglial tissues(Liu et al. 2015; Repalli 2014; Kreisl et al. 2013; Veenman et al. 2007). TSPO has also been suggested to be involved in other neurological syndromes, such as chronic depression(Richards et al. 2018) and amyotrophic lateral sclerosis(Zürcher et al. 2015), and has also been associated with certain psychological conditions, such as depression, panic disorder and bipolar disorder (Richards et al. 2006; Colasanti et al. 2013; Nakamura et al. 2006). Interestingly, a mutant of TSPO, A147T has been observed to be correlated with occurence of above psychological conditions. This is interesting due to the fact that the location of the mutation is adjacent to the cholesterol-binding CRAC motif (Fig. 1.3, Box 1.2). This mutation has also been associated with reduced pregnenolone production, for which cholesterol translocation is necessary (Costa, Pini, Gabelloni, et al. 2009). In many of these conditions. activity of TSPO corresponds with neuroinflammation, which has been associated with

activation of MPTP, leading to the cell death in traumatic brain injury(Mazzeo et al. 2009). Due to the fact that treatment of TSPO ligands has resulted in significant improvement of the condition, TSPO has been suggested to be a therapeutic target for treatment of these pathologies(Rupprecht et al. 2010; Papadopoulos and Lecanu 2009; Nothdurfter et al. 2012).

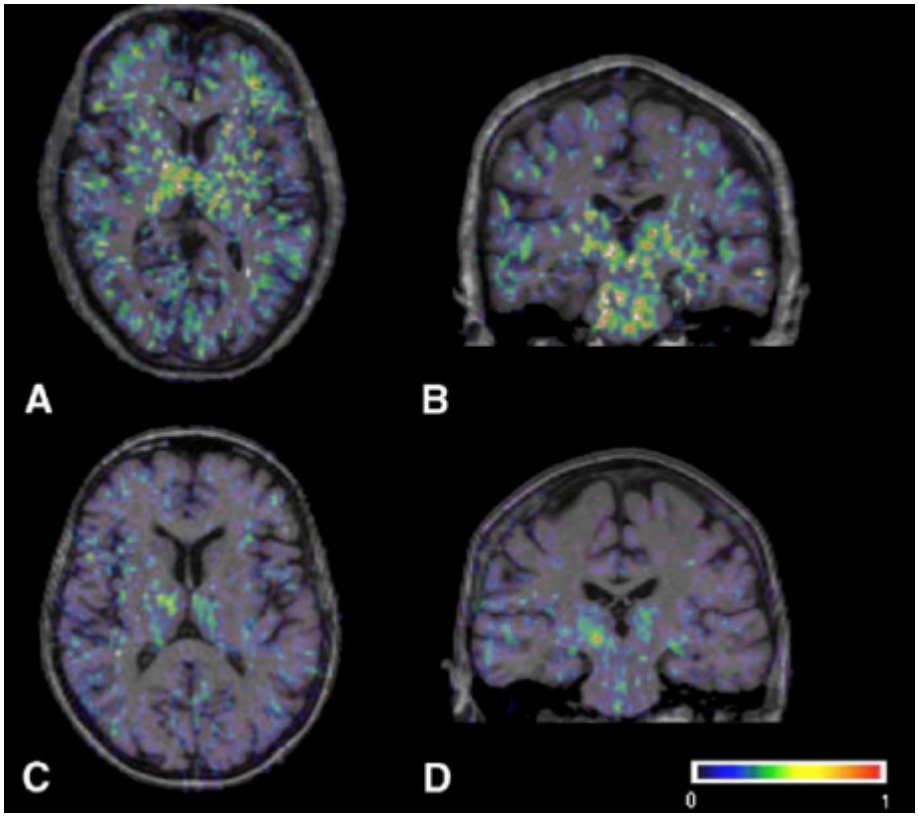


Figure 1.10. Traverse (A) and cross-sections (B) of brain of Parkinson's disease patient and Traverse (C) and cross-sections (D) of brain of healthy control patient showing radiolabelled PK11195 binding. Blue regions indicate tissues with low binding, yellow and red regions indicate regions with high binding (from Gerhard et al. 2006).

1.2.3. Cancers: TSPO has been observed to play a critical role in mitochondria-mediated apoptosis(Veenman and Gavish 2012; Veenman et al. 2008; Veenman et al. 2007), and has also been found to be active in metastatic tissues in various cancers(Zheng et al. 2011; Wu and Gallo 2013; Bhoola et al. 2018; Rechichi et al. 2008). Though there has been growing evidence on the role of TSPO in metastasis, a complete and comprehensive picture is yet to emerge. However, treatment of TSPO ligands to cancerous cell lines has resulted in apoptosis in various tissues(Mendonça-Torres and Roberts 2013; Maaser et al. 2001; Sutter et al. 2002; Sutter et al. 2003) due to which, TSPO has been proposed as therapeutic target for chemotherapy, and many TSPO ligands have been proposed as anticancer agents(Austin et al. 2013; Daniele et al. 2016; J. Li et al. 2016; Werry et al. 2015).

1.3. Structures and Structure-dynamics relationships of TSPO:

1.3.1. Early studies on TSPO structure: Due to the fact that membrane proteins are inherently difficult to purify and study, structure of TSPO at the atomic level was unknown until recently. However, various attempts have been made in past to understand the architecture and fold of the protein. Among the earliest insights came from molecular modelling studies, which suggested that TSPO has a transmembrane (TM) domain consisting of 5 TM alpha helices(Bernassau et al. 1993). Based on the molecular simulations, it was proposed that since the protein represents topology of positively charged receptor surface and uncharged TM domain, TSPO may accommodate cholesterol molecule inside the TM pore without any significant geometric distortions or energy differences(Papadopoulos et al. 1994). This was also observed from docking studies between cholesterol and the cholesterol binding CRAC motif(Jamin et al. 2005). It was further observed from NMR experiments that the ligand binding plays an important role in the stabilization of the TM domain of TSPO(Murail et al. 2008).

Further cryo-electron microscopy studies supported the hypothesis of 5 TM helical architecture(Korkhov et al. 2010). TSPO from *R. sphaeroides* was crystallised and its structure was deduced by means of cryo-electron microscopy in absence of a ligand, with a resolution of 10Å (Fig. 1.11, A; (Korkhov et al. 2010)). It was observed to be a symmetrical homodimer upon solubilization by n-Dodecyl β-D-maltoside detergent and subject to blue native PAGE analysis (Korkhov et al. 2010). Due to the fact that a groove was observed in the interface of two monomers, it was proposed that this groove is the site for binding of the porphyrin ligand, through which the porphyrin transport may take place. It was also observed that two of the helices are wide apart to form another groove. This region was proposed to be second putative pathway for transport of heavy hydrophobic molecules such as cholesterol, akin to the transport mechanism observed in EmrE, Sav1866 etc. Though from this study, structure of TSPO was studied and characterized in near-atomic resolution and TSPO was also observed to be a homodimer, the residues could not be identified from the low-resolution of the cryo-EM images, due to which the exact fold or arrangement of TM helices with respect to each other could not be characterized.

To further characterize the positions of each TM regions from the cryo-EM study by Korkhov et al, a molecular modelling approach was utilised to understand the dominant interfaces of TSPO dimer. The top 2 models, which consisted of interfaces corresponding to TM5 from each monomer, and TM1-3 respectively had significant energy differences compared to other lower-ranking models(Li et al. 2013). A similar study by(Jaremko, Jaremko, Becker, et al. 2014) on the same cryo-EM data characterized by Korkhov et al, revealed the interface to be shared between TM1 and TM2(Jaremko, Jaremko, Becker, et al. 2014). These studies also supported the hypothesis put forth by Yeliseev et al., that the extracellular loop between TM1 and TM2 plays an important role in the dimerisation, as

mutation of the residue W38C in this region led to stabilization of the dimer, which was mediated by formation of disulphide bridges by the mutant cysteine residues(Yeliseev and Kaplan 2000).

Another study exploited the data from cryo-EM studies by Korkhov et al., along with molecular modelling to describe the fold of TSPO. TSPO fold was proposed to have an antiparallel, anticlockwise arrangement of TM helices spanning from TM1 to TM5, when viewed from cytoplasmic side. Contrary to the previous interpretations of cryo-EM maps, the oligomeric interface was predicted to be between TM3 and TM4, based on the inputs from molecular modelling and prediction of evolutionarily coevolving residues(Hinsen et al. 2015). Though this interpretation was in contradiction to the earlier interpretations, it was in better agreement with the crystallographic constraints from the cryo-EM densities compared to previous interpretations. The contradictions in interface were attributed to the difference in sequences between the bacterial and mammalian TSPO, which suggested possible differences in the folds, as well from the differences in dynamics stemming from the lack of bound ligand in the cryo-EM density maps(Hinsen et al. 2015).

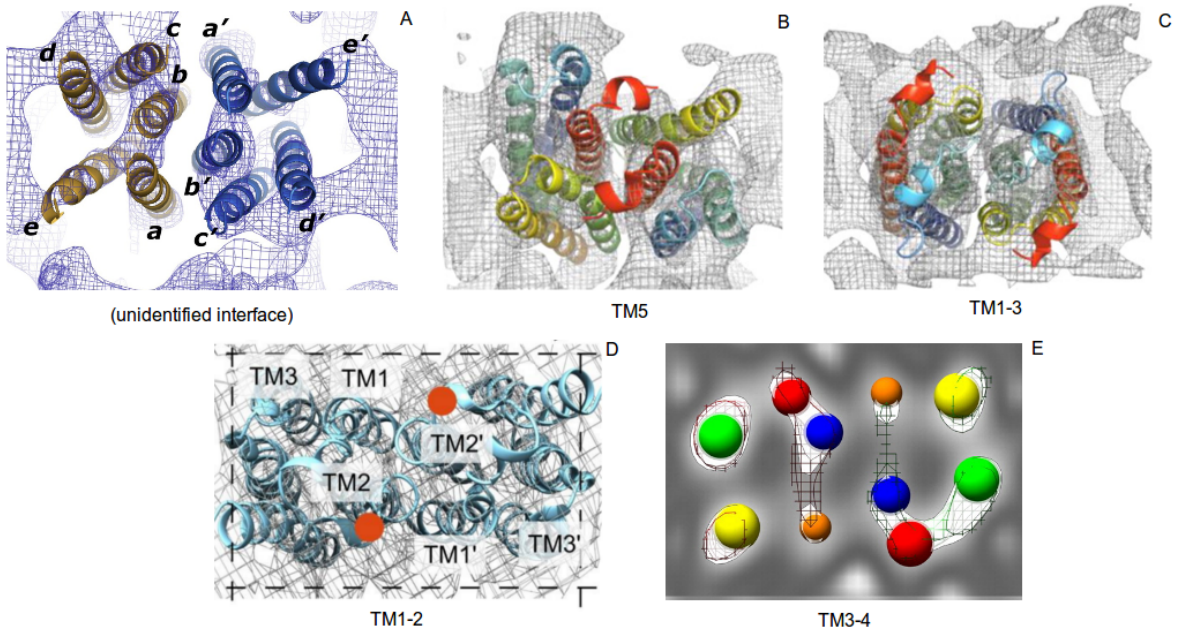


Figure 1.11. Interpretation of TSPO oligomeric interfaces from the cryo-EM map by Korkhov et al. A: Proposed oligomeric interface by Korkhov et al. B: Highest ranking dimer model by Li et al. with interface consisting of TM5. C: Second highest ranking dimer model by Li et al. with interface consisting of TM1-3. D: Interface proposed by Jaremko et al., involving TM1-2. The location of W38C mutation is represented as orange circle. E: Interface proposed by Hinsen et al., involving TM3-4.

1.3.2. Studies on atomic structure of TSPO: Though the study by Korkov et al. may be counted as one of the first attempts to characterize TSPO structure, the low resolution of the cryo-EM map limited the scope of characterization of TSPO structure at the atomic level. However, a major breakthrough was achieved from solving the solution-NMR structure of mouse TSPO(Jaremko,

Jaremko, Giller, et al. 2014). The structure of mouse TSPO was solved in presence of the ligand PK11195 bound to TSPO, reconstituted in dodecylphosphocholine (DPC) micelles. In agreement with the cryo-EM studies (Korkhov et al. 2010), it was observed that TSPO folded into a bundle of five TM helices. However, the fold, or arrangement of TM helices with respect to each other was different from previous predictions (Li et al. 2013, Hinsen et al. 2015). Presence of numerous proline residues, particularly in TM1, 2, 3 & 5 resulted in kinks in the respective TM regions. The presence of charged residues at the C-terminus, and on the residues of cytoplasmic side suggested that these residues may play a role in accommodating the ligand. Though most of the loops between the TM regions are short, the primary exception is the loop between TM1 and TM2 regions corresponding to the residues 28 to 45. This loop consist of 18 residues, which is hypothesized to act as a gate for the entry of ligand(F. Li et al. 2016).

While TSPO was found to have a distinct binding cavity that hosts the PK11195 ligand, this cavity was observed to be distant to the cholesterol-binding CRAC motif (Jaremko et al., 2014b). Due to the fact that binding of PK11195 resulted in significant improvement in assignment of NMR resonances, it was inferred that PK11195 binding not only stabilized tertiary structure of TSPO, but it may also facilitate binding of cholesterol to TSPO (Jaremko, Jaremko, Giller, et al. 2014).

Similarly, structure of A147T mutant of mouse TSPO was also solved using NMR spectroscopy method (Jaremko et al. 2015). Akin to the wild-type, the structure was solved in complex with PK11195 ligand. Upon comparison between wild-type and A147T mutant structures, the Root Mean Square Deviation (RMSD) between the backbone atoms of transmembrane regions was 0.95, indicating high structural similarity. In addition, the dynamics of CRAC motif residues appeared to be unperturbed, despite the mutation in the adjacent A147 residue(Jaremko et al. 2015).

While the NMR structures yielded valuable initial insights into the atomic structure of TSPO, X-ray crystallographic structures of bacterial TSPO were solved in quick succession. Structures of *Rhodobacter sphaeroides* and *Bacillus cereus* TSPO supported the observations from cryo-EM experiments on existence of higher-order oligomeric forms of TSPO(Li, Liu, Zheng, et al. 2015a; Guo et al. 2015). The fold was also found to be similar to that of the mouse TSPO, which could be attributed to strong conservation of residues located in TM regions. The structure of *R. sphaeroides* TSPO was solved for the wild type as well as A139T mutant, which is identical to A147T mutant of human TSPO (Fig. 1.3). The structure of wild-type TSPO was solved as a dimer at 2.5Å resolution, and the mutant was solved at 1.8Å resolution. Interestingly, while the wild-type structure was solved as a dimer involving TM1 and TM3 as the interface, the A139T mutant structure was solved as an asymmetric trimer. In the mutant structure, the interface corresponding to the first pair of monomers consists of TM1 and TM3, and the interface involving the second pair of monomers consist of TM1, 2 and 5. Furthermore, the residues of first extracellular loop in the wild-type structures could not be

assigned electron densities. This could be attributed to the fact that the wild-type structure was solved in absence of any ligand, suggesting that in absence of ligand, the residues of this region may be highly dynamic. In addition, upon comparison of 20 lowest energy conformations from the NMR structure of mouse TSPO, it may be observed that the region corresponding to the first extracellular loop has highest RMSD (Fig. 1.12; Jaremko, Jaremko, Giller, et al. 2014).

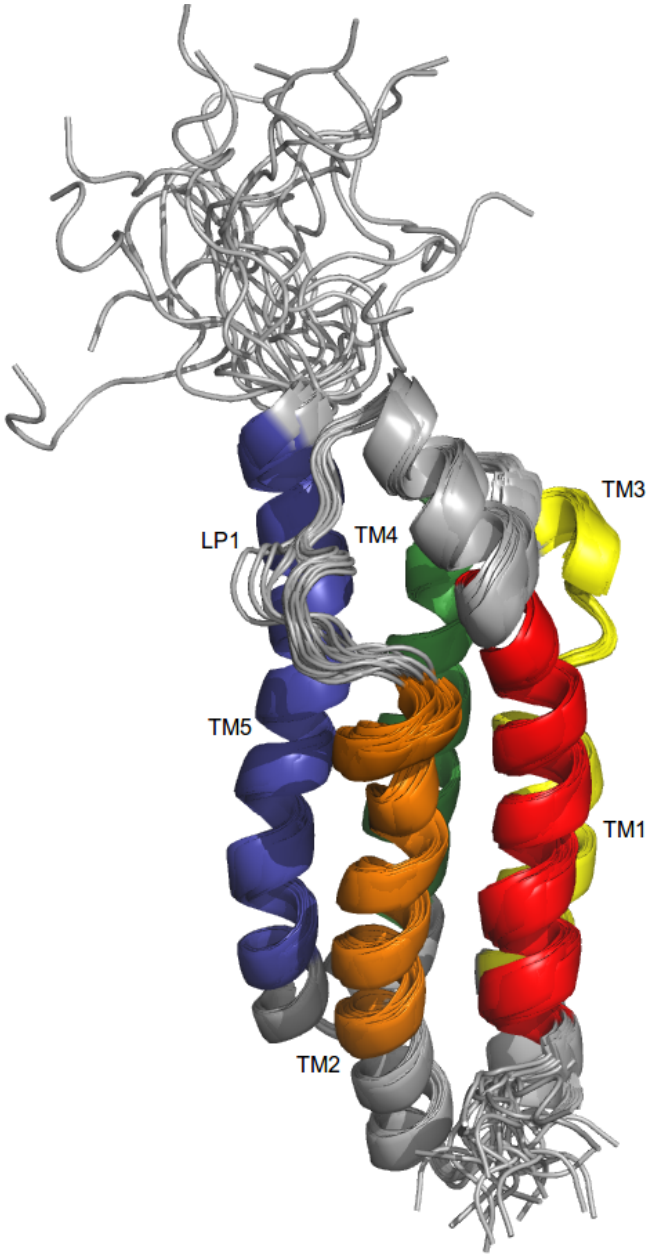


Figure 1.12. NMR structure of mouse TSPO, showing 20 lowest energy conformations. TM regions 1 to 5 are represented as red, orange, yellow, green and blue cartoons respectively and labelled. The extracellular loop 1 is labelled 'LP1', with average RMSD of this region between being 1.78Å. The superimposition was performed using VMD program(Pettersen et al. 2004; Humphrey et al. 1996), and visualised using Pymol program (DeLano 2002).

Interestingly, while the wild-type structure of *R. sphaeroides* TSPO was solved in absence of any ligand, the structure of A139T mutant was solved in presence of the endogenous PP-IX ligand. Though the mutant structure was solved as a trimer, PP-IX was found to be in complex with only one of the monomer chain. On comparison between the *R. sphaeroides* TSPO wild type and mutant structures, differences in conformations adopted were observed among the residues of TM2 and TM5 regions, particularly among the residues of CRAC motif in TM5. These differences correspond to inward tilt of TM2 and TM5, by 7.7° and 6.3° respectively. It was inferred that this tilt resulted in contraction of the binding cavity, which has an adverse impact on the cholesterol binding to TSPO. However, RMSD between the C-alpha atoms of TM2 and TM5 regions are 1.123 and 0.88Å respectively (Fig. 1.13A). Furthermore, it is not clear if this difference in conformation is also influenced by the fact that the mutant structure has PP-IX in complex with TSPO. Furthermore, though the mutant A139T structure is the only TSPO structure in complex with an endogenous ligand, some doubts have been raised regarding the electron densities of the ligand(Wang 2015). As much as 50% of PP-IX atoms are in geometric outlier (Fig. 1.13C), and the ligand density-fitting scores are as high as 11.24 (scores lesser than 2.0 are considered good quality scores,(Wang 2015). In response, the authors of the structure commented that porphyrins with such geometric distortions are frequently encountered in proteins that degrade heme or heme-based derivatives. Since some studies also propose a catalytic role of TSPO in degrading porphyrins (Guo et al. 2015), the authors commented that encountering porphyrins with such geometry is expected(Li, Liu, Zheng, et al. 2015b).

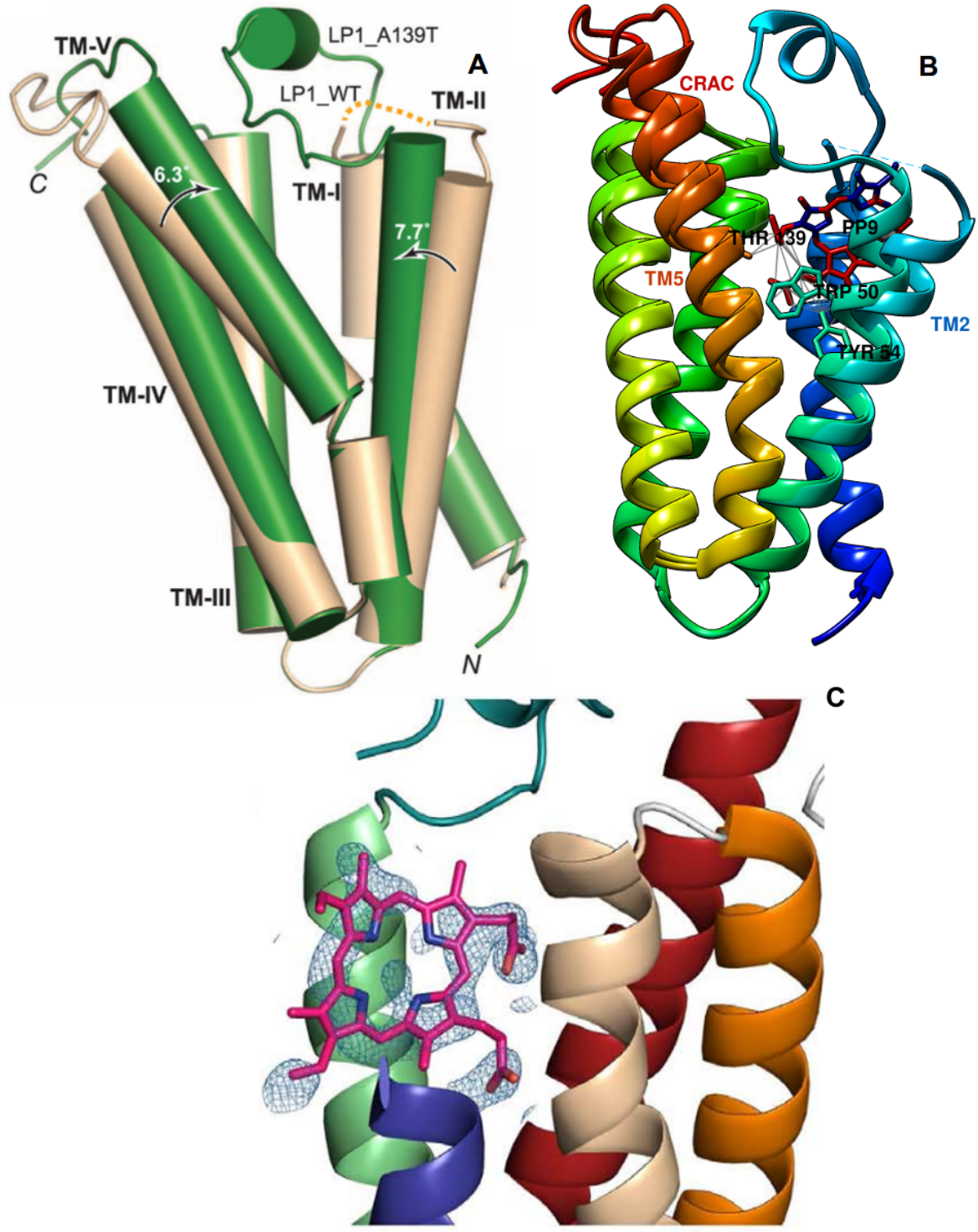


Figure 1.13. A: Superimposition of wild-type and A139T mutant monomer structures of *R. sphaeroides* TSPO, showing tilt of TM2 and 5. Wild type structure is represented as yellow cartoon, mutant structure is represented as green cartoon (Image from Li et al. 2015). B: Ribbon representation of same structures, indicating the contacts between the PP-IX ligand and the residues Trp50, Tyr54 and Thr139. PP-IX is represented as red and blue sticks. C: Ribbon representation of A139T mutant, showing PP-IX and its binding cavity. Corresponding electron densities are represented as mesh (Image from Li et al. 2015).

Apart from *R. sphaeroides*, TSPO structure of of *Bacillus cereus* was also solved(Guo et al. 2015) by means of X-ray crystallography, with resolutions ranging from 4.1 to 1.7Å. Like *R. sphaeroides*, TSPO from *B. cereus* was also found to be in oligomeric form. However, two major differences may be observed from the structures of *B. cereus* TSPO: *1.* While both trimer as well as dimer structures were observed for *R. sphaeroides* TSPO, no trimer structures were observed for *B. cereus* TSPO. *2.* The interface observed for *B. cereus* involved TM2, rather than TM1 & 3, as observed in *R. sphaeroides* TSPO. Furthermore, the buried surface area (BSA) for *B. cereus* dimer is significantly lesser (958Å^2) than that of *R. sphaeroides* wild-type dimer structure (3,061Å^2). Since the BSA is lesser than 1,000Å^2, the stability of the dimer itself may be questionable.

However the differences, a major commonality between the interfaces of two TSPO structures is presence of a motif frequently encountered in TM dimer proteins, termed as 'GxxxG' motif. This motif has been observed in the interfaces of various oligomeric transmembrane proteins, including ATPase-C(Pogoryelov et al. 2012), Glycophorin-A(Lemmon, Flanagan, Hunt, et al. 1992; Lemmon, Flanagan, Treutlein, et al. 1992), TM domain of BCL2/adenovirus E1B 19 kDa protein-interacting protein 3(Sulistijo and Mackenzie 2009) to name a few(Teese and Langosch 2015). In many of the TM dimers involving this motif, the concerned residues of this motif have been observed to meditate oligomerization by means of inter-helical hydrogen bonding among the backbone atoms(Li et al. 2012; Weber and Schneider 2013; Mueller et al. 2014). The motif has a characteristic signature of glycine residues at first and fifth position, with the middle three residues being amino acid residues of any chemical composition, though in addition, aromatic residues such as phenylalanine in the vicinity of these motifs have been observed to enhance the stability of the interactions in bacterial TM dimers(Unterreitmeier et al. 2007). Though such motifs have been observed in TM dimers, a similar motif, having similar signature, with the difference being alanine residues instead of glycine at the terminal positions have been observed, and these motifs are termed as 'AxxxA' motifs(Kleiger et al. 2002). It was then found that this motif can be generalized based on any small amino acid, which includes glycine, alanine, threonine and serine. This motif was proposed to be renamed as 'SmxxxSm' motif, where the first position corresponds to any small amino acid residue (glycine, alanine, serine and threonine)(Duneau et al. 2007; Eilers et al. 2002). However, other groups also cautioned on over-reliance on such sequence-based dimer prediction methods, emphasising that significant number of TM dimer interactions have also been observed that does not involve GxxxG or SmxxxSm motifs(Li et al. 2012).

The dimer interface of *R. sphaeroides* TSPO has two such motifs, both localised on TM3 that have significant contribution towards the dimer interface (71-GQALA-75; 75-AFYAA-79). Similarly, the dimer interface of *B. cereus* TSPO has two such motifs having major contribution towards to dimer

interface (44-GMTIG-48; 48-GMIWA-52). X-ray crystal structures of both the bacterial TSPOs suggest that SmxxxSm motifs may be major contributors of the dimer interface.

While all the three TSPO structures are observed to have similar fold composed of 5 TM regions, comparison of X-ray structure of *B. cereus* with the NMR structure of mouse TSPO revealed significant differences between both the structures, despite the fact that both the structures were solved in complex with the ligand PK11195. RMSD of Cα atoms between 148 residues was as high as 4.9Å. While TM2 and TM5 regions were relatively well superimposed, the RMSD was particularly high between the residues of TM1, 3 and 4(Guo et al. 2015). When TSPO monomers from all the three species are superimposed, the differences in conformation become clear (Fig. 1.14).

RMSD is high between all the TM regions, particularly between the TM3 and TM4 regions (Fig. 1.14, D&E). When the sequence alignment based on the structural superimposition is visualized, it may be observed that the boundaries of TM regions are displaced by approximately 2-3 residues for each TM region (Fig. 1.15). While this displacement may be observed for every TM region, it is unclear whether this displacement is attributed to the conditions under which respective experimental structures were determined, or if it is influenced by inherent differences which may stem from the differences in sequences and overall physico-chemical properties of the protein. While Guo et al., argued that there are significant differences in conformations between bacterial and eukaryotic TSPOs, which is more visible from the comparison of all the TSPO structures (Fig. 1.14, D&E), the root of this difference remains to be understood. One possible reason could be due to the fact that NMR structure was solved in presence DPC detergent micelles, which is often considered as a particularly strong detergent, that may affect the fold of the protein (Li, Liu, Garavito, et al. 2015). However, there may also be a possible role of sequence and physico-chemical nature of the protein in determining conformation of protein. Though influence of sequence on the structure of protein has been widely debated, it is still unclear how much does it contribute with the case of TSPO.

Most of the studies on the structure of TSPO discussed so far have been performed in presence of a ligand, exogenous or endogenous. There are few studies till date that characterize TSPO in absence of any ligand from the structural perspective. This has been attributed to the fact that TSPO itself is a difficult protein to isolate unless it is bound to a ligand. In one of the earlier attempts to study the TSPO structure, it was inferred that ligand binding stabilized the tertiary structure of TSPO(Murail et al. 2008). In an attempt to understand mouse TSPO structure in absence of ligand by means of NMR spectroscopy, it was observed that TSPO is highly flexible in absence of a ligand(Ł. Jaremko et al. 2015). This was evident from the fact that for many of the TM regions, no reasonable-quality NMR signals were observed, which suggests a highly dynamic nature of these residues. Furthermore, for the residues which had good-quality NMR signals, a significant reduction in helicity was observed,

indicating deformations of TSPO residues at the level of both secondary, as well as tertiary structure(Ł. Jaremko et al. 2015).

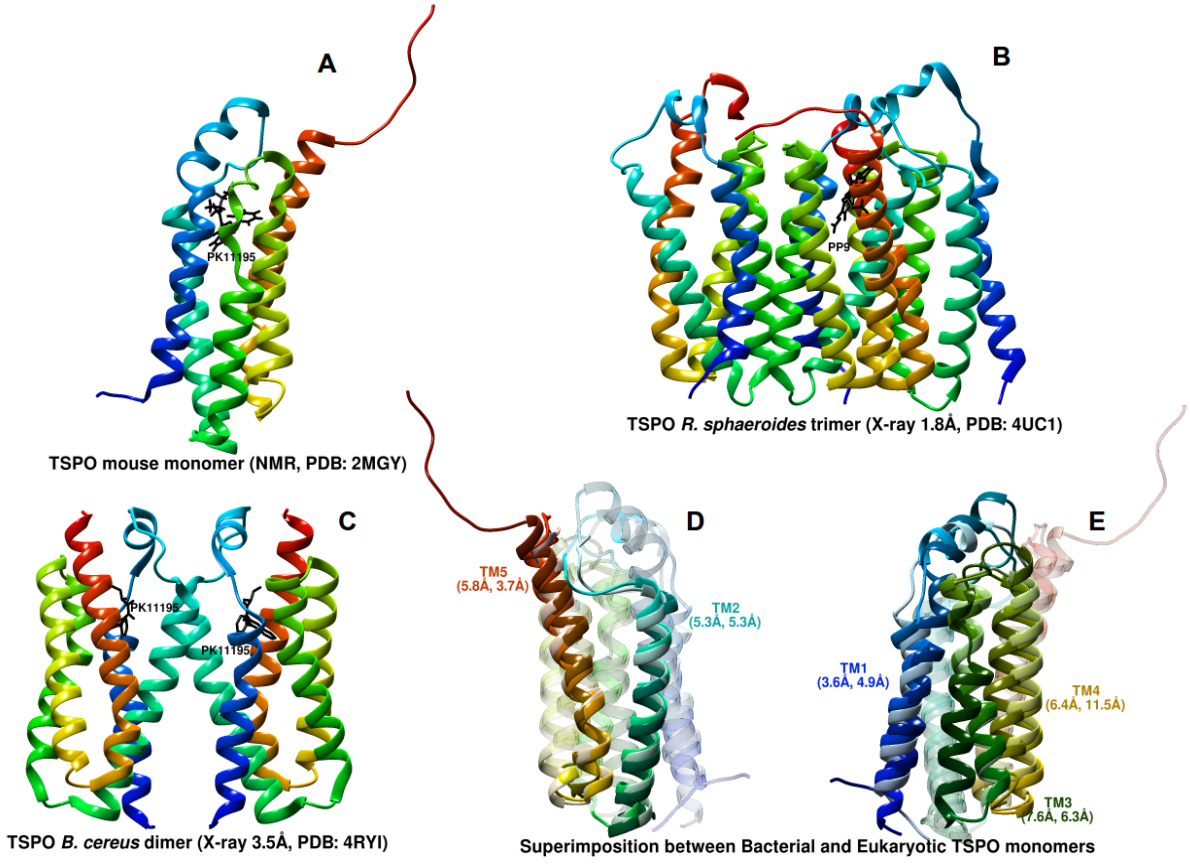


Figure 1.14. Experimental structures of TSPO. A: NMR structure of mouse TSPO (Jaremko et al., 2014). B: X-ray crystallographic structure of *R. sphaeroides* TSPO (Li et al., 2015). C: X-ray crystallographic structure of *B. cereus* TSPO (Guo et al., 2015). Ligands PP-IX (B) and PK11195 (A&C) are labelled accordingly. D: Superimposition between TSPO monomers, showing superimposition between the TM2 and TM5 regions. E: Same as D, showing the superimposition between TM1, 3 & 4 regions. Structures are represented in rainbow colours from N- to C-terminal. Superimposition was performed on C-alpha atoms of the residues of TM regions, based on the TM boundaries from OPM database(Lomize et al. 2012). TM regions are labelled, with RMSD values between the C-alpha atoms of respective TM regions of mouse TSPO and *R. sphaeroides* and *B. cereus* TSPO represented in brackets respectively.

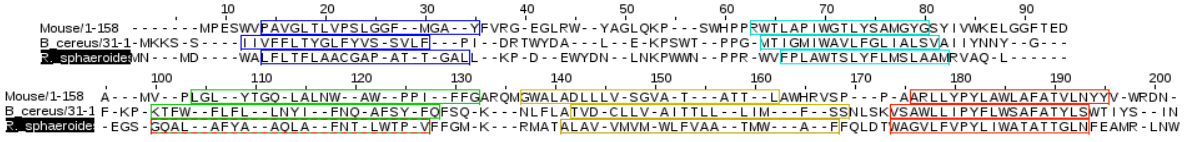


Figure 1.15. Structure-based sequence alignment of TSPO sequences, based on structural superposition of TSPO experimental structures (represented in Figure 1.13, D&E). Residues corresponding to TM1-5 are represented in blue, cyan, green, yellow and red boxes respectively. Structure superposition and sequence alignment was performed using UCSF Chimera program(Pettersen et al. 2004).

TSPO structure	Mouse	*R. sphaeroides*	*B. cereus*
Exp. environment	DPC (Dodecylphosphocholine)	OLC (1-Oleoyl-R-glycerol); (2S)-1-(hexadecanoyloxy)-3-hydroxypropan-2-yl(11Z)-octadec-11-enoate; O-Acetaldehydyl-hexaethylene glycol	DDM(N-dodecyl-β-D-maltopyranoside)
Seq. identity w/mouse TSPO (%)	100	36	30
Stoichiometry	Monomer	Trimer (TM1,3 interface)	Dimer (TM2 interface)
Ligand	PK11195	PP-IX (+1 *apo* form)	PK11195
Mutation	wild-type	A139T mutant (+wild type)	wild-type

Table 1.2. Summary of experimental TSPO structures from different species, depicting experimental environment, stoichiometry and ligands.

1.4. Interacting partners of TSPO

1.4.1. Endogenous ligands of TSPO:

i. Cholesterol: Cholesterol has a characteristic 4-ring structure with an aliphatic tail (Fig. 1.16). While the hydrophobic residues of CRAC motif in TSPO are proposed to interact with the ringed backbone, the residues of enhancement motif, which are more polar chemically compared to CRAC motif, are proposed to interact with the aliphatic tail scaffold(Li, Liu, Valls, et al. 2015).

While many studies suggested an important role of TSPO on cholesterol transport and steroidogenesis, among the earliest studies that suggested direct binding of cholesterol to TSPO was from site directed mutagenesis experiments on the residues of CRAC motif, which suggested existence of a putative motif that binds to cholesterol(Li and Papadopoulos 1998; Li, Yao, et al. 2001). This motif was later characterized as CRAC motif, whose binding with cholesterol was also extensively studied(Jamin et al. 2005; Li, Yao, et al. 2001). It has also been demonstrated that cholesterol has strong interaction with TSPO at nanomolar affinity(Lacapère et al. 2001). Though the exact nature of cholesterol interaction with TSPO, whether it is regulatory or it is transported is still under debate, another motif adjacent to the CRAC motif was recognised in mouse TSPO, which was termed as 'enhancement motif' for the binding of cholesterol(Li, Liu, Valls, et al. 2015). Though CRAC motif is conserved in *R. sphaeroides* TSPO, *R. sphaeroides* TSPO does not transport cholesterol, possibly due to the fact that cholesterol does not exist in bacterial membranes. However, when the enhancement motif was inserted into *R. sphaeroides* TSPO, cholesterol quenching increased significantly, suggesting an indirect role of this motif in cholesterol transport(Li, Liu, Valls, et al. 2015). It was suggested that the residues of this motif bind to the cholesterol tail. While this motif is

conserved in mammalian TSPO sequences, it is variable in bacterial sequences. Since bacterial membranes do not have cholesterol, they consist of a hopanoids, a structural analogue of cholesterol. Hopanoids tend to have a variable tail structure depending on species, this was proposed to be the cause for higher variation of this motif in bacterial TSPO sequences (Li, Liu, Zheng, et al. 2015a; Fan et al. 2012). While there have been contradictory studies that suggest that there is no role of TSPO in cholesterol transport(Selvaraj and Stocco 2015; Morohaku et al. 2014; Tu et al. 2015), recent theoretical studies have proposed another cholesterol interacting motif, whose sequence is palindromic with respect to the CRAC motif(Fantini et al. 2016). This motif was termed 'CARC' motf. Furthermore, authors predicted that this motif interacts with cholesterol with lower energy compared to CRAC motif. Since both the motifs are present in either of the lipid bilayer leaflets, it was suggested that both these motifs enabled TSPO to interact with cholesterol at both the bilayers simultaneously(Fantini et al. 2016; Papadopoulos et al. 2017).

ii. Protoporphyrin-IX (PP-IX): Protoporphyrin-IX (PP-IX) consists of a heme backbone with two carbonyl tails, making it a highly planar molecule (Fig. 1.16). While the heme backbone can chelate with various metal ions, it is unclear which form of PP-IX predominantly binds to TSPO.

PP-IX is the second endogenous ligand, which is hypothesised to be transported via TSPO. In *A. thaliana*, PP-IX has been observed to bind to residue His81 of *A. thaliana* TSPO homologue(Vanhee et al. 2011). Similarly, site-directed mutagenesis experiments across various bacteria have found that PP-IX binds to residue Trp51 and Trp138 (of *B. cereus* TSPO) (Yeliseev and Kaplan 2000; Guo et al. 2015; Ginter et al. 2013). Since some studies have suggested that PP-IX is degraded catalytically by TSPO, Trp138 residue (of *B. cereus* TSPO) was found to act as a catalytic site (Ginter et al. 2013). PP-IX is the only endogenous ligand, for which information on interaction with TSPO is available(Li, Liu, Zheng, et al. 2015a). The binding cavity of PP-IX is common to the cavity of drug ligand PK11195, though PP-IX binds to TSPO at micromolar affinity and PK11195 binds to TSPO at nanomolar affinity(Jaremko, Jaremko, Giller, et al. 2014). Indeed PK11195 has been found to inhibit binding of PP-IX to TSPO(Wendler et al. 2003; Guo et al. 2015). The residues Trp50 and Tyr54, were observed to form strongest contacts with PP-IX in *R. sphaeroides* TSPO, consistent with the experimental studies on PP-IX binding with TSPO(Li, Liu, Zheng, et al. 2015a; Guo et al. 2015).

iii. Diazepam Binding Inhibitor (DBI): DBI is a polypeptide of about 86 residues, earlier characterized in nervous system (Fig. 1.16). It has been observed to have micromolar affinities for TSPO(Besman et al. 1989; Rupprecht et al. 2010). It was later found that DBI is homologous to the Acyl Coenzyme A Binding Protein-3 (ACBP-3), which has been observed to interact with TSPO(Knudsen 1991; Rupprecht et al. 2010). Apart from TSPO, ACBP-3 has also been found to be part of a cholesterol-import complex, suggesting a role of DBI with cholesterol import(Liu et al. 2006). DBI folds into 4 alpha helices, with a binding pocket for Acyl-CoA (van Aalten et al. 2001).

However, the exact mechanism of interaction with TSPO, from physiological as well as from the structural level remains unclear.

1.4.2. Exogenous ligands of TSPO:

i. PK11195: PK11195 (IUPAC name: N-butan-2-yl-1-(2-chlorophenyl)-N-methylisoquinoline-3-carboxamide) is perhaps among the most extensively studied ligands of TSPO. It binds to TSPO with nanomolar affinities, with a distinctive binding pocket(M. Jaremko et al. 2015; Jaremko, Jaremko, Giller, et al. 2014). Apart from high binding affinity, it has been observed to bind with high specificity. Due to this, the radiolabeled variants of PK11195 are extensively used as a biomarker for imaging, as well as diagnosis of various diseases, including several neurodegenerative, psychological diseases and cancers(Su et al. 2015; Bhoola et al. 2018; Rupprecht et al. 2010). It has also been proposed as therapeutic agent for treatment of these diseases(Santidrián et al. 2007).

PK11195 is an isoquinoline derivative, composed of a twin phenyl isoquinoline scaffold, and aliphatic and aromatic substituent that branch from the isoquinoline scaffold. The aromatic group has a chloride ion, giving a negative charge to the molecule, and the aliphatic side chain substituent consists of an amide peptide dihedral, which is capable of torsional rotation. The ligand has two pairs of isomeric forms, arising from dihedral rotations: The first pair of isomer arises from rotation of dihedral associated with the amide scaffold, namely E-form and Z-form (Fig. 1.16). Apart from the amide dihedral, the dihedral corresponding to chlorophenyl group is also capable of torsional rotation, with two isomeric forms(Lee et al. 2012). While PK11195 is observed to bind to TSPO in both the isoforms *in vitro* and *in vivo*, it binds to TSPO preferably in E-form when reconstituted in liposomes(Jaremko, Jaremko, Giller, et al. 2014). The structural properties of the ligand, including the transition between the two isoforms has been characterized by means of NMR experiments, as well as by quantum mechanics (QM) simulations(Lee et al. 2012), which observed the presence of an energy barrier for the rotation between both pairs of isomeric forms. The barrier was higher for rotation of the peptide amide dihedral, compared to chlorophenyl dihedral(Lee et al. 2012). Though the cholesterol and PK11195 binding sites are observed to be spatially separated(Jaremko, Jaremko, Giller, et al. 2014), PK11195 binding has been found to stimulate steroidogenesis(Mukhin et al. 1989), suggesting a role of PK11195 in modulation of cholesterol binding. PK11195 has also been observed to modulate several physiological processes(Santidrián et al. 2007);(Sutter et al. 2003; Sutter et al. 2004; Sutter et al. 2002; Tanimoto et al. 1999; Maaser et al. 2001; Santidrián et al. 2007), though contradictory evidences also exist(Šileikytė et al. 2014; Tu et al. 2015). Several studies point towards the role of PK11195 in stabilizing the secondary structure and tertiary fold of TSPO(Murail et al. 2008; Ł. Jaremko et al. 2015; Jaremko, Jaremko, Giller, et al. 2014).

ii. Ro5-4864: 4'-chlorodiazepam or Ro5-4864 is another ligand of TSPO, found to bind to TSPO at high (micromolar) affinities. The structure consists of 5-carbon, 2-nitrogen seven-member ring, attached to two chlorophenyl rings, giving negative charge to the molecule (Fig. 1.16). This charge may stabilize the binding with extracellular loop of TSPO, consisting of positively charged residues, such as arginine and lysine. Like PK11195, Ro5-4864 is also observed to modulate mitochondrial processes involving TSPO, primarily steroidogenesis and apoptosis(Leonelli et al. 2005; Mills et al. 2008). However, Ro5-4864 does not share a similar binding cavity with that of PK11195. Instead, it is observed that VDAC is also required for Ro5-4864 binding to TSPO(Garnier et. al., 1994). However, this is also put in question, since TSPO is also found to bind to Ro-5864 in absence of VDAC as well (Joseph-Liazun et al. 1997). This binding site has been found to involve the residues of extracellular loop between TM1 and TM2(Farges et al. 1994).

iii. TRO19622, TRO40303: Both the ligands are structural analogues of cholesterol, recently found to bind to TSPO at nanomolar affinities (Fig. 1.16)(Bordet et al. 2007; Schaller et al. 2010). Like cholesterol, TRO19622 (IUPAC name: cholest-4-en-3-one oxime) has 3 cyclohexane and cyclopentane backbone scaffold, and an aliphatic substituent branching from the pentane ring. TRO40303 (IUPAC name: 3,5-Seco-4-nor-cholestan-5-one oxime-3-ol) has 2 cyclohexane rings and a similar aliphatic group branching from cyclopentane. The major difference of both these ligands from cholesterol is that they have an oxime functional group branching from the cyclohexane ring (Fig. 1.16).

TRO19622 was identified as a potential drug candidate for treatment of amyotrophic lateral sclerosis, a neurological disorder involving death of motor neurons. While TRO19622 prevented neuronal death and promoted cell regeneration *in vitro* as well as *in vivo* resulting in delay of onset of the disease(Bordet et al. 2007), TRO40303 was found to inhibit cell death during ischemia reperfusion injury ((Bordet et al. 2007; Schaller et al. 2010). Though TRO19622 was not found to displace radiolabelled PK11195 binding, it displaced radiolabeled cholesterol as well as pregnolenone binding significantly. Both the ligands were observed to interact with the CRAC motif, which is unsurprising from the fact that it is a cholesterol analogue. Interestingly, while both the ligands have similar structural properties as well as similar physiological effects, the difference between the two molecules lies with the binding to VDAC(Bordet et al. 2007; Schaller et al. 2010). While TRO19622 was found to displace steroid binding to VDAC, no such effect was observed with TRO40303 ((Bordet et al. 2007; Schaller et al. 2010)). The difference of these ligand from previously discussed ligands is that while PK11195 and Ro5-4864 have been observed to largely have pro-apoptotic effects, TRO19622 and TRO40303 have anti-apoptotic effects(Bordet et al. 2007; Schaller et al. 2010).

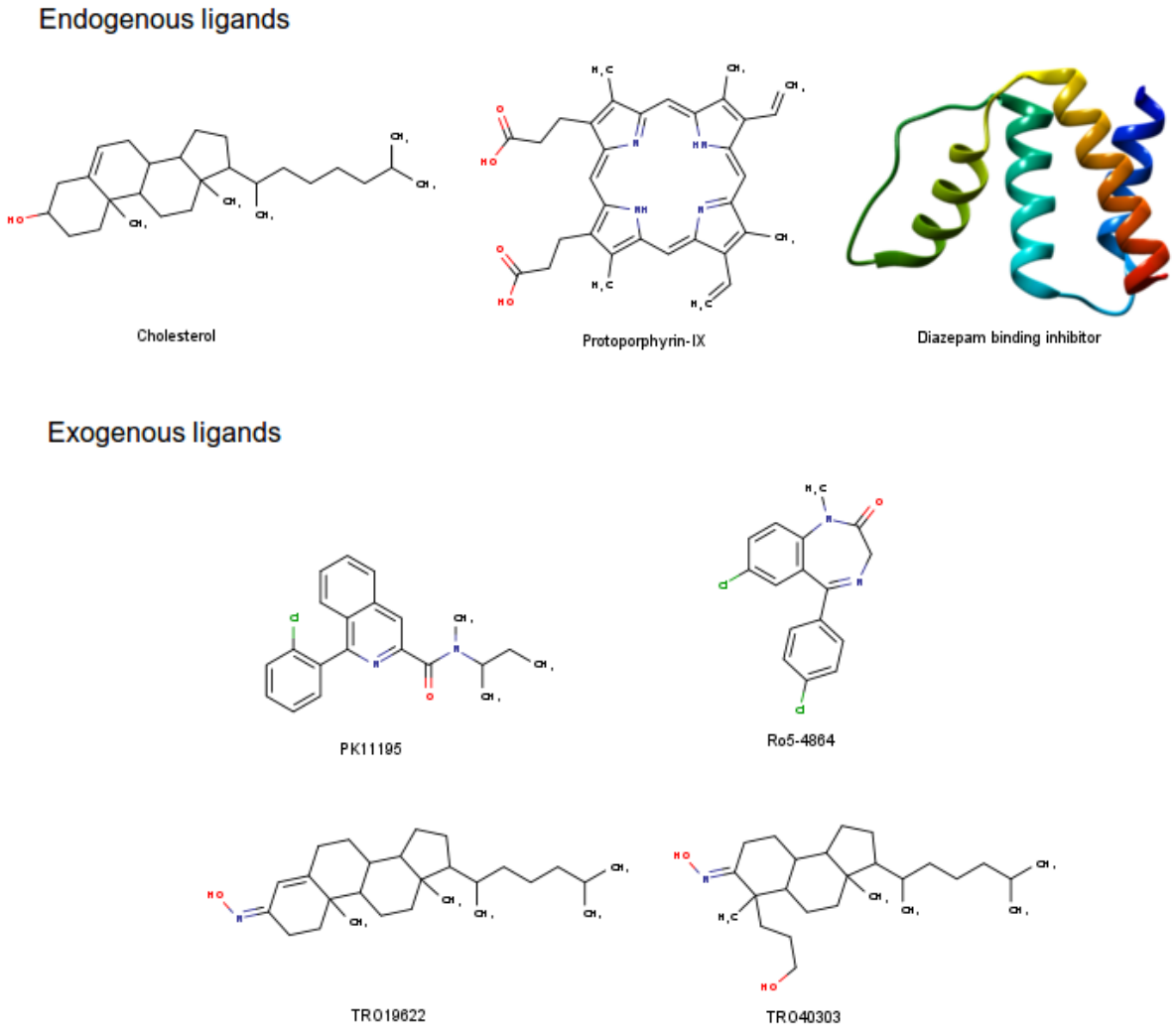


Figure 1.16. Exogenous and Endogenous ligands of TSPO. Ligands were drawn using Marvin Chemdraw program.

1.4.3. Interacting protein partners of TSPO:

i. Voltage Dependent Anion Channel: Among various interaction partners, interaction of TSPO with VDAC has been most extensively studied. VDAC is a 30 kDa channel, which has been demonstrated to transport variety of negatively charged molecules, ranging from ions to nucleic acids. In rat liver mitochondria, TSPO was found to exist as a part of three different subunits, of 18, 30 & 32 kDa molecular weight, which could bind radiolabeled [^{3}H]PK14105, [^{3}H]flunitrazepam and [^{3}H]AHN-086 ligands. The complexes corresponding to 30 & 32kDa molecular weight were found to consist of TSPO with VDAC and ANT respectively(McEnery et al. 1992). Later studies revealed the colocalization of TSPO and VDAC at the contact sites of OMM and IMM(Veenman et al. 2007; Culty et al. 1999). Furthermore, concentrations of TSPO as well as VDAC were modulated by modulation of hormone concentrations(Papadopoulos et al. 2007; Veenman et al. 2007). While isoquinoline-based TSPO ligands such as PK11195 binds exclusively to TSPO, benzodiazepine-based ligands are observed to interact with both TSPO and VDAC, due to the fact that presence of VDAC significantly

increases the binding of benzodiazepines. TSPO interaction with VDAC has been studied in context of various physiological processes, including cholesterol transport into mitochondria, regulation of mitochondrial permeability transition and apoptosis, regulation of mitochondrial energy metabolism, response to injury and inflammation to name a few(Veenman et al. 2007; Rone et al. 2012). While there have been extensive experimental studies characterizing interaction of TSPO with VDAC, no structure of a TSPO-VDAC complex, or knowledge of important interacting residues are available. However, attempts have been made to study TSPO-VDAC interactions. In MA-10 Leydig cells, it was observed from atomic-force microscopy of gold-stained TSPO that each VDAC is surrounded by 4-6 TSPO monomers(Papadopoulos et al. 1994; Papadopoulos, Amri, Boujrad, et al. 1997). Though the study offered some insights into a possible stoichiometric arrangement of TSPO-VDAC complex, the exact interface between the two proteins remained unclear. However, since VDAC has been demonstrated to bind steroids(Budelier et al. 2017), a possibility of TSPO-VDAC interface involving CRAC motif of TSPO exists. Indeed, from the studies characterizing the ligand TRO19622, a structural analogue of cholesterol, it was observed that treatment of TRO19622 resulted in the displacement of radiolabelled steroid binding to VDAC(Bordet et al. 2007). However, the same was not observed in the study characterizing another structural analogue of cholesterol, TRO40303(Schaller et al. 2010). While TRO19622 was studied in the neurons(Bordet et al. 2007), TRO40303 was studied from the cells of cardiac tissue, though the context of both the ligand binding was the same: inflammation(Schaller et al. 2010). Thus, it may be interpreted from both the studies that depending on the tissue or even the metabolic state, TSPO-VDAC interface may differ, adding complexity to this problem. The structure of VDAC consists of a β-barrel architecture, consisting of 18 β strands, and an α helix located at N-terminal (Fig. 1.17)(Ujwal et al. 2008; Hiller et al. 2008; Bayrhuber et al. 2008). Molecular dynamics simulations studies of VDAC revealed an important role of this helix in regulating the flow of conductance of anions through the membrane(Krammer et al. 2011; Krammer et al. 2015; Krammer et al. 2013).

ii. Steroidogenic Acute Regulatory Protein: Apart from TSPO, StAR is another candidate cholesterol transporter protein. Like TSPO, StAR was also first studied in MA-10 leydig tumor cells, where StAR expression was found to increase with stimulation by trophic hormones, which resulted in enhanced steroidogenesis. The role of StAR in cholesterol transport was proposed based on observations that cDNA of this protein is non-existent in individuals with congenital lipid adrenal hyperplasia, a genetic autosomal recessive disorder where synthesis of all adrenal and gonadal steroid hormones is impaired(Lin et al. 1995). Due to the absence of strong evidence supporting the role of TSPO in cholesterol transport into mitochondria, this protein was suggested to be an alternative candidate cholesterol transporter. As a part of this hypothesis, TSPO plays a regulatory role in cholesterol transport, whereas the primary transport of cholesterol is carried out by StAR. In support of this

hypothesis, StAR was observed to be a part of OMM-based cholesterol import complex along with TSPO(Rone et al. 2012). StAR has an alpha/beta fold, with a hydrophobic cavity that can accommodate cholesterol molecule (Fig. 1.17)(Tsujishita and Hurley 2000). One of the proposed transport mechanism involves solubilization of cholesterol and transport into the intermembrane space, akin to many other lipid transporters.

iii. Adenine Nucleotide Translocase (ANT): Unlike other interacting partners of TSPO, ANT is not localized at OMM, it is localized at Inner Mitochondrial Membrane (IMM). Interaction of ANT and TSPO occurs at the contact sites of outer and inner mitochondrial membranes (Fig. 1.17). There is a growing body of supporting information that points towards a role of ANT as a part of multiprotein complex involved in cholesterol transport, apoptosis, response to inflammation to name a few(Veenman et al. 2008; Veenman et al. 2007). However, some studies also cast a doubt on whether ANT is important for such processes (Kokoszka et al. 2004). ANT is a 6 TM helical protein, with 3 extracellular helices(Pebay-Peyroula et al. 2003). Though the interaction of ANT with the mitochondrial multiprotein complex has been well-studied from biochemical experiments, it is not clear if the interaction occurs via TSPO, or if it is mediated by other components of mitochondrial multiprotein complex such as VDAC, Bcl-2, Hexokinase etc(Crompton et al. 1998).

iv. PBR Associated Protein-I (PRAX-I): PRAX-I is a 220 kDa, 1,857 residue cytosolic protein was first observed to interact with TSPO from the two-hybrid assay experiments(Galiègue et al. 1999). The protein distributed largely in the brain tissues, particularly in the hippocampus. The protein was observed to be composed of multiple domains. including proline-rich domain, leucine zipper and Src domains. While the protein was localized in the interface of mitochondria and cytoplasm of rat brain, the function was not well understood. Another study revealed that PRAX-1 is highly expressed in the neurons of central nervous system, particularly in the hippocampus(Galiègue et al. 1999). Furthermore, it was observed that treatment of the peptide SR48968, which is antagonist to NK2 receptor significantly increased PRAX-1 expression, suggesting a role of PRAX-1 in response to stimulants and antidepressants(Chardenot et al. 2002). While PRAX-1 is observed to exist in the interface of cytoplasm and mitochondria, it may interact with TSPO from the side corresponding to C-terminal part of TSPO, and extracellular loop between TM1 and TM2 as the possible interface. However, no atomic structure of this protein is available.

v. PBR-Associated Protein(PAP-7)/Acyl-CoA Binding Domain-3(ACBD3)**:** Though TSPO has been repeatedly observed to have a major role in steroidogenesis, it was unclear exactly how addition of small amounts of stimulatory hormones such as human chorionic gonadotropin, adrenocorticotropic hormone, etc. resulted in large-scale induction of cholesterol transport and steroidogenesis. The answer came from characterization of another TSPO interacting partner, PAP-7. PAP-7 was found from yeast two-cell hybrid experiments, using C-terminal portion of TSPO as bait(Li, Degenhardt, et

al. 2001). It is 52-kDa cytosolic protein, which is expressed in gonadal, adrenal and hippocampal tissues, similar to the expression profile of TSPO. Interestingly, PAP-7 was found to interact with another signalling protein, Protein Kinase-A (PKA), from similar experiments on human lymphocytes, using the latter as bait. Based on initial studies, it was proposed that PAP-7 serves to induce steroidogenesis, by recruiting PKA into mitochondria, which phosphorylates OMM proteins, initiating steroidogenesis(Li, Degenhardt, et al. 2001). This was also supported from the observations that overexpression of PAP-7 resulted in increase in steroid formation (Fig. 1.17)(Hauet et al. 2002; Li, Degenhardt, et al. 2001);(Hauet et al. 2002). A structure-based nomenclature was later proposed and adopted, renaming the protein to Acyl-CoA Binding Domain-3 (ACBD3). Apart from steroidogenesis, ACBD3 was observed to play a role in various other processes, including apoptosis, iron homeostasis, neurogenesis, embryonic development and cancer(Fan et al. 2010), many of which involve TSPO as well. ACBD3 is a multi-domain protein, consisting of acyl-CoA binding domain with high concentration of charged residues, a glutamine rich domain and a GOLD domain. Since ACBD3 is a multi-domain protein, the structures for ACB domain, Q-rich domain and GOLD domains were solved separately(McPhail et al. 2017; Klima et al. 2017). Studies on the structure of this protein revealed that ACBD3 has a stable interaction with Phosphatidylinositol 4-kinase beta (PI4KB) and ACBD3 can recruit PI4KB to membranes both *in vitro* and *in vivo(Klima et al. 2016)*.

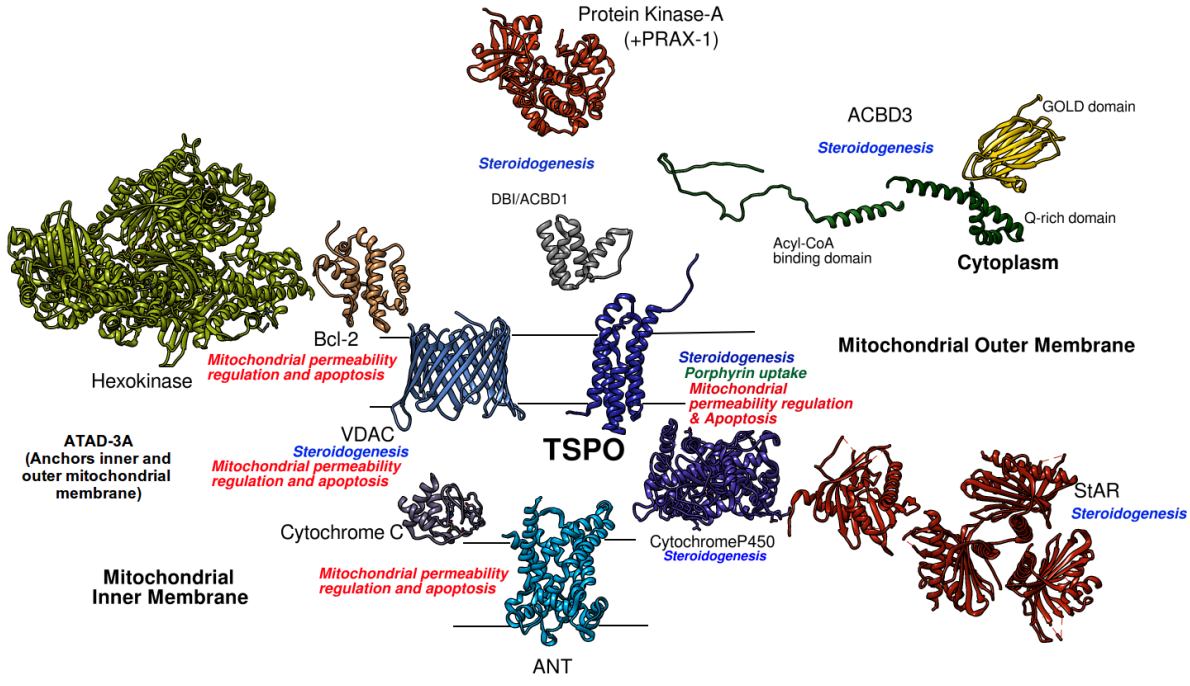


Figure 1.17. Interacting partners of TSPO, with respect to their subcellular localizations and involved physiological processes. Structures were visualized using UCSF Chimera program(Pettersen et al. 2004). Apart from the interacting partners mentioned in preceding paragraphs, other partners include Hexokinase & Bcl-2 (involved in apoptosis as part of MPTP; Veenman et al. 2007), Cytochrome C (which is shuttled out of mitochondria by MPTP during apoptosis, Veenman et al. 2007), ATAD3A (connects inner and outer mitochondrial membranes, involved in steroidogenesis; Rone et al. 2012) and Cytochrome P450 (involved in conversion of cholesterol to pregnenolone; Rone et al. 2012).

Box-1.1: Take-home message:
***Biology of TSPO*:**

- TSPO is an 18 kDa transmembrane protein primarily found in mitochondria, but also reported in golgi apparatus, endoplasmic reticulum, plasma membrane etc.
- It is involved in various physiological processes, such as cholesterol transport, tetrapyrrole metabolism and erythropoeisis, apoptosis, response to injury, regulation of mitochondrial metabolism, regulation of bacterial photosynthesis, response to stress in plants to name a few (Fig. 1.4-1.10).
- These processes are performed in association with various proteins, including Voltage Dependent Anion Channel (VDAC), Adenine Nucleotide Transporter (ANT), Steroidogenic Acute Regulatory Protein (StAR), PBR-associated protein-I (PRAX-I), Acyl-CoA Binding Domain-3 (ACBD3) etc. (Fig. 1.10)
- TSPO binds to various ligand molecules, both endogenous (cholesterol, PP-IX, ACBD1) and exogenous (PK11195, Ro5-4864, Diazepam, TRO19622, TRO40303, FGIN-1-27 etc. Fig. 1.16).
- Many of these ligands bind at high affinities, ranging from micromolar to nanomolar concentrations. Many exogenous ligands bind to TSPO resulting in modulation of physiological processes, such as stimulation of steroidogenesis, collapse of mitochondrial transmembrane potential leading to apoptosis, cytoprotective effects during cell injury etc.
- As a result, presence of TSPO has been used as an effective biomarker for diagnosis of diseases, and TSPO is also proposed to be an attractive therapeutic target of various diseases it is involved in, including malaria, neurodegenerative and psychological disorders, cancer, ischemia reperfusion injury etc.
- A paralog of TSPO, TSPO2 was recently characterized, with role in cholesterol redistribution during erythropoeisis. The blood-specific distrubution of TSPO2 suggests specialization in function. TSPO2 has also been found to be active in the *Plasmodium*-infected red blood cells.

Structural Biology of TSPO:

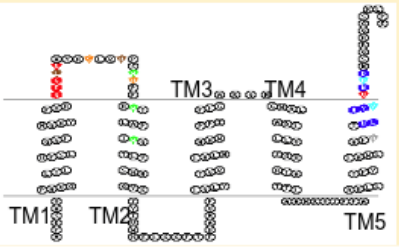
- Structure of TSPO was a mystery until recently. Early experimental and theoretical studies suggested TSPO as transmembrane protein composed of a fold encompassing 5 TM alpha helices (Fig. 1.11), and binding of ligands such as PK11195 further stablized the tertiary structure.
- Atomic-level information on the structure was available recently with solving of NMR structure of mouse TSPO in presence of PK11195 ligand. Apart from the 5-TM architecture, a well-defined PK11195 hydrophobic binding cavity was observed (Fig. 1.12).
- While a well-defined structure was observed in presence of ligand, attempts to study the TSPO structure in absence of ligand led to the observation that TSPO in absence of ligand is highly dynamic, evident from the lack of NMR signals for many residues of TM regions, as well as reduction of helicity of the TM residues.
- X-ray crystal structures of bacterial TSPOs were subsequently available. While similar 5-TM fold was observed, bacterial TSPOs were found in oligomeric stoichiometry, both in dimer, as well as trimer forms (Fig. 1.14). Comparison between eukaryotic and bacterial TSPOs revealed significant differences in conformations (Fig. 1.14).
- Recent NMR studies on mouse TSPO further revealed a role of cholesterol in oligomerization of TSPO, where cholesterol interaction with TSPO was predominant in monomer condition.
- While extensive attempts to understand TSPO have been made, the recent studies on its structure have raised more quesions:

1. ***What are characteristic dynamics of TSPO? How does ligand binding affect the dynamics?***
2. ***Is there a preferred stable TSPO dimer? What role does cholesterol play in the oligomerization dynamics?***
3. ***Do bacterial and mammalian TSPO have similar dynamics? Does the difference in the dynamics stem from the sequence, or is the difference inherently structural?***
4. ***How are differences between TSPO1 and TSPO2 manifested at the level of sequence and structure? How these differences can be correlated to difference in functions?***

Box-1.1: Summary on the state of the art on TSPO from physiological and structural perspective.

Consequences of mutagenesis of TSPO residues:

While TSPO is involved in variety of physiological processes, it still is unclear as to which residues are important for respective roles. However, numerous site-directed mutagenesis studies have shown importance of certain residues, which are distributed throughout the protein. The table below is a summary of residues which have been subjected to mutagenesis studies, and resulting biochemical changes from the mutation.



Blue: CRAC motif residues
Red: Decreased affinity to Ro5-4864
Orange: Decreased affinity to PK11195
Brown: Decreased affinity to Ro5-4864+PK11195
Green: Increase in accumulation of bacteriochlorophyll
Cyan: Reduction in cholesterol uptake & steroidogenesis

Residue No.	Type of mutation	Location of mutation	Equivalent residue in mouse TSPO	Resulting phenotype/physiological effect	Reference	Other remarks
--	C15S	TM1	--	Significant increase in accumulation of bacteriochlorophyll in *R. sphaeroides* (4.93 vs. 11.32 nmol/mg protein)	Yeliseev & Kaplan 2001	
--	W38C	Loop btn. TM1&2	--	Significant decrease in accumulation of bacteriochlorophyll in *R. sphaeroides* (4.93 vs. 8.28 nmol/mg protein), implying increased porphyrin transport by TSPO	Yeliseev & Kaplan 2001	The same mutation resulted in TM1-mediated dimerization by means of 2S-bridges between mutated cysteines
26	V26T	TM1	V26	Protein expressed in very low levels	Farges et al. 1995	
31	L31F	Loop btn. TM1&2*	L31	Decreased binding affinity to Ro5-4864+PK11195	Farges et al. 1995	
32	R32G	Loop btn. TM1&2*	R32	Decreased binding affinity to Ro5-4864	Farges et al. 1995	
36	G36S	Loop btn. TM1&2*	G36	Decreased binding affinity to PK11195	Farges et al. 1995	
39	K39G	Loop btn. TM1&2*	K39	Decreased binding affinity to Ro5-4864+PK11195	Farges et al. 1995	
43	W39F	Loop btn. TM1&2	W43	- Significant increase in accumulation of bacteriochlorophyll in *R. sphaeroides* (4.93 vs. 8.28 nmol/mg protein) - Increased transcription of photosynthesis genes	Yeliseev & Kaplan 2001	Possible PPIX binding site, due to the fact that mutation resulted in accumulation of bacteriochlorophyll, arising from failure to transport porphyrins
43	H43A	Loop btn. TM1&2	H43	Decreased binding affinity to PK11195	Farges et al. 1995	
48	W44F	TM2	W48	- Significant increase in accumulation of bacteriochlorophyll in *R. sphaeroides* (4.93 vs. 9.99 nmol/mg protein) - Increased transcription of photosynthesis genes	Yeliseev & Kaplan 2001	Possible PPIX binding site, due to the fact that mutation resulted in accumulation of bacteriochlorophyll, arising from failure to transport porphyrins
53	W50F	TM2	W53	- Significant increase in accumulation of bacteriochlorophyll in *R. sphaeroides* (4.93 vs. 9.17 nmol/mg protein) - Increased transcription of photosynthesis genes	Yeliseev & Kaplan 2001	
147	A147T	TM5	A147	Reduced cholesterol uptake & steroid synthesis, mutation associated several neurodegenerative diseases, change in relative orientation of TM2 and TM5	Jaremko et al. 2014, Li et al. 2015, Costa et al. 2009	Structure of this mutant available (PDB: 2N02, 4UC1)
153	Y153S	TM5	Y153	Significantly reduced uptake of cholesterol	Li & Papadopoulos 1998	Part of CRAC motif
154	V154M	TM5	V154	Decreased binding affinity to Ro5-4864	Farges et al. 1995	
156	R156L	TM5	R156	Significantly reduced uptake of cholesterol	Li & Papadopoulos 1998	Part of CRAC motif
28,29,30	G28A, E29A, G30A	TM1	G28, E29, G30	Decreased binding affinity to Ro5-4864	Farges et al. 1995	

Box-1.2: Table depicting consequences of mutagenesis of various TSPO residues.

Objectives of this thesis:

Primary questions asked/ objectives of this study:

1) ***How does PK11195 ligand binding impact the dynamics of TSPO?***

   Objectives:

   a. To compare dynamics of TSPO in presence and absence of PK11195 ligand, and study the effect of PK11195 binding on dynamics of TSPO, by means of all atom molecular dynamics.
   b. To investigate correlated dynamics between TSPO residues in *apo*- and *holo*-forms, and trace possible allosteric dynamics.

2) ***Can two TSPO monomers aggregate together as a dimer?***

   Objectives:

   a. To study the dynamics of TSPO dimerization, characterize interfaces and evaluate their interactions, from structural, biophysical and evolutionary perspectives, by coarse-grained molecular dynamics simulations of a pair of separated TSPO monomers.
   b. To study relationship between oligomerization dynamics and cholesterol interaction with TSPO

3) ***Do structural differences between eukaryotic and bacterial TSPOs translate to differences in dynamics?***

   Objective:

   a. To understand similarities/ differences in structural dynamics of TSPO from different species by means of coarse-grained molecular dynamics simulations of TSPO monomers.

4) ***What are the differences between TSPO1 and 2 on the basis of sequence, structure, dynamics and phylogeny?***

   Objectives:

   a. To trace TSPO2 evolution with respect to broader TSPO evolution in animals by construction of phylogenetic trees.
   b. To understand similarities/differences between TSPO1 and TSPO2 in terms

      of structures as well as dynamics.

Chapter-2: Materials and Methods

Table of contents

2.1. A brief introduction to molecular dynamics simulations 54
2.1.1. Basic principles 55
2.1.2. Force fields 56
2.1.3. Methods in molecular dynamics 59
2.1.4. Coarse-grained force fields 62
2.2. Analysis methods of molecular dynamics simulations 64
2.2.1. Selection of residues for analysis 64
2.2.2. Principal component analysis 66
2.2.3. Protein structural networks 67
2.2.4. Protein-protein docking 68
2.3. Methods employed for comparison of TSPO paralogs 71
2.3.1. Multiple sequence alignment 71
2.3.2. Phylogenetic tree construction 72
2.3.3. Normal mode analysis 73

This chapter describes the methods used and the underlying rationale for achieving various objectives listed above. The specific protocol and parameters are provided in the respective chapters.

2.1. A brief introduction to molecular dynamics simulations:

With the advent of experimental techniques such as NMR spectroscopy, small angle X-ray scattering etc., it has been conclusively proven that proteins are dynamic entities (Linderstromlang 1955; Hvidt and Linderstrøm-Lang 1954; Benson and Linderstrom-Lang 1959; Karplus and McCammon 2002; Englander et al. 1997). Though these methods are quite effective in characterizing protein dynamics, they are often insufficient to gain satisfactory insight into the nature of protein dynamics and the relation to its function. Molecular dynamics simulations have been proven repeatedly to act as a bridge between the hypotheses pertaining to protein structures, dynamics, functions, with the observations and insights gained from experimental studies (Karplus and McCammon 2002; Karplus and Kuriyan 2005).

Molecular Dynamics (MD) simulations involve imposition of forces/velocities to molecules, with calculation and integration of energies resulting from the former, for a given time step. This is performed on a protein in an environment similar to its actual physiological environment. In case of water-soluble proteins, this involves solvation of protein in water, in addition to ions. In case of membrane proteins, the degree of complexity increases, which involves the protein embedded to a lipid bilayer (for transmembrane proteins) or interacting with it, in addition to solvent and ions. In the seminal days of MD simulations (1970s-80s), they were primarily employed for the study of physical and chemical properties of different molecules. With progress of time, MD simulation methods and computational power, it was realized that this technique can also be applied to study the structure and dynamics of biological macromolecules, such as proteins, nucleic acids, etc.

However, one of the earliest MD simulations on proteins were at the scale of picoseconds (McCammon et al. 1977). At the scale of femtoseconds, which usually corresponds to a single time step of a MD simulation, inter-atomic level motions such as bond vibrations occur. At the level of picoseconds, residue-level motions, such as rotation and elastic motions of certain side chain bonds may be observed. While at the level of nanoseconds, motions involving secondary structure elements, such as helices, strands and loops are observed, ultimately, structural changes involving tertiary or quaternary structure take place at the level of several microseconds or milliseconds. Thus, the computational power at the earliest days of studying proteins using molecular dynamics was not adequate to sample global conformational changes, though these pioneering studies paved way for tremendous development of MD simulation methodologies that enable such sampling today.

2.1.1. Basic principles:

In a molecular dynamics simulation, velocities are applied on a system of particles, and forces are calculated as a negative gradient of the potential energy:

$$m_i \frac{d^2 r_i}{dt^2} = f_i \qquad f_i = -\frac{\delta}{\delta r_i} U \qquad (1)$$

Where m_i is the mass of the particles, r_i are the positions of the atoms, f_i are the forces, U is the potential energy associated with displacement.

Potential energy is calculated as a cumulative sum of energies arising from bonded and non-bonded interactions.

i. Bonded interactions: In any molecule, there are three types of bonded interactions: bond-stretching/contraction (involving flux in *bond lengths*), bond-bending (involving change in *bond angles* between three atoms) and *dihedral angles* (involving rotation of torsions formed by four atoms). In certain cases involving planar bonds, particularly for aromatic rings, dihedral angles are not sufficient to accurately describe the forces. In such cases, an additional term is used, called *improper dihedrals*.

$$U_{bonded} = \Sigma \tfrac{1}{2} k_b (r_0 - r_t)^2 + \Sigma \tfrac{1}{2} k_a (\theta_0 - \theta_t)^2 + \Sigma \tfrac{U_n}{2} [1 + cos(n\phi - \delta)] + \Sigma U_{imp} \qquad (2)$$

Bond length **Bond angle** **Dihedral angle** **Improper dihedral**

Where the term $(r_0 - r_t)$ represents the displacement of particles between equilibrium positions (derived from the force field) and positions at the given time *t*, the term $(\theta_0 - \theta_t)$ represents the change in bond angle between three particles between initial time and time *t*. ϕ represents torsional angle and δ the initial phase between the four atoms, n a positive integer between 0 and 2π, and U_n is the height of the potential barrier. The improper dihedral term U_{imp} is calculated as:

$$U_{imp} = \sum \frac{k_{imp}}{2} [1 + cos(2\omega - \pi)] \qquad (3)$$

ω represents the improper angle that deviates from planarity. k_a, k_b and k_{imp} are strength parameters for bond stretching, bond bending and torsional rotation associated with improper dihedrals, which is derived from the force fields used in the calculation.

i. Non-bonded interactions: Non-bonded interactions are calculated as a sum of Van der Waals and electrostatic interactions, which are represented by Lennard Jones (LJ) potential and Coulombic potential respectively:

$$U_{non-bonded} = 4\varepsilon[(\frac{\sigma}{r})^{12} - (\frac{\sigma}{r})^{6}] + \frac{q_1 q_2}{4\pi\varepsilon_0 r} \quad (4)$$

LJ potential Coulombic potential

Where r is the distance between the two given particles, σ is the distance at which potential between the two given particles is zero, ε is the depth of the potential energy minima between the two given particles. q_1 and q_2 are the charges of the particles and ε_0 is the permittivity of vacuum media.

Thus, the total energy of the system is described as a sum of both the bonded and non-bonded terms (Fig. 2.1):

$$U = U_{bonded} + U_{non-bonded} \quad (5)$$

2.1.2. Force Fields:

Force field is a model describing the interactions between atoms and molecules in MD simulation. The parameters employed in force field are usually derived either from quantum mechanics (QM) calculations, or by fitting to data from experimental studies, such as NMR, infrared spectroscopy, X-ray scattering etc. Among the earliest force fields were developed in 1980's from *ab initio* studies of condensed matter (Gavezzotti 2007; González 2011). The calculation of total energy (U) of the system from application of forces is performed based on the parameters dictated by the force field. While the force field consists of bonded and non-bonded parameters, as described in equations (2)-(5), the bonded parameters between the atoms are defined by means of harmonic potentials. The exception here is the dihedral potential, which is not defined by harmonic potential. Most of the force fields used for biomolecular simulations employ such potentials. However, a major limitation of this representation is that while this model described 2-body interactions, it cannot represent formation or breaking of bonds. In other words, while bonds may be stretched, bent or rotated, bonds cannot be broken or formed in empirical force fields that operate based on Newtonian dynamics. To study such chemical reactions, QM simulations, or *ab initio* molecular dynamics are employed. Sometimes, a combination of QM and molecular dynamics are also employed to study such phenomena. While using empirical models is not necessarily a true representation of the forces between the atoms and molecules, this model has been observed to be in agreement with data from various experimental studies (Karplus and McCammon 2002). The earliest force fields were primarily developed to study different properties of small organic molecules. Since then there has been an explosion of variety of force fields, with large number of modifications and improvements. Among the most popular force fields include CHARMM, OPLS and AMBER.

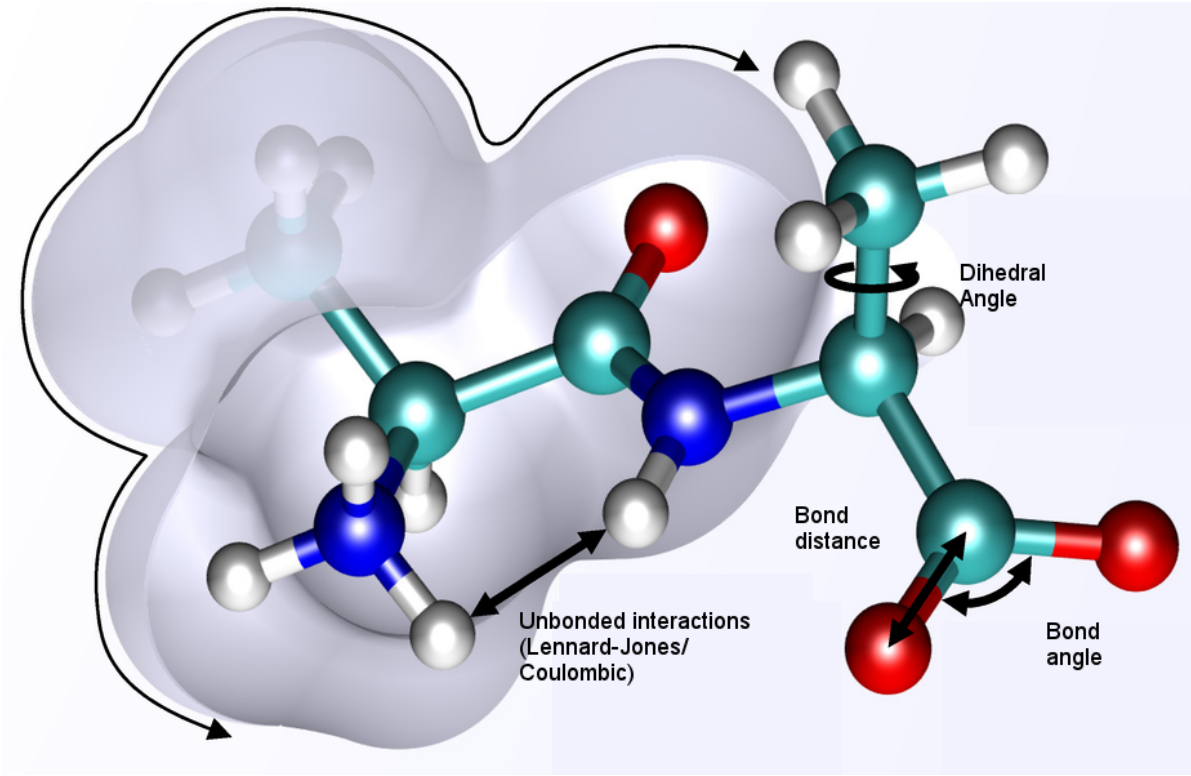


Fig. 2.1. Illustration of different interactions calculated by a force field (Source: Wikimedia commons).

i. CHARMM: CHARMM (Chemistry at Harvard using Molecular Mechanics) is one of most widely used force fields for simulations of proteins. Its primary characteristic is that the parameters for non-bonded interactions were modeled to match the properties of molecules in gaseous phase. Subsequent variations were developed for various biomolecules, such as nucleic acids, lipids, etc (Pastor and Mackerell, 2011; Xu et al., 2016).

ii. OPLS: OPLS (Optimized Potentials for Liquid Simulations) was developed for organic liquids initially, with emphasis on reproducing thermodynamic properties of organic liquids. The force field was developed with strong emphasis on non-bonded potentials, though further variants for proteins included bonded parameters derived from AMBER force fields.

iii. AMBER: AMBER (Assisted Model Building with Energy Refinement) force field was developed as a part of AMBER molecular dynamics package (Weiner et al. 1984). While CHARMM and OPLS force fields in their earliest versions did not have hydrogen atoms, or they were defined explicitly only if they were polarized, AMBER was the first force field to have explicit definitions for both polar as well as non-polar hydrogens. Electrostatic interactions were derived based on fitting of partial charges to electrostatic potentials observed from QM calculations. VdW interactions on the other hand were derived from fitting of parameters to match with the interactions observed from amide crystal data, as well as from

liquid-state simulations (Hagler and Lifson 1974; Ponder and Case 2003). Bonded parameters were derived by matching the bond lengths and angles observed from crystal structures, and were also adopted to fit with those observed from different normal mode frequencies. Torsional parameters were derived at match the potential barriers attributed to torsional rotations observed from the QM calculations. Similar to the OPLS force field, the force constant is heavily derived from the non-bonded interactions.

Further developments of the force field involved imposition of limits on charges a heavy atom can assume, to address the problems pertaining to the charges assumed by heavy atoms. It was observed that heavy atoms of the same protein tend to assume different charges, in different environments. This was particular in case of atoms in buried environment. This imposed limitation was termed as Restrained Electrostatic Potential Fit (RESP), and was implemented in the AMBER94 force field, and subsequently refined and extended to AMBER99 force field (Bayly et al. 1993; Cornell et al. 1995; Wang et al. 2000). However, some problems still persisted, particularly at the level of secondary structures. It was observed that AMBER force fields, particularly AMBER94 and AMBER99 force fields tend to over-emphasise the stability of alpha helices (García and Sanbonmatsu 2002). This was even observed in structures that were predominantly β-strands and hairpins (Okur et al. 2003). This problem was rectified in the next development of AMBER force field, which was termed AMBER99sb. The dihedral terms of peptide backbone were tweaked to match the parameters observed from QM calculations of glycine and alanine tetrapeptides (Hornak et al. 2006). The conformations of tetrapeptides observed from QM calculations were in better agreement with the conformations observed in proteins from PDB. Indeed, this change resulted in significantly better agreement with the observations from NMR experiments (Showalter and Brüschweiler 2007). Further enhancements included improvement in side-chain torsion potentials, where χ_1 torsion angle parameters were adjusted to match those observed from QM simulations of amino acids (Lindorff-Larsen et al. 2010). This adjustment resulted in improved agreement with the side chain torsion distributions observed from NMR experiments (Lindorff-Larsen et al. 2010).

Among the most recent improvements involved adjustment of both backbone and side-chain dihedral parameters of proline and hydroxyproline residues, in order to fit with the data from NMR spin-relaxation experiments (Aliev et al. 2014). This improvement, termed as AMBER99SB-ILDNP was used to describe TSPO in this study, due to its success in accurate representation of proline dynamics (Aliev et al. 2014). Apart from the fact that TSPO is a tryptophan-rich protein, it also has large amount of proline residues distributed throughout different TM regions. It is well-observed fact that presence of prolines in α-helical TM proteins results in kinks in the TM helices, which influence dynamics of TM regions, and also

the function of protein in certain cases (von Heijne 1991). Thus, it is imperative to accurately represent the dynamics of prolines, for which AMBER99SB-ILDNP has been demonstrated to successfully perform, and this force field was selected for all-atom simulations of TSPO.

iv. Slipids: With the rapid development of force field parameters for proteins, it was realized that a similar force field is necessary for accurate description of other biomolecules as well, such as nucleic acids, lipids etc. This is particularly true for simulations of TM proteins such as TSPO. One of the popular force fields developed for lipids, which is used in conjunction with AMBER force fields is Slipid force field. While initially it was developed for saturated lipid molecules (Jämbeck and Lyubartsev 2012b), it was later extended to unsaturated lipids as well as poly-unsaturated molecules, such as DOPC (used in this study, (Jämbeck and Lyubartsev 2012a). The bonded parameters, and non-bonded hydrophobic parameters for the both headgroup and glycerol groups were derived based on Charmm36 force field (Klauda et al. 2010). For the hydrophobic tail, the parameters were derived from adjustment of parameters to match density and heats of vapourization observed from experimental studies. Similar approach was adopted for parameterization of unsaturated bonds, with parameterization performed on *cis*-5-Decene (Jämbeck and Lyubartsev 2012b). MD simulations of lipid bilayers with these parameters showed significant agreement with regard to bilayer-specific properties, such as Deuterium-order parameters, membrane thickness, area per lipid to name a few. Thus, this force field was chosen in this study to describe the DOPC lipid molecules in all-atom simulations.

2.1.3. Methods in molecular dynamics:

Below is a description of various methods and algorithms used as a part of preparatory steps as well as production simulations. The methods described below are implemented in GROMACS MD simulation program (version 4.6.7; Berendsen et al. 1995, Pall et al. 2015, van der Spoel et al. 2014).

i. Energy minimization by Steepest descent method: The starting structure of the protein-bilayer system may contain many geometrical clashes which may have arisen from overlap of atomic positions. Hence, it is imperative to eliminate such steric hindrances, which may distort the dynamics and result in unrealistic motions of particles in the system. To achieve this goal, the system is subjected to energy minimization, through which the system is brought to a local minimum. Steepest descent algorithm is a robust and among the most widely used energy minimization methods. The method involves definition of coordinate vector $\boldsymbol{r}$, which corresponds to the 3N coordinates of all the particles in the system. For the initial step, a maximum displacement h_0 is implemented, maximum force F_n and potential energy of the system are calculated. With every displacement, new positions r_{n+1} are calculated as:

$$r_{n+1} = r_n + \frac{F_n}{max(|F_n|)} h_n \qquad (6)$$

Where h_n is the maximum displacement, F_n represents the force of the system calculated as a negative gradient of potential energy, ($|F_n|$) is the largest force on the system as a negative gradient of potential energy. Forces and energies are calculated again for new positions for the next step. If $U_{n+1} < U_n$, the positions are accepted, and next step is implemented with $h_{n+1}=1.2h_n$. Here, U_n represents the potential energy at the first step, U_{n+1} represent the potential energy at the next step. If the opposite is true, the positions are rejected, and the step is performed again with a with new h_n value of $h_n = 0.2h_n$. The minimization continues once the specified number of steps (50,000 in this study) is reached, or if the F_n is lesser than the specified value (100 kJ/mol nm in this study). Steepest descent minimization method was employed as a preparatory step for both all-atom as well as coarse-grained simulations (chapters 3, 4 & 5).

ii. Integration of Newton's equations by Leap-frog integrator: In a MD simulation, Newton's equations of motions are calculated as a negative gradient of potential energy (Equation 1). Since this has to be performed over millions of steps, a need arises for an algorithm that integrates the equations of motions. Leap-frog integrator is among the most commonly used algorithm in MD simulations for this purpose. The algorithm uses positions *r* at time *t* & velocities *v* at time *t* - ½*Δt*. It updates positions and velocities using forces *F(t)*:

$$v(t + ½\Delta t) = v(t - ½\Delta t) + \frac{\Delta t}{m} F(t) \qquad (7)$$

$$r(t + \Delta t) = r(t) + \Delta t\, v(t + ½\Delta t) \qquad (8)$$

The updated positions and velocities are used to calculate forces and acceleration for the next step, which is repeated over the specified number of steps in a simulation.

iii. Application of position restraints for equilibration of system: Position restraints are applied on protein during the equilibration steps in all the performed simulations (Chapters 3,4 & 5). The purpose is to impose an energetic penalty on the heavy atoms of protein, so that their initial positions are conserved, while at the same time, the other particles in the system, including lipids and solvent equilibrate and stabilize. They are often used to avoid drastic rearrangements of critical parts of the protein, or restrain protein motions that are subjected to larger solvent forces, when membrane and solvent is not yet equilibrated. To impose restraints, the following term is applied:

$$V_{pr(r_i)} = \frac{1}{2} k_{pr} |r_i - R_i|^2 \qquad (9)$$

The particles are restrained to fixed reference positions R_i. This potential is applied on every axis (x, y and z) of a heavy atom. The term $V_{pr(r_i)}$ is a user-defined force-constant, such that r_i does not deviate

from the original position R_i.

iv. Maintenance of temperature by temperature coupling: MD simulations of biomolecules are performed in NPT ensemble (where there are constant number of particles, constant volume & temperature of the system). Since the temperature needs to be maintained constant, a thermostat is applied. The goal of implementing thermostat is to ensure that fluctuations in the average temperature of the system are minimized. The following temperature coupling algorithms were used in this study:

a. Berendsen & V-rescale: Berendsen algorithm corrects local variation in temperature according to the following equation:

$$\frac{dT}{dt} = \frac{T_0 - T}{\tau} \qquad (10)$$

Where τ is the time constant, T_0 is the temperature of the system at preceding step, T is the temperature at succeeding step. The deviation decays exponentially with time constant. Since this algorithm serves to eliminate kinetic energy arising from frictional forces, it is ideal to use this during equilibration steps. However, this method tends to produce a non-canonical ensemble. Thus, for better sampling, V-rescale method is used in this study for simulations of TSPO coarse-grained systems (chapter 4 & 5). V-rescale is essentially Berendsen thermostat, with an additional kinetic energy and degrees of freedom parameter:

$$dK = (K_0 - K)\frac{dt}{\tau_T} + 2\sqrt{\left(\frac{KK_0}{N_f}\right)\frac{dW}{\sqrt{\tau_T}}} \qquad (11)$$

Where K is the kinetic energy of the system, N_f is the number of degrees of freedom and dW is the Wiener process, which depicts temperature fluctuations in a stochastic manner.

b. Nosé-Hoover thermostat: Nosé-Hoover method introduces a friction term into the equations of motion. The frictional force is proportional to the product of particle's velocity and friction parameter:

$$\frac{d^2 r_i}{dt} = \frac{F_i}{m_i} - \frac{P_\varepsilon}{Q}\frac{dr_i}{dt} \qquad (12)$$

The friction parameter is defined by $\frac{P_\varepsilon}{Q} = (T - T_0)$. Reference temperature is denoted as T_0, instantaneous temperature is denoted by T. This thermostat was used for simulations of TSPO-bilayer systems in all-atomic configuration (chapter 3).

v. Pressure Coupling: Similar to temperature, pressure is also kept constant in a

non-canonical ensemble, also referred to as 'NPT ensemble' (abbreviation for constant number of particles, pressure and temperature). The following methods were implemented in semi-isotropic mode, due to the fact that this study was performed on a protein-bilayer system. When semi-isotropic pressure coupling is implemented, uniform pressure is implemented in *x*- and *y*- axes, but differently implemented in *z*- axis, or the axis of membrane. Two types of pressure coupling methods are used:

a. ***Berendsen:*** This method rescales coordinates and box vectors every step, with time-constant τ, which has first order kinetic relaxation effect towards reference pressure P_0:

$$\frac{dP}{dT} = \frac{P_0 - P}{\tau_p} \qquad (13)$$

b. ***Parrinello Rahman***: Parrinello Rahman approach is similar to Nosé-Hoover approach in the sense that it introduces a barostat term to the equations of motions:

$$\frac{dP^2}{dt^2} = V W^{-1} b'^{-1} (P - P_{ref}) \qquad (14)$$

Where V is volume of the simulation box, W is the matrix parameter determining coupling strength. The Berendsen and Parrinello-Rahman methods were used for coarse-grained (chapter 4 & 5) and all-atom simulations (chapter 3) respectively.

2.1.4. Coarse-grained force fields

Martini coarse-grained force field: While MD simulations enable exploration of large swathes of conformational landscapes which were hitherto not possible with experimental techniques such as X-ray crystallography and NMR spectroscopy, the primary limitation of simulations is that the conformational exploration comes with a high computational cost. Larger the protein, higher number of degrees of freedom associated with the protein dynamics, thus higher the computational cost of the conformational exploration. While such exploration has become possible in the recent past with significant improvements in computation, as well as with the simulation programs, the computational cost still remains a limitation, particularly with sampling of the dynamics of higher-order oligomers and protein complexes. To overcome this limitation, a more simplistic model of representation of atoms was developed, which involved representation of centre of mass of two or more atoms as a bead, having physico-chemical properties equivalent to the constitutive particles. Martini is one such model, which has been employed successfully to many biomolecular simulations. Martini force field was originally parameterized on lipids (Marrink et al. 2004; Marrink et al. 2007), then later extended to proteins (Monticelli et al. 2008). In Martini force field, every four heavy atoms are represented as a

bead (Fig. 2.2). In case of atoms that are part of aromatic rings, every two atoms are represented as a bead, for higher accuracy of representation of their motions. The model has four types of interaction sites, based on hydrogen-bonding capabilities: donors, acceptors, both, none. The atom types are grouped into each of these groups. depending on their hydrogen-bonding capabilities (Monticelli et al. 2008).

While both the bonded and non-bonded parameters are determined in a similar way as all-atomic force fields (equations 1-5), there are some differences. For measuring LJ potentials, the value of potential minima ε varies between 2.0-5.6 kJ/mol depending on the polarity of the interacting pairs of particles, to appropriately mimic polar and non polar interactions. For interactions between charged particles, an additional parameter, relative dielectric constant (ε_{rel}) is set to 15. The cutoff distance above which energies are not calculated is 1.2nm, against 0.9nm used in all-atom force fields. Furthermore, non-bonded interactions between neighbouring atoms are not considered. The major change for bonded interactions is that the force constant *k* is kept low, to mimic collective motions occurring at the all-atomic level (Monticelli et al. 2008).

Since the force field was parameterized on a model transmembrane peptide, it is an ideal model for the simulations of TSPO in bilayer. Martini force field particularly places emphasis on accurate modelling of secondary structure at the coarse-grained level. The parameters for amino acid backbones are modelled based on the definitions calculated from DSSP program (Kabsch and Sander 1983). DSSP assigns eight type of secondary structures, namely α-helix, β-strand, extended, bend, turn, 3_{10} helix, π-helix and unstructured, depending on the backbone C-alpha atom geometry. While bond-lengths of all the bonds involving backbone beads are set to 3.5Å, force constants for bond angles and dihedral angles vary depending on the secondary structure. This change is implemented only if four given dihedral-forming residues are part of the same secondary structure (Monticelli et al. 2008).

Further improvements were made to the force field by tweaking parameters of phenylalanine and proline residues, which were criticized to be too hydrophobic in the initial versions of the force field, and introduction of embedded charges to charged residues that are buried/embedded in the bilayer, and modification of backbone parameters resulting in improved stability of α-helices (de Jong et al. 2013). The improved force field, named Martini22 is used to describe both TSPO and the lipids as coarse-grained representation in this study (chapter 3, 4 & 5). In addition, a water model was also developed, which represented four water molecules as a three-bead model (Yesylevskyy et al. 2010).

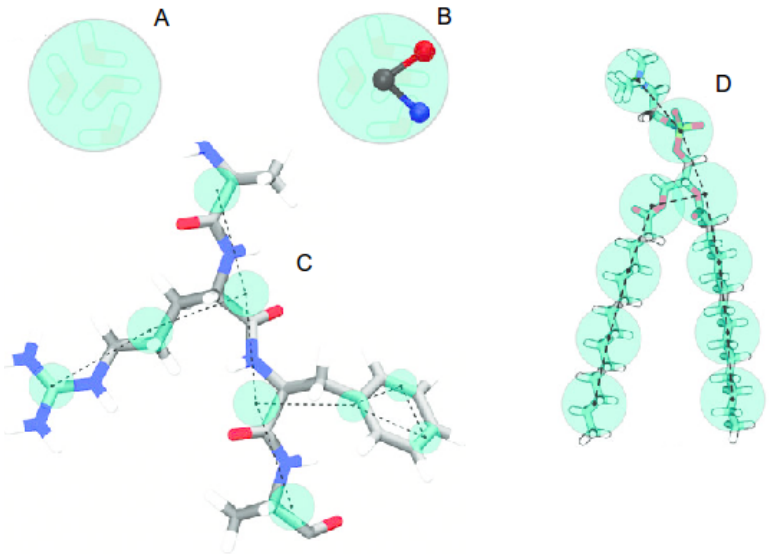


Fig. 2.2. Illustration comparing all-atom structures of biomolecules with respective coarse-grained models (illustrated as green beads): A: Water molecule, B: Polarisable water molecule, C: Peptide fragment, D: Lipid molecule (Marrink and Tieleman 2013).

2.2. Analysis methods of molecular dynamics simulations:

2.2.1. Selection of residues for analyses: All the following analyses (all atom and coarse-grained simulations, chapters 3, 4 & 5) were conducted on the C-alpha atoms, or backbone beads (in case of coarse-grained simulations) of the TM regions, unless mentioned otherwise. The residue selections are outlined in table 2.1.

TM region	Mouse	*R. sphaeroides*	*B. cereus*
1	7-26	7-24	8-25
2	46-63	45-62	45-62
3	82-101	71-90	71-90
4	106-124	100-118	100-118
5	134-157	124-143	124-143

Table 2.1. Table describing boundaries of TM regions, which were employed for all the analysis of TSPO simulations. The TM boundaries are based on the TM definitions from OPM database (Lomize et al. 2012; Lomize et al. 2011).The definitions were tinkered so that the number of residues across TSPO of all the three species are equal.

Since the primary objective of this thesis is to understand the global dynamics of TSPO in various contexts, inclusion of side chains for the analyses would result in addition of additional variable, and increase in degrees of freedom and complexity of analysis. Indeed, most of the motions involving conformational transitions at the level of secondary structure and tertiary structure are accurately captured in the motions of backbone atoms. Thus, all the analyses were performed on the C-alpha atoms. For the coarse-grained trajectories, analyses were performed on the backbone beads, which is representative of backbone atoms and their dynamics.

Because TSPO is a transmembrane protein with five membrane-spanning alpha helices, the residues of these regions tend to have lesser degrees of freedom, due to the fact that they are embedded in the membrane. In contrast, the loop regions have higher degrees of freedom, which result in large number of non-directional, non-collective motions, that may hinder accurate description of large-scale collective motions occurring at microsecond scales. Indeed, in case of TSPO, when principal component analyses were performed on all the C-alpha atoms of the TSPO super-ensemble containing TSPO conformations in *apo*- and *holo*- forms, there was no clear separation between the TSPO *holo*- and *apo*- state conformations (Annexure-1). On the other hand, similar analyses on the residues of TM regions revealed a clear differentiation of conformations explored, depending on the *holo*- or *apo*- state of TSPO. Considering these factors, all the analyses were performed on the C-alpha atoms of TM regions.

Definition of the boundaries of TM regions is another important factor that dictates the outcome of the interpretation of simulations. The boundaries of TM regions are adopted from the boundaries defined for TSPO in the OPM database (Lomize et al. 2012). OPM (Organization of Proteins in Membrane) is a database containing the putative boundaries of TM regions of various integral, as well as peripheral membrane proteins. The boundaries were determined from calculation of membrane transfer energies

for various transmembrane proteins available from Protein Data Bank (PDB). For calculation of membrane transfer energy, atomic solvation parameters for N, O, S, sp2 and sp3 carbon atoms were derived from partition coefficients of water-decadiene, which mimics with that of a bilayer and water-hexadecane, which mimics that of a detergent (Lomize et al. 2006). These parameters were used to calculate membrane-transfer energies of TM proteins, whose structures are available from PDB. Membrane transfer energy was calculated for a TM protein by tweaking four variables associated with the protein: tilt angle with respect to *z*-axis, hydrophobic thickness, translation of the protein along the *z*-axis and rotation of the protein with respect to *z*-axis. The combination of variables that result in lowest transfer energy is considered as the orientation of the given protein. In addition, to account for the fact that several TM proteins have an internal cavity consisting of polar and charged residues, atoms lining such cavities are not considered for calculation of energies (Lomize et al. 2006).

2.2.2. Principal Component Analysis (PCA): While MD simulations of proteins involve exploration across conformational landscape, it is difficult to determine the primary directions of exploration just by direct visualization and observation. This is due to the fact that there are complex side-chain dynamics, which further complicate interpretation of conformational exploration, and is particularly true for proteins with higher-order structure involving more than one monomer. To address this problem, PCA is employed to understand the directions of dynamics of proteins, and this technique was employed to understand TSPO dynamics in various contexts (chapters 3, 4 & 5). PCA is a statistical tool that involves reduction of number of dimensions associated with the given data. In this case, the data is the conformational coordinates explored by the protein during the course of simulation.

$$C_{ij} = \left\{(r_i - \{r_i\}))(r_j - \{r_j\})\right\} \qquad (15)$$

Where C_{ij} is the element of the covariance matrix C, r_i and r_j represent positions associated with atoms i and j respectively. PCA on the MD trajectories typically involves construction of a covariance matrix between each and every pair of atoms of the protein. In this study, the covariance matrix was constructed on the C-alpha atoms of the residues of TM regions, as described above. The covariance matrix is diagonalized and decomposed into eigenvectors or principal components, which represent primary directions of conformational exploration (David and Jacobs 2014).

$$C = E\Lambda E^{-1} \qquad (16)$$

Where E represents the matrix of eigenvectors, and Λ is the diagonal matrix consisting of corresponding eigenvalues. Every principal component (PC) consists of a corresponding eigenvalue, which represents the contribution of the particular eigenvector to the total dynamics, which is represented in terms of number of trajectory frames. Since each PC represents a unique direction of

conformational exploration, the conformational changes corresponding to the given principal component may be extracted from the simulation, which enables direct visualization of the directions of dynamics (David and Jacobs 2014). Furthermore, the trajectory frames may be projected onto the variable space of the respective principal components. On projection onto the variable space of two dominant PCs, one may observe formation of distinct clusters, which correspond to subset of conformations that represents the particular subspace, which enables visualization and easier interpretation of the essential dynamics of the protein. Typically, about 80-90% of the protein dynamics are captured in 5 principal components (Amadei et al. 1993; David and Jacobs 2014). Due to this reason, PCA is a popular tool to understand the dominant directions of the protein dynamics.

2.2.3. Protein Structural Networks (PSN): With a growing body of evidence support a dynamic nature of proteins, a need was felt to characterize protein dynamics in a more intuitive manner. At the same time, clustering and graph theory was gaining cognizance as a suitable method to understand relationships in various fields. Thus, an attempt was made to understand protein structures and their dynamics by application of graph theory and network science. This exercise involved visualization of proteins as a network of residues, which are connected to each other by the means of edges, representing the interactions between the residues. Stronger the interactions between a given pair of residues, stronger the edge. Studies have demonstrated that residues that form the centre of the network, i.e., the residues having maximum number of connections have been found to be functionally important, often acting as substrate binding residues, or as a part of allosteric pathway. Since these residues form the pivot of the network, they are often part of shortest paths of communication between the residues. Furthermore, it has also been observed that these residues are critical for maintaining the integrity of the pathway, which underscores their strong evolutionary conservation.

To deduce the structure of the PSN, Girvan & Newman algorithm was used in this study (chapters 3 & 4) (Girvan and Newman 2002). Instead of characterization of node centrality, the method relies of construction of the network by defining centrality of edges between communities of nodes. The edge having highest centrality is identified in the network, and is removed. The centrality is recalculated, and the edge having highest centrality is identified again, and removed. This procedure is repeated until no edges are remaining, which result in clear differentiation of the communities and the edges that are important for connections between the communities. The method has been found to be a reliable model for representation of network in various instances, including proteins (Vishveshwara et al. 2009; Vishveshwara et al. 2002).

PSNs in conjunction with molecular dynamics simulations were first implemented on the Bio3d molecular dynamics analysis package (Grant et al. 2006). Here, the network is determined from a matrix consisting of correlations between every residue pair, and PSN constructed from the correlation

values of residue pairs. This method has been demonstrated to successfully characterize residue interaction networks and allosteric pathways. On the TSPO trajectory ensembles, direct cross correlation matrix (DCCM) was constructed on the C-alpha atoms of the residues of TM region. DCCM describes correlations between the dynamics of two given residues. The correlation is high if there are high degree of similarity in the residue motions, and vice-versa. Since the goal is to characterize long-range pathways and connections, residue pairs that are in contact with each other were filtered from the matrix. This was performed by filtering of the matrix from the protein contact map, with a cut-off value. However, the cut-off value adopted for construction of matrix has varied in previous studies, depending on the original objective of the study. To characterize long-range interactions, a cutoff value of 7Å was used in one study. In another study, a cutoff value of 10Å was used. Since DCCM was constructed here on the C-alpha atoms, a more stringent threshold of 15Å was employed to filter the correlations that involve contacts between the side-chain atoms (Grant et al. 2006).

For construction of the network, a cutoff value of correlation is applied, below which a given correlated pair is not considered for construction of PSN. To filter the strongest connections between the communities, a stringent correlation threshold was employed. Only the residue pairs having Pearson's correlation of 0.5 and above were used to construct the network. The network consists of residues represented as nodes, which are connected by edges. Group of residues that are closely connected, based on the correlation with their dynamics are grouped into communities. Between two largest communities, the allosteric pathway is traced, and visualized.

While this pathway represents allosteric long-range communication from the protein dynamics perspective, we attempted to reconstruct the same from the biophysical approach. To achieve this, pairwise interaction energies, both hydrophobic and electrostatic interaction energies were calculated between all the residue pairs of TSPO trajectory ensembles in presence of PK11195. A matrix was constructed based on the pairwise interaction energies, and structural network was constructed based on this matrix using Grinn program (Serçinoglu and Ozbek 2018). The similarity of networks was evaluated by comparison of node centralities of both the networks (constructed from DCCM versus pairwise interaction energies), and the shortest pathway was mapped between the same set of residues which were identified components of allosteric pathway.

2.2.4. Protein-Protein Docking: Many of the physiological processes in the cell are achieved by the means of complex protein-protein interactions. Despite the importance of such interactions, their study remains a major bottleneck in the field of biochemistry and molecular biology. Currently, there are only a handful experimental techniques that enable characterization of protein-protein interactions, such as yeast two-hybrid assays, fluorescence resonance transfer, mass spectroscopy, cryo electron

microscopy, small X-ray scattering to name a few. Furthermore, none of the above-mentioned techniques enable full characterization of the protein interactions at the atomic level.

To mitigate this limitation, computational tools have proven to be an invaluable asset. Protein-protein interactions have been successfully characterized by means of MD simulations (Cuendet and Michielin 2008; Rakers et al. 2015). However, a major limitation of MD simulations is the computational cost, which makes it very difficult to study protein-protein interactions at level of desired sampling, particularly in case of large protein complexes. Protein-protein docking methods have been proved to be an attractive alternative for such objectives. A typical protein-protein docking protocol involves the following steps:

a. Search stage involving generation of a number of complexes with diverse conformations and interfaces.
b. Optimization of side-chain conformations (applicable to flexible-docking methods)
c. Filtering physico-chemically/physiologically unfeasible complexes, scoring, clustering and ranking of candidate complexes.

The distinct advantage of the protein-docking methods over MD simulations is that it has the ability to sample diverse range of interfaces between a given pair of proteins and generate a large number of candidate complexes, without the computational cost associated with MD simulations. With a robust scoring method, protein-docking enables generation of physiologically relevant complexes, interfaces of which can also be corroborated by means of experimental studies.

While the search stage involves generation of candidate complexes, involving the use of methods such as fast fourier transform, shape complementarity search and grid-based search methods, scoring stage involves ranking of the complexes according to certain physicochemical parameters. This often includes scoring by electrostatic interactions, interface area, VdW interactions etc. While various docking programs use different types of scoring functions, use of inappropriate scoring function may result in generation of physiologically unfeasible complexes. This is particularly true for TM proteins such as TSPO and VDAC, due to the fact that many of the currently available docking programs use scoring functions that were optimized using globular proteins. In a study that compared various such docking programs for their feasibility to use on TM proteins, it was observed that most of the docking programs, including ZDOCK, ClusPro, PatchDock, SymmDock etc. failed to generate correct TM complexes. In other words, the complexes generated by these programs were found to have RMSD as high as 11-13Å with respect to the experimentally determined complexes (Kaczor et al. 2013). Among these programs, GRAMM-X docking method was found to be best performing, in terms of RMSD with respect to experimental structures. GRAMM-X uses fast-fourier transform method for obtaining candidate complexes (Tovchigrechko and Vakser 2006). For scoring, it relies primarily on a modified

inter-atomic LJ potentials:

$$V_{ij} = \frac{1}{\alpha\sigma_{ij}^{6}+r^{6}}\left(\frac{4\varepsilon_{ij}\sigma_{ij}^{12}}{\alpha\sigma_{i,j}^{6}+r^{6}} - 4\varepsilon_{ij}\sigma_{ij}^{6}\right) \qquad (17)$$

ε represent depth of the potential well associated with the interaction, σ represent the finite distance at which the interaction between atom pairs is zero. α describes 'softness' of the interaction, or the minimum distance at which LJ potential approaches a finite value (Liu et al. 1996). The values of α, ε and σ were set to be 0.4, 0.33nm and 0.5 respectively, which in turn, were obtained from benchmarking docking calculations, representing closest properties for a typical protein in unbound conformation (Tovchigrechko and Vakser 2006). Apart from LJ potential term, the program also relies on evolutionary conservation term of the predicted interface, frequency of residue preference in the interface, binding free energy, atomic contact energy and the volume of the minimum (Tovchigrechko and Vakser 2006). However, it was also observed that the performance of GRAMM-X was due to the fact that the in certain complexes, particularly bacteriorhodopsin monomers used for docking were in the same orientation as that of the dimer complex. When the method was tested on the protein with monomer conformation, the RMSD increased significantly (Kaczor et al. 2013).

While there were no docking programs tailored for membrane proteins, a docking method was developed specifically for alpha-helical membrane proteins recently, called Memdock (Hurwitz et al. 2016). Memdock uses a modification of PatchDock program, called Mem-PatchDock for generation of complexes, which imposes lipid-bilayer environment during generation of complexes (Hurwitz et al. 2016; Schneidman-Duhovny et al. 2005). Scoring is performed by a combination of various terms:

$$E_{total} = aE_{env} + bE_{HB} + cE_{aliph} + dE_{\pi\text{-}\pi} + eE_{repVdW} + fE_{attVdW} + gE_{repElect} + hE_{attElect} \qquad (18)$$

Where E_{HB} represents net energy from hydrogen bonding, E_{aliph} represents aliphatic interaction energy, $E_{\pi\text{-}\pi}$ represents interaction energy from π-π stacking interactions, E_{repVdW} and E_{attVdW} represent repulsive and attractive VdW terms respectively, $E_{repElect}$ and $E_{attElect}$ represent repulsive and attractive electrostatic terms respectively. The term E_{env} represent a form of membrane transfer energy, which is calculated by:

$$E_{env} = \Sigma_i - ln\left(\frac{N(aa_i|L,B)}{N(aa_i)}\right) \qquad (19)$$

Where $N(aa_i|L,B)$ is the frequency of residue type aa_i in the given layer (L,B). The layers are demarcated as shown in Table 2.2, $N(aa_i)$ represents frequency of residue type aa_i in all the layers. Each layer corresponds to specific z-coordinates of the given protein residue, and frequency of residue

in a specific layer is calculated from available dataset of membrane proteins from PDB.

Layer-1	-12Å < Z-coordinate of the residue <12Å
Layer-2	12Å < Z < 18Å
Layer-3	18Å Z < 24Å
Layer-4	2Å4 < Z
Layer-5	-18Å < Z < -12Å
Layer-6	-24Å < Z < -18Å
Layer-7	Z < -24Å

Table 2.2. Table representing boundaries of each layers, in terms of z-coordinates. The above represented layers are used to calculate frequencies of amino acid residues occurring at respective z-coordinates, which in turn are used to calculate membrane transfer energies for a given docked complex in Memdock web server (Hurwitz et al. 2016; Schneidman-Duhovny et al. 2005)).

The docking method was standardized and validated by docking chains of experimentally determined complexes and comparison with experimentally determined complexes, as well as the complexes obtained from GRAMM-X method. It was observed that Memdock performed significantly better compared to GRAMM-X, and was closer to the experimental complex in terms of RMSD (Hurwitz et al. 2016). Due to clear advantage of Memdock over other docking programs, Memdock was used to dock TSPO monomer conformations and compare the interfaces of best-ranking complexes with those observed in simulations (Chapter 4).

2.3. Methods employed for comparison of TSPO paralogues:

2.3.1. Multiple Sequence Alignment: Multiple Sequence Alignment (MSA) is the method of aligning more than two sequences that are homologous to each other, to varying degrees. With the advent of protein and DNA sequencing projects in the 1980's and 90's, and subsequent explosion in the availability of protein sequences from genome sequencing projects, there was a need for a method to identify homologies and variations among different sequences. With this regard, MSA proved to be a valuable tool. MSA involves aligning multiple sequences together, and scoring the positions, depending on whether there is conservation or substitution, using a scoring matrix. In positions containing gaps, penalties are introduced. Various algorithms have been developed to construct MSAs, including Muscle, ClustalW, MAFFT to name a few (Edgar 2004; Thompson et al. 2002; Katoh et al. 2002). Most of them rely on scoring matrices for alignment of sequences, such as BLOSUM or PAM series of matrices (Henikoff and Henikoff 1992; Dayhoff 1972). Scoring matrix is among the most important deciding factors on the quality of the MSA. Various scoring schemes have been developed, that assign specific scores for matches and mismatches for given positions. However, since these matrices are derived from large datasets involving thousands of protein sequences, they are often generalistic, and may give rise to inconsistencies in the alignment.

Recently, a new MSA algorithm was developed, called PROMALS3D (Pei et al. 2008; Pei and Grishin 2014). Instead of relying on generalized scoring matrices, it relies on the principle that

secondary structural elements are critical to maintain structural integrity and function, and are often more conserved than unstructured elements. The methodology of this method, employed in chapter-6 is as follows:

a. With the given number of initial sequences, clustering is performed to eliminate highly identical sequences (identity > 95%).
b. With the clustered sequences, sequence search is performed using PSI-BLAST to obtain additional homologs.
c. From the additional homologs, a sequence profile is built based on the sequence conservation.
d. Apart from the sequence profile, a secondary structure profile is built, based on secondary structure assignment using PSIPRED program (Jones 1999).
e. Constraints are derived from the sequence, secondary structure profiles, as well as from the 3D structure homologous to the sequence, which is user-defined. In the case of TSPO, the NMR structure of mouse TSPO (PDB ID: 2MGY) was used as the input 3D structure. The constraints assign additional weights to certain residues that are more conserved that others. Similarly, additional weight is assigned if the given position has a secondary structure.
f. The alignment is performed, scored and refined, based on the derived constraints.

2.3.2. Phylogenetic tree construction: Construction of phylogenetic trees based on information from DNA/protein sequences has become a primary means to determine evolutionary relationships of the organism with respect to other organisms. This method was used in this study for understanding evolutionary relationships between TSPO1 and its paralog TSPO2 (chapter 6). The phylogenetic tree construction operates on the principle that proteins evolve into orthologs (homologous proteins that are product of speciation) or paralogs (homologous proteins that are products of gene duplication in same organism) by means of accumulated mutations/changes in sequence. This principle was pioneered by Kimura (Kimura 1979). Various methods of construction of phylogenetic trees and their interpretation has been developed. In this study, maximum parsimony and maximum likelihood methods were employed to determine evolutionary relationships between TSPO1 and TSPO2 sequences.

Maximum parsimony method relies on the principle that evolution between two sequences occurs via minimum number of substitutions. In other words, the most likely path of transition between two given sequences is the shortest one. This principle was proposed from the pioneering work by Eck, Dayhoff & Fitch (Fitch 1977). On the other hand, maximum likelihood method is based on the statistical likelihood of substitution in a given MSA position. Though this method is more robust and reliable, it is computationally expensive. We exploited both the methods to obtain a reliable model of

TSPO evolution. An initial tree was generated by maximum parsimony method, which was used as an input tree, for reconstruction using maximum likelihood method. Phylogenetic tree construction involves three steps:

a. Clustering of sequences to generate an initial tree
b. Filtering and refinement of tree
c. Statistical test for reliability of tree by bootstrapping

PhyML program was used for performing above steps (Guindon et al. 2005; Guindon et al. 2010). For the first step, it uses an input multiple sequence alignment and input tree to construct a starting tree, from which an initial consensus tree is generated. For filtering the tree, PhyML uses sub-tree pruning and regrafting method. This involves elimination of trees that have less likelihood, with respect to number of parsimonious moves required for substitution in a position. The trees that involve least number of steps for substitution of given two sequences are selected, and nearest neighbour searching is performed to further optimize tree (Guindon et al. 2005; Guindon et al. 2010). To test the reliability of the consensus tree, bootstrapping is performed. Bootstrapping is a statistical method that involves swapping the sequence positions randomly, and reconstruction of the phylogenetic tree based on the new sequences with swapped positions. Since 100 bootstrap replicates were performed for construction of TSPO phylogenetic tree, 100 such steps were performed. The number of steps where the tree obtained is identical with respect to the initial tree is displayed for every node. Generally, a node with bootstrap value of 50% or above is considered reliable.

As it has been observed from previous studies that TSPO2 diverged from TSPO1 during the course of evolution of birds and mammals, and experimental studies suggest that TSPO2 lost drug-binding abilities that are conserved in TSPO1 (Fan et al. 2009), it is interesting to study the differences between the putative common ancestral sequence of TSPO1 and 2, and present TSPO1 and 2 sequences. For this purpose, FASTML web server was used (Ashkenazy et al. 2012). FASTML is a maximum-likelihood based ancestral sequence reconstruction method, that utilized both substitution as well as data from insertions & deletions to generate a list of ancestral sequences with most likelihood (Ashkenazy et al. 2012). While many of similar programs take only the data from substitutions into account, FASTML uses the information from insertions and deletions as binary data. The most likely ancestral sequences are then constructed using continuous-time Markov process.

2.3.3. Normal mode analysis:While molecular dynamics is a popularly used method to study protein dynamics, there are some limitations. The conspicuous limitation is that it consumes large amount of computational resources. This is particularly true for large protein complexes and multi-domain proteins. However, this barrier can be overcome by another method called normal mode analysis (NMA). Similar to molecular dynamics, NMA involves computation of forces on atoms in molecule,

where atoms are connected by springs (which are represented by beads), forming an elastic network. Apart from that, there are significant differences. Seminal work by Tirion (Tirion 1996) proposed that in large biomolecules such as proteins, characterization of single parameter, backbone dihedral potential is enough to describe global slow dynamics of proteins. Such characterization has been found to have good agreement with dynamics observed from experimental studies in various instances (Na et al. 2018; Kundu et al. 2002; Yang et al. 2007), and are used in this study for comparison of slow-mode dynamics of TSPO1 and TSPO2. In a normal mode calculation, overall potentials are calculated as:

$$V_{ij} = \frac{1}{2}\Sigma K\,\phi(|r_{i,j}| - |r^0_{i,j}|)^2 \qquad (20)$$

The potentials are calculated as Hookean potential, which are harmonic in nature. Based on this potential, Hessian matrix is calculated and decomposed into eigenvectors with eigenvalues which represent frequencies of dynamics. The forces are computed as:

$$F_{ij} = \frac{d^2 V_{ij}}{dq_i dq_j} \qquad (21)$$

Where q_i and q_j represent position vectors of two given atoms. The elements of Hessian matrix consists of forces calculated as depicted in equation (21).

Diagonalization of Hessian matrix gives eigenvectors that represent dynamics at various frequencies. The eigenvectors with lowest frequencies exhibit dynamics that represent global conformational changes of the protein. Several variants of NMA were developed, one such variant is Anisotropic Network Model (ANM) (Atilgan et al. 2001; Bahar et al. 2010). ANM is a variation of coarse-grained normal mode analysis so that equation (21) is represented as:

$$\frac{d^2 V_{ij}}{dq_i dq_j} = \frac{\gamma_{ij}(x_j - x_i)(y_j - y_i)}{r^2_{ij}} \qquad (22)$$

However, NMA makes certain assumptions, of which one needs to be mindful of: The method assumes that the initial coordinates q are at an energy minimum, and thus the dynamics sampled by this method represent dynamics around this minimum. However, the method leads to several inaccuracies if the initial coordinates are far from the local minimum. The potentials are calculated and matrix is diagonalized depending on a cut-off value. This cut-off value dictates whether the given pair of atoms are considered as connected by springs. Typically, a cut-off value of 12-15Å between C-alpha atoms is considered a reasonable potential to reproduce experimentally observed dynamics (Atilgan et al. 2001), and a cut-off of 13Å was used in this study (chapter 6). Construction of ANM were performed using NMWiz tool of ProDy MD analysis program (Bakan et a. 2011).

Chapter-3:

Impact of PK11195 on TSPO dynamics

3.1. Introduction 78
3.2. Results 89
3.2.1. The structure of mouse TSPO in bilayer environment is stable, but slightly differs from the NMR structure with DPC micelles 89
3.2.2. PK11195 ligand impacts TSPO secondary structure 92
3.2.3. PK11195 impacts global dynamics of TSPO 94
3.2.4. Long range correlated motions are activated in TSPO bound to PK11195 99
3.2.5. The long-range motions correspond to a possible allosteric pathway from the conserved to variable regions 101
3.2.6. PK11195 adopts a wide range of conformations when it is bound to TSPO 103
3.2.7. PK11195-interacting residues around the binding cavity have strong evolutionary conservation 108
3.3. Discussion 109
3.3.1. Common binding sites of PK11195 and PP-IX ligands 109
3.3.2. Dynamics of PK11195 ligand 110
3.3.3. Helical rotations of TSPO 110
3.3.4. Effect of PK11195 ligand on TSPO dynamics 111
3.3.5. Functional implications of allosteric communication 116
3.4. Conclusions and Future directions 117

IMPACT OF PK11195 LIGAND ON THE TSPO DYNAMICS

Rajas M Rao, Aurore Vaitinadapoulé, Julien Diharce, Mariano Ostuni, Frédéric Cadet and Catherine Etchebest

ABSTRACT

Translocator protein (TSPO) is an 18 kDa transmembrane protein, localized primarily on outer mitochondrial membrane. It has been found to be involved in various physiological processes and pathophysiological conditions. Though studies on its structure have been performed only recently, there is little information on the nature of dynamics. We studied the dynamics of mouse TSPO protein by means of molecular dynamics simulations, in presence as well as in absence of the diagnostic ligand PK11195. We studied the nature of conformational exploration in both the modes, and we found the presence of coordinated motions between the residues. These motions were found to be part of a putative allosteric pathway, where the residues with evolutionary conservation and possible functional significance transfer the entropy to less important residues , which are evolutionarily variable. We also described the dynamics of PK11195, and PK11195-TSPO interactions. This study provides new insights into the way in which allosteric communication occurs across the protein residues, apart from shedding some interesting perspectives on possible role of TSPO residues and their dynamics from the structural, functional and evolutionary standpoint.

3.1. INTRODUCTION

Translocator Protein (TSPO) is an 18 kDa transmembrane protein that is localized primarily in the outer mitochondrial membrane(Hirsch et al. 1989; Olson et al. 1992), but also has been found in various other locations, such as plasma membrane(Bouyer et al. 2011), endoplasmic reticulum(Fan et al. 2009) and in nuclear fractions(Hardwick et al. 1999). Recently, a paralog of TSPO named TSPO2, was identified in red blood cell membrane(Fan et al. 2009). TSPO2 shares 40% identity with TSPO1, the most abundant isoform (among the mouse paralogs).

TSPO has been found to be involved in many physiological processes, including cholesterol transport and steroidogenesis(Li and Papadopoulos 1998; Rone et al. 2012), tetrapyrrole uptake & metabolism(Verma et al. 1987; Rampon et al. 2009; Veenman et al. 2016), apoptosis(Azarashvili et al. 2014; Veenman et al. 2007), regulation of cellular respiration(Liu et al. 2017), regulation of photosynthesis in green photosynthetic bacteria(Yeliseev and Kaplan 1999), response to abiotic and hormone-induced stress in plants(Vanhee et al. 2011) to name a few. Among these, steroidogenesis and cholesterol transport have been the most extensively studied processes. Its role was further confirmed by cholesterol-binding studies to TSPO, where it was found that cholesterol binds to TSPO at nanomolar affinities(Li and Papadopoulos 1998). Subsequently, the TSPO residues that interact with cholesterol were characterized at the C-terminal of the protein, and were named ‘cholesterol recognition and consensus’ motif or CRAC motif (motif: –L/V-(X)1–5-Y-(X)1–5-R/K-)(Li and Papadopoulos 1998; Li, Yao, et al. 2001; Jamin et al. 2005). The CRAC motif consist of a distinct residue signature involving hydrophobic and positively charged residues. Furthermore, import of cholesterol into mitochondria, in which TSPO has been demonstrated to have a major role, was found to be a limiting factor in steroidogenesis, underscoring the importance of this protein(Li and Papadopoulos 1998).

Apart from CRAC motif, numerous other residues were also found to play an important role in cholesterol interaction/translocation. One such residue is Ala147 (in mouse TSPO). While this residue is in immediate vicinity of CRAC motif, mutation of this residues to threonine was found to be associated with reduced pregnenolone production(Costa, Pini, Gabelloni, et al. 2009). Similarly, another motif near to CRAC motif, called ‘enhancement motif’ or LAF motif (144-LAF-146) was found to amplify cholesterol interaction and uptake by TSPO. In bacterial TSPO (*Rhodobacter sphaeroides*), the CRAC motif is conserved, yet it cannot uptake cholesterol. Upon transfection of this motif, cholesterol uptake by *R. sphaeroides* cells increased 1,000-fold(Li, Liu, Valls, et al. 2015). Apart from the above two sets of residues, another alternate cholesterol binding motif was characterized recently. Like CRAC motif, this motif also existed in C-terminal region of TSPO, and consisted of hydrophobic as well as positively charged amino acid residues(Di Scala et al. 2017;

Fantini et al. 2016; Papadopoulos et al. 2018). Theoretical binding studies further revealed that the residues of this motif interact with cholesterol at lower energy compared to CRAC motif. Since this motif was arranged in palindromic manner with respect to CRAC motif, this motif was named ‘CARC motif’(Fantini et al. 2016).

TSPO has also been found to be involved in porphyrin metabolism [8] and erythropoiesis [6,7]. Protoporphyrin-IX (PP-IX), an intermediate in the tetrapyrrole metabolism binds to TSPO, and is one of its endogenous ligands(Li, Liu, Zheng, et al. 2015a). Apart from PP-IX, PK11195 is another exogenous drug ligand, that binds to TSPO with nanomolar affinities. In combination with high specificity, it is also observed to modulate the physiological processes involving TSPO, such as steroidogenesis(McCauley et al. 1995; Mukhin et al. 1989), apoptosis(Veenman et al. 2007), regulation of mitochondrial permeability(Kugler et al. 2008; Veenman et al. 2007) to name a few. PK11195 is an isoquinoline derivative, with an isoquinoline skeleton with a chlorophenyl substituent and amide moieties (Fig. 1.16, chapter-1). While PK11195 exists in two rotameric forms (E- and Z-), characterized by dihedral rotation of the amide moiety, it is known to exist in complex with TSPO as E- rotamer(Jaremko, Jaremko, Giller, et al. 2014).

Since TSPO is involved in various physiological processes, it is also implicated in various pathologies. TSPO is found to be involved in malaria(Bouyer et al. 2011; Marginedas-Freixa et al. 2016), neuroinflammation, reperfusion injury(Morin et al. 2016), various psychiatric and neurological disorders(Rupprecht et al. 2010) to name a few. Because of its ability to bind a variety of drug molecules, this protein is considered as an attractive drug target.

Besides ligands, TSPO is also found in association with many protein partners, including voltage-gated anion channel (VDAC), adenine nucleotide transporter (ANT)(McEnery et al. 1992), PRAX-I(Galiègue et al. 1999), StAR to name a few(West et al. 2001). Nevertheless, despite more than 30 years of intensive studies, the physiological role of TSPO remains debated, with recent controversial results and interpretations(Morohaku et al. 2014; Tu et al. 2014; Zhao et al. 2016). This has made TSPO an enigmatic protein of great interest.

TSPO structures of different species were recently solved by solution NMR method for the mouse form(Jaremko, Jaremko, Giller, et al. 2014), and using X-Ray crystallography for bacterial forms, (*Rhodobacter Sphaeroides (Rs) and Bacillus Cereus (Bc)*) [15,27](Li, Liu, Zheng, et al. 2015a; Guo et al. 2015). The atomic structure of mouse TSPO confirmed the 5-TM fold of TSPO(Jaremko, Jaremko, Giller, et al. 2014; M. Jaremko et al. 2015). The structure was solved in presence of the ligand PK11195, which bound to TSPO in a distinct binding cavity. However, it was found that this binding cavity is located at a distance from the CRAC motif, and it was further proposed that PK11195 binding stabilized the TSPO structure, which in turn facilitated cholesterol transport(Jaremko,

Jaremko, Giller, et al. 2014).

On the other hand, *B. cereus* structure was solved in different conditions, in presence and absence of PK11195 ligand, and for both monomer (PDB ID: 4RYO) and dimer (PDB ID: 4RYI) stoichiometries (Guo et al. 2015). *R. sphaeroides* structures were obtained for the wild type and the A139T mutant, both structures solved as higher order oligomers. The structure of the mutant bound to the endogenous ligand PP-IX was solved as a trimer, and the wild-type structure in absence of a ligand was solved as a dimer(Li, Liu, Zheng, et al. 2015a). Comparison of X-ray structures of both the bacterial TSPO show significant similarities in the transmembrane domain, despite relatively low identity between both the sequences (22%). However, subtle differences exist between the two species, depending on the oligomerization states of a given structures. On the other hand, comparison of the NMR structure with X-ray structures show that though fold remains similar, significant differences between mouse and bacterial conformations exist(Li, Liu, Garavito, et al. 2015; Guo et al. 2015; Li, Liu, Zheng, et al. 2015a). This might be attributed to differences in sequence, to different experimental conditions or to the nature of experimental techniques themselves(Li, Liu, Garavito, et al. 2015). The last point has been raised in particular since the NMR structure was solved in the presence of dodecylphosphocholine (DPC) detergents, which may weaken the stability of the protein and alter its 3D structure(Li, Liu, Garavito, et al. 2015). However, another explanation might also reside in the dynamics of the protein.

Accumulating evidence show that this property is a plain actor of the biological function, completing the well-known sequence-structure-function tryptic. Unfortunately, the protein dynamics is difficult to study at an atomistic scale. In these regards, NMR experiments are considered as the most appropriate ones for examining the dynamics of a protein, even a membrane protein. As an example, in the case of *m*TSPO, solid-state NMR experiments(Ł. Jaremko et al. 2015) performed in absence of PK11195 ligand allowed to highlight the residues that contribute the most to the dynamics of the protein. The absence of NMR signals enabled identification of these residues and for those exhibiting NMR signals, a reduction in helicity content was also observed(Ł. Jaremko et al. 2015).

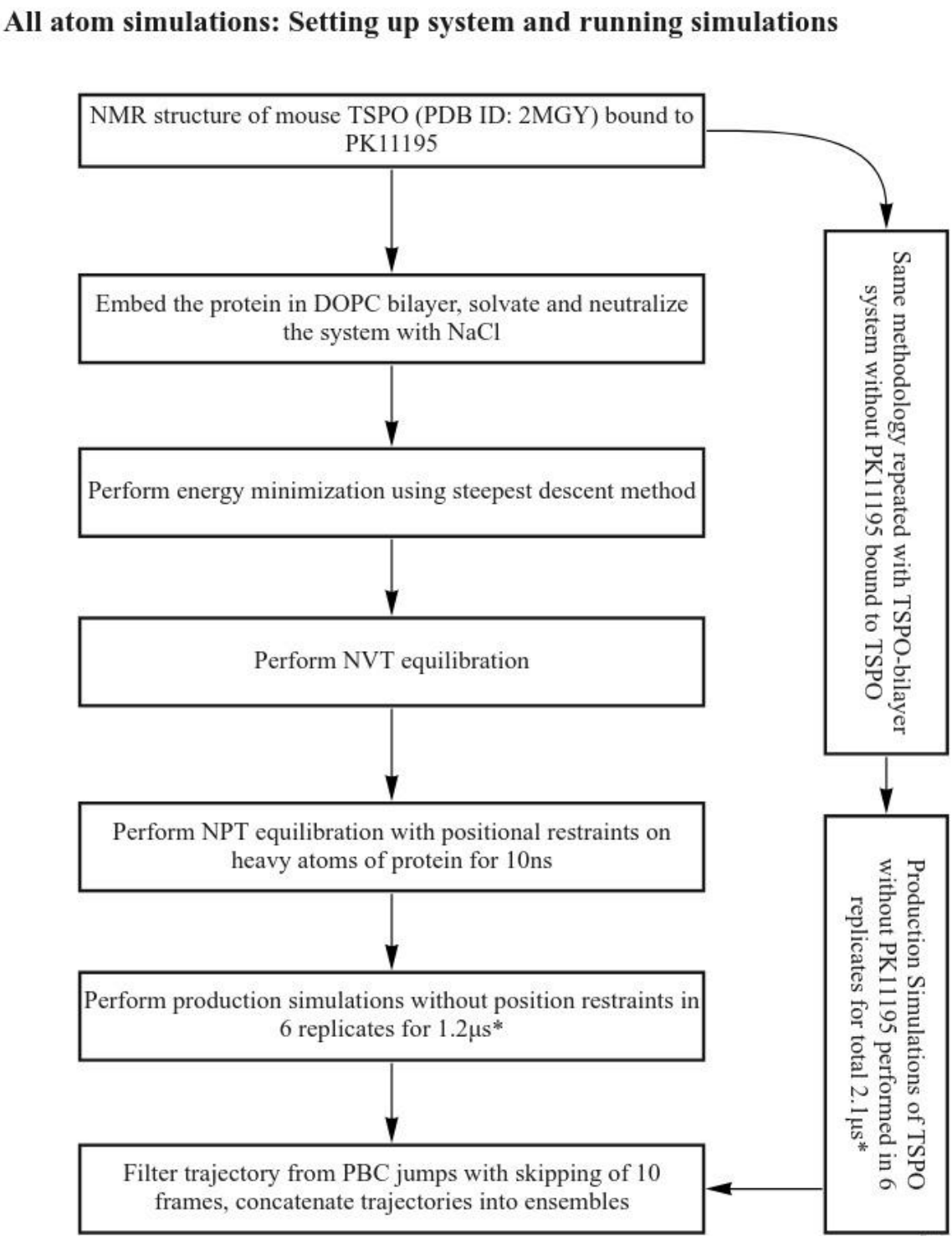
Overall, these observations suggest that TSPO is highly dynamic in absence of PK11195(Ł. Jaremko et al. 2015) or in other words, PK11195 helps in stabilizing the structure(Jaremko, Jaremko, Giller, et al. 2014; Murail et al. 2008). Nevertheless, even though NMR technique is a valuable source of information, it requires expensive and even intractable protocols to provide a complete picture of the dynamics of a membrane protein at an atomistic scale. A worth alternative approach, which is frequently coupled to NMR but also complements structural information from X-ray crystallography, involves use of computational approaches. Many studies that describe intrinsic dynamics of various proteins, and their relation to its function have been performed using molecular dynamics

approaches(Karplus and McCammon 2002; Karplus and Kuriyan 2005).

Accordingly, in order to provide insights into the dynamics of TSPO, we have performed all-atomistic molecular dynamics simulations of NMR structure of mouse TSPO. We chose the mouse sequence, as it is closer to the human sequence (81% identity). The protein was embedded in a DOPC lipid bilayer, in the presence and in absence of PK11195 ligand. We conducted a large set of simulations to obtain a suitable conformational sampling for an aggregate of 3.1μs in both cases. We examined various aspects of TSPO dynamics, including representative dynamics for each binding state and long range correlated motions. As PK11195 ligand, which is located in a cavity bordered by TSPO residues has been proposed to impact stability and the dynamics of TSPO (Murail 2009, Jaremko 2014, Jaremko 2015)(Murail et al. 2008; Jaremko, Jaremko, Giller, et al. 2014; Ł. Jaremko et al. 2015), we also examined in detail the dynamics of PK11195 and nature of its interactions with TSPO residues.

All these results are then discussed at the light of biochemical and biophysical data available and also in terms of evolution. Overall, our results highlight new important features of TSPO that could connect functional motifs and unsuspected key regions of the protein.

The methodology of simulations and their analyses are described in the flow charts below, and detailed description of methods are described in chapter-2.



Box 3.1. Flow chart describing preparatory steps of all atom simulations of TSPO in apo- and holo-forms

Parameters for AA simulations
System:

- ***Force field for TSPO***: ***AMBER99sb-ILDNp*** (Aliev et. al., 2014)
- ***Force field for PK11195***: Parameterzed by Vaitinadapoulé 2015 from QM calculations using Antechamber program of Amber14 MD package (Wang et. al., 2016)
- ***Lipid type***: DOPC
- ***Force field for lipid:*** Slipid
- ***Water model***: TIP3P
- ***Type of ions***: NaCl (0.1M)
- ***Dimensions of simulation box:*** 10x10x10 nm

Energy minimization parameters:

- ***Method***: ***Steepest descent****
- ***No. of steps:*** 50,000
- ***Fn*** = 100 kJ/mol nm
-

Initial equilibration at NVT ensemble:

- Position restraint (force constant): 1,000 kJ/mol nm²
- Time: 25ps (50,000 steps)
- Time-step: 0.5 fs
- Temperature: 300K
- Thermostat: ***V-rescale***
-

Equilibration at NPT ensemble:

- Position restraint (force constance): 1,000, 500, 100, 10, 0 kJ/mol nm² in 5 simulations
- Time: 2ns x 5 simulations
- Time-step: 0.002 ps
- Temperature: 300K
- Thermostat: ***Nose-Hoover***
- Barostat: ***Parinello-Rahman***, with semiisotropic coupling

Production MD simulation

Same parameters as of equilibration at NPT ensemble, except without position restraints

Box 3.2. Parameters for preparatory steps of all atom simulations of TSPO. All the preparatory steps were performed using GROMACS 4.6.7 program(Páll et al. 2015; Berendsen et al. 1995). The details of simulations are described in Table 3.1.

Replicate	Time simulated (ns)	Velocity seed	Water model	Lipid	System composition (excluding protein)		
					Lipids	Water	Ions
1	600	173529	TIP3P	DOPC	363	26,212	Na: 45 Cl: 49
2	600	-3					
3	130	-5					
4	130	-7					

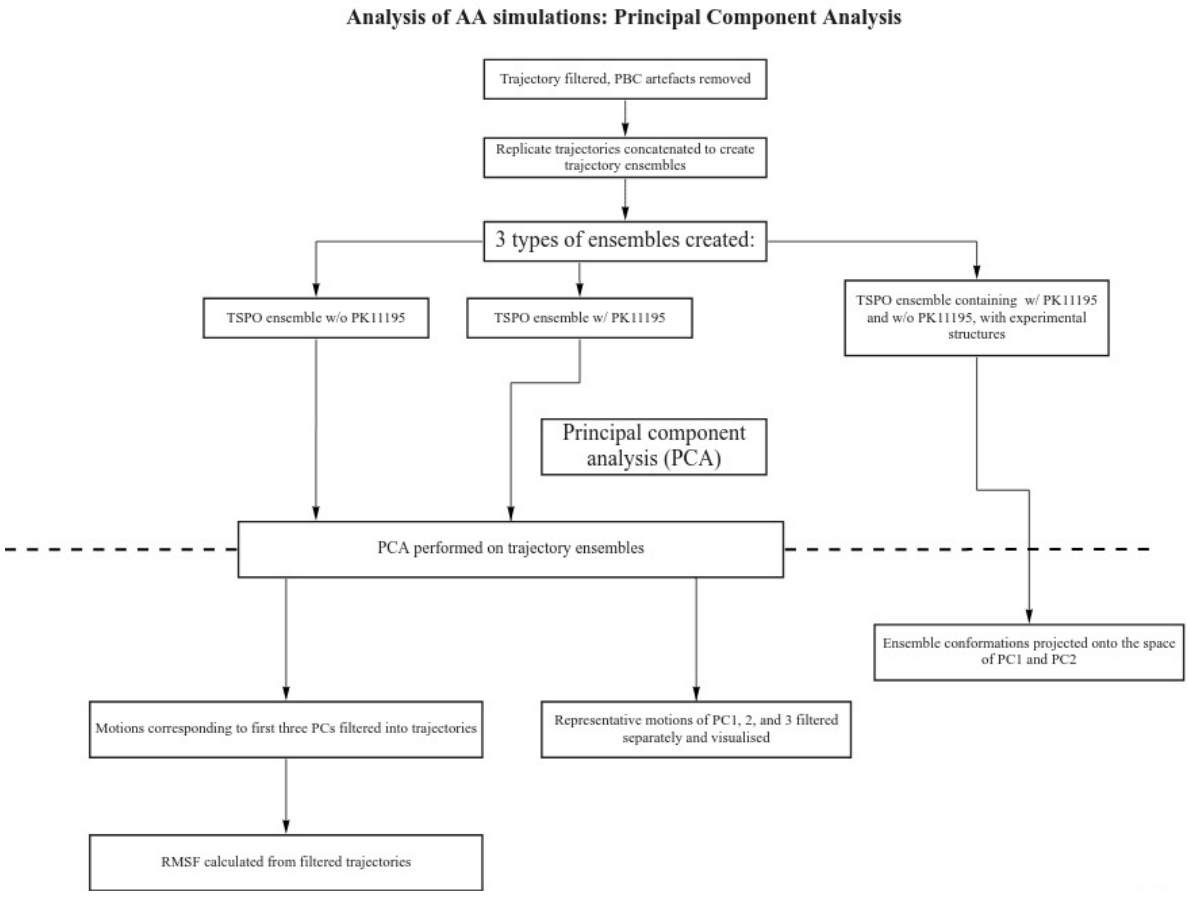
Table 3.1: Summary of TSPO simulations performed in complex with PK11195 ligand.

Replicate	Time simulated (ns)	Velocity seed	Water model	Lipid	Systemcomposition (excluding protein)			PK11195 Parameters
					Lipids	Water	Ions	
1	500	173529	TIP3P	DOPC	360	25,874	Na: 45 Cl: 48	Developed using Antechamber program of Amber MD package (Wang et. al., 2000)
2	500	-3						
3	130	-5						
4	130	-7						

Table 3.2. Summary of TSPO simulations performed in absence of PK11195 ligand, with NMR structure(PDB: 2MGY) as the starting structure.

Replicate	Time simulated (ns)	Velocityseed	Water model	Lipid	System compostion (excluding protein)			CG snapshot (μs)
					Lipids	Water	Ions	
1	420	-3	TIP3P	DOPC	364	6,451	Na: 45 Cl: 49	5.73
2	138	-5						4.98

Table 3.3. Summary of TSPO simulations performed in absence of PK11195 ligand, with snapshots from coarse-grained simulation as starting points. Coarse-grained model of TSPO was created, described by Martini 2.2 force field(de Jong et al. 2013; Monticelli et al. 2008; Marrink et al. 2007), and simulated with steps and parameters described in chapter-5 for 10μs each, in two replicates. RMSD-based clustering was performed on TSPO (3.5Å threshold), and representative structure with highest population was chosen as the starting structure. The trajectory snapshot closest to the representative structure in terms of RMSD was identified, the snapshot extracted and backmapped to all-atomic form using Backward program(Wassenaar et al. 2014).



Box 3.3. Flow chart describing analysis of TSPO dynamics using principal component analysis.

TM region	Residues (used for PCA)	Residues (used for Protein structural network)
1	7-26	6-36
2	46-63	46-69
3	82-101	76-100
4	106-124	104-126
5	134-157	131-157

Table 3.4. Definitions of TM regions used for PCA and protein structural network. The definitions are adopted from OPM database(Lomize et al. 2012; Lomize et al. 2006), which contains boundaries of residues embedded in membrane that are calculated theoretically. Residues used for construction of protein structural network represent the boundaries of TM helices. The boundaries of TM helices were considered for construction of protein structural network in order to account for the information that may be lost by omission of the residues that are helical but not part of TM regions determined by OPM database.

Additional information- Preparation for simulation:

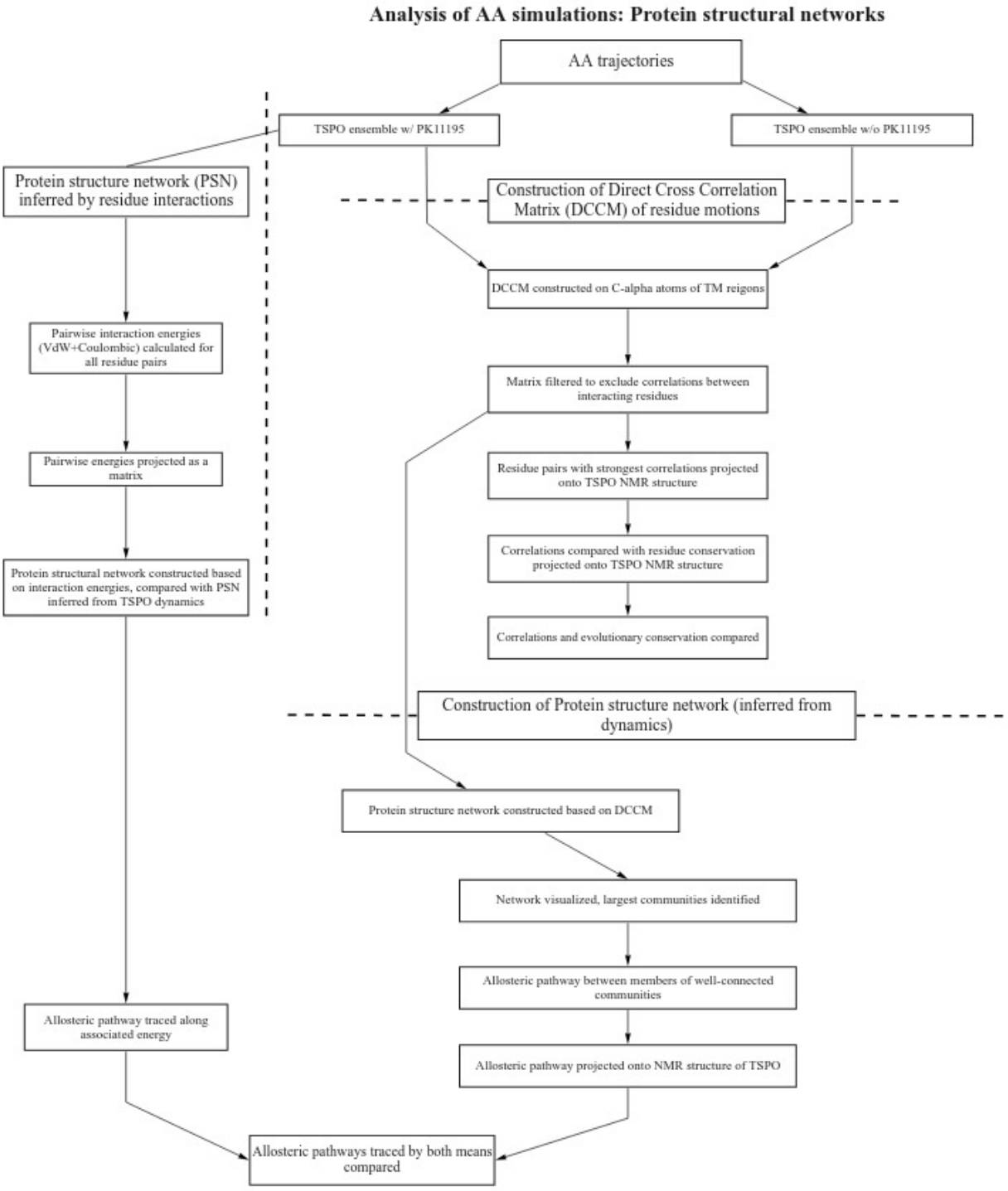
- ***Preparation of AA system based on snapshot of CG simulation*:** Since one set of TSPO simulations in absence of PK11195 were performed with NMR structure (PDB: 2MGY) as the starting point, we performed another set of simulations with a different conformation as a starting point, to investigate the dynamics of TSPO from a different starting point, and also to understand if there is different conformational exploration from a different starting point. For this, coarse-grained simulations were performed with NMR structure as the starting point (identical system as described in table 3.3) in 2 replicates for 10μs each. Martini 2.2 force field (de Jong et al. 2013; Monticelli et al. 2008; Marrink et al. 2007) was used to describe the system in coarse-grained simulation (complete set of parameters described in chapter-5).

  A representative conformation from the simulation was filtered by RMSD-based clustering method using g_cluster tool of GROMACS (Berendsen et al. 1994; Pall et al. 2015). The trajectory snapshot closest to representative conformation (with respect to RMSD; time-points described in table 3.3) was identified, extracted and the system was back-mapped to all-atomic form using Backward program (Wassenaar et al. 2014). The back-mapped system was used as starting point for another two all-atomic simulations.

Additional information-Tools used for analyses:

- ***trjconv*** (GROMACS 4.6.7) for removing PBC artifacts (Berendsen et al., 1994; Pall et al., 2015)
- ***trjcat*** (GROMACS 4.6.7) for concatenating trajectories (Berendsen et al., 1994; Pall et al., 2015)
- ***Residues considered for PCA***:C-alpha atoms of residues 7-26, 46-63, 82-101,106-124,134-157, corresponding to TM regions 1-5 respectively (Table 3.1). Definitions of TM boundaries based on TSPO orientation with respect to membrane, retrieved from OPM database (Lomize et al., 2006, Lomize et al., 2012)
- ***ProDy*** for PCA, filtering of representative motions of PC1-PC3, projection of conformations onto the subspace of PC1 and PC2 (Bakan et al., 2011)
- ***g_covar*** and ***g_anaeig*** (GROMACS 4.6.7) for PCA and filtering of trajectory consisting of motions of PC1-PC3 (Berendsen et al., 1994; Pall et al., 2015)
- ***g_rmsf*** for calculation of RMSF of filtered trajectory (Berendsen et al., 1994; Pall et al., 2015)
- ***Analyses of helical rotations***: Rotations of TM helices were calculated using Simulaid program (Mezei et al., 2010). In particular, rotation of helices along the axis of membrane (z-axis) was calculated. Calculations were performed on the TM helices 1-5, corresponding to the residues 6-36, 46-69, 76-100,104-126 and 131-157 (corresponding to boundaries of TM helices of mouse NMR structure). For calculation of rotations, residues 21, 55, 89, 115 and 135 of TM 1-5 respectively were considered as midpoints.
- ***Analyses of secondary structure***: Secondary structure conservation of TSPO trajectories in different conditions were performed using do_dssp program of GROMACS (Berendsen et al., 1994; Pall et al., 2015), using DSSP program version 1.0 (Kabsch and Sander, 1983)

Box 3.4. Additional information relating to the preparatory steps and PCA of AA simulations.



Box 3.5. Flow chart describing steps followed for building TSPO protein structural network.

Additional information: Tools used for analysis (DCCM, protein structural networks)

- ***DCCM:*** DCCM was constructed using Bio3d MD analysis package (Grant et al., 2006). DCCM constructed on the C-alpha atoms of residues 6-36, 46-69, 76-100, 104-126, 131-159, corresponding to TM helices 1-5. The matrix was filtered to exclude correlations between residues in contact. The filtering was performed from protein contact map, with cutoff length between C-alpha atoms below which given pair of atoms are considered as a contact was 15Å
- ***Mapping of evolutionary conservation***: Conservation was mapped onto the NMR structure of mouse TSPO using Consurf web server (Glazer et al., 2003, Ashkenazy et al., 2016). Evolutionary conservation of TSPO was calculated from multiple sequence alignment of 500 TSPO sequences from eukaryotes, bacteria as well as archaea, and normalized conservation scores were extracted. Since the conservation values vary at the local level depending upon the buriedness of the residue, five-point running average of the score was calculated for each residue and average scores were plotted.
- ***Analysis of PK11195 dynamics***: Dynamics of PK11195 ligand were analysed for individual trajectories by calculation of RMSD on all the atoms of PK11195 (including hydrogens). The nature of TSPO-PK11195 interactions were studied based on interatomic distances of PK11195 and TSPO atoms on the representative structures of each simulation, using Lifplot program (Wallace et al., 1996). The representative structures were extracted by RMSD-based clustering method (RMSD cutoff above which trajectory frames are grouped in different clusters being 3Å)
- ***Protein structural network & tracing allosteric dynamics***: Protein structural network was constructed using Bio3D MD analysis package (Grant et al., 2006). Allosteric dynamics between residues of two communities having strongest connection were traced using the same tool, and visualised using WISP program (Van Wart et al., 2014)
- ***Calculation of pairwise interaction energies, interaction matrix, structural network***: Pairwise interaction energies (VdW as well as Coulombic potentials) were calculated using GROMACS 5.1.1 and GRINN programs (Berendsen et al., 1994, Pall et al., 2014, Serçinoglu & Ozbek, 2018). Interaction energies were calculated only for the residues whose centres of mass were within distance of 20Å for greater than 60% of the ensemble, and minimum distance between the residues were less than 15Å.

  The interaction energies were projected as a matrix and protein structural network was calculated from the matrix using GRINN program (Serçinoglu & Ozbek, 2018), and correlations of interaction energies (correlation of the fluctuations of energies in the ensemble) were calculated. The cutoff energy above which correlations were calculated is 1 kcal/mol. The residue betweenness centralities were compared from the protein structural network constructed using interaciton energy matrix and DCCM. Upon confirmation that there is significant correlation between the centrality plots, the allosteric pathway was traced using the same tool.

Box 3.6. Additional information describing tools used for construction of protein structural networks

3.2. RESULTS

3.2.1. The structure of mouse TSPO in a bilayer environment is stable but slightly differs from the NMR structure with DPC micelles.

While the first atomic structure of TSPO provided a significant insight into the structural biology of TSPO, subsequent studies on bacterial TSPO by means of X-ray crystallographic studies have found that there are large differences between the conformations of mouse NMR structures and bacterial X-ray crystallographic structures (RMSD of 4.90Å between PDB id 4RYQ vs. 2MGY; on the C-alpha atoms of 148 residues common to both the structures)(Guo et al. 2015; Li, Liu, Garavito, et al. 2015; Li, Liu, Zheng, et al. 2015a). Though the identity between both the TSPO is relatively low (25%), it was hypothesised that use of strong detergent dodecylphosphocholine (DPC) may have contributed to the high difference in RMSD(Li, Liu, Garavito, et al. 2015).

This gives rise to the question, is the structure observed from NMR studies stable, or if it may be an unstable conformation as a result of interaction with DPC micelles? We first addressed the question of stability of the structure when placed in a bilayer environment. We chose DOPC lipids because they are found in a Lα fluid phase at room temperature and their hydrophobic thickness (~27 Å)(Pan et al. 2008) is compatible with that of the TSPO TM domain (28.6 Å) as defined in OPM database(Lomize et al. 2012; Lomize et al. 2006) (Table 3.4). The orientation of the mTSPO/PK11195 complex in the DOPC bilayer taken from OPM database is shown in Figure 3.1(Lomize et al. 2012; Lomize et al. 2011; Lomize et al. 2006).

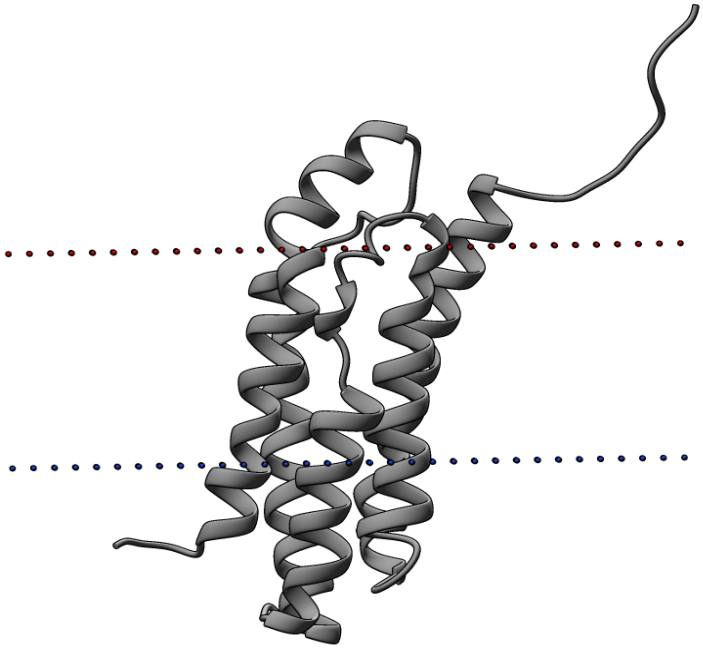


Fig. 3.1. Orientation of TSPO in membrane.
The orientation is adopted from OPM database, which predicts membrane boundaries of different transmembrane proteins.

From measurement of helical rotation angles, we observed significant changes compared to the initial NMR structure (see below for a detailed description). As an example, the rotation angles of each TM helix around its own helical axis show clockwise or anticlockwise variations as large as 90° with respect to the initial orientations, depending on replicates (Fig. 3.2). On average, the largest changes are observed for TM3 (20°), TM1 and TM5 (15°). As a result of these helical rotations, the residues facing the lipids or the protein interior may change. Though bilayer itself may contribute to these changes, in absence of results obtained with a different lipid environment, it is difficult to disentangle its proper role from the intrinsic dynamics of the protein.

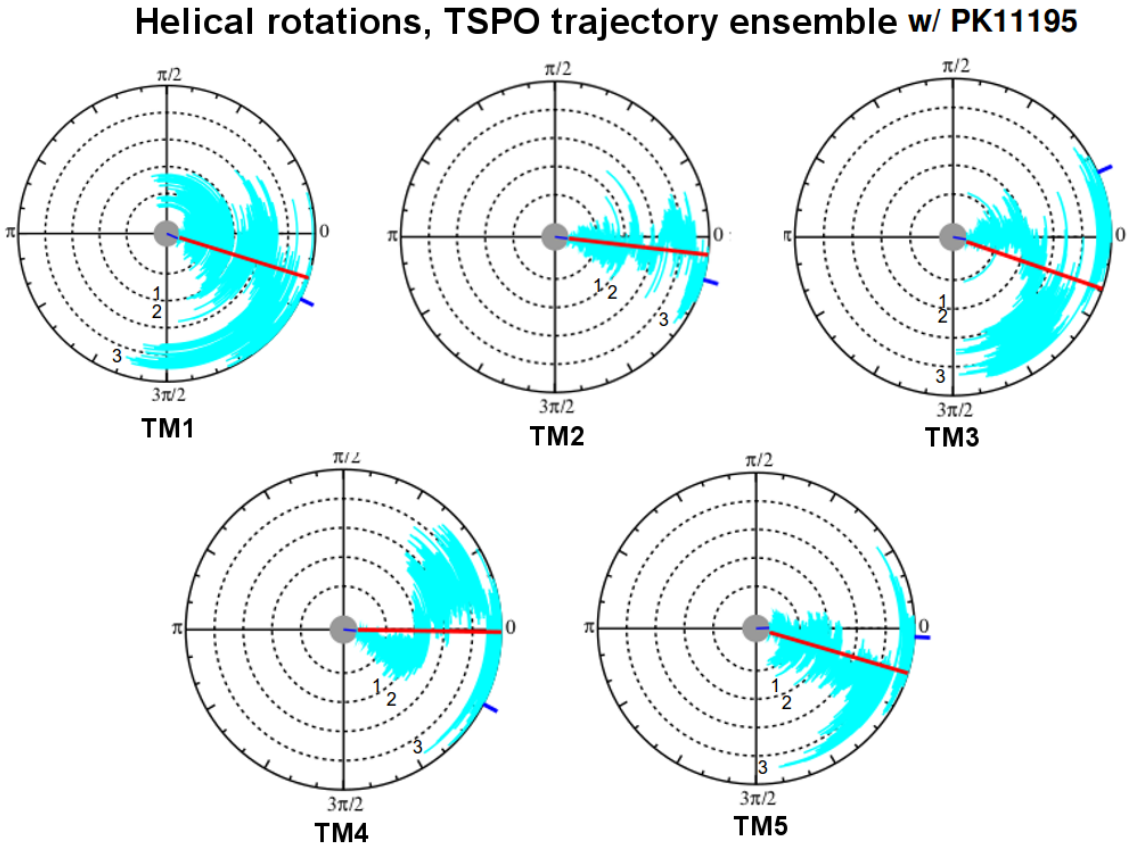


Fig. 3.2. TSPO in lipid bilayer undergoes helical rotations.
Plot showing rotation angles of TM helices of TSPO trajectory ensemble with PK11195. Instantaneous rotation angle is represented by cyan line, average rotation is represented by red line, rotation angle at time t=0 is represented by navy blue tick at the centre of the plot, rotation angle at the end of the simulation is represented by navy blue tick at the periphery of the circle. The time-points where the replicate simulation ends are marked as respective numbers in the plot Rotation angles were calculated for helices using TRAJELIX tool of Simulaid MD analysis package(Mezei 2010; Mezei and Filizola 2006), whose boundaries were defined from the TM helical boundaries of NMR structure (residues 6-36, 46-69,76-100,104-126,131-159 for TM helices 1-5 respectively).

In summary, while we observed large deviations in helical rotations, the average rotation angle is relatively small despite the fact that rotation angles for certain cases were as high as 90° (TM1, TM3 & TM5, Fig. 3.2), suggesting that the overall fold is conserved.

For the ensemble of simulations, the Cα-RMSD for the whole protein with respect to the NMR structure is very large, extending from 6 to 9.5 Å (Fig. 3.3). These values arise from large displacements of the intra and extracellular regions, in particular the extracellular C-terminus end. When restricted to the sole TM domain as defined in Table 3.1, the values are significantly reduced, ranging from 2 to 3 Å. (Fig. 3.3).

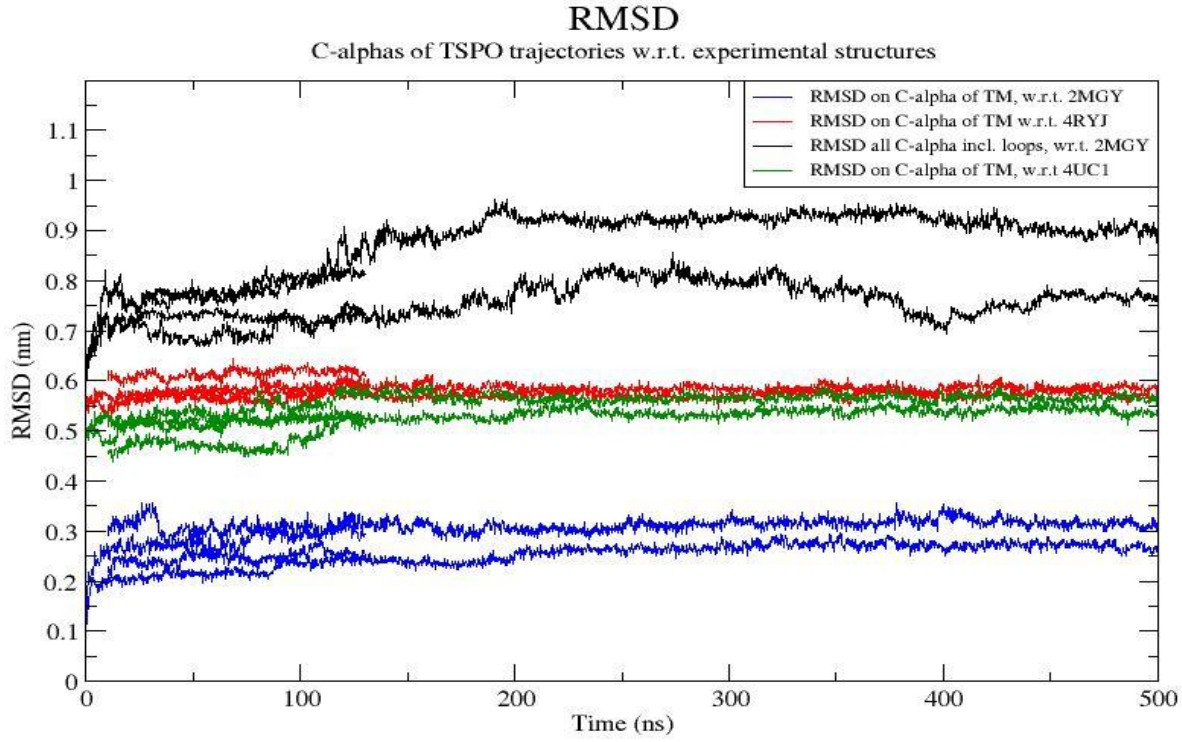


Fig. 3.3. TSPO conformations have high RMSD with respect to experimental structures.
RMSD was calculated on the TSPO conformations from simulations in presence of PK11195 with NMR structure of mouse TSPO as reference (PDB: 2MGY), as well as X-ray structure of *R. sphaeroides* and *B. cereus* as reference (PDB: 4UC1 and 4RYJ respectively). One set of calculations were performed on C-alpha atoms of residues 6-157 (Proteins including N- and C- termini) and the other set of calculations were performed on C-alpha atoms of residues of TM domain (described in Table 3.1).

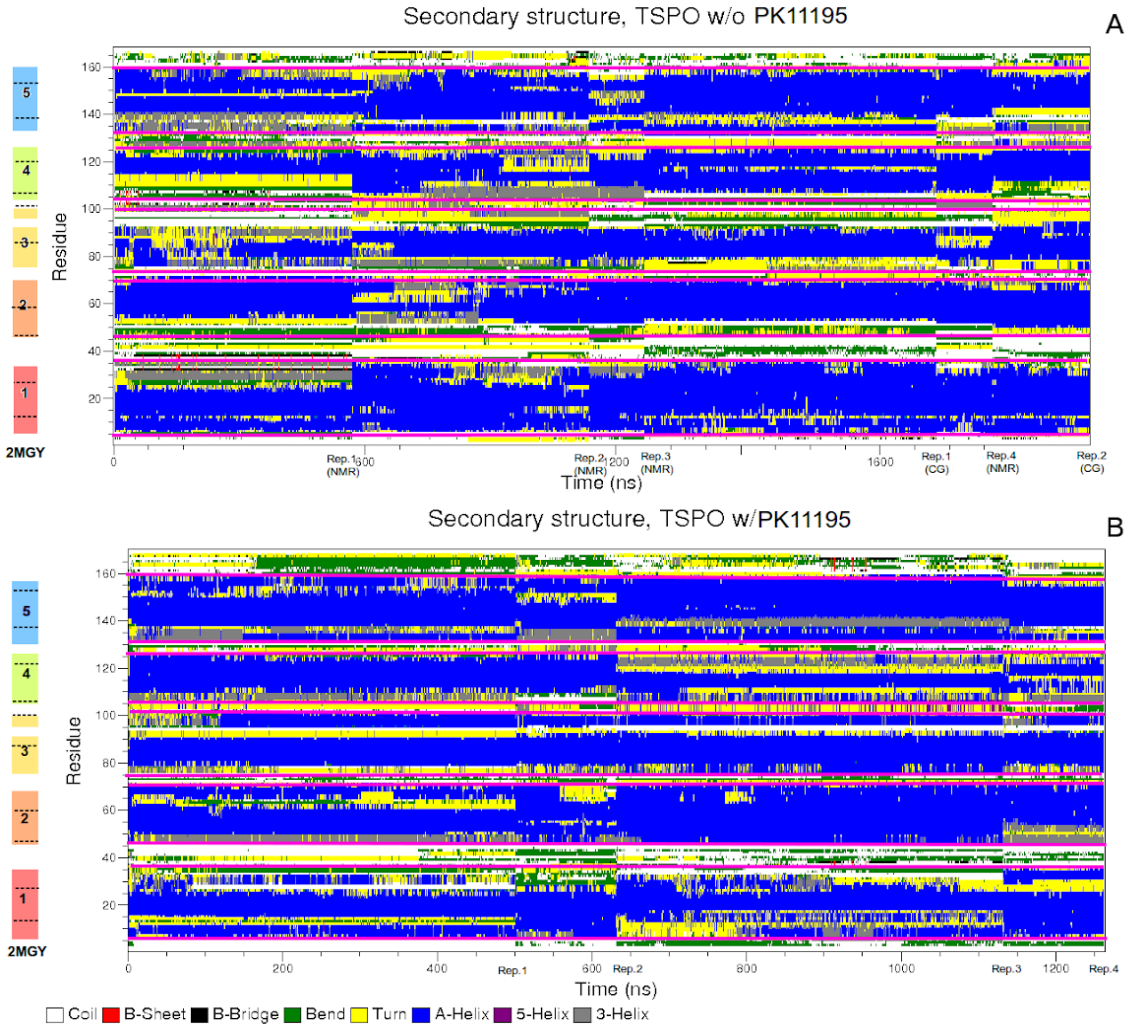
Additionally, we have noted that the conformations observed in X-ray experiments for two bacterial sequences are not necessarily visited during the dynamics of the mouse sequence. Indeed, the Cα-RMSD values between the conformations of the TM domain of mTSPO/PK11195 along the dynamics and that of Rs- and Bc-TSPO X-ray structures are still large, ranging from 4.5 to 6.2 Å.

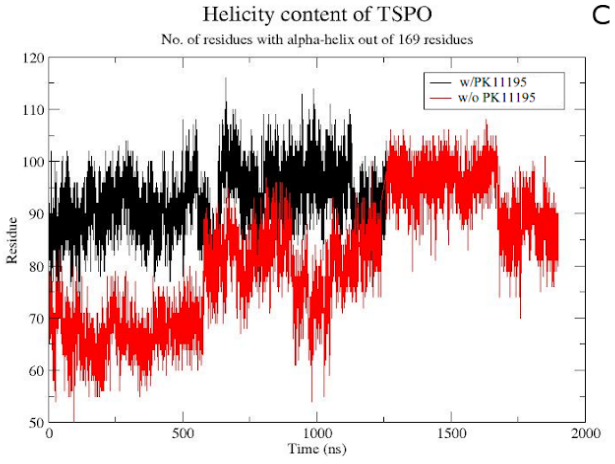
The next question we addressed has concerned the role of the ligand itself. Indeed, most of the experimental structures and the NMR structure as well, haves been solved in the presence of PK11195 molecule. In the NMR structure, the ligand is initially located in a cavity mainly delineated by hydrophobic residues (e.g. Trp, Leu, Phe) that belong to the five TM helices. In order to evaluate the putative stabilizing role of the ligand on the mouse structure, we performed MD simulations in DOPC without ligand and compare structural and dynamics properties in the two situations, *i.e.* with and without PK11195. The results of this comparison are detailed in the next sections.

3.2.2. PK11195 ligand impacts TSPO secondary structure

The TSPO secondary structure variation without and with ligand are illustrated in Figure 3.4. On average, TSPO in absence of PK11195 is found to have lower alpha-helical content with an average number of alpha-helical residues of 83 over 169 residues (49%) as compared to TSPO in presence of PK11195 (54%). The corresponding standard deviations are double in absence of PK11195, which illustrates the flux in protein dynamics, even in the most structured regions.

Regions whose secondary structure was regained upon presence of the ligand include residues 25-35 of TM1, C-terminal region of TM3 (residues 90-100), N-terminal region of TM4 (residues 104-110) and residues 130-140 and 155-160, of TM5, the latter being a part of the cholesterol-binding CRAC motif. These regions are highly dynamic and may regain or lose their secondary structure temporarily. These MD results are in broad agreement with experimental results that showed a stabilizing effect of PK11195 binding on TSPO secondary and tertiary structure(Murail et al. 2008). Interestingly, the regions sensitive to ligand binding in MD simulations were also identified with the solid-state NMR studies(M. Jaremko et al. 2015)**.** Indeed, these studies have shown that certain TM regions, which include TM3, C-terminus of TM-1, and certain residues of CRAC motif undergo disruptions in the secondary structure in the absence of ligand.





(***legend on the next page***)

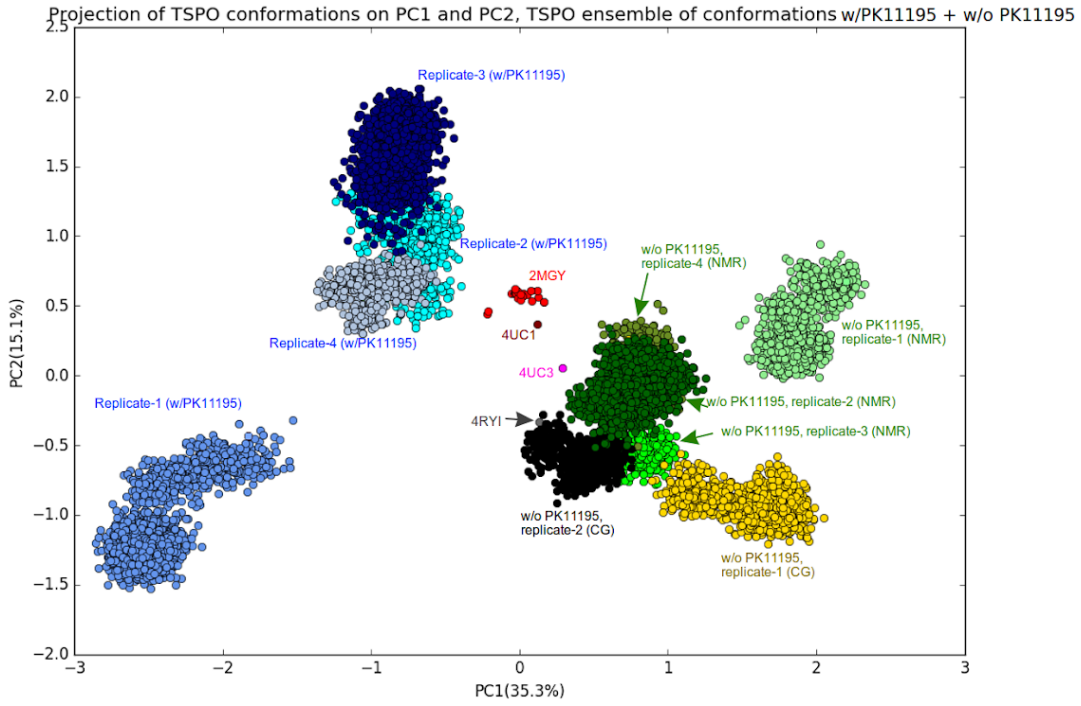
Fig. 3.4. PK11195 binding stabilized TSPO secondary structure.
Secondary structure analyses of TSPO versus simulation time, in absence of PK11195 ligand **(A)**, and in presence of PK11195 ligand **(B)**. TM helical regions of NMR structure of mouse TSPO (PDB: 2MGY) is represented as coloured bars adjacent to the plot. The markers on y-axis (Rep.1, Rep.2...) represent the time-point where the individual replicate trajectory terminates and the next one begins. Approximate boundaries of TM helices in NMR structure (PDB: 2MGY) are represented as pink lines superimposed on (A) and (B).
C: Helicity content of TSPO versus time. Average number of residues with alpha helices for TSPO without PK11195 was found to be 83 residues, with standard deviation of 12.53, and average number of residues with alpha helices for TSPO with PK11195 was found to be 93 residues, with standard deviation of 6.07. Secondary structure calculations were performed on simulation trajectories using DSSP program(Kabsch and Sander 1983)) within GROMACS software(Berendsen et al. 1995; Páll et al. 2015).

3.2.3. PK11195 impacts global dynamics of TSPO

As we observed a stabilization effect of PK11195 on TSPO secondary structures,, we examined thoroughly how the *dynamics* of the protein was impacted. Since principal component analysis (PCA) highlights main directions of motions and helps to identify representative conformations, we performed PCA on the MD conformations. We first analyzed conformations as a whole set regardless of the presence, or absence of the ligand. When considering all the residues except the N-and C-terminal loops, (residues 1-6 and 158-169), PC1 to PC3 for this core region (herein referred to as "core") cumulatively account for 60.1% of the total covariance. The distinction between the two main sets of conformations, namely with and without PK11195 is not straightforward, but most holo conformers are distributed on the right half part of PC1/PC2 plane while the apo conformers are mostly spread over the left half part. Interestingly, when focusing on the TM-domain as defined in Table 3.1, the contribution of the three first PCs to the total covariance slightly increases to 61.5%. Importantly, the projections of the whole set of TSPO conformations on the PC1/PC2 space show that holo conformations are now clearly separated from apo conformations along the PC1. The second component PC2 traps features that seem to be distinct from ligand binding consequences.

In summary, the presence of PK11195 has an influence on the conformational sampling of the TM domain and on the dynamics of the protein. Additionally, the extent of sampling exemplified by the surface area covered by the TSPO conformers appears smaller in presence of PK11195 as compared in absence of PK11195 (Fig. 3.5). Clustering of the conformations onto the PC1-PC2 space suggests that the apo-conformations are more loosely clustered than the holo conformers. Thus, it seems that PK11195 tends to restrict the dynamics of the protein, or at least trap the protein in a local minimum before it may escape and visit alternative conformations. This may explain why the presence of PK11195 is required to solve the structure, as observed experimentally.

A



B

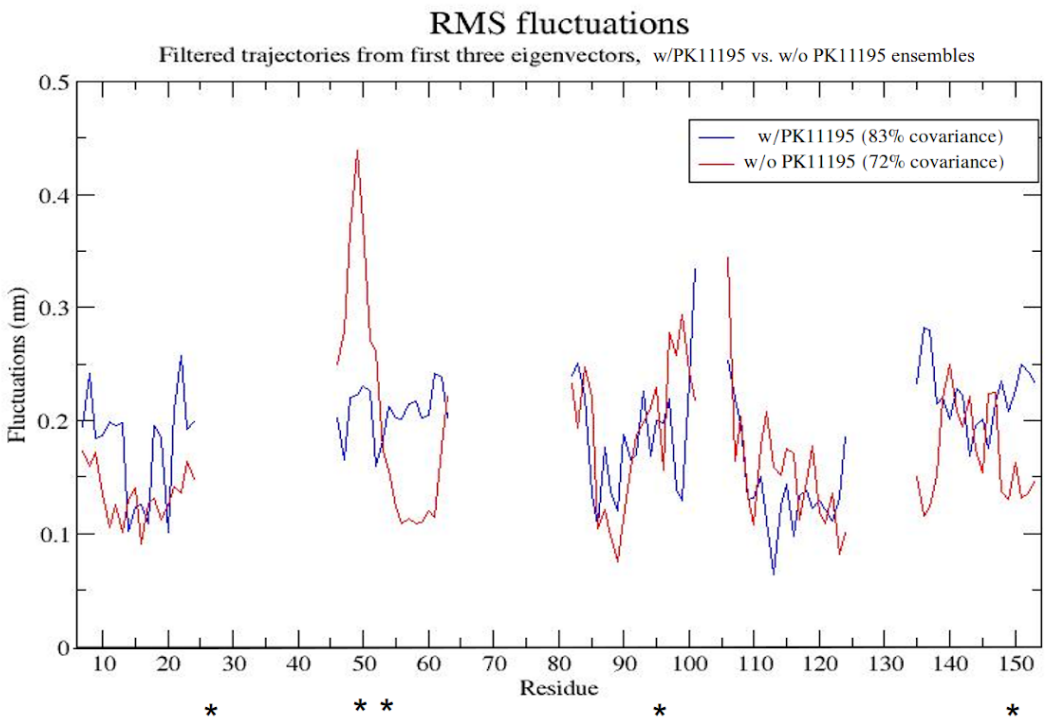


Fig. 3.5. PK11195 influences TSPO dynamics. (*full legend on next page*)

A: Projections of conformations of TSPO trajectory ensemble consisting of conformations in absence, presence of PK11195, and conformations from the NMR structure, on the first (34.6% of covariances) and second (14.1% of covariances) eigenvectors. The green points on the plot correspond to TSPO conformations in absence of PK11195 from simulations where NMR structure of TSPO was used as starting structure, yellow and black points on the plot correspond to TSPO conformations in absence of PK11195 from simulations where back-mapped representative structure of Coarse-Grained conformations were used as starting conformations. Grey point corresponds to X-ray structure of *B. Cereus* (PDB: 4RYI), Pink point corresponds to X-ray structure of *R. sphaeroides* (wild type) and Maroon point corresponds to the X-ray structure of *R. sphaeroides* (A139T mutant). The plot was generated from the projection of TSPO confornations on conformational space of first and second Principal Components, using Prody MD analysis package(Bakan et al. 2011).
B: Root Mean Square Fluctuations (RMSF) of filtered TSPO trajectories of first to third principal components in presence versus in absence of PK11195. The trajectories corresponding to first three principal components were filtered using GROMACS program(Berendsen et al. 1995; Páll et al. 2015). PK11105 interacting residues are marked '*' below y-axis.

In a second step, we separately examined the main motions performed by TSPO in holo and apo states by calculating PCA of TSPO trajectories for each state and quantified similarities in directions by calculating the scalar product of PCs obtained from the two sets of simulations.

The two largest eigenvectors calculated for the core TM domain account for 61.6 %, and 67% of the covariance for the apo and holo form, respectively. These values increase to 72% and 83% respectively, when third largest eigenvector is also considered Trajectories were further filtered along these three largest eigenvectors. Overall, the RMSF of the resulting trajectories (Fig. 3.5, B) does not show major differences in TSPO fluctuation profiles in either of the states. Significant differences are only observed between the two states for residues that are located around the binding cavity *i.e.* they are less flexible when the ligand is removed. Yet, some residues show larger fluctuations in presence of PK11195, e.g. residues 6-8 in TM1, residues to 55-62 in TM2, and residues 134-138 and 150-157 of TM5 (Fig. 3.5, B). Interestingly in the last case, residues that show larger differences belong to the consensus motif 'CARC' that has ability to bind to cholesterol(Fantini et al. 2016; Papadopoulos et al. 2018) and to the well-known cholesterol-binding 'CRAC' motif described experimentally(Li, Yao, et al. 2001; Li and Papadopoulos 1998)(Fig. 3.5,B). Importantly, apart from the amplitude of dynamics, the direction of dominant of motions, *i.e.* PC1 (accounting for 30% and 53% of total dynamics in TSPO in absence and presence of PK11195 respectively), significantly differ between the two states (Fig. 3.6). While residues around the PK11195 binding cavity are more dynamic in absence of ligand, which may stem from the loss of secondary structure of these residues, the most conspicuous difference comes from the dynamics of TM2. This region has global dynamics in presence of PK11195 ligand (Fig. 3.6, B), whereas such dynamics are absent in this region in the absence of PK11195 ligand. Apart from TM2, the region corresponding to the residues 140 to 146 are dynamic in

absence of PK11195, whereas the surrounding regions, including CRAC motif as well as putative cholesterol binding CARC motif become dynamic in presence of PK11195 (Fig. 3.6, B).

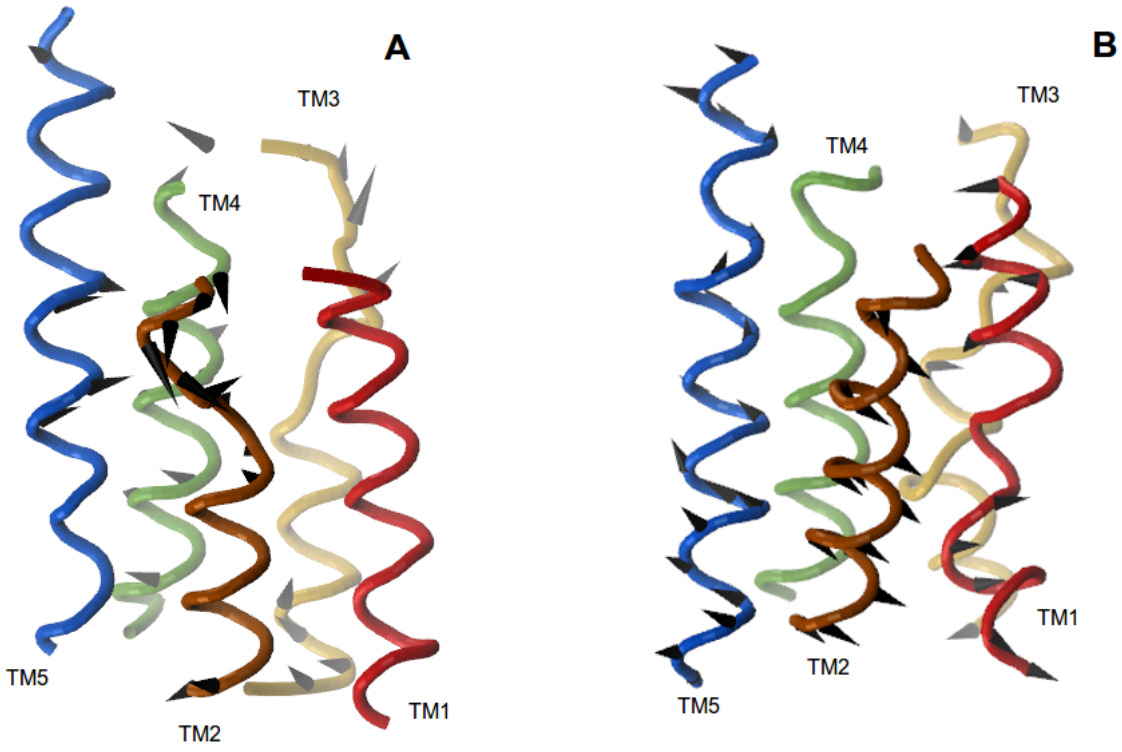
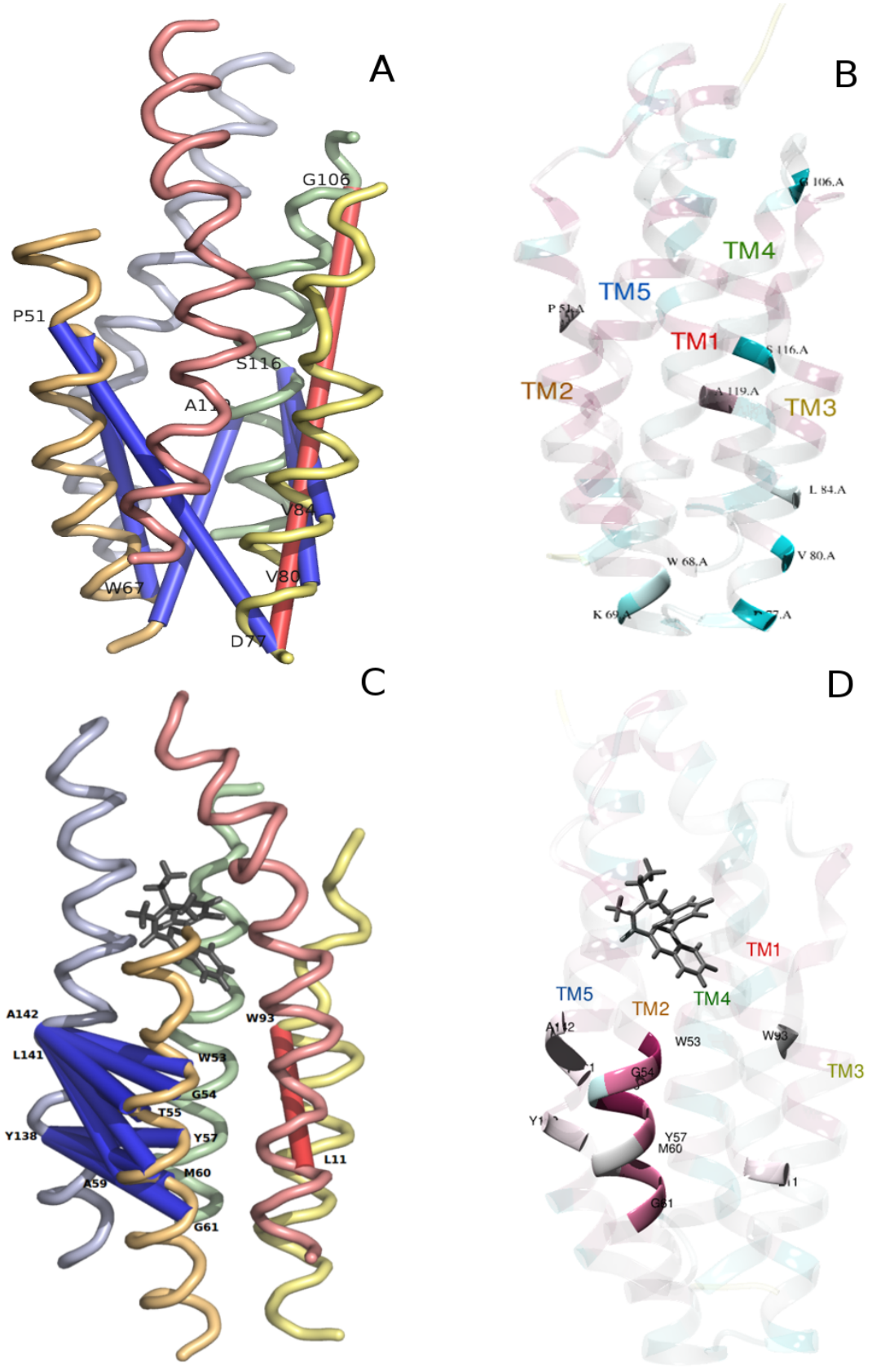


Fig. 3.6. TM regions of TSPO have different dynamics in holo- and apo- forms.
Porcupine plot showing dynamics of PC1 of TSPO in absence of PK11195 ligand (**A**) versus dynamics of PC1 of TSPO in presence of PK11195 ligand (**B**). The PC1 in both the conditions account for 30% and 53% of total dynamics respectively. Residues having fluctuations > 2.0Å are represented with arrows showing directions of their fluctuations.

Since PK11195 binding appears to activate global dynamics, particularly in TM2 region, we attempted to answer the question of if such dynamics involve correlated motions with other regions. Consequently, to understand this phenomenon, we calculated the Dynamic Cross Correlations Matrix (DCCM) on Cα atoms of the TM domain for the ensemble of trajectories in from each state. The strongest correlations between the corresponding residue pairs were projected onto the initial NMR TSPO structure.



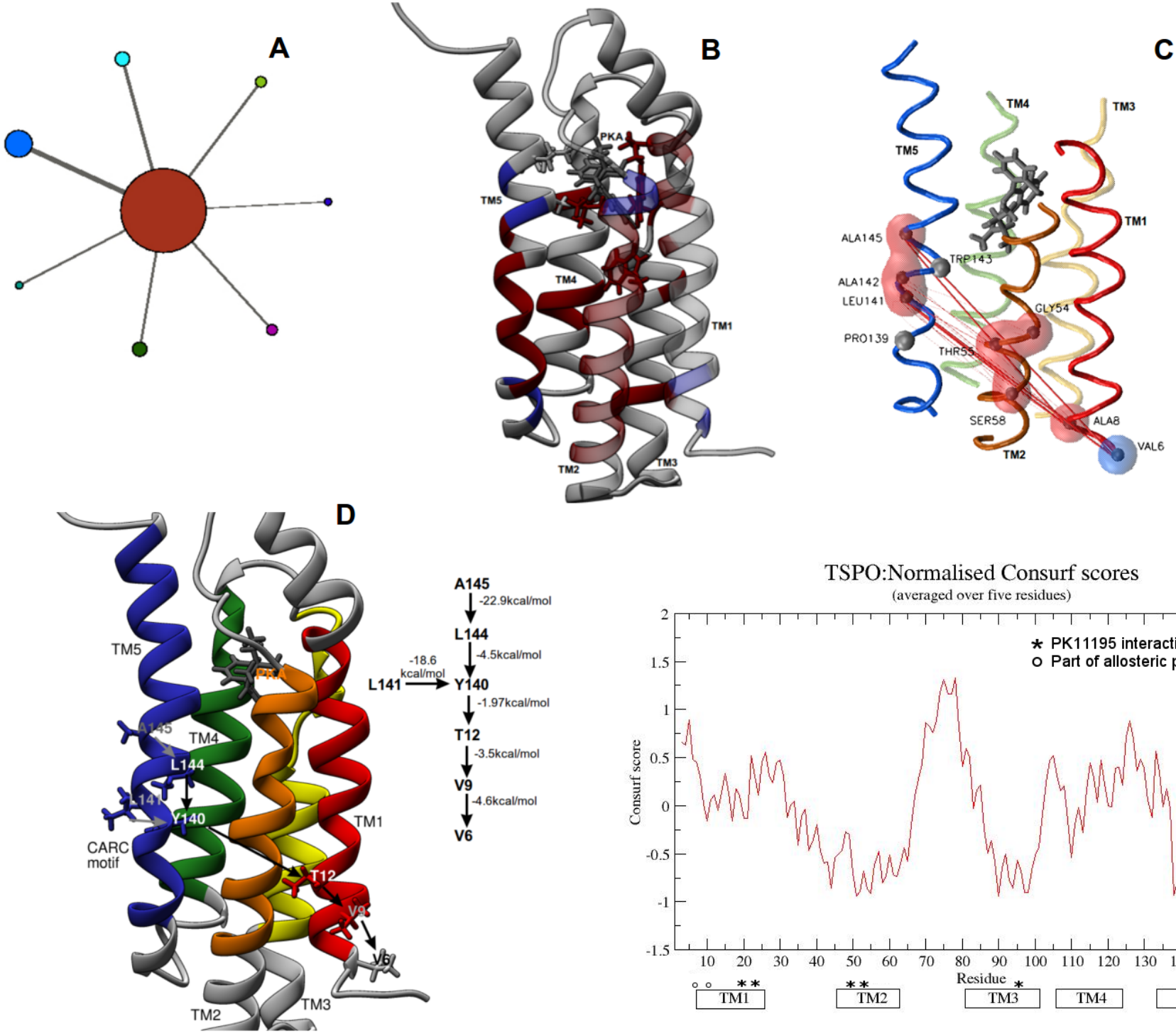
(***legend on the next page***)

Fig. 3.7. Long range correlated motions between conserved residues are activated in TSPO in complex with PK11195.
A: Representations of long-range residue motions having high correlations in TSPO without PK11195. Blue cylinders represent correlations -0.7, red cylinders represent correlation +0.7.
B: Evolutionary conservation of residue pairs having motions with high correlations in TSPO without PK11195.
C: Representations of long-range residue motions having high correlations in TSPO with PK11195. Blue cylinders represent correlations -0.76, red cylinders represent correlation +0.76.
D: Evolutionary conservation of residue pairs having motions with high correlations in TSPO with PK11195. Matrix containing correlation values of residue motions was calculated using Bio3D MD analyses package(Grant et al. 2006), and residue pairs with highest correlation was projected onto the NMR structure of TSPO, and visualised using Pymol(DeLano 2002). TM helices 1 to 5 are represented as red, orange, yellow, green and blue tubes respectively, PK11195 ligand is represented as grey sticks. Conservation scores of TSPO residues were generated using Consurf web server(Glaser et al. 2003; Ashkenazy et al. 2016), and projected onto NMR structure of TSPO using UCSF Chimera(Pettersen et al. 2004). Cyan-coloured residues represent highly variable residues, magenta-coloured residues represent highly conserved residues.

3.2.4. Long-range correlated motions are activated in TSPO bound to PK11195:

The DCCM for each state is given in Fig. 3.7. In absence of PK11195, we observed that the residue pairs P51-W67, P51-D77, W68-A119, S116 with V80 and V84 had correlation value of -0.7, and the residue pair D77 and G106 had correlation value of +0.7 (Fig. 3.7,A). These residues are localised on the TM2, 3 and 4 regions. Most of these residues, with the exception of A119 were found to be highly variable in terms of evolutionary conservation (Fig. 3.7,B).

On the other hand, in presence of PK11195, the residue pairs Y138 with Y57-G61, L141 with T55, and A142 with W53-S58 and M60 have correlation value of -0.76, and the residue pair L11 and W93 have correlation value of +0.76. The residue pairs having negative correlation are located at TM2 and TM5, and the pairs having positive correlation are located at TM1 and TM3 regions (Fig. 3.7, C). Contrary to TSPO without PK11195, many residues having strong correlations in their motions also have strong evolutionary conservation, particularly, the residues W53, L56, Y57, M60 and G61 (Fig. 3.7, D; Consurf scores of -1.49, -1.39, -1.17, -1.03 and -1.08 respectively). Furthermore, the residues Y138 to L141 in TM5 (Consurf scores of -0.29 and-0.26 respectively) are part of the possible cholesterol-binding CARC motif(Fantini et al. 2016; Papadopoulos et al. 2017). The fact that these residues are well conserved might mean that such motions may also be conserved across TSPO of various species.



(legend on the next page)

Fig. 3.8. ***TSPO residues are part of a protein structural network in presence of PK11195 ligand, and an allosteric communication pathway was identified involving TSPO residues.***
A: Schema showing protein structural network of TSPO in presence of PK11195.
B: TSPO structure showing residues part of the central node (coloured red) and largest peripheral node (coloured blue). PK11195 interacting residues are represented as sticks.
C: TSPO structure showing allosteric communication pathway among the residues of TM2, TM5 and TM1, deduced from correlation matrix of residue motions. Five shortest suboptimal paths were traced between the residues of central node and the residues of largest peripheral node (represented as transparent red and blue spheres respectively). The suboptimal paths are represented as red lines between the residues. The width of the line represents length of the path, with bold lines representing shortest paths.
D: TSPO structure showing the allosteric communication pathway, deduced from correlation matrix of non-bonded interactions of TSPO residues. The shortest path between the residue A145 to V6, and residue L141 to V6 were traced. The non-bonded interaction energy between participating residues is depicted in the flowchart adjacent to the figure (**D**).
E: Plot showing normalised conservation scores of TSPO residues, represented as running average over five residues. Since the absolute score varies between adjacent residues depending on the solvent accessibility of the residue, running averages were calculated. TM regions are indicated as horizontal bars below the *y-axis*, PK11195-interacting residues are represented as '*' symbol, and residues involved in allosteric communication are represented as '°' symbol.

3.2.5. The long-range motions correspond to a possible allosteric pathway from the conserved to variable regions: Since the presence of strong correlated motions suggest presence of possible dynamics of allosteric nature, we constructed structural network of TSPO trajectory ensemble in presence of PK11195, based on the correlation matrix of residue motions. We found that the residues of TM2 and TM5 that have strong correlations in their motions (Fig. 3.7, C) are part of a single central community (Fig. 3.8, A, coloured red). This community has strong connection in particular with community comprising of residues 6-9 of TM1 (Fig. 3.8, A, coloured blue). We also projected the residues of both the communities onto the structure of TSPO, and we observed that many of the binding residues (with the exception of A147 and L150) are part of the central community (Fig. 3.8, B). Apart from the PK11195-interacting residues, residues of TM2 and residues of N-terminal part of TM5, certain residues of TM1 (residue A8), TM3 and TM4 are part of the central community (Fig. 3.8, B). On the other hand, the primary peripheral node comprises of residues of N-terminal of TM1 (residues V6 and V9), residues of TM2 (R46 and W47) and TM5 (F146 and T148) (Fig. 3.8, B). Since the central community includes majority of PK11195-binding residues, we considered this community as the source of allosteric dynamics, and the residues of peripheral community as the sink of the allosteric dynamics.

We traced the five shortest sub-optimal paths of allosteric communication between the residues of central community and residues of the largest peripheral community, which is also the community with the strongest connection with the central node (Fig. 3.8, C). The shortest suboptimal path (Fig.

3.8, represented as thick red cylinders) involves residues of TM5 and terminates at the N-terminal of TM1 (A145 to V6 and L141 to V6, Fig. 3.8, C). Apart from the two shortest suboptimal paths, we also observed longer suboptimal pathway involving the residues of TM2, TM5, to TM1 comprising of the residues V6 to V9 (Fig. 3.8, C). We also were interested to know if there are differences in the evolutionary conservation of the residues of two communities, since the residues of central community also include many PK11195-interacting residues. Interestingly, we found that many residues of the central community, including the residues of TM5 involved in the pathway as well as PK11195-interacting residues have strong average conservation scores (Fig. 3.8, E). In contrast, we observed that the residues of peripheral community, particularly those at the N-terminal of TM1 have weak evolutionary conservation scores (Fig. 3.8, E). Since we performed protein structure network analysis on the residues of TM helices, rather than the TM regions defined by OPM database, we performed similar analysis on the residues of TM regions defined by OPM database (Lomize et. al., 2006; Lomize et. al., 2012) using same parameters. On comparison of both the allosteric pathways, we found minor differences in the pathway. In the pathway constructed on the TM boundaries defined by OPM database, the pathway passed through residues G18 of TM1, in addition to the residues of TM5 as well as TM1 (Annexure-2).

While the above results describe the allosteric communication pathway from protein dynamics perspective, we also attempted to describe the same from residue interaction perspective. For that purpose, we calculated matrix of non-bonded interactions between all the residues, and constructed a protein energy network based on this matrix. To ensure that there is similarity between the networks obtained from residue motions versus the residue interactions, we calculated betweenness centralities for both the matrices, and we calculated Pearson's correlation between both the data. We found low correlation value of 0.19 between both the data (Fig. 3.9). The primary reason for low correlation is the fact that we have constructed protein structural network based on motions of C-alpha atoms, whereas we calculated protein energy network based on interaction energies of all the atoms in the residues. We attribute low correlation between the data to this factor. However, we observed that the correlation in betweenness centralities for the residues of TM1 and TM5, where the allosteric pathway was observed is relatively high compared to the correlation (0.32, Fig. 3.9), suggesting that allosteric dynamics observed from residue motions between these regions may also be observed from the residue interactions. Using the protein energy network, we traced the shortest pathway between the TM5 residues (L141 and A145) and TM1 residue V6. We observed that the shortest pathway between L141 and V6 involves the residues Y140, T12 and V9, with distance of 3.48Å and the average interaction energy of -7.1kcal/mol (Fig. 3.8, D). Similarly, the shortest path between the residue A145 and V6 constitute the residues L144, Y140, T12 and V9, with distance of 4.3Å and the average interaction energy between the residues of the pathway being -7.5kcal/mol (Fig. 3.8, D).

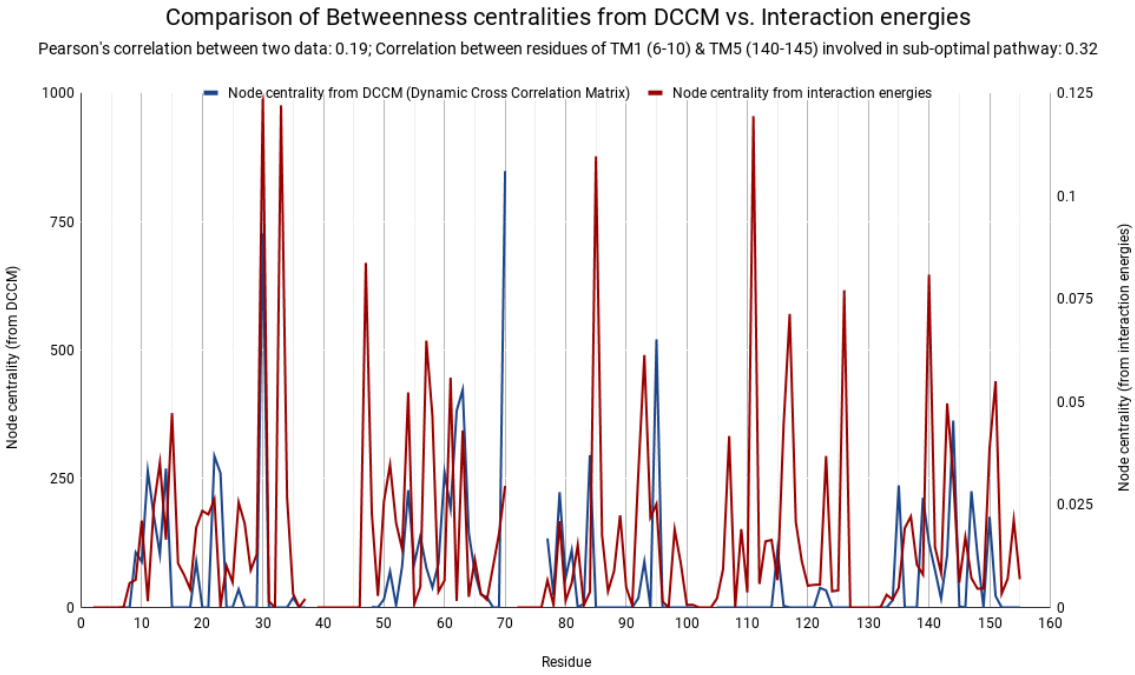


Fig. 3.9. Plot showing comparison of betweenness centralities of protein structural network of TSPO in presence of PK11195 (determined from correlations of residue motions), versus protein energy network of TSPO in presence of PK11195, determined from non-bonded interaction energy matrix.

3.2.6. PK11195 adopts a wide range of conformations when it is bound to TSPO:

Since we noticed significant changes in dynamics of TSPO residues in the holo state, we examined the dynamics of the PK11195 ligand itself. We calculated RMSD of the PK11195 ligand, and also visualized its motions when it is bound to TSPO. We found that RMSD itself is high for PK11195 ligand in all the replicates, suggesting that PK11195 ligand itself is dynamic, even when it is in complex with TSPO (Fig. 3.10).

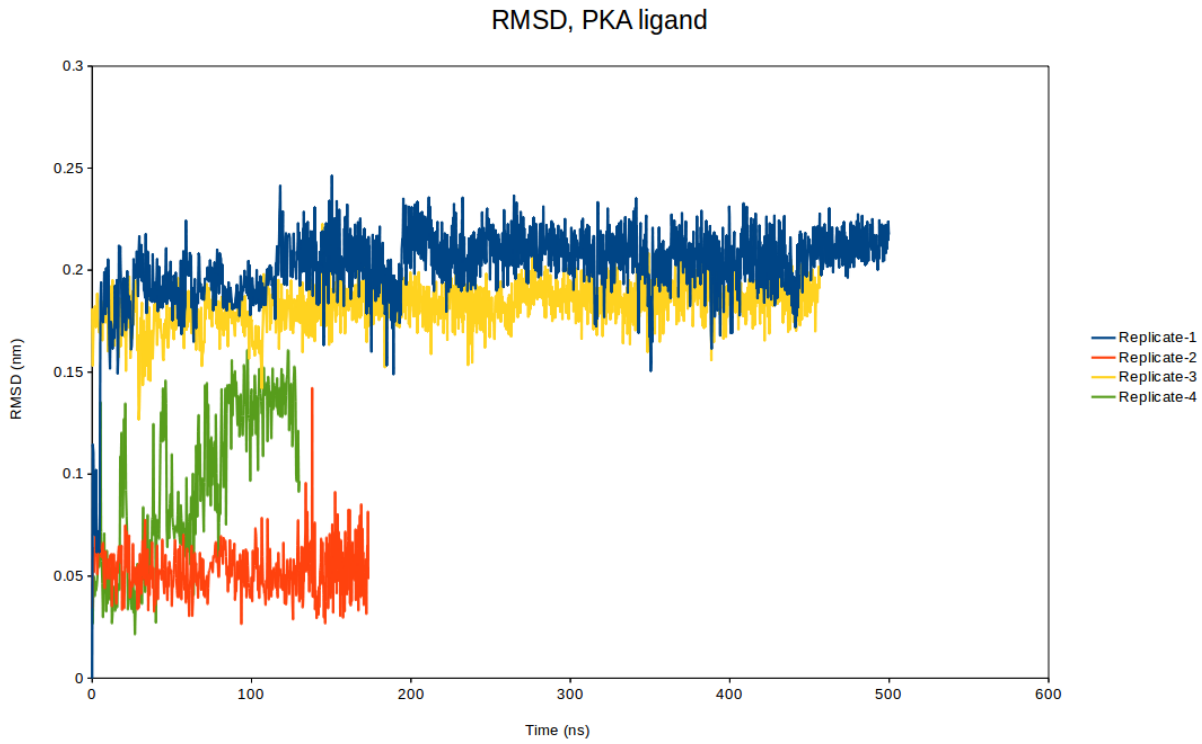


Fig. 3.10. PK11195 is dynamic in complex with TSPO.
Plot showing RMSD of PK11195 ligand in four replicate trajectories of TSPO-PK11195 complex. RMSD was calculated using g_rms tool of GROMACS. To understand the nature and extent of dynamics of the PK11195 ligand bound to TSPO. RMSD of the PK11195 ligand was calculated and plotted for the replicate trajectories of TSPO-PK11195 complex, using g_rms tool of GROMACS.

Two replicates, replicate-1 and replicate-3 were found to have high RMSD (>1.5Å). To understand the reason for the high RMSD, we filtered representative structures of TSPO from each replicate simulation using RMSD-based clustering method, and we visualized conformations of PK11195 from representative structures. Both the replicates having high RMSD were found to adopt Z-rotameric form, as opposed to E-rotamer form found in the PK11195 in complex with the NMR structure (Fig. 3.11). This suggests that PK11195 can exist in either of the rotameric form in complex with TSPO. While RMSD of PK11195 for three of the replicate trajectories reached towards a plateau, RMSD for the fourth replicate trajectory experienced significant changes (Fig. 3.10). Interestingly, we found that in this replicate simulation, PK11195 has transitioned from E-isomer to Z-isomeric form (Fig. 3.11, 3.12), although significant energy transition barrier exist between these two states. Note that we carefully checked PK11195 force field parameters by comparing energies associated with amide dihedral rotations calculated with our parameters with that of quantum mechanics calculations available in the literature(Lee et al. 2012) (Fig. 3.13).

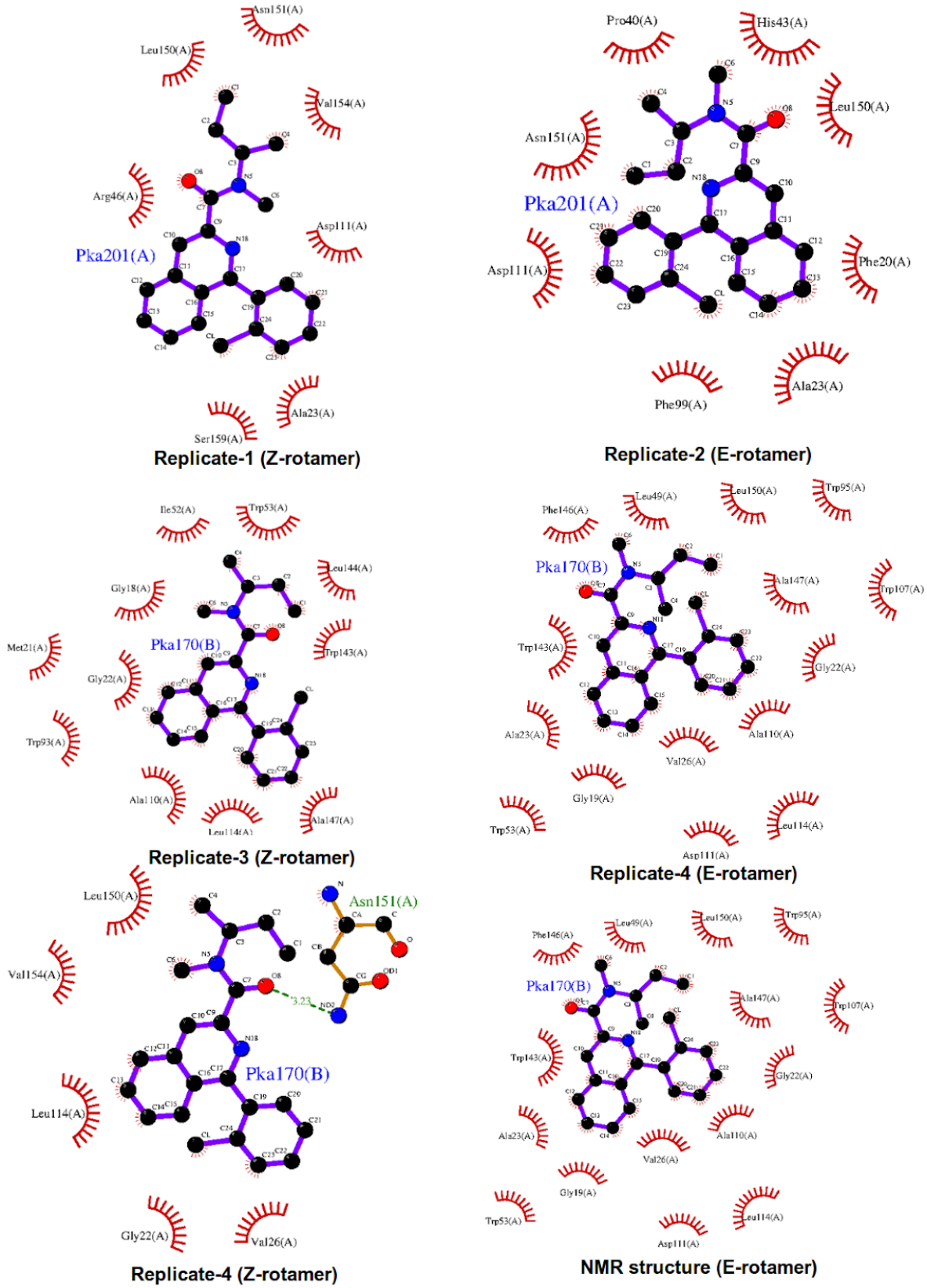


Fig. 3.11. TSPO-PK11195 interactions are hydrophobic in nature.
Schema showing TSPO-PK11195 interactions from different simulations, along with the NMR structure of mouse TSPO (PDB: 2MGY). The interactions are mapped from the representative TSPO-PK11195 structures from each simulation based on interatomic distances, using Ligplot

program(Wallace et al. 1995). The representative TSPO structure was filtered using g_cluster RMSD-based clustering tool of GROMACS(Berendsen et al. 1995; Páll et al. 2015) (Cutoff: 3Å).

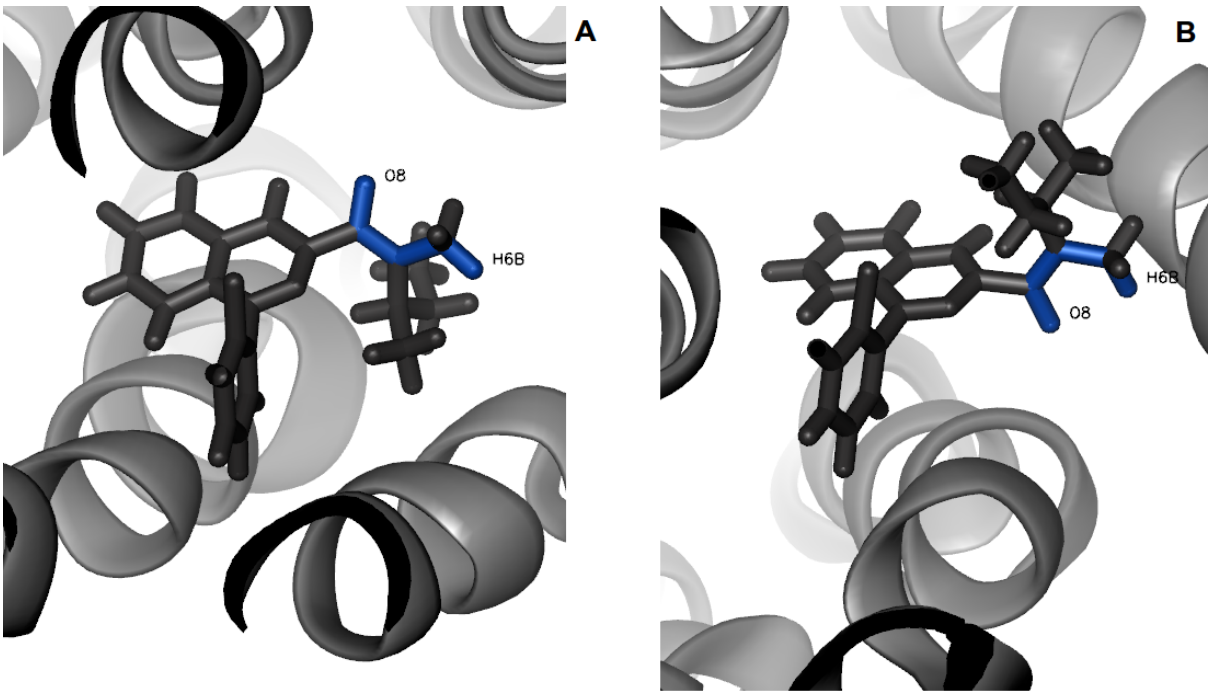


Fig. 3.12. PK11195 transitions between E- and Z- rotamer forms.
Snapshots of TSPO simulation-4 in presence of PK11195, showing transition between E-rotamer form (**A**) at the beginning of the simulation to Z-rotamer form (**B**) at the end of the simulation. The atoms involved in the dihedral rotation are represented as blue sticks.

Despite high RMSD, many TSPO residues that initially border the binding cavity remain in contact with PK11195 for most of the simulations. However, we observed that the PK11195 itself does not have a preferred orientation in the binding cavity, and switches between various conformations, depending on the sampling (Fig. 3.11, 3.12).

Ultimately, we attempted to characterize the nature of interactions between TSPO and PK11195 ligand. To do the same, we filtered representative structures of TSPO from each replicate simulation using RMSD-based clustering method, and we interpreted the nature of TSPO-PK11195 interactions from representative structures based on interatomic distances. We found that TSPO-PK11195 interactions are largely hydrophobic in nature (Fig. 3.11), with one exqception in replicate-4, which has a hydrogen-bond interaction between carbonyl moiety of PK11195 and asparagine residue of TM5 (Fig. 3.11). Apart from that, we have also observed PK11195 ligand in both E- and Z- rotameric forms in different replicate, indicating that PK11195 can exist in both the rotameric forms in complex with TSPO. The interplay between protein residues fluctuations and PK11195 motions was further examined and is detailed in the next section.

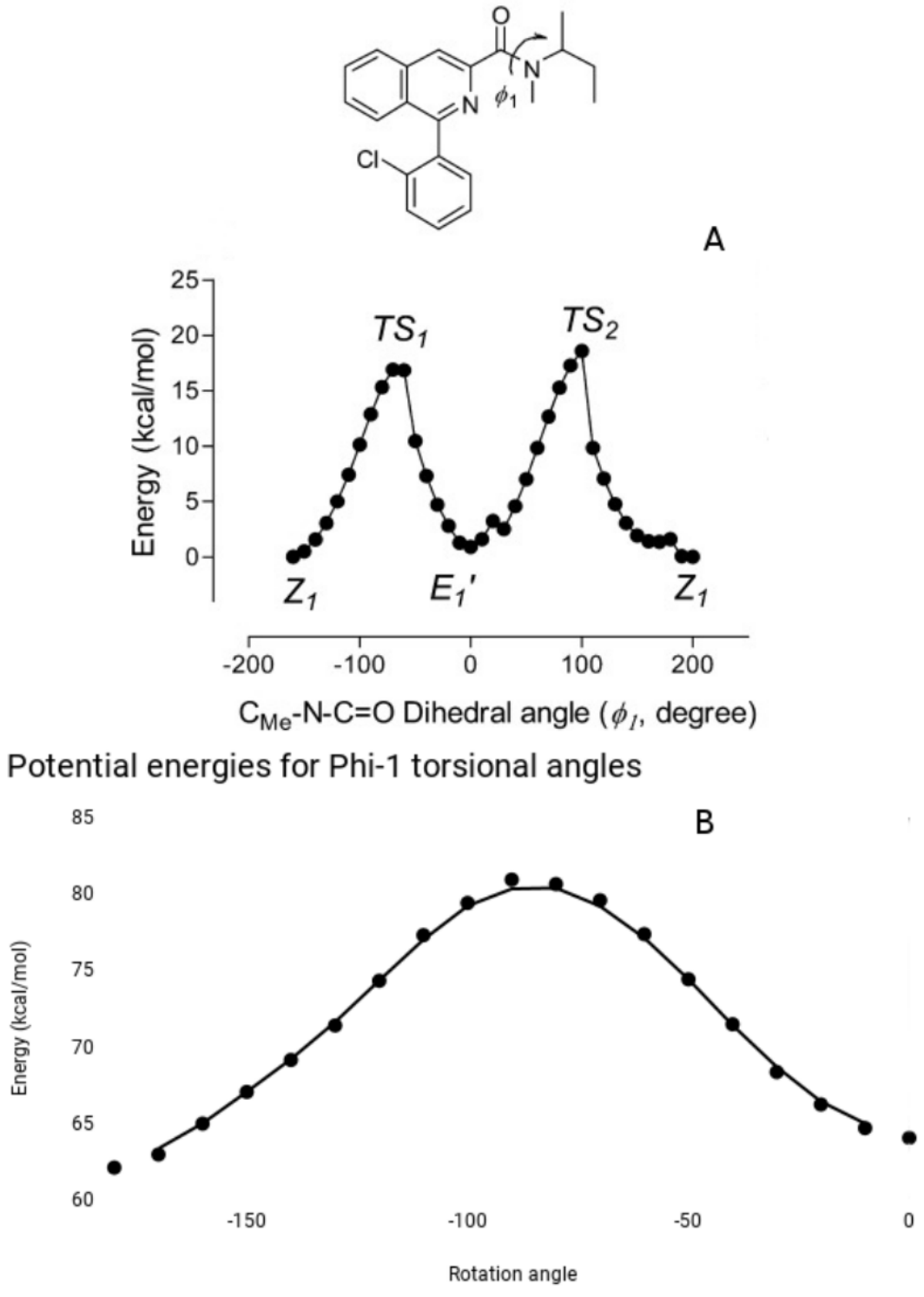


Fig. 3.13. A: Plot showing rotational energies (ϕ1) of CMe-N-C=O dihedral angle of PK11195, adapted from Lee et. al.(Lee et al. 2012)

B: Plot showing rotational energies (ϕ1) of CMe-N-C=O dihedral angle of PK11195, calculated by steepest descent energy minimization method, using GROMACS program(Berendsen et al. 1995; Páll et al. 2015).

3.2.7. PK11195-interacting residues around the binding cavity have strong evolutionary conservation:. We studied the TSPO-PK11195 interactions by calculating the number of residues at the vicinity (<0.5nm) of PK11195 ligand, which may be classified as a contact. We observed that some residues in contact with PK11195 in the solution-NMR structure of mouse TSPO [28] (Fig. 3.14) remain in contact for >90% of the total simulation time (1.2μs). This is consistent with earlier observations that PK11195 binds to TSPO with high specificity(Delavoie et al. 2003). Importantly, we observed that most of these residues, with the exception of V26, also have strong evolutionary conservation (Table 3.5). The residues having highest conservation were L150 (conservation score of -1.52), W53 (-1.49), A147 (-1.489) and W95 (-1.44). We also observed that two of the CRAC motif residues, L150 and N151 are in contact with PK11195 for 98.1% and 97.5% of the simulation time respectively (Table 3.5).

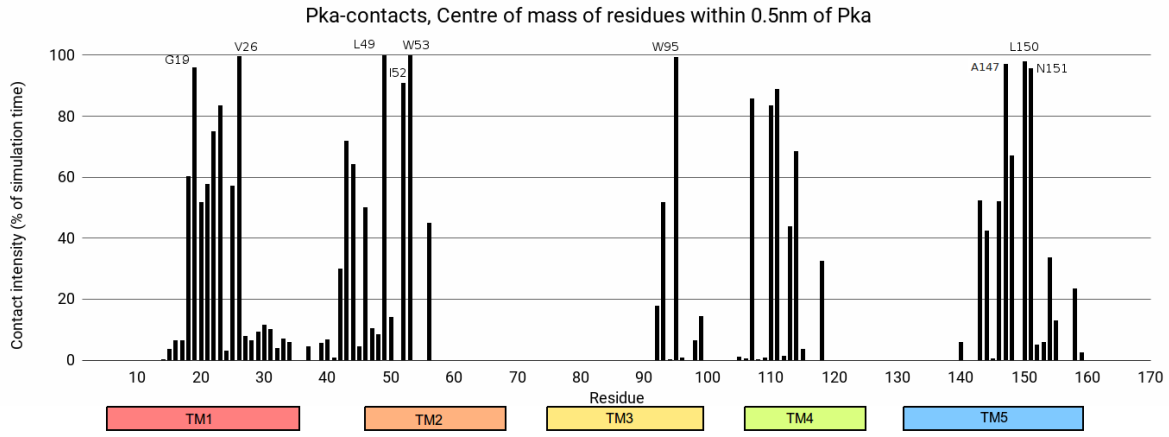


Fig. 3.14. Certain residues have dominant interactions with PK11195.
Plot showing interaction-time of TSPO residues with PK11195 ligand as percentage of time in contact, out of total simulation time (1.23μs) for TSPO in presence of PK11195 ligand. Residues were considered to be in contact with PK11195 if their centre of mass was within 5Å distance of PK11195. Residues in contact with PK11195 for >90% of simulation time are labelled. Residue-PK11195 contacts were calculated using GROMACS program(Berendsen et al. 1995; Páll et al. 2015)

Due to the fact that PK11195 is an exogenous drug molecule, we hypothesized that PK11195 and PP-IX may share a common binding site. To test this hypothesis, we compared the binding sites of PK11195 and PP-IX ligands in TSPO structures of mouse and *R. sphaeroides* (identity of 36% between sequences of both the structures) respectively by performing structural alignment of both the structures. We observed that indeed, both the ligands share a common binding site (Fig. 3.15). Furthermore, from the pairwise alignment of TSPO sequences, we observed that out of 3 residues that bind to PK11195 as well as PP-IX, two of them are identical in both bacterial as well as mouse sequence (Fig. 3.15, B).

Residue	Position	Consurf Score	% contact time
L	49	-1.296	100
W	53	-1.49	99.83
V	26	-0.718	99.68
W	95	-1.44	99.5
L	150	-1.52	98.14
A	147	-1.489	97.02
G	19	-1.123	95.92
N	151	-1.266	95.78
I	52	-0.929	90.86

Table 3.5. Table showing residues having contact (<5.0Å distance of PK11195) with PK11195 ligand for >90% of total simulation time with PK11195 (1.23µs), and corresponding conservation scores. Residue-PK11195 contacts were calculated using GROMACS program, conservation scores were calculated using Consurf web-server(Glaser et al. 2003; Ashkenazy et al. 2016).

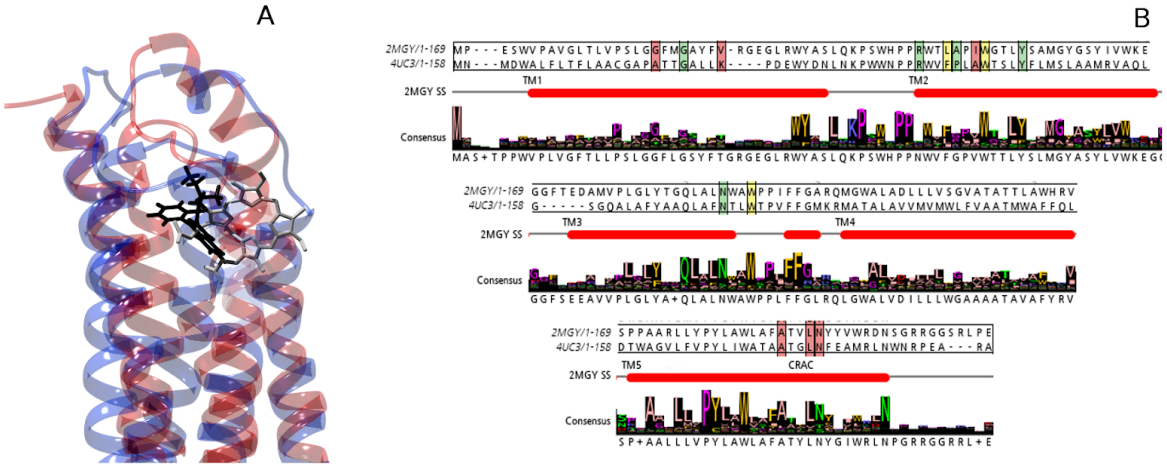


Fig. 3.15. PK11195 and PP-IX share a common binding cavity.
A: Structural superposition of structures of mouse TSPO (PDB ID: 2MGY) and *R. sphaeroides* TSPO (PDB ID: 4UC1), showing common binding site of PK11195 and PP-IX ligands, represented as black and grey sticks respectively. Sequence alignment of mouse and *R. sphaeroides* TSPO sequences showing conservation of PK11195 and PP-IX interacting residues. Residues interacting with PK11195 are shaded red, those interacting with PP-IX are shaded green and residues interacting with both PK11195 as well as PP-IX are shaded yellow.

3.3. DISCUSSION

3.3.1. Common binding sites of PK11195 and PP-IX ligands

We compared the binding of PK11195 in mouse TSPO with binding of Protoporphyrin-IX (PP-IX) in *R. sphaeroides* TSPO, to investigate if there exists any commonality between the binding cavities, as the structure of TSPO in complex with PP-IX is available. We performed superposition and structural alignment on NMR structure of mouse TSPO (PDB: 2MGY) and X-ray crystallographic structure of

R. sphaeroides TSPO (PDB: 4UC1), with which PP-IX ligand is bound. From this structural alignment, we observed that indeed, PK11195 and PP-IX share a common binding cavity (Fig. 3.15). We also tried to understand if the conservation in binding cavity translates to conservation of binding residues, and we found that out of three residues that interact with PK11195 as well as PP-IX, two were identical in both mouse and *Rhodobacter* sequences (residues W53 and W95, Fig. 3.15, B). This suggests that both these residues in mammalian TSPOs may also interact with PP-IX was well. Since both the residues interact with PP-IX as well as PK11195, these residues may be subject of competitive inhibition of PP-IX binding by PK11195, which has been supported from earlier experimental studies(Wendler et al. 2003; Guo et al. 2015). From these results, we identified the structural basis for inhibition of inhibition of PP-IX binding by PK11195 ligand.

3.3.2. Dynamics of PK11195 ligand

While preceding studies on TSPO structure have discussed extensively about TSPO structure, there has been very little discussion about the dynamics of the ligand PK11195 itself, despite the fact that this ligand is among the first choice for performing experiments with TSPO, and its immense pharmaceutical value. From studies on the NMR structure of TSPO, since PK11195 ligand was in complex with TSPO in the E-form, it was inferred that this may be the primary isomeric state of the ligand when in complex with TSPO(Jaremko, Jaremko, Giller, et al. 2014). We observed that the hydrophobic interactions dominate in the TSPO-PK11195 complex (Fig. 3.13). This is not unexpected since the PK11195 binding cavity is dominated with hydrophobic and aromatic residues. From our simulations, we found that not only PK11195 ligand is dynamic when in complex with TSPO meaning that it can rotate and translate within the cavity (Fig. 3.10), but the ligand is also seen to undergo change in conformation from E- to Z- form (Fig. 3.11). To ensure that the PK11195 parameters used in our simulations are able to reproduce the energies observed with the torsional rotation of amide scaffold of PK11195 ligand, whose rotation involves change in E- to Z- form, we compared the potential energies associated with different dihedral angles of this scaffold with the energies observed from quantum mechanics calculations(Lee et al. 2012). Since we observed broad agreement with the energies from quantum mechanics calculations (Fig. 3.12), we infer that the PK11195 parameters used in our simulation are valid. With that, we find change in the isomer state while in complex to be interesting, since most of the experimental studies describe high affinity and specificity of PK11195 binding to TSPO. The physiological and pharmacological effects of this change in isomeric state however, remains to be seen if some effects exist..

3.3.3. Helical rotations of TSPO

Comparison of mouse and bacterial experimental structures revealed large differences(Li, Liu, Garavito, et al. 2015; Li, Liu, Valls, et al. 2015; Guo et al. 2015), though the corresponding sequence identities (36% between mouse and *R. sphaeroides* and 22% between mouse and *B. cereus*) suggest

that they may have more similar conformations. Some studies attributed this disparity in conformations to possible differences in conditions where NMR structure of mouse TSPO was solved, where it was suggested that due to use of strong detergents like DPC, certain mouse TSPO residues become exposed that are otherwise buried(Li, Liu, Garavito, et al. 2015). To understand if such disparities persist in our simulations, or if they are unique to the structure, we computed helical rotations of every TM helix from the TSPO ensemble in presence of PK11195.

Though we found large differences for TM helices 1, 3 and 4 (Fig. 3.2), the average rotation angle for all the TM helices remained relatively low throughout the ensemble. The high rotation angles were in turn observed during certain time-points of the ensemble, which correspond to specific replicates, suggesting that the rotations may be dependent on the sampling. The observed differences between bacterial and mouse TSPO conformations observed from experiments and our simulations could be attributed to the following reasons: (i) the fact that TSPO is intrinsically dynamic, or (ii) due to different experimental conditions having trapped different conformations, (iii) or both. We suggested that comparative MD simulations of both bacterial and mammalian TSPO could be performed to address these hypotheses (discussed in detail in chapter-5).

3.3.4. Effect of PK11195 ligand on TSPO dynamics

Since earlier studies have pointed towards a central role of PK11195 in stabilization of secondary and tertiary structure of TSPO(Murail et al. 2008; Jaremko, Jaremko, Giller, et al. 2014), how exactly the ligand influences TSPO dynamics is not known. One attempt has been made by Jaremko et. al.(Ł. Jaremko et al. 2015) to study TSPO in absence of PK11195, and it was observed that many residues did not elicit significant resonance signals. Furthermore, many residues for which significant signals were detected were found to have reduced helicity content. In terms of secondary structure, our results are in broad agreement with the observations made by Jaremko et. al.(Ł. Jaremko et al. 2015). We observed that TSPO in presence of PK11195 has higher overall helicity content of 93 (54%), in contrast to lower helicity content of 83 (49%) in absence of PK11195 (Fig. 3.4, A&B). We also observed that TSPO in absence of PK11195 has larger variations in secondary structure as opposed to TSPO in presence of PK11195. This is particularly evident from the fact that the standard deviation for average helicity content is higher for TSPO without PK11195 (12.53 residues), compared to TSPO with PK11195 (6.07 residues). We infer that while TSPO in absence of PK11195 experiences loss of secondary structure in some regions, it also has flexibility to explore wider set of secondary structural conformations.

Regions observed to have distortions in secondary structure, *viz.* the residues 25-35 of TM1 region, C-terminal region of TM3 consisting of residues 90-100 and certain regions of TM5 namely residues 130-140 and 155-160 from our simulations are similar regions where reduction in helicity content were observed by Jaremko et. al. (Fig. 3.4)(Ł. Jaremko et al. 2015). It is also worth noting that these

regions are also in the vicinity of PK11195 binding cavity. While Jaremko et. al. were unsuccessful in obtaining NMR signals for TM4(Ł. Jaremko et al. 2015), we observed that significant portion of this region undergoes perturbations in secondary structure in absence of PK11195 (Fig. 3.4, A). However, we also noted the following observations: (i) many of the above mentioned regions that had perturbations in secondary structure also regained alpha-helical configuration for certain duration of simulations (Fig. 3.4, A). (ii) there exists certain time points in simulations in presence of PK11195 where these residues, particularly the residues of the C-terminal region of TM1 have perturbations in secondary structure, despite the presence of PK11195 (Fig. 3.4, B). Irrespective of the binding state, we found that some regions, located in the TM core were found to have disruptions in secondary structure. In particular, the regions corresponding to the residues 5-10 (TM1), residues 90-95 (TM3) and residues 135-140 (TM5) had perturbations in secondary structure. This could be attributed to the fact that these regions contain proline residues, which induces kinks in TM helices(von Heijne 1991; Tieleman et al. 2001). Similar observations were made from the solution NMR structure of mouse TSPO as well(Jaremko, Jaremko, Giller, et al. 2014).

While we were able to understand the impact of PK11195 ligand on the secondary structure of TSPO, we further tried to understand its impact on the tertiary structure and dynamics of TSPO, for which we performed PCA on the C-alpha atoms of TM domain.

PCA on the TSPO ensemble containing conformations with, as well as without PK11195 showed conformational explorations in two states. We observed that with respect to conformational space of PC1 (accounting for 33.6% of covariances), TSPO in absence of PK11195 has wider coverage of conformational space, when compared with TSPO conformations of replicates-2, 3 and 4 with PK11195 (Fig. 3.5, A). However, we also observed that the replicate-1 with PK11195 explored conformational space different to those explored by holo form of TSPO from other replicates (Fig. 3.5, A). As this may bias the above observations, we calculated average RMSF of C-alpha atoms of TSPO from all the replicates with PK11195. We found that the residues of TM regions have relatively low variation in RMSF among different replicates compared to the loop regions (Fig. 3.16), suggesting that the differences with the first replicate of TSPO in presence of PK11195 may indeed stem from the possibility that TSPO in this simulation may be exploring a different conformational landscape.

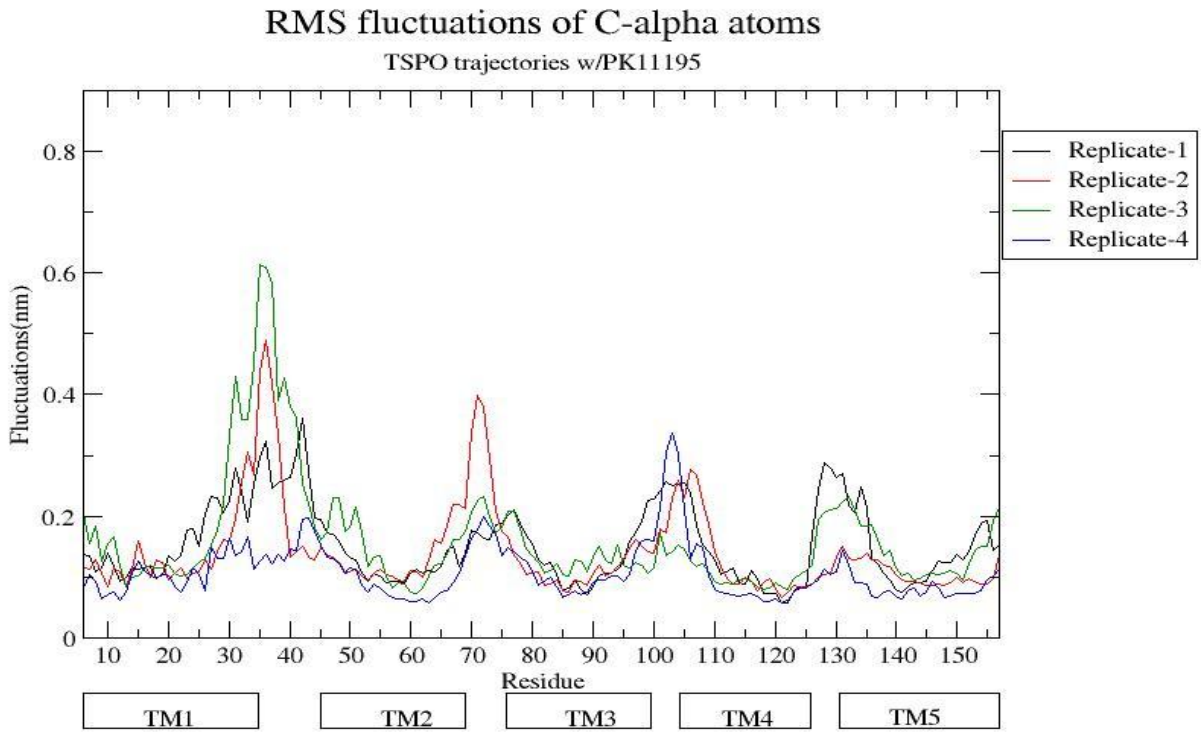


Fig. 3.16. Plot showing RMSF of TSPO simulations in presence of PK11195.

We also calculated RMSF on the filtered trajectories (representing motions of PC1-PC3, accounting for 72% and 83% of covariances respectively) of trajectory ensembles of holo and apo forms of TSPO. We observed that the residues of TM1 (20-30), TM2, TM3 (87-95), TM5 (135-140 and 150-159) have higher fluctuations in presence of PK11195 as opposed to TSPO without PK11195 (Fig. 3.5, B; Fig. 3.6). The exception here is that the residues of TM2 (46 to 52), TM4 (110-120) and TM5 (140-150), which are more dynamic in the absence of PK11195 are rigid in presence of PK11195 (Fig. 3.5, B; Fig. 3.6). Interestingly, the residues of TM5 where the differences in fluctuations are localized, is the region with functional significance. The residues 135-140, which are observed to have higher fluctuations in presence of PK11195 (Fig. 3.5, B), are reported to be a part of cholesterol-binding motif 'CARC' (135-RLLYPYL-141)(Fantini et al. 2016; Papadopoulos et al. 2017). Residues 150-159 are part of cholesterol-binding motif 'CRAC' (149-VLNYYVWR-156)(Li, Yao, et al. 2001; Jamin et al. 2005; Li and Papadopoulos 1998), which has been demonstrated to bind to TSPO with nanomolar affinity(Lacapère et al. 2001). Theoretical studies on both the motifs in TSPO indicated that CARC motif has lower binding energy (-62.4 kJ/mol) compared to CRAC motif (-44.4 kJ/mol)(Fantini et al. 2016). However, while CRAC motif binds to TSPO in nanomolar affinity(Lacapère et al. 2001), there is no experimental evidence yet for binding of cholesterol to CARC motif in TSPO.

On the other hand, residues 140-150, which is found to be more dynamic in absence of PK11195, has a motif consisting of residues 144-LAF-146, which has been found to increase binding affinity of

bacterial TSPO with cholesterol, where this motif is absent(Li, Liu, Valls, et al. 2015). The changes in TM5 region upon PK11195 binding are indeed interesting, though it is difficult to determine what implications do such changes impact TSPO function, in particular, its cholesterol translocation ability. Early experimental studies have suggested that treatment of PK11195 ligand to mitochondria stimulated steroidogenesis(Calogero et al. 1990; Veenman et al. 2007; Barnea et al. 1989), suggesting that PK11195 ligand aided cholesterol translocation. While the structural mechanism for this translocation is unknown, we propose that the increased flexibility of CRAC and CARC motifs observed on TSPO in presence of PK11195 from our simulations may be the basis for PK11195-mediated cholesterol translocation. Since both the motifs have increased flexibility, we propose that this increased flexibility may be accompanied by enhanced accessibility of cholesterol-interacting residues. Future theoretical as well as experimental studies could shed more light into the possible mechanisms.

To explore the possible coordinated motions in both the states of TSPO, we compared the long-range residue motions that have high correlations in TSPO from both the states. We found that in the apo state of TSPO, there are motions among the residues of TM2, 3 and 4, that had correlation of -0.7 (motions of residue pairs D77 and G106 have correlation of +0.7, Fig. 3.7, A). However, we observed that these residue positions are highly variable with respect to their evolutionary conservation (Fig. 3.7, B). In contrast, the residues of TM2 (W53-G61) and TM5 (Y138-A142) have strong correlated motions (-0.76) in the holo state of TSPO (Fig. 3.7, C). Furthermore, these residues also have strong evolutionary conservation (Fig. 3.7, D), which suggests that the motions themselves may be conserved across the TSPO from various phylogenetic groups. Due to the fact that the residues involved in these motions are located at a distance from PK11195 binding cavity (Fig. 3.7, C), and these motions were not observed in TSPO in absence of PK11195, we infer that these motions may be allosteric in nature. In addition, these dynamics in simulations with PK11195 seems to involves an important fraction of the TSPO structure (CARC and LAF motifs of TM5), compared to the one calculated in simulations without PK11195, which are more localized (Fig 3.7, A-D).

Some residues of TM2 that have allosteric motions (G54, T55, Y57, S58 and G61) are exposed to the surface. Of these residues, G54, Y57 and G61 have strong evolutionary conservation scores of -1.063, -1.173 and -1.078. This region may also be involved as interface with other interacting proteins, with which TSPO is observed to interact with. Activation of these motions by PK11195 gives an interesting perspective on TSPO's interactions with its partners, given that PK11195 has been shown to modulate many physiological processes involving TSPO, including catalytic degradation of PP-IX(Guo et al. 2015; Ginter et al. 2013), ROS production(Jayakumar et al. 2002; Gavish and Veenman 2018), regulation of F_0-F_1ATP synthase phosphorylation and thereby respiration(Gavish and Veenman 2018; Krestinina et al. 2009), cell cycle(Gavish and Veenman 2018). Furthermore, PK11195 is also found to

have therapeutic effects in various pathophysiological conditions, including *Plasmodium falciparum* infection in malaria(Bouyer et al. 2011; Marginedas-Freixa et al. 2016), neuroinflammation and brain injury(Rupprecht et al. 2010), reperfusion injury(Morin et al. 2016).

Since the correlated motions in presence of PK11195 were found at a distance from the PK11195 binding site, we hypothesised presence of an allosteric pathway involving residues having correlated motions. To investigate further in this direction, we constructed the protein structural network based on the correlations in the motions of residue pairs, and traced the pathway of allosteric communication from this network. From the structural network, we observed that the residues of TM2 and TM5 form the central node, which are connected to the peripheral communities. Upon tracing the pathway between the central node and three largest peripheral communities, we observed that while the residues of TM2 act as the main source of pathway, the pathway involves the residues of TM5, and terminates at TM1 (Fig. 3.8, C). These results reveal an interesting mechanism by which entropy is transferred across spatially separated residues in presence of a bound ligand, which is also reflected in the changes in the local residue fluctuations (Fig. 3.5, B).

While we observed a possible allosteric pathway from the perspective of protein dynamics, we also investigated if such a pathway exists in terms of residue interaction energies. For this, we attempted to characterize the allosteric pathway from biophysical perspective. We constructed similar pathway from a matrix of pairwise interaction energies between all the residues of TSPO instead of matrix of correlated residue motions. The shortest suboptimal pathway traced from the pairwise interaction energy matrix involved residues L141, Y140, T12, V9 and V6 residues (with length of 3.48Å). Since we observed the presence of allosteric pathway from both biophysical as well as protein dynamics perspective, this is the first study to our knowledge that characterises allostery in proteins from both structural as well as biophysical perspectives.

Similar mode of allosteric transfer, characterized by transfer of entropy between the binding site and allosteric site have been observed in the Ubiquitin system(Hacisuleyman and Erman 2017). Since we observed that the residues of TM2, including PK11195 binding residues L49 and W53 have low RMSF in presence of PK11195 ligand and vice-versa in absence of PK11195 (Fig. 3.5, B), and N-terminal residues of TM1 have high RMSD in presence of PK11195 and vice-versa in absence of PK11195 (Fig. 3.5, B), we infer that the former may constitute source of the allosteric dynamics, and the latter may be the sink, or the endpoint of the pathway, where the source residues transfer the energy towards the sink residues. We have also observed this from the pathway traced based on interaction energies, where there is relatively high interaction energy (-1.97kcal/mol) between Y140 (entropy source) and T12 (entropy sink).

3.3.5. Functional implications of allosteric communication

We make some hypothesis on the possible functional ramifications of the allosteric communication pathway that we traced. The two shortest suboptimal paths were observed between the residues of TM5 and TM1. Residues L141, A142 and A145 were involved in this pathway, and participation of residues L144 and Y140 was also observed from the pathway traced from residue interactions. In TM5, the pathway passes through two regions that have physiological importance: L141 and Y140 are part of the putative cholesterol-binding CARC motif(Fantini et al. 2016; Papadopoulos et al. 2018), and residues L144 and A145 are part of the enhancement motif(Li, Liu, Valls, et al. 2015). However, residues of TM1 that are part of this pathway are not associated with any physiological function till date. Moreover, while TM5 residues have stronger average conservation, TM1 residues have weak conservation and consist of variable residues. We propose two hypotheses as the functional basis of allosteric communication in TSPO:

1. Since this pathway occurs at a distance from the binding site, the TM5 and TM1 residues may be considered as allosteric sites. We propose that this allosteric pathway may modulate protein-protein interactions involving TSPO. Since the physiological significance of TM1 is unknown and N-terminal residues of TM1 are located at the interface of membrane and solvent, we propose that this region may be involved in protein-protein interaction, and the allosteric communication may modulate protein-protein interactions involving TM1. Since N-terminal of TSPO faces intermembrane space of mitochondria, there are a few number of candidate proteins at this region that may interact with TSPO.

One of them is adenine nucleotide transporter (ANT). Activity of TSPO, VDAC and ANT has been characterised in context of apoptosis and plasmodial infection of erythrocytes(Veenman et al. 2008; Veenman et al. 2007; Bouyer et al. 2011; Marginedas-Freixa et al. 2016). Furthermore, PK11195 binding has also been demonstrated to modulate the above processes(Santidrián et al. 2007; Mendonça-Torres and Roberts 2013; Kugler et al. 2008). However, while ANT is shown to interact with VDAC, it is unclear if it interacts with TSPO as well.

Other potential partner is Steroidogenic Acute Regulatory protein (StAR). Together with TSPO, VDAC and other proteins, StAR is shown to form a cholesterol import complex at mitochondria(Rone et al. 2012). FRET experiments have also demonstrated that both the proteins are present in proximity to each other of TSPO with StAR(West et al. 2001). Since TSPO interacts with other proteins via its C-terminal(Li, Degenhardt, et al. 2001), and StAR has been proposed to localize at the intermembrane space(Tsujishita and Hurley 2000; Mathieu et al. 2002), it could be possible that it may interact with proteins via its N-terminal as well. Indeed, since TSPO and StAR are known to be a part of mitochondrial cholesterol import complex(Liu et al. 2006; Rone et al. 2012; Papadopoulos et al. 2007), we propose cholesterol binding by TSPO and possible TSPO-StAR interaction via N-terminal

region may involve this allosteric pathway. However, though TSPO-StAR interaction is demonstrated, the interface of interaction remains unknown. Furthermore, contradictory models of cholesterol transport have also been proposed, that does not involve localization of StAR at intermembrane space(Bose et al. 2002).

2. An alternative hypothesis for this pathway is that it could be a mechanism of stabilization of PK11195 binding site, where the binding site residues are stabilized at the expense of allosteric sites. Studies on Ubiquitin-Human polymerase iota system have found that communication between binding site and allosteric site occurs via a similar allosteric communication, where the entropy is transferred from the 'source' region involving binding site to the 'sink' region involving allosteric site, following Schreiber's entropy transfer mechanism(Hacisuleyman and Erman 2017; Schreiber 2000). In our simulations, we have found that several PK11195-binding residues (particularly at TM2) have lower fluctuations in presence of PK11195. Since the binding site residues as well as TM5 residues that are part of the pathway belong to the same central node (Fig. 3.8, B), we hypothesized that the communication between the central and peripheral node may be a mechanism of entropy transfer. Though we are unable to directly quantify the entropy associated with the residues involved in the pathway and the binding site, we have observed that the residues of the central node have lower RMSF in presence of PK11195 compared to RMSF in absence of PK11195 (Fig. 3.5, B), and the residues of peripheral node (at TM1 region) have higher RMSF in presence of PK11195. Since TM1 residues have very low conservation (Fig. 3.8, E), we hypothesize that the entropy transfer between the conserved binding site and variable allosteric site may be a mechanism for stabilization of binding site. However, we stress at this point that both the hypothesis needs to be evaluated in future, by both experimental as well as theoretical studies. These include quantifying the entropy associated with fluctuations of residues from each state, generation of models of mutated pathway residues and simulations of their models to name a few.

3.4. CONCLUSIONS AND FUTURE DIRECTIONS

Nature of the TSPO structural dynamics with respect to ligand binding has not been understood properly. We described residues and regions which experience changes in dynamics, depending on whether TSPO is bound or unbound to PK11195. In particular, we observed changes in residue fluctuations in TM2 and TM5 in both the holo and apo modes. Due to the fact that many of these residues are considered to be functionally significant with regards to binding cholesterol (being part of CARC as well as LAF motifs), we infer that the changes in fluctuations may also be functionally significant. We also observed that in presence of PK11195, residues of TM2 and TM5 have high correlations in their residue motions. These residues were also found to be evolutionarily conserved. Since the location of the correlated motions was found at a distance from the PK11195-binding cavity,

and we did not observe such motions in absence of PK11195, we hypothesized that these motions may be allosteric in nature. We further described the pathway and the residues involved in the allostery, which involved the residues of TM5 and TM1. We also observed that PK11195 ligand itself is dynamic when in complex with TSPO, though many of its interactions with TSPO residues were preserved. Interestingly, in one simulation, we found that PK11195 underwent change in its isomeric form from E-form to Z-form. When we compared the binding cavities of PK11195 and PP-IX. we found that both the ligands share a common binding cavity. Further analyses of residue positions interacting with PK11195 and PP-IX revealed that some of the residues interact with both PK11195 and PP-IX, and these residues have strong evolutionary conservation. These results led us to hypothesize that PK11195 may inhibit the binding of PP-IX ligand to TSPO, which has been supported by previous experimental studies.

While TM5 and TM1 residues were observed to be a part of the allosteric pathway, we propose that this pathway may modulate protein-protein interactions of TSPO, or the pathway may also be a mechanism of binding site stabilization, where the entropy could be transferred towards the TM1 residues. While we characterized the above allosteric communication pathway propose some functional implications for the same, we stress that further experimental as well as theoretical studies are needed to have a greater understanding of TSPO structure-dynamics-function relationships. Since we observed rotations in TM regions, X-ray crystallography studies in mammalian TSPO forms could also provide valuable insights on the conservation of TSPO fold. Since we proposed functional implications of the PK11195-induced allosteric pathway, experimental studies involving mutagenesis of residues involved in the pathway may be a good way to revisit this hypothesis. In addition, yeast-two hybrid screening studies could also be performed using N-terminal region of TSPO as bait to characterize TSPO-StAR interactions. Since we also proposed a possible stabilization mechanism of binding site via the allosteric pathway, we propose that the entropy associated with the fluctuations of residues involved in allosteric pathway could also be calculated, using methodology similar to Hacisuleyman et al(Hacisuleyman and Erman 2017). Additionally, the isomeric transition of PK11195 in complex with TSPO is a surprising, yet interesting result. The change in TSPO dynamics as a result of change in isomeric form, and behaviour of the ligand itself, and possible pharmacological implications from this transition are some of the interesting questions that may be pursued in future. Finally, since we observed common binding cavity and binding residues, we propose that the dynamics observed in our simulations may also be observed in presence of PP-IX ligand in complex with TSPO. This hypothesis may also be evaluated in future by means of similar molecular dynamics studies.

Chapter-4: Dynamics of Translocator protein oligomerization

4.1. Introduction 121
4.2. Results 129
4.2.1. TSPO monomers assemble to form dimers in most of the simulations 129
4.2.2. Dimerization can occur through different interfaces but some TM pairings are preferred 131
4.2.3. Interactions between aromatic residues were the most dominant at the interface 136
4.2.4. A putative cholesterol translocation mechanism was observed 139
4.2.5. Different long-range communication pathways were found 141
4.3. Discussion 144
4.3.1. Asymmetric interfaces and existence of higher order oligomers 144
4.3.2. Importance of aromatic residues in a tryptophan-rich protein 145
4.3.3. Putative translocation mechanism of cholesterol 146
4.3.4. Prevalence of long-range communications in varied dimer interfaces 147
4.4. Conclusions 148

DYNAMICS OF TRANSLOCATOR PROTEIN OLIGOMERIZATION

Rajas Rao, Julien Diharce, Bérénice Dugué, Mariano Ostuni, Fréderic Cadet, Catherine Etchebest

ABSTRACT

Translocator protein (TSPO) has been extensively studied and debated mitochondrial membrane protein. Yet, its exact role continues to be enigmatic. The situation remains the same from structural perspective as well, particularly on the nature of its oligomeric interface, despite availability of atomic structures from different species. We attempted to study dynamics of TSPO from the perspective of oligomerization. We performed several coarse-grained molecular dynamics simulations on a pair of TSPO monomers that were separated in a bilayer, as well as on dimer models with different interfaces. We identify stable TSPO dimers with diverse interfaces, some of which being consistent with earlier observations on the putative TSPO oligomer interfaces. We also identified allosteric pathways involving in dimers that involve dimer interfaces. Concurrently, a cholesterol molecule interacting with the TSPO was able to translocate through the bilayer, which could be a putative mechanism of cholesterol transport. For most of the stable interfaces, interactions between the aromatic residues dominate diverse oligomeric interfaces. Our observations shed new light on TSPO oligomerization and bring new perspectives on its dynamics, as well its interactions with protein and ligand partners.

Keywords: Translocator protein, transmembrane protein interactions, allosteric pathway

4.1. INTRODUCTION

Translocator protein (TSPO) is an 18 kDa transmembrane (TM) protein that is primarily found in outer mitochondrial membrane, though it has also been reported in other organelles, such as plasma membrane (Bouyer et al. 2011), nuclear envelope (Hardwick et al. 1999) & endoplasmic reticulum (Fan et al. 2009). 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 (Braestrup and Squires 1977). It was named as peripheral benzodiazepine receptor (PBR), due to the fact that it appeared to be predominantly concentrated in peripheral nervous system. However, it was later found to be involved in transport of cholesterol into the mitochondria (Krueger and Papadopoulos 1990). Successive experiments later found that PBR has a major role in translocation of cholesterol into the mitochondria (Papadopoulos, Amri, Li, et al. 1997; Papadopoulos et al. 1991; Li, Yao, et al. 2001), due to which it was renamed as TSPO (Papadopoulos, Amri, Boujrad, et al. 1997; Papadopoulos et al. 2006).

Because many of the enzymes involved in steroid synthesis are located in the inner matrix of mitochondria, import of cholesterol into mitochondria for synthesis of steroidogenic hormones is the rate-limiting step (Papadopoulos et al. 2007; Jefcoate 2002). Treatment with various TSPO ligands, such as PK11195, Ro5-4864 etc. resulted in modulation of this step (Papadopoulos et al. 1991; Krueger and Papadopoulos 1990; Mukhin et al. 1989), which led to the hypothesis that TSPO is the protein that enables cholesterol transport. Early studies on molecular modelling of TSPO further supported this hypothesis (Papadopoulos et al. 1994), where it was observed that the transmembrane pore was large enough to accommodate a cholesterol molecule. 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 (Papadopoulos, Amri, Li, et al. 1997). Located at the C-terminal of the protein in TM5 region, the motif was largely hydrophobic in nature, and consisted of residues L/V-(X)1–5-Y-(X)1–5-R/K (Li, Yao, et al. 2001; Papadopoulos, Amri, Li, et al. 1997). The presence of a positively charged residue at the end of the motif suggested that this residue may act as a tether that interacts with the hydroxyl group of cholesterol. Similarly, in another study, incorporation of TSPO in *E. coli* cells resulted in increased cholesterol uptake. Since *E. coli* does not have cholesterol, increased uptake of cholesterol in TSPO transfected cells further reinforced the idea that TSPO is indispensable for cholesterol translocation (Li and Papadopoulos 1998). 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 (Fan et al. 2018). Furthermore, an alternate cholesterol-binding motif adjacent to the CRAC motif was characterized recently, also located in TM5 region (Papadopoulos et al. 2018;

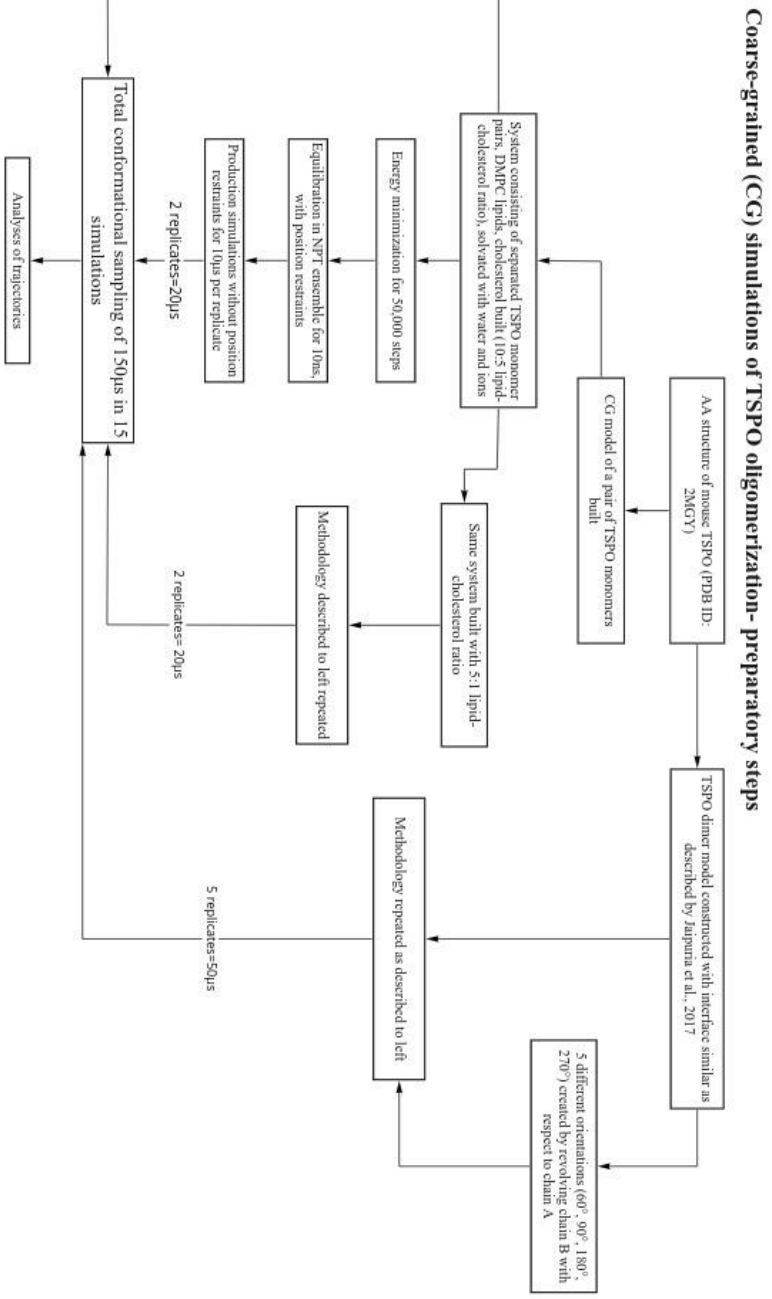
Fantini et al. 2016). This motif has a sequence pattern palindromic to CRAC motif, and thus was called 'CARC' motif. Though theoretical studies have suggested that this motif has capability to bind to cholesterol at higher affinities compared to CRAC motif (Fantini et al. 2016), CARC motif has not yet been demonstrated to bind to cholesterol experimentally.

While the role of TSPO in cholesterol transport has been well studied, recent studies dispute this dogma (Selvaraj and Stocco 2015). *In vivo* experiments on TSPO knockout mice revealed that knockout of TSPO resulted in no significant change in testosterone production, gametogenesis and reproduction, all of which are highly dependent on the cholesterol import into mitochondria (Morohaku et al. 2014). This suggested that TSPO has no role to play in steroidogenesis, contrary to the conclusions drawn by previous studies. In another study, it was revealed that there was no effect of TSPO knockout on the steroidogenesis in mice (Tu et al. 2014). Despite significant advances in knowledge on the role of TSPO in cholesterol translocation, its exact function continues to remain a mystery. However, apart from cholesterol transport, TSPO has also been observed to play a role in various other processes, such as transport of porphyrin (Zeno et al. 2012), erythropoiesis (Rampon et al. 2009), reactive oxygen species (ROS)- mediated apoptosis (Kugler et al. 2008; Maaser et al. 2001; Veenman and Gavish 2012; Veenman et al. 2008), neuroinflammation (Liu et al. 2014; Chechneva and Deng 2016; Bonsack et al. 2016) to name a few.

Until recent past, the structural knowledge of TSPO at atomic level also remained unknown. Earlier molecular modelling studies on TSPO suggested that it folds into a five TM domain (Papadopoulos, Amri, Boujrad, et al. 1997; Bernassau et al. 1993), which was confirmed from cryo-electron microscopy studies (Korkhov et al. 2010; Hinsen et al. 2015; Jaremko, Jaremko, Becker, et al. 2014), as well as from initial atomic structures solved by NMR spectroscopy (Jaremko, Jaremko, Giller, et al. 2014). Soon after NMR structure of mouse TSPO was available as a monomer, further X-ray crystal structures of bacterial TSPO revealed the existence of different TSPO oligomers (Li, Liu, Zheng, et al. 2015a; Guo et al. 2015), confirming that TSPO may be found in oligomeric arrangement as suggested by earlier cryo-electron microscopy studies (Papadopoulos et al. 1994; Lacapère et al. 2001). For instance, the X-ray crystal structure of *Bacillus cereus* TSPO was found to be a homodimer (Guo et al. 2015). The X-ray crystal structure of TSPO from *Rhodobacter sphaeroides* showed that TSPO may exist as a trimer as well as dimer for the A139T mutant (PDB ID: 4UC2 and 4UC1 respectively), and as a dimer for the wild type (Li, Liu, Zheng, et al. 2015a). The dimer interface in both the bacterial dimers are different. While *B. cereus* dimer involves TM2 from each monomer as the dimer interface, *R. sphaeroides* dimer interface is shared between TM1 and TM3 of each monomer. In the trimer structure of A139T mutant, a first interface involves TM1 and TM3 from each monomer chain, whereas the second interface is shared by TM1, 2 and 5 as the contributing regions.

On the other hand, recent solid-state NMR experiments revealed a different picture for the mouse TSPO: TSPO was found to exist in a dynamic equilibrium between monomer and dimer, with the dimer interface shared by TM3 region of each monomer (Jaipuria et al. 2017). The change in stoichiometric orientation was found to be dependent on the interaction with cholesterol, where the interaction with cholesterol was proposed to trigger the dissociation of dimer into monomer orientation. In conclusion, though the assembly of TSPO monomers into quaternary structures are relatively well-documented, there are several contradictions on the nature of oligomeric interface and the prevalent stoichiometric arrangement, depending on the sequence and/or the experimental conditions, in particular, the cholesterol concentration. As a consequence, even though is unequivocal that TSPO has the capability to oligomerize, the interface of the oligomer continues to be elusive.

Accordingly, in this study, we aim to answer the following questions related to the dynamics of TSPO oligomerization: Can a pair of TSPO monomers separated by a certain distance in a lipid bilayer diffuse and assemble into a dimer? if yes, what is the nature of the interface and the stabilizing forces. We attempted to address these questions by means of coarse-grained molecular dynamics simulations on a pair of TSPO monomers, as well as TSPO dimer models. We performed 11 simulations of 10μs each, and we examined the dynamics of dimerization, interfaces and the nature of dimer interactions. The methodology followed is described in the flow charts below, and discussed in detail in chapter-2.



Box. 4.1. Flow chart describing methodology followed for preparatory steps of TSPO dimerization and dimer simulations

Parameters for CG simulations of TSPO oligomerization

Energy minimization parameters:

Method: ***Steepest descent****

No. of steps: 50,000

Equilibration at NPT ensemble:

Position restraint (force constance): 1,000, 500, 100, 10, 0 kJ/mol nm^2 in 5 simulations

Time: 1ns x 5 simulations

Time-step: 0.01 ps

Temperature: 300K

Thermostat: ***Berendsen***

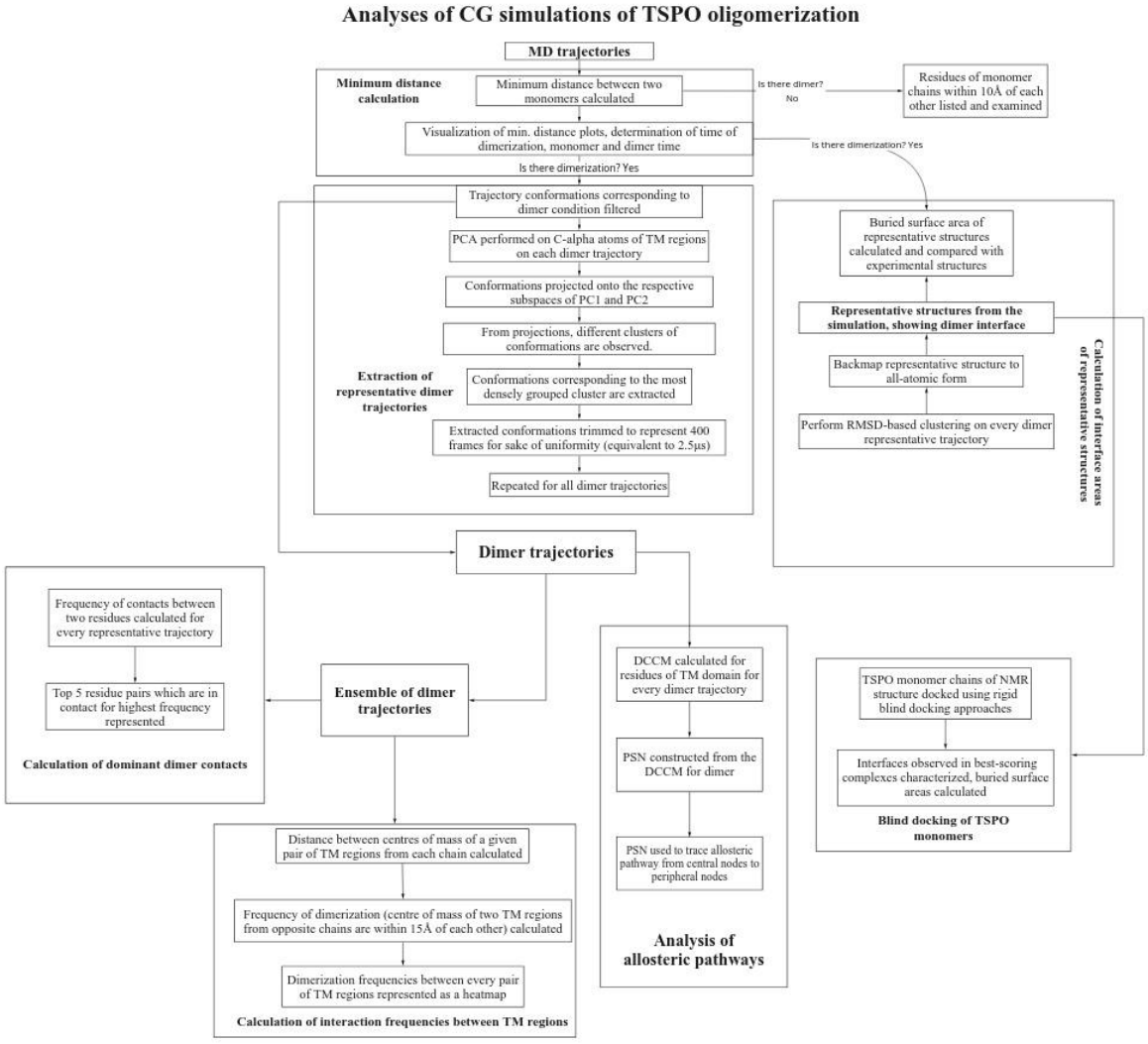
Barostat: ***Parinello-Rahman***, with semiisotropic coupling

Production MD simulation

Same parameters as of equilibration at NPT ensemble, except without position restraints, V-rescale method as thermostat

* Methods marked in ***bold+italic*** are described in detail in chapter-2.

Box. 4.2. Parameters used for preparatory steps of TSPO dimerization and dimer model simulations.



Box. 4.3. Flow chart describing methodology used for analysis of TSPO dimer simulations

Additional information: Tools and parameters used for analyses:

- ***Calculation of minimum distance:*** Minimum distance between two TSPO monomer chains was calculated using g_mindist program of GROMACS 4.6.7 (Berendsen et al., 1994, Pall et al., 2015) for every simulation. If the minimum distance was less than 0.5nm (5Å), then the two TSPO monomers was considered to oligomerized.
- ***Analysis of residues in proximity to each other***: For simulations where minimum distance did not decrease below 0.5nm, residues of both monomers in proximity to each other (1.0nm) during the simulation were calculated using g_select tool of GROMACS (Berendsen et al., 1994, Pall et al., 2015), and the data was plotted as no. of frames given residue is in proximity to the other monomer chain.
- ***Extraction of representative dimer trajectories***: To extract subset of trajectories that have stable interface and representative of the simulation, PCA was performed on each trajectory, and conformations were projected onto the subspace of PC1 and PC2, using Bio3d MD analysis program (Grant et al., 2006). The conformations from densely populated cluster were extracted, and trimmed to ensure that number of frames is 400 (equivalent to 2.5μs of sampling), for the sake of uniformity.
- ***Calculation of interface areas/buried surface areas of representative trajectories***: On the representative dimer trajectories, RMSD-based clustering was performed using g_cluster tool of GROMACS, with cutoff of 0.3nm. The representative structure with largest population was back-mapped to all-atomic form using Backward tool (Wassenaar et al., 2014). Interface area was calculated on the representative structures from each trajectory using the following formula:

$$IA = (ASAa + ASAb) - ASAd$$

- Where IA is the interface area, ASAa is the solvent-accessible surface area for the monomer chain-A, ASAb is the solvent-accessible surface area for the monomer chain-B, and ASAd is the solvent accessible surface area for the dimer. The accessible surface areas were calculated using Naccess program (Hubbard and Thornton, 1993). The same was calculated for experimental TSPO oligomer structures, and compared with representative structures.
- ***Docking of TSPO monomers:*** Docking of TSPO monomers was performed using Memdock web server (Hurwitz et al., 2016). The complexes were scored using Mem-PatchDock scoring method (Hurwitz et al., 2016, Schneidman-Duhovny et al., 2005). Interfaces of 20 best-scoring complexes were noted, and interface areas were calculated.
- ***Calculation of interaction frequencies between TM regions:*** To understand the prevalence of interaction between given two TM regions of opposite monomer chains, distance between centre of mass of TM regions were calculated using g_dist tool of GROMACS (Berendsen et al., 1994, Pall et al., 2015). The frequency of interactions (where distance between centre of mass of TM regions < 15Å) was calculated and normalised against total number of frames in the ensemble. The frequencies were calculated for all the TM regions and projected onto a heatmap.
- ***Tracing allosteric pathway pathway in dimers***: Allosteric pathways in dimer simulations were characterised from protein structure network, which was constructed from the DCCM for every dimer trajectory. The correlation cutoff used for construction of matrix vary between 0.4 to 0.8, depending on the trajectory. Allosteric pathway construction was performed using Bio3d MD analysis package (Grant et al., 2006)
- ***Calculation of dominant dimer contacts:*** To understand the residue pairs that dominate the interactions, a protein contact map was calculated for every frame of each dimer trajectory, and frequency of contacts was calculated as the percentage of ensemble where two given residues are in contact with each other. The cutoff value below which a give n residue pair is considered contact is 5Å.

miro

Box 4.4. Additional information on tools and parameters used for analysis of TSPO dimer simulations.

Simulation	No. of particles	Cholesterol concentration	Orientation of chain B with respect to chain A	Velocity seed
1	TSPO: 2 (separated) DMPC: 414 Cholesterol: 206 Water: 10128 Na: 108 Cl: 116	33.00%	0°	-1
2	TSPO: 2 (separated) DMPC: 414 Cholesterol: 206 Water: 10128 Na: 108 Cl: 116	33.00%	0°	-3
3	TSPO: 2 (separated) DMPC: 382 Cholesterol: 191 Water: 10033 Na: 69 Cl: 77	33.00%	90°	-11
4	TSPO: 2 (separated) DMPC: 382 Cholesterol: 191 Water: 10033 Na: 69 Cl: 77	33.00%	90°	15
5	TSPO: 2 (separated) DMPC: 384 Cholesterol: 192 Water: 10023 Na: 69 Cl: 77	33.00%	180°	-3
6	TSPO: 2 (separated) DMPC: 384 Cholesterol: 192 Water: 10023 Na: 69 Cl: 77	33.00%	180°	-9
7	TSPO: 2 (separated) DMPC: 385 Cholesterol: 192 Water: 10005 Na: 69 Cl: 77	33.00%	270°	-3
8	TSPO: 2 (separated) DMPC: 385 Cholesterol: 192 Water: 10005 Na: 69 Cl: 77	33.00%	270°	-3
9	TSPO: 2 (separated) DMPC: 414 Cholesterol: 207 Water: 9924 Na: 68 Cl: 76	33.00%	270°	-5
10	TSPO: 2 (separated) DMPC: 545 Cholesterol: 124 Water: 9880 Na: 68 Cl: 76	16.00%	0°	-3
11	TSPO: 2 (separated) DMPC: 545 Cholesterol: 124 Water: 9880 Na: 68 Cl: 76	16.00%	0°	-5
12	TSPO: 2 (dimer model with interface similar to NMR model) DMPC: 218 Cholesterol: 109 Water: 7066 Na: 48 Cl: 56	33.00%	0°	173529
13	TSPO: 2 DMPC: 407 Cholesterol: 204 Water: 9816 Na: 68 Cl: 76	33.00%	60° (rotated)	173529
14	TSPO: 2 DMPC: 407 Cholesterol: 204 Water: 9816 Na: 68 Cl: 76	33.00%	90° (rotated)	173529
15	TSPO: 2 DMPC: 410 Cholesterol: 205 Water: 9944 Na: 69 Cl: 77	33.00%	180° (rotated)	173529
16	TSPO: 2 DMPC: 409 Cholesterol: 204 Water: 9502 Na: 65 Cl: 73	33.00%	270° (rotated)	173529

Table 4.1. Summary of TSPO dimerization and TSPO dimer model simulations.

4.2. RESULTS

4.2.1. TSPO monomers assemble to form dimers in most of the simulations

Inspired by the experimental work of Jaipura et al. (Jaipuria et al. 2017) who examined dimerization process of TSPO in DMPC/Cholesterol liposomes with solid-state NMR experiments, we performed simulations in the same lipid-cholesterol ratio and checked if TSPO monomers initially separated at a certain distance (55-60Å) in DMPC/cholesterol bilayer can diffuse and assemble together to form dimers. We started with different rotational orientations to avoid any bias due to insufficient sampling. Indeed, rotational reorientation of monomers generally needs longer time compared to translational diffusion and may not be fully achieved in the limited simulation time.

Eight replicates of 10μs each were thus run to get enough statistics on the pairing. 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 0.5 nm. In six cases, as the simulations progressed, we observed that TSPO monomers come in close contact (Fig. 4.1). Though long-lifetime dimers occurred in half simulations (4 among 8), dimers of short-lifetime may also form but finally dissociate (Fig. 4.1, red lines). This result indicates that the dimer formation does not occur incidentally but needs specific interactions to be stabilized.

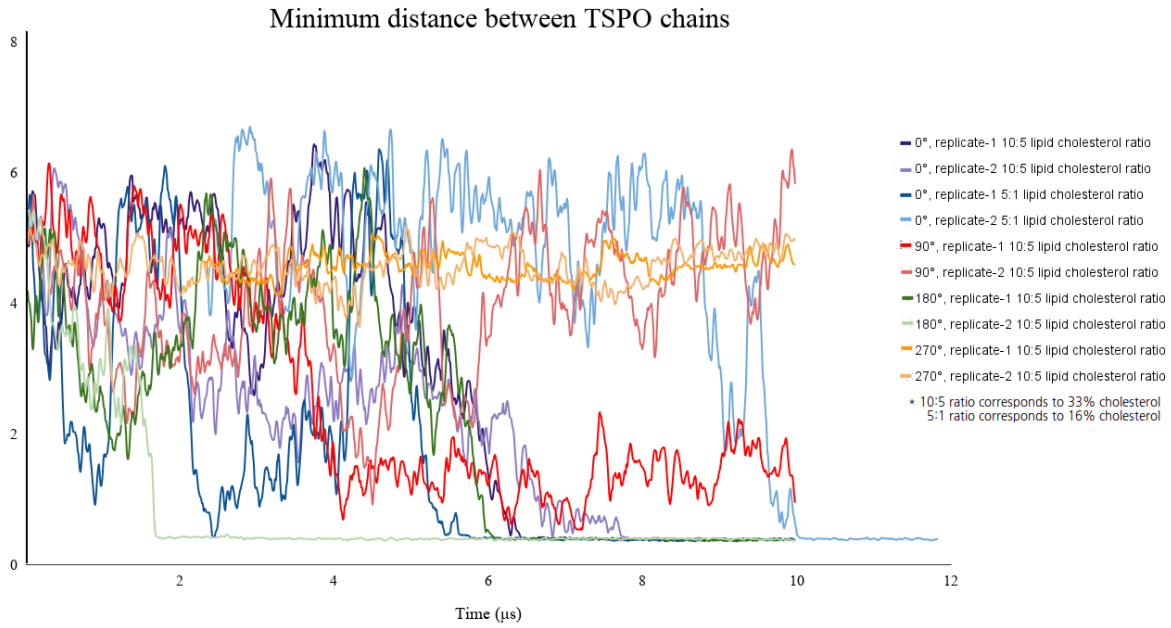


Figure 4.1. Plot showing minimum distance between two TSPO chains. Minimum distance between TSPO chains in each simulation is represented by different colour. y-axis represents Minimum distance (in nm), while the x-axis represents the simulation time (in μs).

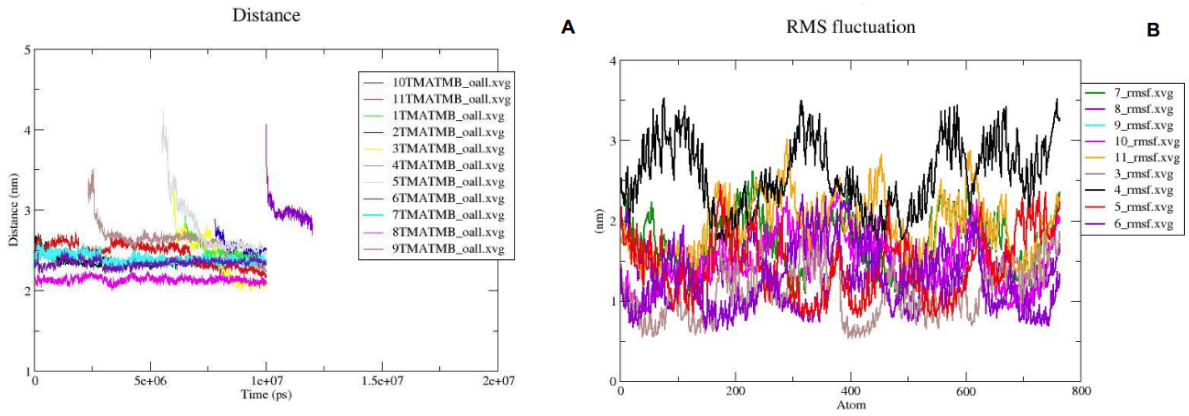


Figure 4.2: Properties of simulated dimer as a function of time. **A:** Distance between center of mass of TM domain of each monomer **B:** Fluctuations of residues of TM domains for all the replicates. Numbers 1-11 represent the following simulations:
1: 0° starting orientation at 33% cholesterol, replicate-1
2: 0° starting orientation at 33% cholesterol, replicate-2
3: 180° starting orientation at 33% cholesterol, replicate-1
4: 0° starting orientation at 33% cholesterol, replicate-2
5: 0° starting orientation at 16% cholesterol, replicate-1
6: 0° starting orientation at 16% cholesterol, replicate-2
7: Dimer model-1 (interface based on NMR studies (Jaipuria et al. 2017))
8: Dimer model-2 (chain-B oriented 60° with respect to model-1)
9:Dimer model-3 (chain-B oriented 90° with respect to model-1)
10:Dimer model-4 (chain-B oriented 180° with respect to model-1)
11:Dimer model-5 (chain-B oriented 270° with respect to model-1).

Besides this whole dimerization process, we also started with pre-formed dimers and examined their conformational stabilities. Five simulations corresponding to different relative orientations of the two monomers were run for a total of 50 µs. Whatever the simulations, the two monomers remained close to one another (Fig. 4.2, A) but interacting regions vary along the simulation. The fluctuations are rather large (Fig. 4.2, B), 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. Interestingly, the dimer with the smallest fluctuations is the one presenting the smallest distance between the center of mass of the two transmembrane domains. In comparison, the dimer starting from the NMR model shows higher fluctuations and a larger distance between the center of mass of the two transmembrane domains. Clearly, stable alternative conformations do exist beside the NMR model.

We also started to examine systems with a reduced ratio of DMPC:Cholesterol, i.e. 16%, which is closer to a physiological ratio, but no significant dependence of cholesterol concentration with the dynamics was observed (Fig. 4.2, B). However, this result needs to be further confirmed by an enhanced sampling, since only a few simulations were run compared to the 33% concentration.

To go further, we examined the number and nature of TM residues that participate in the interface, *i.e.* residues of one monomer in contact with residues of the other monomer. We chose a distance criterion to define a contact, the distance threshold was first chosen as 5.6Å, which roughly corresponds to the sum of the radius of two CG beads, *i.e.* 2.8Å. The corresponding contact numbers are given in Fig. 4.3, A, for the all sets of simulation. The plots show 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. 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. Since the choice of a threshold might be arbitrary, we checked its impact using a slightly increased value, *i.e.* 6 Å (Fig. 4.3, B). 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.

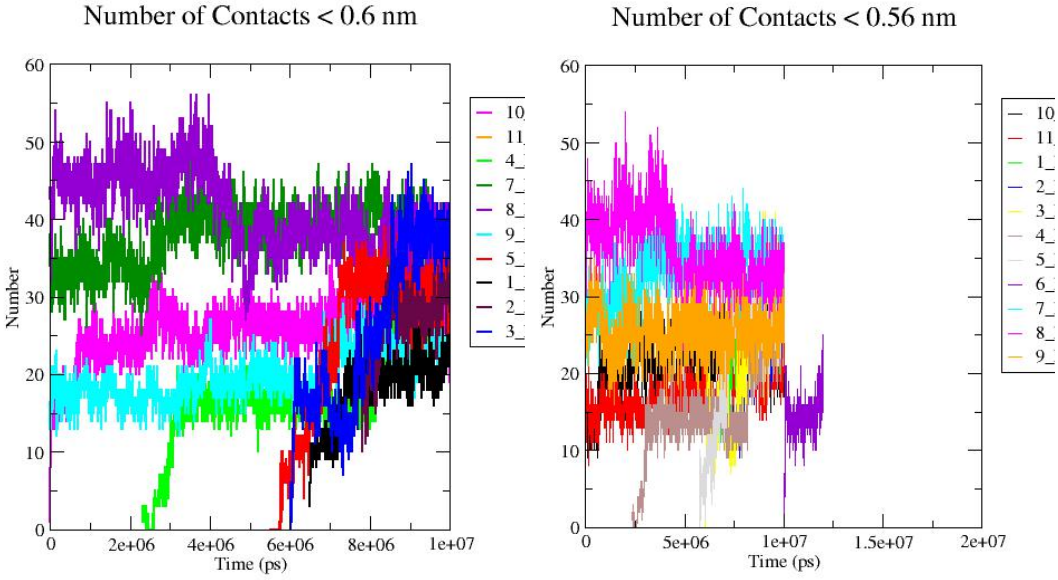


Figure 4.3. A: Number of residues within a distance < 5.6 Å of residues of the other chain; **B:** Number of residues within a distance < 6 Å of residues of the other chain. Numbers 1-11 indicate replicate simulations, as described in the caption of Fig. 4.2.

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

Since the dimers show significant dynamics, the nature of residues at the dimer interface may vary. We first analyzed what are the transmembrane segments involved in the contact using the same distance as above. We thus extracted from each simulation a set of conformations regularly time spaced. From this ensemble, we calculated the distribution of residues in contact, i.e. within a distance < 5.6 Å.

The results are depicted in Fig. 4.4, which shows that residues of TM3 are by far, the most frequently involved in the contacts, while TM1 and TM4 residues participate but to a lesser extent. In contrast TM5 and TM2 seem rarely involved.

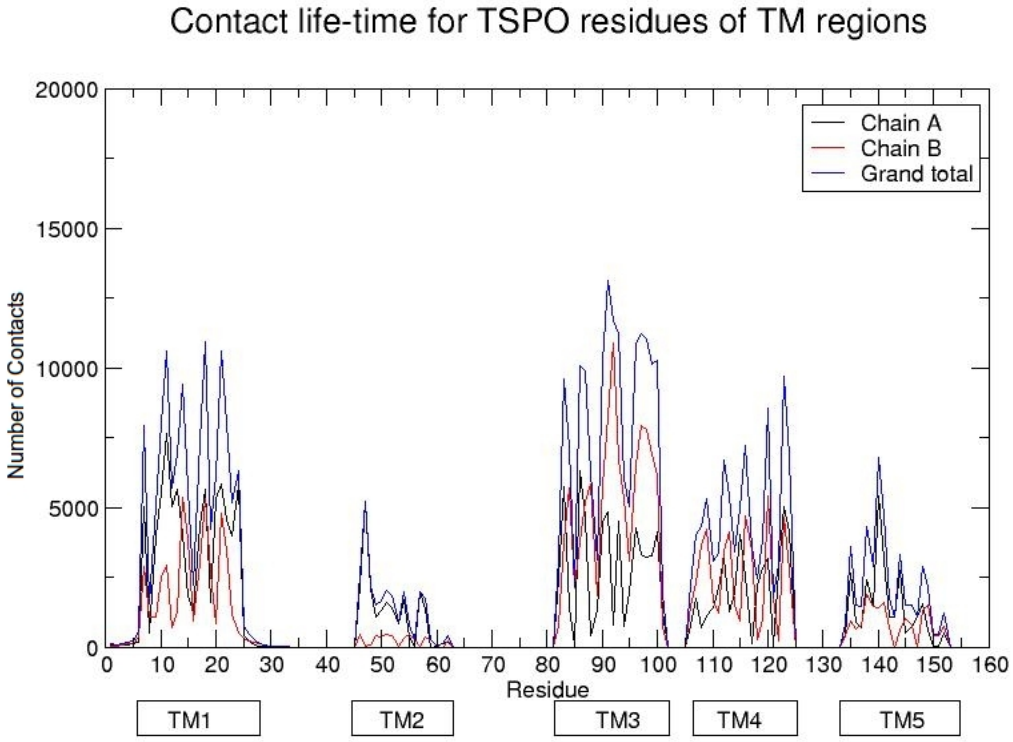


Figure 4.4: Total contact numbers for each residue calculated for all the frames saved from the whole set of trajectories. Contacts are defined on a distance-based threshold < 5.6 Å.

Very similar results were obtained with different thresholds and different measures tested, e.g., from distances between centers of mass of transmembrane segments etc. All these tests confirmed that the most frequent transmembrane segment involved in the interface is TM3 in partnership with TM1, TM5 in partnership with TM4, and TM3 with TM5 in this decreasing order (see Table 4.2).

Moreover, further analyses of the complexes have shown that the number of TM regions participating in the interface can vary between 1 TM from each chain, 1TM from one chain with 2TM regions from the other chain and 2 TM regions from each chain. Actually, we found that most of the interfaces are formed by pairs of pairs of TM (2 TM:2TM) on average, *i.e.* ~ 4 TM are involved in the interface.

	1	2	3	4	5
1	0.00	0.03	0.20	0.09	0.13
2	0.03	0.00	0.08	0.10	0.01
3	0.20	0.08	0.08	0.12	0.13
4	0.09	0.10	0.12	0.03	0.16
5	0.13	0.01	0.13	0.16	0.06

Table 4.2. Heatmap representing the frequency of interactions between different TM regions of TSPO. Interaction instances were calculated based on calculation of minimum distance between the different TM regions of opposite chains. Centre of mass of TM regions that are less than 15.0Å distance of each other were considered to be in interaction. Total number of instances satisfying this condition were calculated, and frequency is represented as the ratio of instances satisfying the conditions of interaction with total number of instances (conformations from the trajectory ensemble.

Despite a large occurrence of dimerization, some simulations were unsuccessful. We went further in the analysis of these failures and examined the residues of monomer chains in proximity (1.0 nm) to each other for the two simulations at 90° starting orientation. We observed that for the first simulation, the residues of the loop between TM1&2 were in contact with the residues of TM4 (Fig. 4.5). For the second simulation, the residues of CRAC motif from each monomer chain were in contact with each other (Fig. 4.5). Since these regions contain charged positive residues, electrostatic repulsion may have prevented the dimerization.

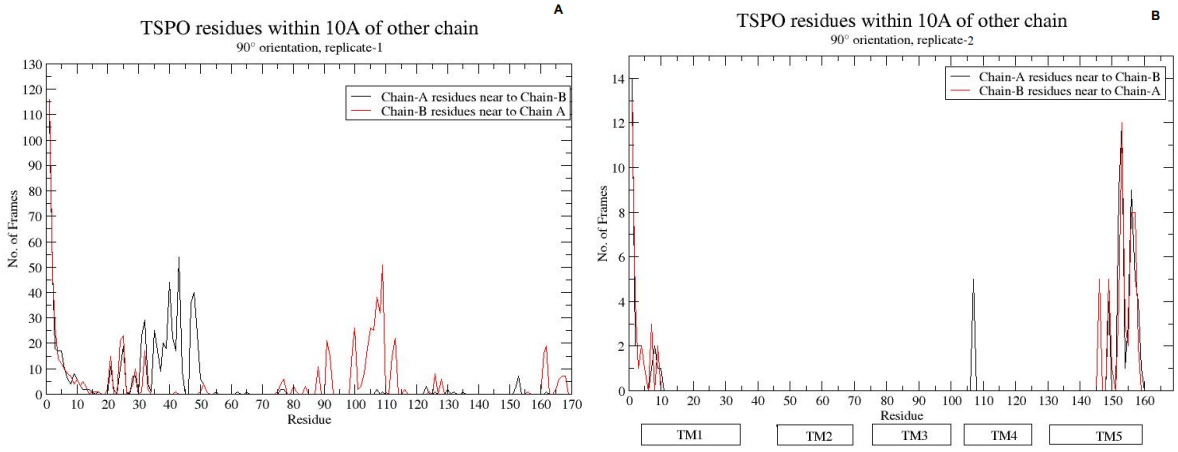
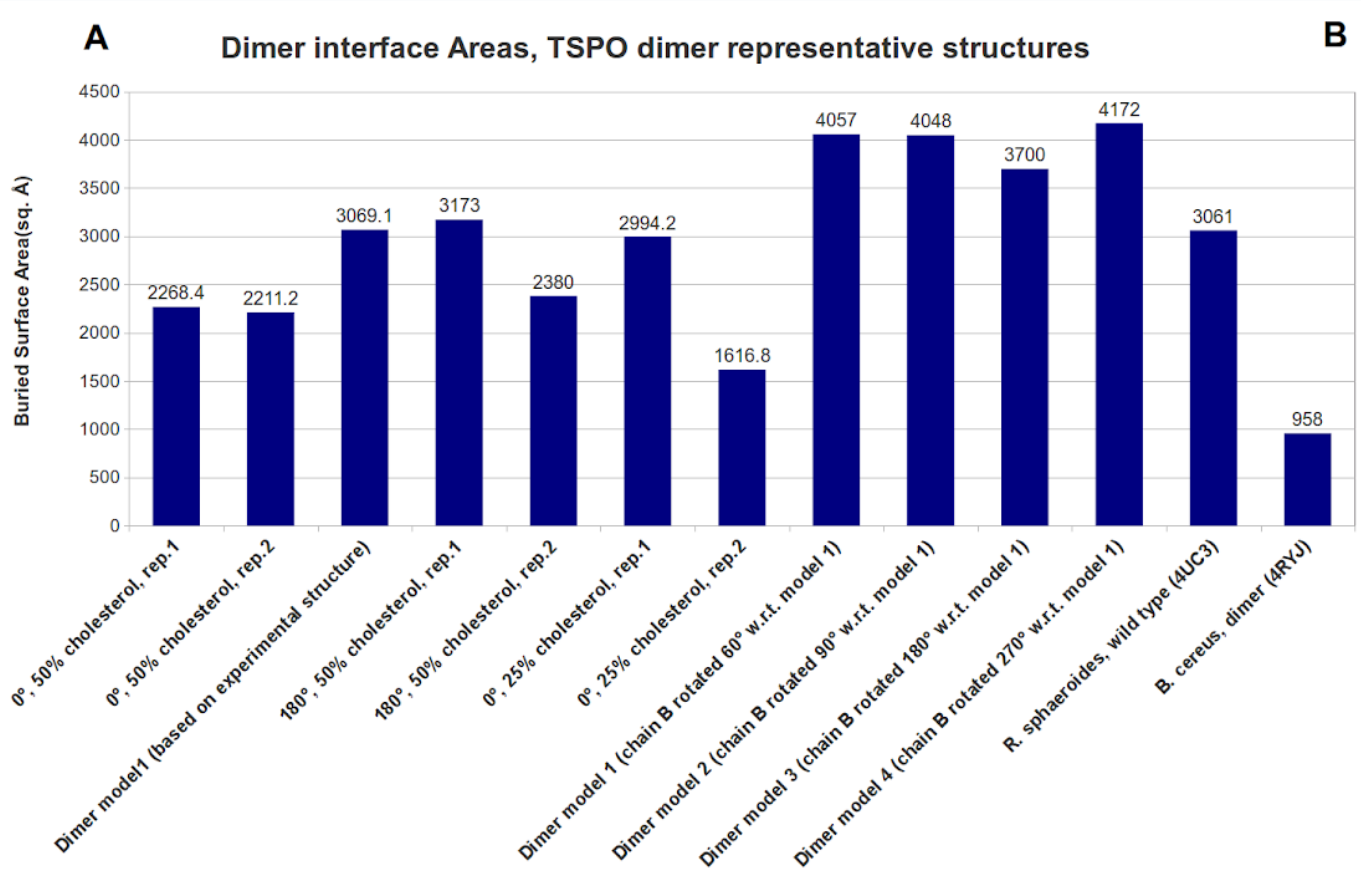


Figure 4.5. Plot showing TSPO residues of each chain within 10Å distance of each other, for the replicates 1 (A) and 2 (B) of simulations with 90° initial orientation of chains with respect to the first orientation. The proximate residues were calculated using g_cluster tool of GROMACS [61,62].

(legend on next page)



B

	Simulation	Representative structure	Dimer interface area (Å²)
1	0° orientation, 33% cholesterol, replicate-1 (TM4+TM5) (TM4+TM5)		2,268
2	0° orientation, 33% cholesterol, replicate-2 (TM1+TM3) (TM2+TM5)		2,211
3	180° orientation, 33% cholesterol, replicate-1 (TM4+TM5) (TM4+TM5)		3,173
4	180° orientation, 33% cholesterol, replicate-2 (TM2+TM5) (TM4+TM5)		2,380
5	0° orientation, 16% cholesterol, replicate-1 (TM1+TM3) (TM4+TM5)		2,994
6	0° orientation, 16% cholesterol, replicate-2 (TM1) (TM5)		1,616
7	TSPO dimer model-1 similar to NMR experimental model (TM1+TM3) (TM1+TM3)		3,069
8	**TSPO dimer model-2 (chain B rotated 60° w.r.t. model-1) (TM1 TM3) (TM3 TM4)**		**4,057**
9	**TSPO dimer model-3 (chain B rotated 90° w.r.t. model-1) (TM1+TM2) (TM3+TM4)**		**4,048**
10	TSPO dimer model-4 (chain B rotated 180° w.r.t. model-1) (TM+TM3) (TM4+TM5)		3,700
11	**TSPO dimer model-5 (chain B rotated 270° w.r.t. model-1) (TM4) (TM3+TM4)**		**4,172**

Figure 4.6. A: Chart showing dimer interface areas of representative TSPO dimer structures. Buried surface areas were calculated as the difference between the twice the accessible surface areas of TSPO in monomeric form and accessible surface area of TSPO dimer. Accessible surface areas were calculated using Naccess program (Hubbard and Naccess 1993). **B:** Table showing representative TSPO dimer conformations and their interfaces obtained from each trajectory. The representative conformations were obtained using the PCA and RMSD-based clustering of the ensemble of TSPO dimer conformations from each trajectory, and back-mapped to all-atomic representation. Dimer interface areas were calculated on the all-atomic representative conformations using Naccess program (Hubbard and Naccess 1993).

All these calculations were based on the coarse-grain model that may preclude from detecting fine specificities. Consequently, we turned to an all-atom description but since the back-mapping procedure is cpu-time consuming, we first extracted representative conformations from respective trajectories using methodology described in Box 4.2. These representative CG conformations were then back mapped to get all-atom structures that were further explored in terms of accessible surface area. The buried surface area of each complex conformation was defined as the sum of the accessible surface of each isolated monomer minus the accessible surface area of the complex. The values range between a few hundred $Å^2$ to five thousand $Å^2$, the average being ~ 3000 $Å^2$. Values of interface areas calculated for representative structures obtained after filtering and clustering of conformations are represented in Fig. 4.6. The average value is comparable to that calculated for X-ray structure of *R. sphaeroides* TSPO dimer (3061 $Å^2$, Fig. 4.6, A). Surprisingly, this value is significantly larger than the corresponding value for *B. cereus* TSPO (958 $Å^2$), which turns out to be closer to the size of unspecific contacts within crystals (Guo et al. 2015). From visualization of representative structures, we also found many interfaces that are previously not described in previously described atomic structures of TSPO (Fig. 4.6, B). Furthermore, many of these interfaces are asymmetrical, though we also observed few symmetrical interfaces as well (Fig. 4.6, B). This is particularly interesting considering that TSPO dimer structures have been solved with differing interfaces from different species (Guo et al. 2015; Li, Liu, Zheng, et al. 2015a; Jaipuria et al. 2017).

Besides MD simulations, we also tested a crude rigid-body docking procedure, *i.e.* starting from mouse monomer NMR structure, without any further refinement. We chose Memdock method since it has been specifically developed for modelling membrane complexes. Interestingly, some interfaces observed from the simulations were also retrieved in the blind-docked complexes of TSPO homo dimer (Table 4.3). Among the top 20 docked complexes of TSPO dimers, 13 complexes were found to have an interface that involves the same set of TM helices as observed in the simulations (Table 4.3). These include the interfaces involving TM 1 & 3 with TM 4 & 5 (5 of 13 complexes), TM 1 & 3 with TM 1 & 3 (4 of 13 complexes), and TM 1 & 3 with TM 2 & 5 (4 of 13 complexes).

Dimer model (rank)	Memdock score	Interface TM regions (Chain-A)	Interface TM regions (Chain-B)	Buried Surface Area (Å^2)
1	-1969.4	1,2	1,3	4943.9
2	-1925.4	1,3	3,4,5	5435
3	**-1923.4**	**1,3**	**4,5**	**5339.2**
4	**-1912.9**	**1,3**	**2,5**	**5664.1**
5	-1904.2	1,2	2,5	5265
6	**-1902.8**	**1,3**	**4,5**	**5316.1**
7	-1901.9	1,2	4,5	5566.9
8	**-1900.4**	**1,3**	**1,3**	**5421.4**
9	-1895.4	1,2	2,5	4860.3
10	**-1889.2**	**1,3**	**2,5**	**5520.9**
11	**-1885.2**	**1,3**	**4,5**	**5783**
12	**-1883.8**	**1,3**	**1,3**	**5414.7**
13	**-1882.7**	**1,3**	**1,3**	**5179.4**
14	-1882.3	1,3	3,4,5	5378.4
15	**-1878.9**	**1,3**	**2,5**	**5141.9**
16	**-1878.0**	**1,3**	**4,5**	**5402.3**
17	**-1875.5**	**1,3**	**2,5**	**5313.8**
18	**-1873.1**	**1,3**	**4,5**	**5789.2**
19	**-1871.7**	**1,3**	**1,3**	**5270.7**
20	-1867.9	3	1,3	4783.1

Table 4.3. Best-scoring TSPO homo-dimer complexes obtained from blind-docking of NMR structure of mouse TSPO. Complexes whose interfaces are similar to those observed from the ensemble of simulations are highlighted in bold. Docking was performed on NMR structures of mouse TSPO (PDB: 2MGY) using Memdock web server (Hurwitz et al. 2016), and interfaces of 20 best-ranking complexes were annotated based on residue-contacts. Buried surface areas were calculated as the difference between twice the accessible surface area of monomer chain and accessible surface area of TSPO dimer. Accessible surface areas were calculated using Naccess program.

Even though it does not constitute a proof, it does bring some support to the results got from MD simulations since the two approaches are significantly different.

4.2.3. Interactions between aromatic residues were the most dominant at the interface

In various observed interfaces, we attempted to understand the nature of interactions occurring between the residues of TSPO chains. We calculated contact maps for the whole set of trajectories and for the representative set of conformations using different thresholds, i.e. 5.0 Å, 5.6 Å or 6.0 Å, and also computed the frequencies associated with the residue contacts. Depending on this value, the ranking of preferred pairs of residues may slightly change. Nevertheless, we systematically found that among the ten most abundant pairs, interactions between aromatic residues and Leu-Leu dominated

Simulation	Chain A residue	Chain B residue	Interaction frequency (% of trajectory where given pairs are <5.6 Å distance)
0° rotation, 33% cholesterol, replicate-1	Tyr140 Leu137 Tyr140 Tyr138 His127	Tyr140 Leu124 Ala119 Thr123 Arg128	99.14% 94.15% 93.72% 93.01% 90.30%
0° rotation, 33% cholesterol, replicate-2	Phe100 Leu91 Trp93 Phe100 Ala90	Tyr153 Phe146 Leu49 Leu150 Pro51	94.56% 94.10% 94.10% 92.74% 92.52%
180° rotation, 33% cholesterol, replicate-1	Tyr140 Leu141 Trp107 Trp107 Ala108	Tyr138 Thr123 Tyr152 Arg166 Arg166	97.25% 79.25% 72.50% 62.13% 58.13%
180° rotation, 33% cholesterol, replicate-2	Thr48 Pro51 Thr48 Thr48 Thr55	Asn92 Thr120 Leu88 Thr120 His127	91.61% 91.15% 90.31% 88.48% 86.40%
0° rotation, 16% cholesterol, replicate-1	Met79 Ile98 Phe25 Pro15 Phe74	Tyr140 Tyr152 Tyr153 Leu141 Tyr138	81.44% 75.67% 75.22% 74.22% 74.22%
0° rotation, 16% cholesterol, replicate-2	Tyr152 Tyr153 Tyr153 Arg135 Tyr140	Phe100 Phe100 Phe25 Val6 Val6	99.00% 95.26% 93.77% 93.52% 92.77%
Dimer model-1 (Interface based on Jaipuria et. al.)	Trp93 Leu91 Pro97 Trp93 Pro97	Pro15 Val14 Pro96 Trp93 Trp95	98.15% 96.70% 95.20% 95.15% 95.15%
Dimer model-2 (Chain B 60° w.r.t. dimer model 1)	Trp68 Thr86 Thr12 Phe20 Met79	Trp126 Gly87 Leu91 Phe100 Val80	94.65% 93.85% 92.20% 91.00% 89.56%
Dimer model-3 (Chain B 90° w.r.t. dimer model 1)	Leu13 Tyr57 Leu17 Val13 Val9	Gln88 Ser116 Leu91 Tyr85 Thr120	98.10% 96.30% 93.95% 93.85% 93.75%
Dimer model-4 (Chain B 180° w.r.t. dimer model 1)	Tyr152 Asp111 Arg135 Leu141 Trp155	Trp33 Arg103 Val14 Phe25 Trp33	99.00% 96.25% 95.40% 95.15% 95.10%
Dimer model-5 (Chain B 270° w.r.t. dimer model 1)	Trp155 Arg128 Asp111 Leu144 Val129	Trp33 Trp93 Phe100 Pro97 Thr86	97.70% 95.00% 93.95% 91.90% 91.20%

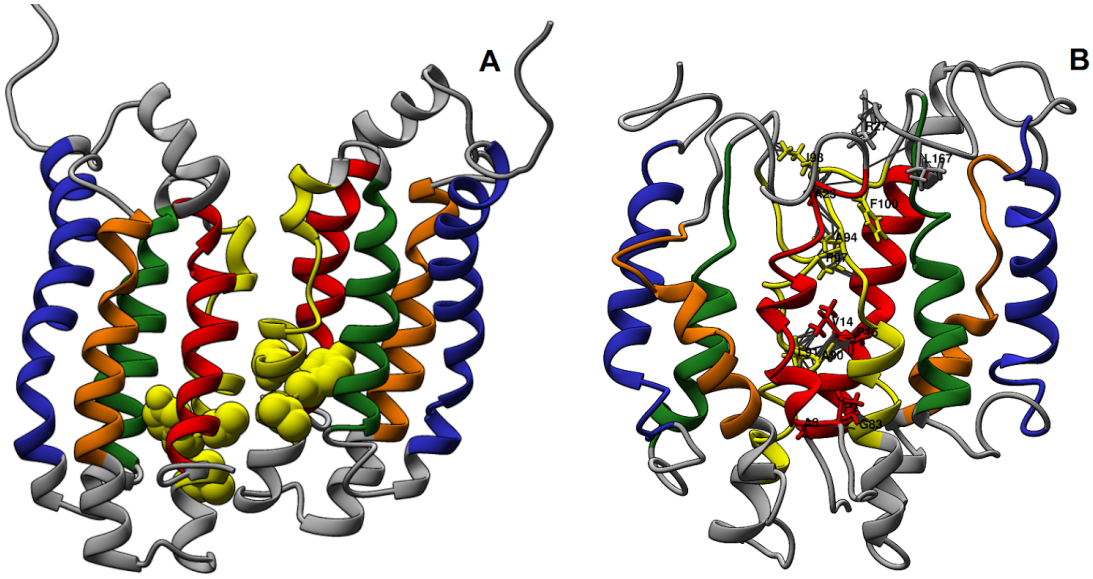
(***legend on the next page***)

Table 4.4. Table showing residue pairs from each TSPO monomer with highest interaction frequencies, as described by proportion of trajectory where given residue pairs are in <5.6Å distance of each other. Interaction frequencies were calculated using gmx mdmat tool of GROMACS and CONAN program (Mercadante et al. 2018; Berendsen et al. 1995; Páll et al. 2015) .

the dimer interactions. Among them, the Phenylalanine residues are surprisingly well-represented, despite their relatively low abundance (5 residues in mouse TSPO). As an example, the five residue pairs that are in contact for largest part of the simulation are given in Table 4.4. Examining the location of these aromatic residues, we also observed that certain residues have higher representation in strongest interacting pairs. Some of them are Tyr140, Tyr152, Phe100, Trp107 and Phe25 (Table 4.4).

Surprisingly, the interactions between small residues, such as glycine, alanine, threonine and serine are less represented, despite the fact that TSPO has abundance of 'SmxxxSm' motifs, known to mediate dimerization in various TM proteins (Zhou et al. 2001; Lemmon, Flanagan, Treutlein, et al. 1992; Duneau et al. 2007). Moreover, when looking at the distribution of accessible residues in the TM domain, surface-exposed small residues were higher in number compared to aromatic ones (36 vs. 18; residues with relative accessible area > 20 in NMR structure; PDB: 2MGY), but are less represented as contacts in our simulations. Consequently, based on an extremely crude calculations, a larger number of small residues pairing would have been expected.

As an additional support to this observation, when we started with a model based on solid-state NMR data, which involves a GxxxG dimerization motif (residues 83-87 of TM3) (Jaipuria et al. 2017), this dimer interface is not retained along the simulation. On the other hand, participation of other residues of TM3, as well as residues of TM1 significantly increased during the course of the simulation.



(***Full legend on next page***)

Figure 4.7. 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 involving TM3 region, based on the interface described by Jaipuria et. al. TM regions 1 to 5 are represented as red, orange, yellow, green and blue ribbons respectively. The contacts between interacting residues are represented by gray cylinders.

There are different possible reasons why the experimentally determined interface was less favored in our simulations: 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 and iii) the interface that involve about 5 residues (Jaipuria et al. 2017), have very low interface area, which may explain why initial interface of TM3 changed to a larger interface involving TM1 and TM3 (which might also be related to the CG model). Interestingly, the shared interface between TM1 & TM3 from our simulations is very similar to the dimer interface observed in *R. sphaeroides* (Li, Liu, Zheng, et al. 2015a) iv) the absence of ligand in our simulations might have affect the monomer dynamics and *per se* might have modified the dimer formation. However, it is important to note that our results are also in agreement with recent molecular modelling studies, where it was observed that a shared interface between TM1 & 3 is more stable than that of TM3 observed from NMR studies (Zeng et al. 2018).

4.2.4. A putative cholesterol translocation mechanism was observed:

As TSPO is assumed to be a cholesterol transporter (although it is a matter of debate), 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 using different thresholds 5 Å, 5.6 Å or 6Å). For the all set of frames (14785) and whatever the threshold chosen, on average, we did not notice any specific location with respect to the two well-known motifs supposed to bind cholesterol, CRAC (residues 149-157) and its reverse counterpart CARC (residues 135-141) located in TM5, compared to the rest of the residues. The large concentration of cholesterol (33%) might hide specific interaction since saturation effect may take place. The simulations with a lower concentration (16%) need longer times and more replicates to bring robust conclusions. Nevertheless, we tried to go further by scrutinizing each simulation separately, focusing specifically on the CARC and CRAC motifs (Fig. 4.8). We chose to examine the simulation based on NMR model (Jaipuria et al. 2017), since this model is the closest model that describe TSPO dimer with respect to cholesterol translocation. While we observed that there are variations in number of cholesterol molecules interacting with CRAC motif throughout the simulation (Fig. 4.8, A), we chose to scrutinize the time-points 650ns-2000ns, for two reasons. First, this is the first timepoint in this simulation that involves large reduction in cholesterol interaction with CRAC motif (20 molecules at time point 650ns vs. <2 molecules at time point 2000ns; Fig. 4.8, B), and

second, we chose first timepoint in simulation so that the dimer interface is not too distant from the NMR model described by Jaipuria et. al. Considering this time-point, we found a cholesterol molecule with a long residence time, close to the putative functional motifs.

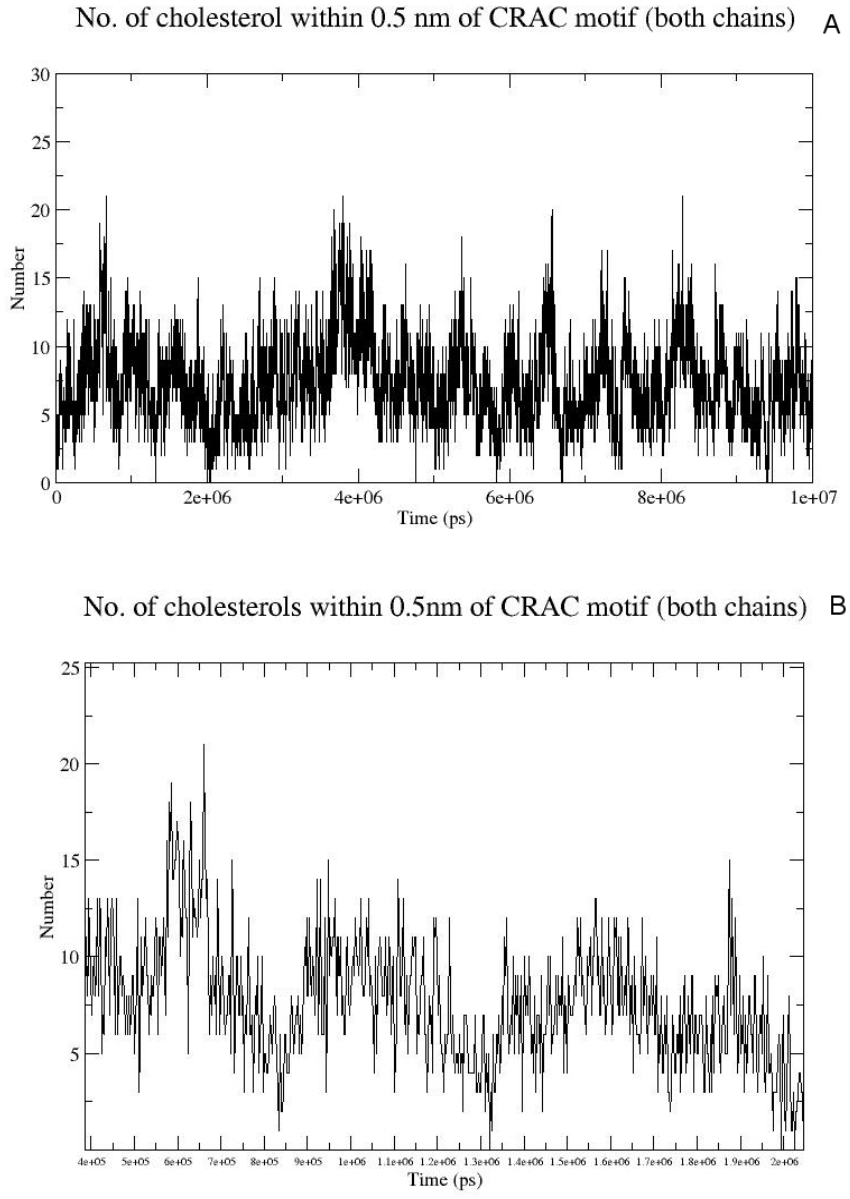


Figure 4.8. A: Plot showing number of cholesterol molecules interacting with CRAC motif in the simulation of TSPO model with interface involving TM3 region. **B**: Plot at **(A)** zoomed to show the decrease of cholesterol binding between time points 650ns-2,000ns. The interactions were calculated using g_select tool of GROMACS (Berendsen et al. 1995; Páll et al. 2015).

While many cholesterol molecules that were close to CRAC motif diffused away along the *x*-*y* plane, we found one cholesterol molecule in close contact with CRAC motif showing an interesting behavior. Upon visualization of the trajectory corresponding to these time points, we found that the cholesterol molecule was translocating between the lower and upper leaflet, by means of a flipping mechanism (Fig. 4.9). Based on the very first observations, a putative mechanism of cholesterol

transport, would consist in using leucine residues of TM5 as hooks, to help cholesterol to flip between the two leaflets. However, thorough analyses of other simulations are necessary (and are on-going) to check whether this mechanism is indeed generic.

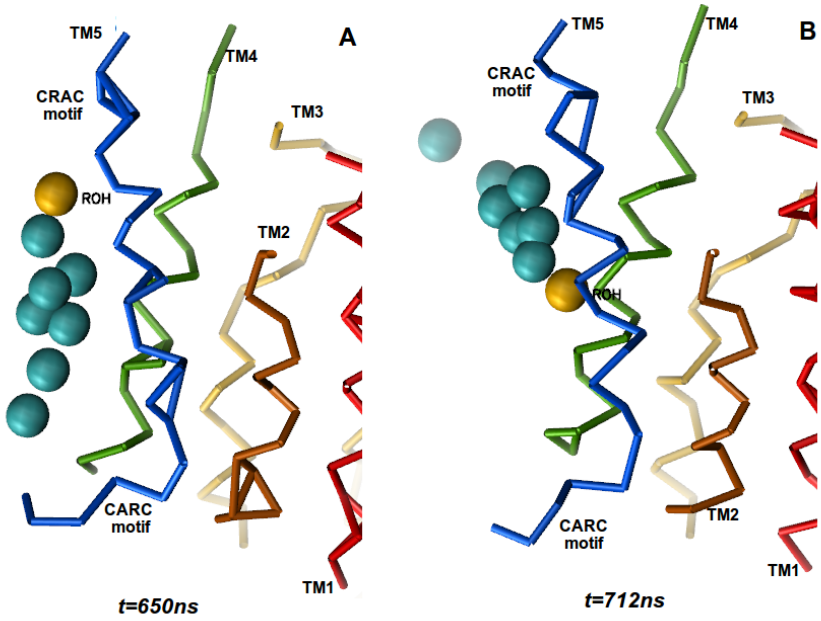
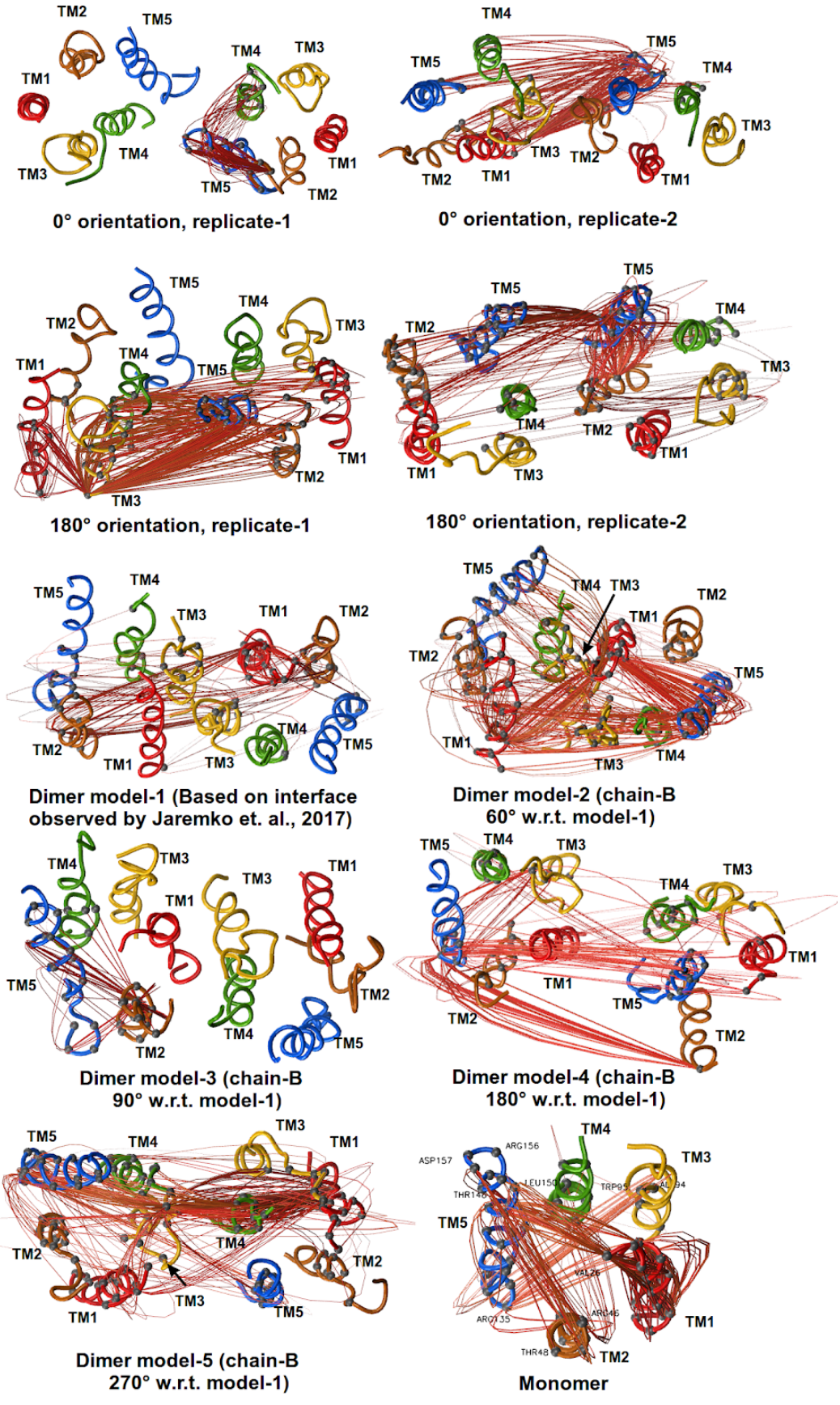


Figure 4.9. A: Snapshot of TSPO and interacting cholesterol molecule at time point 650ns; **B**: Snapshot of TSPO and interacting cholesterol molecule at time point 712ns. TM regions 1-5 are represented as red, orange, yellow, green and blue dynamic bonds respectively, interacting cholesterol molecule is represented as blue VdW beads.

4.2.5. Different long-range communication pathways were found:

Fluctuation patterns calculated for the whole set of trajectories have shown that the TSPO dimer complex is highly dynamic. The question on how these motions impact key functional regions naturally needs to be addressed. Accordingly, we turned to protein structural network (PSN) analysis that has proved to be a useful approach to pinpoint long-range interaction pathways in protein dynamics. In order to investigate the impact of the dimerization on this pathway, we calculated and compared the PSN for an ensemble of monomer conformations before dimerization and for dimers from all the simulations. We observed that in the dimer conformation, the allosteric communication indeed depends on the interface. Two types of allosteric communication were found (Fig. 4.10): A first pathway involves the two monomer chains and the communication takes place via the dimer interface. The second pathway is restricted to a single monomer chain, i.e. long-range effect involves residues of a single monomer and not the residues of the other chain. Interestingly, the inter-chain communication was dominant (observed in 7 out of 9 dimer trajectories analyzed).



(*legend in the next page*)

Figure 4.10. TSPO dimer structures indicating allosteric communication pathways in dimer orientations from different simulations in 33% cholesterol concentration, compared with an ensemble of monomer conformations taken from simulations at 0° and 180° orientations. TM regions 1 to 5 are represented as red, orange, yellow, green and blue ribbons respectively. Allosteric communication pathways were traced from the matrix containing correlations of residue motions from different simulations, using Bio3d MD analysis package (Grant et al. 2006).

On comparison between allosteric pathways in dimer with the ensemble of TSPO monomer (before the dimerization), we found that several communication pathways observed in monomer were also observed in dimer, though the TM regions involved in the communication differed from each interface. For instance, in monomer, residues of TM1 and TM5 were found to be part of an allosteric pathway. Similar residues were also found to be part of a similar pathway in the simulation of dimer model-2. We made similar observations for pathway between TM2-TM5 as well as TM3-TM5.

In the simulations where the interface was similar to the model described from NMR experiments (Jaipuria et al. 2017), we found that though in in one of the simulation, allosteric communication did not involve TM5 region (dimer model-1; Fig. 4.10), we also observed that in dimer model-2, which has similar interface as dimer model-1 (involving TM1 and TM3 regions from each monomer) has allosteric communication involving TM1, TM3 (interface regions) and TM5, similar to the pathway observed by Jaipuria et al.

Concerning the time-points in which we observed CRAC/CARC accompanying cholesterol translocation (Fig. 4.10), we also observed an allosteric pathway spanning the dimer interface and the TM5 region, similar to what was observed from the NMR experiments (Fig. 4.11) (Jaipuria et al. 2017), even though different sets of residues were involved. 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 involved in this pathway (residues 135 and 136 of TM5) [14,15]. Thus, we may hypothesize that these dynamics may also be associated with diffusion of cholesterol between the two leaflets (Fig. 4.10).

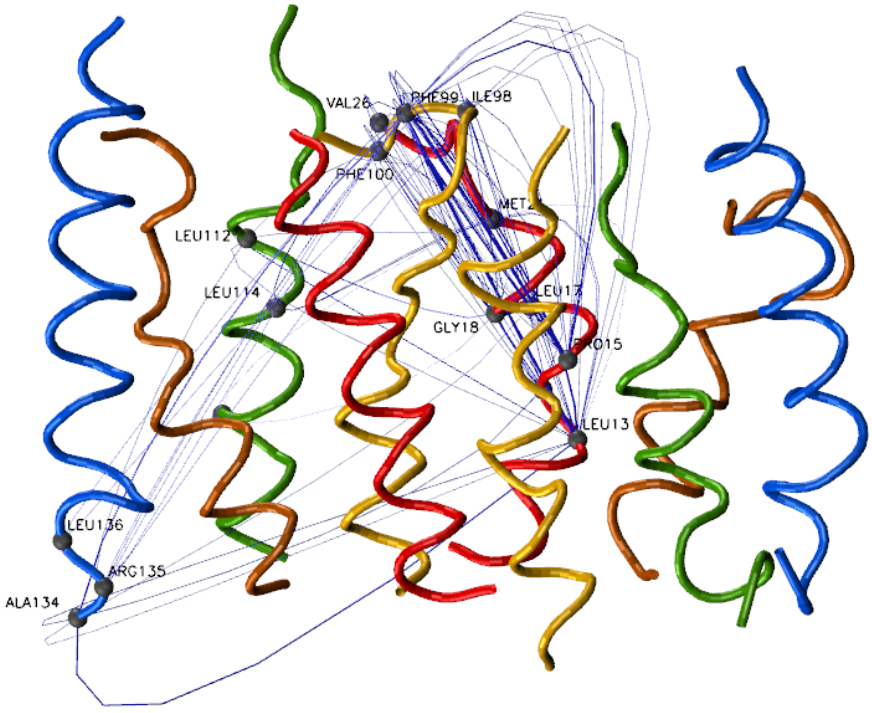


Figure 4.11. Representation of long-range communication pathway between the dimer interface and residues of TM5 region, occuring between the time points 650ns-2000ns of TSPO dimer model-1, where a putative translocation of cholesterol was also observed.

4.3. DISCUSSION

4.3.1. Asymmetric interfaces and existence of higher order oligomers

X-ray structures and NMR studies revealed existence of TSPO in different oligomeric orientations (Guo et al. 2015; Li, Liu, Zheng, et al. 2015a). Upon visualisation of representative structures from our simulations, we found that TSPO dimers exist with different combinations of TM regions as interfaces (Fig. 4.6, B). We observed many interfaces that were previously not observed in any of the experimental studies, suggesting that our simulations have sampled wider interaction space than interfaces sampled from previous experimental studies. The presence of multiple preferred complexes suggests that the interaction between the TSPO monomers may be opportunistic, or it may hint towards existence of TSPO as higher-order oligomers. This has already been confirmed from the X-ray crystallographic studies of *R. sphaeroides*, where the trimer interface involved TM regions 1 and 3 for first pair of dimers, and TM regions 1, 2 and 5 for the second pair of dimers (Li, Liu, Zheng, et al. 2015a). However, earlier Cryo-EM studies could not confirm the presence of TSPO as a higher-order oligomer (Hinsen et al. 2015; Korkhov et al. 2010; M. Jaremko et al. 2015). Furthermore, presence of higher-order TSPO oligomers have been documented in previous experimental studies (Lacapère et al. 2001; Papadopoulos et al. 1994), though they have not been characterised yet at an atomic scale, apart from the crystal structure of *R. sphaeroides* TSPO (Li, Liu, Zheng, et al. 2015a).

While we have observed different dimer interfaces, we also observed some interfaces that were similar to the interfaces observed from X-ray studies. X-ray structures of *R. sphaeroides* TSPO was found to be a trimer, with the first interface being shared between TM1 &3 of both monomers, and the second interface shared between TM1&2 of first monomer with TM5 of second monomer (Li, Liu, Zheng, et al. 2015a). We observed presence of both the interfaces in two separate simulations. In contrast, the interface of *B.cereus* TSPO dimer, involving TM2 from each monomer was not observed from our simulations (Guo et al. 2015). On the other hand, the dimer model of TSPO dimer observed from NMR studies (Jaipuria et al. 2017), which we attempted to reproduce did not persist in our simulations (Jaipuria et al. 2017) but we found a similar allosteric mechanism exists between the dimer interface and the TM5 region. Overall, these observations show that our simulations are supported by some experimental observations but also that dimer formation can be accomplished through a wider range of possibilities.

4.3.2. 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 (Jaipuria et al. 2017). A similar motif is also part of the dimer interface in *B. cereus* TSPO (Guo et al. 2015). GxxxG motifs have been observed to be part of dimer interface in a variety of TM proteins (Sulistijo and Mackenzie 2009; Lemmon, Flanagan, Hunt, et al. 1992; Lemmon, Flanagan, Treutlein, et al. 1992; Pogoryelov et al. 2012; Li et al. 2012), such that they have been often hypothesized as the primary drivers of interactions between TM helices (Kleiger et al. 2002; Mueller et al. 2014). This motif, where Glycine is the key element, has been extended to include alternate residues to glycine, which include small residues such as alanine, serine or threonine at either of the ends, with three residues in between, which may be represented by residues of any chemical nature. The general motif is referred to as 'SmxxxSm' motif (Eilers et al. 2002; Duneau et al. 2007) and has been observed to be abundant in TM proteins. The interaction by such motifs involve hydrogen bonds between the backbone atoms of the small residues (Mueller et al. 2014; Li et al. 2012; Weber and Schneider 2013). Aromatic residues have also been documented to mediate TM oligomerization. In certain proteins, tryptophan residues have been found to be a primary driver of oligomerization, by means of planar π-π interactions (Zoued et al. 2018; Ridder et al. 2005). Other aromatic residues such as phenylalanine and tyrosine have also been found to be part of oligomer interactions, where aromatic residues supplement the oligomerization by SmxxxSm motifs (Unterreitmeier et al. 2007).

In case of TSPO, despite the significant presence of SmxxxSm motifs, they are found to be less involved in the dimerisation interface (Table 4.4). In contrast, the interface was preferentially filled of large residues, mainly aromatics residues. Among the aromatic-aromatic interactions, Tyr138 and Tyr140, both of which are part of an alternate cholesterol binding motif called 'CARC' motif (Fantini

et al. 2016; Papadopoulos et al. 2017) were found to stabilize the interface. Although similar interactions might occur between tyrosines located at other TM regions, particularly TM1 and TM2, we did not observe such interactions in our set of simulations. Interestingly, in a previous experimental study (Delavoie et al. 2003), upon treatment with ROS, TSPO was observed to form oligomers that were mediated by formation of covalent dityrosine interactions. However, accessibility of CRAC as well as CARC motifs are affected in such oligomeric arrangement. Since we observed another cluster of dimers involving TM3 as the interface, we hypothesize that the dityrosine-mediated interface involving TM5 may be involved in response to ROS, whereas the interface involving TM 1 and 3 may be involved in cholesterol transport, due to the fact that CRAC as well as CARC motifs are more accessible in such arrangement. Furthermore, we have also observed flipping of cholesterol across the membrane in such an arrangement.

The fact that NMR interfaces are not retained may be due to different reasons. A first one 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 shown that the presence of a ligand affects the dynamics of a monomer, 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. However, the nature of the ligand itself will be a matter of choice, which will influence the results.

A second reason may be the choice of the coarse-grained model, the Martini force field, which could have favored some interactions, in particular between aromatic residues and consequently, have biased the results. This has also been observed in previous studies involving Martini force fields (Javanainen et al. 2017). Thus, tests with another CG force field should be conducted to confirm the present findings. The SIRAH force field (Machado et al. 2019), which is based on very different considerations compared to Martini Force Field will be an appropriate choice. As we rebuilt all-atom force structures from the most representative conformations extracted from the simulations, we also plan to explore the stability of these structures by running longer simulations.

4.3.3. Putative translocation mechanism of cholesterol

Another interesting observation made from our studies is the presence of a putative translocation mechanism, where the cholesterol molecule uses TM5 region as a tether to flip between the upper and lower lipid bilayer leaflets (Fig. 4.9). While this is indeed an interesting and an exciting observation, existence of this mechanism needs to be corroborated rigorously in future studies, and this also raises many questions that may be explored in future. It remains to be seen if a cholesterol in a similar bilayer system with similar lipid-cholesterol composition (except TSPO) can demonstrate such

flipping mechanism even in the absence of TSPO. If it fails to do so, it may be hypothesized that interaction with TSPO may possibly reduce energy barrier associated with flipping between two leaflets. Furthermore, it also needs to be observed if such translocation can be observed in all-atomistic simulations, and this mechanism is possible with other interfaces observed from our simulations. Ultimately, the role of TSPO ligands also cannot be discounted. Since several experimental studies have demonstrated an effect of TSPO ligands on cholesterol translocation (Krueger and Papadopoulos 1990; Papadopoulos et al. 1991; McCauley et al. 1995), question of whether this flipping mechanism can be observed in ligand-bound environments could also be explored.

4.3.4. Prevalence of long-range communications in varied dimer interfaces

We initially hypothesized that since we observed different interfaces, each of the interface may have unique pathway. In agreement, we indeed observed unique pathways depending on the interface. However, we found a few common denominators across all the pathways. One, we found that the allostery involving inter-chain communication involves transfer of the information via the dimer interface. Two, interestingly, we found that TM5 region is involved in such communications in all the simulations, with the exception of the simulation of dimer-model-1. Though the interface in this model is similar to that observed by Jaremko et al., we also observed that in dimer model-2, which has same interface as dimer model-1 (involving TM1 and TM3 regions from each monomer) has allosteric communication involving TM1, TM3 (interface regions) and TM5, similar to the pathway observed by Jaipuria et. al.

Though we observed inter-dimer communication in majority of our simulations, we found that intra-dimer communication without the involvement of the other monomer is also present in certain cases. The question of if such dichotomy of modes of allostery (inter-dimer communication vs. monomer communication) reflects different energy landscape of TSPO dimer is unclear at this moment. Furthermore, it is also unclear if relative dominance of inter-dimer communication may reflect bias in the sampling performed.

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. However, it remains to be seen what is the role of presence of cholesterol in such communication pathways, since the simulations were performed in high concentration of cholesterol (33%). Furthermore, since all the simulations were performed in absence of any ligand, it would also be interesting to understand what effect does the binding of a ligand have in such communication pathways. The above questions could be addressed in future by performing similar molecular dynamics studies in different environments, i.e., in absence of any cholesterol and in presence of a ligand bound to TSPO.

4.4. CONCLUSIONS

TSPO is a protein from which we do not possess any certainty of its real role, but it is related in some way to important biological processes, making the study of its dynamic behavior very important. Particularly, the question about the dimerization and oligomerization has to be addressed. Here, we study dimerization of TSPO from different perspectives, including dimer interface and long-range communication, which is the first one conducted in such an exhaustive way. It brings new insights in the process and provide key elements about the nature of interactions that contribute to the formation of the dimers. Indeed, several interfaces were observed, even if TM3 is the most retrieved region at the interface of the dimer formed during simulations. The existence of several possible interactions could open the way to the possibility of existence of higher-order TSPO oligomers. The results we obtained also confirms that TSPO has noteworthy dynamic properties that make it a very difficult protein to study. For example, we observed presence of different allosteric pathways depending on the nature of the dimer interface, with common denominator being presence of long-range communications involving the interface in many of our simulations. Among the numerous pathways, we also observed a pathway involving the dimer interface (TM3) and putative cholesterol binding regions (CARC motif of TM5), not different from the pathway described by Jaipuria et. al. Despite the dynamic nature of TSPO, we were able to capture a putative translocation mechanism of cholesterol between two lipid bilayer leaflets. Our findings open up new set of questions on TSPO dynamics-function relationships, which could be explored in future studies.

Chapter 5: Comparative dynamics of mammalian and bacterial TSPO

Table of contents

5.1. Introduction 150
5.2. Results 155
5.2.1. Bacterial and eukaryotic TSPO explored unique sets of conformations 155
5.2.2. Similar dynamics were observed between certain regions of bacterial and mouse TSPO despite difference in conformational exploration 159
5.2.3. PP-IX binding residues are conserved across bacteria as well as eukaryotes 162
5.3. Discussion 165
5.4. Conclusions and future directions 167

5.1. Introduction

TSPO is a well-conserved protein, with its homologues conserved throughout the species of all the three taxonomic domains (Fig. 5.1). Among the bacterial TSPO whose structures are available (*R. sphaeroides* and *B. cereus*), the sequence identity of the respective structures with mouse TSPO are 36% and 24% respectively. Though the sequence identity is not high enough to be considered as close homologues, the sequences are identical enough to be considered homologous at the level of tertiary structure. Furthermore, this is consistent with the existing dogma that though sequences are not necessarily conserved, conservation exists among remote homologs at the level of secondary and tertiary structure.

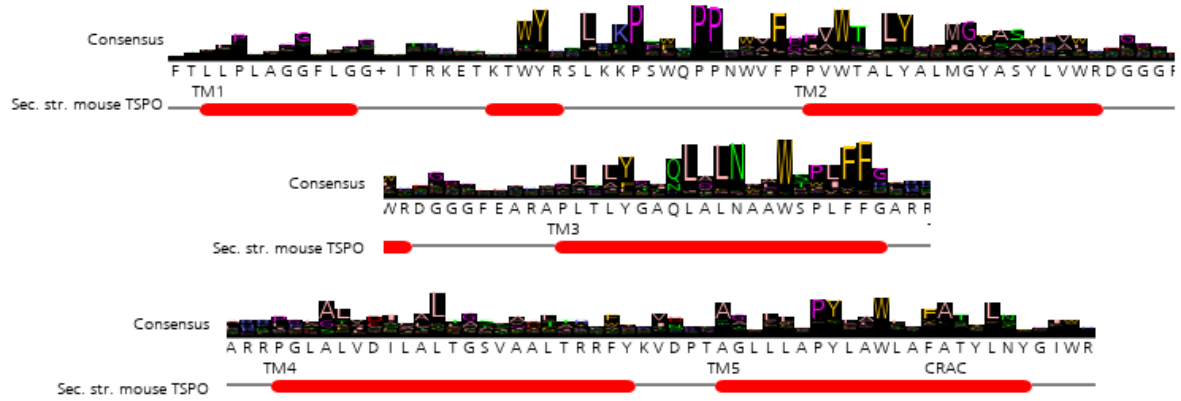


Figure 5.1. Consensus logo of sequences comprising TSPO domain, aligned with the TM boundaries of mouse TSPO sequence. The consensus logo was generated and visualised using Jalview program (Waterhouse et al. 2009), based on multiple sequence alignment of 517 TSPO sequences (including mouse TSPO sequence), used to classify the TSPO domain, retrieved from Pfam database (Bateman et al. 2004; Finn et al. 2016; El-Gebali et al. 2019). Secondary structure of mouse TSPO sequence was calculated and projected on the consensus plot using JPred plugin of Jalview program (Drozdetskiy et al. 2015; Waterhouse et al. 2009).

However, on comparison of the experimental TSPO structures with each other, a rather contradictory picture emerges. Superposition of mouse TSPO structure with those of bacterial TSPO structures reveal significantly high RMSD between the two structures (Fig. 5.2). Between the mouse and *R. sphaeroides* TSPO, RMSD is 4.8 Å, and RMSD between mouse and *B. cereus* TSPO is 5.34 Å (Fig. 5.2). On the other hand, RMSD between the two bacterial structures is 3.89 Å, though the sequence identity between the two bacterial TSPO is 22%. This was also observed previously (Guo et al. 2015) from comparison between the structures of *B. cereus* and mouse TSPO, as well as between *R. sphaeroides* and mouse TSPO ((Guo et al. 2015; Li, Liu, Zheng, et al. 2015a; Ginter et al. 2013; Wendler et al. 2003; Vanhee et al. 2011)).

Furthermore, it was also found that the mode of PK11195 binding is different between the two structures, though the binding cavity itself is well-conserved. That being said, the differences may also arise from the fact that the bacterial and mammalian structures were solved from different experimental techniques. In both the TSPOs, different detergents were used to solubilize the protein.

While the structure of mouse TSPO was solved by solubilizing in Dodecylphosphocholine (DPC) detergent (Jaremko, Jaremko, Giller, et al. 2014), the structure of *B. cereus* TSPO was solved by solubilization in N-octyl-β-D-glucopyranoside (β-OG) detergent (Guo et al. 2015). It has been previously suggested that use of DPC detergent used to solve the NMR structure may have modified TSPO structure, providing a possible explanation to the high RMSD between mouse and *R. sphaeroides* TSPO structures (Li, Liu, Garavito, et al. 2015). The differences are also aggravated by the lack of adequate TSPO structures. While the TSPO structures of *B. cereus* are available as monomers and dimer in *apo*- and *holo*-forms, there are no monomer structures available for *R. sphaeroides* TSPO. On the other hand, there are no dimer and ligand-unbound structures available for mouse TSPO. This further complicates the comparison of TSPO structures.

Thus, the following questions arise: Are there intrinsic differences in the structures of mammalian and bacterial TSPO, or can the differences be attributed to the experimental techniques? Do the differences in structure also translate to differences in dynamics? Do the differences also extend to their functions, due to the fact that bacterial TSPO have been primarily associated with porphyrin metabolism, whereas mammalian TSPO have been associated with variety of other functions, including cholesterol transport apart from porphyrin metabolism? We attempted to address these questions by performing coarse-grained molecular dynamics of bacterial as well as mammalian TSPO, and comparing the dynamics of both the TSPO. In addition to the simulations with experimental structures as the starting point, we constructed a homology model of mouse TSPO using *R. sphaeroides* A139T mutant TSPO structure (PDB:4UC1), since this structure is solved in relatively high resolution (1.8Å), higher sequence identity compared to *B. cereus* TSPO. We performed simulations on this homology model, since NMR structure has high RMSD compared to X-ray crystallographic structures, we attempted to answer if simulations are performed on a X-ray structure-based homology model of mouse TSPO, will the differences in conformations found in NMR vs. X-ray crystal structures be reflected in the conformational space explored by X-ray crystal structure-based homology model of mouse TSPO as well? The protocol followed for the simulations and their analysis is shown as a flowchart in following pages:

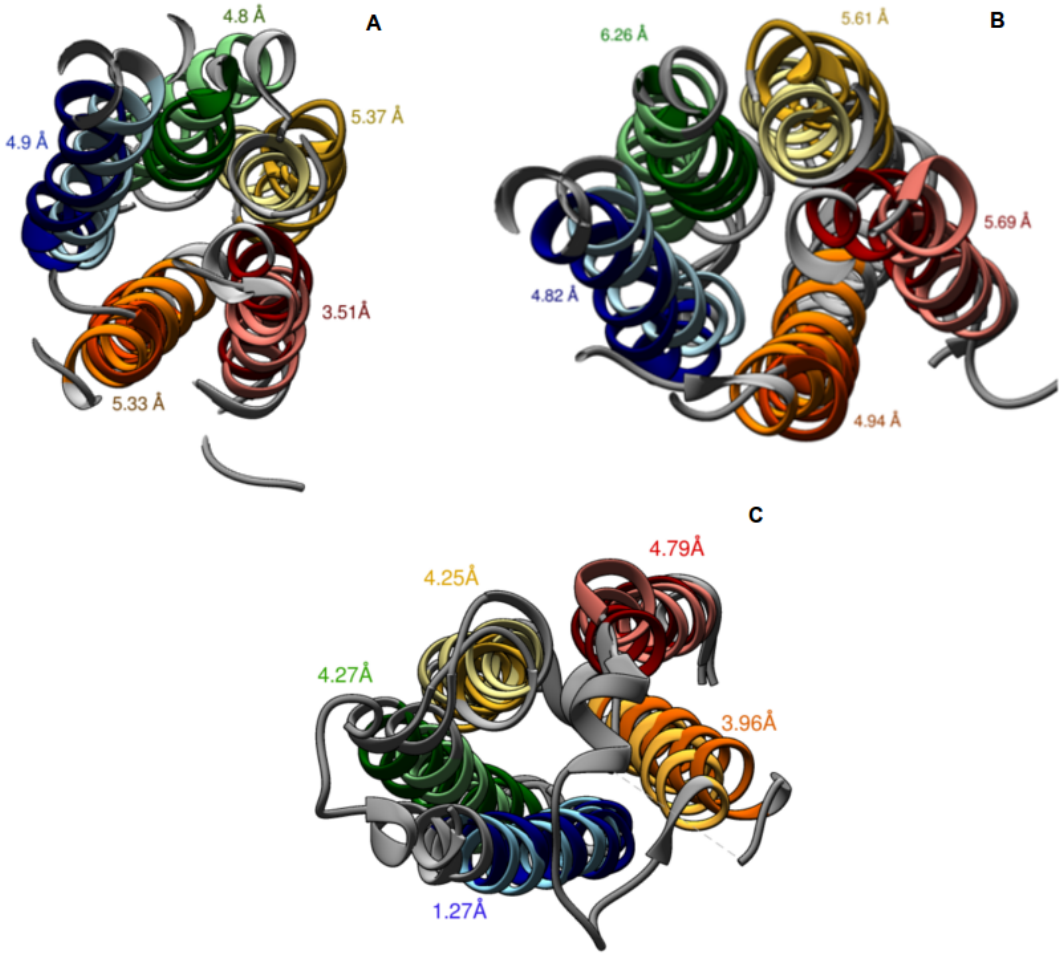
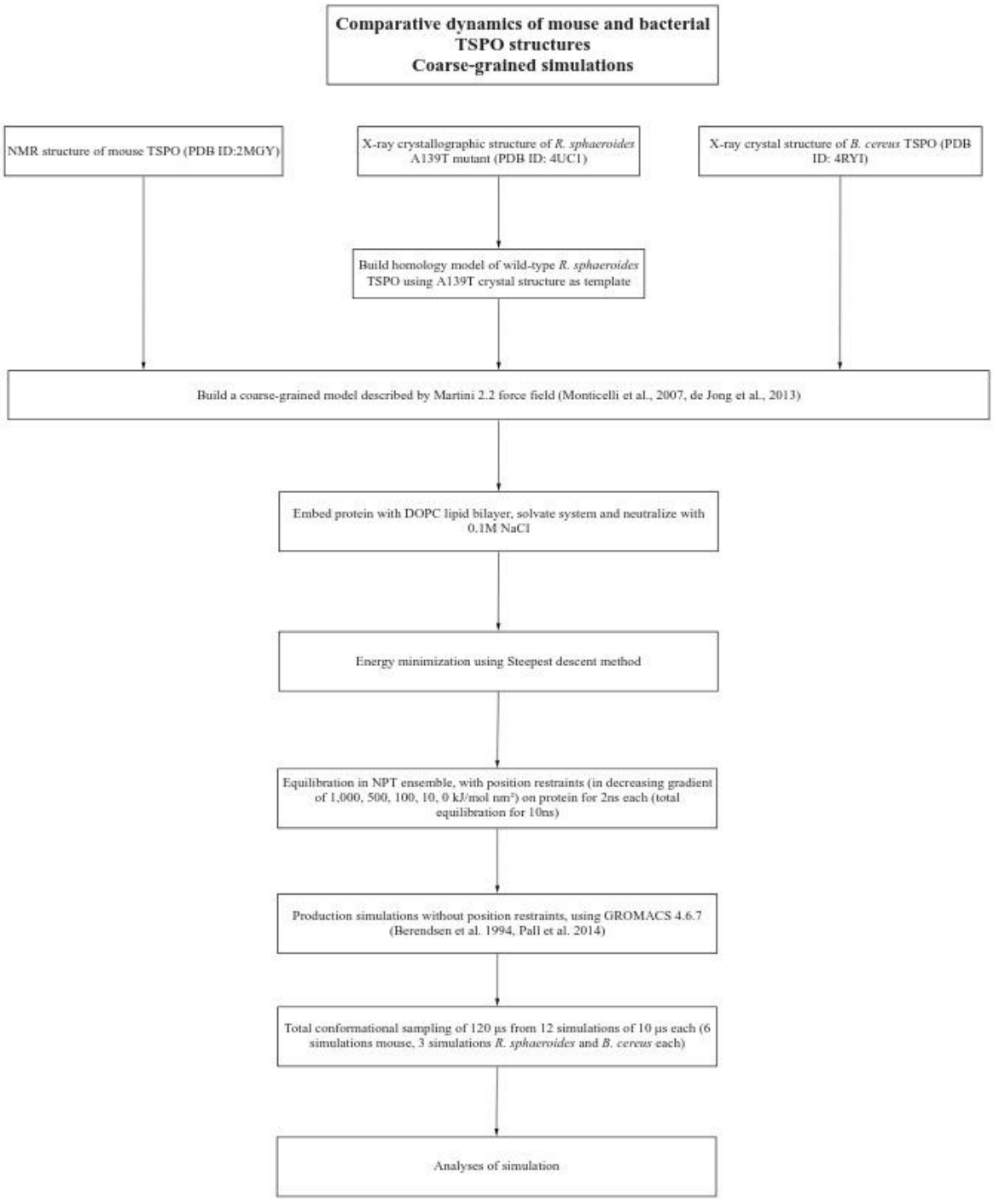
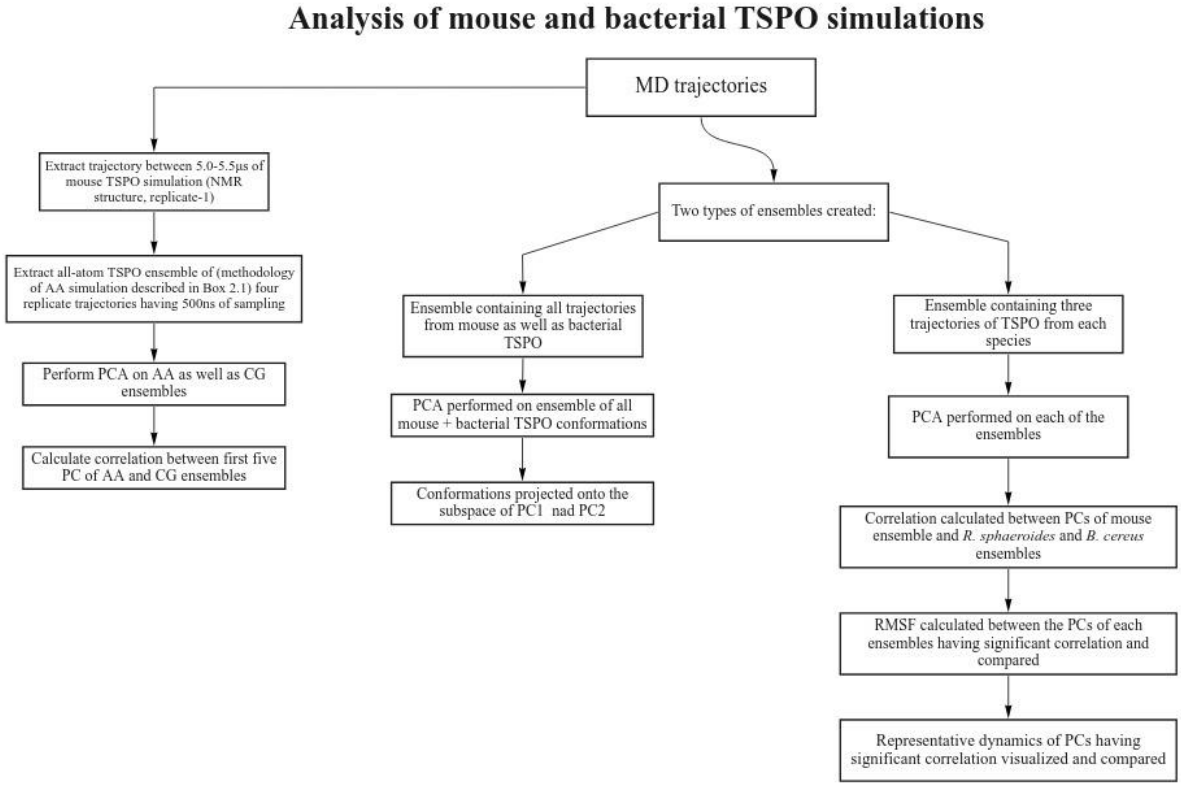


Figure 5.2. A: Structural superimposition of mouse (PDB ID: 2MGY) and *R. sphaeroides* (PDB ID: 4UC1) TSPO structures. **B:** Structural superimposition of mouse (PDB: 2MGY) and *B. cereus* (PDB ID: 4RYO) TSPO structures. RMSD between the residues of TM domain from both superpositions are 4.84 Å and 5.34 Å respectively, and sequence identity between the mouse and bacterial sequences are 36% and 24% respectively. TM regions of mouse TSPO are represented in dark shades, and that of bacterial TSPO are represented in light shades. **C:** Structural superimposition of *R. sphaeroides* TSPO (PDB ID: 4UC3) and *B. cereus* TSPO (PDB ID: 4RYI). RMSD between residues of TM domain is 3.89Å, and identity between the two sequences is 22%. TM regions of *R. sphaeroides* TSPO are represented in dark shades, those of *B. cereus* are represented as light shades. TM1-5 of all the structures are represented as red, orange, yellow, green and blue cartoons respectively, and RMSD between residues of respective TM regions are shown adjacent to TM regions. All the superpositions were performed on the C-alpha atoms of the residues of TM regions 1-5. Superposition and RMSD calculations were performed using UCSF Chimera program (Pettersen et al. 2004).



Box 5.1: Flowchart depicting preparatory steps performed for coarse-grained molecular dynamics simulations of mammalian and bacterial TSPO. A detailed description of the methods employed are provided in chapter 2.



Box 5.2: Flowchart describing analysis performed to compare dynamics of mammalian and bacterial TSPO. The methods used for analysis is described in detail in chapter-2.

Parameters for simulations

System:

- Forcefield: Martini 2.2
- Lipid type: DOPC
- Type of ions: NaCl (0.1M)
- Dimensions of simulation box: 10x10x10nm

Energy minimization parameters:

- Method: Steepest descent
- No. of steps: 50,000

Equilibration:

- Position restraints (force constant): 1000, 500, 100, 10, 0 kJ/mol nm²
- Equilibration time: 1ns x 5 simulations
- Temperature: 300K
- Thermostat: Berendsen
- Barostat: Parinello-Rahman with semiisotropic pressure coupling

Production MD simulation:

- Same parameters as above except position restraints

Box 5.3. Parameters used for coarse-grained simulations of mouse and bacterial TSPO.

5.2. Results

5.2.1. Bacterial and eukaryotic TSPO explored unique sets of conformations

In an attempt to answer the above-mentioned questions, we performed coarse-grained molecular dynamics simulations of mouse TSPO (PDB ID: 2MGY), *R. sphaeroides* TSPO and *B. cereus* TSPO (PDB ID: 4RYI) structures. Since wild-type *R. sphaeroides* TSPO has many missing residues, the wild-type *R. sphaeroides* TSPO structure was modelled from the X-ray structure of *R. sphaeroides* A139T mutant (PDB ID: 4UC1). We used A139T mutant structure to construct the model since the mutant structure is available at a higher resolution (1.8Å) compared to the wild-type structure (2.5Å).

We also performed simulations on a homology model of mouse TSPO, which was constructed from X-ray crystallographic structure of *R. sphaeroides* TSPO A139T mutant (PDB: 4UC1, 36% identity with mouse TSPO) as a template. All the simulations together account for a total sampling of 120μs. A summary of trajectories is described in Table 5.1. To ensure that the coarse-grained model used to describe TSPO structures have similarity with that of all-atom models, the dynamics of mouse TSPO trajectory in absence of a ligand was compared with that of equivalent all-atom trajectory (ensemble of TSPO conformations from simulations in absence of PK11195 ligand; ensemble consist of 4 replicate simulations: simulation-1, 2, 3 & 5, each 125ns, table 3.1, chapter-3) simulated with Amber99-SB ILDNP force field (Aliev et al. 2014). The coarse-grained trajectory used for comparison are the sub-trajectories corresponding to 2.5μs-3.0μs and 5μs-5.5μs time-points of replicate-1 of mouse TSPO trajectory (simulation from NMR structure, table 5.1). Upon calculation of RMSD for the coarse-grained trajectory, on the backbone beads of residues of TM regions (residues listed on table 5.2), we observed that RMSD is in flux throughout the simulation (Fig. 5.3A). RMSD however, remained at plateau at time-points 2.5μs-3.0μs and 5μs-5.5μs (Fig. 5.3A). We performed PCA on all atom as well as both the coarse-grained sub-trajectories, and compared the first five principal components of all atom trajectory versus coarse-grained trajectories, by calculating Pearson's correlation values between respective principal components, and projected it as a heatmap. We observed that despite differences between coarse-grained and all-atom models, there are overlaps between certain principal components of both all-atom and coarse-grained trajectories (Figure 5.3B, C), indicating that there are similarities between the motions of these principal components. It may also be noted that top principal components (PC1 of all atom trajectory with PC2 of coarse-grained trajectory) also have correlation values of above 0.3 (Fig. 5.3 B, C).

To understand if similarities exist on the conformations explored by bacterial and mammalian TSPO, an ensemble was created consisting of trajectories of both bacterial and mammalian TSPO. PCA was performed on this ensemble, to understand if there is an overlap in the conformations explored by bacterial and mammalian TSPO on the subspace of two largest principal components. PCA was performed on the backbone beads of the common core of TM regions of TSPO from three species.

Organism	Simulation	No. of particles	Time of simulation (μs)	Velocity seed
Mouse	1	TSPO: 1 DOPC: 364 Water: 6459 Na: 45 Cl: 49	10	173529
	2		10	-1
	3		10	-3
	4		10	-5
R. sphaeroides	1	TSPO: 1 DOPC: 311 Water: 4744 Na:34 Cl: 35	10	-3
	2		10	-5
	3		10	-7
B. cereus	1	TSPO: 1 DOPC: 311 Water: 4738 Na:33 Cl: 37	10	-3
	2		10	-5
	3		10	-7
Mouse TSPO homology model constructed with R. sphaeroides TSPO as template (PDB: 4UC1)	2	TSPO: 1 DOPC: 311 Water: 4744 Na:32 Cl: 36	10	-3
	3		10	-5

Table 5.1. Table showing summary of simulations performed with preparatory steps.

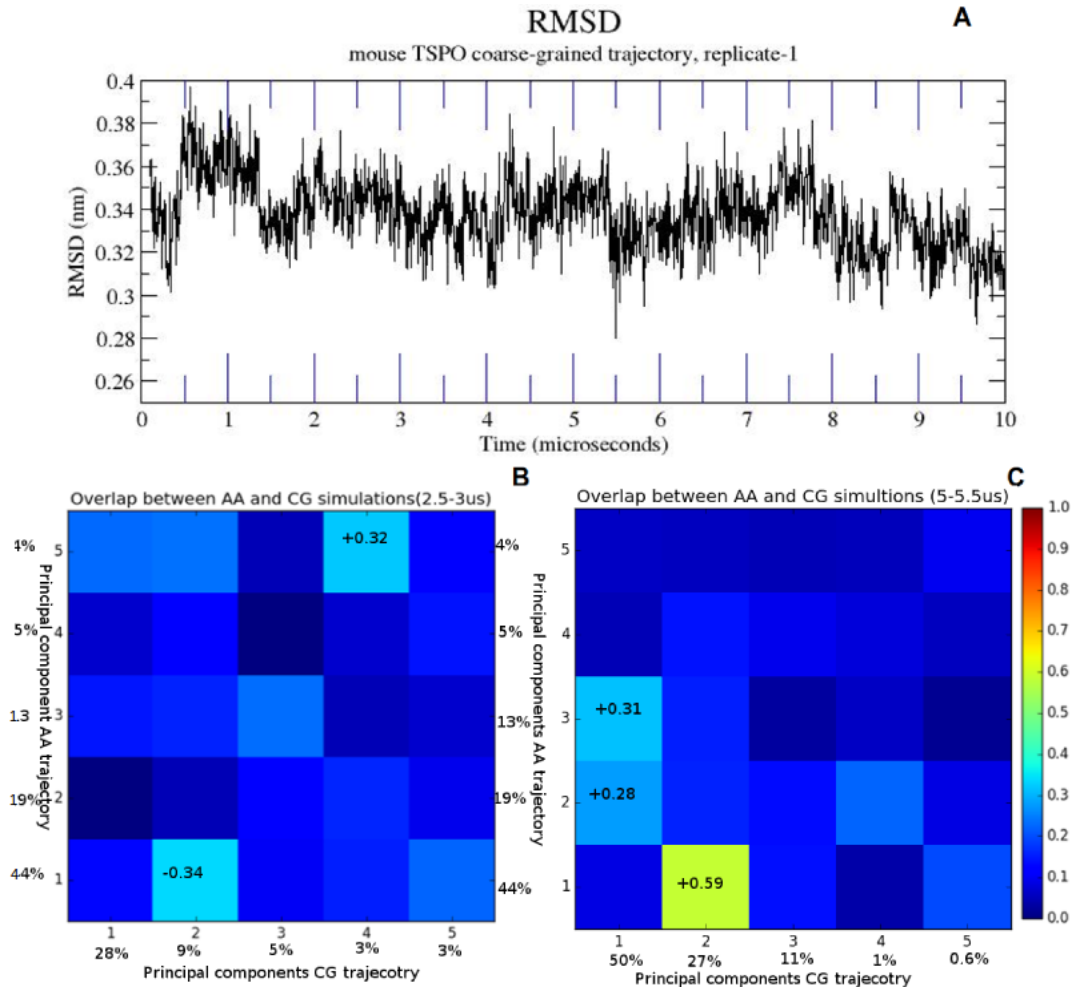


Figure 5.3. A: RMSD of coarse-grained mouse TSPO (NMR structure as starting structure) trajectory replicate-1 (table 5.1). RMSD was calculated on the backbone beads of residues of TM regions (described in table 5.2). Calculation was performed using g_rms tool of GROMACS program (Berendsen et. al., 1994, Pall et. al., 2015). **B:** Heatmap showing correlations between the top principal components representing TSPO coarse-grained dynamics and equivalent all-atom dynamics. PCA was performed on all-atom ensemble of four trajectories with sampling of 500ns. The same was performed on 2.5µs-3.0µs time-points of a coarse-grained trajectory of 10µs. **C:** Heatmap showing correlations between top principal components of coarse-grained trajectory (5.0µs-5.5µs time-points of replicate-1 of mouse TSPO trajectory, with NMR structure as starting point, as described in table 5.1). PCA was performed and correlations calculated using ProDy MD analysis package (Bakan et al. 2011).

The transmembrane common core was determined based on the membrane boundaries described for TSPO structures of respective species from OPM database (Lomize et al. 2012; Lomize et al. 2011). The boundaries were tinkered so that the number of residues across the TSPO of all the three species are the same (Annexure-3). The boundaries used for analysis are described in Table 5.2. We found that there was no overlap of mouse and either of bacterial conformations projected onto the space of first two principal components (Fig. 5.4). Thus, we found that depending on the species, TSPO explored unique sets of conformations (Fig. 5.4). However, we also observed overlap to a certain extent

between the conformations of mouse TSPO model based on *R. sphaeroides* structure, and the conformations of *R. sphaeroides* trajectories (Fig 5.4).

TM region	**Mouse**	***R. sphaeroides***	***B. cereus***
TM1	7-24	7-24	8-25
TM2	46-63	45-62	45-62
TM3	82-101	71-90	75-94
TM4	106-124	100-118	104-122
TM5	134-153	124-143	127-146

Table 5.2. Table describing the boundaries of TM regions, which were employed for all the analyses of simulations of bacterial and mouse TSPO. The TM boundaries are based on the TM definitions from OPM database (Lomize et al. 2012; Lomize et al. 2011). The definitions were tinkered so that number of residues across all the three TSPO are equal. The difference between TM definitions of OPM database and TM definitions used here are described in Annexure-3**.**

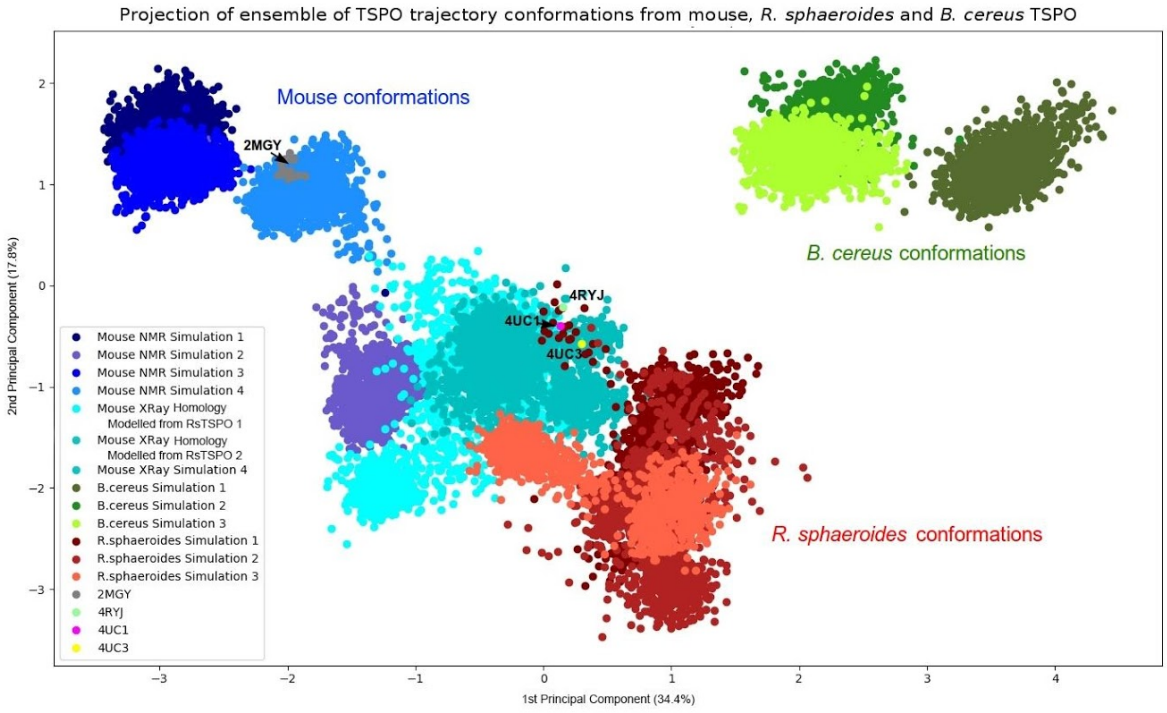


Figure 5.4. Plot showing projections of mammalian and bacterial TSPO conformations onto the conformational space of PC1 and PC2. Conformations corresponding to mouse TSPO (NMR structure), mouse TSPO (homology model constructed from *R. sphaeroides* TSPO A139T mutant, PDB: 4UC1), *R. sphaeroides* and *B. cereus* TSPO are represented by different shades of blue, cyan, red and green respectively. The conformations of experimental structures of mouse, *R. sphaeroides* wild-type and mutant, *B. cereus* TSPO are represented as gray, yellow, magenta and lime points respectively. PCA on the trajectory ensemble was performed using Prody MD analysis package (Bakan et al. 2011).

5.2.2. Similar dynamics were observed between certain regions of bacterial and mouse TSPO despite difference in conformational exploration.

While we observed that different sets of conformations were explored by TSPO from each species (Fig. 5.4), we attempted to understand if there is a consensus on the direction of essential dynamics, that may be localized to certain regions. To address this question, we constructed species-specific ensembles, by concatenating trajectories from TSPO of respective species (representing 30μs of sampling with each ensemble replicates 3 & 4 from mouse TSPO using NMR structure as starting point, replicate-2 using mouse homology model as starting point; Table 5.1), and we performed PCA on each of the species-specific ensembles. We also compared top five principal components of mouse TSPO with bacterial TSPO by calculating Pearson's correlation between the top principal components. Though we did not find correlations between all the PCs, we did observe overlap to a certain extent between PCs of mouse and bacterial TSPO (Fig. 5.5).

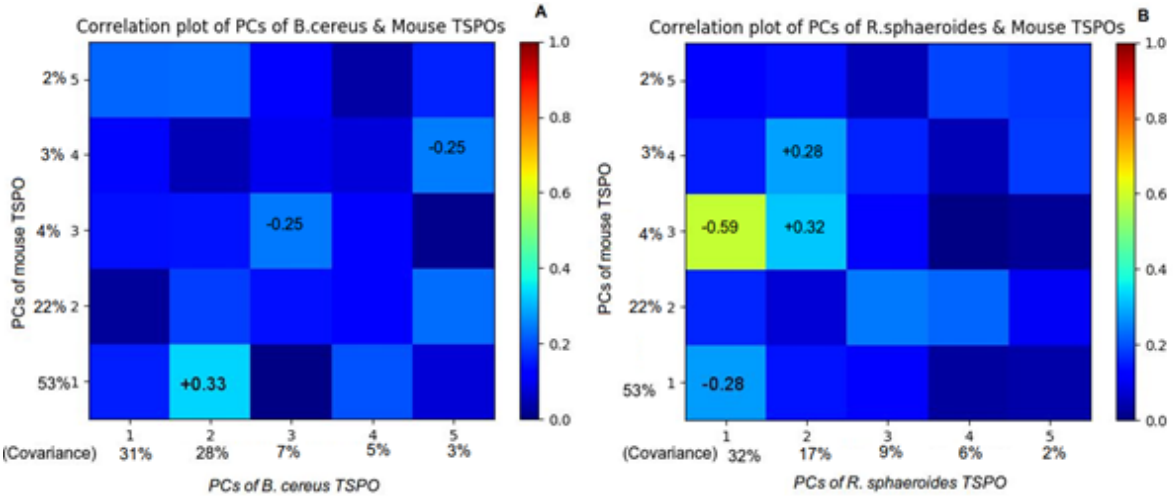


Figure 5.5. A: Heatmap showing correlations between PC 1-5 of mouse TSPO with PC 1-5 of *B. cereus* TSPO. **B:** Heatmap showing correlations between the motions of PC 1-5 of mouse TSPO and *R. sphaeroides* TSPO. The values on the plot represent Pearson's correlation value between the respective principal components. Correlation values > 0.25 are shown on the respective blocks of the heatmap. Percentage covariance of the PC is represented along the axis of the respective PC. PCA was performed and correlations calculated using ProDy MD analysis package (Bakan et al. 2011).

The first PC of mouse TSPO ensemble has correlation of +0.33 with the second PC of *B. cereus* TSPO ensemble, both accounting for 53% and 29% of covariances respectively (Fig. 5.5). Similarly, the first PC of mouse TSPO ensemble has correlation of -0.28 with the first PC of *R. sphaeroides* TSPO, both accounting for 53% and 32% of covariances respectively (Fig. 5.5). Apart from these two PC, other pairs of PC are also observed to have overlaps. Notable example is the PC1 of *R. sphaeroides* ensemble has high correlation value of –0.59 with PC3 of mouse ensemble, and PC2 of *R. sphaeroides* ensemble has overlap value of +0.32 with PC3 of mouse ensemble. However, mouse PC3 has a small contribution towards the overall dynamics (4% of covariance).

To understand if the correlations between top PC are localized to certain residues/ TM regions, the position vectors corresponding to PC pairs having high correlations were extracted, and RMSF was calculated. We performed this comparison between the motions of PC1 of mouse (53% covariance) versus PC2 of *B. cereus* (28% covariance, +0.33 correlation) ensembles. Similarly, we also compared the motions of PC1 of mouse (53% covariance) versus PC1 of *R. sphaeroides* (32% covariance, -0.28 correlation) ensembles. Though we did observe that certain pairs of PCs have higher correlations than the above two pairs, they have low contributions towards the overall dynamics. For instance, though correlation between PC1 (*R. sphaeroides* ensemble) and PC3 (mouse ensemble) is considerably high at -0.59, we did not compare this pair due to the fact that mouse PC3 has low contribution (4% covariance) towards overall dynamics of the ensemble (Fig. 5.5, A&B). For the same reason, we did not compare the dynamics of PC pairs involving PC3 and above, since such PC have lower contribution to the overall dynamics compared to PC1 and PC2 (Fig. 5.5, A&B).

On comparison of RMSF between correlated PCs of mouse (PC1) and *B. cereus* (PC2) ensembles, we observed similarity in fluctuations for the residues of TM4 and TM5 (Fig. 5.6, A). Similarly, between the correlated PCs of mouse (PC1) and *R. sphaeroides* (PC1) ensembles, similarity in fluctuations of TM2 and TM5 residues were observed (Fig. 5.6, B). Interestingly, some of the residues in TM2 and TM5 regions are observed to have physiological significance. For instance, some of the residues of TM5 which have similar fluctuations consist of the CARC motif residues, which is an alternate cholesterol-binding motif (Fantini et al. 2016; Papadopoulos et al. 2017). Apart from that, the region also has the PP-IX interacting tryptophan residue, that is proposed to be involved in catalytic degradation of PP-IX (Guo et al. 2015; Ginter et al. 2013). Though the residues of TM2 are not part of any functional motif, they consist of PK11195 binding and PP-IX binding residues (Fig. 5.6, B; Fig. 5.7, A)(Li, Liu, Garavito, et al. 2015; Guo et al. 2015; Li, Liu, Zheng, et al. 2015a; Jaremko, Jaremko, Giller, et al. 2014).

Since we observed that the first principal components of mouse and *R. sphaeroides* TSPO ensembles (accounting for 53% and 32% of total dynamics respectively) have correlation of below -0.25, we chose to compare the dynamics of respective PC1 by means of visualization of representative dynamics. When we visualized the dynamics of respective PC1, indeed we found similarities in fluctuations of above residues (Fig. 5.7). In both the TSPO, the region before residues Pro139 and Tyr140 (mouse TSPO) were dynamic in mouse, as well as in equivalent residues of *R. sphaeroides* TSPO (Fig. 5.7, A & B). Similarly, N-terminal of TM2 residues were dynamic both in mouse TSPO as well as *R. sphaeroides* TSPO, and the residues immediately below N-terminal, are rigid in both the TSPO (Fig. 5.7, A & B). This is also reflected in the RMSF plot of both the PC, where N-terminal region has greater fluctuations than the succeeding region (Fig. 5.6, B). In TM5, residues Pro139 and Tyr140 of mouse TSPO are more deformed, compared to equivalent residues of *R. sphaeroides* TSPO

(Fig. 5.7, A & B). In TM2, the N-terminal residues, though are dynamic, have different directions of motions. While these residues move towards TM5 in mouse (Fig. 5.7, A), they move away from TM5 in *R. sphaeroides* (Fig. 5.7, B). On the other hand, residues from Trp53 to Trp57 in mouse TSPO, and equivalent residues of *R. sphaeroides* TSPO are rigid (Fig. 5.7).

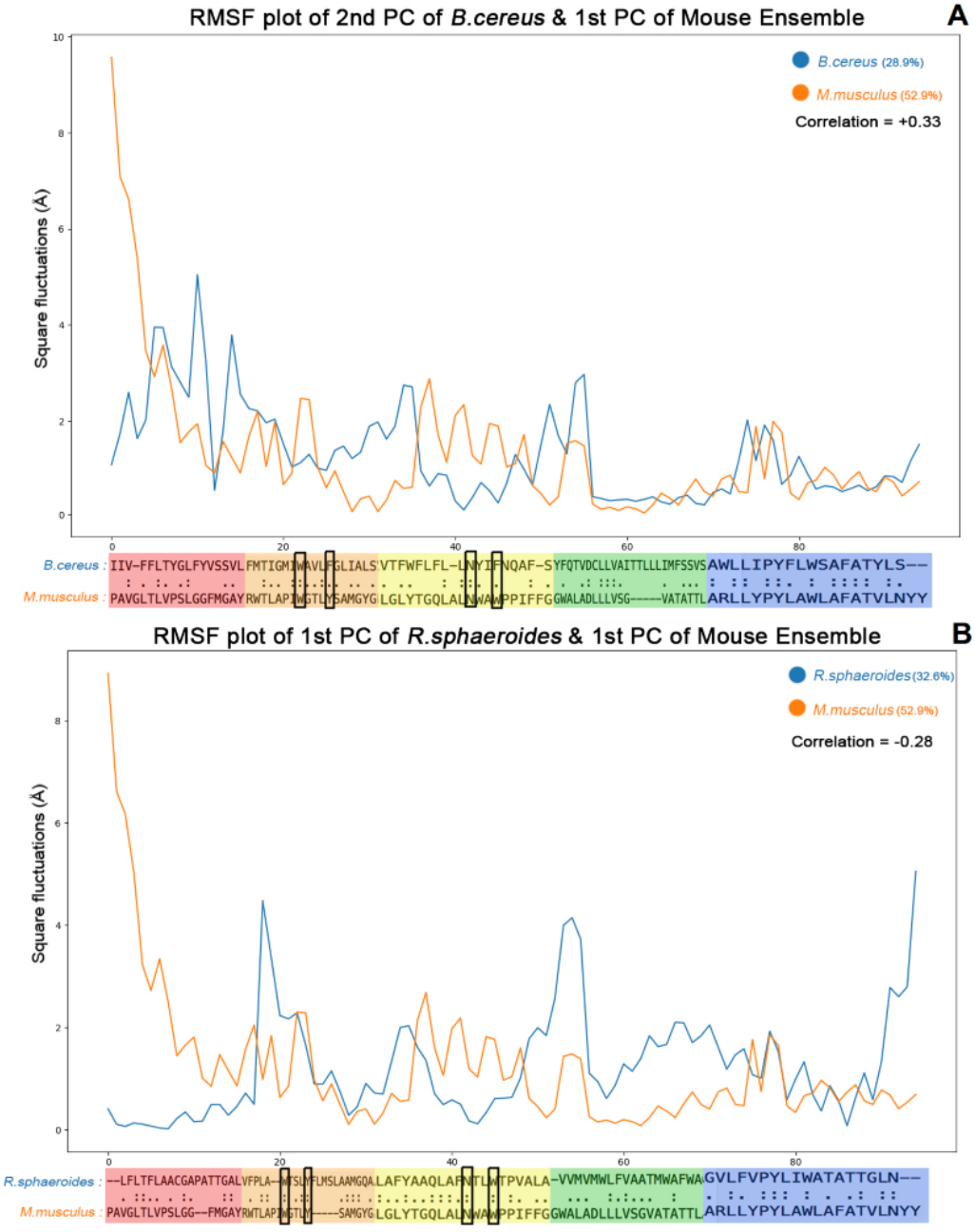


Figure 5.6. A: Plot showing RMSF from the trajectory representing dynamics of PC1 from the mouse ensemble and PC2 from *B. cereus* TSPO ensemble. **B:** Plot showing RMSF from the trajectory corresponding to PC1 of mouse and PC1 of *R. sphaeroides* TSPO ensembles. The sequence alignment of respective TSPO showing respective residue positions labelled on *y-axis* is shown as a pairwise alignment below the *y-axis*. PCA and RMSF calculations were performed using ProDy MD analysis package (Bakan et al. 2011). Pearson's correlation between the RMSF values are represented, and covariances of respective PC are represented in the legend. The residues marked with black squares indicated PP-IX binding positions in X-ray structure of *R. sphaeroides* TSPO.

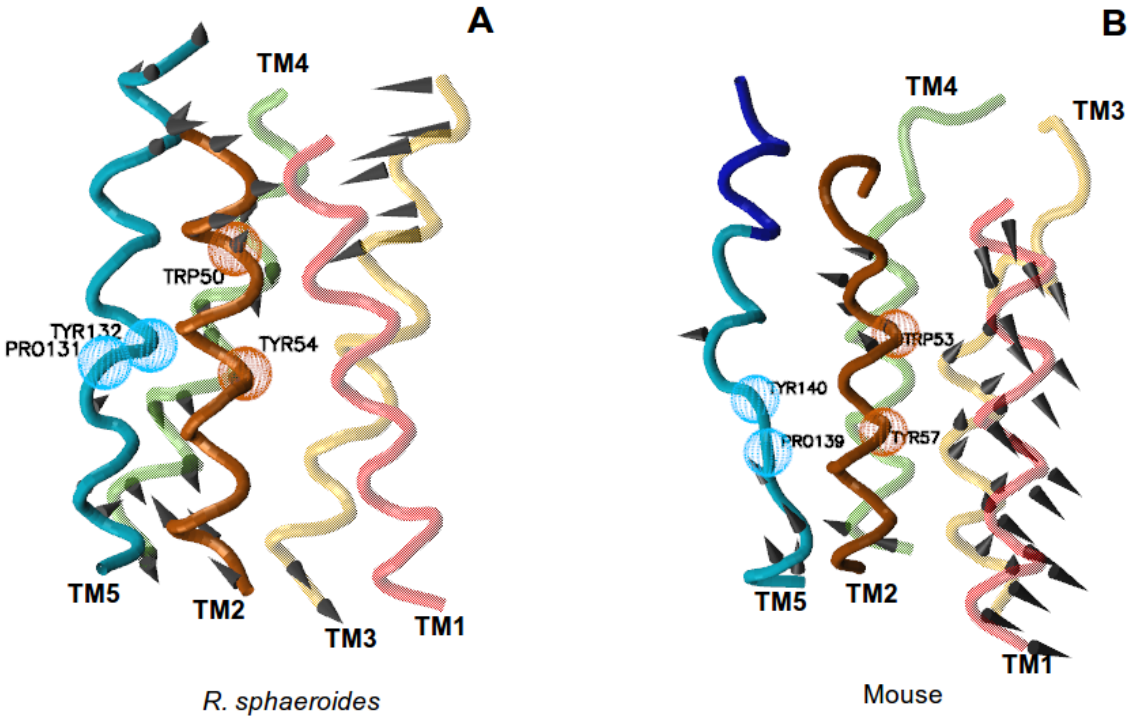


Figure 5.7. Porcupine plot showing the dynamics of first principal components from *R. sphaeroides* (32% covariance) **(A)** and mouse (52% covariance) **(B)** TSPO simulations. TM regions 1 to 5 are represented as red, orange, yellow, green and sky-blue ribbons respectively. CRAC motif of mouse TSPO is represented as dark blue ribbon. TM2 residues that have similar dynamics in both TSPO are represented as orange beads; TM5 residues having similar dynamics in both TSPO are represented as sky blue beads. PCA was performed using ProDy MD analysis package (Bakan et al. 2011), porcupine plot was visualised using VMD program (Humphrey et al. 1996).

5.2.3. PP-IX binding residues are conserved across bacteria as well as eukaryotes

From comparisons of mammalian and bacterial simulations, we found that though each TSPO explored unique sets of conformations (Fig. 5.4), there are similarities in dynamics of certain TM regions (Fig. 5.5, 5.6, 5.7). Interestingly, the regions that are observed to have similarity between mouse and *R. sphaeroides* TSPO (TM2 and TM5) also have residues that are previously shown to have physiological significance. TM2 constitutes residues that are experimentally shown to bind to PP-IX (Guo et al. 2015), whereas TM5 region, having similar dynamics in both the TSPO, constitute the CRAC motif, an alternate cholesterol-binding motif that has been characterized from computational studies (Fantini et al. 2016).

Though TSPO-cholesterol interaction does not occur in bacterial TSPO, TSPO-PP-IX interaction has been studied both in eukaryotes and bacteria. To understand if the porphyrin binding residues found in *R. sphaeroides* are conserved in eukaryotes as well, we projected TSPO residue conservation onto the structure of *R. sphaeroides* A139T mutant structure (Li et. al., 2015), which is solved in complex with PP-IX. The residue conservation scores in turn were calculated based on multiple sequence alignment

of TSPO sequences from bacteria, archaea as well as eukaryotes, and projected onto the TSPO structure. We also evaluated primary porphyrin interacting residues by calculating overlap of VdW radii between PP-IX and TSPO residues, and comparing residues with largest VdW radii overlap with their respective conservation scores. We used overlap of VdW radii overlap as a proxy for PP-IX TSPO interaction for the following reason: Due to the nature of PP-IX ligand (Fig. 1.9, Chapter-1), the interaction with TSPO is a mixture of hydrophobic interaction involving aromatic moieties and hydrophilic in nature. Since both the interactions have direct correlation to a certain extent with the interacting atom pairs, larger overlap in VdW radii suggest stronger interactions. To support this assumption, we also mapped TSPO-PP-IX contacts and classified their nature based on interatomic distances, using LIGPLOT program (Wallace et al. 1995), which interprets interactions based on interatomic distances. This comparison is summarised in Table 5.3.

Residue in *R. sphaeroides*	**Equivalent residue in mouse TSPO**	**TM region**	**Overlap of VdW radii (Å)**	**Distance (Å)**	**Conservation score**
Trp87	***Trp95***	TM3	0.56	3.07	-1.1
Trp50	***Trp53***	TM2	0.55	2.97	-1.1
Tyr54	Tyr57	TM2	0.46	3.18	-1.01
Arg43	Arg46	TM2	0.52	2.88	-0.24
Asn84	Asn90	TM3	0.39	2.91	-1.06
Trp135*	Trp143	TM5	--	--	-1.09
Ala139*	***Ala147***	TM5	--	--	-1.09

Table 5.3. Table showing residues of *R. sphaeroides* TSPO having strongest contacts (in terms of VdW radii overlap) with PP-IX ligand in the X-ray structure of A139T mutant (PDB: 4UC1, Li et al., 2015). VdW radii overlap was calculated using UCSF Chimera visualization program (Pettersen et al. 2004). Residues shown in bold and italic are the residues whose mouse equivalent residues are PK11195 ligand-interacting residues. Residues marked in '*' are TM5 residues that have been demonstrated to interact with PP-IX from experimental studies ((Guo et al. 2015; Li, Liu, Zheng, et al. 2015a; Ginter et al. 2013; Wendler et al. 2003; Vanhee et al. 2011)), but are not observed to have large overlap of VdW radii with PP-IX (having VdW radii overlap < 0) in the X-ray crystal structure. Conservation scores were calculated using Consurf web server (Ashkenazy et al. 2016; Glaser et al. 2003). Negative conservation scores signify evolutionary conservation, positive conservation scores signify variation in the given position.

We found that residues of TM2 (Arg43, Trp50 and Tyr54 in *R. sphaeroides* TSPO) and TM3 (Asn84 and Trp87) are the primary contributors to PP-IX binding in *R. sphaeroides* TSPO structure (Table 5.3). Furthermore, all of the above-mentioned contacts, with the exception of Arg43, are hydrophobic in nature, suggesting that overlap of VdW radii could be used to infer dominant contacts between TSPO and PP-IX. This is also in agreement with earlier experimental studies on PP-IX binding with *R. sphaeroides* TSPO, where mutagenesis of Trp50 residue had a direct impact on the bacterial photosynthesis, a process that involve PP-IX binding directly (Yeliseev and Kaplan 2000). However, it also needs to be noted that several other residues, such as Glu29, Leu34, Lys36,Trp30, Trp44, Gln80,

Arg96, which were found to be important for PP-IX binding from the same study (Yeliseev and Kaplan 2000) are not dominant contacts with PP-IX in the experimental structure ((Yeliseev and Kaplan 2000), Fig. 5.8, A & B). Thus, it could be possible that PP-IX may adopt other conformations in association with other residues, since A139T mutant structure is observed to have differences in conformation of TM2 and TM5 (Li, Liu, Zheng, et al. 2015a). It is also worth noting that Ala139 residue is shown to bind to PP-IX in experimental studies of bacterial TSPO homologue (Ginter et al. 2013). In humans, the mutation of this residue to threonine is associated with certain neurological disorders, and is also shown to affect cholesterol translocation (Costa, Pini, Gabelloni, et al. 2009; Colasanti et al. 2013). However, it is unclear if this is related to PP-IX binding, and if so, how is it related.

Interestingly, we observed that most of the PP-IX interacting residues from the X-ray crystal structure have low conservation scores, suggesting that they are well-conserved across bacteria as well as eukaryotes (Table 5.3, Fig. 5.8). This can also be observed from the pairwise alignment of *R. sphaeroides* and mouse TSPO sequences (projected with respect to RMSF plot of PC1, Fig. 5.6, B), where many of the PP-IX interacting residues are identical in both the sequences. This suggests that these residues may indeed be critical for TSPO interaction with PP-IX, regardless of the organism.

Consistent with the observations from conservation scoring of the residues, we observe that some PP-IX binding residues have similar dynamics in mouse and *R. sphaeroides* TSPO (Fig. 5.6, B). However, it may also be observed that there are differences in dynamics of TM3 region that host PP-IX interacting residues, despite conservation of the residues (Fig. 5.6, B; Fig. 5.7). Apart from TM2 and TM3, we found similarities in dynamics of TM5, which constitute residues that are near to PP-IX binding residue (Trp143 of mouse TSPO). Since we earlier observed that both TM2 and TM5 regions are conserved (table 3.5, chapter-3), we suggest that the common motions observed from our simulations may represent common fold-level dynamics of TSPO.

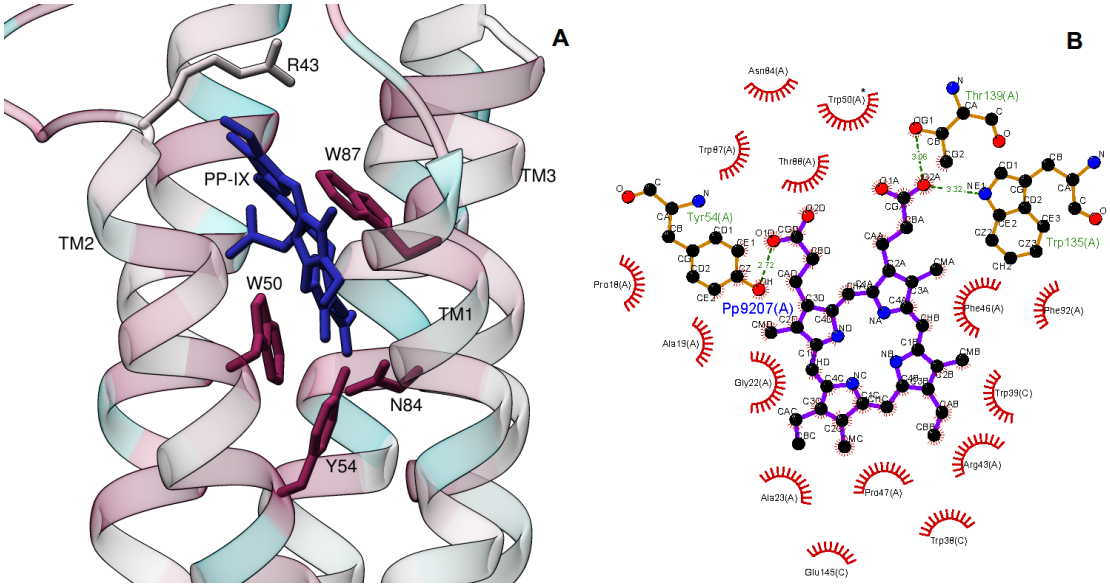


Figure 5.8. A: X-ray crystal structure of *R. sphaeroides* TSPO (PDB ID: 4UC1, (Li, Liu, Zheng, et al. 2015a)) in complex with PP-IX ligand showing protein ligand contacts and corresponding residue conservation. PP-IX is coloured blue and TSPO residues are represented as pink to cyan spectrum, signifying residue conservation scores. Residues coloured pink represent strong evolutionary conservation (low scores), residues coloured blue represents strong variations (high scores). Residue conservation scores were calculated using Consurf web server (Ashkenazy et al. 2016; Glaser et al. 2003). **B:** Illustration of PP-IX interacting residues of X-ray crystal structure of *R. sphaeroides* TSPO (PDB ID: 4UC1), showing different types of interactions as interpreted by interatomic distances. Hydrophobic contacts are represented as red spheres, hydrophilic contacts (hydrogen bonds) between TSPO and PP-IX atoms are represented by dashed lines. PP-IX and residues having hydrophilic contacts with PP-IX are represented by ball and sticks. Contacts were classified and illustrated using LIGPLOT program (Wallace et al. 1995)

5.3. Discussion

Though mouse and bacterial TSPO have reasonable sequence identities (36% between mouse TSPO and *R. sphaeroides* TSPO; 22% identity between mouse and *B. cereus* TSPO), RMSD between the respective experimental structures is very high (4.84Å between mouse TSPO NMR structure and *R. sphaeroides* wild-type TSPO, Fig. 5.1; 5.34Å between mouse TSPO NMR structure and *B. cereus* TSPO, Fig. 5.1), an observation also made from structural studies of bacterial TSPO (Guo et al. 2015; Li, Liu, Zheng, et al. 2015a). We initially hypothesized that despite the differences at the level of structure, mammalian and bacterial TSPO may have similar dynamics, since the identity between mammalian and bacterial sequences suggest similar dynamics that are part of a common fold. To test this hypothesis, we performed coarse-grained simulations of experimental TSPO structures from both mouse and bacteria, and compared their dynamics by means of PCA. Contrary to the initial hypothesis, we found that mammalian and bacterial TSPO explored different sets of conformations (Fig. 5.4), though there was also an overlap in the conformations explored between *R. sphaeroides* and mouse TSPO to a small extent. Then we attempted to understand if there are similarities in dynamics of dominant PC of mammalian versus bacterial TSPO. We observed similarities in dynamics of

dominant principal components of *B. cereus* TSPO and mouse TSPO in the regions of TM4 and TM5 (Fig. 5.6, A), as well as between the TSPO of *R. sphaeroides* and mouse in the regions of TM2 and TM5 (Fig. 5.6, B). The TM2 and TM5 regions having similarities between *R. sphaeroides* and mouse TSPO also have PP-IX binding residues and a putative cholesterol binding motif (Fig. 5.7). Furthermore, we observed that many of the residues in contact with PP-IX in *R. sphaeroides* TSPO structure are conserved across all the TSPO homologues (Table 5.3, Fig. 5.6, B, Fig. 5.8).

The similarity between the fluctuations of TM2 residues of *Rhodobacter* and mouse TSPO, and conservation of PP-IX binding residues gives rise to the question: Does the similarity in TM2 dynamics suggest towards a common physiological role of TM2 across prokaryote and eukaryote TSPO? Similar to eukaryotic TSPO, bacterial TSPO is also documented to be involved in various functions. One of the well-studied function is the negative regulation of bacterial photosynthesis in members of the genus *Rhodobacter* (Yeliseev and Kaplan 1995; Zeng and Kaplan 2001), which is closely related to its interaction with tetrapyrroles (Zeng and Kaplan 2001) and its role in response to stress (Busch and Montgomery 2017; Busch and Montgomery 2015; Leneveu-Jenvrin et al. 2015; Leneveu-Jenvrin et al. 2014). Similarly, in mammals, TSPO-porphyrin interactions occur in context of programmed cell death, as well as response to stress (Zeno et al. 2012). Thus, there exists a possibility that interaction TSPO-tetrapyrrole interaction may be conserved across various TSPO homologues, as proposed from earlier discussions (Fan et al. 2012).

However, there are two divergent hypotheses on the physiological role of TSPO with regard to porphyrin binding. One set of studies suggest that TSPO may be involved in transport and trafficking of porphyrin (Zeno et al. 2012). Another set of recent studies on bacterial TSPO revealed that other bacterial TSPO homologues have a catalytic activity, which has been documented experimentally in bacterial TSPO, including *R. sphaeroides, B. cereus and Chlorobium tepidum* to name a few (Ginter et al. 2013; Guo et al. 2015). While vertebrate TSPO homologues have been shown to be involved in the transport of porphyrin (Marginedas-Freixa et al. 2016; Rampon et al. 2009; Zeno et al. 2012). Porphyrin transport in bacterial TSPO homologues is not well-characterized. On the other hand, while some studies indicate that bacterial TSPO homologues have ability to bind to variety of porphyrins (Busch et al. 2017), some bacterial TSPO homologues have been recently found to catalyse degradation of porphyrins (Guo et al. 2015; Ginter et al. 2013). Furthermore, it was also suggested that TSPO homologue in cyanobacteria *Fremyella diplosiphon* is unlikely to function as a porphyrin transporter, as TSPO mutant *F. diplosiphon* cells could uptake hemins in iron-starved conditions (Busch et al. 2017). In other instances, TSPO has been proposed to play a regulatory role by TSPO-PP-IX interactions during stress conditions in plants (Balsemão-Pires et al. 2011; Vanhee et al. 2011). While there appears to be a divergence on the exact roles of TSPO in porphyrin metabolism in prokaryotes and eukaryotes, porphyrin binding has been well-documented experimentally in both

bacterial as well as eukaryotic TSPO (Guo et al. 2015; Li, Liu, Zheng, et al. 2015a; Ginter et al. 2013; Wendler et al. 2003; Vanhee et al. 2011).

Our results show that TSPO regions that host PP-IX binding residues have similar dynamics in mouse as well as *R. sphaeroides* TSPO (Fig. 5.6, 5.7). However, we also note that though *B. cereus* TSPO is also observed to bind to PP-IX experimentally (Guo et al. 2015), TM2 region of *B. cereus* TSPO have different dynamics compared to mouse and *R. sphaeroides* TSPO (Fig. 5.4). Furthermore, it may also be noted that TM3 has different dynamics in all the three TSPO (Fig. 5.6), though some of its PP-IX binding residues have strong evolutionary conservation (Fig. 5.8, A). These observed differences could be due to the fact that the TSPO in our simulations may be exploring different sets of energetic landscapes, that may have resulted in differences in conformations, or the differences in above regions could indeed be intrinsic in nature. This could be conclusively addressed in future by similar simulations, that have larger sampling. That being said, across all the three TSPO, the N-terminal region of TM5 have similar RMSF (Fig. 5.6), and also have similar dynamics to a certain extent, characterized by rigidity of Pro139 and Tyr140 residues (mouse TSPO, Fig. 5.7). Since many residues in these two regions, including PP-IX binding residues are conserved (table 3.5, chapter-3; Fig. 5.8) and have similar dynamics (Fig. 5.6, 5.7), we also suggest that dynamics of these two regions may represent the common fold level motions of TSPO. Our observations of similarity of TM2 and TM5 dynamics across bacterial and mammalian TSPO are also in agreement with earlier studies, that have proposed an important role of TM2 and TM5 residues in PP-IX binding and catalysis (Yeliseev and Kaplan 2000; Ginter et al. 2013; Guo et al. 2015). We provide a structural and dynamic basis of importance of these regions in PP-IX binding. That being said, studying dynamics of TSPO in context of PP-IX binding is an interesting problem, and is an interesting direction to proceed in future. The dynamics of TM2 and TM5 can be studied across TSPO of different species in different modes, being ligand-unbound form versus PP-IX bound form by means of all-atom molecular dynamics, and binding free energy calculations.

5.4. Conclusions and future directions

To conclude, while there are differences in the conformations explored by mouse and bacterial TSPO in our simulations, we also observed commonality in dynamics of both TSPO to a certain extent (Fig. 5.4.), which were captured in the motions of the top principal components (Fig. 5.6, 5.7). Interestingly, some of the regions which have similar dynamics between both the TSPO consist of PP-IX interacting residues, which are conserved between bacterial as well as mammalian TSPO (Fig. 5.8). From these observations, we suggest that the commonality in dynamics may indicate the common fold-level dynamics of TSPO. Though bacterial TSPO have been largely associated with a catalytic role with respect to PP-IX (Zeng and Kaplan 2001; Busch and Montgomery 2015; Ginter et al. 2013; Guo et al. 2015), and eukaryotic TSPO have been proposed to have transporter role (Marginedas-Freixa et al.

2016; Rampon et al. 2009; Zeno et al. 2012) as well as catalytic role (Guo et al. 2015), conservation of PP-IX binding sites, as well as similarity in their dynamics point towards a common role of bacterial as well as eukaryotic TSPO in PP-IX binding. However, further experimental as well as theoretical studies are essential to improve the understanding of the possible divergence in TSPO physiology with respect to tetrapyrroles. While earlier experimental studies on bacterial TSPO have suggested that PP-IX binding residues of TM5 are important for the catalytic role of TSPO (Trp135 and Ala139), due to the fact that mutation of these residues have resulted in loss of catalytic ability of TSPO to degrade PP-IX (Ginter et al. 2013; Guo et al. 2015), it would be interesting to understand the exact role these residues play in PP-IX catalysis. Furthermore, it would be worth pondering if such functions may exist in mammalian TSPO as well, due to the fact that both the residues are conserved in mammalian TSPO as well. In particular, when Ala147 residue, whose *R. sphaeroides* equivalent is Ala139, is mutated to threonine, the presence of this mutation is associated with several neurological disorders (Colasanti et al. 2013; Costa, Pini, Martini, et al. 2009). Future studies in this direction would greatly enhance our understanding of the role of TSPO in PP-IX metabolism and the role of the latter in several TSPO-associated pathophysiological conditions.

Chapter-6: Structural and Evolutionary comparison of TSPO paralogs

Table of contents

6.1. Introduction 170
6.2. Results 171
6.2.1. TSPO2 diverged from TSPO1 during evolution of early vertebrates 171
6.2.2. Many PP-IX binding and cholesterol interacting residues are not conserved in TSPO2 175
6.2.3. Large differences were observed between TSPO1 and TSPO2 at the level of structures 178
6.2.4. Both differences and similarities were observed in the TSPO1 vs. TSPO2 dynamics 181
6.3. Discussion 184
6.4. Conclusions and future directions 189

6.1. Introduction

Recently, while searching for remote homologs of TSPO, another isoform of TSPO was reported. This isoform was named TSPO2, and this isoform was found to be expressed specifically in differentiating blood cells in higher-order mammals (Fan et al. 2009). While TSPO (also referred to as TSPO1) is predominantly found in mitochondrial outer membrane, TSPO2 was primarily expressed in the nuclear and endoplasmic reticulum membranes (Fan et al. 2009; Simos et al. 1996). Phylogenetic analysis of TSPO1 and TSPO2 sequences from higher vertebrates revealed that TSPO2 may have arisen from TSPO1 by a gene duplication event, which may have occurred before the divergence of birds and mammals, about 300 million years ago (Fan et al. 2009). While it appears that TSPO2 may have a tissue-specific role based on its distribution, early experimental studies led to some interesting hypotheses. TSPO2 was found to have underwent decrease in high-affinity ligand binding properties, particularly binding to the ligand PK11195 at nanomolar affinities, that was characteristic property of TSPO1. Though drug-binding properties at high affinity were reduced, TSPO2 inherited cholesterol binding properties, and it was found to play a role in cholesterol redistribution during the process of erythropoiesis (Fan et al. 2009).

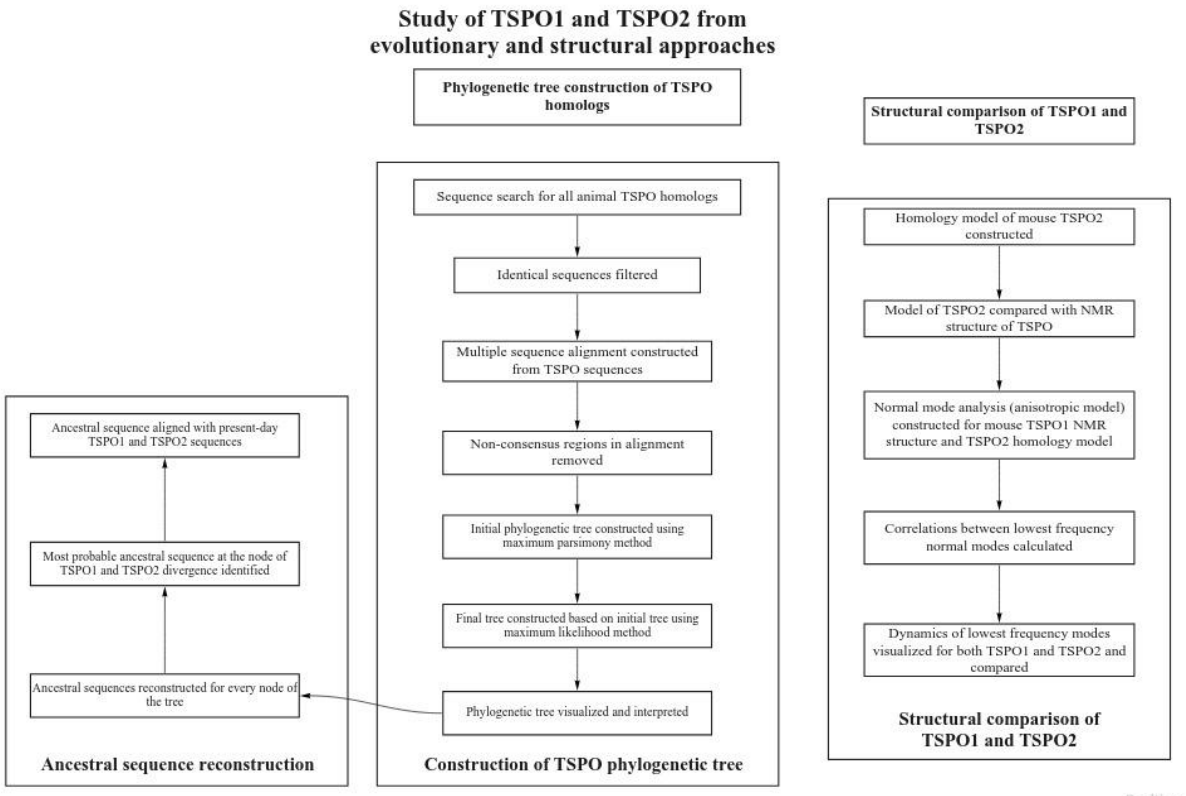
During the final steps of erythropoiesis, expulsion of nucleus is an essential step, which is a prerequisite for formation of mature enucleated erythrocytes. However, presence of free intracellular cholesterol inhibits this process (Fan et al. 2009; Holm et al. 2002). Thus, to initiate nucleus expulsion, TSPO2 serves to sequester cholesterol, which paves way for further maturation of erythrocytes (Fan et al. 2009). On the other hand, the role of TSPO1 has been largely associated with cholesterol translocation into mitochondria, regulation of mitochondrial permeability, apoptosis and inflammation to name a few. Though TSPO1 is also found to be expressed in cells undergoing erythropoiesis, its expression was found to decrease as erythropoiesis progressed, whereas at the same time, expression of TSPO2 increased (Marginedas-Freixa et al. 2016). Though TSPO1 is observed in differentiating erythropoietic cells, its exact role in such tissues remain poorly understood. These studies highlight a niche-role of TSPO2 in regulating cholesterol concentration during erythropoiesis. However, TSPO2 has also been found to be active in mature erythrocytes, along with VDAC, a well-studied TSPO interaction partner (Bouyer et al. 2011; Marginedas-Freixa et al. 2016). In the mature erythrocytes, TSPO2 was isolated as electrophoretic bands of different molecular weights. Western blot experiments indicated that apart from observing TSPO2-antibody complex at 18 kDa range, which corresponds to its molecular weight, it was also observed in bands of 36 kDa, 72 kDa and 800 kDa range in mature erythrocytes (Marginedas-Freixa et al. 2016). While the former two bands may correspond to an oligomeric form of TSPO2, similar to its cousin TSPO1, 800 kDa band corresponds to a larger complex, which consist of VDAC and ANT (Rone et al. 2012; Marginedas-Freixa et al. 2016).

Interestingly, apart from normal erythrocytes, TSPO2 was found to be active in erythrocytes infected with the malarial parasite *Plasmodium falciparum* along with its interacting partner VDAC and ANT (Bouyer et al. 2011; Marginedas-Freixa et al. 2016). Treatment of infected cells with TSPO-binding drugs such as PK11195, Ro5-4864 and diazepam resulted inhibition of *P. falciparum* growth in infected erythrocytes (Bouyer et al. 2011). A more recent study found that TSPO2 is present in healthy erythrocytes as well (Marginedas-Freixa et al. 2016). However, in erythrocytes, TSPO2 was attributed to transport of porphyrins, in contrast to its role of cholesterol redistribution in differentiating blood cells (Marginedas-Freixa et al. 2016; Fan et al. 2009). It was observed that transport of Zn-PP-IX led to accumulation of ROS. In erythrocytes infected with *P. falciparum*, treatment of TSPO ligands PK11195, Ro5-4684 and SSR-180575 significantly accelerated Zn-PP-IX uptake, and accumulation of ROS. This led to inhibition of both intraerythrocytic as well as intererythrocytic parasite growth (Marginedas-Freixa et al. 2016). Thus, akin to TSPO1, TSPO2 appears to have conserved both cholesterol binding, as well as porphyrin uptake property. However, while initial studies indicated that TSPO2 lost its drug-binding properties ((Marginedas-Freixa et al. 2016; Fan et al. 2009), studies on TSPO2 in mature erythrocytes indicate that TSPO2 binds drug ligands (Marginedas-Freixa et al. 2016). Furthermore, the difference between the two paralogs at the level of sequence are not unequivocal. An attempt was made to understand the differences between TSPO1 and TSPO2 at the level of sequences. It was found that there are significant differences in the sequence of second extracellular loop (Fan et al. 2012). In TSPO1, this region corresponds to mitochondrial localization signal, which appears to have lost in TSPO2. This was proposed to be the rationale behind the organelle-specific distribution of TSPO1 and TSPO2. Though it has been indicated that there are differences between TSPO1 and TSPO2 in the second extracellular loop, the differences in binding-cavity residues, which determines binding affinity of various endogenous as well as exogenous ligands are unclear. Furthermore, while we know some aspects of physiological role of TSPO2, there is no information on TSPO2 from a structural perspective.

6.2. Results

6.2.1. TSPO2 diverged from TSPO1 during evolution of early vertebrates

To address these knowledge gaps, we attempted to trace the TSPO1 and TSPO2 evolution based on sequence alignment of TSPO1 and TSPO2 sequences from all the multicellular animals (metazoans). For the same, we constructed a phylogenetic tree of 196 TSPO sequences from various metazoan species. The exact protocols, methods and analysis used in this chapter are described by flow charts below. The methods employed are discussed in detail in chapter-2.

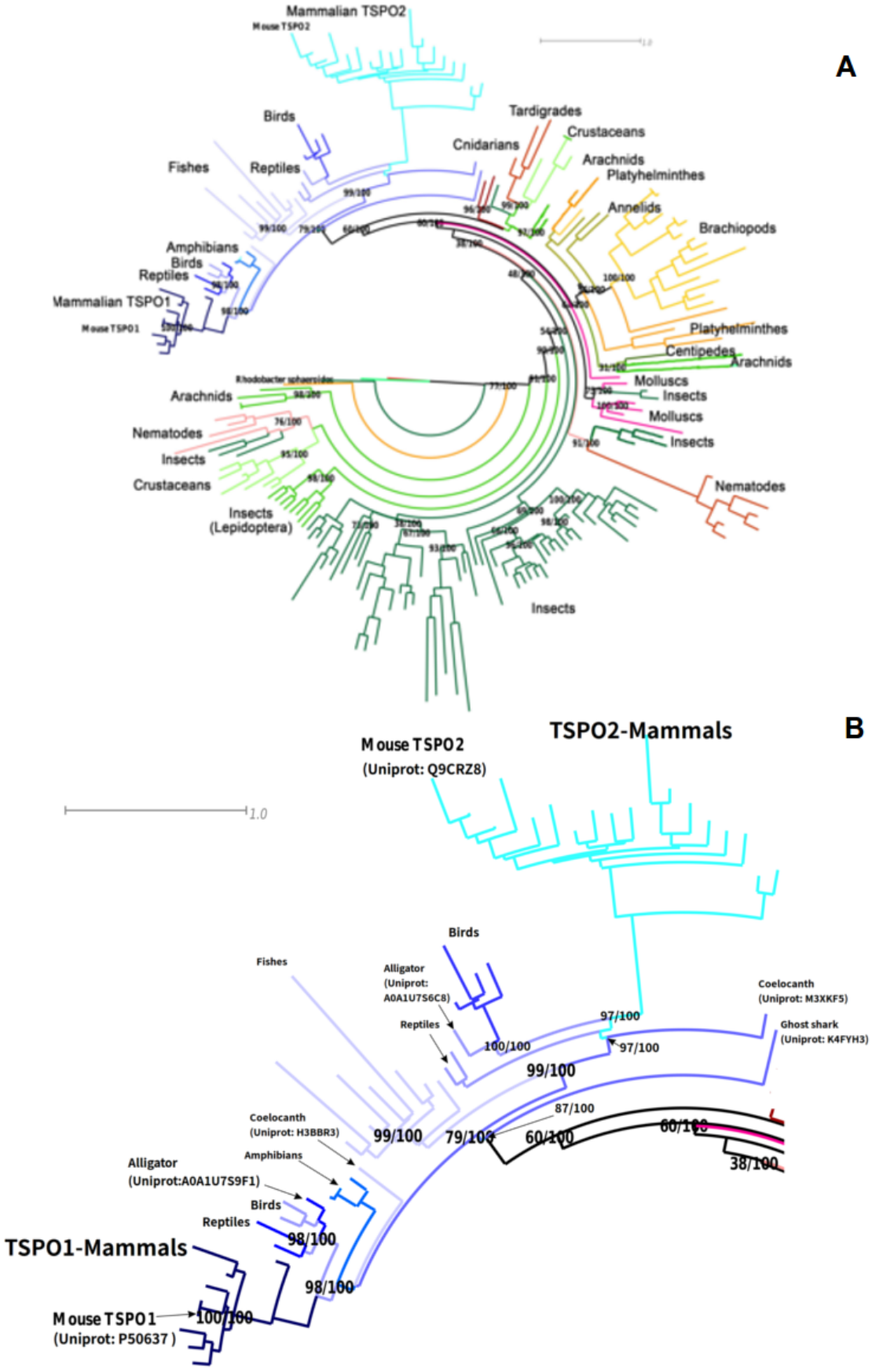


Box 6.1. Flow chart showing methods used for comparison of TSPO1 and TSPO2 at sequence and structural level

We made some interesting observations from the consensus phylogenetic tree: The sequences of TSPO1 and TSPO2 clustered into distinct branches, in agreement with earlier studies on evolution of TSPO2 (Fig. 6.1, A, (Fan et al. 2009)). However, on closer observation of the cluster consisting of TSPO2 sequences, we observed that both the branches have sequences from lower vertebrates, including birds, reptiles and fishes that are distributed in both theTSPO1 and TSPO2 clades (Fig. 6.1, B). Previous study on sequence-based evolutionary relationships of TSPO1 and TSPO2 sequences proposed that TSPO1 and TSPO2 may have diverged before evolution of birds and mammals, about 300 million years ago (Fan et al. 2009). Since we observed reptilian and fish sequences clustered with bird and mammalian TSPO2 sequences, we attempted to support our hypothesis that observed sequences may be paralogous. To do the same, we performed pairwise sequence alignment of *Alligator sinensis* and *Latimeria chalumnae* (Indian ocean coelacanth) TSPO sequences from TSPO1 and TSPO2 cluster (Fig. 6.2, B). We observed that indeed, there are low identities between the respective TSPO sequences (53.8% and 63.1% between *Alligator* and *Latimeria* TSPO sequences), given the fact that both the pairs of sequences are from the same species (Fig. 6.2, B). However, we observed higher identity between these sequences compared to the identity between mouse TSPO1 and TSPO2 sequences (40%). We thus infer that the TSPO sequences from each of the cluster may indeed be different isoforms, or paralogous. Interestingly, the *Latimeria chalumnae*, or Coelacanth is

regarded as a ‘living fossil’, and has evolved into its present form about 400 million years ago (Johanson et al. 2006), and the outgroup of the TSPO1 and TSPO2 clade is a TSPO sequence of *Callorhinchus milii* (Australian ghost shark). Though we observed presence of coelacanth TSPO sequences in TSPO1 and TSPO2 clades, we could not observe similar distribution of other fish TSPO sequences. The bony-fish (class: Actinopterygii) TSPO sequences are clustered in TSPO2 clade and amphibian TSPO sequences are in turn clustered in the TSPO1 clade (Fig. 6.1, B). Since divergence of bony fishes and cartilaginous fishes (sharks, class: Chondrichthyes) occurred about 400 million years ago, the divergence of TSPO1 and TSPO2 may have occurred at least 400 million years ago as well, earlier than the previous estimates (Fan et al. 2009).

We also observed that there are significant differences in the rates of divergence of TSPO1 and TSPO2 sequences. TSPO2 sequences were found to have higher divergence compared to TSPO1 sequences (Fig. 6.1, B), which is in agreement with the earlier conclusions from computational studies of TSPO2 (Fan et al. 2009; Fan et al. 2012). This gave rise to the next question: Since there is high divergence of TSPO2 sequences, how does this divergence account for cholesterol and porphyrin-binding capabilities, since previous experimental studies have pointed towards differentiated roles of TSPO1 and TSPO2, and possible loss of function in TSPO2? To gain understanding into this, we calculated putative ancestral sequence from the node before the divergence of TSPO1 and TSPO2, and compared it with present-day TSPO1 and TSPO2 sequences.



(*Legend on next page*)

Fig. 6.1. A: Phylogenetic tree showing evolution of animal TSPO sequences. Phylogenetic tree construction was performed by combination of maximum parsimony and maximum likelihood methods with 100 bootstrap replicates. TSPO sequences from *R. sphaeroides* was used as the taxonomic outgroup. B: The same phylogenetic tree zoomed into the vertebrate clade showing divergence of TSPO1 and TSPO2 sequences. Each branch colour represents distinct taxonomic group, which are labelled at the respective edges. The numbers on the node represent bootstrap values.

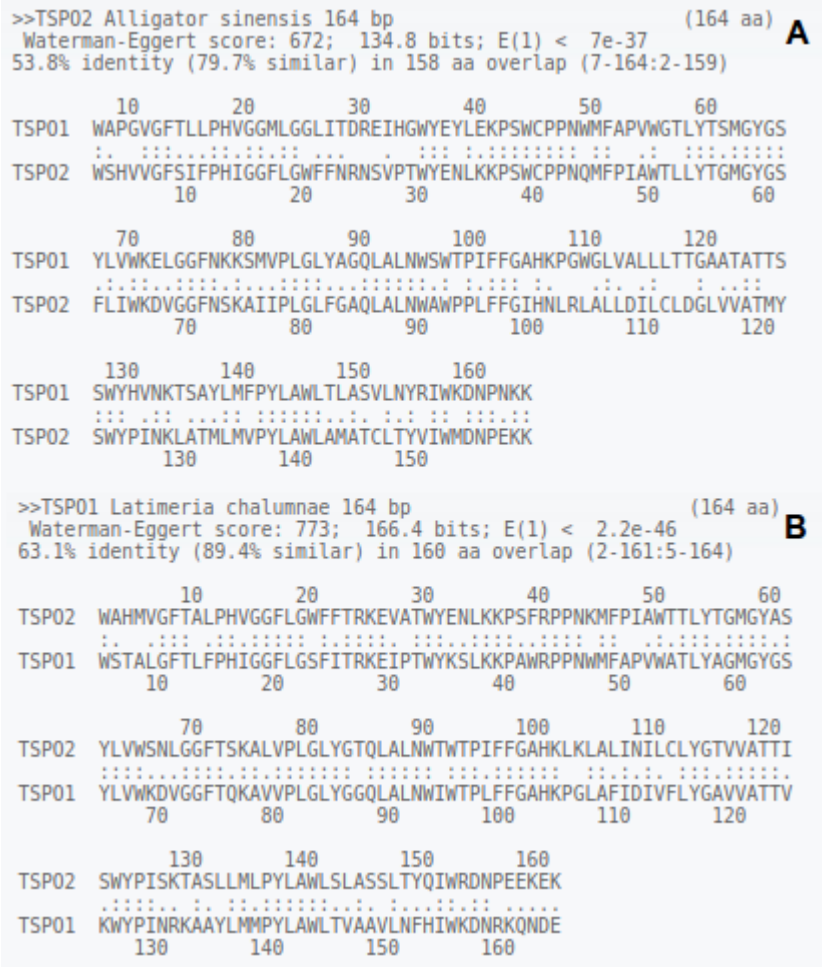


Fig. 6.2. Pairwise sequence alignment between the TSPO sequences of Chinese alligator (**A**) (*Alligator sinensis*), with 53.8% of sequence identity and TSPO sequences of Coelacanth (**B**) (*Latimeria chalumnae*), with 63.1% sequence identity. Both the sequences were retrieved from the Uniprot database, and aligned using LALIGN server.

6.2.2. Many PP-IX binding and cholesterol binding residues are not conserved in TSPO2

From the construction of TSPO phylogenetic tree, we learned that TSPO1 and TSPO2 diverged at the time of early vertebrate evolution. We also observed that TSPO2 has a higher rate of divergence than TSPO1. Since TSPO2 has been reported to have undergone loss of function (Fan et al. 2009), we wanted to understand if this loss of function directly associated with its rate of divergence? More precisely, is the divergence of TSPO2 reflected in the divergence of PP-IX and cholesterol binding residues? As an alternate hypothesis, is there a general conservation of binding residues in TSPO1, due to which it has retained its functions? To get answers to the above questions, we reconstructed

common ancestral sequence of TSPO1 and TSPO2, and compared them with present day TSPO1 and TSPO2 sequences. The divergence of both the TSPO has been well-studied by means of horizontal studies including comparison of biochemical properties both the TSPO paralogs and comparison of present-day sequences. However, studying the same from a vertical approach, involving reconstruction of ancestral sequences provides another perspective on its functional divergence, which often occurs from a series of mutations and gene duplications. In this case, identification of divergent residues not only provides an understanding into the functional divergence, but it also provides an insight into the ways in which specific groups of residues confer function to the protein. Furthermore, similar studies have already provided insights into how functional divergence occurs and specificity evolves in members of proteins belonging to different families, such as opsins and steroid receptors to name a few (Harms and Thornton 2010; Thornton et al. 2003; Yokoyama and Radlwimmer 1999).

We reconstructed the most probable ancestral sequence to the TSPO1 and TSPO2 clades (methodology described in box 6.1, and in chapter-2), and compared it with present day human and mouse TSPO1 and TSPO2 sequences by aligning all the sequences. In particular, we compared interacting residues of mouse PP-IX (inferred from alignment of mouse TSPO sequence with *R. sphaeroides* TSPO, from Fig. 5.6, B of chapter-5) and cholesterol. Among PP-IX binding residues, we found mixed extent of conservation across TSPO1 and 2 sequences. Few PP-IX binding residues are conserved across ancestral TSPO, TSPO1 and TSPO2 sequences (Tyr57 and Trp143 of mouse TSPO, represented in the alignment positions 58 and 148 respectively), and we observed that many of the PP-IX binding residues are not conserved in TSPO2 sequences (Fig. 6.3). Since PP-IX and PK11195 are found to displace binding of each other (Wendler et al. 2003; Guo et al. 2015), we found that many of PK11195 binding residues (Leu49, Trp53, Trp95 and Leu149 of mouse TSPO, represented in the sequence alignment on positions 49, 53, 96 and 151 respectively, Fig. 6.3) are also not conserved in TSPO2, in agreement with previous experiments, where PK11195 interaction ability was observed to have lost in TSPO2 (Fan et al. 2009).

Similar to PP-IX binding residues, we observed mixed degree of conservation in CRAC motif as well. Experimental studies that characterized CRAC motif in TSPO identified residues Tyr152, Tyr153 and Arg156 of mouse TSPO to be critical for binding with cholesterol (Li, Yao, et al. 2001; Jamin et al. 2005). In agreement with experimental studies, we observed conservation across the above-mentioned residues of CRAC motif (residue positions 153, 156-159, Fig. 6.3, (Fan et al. 2009)), implying a common cholesterol binding property of both TSPO1 and TSPO2. But we observed that some CRAC motif residues are not conserved in TSPO2 sequences. The position corresponding to Tyr153 residue is variable in both TSPO1 and TSPO2 sequences. Apart from that, two asparagine residues (on positions 152 & 159 of the alignment, Fig. 6.3) are conserved in TSPO1 sequences, but are variable

across TSPO2 sequences. As previously discussed, another alternative cholesterol binding motif exists in TM5, called 'CARC' motif (Fantini et al. 2016; Papadopoulos et al. 2018). We found that while CRAC motif in general is conserved in TSPO1 and TSPO2, the vital residues of CARC motif (Arg135 and Tyr138 of mouse TSPO, positions 140 and 143 in the alignment respectively) are not conserved in TSPO2 sequences, though it is conserved in TSPO1 homologs.

Apart from this, two other sets of residues in TM5 are considered to be important in the context of cholesterol binding. Preceding to the CRAC motif, residue Ala147 in mouse TSPO has been observed to influence cholesterol binding in TSPO, where its substitution to threonine is associated with reduced pregnenolone production (Costa, Pini, Gabelloni, et al. 2009), despite the fact that it is not a part of CRAC motif. While this residue is substituted to threonine in some human individuals, the substitution is observed to be prevalent in TSPO2 sequences (position 152 in the alignment, Fig. 6.3). Apart from Ala147, another motif adjacent to Ala147 was discovered, that played an important role in enhancement of cholesterol binding. In mouse TSPO1 sequences, this motif consist of residues 144 to 146 (144-LAF-146), and was called "enhancement motif". In bacterial cells where LAF motif containing bacterial TSPO sequences were transfected, cholesterol binding was observed to increase 1000-fold (Li, Liu, Valls, et al. 2015). This observation was inferred to be significant because bacterial TSPO sequences do not have this motif, and are unable to translocate cholesterol. We also observed that while this motif is conserved in TSPO1 sequences, it is variable in ancestral as well as TSPO2 sequences (positions 149 to 151 in the alignment, Fig. 6.3).

The past study on comparison between TSPO1 and TSPO2 sequences pinpointed differences in the TM3 and TM4 regions (Fan et al. 2012). However, we observed that the differences span all the TM regions, including TM3 and TM4 (Fig. 6.3). However, in agreement with previous sequence-based comparisons, the residues of second extracellular loop (residues positions 79 to 84, Fig. 6.3) have significant variation (Fan et al. 2012). While residues of this region are largely charged in TSPO1 sequences, the residues of the same region are uncharged in TSPO2.

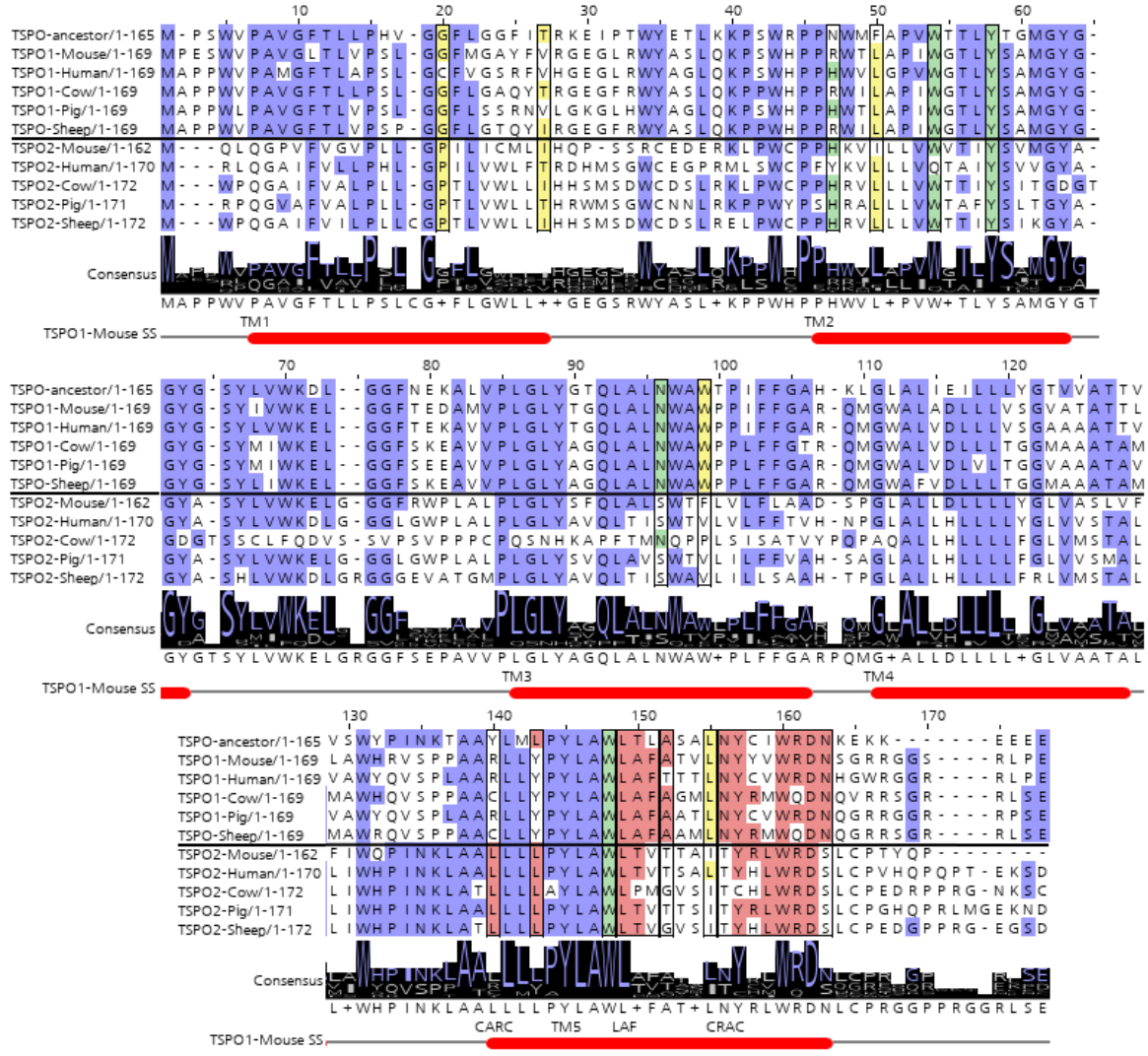


Fig. 6.3. Multiple sequence alignment of TSPO1 and 2 sequences with the most probable ancestral sequence from which divergence of TSPO1 and 2 occurred. The residues of CRAC motif are highlighted as red, residue positions equivalent to residues interacting with PP-IX, as determined from the X-ray crystallographic structure of the *R. sphaeroides* TSPO (PDB: 4UC1) are highlighted green, and the residues of mouse TSPO1 interacting with PK11195 in NMR structure (PDB ID: 2MGY) are highlighted yellow. Other conserved residue positions are highlighted blue. Multiple sequence alignment was performed using MAFFT program (G-INSi method) (Katoh et al. 2002), and visualized using Jalview (Waterhouse et al. 2009) .

6.2.3. Large differences were observed between TSPO1 and TSPO2 at the level of structures

Another major knowledge gap regarding TSPO2 is its lack of atomic structures. We attempted to address this gap by constructing a multi-template homology model of mouse TSPO2 using the NMR structure of mouse TSPO1 (PDB ID: 2MGY), wild-type crystal structure of *R. sphaeroides* TSPO (PDB ID: 4UC3) and *B. cereus* TSPO (PDB ID: 4RYI) as templates. We previously observed that there are significant differences in conformations of mouse and bacterial experimental structures (Fig. 5.1, chapter-5). It was earlier discussed that the differences in the conformations of experimental structures may have arisen from use of detergents such as DPC to solubilize the protein in the case of

mouse TSPO, which may have resulted in drastic changes in its conformation (Li, Liu, Garavito, et al. 2015). In addition, MD simulation studies on mitochondrial membrane proteins in DPC micelles have indicated that DPC micelles significantly alter the fold of the protein (Kurauskas et al. 2018). To account for such discrepancies, we constructed a multi-template model, using both NMR as well as X-ray crystallographic structures as templates. We attempted to understand the differences between TSPO1 and TSPO2 at two levels: from the perspective of structure and at the level of dynamics. To understand the differences between TSPO1 and TSPO2 at the level of structure, we compared the homology model of TSPO2 with the NMR structure of TSPO1. On superimposition on the C-alpha atoms of the residues of TM domain (TM domain described in table 6.1), we observed high RMSD between the residues of TM regions (Table 6.1), though the fold of TSPO2 was similar to that of TSPO1 (Fig. 5.4). This however, may be the direct consequence of using multiple templates for modelling the structure of TSPO2.

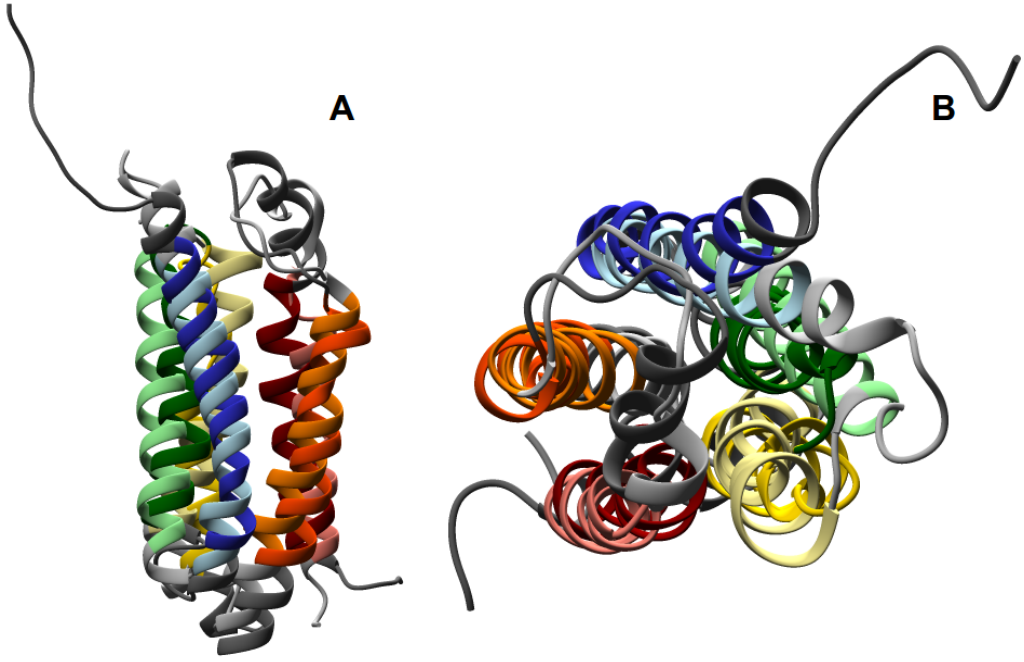


Figure 6.4. Superposition of mouse TSPO1 and TSPO2 structures. **A:** Horizontal view **B:** Vertical view. TM regions 1 to 5 are represented by red, orange, yellow, geen and blue ribbons respectively. TM regions of TSPO1 are represented in dark shades, those of TSPO2 are represented in light shades. Homology model of TSPO2 was constructed using TSPO1 structures of mouse, *R. sphaeroides* and *B. cereus* as templates. Structures were superimposed and visualized using UCSF Chimera program (Pettersen et al. 2004).

TM Region	***Residues consisting of TM region for TSPO1***	***Residues consisting of TM region for TSPO2***	***RMSD (Å)***
1	7-26	6-25	4.85
2	45-63	43-61	4.8
3	80-102	78-99	6.75

4	103-124	103-124	5.86
5	134-154	128-149	3.64

Table 6.1. Table showing RMSD between TM regions of mouse TSPO1 NMR structure and mouse TSPO2 homology model. RMSD was calculated from superimposition of TSPO1 and TSPO2 structures. The definitions of TM boundaries of TSPO1 are based on the TM boundaries in OPM database (Lomize et al. 2012; Lomize et al. 2011), the definitions of TM boundaries for TSPO2 were calculated using OREMPRO web server (Postic et al. 2016).

While TM5 region has relatively low RMSD, all the other TM regions have high RMSD (Table 6.1). The high RMSD for the TM regions could be attributed to the fact that there are significant structural differences between the templates themselves (Fig. 5.2), which are manifested in the homology model of TSPO2. On the other hand, the relatively low RMSD for TM5 region may be attributed to the fact that it has many identical residues common to TSPO2, as well as bacterial templates. Thus, TM5 conformations may be conserved in both TSPO1 and TSPO2 (Fig. 5.4A).

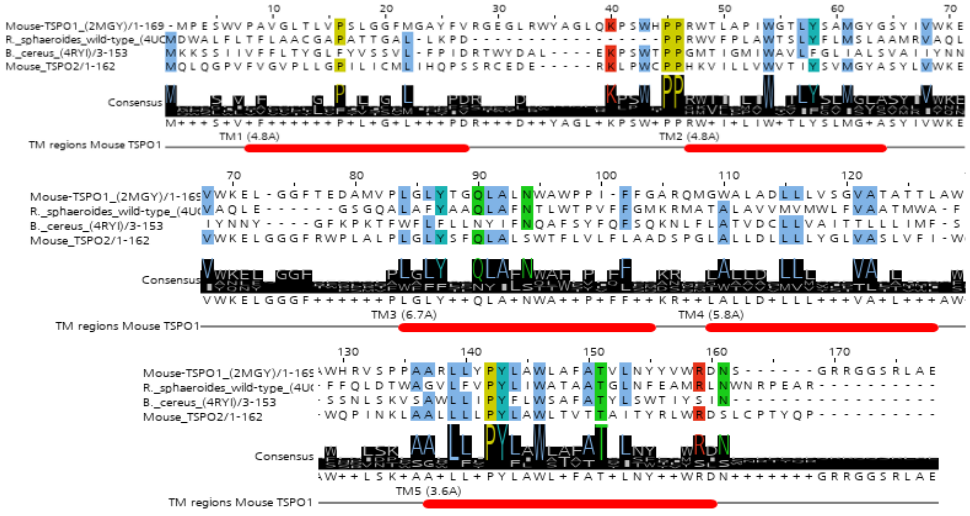


Figure 6.5. Alignment of TSPO1 sequences of mouse*, R. sphaeroides* and *B. cereus*, whose structures were used as template to construct homology model of TSPO2, along with the sequence of mouse TSPO2. Alignment blocks having identity of >75% are represented by Clustal colours, TM region boundaries of mouse TSPO NMR structure (adopted from OPM database) and RMSD with TSPO2 homology model are represented as red cylinders below the alignment.

However, it may also be noted that other TM regions also have residues that are conserved across all the TSPO structures, as well as in TSPO2 (Fig. 6.5). In particular, the TM3 has conserved residues in all the TSPO sequences, yet this region has high RMSD of 6.7Å (Fig. 6.7, Table 6.1). Similarly, TM2 also has conserved residues, yet it has significantly high RMSD.

6.2.4. Both differences and similarities were observed in the TSPO1 vs. TSPO2 dynamics

While we observed that significant differences at the level of structure exists between TSPO1 and TSPO2, we pondered if the differences in structure translates into difference in dynamics. Or as an alternate hypothesis, if the conserved regions have similar dynamics, despite assuming different conformations. To answer these questions, we constructed an anisotropic network model of NMR structure of TSPO1 and TSPO2 structures.

Initially, we attempted to compare different modes based on their correlation. However, we observed that while diagonal modes (1st, 2nd 3rd modes of TSPO1 vs TSPO2) do not have high correlations in their motions, many non-diagonal modes have high correlations (Fig. 6.6). Since comparison of selected modes at the cost of others may result in introduction of bias into the comparison, we attempted to compare the dynamics of normal modes that have high collectivity in their motions. This is due to the fact that in MD simulations, often the top principal components that describe majority of protein dynamics also have high collectivity in their motions of >0.5 (Brüschweiler and Case 1994; David and Jacobs 2014; Brüschweiler 1995), and dynamics at the level of tertiary structure involve such collective motions (Brüschweiler and Case 1994). We calculated collectivity of 20 lowest frequency modes of TSPO1 and TSPO2 (Table 6.2)

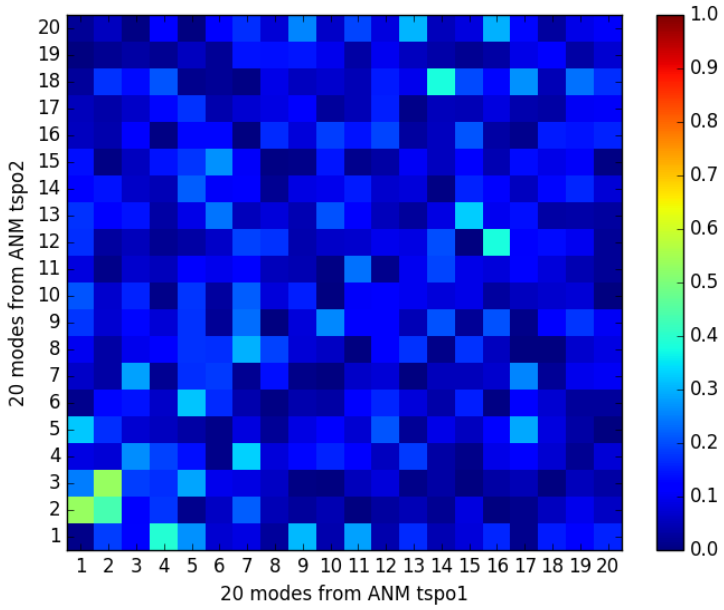


Fig. 6.6. Heatmap showing correlation between the dynamics of non-trivial lowest-frequency normal modes of TSPO1 and TSPO2. Correlations are represented by rainbows (from blue to red), in increasing order of correlations. Normal modes for both TSPO1 and TSPO2 were calculated based on anisotropic network model, with cutoff being 13Å and force constant being 1.0. Normal mode calculations and Pearson's correlations between the normal modes were calculated using Prody molecular dynamics analysis package (Bakan et al. 2011).

Normal mode	Collectivity of TSPO1	Collectivity of TSPO2
1	0.10	0.28
2	0.12	0.10
3	0.19	0.38
4	0.22	0.20
5	0.36	0.12
6	0.11	0.14
7	**0.50**	0.17
8	0.31	0.28
9	0.31	0.41
10	0.27	0.12
11	0.48	0.12
12	0.26	0.30
13	0.33	0.32
14	0.35	0.16
15	**0.51**	0.24
16	0.34	0.38
17	0.29	0.22
18	0.20	**0.55**
19	0.34	0.24
20	0.39	**0.54**

Table 6.2. Table describing degree of collectivity of residue dynamics of TSPO1 and TSPO2 normal modes. Modes used for comparison are represented in bold. Normal modes and their collectivities were calculated using ProDy MD analysis program (Bakan et al. 2011).

We selected two modes from TSPO1 and TSPO2 each that have highest collectivities (>0.5), and compared their dynamics and fluctuations. On comparison of TSPO1 vs. TSPO2 dynamics, the conspicuous difference we observed was that there are collective motions between the residues at upper halves of TM2 and TM5 regions in TSPO1. On the other hand, such motions were absent in both the normal modes of TSPO2, though there were individual fluctuations at the residues of CRAC motif of TM5 (Fig. 6.7). Since the scale of the motions in TSPO2 are localized to a single residue, we hypothesize that such motions may represent deformations in the region. Similarly, we also observed presence of dynamics in the N-terminal region of TM1 in TSPO2, which were not observed in TSPO1. Interestingly, the dynamics appear to occur in the direction of membrane axis, suggesting possible helix stretch/contraction motions. On the other hand, we also observed some similarities in the dynamics. We found that the residues surrounding Trp143 (of mouse TSPO1) are rigid. While the residues that lie in further periphery to this residue (including CRAC & CARC motifs, located at N- and C-terminal regions of TM5) are dynamic, the region surrounding Trp143 residue, in the middle of TM5 are rigid in all the visualized normal modes of TSPO1 and TSPO2 (Fig. 6.7). This suggest that this region may act as a hinge, while the residues surrounding it are flexible.

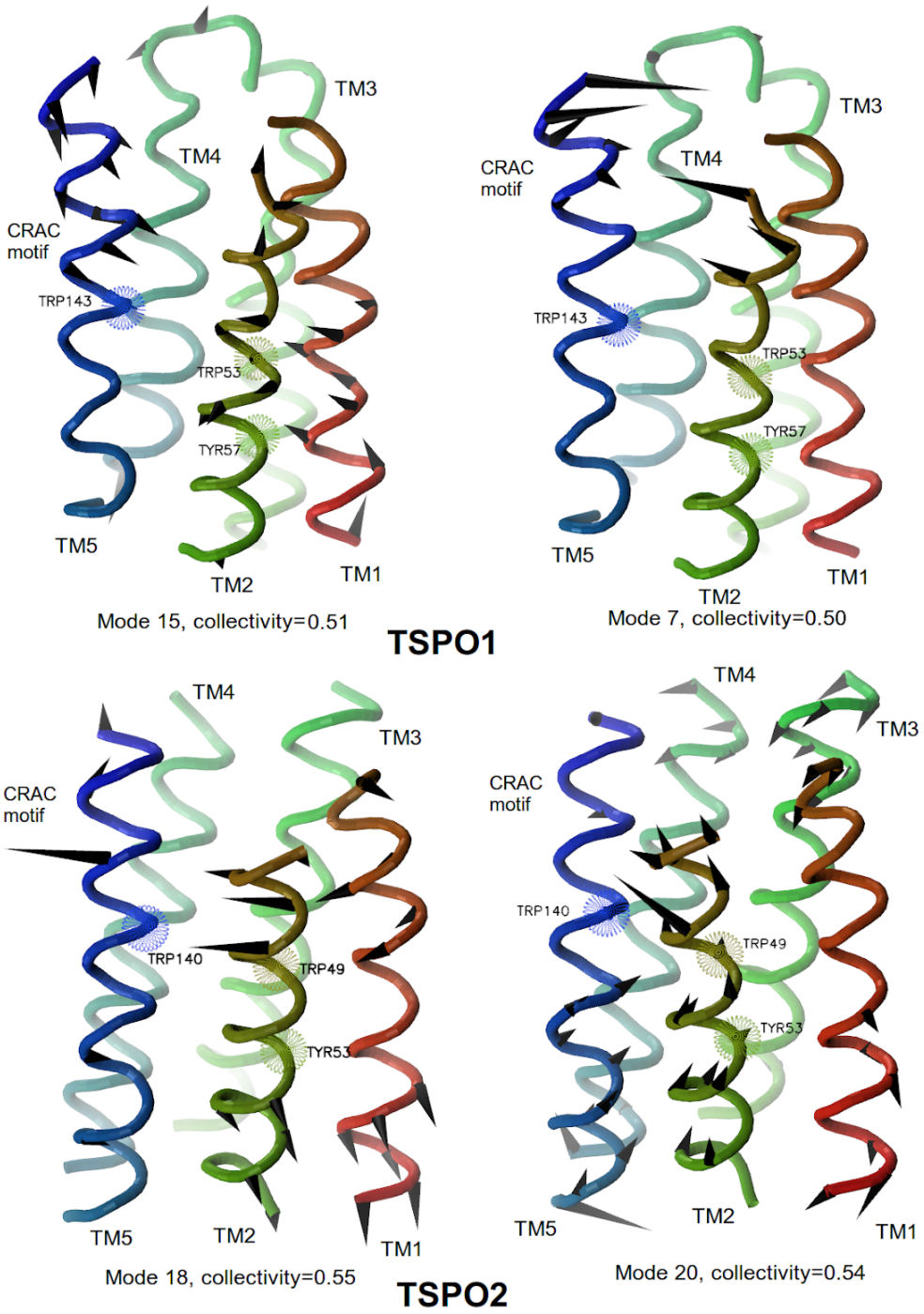


Fig. 6.7. Porcupine plot of mouse TSPO1 and mouse TSPO2 coarse-grained models showing dynamics of normal modes having high collectivity (>0.5). TSPO1 and TSPO2 residues are represented as tubes, coloured rainbow gradient from N- to C- terminal (red to blue). Residues of TSPO1 experimentally shown to bind to PP-IX (Li, Liu, Zheng, et al. 2015a; Guo et al. 2015; Yeliseev and Kaplan 2000), and equivalent residues of TSPO2 are shown as dotted beads. Normal modes were calculated using ProDy MD analysis program (Bakan et al. 2011), and porcupine plots were generated using normal mode wizard tool of ProDy (Bakan et al. 2011).

6.3. Discussion

TSPO2 was first characterized as a result of search for remote TSPO homologs, and it was proposed that TSPO1 and TSPO2 diverged before the divergence of birds and mammals, about 300 million years ago (Fan et al. 2009). We attempted to further understand the divergence in context of broader diversity of TSPO sequences in animals. Furthermore, recent explosion in the number of sequenced animal genomes warranted this study, to account for sequences that were recently annotated. We found that the divergence of TSPO1 and TSPO2 occurred much earlier than previously thought, since we found presence of reptilian as well as fish sequences in both the TSPO1 and TSPO2 clades (Fig. 6.1, B). We compared TSPO sequences of the same species that were found in TSPO1 and TSPO2 clades, and found that the sequences have low identity. Given that the sequences were from the same species, the low sequence identity suggests that the sequences are indeed paralogous (Fig. 6.2). Interestingly, the fish species (*Latimeria chalumnae,* or Coelacanth), whose TSPO sequences are distributed in both the clades is referred to as a 'living fossil'. It is suggested that this species evolved into its present form over 400 million years ago (Johanson et al. 2006). Furthermore, a TSPO sequence of a cartilaginous fish (*Callorhinchus milii* or Australian ghost shark) is clustered as an outgroup to both the TSPO1 and TSPO2 clades. Then, we hypothesized that the divergence of TSPO1 and TSPO2 may have occurred at the time of divergence of bony fishes from cartilaginous fishes, about 400 to 420 million years ago, suggesting that the divergence of TSPO1 and TSPO2 occurred during the early evolution of vertebrates. However, we also observed that there are TSPO sequences from fishes that are distributed uniquely in TSPO2 clade (Fig. 6.2,B). Similarly, TSPO sequences from amphibians are uniquely distributed in the TSPO1 clade. At this point, it is unclear whether such distribution is a result of loss of function of the other TSPO gene or if it could be a result of poor annotation of sequences in the genome. If it is the former, then it would be interesting to understand what are functional implications of such probable loss of functions.

However, we consider presence of reptilian and fish TSPO1 and TSPO2 sequences to be significant. Invertebrates and primitive vertebrates, such as jawless fishes have no differentiated blood cells, but all the vertebrates from cartilaginous fishes have well-differentiated red blood cells and lymphocytes. To explain this sudden divergence, there is a well-accepted hypothesis that a large-scale duplication of many gene families led to differentiation of blood cells (Anderson et al. 2001; Holland et al. 1994). This hypothesis also suggest that such duplications were the cause of evolution of many features that were first seen in vertebrates, including blood cell differentiation (Holland et al. 1994; Pébusque et al. 1998). Since we observed divergence of TSPO at around similar time, we propose that the divergence of TSPO may be a part of series of gene duplications that occured during early vertebrate evolution. More importantly, it suggests that TSPO1 and TSPO2 may play a more significant role in erythropoiesis, in agreement with experimental studies that point in the same direction (Fan et al.

2009; Fan et al. 2012). However, this hypothesis needs further support, due to the fact that the TSPO phylogenetic tree presented in this study may not be a complete one, which could be attributed to paucity of fully sequenced genomes, particularly from fishes and early tetrapods. This paucity could also have resulted in discrepancy in the distribution of fish and amphibian TSPO sequences observed in this tree.

Since we also observed significant difference in the rates of divergence of mammalian TSPO1 and TSPO2 sequences (Fig. 6.2, B), and preceding experimental studies have also discussed about potential loss of function of TSPO2 (Fan et al. 2009; Fan et al. 2012)), we wondered if the divergence is carried over to the cholesterol and PP-IX binding residues. We also attempted to answer the question of what is the extent of divergence of TSPO1 and TSPO2 sequences with respect to putative common ancestor of both the sequences. To answer these questions, we compared present-day TSPO1 and TPSO2 sequences with the most-probable TSPO sequence that is ancestral to both present-day TSPO1 and TSPO2 sequences.

We observed that many of the PK11195 binding residues are not conserved in TSPO2 sequences, while they are conserved in TSPO1 sequences. In particular, the residues Trp53 and Trp95, that are conserved in TSPO1 sequences, have diverged in mouse TSPO2. We have observed earlier that both these residues are also dominant interacting partners with PK11195, as observed from AA simulations of TSPO in presence of PK11195 ligand (Fig. 3.14, table 3.5, chapter-3). Since both the residues are tryptophan, they may have contributed to PK11195 binding by means of non-bonded planar interactions with isoquinoline scaffold of PK11195. In mouse TSPO2, though residue equivalent to Trp53 is conserved, residue equivalent to Trp95 is substituted to phenylalanine residue (Fig. 6.4). In human TSPO2 on the other hand, both the residues are substituted to non-aromatic residues (Fig. 6.4). In TSPO2, these substitutions may have an adverse impact on the PK11195 binding ability of TSPO2, as observed from experimental studies (Fan et al. 2009). However, evidence on the PK11195 binding ability of TSPO2 is rather mixed. While earliest studies showed loss of PK11195 binding ability of mouse TSPO2, more recent study on TSPO2 in red blood cells indicated that human TSPO2 in erythrocytes can bind to PK11195, albeit in micromolar affinities as opposed to nanomolar affinities in human TSPO1 (Marginedas-Freixa et al. 2016; Bouyer et al. 2011). While the ability of TSPO2 to interact with PK11195 remains to be conclusively established, we provide a structural basis by which PK11195 binding ability may be affected in TSPO2. On the other hand, TSPO2 was found to bind to Zn-PP-IX in red blood cells (Marginedas-Freixa et al. 2016).

For the PP-IX binding residues, while we found conservation of few residues (Tyr53 and Trp140 in mouse TSPO2), we observed that many of other PP-IX binding residues have diverged in TSPO2 sequence (Fig. 6.4). Conservation of PP-IX interacting residues despite substitution of other residues suggest that the two conserved PP-IX interacting TSPO2 residues, Tyr53 and Trp143 in mouse TSPO2

may play a critical role in the PP-IX interaction. This hypothesis is also in agreement with experimental studies on PP-IX binding in bacteria *R. sphaeroides*, where it was found that mutagenesis of equivalent bacterial tryptophan residue (Trp138 in *R. sphaeroides*) to phenylalanine resulted in loss of PP-IX binding ability (Yeliseev and Kaplan 2000; Guo et al. 2015).

While we observed a general conservation of cholesterol-binding CRAC motif in both TSPO1 and TSPO2 sequences, we found that some residues are substituted in TSPO2 sequences. In particular, we found that two asparagine residues that flank both the ends of the CRAC motif are not conserved in TSPO2 (Fig. 6.4). It is difficult to conclude at this point if the substitutions may have had an impact on the cholesterol-translocating ability of TSPO across the membrane. In addition, we also observed significant variations in other residues that are associated with cholesterol interaction. For instance, we found that the CARC motif, which is suggested to be an alternate cholesterol binding motif in TSPO1 (Fantini et al. 2016) is not conserved in TSPO2 sequences. Both the arginine and tyrosine residues that define such cholesterol-binding motifs (Fantini et al. 2016; Di Scala et al. 2017) have diverged in TSPO2 sequences (Fig. 6.4). It was earlier found that both CRAC and CARC motifs coexist in many membrane proteins other than TSPO (Di Scala et al. 2017; Fantini et al. 2016) (Di Scala et al. 2017). Their absence in TSPO2 is surprising, given the fact that such an arrangement is observed in many other transmembrane proteins, such as G-protein coupled receptors, Glutamate receptor 5, Type-3 somatostatin receptor that are known to interact with cholesterol (Fantini et al. 2016; Di Scala et al. 2017). Since this arrangement is primarily observed in neurotransmitter and hormone receptors, it remains to be seen if the loss of CARC motif is another evidence for loss of hormone-mediated cholesterol transport, or if this loss is more related to localization, where possible differences in membrane environment (mitochondrial outer membrane vs. endoplasmic reticulum) may have influenced divergence of this motif.

Ala147 residue, which is considered to be important for cholesterol translocation is observed to be substituted to threonine in TSPO2 sequences. Earlier experimental studies found that substitution of Ala147 to threonine in TSPO1 resulted in reduced pregnenolone synthesis, for which transport of cholesterol into mitochondria is necessary (Costa, Pini, Gabelloni, et al. 2009). This substitution has also been found to be associated with prevalence of breast cancer (Hardwick et al. 1999), as well as with occurrence of certain neurological conditions (Costa, Pini, Martini, et al. 2009; Colasanti et al. 2013). Furthermore, studies into the structure of mutant form revealed changes in conformation of TM2 and TM5 regions, which was thought to have affected cholesterol binding capability of TSPO1 (Li, Liu, Zheng, et al. 2015a). In humans, the residue is conserved in certain individuals, and is substituted to threonine in others. In TSPO2 however, it remains to be seen what impact does this substitution have in terms of cholesterol binding affinity, and on the general dynamics of the protein.

Apart from CARC motif and Ala147 residue, another group of residues that are found to have undergone significant change is called the 'LAF' motif (144-LAF-146 in mouse TSPO1; Fig. 6.4). This motif was found to increase TSPO affinity to cholesterol binding and uptake in bacterial TSPO. Bacterial TSPO sequences originally lack this motif, and do not have the capability of cholesterol transport, despite having residues identical to CRAC motif at the C-terminal. Introduction of this motif in bacterial sequences resulted in increased uptake of cholesterol at least by 1000-fold (Li, Liu, Valls, et al. 2015). While this motif is conserved in all the TSPO1 proteins, it is substituted to different residues in TSPO2 proteins. It was also observed that while this motif is conserved in mammalian TSPO sequences, it is not conserved in other vertebrate TSPO sequences (Li, Liu, Valls, et al. 2015). Our results are in good agreement with this, since we found that this motif is also absent in the ancestral TSPO sequence (Fig. 6.4).

The above mentioned changes in residues associated with cholesterol interaction in TSPO2 gives rise to the question: how do these changes affect cholesterol-binding ability of TSPO2? First studies on TSPO2 has a role in sequestering free intracellular cholesterol. Since both CARC and LAF motifs are absent, or substituted in TSPO2, we hypothesize that TSPO2 may have lost the cholesterol-transport ability though it retained cholesterol-binding ability, due to the fact that particularly the latter has been shown to greatly enhance the cholesterol binding and transport ability of TSPO (Li, Liu, Valls, et al. 2015).

Apart from comparing TSPO1 and TSPO2 at the level of sequence, we also made an attempt on comparison of the same from structural perspective. Since no experimental structure of mouse TSPO2 is available till date, we constructed a homology model of TSPO2, using NMR structure of mouse TSPO1 as well as X-ray structures of bacterial TSPO as templates. On comparison of TSPO1 NMR structure with the TSPO2 model, we found large RMSD between all the TM regions (Table 6.1, Fig. 6.5). While it is difficult to pinpoint the exact cause behind such large differences, there are two possible explanations: One, the fact that there are significant differences between experimental NMR and X-ray structures, despite having reasonable sequence identities (36% between mouse TSPO and *R. sphaeroides* TSPO) led to inference by some groups that the high RMSD may be attributed to the fact that structure of mouse TSPO was solved in presence of dodecylphosphocholine (DPC), a particularly strong detergent (Li, Liu, Garavito, et al. 2015). It was hypothesized that use of such detergent may have resulted in drastic changes in conformation. Two, despite identity of about 25-40% between mammalian and bacterial sequences, the differences in sequence may have resulted in differences in dynamics. In support of this, we have observed that mammalian and bacterial TSPO explored different sets of conformations (Fig. 5.4, chapter-5).

Finally, we attempted to understand if the differences in conformations are also reflected in the dynamics of TSPO1 and TSPO2. To answer this, we performed normal mode analysis on both the

TSPO paralogs, and compared the modes of each TSPO with high collectivity in their motions. Though we observed differences in TSPO1 vs. TSPO2 dynamics that were localized to different regions (Fig. 6.7), we observed that there were also similarities in dynamics in certain regions. In particular, the region of TM5 between CRAC and CARC motifs was found to be rigid in all the observed normal modes of TSPO1 and TSPO2, when either CRAC, or CARC motif was found to be dynamic (Fig. 6.7). Since both the flanking regions are dynamic in contrast to rigidity of this region, this region could be exhibiting hinge-like dynamics. This region is particularly interesting due to the fact that it has the residue Trp143 (in mouse TSPO1), that is one of the primary binding residues of PP-IX (Yeliseev and Kaplan 2000; Guo et al. 2015), as well as the catalytic degradation site if PP-IX in bacterial TSPO (Ginter et al. 2013; Guo et al. 2015). In support of our observations from this comparison, we observed similar dynamics of TSPO1 from MD simulations (chapter-3), where TSPO1 is complexed with PK11195 ligand. We observed that the residues surrounding Trp143 (residues 140-145) are relatively rigid (RMSD < 0.2nm), and flanking residues are more dynamic (RMSD > 0.2nm; Fig. 3.5, B, chapter-3). Furthermore, such hinge-like sites have been observed to be critical for the biological function in various proteins (Isin et al. 2012; Tieleman et al. 2001). The fact that such this region also had similar dynamics in bacterial TSPO (Chapter-5, Fig. 5.6) suggest that this region may exhibit dynamics that represent characteristic TSPO-fold dynamics. In support of this hypothesis, we have observed that this region in conserved across mammalian as well as bacterial TSPO (Fig. 6.5), the experimental mammalian and bacterial TSPO structures have low RMSD around this region (Fig. 5.2) and have similar dynamics (Fig. 5.6), and there is similarity of normal-mode dynamics between TSPO1 and TSPO2 for this region (Fig. 6.7). However, if such dynamics also exists in TSPO2 needs to be seen in future, by means of molecular dynamics simulations of TSPO2 models.

As an alternate hypothesis, the rigidity of this region could facilitate TSPO-tetrapyrrole interactions. TSPO-tetrapyrrole interactions is a common property observed experimentally in all the forms of TSPO, including mammalian TSPO1 (Guo et al. 2015; Verma et al. 1987), TSPO2 (Marginedas-Freixa et al. 2016), bacterial TSPO (Guo et al. 2015; Ginter et al. 2013; Yeliseev and Kaplan 1999) as well as plant TSPO (Vanhee et al. 2011). Since we observed rigidity of this region in normal modes of TSPO1 and TSPO2 (Fig. 6.7), and we also observed similar dynamics of this regions between mammalian and bacterial TSPO (Fig. 5.6, Fig. 5.7), we hypothesize that the dynamics of this region may be critical for TSPO-tetrapyrrole interactions. However, though there is support from experimental studies on importance of Trp143 residue for tetrapyrrole interactions (Ginter et al. 2013; Guo et al. 2015), we also propose that this hypothesis needs to be further evaluated both by means of experimental as well as molecular modelling studies.

6.4. Conclusion and future directions

TSPO2 was first characterized as a paralog of TSPO1 that diverged about 300 million years ago. From the phylogenetic tree of TSPO sequences, we found presence of TSPO2 sequences from early vertebrates, such as reptiles and fishes, implying that the divergence of TSPO may have occurred much earlier, at about 400 million years ago. Since this divergence also coincide with divergence of many other gene families involved in blood cell development, we hypothesise that TSPO may have an important role in erythropoiesis. However, this hypothesis is limited by paucity of fully sequenced genomes, and may be revisited in future. To understand the extent of conservation in various ligand-binding residues in TSPO1 and TSPO2, we compared the present day TSPO1 and TSPO2 sequences with putative ancestral TSPO sequences. We observed divergence in PK11195, PP-IX as well as cholesterol binding residues, though we also observed that some PP-IX and cholesterol binding residues are strongly conserved across both the TSPO paralogs, as well as in the ancestral sequence, in agreement with the previous hypothesis of specialization of physiological function of TSPO2. However, the impact on the divergence in certain residues of binding site on the binding capability of TSPO-ligands would be an interesting direction to proceed further, by means of computational studies, including molecular dynamics simulations in complex with various ligands, as well as binding free-energy studies of said ligands with TSPO.

From the structural perspective, we also found significant differences in conformations adopted between TSPO1 and TSPO2, which were also observed to a certain extent from the normal mode dynamics of both the proteins. While we observed differences in dynamics, we also observed some overlaps to a certain extent, particularly in TM5 regions, which displayed hinge-like dynamics in both TSPO1 and TSPO2. Because we also observed previously that this region has similar dynamics between mammalian and bacterial TSPO, we hypothesise that this region may exhibit characteristic dynamics of the TSPO fold. It would be interesting to see if such dynamics could be observed in TSPO2 by means of MD simulations as well. We propose that the above observations can be the basis of further experimental as well as theoretical studies on the structure-dynamics-function relationships of TSPO1 and TSPO2.

Concluding remarks

TSPO is a transmembrane protein found largely in outer mitochondrial membranes. While earliest studies characterized binding of various ligands to the protein, its functional role remained unknown. Later functional studies suggested a possible role of TSPO in steroidogenesis and regulation of mitochondrial metabolism. Since then, large number of studies have helped in better understanding TSPO and its role and place in the cellular and organismal homeostasis. Some of them include mutagenesis experiments that showed its involvement in steroidogenesis, characterization of its interaction with other mitochondrial proteins such as VDAC, ANT etc., characterization of a cholesterol-binding motif to name a few. Apart from steroidogenesis, TSPO was also found to play a role in apoptosis, erythropoiesis and inflammation, apart from various other physiological processes in other organisms. However, much remains to be understood about TSPO from the physiological perspective. Furthermore, some recent studies also contradict the existing consensus that TSPO is involved in cholesterol translocation, adding to the mystery and ambiguity surrounding the protein.

TSPO remained an enigma from the structural perspective as well. Though initial cryo-EM studies suggested a five TM domain arrangement of the protein with oligomeric stoichiometry, atomic structures were solved only recently. Initial structures in complex with ligands PK11195 and PP-IX have contributed immensely to our understanding of TSPO. However, several ambiguities also exist. Comparison of bacterial and mammalian TSPO structures, both of which were solved in presence of high-affinity binding ligand PK11195 revealed significantly high differences in conformations, though it had an identical fold, and sequence identity of 24% (between mouse TSPO & *B. cereus* TSPO). Apart from this, an attempt to study the structure of TSPO in absence of a ligand revealed that the protein is highly dynamic accompanied by loss of secondary structure. In addition to that, the understanding of oligomeric arrangement is unclear. Earlier experimental studies suggested that TSPO exist in multi-oligomer orientations. However, recent atomic structures showed a dimer and trimeric arrangements, where different interfaces were observed. Thus, we initiated this study with the following questions: What is the impact of ligand binding on the TSPO dynamics? If a pair of separated TSPO monomers dimerizes, is there a preferred orientation for a TSPO dimer? Are there differences in dynamics of bacterial and mammalian TSPO? To answer these questions, we used a combination of molecular modelling and simulation techniques to obtain insights into TSPO dynamics from various perspectives.

To answer the first question, we performed all atom molecular dynamics simulations of TSPO embedded in lipid bilayers. We performed two distinct simulations, one where the ligand PK11195 is bound to TSPO with the same conformation as observed in NMR structure, and the other system

being TSPO in absence of any ligand. We compared the dynamics of *holo*- and *apo*- state TSPO by means of PCA, and we observed significant differences in the dynamics between *holo*- and *apo*- forms of TSPO with respect to the dynamics of TM domain. However, there are overlaps in conformations to a certain extent if extracellular loops are also considered. Earlier studies on TSPO structure suggested that PK11195 binding stabilised TSPO secondary as well as tertiary structure. Our observations were in contradiction to this hypothesis, though PK11195 binding did stabilize the secondary structure in our simulations. Having said that, we also observed some interesting differences in TSPO dynamics in *holo*- form, that were not observed in *apo*- form. In ligand-bound form, we observed motions between residues of TM2 and TM5 that have dynamics with high correlation. We also found that these residues have strong evolutionary conservation, and some of these residues are also exposed to the surface. It could be possible that these residues may play a more central role in protein-protein interactions and/or, protein-lipid interactions. Since there are no targeted mutagenesis studies on these residues, we propose that this would be an interesting direction to proceed in future.

From this observation, we pondered whether the above-mentioned residues may be part of an allosteric dynamic pathway, since we observed that these residues are at a distance from the PK11195 binding site, and the motions involving these residues were absent in *apo*-form. To answer this question, we constructed protein structural network based on the correlated residue motions. Based on this network, we observed an allosteric pathway that spans the residues of TM5 to N-terminal region of TM1. While it is interesting to observe such an allosteric pathway that involve potentially significant residues, it is imperative that further studies are essential to obtain more insights into this pathway, and its functional ramifications. While we made this observation in context of protein dynamics, we were also able to support our observations from a biophysical perspective as well, where we found some motions correlated an allosteric pathway between the residues of TM2 and the N-terminus of TM1. Those particular motions could be associated to an allosteric pathway in TSPO. To our knowledge, this is the first study that describe an such correlated motions that could constitute allosteric dynamics, combining approaches from protein dynamics as well as protein interactions from biophysical perspective.

Apart from the dynamics of TSPO, we also studied the dynamics of PK11195 ligand, and we found that the ligand is also dynamic, and adopted different conformations involving both the rotameric forms when in complex with TSPO. We also observed transition between the different isomeric forms, which involved rotation of amide scaffold of the ligand. While we were able to characterise PK11195 dynamics in complex with TSPO, it would be interesting to study what impact do such dynamics have on the TSPO dynamics, as well as on pharmacological activity of PK11195. Furthermore, our

observations could be exploited in future to study binding dynamics of other TSPO ligands, as well as to develop new generation of TSPO ligands.

While our observations into PK11195-influenced TSPO dynamics shed some interesting insights into TSPO dynamics, there are numerous directions to proceed in future from this point. For instance, while we have studied dynamics of TSPO either in complex with PK11195, or in absence of any ligand, a different set of simulations could be performed in future, where the PK11195 ligand is at a short distance from TSPO. Such simulations will provide interesting insights on how TSPO accommodates PK11195 ligand and the role of extracellular loop between TM1 and TM2 in the same, since it has been suggested before that this loop may act as a gate to accommodate ligands. Apart from that, It would also be interesting to study how ligands that bind to different binding sites, such as Ro5-4864, TRO19622 and TRO40303 influence TSPO dynamics, and what are the contribution of different residues, as well as ligand scaffolds to the binding, by means of binding free energy calculations.

While TSPO oligomerization has been studied from both biochemical as well as structural biology experiments, an uncertainty still exists about the nature of TSPO oligomer. Different atomic structures of TSPO dimer solved till date have different interfaces. On the other hand, early experimental studies have shown that TSPO may exist in as multimeric states, involving up to 6-mer stoichiometries. We attempted to characterize dynamics of TSPO oligomerization. While we observed TSPO dimerization in many of our simulations, we also observed repulsion of TSPO monomers in certain simulations, suggesting that certain conditions may not favour TSPO dimerization.

Similarly, we also found that certain TM regions, particularly TM3, are more likely to be associated with the dimer interface, which is consistent with observation of TM3 regions as dimer interface from studies on TSPO structure in mouse and *R. sphaeroides*.

While TSPO consist of many SmxxxSm motifs that are observed to drive dimerization in many TM proteins, we found that aromatic residues, rather than such motifs drive TSPO oligomerization to a large extent. Since TSPO contains large number of SmxxxSm motifs, and Martini force field used in our simulations has been previously found to introduce a bias towards aggregation of proteins in membranes, all-atomistic simulations are needed to be performed on similar TSPO-dimer systems to fully understand the major players that stabilize TSPO dimerization. Interestingly, we observed certain dityrosine pairs were observed to form contacts for large part of simulations in multiple instances.

This observation is interesting due to the fact that TSPO oligomerization has been experimentally described to involve tyrosine residues by means of covalent interactions from earlier studies. Furthermore, it was found that deletion of tyrosine residues did impact the oligomerization and we

hypothesise that the interface that corresponded to such oligomerization may be similar to the dityrosine mediated dimer interface observed in our simulations. We thus propose that targeted mutagenesis experiments of these residues may further reveal the importance of these residues in ROS-induced TSPO oligomerization.

We also performed simulations on a TSPO dimer model, that had initial conformation similar as observed in NMR studies. Though we started with this model, we observed changes in the interface, which witnessed stronger participation of residues from TM1 as well as C-terminal of TM3. Interestingly, we also observed flipping of a cholesterol molecule from this simulation. As a part of this mechanism, the cholesterol molecule uses TM5 region of TSPO as an anchor and flips between lipid bilayer leaflets. We propose that this may constitute one of the TSPO-mediated cholesterol translocation mechanism. While we observed this translocation in coarse-grained simulation, further insights could be revealed by performing all-atom simulations on the TSPO dimer with the same interface and bilayer environment. Apart from that, future all-atom simulations of dimers observed in this study could be performed, and binding free energy analysis could be performed to understand the contribution of TSPO residues to the dimer interface.

Since NMR studies on the TSPO dimer described an allosteric regulation, we attempted to find out if similar long-range communication exist in our simulations. We not only found presence of such communications despite diversity in interfaces, we also found that such communication involves transfer of informations via the dimer interface. Furthermore, in certain simulations, we also found such communications involving the dimer interface (TM3) and residues of cholesterol-binding motifs in TM5, consistent with observations from NMR studies of TSPO dimer. Presence of such long-range communication pathways adds additional dimension to the dynamic nature of TSPO. How does this in turn impact TSPO dynamics and function remains to be understood.

Since there are significant differences between the experimental structures of mammalian and bacterial TSPO, we attempted to answer the question of whether the differences also translate to the dynamics, or if there is an overlap in the conformations explored by the mammalian and bacterial TSPO. We performed coarse-grained simulations of mammalian and bacterial TSPO, and we created an ensemble consisting of mammalian as well as bacterial TSPO trajectories. We compared the conformations explored by each TSPO by means of PCA. We found that eukaryotic and bacterial TSPO explored unique sets of conformations, though we found similarities in dynamics of residues in certain regions. We also observed that the residues having similar dynamics in both eukaryotic and bacterial TSPO are porphyrin-binding residues, and these residues are also well-conserved evolutionarily. Though there appears to be ambiguity on the biochemical processes associated with PP-IX binding to TSPO, our observations are consistent with existing knowledge on TSPO-PP-IX

interactions, that emphasise towards an important role of TM2 and TM5 residues on either of porphyrin catalysis or transport role of TSPO. An interesting direction to proceed in this respect would be to perform all-atom simulations of mammalian as well as bacterial TSPO in complex with PP-IX ligand. Since existing TSPO structure in complex with PPIX is a mutant structure, where TM5 region is not in interaction with PP-IX, TSPO PP-IX complexes can be generated by performing docking experiments, and using docked complexes to perform all-atom molecular dynamics simulations, as well as binding free-energy calculations to understand the contribution of different residues to PP-IX binding.

Another question we attempted to answer was regarding the divergence of TSPO1 and TSPO2. We attempted to trace the divergence of TSPO2 from phylogenetic perspective. From phylogenetic tree construction of TSPO sequences, we found existence of TSPO1 and TSPO2 sequences from reptiles and fishes, suggesting that the divergence of TSPO1 & TSPO2 may have occurred during early vertebrate evolution. However, the limitation of this hypothesis is the paucity of sequenced vertebrate genomes, due to which this hypothesis needs confirmation in future. We also reconstructed most probable ancestral sequence of present-day TSPO1 & TSPO2, and compared it with present-day TSPO1 & TSPO2 sequences. This comparison revealed significant differences in the residues that are shown to bind to PP-IX, PK11195 and cholesterol molecules. Apart from this, we also observed divergence in the residues associated with cholesterol binding, such as CARC motif, LAF motif and Ala147 residue. On account of these observations, we hypothesised that TSPO2 may lack cholesterol transport activity, though it retained cholesterol binding activity. However, this hypothesis needs to be supported in future by further experimental as well as computational studies.

We then attempted to understand the differences between TSPO1 and TSPO2 from structural point of view. For this, we constructed a multi-template homology model of TSPO2, and compared it with the NMR structure of mouse TSPO1, and we observed high RMSD between TSPO1 and TSPO2 structures. To answer the question of whether the difference in structures also carried over to the dynamics, we constructed normal modes of two structures and compared the same. We observed both differences as well as similarities between the normal mode dynamics of both structures. While we observed collective dynamics between TM2 and TM5 regions of TSPO1 but not TSPO2, we observed helix-stretching motions in TM1 region of TSPO2, but not TSPO1. However, we also observed similarities in the TSPO1 and TSPO2. We observed that in both the proteins, residues neighbouring to Trp143 residue (mouse TSPO1) of TM5 were rigid, and the residues at the periphery of this region were dynamic, indicating hinge-like dynamics of this region. Similar hinge-like dynamics may suggest characteristic TSPO fold motions, since this region has high identity among TSPO1 & TSPO2

sequences, as well as between mammalian and bacterial TSPO. Furthermore, residue Trp143 is also considered to be an important PP-IX binding residue.

Though we found that certain regions have similar dynamics in both TSPO1 and 2 to some extent, we observed that there are considerable differences in dynamics of TM1 and TM3 region between TSPO1 and TSPO2. While it is not clear whether such differences in sequence conservation and dynamics have a significant impact in PP-IX binding, it would be interesting to further study the effect of changed conformations and dynamics on the porphyrin binding ability. This could be done by performing MD simulations on wild-type TSPO2 in complex with PP-IX ligand obtained from docking studies, as well as mutant variants of TSPO2, where PP-IX binding residues common to TSPO1 and TSPO2 are substituted with a residue having different chemical properties. From these simulations, it would also be interesting to study the nature of dynamics of TM5 region, and test if hinge-like dynamics observed in this study may be observed in simulations as well.

We initiated the study and discussion on TSPO in this thesis in the context of involvement of TSPO2 and its partner proteins in malarial infection. However, understanding of TSPO dynamics so far has been insufficient, though the atomic structures of TSPO were solved recently. By obtaining deeper insights on the role of TSPO dynamics from ligand-binding as well as oligomerization perspectives, we propose that the knowledge obtained from this study could be exploited to further understand the dynamics of TSPO at the level of tertiary structure involving other interacting partners such as VDAC and ANT. Insights from such studies, as well as from this thesis on the role of PK11195 drug in influencing TSPO dynamics will be critical to develop novel methodologies to treat the malarial infection, for which a reliable therapeutic approach still remains elusive.

Furthermore, TSPO has also been suggested as a target for various other pathophysiological conditions, including cancers, neurodegenerative diseases such as Parkinson's disease, dementia etc., as well as a therapeutic target to mitigate the effects of inflammation in various tissues. Though ligands binding to TSPO with high affinity have been proposed as drugs for these conditions, the structural mechanism by which these drugs bring about such therapeutic effect is largely unknown, which is critical for development of drugs targeting these conditions. This work describes the role of one such ligand, PK11195 in modulating TSPO dynamics, and the knowledge gained from this work will be critical for further development of therapeutic approaches, where TSPO is targeted.

On the other hand, porphyrins and ROS are observed in context of the above-mentioned pathophysiological conditions. While their presence indicates an important role of TSPO-PPIX interactions in such conditions, exact nature of TSPO-PPIX interaction itself is not well understood. Our observations on possible involvement of certain TSPO residues in PP-IX binding and interaction may be further exploited to study and characterize TSPO-PPIX interactions, which contributes

significantly to understanding of the role of porphyrins in various pathophysiological conditions including malaria, for which there is a need to develop novel approaches to treat the disease.

The broader objective of this thesis was to exploit the knowledge gained from study of structural dynamics to design new generation of antimalarial drugs that target TSPO. The insights obtained from this thesis provides a great deal of knowledge which would be critical for this objective. Since comparison of TSPO2 with TSPO1 revealed certain important conserved residues, these residues could be used as a target for the new generation of drugs that can target both TSPO1 as well as TSPO2. Understanding of TSPO dynamics in presence of a ligand, and residues involved in allosteric communications could also be considered targets for novel drugs, since allosteric sites have been exploited as drug targets in other proteins as well. Knowledge on TSPO dynamics and interactions from this thesis will also be critical for studying interactions with VDAC, which is the primary interacting partner of TSPO and its 'partner in crime' in the malarial infections. In summary, the knowledge gained on the various facets of TSPO dynamics could be used in various ways as described above to develop TSPO drugs that not only target in malarial infections, but also in various other pathologies such as cancers, porphyria and neurodegenerative diseases that have active participation of TSPO.

References

Chapter-1:

Agnello, D., Carvelli, L., Muzio, V., Villa, P., Bottazzi, B., Polentarutti, N., Mennini, T., Mantovani, A., and Ghezzi, P. (2000). Increased peripheral benzodiazepine binding sites and pentraxin 3 expression in the spinal cord during EAE: relation to inflammatory cytokines and modulation by dexamethasone and rolipram. J. Neuroimmunol. *109*, 105–111.

van Agtmael, M.A., Eggelte, T.A., and van Boxtel, C.J. (1999). Artemisinin drugs in the treatment of malaria: from medicinal herb to registered medication. Trends Pharmacol. Sci. *20*, 199–205.

Amaratunga, C., Sreng, S., Suon, S., Phelps, E.S., Stepniewska, K., Lim, P., Zhou, C., Mao, S., Anderson, J.M., Lindegardh, N., et al. (2012). Artemisinin-resistant Plasmodium falciparum in Pursat province, western Cambodia: a parasite clearance rate study. Lancet Infect. Dis. *12*, 851–858.

Austin, C.J.D., Kahlert, J., Kassiou, M., and Rendina, L.M. (2013). The translocator protein (TSPO): a novel target for cancer chemotherapy. Int. J. Biochem. Cell Biol. *45*, 1212–1216.

Banati, R.B., Middleton, R.J., Chan, R., Hatty, C.R., Kam, W.W.-Y., Quin, C., Graeber, M.B., Parmar, A., Zahra, D., Callaghan, P., et al. (2014). Positron emission tomography and functional characterization of a complete PBR/TSPO knockout. Nat. Commun. *5*, 5452.

Bayrhuber, M., Meins, T., Habeck, M., Becker, S., Giller, K., Villinger, S., Vonrhein, C., Griesinger, C., Zweckstetter, M., and Zeth, K. (2008). Structure of the human voltage-dependent anion channel. Proc Natl Acad Sci USA *105*, 15370–15375.

Bénavidès, J., Guilloux, F., Allam, D.E., Uzan, A., Mizoule, J., Renault, C., Dubroeucq, M.C., Guérémy, C., and Le Fur, G. (1984). Opposite effects of an agonist, R05-4864, and an antagonist, PK 11195, of the peripheral type benzodiazepine binding sites on audiogenic seizures in DBA/2J mice. Life Sci. *34*, 2613–2620.

Bernassau, J.M., Reversat, J.L., Ferrara, P., Caput, D., and Lefur, G. (1993). A 3D model of the peripheral benzodiazepine receptor and its implication in intra mitochondrial cholesterol transport. J. Mol. Graph. *11*, 236–244, 235.

Besman, M.J., Yanagibashi, K., Lee, T.D., Kawamura, M., Hall, P.F., and Shively, J.E. (1989). Identification of des-(Gly-Ile)-endozepine as an effector of corticotropin-dependent adrenal steroidogenesis: stimulation of cholesterol delivery is mediated by the peripheral benzodiazepine receptor. Proc Natl Acad Sci USA *86*, 4897–4901.

Bhoola, N.H., Mbita, Z., Hull, R., and Dlamini, Z. (2018). Translocator protein (TSPO) as a potential biomarker in human cancers. Int. J. Mol. Sci. *19*.

Bono, F., Lamarche, I., Prabonnaud, V., Le Fur, G., and Herbert, J.M. (1999). Peripheral benzodiazepine receptor agonists exhibit potent antiapoptotic activities. Biochem. Biophys. Res. Commun. *265*, 457–461.

Bonsack, F., Alleyne, C.H., and Sukumari-Ramesh, S. (2016). Augmented expression of TSPO after intracerebral hemorrhage: a role in inflammation? J. Neuroinflammation *13*, 151.

Bordet, T., Buisson, B., Michaud, M., Drouot, C., Galéa, P., Delaage, P., Akentieva, N.P., Evers, A.S., Covey, D.F., Ostuni, M.A., et al. (2007). Identification and characterization of cholest-4-en-3-one, oxime (TRO19622), a novel drug candidate for amyotrophic lateral sclerosis. J. Pharmacol. Exp. Ther. *322*, 709–720.

Bose, M., Whittal, R.M., Miller, W.L., and Bose, H.S. (2008). Steroidogenic activity of StAR requires contact with mitochondrial VDAC1 and phosphate carrier protein. J. Biol. Chem. *283*, 8837–8845.

Bouyer, G., Cueff, A., Egée, S., Kmiecik, J., Maksimova, Y., Glogowska, E., Gallagher, P.G., and Thomas, S.L.Y. (2011). Erythrocyte peripheral type benzodiazepine receptor/voltage-dependent anion channels are upregulated by Plasmodium falciparum. Blood *118*, 2305–2312.

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

Budelier, M.M., Cheng, W.W.L., Bergdoll, L., Chen, Z.-W., Janetka, J.W., Abramson, J., Krishnan, K., Mydock-McGrane, L., Covey, D.F., Whitelegge, J.P., et al. (2017). Photoaffinity labeling with cholesterol analogues precisely maps a cholesterol-binding site in voltage-dependent anion channel-1. J. Biol. Chem. *292*, 9294–9304.

Cantoni, L., Rizzardini, M., Skorupska, M., Cagnotto, A., Codegoni, A., Pecora, N., Frigo, L., Ferrarese, C., and Mennini, T. (1992). Hepatic protoporphyria is associated with a decrease in ligand binding for the mitochondrial benzodiazepine receptors in the liver. Biochem. Pharmacol. *44*, 1159–1164.

Chang, R.S., and Snyder, S.H. (1978). Benzodiazepine receptors: labeling in intact animals with [3H] flunitrazepam. Eur. J. Pharmacol. *48*, 213–218.

Chardenot, P., Roubert, C., Galiègue, S., Casellas, P., Le Fur, G., Soubrié, P., and Oury-Donat, F. (2002). Expression profile and up-regulation of PRAX-1 mRNA by antidepressant treatment in the rat brain. Mol. Pharmacol. *62*, 1314–1320.

Chechneva, O.V., and Deng, W. (2016). Mitochondrial translocator protein (TSPO), astrocytes and neuroinflammation. Neural Regen. Res. *11*, 1056–1057.

Chelli, B., Lena, A., Vanacore, R., Da Pozzo, E., Costa, B., Rossi, L., Salvetti, A., Scatena, F., Ceruti, S., Abbracchio, M.P., et al. (2004). Peripheral benzodiazepine receptor ligands: mitochondrial transmembrane potential depolarization and apoptosis induction in rat C6 glioma cells. Biochem. Pharmacol. *68*, 125–134.

Chen, M.-K., and Guilarte, T.R. (2008). Translocator protein 18 kDa (TSPO): molecular sensor of brain injury and repair. Pharmacol. Ther. *118*, 1–17.

Colasanti, A., Owen, D.R., Grozeva, D., Rabiner, E.A., Matthews, P.M., Craddock, N., and Young, A.H. (2013). Bipolar Disorder is associated with the rs6971 polymorphism in the gene encoding 18 kDa Translocator Protein (TSPO). Psychoneuroendocrinology *38*, 2826–2829.

Costa, B., Pini, S., Gabelloni, P., Da Pozzo, E., Abelli, M., Lari, L., Preve, M., Lucacchini, A., Cassano, G.B., and Martini, C. (2009). The spontaneous Ala147Thr amino acid substitution within the translocator protein influences pregnenolone production in lymphomonocytes of healthy individuals. Endocrinology *150*, 5438–5445.

Crompton, M., Virji, S., and Ward, J.M. (1998). Cyclophilin-D binds strongly to complexes of the voltage-dependent anion channel and the adenine nucleotide translocase to form the permeability

transition pore. Eur. J. Biochem. *258*, 729–735.

Culty, M., Li, H., Boujrad, N., Amri, H., Vidic, B., Bernassau, J.M., Reversat, J.L., and Papadopoulos, V. (1999). In vitro studies on the role of the peripheral-type benzodiazepine receptor in steroidogenesis. J. Steroid Biochem. Mol. Biol. *69*, 123–130.

Daniele, S., La Pietra, V., Barresi, E., Di Maro, S., Da Pozzo, E., Robello, M., La Motta, C., Cosconati, S., Taliani, S., Marinelli, L., et al. (2016). Lead Optimization of 2-Phenylindolylglyoxylyldipeptide Murine Double Minute (MDM)2/Translocator Protein (TSPO) Dual Inhibitors for the Treatment of Gliomas. J. Med. Chem. *59*, 4526–4538.

Das, S., Saha, B., Hati, A.K., and Roy, S. (2018). Evidence of Artemisinin-Resistant Plasmodium falciparum Malaria in Eastern India. N. Engl. J. Med. *379*, 1962–1964.

Daugherty, D.J., Selvaraj, V., Chechneva, O.V., Liu, X.-B., Pleasure, D.E., and Deng, W. (2013). A TSPO ligand is protective in a mouse model of multiple sclerosis. EMBO Mol. Med. *5*, 891–903.

Davey, M.E., and de Bruijn, F.J. (2000). A homologue of the tryptophan-rich sensory protein TspO and FixL regulate a novel nutrient deprivation-induced Sinorhizobium meliloti locus. Appl. Environ. Microbiol. *66*, 5353–5359.

DeLano, W.L. (2002). Pymol: An open-source molecular graphics tool. CCP4 Newsletter On Protein Crystallography *40*, 82–92.

Duneau, J.-P., Vegh, A.P., and Sturgis, J.N. (2007). A dimerization hierarchy in the transmembrane domains of the HER receptor family. Biochemistry *46*, 2010–2019.

Eilers, M., Patel, A.B., Liu, W., and Smith, S.O. (2002). Comparison of helix interactions in membrane and soluble alpha-bundle proteins. Biophys. J. *82*, 2720–2736.

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

Fan, J., Liu, J., Culty, M., and Papadopoulos, V. (2010). Acyl-coenzyme A binding domain containing 3 (ACBD3; PAP7; GCP60): an emerging signaling molecule. Prog. Lipid Res. *49*, 218–234.

Fan, J., Lindemann, P., Feuilloley, M.G.J., and Papadopoulos, V. (2012). Structural and functional evolution of the translocator protein (18 kDa). Curr. Mol. Med. *12*, 369–386.

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

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

Farges, R., Joseph-Liauzun, E., Shire, D., Caput, D., Le Fur, G., and Ferrara, P. (1994). Site-directed mutagenesis of the peripheral benzodiazepine receptor: identification of amino acids implicated in the binding site of Ro5-4864. Mol. Pharmacol. *46*, 1160–1167.

Frank, W., Baar, K.-M., Qudeimat, E., Woriedh, M., Alawady, A., Ratnadewi, D., Gremillon, L., Grimm, B., and Reski, R. (2007). A mitochondrial protein homologous to the mammalian peripheral-type benzodiazepine receptor is essential for stress adaptation in plants. Plant J. *51*, 1004–1018.

Galiègue, S., Jbilo, O., Combes, T., Bribes, E., Carayon, P., Le Fur, G., and Casellas, P. (1999).

Cloning and characterization of PRAX-1. A new protein that specifically interacts with the peripheral benzodiazepine receptor. J. Biol. Chem. *274*, 2938–2952.

Garnier, M., Dimchev, A. B., Boujrad, N., Price, J. M., Musto, N. A. and Papadopoulos, V. (1994). In vitro reconstitution of a functional peripheral-type benzodiazepine receptor from mouse Leydig tumor cells. Mol. Pharmacol. *45*, 201-211.

Gatliff, J., and Campanella, M. (2016). TSPO: kaleidoscopic 18-kDa amid biochemical pharmacology, control and targeting of mitochondria. Biochem. J. *473*, 107–121.

Gatliff, J., East, D., Crosby, J., Abeti, R., Harvey, R., Craigen, W., Parker, P., and Campanella, M. (2014). TSPO interacts with VDAC1 and triggers a ROS-mediated inhibition of mitochondrial quality control. Autophagy *10*, 2279–2296.

Gerhard, A., Pavese, N., Hotton, G., Turkheimer, F., Es, M., Hammers, A., Eggert, K., Oertel, W., Banati, R.B., and Brooks, D.J. (2006). In vivo imaging of microglial activation with [11C](R)-PK11195 PET in idiopathic Parkinson's disease. Neurobiol. Dis. *21*, 404–412.

Ginsburg, H., and Stein, W.D. (2005). How many functional transport pathways does Plasmodium falciparum induce in the membrane of its host erythrocyte? Trends Parasitol. *21*, 118–121.

Ginsburg, H., Krugliak, M., Eidelman, O., and Cabantchik, Z.I. (1983). New permeability pathways induced in membranes of Plasmodium falciparum infected erythrocytes. Mol. Biochem. Parasitol. *8*, 177–190.

Ginsburg, H., Kutner, S., Krugliak, M., and Cabantchik, Z.I. (1985). Characterization of permeation pathways appearing in the host membrane of Plasmodium falciparum infected red blood cells. Mol. Biochem. Parasitol. *14*, 313–322.

Ginter, C., Kiburu, I., and Boudker, O. (2013). Chemical catalysis by the translocator protein (18 kDa). Biochemistry *52*, 3609–3611.

Guillaumot, D., Guillon, S., Déplanque, T., Vanhee, C., Gumy, C., Masquelier, D., Morsomme, P., and Batoko, H. (2009). The Arabidopsis TSPO-related protein is a stress and abscisic acid-regulated, endoplasmic reticulum-Golgi-localized membrane protein. Plant J. *60*, 242–256.

Guo, Y., Kalathur, R.C., Liu, Q., Kloss, B., Bruni, R., Ginter, C., Kloppmann, E., Rost, B., and Hendrickson, W.A. (2015). Protein structure. Structure and activity of tryptophan-rich TSPO proteins. Science *347*, 551–555.

Gut, P. (2015). Targeting mitochondrial energy metabolism with TSPO ligands. Biochem. Soc. Trans. *43*, 537–542.

Halestrap, A.P., McStay, G.P., and Clarke, S.J. (2002). The permeability transition pore complex: another view. Biochimie *84*, 153–166.

Hansson, M.J., Llwyd, O., Morin, D., de Paulis, D., Arnoux, T., Gouarné, C., Koul, S., Engblom, H., Bordet, T., Tissier, R., et al. (2015). Differences in the profile of protection afforded by TRO40303 and mild hypothermia in models of cardiac ischemia/reperfusion injury. Eur. J. Pharmacol. *760*, 7–19.

Hauet, T., Liu, J., Li, H., Gazouli, M., Culty, M., and Papadopoulos, V. (2002). PBR, StAR, and PKA: partners in cholesterol transport in steroidogenic cells. Endocr. Res. *28*, 395–401.

Health, O.W. (2017). World malaria report 2017.

Hiller, S., Garces, R.G., Malia, T.J., Orekhov, V.Y., Colombini, M., and Wagner, G. (2008). Solution

structure of the integral human membrane protein VDAC-1 in detergent micelles. Science *321*, 1206–1210.

Hinsen, K., Vaitinadapoule, A., Ostuni, M.A., Etchebest, C., and Lacapere, J.-J. (2015). Construction and validation of an atomic model for bacterial TSPO from electron microscopy density, evolutionary constraints, and biochemical and biophysical data. Biochim. Biophys. Acta *1848*, 568–580.

Hirsch, J.D., Beyer, C.F., Malkowitz, L., Beer, B., and Blume, A.J. (1989). Mitochondrial benzodiazepine receptors mediate inhibition of mitochondrial respiratory control. Mol. Pharmacol. *35*, 157–163.

Humphrey, W., Dalke, A., and Schulten, K. (1996). VMD: visual molecular dynamics. J. Mol. Graph. *14*, 33–38, 27.

Inoue, O., Yamasaki, T., Hashimoto, K., and Kojima, M. (1985). [Evaluation of 3H-PK11195 as a radioligand for the in vivo study of peripheral benzodiazepine receptor]. Kaku Igaku *22*, 1385–1389.

Jamin, N., Neumann, J.-M., Ostuni, M.A., Vu, T.K.N., Yao, Z.-X., Murail, S., Robert, J.-C., Giatzakis, C., Papadopoulos, V., and Lacapère, J.-J. (2005). Characterization of the cholesterol recognition amino acid consensus sequence of the peripheral-type benzodiazepine receptor. Mol. Endocrinol. *19*, 588–594.

Jaremko, L., Jaremko, M., Becker, S., and Zweckstetter, M. (2014a). Toward the functional oligomerization state of tryptophan-rich sensory proteins. Protein Sci. *23*, 1154–1160.

Jaremko, L., Jaremko, M., Giller, K., Becker, S., and Zweckstetter, M. (2014b). Structure of the mitochondrial translocator protein in complex with a diagnostic ligand. Science *343*, 1363–1366.

Jaremko, M., Jaremko, Ł., Jaipuria, G., Becker, S., and Zweckstetter, M. (2015a). Structure of the mammalian TSPO/PBR protein. Biochem. Soc. Trans. *43*, 566–571.

Jaremko, Ł., Jaremko, M., Giller, K., Becker, S., and Zweckstetter, M. (2015b). Conformational flexibility in the transmembrane protein TSPO. Chem. Eur. J *21*, 16555–16563.

Joseph-Liazun, E., Farges, E., Delmas, P., Ferrara, P., Loison, G. (1997). The Mr 18,000 subunit of the peripheral-type benzodiazepine receptor exhibits both benzodiazepine and isoquinoline carboxamide binding sites in the absence of the voltage-dependent anion channel or the adenine nucleotide carrier, J. Biol. Chem. *272* 28102–28106.

Kleiger, G., Grothe, R., Mallick, P., and Eisenberg, D. (2002). GXXXG and AXXXA: common alpha-helical interaction motifs in proteins, particularly in extremophiles. Biochemistry *41*, 5990–5997.

Klein, E.Y. (2013). Antimalarial drug resistance: a review of the biology and strategies to delay emergence and spread. Int. J. Antimicrob. Agents *41*, 311–317.

Klima, M., Tóth, D.J., Hexnerova, R., Baumlova, A., Chalupska, D., Tykvart, J., Rezabkova, L., Sengupta, N., Man, P., Dubankova, A., et al. (2016). Structural insights and in vitro reconstitution of membrane targeting and activation of human PI4KB by the ACBD3 protein. Sci. Rep. *6*, 23641.

Klima, M., Chalupska, D., Różycki, B., Humpolickova, J., Rezabkova, L., Silhan, J., Baumlova, A., Dubankova, A., and Boura, E. (2017). Kobuviral Non-structural 3A Proteins Act as Molecular Harnesses to Hijack the Host ACBD3 Protein. Structure *25*, 219–230.

Knudsen, J. (1991). Acyl-CoA-binding and transport, an alternative function for diazepam binding

inhibitor (DBI), which is identical with acyl-CoA-binding protein. Neuropharmacology *30*, 1405–1410.

Kokoszka, J.E., Waymire, K.G., Levy, S.E., Sligh, J.E., Cai, J., Jones, D.P., MacGregor, G.R., and Wallace, D.C. (2004). The ADP/ATP translocator is not essential for the mitochondrial permeability transition pore. Nature *427*, 461–465.

Korkhov, V.M., Sachse, C., Short, J.M., and Tate, C.G. (2010). Three-dimensional structure of TspO by electron cryomicroscopy of helical crystals. Structure *18*, 677–687.

Krammer, E.-M., Homblé, F., and Prévost, M. (2011). Concentration dependent ion selectivity in VDAC: a molecular dynamics simulation study. PLoS ONE *6*, e27994.

Krammer, E.-M., Homblé, F., and Prévost, M. (2013). Molecular origin of VDAC selectivity towards inorganic ions: a combined molecular and Brownian dynamics study. Biochim. Biophys. Acta *1828*, 1284–1292.

Krammer, E.-M., Vu, G.T., Homblé, F., and Prévost, M. (2015). Dual mechanism of ion permeation through VDAC revealed with inorganic phosphate ions and phosphate metabolites. PLoS ONE *10*, e0121746.

Kreisl, W.C., Lyoo, C.H., McGwier, M., Snow, J., Jenko, K.J., Kimura, N., Corona, W., Morse, C.L., Zoghbi, S.S., Pike, V.W., et al. (2013). In vivo radioligand binding to translocator protein correlates with severity of Alzheimer's disease. Brain *136*, 2228–2238.

Krueger, K.E., and Papadopoulos, V. (1990). Peripheral-type benzodiazepine receptors mediate translocation of cholesterol from outer to inner mitochondrial membranes in adrenocortical cells. J. Biol. Chem. *265*, 15015–15022.

Kugler, W., Veenman, L., Shandalov, Y., Leschiner, S., Spanier, I., Lakomek, M., and Gavish, M. (2008). Ligands of the mitochondrial 18 kDa translocator protein attenuate apoptosis of human glioblastoma cells exposed to erucylphosphohomocholine. Cell. Oncol. *30*, 435–450.

Kunduzova, O.R., Escourrou, G., De La Farge, F., Salvayre, R., Séguélas, M.-H., Leducq, N., Bono, F., Herbert, J.-M., and Parini, A. (2004). Involvement of peripheral benzodiazepine receptor in the oxidative stress, death-signaling pathways, and renal injury induced by ischemia-reperfusion. J. Am. Soc. Nephrol. *15*, 2152–2160.

Lacapère, J.J., Delavoie, F., Li, H., Péranzi, G., Maccario, J., Papadopoulos, V., and Vidic, B. (2001). Structural and functional study of reconstituted peripheral benzodiazepine receptor. Biochem. Biophys. Res. Commun. *284*, 536–541.

Lavisse, S., Guillermier, M., Hérard, A.-S., Petit, F., Delahaye, M., Van Camp, N., Ben Haim, L., Lebon, V., Remy, P., Dollé, F., et al. (2012). Reactive astrocytes overexpress TSPO and are detected by TSPO positron emission tomography imaging. J. Neurosci. *32*, 10809–10818.

Lee, Y.-S., Siméon, F.G., Briard, E., and Pike, V.W. (2012). Solution structures of the prototypical 18 kDa translocator protein ligand, PK 11195, elucidated with 1H/13C NMR spectroscopy and quantum chemistry. ACS Chem. Neurosci. *3*, 325–335.

Lemmon, M.A., Flanagan, J.M., Hunt, J.F., Adair, B.D., Bormann, B.J., Dempsey, C.E., and Engelman, D.M. (1992a). Glycophorin A dimerization is driven by specific interactions between transmembrane alpha-helices. J. Biol. Chem. *267*, 7683–7689.

Lemmon, M.A., Flanagan, J.M., Treutlein, H.R., Zhang, J., and Engelman, D.M. (1992b). Sequence specificity in the dimerization of transmembrane alpha-helices. Biochemistry *31*, 12719–12725.

Leonelli, E., Yague, J.G., Ballabio, M., Azcoitia, I., Magnaghi, V., Schumacher, M., Garcia-Segura, L.M., and Melcangi, R.C. (2005). Ro5-4864, a synthetic ligand of peripheral benzodiazepine receptor, reduces aging-associated myelin degeneration in the sciatic nerve of male rats. Mech. Ageing Dev. *126*, 1159–1163.

Levin, E., Premkumar, A., Veenman, L., Kugler, W., Leschiner, S., Spanier, I., Weisinger, G., Lakomek, M., Weizman, A., Snyder, S.H., et al. (2005). The peripheral-type benzodiazepine receptor and tumorigenicity: isoquinoline binding protein (IBP) antisense knockdown in the C6 glioma cell line. Biochemistry *44*, 9924–9935.

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

Li, E., Wimley, W.C., and Hristova, K. (2012). Transmembrane helix dimerization: beyond the search for sequence motifs. Biochim. Biophys. Acta *1818*, 183–193.

Li, F., Xia, Y., Meiler, J., and Ferguson-Miller, S. (2013). Characterization and modeling of the oligomeric state and ligand binding behavior of purified translocator protein 18 kDa from Rhodobacter sphaeroides. Biochemistry *52*, 5884–5899.

Li, F., Liu, J., Valls, L., Hiser, C., and Ferguson-Miller, S. (2015a). Identification of a key cholesterol binding enhancement motif in translocator protein 18 kDa. Biochemistry *54*, 1441–1443.

Li, F., Liu, J., Zheng, Y., Garavito, R.M., and Ferguson-Miller, S. (2015b). Protein structure. Crystal structures of translocator protein (TSPO) and mutant mimic of a human polymorphism. Science *347*, 555–558.

Li, F., Liu, J., Garavito, R.M., and Ferguson-Miller, S. (2015c). Evolving understanding of translocator protein 18 kDa (TSPO). Pharmacol. Res. *99*, 404–409.

Li, F., Liu, J., Zheng, Y., Garavito, R.M., and Ferguson-Miller, S. (2015d). Response to Comment on “Crystal structures of translocator protein (TSPO) and mutant mimic of a human polymorphism”. Science *350*, 519.

Li, F., Liu, J., Liu, N., Kuhn, L.A., Garavito, R.M., and Ferguson-Miller, S. (2016a). Translocator Protein 18 kDa (TSPO): An Old Protein with New Functions? Biochemistry *55*, 2821–2831.

Li, H., Yao, Z., Degenhardt, B., Teper, G., and Papadopoulos, V. (2001a). 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 *98*, 1267–1272.

Li, H., Degenhardt, B., Tobin, D., Yao, Z.X., Tasken, K., and Papadopoulos, V. (2001b). Identification, localization, and function in steroidogenesis of PAP7: a peripheral-type benzodiazepine receptor- and PKA (RIalpha)-associated protein. Mol. Endocrinol. *15*, 2211–2228.

Li, J., Zhang, Z., Lv, L., Qiao, H., Chen, X., and Zou, C. (2016b). A bispecific antibody (ScBsAbAgn-2/TSPO) target for Ang-2 and TSPO resulted in therapeutic effects against glioblastomas. Biochem. Biophys. Res. Commun. *472*, 384–391.

Lin, D., Sugawara, T., Strauss, J.F., Clark, B.J., Stocco, D.M., Saenger, P., Rogol, A., and Miller, W.L. (1995). Role of steroidogenic acute regulatory protein in adrenal and gonadal steroidogenesis. Science *267*, 1828–1831.

Lindemann, P., Koch, A., Degenhardt, B., Hause, G., Grimm, B., and Papadopoulos, V. (2004). A novel Arabidopsis thaliana protein is a functional peripheral-type benzodiazepine receptor. Plant Cell Physiol. *45*, 723–733.

Liu, B., Le, K.X., Park, M.-A., Wang, S., Belanger, A.P., Dubey, S., Frost, J.L., Holton, P., Reiser, V., Jones, P.A., et al. (2015). In Vivo Detection of Age- and Disease-Related Increases in Neuroinflammation by 18F-GE180 TSPO MicroPET Imaging in Wild-Type and Alzheimer's Transgenic Mice. J. Neurosci. *35*, 15716–15730.

Liu, G.-J., Middleton, R.J., Hatty, C.R., Kam, W.W.-Y., Chan, R., Pham, T., Harrison-Brown, M., Dodson, E., Veale, K., and Banati, R.B. (2014). The 18 kDa translocator protein, microglia and neuroinflammation. Brain Pathol. *24*, 631–653.

Liu, G.-J., Middleton, R.J., Kam, W.W.-Y., Chin, D.Y., Hatty, C.R., Chan, R.H.Y., and Banati, R.B. (2017). Functional gains in energy and cell metabolism after TSPO gene insertion. Cell Cycle *16*, 436–447.

Liu, J., Rone, M.B., and Papadopoulos, V. (2006). Protein-protein interactions mediate mitochondrial cholesterol transport and steroid biosynthesis. J. Biol. Chem. *281*, 38879–38893.

Lockhart, A., Davis, B., Matthews, J.C., Rahmoune, H., Hong, G., Gee, A., Earnshaw, D., and Brown, J. (2003). The peripheral benzodiazepine receptor ligand PK11195 binds with high affinity to the acute phase reactant alpha1-acid glycoprotein: implications for the use of the ligand as a CNS inflammatory marker. Nucl. Med. Biol. *30*, 199–206.

Lomize, M.A., Pogozheva, I.D., Joo, H., Mosberg, H.I., and Lomize, A.L. (2012). OPM database and PPM web server: resources for positioning of proteins in membranes. Nucleic Acids Res. *40*, D370-6.

Maaser, K., Höpfner, M., Jansen, A., Weisinger, G., Gavish, M., Kozikowski, A.P., Weizman, A., Carayon, P., Riecken, E.O., Zeitz, M., et al. (2001). Specific ligands of the peripheral benzodiazepine receptor induce apoptosis and cell cycle arrest in human colorectal cancer cells. Br. J. Cancer *85*, 1771–1780.

Marangos, P.J., Patel, J., Boulenger, J.P., and Clark-Rosenberg, R. (1982). Characterization of peripheral-type benzodiazepine binding sites in brain using [3H]Ro 5-4864. Mol. Pharmacol. *22*, 26–32.

Marginedas-Freixa, I., Hattab, C., Bouyer, G., Halle, F., Chene, A., Lefevre, S.D., Cambot, M., Cueff, A., Schmitt, M., Gamain, B., et al. (2016). TSPO ligands stimulate ZnPPIX transport and ROS accumulation leading to the inhibition of P. falciparum growth in human blood. Sci. Rep. *6*, 33516.

Mazzeo, A.T., Beat, A., Singh, A., and Bullock, M.R. (2009). The role of mitochondrial transition pore, and its modulation, in traumatic brain injury and delayed neurodegeneration after TBI. Exp. Neurol. *218*, 363–370.

McEnery, M.W., Snowman, A.M., Trifiletti, R.R., and Snyder, S.H. (1992). Isolation of the mitochondrial benzodiazepine receptor: association with the voltage-dependent anion channel and the adenine nucleotide carrier. Proc Natl Acad Sci USA *89*, 3170–3174.

McPhail, J.A., Ottosen, E.H., Jenkins, M.L., and Burke, J.E. (2017). The Molecular Basis of Aichi

Virus 3A Protein Activation of Phosphatidylinositol 4 Kinase IIIβ, PI4KB, through ACBD3. Structure *25*, 121–131.

Mendonça-Torres, M.C., and Roberts, S.S. (2013). The translocator protein (TSPO) ligand PK11195 induces apoptosis and cell cycle arrest and sensitizes to chemotherapy treatment in pre- and post-relapse neuroblastoma cell lines. Cancer Biol. Ther. *14*, 319–326.

Mills, C., Makwana, M., Wallace, A., Benn, S., Schmidt, H., Tegeder, I., Costigan, M., Brown, R.H., Raivich, G., and Woolf, C.J. (2008). Ro5-4864 promotes neonatal motor neuron survival and nerve regeneration in adult rats. Eur. J. Neurosci. *27*, 937–946.

Morin, D., Musman, J., Pons, S., Berdeaux, A., and Ghaleb, B. (2016). Mitochondrial translocator protein (TSPO): From physiology to cardioprotection. Biochem. Pharmacol. *105*, 1–13.

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

Mueller, B.K., Subramaniam, S., and Senes, A. (2014). A frequent, GxxxG-mediated, transmembrane association motif is optimized for the formation of interhelical Cα-H hydrogen bonds. Proc Natl Acad Sci USA *111*, E888-95.

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

Murail, S., Robert, J.-C., Coïc, Y.-M., Neumann, J.-M., Ostuni, M.A., Yao, Z.-X., Papadopoulos, V., Jamin, N., and Lacapère, J.-J. (2008). Secondary and tertiary structures of the transmembrane domains of the translocator protein TSPO determined by NMR. Stabilization of the TSPO tertiary fold upon ligand binding. Biochim. Biophys. Acta *1778*, 1375–1381.

Nakamura, K., Yamada, K., Iwayama, Y., Toyota, T., Furukawa, A., Takimoto, T., Terayama, H., Iwahashi, K., Takei, N., Minabe, Y., et al. (2006). Evidence that variation in the peripheral benzodiazepine receptor (PBR) gene influences susceptibility to panic disorder. Am. J. Med. Genet. B Neuropsychiatr. Genet. *141B*, 222–226.

Nakazawa, F., Alev, C., Shin, M., Nakaya, Y., Jakt, L.M., and Sheng, G. (2009). PBRL, a putative peripheral benzodiazepine receptor, in primitive erythropoiesis. Gene Expr. Patterns *9*, 114–121.

Nothdurfter, C., Baghai, T.C., Schüle, C., and Rupprecht, R. (2012). Translocator protein (18 kDa) (TSPO) as a therapeutic target for anxiety and neurologic disorders. Eur. Arch. Psychiatry Clin. Neurosci. *262 Suppl 2*, S107-12.

Obame, F.N., Zini, R., Souktani, R., Berdeaux, A., and Morin, D. (2007). Peripheral benzodiazepine receptor-induced myocardial protection is mediated by inhibition of mitochondrial membrane permeabilization. J. Pharmacol. Exp. Ther. *323*, 336–345.

Olson, J.M., Junck, L., Young, A.B., Penney, J.B., and Mancini, W.R. (1988). Isoquinoline and peripheral-type benzodiazepine binding in gliomas: implications for diagnostic imaging. Cancer Res. *48*, 5837–5841.

Olson, J.M., McNeel, W., Young, A.B., and Mancini, W.R. (1992). Localization of the peripheral-type benzodiazepine binding site to mitochondria of human glioma cells. J Neurooncol *13*, 35–42.

Papadopoulos, V., and Lecanu, L. (2009). Translocator protein (18 kDa) TSPO: an emerging

therapeutic target in neurotrauma. Exp. Neurol. *219*, 53–57.

Papadopoulos, V., Mukhin, A.G., Costa, E., and Krueger, K.E. (1990). The peripheral-type benzodiazepine receptor is functionally linked to Leydig cell steroidogenesis. J. Biol. Chem. *265*, 3772–3779.

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

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

Papadopoulos, V., Amri, H., Boujrad, N., Cascio, C., Culty, M., Garnier, M., Hardwick, M., Li, H., Vidic, B., Brown, A.S., et al. (1997a). Peripheral benzodiazepine receptor in cholesterol transport and steroidogenesis. Steroids *62*, 21–28.

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

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

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

Papadopoulos, V., Fan, J., and Zirkin, B. (2017). Translocator protein (18 kDa): an update on its function in steroidogenesis. J. Neuroendocrinol. *30*.

Paradis, S., Leoni, V., Caccia, C., Berdeaux, A., and Morin, D. (2013). Cardioprotection by the TSPO ligand 4'-chlorodiazepam is associated with inhibition of mitochondrial accumulation of cholesterol at reperfusion. Cardiovasc. Res. *98*, 420–427.

Pastorino, J.G., Simbula, G., Gilfor, E., Hoek, J.B., and Farber, J.L. (1994). Protoporphyrin IX, an endogenous ligand of the peripheral benzodiazepine receptor, potentiates induction of the mitochondrial permeability transition and the killing of cultured hepatocytes by rotenone. J. Biol. Chem. *269*, 31041–31046.

Pebay-Peyroula, E., Dahout-Gonzalez, C., Kahn, R., Trézéguet, V., Lauquin, G.J.-M., and Brandolin, G. (2003). Structure of mitochondrial ADP/ATP carrier in complex with carboxyatractyloside. Nature *426*, 39–44.

Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C., and Ferrin, T.E. (2004). UCSF Chimera—a visualization system for exploratory research and analysis. J. Comput. Chem. *25*, 1605–1612.

Pogoryelov, D., Klyszejko, A.L., Krasnoselska, G.O., Heller, E.-M., Leone, V., Langer, J.D., Vonck, J., Müller, D.J., Faraldo-Gómez, J.D., and Meier, T. (2012). Engineering rotor ring stoichiometries in the ATP synthase. Proc Natl Acad Sci USA *109*, E1599-608.

Raeburn, D., Miller, L.G., and Summer, W.R. (1988). Peripheral type benzodiazepine receptor and airway smooth muscle relaxation. J. Pharmacol. Exp. Ther. *245*, 557–562.

Raghavendra Rao, V.L., Dogan, A., Bowen, K.K., and Dempsey, R.J. (2000). Traumatic brain injury leads to increased expression of peripheral-type benzodiazepine receptors, neuronal death, and activation of astrocytes and microglia in rat thalamus. Exp. Neurol. *161*, 102–114.

Rampon, C., Bouzaffour, M., Ostuni, M.A., Dufourcq, P., Girard, C., Freyssinet, J.M., Lacapere, J.-J., Schweizer-Groyer, G., and Vriz, S. (2009). Translocator protein (18 kDa) is involved in primitive erythropoiesis in zebrafish. FASEB J. *23*, 4181–4192.

Rechichi, M., Salvetti, A., Chelli, B., Costa, B., Da Pozzo, E., Spinetti, F., Lena, A., Evangelista, M., Rainaldi, G., Martini, C., et al. (2008). TSPO over-expression increases motility, transmigration and proliferation properties of C6 rat glioma cells. Biochim. Biophys. Acta *1782*, 118–125.

Repalli, J. (2014). Translocator protein (TSPO) role in aging and Alzheimer's disease. Curr. Aging Sci. *7*, 168–175.

Rey, C., Mauduit, C., Naureils, O., Benahmed, M., Louisot, P., and Gasnier, F. (2000). Up-regulation of mitochondrial peripheral benzodiazepine receptor expression by tumor necrosis factor alpha in testicular leydig cells. Possible involvement in cell survival. Biochem. Pharmacol. *60*, 1639–1646.

Richards, E.M., Zanotti-Fregonara, P., Fujita, M., Newman, L., Farmer, C., Ballard, E.D., Machado-Vieira, R., Yuan, P., Niciu, M.J., Lyoo, C.H., et al. (2018). PET radioligand binding to translocator protein (TSPO) is increased in unmedicated depressed subjects. EJNMMI Res. *8*, 57.

Rojas, S., Martín, A., Arranz, M.J., Pareto, D., Purroy, J., Verdaguer, E., Llop, J., Gómez, V., Gispert, J.D., Millán, O., et al. (2007). Imaging brain inflammation with [(11)C]PK11195 by PET and induction of the peripheral-type benzodiazepine receptor after transient focal ischemia in rats. J. Cereb. Blood Flow Metab. *27*, 1975–1986.

Rone, M.B., Midzak, A.S., Issop, L., Rammouz, G., Jagannathan, S., Fan, J., Ye, X., Blonder, J., Veenstra, T., and Papadopoulos, V. (2012). Identification of a dynamic mitochondrial protein complex driving cholesterol import, trafficking, and metabolism to steroid hormones. Mol. Endocrinol. *26*, 1868–1882.

Rupprecht, R., Papadopoulos, V., Rammes, G., Baghai, T.C., Fan, J., Akula, N., Groyer, G., Adams, D., and Schumacher, M. (2010). Translocator protein (18 kDa) (TSPO) as a therapeutic target for neurological and psychiatric disorders. Nat. Rev. Drug Discov. *9*, 971–988.

Santidrián, A.F., Cosialls, A.M., Coll-Mulet, L., Iglesias-Serret, D., de Frias, M., González-Gironès, D.M., Campàs, C., Domingo, A., Pons, G., and Gil, J. (2007). The potential anticancer agent PK11195 induces apoptosis irrespective of p53 and ATM status in chronic lymphocytic leukemia cells. Haematologica *92*, 1631–1638.

Schaller, S., Paradis, S., Ngoh, G.A., Assaly, R., Buisson, B., Drouot, C., Ostuni, M.A., Lacapere, J.-J., Bassissi, F., Bordet, T., et al. (2010). TRO40303, a new cardioprotective compound, inhibits mitochondrial permeability transition. J. Pharmacol. Exp. Ther. *333*, 696–706.

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

Šileikytė, J., Blachly-Dyson, E., Sewell, R., Carpi, A., Menabò, R., Di Lisa, F., Ricchelli, F., Bernardi, P., and Forte, M. (2014). Regulation of the mitochondrial permeability transition pore by the outer

membrane does not involve the peripheral benzodiazepine receptor (Translocator Protein of 18 kDa (TSPO)). J. Biol. Chem. *289*, 13769–13781.

Snyder, S.H., Verma, A., and Trifiletti, R.R. (1987). The peripheral-type benzodiazepine receptor: a protein of mitochondrial outer membranes utilizing porphyrins as endogenous ligands. FASEB J. *1*, 282–288.

Su, Z., Roncaroli, F., Durrenberger, P.F., Coope, D.J., Karabatsou, K., Hinz, R., Thompson, G., Turkheimer, F.E., Janczar, K., Du Plessis, D., et al. (2015). The 18-kDa mitochondrial translocator protein in human gliomas: an 11C-(R)PK11195 PET imaging and neuropathology study. J. Nucl. Med. *56*, 512–517.

Sulistijo, E.S., and Mackenzie, K.R. (2009). Structural basis for dimerization of the BNIP3 transmembrane domain. Biochemistry *48*, 5106–5120.

Sutter, A.P., Maaser, K., Höpfner, M., Barthel, B., Grabowski, P., Faiss, S., Carayon, P., Zeitz, M., and Scherübl, H. (2002). Specific ligands of the peripheral benzodiazepine receptor induce apoptosis and cell cycle arrest in human esophageal cancer cells. Int. J. Cancer *102*, 318–327.

Sutter, A.P., Maaser, K., Barthel, B., and Scherübl, H. (2003). Ligands of the peripheral benzodiazepine receptor induce apoptosis and cell cycle arrest in oesophageal cancer cells: involvement of the p38MAPK signalling pathway. Br. J. Cancer *89*, 564–572.

Sutter, A.P., Maaser, K., Grabowski, P., Bradacs, G., Vormbrock, K., Höpfner, M., Krahn, A., Heine, B., Stein, H., Somasundaram, R., et al. (2004). Peripheral benzodiazepine receptor ligands induce apoptosis and cell cycle arrest in human hepatocellular carcinoma cells and enhance chemosensitivity to paclitaxel, docetaxel, doxorubicin and the Bcl-2 inhibitor HA14-1. J. Hepatol. *41*, 799–807.

Szabó, I., De Pinto, V., and Zoratti, M. (1993). The mitochondrial permeability transition pore may comprise VDAC molecules. II. The electrophysiological properties of VDAC are compatible with those of the mitochondrial megachannel. FEBS Lett. *330*, 206–210.

Taketani, S., Kohno, H., Furukawa, T., and Tokunaga, R. (1995). Involvement of peripheral-type benzodiazepine receptors in the intracellular transport of heme and porphyrins. J. Biochem. *117*, 875–880.

Tanimoto, Y., Onishi, Y., Sato, Y., and Kizaki, H. (1999). Benzodiazepine receptor agonists modulate thymocyte apoptosis through reduction of the mitochondrial transmembrane potential. Jpn J Pharmacol *79*, 177–183.

Teese, M.G., and Langosch, D. (2015). Role of gxxxg motifs in transmembrane domain interactions. Biochemistry *54*, 5125–5135.

Torres, S.R., Fröde, T.S., Nardi, G.M., Vita, N., Reeb, R., Ferrara, P., Ribeiro-do-Valle, R.M., and Farges, R.C. (2000). Anti-inflammatory effects of peripheral benzodiazepine receptor ligands in two mouse models of inflammation. Eur. J. Pharmacol. *408*, 199–211.

Tsujishita, Y., and Hurley, J.H. (2000). Structure and lipid transport mechanism of a StAR-related domain. Nat. Struct. Biol. *7*, 408–414.

Tu, L.N., Morohaku, K., Manna, P.R., Pelton, S.H., Butler, W.R., Stocco, D.M., and Selvaraj, V. (2014). Peripheral benzodiazepine receptor/translocator protein global knock-out mice are viable with no effects on steroid hormone biosynthesis. J. Biol. Chem. *289*, 27444–27454.

Tu, L.N., Zhao, A.H., Stocco, D.M., and Selvaraj, V. (2015). PK11195 effect on steroidogenesis is not mediated through the translocator protein (TSPO). Endocrinology *156*, 1033–1039.

Ujwal, R., Cascio, D., Colletier, J.-P., Faham, S., Zhang, J., Toro, L., Ping, P., and Abramson, J. (2008). The crystal structure of mouse VDAC1 at 2.3 A resolution reveals mechanistic insights into metabolite gating. Proc Natl Acad Sci USA *105*, 17742–17747.

Unterreitmeier, S., Fuchs, A., Schäffler, T., Heym, R.G., Frishman, D., and Langosch, D. (2007). Phenylalanine promotes interaction of transmembrane domains via GxxxG motifs. J. Mol. Biol. *374*, 705–718.

Vaitinadapoulé, A. (2015). TSPO and VDAC, two Membrane Proteins of Red Blood Cell involved in Malaria. A bioinformatics and Molecular Modelling Study. Université de la Réunion. Doctorate thesis

van Aalten, D.M.F., Milne K.G., Zou, J.Y., Kleywegt., G. J., Bergfors, T., Ferguson, M. A.J., Knudsen, J. and Jones T. A. (2001). Binding site differences revealed by crystal structures of Plasmodium falciparum, and Bovine Acyl-CoA Binding Protein. J. Mol. Biol. 309, 181-182.

Vanhee, C., Zapotoczny, G., Masquelier, D., Ghislain, M., and Batoko, H. (2011). The Arabidopsis multistress regulator TSPO is a heme binding membrane protein and a potential scavenger of porphyrins via an autophagy-dependent degradation mechanism. Plant Cell *23*, 785–805.

Veenman, L., and Gavish, M. (2012). The role of 18 kDa mitochondrial translocator protein (TSPO) in programmed cell death, and effects of steroids on TSPO expression. Curr. Mol. Med. *12*, 398–412.

Veenman, L., Papadopoulos, V., and Gavish, M. (2007). Channel-like functions of the 18-kDa translocator protein (TSPO): regulation of apoptosis and steroidogenesis as part of the host-defense response. Curr. Pharm. Des. *13*, 2385–2405.

Veenman, L., Shandalov, Y., and Gavish, M. (2008). VDAC activation by the 18 kDa translocator protein (TSPO), implications for apoptosis. J. Bioenerg. Biomembr. *40*, 199–205.

Veenman, L., Vainshtein, A., Yasin, N., Azrad, M., and Gavish, M. (2016). Tetrapyrroles as Endogenous TSPO Ligands in Eukaryotes and Prokaryotes: Comparisons with Synthetic Ligands. Int. J. Mol. Sci. *17*.

Verma, A., and Snyder, S.H. (1989). Peripheral type benzodiazepine receptors. Annu. Rev. Pharmacol. Toxicol. *29*, 307–322.

Verma, A., Nye, J.S., and Snyder, S.H. (1987). Porphyrins are endogenous ligands for the mitochondrial (peripheral-type) benzodiazepine receptor. Proc Natl Acad Sci USA *84*, 2256–2260.

Vowinckel, E., Reutens, D., Becher, B., Verge, G., Evans, A., Owens, T., and Antel, J.P. (1997). PK11195 binding to the peripheral benzodiazepine receptor as a marker of microglia activation in multiple sclerosis and experimental autoimmune encephalomyelitis. J. Neurosci. Res. *50*, 345–353.

Walter, R.B., Pirga, J.L., Cronk, M.R., Mayer, S., Appelbaum, F.R., and Banker, D.E. (2005). PK11195, a peripheral benzodiazepine receptor (pBR) ligand, broadly blocks drug efflux to chemosensitize leukemia and myeloma cells by a pBR-independent, direct transporter-modulating mechanism. Blood *106*, 3584–3593.

Wang, J. (2015). Comment on “Crystal structures of translocator protein (TSPO) and mutant mimic of a human polymorphism”. Science *350*, 519.

Weber, M., and Schneider, D. (2013). Six amino acids define a minimal dimerization sequence and

stabilize a transmembrane helix dimer by close packing and hydrogen bonding. FEBS Lett. *587*, 1592–1596.

Wendler, G., Lindemann, P., Lacapère, J.-J., and Papadopoulos, V. (2003). Protoporphyrin IX binding and transport by recombinant mouse PBR. Biochem. Biophys. Res. Commun. *311*, 847–852.

Werry, E.L., Barron, M.L., and Kassiou, M. (2015). TSPO as a target for glioblastoma therapeutics. Biochem. Soc. Trans. *43*, 531–536.

Wu, X., and Gallo, K.A. (2013). The 18-kDa translocator protein (TSPO) disrupts mammary epithelial morphogenesis and promotes breast cancer cell migration. PLoS ONE *8*, e71258.

Yeliseev, A.A., and Kaplan, S. (1995). A sensory transducer homologous to the mammalian peripheral-type benzodiazepine receptor regulates photosynthetic membrane complex formation in Rhodobacter sphaeroides 2.4.1. J. Biol. Chem. *270*, 21167–21175.

Yeliseev, A.A., and Kaplan, S. (1999). A novel mechanism for the regulation of photosynthesis gene expression by the TspO outer membrane protein of Rhodobacter sphaeroides 2.4.1. J. Biol. Chem. *274*, 21234–21243.

Yeliseev, A.A., and Kaplan, S. (2000). TspO of rhodobacter sphaeroides. A structural and functional model for the mammalian peripheral benzodiazepine receptor. J. Biol. Chem. *275*, 5657–5667.

Yeliseev, A.A., Krueger, K.E., and Kaplan, S. (1997). A mammalian mitochondrial drug receptor functions as a bacterial “oxygen” sensor. Proc Natl Acad Sci USA *94*, 5101–5106.

Zhao, A.H., Tu, L.N., Mukai, C., Sirivelu, M.P., Pillai, V.V., Morohaku, K., Cohen, R., and Selvaraj, V. (2016). Mitochondrial translocator protein (TSPO) function is not essential for heme biosynthesis. J. Biol. Chem. *291*, 1591–1603.

Zhao, Y.-Y., Yu, J.-Z., Li, Q.-Y., Ma, C.-G., Lu, C.-Z., and Xiao, B.-G. (2011). TSPO-specific ligand vinpocetine exerts a neuroprotective effect by suppressing microglial inflammation. Neuron Glia Biol. *7*, 187–197.

Zheng, J., Boisgard, R., Siquier-Pernet, K., Decaudin, D., Dollé, F., and Tavitian, B. (2011). Differential expression of the 18 kDa translocator protein (TSPO) by neoplastic and inflammatory cells in mouse tumors of breast cancer. Mol. Pharm. *8*, 823–832.

Zürcher, N.R., Loggia, M.L., Lawson, R., Chonde, D.B., Izquierdo-Garcia, D., Yasek, J.E., Akeju, O., Catana, C., Rosen, B.R., Cudkowicz, M.E., et al. (2015). Increased in vivo glial activation in patients with amyotrophic lateral sclerosis: assessed with [(11)C]-PBR28. Neuroimage Clin. *7*, 409–414.

Chapter-2:

Aliev, A.E., Kulke, M., Khaneja, H.S., Chudasama, V., Sheppard, T.D., and Lanigan, R.M. (2014). Motional timescale predictions by molecular dynamics simulations: case study using proline and hydroxyproline sidechain dynamics. Proteins *82*, 195–215.

Amadei, A., Linssen, A.B., and Berendsen, H.J. (1993). Essential dynamics of proteins. Proteins *17*, 412–425.

Ashkenazy, H., Penn, O., Doron-Faigenboim, A., Cohen, O., Cannarozzi, G., Zomer, O., and Pupko, T. (2012). FastML: a web server for probabilistic reconstruction of ancestral sequences. Nucleic Acids Res. *40*, W580-4.

Atilgan, A.R., Durell, S.R., Jernigan, R.L., Demirel, M.C., Keskin, O., and Bahar, I. (2001). Anisotropy of fluctuation dynamics of proteins with an elastic network model. Biophys. J. *80*, 505–515.

Bahar, I., Lezon, T.R., Bakan, A., and Shrivastava, I.H. (2010). Normal mode analysis of biomolecular structures: functional mechanisms of membrane proteins. Chem. Rev. *110*, 1463–1497.

Bayly, C.I., Cieplak, P., Cornell, W., and Kollman, P.A. (1993). A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges: the RESP model. J. Phys. Chem. *97*, 10269–10280.

Benson, E.E., and Linderstrom-Lang, K. (1959). Deuterium exchange between myoglobin and water. Biochim. Biophys. Acta *32*, 579–581.

Cornell, W.D., Cieplak, P., Bayly, C.I., Gould, I.R., Merz, K.M., Ferguson, D.M., Spellmeyer, D.C., Fox, T., Caldwell, J.W., and Kollman, P.A. (1995). A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules. J. Am. Chem. Soc. *117*, 5179–5197.

Cuendet, M.A., and Michielin, O. (2008). Protein-protein interaction investigated by steered molecular dynamics: the TCR-pMHC complex. Biophys. J. *95*, 3575–3590.

David, C.C., and Jacobs, D.J. (2014). Principal component analysis: a method for determining the essential dynamics of proteins. Methods Mol. Biol. *1084*, 193–226.

Dayhoff, M.O. (1972). A model of evolutionary change in proteins. Atlas of Protein Sequence and Structure *5*, 89–99.

Edgar, R.C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. *32*, 1792–1797.

Englander, S.W., Mayne, L., Bai, Y., and Sosnick, T.R. (1997). Hydrogen exchange: the modern legacy of Linderstrøm-Lang. Protein Sci. *6*, 1101–1109.

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

Fitch, W.M. (1977). On the problem of discovering the most parsimonious tree. Am. Nat. *111*, 223–257.

García, A.E., and Sanbonmatsu, K.Y. (2002). Alpha-helical stabilization by side chain shielding of backbone hydrogen bonds. Proc Natl Acad Sci USA *99*, 2782–2787.

Gavezzotti, A. (2007). Molecular aggregation: structure analysis and molecular simulation of crystals

and liquids (OUP Oxford).

Girvan, M., and Newman, M.E.J. (2002). Community structure in social and biological networks. Proc Natl Acad Sci USA *99*, 7821–7826.

González, M.A. (2011). Force fields and molecular dynamics simulations. JDN *12*, 169–200.

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

Guindon, S., Lethiec, F., Duroux, P., and Gascuel, O. (2005). PHYML Online--a web server for fast maximum likelihood-based phylogenetic inference. Nucleic Acids Res. *33*, W557-9.

Guindon, S., Dufayard, J.-F., Lefort, V., Anisimova, M., Hordijk, W., and Gascuel, O. (2010). New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Syst. Biol. *59*, 307–321.

Hagler, A.T., and Lifson, S. (1974). Energy functions for peptides and proteins. II. The amide hydrogen bond and calculation of amide crystal properties. J. Am. Chem. Soc. *96*, 5327–5335.

von Heijne, G. (1991). Proline kinks in transmembrane alpha-helices. J. Mol. Biol. *218*, 499–503.

Henikoff, S., and Henikoff, J.G. (1992). Amino acid substitution matrices from protein blocks. Proc Natl Acad Sci USA *89*, 10915–10919.

Hornak, V., Abel, R., Okur, A., Strockbine, B., Roitberg, A., and Simmerling, C. (2006). Comparison of multiple Amber force fields and development of improved protein backbone parameters. Proteins *65*, 712–725.

Hurwitz, N., Schneidman-Duhovny, D., and Wolfson, H.J. (2016). Memdock: an α-helical membrane protein docking algorithm. Bioinformatics *32*, 2444–2450.

Hvidt, A., and Linderstrøm-Lang, K. (1954). Exchange of hydrogen atoms in insulin with deuterium atoms in aqueous solutions. Biochim. Biophys. Acta *14*, 574–575.

Jämbeck, J.P.M., and Lyubartsev, A.P. (2012a). Derivation and systematic validation of a refined all-atom force field for phosphatidylcholine lipids. J. Phys. Chem. B *116*, 3164–3179.

Jämbeck, J.P.M., and Lyubartsev, A.P. (2012b). An Extension and Further Validation of an All-Atomistic Force Field for Biological Membranes. J. Chem. Theory Comput. *8*, 2938–2948.

Jones, D.T. (1999). Protein secondary structure prediction based on position-specific scoring matrices. J. Mol. Biol. *292*, 195–202.

de Jong, D.H., Singh, G., Bennett, W.F.D., Arnarez, C., Wassenaar, T.A., Schäfer, L.V., Periole, X., Tieleman, D.P., and Marrink, S.J. (2013). Improved parameters for the MARTINI coarse-grained protein force field. J. Chem. Theory Comput. *9*, 687–697.

Kabsch, W., and Sander, C. (1983). Secondary structure definition by the program DSSP. Biopolymers *22*, 2577–2637.

Kaczor, A.A., Selent, J., Sanz, F., and Pastor, M. (2013). Modeling complexes of transmembrane proteins: systematic analysis of protein–protein docking tools. Mol. Inform. *32*, 717–733.

Karplus, M., and Kuriyan, J. (2005). Molecular dynamics and protein function. Proc Natl Acad Sci USA *102*, 6679–6685.

Karplus, M., and McCammon, J.A. (2002). Molecular dynamics simulations of biomolecules. Nat. Struct. Biol. *9*, 646–652.

Katoh, K., Misawa, K., Kuma, K., and Miyata, T. (2002). MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Res. *30*, 3059–3066.

Kimura, M. (1979). The neutral theory of molecular evolution. Sci. Am. *241*, 98–100, 102, 108 passim.

Klauda, J.B., Venable, R.M., Freites, J.A., O'Connor, J.W., Tobias, D.J., Mondragon-Ramirez, C., Vorobyov, I., MacKerell, A.D., and Pastor, R.W. (2010). Update of the CHARMM all-atom additive force field for lipids: validation on six lipid types. J. Phys. Chem. B *114*, 7830–7843.

Kundu, S., Melton, J.S., Sorensen, D.C., and Phillips, G.N. (2002). Dynamics of proteins in crystals: comparison of experiment with simple models. Biophys. J. *83*, 723–732.

Linderstromlang, K. (1955). Deuterium exchange between peptides and water. Chemistry & Industry 503.

Lindorff-Larsen, K., Piana, S., Palmo, K., Maragakis, P., Klepeis, J.L., Dror, R.O., and Shaw, D.E. (2010). Improved side-chain torsion potentials for the Amber ff99SB protein force field. Proteins *78*, 1950–1958.

Liu, H., Mark, A.E., and van Gunsteren, W.F. (1996). Estimating the Relative Free Energy of Different Molecular States with Respect to a Single Reference State. J. Phys. Chem. *100*, 9485–9494.

Lomize, A.L., Pogozheva, I.D., Lomize, M.A., and Mosberg, H.I. (2006). Positioning of proteins in membranes: a computational approach. Protein Sci. *15*, 1318–1333.

Lomize, A.L., Pogozheva, I.D., and Mosberg, H.I. (2011). Anisotropic solvent model of the lipid bilayer. 2. Energetics of insertion of small molecules, peptides, and proteins in membranes. J. Chem. Inf. Model. *51*, 930–946.

Lomize, M.A., Pogozheva, I.D., Joo, H., Mosberg, H.I., and Lomize, A.L. (2012). OPM database and PPM web server: resources for positioning of proteins in membranes. Nucleic Acids Res. *40*, D370-6.

Marrink, S.J., and Tieleman, D.P. (2013). Perspective on the Martini model. Chem. Soc. Rev. *42*, 6801–6822.

Marrink, S.J., de Vries, A.H., and Mark, A.E. (2004). Coarse grained model for semiquantitative lipid simulations. J. Phys. Chem. B *108*, 750–760.

Marrink, S.J., Risselada, H.J., Yefimov, S., Tieleman, D.P., and de Vries, A.H. (2007). The MARTINI force field: coarse grained model for biomolecular simulations. J. Phys. Chem. B *111*, 7812–7824.

McCammon, J.A., Gelin, B.R., and Karplus, M. (1977). Dynamics of folded proteins. Nature *267*, 585–590.

Monticelli, L., Kandasamy, S.K., Periole, X., Larson, R.G., Tieleman, D.P., and Marrink, S.-J. (2008). The MARTINI coarse-grained force field: Extension to proteins. J. Chem. Theory Comput. *4*, 819–834.

Na, H., Ben-Avraham, D., and Tirion, M.M. (2018). Slow normal modes of proteins are accurately reproduced across different platforms. Phys. Biol. *16*, 016003.

Okur, A., Strockbine, B., Hornak, V., and Simmerling, C. (2003). Using PC clusters to evaluate the

transferability of molecular mechanics force fields for proteins. J. Comput. Chem. *24*, 21–31.

Papaleo, E., Mereghetti, P., Fantucci, P., Grandori, R., and De Gioia, L. (2009). Free-energy landscape, principal component analysis, and structural clustering to identify representative conformations from molecular dynamics simulations: the myoglobin case. J. Mol. Graph. Model. *27*, 889–899.

Pei, J., and Grishin, N.V. (2014). PROMALS3D: multiple protein sequence alignment enhanced with evolutionary and three-dimensional structural information. Methods Mol. Biol. *1079*, 263–271.

Pei, J., Kim, B.-H., and Grishin, N.V. (2008). PROMALS3D: a tool for multiple protein sequence and structure alignments. Nucleic Acids Res. *36*, 2295–2300.

Ponder, J.W., and Case, D.A. (2003). Force fields for protein simulations. Adv. Protein Chem. *66*, 27–85.

Rakers, C., Bermudez, M., Keller, B.G., Mortier, J., and Wolber, G. (2015). Computational close up on protein-protein interactions: how to unravel the invisible using molecular dynamics simulations? WIREs Comput Mol Sci *5*, 345–359.

Sang, P., Yang, L.-Q., Ji, X.-L., Fu, Y.-X., and Liu, S.-Q. (2014). Insight derived from molecular dynamics simulations into molecular motions, thermodynamics and kinetics of HIV-1 gp120. PLoS ONE *9*, e104714.

Schneidman-Duhovny, D., Inbar, Y., Nussinov, R., and Wolfson, H.J. (2005). PatchDock and SymmDock: servers for rigid and symmetric docking. Nucleic Acids Res. *33*, W363-7.

Serçinoglu, O., and Ozbek, P. (2018). gRINN: a tool for calculation of residue interaction energies and protein energy network analysis of molecular dynamics simulations. Nucleic Acids Res. *46*, W554–W562.

Showalter, S.A., and Brüschweiler, R. (2007). Validation of molecular dynamics simulations of biomolecules using NMR spin relaxation as benchmarks: application to the AMBER99SB force field. J. Chem. Theory Comput. *3*, 961–975.

Thompson, J.D., Gibson, T.J., and Higgins, D.G. (2002). Multiple sequence alignment using ClustalW and ClustalX. Curr. Protoc. Bioinformatics *Chapter 2*, Unit 2.3.

Tirion, M.M. (1996). Large Amplitude Elastic Motions in Proteins from a Single-Parameter, Atomic Analysis. Phys. Rev. Lett. *77*, 1905–1908.

Tovchigrechko, A., and Vakser, I.A. (2006). GRAMM-X public web server for protein-protein docking. Nucleic Acids Res. *34*, W310-4.

van der Spoel, D., Lindahl, E., Hess, B., GROMACS development team. (2014). GROMACS User Manual version 4.6.7, www.gromacs.org.

Vishveshwara, S., Brinda, K.V., and Kannan, N. (2002). Protein Structure: Insights from graph theory. J. Theor. Comput. Chem. *01*, 187–211.

Vishveshwara, S., Ghosh, A., and Hansia, P. (2009). Intra and inter-molecular communications through protein structure network. Curr. Protein Pept. Sci. *10*, 146–160.

Wang, J., Cieplak, P., and Kollman, P.A. (2000). How well does a restrained electrostatic potential (RESP) model perform in calculating conformational energies of organic and biological molecules? J.

Comput. Chem. *21*, 1049–1074.

Weiner, S.J., Kollman, P.A., Case, D.A., Singh, U.C., Ghio, C., Alagona, G., Profeta, S., and Weiner, P. (1984). A new force field for molecular mechanical simulation of nucleic acids and proteins. J. Am. Chem. Soc. *106*, 765–784.

Yang, L.-W., Eyal, E., Chennubhotla, C., Jee, J., Gronenborn, A.M., and Bahar, I. (2007). Insights into equilibrium dynamics of proteins from comparison of NMR and X-ray data with computational predictions. Structure *15*, 741–749.

Yesylevskyy, S.O., Schäfer, L.V., Sengupta, D., and Marrink, S.J. (2010). Polarizable water model for the coarse-grained MARTINI force field. PLoS Comput. Biol. *6*, e1000810.

Chapter-3:

Ashkenazy, H., Abadi, S., Martz, E., Chay, O., Mayrose, I., Pupko, T., and Ben-Tal, N. (2016). ConSurf 2016: an improved methodology to estimate and visualize evolutionary conservation in macromolecules. Nucleic Acids Res. *44*, W344-50.

Azarashvili, T., Baburina, Y., Grachev, D., Krestinina, O., Papadopoulos, V., Lemasters, J.J., Odinokova, I., and Reiser, G. (2014). Carbenoxolone induces permeability transition pore opening in rat mitochondria via the translocator protein TSPO and connexin43. Arch. Biochem. Biophys. *558*, 87–94.

Bakan, A., Meireles, L.M., and Bahar, I. (2011). ProDy: protein dynamics inferred from theory and experiments. Bioinformatics *27*, 1575–1577.

Barnea, E.R., Fares, F., and Gavish, M. (1989). Modulatory action of benzodiazepines on human term placental steroidogenesis in vitro. Mol. Cell. Endocrinol. *64*, 155–159.

Berendsen, H.J.C., van der Spoel, D., and van Drunen, R. (1995). GROMACS: A message-passing parallel molecular dynamics implementation. Comput. Phys. Commun. *91*, 43–56.

Bose, H.S., Lingappa, V.R., and Miller, W.L. (2002). The steroidogenic acute regulatory protein, StAR, works only at the outer mitochondrial membrane. Endocr. Res. *28*, 295–308.

Bouyer, G., Cueff, A., Egée, S., Kmiecik, J., Maksimova, Y., Glogowska, E., Gallagher, P.G., and Thomas, S.L.Y. (2011). Erythrocyte peripheral type benzodiazepine receptor/voltage-dependent anion channels are upregulated by Plasmodium falciparum. Blood *118*, 2305–2312.

Calogero, A.E., Kamilaris, T.C., Bernardini, R., Johnson, E.O., Chrousos, G.P., and Gold, P.W. (1990). Effects of peripheral benzodiazepine receptor ligands on hypothalamic-pituitary-adrenal axis function in the rat. J. Pharmacol. Exp. Ther. *253*, 729–737.

Costa, B., Pini, S., Gabelloni, P., Da Pozzo, E., Abelli, M., Lari, L., Preve, M., Lucacchini, A., Cassano, G.B., and Martini, C. (2009). The spontaneous Ala147Thr amino acid substitution within the translocator protein influences pregnenolone production in lymphomonocytes of healthy individuals. Endocrinology *150*, 5438–5445.

DeLano, W.L. (2002). Pymol: An open-source molecular graphics tool. CCP4 Newsletter On Protein Crystallography *40*, 82–92.

Delavoie, F., Li, H., Hardwick, M., Robert, J.-C., Giatzakis, C., Péranzi, G., Yao, Z.-X., Maccario, J., Lacapère, J.-J., and Papadopoulos, V. (2003). In vivo and in vitro peripheral-type benzodiazepine receptor polymerization: functional significance in drug ligand and cholesterol binding. Biochemistry *42*, 4506–4519.

Di Scala, C., Baier, C.J., Evans, L.S., Williamson, P.T.F., Fantini, J., and Barrantes, F.J. (2017). Relevance of CARC and CRAC Cholesterol-Recognition Motifs in the Nicotinic Acetylcholine Receptor and Other Membrane-Bound Receptors. Curr. Top. Membr. *80*, 3–23.

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

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

Galiègue, S., Jbilo, O., Combes, T., Bribes, E., Carayon, P., Le Fur, G., and Casellas, P. (1999).

Cloning and characterization of PRAX-1. A new protein that specifically interacts with the peripheral benzodiazepine receptor. J. Biol. Chem. *274*, 2938–2952.

Gavish, M., and Veenman, L. (2018). Regulation of mitochondrial, cellular, and organismal functions by TSPO. Adv. Pharmacol. *82*, 103–136.

Ginter, C., Kiburu, I., and Boudker, O. (2013). Chemical catalysis by the translocator protein (18 kDa). Biochemistry *52*, 3609–3611.

Glaser, F., Pupko, T., Paz, I., Bell, R.E., Bechor-Shental, D., Martz, E., and Ben-Tal, N. (2003). ConSurf: identification of functional regions in proteins by surface-mapping of phylogenetic information. Bioinformatics *19*, 163–164.

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

Guo, Y., Kalathur, R.C., Liu, Q., Kloss, B., Bruni, R., Ginter, C., Kloppmann, E., Rost, B., and Hendrickson, W.A. (2015). Protein structure. Structure and activity of tryptophan-rich TSPO proteins. Science *347*, 551–555.

Hacisuleyman, A., and Erman, B. (2017). Entropy Transfer between Residue Pairs and Allostery in Proteins: Quantifying Allosteric Communication in Ubiquitin. PLoS Comput. Biol. *13*, e1005319.

Hardwick, M., Fertikh, D., Culty, M., Li, H., Vidic, B., and Papadopoulos, V. (1999). 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. *59*, 831–842.

von Heijne, G. (1991). Proline kinks in transmembrane alpha-helices. J. Mol. Biol. *218*, 499–503.

Hirsch, J.D., Beyer, C.F., Malkowitz, L., Beer, B., and Blume, A.J. (1989). Mitochondrial benzodiazepine receptors mediate inhibition of mitochondrial respiratory control. Mol. Pharmacol. *35*, 157–163.

Jamin, N., Neumann, J.-M., Ostuni, M.A., Vu, T.K.N., Yao, Z.-X., Murail, S., Robert, J.-C., Giatzakis, C., Papadopoulos, V., and Lacapère, J.-J. (2005). Characterization of the cholesterol recognition amino acid consensus sequence of the peripheral-type benzodiazepine receptor. Mol. Endocrinol. *19*, 588–594.

Jaremko, L., Jaremko, M., Giller, K., Becker, S., and Zweckstetter, M. (2014). Structure of the mitochondrial translocator protein in complex with a diagnostic ligand. Science *343*, 1363–1366.

Jaremko, M., Jaremko, Ł., Jaipuria, G., Becker, S., and Zweckstetter, M. (2015a). Structure of the mammalian TSPO/PBR protein. Biochem. Soc. Trans. *43*, 566–571.

Jaremko, Ł., Jaremko, M., Giller, K., Becker, S., and Zweckstetter, M. (2015b). Conformational flexibility in the transmembrane protein TSPO. Chem. Eur. J *21*, 16555–16563.

Jayakumar, A.R., Panickar, K.S., and Norenberg, M.D. (2002). Effects on free radical generation by ligands of the peripheral benzodiazepine receptor in cultured neural cells. J. Neurochem. *83*, 1226–1234.

de Jong, D.H., Singh, G., Bennett, W.F.D., Arnarez, C., Wassenaar, T.A., Schäfer, L.V., Periole, X., Tieleman, D.P., and Marrink, S.J. (2013). Improved parameters for the MARTINI coarse-grained protein force field. J. Chem. Theory Comput. *9*, 687–697.

Kabsch, W., and Sander, C. (1983). Secondary structure definition by the program DSSP. Biopolymers *22*, 2577–2637.

Karplus, M., and Kuriyan, J. (2005). Molecular dynamics and protein function. Proc Natl Acad Sci USA *102*, 6679–6685.

Karplus, M., and McCammon, J.A. (2002). Molecular dynamics simulations of biomolecules. Nat. Struct. Biol. *9*, 646–652.

Krestinina, O.V., Grachev, D.E., Odinokova, I.V., Reiser, G., Evtodienko, Y.V., and Azarashvili, T.S. (2009). Effect of peripheral benzodiazepine receptor (PBR/TSPO) ligands on opening of Ca2+-induced pore and phosphorylation of 3.5-kDa polypeptide in rat brain mitochondria. Biochemistry Mosc *74*, 421–429.

Kugler, W., Veenman, L., Shandalov, Y., Leschiner, S., Spanier, I., Lakomek, M., and Gavish, M. (2008). Ligands of the mitochondrial 18 kDa translocator protein attenuate apoptosis of human glioblastoma cells exposed to erucylphosphohomocholine. Cell. Oncol. *30*, 435–450.

Lacapère, J.J., Delavoie, F., Li, H., Péranzi, G., Maccario, J., Papadopoulos, V., and Vidic, B. (2001). Structural and functional study of reconstituted peripheral benzodiazepine receptor. Biochem. Biophys. Res. Commun. *284*, 536–541.

Lee, Y.-S., Siméon, F.G., Briard, E., and Pike, V.W. (2012). Solution structures of the prototypical 18 kDa translocator protein ligand, PK 11195, elucidated with 1H/13C NMR spectroscopy and quantum chemistry. ACS Chem. Neurosci. *3*, 325–335.

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

Li, F., Liu, J., Valls, L., Hiser, C., and Ferguson-Miller, S. (2015a). Identification of a key cholesterol binding enhancement motif in translocator protein 18 kDa. Biochemistry *54*, 1441–1443.

Li, F., Liu, J., Zheng, Y., Garavito, R.M., and Ferguson-Miller, S. (2015b). Protein structure. Crystal structures of translocator protein (TSPO) and mutant mimic of a human polymorphism. Science *347*, 555–558.

Li, F., Liu, J., Garavito, R.M., and Ferguson-Miller, S. (2015c). Evolving understanding of translocator protein 18 kDa (TSPO). Pharmacol. Res. *99*, 404–409.

Li, H., Yao, Z., Degenhardt, B., Teper, G., and Papadopoulos, V. (2001a). 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 *98*, 1267–1272.

Li, H., Degenhardt, B., Tobin, D., Yao, Z.X., Tasken, K., and Papadopoulos, V. (2001b). Identification, localization, and function in steroidogenesis of PAP7: a peripheral-type benzodiazepine receptor- and PKA (RIalpha)-associated protein. Mol. Endocrinol. *15*, 2211–2228.

Liu, G.-J., Middleton, R.J., Kam, W.W.-Y., Chin, D.Y., Hatty, C.R., Chan, R.H.Y., and Banati, R.B. (2017). Functional gains in energy and cell metabolism after TSPO gene insertion. Cell Cycle *16*, 436–447.

Liu, J., Rone, M.B., and Papadopoulos, V. (2006). Protein-protein interactions mediate mitochondrial

cholesterol transport and steroid biosynthesis. J. Biol. Chem. *281*, 38879–38893.

Lomize, A.L., Pogozheva, I.D., Lomize, M.A., and Mosberg, H.I. (2006). Positioning of proteins in membranes: a computational approach. Protein Sci. *15*, 1318–1333.

Lomize, A.L., Pogozheva, I.D., and Mosberg, H.I. (2011). Anisotropic solvent model of the lipid bilayer. 2. Energetics of insertion of small molecules, peptides, and proteins in membranes. J. Chem. Inf. Model. *51*, 930–946.

Lomize, M.A., Pogozheva, I.D., Joo, H., Mosberg, H.I., and Lomize, A.L. (2012). OPM database and PPM web server: resources for positioning of proteins in membranes. Nucleic Acids Res. *40*, D370-6.

Marginedas-Freixa, I., Hattab, C., Bouyer, G., Halle, F., Chene, A., Lefevre, S.D., Cambot, M., Cueff, A., Schmitt, M., Gamain, B., et al. (2016). TSPO ligands stimulate ZnPPIX transport and ROS accumulation leading to the inhibition of P. falciparum growth in human blood. Sci. Rep. *6*, 33516.

Marrink, S.J., Risselada, H.J., Yefimov, S., Tieleman, D.P., and de Vries, A.H. (2007). The MARTINI force field: coarse grained model for biomolecular simulations. J. Phys. Chem. B *111*, 7812–7824.

Mathieu, A.P., Fleury, A., Ducharme, L., Lavigne, P., and LeHoux, J.G. (2002). Insights into steroidogenic acute regulatory protein (StAR)-dependent cholesterol transfer in mitochondria: evidence from molecular modeling and structure-based thermodynamics supporting the existence of partially unfolded states of StAR. J. Mol. Endocrinol. *29*, 327–345.

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

McEnery, M.W., Snowman, A.M., Trifiletti, R.R., and Snyder, S.H. (1992). Isolation of the mitochondrial benzodiazepine receptor: association with the voltage-dependent anion channel and the adenine nucleotide carrier. Proc Natl Acad Sci USA *89*, 3170–3174.

Mendonça-Torres, M.C., and Roberts, S.S. (2013). The translocator protein (TSPO) ligand PK11195 induces apoptosis and cell cycle arrest and sensitizes to chemotherapy treatment in pre- and post-relapse neuroblastoma cell lines. Cancer Biol. Ther. *14*, 319–326.

Mezei, M. (2010). Simulaid: a simulation facilitator and analysis program. J. Comput. Chem. *31*, 2658–2668.

Mezei, M., and Filizola, M. (2006). TRAJELIX: a computational tool for the geometric characterization of protein helices during molecular dynamics simulations. J. Comput. Aided Mol. Des. *20*, 97–107.

Monticelli, L., Kandasamy, S.K., Periole, X., Larson, R.G., Tieleman, D.P., and Marrink, S.-J. (2008). The MARTINI coarse-grained force field: Extension to proteins. J. Chem. Theory Comput. *4*, 819–834.

Morin, D., Musman, J., Pons, S., Berdeaux, A., and Ghaleh, B. (2016). Mitochondrial translocator protein (TSPO): From physiology to cardioprotection. Biochem. Pharmacol. *105*, 1–13.

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

Mukhin, A.G., Papadopoulos, V., Costa, E., and Krueger, K.E. (1989). Mitochondrial benzodiazepine

receptors regulate steroid biosynthesis. Proc Natl Acad Sci USA *86*, 9813–9816.

Murail, S., Robert, J.-C., Coïc, Y.-M., Neumann, J.-M., Ostuni, M.A., Yao, Z.-X., Papadopoulos, V., Jamin, N., and Lacapère, J.-J. (2008). Secondary and tertiary structures of the transmembrane domains of the translocator protein TSPO determined by NMR. Stabilization of the TSPO tertiary fold upon ligand binding. Biochim. Biophys. Acta *1778*, 1375–1381.

Olson, J.M., McNeel, W., Young, A.B., and Mancini, W.R. (1992). Localization of the peripheral-type benzodiazepine binding site to mitochondria of human glioma cells. J Neurooncol *13*, 35–42.

Páll, S., Abraham, M.J., Kutzner, C., Hess, B., and Lindahl, E. (2015). Tackling Exascale Software Challenges in Molecular Dynamics Simulations with GROMACS. In Solving Software Challenges for Exascale, S. Markidis, and E. Laure, eds. (Cham: Springer International Publishing), pp. 3–27.

Pan, J., Tristram-Nagle, S., Kucerka, N., and Nagle, J.F. (2008). Temperature dependence of structure, bending rigidity, and bilayer interactions of dioleoylphosphatidylcholine bilayers. Biophys. J. *94*, 117–124.

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

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

Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C., and Ferrin, T.E. (2004). UCSF Chimera—a visualization system for exploratory research and analysis. J. Comput. Chem. *25*, 1605–1612.

Rampon, C., Bouzaffour, M., Ostuni, M.A., Dufourcq, P., Girard, C., Freyssinet, J.M., Lacapere, J.-J., Schweizer-Groyer, G., and Vriz, S. (2009). Translocator protein (18 kDa) is involved in primitive erythropoiesis in zebrafish. FASEB J. *23*, 4181–4192.

Rone, M.B., Midzak, A.S., Issop, L., Rammouz, G., Jagannathan, S., Fan, J., Ye, X., Blonder, J., Veenstra, T., and Papadopoulos, V. (2012). Identification of a dynamic mitochondrial protein complex driving cholesterol import, trafficking, and metabolism to steroid hormones. Mol. Endocrinol. *26*, 1868–1882.

Rupprecht, R., Papadopoulos, V., Rammes, G., Baghai, T.C., Fan, J., Akula, N., Groyer, G., Adams, D., and Schumacher, M. (2010). Translocator protein (18 kDa) (TSPO) as a therapeutic target for neurological and psychiatric disorders. Nat. Rev. Drug Discov. *9*, 971–988.

Santidrián, A.F., Cosialls, A.M., Coll-Mulet, L., Iglesias-Serret, D., de Frias, M., González-Gironès, D.M., Campàs, C., Domingo, A., Pons, G., and Gil, J. (2007). The potential anticancer agent PK11195 induces apoptosis irrespective of p53 and ATM status in chronic lymphocytic leukemia cells. Haematologica *92*, 1631–1638.

Schreiber, T. (2000). Measuring information transfer. Phys. Rev. Lett. *85*, 461–464.

Tieleman, D.P., Shrivastava, I.H., Ulmschneider, M.R., and Sansom, M.S. (2001). Proline-induced hinges in transmembrane helices: possible roles in ion channel gating. Proteins *44*, 63–72.

Tsujishita, Y., and Hurley, J.H. (2000). Structure and lipid transport mechanism of a StAR-related domain. Nat. Struct. Biol. *7*, 408–414.

Tu, L.N., Morohaku, K., Manna, P.R., Pelton, S.H., Butler, W.R., Stocco, D.M., and Selvaraj, V.

(2014). Peripheral benzodiazepine receptor/translocator protein global knock-out mice are viable with no effects on steroid hormone biosynthesis. J. Biol. Chem. *289*, 27444–27454.

Vanhee, C., Zapotoczny, G., Masquelier, D., Ghislain, M., and Batoko, H. (2011). The Arabidopsis multistress regulator TSPO is a heme binding membrane protein and a potential scavenger of porphyrins via an autophagy-dependent degradation mechanism. Plant Cell *23*, 785–805.

Veenman, L., Papadopoulos, V., and Gavish, M. (2007). Channel-like functions of the 18-kDa translocator protein (TSPO): regulation of apoptosis and steroidogenesis as part of the host-defense response. Curr. Pharm. Des. *13*, 2385–2405.

Veenman, L., Shandalov, Y., and Gavish, M. (2008). VDAC activation by the 18 kDa translocator protein (TSPO), implications for apoptosis. J. Bioenerg. Biomembr. *40*, 199–205.

Veenman, L., Vainshtein, A., Yasin, N., Azrad, M., and Gavish, M. (2016). Tetrapyrroles as Endogenous TSPO Ligands in Eukaryotes and Prokaryotes: Comparisons with Synthetic Ligands. Int. J. Mol. Sci. *17*.

Verma, A., Nye, J.S., and Snyder, S.H. (1987). Porphyrins are endogenous ligands for the mitochondrial (peripheral-type) benzodiazepine receptor. Proc Natl Acad Sci USA *84*, 2256–2260.

Wallace, A.C., Laskowski, R.A., and Thornton, J.M. (1995). LIGPLOT: a program to generate schematic diagrams of protein-ligand interactions. Protein Eng. *8*, 127–134.

Wassenaar, T.A., Pluhackova, K., Böckmann, R.A., Marrink, S.J., and Tieleman, D.P. (2014). Going Backward: A Flexible Geometric Approach to Reverse Transformation from Coarse Grained to Atomistic Models. J. Chem. Theory Comput. *10*, 676–690.

Wendler, G., Lindemann, P., Lacapère, J.-J., and Papadopoulos, V. (2003). Protoporphyrin IX binding and transport by recombinant mouse PBR. Biochem. Biophys. Res. Commun. *311*, 847–852.

West, L.A., Horvat, R.D., Roess, D.A., Barisas, B.G., Juengel, J.L., and Niswender, G.D. (2001). Steroidogenic acute regulatory protein and peripheral-type benzodiazepine receptor associate at the mitochondrial membrane. Endocrinology *142*, 502–505.

Yeliseev, A.A., and Kaplan, S. (1999). A novel mechanism for the regulation of photosynthesis gene expression by the TspO outer membrane protein of Rhodobacter sphaeroides 2.4.1. J. Biol. Chem. *274*, 21234–21243.

Zhao, A.H., Tu, L.N., Mukai, C., Sirivelu, M.P., Pillai, V.V., Morohaku, K., Cohen, R., and Selvaraj, V. (2016). Mitochondrial translocator protein (TSPO) function is not essential for heme biosynthesis. J. Biol. Chem. *291*, 1591–1603.

Chapter-4:

Berendsen, H.J.C., van der Spoel, D., and van Drunen, R. (1995). GROMACS: A message-passing parallel molecular dynamics implementation. Comput. Phys. Commun. *91*, 43–56.

Bernassau, J.M., Reversat, J.L., Ferrara, P., Caput, D., and Lefur, G. (1993). A 3D model of the peripheral benzodiazepine receptor and its implication in intra mitochondrial cholesterol transport. J. Mol. Graph. *11*, 236–244, 235.

Bonsack, F., Alleyne, C.H., and Sukumari-Ramesh, S. (2016). Augmented expression of TSPO after intracerebral hemorrhage: a role in inflammation? J. Neuroinflammation *13*, 151.

Bouyer, G., Cueff, A., Egée, S., Kmiecik, J., Maksimova, Y., Glogowska, E., Gallagher, P.G., and Thomas, S.L.Y. (2011). Erythrocyte peripheral type benzodiazepine receptor/voltage-dependent anion channels are upregulated by Plasmodium falciparum. Blood *118*, 2305–2312.

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

Chechneva, O.V., and Deng, W. (2016). Mitochondrial translocator protein (TSPO), astrocytes and neuroinflammation. Neural Regen. Res. *11*, 1056–1057.

Delavoie, F., Li, H., Hardwick, M., Robert, J.-C., Giatzakis, C., Péranzi, G., Yao, Z.-X., Maccario, J., Lacapère, J.-J., and Papadopoulos, V. (2003). In vivo and in vitro peripheral-type benzodiazepine receptor polymerization: functional significance in drug ligand and cholesterol binding. Biochemistry *42*, 4506–4519.

Duneau, J.-P., Vegh, A.P., and Sturgis, J.N. (2007). A dimerization hierarchy in the transmembrane domains of the HER receptor family. Biochemistry *46*, 2010–2019.

Eilers, M., Patel, A.B., Liu, W., and Smith, S.O. (2002). Comparison of helix interactions in membrane and soluble alpha-bundle proteins. Biophys. J. *82*, 2720–2736.

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

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

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

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

Guo, Y., Kalathur, R.C., Liu, Q., Kloss, B., Bruni, R., Ginter, C., Kloppmann, E., Rost, B., and Hendrickson, W.A. (2015). Protein structure. Structure and activity of tryptophan-rich TSPO proteins. Science *347*, 551–555.

Hardwick, M., Fertikh, D., Culty, M., Li, H., Vidic, B., and Papadopoulos, V. (1999). 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. *59*, 831–842.

Hinsen, K., Vaitinadapoule, A., Ostuni, M.A., Etchebest, C., and Lacapere, J.-J. (2015). Construction and validation of an atomic model for bacterial TSPO from electron microscopy density, evolutionary constraints, and biochemical and biophysical data. Biochim. Biophys. Acta *1848*, 568–580.

Hubbard, S.J., and Naccess, T. (1993). Computer program. Department of Biochemistry and Molecular Biology, University College, London.

Hurwitz, N., Schneidman-Duhovny, D., and Wolfson, H.J. (2016). Memdock: an α-helical membrane protein docking algorithm. Bioinformatics *32*, 2444–2450.

Jaipuria, G., Leonov, A., Giller, K., Vasa, S.K., Jaremko, Ł., Jaremko, M., Linser, R., Becker, S., and Zweckstetter, M. (2017). Cholesterol-mediated allosteric regulation of the mitochondrial translocator protein structure. Nat. Commun. *8*, 14893.

Jaremko, L., Jaremko, M., Becker, S., and Zweckstetter, M. (2014a). Toward the functional oligomerization state of tryptophan-rich sensory proteins. Protein Sci. *23*, 1154–1160.

Jaremko, L., Jaremko, M., Giller, K., Becker, S., and Zweckstetter, M. (2014b). Structure of the mitochondrial translocator protein in complex with a diagnostic ligand. Science *343*, 1363–1366.

Jaremko, M., Jaremko, Ł., Jaipuria, G., Becker, S., and Zweckstetter, M. (2015). Structure of the mammalian TSPO/PBR protein. Biochem. Soc. Trans. *43*, 566–571.

Javanainen, M., Martinez-Seara, H., and Vattulainen, I. (2017). Excessive aggregation of membrane proteins in the Martini model. PLoS ONE *12*, e0187936.

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

Kleiger, G., Grothe, R., Mallick, P., and Eisenberg, D. (2002). GXXXG and AXXXA: common alpha-helical interaction motifs in proteins, particularly in extremophiles. Biochemistry *41*, 5990–5997.

Korkhov, V.M., Sachse, C., Short, J.M., and Tate, C.G. (2010). Three-dimensional structure of TspO by electron cryomicroscopy of helical crystals. Structure *18*, 677–687.

Krueger, K.E., and Papadopoulos, V. (1990). Peripheral-type benzodiazepine receptors mediate translocation of cholesterol from outer to inner mitochondrial membranes in adrenocortical cells. J. Biol. Chem. *265*, 15015–15022.

Kugler, W., Veenman, L., Shandalov, Y., Leschiner, S., Spanier, I., Lakomek, M., and Gavish, M. (2008). Ligands of the mitochondrial 18 kDa translocator protein attenuate apoptosis of human glioblastoma cells exposed to erucylphosphohomocholine. Cell. Oncol. *30*, 435–450.

Lacapère, J.J., Delavoie, F., Li, H., Péranzi, G., Maccario, J., Papadopoulos, V., and Vidic, B. (2001). Structural and functional study of reconstituted peripheral benzodiazepine receptor. Biochem. Biophys. Res. Commun. *284*, 536–541.

Lemmon, M.A., Flanagan, J.M., Treutlein, H.R., Zhang, J., and Engelman, D.M. (1992a). Sequence specificity in the dimerization of transmembrane alpha-helices. Biochemistry *31*, 12719–12725.

Lemmon, M.A., Flanagan, J.M., Hunt, J.F., Adair, B.D., Bormann, B.J., Dempsey, C.E., and Engelman, D.M. (1992b). Glycophorin A dimerization is driven by specific interactions between transmembrane alpha-helices. J. Biol. Chem. *267*, 7683–7689.

Li, H., and Papadopoulos, V. (1998). Peripheral-type benzodiazepine receptor function in cholesterol

transport. Identification of a putative cholesterol recognition/interaction amino acid sequence and consensus pattern. Endocrinology *139*, 4991–4997.

Li, E., Wimley, W.C., and Hristova, K. (2012). Transmembrane helix dimerization: beyond the search for sequence motifs. Biochim. Biophys. Acta *1818*, 183–193.

Li, F., Liu, J., Zheng, Y., Garavito, R.M., and Ferguson-Miller, S. (2015). Protein structure. Crystal structures of translocator protein (TSPO) and mutant mimic of a human polymorphism. Science *347*, 555–558.

Li, H., Yao, Z., Degenhardt, B., Teper, G., and Papadopoulos, V. (2001). 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 *98*, 1267–1272.

Liu, G.-J., Middleton, R.J., Hatty, C.R., Kam, W.W.-Y., Chan, R., Pham, T., Harrison-Brown, M., Dodson, E., Veale, K., and Banati, R.B. (2014). The 18 kDa translocator protein, microglia and neuroinflammation. Brain Pathol. *24*, 631–653.

Maaser, K., Höpfner, M., Jansen, A., Weisinger, G., Gavish, M., Kozikowski, A.P., Weizman, A., Carayon, P., Riecken, E.O., Zeitz, M., et al. (2001). Specific ligands of the peripheral benzodiazepine receptor induce apoptosis and cell cycle arrest in human colorectal cancer cells. Br. J. Cancer *85*, 1771–1780.

Machado, M.R., Barrera, E.E., Klein, F., Sóñora, M., Silva, S., and Pantano, S. (2019). The SIRAH force field 2.0: Altius, Fortius, Citius. J. Chem. Theory Comput.

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

Mercadante, D., Gräter, F., and Daday, C. (2018). CONAN: A Tool to Decode Dynamical Information from Molecular Interaction Maps. Biophys. J. *114*, 1267–1273.

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

Mueller, B.K., Subramaniam, S., and Senes, A. (2014). A frequent, GxxxG-mediated, transmembrane association motif is optimized for the formation of interhelical Cα-H hydrogen bonds. Proc Natl Acad Sci USA *111*, E888-95.

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

Páll, S., Abraham, M.J., Kutzner, C., Hess, B., and Lindahl, E. (2015). Tackling Exascale Software Challenges in Molecular Dynamics Simulations with GROMACS. In Solving Software Challenges for Exascale, S. Markidis, and E. Laure, eds. (Cham: Springer International Publishing), pp. 3–27.

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

Papadopoulos, V., Boujrad, N., Ikonomovic, M.D., Ferrara, P., and Vidic, B. (1994). Topography of

the Leydig cell mitochondrial peripheral-type benzodiazepine receptor. Mol. Cell. Endocrinol. *104*, R5-9.

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

Papadopoulos, V., Amri, H., Boujrad, N., Cascio, C., Culty, M., Garnier, M., Hardwick, M., Li, H., Vidic, B., Brown, A.S., et al. (1997b). Peripheral benzodiazepine receptor in cholesterol transport and steroidogenesis. Steroids *62*, 21–28.

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

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

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

Pogoryelov, D., Klyszejko, A.L., Krasnoselska, G.O., Heller, E.-M., Leone, V., Langer, J.D., Vonck, J., Müller, D.J., Faraldo-Gómez, J.D., and Meier, T. (2012). Engineering rotor ring stoichiometries in the ATP synthase. Proc Natl Acad Sci USA *109*, E1599-608.

Rampon, C., Bouzaffour, M., Ostuni, M.A., Dufourcq, P., Girard, C., Freyssinet, J.M., Lacapere, J.-J., Schweizer-Groyer, G., and Vriz, S. (2009). Translocator protein (18 kDa) is involved in primitive erythropoiesis in zebrafish. FASEB J. *23*, 4181–4192.

Ridder, A., Skupjen, P., Unterreitmeier, S., and Langosch, D. (2005). Tryptophan supports interaction of transmembrane helices. J. Mol. Biol. *354*, 894–902.

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

Sulistijo, E.S., and Mackenzie, K.R. (2009). Structural basis for dimerization of the BNIP3 transmembrane domain. Biochemistry *48*, 5106–5120.

Tu, L.N., Morohaku, K., Manna, P.R., Pelton, S.H., Butler, W.R., Stocco, D.M., and Selvaraj, V. (2014). Peripheral benzodiazepine receptor/translocator protein global knock-out mice are viable with no effects on steroid hormone biosynthesis. J. Biol. Chem. *289*, 27444–27454.

Unterreitmeier, S., Fuchs, A., Schäffler, T., Heym, R.G., Frishman, D., and Langosch, D. (2007). Phenylalanine promotes interaction of transmembrane domains via GxxxG motifs. J. Mol. Biol. *374*, 705–718.

Veenman, L., and Gavish, M. (2012). The role of 18 kDa mitochondrial translocator protein (TSPO) in programmed cell death, and effects of steroids on TSPO expression. Curr. Mol. Med. *12*, 398–412.

Veenman, L., Shandalov, Y., and Gavish, M. (2008). VDAC activation by the 18 kDa translocator protein (TSPO), implications for apoptosis. J. Bioenerg. Biomembr. *40*, 199–205.

Weber, M., and Schneider, D. (2013). Six amino acids define a minimal dimerization sequence and stabilize a transmembrane helix dimer by close packing and hydrogen bonding. FEBS Lett. *587*,

1592–1596.

Zeng, J., Guareschi, R., Damre, M., Cao, R., Kless, A., Neumaier, B., Bauer, A., Giorgetti, A., Carloni, P., and Rossetti, G. (2018). Structural prediction of the dimeric form of the mammalian translocator membrane protein TSPO: A key target for brain diagnostics. Int. J. Mol. Sci. *19*.

Zeno, S., Veenman, L., Katz, Y., Bode, J., Gavish, M., and Zaaroor, M. (2012). The 18 kDa mitochondrial translocator protein (TSPO) prevents accumulation of protoporphyrin IX. Involvement of reactive oxygen species (ROS). Curr. Mol. Med. *12*, 494–501.

Zhou, F.X., Merianos, H.J., Brunger, A.T., and Engelman, D.M. (2001). Polar residues drive association of polyleucine transmembrane helices. Proc Natl Acad Sci USA *98*, 2250–2255.

Zoued, A., Duneau, J.-P., Durand, E., España, A.P., Journet, L., Guerlesquin, F., and Cascales, E. (2018). Tryptophan-mediated Dimerization of the TssL Transmembrane Anchor Is Required for Type VI Secretion System Activity. J. Mol. Biol. *430*, 987–1003.

Chapter-5:

Aliev, A.E., Kulke, M., Khaneja, H.S., Chudasama, V., Sheppard, T.D., and Lanigan, R.M. (2014). Motional timescale predictions by molecular dynamics simulations: case study using proline and hydroxyproline sidechain dynamics. Proteins *82*, 195–215.

Ashkenazy, H., Abadi, S., Martz, E., Chay, O., Mayrose, I., Pupko, T., and Ben-Tal, N. (2016). ConSurf 2016: an improved methodology to estimate and visualize evolutionary conservation in macromolecules. Nucleic Acids Res. *44*, W344-50.

Bakan, A., Meireles, L.M., and Bahar, I. (2011). ProDy: protein dynamics inferred from theory and experiments. Bioinformatics *27*, 1575–1577.

Balsemão-Pires, E., Jaillais, Y., Olson, B.J.S.C., Andrade, L.R., Umen, J.G., Chory, J., and Sachetto-Martins, G. (2011). The Arabidopsis translocator protein (AtTSPO) is regulated at multiple levels in response to salt stress and perturbations in tetrapyrrole metabolism. BMC Plant Biol. *11*, 108.

Bateman, A., Coin, L., Durbin, R., Finn, R.D., Hollich, V., Griffiths-Jones, S., Khanna, A., Marshall, M., Moxon, S., Sonnhammer, E.L.L., et al. (2004). The Pfam protein families database. Nucleic Acids Res. *32*, D138-41.

Busch, A.W.U., and Montgomery, B.L. (2015). The Tryptophan-Rich Sensory Protein (TSPO) is Involved in Stress-Related and Light-Dependent Processes in the Cyanobacterium Fremyella diplosiphon. Front. Microbiol. *6*, 1393.

Busch, A.W.U., and Montgomery, B.L. (2017). Distinct light-, stress-, and nutrient-dependent regulation of multiple tryptophan-rich sensory protein (TSPO) genes in the cyanobacterium Fremyella diplosiphon. Plant Signal. Behav. *12*, e1293221.

Busch, A.W.U., WareJoncas, Z., and Montgomery, B.L. (2017). Tryptophan-Rich Sensory Protein/Translocator Protein (TSPO) from Cyanobacterium Fremyella diplosiphon Binds a Broad Range of Functionally Relevant Tetrapyrroles. Biochemistry *56*, 73–84.

Colasanti, A., Owen, D.R., Grozeva, D., Rabiner, E.A., Matthews, P.M., Craddock, N., and Young, A.H. (2013). Bipolar Disorder is associated with the rs6971 polymorphism in the gene encoding 18 kDa Translocator Protein (TSPO). Psychoneuroendocrinology *38*, 2826–2829.

Costa, B., Pini, S., Gabelloni, P., Da Pozzo, E., Abelli, M., Lari, L., Preve, M., Lucacchini, A., Cassano, G.B., and Martini, C. (2009a). The spontaneous Ala147Thr amino acid substitution within the translocator protein influences pregnenolone production in lymphomonocytes of healthy individuals. Endocrinology *150*, 5438–5445.

Costa, B., Pini, S., Martini, C., Abelli, M., Gabelloni, P., Landi, S., Muti, M., Gesi, C., Lari, L., Cardini, A., et al. (2009b). Ala147Thr substitution in translocator protein is associated with adult separation anxiety in patients with depression. Psychiatr. Genet. *19*, 110–111.

Drozdetskiy, A., Cole, C., Procter, J., and Barton, G.J. (2015). JPred4: a protein secondary structure prediction server. Nucleic Acids Res. *43*, W389-94.

El-Gebali, S., Mistry, J., Bateman, A., Eddy, S.R., Luciani, A., Potter, S.C., Qureshi, M., Richardson, L.J., Salazar, G.A., Smart, A., et al. (2019). The Pfam protein families database in 2019. Nucleic Acids Res. *47*, D427–D432.

Fan, J., Lindemann, P., Feuilloley, M.G.J., and Papadopoulos, V. (2012). Structural and functional

evolution of the translocator protein (18 kDa). Curr. Mol. Med. *12*, 369–386.

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

Finn, R.D., Coggill, P., Eberhardt, R.Y., Eddy, S.R., Mistry, J., Mitchell, A.L., Potter, S.C., Punta, M., Qureshi, M., Sangrador-Vegas, A., et al. (2016). The Pfam protein families database: towards a more sustainable future. Nucleic Acids Res. *44*, D279-85.

Ginter, C., Kiburu, I., and Boudker, O. (2013). Chemical catalysis by the translocator protein (18 kDa). Biochemistry *52*, 3609–3611.

Glaser, F., Pupko, T., Paz, I., Bell, R.E., Bechor-Shental, D., Martz, E., and Ben-Tal, N. (2003). ConSurf: identification of functional regions in proteins by surface-mapping of phylogenetic information. Bioinformatics *19*, 163–164.

Guo, Y., Kalathur, R.C., Liu, Q., Kloss, B., Bruni, R., Ginter, C., Kloppmann, E., Rost, B., and Hendrickson, W.A. (2015). Protein structure. Structure and activity of tryptophan-rich TSPO proteins. Science *347*, 551–555.

Humphrey, W., Dalke, A., and Schulten, K. (1996). VMD: visual molecular dynamics. J. Mol. Graph. *14*, 33–38, 27.

Jaremko, L., Jaremko, M., Giller, K., Becker, S., and Zweckstetter, M. (2014). Structure of the mitochondrial translocator protein in complex with a diagnostic ligand. Science *343*, 1363–1366.

Leneveu-Jenvrin, C., Connil, N., Bouffartigues, E., Papadopoulos, V., Feuilloley, M.G.J., and Chevalier, S. (2014). Structure-to-function relationships of bacterial translocator protein (TSPO): a focus on Pseudomonas. Front. Microbiol. *5*, 631.

Leneveu-Jenvrin, C., Bouffartigues, E., Maillot, O., Cornelis, P., Feuilloley, M.G.J., Connil, N., and Chevalier, S. (2015). Expression of the translocator protein (TSPO) from Pseudomonas fluorescens Pf0-1 requires the stress regulatory sigma factors AlgU and RpoH. Front. Microbiol. *6*, 1023.

Li, F., Liu, J., Zheng, Y., Garavito, R.M., and Ferguson-Miller, S. (2015a). Protein structure. Crystal structures of translocator protein (TSPO) and mutant mimic of a human polymorphism. Science *347*, 555–558.

Li, F., Liu, J., Garavito, R.M., and Ferguson-Miller, S. (2015b). Evolving understanding of translocator protein 18 kDa (TSPO). Pharmacol. Res. *99*, 404–409.

Lomize, A.L., Pogozheva, I.D., and Mosberg, H.I. (2011). Anisotropic solvent model of the lipid bilayer. 2. Energetics of insertion of small molecules, peptides, and proteins in membranes. J. Chem. Inf. Model. *51*, 930–946.

Lomize, M.A., Pogozheva, I.D., Joo, H., Mosberg, H.I., and Lomize, A.L. (2012). OPM database and PPM web server: resources for positioning of proteins in membranes. Nucleic Acids Res. *40*, D370-6.

Marginedas-Freixa, I., Hattab, C., Bouyer, G., Halle, F., Chene, A., Lefevre, S.D., Cambot, M., Cueff, A., Schmitt, M., Gamain, B., et al. (2016). TSPO ligands stimulate ZnPPIX transport and ROS accumulation leading to the inhibition of P. falciparum growth in human blood. Sci. Rep. *6*, 33516.

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

Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C., and Ferrin,

T.E. (2004). UCSF Chimera—a visualization system for exploratory research and analysis. J. Comput. Chem. *25*, 1605–1612.

Rampon, C., Bouzaffour, M., Ostuni, M.A., Dufourcq, P., Girard, C., Freyssinet, J.M., Lacapere, J.-J., Schweizer-Groyer, G., and Vriz, S. (2009). Translocator protein (18 kDa) is involved in primitive erythropoiesis in zebrafish. FASEB J. *23*, 4181–4192.

Vanhee, C., Zapotoczny, G., Masquelier, D., Ghislain, M., and Batoko, H. (2011). The Arabidopsis multistress regulator TSPO is a heme binding membrane protein and a potential scavenger of porphyrins via an autophagy-dependent degradation mechanism. Plant Cell *23*, 785–805.

Wallace, A.C., Laskowski, R.A., and Thornton, J.M. (1995). LIGPLOT: a program to generate schematic diagrams of protein-ligand interactions. Protein Eng. *8*, 127–134.

Waterhouse, A.M., Procter, J.B., Martin, D.M.A., Clamp, M., and Barton, G.J. (2009). Jalview Version 2--a multiple sequence alignment editor and analysis workbench. Bioinformatics *25*, 1189–1191.

Wendler, G., Lindemann, P., Lacapère, J.-J., and Papadopoulos, V. (2003). Protoporphyrin IX binding and transport by recombinant mouse PBR. Biochem. Biophys. Res. Commun. *311*, 847–852.

Yeliseev, A.A., and Kaplan, S. (1995). A sensory transducer homologous to the mammalian peripheral-type benzodiazepine receptor regulates photosynthetic membrane complex formation in Rhodobacter sphaeroides 2.4.1. J. Biol. Chem. *270*, 21167–21175.

Yeliseev, A.A., and Kaplan, S. (2000). TspO of rhodobacter sphaeroides. A structural and functional model for the mammalian peripheral benzodiazepine receptor. J. Biol. Chem. *275*, 5657–5667.

Zeng, X., and Kaplan, S. (2001). TspO as a modulator of the repressor/antirepressor (PpsR/AppA) regulatory system in Rhodobacter sphaeroides 2.4.1. J. Bacteriol. *183*, 6355–6364.

Zeno, S., Veenman, L., Katz, Y., Bode, J., Gavish, M., and Zaaroor, M. (2012). The 18 kDa mitochondrial translocator protein (TSPO) prevents accumulation of protoporphyrin IX. Involvement of reactive oxygen species (ROS). Curr. Mol. Med. *12*, 494–501.

Chapter-6:

Anderson, M.K., Sun, X., Miracle, A.L., Litman, G.W., and Rothenberg, E.V. (2001). Evolution of hematopoiesis: Three members of the PU.1 transcription factor family in a cartilaginous fish, Raja eglanteria. Proc Natl Acad Sci USA *98*, 553–558.

Bakan, A., Meireles, L.M., and Bahar, I. (2011). ProDy: protein dynamics inferred from theory and experiments. Bioinformatics *27*, 1575–1577.

Bouyer, G., Cueff, A., Egée, S., Kmiecik, J., Maksimova, Y., Glogowska, E., Gallagher, P.G., and Thomas, S.L.Y. (2011). Erythrocyte peripheral type benzodiazepine receptor/voltage-dependent anion channels are upregulated by Plasmodium falciparum. Blood *118*, 2305–2312.

Brüschweiler, R. (1995). Collective protein dynamics and nuclear spin relaxation. J. Chem. Phys. *102*, 3396.

Brüschweiler, R., and Case, D.A. (1994). Collective NMR relaxation model applied to protein dynamics. Phys. Rev. Lett. *72*, 940–943.

Colasanti, A., Owen, D.R., Grozeva, D., Rabiner, E.A., Matthews, P.M., Craddock, N., and Young, A.H. (2013). Bipolar Disorder is associated with the rs6971 polymorphism in the gene encoding 18 kDa Translocator Protein (TSPO). Psychoneuroendocrinology *38*, 2826–2829.

Costa, B., Pini, S., Gabelloni, P., Da Pozzo, E., Abelli, M., Lari, L., Preve, M., Lucacchini, A., Cassano, G.B., and Martini, C. (2009a). The spontaneous Ala147Thr amino acid substitution within the translocator protein influences pregnenolone production in lymphomonocytes of healthy individuals. Endocrinology *150*, 5438–5445.

Costa, B., Pini, S., Martini, C., Abelli, M., Gabelloni, P., Landi, S., Muti, M., Gesi, C., Lari, L., Cardini, A., et al. (2009b). Ala147Thr substitution in translocator protein is associated with adult separation anxiety in patients with depression. Psychiatr. Genet. *19*, 110–111.

David, C.C., and Jacobs, D.J. (2014). Principal component analysis: a method for determining the essential dynamics of proteins. Methods Mol. Biol. *1084*, 193–226.

Di Scala, C., Baier, C.J., Evans, L.S., Williamson, P.T.F., Fantini, J., and Barrantes, F.J. (2017). Relevance of CARC and CRAC Cholesterol-Recognition Motifs in the Nicotinic Acetylcholine Receptor and Other Membrane-Bound Receptors. Curr. Top. Membr. *80*, 3–23.

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

Fan, J., Lindemann, P., Feuilloley, M.G.J., and Papadopoulos, V. (2012). Structural and functional evolution of the translocator protein (18 kDa). Curr. Mol. Med. *12*, 369–386.

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

Ginter, C., Kiburu, I., and Boudker, O. (2013). Chemical catalysis by the translocator protein (18 kDa). Biochemistry *52*, 3609–3611.

Guo, Y., Kalathur, R.C., Liu, Q., Kloss, B., Bruni, R., Ginter, C., Kloppmann, E., Rost, B., and Hendrickson, W.A. (2015). Protein structure. Structure and activity of tryptophan-rich TSPO proteins. Science *347*, 551–555.

Hardwick, M., Fertikh, D., Culty, M., Li, H., Vidic, B., and Papadopoulos, V. (1999). 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. *59*, 831–842.

Harms, M.J., and Thornton, J.W. (2010). Analyzing protein structure and function using ancestral gene reconstruction. Curr. Opin. Struct. Biol. *20*, 360–366.

Holland, P.W., Garcia-Fernàndez, J., Williams, N.A., and Sidow, A. (1994). Gene duplications and the origins of vertebrate development. Dev. Suppl. 125–133.

Holm, T.M., Braun, A., Trigatti, B.L., Brugnara, C., Sakamoto, M., Krieger, M., and Andrews, N.C. (2002). Failure of red blood cell maturation in mice with defects in the high-density lipoprotein receptor SR-BI. Blood *99*, 1817–1824.

Isin, B., Tirupula, K.C., Oltvai, Z.N., Klein-Seetharaman, J., and Bahar, I. (2012). Identification of motions in membrane proteins by elastic network models and their experimental validation. Methods Mol. Biol. *914*, 285–317.

Jamin, N., Neumann, J.-M., Ostuni, M.A., Vu, T.K.N., Yao, Z.-X., Murail, S., Robert, J.-C., Giatzakis, C., Papadopoulos, V., and Lacapère, J.-J. (2005). Characterization of the cholesterol recognition amino acid consensus sequence of the peripheral-type benzodiazepine receptor. Mol. Endocrinol. *19*, 588–594.

Johanson, Z., Long, J.A., Talent, J.A., Janvier, P., and Warren, J.W. (2006). Oldest coelacanth, from the Early Devonian of Australia. Biol. Lett. *2*, 443–446.

Katoh, K., Misawa, K., Kuma, K., and Miyata, T. (2002). MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Res. *30*, 3059–3066.

Kurauskas, V., Hessel, A., Ma, P., Lunette, P., Weinhaupl, K., Imbert, L., Brutscher, B., King, MS., Sounier, R., Dolce, V., Kunji, E., R., S., Capobianco, L., Chipot, C., Dehez, F., Bersch, B., Schanda, P. (2018). How Detergent Impacts Membrane Proteins: Atomic-Level Views of Mitochondrial Carriers in Dodecylphosphocholine. Phys. Chem. Lett. *9*, 933-938.

Li, F., Liu, J., Valls, L., Hiser, C., and Ferguson-Miller, S. (2015a). Identification of a key cholesterol binding enhancement motif in translocator protein 18 kDa. Biochemistry *54*, 1441–1443.

Li, F., Liu, J., Garavito, R.M., and Ferguson-Miller, S. (2015b). Evolving understanding of translocator protein 18 kDa (TSPO). Pharmacol. Res. *99*, 404–409.

Li, F., Liu, J., Zheng, Y., Garavito, R.M., and Ferguson-Miller, S. (2015c). Protein structure. Crystal structures of translocator protein (TSPO) and mutant mimic of a human polymorphism. Science *347*, 555–558.

Li, H., Yao, Z., Degenhardt, B., Teper, G., and Papadopoulos, V. (2001). 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 *98*, 1267–1272.

Lomize, A.L., Pogozheva, I.D., and Mosberg, H.I. (2011). Anisotropic solvent model of the lipid bilayer. 2. Energetics of insertion of small molecules, peptides, and proteins in membranes. J. Chem. Inf. Model. *51*, 930–946.

Lomize, M.A., Pogozheva, I.D., Joo, H., Mosberg, H.I., and Lomize, A.L. (2012). OPM database and PPM web server: resources for positioning of proteins in membranes. Nucleic Acids Res. *40*, D370-6.

Marginedas-Freixa, I., Hattab, C., Bouyer, G., Halle, F., Chene, A., Lefevre, S.D., Cambot, M., Cueff, A., Schmitt, M., Gamain, B., et al. (2016). TSPO ligands stimulate ZnPPIX transport and ROS accumulation leading to the inhibition of P. falciparum growth in human blood. Sci. Rep. *6*, 33516.

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

Pébusque, M.J., Coulier, F., Birnbaum, D., and Pontarotti, P. (1998). Ancient large-scale genome duplications: phylogenetic and linkage analyses shed light on chordate genome evolution. Mol. Biol. Evol. *15*, 1145–1159.

Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C., and Ferrin, T.E. (2004). UCSF Chimera—a visualization system for exploratory research and analysis. J. Comput. Chem. *25*, 1605–1612.

Postic, G., Ghouzam, Y., and Gelly, J.-C. (2016). OREMPRO web server: orientation and assessment of atomistic and coarse-grained structures of membrane proteins. Bioinformatics *32*, 2548–2550.

Rone, M.B., Midzak, A.S., Issop, L., Rammouz, G., Jagannathan, S., Fan, J., Ye, X., Blonder, J., Veenstra, T., and Papadopoulos, V. (2012). Identification of a dynamic mitochondrial protein complex driving cholesterol import, trafficking, and metabolism to steroid hormones. Mol. Endocrinol. *26*, 1868–1882.

Simos, G., Maison, C., and Georgatos, S.D. (1996). Characterization of p18, a component of the lamin B receptor complex and a new integral membrane protein of the avian erythrocyte nuclear envelope. J. Biol. Chem. *271*, 12617–12625.

Thornton, J.W., Need, E., and Crews, D. (2003). Resurrecting the ancestral steroid receptor: ancient origin of estrogen signaling. Science *301*, 1714–1717.

Tieleman, D.P., Shrivastava, I.H., Ulmschneider, M.R., and Sansom, M.S. (2001). Proline-induced hinges in transmembrane helices: possible roles in ion channel gating. Proteins *44*, 63–72.

Vanhee, C., Zapotoczny, G., Masquelier, D., Ghislain, M., and Batoko, H. (2011). The Arabidopsis multistress regulator TSPO is a heme binding membrane protein and a potential scavenger of porphyrins via an autophagy-dependent degradation mechanism. Plant Cell *23*, 785–805.

Verma, A., Nye, J.S., and Snyder, S.H. (1987). Porphyrins are endogenous ligands for the mitochondrial (peripheral-type) benzodiazepine receptor. Proc Natl Acad Sci USA *84*, 2256–2260.

Waterhouse, A.M., Procter, J.B., Martin, D.M.A., Clamp, M., and Barton, G.J. (2009). Jalview Version 2--a multiple sequence alignment editor and analysis workbench. Bioinformatics *25*, 1189–1191.

Wendler, G., Lindemann, P., Lacapère, J.-J., and Papadopoulos, V. (2003). Protoporphyrin IX binding and transport by recombinant mouse PBR. Biochem. Biophys. Res. Commun. *311*, 847–852.

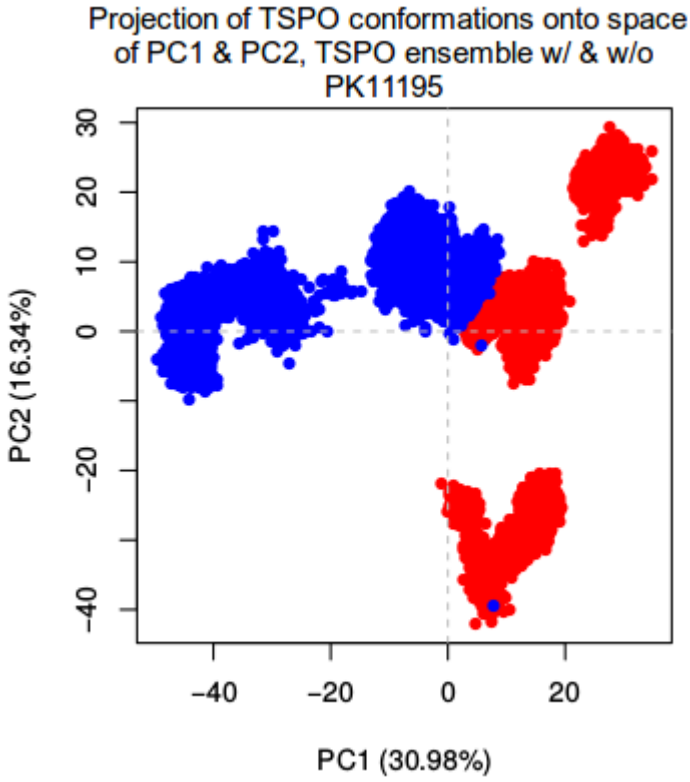
Yeliseev, A.A., and Kaplan, S. (1999). A novel mechanism for the regulation of photosynthesis gene expression by the TspO outer membrane protein of Rhodobacter sphaeroides 2.4.1. J. Biol. Chem. *274*, 21234–21243.

Yeliseev, A.A., and Kaplan, S. (2000). TspO of rhodobacter sphaeroides. A structural and functional

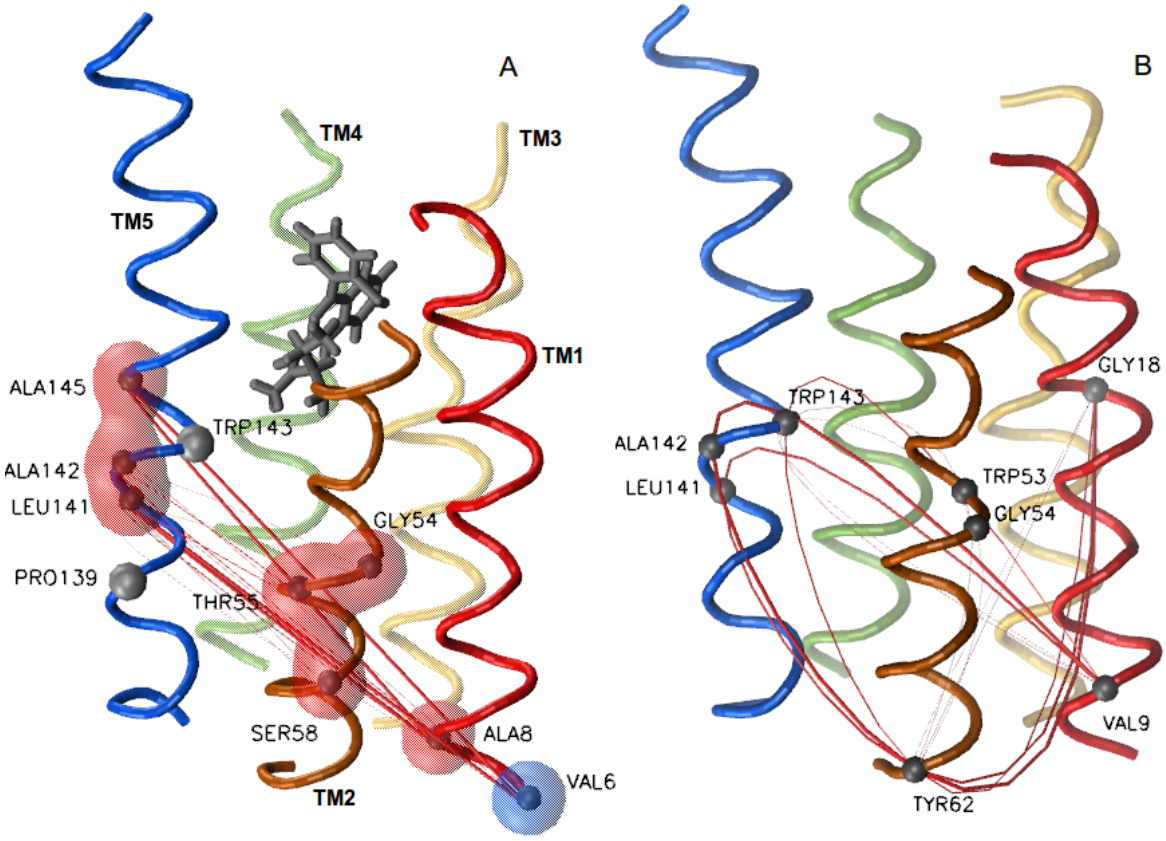
model for the mammalian peripheral benzodiazepine receptor. J. Biol. Chem. *275*, 5657–5667.

Yokoyama, S., and Radlwimmer, F.B. (1999). The molecular genetics of red and green color vision in mammals. Genetics *153*, 919–932.

Annexure



Annexure-1 PCA on TSPO conformations, residues 6-157 (including loops): Projection of ensemble of all TSPO conformations in presence of PK11195 (blue) as well as in absence of PK11195 (red), projected onto the subspace of PC1 and PC2. The PCA was performed on all the residues, including TM regions as well as extracellular loops, except N- and C-terminal loop residues.



Annexure-2: Allosteric pathway analysis on residues of TM helices vs. the residues of TM regions. A: Structure of TSPO showing allosteric pathway on TSPO in presence of PK11195. Analysis were performed on the residues of TM helices (residues 6-36;46-69;76-100;104-126;131-157, representing TM helices 1-5 respectively); B: Structure of TSPO showing allosteric pathway on TSPO in presence of PK11195. Analysis were performed on the residues of TM regions (residues 7-26,46-63,82-101,106-124,134-157, representing TM regions 1-5 respectively). Boundaries of TM regions are adopted from OPM database (Lomize et. al., 2006; Lomize et. al., 2012).

	Mouse		*R. sphaeroides*		*B. cereus*	
	OPM boundaries	TM regions for analysis	OPM boundaries	TM regions for analysis	OPM boundaries	TM regions for analysis
TM1	7-26	7-24	5-24	7-24	7-25	8-25
TM2	46-63	46-63	44-65	45-62	44-65	45-62
TM3	82-102	82-101	71-90	71-90	73-95	75-94
TM4	106-124	106-124	99-120	100-118	102-122	104-122
TM5	134-153	134-153	124-144	124-143	127-144	127-146

Annexure-3: Comparison of TM boundaries of TSPO structures of mouse (PDB: 2MGY), *R. sphaeroides* (PDB: 4UC1) and *B. cereus* (4RYI) from OPM database (Lomize et. al., 2006; Lomize et. al., 2012)

Abstract

Despite availability of different drugs, malaria remains a serious health challenge in tropical countries, and is rapidly making inroads in the countries with temperate climate. To compound the problem, resistance to the antimalarial drugs by the Plasmodial parasite has been documented in many instances, underscoring the need for novel therapeutic approaches for the disease. However, discovery of involvement of a transmembrane protein Translocator protein (TSPO) in the Plasmodium-infected erythrocytes showed the potential of this protein as a target for novel antimalarial drugs. TSPO is an 18 kDa transmembrane protein found in various organelles of cell and involved in variety of processes, including steroidogenesis and cholesterol transport, tetrapyrrole transport/catalysis, erythropoiesis. apoptosis, response to stress to name a few. As a result, it is also involved in various diseases apart from malaria, and is considered a potential drug target. Though recent studies on its atomic structure provided some interesting insights into the structure, several questions on the structure-dynamics-function relationship of TSPO remained unanswered. In this thesis, we attempted to answer some of these questions using various bioinformatics and molecular modelling approaches. We describe how does the ligand PK11195 influence the dynamics of TSPO using all-atom molecular dynamics simulations, and how does PK11195 itself behave in complex with TSPO. We also studied different aspects of TSPO oligomerization using coarse-grained molecular dynamics simulations, and explored how do mammalian TSPO dynamics compare with those of bacterial homologs. Ultimately, we attempted to make a comprehensive comparison between the two paralogs of TSPO in mammals, TSPO1 and TSPO2, from sequence, structural as well as phylogenetic perspectives. Our results give interesting insights into the structure-dynamics-function relationships of TSPO, and pave way to some new hypotheses on the same, which could be exploited to develop new generation of antimalarial drugs that target TSPO.

Keywords:

Translocator protein, molecular dynamics, PK11195, oligomerization.

Résumé

Le paludisme reste un grave problème de santé, bien que différents médicaments sont disponibles pour le traitement. En outre, il se propage à l’une vitesse alarmante dans les pays avec climat tempéré. Pour aggraver le problème, la résistance aux antipaludiques du parasite Plasmodium a été documentée dans de nombreux cas, soulignant la nécessité de nouvelles approches thérapeutiques pour la maladie. Cependant, la découverte de l'implication d'une protéine transmembranaire, la protéine translocateur (TSPO), dans les érythrocytes infectés par Plasmodium a montré le potentiel de cette protéine en tant que cible pour de nouveaux médicaments antipaludiques. La TSPO est une protéine transmembranaire de 18 kDa présente dans divers organites de cellules et impliquée dans divers processus, notamment la stéroïdogenèse et le transport du cholestérol, le transport / catalysme du tétrapyrrole, l'érythropoïèse. apoptose, réponse au stress, et cetera. En conséquence, il est également impliqué dans diverses maladies en dehors du paludisme et est considéré comme une cible potentielle pour un médicament. Bien que des études récentes sur sa structure atomique aient fourni des informations intéressantes sur la structure, plusieurs questions sur la relation structure-dynamique-fonction de TSPO sont restées.Dans cette thèse, nous avons tenté de répondre à certaines de ces questions en utilisant diverses approches de bioinformatique et de modélisation moléculaire. Nous décrivons comment le ligand PK11195 influence la dynamique de TSPO en utilisant des simulations de dynamique moléculaire tout atome et comment PK11195 se comporte-t-il en complexe avec TSPO. Nous avons également étudié les aspects différents de l’oligomérisation de TSPO à l’aide de simulations de dynamique moléculaire à grains grossiers et exploré comment la dynamique de TSPO de mammifère se compare à celle d’homologues bactériens. Finalement, nous avons tenté de faire une comparaison complète entre les deux paralogues de TSPO chez les mammifères, TSPO1 et TSPO2, du point de vue de la séquence, tant structurel que phylogénétique. Nos résultats fournissent des informations intéressantes sur les relations structure-dynamique-fonction de TSPO et ouvrent la voie à de nouvelles hypothèses similaires, qui pourraient être exploitées pour développer une nouvelle génération d'antipaludiques ciblant TSPO.

Mots clés:

Protéine translocateur, dynamique moléculaire, PK11195, oligomérisation.

UNIVERSITÉ DE LA RÉUNION

Pôle Recherche
Ecoles doctorales

LETTRE D'ENGAGEMENT DE NON-PLAGIAT

Je, soussigné(e) Rajas RAO en ma qualité de doctorant(e) de l'Université de La Réunion, déclare être conscient(e) que le plagiat est un acte délictueux passible de sanctions disciplinaires. Aussi, dans le respect de la propriété intellectuelle et du droit d'auteur, je m'engage à systématiquement citer mes sources, quelle qu'en soit la forme (textes, images, audiovisuel, internet), dans le cadre de la rédaction de ma thèse et de toute autre production scientifique, sachant que l'établissement est susceptible de soumettre le texte de ma thèse à un logiciel anti-plagiat.

Fait à ~~Saint-Denis~~ Paris le : 5-4-2019

Signature :

Extrait du Règlement intérieur de l'Université de La Réunion
(validé par le Conseil d'Administration en date du 11 décembre 2014)

Article 9. Protection de la propriété intellectuelle – Faux et usage de faux, contrefaçon, plagiat

L'utilisation des ressources informatiques de l'Université implique le respect de ses droits de propriété intellectuelle ainsi que ceux de ses partenaires et plus généralement, de tous tiers titulaires de tes droits.
En conséquence, chaque utilisateur doit :
- utiliser les logiciels dans les conditions de licences souscrites ;
- ne pas reproduire, copier, diffuser, modifier ou utiliser des logiciels, bases de données, pages Web, textes, images, photographies ou autres créations protégées par le droit d'auteur ou un droit privatif, sans avoir obtenu préalablement l'autorisation des titulaires de ces droits.

La contrefaçon et le faux
Conformément aux dispositions du code de la propriété intellectuelle, toute représentation ou reproduction intégrale ou partielle d'une œuvre de l'esprit faite ans le consentement de son auteur est illicite et constitue un délit pénal.
L'article 444-1 du code pénal dispose : « Constitue un faux toute altération frauduleuse de la vérité, de nature à cause un préjudice et accomplie par quelque moyen que ce soit, dans un écrit ou tout autre support d'expression de la pensée qui a pour objet ou qui peut avoir pour effet d'établir la preuve d'un droit ou d'un fait ayant des conséquences juridiques ».
L'article L335_3 du code de la propriété intellectuelle précise que : « Est également un délit de contrefaçon toute reproduction, représentation ou diffusion, par quelque moyen que ce soit, d'une œuvre de l'esprit en violation des droits de l'auteur, tels qu'ils sont définis et réglementés par la loi. Est également un délit de contrefaçon la violation de l'un des droits de l'auteur d'un logiciel (…) ».

Le plagiat est constitué par la copie, totale ou partielle d'un travail réalisé par autrui, lorsque la source empruntée n'est pas citée, quel que soit le moyen utilisé. Le plagiat constitue une violation du droit d'auteur (au sens des articles L 335-2 et L 335-3 du code de la propriété intellectuelle). Il peut être assimilé à un délit de contrefaçon. C'est aussi une faute disciplinaire, susceptible d'entraîner une sanction.
Les sources et les références utilisées dans le cadre des travaux (préparations, devoirs, mémoires, thèses, rapports de stage…) doivent être clairement citées. Des citations intégrales peuvent figurer dans les documents rendus, si elles sont assorties de leur référence (nom d'auteur, publication, date, éditeur…) et identifiées comme telles par des guillemets ou des italiques.

Les délits de contrefaçon, de plagiat et d'usage de faux peuvent donner lieu à une sanction disciplinaire indépendante de la mise en œuvre de poursuites pénales.