
**Multiscale modelling and structure-guided
development of small molecules tackling
neuroinflammation and inflammageing**



This thesis is submitted to the School of Natural and
Environmental Sciences, Newcastle University, for the degree
of Doctor of Philosophy

Danlin Liu

September 2021

Abstract

Early diagnosis and treatment of neurodegenerative diseases has gained attention, given their increasing prevalence. Multiple proteins are currently being investigated as novel targets for drug development, including $\alpha 7$ nicotinic receptors and mitochondrial translocator protein TSPO. These proteins play an important role in neuroinflammation, which makes them attractive for development of drugs and diagnostics. However, small molecule development has been hampered by existence of a human-specific dup $\alpha 7$ isoform, and an A147T-TSPO polymorph, which present a challenge for development of potent and selective ligands.

In this work, I characterised the structure and dynamics of the most plausible functional $\alpha 7$ pentamers bearing dup $\alpha 7$ subunits. The receptors have been modelled and assessed using multiscale molecular dynamics (MD) simulations and enhanced sampling techniques. The energy landscapes of the pentamers with different stoichiometries showed that receptors with a low ratio of dup $\alpha 7$ / $\alpha 7$ remained functional. Sensitivity of dup $\alpha 7$ receptors to an antagonist (α -BTX) and amyloid A β_{42} has also been assessed.

Further, putative “druggable” binding sites at dup $\alpha 7$ receptors were mapped, and interactions between dup $\alpha 7$ and small molecules were explored using a combination of solvent mapping, MD simulations, and molecular docking. Results indicated that neither established orthosteric agonist nor allosteric ligands can bind to dup $\alpha 7$ /dup $\alpha 7$ interfaces, however, $\alpha 7$ /dup $\alpha 7$ interfaces remain “druggable”. In addition, several novel allosteric sites were detected on $\alpha 7$ /dup $\alpha 7$ receptors. The final part of this work focused on development of novel tracers for A147T-TSPO variant. Using a combination of molecular modelling, MD simulations, and structure-guided drug design, I have evaluated plausible binding modes of established TSPO ligands to A147T-TSPO. Results explain the origins of diminished affinity of some established TSPO ligands to A147T-TSPO. Moreover, I have identified the position of fluorine atom, which is a derivative of DPA-714 compound to bind to A147T-TSPO with sub-nanomolar affinity. The compound, denoted as MKD, is feasible for the radiosynthesis.

Declaration

The research described within this thesis was performed between September 2017 and September 2021 in the Computational Medicinal Chemistry Laboratory, Bedson Building, School of Natural and Environmental Sciences, Newcastle University, Newcastle-upon-Tyne, UK, NE1 7RU.

All research described within this thesis is original in content and does not incorporate any material or ideas previously published or presented by other authors except where due reference is given.

No part of this thesis has been previously submitted for a degree, diploma or any other qualification at any other academic institution.

若批评无自由，则赞美无意义。

Acknowledgements

First of all, I would like to thank my supervisor, Dr Agnieszka Bronowska. I still can remember the words she said four years ago: ‘Danlin, would you like to be a PhD student?’. After that conversation, I joined this lovely group and started my life as a PhD student. She supported every idea I have had and encouraged me to try them out. She makes science look interesting, makes complex subjects clear and straightforward, and helps me open my mind. Her genius, fantastic creativity, and excitement for science have influenced me a lot in doing research. In terms of life, she gave me complete freedom and respected my routines and habits. When I had trouble in life or felt confused, anxious, or sad, she was generous in helping me, guiding me and solving these problems. She will always be an essential person in my life.

Additionally, I would like to thank my colleagues in the group for their support, help and friendship throughout these years, especially Dr João de Souza. He has taught me a lot and given me much advice, not only in science but also in life. His patience in answering all my questions helped me grow from having almost no experience in computational chemistry. Also, all the current and former members of the group - Quico, Sylvia, Matt, Ayaz, Ruidi, Rhys, Mete, Piotr, Weikang, Lanyu, Shangze, Haoyuan, and Yancong. This group is so lovely and happy, thanks to their dedication and hard work.

I should also thank my computer machine for helping me do much work and for being with me throughout my PhD.

Finally, I would like to thank the most important people in my life, my parents, Meiyan and Xufeng. They accompanied me with love all the time. They taught me, nurtured me and built my outlook on life, the world and my values. They supported me in all my decisions and funded my four years of PhD studies. I feel very, very grateful to them. They are excellent parents. In these particular years, they have been the biggest motivation for me to complete my PhD studies. Without them, I would not be the person I am now. Thank you, my dear mum and dad.

Abbreviations

| | | |
|----------|-----------|---|
| A | ATP | Adenosine triphosphate |
| | AD | Alzheimer's disease |
| | A β | Amyloid β |
| | APP | Amyloid precursor protein |
| | Ach | Acetylcholine |
| | Ac CoA | Acetyl coenzyme A |
| | AChE | Acetylcholinesterase |
| | AChR | Acetylcholine receptor |
| | AC | Adenylyl cyclase |
| | ASC | Adaptor-apoptosis-associated speck-like protein |
| | AMPK | Adenosine monophosphate-activated protein kinase |
| | AChBP | Acetylcholine binding protein |
| | AKT | Protein kinase B |
| | ANT | Adenine nucleotide translocation |
| | ADHD | Attention deficit hyperactivity disorder |
| | ALS | Amyotrophic lateral sclerosis |
| | ADME | Absorption, distribution, metabolism, and excretion |

| | | |
|----------|-------|---|
| | AChEI | Acetylcholine esterase inhibitors |
| | Å | Angström |
| | AMBER | Assisted Model Building with Energetic Refinement |
| | ANC | Adenine nucleotide carrier |
| B | BBB | Blood-brain barrier |
| | BuChE | Butyrycholinesterase |
| | BTX | Bungarotoxin |
| | BLAST | Basic Local Alignment Search Tool |
| C | C/Cys | Cysteine |
| | CRP | C-reactive protein |
| | CNS | Central nervous system |
| | CSF | Cerebrospinal fluid |
| | CHIT1 | Chitotriosidase |
| | ChAT | Cytosolic enzyme choline acetyltransferase |
| | CHT1 | Choline transporter |
| | CHS | Cholesteryl hemisuccinate |
| | cAMP | Cyclic Adenosine Monophosphate |
| | CVD | Cardiovascular disease |

| | | |
|----------|---------|---|
| | CAP | Cholinergic anti-inflammatory pathway |
| | CD | Circular dichroism |
| | CoA | Coenzyme |
| | CCL-2/7 | Chemokines 2/7 |
| | CHARMM | Chemistry at HARvard Macromolecular Mechanics |
| | Cryo-EM | Cryogenic Electron Microscopy |
| | CADD | Computer-aided drug design |
| | CG | Coarse-gained |
| | COM | Centre-of-mass |
| D | D/Asp | Aspartic acid |
| | DAG | Diacylglycerol |
| | DOF | Degree of freedom |
| | DPPC | Dipalmitoylphosphatidylcholine |
| | Dup | Duplication |
| E | E/Glu | Glutamic acid |
| | EC | Extracellular |
| | EM | Electron microscopy |
| | ELIC | Erwinia ligand-gate ion channel |

| | | |
|----------|---------|---|
| | ERK | Extracellular signal-regulated kinases |
| F | F/Phe | Phenylalanine |
| | Flt-1 | Fms Related Receptor Tyrosine Kinase 1 |
| | FF | Force field |
| | FFT | Fast Fourier Transformation |
| G | G/Gly | Glycine |
| | GPCR | G protein-coupled receptor |
| | GIRK | G-protein-activated inwardly rectifying potassium channel |
| | Glu | Glutamic acid |
| | GABA | γ -aminobutyric acid |
| | GlyR | glycine receptors |
| | GPU | Graphics Processing Unit |
| | GROMACS | GRONingen MACHine for Chemical Simulations |
| H | H/His | Histidine |
| | HF | Heart failure |
| | HO-1 | Heme oxygenase-1 |
| | 5-HT | 5-hydroxytryptamine |
| | HD | Huntington's disease |

| | | |
|----------|------------------|---------------------------------------|
| | H-bond | Hydrogen bond |
| I | I/Ile | Isoleucine |
| | IL | interleukin |
| | IFN- γ | Interferon-gamma |
| | ICAM-1 | Intercellular Adhesion Molecule-1 |
| | IP3 | Inositol-1,4,5-trisphosphate |
| | ICM | Ischaemic cardiomyopathy |
| | IC | Intracellular |
| | ITC | Isothermal titration calorimetry |
| | I κ B | Inhibitor kappa B |
| | IC ₅₀ | Half Maximal Inhibitory Concentration |
| J | JAK | Janus kinase |
| K | K/Lys | Lysine |
| | kcal | Kilocalories |
| | K _d | Dissociation constant |
| L | L/Leu | Leucine |
| | LPS | Lipopolysaccharide |
| | LTP | Long-term potentiation |

| | | |
|----------|-------|--|
| | LGIC | Ligand-gated ion channel |
| | LJ | Lennard-Jones |
| | LE | Ligand efficiency |
| | LEE | Lipophilic ligand efficiency |
| | Lob | Lobeline |
| | logP | Partition coefficient |
| M | M/Met | Methionine |
| | MMP | Matrix metalloproteinase |
| | MS | Multiple sclerosis |
| | MPTP | 1-methyl-4-phenyl-1,2,3,6-tetrahydropyridine |
| | MPTP | Mitochondrial permeability transition pore |
| | MPT | Mitochondrial permeability transition |
| | mAChR | Muscarinic acetylcholine receptor |
| | MAPK | Mitogen-activated protein kinase |
| | mRNA | Messenger ribonucleic acid |
| | MI | Myocardial infarction |
| | MD | Molecular Dynamics |
| | MM | Molecular mechanics |

| | | |
|----------|-------------|---|
| | MW | Molecular weight |
| N | nAChR | Nicotinic acetylcholine receptor |
| | NFT | Neurofibrillary tangles |
| | NLRP3 | NOD-like receptor family, pyrin domain containing 3 |
| | NMCM | Neonatal mouse cardiomyocytes |
| | NNCS | Non-neuronal cholinergic system |
| | Nic | Nicotine |
| | N/Asn | Asparagine |
| | NMR | Nuclear Magnetic Resonance |
| | NPT | Isothermal-Isobaric Ensemble |
| | ns | Nanosecond |
| | NVT | Canonical Ensemble |
| | nM | Nanomolar |
| O | 6-OHDA | 6-hydroxydopamine |
| P | P/Pro | Proline |
| | PD | Parkinson's disease |
| | PLC β | Phospholipase C β |
| | PKB | Protein kinase B |

| | | |
|----------|-------|---|
| | PKC | Protein kinase C |
| | PI3K | Phosphoinositide 3-kinases |
| | PARP | Poly ADP-ribose polymerase |
| | PET | Positron emission tomography |
| | PBR | Peripheral type benzodiazepine receptor |
| | PBC | Periodic Boundary Condition |
| | PC | Principal Component |
| | PCA | Principal Component Analysis |
| | PDB | Protein Data Bank |
| | PMF | Potential of Mean Force |
| | PPI | Protein-protein interactions |
| | ps | Picosecond |
| | P | Pressure |
| | PME | Particle Mesh Ewald |
| | pM | Picomolar |
| Q | Q/Gln | Glutamine |
| | QM | Quantum Mechanics |
| R | R/Arg | Arginine |

| | | |
|----------|---------------|---|
| | ROS | Reactive oxygen species |
| | RNS | Reactive nitrogen species |
| | RMSD | Root-mean Square Deviation |
| | RMSF | Root-mean Square Fluctuation |
| S | S/Ser | Serine |
| | Syn | Synuclein |
| | SP | Senile plaques |
| | STAT | Signal transduction and activator of transcription |
| | SNP | single-nucleotide polymorphism\ |
| | SAR | Structure-activity relationship |
| | SBDD | Structure-based drug design |
| | SASA | Solvent Accessible Surface Area |
| T | T/Thr | Threonine |
| | TIP3P | Transferable Intermolecular Potential with 3 Points |
| | TSPO | Translocator Protein |
| | TNF- α | Tumour necrosis factor α |
| | TM | Transmembrane |
| | TPSA | Topological polar surface area |

| | | |
|----------|-------|-------------------------------------|
| U | US | Umbrella Sampling |
| V | v/v | volume/volume |
| | VACHT | Vesicular acetylcholine transporter |
| | VDCC | Voltage-gated calcium channels |
| | VDAC | Voltage-dependent anion channel |
| | V/Val | Valine |
| | VdW | Van der Waals |
| W | W/Trp | Tryptophan |
| | WHAM | Weighted Histogram Analysis Method |
| | WT | Wild Type |
| Y | Y/Tyr | Tyrosine |

List of publications related to this work

D. Liu*, G. Richardson*, F. M. Benli *, C. Park, J. V. de Souza, A. K. Bronowska, I. Spyridopoulos, *Inflammageing in the cardiovascular system: mechanisms, emerging targets, and novel therapeutic strategies.*, *Clin Sci*. 2020 Sep 18;134(17):2243-2262. doi: 10.1042/CS20191213.

D. Liu *, J. V. de Souza*, and A. K. Bronowska, *Structure, dynamics, and small molecule ligand recognition of human-specific CHRFAM7A (Dupa7) nicotinic receptor linked to neuropsychiatric disorders.*, *IJMS*, 2021, 22(11):5466

D. Liu, J. V. de Souza, A. K. Bronowska. *Conformational analysis and small molecule ligand recognition of the human-specific CHRFAM7A nicotinic receptor implicated in neuropsychiatric disorders*. ABSTRACTS OF PAPERS OF THE AMERICAN CHEMICAL SOCIETY. 1155 16TH ST, NW, WASHINGTON, DC 20036 USA: AMER CHEMICAL SOC, 2019, 258.

D. Liu *, O. M. Gray, R. Chen, J. V. de Souza, A. Ahmad, P. Banks, M. A. Carroll, K. Madden, A. K. Bronowska, *Structure-guided development and preclinical evaluation of MKL-08, a novel high-affinity A147T TSPO ligand, as a PET radiotracer of neuroinflammation* (to be submitted to *Curr Topics Med Chem*)

D. Liu *, J. V. de Souza*, K. Madden and A. K. Bronowska, *In silico assessment of the CHRFAM7A (Dupa7) nicotinic receptor “druggability”*. (to be submitted to *IJMS*)

*Equal contribution

Table of Contents

| | |
|---|--------------|
| Abstract | ii |
| Acknowledgements | vi |
| Abbreviations | vii |
| List of publications related to this work..... | xvii |
| Table of Contents | xviii |
| List of Figures | xxiv |
| List of Tables | xxix |
| Thesis structure..... | 1 |
| Chapter 1 Introduction to neuro-immune interactions..... | 3 |
| 1.1 Inflammageing, neuroinflammation and neurodegeneration | 3 |
| 1.2 Biomarkers of neuroinflammation | 6 |
| 1.3 Cholinergic receptors in neuroinflammation and neurodegeneration | 7 |
| 1.4 TSPO in neuroinflammation..... | 7 |
| 1.5 Cholinergic transmission and acetylcholine receptors in neuroinflammation..... | 8 |
| 1.6 Nicotinic receptors and inflammageing | 14 |
| 1.7 Inflammation-driven cardiovascular dysregulation and cardiovascular cholinergic system in the regulation of immune function..... | 15 |
| Chapter 2 Structural biology and pharmacology of nicotinic receptors | 21 |
| 2.1 Nicotinic acetylcholine receptors: subunit structure, diversity, and receptor specialisation | 21 |

| | |
|--|-----------|
| 2.1.1 Subunits | 23 |
| 2.1.2 Structural biology of nicotinic receptors..... | 25 |
| 2.1.3 Agonist (orthosteric) binding sites..... | 27 |
| 2.1.4 Allosteric binding sites and allostery of nicotinic receptors | 29 |
| 2.1.5 Ligands of nicotinic receptors..... | 31 |
| 2.1.6 Channel gating..... | 31 |
| 2.1.7 Diversity | 32 |
| 2.2 $\alpha 7$ & dup $\alpha 7$ | 33 |
| 2.3 Nicotinic receptors in neurodegenerative disease: interactions with amyloid- β | 35 |
| 2.4 TSPO as inflammation biomarker to validate $\alpha 7$ results | 36 |
| 2.4.1 Structure | 36 |
| 2.4.2 Function..... | 38 |
| 2.5 TSPO and nicotinic receptors as targets for drugs and imaging agents | 39 |
| 2.5.1 Nicotinic receptors as targets..... | 39 |
| 2.5.2 TSPO as a target | 41 |
| 2.6 Challenges in the development of ligands targeting nicotinic receptors and TSPO | 42 |
| Chapter 3 Theoretical background | 44 |
| 3.1 Computational methods for a drug discovery pipeline..... | 45 |
| 3.2 Homology modelling..... | 47 |
| 3.3 Molecular mechanics..... | 49 |
| 3.4 Force field..... | 50 |

| | |
|---|----|
| 3.5 Molecular dynamics | 53 |
| 3.5.1 Theory behind molecular dynamics | 53 |
| 3.5.2 Conditions in molecular dynamics simulation | 56 |
| 3.5.3 Stages of molecular dynamic simulation | 60 |
| 3.6 Coarse grain simulations | 61 |
| 3.6.1 Coarse-grain modelling | 62 |
| 3.6.2 Martini force field | 63 |
| 3.6.3 Elastic network modelling | 64 |
| 3.6.4 Coarse-grained protein simulation running steps | 66 |
| 3.7 Umbrella sampling | 67 |
| 3.8 Druggability assessment: FTMap solvent mapping | 69 |
| 3.9 Molecular docking | 70 |
| 3.10 Scoring function for SeeSAR - HYDE | 73 |
| 3.11 Protein-protein docking | 75 |
| 3.12 Methods for analysis of molecule dynamics | 76 |
| 3.12.1 Root-mean-square fluctuation | 76 |
| 3.12.2 Root-mean-square deviation | 77 |
| 3.12.3 Cluster analysis | 78 |
| 3.12.4 Principal component analysis | 79 |
| 3.12.5 Hydrogen bonds analysis | 81 |

Chapter 4 Structure and dynamics of human-specific CHR_{FAM7A} (Dup α 7) nicotinic

| | |
|---|------------|
| receptor linked to neuropsychiatric disorders | 83 |
| 4.1 Results..... | 83 |
| 4.1.1 Human-specific dup α 7 nicotinic receptor | 83 |
| 4.1.2 Stoichiometry studies for different combinations of dup α 7/ α 7 | 86 |
| 4.1.3 The effect of dup α 7 subunits on Ca ²⁺ conductance | 92 |
| 4.2 Discussion..... | 94 |
| 4.3 Materials and methods..... | 97 |
| 4.3.1 Molecular modelling of α 7 and dup α 7 receptors | 97 |
| 4.3.2 Coarse-grain molecular dynamics simulations..... | 97 |
| 4.3.3 Atomistic molecular dynamics (MD) simulations..... | 98 |
| 4.3.4 Umbrella sampling (US) simulations | 100 |
| Chapter 5 Mapping of putative allosteric binding sites of α7 and dupα7 nicotinic receptors | 101 |
| 5.1 Results..... | 102 |
| 5.1.1 The effect of dup α 7 subunits on macromolecular ligand binding | 102 |
| 5.1.2 Conformational changes in orthosteric (agonist) binding sites | 105 |
| 5.1.3 Prediction of allosteric binding sites using FTMap | 114 |
| 5.1.4 “Druggability” of the allosteric binding sites probed by UCSF DOCK6..... | 118 |
| 5.2 Discussion..... | 121 |
| 5.2.1 Dup α 7/ α 7 ratio affects the ligand binding and may be linked to the nicotinic translational gap | 121 |

| | |
|---|------------|
| 5.2.2 The effect of dup α 7 subunits on orthosteric binding sites..... | 122 |
| 5.2.3 The effects of dup α 7 subunits on allosteric binding sites..... | 123 |
| 5.3 Materials and Methods | 124 |
| 5.3.1 Protein preparation | 124 |
| 5.3.2. Small molecule preparation and the construction of the target-focused small molecular library | 125 |
| 5.3.3 Atomistic molecular dynamics (MD) simulations | 125 |
| 5.3.4 Binding site mapping by FTMap | 126 |
| 5.3.5 Molecular docking of A β 42 and α -Bungarotoxin (α -BTX) | 127 |
| 5.3.6 Molecular docking of small molecules | 127 |
| Chapter 6 Addressing “druggability” of different polymorphs of 18-kDa translocator protein (TSPO)..... | 128 |
| 6.1 Results..... | 129 |
| 6.1.1 Compound prioritisation | 130 |
| 6.1.2 The drug-like possibility of compounds..... | 134 |
| 6.1.3 Molecular dynamics studies of ligand-TSPO complexes..... | 136 |
| 6.2 Discussion..... | 142 |
| 6.3 Materials and Methods | 145 |
| 6.3.1 Homology modelling of human TSPO | 145 |
| 6.3.2 Molecular docking and structure-guided ligand design..... | 145 |
| 6.3.3 Atomistic molecular dynamics (MD) simulations..... | 146 |

| | |
|---|------------|
| Conclusions and future directions..... | 148 |
| Appendix..... | 152 |
| References..... | 188 |

List of Figures

| | |
|---|----|
| Figure 1: The immune factors contributing to ageing. | 4 |
| Figure 2: Role of nonneuronal cells in inflammation..... | 5 |
| Figure 3: Physiology of the cholinergic synapse..... | 9 |
| Figure 4: The signalling pathway of mAChRs..... | 11 |
| Figure 5: The structure and signalling pathway of mAChRs and nAChRs..... | 12 |
| Figure 6: Proposed mechanism of $\alpha 7$ -mediated inhibition of NLRP3 inflammasome. | 15 |
| Figure 7: Many different factors are associated with cardiovascular disease. | 16 |
| Figure 8: The comparison between the neuronal cholinergic system and the cardiac nonneuronal cholinergic system. | 18 |
| Figure 9: The relationship between immune cells and the myocardial cholinergic system. .. | 19 |
| Figure 10: The topology and a structure example of nAChR..... | 22 |
| Figure 11: The topology of a neuronal nAChR subunit and a structure example of nAChR. | 23 |
| Figure 12: Different types of AChRs pentamers. | 25 |
| Figure 13: Architecture of the $\alpha 3\beta 4$ receptor..... | 27 |
| Figure 14: Structure of $\alpha 7$ AChBP and location of the agonist binding site..... | 28 |
| Figure 15: The location of binding sites in different AChRs agonist binding sites is depicted as red triangles..... | 29 |
| Figure 16: Orthosteric and allosteric binding sites mapped at $\alpha 7$ -AChBP..... | 30 |
| Figure 17: Agonists of nicotinic receptors..... | 31 |
| Figure 18: A) Cross-section of the acetylcholine receptor TM domain. B) The closed-to-open | |

| | |
|---|----|
| transition for the Torpedo nicotinic receptor. | 32 |
| Figure 19: The human $\alpha 7$ nAChR, FAM7A & dup $\alpha 7$ nAChR..... | 34 |
| Figure 20: The structure of TSPO. | 37 |
| Figure 21: The mechanism of complex-forming by TSPO and other proteins on the mitochondrial membrane. | 38 |
| Figure 22: Drug development pipeline..... | 44 |
| Figure 23: Computer-aided drug design used in drug discovery. | 46 |
| Figure 24: The process of homology modelling..... | 48 |
| Figure 25: Four critical contributions to a molecular mechanics force field: bond stretching, angle bending, torsional terms, and non-bonded interactions. | 51 |
| Figure 26: The integration cycles of a molecular dynamics simulation..... | 55 |
| Figure 27: Molecular dynamics Velocity Verlet algorithm..... | 56 |
| Figure 28: The system model for MD simulation. | 58 |
| Figure 29: 3D representation of periodic boundary conditions. | 59 |
| Figure 30: Application range for different molecular modelling..... | 62 |
| Figure 31: The potential energy landscape of all-atom and coarse-grain simulations. | 63 |
| Figure 32: Coarse-grained representation of all-atom structure (lipid, protein, water) and all standard proteins residues..... | 64 |
| Figure 33: The elastic network model. | 65 |
| Figure 34: The steps of coarse-grained simulation and the follow-up by the all-atom MD simulation. | 66 |
| Figure 35: The representation of the PMF curve and its related sampled window. | 69 |

| | |
|---|----|
| Figure 36: Principles of FTMap. | 70 |
| Figure 37: Visualisation of the DOCK6 box in UCSF Chimera | 72 |
| Figure 38: Docking the ligands to the receptor | 73 |
| Figure 39: The example of SeeSAR analysis. | 75 |
| Figure 40: The example of $\alpha 7$ nAChR RMSF results..... | 77 |
| Figure 41: The example of two systems RMSD results. | 78 |
| Figure 42: Two clustered structures for 5-dup $\alpha 7$ nAChR 100ns MD trajectory..... | 79 |
| Figure 43: 2D Projection of PCA. A different colour separates each experiment result. | 81 |
| Figure 44: The conditions for H-bond exiting between two atoms. | 81 |
| Figure 45: The example of H-bond results. | 82 |
| Figure 46: Sequence alignment of $\alpha 7$ /dup $\alpha 7$ EC domains and its eight possible models..... | 85 |
| Figure 47: The extracellular EC domain conformation of $\alpha 7$ subunit (residues 1–180)..... | 86 |
| Figure 48: (A) RMSD results of eight complete transmembrane models during 1 μ s CG MD simulation. (B) Total potential energy vs. time results of eight complete transmembrane models during 1 μ s MD simulation. | 88 |
| Figure 49: (A) RMSD results of eight EC domain models during 100 ns MD simulation. (B) RMSD of eight full-length receptor models during 100 ns MD simulation. (C) Total potential energy vs. time of eight EC domains during 100 ns MD simulation. (D) Total potential energy vs. time of eight complete transmembrane structures during 100 ns MD simulation. | 89 |
| Figure 50: H-bond results of EC domain models and complete models. | 90 |
| Figure 51: PCA results for the EC domain / fully modelled of eight dup $\alpha 7$ models. | 91 |
| Figure 52: Conformational changes in the EC domain of the $\alpha 7$ / AB-Dup model during the | |

| | |
|---|-----|
| 100ns MD simulation. | 92 |
| Figure 53: The umbrella sampling result of eight models. | 93 |
| Figure 54: 3D structure of α -BTX and A β ₄₂ | 103 |
| Figure 55: Structure of α 7 nAChR and the position of the orthosteric binding site..... | 107 |
| Figure 56: A) RMSD and B) PCA results of four EC domain dimers during 100 ns MD simulation. | 109 |
| Figure 57: Molecular docking results of Ach binding to four distinct EC dimers. | 110 |
| Figure 58: Molecular docking results of Ach and Lob agonists and four arrangements of dup α 7 nicotinic receptor. | 111 |
| Figure 59: Mapping of α 7 nAChR binding sites via FTMap. | 115 |
| Figure 60: The binding site's results of four types of dup α 7 nAChR using FTMap. | 116 |
| Figure 61: An overview of the orthosteric pocket and four allosteric binding sites discovered for the canonical α 7 nAChR. | 118 |
| Figure 62: The binding sites reported by the blind docking for four dup α 7 nAChRs..... | 120 |
| Figure 63: The detected binding sites of A147T TSPO..... | 130 |
| Figure 64: The structure of PK11195, DPA714, MKL and MKD compounds. | 131 |
| Figure 65: Diagram of TSPO in complex with MKL (A) or MKD (B). | 132 |
| Figure 66: The predicted binding modes of four ligands described in this study in the A147T TSPO binding pocket..... | 134 |
| Figure 67: A) RMSD calculation of ligands fit A147T TSPO variant backbone in complexes during 100 ns MD simulation; B) Two complexes of A147T TSPO variant and MKL ligand, the configuration of blue complex sampled around 1 ns and the configuration of blue complex sampled around 60 ns. C) RMSD calculation of A147T TSPO variant backbones in complexes | |

| | |
|--|-----|
| during 100 ns MD simulation. | 137 |
| Figure 68: Principal component 2D projection calculation of four ligands fit A147T TSPO variant backbone in the last 50ns during 100 ns of the atomistic MD simulation. | 138 |
| Figure 69: Analysis of the residual flexibility of A147T TSPO. | 139 |
| Figure 70: H-bond formation and contacts between ligands and A147T TSPO detected after MD simulations. | 140 |
| Figure 71: The conformation of 2MGY (blue) and 2N02 (purple). | 143 |

List of Tables

| | |
|--|-----|
| Table 1: The classification of nAChR subunits. $\alpha 7$, which is a subject of this study, is highlighted in bold..... | 24 |
| Table 2: Selected small molecule ligands of nAChR | 40 |
| Table 3: Stages of a typical MD simulation | 60 |
| Table 4: Sampling results Ca^{2+} permeability energetics..... | 94 |
| Table 5: Binding affinity ranges of $\text{A}\beta_{42}$ or α -BTX to orthosteric binding sites. All affinity ranges were calculated in SeeSAR using HYDE. | 104 |
| Table 6: Ach ligand focus on docking with $\alpha 7$ and four dup α models | 112 |
| Table 7: Lob ligand focus on docking with $\alpha 7$ and four dup α models | 113 |
| Table 8: Properties of ligands detected using SwissADME..... | 135 |
| Table 9: H-bond and contacts between ligands and A147T TSPO..... | 140 |
| Table 10: Binding affinity of A147T TSPO and ligand calculated via SeeSAR | 141 |

Thesis structure

The first two chapters describe the pathogenic mechanisms of neuroinflammation, neurodegeneration and inflammation generation related to two different transmembrane receptors, the nicotinic acetylcholine receptor (nAChR) and the translocator protein (TSPO), which involve interactions between the immune system and the neural and non-neural systems. The subsequent chapter provides a detailed theoretical background to the computational methods used in this research. Finally, the results obtained for this research are then discussed in depth in the following three chapters

Chapter 1: Introduction to neuro-immune interactions

This chapter focuses on explaining and exemplifying the biological links between neuroinflammation, neurodegeneration and inflammatory diseases and the immune system, the cholinergic system, and their physiologies and pathologies are discussed in depth. Moreover, it describes the critical role of cholinergic receptors and transporter proteins play within a cellular system. This description is used to show how nAChR and TSPO are essential targets for clinical treatment.

Chapter 2: Structural biology and pharmacology of nicotinic receptors

This chapter provides a comprehensive overview of the biological properties of the nAChR receptor family and TSPO, such as structure, function, and expression profiles. Their corresponding agonists and antagonists are also analysed. Finally, we discussed therapeutical procedures which use these receptors as targets and the stages of their development.

Chapter 3: Theoretical background

Chapter 3 describes the theory behind the computational methods used for this work: homology modelling, molecular dynamics simulations, umbrella sampling, molecular docking, and data analysis.

Chapter 4: Structure and dynamics of human-specific CHRFAM7A (dup α 7) nicotinic receptor linked to neuropsychiatric disorder

The focus of this chapter is describing the most likely $\alpha 7$ / $\text{dup}\alpha 7$ arrangements and stoichiometry at the atomistic level of detail. The eight homology models were tested by molecular dynamics simulations (all-atom and coarse-grained) and umbrella sampling simulation. The results provide the link between the structure, dynamics, and function of human-specific $\text{dup}\alpha 7$ receptors.

Chapter 5: Mapping of putative allosteric binding sites of $\alpha 7$ and $\text{dup}\alpha 7$ nicotinic receptors

This chapter focuses on three aspects: 1) How such $\text{dup}\alpha 7/\alpha 7$ interfaces or $\text{dup}\alpha 7/\text{dup}\alpha 7$ interfaces would interact with α -bungarotoxin (α -BTX) and amyloid $A\beta_{42}$; 2) Identification of orthosteric and the druggable allosteric binding sites; 3) The effect of $\text{dup}\alpha 7$ on binding affinity and binding sites generation detected by molecular docking calculations with several known agonists used (acetylcholine (Ach), nicotine (Nic), lobeline (Lob) and ligand library).

Chapter 6: Addressing “druggability” of different polymorphs of 18-kDa translocator protein (TSPO)

Chapter 6 focuses on assessing the druggability and affinity of small ligands developed in Newcastle University (MKL, MKD) to TSPO compared to two established TSPO radioligands (PK11195, DPA714). Their difference was studied and rationalised to develop future tracers for neurological and immune disorders.

Chapter 7: Conclusion and future directions

This chapter concludes the projects, describes how this work addressed the aims, discusses the limitations and implications and significance of this work findings, and the plan for further research.

Chapter 1 Introduction to neuro-immune interactions

The nervous system plays a crucial role in the maintenance of immune homeostasis. Both systems interact with each other and modulate one another. These interactions are established and maintained primarily by the acetylcholine receptor-mediated activation of intracellular signalling pathways and by the expression of the nonneuronal cholinergic system by immune cells¹⁻¹³. The cholinergic system mainly regulates adaptive and innate immune cells through $\alpha 7$ nicotinic acetylcholine receptors ($\alpha 7$ nAChRs)^{3-9, 14}. In addition, the 18 kDa mitochondrial translocator protein (TSPO), which is involved in the synthesis and transport of cholesterol, modulates neuroinflammation. This chapter summarises several important roles exerted by $\alpha 7$ receptors and TSPO in neuroinflammation, neurodegeneration, and inflammageing.

1.1 Inflammageing, neuroinflammation and neurodegeneration

With the extension of life, the incidence rate of many age-associated chronic diseases increases in current society. Understanding the pathogenesis of these disorders and providing effective treatment has become a public health priority.

Inflammageing is a classical chronic inflammation related to advanced age. It is described as a low grade, chronic, controlled, asymptomatic and systemic inflammatory state¹⁵. Inflammageing is a complex issue involving the nervous system, the local vascular system, the immune system, and molecular mediators^{1,2}. It is characterised by a complicated balance between pro-inflammatory and anti-inflammatory responses¹⁶. Although the underlying causes of inflammageing are unknown, dysregulation and impairment of the immune system are certainly involved (Figure 1). As ageing persists, the effectiveness of the individual's adaptive immune system declines, while the activity of the innate immune system increases, which affects the growth of the number of immune cells (such as natural killer cells, macrophages, lymphocytes), and the production of pro-inflammatory cytokines, especially interleukin 6 (IL-6), tumour necrosis factor α (TNF- α), and C-reactive protein (CRP)¹⁶⁻¹⁸. These molecules are likely to damage cells, eventually leading to the development of chronic inflammation¹⁶⁻¹⁸.

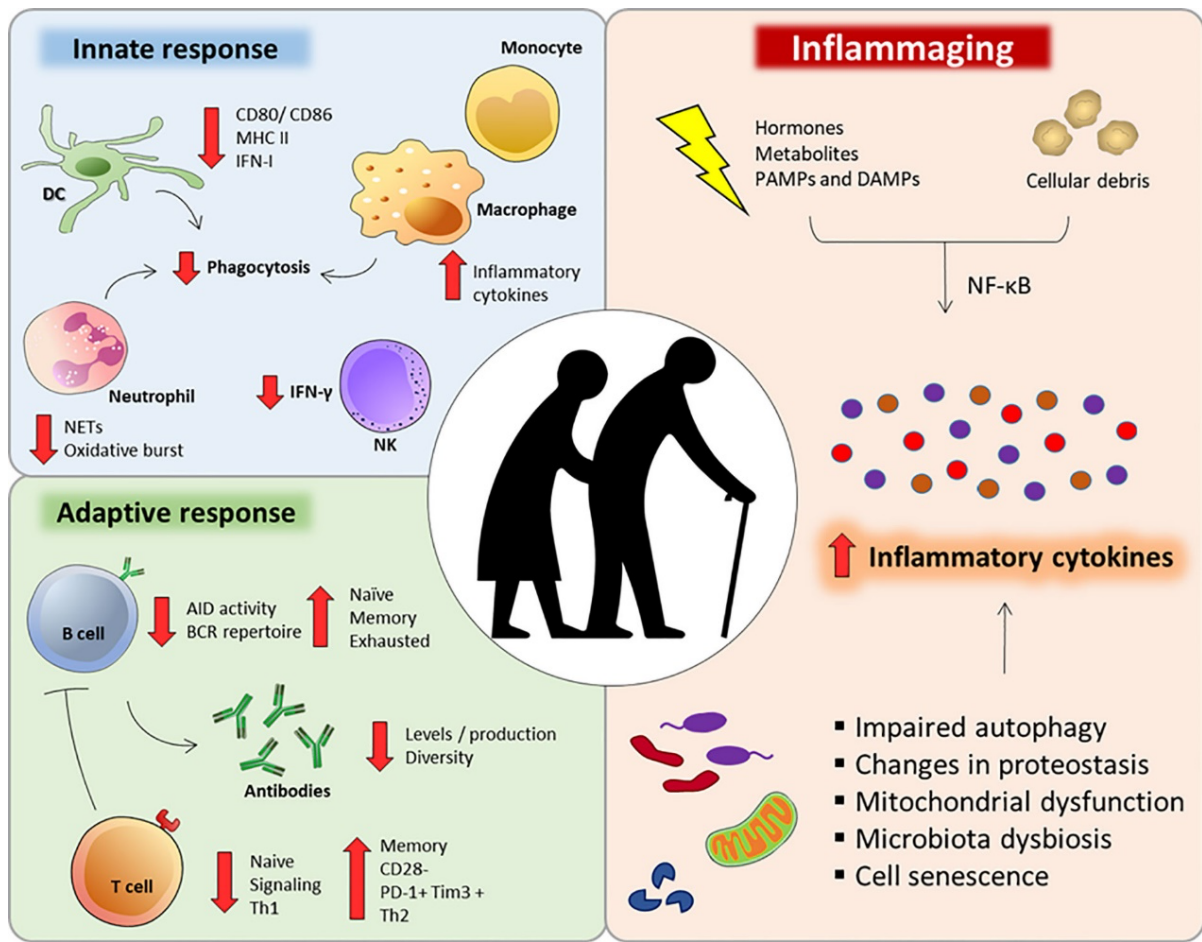


Figure 1: The immune factors contributing to ageing. Ageing impedes both innate and adaptive immune systems, dysregulating the release of inflammatory cytokines and chemokines, leading to changes in proteostasis, mitochondrial and autophagic dysregulation, and cellular senescence. Extracted from ¹⁹.

An increasing number of studies show that inflammaging accelerates the biological ageing process and worsens many age-related diseases such as Alzheimer's disease (AD)²⁰, atherosclerosis²¹ and heart disease²². Inflammaging is coupled to two related pathophysiological inflammatory processes relevant to the scope of this thesis: neuroinflammation and neurodegeneration.

When inflammation occurs in the central nervous system (CNS), it is also known as neuroinflammation. Chronic inflammation occurs due to the continued activation of glial cells and the recruitment of other immune cells to the brain (Figure 2). Apart from ageing¹², many other factors contribute to chronic neuroinflammation, involving autoimmunity²³ and physical injuries: traumatic brain injury²⁴ and spinal cord injury²⁵.

From the findings of Gehrman's group, microglia are the primary innate immune cells of CNS. Microglia migrate to the site of infection/injury and participate in the immune response by altering morphology and secreting cytokines, chemokines, prostaglandins and reactive oxygen species, which destroy pathogens and remove damaged cells²⁶. In addition, they contribute to the resolution of the inflammatory response by producing anti-inflammatory cytokines²⁶. However, with the continued neuroglial cell activation and the production of inflammatory cytokines increase, the blood-brain barrier (BBB) as a protector of the CNS is likely to be damaged, resulting in peripheral immune cells (such as macrophages and lymphocytes) being available to migrate to the brain, thereby perpetuating the immune response. This perpetuation exacerbates the inflammatory environment in the brain, ultimately leading to chronic neuroinflammation and neurodegeneration^{23, 27}. Hence, many studies show that neuroinflammation is closely linked to neurodegenerative disorders.

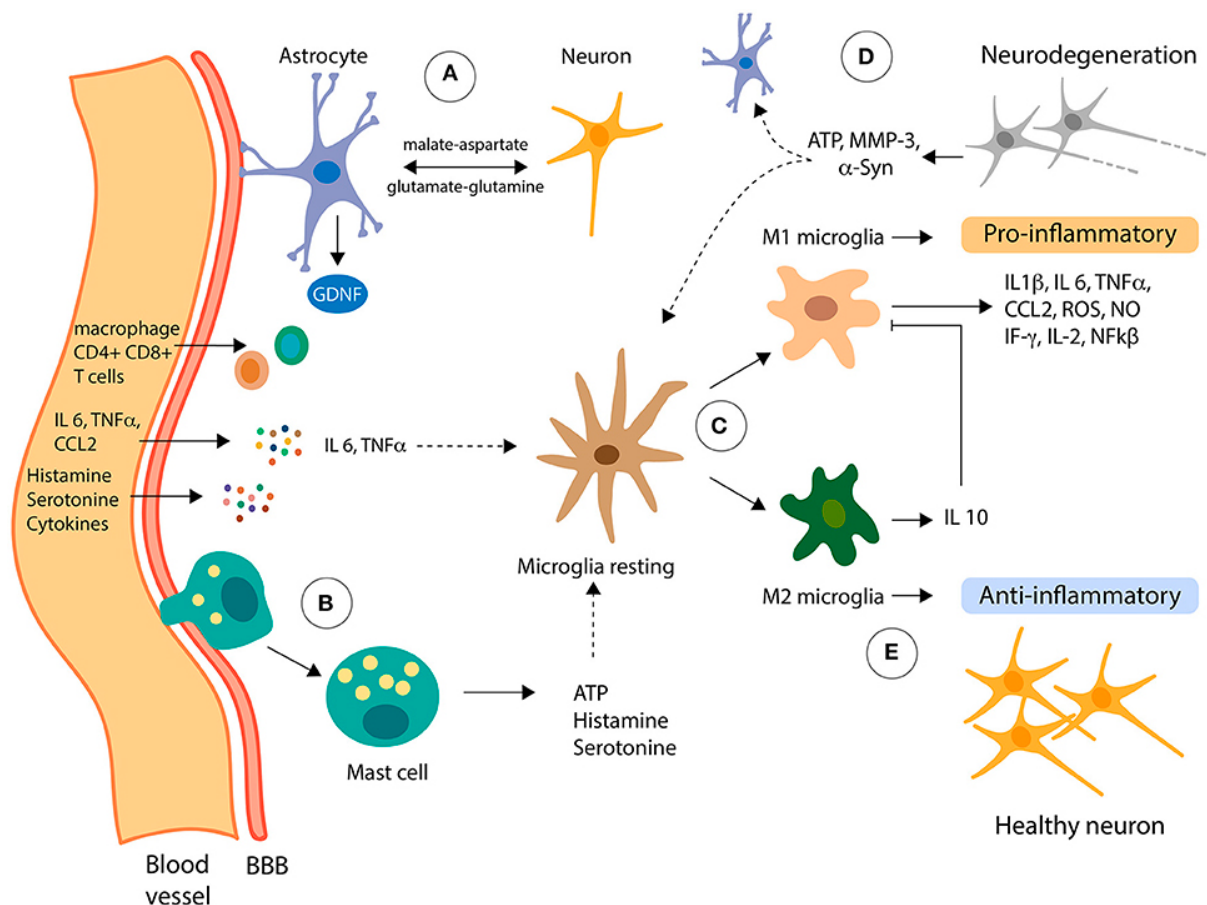


Figure 2: Role of nonneuronal cells in inflammation. (A) Astrocytes are the most abundant cells in the central nervous system. They receive signals from the periphery and within the

central nervous system and provide metabolic support to neurons. (B) Mast cells can induce changes in microglia through the delivery of pro-inflammatory effectors. (C) Microglia are usually activated in two classical phenotypes, M1 and M2. (D) LPS and IFN γ can polarize microglia to the M1 phenotype to secrete pro-inflammatory cytokines, which accelerates neurodegeneration. In addition, neuronal failure can release molecules such as α -Syn, ATP and MMP-3, which cross-signal with astrocytes and microglia, resulting in increased toxicity of neuroinflammation. (E) IL 4 and IL 13 induce microglia activation to the M2 phenotype, which contributes to anti-inflammation in the central nervous system by releasing IL 10 cytokines and downregulating M1 function. Extracted from ²⁸.

Neurodegeneration is the progressive degeneration of neuronal structure and function. It causes a significant decline in a wide range of cognitive performance, including memory and information processing speed. Many neurodegenerative diseases occur as a result of neurodegenerative processes, and the most specific diseases are Alzheimer's disease (AD), Parkinson's disease (PD) and multiple sclerosis (MS)²⁹.

1.2 Biomarkers of neuroinflammation

Since there is an established link between neuroinflammation and neurodegenerative diseases, it is essential to investigate the mechanisms of neuroinflammation to identify effective treatments for such diseases. However, the capability to probe the inflammatory biochemical processes directly in the brain is a severe limitation. Therefore, the study of neuroinflammatory biomarkers is an essential aim of brain pathology research.

As is known from many studies, neuroinflammatory diseases involve glial cells and immune cells and other inflammatory mediators, such as pro-inflammatory cytokines or reactive oxygen species. Some researchers suggest that these molecules released into the blood and cerebrospinal fluid in response to brain inflammation can be used as neuroinflammatory markers to help elucidate mechanisms, diagnose disease, monitor treatment and demonstrate target engagement in clinical trials³⁰. Janelidze's and Piero's studies measured cerebrospinal fluid (CSF) biomarkers such as YKL-40, ICAM-1, CHIT1 and Flt-1 to establish the relationship between those biomarkers and the levels of A β and tau aggregated in AD patient brain^{31, 32}. The reason to select these specific biomarkers was their association with the activation of microglia and astrocytes and neuroinflammation. These studies indicated that higher levels of those

specific biomarkers were associated with an increased risk of developing AD^{31, 32}. This result may be helpful to predict and monitor AD.

In addition, some studies have shown that low levels of A β ₄₂^{33, 34} or high levels of phosphorylated tau proteins³⁵ or alterations of neuronal nAChR³⁶ are present in the brains of patients with neuroinflammatory diseases, these of which could also be selected as essential biomarkers of neuroinflammation. Furthermore, many studies and clinical data have identified other proteins such as TSPO as being involved in several physiological processes and closely related to neuroinflammation. As TSPO expression is upregulated in response to injury or degenerative disease, its expression can be used as a sensitive biomarkers to study the mechanisms of neurodegenerative diseases^{37,38}. Moreover, most TSPO ligands were developed as neuroimaging agents and diagnostic tools for brain inflammation.³⁹

1.3 Cholinergic receptors in neuroinflammation and neurodegeneration

In recent years, an increasingly strong connection has been found between chronic neuroinflammation and impaired memory, especially in AD. Neuronal cholinergic receptors maintain normal cognitive function and regulate cellular signalling, and central cholinergic neuronal degeneration is implicated in impaired learning, memory, sleep regulation, and attention^{12, 40, 41}. Gamage has described how stimulation of $\alpha 7$ nACh receptors is neuroprotective by decreasing neuroinflammation via decreasing the release of pro-inflammatory cytokines⁴². There is also evidence for astroglial $\alpha 7$ nACh receptor stimulation mediating anti-inflammatory and antioxidant effects^{6, 40, 43}. This evidence shows that targeting $\alpha 7$ nACh receptors could diminish neuroinflammation, which is relevant to treating several neurodegenerative diseases mentioned before.

1.4 TSPO in neuroinflammation

Apart from cholinergic receptors, the translocator proteins (TSPOs) are another transmembrane protein indicated to be associated with neuroinflammation. TSPOs have been found on the outer mitochondrial membrane and expressed in the microglia cells in the brain³⁹. In normal conditions, the levels of TSPO are low in the nervous system but are significantly raised in

microglia and astrocytes in response to neurodegenerative diseases, such as AD⁴⁴, multiple sclerosis⁴⁵ and PD⁴⁶. In addition, Ji and Maeda indicated that the expression of TSPO in microglia is associated with substantial neuronal loss, whereas its expression in astrocytes is associated with reduced neuronal damage.⁴⁷ This will be discussed with more focus in TSPO in Chapter 2.5.

1.5 Cholinergic transmission and acetylcholine receptors in neuroinflammation

The cholinergic system has been implicated in emotion, memory, cognition, attention control, movement, immune function, and anti-inflammatory and neuroinflammation¹¹⁻¹³. All cholinergic neurons involved in many brain functions and neurodegenerative disorders utilise acetylcholine (ACh) as the principal neurotransmitter⁴⁸. The mechanism of cholinergic transmission is shown in Figure 3. ACh is synthesised from choline and acetyl coenzyme A (Ac-CoA) by the cytosolic enzyme choline acetyltransferase (ChAT) and transported to storage vesicles via the vesicular acetylcholine transporter (VACHT). As the cholinergic neuron gets stimulated by action potentials and the intracellular concentration of free calcium ions rises, the stored ACh is released from the nerve terminals into the synaptic gap via exocytosis. Released ACh subsequently binds to pre and post-synaptic acetylcholine receptors (AChRs), which triggers various biological effects. The ending of the ACh function is via the hydrolysis by acetylcholinesterase (AChE) or butyrylcholinesterase (BuChE) to generate free choline. Some of this choline is transported back to the cholinergic nerve terminals via the choline transporter 1 (CHT1) as a substrate for synthesising new ACh^{41, 48, 49}.

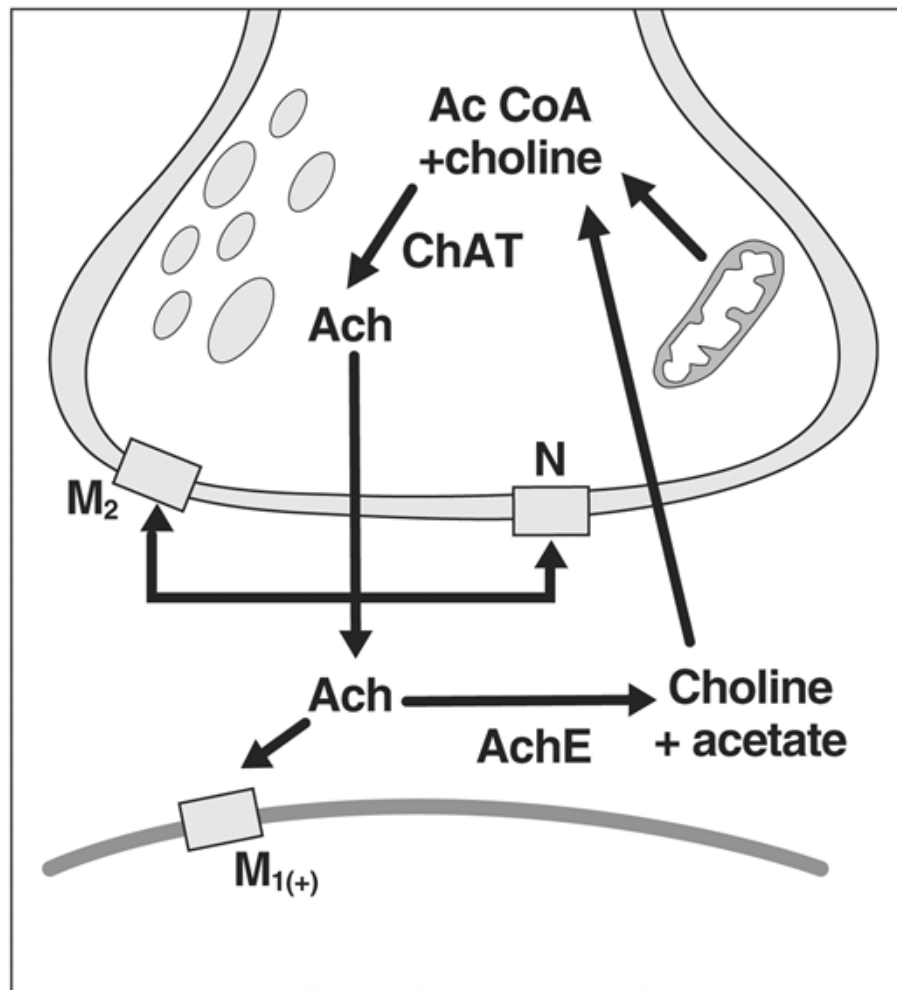


Figure 3: Physiology of the cholinergic synapse. Choline and Ac CoA are the critical substrates for the synthesis of ACh catalysed by ChAT. After ACh is released into the synapse and binds to presynaptic AChR receptors (muscarinic – M and nicotinic – N), the activated postsynaptic receptor (M1) achieves transmitting a signal from one neuron to the other. Then ACh is hydrolysed to choline and acetate by AChE in the synaptic cleft. Some of the choline is returned by the uptake mechanism and recycled by the neuron. Extracted from ⁴⁰.

Numerous studies have demonstrated that there is a strong link between neuroinflammation and abnormalities in the cholinergic system. In the brain of AD and PD patients, some groups found the loss of ChAT activity and decreased cholinergic neurons⁵⁰. These phenomena are thought to be associated with cognitive impairment. In addition, the activity of AChE in typical neurons has also been found to decrease, whereas the activity of AChE is displaced from neurons within the senile plaques (SP) and neurofibrillary tangles (NFT) increase, which promote to hydrolyse ACh and cause the activity of ACh decrease. The degree of reduction in ACh is positively correlated with the severity of dementia⁵⁰. In MS brain, ChAT activity is decreased whereas

both AChE and BuChE activity increase, and lower Ach levels^{51, 52}. The cholinergic agonists, antagonists, allosteric modulators and other cholinergic drugs (such as AChE inhibitor) have been shown to alter memory and cognitive abilities in humans to provide treatment therapies for neurodegenerative diseases⁵³.

Acetylcholine receptors also play an essential role to achieve rapid transmission in the cholinergic system¹². There are two types of acetylcholine receptors. One is muscarinic acetylcholine receptors (mAChRs, M-type) involved in the metabotropic cholinergic system via activation of metabotropic cholinergic system⁵⁴. The other one is nicotine acetylcholine receptors (nAChRs, N-type), which is involved in the ionotropic cholinergic system (ligand-gated ion channels)⁵⁵.

Muscarinic receptors belong to a class of G protein-coupled receptors (GPCRs) that have seven transmembrane regions⁵⁴. There are five distinct subtypes, denoted as M1-5, respectively. Different types of mAChR are coupled to different G proteins. For example, the M1, M3 and M5 receptors are coupled to G α_q protein, while M2 and M4 are coupled with G α_i/o protein⁵⁶. As Figure 4 shows, mAChRs have been located both pre and postsynaptically throughout the brain. M1 subtypes have been found the most abundantly expressed in CNS, followed by M2 and M4 subtypes. M3 and M5 are the least expressed⁴⁸. Different mAChRs can modulate the activity of a range of phospholipases, ion channels, protein kinases and other signalling molecules. The function of the muscarinic receptors is shown in Figure 4, wherein it may activate ion channels and the secondary messengers via intracellular G protein and other enzymes, respectively, leading to a series of intracellular events^{1, 2}.

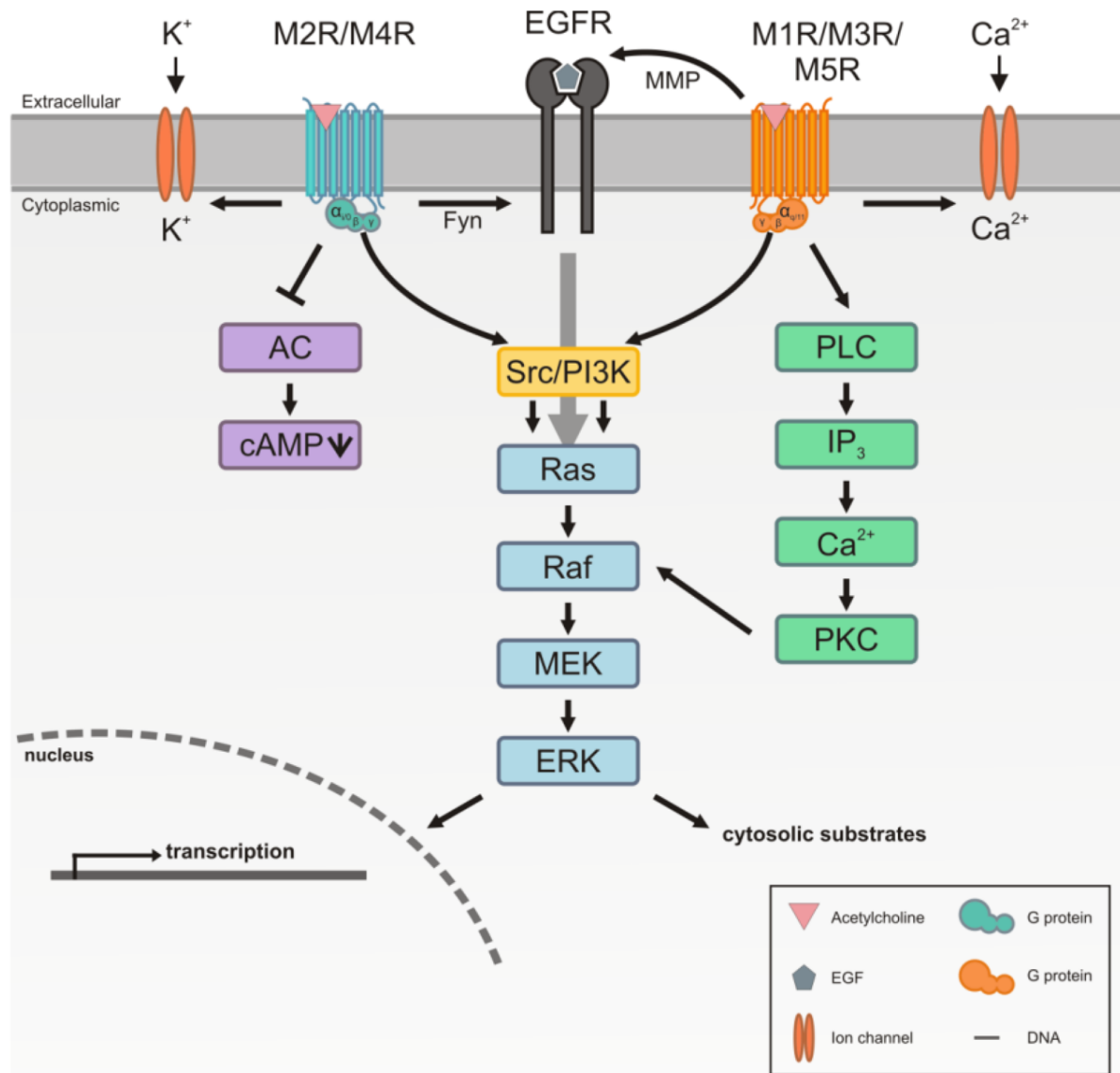


Figure 4: The signalling pathway of mACHRs. Each mACHRs subtype is a seven-transmembrane receptor. The M1, M3, and M5 mACHRs preferentially couple to the Gq/G11-type G-proteins to stimulate the phospholipase C (PLC β), which results in the release of inositide phosphate (IP₃) and diacylglycerol (DAG), leading to increased intracellular Ca²⁺ concentration. The M2 and M4 mACHRs selectively activate Gi/Go-type G-proteins, thereby negatively modulating adenylyl cyclase (AC), reducing intracellular concentration of cyclic AMP (cAMP), and prolonging K⁺ channel opening. Extracted from ².

It is well documented that mACHRs also activate diverse signal transduction pathways^{2, 5, 57} and mediate the production and synthesis of proinflammatory cytokines^{50, 58, 59} (Figure 5). Historically, since the discovery of T-lymphocyte cytotoxicity via the muscarinic cholinergic system and a cholinergic anti-inflammatory pathway emerged⁶⁰, many researchers suspected

the relationship between mAChR and neuroinflammation. In the brain of AD patients, the lower expression of M2 subtypes of mAChR and dysfunctional M1 mAChR have been found⁴⁰. It has been shown that M1 mAChR knockout mice are severely impaired in working memory and memory⁶¹. These experiments confirm that mAChR abnormalities may contribute to neuroinflammation. M1-specific agonists were administered to several animal models of amnesia. This result showed that they promoted the induction of LTP and improved cognitive function⁶². Therefore, mAChR could be considered as a target for therapies of diseases with underlying neuroinflammation.

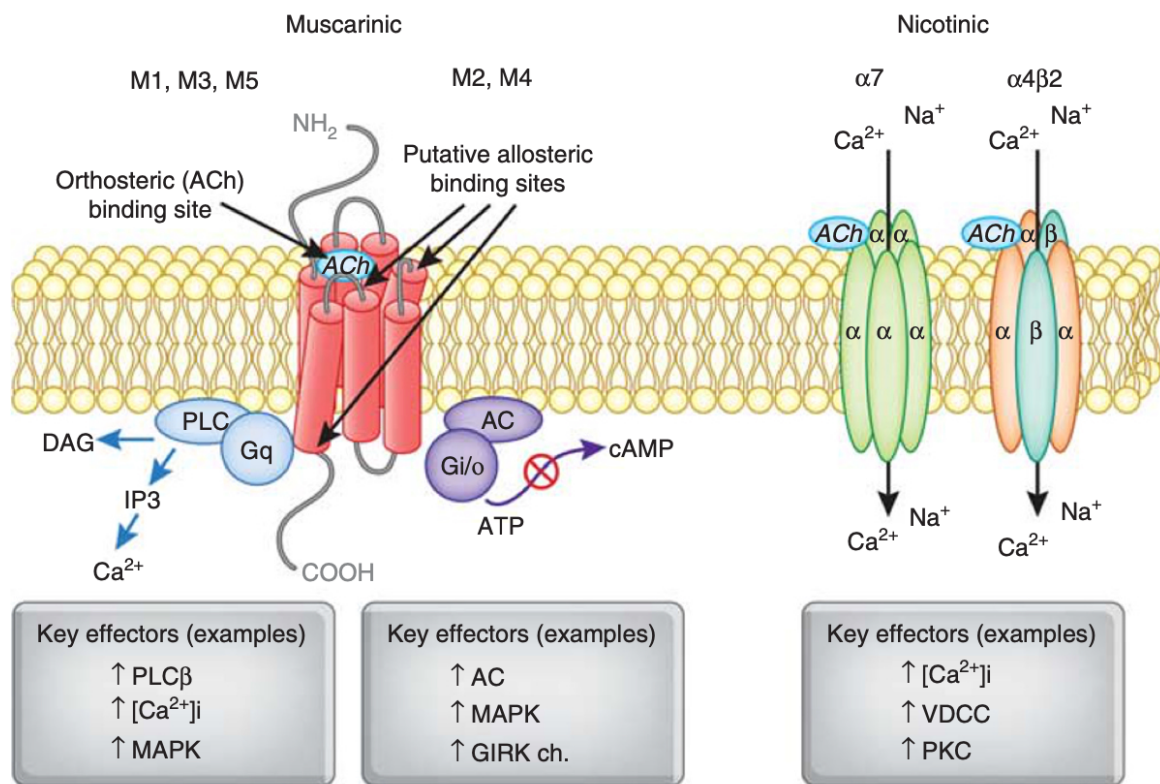


Figure 5: The structure and signalling pathway of mAChRs and nAChRs. The function of mAChRs is in the left and nAChRs in the right. The most abundant neuronal nAChRs types are the heteromeric $\alpha 4 \beta 2$ receptor and homomeric $\alpha 7$ receptor. The action of both these two types of nAChRs can increase the release of Ca^{2+} , provide a broader control of synaptic plasticity and neurotransmitter release, as well as gene transcription. PLC β , phospholipase C β ; DAG, 1,2-diacylglycerol; IP₃, inositol-1,4,5-trisphosphate; AC, adenylate cyclase; GIRK, G-protein-activated inwardly rectifying potassium channel; MAPK, mitogen-activated protein kinase; cAMP, cyclic adenosine monophosphate; ATP, adenosine triphosphate; VDCC, voltage-gated calcium channels; PKC, protein kinase C. Extracted from ¹.

Nicotinic acetylcholine receptors are pentameric ion channels embedded in the cell membrane, relying on a ligand-gated ion channel (LGICs) mechanism for signalling. As shown in Figure 5, when binding with a chemical messenger (such as Ach), the conformation of the receptor changes and causes the central channel pore to open. The pore allows positively charged ions (Ca^{2+} , K^{+} , Na^{+}) to move across it. In particular, Na^{+} enters and K^{+} exits, following the electrochemical gradient^{1,9}.

Similarly, activation of $\alpha 7$ nAChR can also mediate the release of neurotransmitters such as glutamic acid (Glu), dopamine. Published data point towards nicotinic receptors, especially $\alpha 7$ nAChR, mediating cholinergic regulation of inflammation in the brain. For example, this regulation goes through the Janus kinase (JAK) / signal transduction and activator of transcription (STAT), phosphatidylinositol-3-kinase (PI3K) / protein kinase B (PKB) and nuclear factor NF- κ B signalling pathways. Furthermore, $\alpha 7$ nAChRs also could promote a microglial transition to M2 type, inhibit the pro-inflammatory factors (such as IL-1, IL-1 β , IL-6) and tumour necrosis factor (TNF- α), promote the expression of the anti-inflammatory factors IL-4 and IL-10, reduce the inflammatory response induced by LPS and produce a cerebroprotective effect³⁻⁹.

Lower expression of nAChR in the brain affects normal function, ultimately leading to neurodegenerative diseases, which has been confirmed by several studies that showed the expression of $\alpha 4\beta 2$ and $\alpha 7$ nAChR decreased in the brains of AD, PD and MS patients^{5, 63, 64}. The interaction of $\alpha 7$ nAChR and beta-amyloid (A β) has also been demonstrated involving the generation of AD³. This interaction will be described in Chapter 2. Notably, some studies suggested that $\alpha 7$ nAChRs regulate the calcium transmission and release of the Ach, which plays a vital role in cognition and memory. Besides, $\alpha 7$ nAChRs and $\alpha 4$ nAChRs expression has been found markedly decreased in the AD brain⁶⁵.

Furthermore, $\alpha 7$ nAChR was also found to have the ability to control the number of dendritic cells and T-cell accumulation in the immune system in MS patients. This claim was further confirmed in experiments where T cell differentiation and responses were suppressed through exposure to the nAChR agonist nicotine⁶⁶. As a result, nAChR has become an essential target

for investigating the treatments of neuroinflammatory and neurodegenerative diseases. Many nAChR agonists and antagonists have successfully relieved and treated AD and PD in preclinical models and patients. For example, one of the most famous drugs to treat mild to moderate AD is Galantamine. It is used to increase the utilization of acetylcholine by inhibiting acetylcholinesterase hydrolysis and active nAChR from releasing acetylcholine when Galantamine binds to the allosteric binding sites nAChR⁶⁷.

1.6 Nicotinic receptors and inflammageing

The previous sections discussed how nicotinic receptors could regulate the immune system, which causes the generation of inflammation in the nervous system or the cardiovascular system. It can be found expressed in both the nervous and non-nervous systems *in vivo*, where it interacts with various proteins, affecting a wide range of physiological activities⁵⁵.

Mounting evidence suggests that $\alpha 7$ receptor signalling can indirectly inhibit NLRP3 inflammatory activity by preventing mitochondrial DNA (which is NLRP3 ligand) release and reducing mitochondrial damage⁶⁸. By the studies *in vitro* and *in vivo*, several groups indicated that Ach and nicotinic receptor agonists (PNU282987) showed significant inhibition for the NLRP3 inflammasome activation^{68, 69}. (Figure 6) Ke and Shao's⁶⁹ study indicated that the activating $\alpha 7$ receptor by PNU282987 can block the NLRP3 interacted with β -arrestin-1 and thus reduced the production of inflammatory cytokines (IL-1 β and IL-18), whereas deletion of $\alpha 7$ receptors enhances NLRP3 inflammasome activation via $\alpha 7$ knockout mice studies.

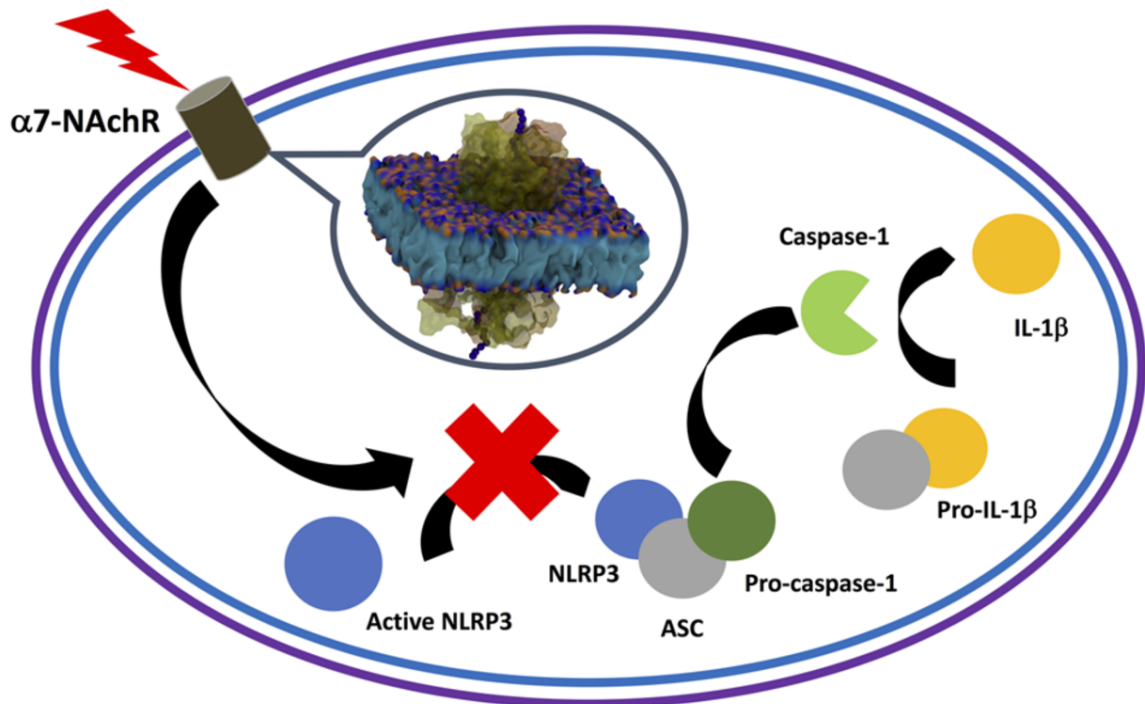


Figure 6: Proposed mechanism of $\alpha 7$ -mediated inhibition of NLRP3 inflammasome. Activation of $\alpha 7$ receptors (represented as a red bolt), either by neurotransmitter (ACh), agonist, or positive allosteric modulator, inhibits the formation of NLRP3 inflammasome. The inset shows the molecular structure of human $\alpha 7$; the pentamer is rendered as the molecular surface and coloured ochre. The lipid bilayer is coloured blue (fatty acid tails), purple and fuchsia (lipid headgroups). The flux of ions occurring upon the activation is represented as blue spheres (the central pore). Made by author⁷⁰.

Nicotinic receptors also interact with G protein and amyloids, particularly β -amyloid. More details on these interactions are covered in Chapter 2.

1.7 Inflammation-driven cardiovascular dysregulation and cardiovascular cholinergic system in the regulation of immune function

From the introduction in the previous section, it is clear that chronic inflammation is coupled with the ageing process. Chronic inflammation occurs in the CNS, leading to neurodegenerative disease and coronary artery disease, affecting the heart's function and causes cardiovascular disease (CVD), such as artery disease, aortic aneurysm, and cerebrovascular events. The main events in CVD are atherosclerosis and heart failure (HF)⁷⁰, which are specific diseases associated with ageing. As the heart ages, it undergoes myocardial remodelling as expressed by

endothelial stenosis, vasomotor dysfunction and stiffness, cardiomyocyte hypertrophy and myocardial fibrosis, which leads to increased ventricular stiffness, impairment of cardiac function and ultimately, a variety of cardiac diseases^{71, 72}. With the analysis of anti-inflammation drugs, it is indicated that inflammation can be a causal factor in the development of the atherosclerotic process and the deterioration of HF (Figure 7).

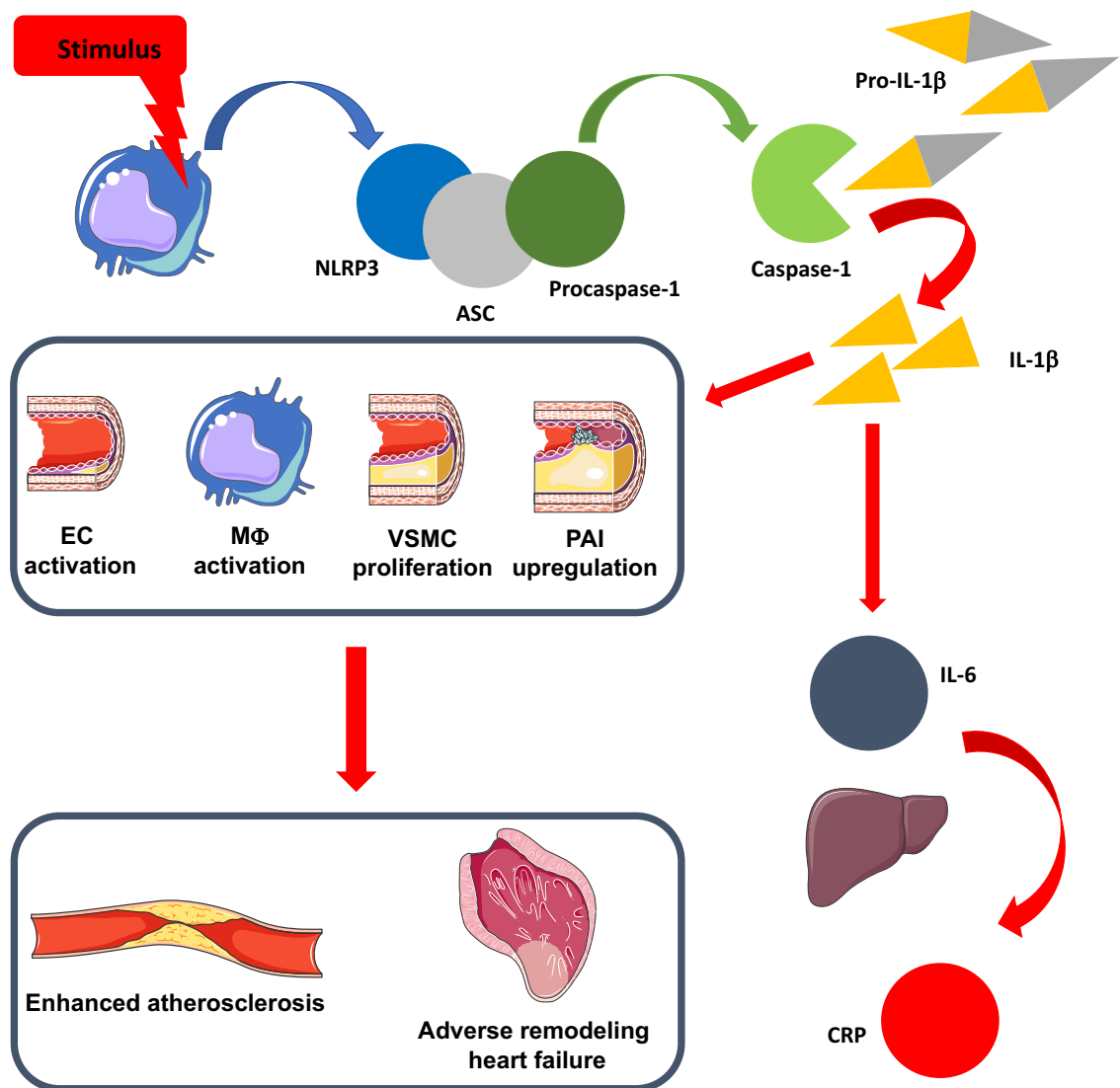


Figure 7: Many different factors are associated with cardiovascular disease. The local production of IL-1 β is triggered by the inflammatory stimulus (red bolt), which subsequently activates the NLRP3 inflammasome. The NLRP3 inflammasome consists of three major components: the sensor NLRP3 protein (blue), the adaptor-apoptosis-associated speck-like protein (ASC) (light grey), and the effector pro-caspase-1 (bottle green). The activated NLRP3 interacts with ASC, and pro-caspase-1 binds to ASC to assemble into a large cytosolic complex, which triggers the activation of caspase-1. Active caspase-1 (light green)

cleaves the pro-inflammatory IL-1 β (yellow) from its precursor to their biologically active forms, which triggers or enhances pathological events such as atherosclerosis and adverse cardiac remodelling, and – in consequence – HF. In parallel, IL-1 β induces IL-6 that stimulates C-reactive protein synthesis in the liver. IL-6 does not cause IL-1-related cardiovascular pathology but serves CRP serum levels as a surrogate marker for inflammation. Made by author⁷⁰.

Available literature data suggest that the cholinergic anti-inflammatory pathway (CAP) can moderate the inflammatory response in ischaemic cardiomyopathy (ICM)⁷³. Following excitation of CAP by injection of $\alpha 7$ nicotinic acetylcholine receptors agonists into rats, adenosine monophosphate-activated protein kinase (AMPK) signalling, widely thought to play a cardioprotective role in various cardiomyopathies, was activated, thereby activating ventricular remodelling, reducing inflammatory cytokines and maintaining the integrity of ischaemic cardiac ultrastructure⁷³.

Furthermore, the nonneuronal nicotinic $\alpha 7$ receptor, which plays an essential role in the cholinergic anti-inflammatory pathway, also has a cardioprotective effect⁷⁰. It can be achieved by $\alpha 7$ receptor-mediated immune cells via decrease the production of pro-inflammatory cytokine (IL-1 β , IL-6 and TNF- α). The experiments of Baez-Pagan's group showed that nicotinic $\alpha 7$ agonists (e.g. Ach, nicotine) revealed a reduction in the formation of pro-inflammatory cytokines⁷⁴. Moreover, the results from the studies by Xing and colleagues showed low-dose nicotine (nicotinic receptor agonist) promotes autophagy and accelerates autophagic flux in neonatal mouse cardiomyocytes (NMCM) through upregulation of heme oxygenase-1 (HO-1) while inhibiting NMCM apoptosis. These results provide the nicotinic $\alpha 7$ receptor as a new target for developing the treatment of cardiovascular disease.

The study by Ke and co-workers showed that activating $\alpha 7$ nAChR receptor leads to the NLRP3 inflammasome inhibition via regulation of β -arrestin-1 in the monocyte/microglia system⁷⁰. Another study showed that activation of $\alpha 7$ nAChR receptor inhibited NLRP3 inflammasome activation by preventing mitochondrial DNA release⁷⁰. These findings collectively reveal a cholinergic receptor-mediated anti-inflammatory pathway, and wherein $\alpha 7$ nAChR receptors have blocked NLRP3 activation. Hence, this data highlights the importance of $\alpha 7$ nAChR receptors as cardiovascular biomarkers and indicates these receptors as potential drug targets

for treating cardiovascular disorders.

The cardiovascular cholinergic system includes neuronal and nonneuronal types⁷⁵ (Figure 8). The neuronal cholinergic system releases neuronal ACh in the heart, consisting of preganglionic parasympathetic pathways, intracardiac parasympathetic ganglia and postganglionic parasympathetic neurons⁷⁵. Cardiomyocytes contribute to the nonneuronal cholinergic system (NNCS) and immune cells, which poses the machinery (ChAT, VAChT, AChE, nAChR, mAChR and CHT1) that can perform the entire process of synthesis, storage, reuptake and degradation of non-neuronal ACh independently^{75, 76}. As described in the previous sections, ACh signalling, controlled by the neuronal cholinergic system, can regulate systemic inflammatory responses. Results of several recent studies supported that non-neuronal ACh released from immune cells in the myocardium participate in local cardiac immune responses by binding with the nAChR via autocrine/paracrine action⁷⁷.

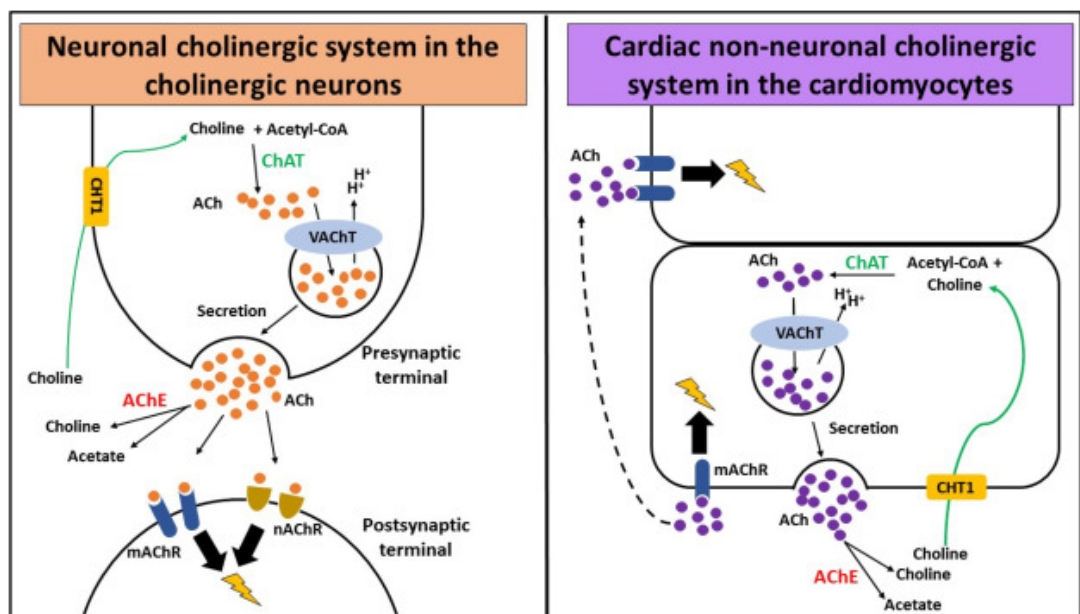


Figure 8: The comparison between the neuronal cholinergic system and the cardiac nonneuronal cholinergic system. This schematic illustrates the synthesis, storage, transmission, reuptake and degradation of ACh in cardiac myocytes and cholinergic neurons. In contrast to the neuronal cholinergic system, nonneuronal ACh released from cardiac myocytes binds to mAChRs in an autocrine/paracrine manner to mediate signalling. Extracted from ⁷⁶.

The ACh released by the cholinergic system plays a crucial role in modulating the immune function by activating the inflammatory reflex, also known as the cholinergic anti-inflammatory

pathway⁷⁸. This pathway has been shown to regulate the innate immune response to protect the heart against the localized inflammatory response via the $\alpha 7$ nAChR mediated¹⁰. The activation of CAP depends on the vagus nerve and the release of nonneuronal Ach¹⁰. Many studies suggested that the NNCS can regulate the function of immune cells in myocardial inflammation through this pathway (Figure 9). In the NNCS, the T cell-derived Ach activate $\alpha 7$ nAChR on macrophages by acting in an autocrine/paracrine manner and then inhibit the secretion of pro-inflammatory cytokine (TNF- α , IL-1 α and IL-6)¹⁴. Moreover, in rat models, injection with AChE inhibitors and $\alpha 7$ nAChR agonist PNU-282987 affect the mobilization of cardiac immune cells and downregulate the expression of the proinflammatory mediators⁷⁷. These pieces of evidence indicated the relationship between the cholinergic system and immune function.

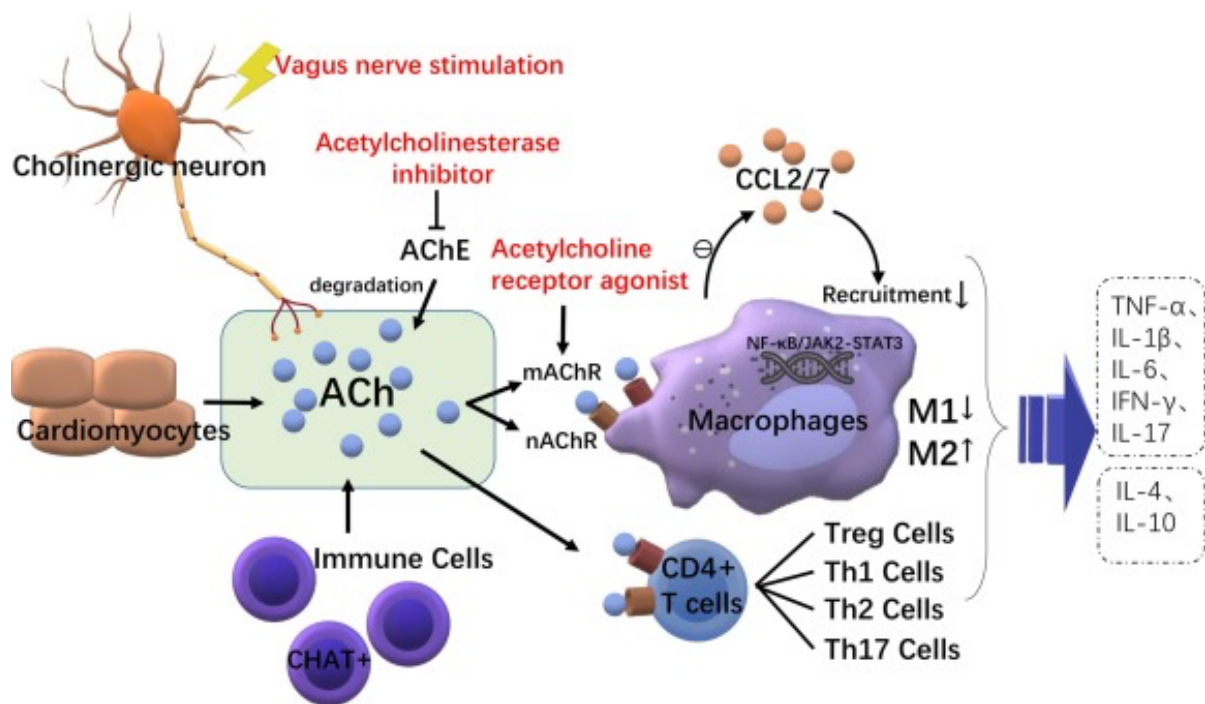


Figure 9: The relationship between immune cells and the myocardial cholinergic system. The function of immune cells in myocardial inflammation is regulated by the cholinergic system and is associated with many cardiovascular diseases. Acetylcholine, derived from the vagal secretory and nonneuronal cholinergic systems, acts on acetylcholine receptors on macrophages to reduce the chemokine CCL-2/7, thereby reducing the recruitment of inflammatory macrophages. In addition, the functional state of macrophages is changed from M1 to M2. Employing electrical activation of the vagus nerve, acetylcholinesterase inhibitors, and acetylcholine receptor agonists, the level of acetylcholine in the microenvironment can be increased, and acetylcholine receptors activated, which in turn modulates the immune system by affecting immune cells and reducing inflammatory factors, thus serving as a

therapeutic strategy for myocardial inflammation. Extracted from ⁷⁷.

More detail on this topic has been provided in the review paper entitled “Inflammageing in the cardiovascular system: mechanism, emerging targets, and novel therapeutic strategies”⁷⁰ that I have co-authored.

Chapter 2 Structural biology and pharmacology of nicotinic receptors

As outlined in Chapter 1, neurodegenerative diseases, such as AD and PD, are neurological disorders characterized by the progressive degeneration of the structure and function of neurons in the brain and spinal cord²⁹. Considering their crucial role in CNS and involvement in those diseases, nAChRs can be a valuable target for drug development. This chapter will focus on the specificity of nicotinic receptors, their structural and pharmacological diversity, and their interactors, which modulate their functional responses.

2.1 Nicotinic acetylcholine receptors: subunit structure, diversity, and receptor specialisation

Nicotinic acetylcholine receptors⁷⁹⁻⁸⁴ are ligand-gated ion channels, specifically sensitive to nicotine: these channels control Na⁺, K⁺, Ca²⁺ transport. They belong to the members of the ‘cysteine-loop’ family of ligand-gated ion channels (LGICs). This family also includes 5-hydroxytryptamine (5-HT₃), γ -aminobutyric acid (GABA_A) and glycine (GlyR) receptors^{79, 83}. In the extracellular domain of these family members, there is a disulfide bond build by two cysteines separated by 13 intervening amino acids⁸⁵⁻⁸⁸ (Figure 10). These receptors play an essential role in regulating rapid synaptic neurotransmission in the nervous and non-neuronal systems. In my study, the research is focused on the specific sub-family of nicotinic acetylcholine receptors denoted as subunit alpha-7 ($\alpha 7$).

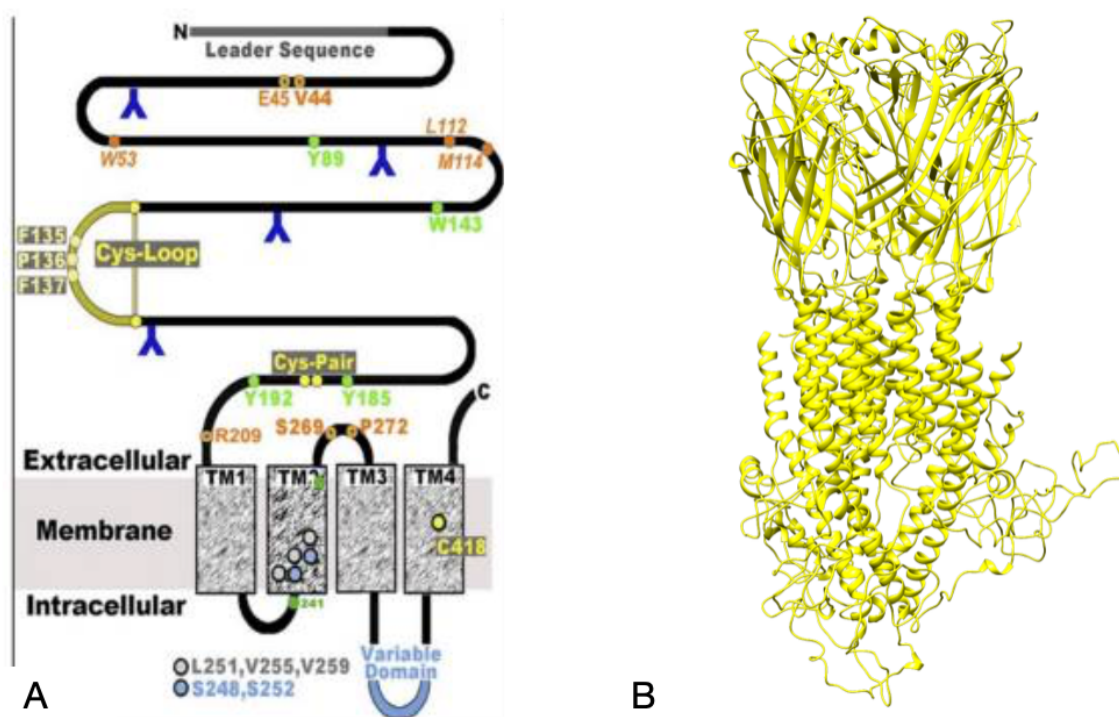


Figure 10: The topology and a structure example of nAChR. A) The topology of nAChRs. Cys-loop is one of the motifs of nAChRs, composed of two disulfide-linked cysteines separated by 13 amino acids. Subunits that have the Cys-Cys pair are α subunits. Residues in green are the part of the α subunit agonist-binding pocket, and orange residues are the part of the agonist-binding site in the β or negative face subunit. Orange residues with black dots relate to the gating channel. Extracted from ⁵⁵; B) The structure of the $\alpha 7$ nAChR, made by the author.

In vertebrates, nicotinic receptors are distributed in many tissues, including muscle, CNS, and peripheral nervous system^{55, 89-91}. Based on their primary sites of expression, they are divided into two subtypes: one is muscle-type nicotinic receptors, and the other is neuronal-type nicotinic receptors^{55, 92, 93}. Muscle-type receptors (nAChR_{ms}) are located at the neuromuscular junction and mediate the transfer of transmitters between the nerves and muscle⁹³. Neuronal nicotinic receptors (nAChR_{ns}) are located in the postsynaptic membrane of the autonomic ganglion and the central nervous system, where they are responsible for transmitting outgoing signals from the presynaptic to the postsynaptic neurons and releasing neurotransmitters to regulate excitability^{55, 92}. Both nAChR_{ms} and nAChR_{ns} have some common characteristics in the amino acid composition and the structure of the receptor. As shown in Figure 11, $\alpha 7$ receptors show the structure composed of five identical subunits symmetrically arranged

around the axis of the channel. Each subunit comprises the structure as follows:

- 1) a long hydrophilic extracellular (EC) NH₂-terminal domain of ~200 amino acids;
- 2) three α -helical hydrophobic transmembrane (TM) domains (M1-M3);
- 3) a cytoplasmic loop of variable size and amino acid sequence between M3 and M4 in intracellular (IC) domain;
- 4) a fourth transmembrane region (M4) with a short extracellular COOH-terminal of ~20 amino acids.^{55, 84, 94, 95}

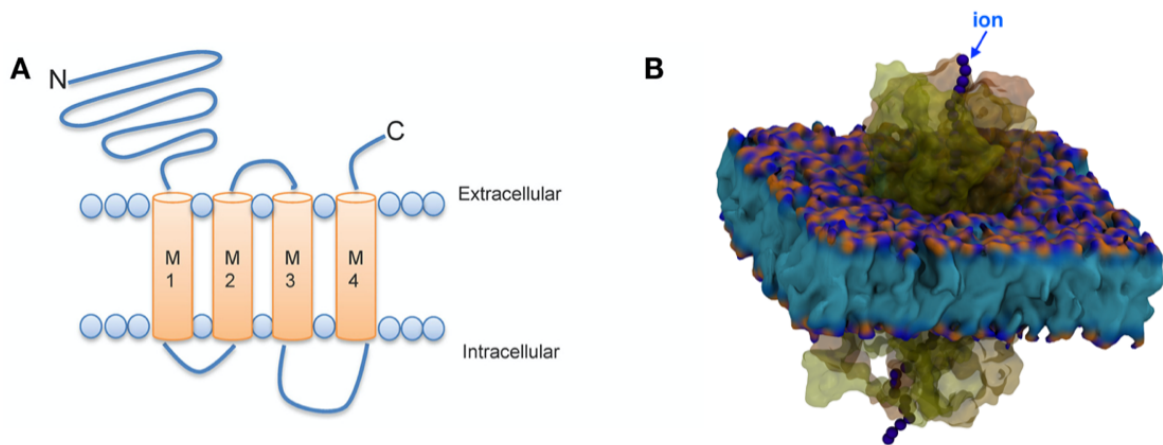


Figure 11: The topology of a neuronal nAChR subunit and a structure example of nAChR. A) Membrane topology of a neuronal nAChR subunit. Each subunit contains an N/C-terminal extracellular domain, four transmembrane domains (M1-4) and an intracellular loop. Extract from ⁹⁵. B) The way that Ca²⁺ passes the $\alpha 7$ nAChR when embedded into the membrane. Made by the author.

Besides the canonical $\alpha 7$, there is a human-specific dup $\alpha 7$ as one of the primary nicotinic receptors in the brain involved in several neuropsychiatric disorders^{96, 97}, and many studies have found the reason that may cause diseases is that it can be duplicated and form a hybrid gene with exons A-E of FAR7a, known as CHRFAM7A, which is unique to humans⁹⁸⁻¹⁰⁰. It is valuable to analyse its physiological and pharmacological properties. More information will be expanded in detail in section 2.2.

2.1.1 Subunits

To date, seventeen distinct nAChR subunits ($\alpha 1$ -10, $\beta 1$ -4, γ , δ , ϵ) have been identified^{86, 101}. Those seventeen subunits are divided into four subfamilies (I - IV)¹⁰². The four subunit types

($\alpha 9$, $\alpha 10$ and $\alpha 7$, $\alpha 8$) in subfamilies I and II can form not only homopentameric receptors but also heteropentamers^{91, 103}. The $\alpha 8$ subunit only appears in avian species, and it is not present in humans and other mammals^{90, 91, 103, 104}. A variety of heteropentamers exist in subfamily III ($\alpha 2$ -6, $\beta 2$ -4). Those three subfamilies (I - III) are widely expressed in neurons, whereas the subfamily IV receptors ($\alpha 1$, $\beta 1$, γ , δ and ϵ) are expressed in muscle^{55, 103}. The classification of nAChR subunits characterised to date is showed in Table 1.

Table 1 : The classification of nAChR subunits. $\alpha 7$, which is a subject of this study, is highlighted in bold.

| Neuronal type | | | | | Muscle type |
|--------------------------|---|---|-----------------------|------------------------|---|
| I | II | III | | | IV |
| | | 1 | 2 | 3 | |
| $\alpha 9$, $\alpha 10$ | $\alpha 7$, $\alpha 8$ | $\alpha 2$, $\alpha 3$, $\alpha 4$, $\alpha 6$ | $\beta 2$, $\beta 4$ | $\beta 3$, $\alpha 5$ | $\alpha 1$, $\beta 1$, γ , δ , ϵ |

When compared to multiple neuronal nAChRs structures, the muscle-type receptors are much less diverse, and only two subunits form mAChRs receptors: ($\alpha 1$)₂ $\beta 1\delta\gamma$ and ($\alpha 1$)₂ $\beta 1\delta\epsilon$. In neurons, the nicotinic receptors form mostly homopentamers or heteropentamers consisting of two different subunits. For instance, the $\alpha 7$ receptors are composed of five identical $\alpha 7$ subunits, and another common neuronal type, $\alpha 4\beta 2$ that also abundant in CNS, is made up of two $\alpha 4$ and three $\beta 2$ subunits^{105, 106}. The structure of typical homopentameric and heteropentameric receptors are shown in Figure 12.

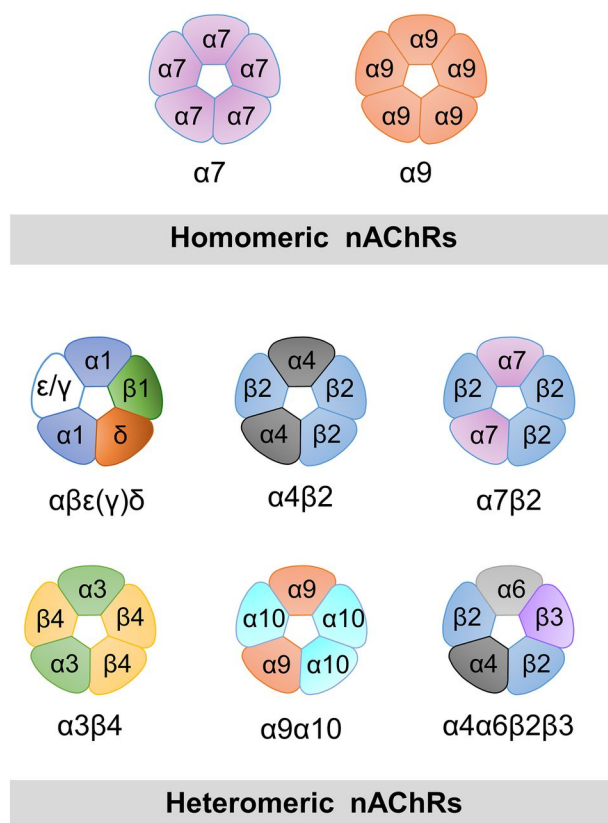


Figure 12: Different types of AChRs pentamers. Upper panel: homomeric nAChRs of $\alpha 7$ and $\alpha 9$; both receptors consist of five identical subunits. Lower panel: heteromeric nAChRs of $(\alpha 1)2\beta 1\delta\epsilon$, $\alpha 4\beta 2$, $\alpha 7\beta 2$, $\alpha 3\beta 4$, $\alpha 4\alpha 6\beta 2\beta 3$, $\alpha 9\alpha 10$, which are composed of different subunits. Extracted from ¹⁰⁷.

2.1.2 Structural biology of nicotinic receptors

The detailed knowledge of the structural biology of nicotinic receptors is critical for the structure-guided development of novel drugs. However, the information on the atomistic resolution structure of nAChRs is limited due to the size and transmembrane character of nicotinic receptors, low expression yield in a heterologous system, poor biochemical stability when extracted from the plasma membrane, and post-translational modification, since some nAChRs are heavily glycosylated. These technical hurdles impeded the collection of high-resolution structures of nAChRs. Until recently, most mechanistic insights on nAChRs were based on homology modelling techniques, utilising the following experimental homologous structures^{108, 109}: 1) The X-ray crystallographic structures of the pentameric acetylcholine binding protein (AChBP) at 3.3 Å resolution (PDB code: 1I9B)^{110, 111}; 2) The electron microscopy (EM) models of *Torpedo* acetylcholine receptor at 4.0 Å resolution (PDB code:

2BG9)¹¹²⁻¹¹⁴; 3) The X-ray crystallographic structures of PH-gated prokaryotic homologues of nAChRs at 3.1 Å resolution^{115, 116}. However, the shortcomings of these structural models cannot be ignored, such as the (1) structure includes only extracellular domains, and the (3) one is unstable once the agonist ivermectin is removed¹¹⁷.

Fortunately, the high-resolution structural information embedded in the membrane became available with the technical development obtained in the last few years. Recently, near-atomic resolution cryogenic Electron Microscopy (cryo-EM) and X-ray structures of several nAChRs have been published, such as the human $\alpha 3\beta 4$ ganglionic nAChR¹¹⁸ (PDB code: 6PV7, Figure 13), the extracellular domain of $\alpha 9$ with ligands¹¹⁹ (PDB code: 4D01), and the human $\alpha 4\beta 2$ ^{81, 120} (PDB code: 5KXI). Besides this, there are also recent X-ray works on ELIC channels such as 6HJX¹²¹, 6SSP¹²², and cryoEM 5-HT3 receptor¹²³ (PDB code: 6HIO) produced.

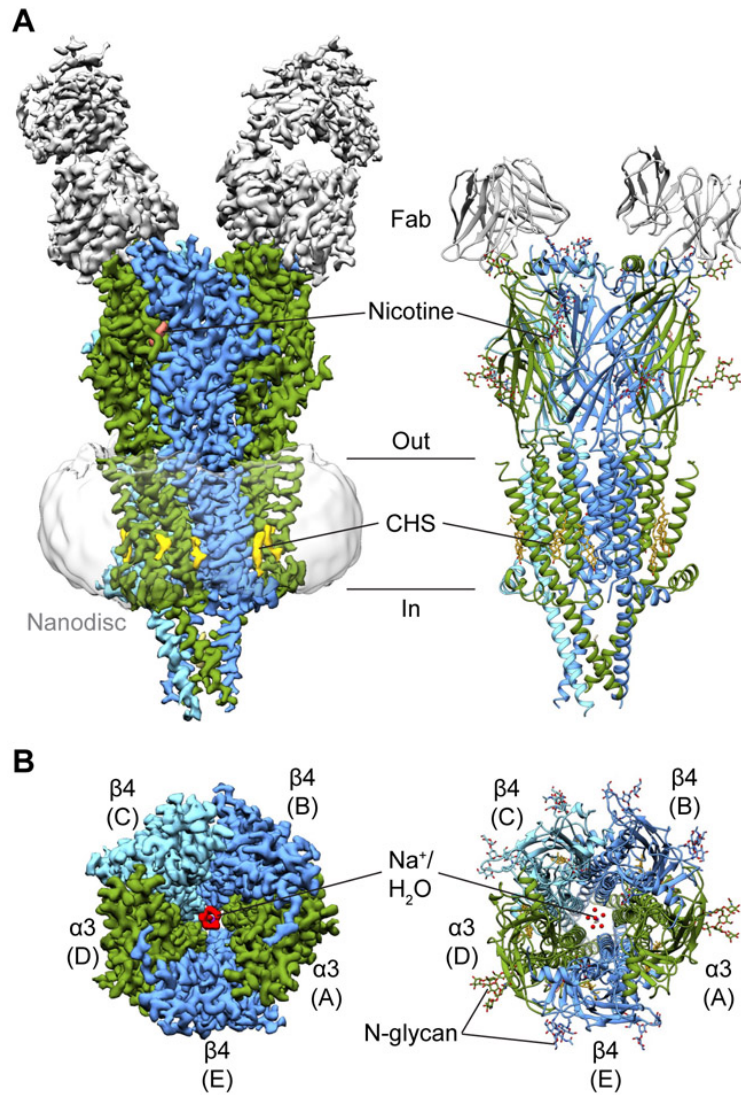


Figure 13 : Architecture of the $\alpha 3\beta 4$ receptor. (A) Side views of the cryo-EM map and atomic model of the $\alpha 3\beta 4$ -nicotine complex. (B) Top views of the cryo-EM map and atomic model of the $\alpha 3\beta 4$ -nicotine complex. $\alpha 3$ subunits are coloured in green, $\beta 4$ subunits in blue, fabs in grey, nicotine in salmon, cholesteryl hemisuccinate (CHS) in yellow, water in red, and sodium in purple. Extract from ¹¹⁸.

However, there is still no experimental structure of a mammalian $\alpha 7$ receptor, and especially, information on partially duplicated human-specific dup $\alpha 7$ is missing. Therefore, the first challenge in studies on human $\alpha 7$ receptors and structure-based drugs specific to $\alpha 7$ and dup $\alpha 7$ receptors is building reliable models.

2.1.3 Agonist (orthosteric) binding sites

A positionally-conserved agonist (orthosteric) binding site is located at the interface of two EC

domains of two adjacent receptor subunits (Figure 14). This site is comprised of residues within the C-loop ($\alpha 1$, $\alpha 2$, $\alpha 3$, $\alpha 4$, $\alpha 6$, $\alpha 7$ and $\alpha 9$ receptors). Many studies indicated that C-loop is a crucial binding part for ligand binding, and its dynamics influences the activity of AChRs. In addition to C-loop residues, many hydrophobic aromatic amino acids are required for the binding site formation, such as W53, Y91, W145, Y184, Y191¹²⁴⁻¹²⁷ in $\alpha 7$ receptors (Figure 14).

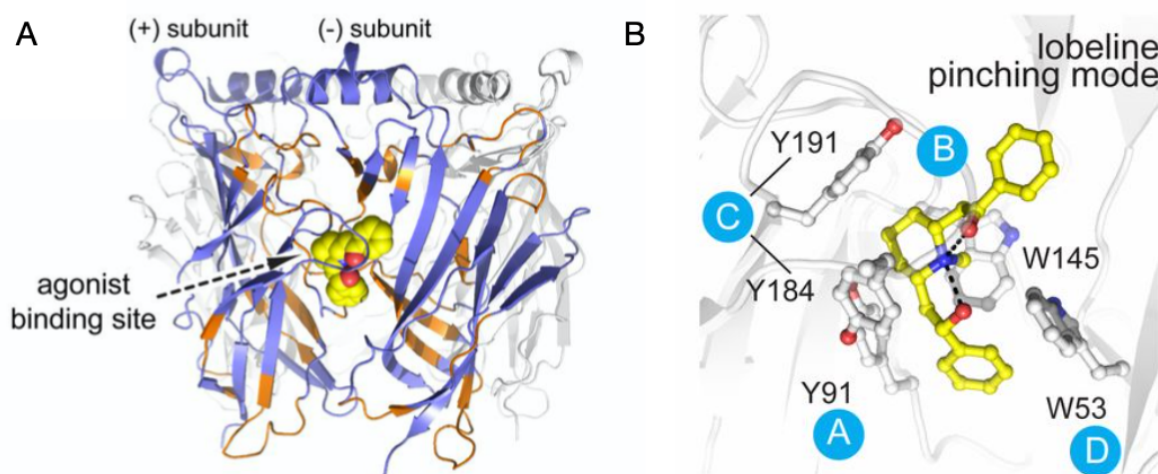


Figure 14: Structure of $\alpha 7$ AChBP and location of the agonist binding site. (A) Side view of the $\alpha 7$ nAChR structure with the agonist lobeline; the binding site is localized at the subunit interface of the (+) and (-) adjacent subunits. Lobeline is shown in yellow sphere representation. (B) The residues of the agonist binding site. The blue circles indicate the corresponding residue of the binding site, termed A, B, and C on the (+) subunit and D on the (-) subunit. Extracted from ¹²⁷.

Most α type subunits (e.g., $\alpha 7$ in $\alpha 7$ nAChR) provide the agonist binding sites, whereas β type subunits act as structural scaffold subunits contributing to the stability of the receptor conformation. For example, $\beta 1$ and $\beta 3$ subunits do not directly participate in the binding site formation⁵⁵. The location and number of binding sites in different types of AChRs are shown in Figure 15. As for the neuronal-type homogeneous model, the pentamer consist of $\alpha 7$, $\alpha 8$, $\alpha 9$ subunits have five identical same agonist binding sites. In general, the ligand affinity associates with the identity of the α subunit hydrophobic residue, whereas the ligand selectivity depends on the adjacent subunit residues.⁵⁵

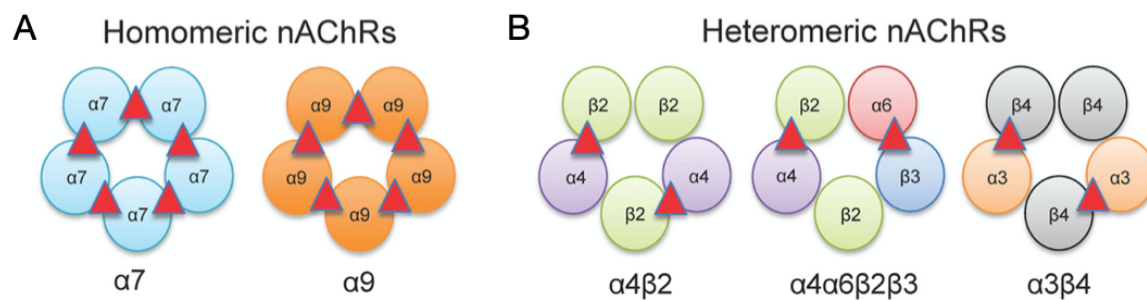


Figure 15: The location of binding sites in different AChRs agonist binding sites is depicted as red triangles. Extract from ⁹⁵.

2.1.4 Allosteric binding sites and allosteric of nicotinic receptors

Most agonists, including neurotransmitters and drugs, bind at the orthosteric sites to activate nAChRs, leading the channel opening and synaptic neurotransmission. Therefore, the prolonged activation at the orthosteric site of nAChRs usually not only leads to tolerance development or desensitization to the drug¹²⁸. Recently, binding studies of the nAChRs focused on allosteric sites that may benefit new drug discovery due to fewer side effects. The location of putative allosteric binding sites of $\alpha 7$ nAChR is shown in Figure 16.

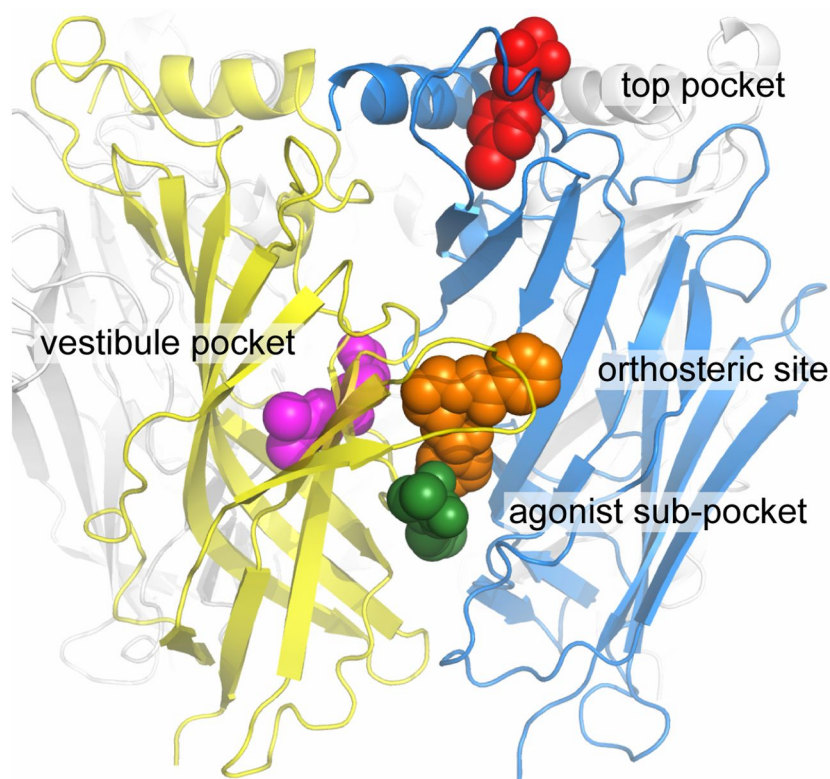


Figure 16: Orthosteric and allosteric binding sites mapped at $\alpha 7$ -AChBP. The orthosteric agonist and several different allosteric binders are depicted as spheres. Two different subunits are rendered as ribbons and coloured yellow and blue, respectively. Extracted from ¹²⁷.

Numerous studies on allosteric sites reported that many allosteric sites and their structural diversity depend on the type of subunit⁵¹. This results in a large number of allosteric sites producing greater specificity compared to orthosteric sites¹²⁹. By not binding to the orthosteric site, allosteric binders can enhance or inhibit agonist activation (positive or negative allosteric modulators). Allosteric ligands activate nAChRs only in the presence of an endogenous ligand such as Ach¹³⁰, modifying agonist efficacy by reducing or increasing the energy barrier between the closed and open states. Studies have shown that allosteric ligands are associated with better selectivity, lower tolerance, fewer side effects and lower toxicity⁸⁰. Thus, the discovery of allosteric binding sites and the design of allosteric ligands offer new directions for drug development.

The discovery of allosteric binding sites of nAChR is still a challenge due to the experimental difficulties in detecting and verifying allosteric effects and binding sites. These difficulties come from the complexity of protein systems and the limitation of experimental techniques.

For example, X-ray crystallography requires well-diffracting crystals; solution NMR is limited to small proteins (less than 35kDa), cryo-EM may have resolution and size limitations. There are shortcomings for biophysical assays, i.e., isothermal titration calorimetry (ITC) and circular dichroism (CD) require a significant amount of recombinant protein available. Therefore, computational methods have provided accurate models for identifying allosteric binding sites and their possible ligands, such as FTMap. I will discuss this method specified in the following Chapter 3.

2.1.5 Ligands of nicotinic receptors

The ligands of nAChR are highly structurally diverse, including many high-affinity small molecule natural ligands (e.g., neurotransmitter acetylcholine, nicotine, cytisine, and epibatidine) (Figure 17)¹³¹, and small synthetic ligands (e.g., varenicline, morantel), most of these small ligands are as agonists. Also, large natural products (e.g., methyllycaconitine¹³², tubocurarine¹³³) and peptide toxins (α -bungarotoxin) used to be antagonists or allosteric modulations.

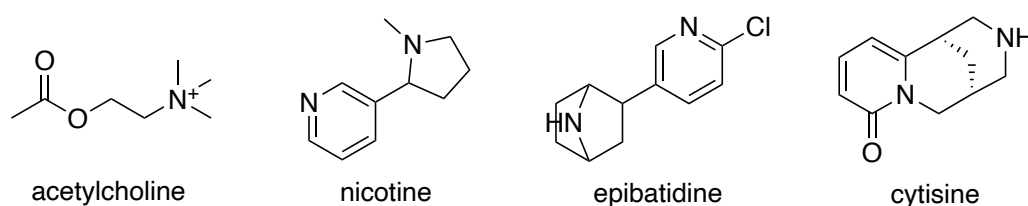


Figure 17: Agonists of nicotinic receptors.

Despite a large number of known nAChR ligands, there is a continuing demand for new nAChR ligands, especially those with new subunit selectivity and activity profiles, for example, dupa7, which has no small molecule or peptide ligand that is known to interfere with it.

2.1.6 Channel gating

The structure of nAChR, shown in Figure 18, has five subunits arranged in a ring to form a pentamer, and the middle part is hollow and works as an ion channel. Ligands regulate Channel-gating (agonists, allosteric modulators) binding to the EC part and causing conformational changes propagating to the receptor's transmembrane (TM) domain. Figure 18A shows the

cross-section of the TM domain: it consists of a total of 20 α -helices of five subunits. The five TM helices² are arranged radially on the inner side to form the pore, and the other helices (M1, M2, M3) are arranged on the outside to protect the inner loop¹³⁴. When the receptor is in the static state, ions (e.g. K^+ and Ca^{2+}) cannot pass through the channel due to the channel closed and the only barrier blockage¹³⁵. (Figure 18)

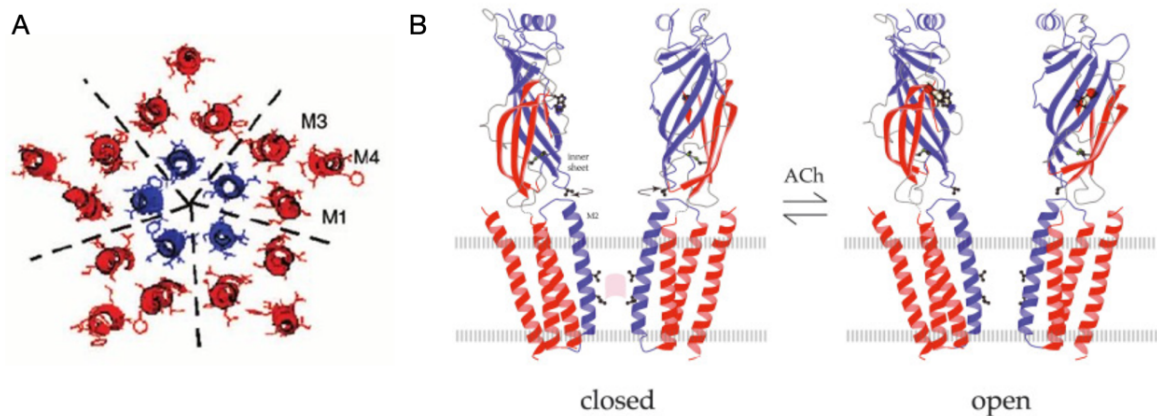


Figure 18: A) Cross-section of the acetylcholine receptor TM domain (PDB code: 2BG9). Inner helix 2 in blue colour and helix 1, 3, 4 are signed by M1, M3, M4 with red colour. A dashed line separates the subunit. Extracted from ¹³⁶. B) The closed-to-open transition for the Torpedo nicotinic receptor. Ach binding causes the inner β EC sheets (blue arrow) to rotate clockwise. This motion is transmitted to the inner M2 helix (blue), which breaks the channel gate (pink) made by leucine (α L251) and valine (α L255) side chains, and opens the channel. The membrane bilayer is showed as broken grey lines corresponding to lipid head groups. Extracted from ¹³⁷.

As shown in Figure 18B, when the EC domain is exposed to the agonist (such as Ach), the binding of the agonist to the interface of the adjacent subunits triggers conformational changes, opening the channel. During this process, the inner EC β -sheets turn around, the channel axis rotates 15 degrees and drives the rotation of TM helix 2, leading to the pore opening^{137, 138}. These conformational changes propagate from the binding site toward the interface between the extracellular and transmembrane domains, called the coupling region. These conformational changes contribute to the molecular mechanism of the receptor activation.

2.1.7 Diversity

There are multiple types of nAChRs characterized to date; those receptors share overall

architecture and have similar orthosteric sites, but they also show functional diversity and specialisation. For example, $\alpha 3$ is a basic element in all nAChRs of autonomic ganglion cells, which plays an essential role in the fast conduction of neurotransmitters in the autonomic nervous system¹³⁹. $\beta 3$ lacks the necessary residue for agonist binding¹⁴⁰. The pentamer assembled by $\alpha 4$ and $\beta 2$ subunits or $\alpha 7$ subunits have attracted the attention of researchers and have the highest value for studying because they are involved in different central nervous system functions (such as human cognition, pain, learning, memory) and implicated in several neurological and autoimmune diseases. These two $\alpha 4\beta 2$ -containing nAChRs and $\alpha 7$ nAChRs are the mainly nAChRs type in the central nervous system¹⁴¹, where they are responsible for transmitting outgoing signals from the presynaptic to the postsynaptic neurons. $\alpha 4\beta 2$ -nAChRs have high affinity to nicotine ($K_i = 1.05$ nM) and acetylcholine (ACh) and slow desensitization¹⁴². On the contrary, $\alpha 7$ receptors show low affinity to nicotine ($K_i = 4000$ nM) and ACh and very rapid desensitization but have high affinity to α -bungarotoxin (α -BTX) ($K_i = 0.6$ nM) and high Ca^{2+} permeability^{142, 143}.

2.2 $\alpha 7$ & dup $\alpha 7$

Nicotinic $\alpha 7$ receptors are the only homopentameric AChRs widely expressed in the central nervous system⁹⁶. The function of $\alpha 7$ nAChRs has been found involved in cognition, memory, and immunomodulation, which have emerged as attractive targets for neuropsychiatric disorders, neuroinflammation, neuropathic pain, and autoimmune diseases⁹⁷.

The human $\alpha 7$ subunit is encoded by CHRNA7 gene^{98, 144}, located at human chromosome 15q13-14 and contains ten exons and nine introns. Among them, exon 4, 6 and 7 express the $\alpha 7$ agonist binding site. The gene information is showed in Figure 19. However, the changes in some residues of the CHRNA7 gene produce duplication $\alpha 7$ when undergoing recombination events⁹⁸⁻¹⁰⁰. Exons 5-10 in CHRNA7 are duplicated and fused to exons A-E of FAM7A (family with sequence similarity 7A), resulting in the hybrid gene denoted as CHRFAM7A⁹⁸⁻¹⁰⁰. This product, dup $\alpha 7$, is a truncated subunit, where the N-terminal 146 residues of the ligand binding domain of the $\alpha 7$ receptor have been replaced by 27 residues from FAM7A⁹⁸⁻¹⁰⁰. This duplication is evolutionary new and unique to humans: the genetic sequence exons 5-10 shared

by CHRFA7A and CHRNA7 shows more than 99% sequence identity^{98, 144, 145}. The N-terminal part of the canonical $\alpha 7$ subunit representing N-terminal in the EC domain has been replaced by a 27 amino acid long N-terminal domain in dupa7; the remaining sequences of both protein products are identical. More information about sequence alignment and structure of $\alpha 7$ and dupa7 subunits is provided in Chapters 4 and 5.

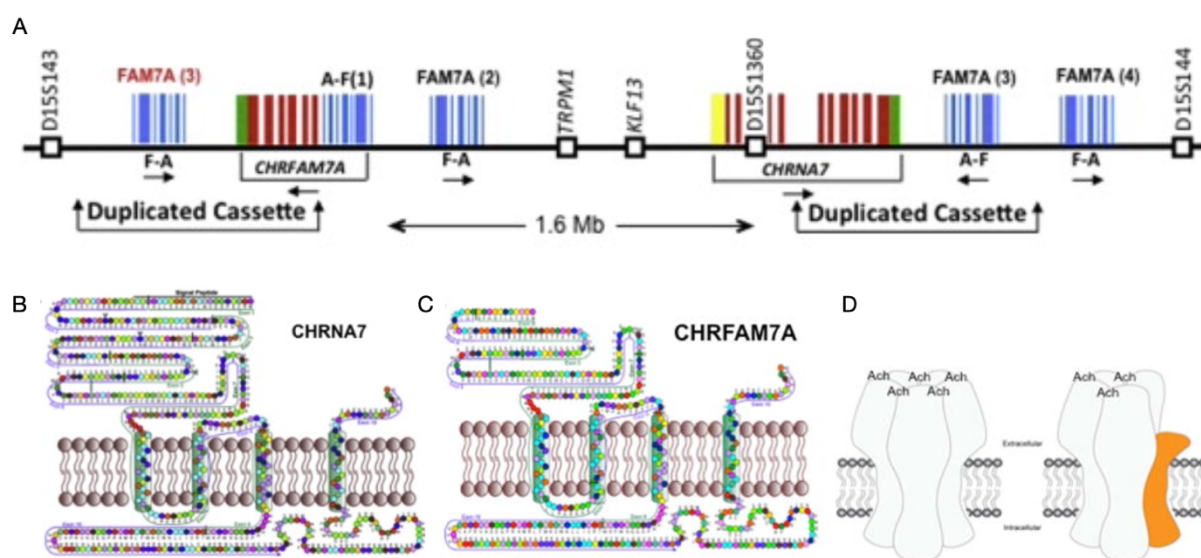


Figure 19: The human $\alpha 7$ nAChR, FAM7A & dupa7 nAChR. A) Formation of the chimeric gene of dupa7 CHRFA7A; B) The gene product of $\alpha 7$ subunit; C) The gene product of dupa7 subunit; D) The structure of the canonical $\alpha 7$ pentamer (left) and possible structure of dupa7 pentamer (right), dupa7 subunit is signed by orange colour. Extracted from⁹⁸.

It has been reported that a small amount of CHRFA7A represents 10–20% of the $\alpha 7$ sequence in the mRNA of the human brain¹⁴⁶. The dupa7 subunit exists in the form of co-assemblies with native $\alpha 7$ subunits to form functional nAChRs. Due to the loss of exons 1-4, the dupa7 subunit losses one glycosylation site but still retain the cysteine bridge and vicinal cysteines of the agonist binding site⁹⁸. The number and position of dupa7 subunits assembled in nAChR reduce the number of agonist binding sites, resulting in adverse effects on the function of nAChR. Dup a7 nAChR not only affects the overall structure but Costantini and co-workers showed that dupa7 interferes with the binding of α -bungarotoxin (α -BTX) at the receptors¹⁴⁷.

It has been postulated that dupa7 acts as a dominant-negative inhibitor of $\alpha 7$ function, suggesting its role in human cognition and immune responses by perturbing normal (canonical)

$\alpha 7$ activities. This perturbation is likely to impact neurological disorders, including Alzheimer's disease^{97, 148} and schizophrenia¹⁴⁹, and affect the cholinergic immunomodulation^{150, 151}. Kunii and coworkers observed that *CHRFAM7A* was upregulated in the brains of patients with schizophrenia and bipolar disorder, with an increased ratio of *CHRFAM7A*/*CHRNA7*¹⁵². On the other hand, this interaction can also be used to treat neurological disorders, most likely *via* regulating the functions of the $\alpha 7$ receptor by direct interactions to mitigate A β uptake¹⁴⁸. More recently, Szigeti and coworkers reported functional readouts for the *CHRFAM7A* alleles for two phenotypic readouts in mild-to-moderate Alzheimer's disease cohort and showed a 3:1 split in the population for *CHRFAM7A* carriers to non-carriers of the functional direct allele¹⁴. To demonstrate the translational gap, the group performed two double-blind pharmacogenetic studies for both first exposure and disease-modifying effect. It concluded that dup $\alpha 7$ accounts for the translational gap in the development of new drugs tackling Alzheimer's disease and that design of future trials need to incorporate *CHRFAM7A* pharmacogenetics¹⁴.

2.3 Nicotinic receptors in neurodegenerative disease: interactions with amyloid- β

Recently, many researchers investigated the biochemical relationship between $\alpha 7$ receptors and amyloid β peptide 1-42 (A β_{1-42}). In AD patient brains, both $\alpha 7$ nAChR and A β_{1-42} were found located in neurons and neuritic plaques, and A β_{1-42} coimmunoprecipitated with $\alpha 7$ ¹⁵³⁻¹⁵⁸. Importantly, these two proteins have a high affinity to interact with each other and form a stable complex, which may inhibit Ach release from the cholinergic neurons and changes in calcium homeostasis, leading to the physiology of neurons influenced caused neurodegeneration^{154, 159}. Studies show that $\alpha 7$ may mediate the cytotoxicity induced by A β_{1-42} and promote deposition of A β_{1-42} in neurons^{154, 159}. Meanwhile, A β_{1-42} can downregulate the expression of $\alpha 7$, which contributes to the cholinergic signalling deficits^{153, 160}. However, the activation of $\alpha 7$ could be beneficial in the low concentration of A β_{1-42} ¹⁵³. All these indicate that $\alpha 7$ and amyloid- β have a close relationship, which is vital to the pathophysiology of AD.

In addition, studies also found that $\alpha 7$ agonists and antagonists such as nicotine and α -BTX can inhibit the formation of A β_{1-42} - $\alpha 7$ complex, which protects against A β_{1-42} induced cytotoxicity and tau phosphorylation *in vitro* and prevents A β_{1-42} -related apoptosis^{154, 161}.

2.4 TSPO as inflammation biomarker to validate $\alpha 7$ results

As showed in previous sections, the link between the activity of nAChRs, especially $\alpha 7$, and positive outcomes in neuroinflammation and neurodegenerative conditions has been established. This renders $\alpha 7$ receptors attractive targets for new therapeutics and diagnostics. However, to exploit $\alpha 7$ receptors in PET in vivo studies, it is required to validate the results against other, more established targets involved in neuroinflammation and neurodegeneration. This section will introduce the TSPO: 18 kDa translocator protein, which is recognised as an essential biomarker for neuroinflammation due to its increased expression in microglia upon inflammation. TSPO is located at the outer mitochondrial membrane^{39, 162-164}, belonging to the tryptophan-rich sensory protein family. It was previously known as peripheral type benzodiazepine receptor (PBR), but this protein has been found to bind to diazepam and expressed throughout the body and brain^{164, 165}. The name was changed to translocator protein due to its capability to transport small molecules into the mitochondria and after discovering chemically diverse molecules bound by TSPO with high affinity.

2.4.1 Structure

The human TSPO is an 18kDa protein located at the outer mitochondrial membrane^{39, 162-164}. Its structure consists of five transmembrane α -helical bundles (TM1-5), forming dimers and higher-order oligomers (Figure 20). Human TSPO contains 169 residues, which include two cysteines of an uncertain regulatory role. Its gene is localized to the 22q13.31 chromosome in the human genome. The transmembrane topology of TSPO shows C-terminus outside the mitochondria (cytosol) and a short N-terminus inside the mitochondria.¹⁶⁶

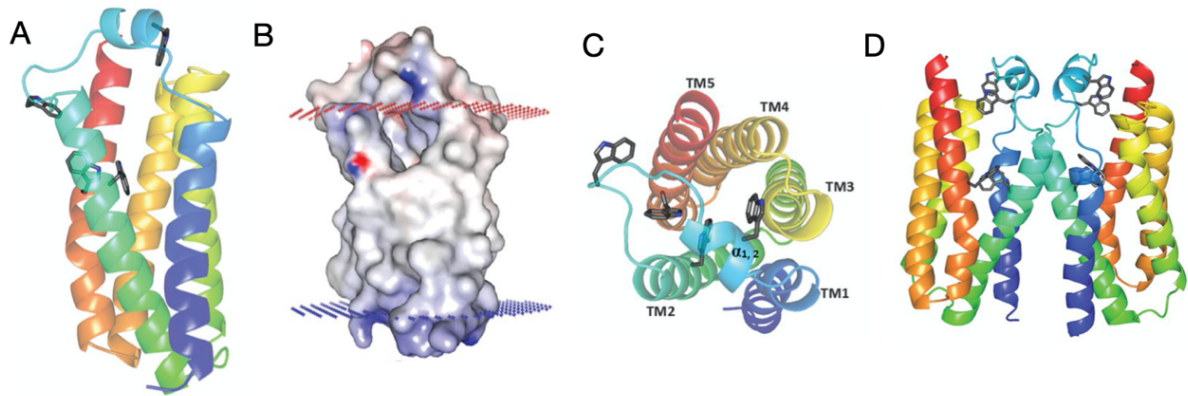


Figure 20: The structure of TSPO. A) The 3D structure of apo TSPO. B) Electrostatic potential at the surface of the TSPO structure. C) Top view of TSPO structure. D) The structure of dimer TSPO. Extracted from ¹⁶⁷.

Several studies have shown that TSPO is associated with cytosolic and other mitochondrial proteins. (Figure 21). This protein is highly expressed in body organs involved in steroid synthesis, including the brain. It forms a heteromeric complex with the 32 kDa voltage-dependent anion channel (VDAC) at sites of contact between the inner and outer mitochondrial membranes and the 30 kDa adenine nucleotide translocation (ANT) on the inner membrane, which functions as a mitochondrial permeability transition pore (MPTP) for the transport of small molecules into the mitochondria¹⁶⁸. By the interaction with VDAC and adenine nucleotide carrier (ANC), the TSPO is able to regulate the mitochondrial function, leading to the regulation of the heart, steroidogenesis, apoptosis and immunomodulation¹⁶⁸. Numerous studies indicate that TSPO is involved in a variety of human diseases, particularly in neurodegeneration, and it plays a crucial role in responses to stress.

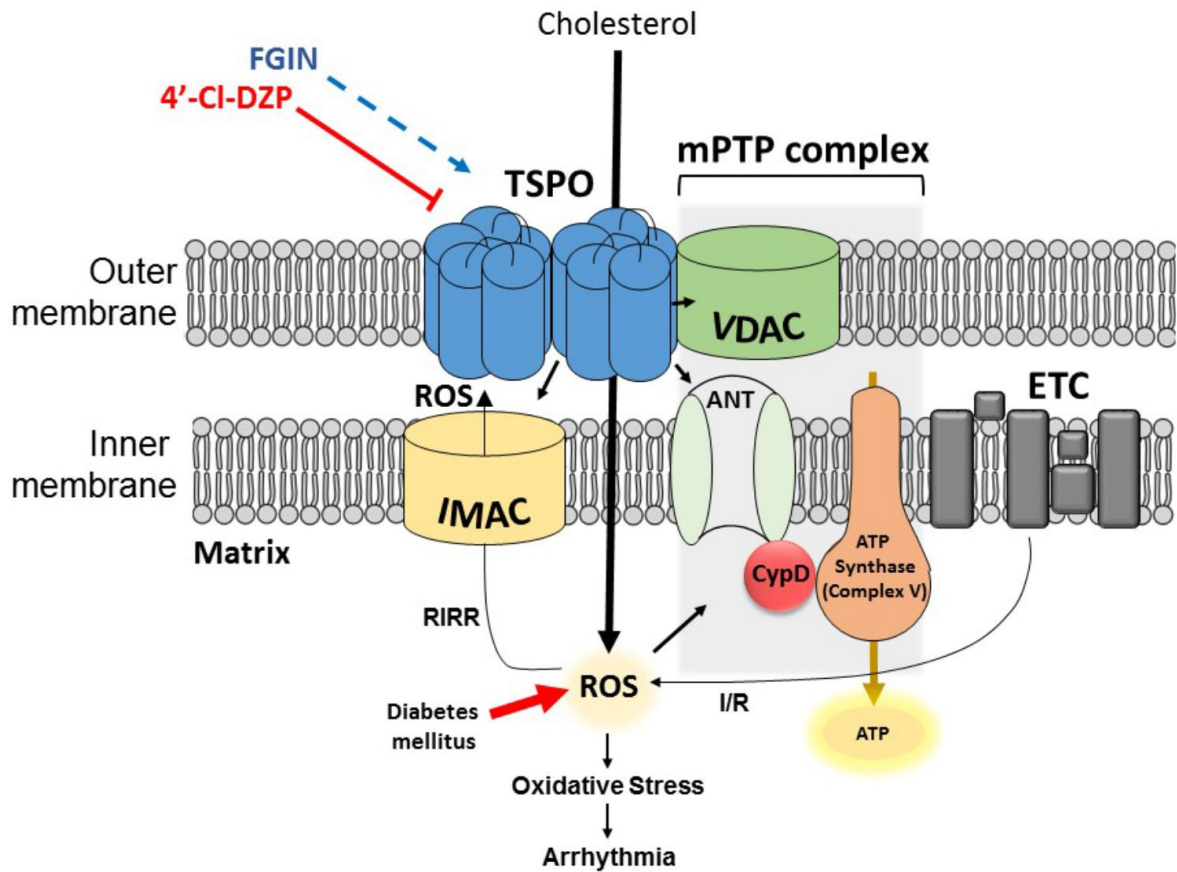


Figure 21: The mechanism of complex-forming by TSPO and other proteins on the mitochondrial membrane. Extracted from ¹⁶⁹.

2.4.2 Function

TSPO has been found in various tissues in humans, such as the adrenal, kidney, heart, testis, ovary, liver, brain, hematopoietic cell and lymphocytes^{163, 170}. TSPO has been involved in cholesterol transport in mitochondrial, steroidogenesis, regulation in the heart, modulation of voltage-dependent calcium channels, apoptosis, microglial activation related to brain damage, regulation of the mitochondrial membrane potential, inflammation, cell growth and differentiation, stress adaptation, cancer cell proliferation and immunomodulation^{39, 163-165, 171-175}. One of the well-known functions of TSPO is intramitochondrial cholesterol transport¹⁷⁶. The inner surface of the channel inside the TSPO helical bundle is hydrophilic and uncharged, which provides a pathway allowing cholesterol molecules to move into the mitochondria^{164, 165}. Besides this, TSPO is also involved in steroid hormone synthesis, mitochondrial respiration, mitochondrial permeability transition (MPT) pore opening and heme transport. The role of

TSPO in those functions have been demonstrated in knockout¹⁷⁷ and antisense experiments¹⁷⁸, whereby the downregulation of TSPO leads to a reduction in steroid synthesis, including endogenous steroid synthesis in CNS. Additionally, the mouse studies have also demonstrated that the up-regulation of TSPO in microglia and astrocytes in CNS is closely associated with microglia activation and directly affects the extent of neuronal damage and inflammation.

Moreover, TSPO has been found widely expressed in the microglia, which play an essential role in neuroinflammation¹⁶⁸. With microglia proliferation, the amount of biogenesis in mitochondrial membranes such as TSPO also increase¹⁷⁹. TSPO contributes to producing the reactive oxygen species (ROS) in the mitochondria, leading to neuroinflammation and contributing to neurodegenerative diseases such as AD. Thereby, TSPO has been used as a biomarker for neurodegenerative diseases such as Alzheimer's disease^{44, 180}, multiple sclerosis¹⁸¹ and Parkinson's disease⁴⁶. For this reason, a broad panel of TSPO ligands have been developed as neuroimaging agents; for example, for positron emission tomography (PET) studies, where they are used as diagnostic tools for neurodegenerative disorders.

2.5 TSPO and nicotinic receptors as targets for drugs and imaging agents

Both nAChRs and TSPO can be targeted therapeutically to treat various disorders with inflammation as the underlying cause. As described in the previous section, they have a great potential to serve as diagnostic biomarkers and are targeted by small radiolabelled molecules in PET studies. This section will outline some clinically approved drugs and imaging agents targeting those proteins.

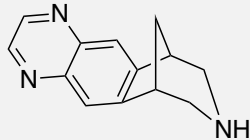
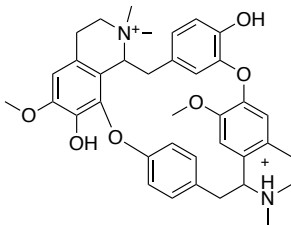
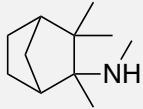
2.5.1 Nicotinic receptors as targets

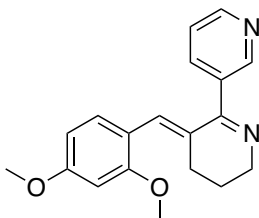
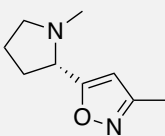
Varenicline¹⁸² was the first clinically approved drug targeting nicotinic receptors. It is an orthosteric agonist used to treat smoking addiction, acting on $\alpha 4/\beta 2$, $\alpha 3$, $\alpha 6$ and $\alpha 7$. As it is an orthosteric and competitive ligand, it reduces the binding of nicotine. The other type of nicotinic agonist used in the clinic is a non-depolarizing neuromuscular blocking agent, tubocurarine¹³³ can be an example. A nicotinic antagonist used in the clinic is mecamylamine¹⁸³, a non-selective and non-competitive nicotinic antagonist. It can cross the blood-brain barrier, and it works as a

treatment for hypertension.

There are also drugs for specific nAChR subtypes. For instance, GTS-21^{129, 184} is $\alpha 7$ nicotinic agonist, under tests for treatment in Alzheimer's disease and schizophrenia and schizoaffective disorders. As for $\alpha 4/\beta 2$ nicotinic agonist, ABT-418¹⁸⁵ has also been found to be active in a limited human trial in attention deficit hyperactivity disorder (ADHD). All the ligands mentioned in this chapter are list in Table 2.

Table 2: Selected small molecule ligands of nAChR

| LIGAND | PHARMACODYNAMIC PROPERTIES | THERAPEUTIC USE | STRUCTURE |
|---------------------|---|---|---|
| VARENICLINE | Partial agonist of nAChRs, such as: $\alpha 4/\beta 2$, $\alpha 3$, $\alpha 6$ and $\alpha 7$ | Treatment of tobacco dependence |  |
| TUBOCURARINE | antagonist of nAChRs | The non-depolarizing neuromuscular blocking agent |  |
| MECAMYLAMINE | Partial antagonist of nAChRs, such as: $\alpha 2$, $\alpha 4$, $\alpha 7$ and $\beta 2$ | Treatment of moderate to severe essential hypertension and simple malignant hypertension. |  |

| | | | |
|----------------|--------------------------------------|--|---|
| GTS-21 | $\alpha 7$ nicotinic agonists | Treatment in AD and schizophrenia and schizoaffective disorders. |  |
| ABT-418 | $\alpha 4/\beta 2$ nicotinic agonist | Treatment in hyperactivity disorder |  |

2.5.2 TSPO as a target

The PK11195 was the first nonbenzodiazepine ligand for TSPO, and (R)-[^{11}C]PK11195 have been used in positron emission tomography (PET) to visualise brain inflammation in patients with neuronal damage¹⁸⁶. Nevertheless, other agents were subsequently developed because of the drawbacks of [^{11}C]PK11195, such as poor bioavailability to brain tissue and a high degree of non-specific (off-target) binding. Among them was [^{11}C]DPA-713, which shows a good blood-brain barrier permeability and a high affinity for human TSPO in the brain (dissociation constant, 7.0 nM)¹⁸⁶. Conversely to the shorter half-life of ^{11}C (20.4 min), the longer half-life of ^{18}F (109.8 min) was suggested to be incorporated into ligands, and this produced [^{18}F]DPA-714¹⁸⁶. In addition, carbon-fluorine bonds have comparable stability, and many drugs are fluorinated to prolong metabolism¹⁸⁷. At the same time, the higher hydrophobicity of the carbon-fluorine bond increases the compound's lipophilicity, which typically increases the permeability of the cell membrane and thus improves the bioavailability of the drug¹⁸⁷. These benefits make the labelling of F the first choice PET isotope for radiopharmaceutical chemistry. Many studies compared the above ligands, and they indicated that DPA-714 show the highest bioavailability in brain tissue with very low non-specific binding¹⁸⁶. However, it is still needed to develop novel TSPO ligands with high affinity, good selectivity profile, and facile radiosynthesis for PET studies for chemical synthesis optimisation.

2.6 Challenges in the development of ligands targeting nicotinic receptors and TSPO

Although the distribution, subtype and function of nicotinic receptors, including human-specific $\text{dup}\alpha 7$, are known, some knowledge gaps remain to be addressed. The exact mechanism and specific contribution of $\text{dup}\alpha 7$ to the biology of $\alpha 7$ receptors remain highly elusive. The structural information on this human-specific receptor is missing, which hampers the efforts to elucidate its effect on $\alpha 7$ pentamers. This knowledge gap motivated the research presented in this dissertation.

Moreover, the TSPO PET signal decreased in successful studies of novel therapeutics in AD preclinical models, suggesting that TSPO could be used to monitor treatment progress in clinical trials¹⁸⁸. Despite these encouraging preclinical results, earlier clinical studies focusing on TSPO produced mixed results. Several imaging studies using the [11C]PK11195 reported higher TSPO brain signal in amyotrophic lateral sclerosis (ALS), AD, PD, and in brains of people identified as at risk of Huntington's disease (HD) compared to controls¹⁸⁸, other studies using second-generation TSPO ligands reported no difference in TSPO PET signal in AD and MS¹⁸⁸. The primary source of failure in these clinical trials was sizeable inter-individual variability in PET signal, caused by *rs6971* single-nucleotide polymorphism (SNP) in the human TSPO gene, which causes a substitution of alanine for threonine at the 147th amino acid (A147T) of TSPO. This polymorphism is present in 30% of Caucasians, 25% of Africans, 4% of Japanese, and 2% of Han Chinese (<http://hapmap.ncbi.nlm.nih.gov>). Patients who are homozygous for wild type TSPO show good response to several 2nd generation PET tracers, heterozygous show moderate response, and patients homozygous for the A147T TSPO variant show poor response.

While DPA-714 represents an improvement in relatively many of the second-generation ligands regarding its binding to the A147T TSPO variant, it still shows reduced binding affinity to that variant. It is crucial to gain a greater understanding of how the binding requirements of A147T TSPO differ from wtTSPO to overcome this problem and its limitations in the synthetic feasibility. One approach to deconvolute factors responsible for the loss of affinity of TSPO ligands to A147T TSPO variant is evaluating the differences in structure and dynamics of DPA-

713, DPA-714 and their analogues bound to wtTSPO and A147T TSPO variant, using structure-based molecular modelling approaches.

Chapter 3 Theoretical background

Drug development can be described as the process of bringing a lead compound (traditionally, a small molecule) into clinical practice. It includes preclinical laboratory research, four clinical phases in humans, and authority approval. The workflow governing the drug discovery pipeline is shown in Figure 22.

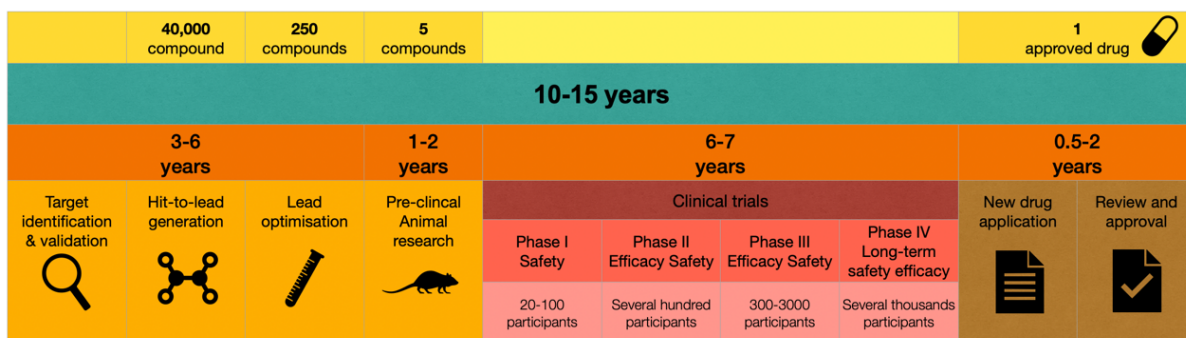


Figure 22: Drug development pipeline. Made by author.

The whole drug design goal is to find new drugs based on existing knowledge of the biological target. Three steps need to be done relating to target at first: target identification, target selection and target validation. Target identification is the step to identify biomolecular targets (most common protein) involved in the disease progression. Target selection is the decision to find a drug with a specific biological effect that is expected to have a therapeutic effect while acting on a specific target, and it is influenced by a complex mixture of scientific, medical, and strategic factors¹⁸⁹. Target validation is the process of proving that manipulating (e.g. enhancing or inhibiting) the selected target can provide therapeutic benefits to patients¹⁹⁰.

Traditionally, G protein-coupled receptors (GPCRs), enzymes, transcription factors and ion channels were popular protein targets for small molecule drugs. For example, the $\alpha 7$ nAChR receptor has been selected and validated as a target for drugs targeting CNS disorders, and approved drugs targeting this receptor include an agonist tropisetron (brand name Navoban) and positive allosteric modulator galantamine.

Following target identification and validation, subsequent steps in drug discovery programs are hit identification and lead discovery. A hit is a compound with a desired activity, and its activity

is confirmed upon testing. A lead series is a cluster of compounds sharing some typical section or chemical moiety, and it has a different potency than the initial hit due to having a different chemical moiety other than this core structure.

After lead discovery, the next step is lead optimisation. The aim of optimisation is to improve the potency and selectivity of ligand, making it capable of enhancing or inhibiting the target function in order to become an effective and safe drug. The ligand needs to be designed to complement the binding site to achieve tight binding to the target. Whether the optimisation is effective can be assessed based on several properties such as bioavailability, metabolic half-life and side effects.

Furthermore, the preclinical test would be done in vitro/vivo to select favourable compounds with beneficial ADME (absorption, distribution, metabolism, and excretion) property and low toxicity. Four phases of clinical trials follow the pre-clinical studies in the laboratory.

3.1 Computational methods for a drug discovery pipeline

The pipeline (Figure 22) shows that a drug from discovery takes more than a decade to become a product approved and marketed by the authorities. The total development is a lengthy and costly process, requiring time, money, and technical expertise such as human resources, experimental materials and research skills. Generally, classical drug development costs millions of pounds, and it takes 10 to 15 years to develop¹⁹¹. Usually, only 1 out of 40,000 tested compounds is approved as a safe drug that produces the efficacy response with minimal side effects. However, with the rapid development of computational technology, the computer-aided drug design (CADD) techniques developed since helped to save time and cost, and at the same time, increased the success rate to develop a drug.

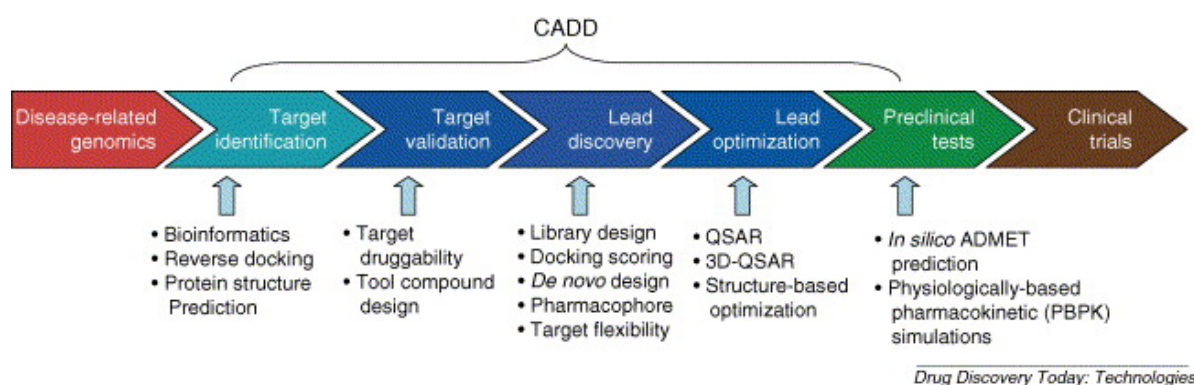


Figure 23: Computer-aided drug design used in drug discovery. Extracted from ¹⁹².

CADD uses computational approaches to discover, identify, optimise and analyse targets and possible drugs. It is an integral component of drug development nowadays, which help to increase the success rate and decrease the consumption of time and cost^{192, 193}. (Figure 23) Generally, the structure of the therapeutic target could be solved through experimental methods, such as X-ray crystallography, cryo-electron microscopy (cryoEM) and NMR¹⁹⁴. However, there are limitations for said methods, such as sample preparation and resolution, which can hamper drug development.

To increase the number of structures, computational molecular modelling uses sequence data (1D), transforming it into structural data (3D), allowing 3D structures of the target. In addition, identification of novel “druggable” binding sites also can be achieved by using computational tools. Moreover, these techniques can be used to predict the interaction between target and compounds. Furthermore, computational methods can help increase hit discovery rate via virtual screening, lead optimisation and predict the binding modes of known inhibitors. Hence, CADD technology has accelerated the development of drugs.

CADD techniques such as homology modelling, atomistic and coarse-grained molecular dynamics (MD) simulations and docking were used in this PhD project. These methods have helped develop proteins with unknown experimental structures, study the intrinsic dynamics of proteins and interactions between receptors and their ligands, identify new binding sites and assess the 'druggability' of ligands. All these methods are described below.

3.2 Homology modelling

Homology modelling¹⁹⁵ is a method for constructing an atomic-resolution model of a target protein based on its amino acid sequence alignment and an experimental three-dimensional structure template of a related homologous protein. This method can efficiently provide an atomic-resolution structural model of a protein of interest when its experimental structure is not solved. This approach avoids the long time and costly experiments to obtain protein structure from methods such as X-ray crystallography, cryo-EM and protein nuclear magnetic resonance (NMR)¹⁹⁴. However, it needs to be emphasized that obtaining the homology model does not replace the experimental structure. It gives an insight more quickly but is not as reliable as an experimental structure and needs validation¹⁹⁴.

Homologous proteins have related primary sequences; therefore, it can be assumed that their tertiary structures should be related, given the associated sequence-structure relationship. To generate reliable, high-quality structural models, the target proteins and template must be closely related, typically, 30% or higher primary sequence identity¹⁹⁶.

The homology modelling workflow is shown in Figure 24, which can be divided into four steps: template selection and identification, target-template alignment, model construction, and model assessment¹⁹⁷.

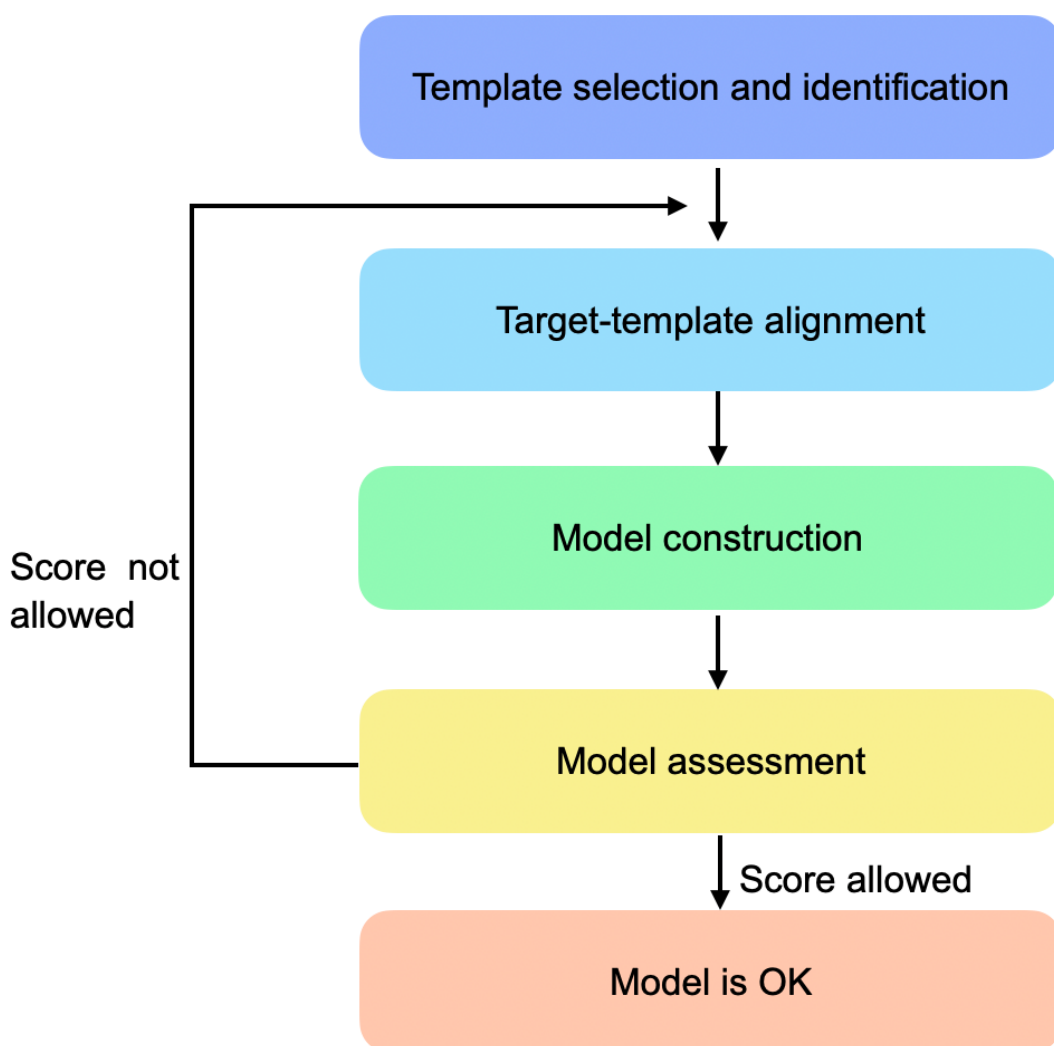


Figure 24: The process of homology modelling. Made by author.

The structural biology information on homologous proteins related to the target can be retrieved from web databases, which the most known one is the RSCB Protein Data Bank (PDB) (<https://www.rcsb.org>)¹⁹⁸. For primary sequences, UniProt (<https://www.uniprot.org>)¹⁹⁹ is one of the most used databases. This database provides protein information, such as sequence and function, and helps select the correct template. The template selection can determine the quality of the model. It needs to consider factors such as overall sequence similarity, percentage of identity in the structurally-conserved regions, resolution of the template, and environmental factors (e.g. crystallization conditions, the presence of small molecules co-crystallized, oligomeric state). To assess the similarity between primary sequences of our target and our

query, Basic Local Alignment Search Tool (BLAST)²⁰⁰ is commonly used. This method performs protein sequence alignment using evolutive similarity matrices, such as BLOSSUM²⁰¹ and PAM²⁰². Hence, with a similar structure to our target, we can build a model. There are several published methods to use: among the most popular, there are fully automated web server-based tools such as SWISS-MODEL. SWISS-MODEL²⁰³ is accessible via Expasy web server (<https://swissmodel.expasy.org>)²⁰⁴. It builds the model and performs the molecular mechanical energy minimization with a fragment assembly approach. It breaks down a protein structure into multiple conserved structural fragments, then picks up these fragments from a template structure and assembles them to build the target structure.¹⁹⁴ Then, the most appropriate target structure can be selected using the QMEAN²⁰⁵ scoring function for the quality estimation. QMEAN^{206, 207} is a composite score that uses the statistical potential of mean force and the agreement of the model with the structural features predicted from the sequence. MODELLER²⁰⁸ is another popular tool widely used for the homology modelling community. It builds target structure by satisfying spatial restraints and can perform other tasks such as multiple alignments of sequence/structure and protein structure optimisation. MODELLER is often combined with other modelling tools, e.g., to model the missing loops that could not be calculated using automated workflows such as SWISS-MODEL. Such a procedure is commonly used: in this project, SWISS-MODEL and MODELLER were used to calculate nicotinic receptors' models and model the missing loops, respectively. In addition, the ISOLDE software²⁰⁹ for structure refinement and Alpha-fold²¹⁰ for AI-driven structure prediction were employed in building the models.

3.3 Molecular mechanics

Molecular mechanics (MM) methods are used to compute the system energy for molecular structural assessment. By large, the physics used in MM is based on classical mechanics. This method is the cornerstone of several methods crucial in modern modelling of biological systems and molecular simulations methods such as molecular dynamics²¹¹, umbrella sampling²¹¹, molecular docking²¹¹ and many others²¹¹. The advantage of MM methods is their speed, applicability to extensive systems containing up to millions of atoms, and accuracy suitable for

those workflows. The key disadvantages are that classical molecular mechanics cannot permit any covalent bond breaking or making – it is not suitable for modelling chemical reactions.

Classical molecular mechanics use the Born-Oppenheimer approximation to calculate the structural descriptors of the systems, such as partial charges and atomic sizes. Because electrons are far lighter than the nucleus, the dynamics of electrons is so fast that they can be considered to react instantaneously to the motion of the nuclei. Based on this, the Born-Oppenheimer approximation considers that the motion of atomic nuclei and electrons in a molecule can be separated. It calculates the potential energy as solely a function of positions of atomic nuclei.

The all-atom classical molecular mechanics method relies on the following principle:

- 1) Nuclei and electrons are simulated as one sphere-like particle
- 2) Each particle is spherical and is assigned a van der Waals radius and a partial charge
- 3) Bonded interactions are modelled using classical potentials, often described as harmonic oscillators or harmonic Fourier series.

The molecular mechanical calculation determines molecular structure and energy using a series of molecular descriptors and functions called a force field.

3.4 Force field

The force field (FF) is a series of descriptors used to model atomic interactions. The basic functional form of potential energy in molecular mechanics contain two terms (Equation 1). One describes bond interactions, including bond lengths, angles and torsions (proper and improper dihedrals). The non-bonded term, on the other hand, describes non-covalent bound atoms. These interactions are often modelled as van der Waals and electrostatic (Coulombic) interactions between atoms (Figure 25).

$$E_{total} = E_{bonded} + E_{non-bonded} \quad (1)$$

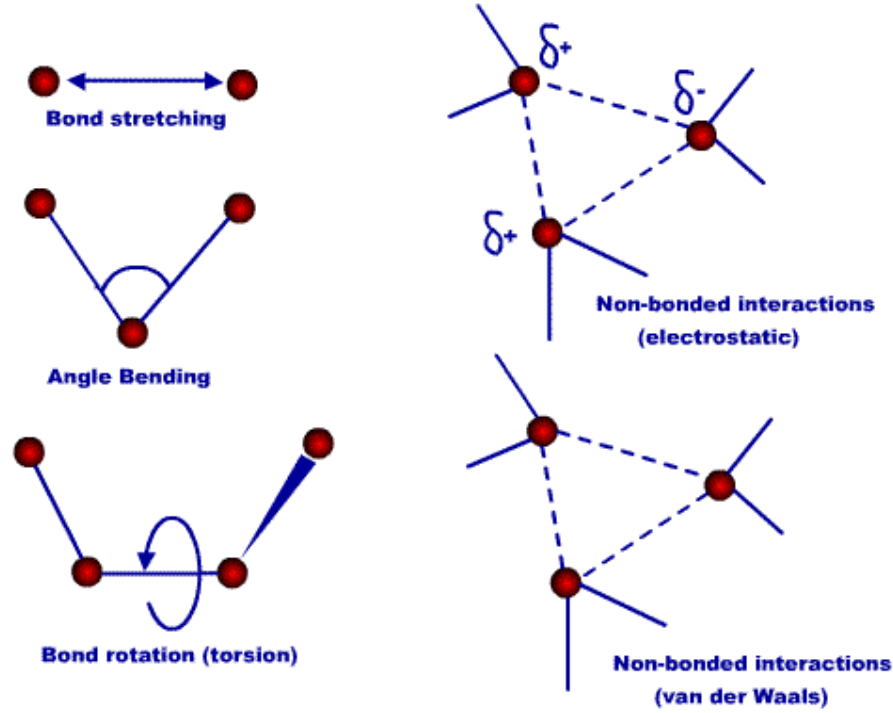


Figure 25: Four critical contributions to a molecular mechanics force field: bond stretching, angle bending, torsional terms, and non-bonded interactions. Extracted from ²¹².

A set of harmonics oscillators models energy in the bonded term of the system of interest as a function of the bond lengths, valence angles and dihedral vibrations²¹³ (Equation 2 and 3):

$$E_{Bonded} = E_{bonds} + E_{angle} + E_{dihedrals} \quad (2)$$

$$E_{Bonded} = \sum_{bonds} k_r (r_i - r_0)^2 + \sum_{angles} k_\theta (\theta_i - \theta_0)^2 + \sum_{dihedrals} k_\phi^{dihe} [1 + \cos(n_i \phi_i - \delta_i)] \quad (3)$$

The first term models all covalent bonds between two bonded atoms. It estimates the bond stretching energy associated with the bond length r_i deviated from the reference value r_0 . The second term is a summation over all bending angles formed by two covalent bonds. Two covalently bound atoms are considered in the bond stretching term, whereas three are in the angle bending term. The contributions of both bond stretching and angle bending are characterized by two parameters: a reference or equilibrium value (r_0 and θ_0) and a force constant (k_i). k_r and k_θ parameters control the stiffness of the bond and angle spring, respectively. These two force constants are proportionality constants, as defined by Hooke's

law. Unique k_r and r_0 parameters are dependent on the particular atom-type pair (e.g., C-O, C-C, C=C.). Often these parameters are obtained from either higher-resolution methods (such as quantum mechanical simulations) or experimental validation (i.e., NMR). The dihedral or torsional energy is also modelled by different means²¹³. Commonly, the used function is a Fourier series, given that the torsional energy landscape is often more complex for modelling than the previous two potentials described. Hence, alongside the equilibrium value and a force constant parameter per term of the series, a phase (δ_i) and a periodicity (n_i) need to be added, to properly model the periodic conditions associated to the torsional landscape.

$$E_{Non-bonded} = E_{van\ der\ Waals} + E_{electrostatic} \quad (4)$$

$$E_{non-bonded} = \sum_{atom\ i} \sum_{j \neq i} 4\epsilon_{i,j} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^6 \right] + \sum_i \sum_{j \neq i} \frac{q_i q_j}{4\pi\epsilon r_{ij}} \quad (5)$$

The non-bonded interactions are calculated for all particles/atoms not bound, albeit in physical proximity. The non-bonded energy calculates the pairwise sum of the energies of all possible interacting atoms, which account for van der Waals interactions and electrostatic interactions. (Equation 4 and 5)

When the distance becomes slightly less than the sum of contact radii of two atoms, the repulsion term occurs due to overlapping electronic orbitals of different atoms. On the contrary, the attractive term happens due to the attractive dispersion force with the longer distance²¹¹. Once these independent two atoms move apart by a few angstroms, the interactions are weakened. These effects are modelled by the classical Lennard-Jones (LJ) 12-6 potential, as shown in Equation 5²¹⁴. The LJ potential is characterized by a repulsive part that models as r_{ij}^{-12} and an attractive part that models as r_{ij}^{-6} . Regarding other parameters: q is the partial charge of the atom i and j , $\epsilon_{i,j}$ is the potential well depth, r_{ij} is the distance between two interacting particles, σ_{ij} is the distance at which the inter-particle potential energy is zero.

The final term in Equation 5 represents the long-range electrostatic interactions. This term model the energy function as the sum of interactions between two partial atomic charges q_i and q_j using Coulomb's law. These charges are described as partial atomic charges to model

quantum level electronic effects into a classical mechanics framework.

Since molecular mechanics-based methods gained popularity over the last 20-30 years, many molecular modelling force fields are in use nowadays. The most popular biomolecular force field for proteins and nucleic acids is AMBER force field²¹⁵, but other force fields are available, such as GROMOS²¹⁶, CHARMM²¹⁵, and OPLS²¹⁷. In this project, AMBER FF99SB-ILDN force field²¹⁸ has been used for the following steps of molecular dynamics (MD) simulations of macromolecules, energy minimisation, conformational analysis, and molecular docking.

3.5 Molecular dynamics

Classical molecular dynamics (MD)^{219, 220} is a particle-based simulation method to study the interaction and time-dependent physical motions and conformational change of atoms and molecules in a system according to Newton's equations of motion. The atoms are treated as hard spheres in atomistic MD simulation, just like in all MM-based methods. The result of the MD simulation is a trajectory that specifies the positions and velocities of the particles in the system over a period of time. The trajectories are determined by solving Newton's equations to calculate the forces acting on each atom, where the potential energy of the system and the force between the interacting particles are calculated using molecular mechanical force field (Chapter 3.4).²²¹ Analysis of these MD trajectories provides a set of various properties of a system of interest, including potential and kinetic energies, conformational ensembles, changes in secondary structure and solvation, radius of gyration and other macroscopic quantities.

3.5.1 Theory behind molecular dynamics

The main idea behind the classical MD is Newton's second law of motion (Equation 6) used to study the dynamics of the system that is obtained via molecular modelling techniques or experimental structure determination.²²¹

$$\vec{F}_i = m_i \vec{a}_i = m_i \cdot \frac{d\vec{v}}{dt} = m_i \cdot \frac{d^2\vec{x}_i}{dt^2} \quad (6)$$

Where F_i is the force acting on the atom i of the system, a_i is the corresponding acceleration, m_i is the mass. \vec{v} and \vec{x}_i are the velocity and position of atom i . As described in this Equation

6, once the forces are acting on each atom in the system, the atoms are translated in space over a period of time. The rate and direction of motion (velocity) are controlled by the forces exerted by the interactions with surrounding particles of the system. Thus, to obtain the information about the dynamics of the system (i.e. the position of each atom in the system in a specified time), the calculation of the forces acting on each atom needs to be performed via calculating the gradient on the potential energy. The forces are calculated via Equation 7:

$$\vec{F}_i = -\vec{\nabla}_i U(t, \vec{r}_1, \vec{r}_2, \vec{r}_3, \dots, \vec{r}_N) \quad (7)$$

Where F_i is the resulting force acting on the atom i of the system, N is the number of atoms in the system, t is a given period time, r is a set of separate pairwise atom distance vectors, U is the total potential energy.

The acceleration of each particle is computed from the known force and mass by the following Equation 8:

$$\vec{a} = \frac{\vec{F}}{m_i} \quad (8)$$

With the acceleration given by Equation 8, the velocities are calculated based on the following relationship, showed in Equation 9:

$$\vec{v}_i(\vec{r}_i t + \Delta t) = \int_t^{t+\Delta t} \vec{a}_i dt \quad (9)$$

Lastly, the positions are calculated from the velocities (Equation 10):

$$\vec{r}_i(\vec{r}_i t + \Delta t) = \int_t^{t+\Delta t} \vec{v}_i dt \quad (10)$$

The MD simulation is performed as shown in Figure 26 by solving the integral in equation 10. Repeating the steps shown there, a set of coordinates will be generated at each small time step, advancing the MD through time.

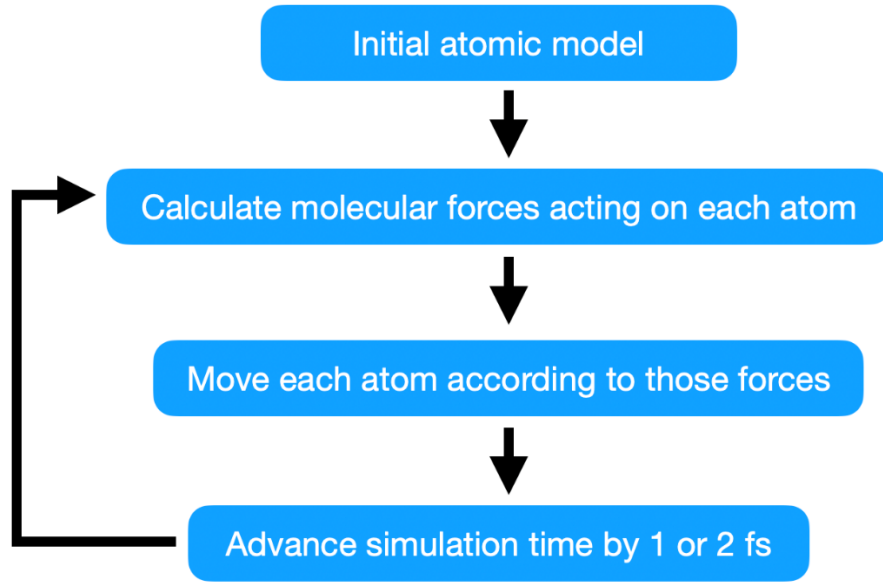


Figure 26: The integration cycles of a molecular dynamics simulation. Extracted from ²²².

The trajectories can be produced by integrating a series of subdivided sub-states of small timesteps using an integration algorithm such as the Velocity Verlet algorithm²¹¹. Velocity Verlet algorithm²²³ is a numerical method that is most commonly used to integrate Newton's equations of motion and calculate the trajectories of a system in atomistic MD simulation. It generates a series of the coordinates and the velocity for intermediate times Δt via (Equation 11 and 12):

$$\vec{x}(t + \Delta t) = \vec{x}(t) + \vec{v}(t)\Delta t + \frac{1}{2} \vec{a}(t)\Delta t^2 \quad (11)$$

$$\vec{v}(t + \Delta t) = \vec{v}(t) + \frac{1}{2} (\vec{a}(t) + \vec{a}(t + \Delta t))\Delta t \quad (12)$$

The process for Velocity Verlet algorithm integration is shown in Figure 27. It is accurate, fast and stable, and easy to implement. It requires only one force evaluation for each time step.

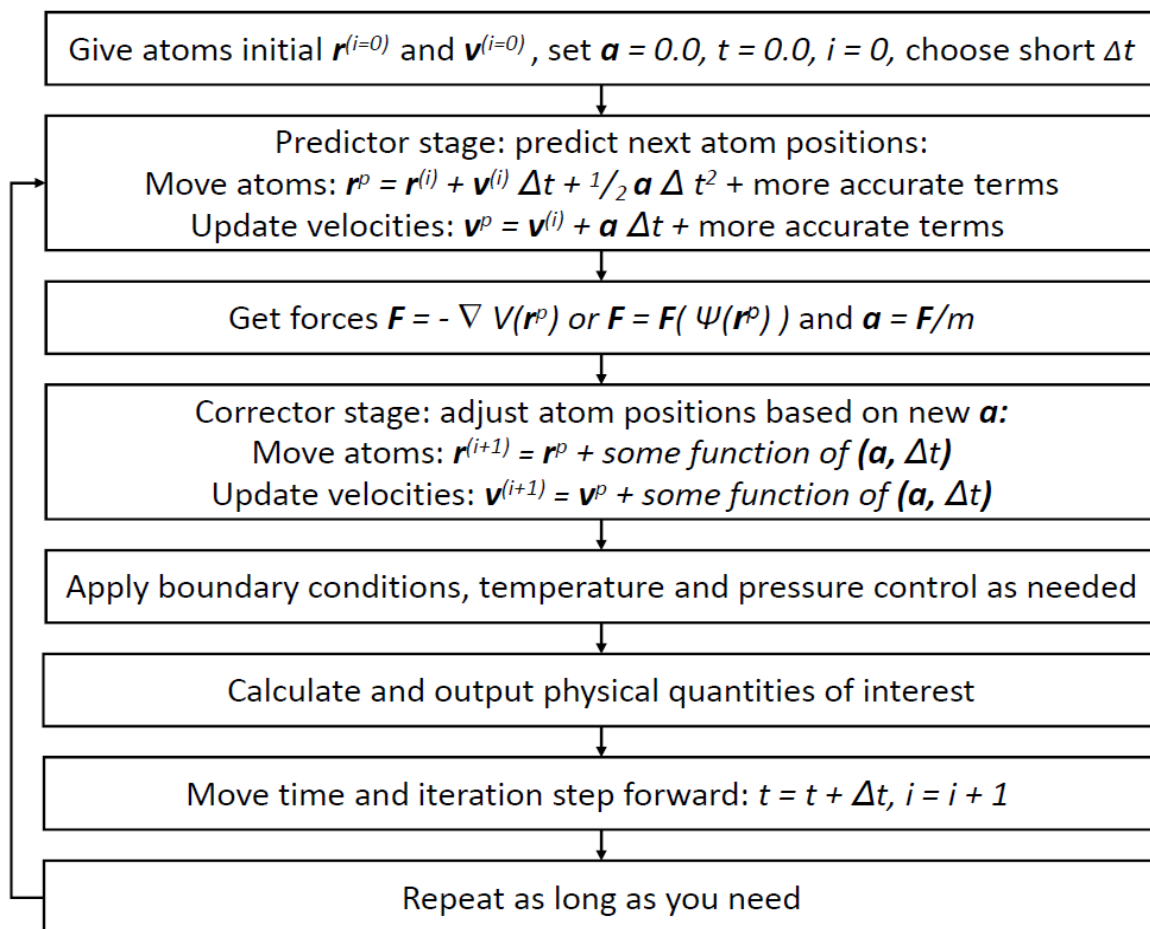


Figure 27: Molecular dynamics Velocity Verlet algorithm. t , simulation time; dt iteration time; For each spatial coordinate of the N simulated atoms (i): x , atom coordinate; F , forces component; a acceleration; m , atom mass; v , velocity.

3.5.2 Conditions in molecular dynamics simulation

Besides the force field described in Chapter 3.4, several other conditions need to be considered in MD simulation. Suitable conditions should allow the calculation to be completed in a reasonable timescale, obtaining the results compared with experimental data. Here are the conditions that should be taken into account in the standard simulations of biomolecular systems:

1. Simulation time scales

The number of particles in the simulation system should be between $10^3 - 10^8$; the range of time step $\Delta t = 1 - 4$ fs, mostly is 2 fs, and the range of time scales ns to μs (typically 100ns - 1 μs),

but a much longer time step can be used in coarse-grained simulations. In all-atom MD, because it used higher frequency transformations, the time step (the rate of calculations - 2fs) is small to be able to obtain a continuous distribution of sampled states through time. Given the fact that CG deals with larger fluctuations, related to a collective motion of atoms represented by the beads, the signal reception frequency can be larger.

2. Solvation effects

MD simulations of biomolecular systems are typically performed in an aqueous solvent. Water molecules fill the entire simulation box, where the protein of interest is often placed in its centre (Figure 28). This organization can mimic the interaction between the molecule of interest and the solvated environment – either aqueous or not. An example of an apolar environment is membrane-bound systems, such as the one depicted in Figure 28. This model is a nicotinic receptor 3D model embedded into the membrane model surrounded by water molecules. There are two ways to model the solute-solvent interaction; one relies on the implicit solvent that can be calculated using a mean-field approach, while another relies on the explicit solvent (TIP3P). Nowadays, the latter is far more used in MD simulation (and is used in this work). Water molecules are described similarly to other molecules for the explicit solvent: as spheres connected by unbreakable bonds. The water molecular interactions with other particles are calculated via the non-bonded term of the force field.

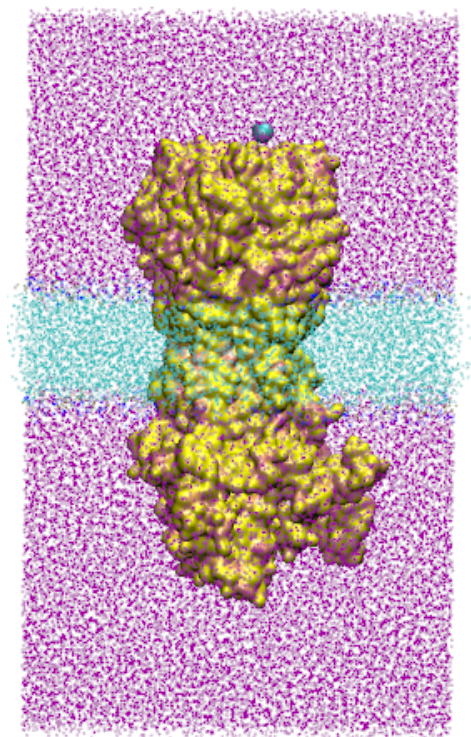


Figure 28: The system model for MD simulation. The $\alpha 7$ nicotinic receptor embedded into the membrane with water fill into the box. Pink dots represented the water atoms; blue dots represented the membrane; the surface of $\alpha 7$ nicotinic receptor was yellow; the Ca^{2+} was represented by blue ball.

3. Periodic boundary conditions

Periodic boundary conditions (PBCs) are a set of boundary conditions that allow the system to loop one side of the simulation back to the other side (Figure 29). They are used to avoid the effects of finite size effect and keep the number of molecules constant in the box. In all kinds of molecular dynamics simulations, the size of the simulation box must be large enough to avoid the interference of boundary conditions.

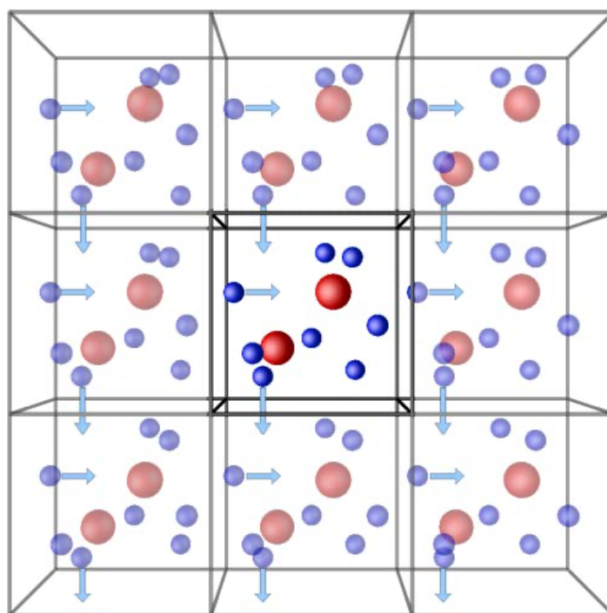


Figure 29: 3D representation of periodic boundary conditions. The central cell represents the simulation box. Blue and red balls represent solvent particles. The blue arrows show the movements of particles. As a particle leaves the simulation box, it goes back to the box from the opposite end. Extracted from <http://isaacs.sourceforge.net/phys/pbc.html#fpbc>.

4. Ensembles

Certain thermodynamics variables need to be set to consider the simulation to be in a statistical mechanical ensemble. Often, the system is isolated, and in adiabatic conditions, i.e., it cannot exchange particles and energy with the environment so that the system has a constant number of particles (N), constant volume (V) and constant energy (E). This is called a microcanonical ensemble (NVE). Also, another set of variables that can be set constant for our systems are the number of particles (N), pressure (P) and temperature (T), which forms an isobaric-isothermal ensemble. These two ensembles are often in biomolecular simulations since it mimics configurations which are used in laboratory conditions.

NVT ensembles are used to alter the temperature of the system by modifying the velocities of particles in the system and maintains the temperature by a thermostat added to exchange the energy of the endothermic and exothermic process. Besides a thermostat added to maintain the temperature, Isothermal-isobaric ensembles (NPT) has a barostat to maintain the pressure by changing the volume of the simulation box. The Berendsen thermostat²²⁴ and the Parrinello-

Rahman barostat²²⁵ are commonly used in NVT and NPT and the ones used in this work.

3.5.3 Stages of molecular dynamic simulation

A typical atomistic MD simulation involves several stages, which are shown in Table 3. Before running the MD simulation, the initial conformation of a system of interest (a protein or a complex) is generated as described in Chapter 3.2.

Table 3: Stages of a typical MD simulation

| Stage | Details |
|--|--|
| Construction of a molecular model | assembly of the molecular scaffold of the protein structure |
| Determination of partial atomic charges and assigning of the force-field | Calculation of the partial atomic charges if needed (e.g., for non-standard residues), and choosing a suitable FF to calculate bonded and non-bonded interactions |
| Simulation system setup | <ol style="list-style-type: none">1. Building the simulation box (shape, size)2. Specification of the boundary conditions3. Solvating the system (adding water and counter-ions) |
| Energy minimisation | Adjust the structure to the best starting configuration for simulation (e.g., removing steric clashes and torsional strains) |

| | |
|--------------------------------------|---|
| Heating and NVT equilibration | Linear heating of the system from an initial T (0 K) to the target temperature (typically 300 K), followed by a short equilibration at a constant temperature. In this stage, atomic position restraints are often used to ensure that the heating process does not disrupt protein structural integrity. |
| Equilibration (NPT) | Equilibration to a target pressure is followed by equilibration at constant pressure. Similar to the heating stage, position restraints are added. |
| Production | Sample structural configurations. Several replicas of MD trajectory would be run for statistically significance. |
| Analysis | Determination of properties of interest from the production trajectory |

3.6 Coarse grain simulations

Some molecular systems may contain many particles, often within the scale of tens of millions of atoms. This can be challenging for studies using all-atom MD methods due to the large number of computer resources required and the long times needed for equilibration and meaningful observables (e.g., significant conformational changes) to occur. Hence, it is necessary to use alternative reliable methods to provide insights into the ‘slow’ (micro-to-millisecond time scale) functional dynamic behaviour of large macromolecules and their complexes. One of such methods is coarse-grain protein modelling.

3.6.1 Coarse-grain modelling

Coarse-grained (CG) modelling²²⁶ uses a simplified representation to represent and simulate the behaviour of complex systems. It is widely used for modelling large biomolecules, such as large proteins, nucleic acids, lipid membranes and certain post-translational modifications (e.g., long carbohydrate chains) in large multimeric complexes. In the coarse-grained framework, functional groups such as aromatic rings or even the whole amino acids are represented as ‘pseudo-atoms’ instead of individual atoms being considered. For example, the protein polypeptide chain is replaced by one or two pseudo atoms (so-called united atoms or molecular beads) per residue. By decreasing the number of degrees of freedom, the coarse-grained model allows the system's dynamics to be quickly calculated and accelerate simulation compared to the all-atom framework. This change of model resolution results in longer simulation times than atomistic models and allows larger models to be effectively simulated. The comparison between coarse-grained framework and other modelling methods, focusing on their spatiotemporal scales, is showed in Figure 30.

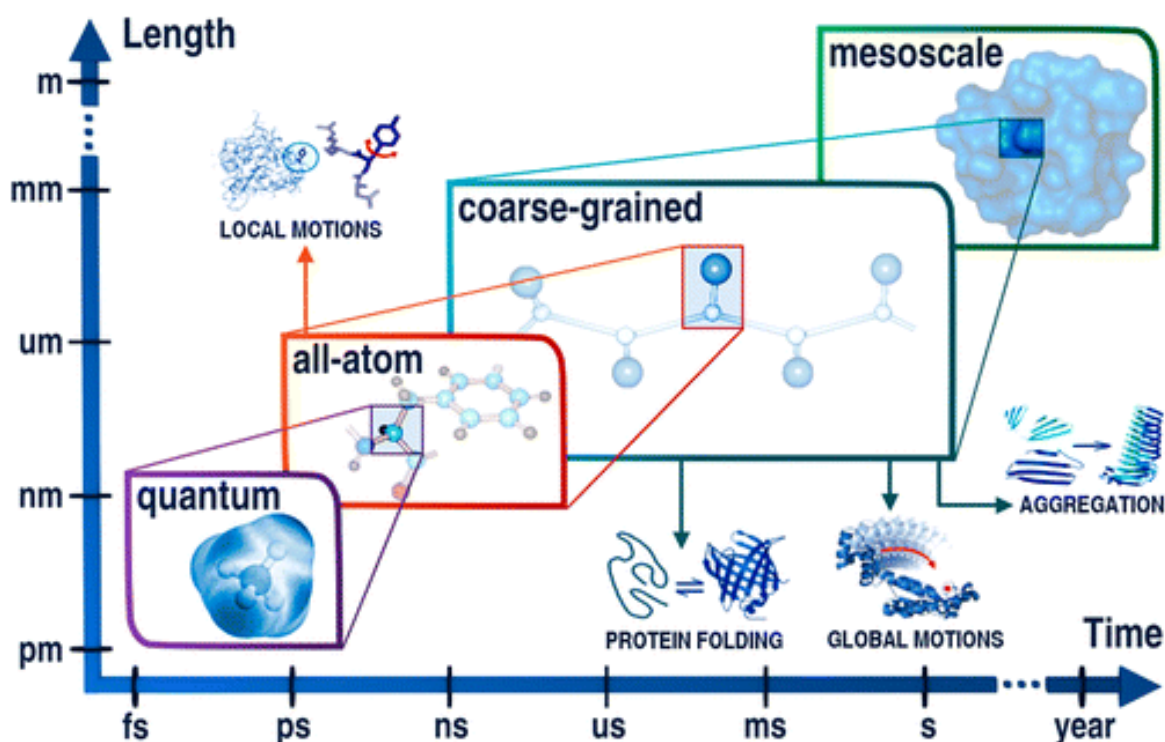


Figure 30: Application range for different molecular modelling. This plot shows the approximate range of time scales correspond to the system size. Extracted from ²²⁶.

3.6.2 Martini force field

Comparing to the all-atom FF counterpart, the coarse-grained force field smoothens out the energy landscape, which is more efficient to sample over multiple minima and avoid traps in the local minima²²⁶ (Figure 31).

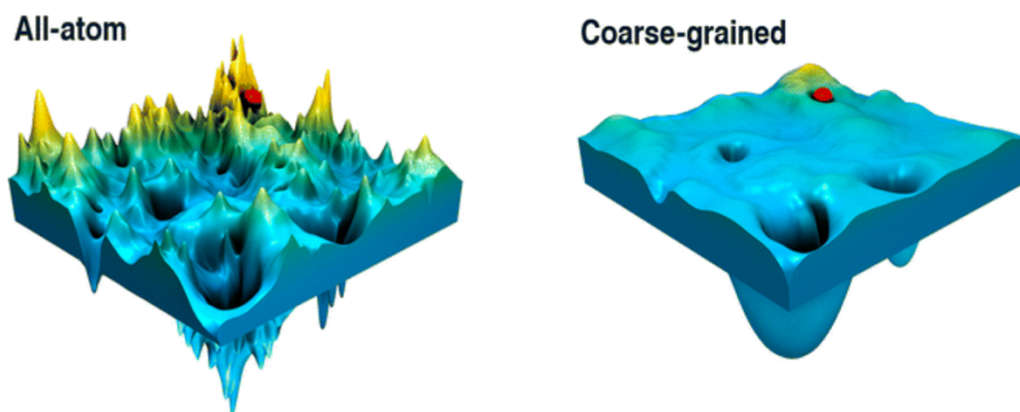


Figure 31: The potential energy landscape of all-atom and coarse-grain simulations. The Z-axis represent free energy, the X-axis and Y-axis represent degrees of freedom, the yellow peak represents the local maximum energy, while the hole represents the local and global minimum energy. The free energy would be calculated for the left figure based on all-atom and the right for the coarse-grained system. This difference enables the energy landscape of global minima to be found efficiently while avoiding the local minima pitfalls, although losing resolution on the landscape. Extracted from ²²⁶.

The Martini force field²²⁷ is one of the most popular coarse-grained force fields chosen for molecular dynamics (MD) simulations of biological systems, well suited for proteins-membrane systems. It was parametrized in a systematic way to reproduce free energy and thermodynamic properties²²⁷. This FF is based on a four-to-one mapping, which means that, on average, four heavy atoms and their respective hydrogens are represented by a single bead (Figure 32). For example, four water molecules correspond to one coarse-grained water bead. To keep the accurate representation of the underlying atomistic structure, four bead categories have been defined to correspond to four main types of interaction, which are polar (P), non-polar (N), apolar (C) and charged (Q).²²⁶⁻²²⁸ Each bead type is divided into hydrogen-bonding capabilities (donor, acceptor, both or none) and five degrees of polarity, given 18 bead types in total.²²⁶⁻²²⁸ In addition, besides normal size four-to-one mapping, it has a small size (S-size)

three-to-one mapping and tiny size (T-size) two-to-one mapping used in the ring structure and nucleic acids, respectively. These mapping schemes provide an accurate and effective way to switch atomistic to coarse-grained systems.

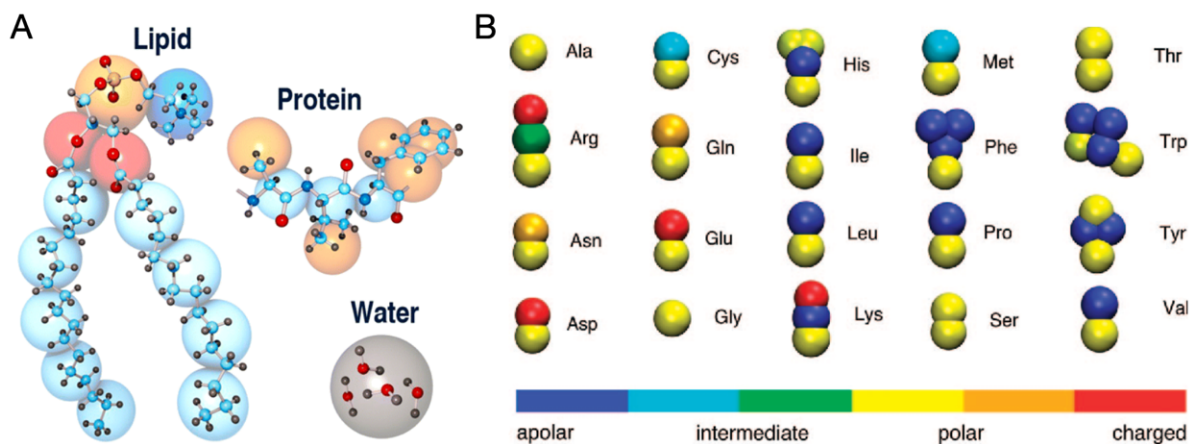


Figure 32: Coarse-grained representation of all-atom structure (lipid, protein, water) and all standard proteins residues. All-atom models are represented in balls and sticks, while coarse-grained models are shown in a large sphere. Different colours of the sphere represent different particle types. Extracted from^{226, 228}.

As for the interaction between coarse-grained beads will be described by a force field containing terms typical of other classical force fields. Bonded interactions are modelled by potential energy functions derived from either higher resolution models (QM/MM) or experimental data.²²⁶ Nonbonded interactions include van der Waals interactions and electrostatic interactions controlled by a Lennard-Jones 12-6 potential and Coulombic potential function, respectively, but the parameters are adjusted depending on the types of interacting beads.²²⁶

3.6.3 Elastic network modelling

The standard coarse-graining generally describes the interactions per-residue (bonded bead), generating a simplified backbone. The fine details of a side chain and protein secondary structures controlled and maintained by H-bonds are lost in the coarse-graining process. Therefore, it may be required to add extra harmonic bonds between non-bonded beads to constrain secondary structure elements (alpha-helices and beta-sheets). The elastic network

modelling^{226, 229, 230} is a minimalist approach that takes the form of an elastic spring network as an internal interaction in biological macromolecules, which help to increase the structural stabilization of the overall shape coarse-grained models, as shown in Figure 33. The elastic network helps to tune the elastic bonds by set the elastic bond force constant and adjust the lower and upper elastic bond cut-off to reduce the flexibility of coarse-graining over the entire protein structure, and describe protein collective motions adequately.^{229, 230} Sen and colleagues²²⁹ used various cut-off radii for multiple scales (e.g., residue, atomic, proton and explicit solvent levels) of the high-resolution protein dataset to define parameters that should obtain the most realistic outcomes, which defined the parameters used in this work. This study indicated that the atomic level elastic network model provides an improved dynamic for a variety of proteins than the coarse-grained model.²²⁹

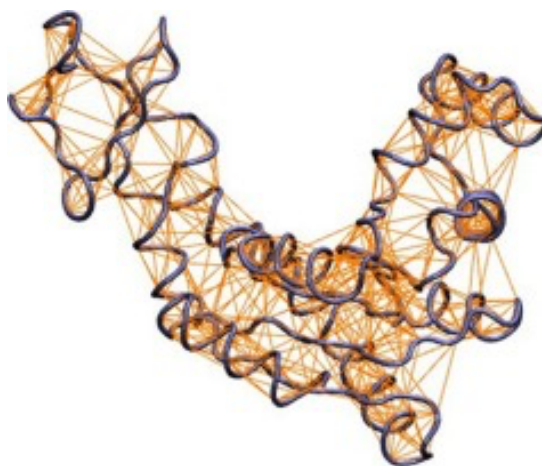


Figure 33: The elastic network model. Orange lines represent the elastic network, while the grey tube is an example protein backbone. Extracted from²³⁰.

Elastic networks are usually applied to coarse-grained models. They are mainly using $C\alpha$ atoms and relying on $C\alpha$ - $C\alpha$ proximity to place the springs. Two parameters are needed to describe the reaction correctly with a proper network: the cut-off distance and the spring constant. The former determines whether two residues are in contact by a spring. The latter defines the strength of interactions²²⁹.

Coarse-grained models with elastic network approximations can describe both inter-and intramolecular interactions in the system, which provide helpful simplified structural models for stabilizing inter-and intramolecular interactions. Since the nicotinic receptor is a substantial

transmembrane protein composed of a series of α -bundles, it required a very long time in all-atom MD simulations, coarse-grained framework with Martini force field has been chosen for my studies reported in this work.

3.6.4 Coarse-grained protein simulation running steps

The setup and running of a coarse-grained protein simulation are similar to an all-atom MD simulation. The main difference comes from how the topology and the molecular parameters are set for the protein. Often, Martini FF used a script developed by the creators of Martini FF that translates the protein from all-atom and parametrizes it in CG. Concerning the steps for equilibration, two additional are required at the beginning. First, the atomistic protein structure needs to convert into the coarse-grained model before the simulation starts. Apart from this, the other steps are the same as for the all-atom MD simulation but using the files associated with the coarse-grained FF. The all-atom information can be recovered at the end of the coarse-grained simulation using Reconstruction Algorithm for Coarse-Grain Structures (RACOGS)²³¹. The process is shown in Figure 34.

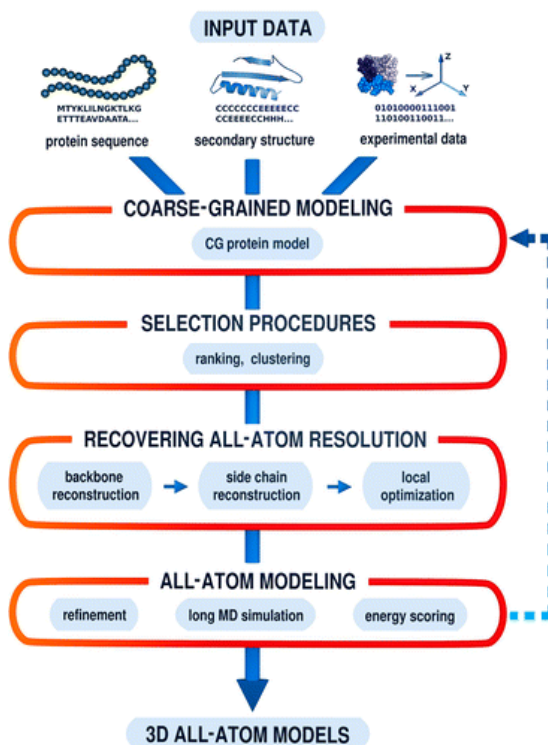


Figure 34: The steps of coarse-grained simulation and the follow-up by the all-atom MD simulation. Extracted from ²²⁶.

3.7 Umbrella sampling

Due to the limitation of MD simulations in the timescale, it is hard to evaluate some significant conformational changes, which requires energy barriers to be crossed. To help with this shortcoming, a series of methods were developed through the years to drive a simulation towards interesting rare states. One of such methods is Umbrella sampling.

Umbrella sampling (US)^{232, 233} is one of the techniques known as biased molecular dynamic by providing an additional free energy term to the system, which ensures sampling along the chosen (one or more dimensional) pre-selected degree of freedom (DOF) of the system. These DOFs, also called reaction coordinates, are a series of conjoined variables that define a motion within the system we want to investigate. Commonly, a continuous parameter describes the different states as it changes from one thermodynamic state to another, defined according to geometric principles, such as distance and torsion.

The intermediate steps of a conformational transition are split into a series of windows, shown in Figure 35. To overcome free energy barriers and accelerate the dynamics, the bias potentials are added in each window to move a system from one thermodynamic state to another by exerting an external, non-equilibrium force on the defined reaction coordinate. Performing MD simulations for each window allows the free energy change to be calculated as (Equation 13):

$$E^b(r) = E^u(r) + \omega_i(\xi) \quad (13)$$

Where b indicates a biased quantity, whilst u indicates an unbiased quantity. ω_i is the bias potential of the window i , which is an additional free energy term, depending on the reaction coordinate ξ , and its strength is defined by the constant force k . This harmonic potential can be calculated via Equation 14:

$$\omega_i(\xi) = \frac{k}{2}(\xi - \xi_i)^2 \quad (14)$$

Where ξ_i is a reference coordinate point. A bias potential is applied in each window to keep the system close to the reference point of window i . This results in the system in each window would be restricted to sampling a narrow phase space along with the reaction coordinates to

allow an overlap of the potential energy distribution between adjacent windows, forming a continuous distribution of states along the sampled coordinate.

To obtain the unbiased free energy $A_i(\xi)$, the unbiased distribution of the reaction coordinate is calculated via Equation 15:

$$P_i^u(\xi) = \frac{\int \exp[-\beta E(r)] \delta[\xi'(r) - \xi] d^N r}{\int \exp[-\beta E(r)] d^N r} \quad (15)$$

Then the unbiased probability $P_i^u(\xi)$ could be determined by Equation 16:

$$P_i^u(\xi) = P_i^b(\xi) \exp[\beta \omega_i(\xi)] < \exp[-\beta \omega_i(\xi)] > \quad (16)$$

The biased probability $P_i^b(\xi)$ can be known from the simulation in each window. Then the free energy of each window can be computed using Equation 17:

$$A_i(\xi) = -\left(\frac{1}{\beta}\right) \ln P_i^b(\xi) - \omega_i(\xi) + F_i \quad (17)$$

Where F_i is a constant that can be calculated using the weighted histogram analysis method (WHAM)²³⁴.

Finally, the free energy changes between the two states along with the reaction coordinates, known as the potential of mean force (PMF), can be calculated by combining the unbiased free energies extracted from a series of windows. An example of the final PMF curve and its related sampled windows are showed in Figure 35. This method was used in this work to study the nAChR ion channel function.

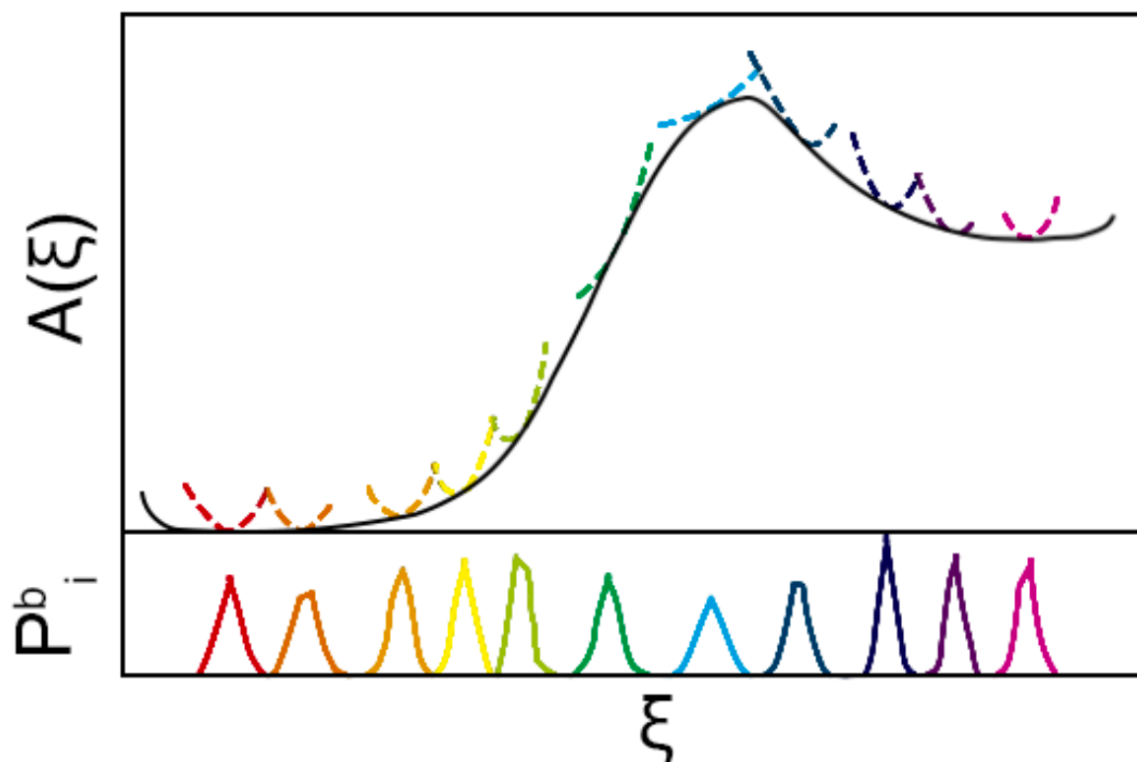


Figure 35: The representation of the PMF curve and its related sampled window. The free energy landscape is represented in black, and the sampled population of each window are represented by thick solid curves and thin dashed curves, respectively. The biased distributions P_i^b obtained from the simulation of each window are shown at the bottom, represented by thin solid curves. Extracted from²³².

3.8 Druggability assessment: FTMap solvent mapping

FTMap^{235, 236} is a mapping web server used to search and predict binding hotspots for proteins based on ‘static’ structure (e.g., PDB format) (Figure 36). After providing a docking pre-processed protein structure in PDB format (all bound ligands and solvent molecules are removed), FTMap uses 16 simple organic molecules (Figure 36b) as probes of diverse sizes, shapes and polarities. They interact at suitable spots on the protein surface and then cluster and sequentially ranks based on their final conformation average interaction score. The algorithm of FTMap is based on the Fast Fourier Transformation (FFT) correlation method²³⁵, which facilitates efficient sampling of billions of small molecule probe positions on translational and rotational grids, and uses rigid-body docking to generate binding positions (Figure 36a).²³⁵ Probe positions are scored according to energy expressions, including van der Waals (VDW) terms, electrostatic interaction terms, cavity terms and structure-based pairwise interaction

potentials.²³⁵ The position containing the most significant number of different probe clusters are the main hotspots, which would be the most likely binding sites of the protein, as shown in Figure 36a as an example. Compared to the same system sampled in real space, the FTMap calculation for an average size protein takes at least 1000 times less time than even on large computing systems.²³⁵ Due to its excellent speed and robustness, FTMap has become one of the most popular methods for detecting allosteric binding sites, playing an essential role in fragment-based drug design.

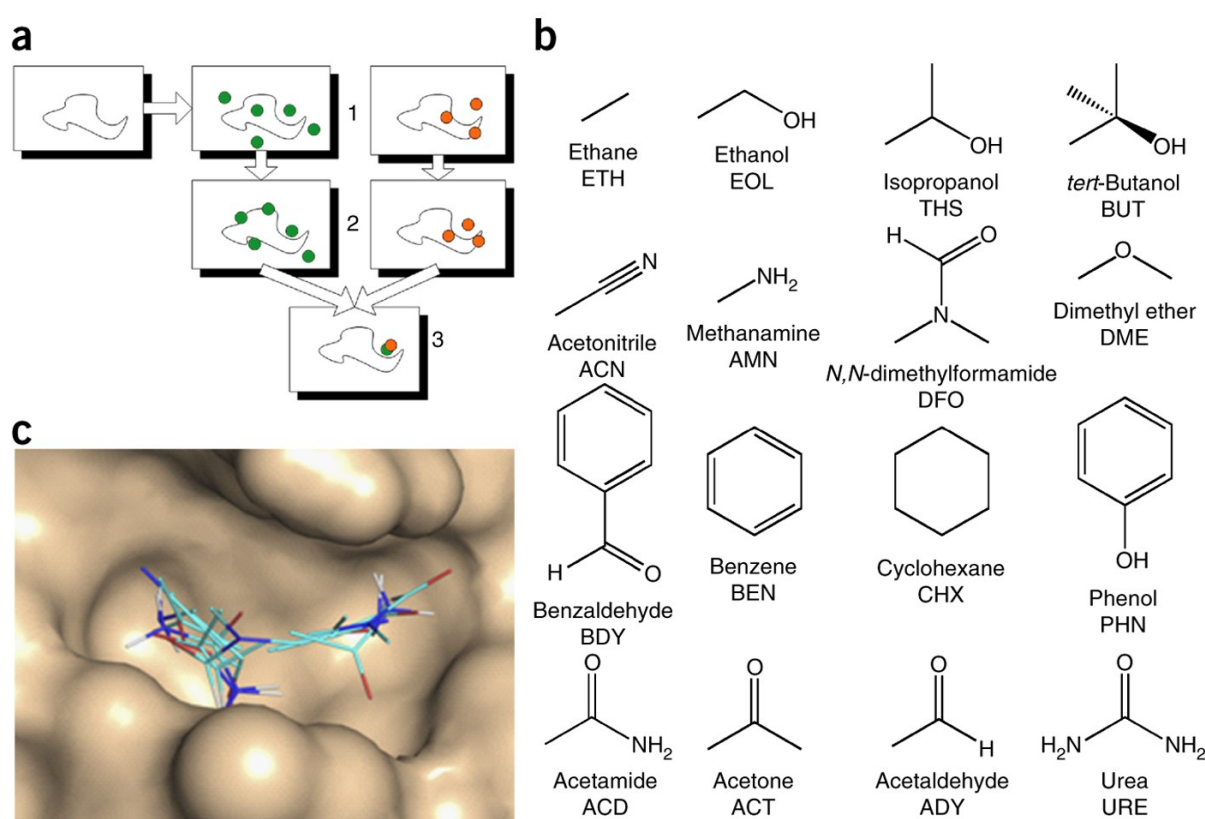


Figure 36: Principles of FTMap. (a) The process of FTMap docking. The small molecule ligands find the favourable binding site by energy-minimisation, and then difference ligands are clustered to generate the ligand clusters, the sites with the most significant number of ligands clusters are the main hot spots. (b) Sixteen simple organic molecules are used as a probe in FTMap. (c) An example of probe clusters in a target pocket. Extract from ²³⁵.

3.9 Molecular docking

Molecular docking^{237, 238} is used to predict the interaction between ligands and proteins at the atomic level, which is the technique widely used in structure-based drug design. The docking process involves two steps: identifying the conformation of the ligand and assessing the affinity

between ligand and target.

Before molecular docking, a selection of binding sites is first required, and then molecular docking generates several different conformations and allows the adoption of the appropriate configuration to fit the predefined region in the receptor.

To score the poses to select the most realistic one, we often use a scoring function. Commonly, this score relies on the binding energies between the receptor and the ligand pose: docking energy scores between the ligand molecule and the target receptor, electrostatic energies, Lennard-Jones potentials and H-bonds. Also, the pose generation can be done via several algorithms, either treating the ligand as a rigid body without torsions or as a fully flexible body²¹¹.

Several software packages are available for docking, and each one has a different scoring function, such as force-field-based scoring function (AutoDock and DOCK) and knowledge-based scoring functions (SeeSAR).

The force field-based scoring function obtains the binding energy by accounting for the sum of the non-bonding interactions, including receptor-ligand interaction energy and internal ligand/protein energy. The former includes the intermolecular electrostatic and Lennard-Jones components (van der Waals interaction), while the latter includes the bond stretching/bending/torsional. To accelerate the pose formation and scoring, often, the receptor is described as a potential grid, reducing it to precalculated potential points in a 3D box (Figure 37). The box size is often defined by the distance from the ligand centre-of-mass (COM) to the edge of the box. The grid can identify whether the ligand atoms overlap significantly in space with the receptor atoms. The grid generation involves calculating Coulomb (electrostatic) and Lennard-Jones (VDW) energies for each point on the grid around the target.

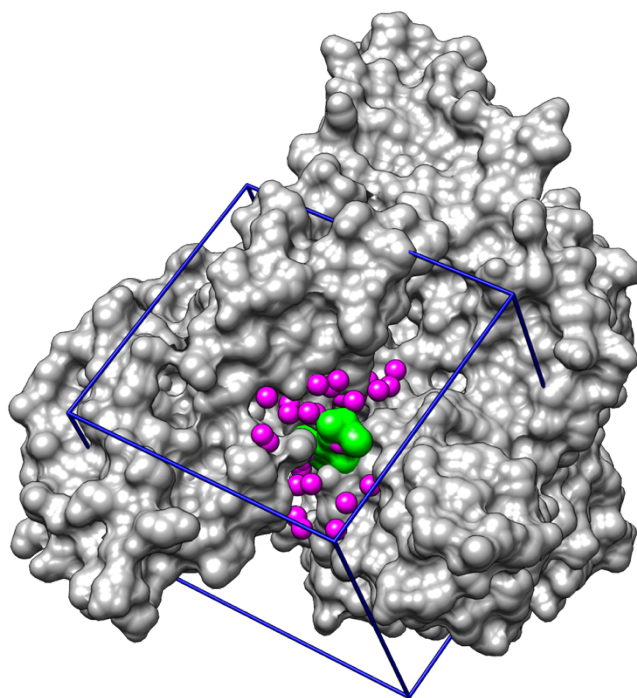


Figure 37: Visualisation of the DOCK6 box in UCSF Chimera²³⁹. The pink beads represent selected spheres; the grey surface represents the receptor. The blue cubic box represents the grid box. The green surface represents the ligand binding site. Extracted from http://ringo.ams.stonybrook.edu/index.php/2019_DOCK_tutorial_3_with_PDBID_3JQZ

Knowledge-based scoring functions obtain interatomic contact and/or distances between ligands and proteins by statistically analysing ligand-protein complexes known structures and binding data. It can be used to screen large complex databases and simulate uncommon interactions such as sulphur-aromatic or cation- π ²³⁷, but has the problem of interactions being underrepresented in a limited training set of structures. One type of knowledge-based scoring function is HYDE; it is used within SeeSAR, based on reproducing experimental structures rather than binding energy. This method will be described in Chapter 3.10.

In this work, UCSF DOCK6 has been selected to be used due to a reliable and reliable pose prediction²⁴⁰. DOCK6 starts by filling the solvent-accessible receptor surface with overlapping spheres to drive the docking localization. From all generated spheres clusters, a subset is selected to represent the binding site. This subset is then assigned to the ligand atoms. (Figure 38) Once matched, energy minimisation is used to optimise the interaction between the ligand and the receptor. During this time, the ligand is considered flexible whilst the receptor is considered rigid. The quality of binding geometries (poses and modes) is often determined

using an AMBER scoring function.

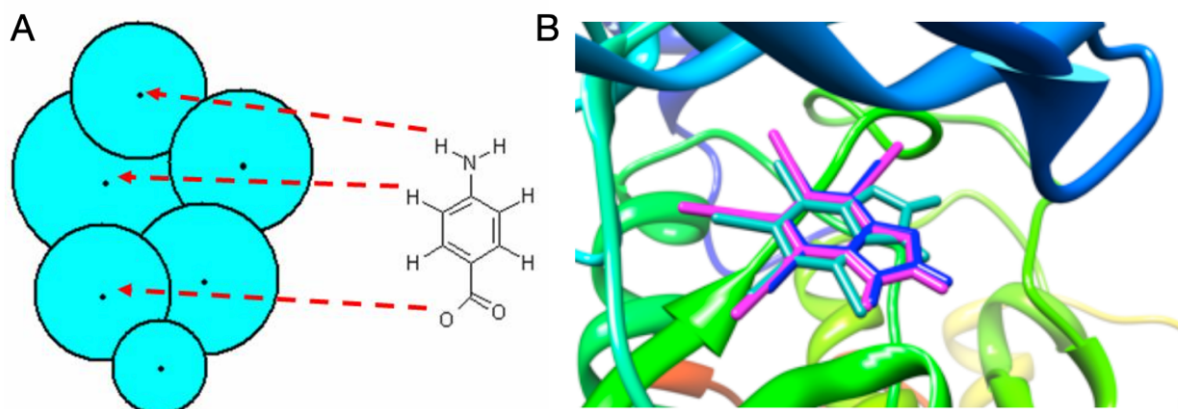


Figure 38: Docking the ligands to the receptor. A) The most extensive rigid substructure of the ligand is paired at the binding site by the alignment of the heavy atom to the receptor sphere. B) The different conformations of ligands bind to the receptor. Extracted from http://dock.compbio.ucsf.edu/DOCK_6/dock6_manual.htm.

3.10 Scoring function for SeeSAR - HYDE

HYDE (HYdrogen bonding and DEhydration algorithm) is the scoring function to evaluate the approximate affinity of a ligand in a binding pocket through calculating the free energies of ligand binding^{241, 242}. The visual output score can help to predict ligand-protein. HYDE binding assessment bases on two factors: desolvation and interaction.

HYDE estimates and sums hydrogen bonding energy and dehydration energy for each atom in protein-ligand complexes (Equation 18)^{241, 242}. The primary feature of this scoring function is the use of logP-derived atomic increments for calculating the change in hydrogen bonding $\Delta G_{H-bonds}^i$ and dehydration energy $\Delta G_{dehydration}^i$ between unbound state and bound state.²⁴¹ The negative value is favourable, whilst the positive value is unfavourable.

$$\Delta G_{HYDE} = \sum_{atoms\ i} (\Delta G_{dehydration}^i + \Delta G_{H-bonds}^i) \quad (18)$$

The logP (logarithm of the octanol/water partition coefficient) plays an essential role in this estimation. A more positive logP value indicates higher lipophilicity and affinity for the organic phase, whereas a more negative logP means a more hydrophilic substance with a higher affinity for the aqueous phase. The dehydration energy calculation relates to whether the atom is

hydrophobic or hydrophilic, the former atom contributes favourably to the overall binding energy, whilst the hydrophilic dehydration is unfavorable²⁴¹. The calculation about hydrophobic atom (Equation 19) includes the change in solvent accessible surface Δacc^i of an atom i , its logP increment $p \log P^i$.

$$\Delta G_{dehydration}^{i,hydrophobic} = -2.3RT \cdot p \log P^i \cdot (acc_{unbound}^i - acc_{bound}^i) \quad (19)$$

For hydrophilic atoms, it changes to the molecular surface area to consider and is assessed by whether there is space for a water molecule in the preferred direction of a hydrogen bond. This is shown in Equation 20:

$$\Delta G_{dehydration}^{i,hydrophilic} = -2.3RT \cdot p \log P^i \cdot f_{bur}^i \cdot f_{water}^i \cdot \sum_{H-bond\ functions\ j} w^j \cdot p \quad (20)$$

The weights w^j is for multiple hydrogen bonds, which can be formed by a single hydrophilic atom, and the factor f_{bur}^i is a scaling factor related to how buried the atoms of a hydrophilic group is in the unbound state. In addition, the other factor f_{water}^i is a correction factor that accounts for the local arrangement of water around the hydrogen bond. This factor is due to the ligand with many adjacent polar groups restricted the arrangement of water, which reduces the loss of hydrogen bonding function for dehydration.

The hydrogen bond energy calculation is similar to the dehydration term, shown in Equation 21:

$$\Delta G_{H-bond}^i = \frac{2.3RT}{F_{sat}(T)} \cdot p \log P^i \cdot f_{bur}^i \cdot \sum_{H-bonds\ j} w^j \cdot f_{dev}^j \quad (21)$$

One of the differences in the factors between equations 20 and 21 is the saturation factor $F_{sat}(T)$. This parameter indicates the incomplete saturation of the water hydrogen bond network at a specific temperature. Another different is f_{dev}^j which determines the geometrical quality of a hydrogen bond j . The reason for considering this is that the energy of a hydrogen bond decreasing with the deviation of hydrogen bonding geometry between donor and acceptor.

SeeSAR²⁴³ is a software developed by BioSolveIT which estimates the binding affinities between protein and ligands using the HYDE score. SeeSAR can visualise and quantify the

HYDE score of each atom in the ligand, as well as the respective contributions from desolvation and protein-ligand interactions (Figure 39). In addition, SeeSAR gives a logarithmic estimate of the binding affinity to reflect the binding situation. According to these visualised data, SeeSAR is a method that greatly facilitates the analysis of the binding affinity of ligand-protein.

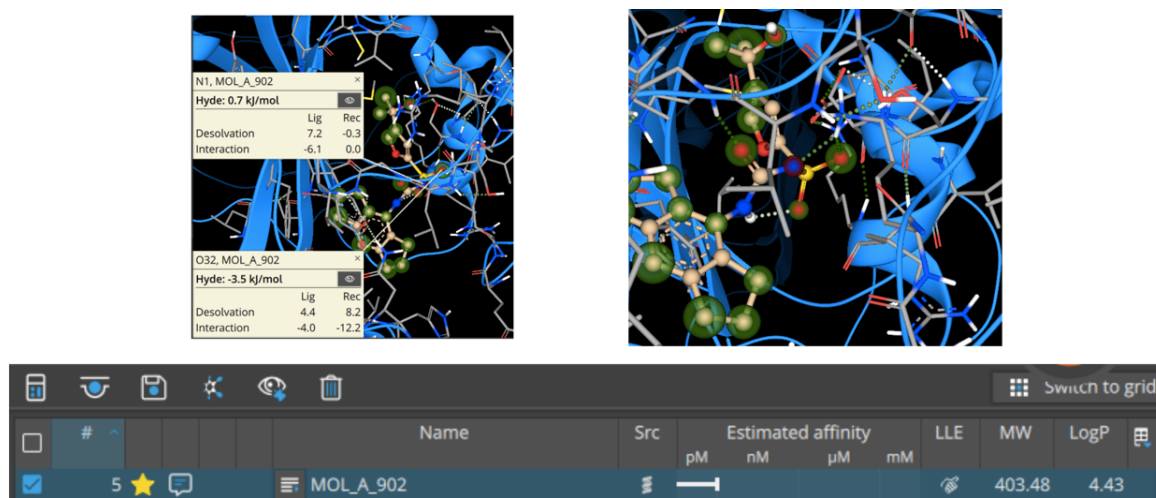


Figure 39: The example of SeeSAR analysis. The protein is represented in blue ribbon. The protein side-chain is signed by a grey bond, with an H atom in blue and an O atom in red. The green circle signs the atom of the ligand with favourable HYDE, whereas with unfavourable is signed by the red circle. The atom with no colour circle means no contribution to ΔG_{HYDE} . H-bonds are signed by the dashed line. The data about estimated affinity, lipophilic ligand efficiency (LEE), molecular weight (MW), logP of ligand are shown in the down figure.

3.11 Protein-protein docking

Since here, the descriptions of methods were focused on small molecules- protein interactions, although protein-protein interactions (PPI) are essential for biological functions in the cell. Different in the process from small molecule ligand-protein docking, protein-protein docking methods predict the protein-protein complex structure, outputting a molecular protein complex. For simplicity and speed, protein-protein docking is often done using both partners as rigid, and the obtained cluster is evaluated by scoring functions developed for this cause.

The ClusPro server^{244, 245} (<https://cluspro.org>) is a popular tool to perform protein-protein docking. After PDB coordinates of two proteins to be docked (denoted as “receptor” and “ligand”) are submitted, the ClusPro would perform the following steps: 1) Sampling millions

of conformations of the two interacting proteins to perform the rigid-body docking, and the ligand would be rotated 70,000 rotations. 2) The most significant cluster corresponding to the most likely complex models can be selected from 1,000 lowest-energy structures generated using RMSD-based clustering^{244, 245}. 3) Optimise the selected structures using energy minimization, and the interaction energy can be calculated via the Fast Fourier Transform (FFT) correlation technique²³⁵. Finally, the highest populated low-energy clusters can be listed as possible protein-protein complexes.

3.12 Methods for analysis of molecule dynamics

3.12.1 Root-mean-square fluctuation

Root-mean-square fluctuation (RMSF) measures the average squared difference between the position of each specific atom i and the reference position. Defined as (Equation 22):

$$RMSF_i = \sqrt{\frac{1}{T} \sum_{t_j=1}^T [r_i(t_j) - r_i^{ref}]^2} \quad (22)$$

Where T is the time over which the average is taken, r_i^{ref} is the reference position for the atom i . Atoms with higher spatial fluctuation or residues with large conformational changes are often analysed by plotting their RMSF values versus some index identifier, such as atomic or residue numbers. (exemplified in Figure 40). Obtaining this information is essential to assess the local flexibility of the systems, especially under different circumstances, such as apo and holo configurations.

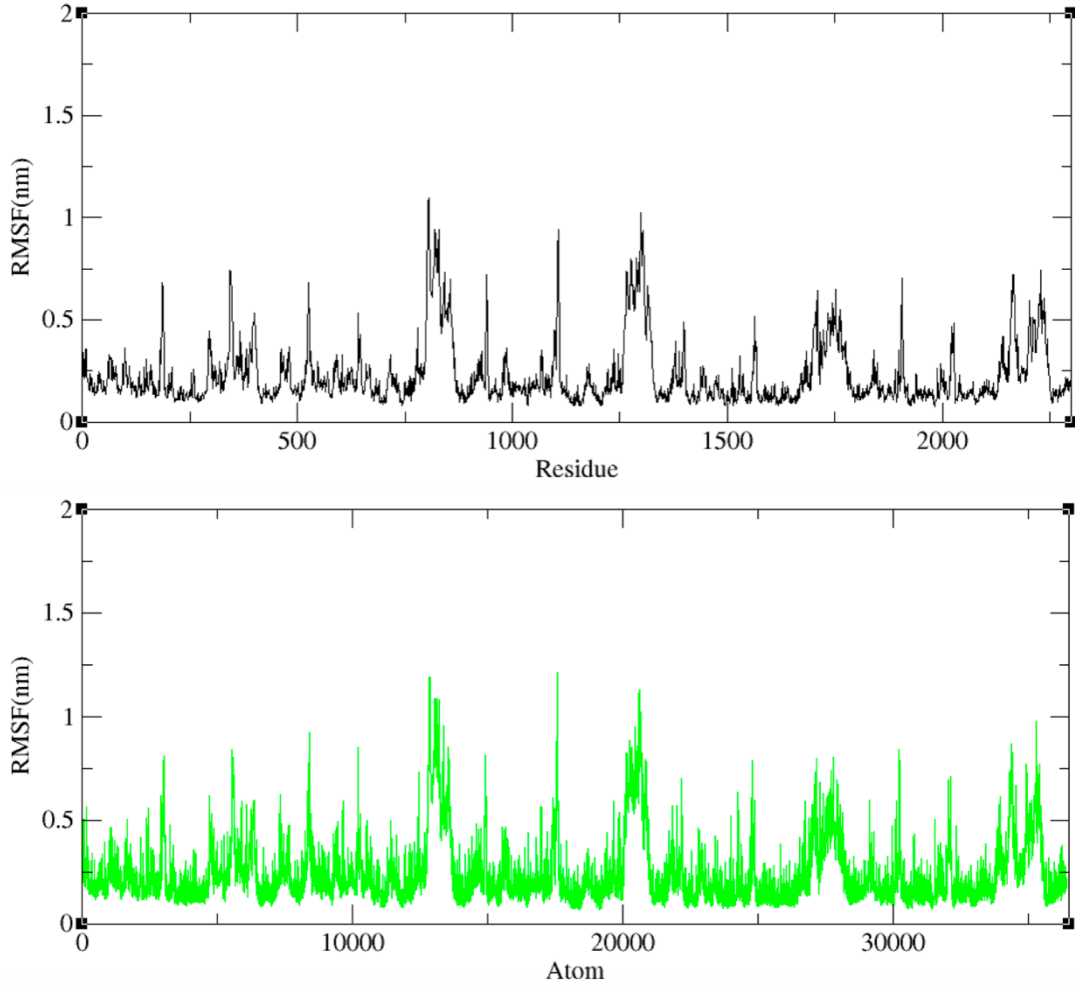


Figure 40: The example of $\alpha 7$ nAChR RMSF results. The top figure is the RMSF result about a total of 2300 residues of $\alpha 7$ nAChR. The bottom figure is the RMSF result about a total of 36400 atoms of $\alpha 7$ nAChR. Made by author.

3.12.2 Root-mean-square deviation

Root-mean-square deviation (RMSD) calculates the average of a combination of atoms to a system from a reference structure through time in an MD generated ensemble (Figure 41). The calculation is as Equation 23

$$RMSD(t) = \sqrt{\frac{1}{M} \sum_{i=1}^N m_i [r_i(t) - r_i^{ref}]^2} \quad (23)$$

Where $M = \sum_i m_i$, N is the number of atoms, $r_i(t)$ is the position of atom i at time t , r_i^{ref} is the reference position for atom i , and m_i is the mass of atom i . RMSD is used to quantify the dynamic stability of the system via analysis of the conformation changes of the system over the

simulation time. Smooth RMSD plots with small standard deviations indicate the stable system, whilst large fluctuations may suggest either a flexible protein or problems with equilibration, which means that the system has not yet reached the energy convergence and requires more extended simulations.

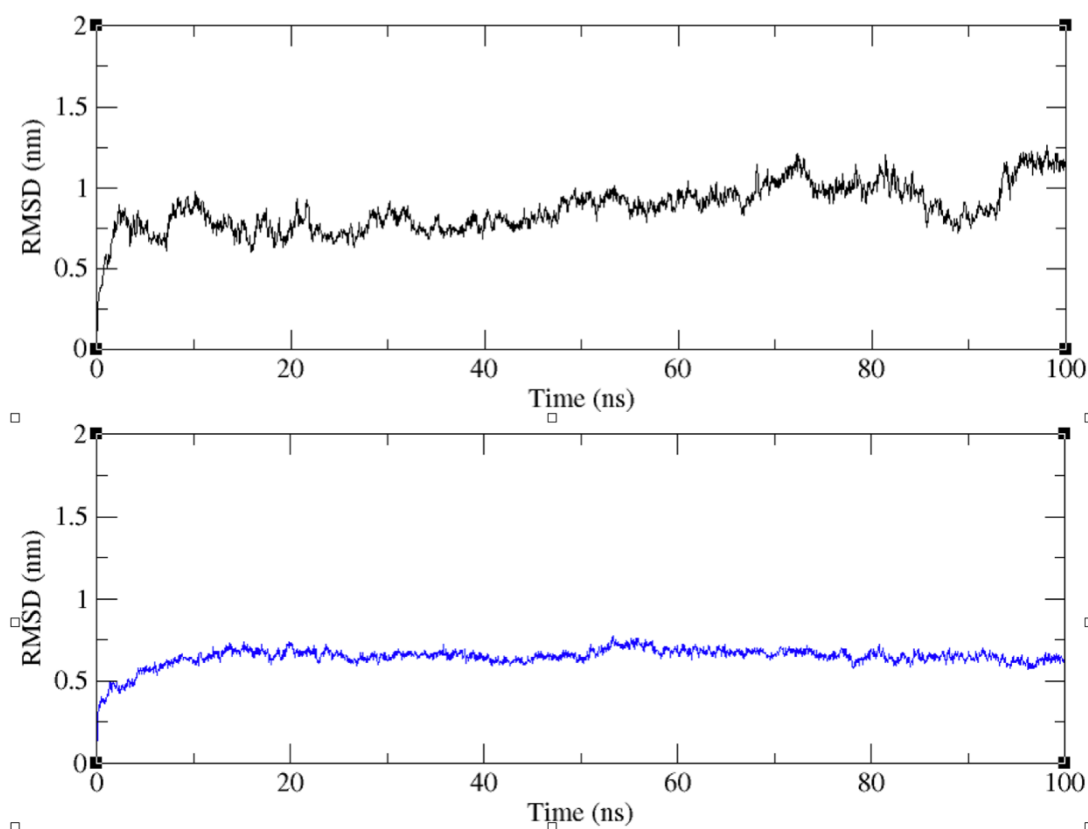


Figure 41: The example of two systems RMSD results. The top figure shows a non-equilibrated RMSD result, while the bottom one shows a flat RMSD result, indicating that the top system shows more fluctuation than another during a 100ns MD simulation. Made by author.

3.12.3 Cluster analysis

Molecular simulations can generate a large amount of conformational data in snapshots to be processed and analysed, which requires a data reduction approach to make it more reasonable. This data reduction can be achieved by cluster analysis, which groups similar conformations, and extracts a representative conformation from each set. In this case, conformations with high similarity form a subset, and representative conformations can be selected from each subset, reducing the data volume for subsequent analysis. The cluster analysis is frequently used to

identify significant protein conformers after MD simulation (Figure 42) and use these conformers to, e.g., hotspot mapping, umbrella sampling simulations, or molecular docking.

In this work, the approach developed by Daura and coworkers²⁴⁶ has been used to generate and sample clusters based on the mutual RMSD. The algorithm of this clustering method calculates a root-mean-square distance between the conformations. The configuration with the highest number of neighbours is identified as a cluster centroid. In sequence, this cluster will be removed from the pool, repeating the cycle for the remaining structures in the pool²⁴⁶. Following these steps, a series of non-overlapping clusters can be generated, each with a single representative state.

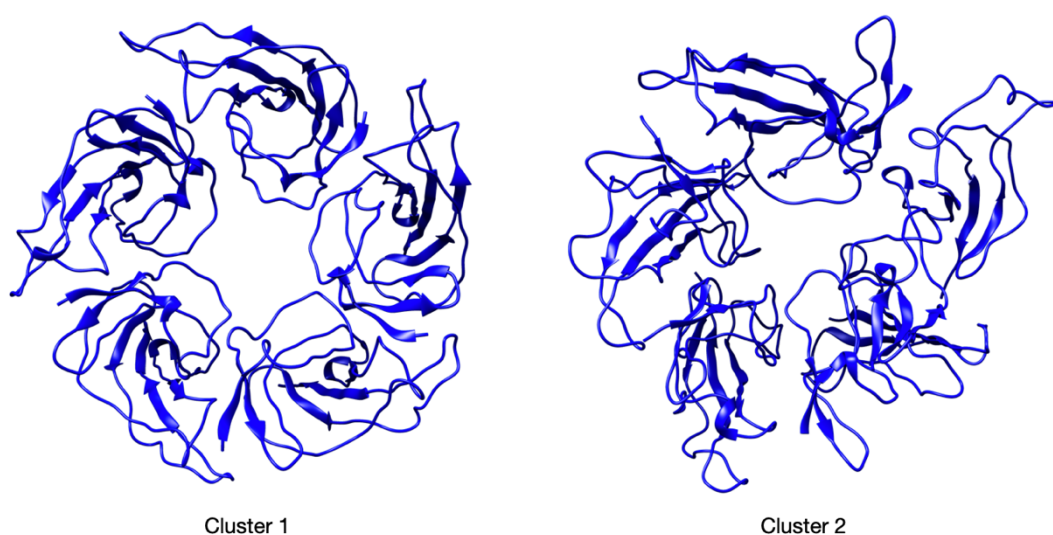


Figure 42: Two clustered structures for 5-dup $\alpha 7$ nAChR 100ns MD trajectory. Made by author.

3.12.4 Principal component analysis

The overarching goal of the principal component analysis (PCA) is to reduce the for more straightforward and more concise analysis. PCA quantifies the system's motion after simulation, identifying low-frequency collective motions, which are functionally relevant, and separating them from fast, functionally irrelevant motions (e.g., side chain methyl group rotations). PCA enables visualization of those selected high-amplitude modes, and thus it makes it easier to assess the primary driving force in the dynamics of the simulated system (Figure 43).

PCA is obtained from the diagonalization of the covariance matrix (σ) of atomic fluctuations

constructed from the MD trajectory. The matrix σ can be calculated in Equation 24:

$$\sigma_{ij} = \langle (q_i - \langle q_i \rangle)(q_j - \langle q_j \rangle) \rangle \quad (24)$$

where q_i, q_j are the mass-weighted Cartesian coordinates. By diagonalizing σ , the $3N$ eigenvalues and eigenvectors can be calculated. These eigenvectors are referred to as the principal components (PCs) of the trajectory. Each references a macro motion filtered from the selected trajectory. The eigenvalues are the magnitude of each eigenvector, representing the sum of the translation per atom for that specific PC. Usually, the eigenvectors are ranked via the magnitude that each one has. Hence, the first principal component corresponds to the largest eigenvalue, and it represents the direction in which the molecule has the most significant relevant motion, and so on for all subsequent PCs. Each PC corresponds to an axis in the n -dimensional space and is orthogonal to all other PCs.

One usual way to analyse and compare macrodynamics using PCA is to observe the 2D projection graphs for PC pairs. The denser region of dots within this 2D scatterplot indicates that this conformation is more likely to be sampled, whereas the sparse regions correspond to a more unstable conformation. Cluster analysis can be performed on the PCA dataset, where a dense region of dots can be grouped into a cluster of similar conformations.

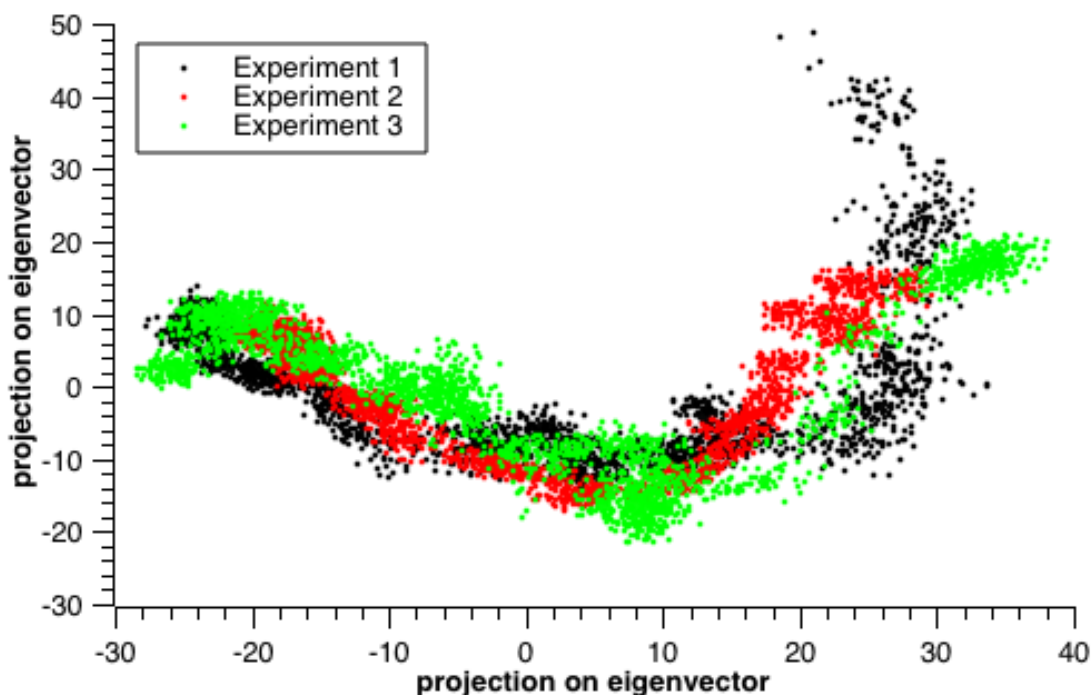


Figure 43: 2D Projection of PCA. A different colour separates each experiment result. Each dot represents a mode of motion of the protein. Made by the author.

3.12.5 Hydrogen bonds analysis

Hydrogen bonds are involved with most interaction modes related to protein structure, protein folding and molecular recognition. Hydrogen bonds provide a significant contribution to establishing the rigid structure of proteins and making intermolecular interactions specific. To determine whether an H-bond can be found between a donor atom/residue and an acceptor atom/residue, a common geometric criterion used is 1) $r \leq r_{HB} = 0.35nm$; 2) $\alpha \leq \alpha_{HB} = 30^\circ$ (Figure 44).

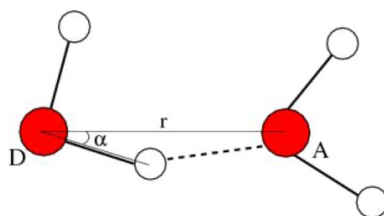


Figure 44: The conditions for H-bond existing between two atoms. The white circle represents the donor atom; the red circle represents the acceptor atom. The dashed line signs H-bond.

As shown in Figure 45, some software such as Chimera²³⁹ and SeeSAR can provide H-bond results visually on screen, making it simple to observe the interactions between two molecules. Also, from MD, the H-bonds are calculated via the autocorrelation function $C(\tau)$. This value is used to calculate the lifetime of the H-bond and is calculated via Equation 25:

$$C(\tau) = \langle S_i(t)S_i(t + \tau) \rangle \quad (25)$$

where $S_i(t) = \{1,0\}$ is the existing function of the H-bond i at time t . The average H-bond lifetime is estimated from the integral of $C(\tau)$, shown as following Equation 26:

$$\tau_{HB} = \int_0^{\infty} C(\tau) d\tau \quad (26)$$

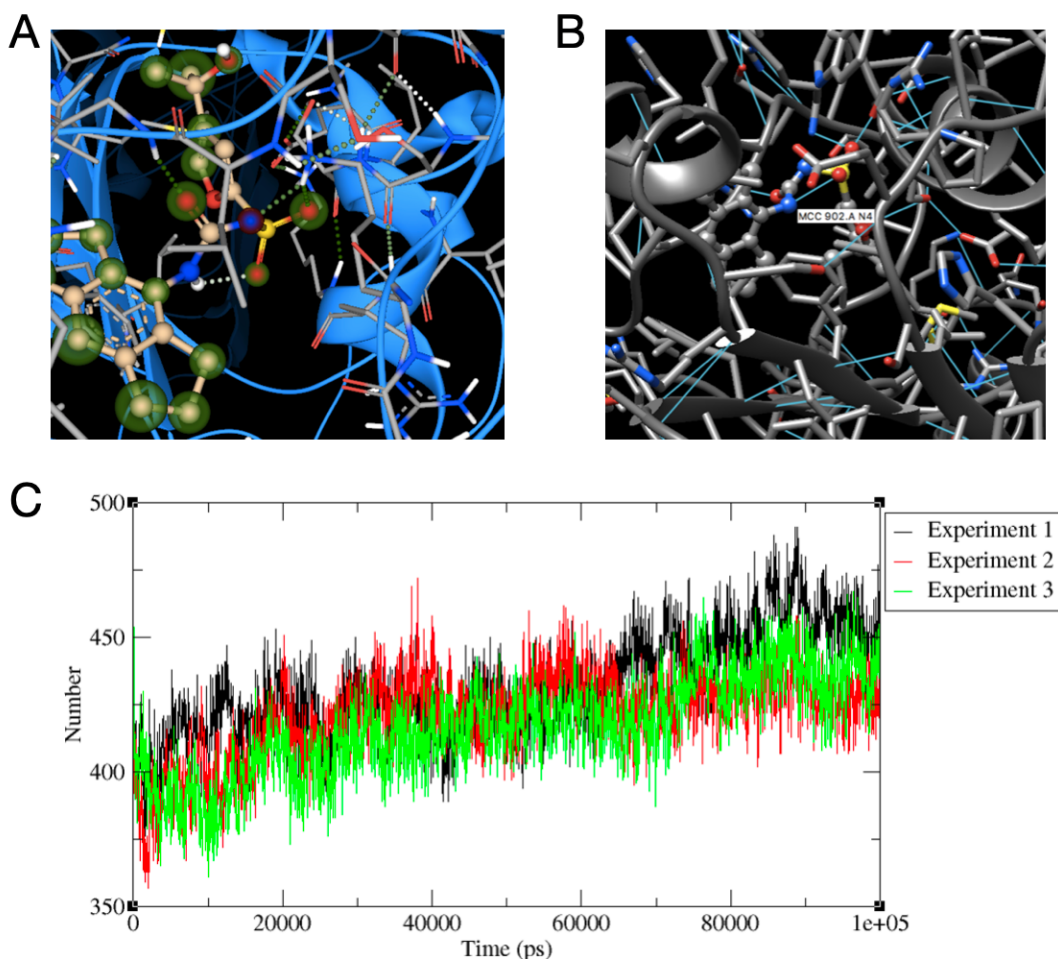


Figure 45: The example of H-bond results. A) The H-bond results are shown via FTMap; B) The H-bond results are shown via Chimera; C) The H-bond results are shown via MD simulation. Made by author.

Chapter 4 Structure and dynamics of human-specific *CHRFAM7A*

(Dup α 7) nicotinic receptor linked to neuropsychiatric disorders

As outlined in Chapter 2, dup α 7 negatively affects the functioning of α 7 receptors associated with neurological disorders, including Alzheimer's diseases and schizophrenia. However, the stoichiometry for the α 7 nicotinic receptor containing dup α 7 monomers and their functional status remain poorly understood. To solve it, my part of the work focused on developing computational models of all possible combinations of wild-type α 7 and dup α 7 pentamers and evaluating their stability via atomistic molecular dynamics and coarse-grain simulations. In this work, I assessed the effect of dup α 7 subunits on the Ca^{2+} conductance using umbrella sampling. The result indicated that receptors comprising of four or more dup α 7 subunits had been found not stable enough to constitute a functional ion channel. The results also showed that models with dup α 7/ α 7 interfaces are more stable and are less detrimental for the ion conductance in comparison to dup α 7/dup α 7 interfaces. My findings show that the optimal stoichiometry of dup α 7/ α 7 functional pentamers should be no more than three dup α 7 monomers, in favour of a dup α 7/ α 7 interface in comparison to a homodimer dup α 7/dup α 7 interface.

4.1 Results

4.1.1 Human-specific dup α 7 nicotinic receptor

In humans, exons 5–10 in the *CHRNA7* gene, which encodes for nicotinic α 7 receptor, may be duplicated and fused to exons A–E of *FAM7A* (family with sequence similarity 7A), resulting in the hybrid gene denoted as *CHRFAM7A*⁹⁸⁻¹⁰⁰. Its product, dup α 7 (**d**uplicated **α 7**), is a truncated variant, where the N-terminal 146 residues of the ligand binding domain of the α 7 receptor have been replaced by 27 residues from FAM7 protein⁹⁸⁻¹⁰⁰, and the remaining sequences of α 7 and dup α 7 are identical^{247, 248} (Figure 46A).

It is postulated that dup α 7 acts as a dominant-negative inhibitor of α 7 function, suggesting its role in human cognition and immune responses by perturbing normal α 7 activities²⁴⁹. However, the exact mechanism and specific contribution of dup α 7 to the biology of α 7 receptors remain

highly elusive. Even though the quaternary arrangements of both nicotinic receptors are known⁹⁵, and reports on overall stoichiometries of heteromeric $\alpha 7$ and dup $\alpha 7$ receptors have been published recently²⁴⁸, atomistic details controlling those assemblies are still missing. Unravelling the molecular mechanisms governing the formation of the most probable dup $\alpha 7/\alpha 7$ pentamers is of high clinical interest, crucial for structure-guided approaches to target those heteromeric receptors, which account for the translational gap research focusing on nicotinic receptors as therapeutic targets for neurodegenerative diseases¹⁴⁸.

(A)

| | |
|-------------|--|
| ACHA7_HUMAN | MRCSPGGVWLALAASLLHVSLLQGEFQRKLYKELVKNYNPLERPVANDSQPLTVYFSLSL |
| CRFM7_HUMAN | -----M-QKYCIYQH----- |
| | 70 80 90 100 110 120 |
| | |
| ACHA7_HUMAN | QIMDVDEKNQVLTTNIWLQMSWTDHYLQWNVSEYPGVKTVRFPDGGIWKPDILLYNSADE |
| CRFM7_HUMAN | -----QFQLLIQHLWIAAN-----CDIADE |
| | 130 140 150 160 170 180 |
| | |
| ACHA7_HUMAN | RFDATFHTNVLVNSSGHCQYLPPGIFKSSCYIDVRWFPFDVQHCKLKFGSWSYGGWSLDL |
| CRFM7_HUMAN | RFDATFHTNVLVNSSGHCQYLPPGIFKSSCYIDVRWFPFDVQHCKLKFGSWSYGGWSLDL |

(B)

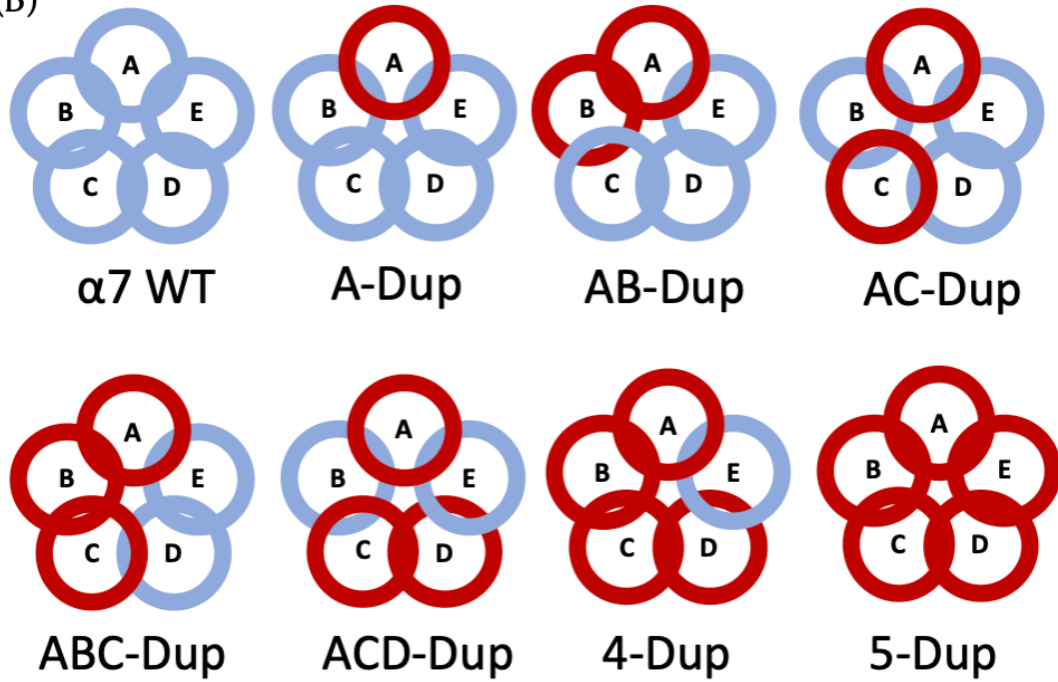


Figure 46: Sequence alignment of $\alpha 7$ /dup $\alpha 7$ EC domains and its eight possible models. (A) Sequence alignment of $\alpha 7$ /dup $\alpha 7$ extracellular (EC) domains (residues 1–180), performed by ClustalW, green are the residues with high similarity and in red the conserved residues. (B) Schematic representation of all eight different model arrangements dup $\alpha 7$ - $\alpha 7$ pentamer, considered in this study: the canonical (WT) $\alpha 7$ subunits are coloured blue; dup $\alpha 7$ subunits are coloured red.

To describe the most likely $\alpha 7$ /dup $\alpha 7$ stoichiometry at an atomistic level of detail and to elucidate the structural, energetic and functional effects of incorporation of dup $\alpha 7$ monomers

into the $\alpha 7$ channels, I combined molecular modelling, multiscale molecular dynamics (MD) simulations (all-atom and coarse-grain), and umbrella sampling (US) simulations of the whole receptors embedded in a DPPC membrane. I have also studied extracellular domains separately using all-atom MD simulations, focusing on their structure and intrinsic dynamics alongside the binding side dynamical behaviour.

4.1.2 Stoichiometry studies for different combinations of *dupa7*/ $\alpha 7$

The three-dimensional molecular models of extracellular domains of the canonical $\alpha 7$ and *dupa7* are shown in Figure 47. In *dupa7*, the N-terminal segment (blue region in Figure 47 of $\alpha 7$ has been replaced a shorter α -helix (Figure 47B). Three β -sheet segments ($\beta 1$ – $\beta 3$ in Figure 47) are missing, yet the following segment (β -sheets 4 to 10) resembles the canonical structure.

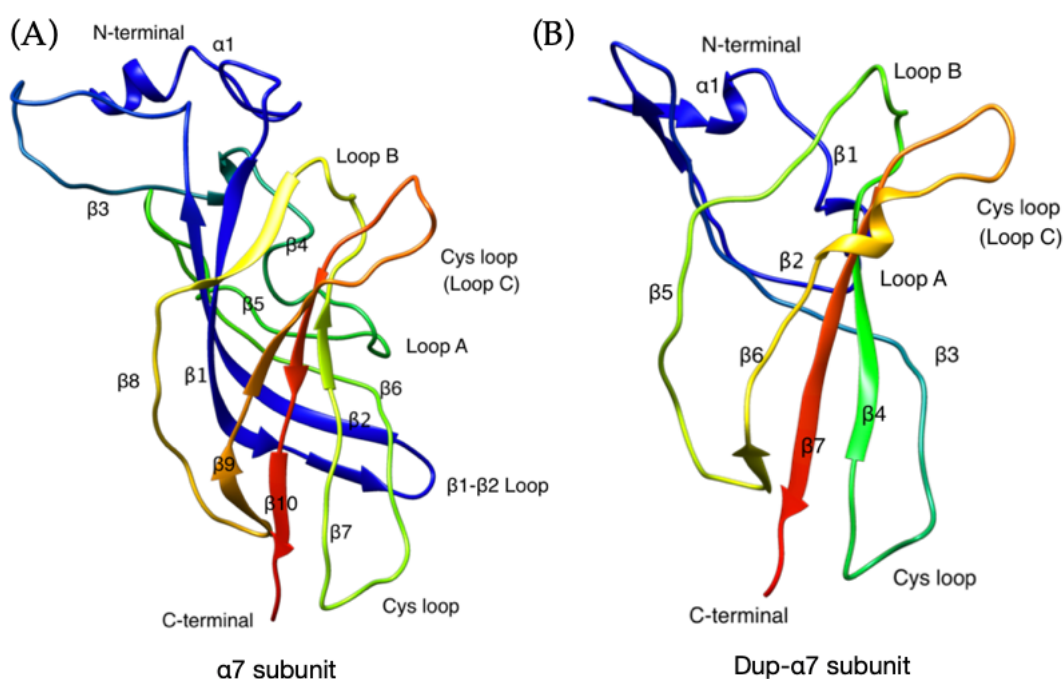


Figure 47: The extracellular EC domain conformation of $\alpha 7$ subunit (residues 1–180). (A) and *dupa7* subunit (B). The structures are coloured by gradient, from blue (N-terminus) to red (C-terminus).

Given that the ‘pentamer’s overall structure containing *dupa7* subunits has not been experimentally solved, I have built and investigated all possible models. These different combined pentamers produced eight models containing seven combinations of *dupa7* nicotinic receptors (Figure 46B). The dimerisation assembles the functional pentamer through only two

interfaces. Given this, four possible dimer combinations can be formed: WT-WT, dup α 7-WT, WT-dup α 7, dup α 7-dup α 7. Those combinations are functionally relevant because the orthosteric binding site is located at the interface between two monomers. Thus, each particular dimer combination will affect the receptor's function.

The eight models were first simulated using a coarse-grain approach by translating the atomistic model to MARTINI beads (Supplementary Figure S1). The RMSD and the total potential energy in the function of time are shown in Figure 48. All eight pentamers showed a similar magnitude of deviation from their starting structures (Figure 48A), with the average values fluctuating between 1 to 1.4 nm, which was achieved by most of the models after 50 ns of unrestrained coarse-grain MD simulation.

The RMSD plots did not show any statistical difference regarding relative motions regarding their starting configuration when a single WT α 7 is swapped for a dup α 7. However, there is a significant difference in total internal potential energy between the different stoichiometries. As expected, the WT α 7 has the lowest internal energy average value, closely followed by A-Dup α 7. Interestingly, both models with two dup α 7 (AB-Dup α 7 and AC-Dup α 7) have undistinguishable average values, which is also the case for both models with three dup α 7 (ABC- and ACD-Dup α 7). These characteristics show that CG models could discern the different 'stoichiometries' internal energy, but it does not have the resolution to discern which internal configuration is more stable. Therefore, I decided to simulate all eight models with an atomistic resolution.

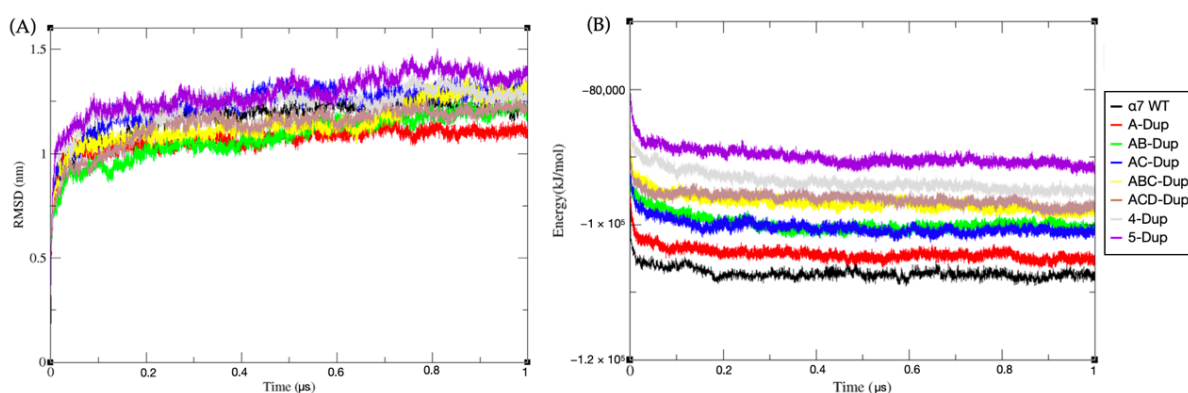


Figure 48: (A) RMSD results of eight complete transmembrane models during 1 μ s CG MD simulation. (B) Total potential energy vs. time results of eight complete transmembrane models during 1 μ s MD simulation. The black line shows the data for $\alpha 7$ WT model, red-A-Dup, green-AB-Dup, blue-AC-Dup, yellow-ABC-Dup, brown-ACD-Dup, grey-4-Dup, and purple-5-Dup. The schematic arrangements of all models are shown in the methods section.

To unravel the most probable stoichiometry and the effect the interfacial interactions have on the system stability, all eight models of different combinations of dup $\alpha 7/\alpha 7$ were simulated for 100 ns (each replica, all simulations performed in triplicates) in an atomistic resolution. This step was made for the whole model embedded in a DPPC membrane and its EC domain separately. The protein RMSD for the trajectories shows that for a higher ratio of dup $\alpha 7/\alpha 7$, the overall average RMSD is higher compared to the starting structure. This characteristic is more pronounced for the EC domain simulations, as shown in Figure 49A and 49B, respectively.

The pentamer simulations embedded in the membrane show RMSDs quickly plateauing for most of the models containing less than three dup $\alpha 7$ subunits. Nonetheless, the ACD-Dup $\alpha 7$ model shows a higher average RMSD than its ABC-Dup $\alpha 7$ counterpart, which may indicate that interfacial dup $\alpha 7$ -dup $\alpha 7$ interactions may be favourable for the molecular assemble.

The pentamers' internal potential energy (Figure 49C–D) shows that the average energy increases as the dup $\alpha 7/\alpha 7$ ratio increases. For the extracellular (EC) domain, the combination containing a single dup $\alpha 7$ subunit has significantly lower energy than all the other models. However, the full-length receptor model simulations show that the A model's internal energy (Figure 49D) fluctuates between similar values to the WT and AB model. The difference in energy between the EC domain and its counterpart containing all three domains may indicate that the transmembrane and intracellular (IC) domains may significantly stabilise the pentamers containing dup $\alpha 7$ subunits. However, it is not accuracy to compare this quantity between different simulations with different numbers of atoms (in general, more atoms will lead to more negative total energies due to attractive long-ranged van der Waals). Ideally, the free energy change associated with sub-unit binding should be calculated, although this is not possible due to computational limitations associated with the system's size. The number of hydrogen bonds formed is a reasonable metric to estimate the stability in the absence of the rigorous free energy

calculations.

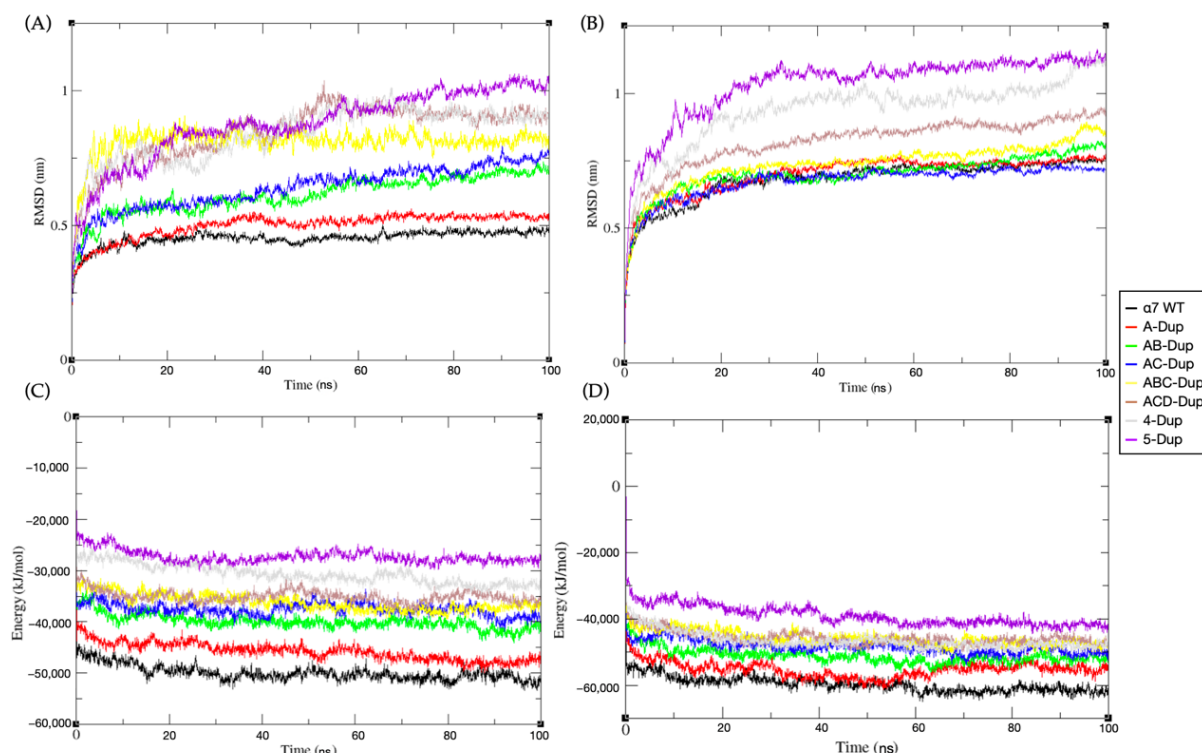


Figure 49: (A) RMSD results of eight EC domain models during 100 ns MD simulation. (B) RMSD of eight full-length receptor models during 100 ns MD simulation. (C) Total potential energy vs. time of eight EC domains during 100 ns MD simulation. (D) Total potential energy vs. time of eight complete transmembrane structures during 100 ns MD simulation. The canonical $\alpha 7$ (WT) model is shown in black, A-Dup $\alpha 7$ -red, AB-Dup $\alpha 7$ -green, AC-Dup $\alpha 7$ -green, ABC-Dup $\alpha 7$ -yellow, ACD-Dup $\alpha 7$ -brown, 4-Dup $\alpha 7$ -grey, and 5-Dup $\alpha 7$ -purple. The schematic arrangements of all models are shown in the methods section.

As expected, the total number of hydrogen bonds formed shows similar behaviour to the calculated potential energy. Specifically, the EC domain trajectories (Figure 50A) show more H-bonds for WT and A-Dup than their counterparts. The pentamers containing a higher number of dup $\alpha 7$ subunits show higher average internal energy and a lower average number of hydrogen bonds. Therefore, a clear correlation was observed between a higher dup $\alpha 7$ / $\alpha 7$ ratio, lower structural energy, and fewer total hydrogen bonds (Figure 50B).

When observing the hydrogen bond formation between interfaces, several key residues were identified. All these sets of residues are listed in Supplementary Tables S1–S16. Investigating WT-WT dimer, we identified six residues often present in the interaction between two adjacent subunits

(N69, N75, R101, P110, D111 and W172). In contrast, the number of interfacial hydrogen bond pairs for WT-dup α 7 interfaces varied between three to four. One of the main differences between the WT-WT interfaces and dup α 7 containing interfaces is that the latter does not have the hydrogen bond between R101 and the P110 of the sequential subunit residues are not present in the sequence of the dup α 7.

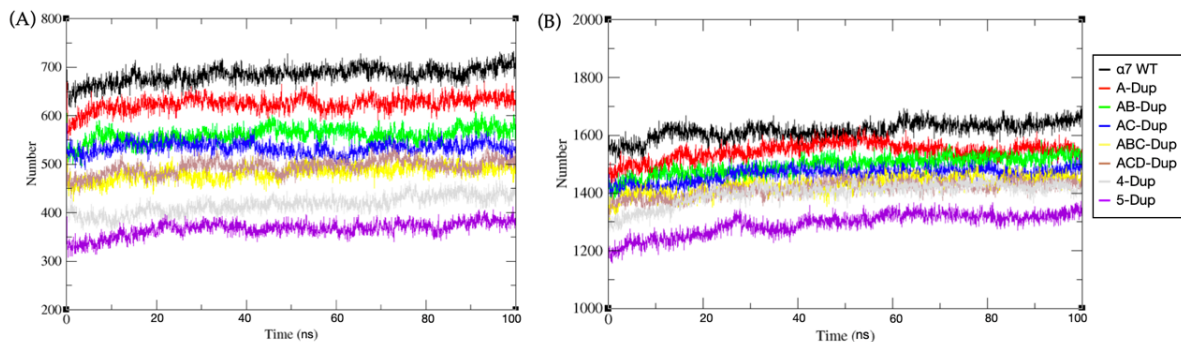


Figure 50: H-bond results of EC domain models and complete models. (A) The total number of hydrogen bonds vs. time of the eight EC domain combinations. (B) The total number of hydrogen bonds vs. time of eight fully models during 100 ns MD simulation. The black line shows α 7 WT model; red-A-Dup α 7, green-AB-Dup α 7, blue-AC-Dup α 7, yellow-ABC-Dup α 7, brown-ACD-Dup α 7, grey-4-Dup α 7, and purple-5-Dup α 7. The schematic arrangements of all models are shown in the methods section.

A higher number of dup α 7 subunits directly changes the molecular dynamics. Figure 51 shows that the trajectory projection on the two most significant principal components is affected by the dup α 7/ α 7 ratio. WT, A-Dup α 7 and the complexes with two dup α 7 (AB and AC) subunits have similar distributions (Figure 51A), especially for the transmembrane complexes (Figure 51B). The most significant difference comes from models with three or higher dup α 7 subunits. Given their higher energy, pentamers with four or five subunits resulted in higher magnitudes on their projection values, achieving a sparser distribution on the two-dimensional principal component space. These complexes show the same behaviour for their extracellular (EC) domain simulations, albeit less pronounced than the complexes containing a lower dup α 7/ α 7 ratio when the EC domain is simulated by itself.

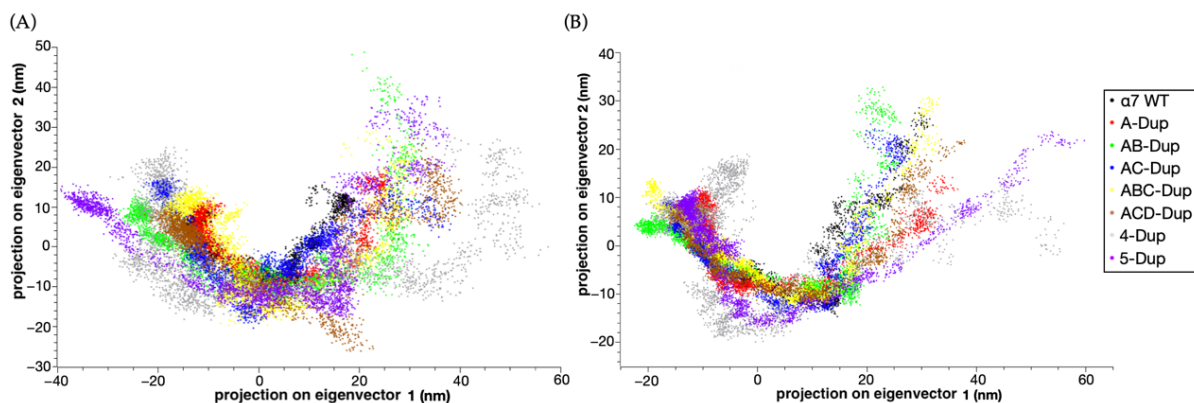


Figure 51: PCA results for the EC domain / fully modelled of eight dup α 7 models. (A) Principal component analysis, showing the 2D projection of eight different models of the extracellular (EC) domain. (B) Principal component 2D projection of eight different fully models during 100 ns of the atomistic MD simulation. The data for α 7 WT model is coloured black; A-Dup α 7-red, AB-Dup α 7-green, AC-Dup α 7-blue, ABC-Dup α 7-yellow, ACD-Dup α 7-brown, 4-Dup α 7-grey, 5-Dup α 7-purple. The schematic arrangements of all models are shown in the methods section.

These differences in the transmembrane model dynamics are directly related to the motions on its EC domain, mainly of loop C and the loops connecting the central β sheets (Figure 52). The motion of Loop C changes depending on its neighbouring subunit for both simulated models, as shown in Figure 52 and Supplementary Figure S2. For WT-WT interfaces (Figure 52A), the loop fluctuates between an open and closed conformation and the entire α 7 model remains stable overall (Supplementary Figure S3). WT-dup α 7 and dup α 7-dup α 7 interfaces (Figure 52B) show a more open conformation and higher local flexibility.

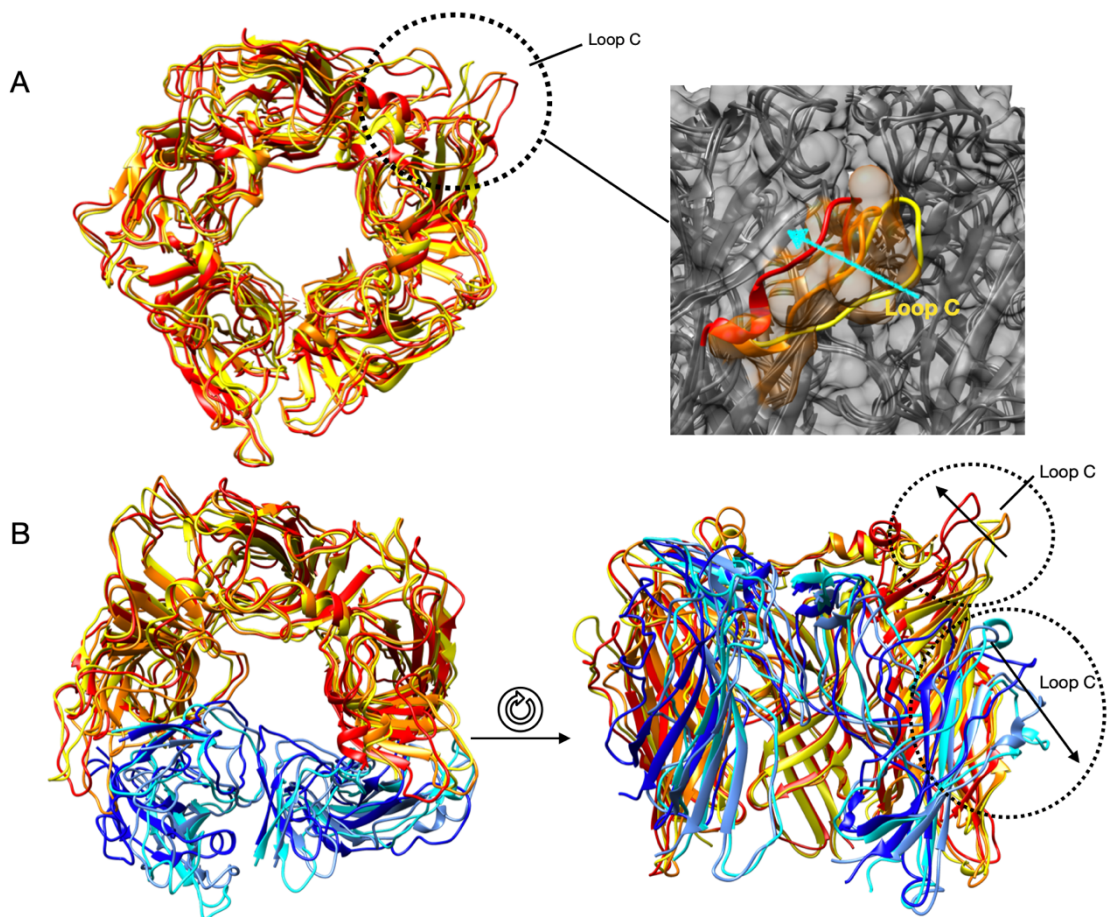


Figure 52: Conformational changes in the EC domain of the $\alpha 7$ / AB-Dup model during the 100ns MD simulation. (A) Left panel: The extracellular view of the three conformations of the $\alpha 7$ model EC domain. Right panel: the Loop C motion. The representative configurations sampled around 15 ns, 50 ns and 85 ns are coloured yellow, orange and red, respectively. (B) The extracellular view of the three conformations of the AB-Dup $\alpha 7$ model EC domain. The representative configurations for the starting conformation's state (around 15 ns) are coloured yellow and cyan; the representative configurations for the state sampled around 50 ns are coloured orange and cornflower blue; the representative configurations for state sampled around 85 ns are coloured red and blue. The dashed lines represent the interfacial loops, which are the areas with the highest fluctuation; the arrows within the dashed circles represent the macro motions within loop C for both WT and dupa7.

4.1.3 The effect of dupa7 subunits on Ca^{2+} conductance

Umbrella sampling simulations were carried out to investigate the effect of different combinations of dupa7/ $\alpha 7$ subunits on the Ca^{2+} intake. As shown in Figure 53, all eight potential mean force (PMF) curves show similar profiles. The energetic differences between the eight combinations started to arise around -5 nm, near the TM region's start. At -8 nm, the ion is at

the tunnel entrance in the EC domain, transitioning to the transmembrane region (TM) around -1.5 nm. At this region, $\alpha 7$ and A-Dup $\alpha 7$ have the lowest energy (-37.5 kcal.mol $^{-1}$). However, an energetic barrier emerges after the TM entrance, higher for AB, ABC, 4-Dup $\alpha 7$ and 5-Dup $\alpha 7$. This energetic difference emerges from how the ions interact with the model, specifically with glutamate E254. The canonical $\alpha 7$ (WT) structure can maintain the beginning of the α -helix located in the TM entrance, where the E254 residue is located (Sequence: 250-LVAEIM-257). This behaviour allows a more favourable interaction with the E254, resulting in more contact points with Ca $^{2+}$ (Supplementary Figure S4 and Supplementary Tables S17 and S18).

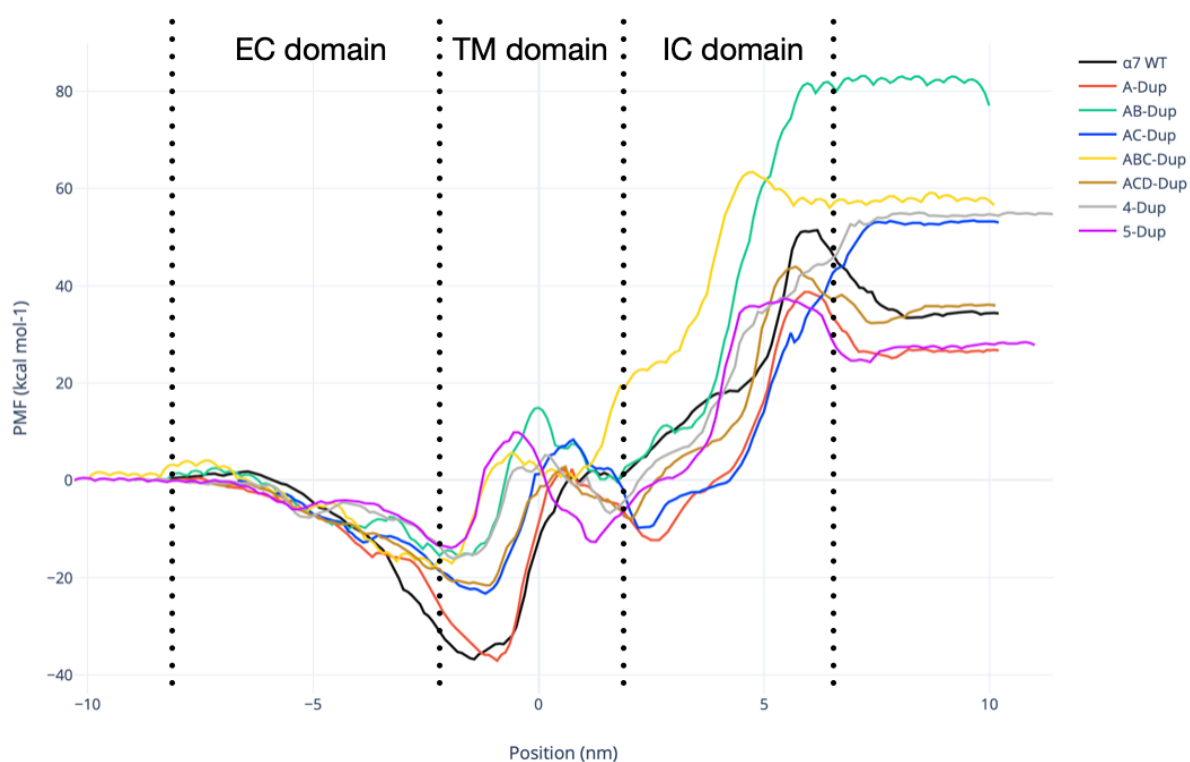


Figure 53: The umbrella sampling result of eight models. Potential of mean force (PMF) calculated for the position of the Ca $^{2+}$ moving through the pentamer axis. The black line shows the PMF obtained for the canonical $\alpha 7$ (WT) receptor, A-Dup $\alpha 7$ -red, AB-Dup $\alpha 7$ -green, AC-Dup $\alpha 7$ -blue, ABC-Dup $\alpha 7$ -yellow, ACD-Dup $\alpha 7$ -brown, 4-Dup $\alpha 7$ -grey, and 5-Dup $\alpha 7$ -purple. The schematic arrangements of all models are shown in the method section. the average associated error for the runs is 5kcal/mol obtained via averaging all errors using bootstrap method in gmx wham. The error bars were omitted for clarity.

The final 2–6 nm section represents the IC domain, which shows that the energy becomes significantly higher when the ion is far away from the TM domain. The AB-Dup $\alpha 7$ shows the

highest energy (81.2 kcal.mol⁻¹) at the bottom of the IC domain, with the lowest 5-Dup α 7 model results (28.3 kcal.mol⁻¹) as shown in Table 4.

Table 4: Sampling results Ca²⁺ permeability energetics.

| <i>Energy/Model (kcal. Mol⁻¹)</i> | α 7 | A-Dup | AB-Dup | AC-Dup | ABC-Dup | ACD-Dup | 4-Dup | 5-Dup |
|---|------------|-------|--------|--------|---------|---------|-------|-------|
| <i>EC \rightarrow TM (around-1.5 nm)</i> | -37.5 | -37.5 | -15.7 | -23.6 | -16.4 | -21.6 | -15.9 | -13.9 |
| <i>Transmembrane domain (around 0 nm)</i> | 0.2 | 1.8 | 14.6 | 7.7 | 4.1 | 2.3 | 4.8 | 9.84 |
| <i>TM \rightarrow IC (around 2 nm)</i> | -0.1 | -12.5 | -0.1 | -10.4 | 1.1 | -7.6 | -6.6 | -12.6 |
| <i>Exiting IC (around 6 nm)</i> | 51.1 | 37.6 | 81.2 | 52.3 | 63.1 | 43.7 | 52.8 | 37.3 |

4.2 Discussion

This work aimed to understand the assembly and stability of different combinations of functional α 7 nicotinic acetylcholine receptors (nAChR), bearing a partial duplicate α 7nAChR. Lasala and coworkers reported that WT α 7 nicotinic receptors could form functional pentamers incorporating dup α 7 subunits. However, the minimum number of WT α 7 had to be two²⁴⁸. Nonetheless, neither the effect of different stoichiometry on α 7 nAChR receptor function nor the structural stability of pentamers containing dup α 7 subunits is known.

These results showed that the arrangement of the extracellular domain for both α 7 and dup α 7 subunits is similar, and the exception is the N-terminal portion of the EC domain. This characteristic agrees with the models published shown by Lasala and coworkers²⁴⁸. Their models showed that changes in the α 1 loop and differences in the configuration of the β -sheets in the EC domain core could be observed between the canonical and receptors with dup α 7 subunits. My data confirmed that these modifications affected the ‘heteropentamers’ stability and showed how the interfaces interact.

I showed how structural equilibration of the receptors with different stoichiometries occurs

through time. For the pentamers with higher dup α 7 content, the RMSD curves take longer to plateau, and they reach higher average values than their starting structure, reflecting their decreased stability and weaker cohesion. These characteristics are found in both resolutions used: coarse-grain and atomistic MD as well. Additionally, similar behaviour when comparing the internal structural energy can be found: the ‘pentamers’ average structural energy steadily increases with the dup α 7/ α 7 ratio. This increase comes from several hydrogen ‘bonds’ disassemble and favourable interactions, both at the interfaces and hydrophobic core. One of the critical differences comes from the absence of the interaction with the R101 in the receptors containing dup α 7 subunits. This residue interacts with several other residues located in the subsequential subunit, and the lack of these interactions directly affects the structural cohesion of the interface.

The truncation directly affects the orthosteric binding site configuration. As we showed, the loop dynamics that works as a gatekeeper entirely depends on the dup α 7/ α 7 interactions.

The canonical α 7 (WT) pentamer simulations showed that loop C moves from an open conformation to a closed conformation. At the WT-dup α 7 interfaces, very different dynamic signatures are found. As discussed in previous works ^{143, 250}, a functional α 7 pentamer requires a single functional orthosteric binding site, albeit this is less sensitive than a fully functional pentamer with up to five orthosteric sites. Given the destabilisation effect that dup α 7 subunit causes on its dimerisation interface and in the orthosteric binding site, the dup α 7 interface should be less sensitive to the orthosteric ligand binding. This effect is evidenced by how the loop C on the dup α 7 subunit moves away from the pentamer. This movement shows that the dup α 7 binding site might remain open, not stabilising the orthosteric ligand in the binding cavity.

The partial duplication and truncation that leads to the dup α 7 protein are located in the extracellular domain (Figure 46A). Hence, this is the region that shows the highest difference between WT and dup α 7 dynamics. However, the transmembrane domain reduces the system’s overall energy by creating a higher number of hydrogen bonds, resulting in a lower average RMSD and RMSD standard deviation compared to the EC domain by itself. Nonetheless, pentamers containing dup α 7 subunits undergo structural changes within the TM domain,

mainly on the entrance region (residues 250–257). The energetic landscape referring to the extracellular to intracellular calcium intake showed that this region plays a vital role in this process.

Residues 250–257 go from an α -helical conformation in the WT pentamer to a short π -helix in most dup α 7 models. This conformation transition destabilises the favourable interactions between the calcium and the residue E254 and its neighbouring residues sampled in the WT simulation. This transition also is an effect that is directly related by the dup α 7/ α 7 ratio: Both WT and A-Dup α 7 models show similar profiles, indicating that a single dup α 7 subunit does not affect the calcium intake substantially. Nonetheless, another dup α 7 subunit increases the system energy substantially for the TM domain entrance region. This energy difference suggests that pentamers bearing two consecutive dup α 7 subunits are unlikely to occur.

These effects are also position-dependent: AB-Dup α 7 model and ACD-Dup α 7 model showed similar energetic profiles within the calcium intake TM. These profiles had lower free energy values than the AC-Dup α 7 model in the TM region. Hence, it indicates that dup α 7 interfaces have a less pronounced effect on the function of the dup α 7- α 7 interface for models with more than one dup α 7 subunit. The 250–257 loop interactions and the calcium are significantly reduced for models with four or five subunits. A higher barrier is found in comparison to the other models.

Hence, this shows that pentamers with four or five dup α 7 subunits are not functionally viable. In addition to the previously published works, the results shown in this work shed light on the plausible stoichiometry of dup α 7/ α 7 subunits. We evaluated stability and functional differences emerging from the positioning of the dup α 7 subunits within the pentamer. Albeit with a weaker interface interaction, the AC-Dup α 7 model showed a lower effect on the calcium transition than the AB-Dup α 7 model. A similar effect was observed when the pentamer had three dup α 7 subunits. The simulations of models with four or five dup α 7 are both too unstable, with an unfavourable effect on the Ach binding site organisation compared to the crystal structure. Additionally, it disrupts the calcium transmission by the pentamer.

In summary, a number higher than three dupa7 subunits is unlikely to be naturally occurring and functional. The most stable stoichiometry is the 1: dupa7-4:WT α 7, which is also one of the combinations with the least negative effect on calcium transmission. With two subunits, we expect to see dupa7 interacting with WT subunits, given that it has a lower effect on its function. Naturally occurring pentamers with three subunits should be less likely than with two, but should also be more favourable the interaction between dupa7/ α 7 interfaces, which we would expect the predominance of an ACD-Dupa7 organisation.

4.3 Materials and methods

4.3.1 Molecular modelling of α 7 and dupa7 receptors

The initial models of pentameric α 7 homopentamers and partially duplicated dupa7 subunits were created using SWISS-MODEL, a fully automated protein structure homology-modelling server, accessible via the ExPASy web server^{203, 204}. The primary sequences of the human canonical α 7 and dupa7 were obtained from the UniProt repository (entries P36544 and Q494W8, respectively). Fifty models were generated and ranked according to their sequence similarity and QMEAN²⁰³ quality scores combined. After a visual inspection procedure for the top 10 ranked molecules, the model based on the crystal structure of α 7-AChBP in complex with lobeline (PDB code: 5AFH) combined with high-resolution (4.3 Å) cryo-EM structure of mouse 5-HT₃ serotonin receptor (PDB code: 6BE1) was chosen for both receptors as the difference comes mainly from the N-terminal region (Figure 46A). Different stoichiometries were generated in the UCSF Chimera molecular modelling and visualisation toolkit²⁵¹ by overlaying the α 7 WT subunit with dupa7 in eight receptor models in total (Figure 46). A disulfide bridge at the conserved Cys-loop was ensured for each model. All models, arising from different stoichiometries of dupa7 and α 7 subunits, were quality checked by UCSF Chimera, having any missing loops modelled by MODELLER interface^{252, 253} within UCSF Chimera, and conformations of interfacial side chains checked for steric clashes.

4.3.2 Coarse-grain molecular dynamics simulations

Coarse-grain simulations of all eight models were done to equilibrate the modelled pentamers

and understand the intrinsic dynamics in a time scale relevant to the ion channel conductance. The atomistic models were translated to MARTINI beads²⁵⁴ and parametrised with the MARTINI 2.2 force field²⁵⁵. The principal protein axis was aligned to the Cartesian Z-axis. The membrane was built using the insane.py script. The membrane model was built using dipalmitoylphosphatidylcholine (DPPC), with 15 nm in the X and Y axis and 25 nm in the Z-axis, resulting in a box of $15 \times 15 \times 25$ nm in dimensions. The system was then solvated, using 90% MARTINI water beads and 10% anti-freeze MARTINI water beads, the anti-freeze beads added avoids the tendency of the water model used in the CG model of the membrane to freeze too easily²⁵⁵. The solvated receptor-membrane systems were energy minimised using the steepest descent algorithm and equilibrated. In the minimisation, the energy step size was set to 0.001 nm, and the maximum number of steps was set to 50,000. The minimisation was stopped when the maximum force fell below 1000 kJ/mol/nm using the Verlet cutoff scheme. Treatment of long-range electrostatic interactions and Van der Waals interactions were set to be shifted to 0 and 0.9 nm, respectively, beyond the cutoff of 1.5 nm. After the energy minimisation, heating to 300 K was performed for 10 ns with a time step of 20 fs. The temperature coupling was set between the protein and the non-protein entities using a Berendsen thermostat, with a time constant of 1 ps and the temperature set to reach 300 K with the pressure coupling. Pressure equilibration was run at 300 K with a semi-isotropic Berendsen barostat and set to 1 bar in an NPT ensemble. Both NVT and NPT had harmonic position restraints were applied to the backbone. The constraint algorithm used was LINCS. The production run was made using the same parameters as NPT, except the backbone position restraints were removed. The production run was made in triplicates of 1 μ s each for each of the eight combinations of $\alpha 7$ /dup $\alpha 7$.

4.3.3 Atomistic molecular dynamics (MD) simulations

Atomistic MD simulations have been carried out to generate ensembles to get a detailed insight into the stoichiometry of nicotinic $\alpha 7$ /dup $\alpha 7$ receptors in atomistic resolution. The simulations were performed for $\alpha 7$ /dup $\alpha 7$ pentamers with different stoichiometries for the models containing all three domains (EC-TM-IC) and only the EC domain to evaluate the effect of the TM-IC domain on the dynamics.

All simulations were performed using Gromacs 2016.3²⁵⁶. The protein was parametrised using the AMBER99SB-ILDN force field, with the DPPC lipid bilayer and TIP3P water model²¹⁸. The $\alpha 7$ and dup $\alpha 7$ models were embedded in a DPPC bilayer lipid molecule, using the computational membrane builder tool in the CHARMM-GUI server (www.charmm-gui.org)²⁵⁷⁻²⁵⁹. Box distance was set to 1 nm, and periodic boundary conditions were applied. The box was solvated and Na⁺ and Cl⁻ ions were added to achieve a 0.1 M concentration and maintain charge neutrality of the unit. The solvated receptor-membrane systems were energy minimised and equilibrated. The minimisation ran using steepest descent for 1000 cycles followed by the conjugate gradient. Energy step size was set to 0.001 nm, and the maximum number of steps was set to 50,000. The minimisation was stopped when the maximum force fell below 1000 kJ/mol/nm using the Verlet cutoff scheme. Treatment of long-range electrostatic interactions was set to Particle Mesh-Ewald (PME)²⁶⁰, and the short-range electrostatic and van der Waals cutoff was set to 1.0 nm. After the energy minimisation, heating to 300 K was performed for 20 ps with a time step of 2 fs and position restraints applied to the backbone in an NVT ensemble. The constraint algorithm used was LINCS, which was applied to all bonds and angles in the protein²⁶¹. The cutoff for non-bonded short-range interaction was set to 1.0 nm with the Verlet cutoff scheme, and long-range electrostatics were set to PME. The temperature coupling was set between the protein and the non-protein entities using a Berendsen thermostat, with a time constant of 0.1 ps and the temperature set to reach 300 K with the pressure coupling off. Pressure equilibration was run at 300 K with a Parrinello–Rahman barostat and set to 1 bar²²⁵ in an NPT ensemble. The equilibration trajectories were set to 5 ns (discarded from the analysis), and the production MD simulations were performed for 100 ns. Each trajectory was run in triplicates.

Analysis of the trajectories was performed using GROMACS tools, including RMSD to assess overall stability, per-residue RMSF to assess the local flexibility, and calculating SASA for solvent-mapping.

For all-atom molecular dynamics simulations, standard workstations and a GPU in-house server were used. These included workstations with Intel i7 7th generation processor, with 16 GB of RAM

with a NVIDIA GTX1080, and a dedicated server with 32 GB OS-RAM, a Xeon 44 core processor with multithreading, and 2x NVIDIA GTX1080Ti GPUs.

4.3.4 Umbrella sampling (US) simulations

Umbrella sampling (steered molecular dynamics)²³² simulations were used to assess the influence of different stoichiometry of receptors on the ‘pentamer’s ion conductance. The energies of the Ca^{2+} ion pulled through the axis of the channel pore (Z-axis) were calculated using the Weighted Histogram Analysis Method (WHAM) method²³⁴ to extract the potential of mean force (PMF). To prevent the channel from moving out of the membrane, the receptor subunits were position-restrained during the pulling simulations, using $1000 \text{ kJ mol}^{-1}\text{nm}^{-2}$. The ion has been placed above the top of the EC domain and pulled downwards along the Z-axis towards the TM and IC domains over 5 ns at a rate of 0.01 \AA/ps . A series of umbrella sampling windows were generated from the pulling trajectory to proceed with the umbrella sampling. The entire pathway covering the range of $[-10, 10] \text{ \AA}$ was divided into 0.7 \AA , totalizing 40 windows.

Chapter 5 Mapping of putative allosteric binding sites of $\alpha 7$ and dup $\alpha 7$ nicotinic receptors

The analysis of stoichiometry, molecular dynamics, conformational changes, and energetics of $\alpha 7$ receptors, described in Chapter 4, indicated that the two most unlikely assemblies of dup $\alpha 7$ receptors contain four and five dup $\alpha 7$ subunits, respectively. Among the rest of the pentamers, the most stable and thus plausible assemblies were A-Dup, containing only one dup $\alpha 7$ subunit, and AC-Dup containing two non-consecutive dup $\alpha 7$ subunits.

Based on these models, protein-protein docking studies were made to evaluate how such dup $\alpha 7/\alpha 7$ interfaces or dup $\alpha 7$ /dup $\alpha 7$ interfaces would interact with α -bungarotoxin (α -BTX), which is an antagonist to $\alpha 7$, and amyloid A β_{42} . The result showed that receptors bearing dup $\alpha 7$ subunits are less sensitive to A β_{42} effects, which may shed light on the translational gap reported for strategies focused on nicotinic receptors in ‘Alzheimer’s disease research. Following on those results, how the formation of $\alpha 7$ /dup $\alpha 7$ heteromeric receptors affected the binding of two known ligands of $\alpha 7$ nicotinic receptors was investigated: the A β_{42} and α -BTX using molecular docking methods.

Another aim of the work described in this chapter was to identify putative allosteric binding sites, which are “druggable” by small molecule ligands on those receptors. Identifying those sites is crucial for developing selective dup $\alpha 7$ ligands, as the orthosteric binding site is assumed to be abrogated in dup $\alpha 7$ subunits.

To map any plausible allosteric binding sites and the extent of changes within the orthosteric binding site, the molecular representations obtained from molecular dynamics of the extracellular (EC) domain of the functional pentamers of the canonical $\alpha 7$ receptor were selected for the solvent mapping via FTMap webserver. The same procedure subjected A-Dup, AB-Dup, AC-Dup and ACD-Dup models to identify specific binding sites.

To study conformational changes within orthosteric sites in pentamers containing dup $\alpha 7$ subunits, several known agonists were used: acetylcholine (Ach), nicotine (Nic), lobeline (Lob). Alongside these known agonists, a target-focused nAChR orthosteric ligand library (1790 small

molecules) was docked and subjected receptors to molecular docking procedure using UCSF DOCK6 software²⁶². Four novel allosteric binding sites have been discovered in this process. One of those sites is located inside the single subunit, and it is formed by $\beta 2$, $\beta 4$, $\beta 5$ and $\beta 6$ loops (Figure 61). The second one is formed between two adjacent subunits, involving residues from the $\alpha 1$ loop and $\beta 4$ loop of one subunit and residues in $\alpha 1$, $\beta 2$ and $\beta 3$ loop of the adjacent subunit. The last two allosteric binding sites were found in the middle of the adjacent subunits: one in the upper pocket above the orthosteric binding site, involving loops B and C in one subunit and loops $\beta 1$, $\beta 2$ in the adjacent subunit. Another allosteric site was found below the orthosteric site, and it involved residues in loops A, $\beta 6$ and $\beta 7$ in one subunit and residues in loops $\beta 1$, $\beta 8$ in the adjacent subunit. As shown in Figure 61, these binding sites may provide a starting point for future drug development to target human-specific dup $\alpha 7$ isoform of nicotinic receptors.

5.1 Results

5.1.1 *The effect of dup $\alpha 7$ subunits on macromolecular ligand binding*

Alongside the effect of the dup $\alpha 7$ / $\alpha 7$ ratio on the pentameric receptor dynamics, we also assessed the effects of dup $\alpha 7$ / $\alpha 7$ ratio on the binding of two macromolecular ligands: α -bungarotoxin (α -BTX) and the amyloid β ($A\beta_{42}$) (Figure 54). α -BTX is a well-established, 74 residues (8 kDa) neurotoxin that binds nicotinic acetylcholine receptors, including $\alpha 7$ subtypes, and acts as a competitive antagonist upon them¹⁴². While the experimental structure of α -BTX- $\alpha 7$ is known (PDB code: 4HQP), the structural information on interactions between receptors containing dup $\alpha 7$ subunits and α -BTX is missing. $A\beta_{42}$ and $\alpha 7$ interact with high affinity^{263, 264}. However, the details of those interactions remain elusive.



Figure 54: 3D structure of α -BTX and $A\beta_{42}$

To gain insight into those interactions, molecular docking calculations were performed. Molecular models of the α -BTX binding sites were built on the dimerisation interface (orthosteric binding sites, inferred from the reported α -BTX- $\alpha 7$ interactions). Hence, evaluation of the binding poses of α -BTX and their binding affinities to all four combinations of receptors: $\alpha 7/\alpha 7$ (canonical), $\alpha 7$ -dup $\alpha 7$, dup $\alpha 7$ - $\alpha 7$, and dup $\alpha 7$ -dup $\alpha 7$.

To identify the binding site for $A\beta_{42}$, an exhaustive scan of the whole extracellular receptor domain using ClusPro²⁴⁴ was performed. The best-scoring poses (8/10 hits) converged to the site, which partially overlaps with the reported α -BTX binding sites, consistent with the reports on $\alpha 7$ receptor activation via orthosteric modality reversing the $A\beta_{42}$ binding²⁶⁴. Therefore, it was concluded that the $A\beta_{42}$ binding site overlaps with that for α -BTX.

The results are shown in Table 5. α -BTX showed very low predicted binding affinity to all the receptors containing dup $\alpha 7$ subunits (Table 5), suggesting that those receptors will be resistant to α -BTX. This observation agrees with the published experimental data showing that CHRFAM7A decreased α -BTX binding as detected by immunohistochemistry and flow cytometry and markedly decreased α -BTX staining neuromuscular junction of CHRFAM7A transgenic mice²⁶⁵.

For $A\beta_{42}$ interactions, the ligand residues that interacted with the receptor binding sites were located in the C-terminal regions of the $A\beta_{42}$ monomer (Table 5). The predicted binding affinity of the $A\beta_{42}$ to the binding site comprising two $\alpha 7$ subunits (the canonical receptor) was in the low nM range, consistent with available experimental data²⁶⁴. As a comparison, α -BTX was

predicted to bind to the canonical $\alpha 7$ binding sites with one order of magnitude lower (high pM range) than $A\beta$. This difference, too, is consistent with the experimental data available.

The $\alpha 7$ receptor residues crucial for maintaining the binding pose include R208, F209, and E211. All those residues are conserved in both $\alpha 7$ and $\text{dup}\alpha 7$ isoforms, albeit their dynamics are different in $\text{dup}\alpha 7$. In particular, R208 and E211 were involved in favourable, stabilising electrostatic interactions with the ligand. F209 is involved in the network of aromatic residues, including two tryptophan residues: W171 and W77. The latter is located at the different subunits, as showed in Figure 46. Despite variation in the N-terminal region of the receptor, this residue is conserved in primary sequence alignment between $\alpha 7$ and $\text{dup}\alpha 7$ (Figure 46A).

Analysing the effects of $\text{dup}\alpha 7$ on the binding affinities, it was observed that more favourable predicted binding affinities for $A\beta$ by both $\alpha 7$ - $\text{dup}\alpha 7$ and $\text{dup}\alpha 7$ - $\text{dup}\alpha 7$ (predicted to be within low-mM range), compared to the canonical $\alpha 7$ - $\alpha 7$ binding site (affinities in low-nM range, Table 5). However, the calculated binding affinity for $A\beta$ to $\text{dup}\alpha 7$ - $\alpha 7$ improved (high nM to low μM range). This difference indicates that the residues involved in the interactions with $A\beta$ and are not conserved between $\text{dup}\alpha 7$ and $\alpha 7$ (S56, S58, L60, and Q61) may be necessary for complex stabilisation. Additionally, it indicates that the $\alpha 7$ secondary structure architecture in this region is critical for the binding of $A\beta_{42}$. This characteristic suggests that receptors containing $\text{dup}\alpha 7$ subunits will be more resistant to $A\beta_{42}$ -related toxicity, supported by recent experimental data ¹⁴⁸.

Table 5: Binding affinity ranges of $A\beta_{42}$ or α -BTX to orthosteric binding sites. All affinity ranges were calculated in SeeSAR using HYDE.

| Binding Site Interface | $A\beta_{42}$ Affinity (K_i calc) | α -BTX Affinity (K_i calc) |
|---|--------------------------------------|--------------------------------------|
| $\alpha 7$ - $\alpha 7$ | Low nM | High pM |
| $\alpha 7$ - $\text{dup}\alpha 7$ | mM | >mM |
| $\text{dup}\alpha 7$ - $\alpha 7$ | High nM to Low μM | >mM |
| $\text{dup}\alpha 7$ - $\text{dup}\alpha 7$ | mM | >mM |

5.1.2 Conformational changes in orthosteric (agonist) binding sites

The structures of the receptor subjected to the analysis were obtained by averaging the two highest-populated clusters of $\alpha 7$ pentamers obtained from 100 ns all-atom MD simulations, which were described in detail in Chapter 5 of this dissertation. As the extracellular location of the orthosteric site is known for nicotinic receptors, high-resolution structures have been resolved by X-ray crystallography (PDB:5AFH)¹²⁷, and the EC domain was the analysis focus.

First, the orthosteric binding sites of the canonical $\alpha 7$ nAChR was mapped using acetylcholine (Ach), nicotine (Nic), and lobeline (Lob) as agonists. Molecular docking reproduced experimentally determined binding modes of Nic and Lob correctly. The predicted binding mode of Ach was validated by comparing the molecular docking results with the crystal structure of related ligand carbamylcholine to acetylcholine binding protein (AChBP) of *L. stagnalis* (PDB code: 1UV6). The results are shown in Figure 55. They show that the molecular docking protocol chosen in this study can reliably reproduce the experimental binding modes for orthosteric ligands inferred from homologous proteins.

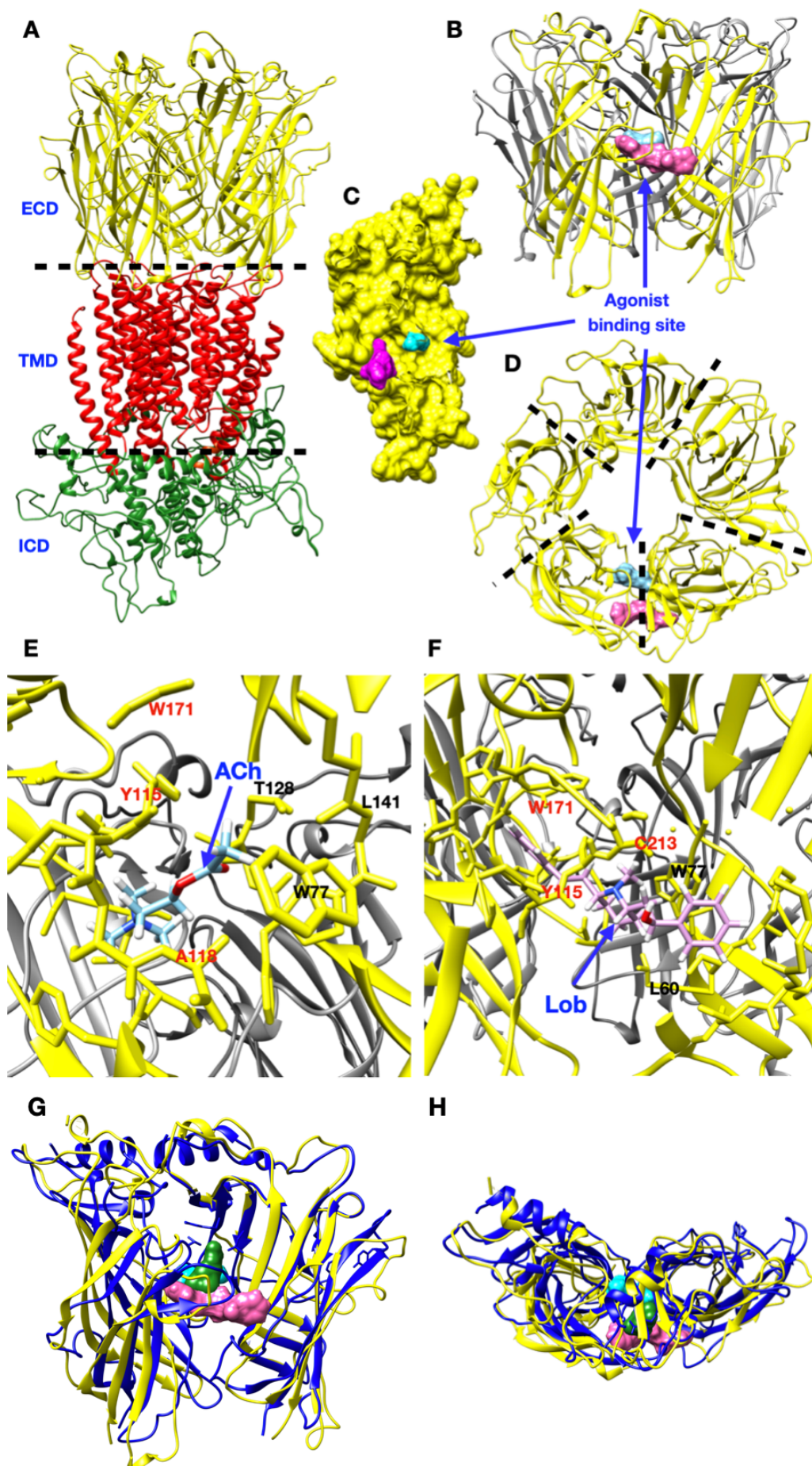


Figure 55: Structure of $\alpha 7$ nAChR and the position of the orthosteric binding site. A) Side view of the whole structure of $\alpha 7$ nAChR pentamer including extracellular domain (ECD, yellow), transmembrane domain (TMD, red) and intercellular domain (ICD, green); B) Side view of the agonist binding site position of $\alpha 7$ nAChR which is located between the adjacent subunits. The blue and pink spheres represented acetylcholine and lobeline, respectively. C) The side view of the complex of single $\alpha 7$ subunit with Ach and Lob ligands. $\alpha 7$ subunit and two ligands are shown in solid surface and cartoon representation. D) Top-down view of $\alpha 7$ nAChR EC domain and the position of agonist binding sites of Ach and Lob. The black dashed lines delineate the subunits; E&F) Detail view of Ach and Lob agonist binding sites. The blue compound is Ach, while the pink compound is Lob. The specific surrounding residues of $\alpha 7$ subunit are represented in yellow colour, and the type of these residues was signed in red and black, different colours separate the residues belong to different $\alpha 7$ subunit, red residues belong to the left-side subunit, while black belong to the right-side subunit; G&H) Side view and top-down view of $\alpha 7$ nAChR overlay with experimental structures (PDB code: 1UV6) showing that the binding mode is reproduced well, the experimental structures are signed in blue colour with carbamylcholine ligand surface (green).

Although the Ach molecule bound to the receptor is positioned closer to the receptor channel, the Lob or Nic occupy the same space and interact with several key residues (Figure 55B-D). The slight difference in the binding mode most likely results from Ach being smaller than other agonists. Figure 55E shows that the “core” of the orthosteric binding site of $\alpha 7$ subunit is formed by residues 113-126 from loop A, W171 from loop B, and W77, T128 and L141 from the $\beta 2$, $\beta 5$ and $\beta 6$ loops of the adjoining $\alpha 7$ subunit. As for the region of the binding site involved in interactions with larger agonists, such as Lob (Figure 55F), it is formed by residues 115-119, residues 167-171 and residues 210-220 belonging to loops A, B and C. Loops A and B create the left, bottom and posterior sides of the binding pocket, while the loop C of this subunit covers the front of the pocket. Residues 56-60, W77 and residues 183-189 from the $\beta 1$, $\beta 2$ and $\beta 8$ loops of the adjacent $\alpha 7$ subunit additionally contribute to that binding region, interacting with the ligand.

This architecture indicates an essential role of residues in the loops A-C and $\beta 2$ -loop in agonist binding, especially the residues 116-119, W77, and W171. This data is consistent with the published experimental data on $\alpha 7$ nAChR and homologous proteins, showing the binding modes of Nic, Lob, and Ach analogue (carbamylcholine) binding modes¹²⁷. These results also indicate that molecular docking utilising DOCK6 is reliable and, as such, may be used for the

analysis of “druggable” binding sites of dup α 7 nAChRs.

In dup α 7 subunit, loop A and three β -sheet segments (β 1, β 2 and β 3) are missing compared to the canonical α 7 subunit. This difference is due to the N-terminal 146 residues of CHRNA7 being replaced by 27 residues from FAM7 protein, which dramatically changes the orthosteric binding site architecture. Therefore, it can be expected that one of the orthosteric binding sites in the dup α 7 receptor will be abrogated, but the effect exerted by the dup α 7 subunit on the distal orthosteric site is challenging to assess. To qualitatively estimate this effect, four types of dimers of dup α 7 nAChR EC domain (WT-WT dimer, WT-dup dimer, dup-WT dimer and dup-dup dimer) were extracted from the MD simulation trajectories and subjected to the analysis. The RMSD results are shown in Figure 56A. As expected, the RMSD results of the WT-WT dimer and dup-WT dimer were similar to one another, and the complexes (including the binding sites) were stable.

WT-dup dimer showed slight fluctuations compared to the canonical (WT-WT) dimer and moved less than the canonical dimer, whilst trajectories obtained for the dup-dup dimer were the least stable. This result is consistent with the results described in Chapter 4, showing that the pentamers with two adjacent dup α 7 subunits are less stable and thus less likely to occur than “mixed” pentamers, wherein the canonical (WT) subunits surround dup α 7 subunits. The PCA results in Figure 56B showed that the distribution of points in the WT-WT dimer and dup-WT dimer map was more concentrated than the relatively widely dispersed points in the WT-dup dimer and dup-dup dimer map. A more intense spread of points corresponds to a more stable structure; in other words, the WT-WT dimer and dup-WT dimer trajectories were more stable than the other two dimers.

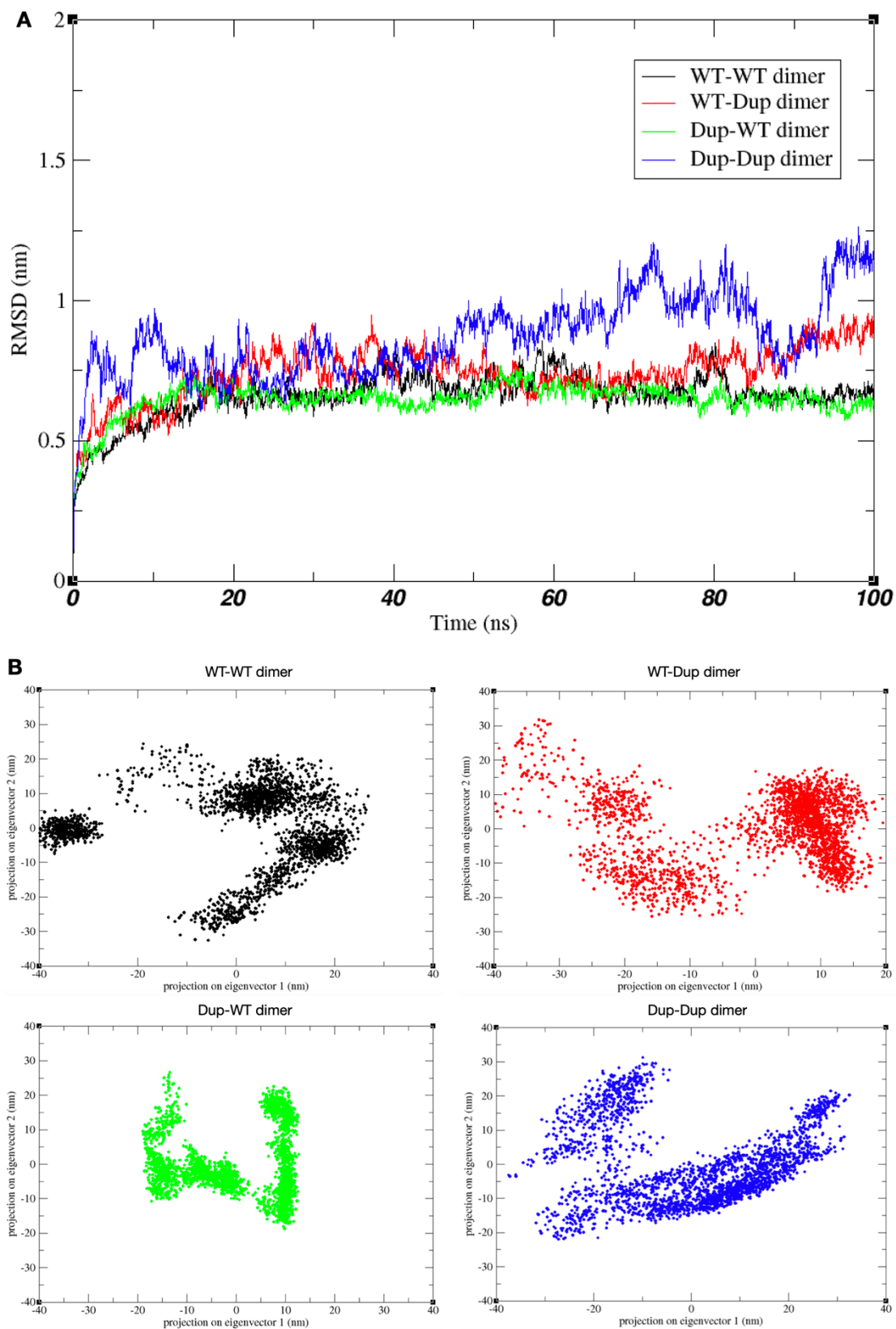


Figure 56: A) RMSD and B) PCA results of four EC domain dimers during 100 ns MD simulation. The canonical $\alpha 7$ WT-WT dimer is shown in black, WT-dup dimer - red, dup-WT dimer - green, dup-dup dimer – blue.

According to the PCA results of the four dimers in Figure 56B, the dots are all densely distributed in two regions which suggests that their structures can be grouped into two clusters of similar conformations. The average of these two clusters could be the most suitable structure of dimer to do the following tests.

Moving to the next docking step, the dimers utilized in molecular docking calculations were built by averaging the two highest-populated clusters from each 100 ns MD trajectory. The results of unbiased molecular docking results are shown in Figure 57. As expected, the orthosteric binding sites in WT-dup dimer and dup-WT dimer are similar to the canonical WT-WT dimer, whereas the dup-dup binding site is strikingly different. The binding mode calculated for Ach bound to the dup-dup dimer showed Ach molecule binding inside each of the dup α 7 subunits and the conformation of the interface of two adjacent dup α 7 subunits abrogated the binding of small molecules. This binding mode indicates that the “druggable” interfacial binding site may still exist in WT-dup and dup-WT dimer but disappears in the dup-dup dimer.

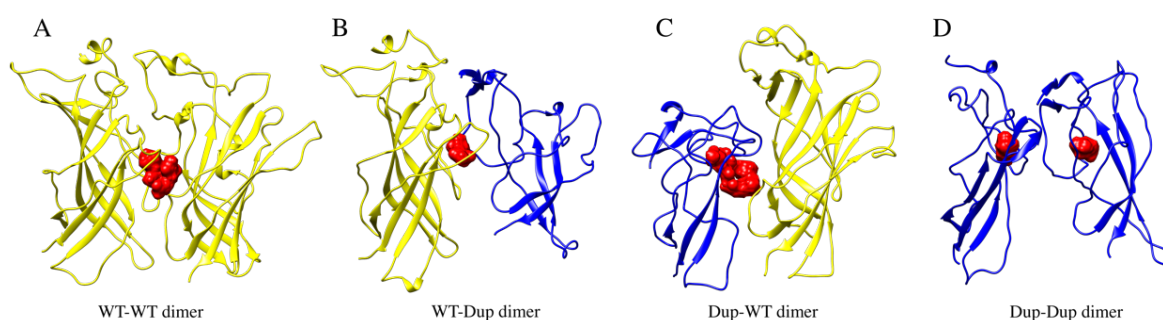


Figure 57: Molecular docking results of Ach binding to four distinct EC dimers. A) WT-WT; B) WT-dup; C) dup-WT; D) dup-dup. Protein is represented as ribbons, while ligand (Ach) is rendered as red spheres. The canonical (WT) α 7 subunit is coloured yellow, whereas the dup α 7 subunit is coloured blue.

In receptor pentamers containing dup α 7 subunits, A-Dup, AB-Dup, AC-Dup, and ACD-Dup represented the plausible arrangements of functional receptors, as described in Chapter 4. Building on these data and the results obtained for dimers, the extracellular domains of pentamers studied were used to identify the number of possible functional orthosteric binding sites existing in each pentameric receptor. UCSF DOCK6 was used to focus on those binding

modes calculated for Ach and Lob binding to dimers. For each receptor, the highest-populated cluster has been selected from the MD trajectory of the full-length receptor. Ach and Lob were docked using the same procedure as described for docking to dimers. The results are shown in Figure 58 and Table 6, and Table 7.

Comparing to the canonical (WT) $\alpha 7$ receptor, the number of mapped interdomain binding sites decreased in receptors bearing dup $\alpha 7$ subunits. The models containing one or two dup $\alpha 7$ subunits (A-Dup, AB-Dup, and AC-Dup) had one fewer interdomain interface than the WT $\alpha 7$ pentamer (Figure 58). ACD-Dup, which contains three dup $\alpha 7$ subunits, had only two interfaces mapped (Figure 58D).

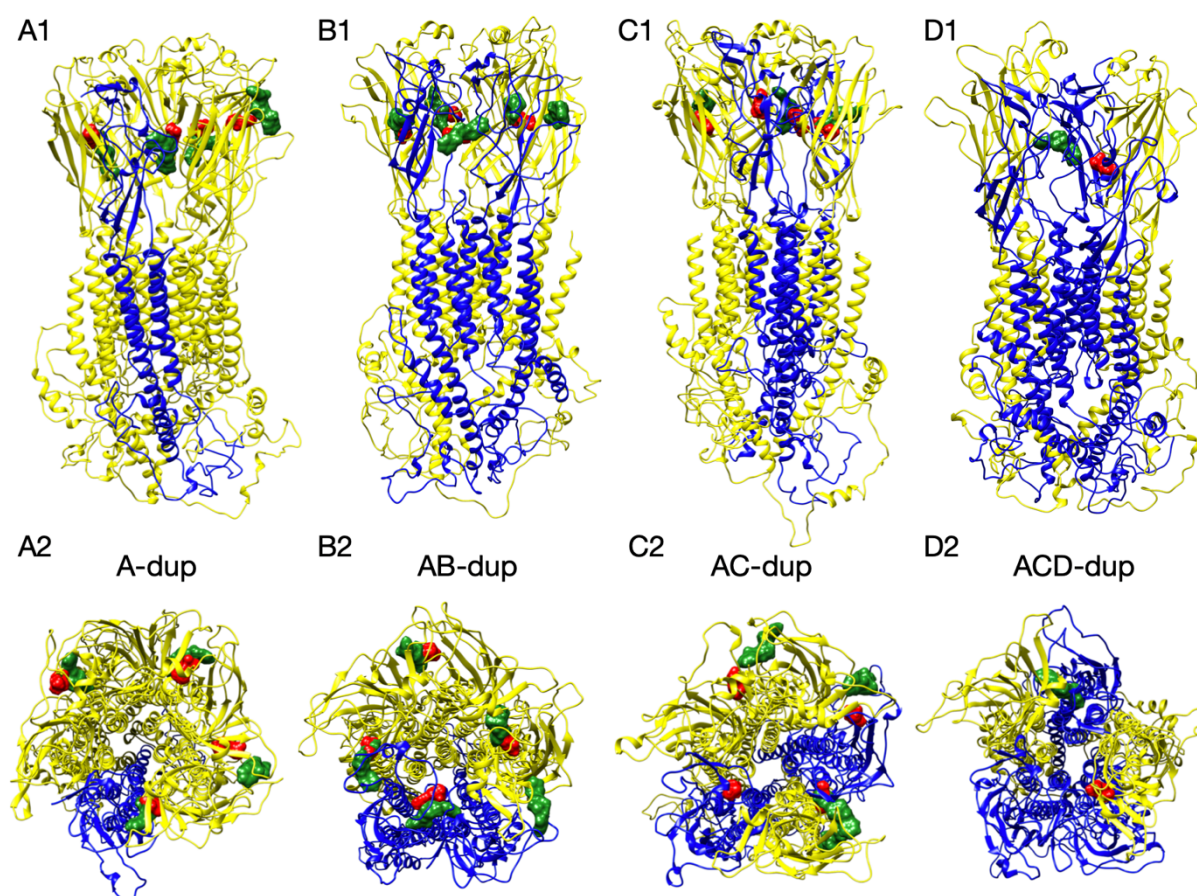


Figure 58: Molecular docking results of Ach and Lob agonists and four arrangements of dup $\alpha 7$ nicotinic receptor. Protein is represented as backbone, while ligands are rendered as red (Ach) and green (Lob) spheres. The canonical (WT) $\alpha 7$ subunit is coloured yellow, whereas the dup $\alpha 7$ subunit is coloured blue. A1&A2) Side view and top-down view of the A-Dup containing a single dup $\alpha 7$ subunit and four docked ligands; B1&B2) Side view and top-down view of the conformation of AB-Dup (two dup $\alpha 7$ subunits), four Ach and five Lob

ligands; C1&C2) Side view and top-down view of the conformation of AC-Dup (two dup α 7 subunits) and four Ach Lob ligands; D1&D2) Side view and top-down view of the conformation of ACD-Dup (three dup α 7 subunits) and one Ach and Lob bound.

Calculated binding scores and cluster sizes from DOCK6 are shown in Table 6 and Table 7. Both Ach and Lob showed moderate affinity, with a substantially more favourable binding for Ach (−190 KJ/mol for Ach; −50 kJ/mol for Lob) in all four dup α 7 models. Even though the calculated scores were less favourable than those calculated for the canonical α 7 receptors (scores: −153 KJ/mol for Ach; −50 KJ/mol for Lob), this may indicate that receptors containing dup α 7 subunits may retain functional orthosteric sites, and the presence of dup α 7 in the pentamer does not preclude the agonist binding. Lob showed a less favourable binding energy score and lower cluster size than Ach for its top hit, and this is due to Lob being considerably larger than Ach and the limited binding size volume.

Table 6: Ach ligand focus on docking with α 7 and four dup α models

| Model | Dimer | Cluster size | Grid Score (kJ/mol) | Electrostatic energy (kJ/mol) | VdW Energy (kJ/mol) |
|----------------|-------------|--------------|---------------------|-------------------------------|---------------------|
| α 7+Ach | WT-WT | 20 | -152.16 | -152.15 | -27.28 |
| A+Ach | AB(Dup-WT) | 12 | -190.38 | -166.37 | -24.00 |
| | BC (WT-WT) | 33 | -182.44 | -161.41 | -21.03 |
| | CD (WT-WT) | 48 | -193.31 | -166.41 | -26.90 |
| | DE (WT-WT) | 25 | -184.16 | -167.00 | -17.16 |
| AB+Ach | AB(Dup-Dup) | 13 | -174.69 | -151.85 | -22.84 |
| | CD (WT-WT) | 16 | -187.47 | -166.78 | -20.69 |
| | DE (WT-WT) | 36 | -171.82 | -150.84 | -20.98 |
| | EA (WT-Dup) | 28 | -186.79 | -158.09 | -28.71 |

| | | | | | |
|---------|-------------|----|---------|---------|--------|
| AC+Ach | BC (WT-Dup) | 15 | -180.12 | -155.85 | -24.27 |
| | CD(Dup-WT) | 26 | -173.63 | -149.78 | -23.85 |
| | DE (WT-WT) | 18 | -174.65 | -155.68 | -18.98 |
| | EA (WT-Dup) | 36 | -188.67 | -162.99 | -25.68 |
| ACD+Ach | DE (Dup-WT) | 54 | -193.33 | -165.89 | -27.44 |

Table 7: Lob ligand focus on docking with $\alpha 7$ and four dup α models

| Model | Dimer | Cluster size | Grid Score (kJ/mol) | Electrostatic energy (kJ/mol) | VdW Energy (kJ/mol) |
|-----------------|--------------|--------------|---------------------|-------------------------------|---------------------|
| $\alpha 7$ +lob | WT-WT | 13 | -50.21 | -10.18 | -40.03 |
| A+lob | AB (Dup-nor) | 20 | -53.80 | -7.33 | -46.47 |
| | BC (WT-WT) | 23 | -50.67 | -8.89 | -41.79 |
| | CD (WT-WT) | 6 | -48.37 | -10.00 | -38.37 |
| | DE (WT-WT) | 10 | -45.08 | -11.91 | -33.17 |
| AB+lob | AB (Dup-Dup) | 8 | -46.73 | -7.52 | -39.21 |
| | BC (Dup-WT) | 5 | -45.88 | -5.66 | -40.22 |
| | CD (WT-WT) | 9 | -45.04 | -5.02 | -40.00 |
| | DE (WT-WT) | 11 | -40.03 | -8.57 | -31.46 |
| | EA (WT-Dup) | 6 | -47.74 | -9.59 | -38.14 |
| AC+Lob | BC (WT-Dup) | 10 | -50.16 | -7.38 | -42.77 |
| | CD (Dup-WT) | 7 | -45.51 | -8.51 | -37.01 |
| | DE (Dup-Dup) | 4 | -51.71 | -7.16 | -44.55 |

| | | | | | |
|---------|-------------|---|--------|-------|--------|
| ACD+Lob | DE (Dup-WT) | 5 | -45.86 | -4.09 | -41.77 |
|---------|-------------|---|--------|-------|--------|

5.1.3 Prediction of allosteric binding sites using FTMap

To find the potential allosteric binding sites at $\alpha 7$ and dup $\alpha 7$ nicotinic receptors, solvent mapping using the FTMap webserver^{235, 236} has been used on outputs from all-atom MD simulations. FTMap is a fast method routinely used to detect new potential “druggable” sites on proteins^{235, 236}. In works presented in this chapter, EC domains of receptor pentamer were used in the FTMap calculations, extracted from 100 ns of all-atom MD simulations using cluster analysis. For each receptor, the centroid of the highest-populated cluster was selected.

The mapping result for the canonical $\alpha 7$ pentamer is shown in Figure 59. FTMap retrieved the orthosteric binding site, which is highlighted as red spheres, between two subunits. FTMap also detected two allosteric binding sites. One of them, rendered as a magenta sphere, is in the top part of the $\alpha 7$ subunit, surrounded by $\alpha 1$ and $\beta 4$ loops, close to the receptor channel (Figure 59), which consists of 39PLER42, 107IWK109 and Y173 residues. Another one is positioned in the centre of $\alpha 7$ subunit, surrounded by four β -sheet loops ($\beta 2$, $\beta 4$, $\beta 5$ and $\beta 6$), and it is highlighted by green spheres closed to L78, M80, 109KPDILL114, F126, 140YLPPGIF146 and F168 residues. To distinguish the pockets, one (magenta) is called the top pocket, while another one (green) is called the vestibule pocket.

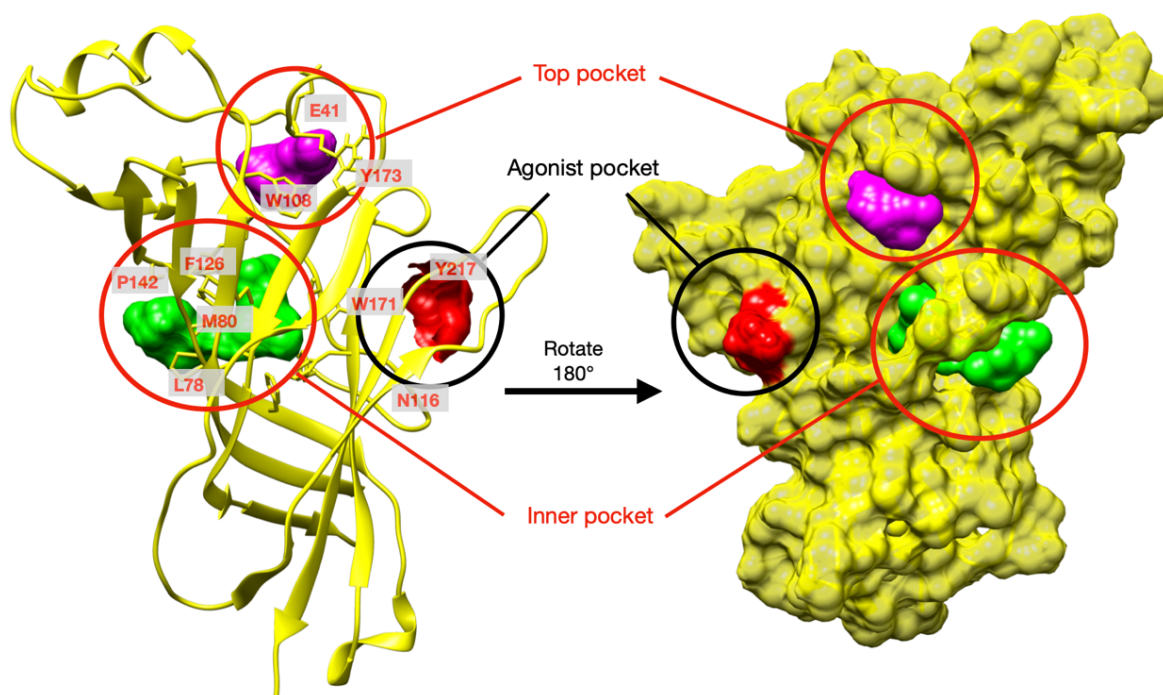


Figure 59: Mapping of $\alpha 7$ nAChR binding sites via FTMap. The left-hand figure is the front view of three binding pockets in the EC domain of the $\alpha 7$ subunit. A new cartoon yellow ribbon represented the $\alpha 7$ subunit. The right-hand figure is the back view. $\alpha 7$ subunit was shown in surface representation. The red spheres circled by black circles correspond to the ligand and occupies the agonist binding site. The magenta and green spheres circled by red circles corresponds to ligands and occupies the top, binding site and vestibule binding site, respectively. The main residues which formed the pockets are signed with red colour.

To map any putative allosteric sites of dup $\alpha 7$ nAChRs, the same workflow has been used. Figure 60 shows the results obtained for A-Dup, AB-Dup, AC-Dup and ACD-Dup receptors. The canonical (WT) $\alpha 7$ subunit is coloured yellow, whereas the dup $\alpha 7$ subunit is coloured blue. The locations of the small molecular FTMap probes are shown as different colour spheres. As for the canonical $\alpha 7$ receptor, the agonist binding site is highlighted by the red sphere. The top and vestibule pockets are shown as pink and green spheres, respectively, consistently with the representation in Figure 60. The blue sphere, representing the top pocket, has been detected for AB-Dup and AC-Dup only (Figure 60B and C). For AB-Dup, this site has been mapped on a WT-WT interface, while for AC-Dup, it has been mapped on a WT-dup $\alpha 7$ interface.

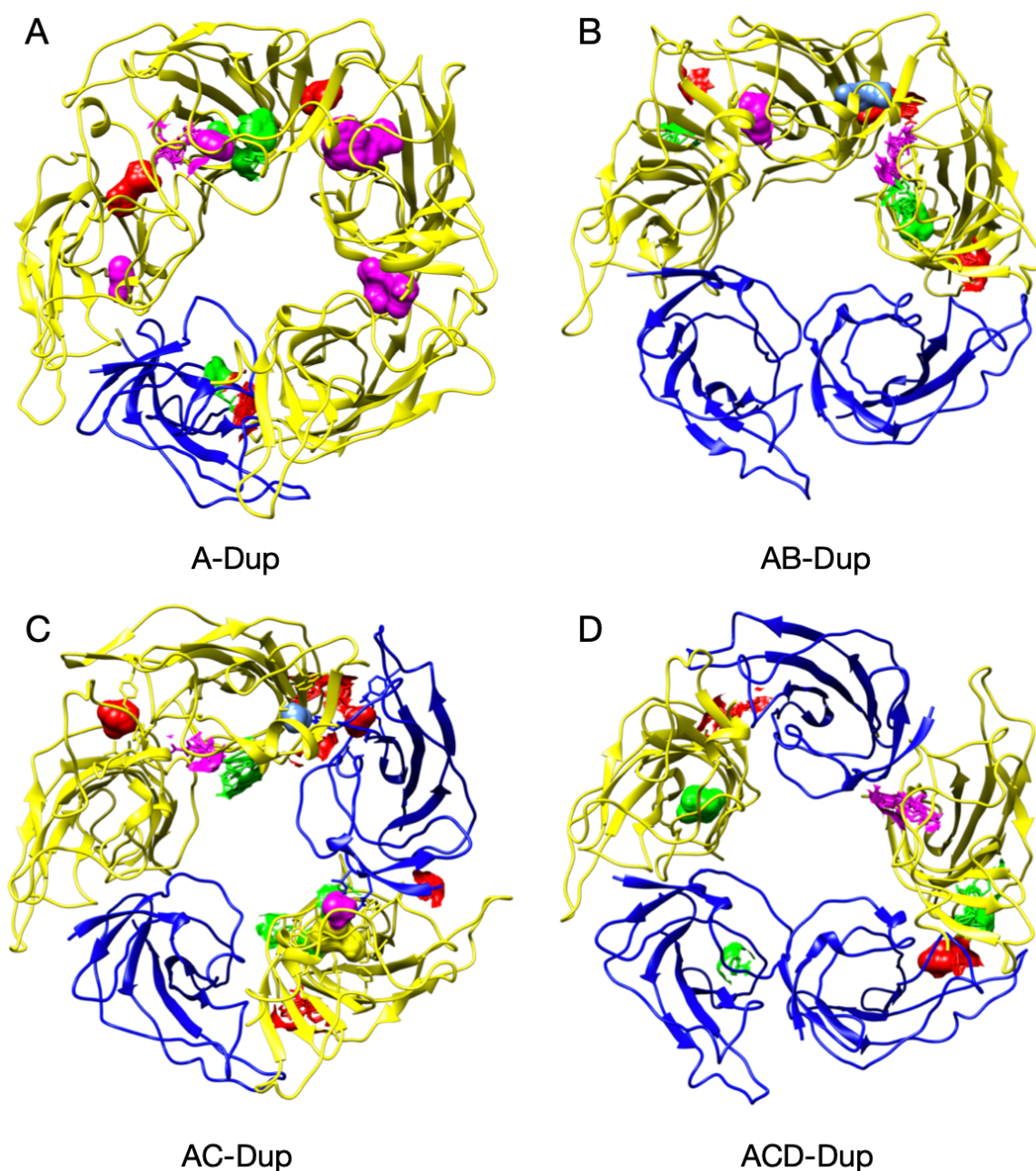


Figure 60: The binding site's results of four types of dup α 7 nAChR using FTMap. Red sphere represented agonist binding sites, pink sphere – top pocket, green sphere – vestibule pocket, blue sphere - top, binding pocket. A) Top-down view of A-Dup FTMap results contains 3 agonist binding sites, 4 top pockets, and 2 vestibule pockets. B) Top-down view of AB-Dup FTMap results contains 3 agonist binding sites, 3 top pockets, and 2 vestibule pockets. C) Top-down view of AC-Dup FTMap results contains 4 agonist binding sites, 3 top pockets, and 2 vestibule pockets. D) Top-down view of ACD-Dup FTMap results, it contains 2 agonist binding sites, 1 top pocket and 3 vestibule pockets.

In the A-Dup model, three agonist binding sites (red), four top pockets and two vestibule

pockets have been identified, with two binding sites associated with the dup α 7 subunit: one of them being an orthosteric binding site located at the dup-WT dimer interface, and another one being an internal pocket situated within the dup α 7 subunit. As for the two dup α 7 models, there are three agonist binding sites in AB-Dup and four in AC-Dup. Considering the potential allosteric binding sites, AB-Dup and AC-Dup contain the same number and type of pockets: two top pockets, two vestibule pockets, and one new pocket. The positions of all these allosteric binding sites are similar to canonical α 7 subunits, and there are no internal sites identified in dup α 7 subunits. The ACD-Dup model without a WT-WT interface displayed one agonist binding site at the interface of each of the two dup-WT dimers, one top pocket within one WT subunit, and three vestibule pockets located at the inside of both standard and dup chain. dup-dup dimer had no orthosteric binding site detected by FTMap.

For canonical α 7 receptors, the positions of the orthosteric binding sites were successfully retrieved by both methods: blind docking using UCSF DOCK6 and FTmap. For the A-Dup model, UCSF DOCK missed one site in the interfaces of BC chain dimer (WT-WT dimer), and for AB-Dup, it missed the binding site in dup-dup dimer. The orthosteric binding sites appeared more often at the dup-WT interface than at the WT-dup interface in UCSF DOCK6 results. No orthosteric binding sites were detected at the dup-dup interface by FTMap. This indicates that both methods have different limitations, and two complementary methods should be used in “druggability” studies of large, flexible, and thus challenging protein complexes such as nicotinic receptors.

Regarding the putative allosteric binding sites, the highest number of top pockets have been found in A-dup, followed by AB-Dup and AC-Dup containing two top pockets, while only one has been retrieved in the ACD-Dup model. This trend has been reversed for the vestibule pockets: three vestibule pockets were detected in the ACD-Dup model, which is the highest number in four dup models; all other models contained two mapped pockets. The top pockets are preferred in the canonical α 7 subunit, smaller numbers in the receptors with more dup α 7 subunits. Conversely, the vestibule pockets can exist in both canonical α 7 subunit and dup α 7 subunit.

5.1.4 “Druggability” of the allosteric binding sites probed by UCSF DOCK6

Building on the results retrieved by FTMap, detected binding sites were used in docking the focused library of 1790 nAChR allosteric. All averaged clusters of EC domain (pentamer) structures obtained from 100 ns MD simulations in DPPC membrane were used as targets. After blind docking, the results were ranked according to the number of clusters (from highest to lowest) and scores (from lowest to highest), respectively, and the top 200 docking poses from each group were selected to analyse ligand binding positions. Hence, potential binding sites were inferred from the results obtained by FTMap and blind docking of agonists Ach and Lob.

As expected, docking results of the focused ligand library to the canonical $\alpha 7$ nAChR retrieved results highly similar to those obtained by FTMap. In addition to retrieving all binding sites detected by FTMap, two new binding sites were discovered using the blind docking approach. These new sites are shown in Figure 61 as purple and orange spheres. Both pockets are close to the orthosteric binding site, one above and another below, and hence they were denoted as upper and down pockets, respectively.

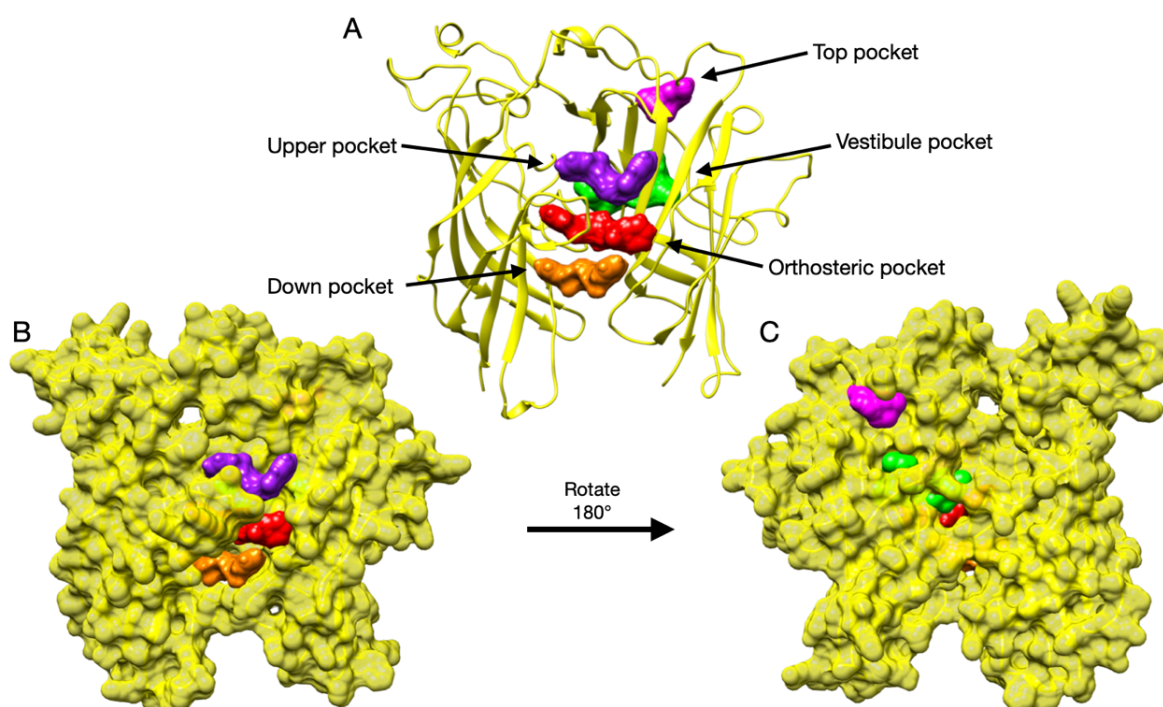


Figure 61: An overview of the orthosteric pocket and four allosteric binding sites discovered for the canonical $\alpha 7$ nAChR. A) $\alpha 7$ nAChR is shown in cartoon representation. Lob and Ach

are rendered as red spheres, occupying the orthosteric site. The pink sphere links to the allosteric top pocket, the green sphere links to the vestibule allosteric pocket. The purple sphere represents the upper pocket, and the orange sphere represents the allosteric down pocket. B) The front view of the three binding sites in $\alpha 7$ WT-WT interface, the model of $\alpha 7$ nAChR is rendered as solid surface and coloured yellow. Only the upper pocket, orthosteric pocket and down pocket can be seen from the front view. C) The back view of the three binding sites at the $\alpha 7$ WT-WT interface. Top (pink), orthosteric (red), and vestibule (green) pockets can be seen from the back view.

In dup $\alpha 7$ receptors, specifically in the A-Dup, AB-Dup and AC-Dup models, the DOCK6 blind docking results (Figure 62) replicated very well the FTMap results, but the results obtained for the ACD-Dup model showed a different number and location of binding sites than the FTMap results. The total number of retrieved orthosteric binding sites were highest in A-Dup and AC-Dup models, followed by AB-Dup. However, the blind docking detected no orthosteric site on the ACD-Dup receptor. Two additional vestibule sites were detected in the A-Dup model, and one additional top pocket and one additional vestibule pocket were identified in the AC-Dup model. These binding sites are shown in Figure 62.

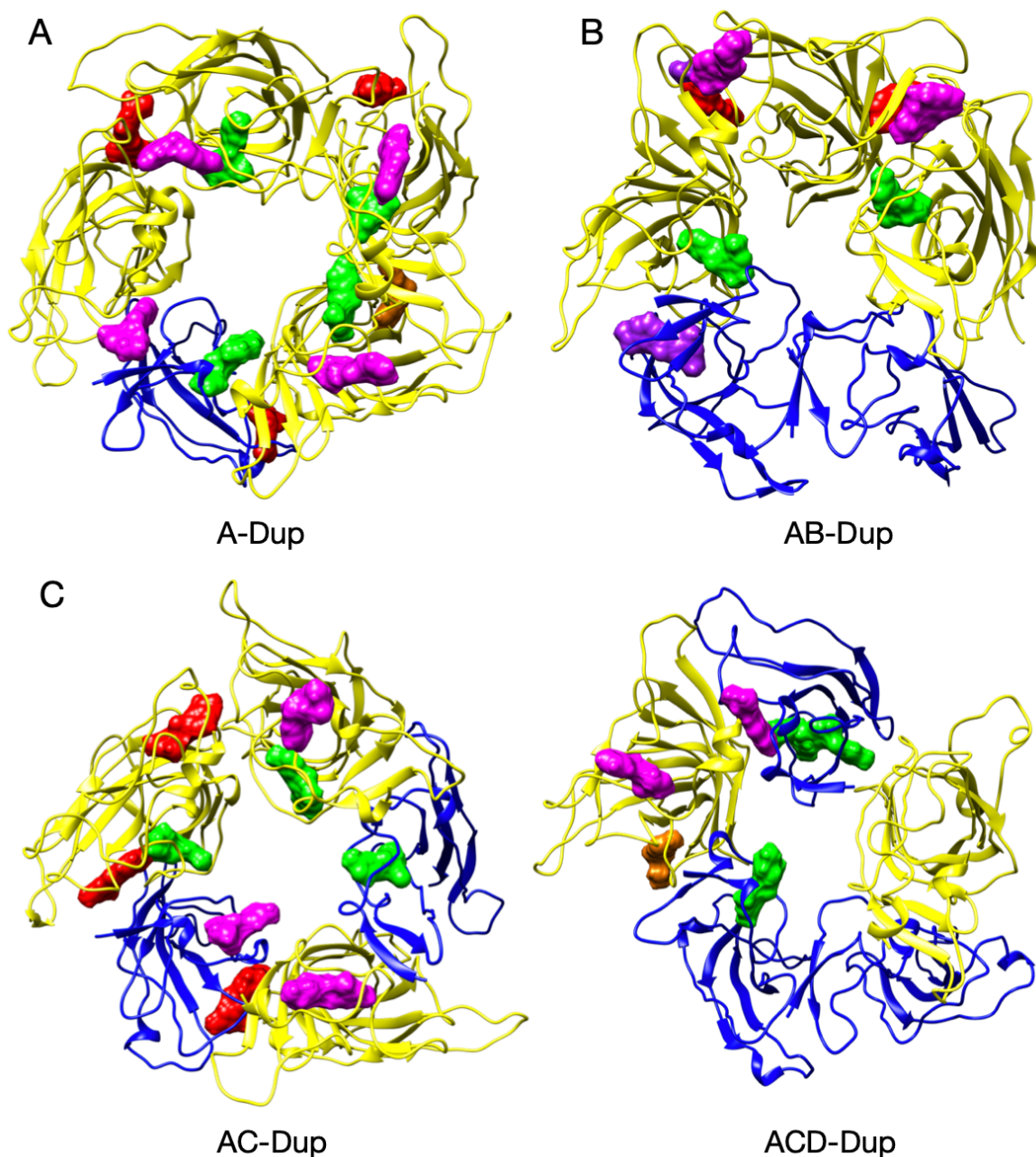


Figure 62: The binding sites reported by the blind docking for four dup α 7 nAChRs. Red spheres represent the orthosteric agonist binding site, pink spheres – top pocket, green – vestibule pocket, orange - down pocket, purple – the upper pocket. A) Top-down view of A-Dup DOCK6 results, it contains 3 agonist binding sites, 4 top pockets, 4 vestibule pockets and 1 down pocket. B) Top-down view of AB-Dup DOCK6 results, it contains 2 agonist binding sites, 2 top pockets, 2 vestibule pockets and 2 upper pockets. C) Top-down view of AC-Dup DOCK6 results, it contains 3 agonist binding sites, 3 top pockets, 3 vestibule pockets. D) Top-down view of ACD-Dup DOCK6 results, it contains 2 top pockets, 2 vestibule pockets and 1 down pocket.

5.2 Discussion

5.2.1 Dupa7/ α 7 ratio affects the ligand binding and may be linked to the nicotinic translational gap

The α 7 receptor has been a promising target for diseases affecting cognition. However, the results gathered in animal studies failed to translate into human clinical trials identifying a translational gap. As *CHRFAM7A* is human-specific, it was not included in those preclinical studies, and effects arising from its distinct structural and dynamic features were not considered. As the *CHRFAM7A* gene is present in different copy number variations in the human genome with high frequency²⁶³, understanding distinct features of dupa7 may offer novel insights when exploring the human α 7 receptor as a drug target.

Recent reports have shown the direct interactions between α 7 receptors and $A\beta_{42}$ ^{156, 158, 266}. These studies strongly suggest that the α 7 receptor can contribute to synaptic dysfunction in Alzheimer's disease as $A\beta$ oligomers can alter neuronal signalling through interactions with nicotinic receptors, particularly with α 7. However, how exactly $A\beta$ interacts with the α 7 receptor and whether human-specific dupa7 increases or decreases those interactions have not been fully understood.

Regarding interactions with disease-linked macromolecular ligands, our results indicate that in mixed functional receptors (i.e., comprising of both α 7 and dupa7 subunits), dupa7/ α 7 interfaces can bind $A\beta_{42}$ with a higher affinity than α 7/dupa7 and dupa7/dupa7, albeit impaired compared to canonical α 7/ α 7 sites. Receptors bearing dupa7 subunits are shown to be insensitive to α -BTX. These results collectively suggest that the receptors bearing dupa7 subunits may be less sensitive to effects exerted by neurotoxin or $A\beta_{42}$. This data is in agreement with the recent study, which focused on the function of *CHRFAM7A* alleles in vitro in two disease-relevant phenotypic readouts: electrophysiology and $A\beta$ uptake, and in the double-blind pharmacogenetic analysis on the effect of therapy using acetylcholine esterase inhibitors (AChEIs), based on *CHRFAM7A* carrier status¹⁴⁸.

Mechanistic insights arising from our work suggest competitive binding between α -BTX (an

orthosteric ligand) and A β ₄₂ to the α 7 receptors. The earlier studies support this characteristic, showing that both orthosteric agonists and antagonists mitigate A β uptake¹⁴⁸. Uptake of A β ₄₂ via α 7 receptors binding induces apoptosis, and orthosteric α 7 agonists mitigate the A β -induced apoptosis in animal models as reported by Szigeti et al. and references therein¹⁴⁸. The results of our study highlight the mechanistic link between receptor structure and A β binding, indicating key differences between α 7 receptors and receptors bearing dup α 7 subunits, which may be translated to the clinic. Moreover, our results suggest that in receptors containing dup α 7 subunits, A β ₄₂ might be competitive to α -BTX, albeit its binding affinity is low, hence the significance of this potential competing is challenging to estimate. Further follow-up studies are needed to validate these findings.

At the time of translation to the clinical trials, virtually all drugs effective in animals have demonstrated a lack of efficacy in humans, showing a robust translational gap. Dup α 7 functional studies are sparse and are lacking in the clinical context. Clinical efforts need to be continued with a trial design incorporating dup α 7 distinct structural biology, pharmacology and pharmacogenetics. Dup α 7 non-carriers account for 25% of the Alzheimer's disease population, which is significant considering an increasing number of AD patients. Our results, which match neuronal toxicity data published¹⁴⁸, suggest that dup α 7 carriers should be protected against A β effects to some extent, and dup α 7 non-carriers should be more acutely affected by A β effects. Therefore, therapeutics that reduce amyloid burden could be effective in non-carriers. Considering the number of AD patients worldwide and AD being essentially an unmet clinical need, these findings pave the way to bring new AD therapeutics into the clinic.

5.2.2 The effect of dup α 7 subunits on orthosteric binding sites

In this work, one of the main focuses was identifying new allosteric binding sites in the canonical α 7 nAChR and dup α 7 nAChR. While experimental studies reported the specific location of the orthosteric binding sites shared by several nicotinic receptors, allosteric binding sites are generally more unique to different receptor types^{79, 127}. Therefore, it was likely that some of those sites may be unique for receptors containing human-specific dup α 7 subunits and thus serve as targets in structure-based ligand design efforts to develop new small molecule

diagnostics and therapeutics in the future.

Two complementary methods were used in binding site mapping: solvent-mapping by FTMap webserver and blind docking (wherein no specific location of the binding site is given) using UCSF DOCK6. Both methods successfully retrieved the location of the orthosteric binding site, known from experimental studies published to date^{79, 127}. In addition, the data indicates that orthosteric binding sites can exist not only at the interface between two canonical $\alpha 7$ subunits (WT-WT) but also at interfaces between dup-WT and WT-dup subunits. Dup-dup interface is highly flexible and unstable, so its orthosteric site is unlikely to hold. The total number of functional orthosteric binding sites mapped for the receptor pentamers decreased when the number of dup $\alpha 7$ subunits in the pentamer increased. This profile is consistent with the published report⁹⁸ suggesting that the dup $\alpha 7$ subunits affect the formation of agonist binding sites due to the absence of specific binding site-forming loops, leading to a dramatic change in the configuration of the interface between two subunits forming the orthosteric site. However, the results show that interfaces containing dup $\alpha 7$ subunits may still retain the functional orthosteric site, as long as the interface remains in the dup-WT, rather than the WT-dup configuration.

5.2.3 The effects of dup $\alpha 7$ subunits on allosteric binding sites

In this work, several putative allosteric sites were identified. Several of those sites are located at the single subunit rather than at the interface between two subunits. There is a precedence of such allosteric binding sites, reported for humanised acetylcholine binding protein (PDB codes: 5OUG and 5OUI), which raises the confidence in those results, yet experimental validation, e.g. by mutagenesis or structural biology studies are needed.

The names and relative positions of four allosteric binding sites identified in this work at the canonical $\alpha 7$ receptor are: 1) The so-called top pocket, which is formed by $\alpha 1$ loop and $\beta 4$ loop and may involve $\alpha 1$, $\beta 2$ and $\beta 3$ loops of the adjacent subunit; 2) The vestibule pocket inside of the subunit, which is formed by four β -sheet loops ($\beta 2$, $\beta 4$, $\beta 5$ and $\beta 6$ loops); 3) The upper pocket, which is interfacial and located above the orthosteric binding site, and it is formed by

loops B and C in one subunit and loops $\beta 1$, $\beta 2$ in the adjacent subunit; 4) The down pocket, which is lower than the orthosteric site, is formed by loops A, $\beta 6$ and $\beta 7$ in one subunit, and loops $\beta 1$, $\beta 8$ in the adjacent subunit. These allosteric binding sites have also been found in the dup $\alpha 7$ receptors, although no allosteric binding sites were found at the dup-dup interface, which is apparent in the ACD-Dup receptor. Results obtained by both FTMap and DOCK6 were consistent and indicated that the “druggable” binding sites appeared in fewer positions in the dup model, especially in receptors containing three duplicated subunits (ACD-Dup). The interface between dup-dup subunits does not seem to form functional binding sites, whether orthosteric or allosteric. This formation is consistent with the data obtained from the all-atom MD simulation, which shows dup subunits' high flexibility. This flexibility can explain the differences between results reported by FTMap and DOCK6 for the ACD-dup receptor model.

The results obtained in this study offer for the first time an insight into the distinct “druggability” of human-specific receptors containing dup $\alpha 7$ subunits. They can assist in interpreting published clinical data, showing different responses of patients bearing dup $\alpha 7$ receptors to nicotinic agents. Although qualitative rather than quantitative, these results may – after the validation (e.g., by site-directed mutagenesis studies) – be utilised in structure-guided development of subunit-selective ligands, which can be rendered as future diagnostics and therapeutics.

5.3 Materials and Methods

5.3.1 Protein preparation

The models of EC domain and whole pentamer of $\alpha 7$ model, dup $\alpha 7$ models (A-Dup, AB-Dup, AC-Dup and ACD-Dup) and dimers (WT-WT, WT-dup, dup-WT, dup-dup) to be used for docking were created via the Gromacs tool ‘Cluster’ from the trajectory from 100 ns MD simulation (Chapter 4). Two most-populated clusters were selected for each trajectory.

5.3.2. Small molecule preparation and the construction of the target-focused small molecular library

Ach and Lob structures were built in UCSF Chimera, and their SMILES strings were obtained from ChEMBL. Missing hydrogen atoms were added, and partial atomic charges were calculated using the AM1-BCC²⁶⁷. The nAChR library of 1793 fragments is built from the ChEMBL website when searching ‘neuronal acetylcholine receptor protein alpha-7 subunit and selecting Homo sapiens as a target. The downloaded ‘all bioactivity data’ and picked allosteric ligands from it. Openbabel (<http://openbabel.sf.net/>) was used to prepare the ligands using MM94 with AM1-BCC charges²⁶⁷. The receptor was prepared using the Dockprep tool in the UCSF Chimera, with charges obtained from AMBER14SB. The docking procedures were as follows: sphgen was used to map binding areas, with selected all spheres generated. The box for docking was generated with a distance from the edge of the spheres of 0.6 nm. The grid for docking was generated using the dock6 grid tool, with a 0.3 spacing distance between points. 100 poses were generated per ligands, each with 100 orientations each. The clustering RMSD was 0.25 nm, with the molecules being ranked via their total AMBER grid score.

5.3.3 Atomistic molecular dynamics (MD) simulations

Atomistic MD simulations have been carried out to generate ensembles to get the average structure of four types of dimers to evaluate the effect of the dup subunit on the dynamics. All simulations were performed using Gromacs 2016.3²⁵⁶. The protein was parametrised using the AMBER99SB-ILDN force field and TIP3P water model²¹⁸. Box distance was set to 1 nm, and periodic boundary conditions were applied. The box was solvated and Na⁺ and Cl⁻ ions were added to achieve a 0.1 M concentration and maintain the charge neutrality of the unit. The solvated receptor-membrane systems were energy minimised and equilibrated. The minimisation ran using the steepest descent. The energy step size was set to 0.2 nm, and the maximum number of steps was set to 50,000. The minimisation was stopped when the maximum force fell below 800 kJ/mol/nm using the Verlet cutoff scheme. Treatment of long-range electrostatic interactions was set to Particle Mesh-Ewald (PME)²⁶⁰, and the short-range electrostatic and van der Waals cutoff was set to 1.0 nm. After the energy minimisation, heating

to 300 K was performed for 5 ps with a time step of 1 fs and position restraints applied to the backbone in an NVT ensemble. The constraint algorithm used was LINCS, which was applied to all bonds and angles in protein²⁶¹. The cutoff for non-bonded short-range interaction was set to 2.0 nm with the Verlet cutoff scheme. Long-range electrostatics were set to PME. The temperature coupling was set between the protein and the non-protein entities using a Berendsen thermostat, with a time constant of 0.1 ps and the temperature set to reach 300 K with the pressure coupling off. Pressure equilibration was run at 300 K with a Parrinello-Rahman barostat and set to 1 bar²²⁵ in an NPT ensemble. The equilibration trajectories were set to 10 ps (discarded from the analysis), and the production MD simulations were performed for 100 ns.

The trajectories were analysed using GROMACS tools, including RMSD to assess overall stability, per-residue RMSF to assess the local flexibility.

For all-atom molecular dynamics simulations, standard workstations and a GPU in-house server were used. These included workstations with Intel i7 7th generation processor, with 16 GB of RAM with a NVIDIA GTX1080, and a dedicated server with 32 GB OS-RAM, a Xeon 44 core processor with multithreading, and 2x NVIDIA GTX1080Ti GPUs.

5.3.4 Binding site mapping by FTMap

Two highest-populated clusters of both EC domain and whole pentamer of each $\alpha 7$ and dup $\alpha 7$ (A-Dup, AB-Dup, AC-Dup, ACD-Dup) protein structure after 100 ns all-atom MD simulations were prepared for FTMap analysis (PDB format with all bound ligands and solvent molecules removed).

FTMap^{235, 236} is a solvent mapping server used to search and predict binding hotspots for proteins. FTMap uses 16 simple organic molecules (e.g. benzene, acetic acid) as probes to dock them, detect all putative binding sites (hotspots), and rank those hotspots by the interaction energy term²³⁵. The top hotspots would be the most likely binding sites for any given protein.

5.3.5 Molecular docking of A β 42 and α -Bungarotoxin (α -BTX)

The analysis of interactions between different $\alpha 7$ /dup $\alpha 7$ receptors and two established macromolecular ligands: A β and α -BTX were performed by molecular docking. A β (PDB code: 6RHY) and α -BTX (PDB code:4HQP) were docked to all four possible combinations of $\alpha 7$ and dup $\alpha 7$ interfaces using ClusPro web server^{244, 268}. The top 10 lowest-energy complexes were selected to further analysis, and the binding affinities were calculated by SeeSAR (www.biosolveit.de)²⁶⁹, using the HYDE scoring function²⁴².

5.3.6 Molecular docking of small molecules

The results of FTMap analysis (A-Dup, AB-Dup, AC-Dup, and ACD-Dup extracellular portions) were subjected to molecular docking of small molecules using UCSF DOCK6²⁴⁰ software. Binding sites selected by FTmap were validated using the DOCK6 accessory 'sphgen' to produce 'negative images' of all binding sites. Subsets of sphgen spheres corresponding to the hotspots identified by FTMap were assigned as binding sites and used for docking. A box of the length of 6 Å from the edge of the identified spheres was created, which was centered around said spheres. A grid spacing of 0.25 Å was using with a bump overlap of 0.5. Flexible docking was performed with an internal energy cut-off of 1000 kcal mol⁻¹. The ligands were oriented to the spheres with a maximum of 1,000 orientations per ligand tested. Initially, validated ligands with known binding modes and potencies, acetylcholine (Ach) and lobeline (Lob), were positive controls. After successful validation, a focused library of 1,793 molecules was docked and virtually screened.

Molecular docking calculations were ran using Newcastle University's Tier 3 HPC facility, HPC Rocket, with an average of 2 nodes per docking run. Each node utilised had 2 Intel Xeon E5-2699 v4 processors and 128 GB of memory.

Chapter 6 Addressing “druggability” of different polymorphs of 18-kDa translocator protein (TSPO)

Neuroinflammation is a pivotal contributor to the pathophysiology of neurodegenerative disorders. The link between stimulation of nicotinic receptors, particularly $\alpha 7$ nAChRs, and reducing the inflammation has been established, validating those receptors as attractive targets for therapeutics and diagnostics, e.g., for positron emission tomography (PET) imaging. However, the existence of human-specific isoforms of $\alpha 7$ receptor, described in previous chapters, and its prevalence in the general population, hampered the development of PET tracers specific to $\alpha 7$, and several complementary targets have been proposed over the last years.

The mitochondrial translocator protein (18 kDa) (TSPO) has been the most extensively studied target in the context of PET imaging²⁷⁰. This highly conserved transmembrane protein is highly upregulated in microglia and astrocytes in response to inflammatory stimulation, and its involvement in neuroinflammation has been validated over past years²⁷¹. TSPO contains five transmembrane α -helices spanning the outer mitochondrial membrane, and its function has been linked to the transport of highly lipophilic small molecules such as cholesterol from the cytosol into the mitochondria²⁷². TSPO forms oligomers, and it also interacts with voltage-dependent anion channel (VDAC) and the adenine nucleotide carrier (ANC) to form a complex involved in regulating mitochondrial homeostasis.

TSPO has been initially identified as a peripheral-type benzodiazepine receptor, but it can bind a broad range of chemically diverse small molecules other than benzodiazepines, including endogenous (e.g., porphyrins and heme, cholesterol) and exogenous (e.g., isoquinoline carboxamides, imidazopyridines) compounds^{39, 273, 274}. Because of its up-regulation in neuroinflammation and a vast range of its ligands suitable as tracers for PET, it can be used as a target for diagnostics in animal models of neurodegenerative diseases and human patients. PET studies with radioligands targeting TSPO have been pivotal in understanding the complex role neuroinflammation plays in disorders affecting the brain.

In clinical practice, the application of radiolabelled TSPO ligands allows for detecting and

tracking neuroinflammatory lesions in patients affected by diseases such as Alzheimer's and Parkinson's. PK11195 and DPA-714 are two established TSPO radioligands¹⁸⁶. Unfortunately, [¹¹C]PK11195, the first and most widely used TSPO radioligand, has severe limitations, while the next-generation TSPO radioligands have suffered from high interindividual variability in binding due to a genetic polymorphism in the TSPO gene (rs6971, corresponding to A147T mutation)²⁷⁵. Radiosynthesis of [¹⁸F]DPA714 has specific synthetic challenges associated, but its scaffold could serve as a parent compound in developing next-generation TSPO tracers²⁷⁶.

In the work presented in this chapter, the scaffold of [¹⁸F]DPA714 was redesigned, changing the location of the fluorine atom from C28 to different positions within the molecule (Figure 64). This change has been done with two aims in mind: (i) facilitate radiosynthesis and (ii) simultaneously improve pharmacodynamics, particularly interactions with the A147T TSPO variant. In the structure-guided procedure, properties such as synthesis facilitation have been prioritized. However, this generated two distinct enantiomers, denoted as MKL and MKD, which required to have their properties (i.e., interactions with TSPO, binding affinity, and pharmacokinetics) evaluated separately. A molecular model of the A147T variant of human TSPO was developed to investigate the difference between the enantiomers. This investigation was made alongside mapping the binding site and predicting the binding mode of DPA714 and its derivatives. Moreover, the binding affinity and conformational changes resulting from the binding were calculated, rationalising their difference.

6.1 Results

6.1.1 Molecular modelling of human TSPO variants and binding site mapping

The A147T TSPO variant (shown in Figure 63A). While the interfacial binding sites are likely important for cholesterol transport, this work focused on the central pocket (yellow). This selection was made in order to be consistent with experimental structures of PK11195 bound to TSPO. The docking reproduced near-native PK11195 binding mode, as well as its affinity range. This pocket comprises residues CYS19, SER23, HIS43, HIS46, LEU49, GLY50, TRP107, VAL110, LEU114, TRP143, LEU150. DPA-714 has been docked to the same pocket, with

reasonably exhaustive docking sampling, and from the 500 poses generated for each of the ten receptor representants (obtained via MD clustering). Five poses were selected, matching the experimental affinity range have been selected for further inspection and were found to form one distinct cluster of poses (Figure 63B). A representative complex from this cluster has been selected for MD simulation to evaluate the stability of the predicted binding mode. The binding affinity has been recalculated after the MD simulation and was found in agreement with the experimentally determined K_i value (DPA-714 with nanomolar affinity for TSPO).²⁷²

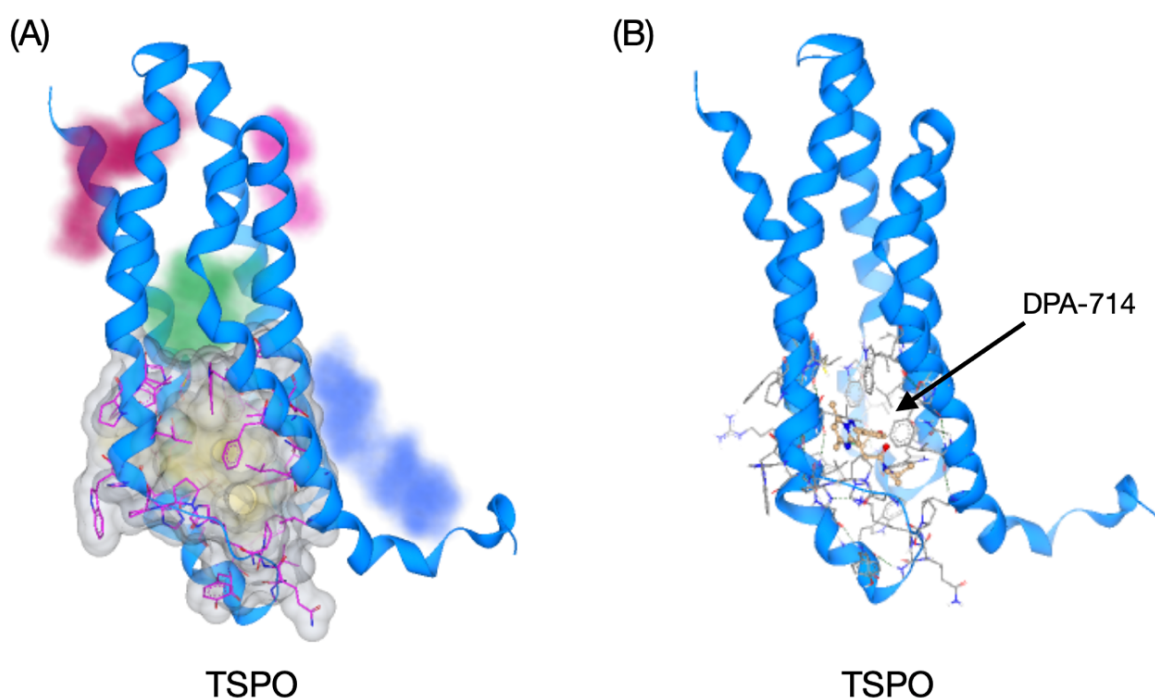


Figure 63: The detected binding sites of A147T TSPO. (A) Five binding sites were detected in the A147T TSPO variant via SeeSAR, shown in the red, pink, green, yellow, and blue region. The yellow region was covered with a grey surface, and the pink residue was selected docking with four ligands. The blue ribbon represented the conformation of A147T TSPO variant. (B) The dock results of the DPA-714 on A147T TSPO variants. The grey side chains belong to the A147T TSPO variant from the TSPO binding pocket.

6.1.1 Compound prioritisation

Since adding fluorine to the structure would facilitate the radiosynthesis of fluorinated DPA-714 derivatives, the position of the fluorine atom was investigated. We prioritised the compounds with the best-predicted affinity and a more straightforward synthetic route for ^{18}F

radioligands. Changing the fluorine (F) position from C28 to C10 atom resulted in two enantiomers, denoted as MKL and MKD (Figure 64), with the best-predicted affinity to both variants of TSPO and favourable synthetic route. The compounds were synthesized by Dr Mike Carroll's group, Newcastle University. In parallel, investigations of their proposed binding modes, particularly for the A147T variant, were carried out in our group.

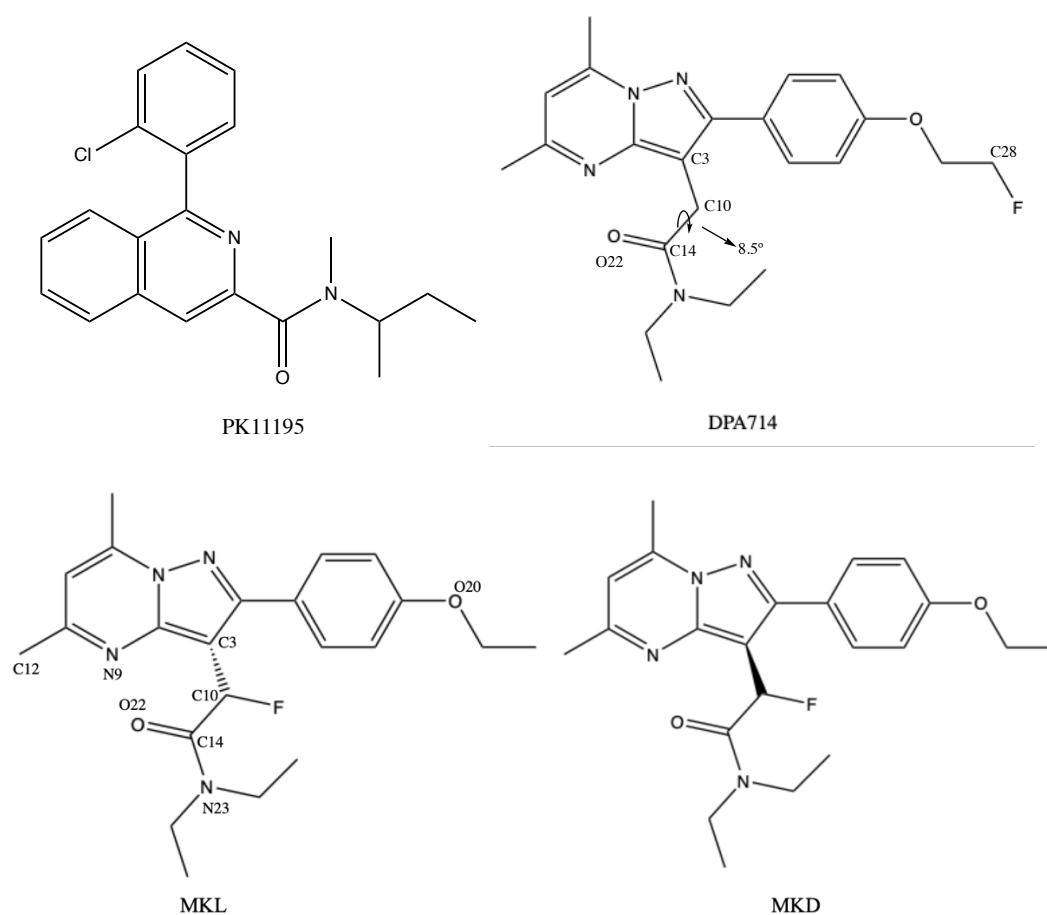


Figure 64: The structure of PK11195, DPA714, MKL and MKD compounds. The necessary atoms are signed in Figure, MKL and MKD share the same atom name and number.

Binding affinities were calculated for all ligands using the predicted pose of DPA-714 as a starting point. Lead optimisation (i.e. transformation from DPA-714 to MKL and MKD) and binding affinity calculations were performed using SeeSAR²⁴³. MKL was ranked as the best compound in terms of affinity (affinity (K_i calc) < pM), outperforming 2 orders of magnitude its stereoisomer, MKD (affinity (K_i calc) = nM). The driving forces behind the increased affinity (such as ligand conformation, fundamental interactions and desolvation energy) were

investigated to understand the plausible reasons for this predicted stereoselectivity. The predicted binding modes, critical interactions, and predicted binding affinities are shown in Figure 65.

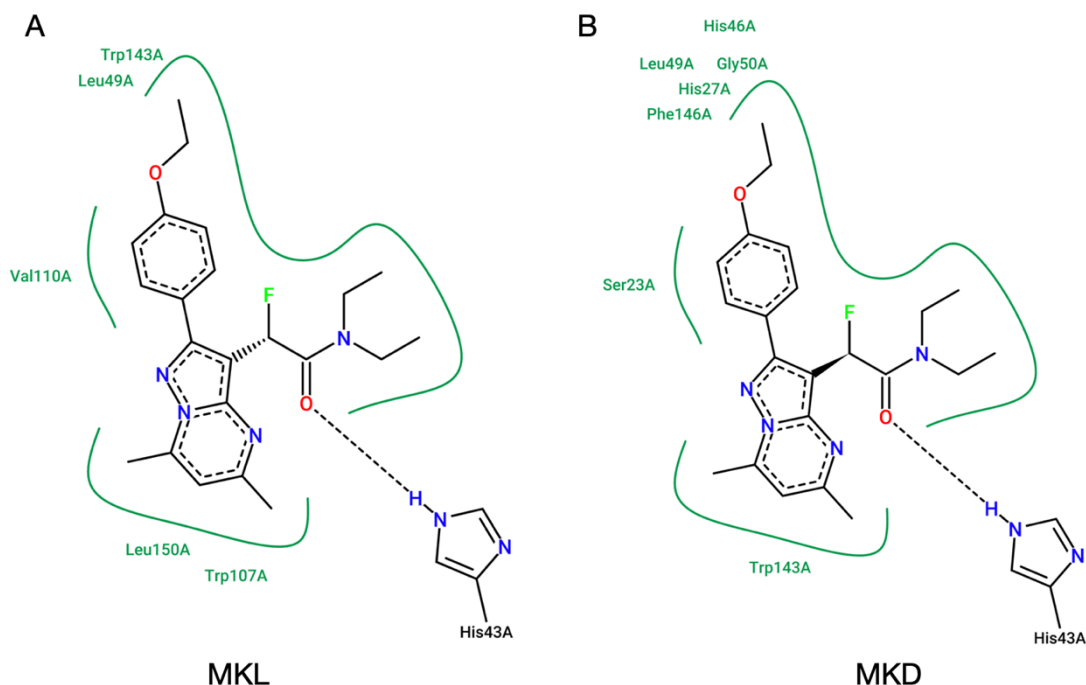


Figure 65: Diagram of TSPO in complex with MKL (A) or MKD (B). The interaction pattern comprises one hydrogen bond (signed as black dashed lines) and hydrophobic contacts (represented by the green labelled residues and green curve along with the contacting hydrophobic ligand parts).

Some differences in the obtained binding mode can be attributed to the torsional space of the parent compound and derivatives. The value of the C3-C10-C14-O22 (following the naming shown in figure 65) torsional angle of DPA-714 “parent” compound” was 8.5°. The torsional angles F-C10-C14-O22 and C3-C10-C14-O22 were 105.6° and 134.1° in MKL and 68.8° and 92.9° in MKD, respectively. This difference in torsions allowed a change of interactions contact points. For instance, as shown in Figure 65, hydrophobic contacts for MKL were different from that of MKD, increasing the stability of binding. Although the interactions with HIS43 were the same for both MDL and MKD, the latter buried itself, hence increasing its contact with G50, H27, F146, L49 and H46.

In terms of hydrogen bonds and desolvation energy (calculated via HYDE score and

represented by coloured “coronas” surrounding each heavy atom of the ligand, as shown in Figure 66, MKL showed the best binding properties of all ligands investigated. The only moiety of MKL with a small unfavourable contribution to the binding affinity was the ether oxygen O20 (Figure 66), whereas DPA714 and MKD showed several moieties with unfavourable contributions calculated. The ether oxygen (O22) and two nitrogens (N9 and N23) contributed unfavourably in DPA714, while in MKD, unfavourable contributions arose from O20, N9 nitrogen, and C12 carbon (Figure 66). The contribution from fluorine was more favourable in MKD (-6.6 kJ/mol) than in MKL (-0.4 kJ/mol). Nevertheless, this difference has been offset by the unfavourable contributions of several moieties in MKL described above.

The O22 in MKD yielded the most favourable desolvation contribution of -1.4 kJ/mol, with -0.6 kJ/mol desolvation contribution for O22 in MKL and unfavourable 2.3 kJ/mol contribution in DPA714. This indicates that the change of fluorine position can affect ‘intrinsic’ factors for the atoms nearby and the desolvation of the region, and it explains why no H-bond has been found in the DPA-714 complex in this region. This explains the lower calculated affinity of DPA714 compared to MKL and MKD. Collectively, this suggests that L-enantiomer (MKL) with a high binding affinity (low pM) would be optimal for the next-generation tracer for TSPO.

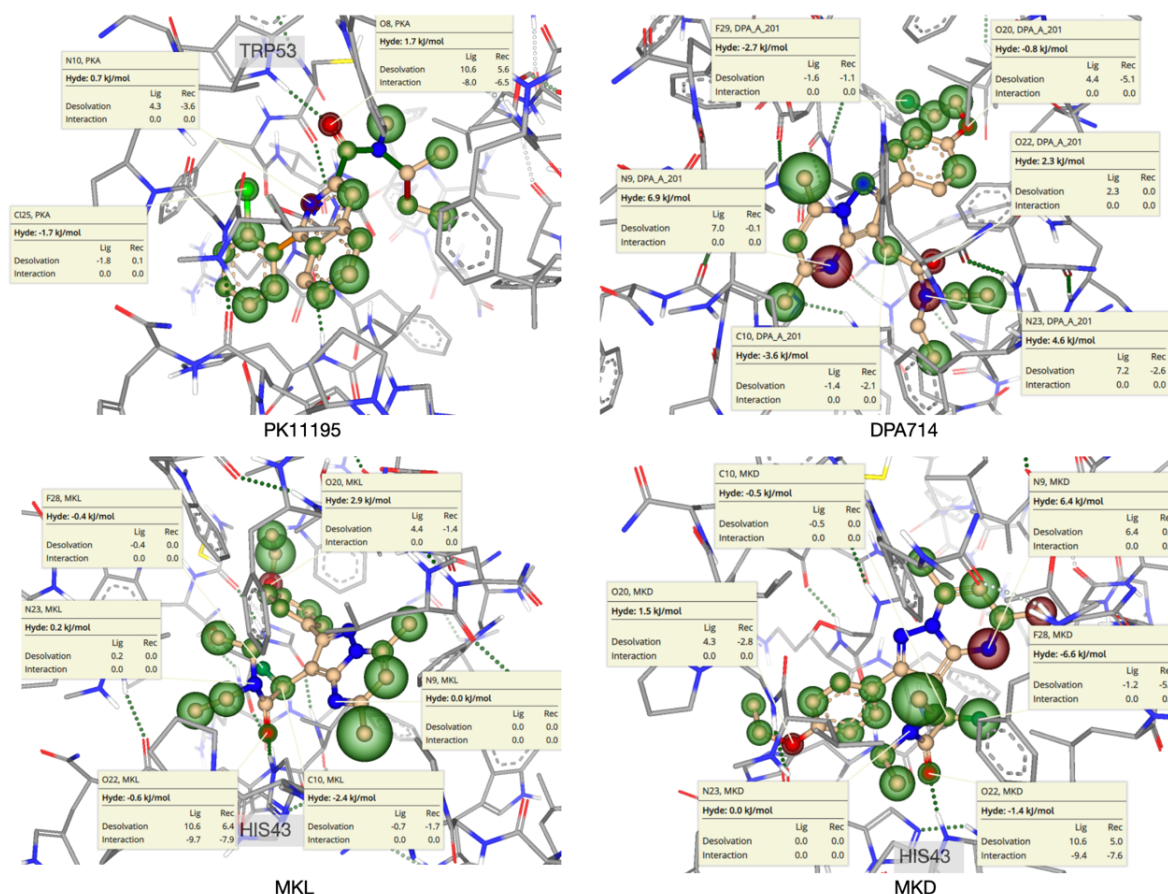


Figure 66: The predicted binding modes of four ligands described in this study in the A147T TSPO binding pocket. TSPO binding site is represented as a grey-coloured residue network. The ligands (PK11195, DPA714, MKL and MKD) are represented by balls and sticks, coloured by a heteroatom, with green (favourable) or red (unfavourable) “coronas” surrounding them. The size of the “corona” reflects the relative contribution to the binding. The dotted lines represent H-bonds (green – favourable).

6.1.2 The drug-like possibility of compounds

To evaluate the drug-likeness of the stereoisomers MKL and MKD, the more physiochemical, pharmacological and ADME properties (absorption, distribution, metabolism, and excretion) such as size, lipophilicity (logP), water solubility (logS), BBB permeant were evaluated using the webserver SwissADME. The results are shown in Table 8. PK11195 and DPA714 were reference substances. PK11195 show all the number about rotatable bonds (5) and h-bond acceptors (2) less than the detection of other three ligands (DPA714, MKL and MKD). As sharing the same central structure, the results of MKL and MKD are the same and almost the same with DPA714, for example, 5 H-bond acceptor, no H-bond donor, poorly water solubility

and all were predicted to cross the BBB, which is a suitable drug-like property required for central nervous system drugs. One slightly difference concerns the rotatable bonds. MKL and MKD contain 8, while DPA714 contains 9, suggesting that the position of fluorine affects the rotation of a bond. This result corresponds to the H-bond results from the UCSF Chimera, whereby all the H acceptors in the established hydrogen bonds are ligands, and the H donors are protein receptors. Investigating the data for the newly designed ligands MKL and MKD, both were predicted with high drug-likeness and satisfied Lipinski's 5 rules: molecular weight no more than 500, hydrogen bond and hydrogen bond acceptor less than 5 and 10, respectively, and logP no more than 5. Therefore, MKL and MKD are suitable compounds to replace DPA714, in particular MKL.

Table 8: Properties of ligands detected using SwissADME.

| | PK11195 | DPA714 | MKL | MKD |
|--------------------------|---------|--------|------|------|
| Rotatable bonds | 5 | 9 | 8 | 8 |
| H-bond acceptors | 2 | 5 | 5 | 5 |
| H-bond donors | 0 | 0 | 0 | 0 |
| Lipophilicity (Log Po/w) | 4.63 | 3.66 | 3.8 | 3.8 |
| BBB permeant | Yes | Yes | yes | Yes |
| Drug-likeness | Yes | Yes | yes | Yes |
| Bioavailability score | 0.55 | 0.55 | 0.55 | 0.55 |

In addition to the SwissADME parameters, the drug-likeness parameters from SeeSAR were evaluated for a consensus. In terms of important drug-like properties such as molecular weight (MW), logP, and total polar surface area (TPSA), the compounds are very similar to the “parent”

compound, which can be expected. MKL and MKD are slightly more lipophilic due to a change of the fluorine position (calculated LogP = 4.29, instead of LogP of 3.77 calculated for DPA-714). TPSA and MW are virtually identical for all three compounds: MW=398.47 and TPSA = 60.25 Å² for DPA-714; MW=398.48 and TPSA = 60.25 Å² for MKL and MKD, respectively. Collectively, these indicate that both MKL and MKD should be suitable for PET tracers, considering their predicted bioavailability (MW, logP) and BBB penetration (TPSA).

6.1.3 Molecular dynamics studies of ligand-TSPO complexes

To evaluate the structural and dynamic properties of A147T TSPO variant complexes with ligands investigated in this chapter and compare the conformational changes exerted by different ligands, all-atom MD simulations of all ligand-protein complexes were carried out.

The RMSDs calculated on heavy atoms of ligands bound to the A147T TSPO variant reach a converged state prior 50 ns for all complexes except MKL, which showed an increase in RMSD during 56-67 ns simulation time. The geometry of this MKL changes to upright after 60 ns can be seen in Figure 67B. PK11195 and MKD had the lowest average RMSDs, which may reflect differences in the biological activities exerted by those ligands: PK11195 is an antagonist, DPA-714 is an agonist, the pharmacological activity of MKL and MKD is yet to be determined. Nevertheless, the differences in ligand RMSDs were small (less than 2Å for ‘extremes’, MKL and MKD), and all complexes were stable, with only minor conformational changes in ligands observed during the simulation and overall binding modes preserved (Figure 67B).

However, the binding of ligands induced conformational changes to the A147T TSPO variant. Figure 67C shows the RMSDs calculated on the A147T TSPO variant backbone over the simulation time, with the results calculated for the reference structure (apoTSPO) included. Of all structures, the lowest average RMSD has been obtained for the complex with PKA1195, a TSPO antagonist. The average RMSD values can be ranked as follows: PKA11195 < apoTSPO < DPA-714 = MKD = MKL. This ranking of average values may represent the effect of the ligands on the phenotypical effect of the A147T TSPO variant. The results suggest that MKL and MKD should form a high-affinity complex and exert a strong dynamic effect on the A147T

TSPO variant.

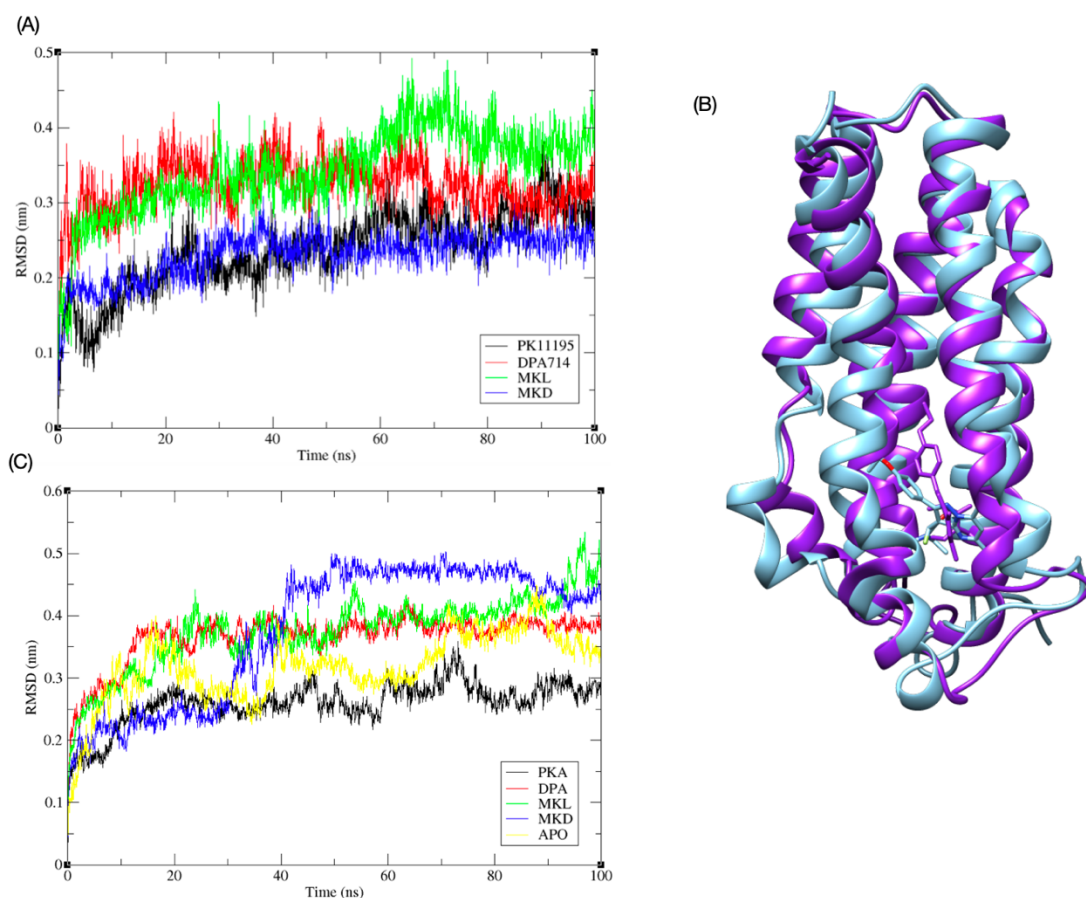


Figure 67: A) RMSD calculation of ligands fit A147T TSPO variant backbone in complexes during 100 ns MD simulation; the black represents PK11195, red – DPA714, green – MKL, and blue – MKD; B) Two complexes of A147T TSPO variant and MKL ligand, the configuration of blue complex sampled around 1 ns and the configuration of blue complex sampled around 60 ns. C) RMSD calculation of A147T TSPO variant backbones in complexes during 100 ns MD simulation. The black represents the A147T TSPO variant backbone in the PKA11195 complex, red – the A147T TSPO variant backbone in the DPA-714 complex, green – the A147T TSPO variant backbone in the MKL complex, blue – the TSPO backbone in the MKD complex. The apo A147T TSPO variant (APO) is shown in yellow as a reference.

The global, correlated motions of the A147T TSPO variant in complexes with ligands and apoTSPO have been assessed by the principal component analysis (PCA, Figure 68). Although all five models showed a slightly similar distribution, MKL has a weaker effect on the dynamics, being the most similar to the apoTSPO distribution.



Figure 68: Principal component 2D projection calculation of four ligands fit A147T TSPO variant backbone in the last 50ns during 100 ns of the atomistic MD simulation. The apo A147T TSPO variant (APO) is shown in blue as a reference. The black represents the A147T TSPO variant backbone in the PKA11195 complex, red – the A147T TSPO variant backbone in the DPA-714 complex, green – the A147T TSPO variant backbone in the MKL complex, purple – the TSPO backbone in the MKD complex.

The fluctuations per residue in TSPO during the simulation were assessed by calculating their root-mean-square deviation. As shown in Figure 69, the results for all ligands-TSPO complexes show a local increase in flexibility in the region spanning residues 92-108, which is close to the binding pocket. PK11195 shows enhanced flexibility in the residues 72-80, forming a short loop at the mitochondrial site. This trend is also visible for apoTSPO, albeit to a lower extent, but not in DPA-714 and its derivatives. The binding of MKD, MKL and DPA-714 increased the flexibility in the cytosolic loop 28-45, although this effect was more negligible for the latter two ligands. MKL and DPA-714 showed increased flexibility in the residues 60-67, which are part of the transmembrane helix at the mitochondrial site. This shows that each of the ligands has a different modulation effect on the variant TSPO.

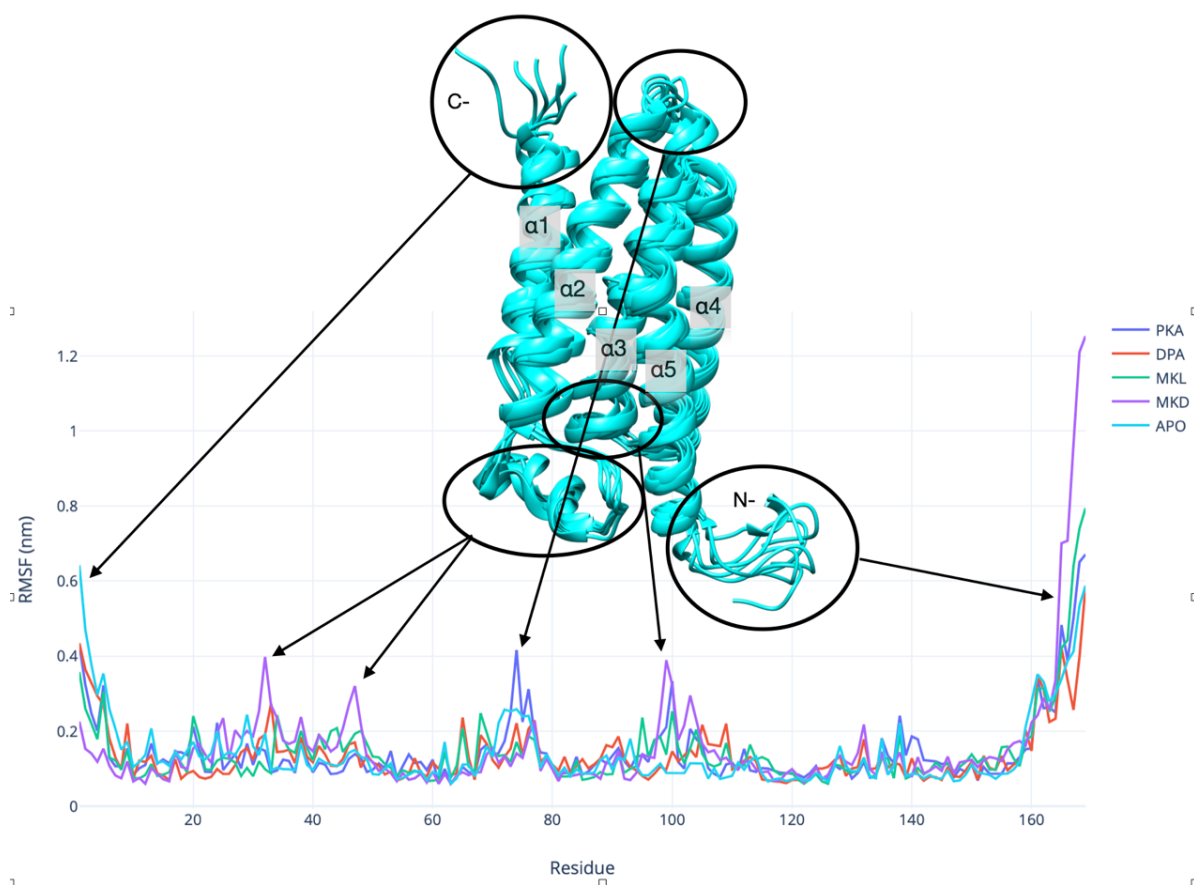


Figure 69: Analysis of the residual flexibility of A147T TSPO. The regions of protein structure in the five circles correspond to the five residue peaks shown in the RMSF plot. In the centre, six representative conformations of the apo A147T TSPO trajectory were sampled every 20ns in a 100ns simulation.

Analysis of H-bonding networks in the complexes and contact points between ligands and TSPO identified key residues in Figure 70 and listed in Table 9. These residues are in agreement with the initial docking results and with available experimental data. MKL complex contains the highest amount of intermolecular H-bonds (2) and hydrophobic contacts (93) between ligand and receptor after MD simulation. These may explain the predicted binding affinity of the ligand. PK11195 and DPA714 scored the same number of H-bonds (1) with A147T TSPO, and the former had 65 contact points, while the latter had 95 contact points. MKD complex appears to form no ligand-protein H-bonds, but it showed the largest number of ligand-protein contact points (107) after MD simulation. The amide oxygen atoms of PK11195, DPA714 and MKL and the N1 atom of MKL are H-bond acceptors. Three residues (LEU49, ASN151 and TRP53) of TSPO were H-bond donors, as showed in Figure 70. Combined with the data shown

in Table 9, all four ligand-TSPO complexes have established more H-bonds and contacts than the primary complexes after MD simulation. Furthermore, all four ligands changed the structure and shifted position compared to the initial ligand binding modes, although all remained in the binding pocket. Hence, MD simulations adjusted the ligand positions and structures to make tighter connections to the proteins.

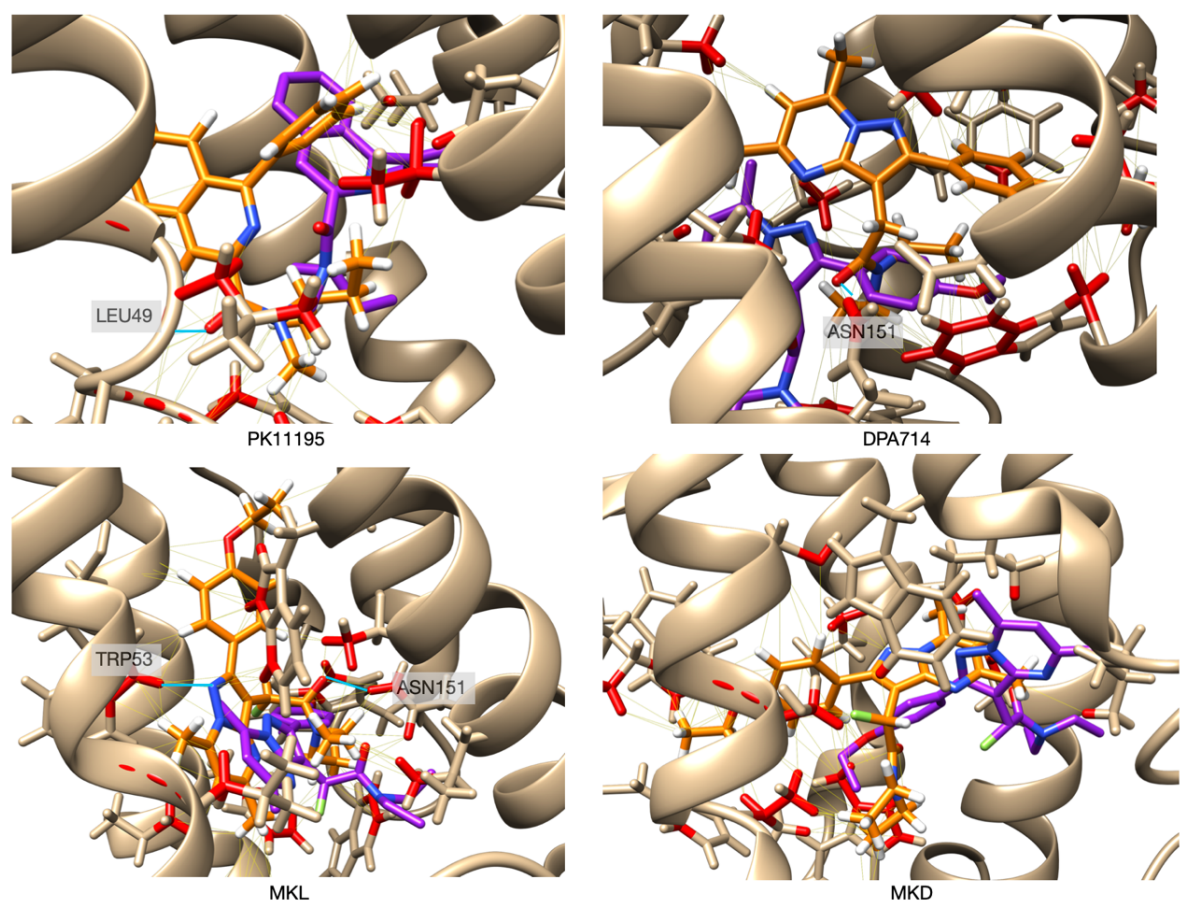


Figure 70: H-bond formation and contacts between ligands and A147T TSPO detected after MD simulations. The four ligands are coloured orange. The initial binding ligands are coloured purple. The blue line represents the H-bond, and the yellow line represents the contacts between ligands and TSPO.

Table 9: H-bond and contacts between ligands and A147T TSPO

| | H-bond | H-bond donor (TSPO) | Contact with TSPO |
|-----------------|--------|---------------------|-------------------|
| Initial PK11195 | 0 | | 72 |
| (MD) PK11195 | 1 | LEU49 - O8 | 65 |

| | | | |
|-----------------------|---|----------------------|-----|
| Initial DPA714 | 0 | | 84 |
| (MD) DPA714 | 1 | ASN151-O22 | 95 |
| Initial MKL | 0 | | 91 |
| (MD) MKL | 2 | TRP53-N1, ASN151-O22 | 93 |
| Initial MKD | 0 | | 84 |
| (MD) MKD | 0 | | 107 |

To validate the protein-ligand binding modes after MD simulation, the ligands binding affinity with A147T TSPO was re-calculated binding scores via SeeSAR, and the results are shown in Table 10. MKL (affinity < pM) had the strongest binding affinity to the A147T TSPO receptor at an initial state, followed by DPA714 and MKD ligands, and finally PK11195 (affinity: high μ M to low mM).

After MD simulation, the binding of the ligands to TSPO was recalculated. DPA still showed a high affinity and became the ligand with the strongest affinity (affinity: high pM to low nM), followed by MKL and MKD, which both showed a weakened affinity with similar results (affinity: high nM to low μ M), the last still being PKA, but with a stronger affinity than its initial state (affinity: low μ M). All these four ligands still showed a high binding affinity with A147 TSPO.

Table 10: Binding affinity of A147T TSPO and ligand calculated via SeeSAR

| Ligands | Initial TSPO Affinity (K_i calc) | MD TSPO Affinity (K_i calc) |
|----------------|---|--|
| PK11195 | High μ M to low mM | Low μ M |

| | | |
|--------|-------------------|------------------------|
| DPA714 | High pM to low nM | High pM to low nM |
| MKL | < pM | High nM to low μ M |
| MKD | Middle nM | High nM to low μ M |

6.2 Discussion

The clinical applicability of current TSPO PET radioligands has been hampered by the lack of druglike binding affinity of many second-generation compounds, including PBR-28, DPA-714, and DAA1106, to a prevalent form of polymorphic TSPO (A147T) compared to wild type human TSPO. The affinity decrease for some ligands can be dramatic (e.g., drop of K_i from 3 nM to 237 nM for PBR-28). For DPA-714, K_i drops from 15 nM for wtTSPO to 66 nM for A147T variant. This makes the scaffold preferred for further development. However, the exact binding mode for this scaffold remains unknown, which disables structure-guided optimisation.

To understand the interactions between PK11195 and A147T TSPO polymorph, and to predict the binding mode of chemically distinct DPA-714, a homology model of human TSPO was built, based on murine TSPO bound to PK11195, and subsequently ran all-atom MD simulations of the apo and holo TSPO embedded in a model membrane. All configurations sampled equilibrated ensembles in a short period.

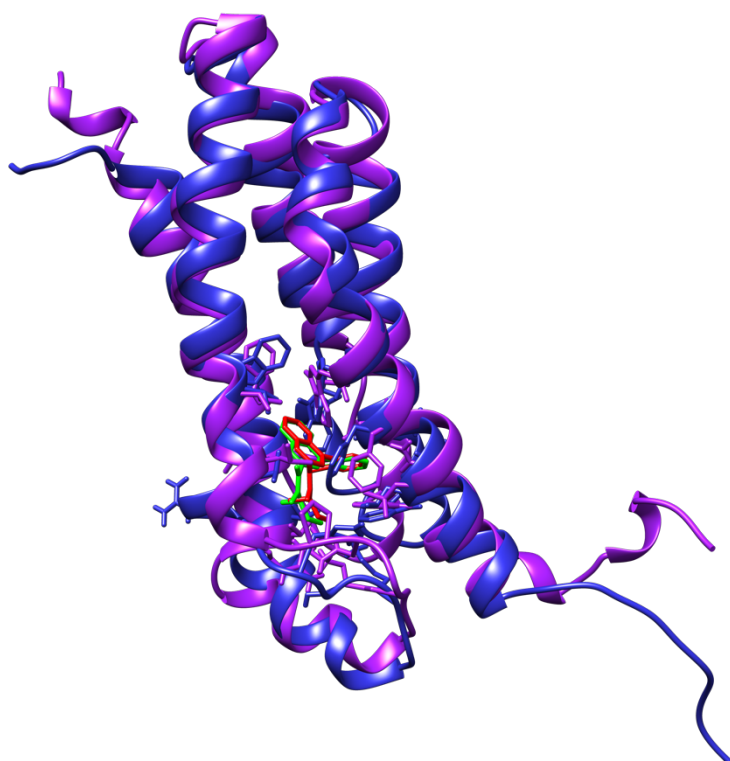


Figure 71: The conformation of 2MGY (blue) and 2N02 (purple). The green PK11195 ligand belongs to 2MGY, and another red PK11195 ligand belongs to 2N02

Through extensive molecular docking calculations performed on the ensemble calculated from the MD simulations, it was found that the binding pocket for DPA-714 and its derivatives is formed by residues CYS19, SER23, HIS43, HIS46, LEU49, GLY50, TRP107, VAL110, LEU114, TRP143, LEU150, ASN151, and that binding site is distinct from the binding site for PK11195 inferred from experimental studies on mouse TSPO. All four compounds investigated in this study showed favourable binding affinity to TSPO (nM range), and experimental binding affinity known for PK11195 and DPA-714 were reproduced well, increasing confidence in calculated binding modes. One enantiomer of designed molecules, denoted as MKL, had better affinity than another enantiomer (MKD), and the difference span two orders of magnitude. Introducing the F atom in the C10 position of MKL affected the binding site's conformation and the ligand's orientation in the TSPO binding pocket and caused the formation of an intermolecular H-bond, leading to an increase in the calculated binding affinity. By performing equilibrium MD simulations of complexes embedded in the explicit membrane, it was shown that two enantiomers bound to TSPO caused different conformational changes at the putative

cholesterol binding site. The region about residues 60-67 is implicated in cholesterol binding. Termini are the most flexible regions in all simulations, which is expected, considering they are unconstrained and solvent-exposed. Overall, ligand-induced changes in flexibility occur at sites distal from the ligand binding site, and those changes, different for agonists and antagonists (PK11195), might contribute to the specific biological activity of ligands evaluated in this work. MKL also showed stabilizing effect on TSPO structural core. This region is implicated in ligand efflux, so an increase in flexibility in this region of TSPO by PK11195 might contribute to the antagonistic activity of this ligand.

While DPA-714 represents an improvement in relatively many of the second-generation ligands regarding its binding to the A147T TSPO variant, it still shows reduced binding affinity to that variant. To overcome this problem and limitations in the synthetic feasibility, it is crucial to understand how the binding requirements of A147T TSPO differ from wtTSPO. One approach to deconvolute factors responsible for the loss of affinity of TSPO ligands to A147T TSPO variant is evaluating the differences in structure and dynamics, DPA-714 and their analogues bound to wtTSPO and A147T TSPO variant, using structure-based molecular modelling approaches.

The solution NMR structures of the murine TSPO (mTSPO), which has 82% sequence identity to human TSPO, do not show any dramatic structural differences between wild type mTSPO and the A147T variant for the structures bound to PK11195 (PDB codes 2MGY and 2N02, respectively). The backbone RMSD is 1.14 Å, and the binding poses of the ligand in both structures are virtually identical (Figure 71). Therefore, it is unlikely that different protein conformations can explain the reduced affinity of 2nd generation TSPO ligands to the A147T variant. Thus, different binding modes and this explanation may require a detailed examination of subtle structural changes at the binding site and changes in intrinsic protein dynamics caused by the binding event. Analysis of those factors is hampered by the absence of the experimental structure of any 2nd generation ligand bound to TSPO published to date.

6.3 Materials and Methods

6.3.1 Homology modelling of human TSPO

Homology models of human TSPO have been created using SWISS-MODEL, a fully automated protein structure homology-modelling webserver^{203, 204}. The primary sequences of the human wild-type (WT) TSPO have been obtained from the UniProt repository (entry P30536). Fifteen models were calculated and ranked according to their sequence similarity and QMEAN²⁰³ quality scores combined. After the visual inspection, the model of WT TSPO based on the solution NMR structure of murine TSPO (UniProt entry P50637) in complex with its high-affinity antagonist PK11195 (PDB code: 2MGY). For the A147T model, the solution NMR structure of murine TSPO bearing A147T mutation and bound to PK11195 (PDB code: 2N02) was chosen as a template. These have been done for several reasons. First, overall sequence conservation between human and mouse TSPO is very high (81% of identical residues), which gives high confidence in the model quality. The template 2N02 represents the experimental structure of the A147T TSPO variant, which reduces uncertainty in the modelling workflow. Finally, both templates contain a high-affinity ligand-bound. Therefore, predicting a putative binding site, a source of yet another uncertainty, is not required. Both models were quality checked by UCSF Chimera²⁵¹, for any missing loops and steric clashes, the PK11195 molecule has been fitted back in, and the resulting complexes have been subjected to 2,500 cycles steepest-descent molecular-mechanical energy minimisations within UCSF Chimera, using Amber99SB force field. Partial atomic charges on PK11195 were calculated using the AM1-BCC quantum mechanical method²⁶⁷.

6.3.2 Molecular docking and structure-guided ligand design

The model of each PK11195-TSPO complex (wtTSPO and A147T) have been imported into SeeSAR (version 10.3), a structure-guided drug design platform, which facilitates the optimisation of molecules from virtual screening to lead-optimisation and scaffold-hopping (www.biosolveit.de). The binding site has been defined using (1) the existing ligand (PK11195) and (2) unoccupied spaces. To assess the suitability of the workflow, PK11195 has been re-

docked and its poses and predicted binding affinity has been calculated and compared with experimental data. In molecular docking, 300 poses have been selected with a high clash tolerance filter.

3D coordinates of the DPA-714 have been constructed in UCSF Chimera, imported to SeeSAR, and both molecules have been subjected to molecular docking (300 poses with high clash tolerance filter) using the binding site defined for PK11195 and HYDE scoring function. The best pose, with the calculated binding affinity matching the experimental value, has been selected in lead optimisation, wherein different parts of the scaffold have been fluorinated, with the input and guidance from the synthetic chemistry team. In total, 50 analogues were inspected. Two compounds denoted as MKL and MKD (Figure 64), with the best-predicted affinity and favourable pharmacokinetic parameters (logP, MW, TPSA), were selected for further studies. The ADME properties of ligands were predicted using the SwissADME (www.swissadme.ch)²⁷⁷.

6.3.3 Atomistic molecular dynamics (MD) simulations

All simulations were performed using Gromacs 2016.3²⁵⁶. The A147T TSPO variant was parametrised using the AMBER99SB-ILDN force field, with the DPPC lipid bilayer and TIP3P water model²¹⁸. The apoTSPO model and four ligand-TSPO complexes were embedded in a DPPC bilayer lipid molecule, using the computational membrane builder tool in the CHARMM-GUI server (www.charmm-gui.org)²⁵⁷⁻²⁵⁹. Box distance was set to 1 nm, and periodic boundary conditions were applied. The box was solvated and Na⁺ and Cl⁻ ions were added to achieve a 0.1 M concentration and maintain the charge neutrality of the unit. The solvated receptor-membrane systems were energy minimised and equilibrated. The minimisation ran using the steepest descent followed by the conjugate gradient. The energy step size was set to 0.01 nm, and the maximum number of steps was set to 50,000. The minimisation was stopped when the maximum force fell below 1000 kJ/mol/nm using the Verlet cutoff scheme. Treatment of long-range electrostatic interactions was set to Particle Mesh-Ewald (PME)²⁶⁰, and the short-range electrostatic and van der Waals cutoff was set to 1.0 nm. After the energy minimisation, heating to 300 K was performed for 5 ps with a time step of 1 fs and

position restraints applied to the backbone in an NVT ensemble. The constraint algorithm used was LINCS ²⁶¹, which was applied to all bonds and angles in the protein. The cutoff for non-bonded short-range interaction was set to 2.0 nm with the Verlet cutoff scheme. Long-range electrostatics were set to PME ²⁶⁰. The temperature coupling was set between the protein and the non-protein entities using a Berendsen thermostat, with a time constant of 0.1 ps and the temperature set to reach 300 K with the pressure coupling off. Pressure equilibration was run at 300 K with a Parrinello-Rahman barostat and set to 1 bar ²²⁵ in an NPT ensemble. The equilibration trajectories were set to 10 ns, and the production MD simulations were performed for 100 ns.

The trajectories were analysed using GROMACS tools, including root-mean-square deviation (RMSD) to assess overall stability, per-residue root-mean-square fluctuation (RMSF) to assess the local flexibility H-bonding network analysis and contact points to assess the interaction between protein and ligand were done via UCSF Chimera.

Conclusions and future directions

The number of people suffering from neuroinflammation and neurodegenerative diseases has increased significantly in current society. Understanding the pathogenesis of these diseases and providing effective treatments and defences has become a crucial task for medical research for clinical treatment. As discussed in this work and the literature, duplicated $\alpha 7$ nAChR and A147T TSPO variants have been confirmed to negatively affecting cellular receptors' function. These reports correlate to the increased expression in microglia upon inflammation, respectively, associated with neurological disorders such as AD, PD, and schizophrenia. Hence, this makes these two proteins important targets and biomarkers for new therapies and diagnostics.

The research shown in this work aimed to identify the conformation, estimate protein-ligand or protein-protein interactions, and detect the potential binding sites of dup $\alpha 7$ nAChR. This work performed a systematic study on all possible combinations of dup $\alpha 7$ / $\alpha 7$ nicotinic receptors, focusing on their structural stability and stoichiometry, to find the most probable functional pentamers bearing dup $\alpha 7$ subunits. Understanding this was essential to understand the effect of dup $\alpha 7$ since it has been regarded as a dominant-negative regulator of $\alpha 7$ receptors. However, reports of functional pentamers bearing dup $\alpha 7$ subunits have been published. All possible combinations were modelled and evaluated using structure-based multiscale computational methods to address the conflicting evidence from published studies. We showed that higher content of dup $\alpha 7$ subunits resulted in less cohesive pentamers, and dup $\alpha 7$ /dup $\alpha 7$ interfaces, corresponding to the orthosteric binding sites, were markedly less stable than dup $\alpha 7$ / $\alpha 7$ interfaces. These indicate that the most likely combinations were pentamers bearing one dup $\alpha 7$ subunit (A-Dup $\alpha 7$ model) or pentamers containing two non-consecutive subunits (AC-Dup $\alpha 7$ model). Pentamers bearing three subunits with the lowest dup $\alpha 7$ /dup $\alpha 7$ interfaces (ACD-Dup $\alpha 7$ model) were also suggested to be functional via analysis of the ion transmission energetic landscape carried out via umbrella sampling simulations. The comparative studies of the energetic landscapes for the pentamers with different stoichiometries showed that receptors with a low ratio of dup $\alpha 7$ / $\alpha 7$ are still functional, even though higher energy barriers are observed for these pentamers. On the other hand, the increase of the number of dup $\alpha 7$ subunits

negatively affected the Ca^{2+} uptake via the receptor.

Our work has also shown that dup $\alpha 7$ interfaces are insensitive to α -bungarotoxin (α -BTX), but not to $\text{A}\beta_{42}$, even though their $\text{A}\beta_{42}$ binding is impaired compared to the canonical $\alpha 7$ receptors. This impairment indicates that receptors containing dup $\alpha 7$ subunits are less sensitive to $\text{A}\beta_{42}$ effects and that dup $\alpha 7$ subunits, despite their impaired agonist binding, may offer protection against detrimental $\text{A}\beta_{42}$ effects.

After analysing stoichiometry, molecular dynamics, conformational changes, and binding affinity, the focus was on finding the 'druggable' allosteric binding sites and evaluating the effect of dup $\alpha 7$ subunit on both agonists allosteric binding sites. Two agonists, Ach and Lob, were selected to dock with these five pentamers using DOCK6. The results showed that agonist binding sites mapped in silico were located at the interface between two adjacent $\alpha 7$ subunits, behind the loop C and surrounding by loop A, Loop B of left-hand subunit and $\beta 1$, $\beta 2$, $\beta 5$ and $\beta 8$ of the right-hand subunit. The dup $\alpha 7$ subunit has not affected the agonist binding sites in dup-WT dimer interfaces but influences the agonist binding site generation in WT-dup and dup-dup interfaces. The same agonist binding site results were also detected via FTMap.

In addition, the allosteric binding sites were predicted using FTMap and identified by docking with an $\alpha 7$ nAChR allosteric ligand library using DOCK6. Four binding sites have been found in $\alpha 7$ nAChR. The vestibule pocket was inside a single chain, while the top pocket, the upper pocket and the down pocket were in the middle of the dimer. The top pocket involved residues from the $\alpha 1$ loop and $\beta 4$ loop of one subunit. The vestibule pocket is formed by $\beta 2$, $\beta 4$, $\beta 5$ and $\beta 6$ loops from the same subunit. The top pocket involved residues from the $\alpha 1$ loop and $\beta 4$ loop of one subunit and may residues in $\alpha 1$, $\beta 2$ and $\beta 3$ loop of the adjacent subunit. The last two allosteric binding sites were found in the middle of the adjacent subunits: one in the upper pocket above the orthosteric binding site, involving loops B and C in one subunit and loops $\beta 1$, $\beta 2$ in the adjacent subunit; the other one below the orthosteric site, and it involved residues in loops A, $\beta 6$ and $\beta 7$ in one subunit and residues in loops $\beta 1$, $\beta 8$ in the adjacent subunit. These four allosteric binding sites have also been found in dup $\alpha 7$ models, but the dup-dup subunit does not exist either agonist or allosteric binding sites, which have been proved by the FTMap

and DOCK6 results of AB-Dup and ACD-dup.

I hope that this work will contribute to the elucidation of the biological roles of dup α 7 subunits, generating models that can be used for a rational drug design. Future research aiming to characterise the function of dup α 7 in the clinical context may result in novel pathways for AD treatment based on early-stage preclinical data. As α 7 receptors are implicated in a broad range of diseases, including cognition, memory, schizophrenia, chronic pain and inflammaging, mechanistic insights into receptors containing dup α 7 subunits will impact these therapeutic areas, including those conditions which currently represent an unmet clinical need.

In sequence, the focus was on shedding light on interactions between ligands and A147T TSPO polymorphs to predict the binding mode of chemically distinct DPA-714. For this, it was first built a homology model of the human A147T TSPO variant, based on murine TSPO bound to PK11195, and subsequently ran all-atom molecular dynamics (MD) simulations of the TSPO embedded in a model membrane. I found that the presence of PK11195 stabilised TSPO overall structure, even in the absence of any dramatic conformational changes. Interestingly, I also found that PK11195 binding reduced the intrinsic dynamics of the N-terminal segment of TSPO. These regions are implicated in cholesterol binding in the inner membrane leaflet. The impact of PK11195 on intrinsic dynamics of TSPO was less pronounced for the C-terminal segment located in the outer membrane leaflet, wherein the ligand binding site is located, compared to other regions and systems.

In this work, two different enantiomers (MKL, MKD) based on the scaffold of DPA-714 were designed and synthesised with the fluorine (F) position changed from C29 to C10 atom. These two compounds have better synthetic feasibility than their parent compound (DPA-714). The target A147T TSPO variant was mapped for binding sites (the predicted pose of DPA-714) and used to investigate the difference between enantiomers and two established TSPO radioligand (PA11195 and DPA-714) via SeeSAR. The MKL ligand was showed as the best compound in terms of affinity (affinity (K_i calc) < pM), had better affinity than another enantiomer (MKD) with calculated affinity (K_i calc) of mid-nM range. Introducing the fluorine atom in the C10 position of MKL affected the binding site's conformation and the ligand's orientation in the

TSPO binding pocket and caused the formation of an intermolecular H-bond, leading to an increase in the calculated binding affinity.

By performing equilibrium MD simulations of complexes embedded in an explicit membrane, it was found that two enantiomers bound to TSPO caused different conformational changes at the putative cholesterol binding site. MKL also showed a more stabilising effect on TSPO structural core. Most pharmacokinetic properties of both enantiomers were otherwise the same and highly similar to properties of the parent compound, DPA-714. These results strongly suggest that residues TRP53 and ASN151 of TSPO play an essential role in stabilising the complex via forming H-bonds with the ligands. Besides these, residues VAL26, LEU49, TRP107, VAL110, LEU150 also as necessary part involved the binding pocket-forming and hydrophobic contacts with ligands. Hence, the drug-likeness of MKL and MKD was predicted: both have suitable lipophilicity and can cross the blood-brain barrier (BBB), which is crucial for central nervous system drugs and radioligands.

As the synthesis of both isomers (bearing non-radioactive fluorine) has now been completed in the group of Dr Mike Carroll and testing in cellular assays are currently in progress (group of Dr Kate Madden), future works should evaluate the binding affinity of MKL and MKD in cell-free assays (e.g., MST or ITC) to validate molecular modelling results. Should those studies confirm MKL (and possibly also MKD) as a potent ligand of A147T TSPO, the following steps should focus on improved radiosynthesis and proof-of-concept tests in human brain tissue.

Appendix

Supporting Information

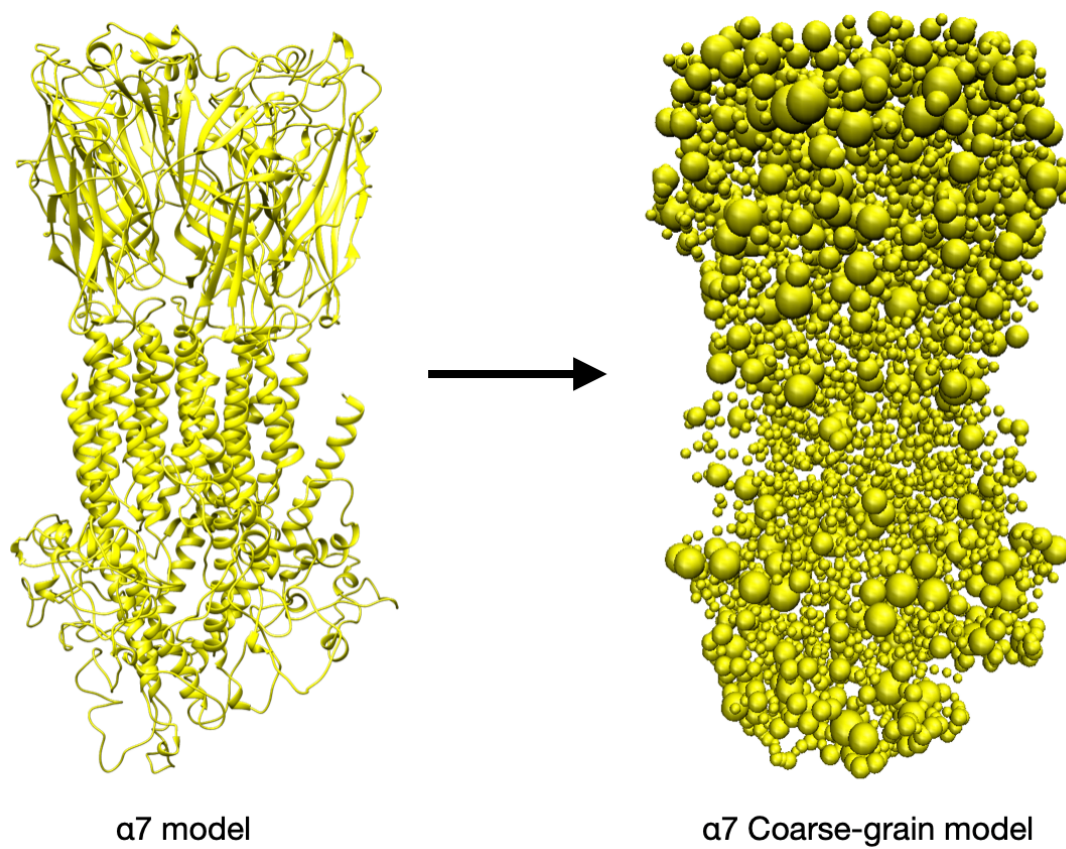


Figure S. 1: The atomistic $\alpha 7$ model translation to MARTINI coarse-grain framework.

Table S. 1: Full list of interfacial hydrogen bond contacts for the canonical $\alpha 7$ receptor (WT) before MD simulations.

| Donor | | Acceptor | | Donor-Acceptor distance (Å) | Donor hydrogen-Acceptor distance (Å) | Total number of contacts |
|------------|------------|----------|-----------|-----------------------------|--------------------------------------|--------------------------|
| AB (WT-WT) | 10 H-bonds | | | | | 171 |
| GLN | 61.A NE2 | SER | 149.B OG | 2.868 | 1.887 | |
| ASN | 75.A ND2 | ARG | 121.B O | 2.983 | 2.166 | |
| ARG | 101.A NE | TYR | 173.B OH | 2.837 | 1.823 | |
| ARG | 101.A NH1 | TYR | 173.B OH | 3.388 | 2.633 | |
| ARG | 101.A NH2 | PRO | 39.B O | 3.345 | 2.583 | |
| ARG | 101.A NH2 | LEU | 40.B O | 3.013 | 2.183 | |
| ASN | 69.B ND2 | MET | 63.A O | 2.815 | 1.808 | |
| ASN | 69.B ND2 | ASP | 64.A OD1 | 2.834 | 1.909 | |
| GLN | 70.B NE2 | PRO | 192.A O | 3.623 | 2.651 | |
| TRP | 171.B NE1 | LEU | 141.A O | 2.842 | 1.936 | |
| BC (WT-WT) | 5 H-bonds | | | | | 219 |
| ASN | 75.B ND2 | GLU | 120.C OE1 | 2.685 | 1.809 | |
| ARG | 101.B NH1 | SER | 172.C O | 3.267 | 2.429 | |
| ASN | 133.B ND2 | TYR | 173.C OH | 3.189 | 2.322 | |
| TRP | 171.C NE1 | LEU | 141.B O | 2.777 | 1.77 | |
| SER | 172.C OG | ASN | 133.B O | 2.843 | 1.908 | |
| CD (WT-WT) | 5 H-bonds | | | | | 164 |
| LYS | 28.C N | GLU | 41.D OE2 | 3.458 | 2.628 | |
| TRP | 77.C NE1 | ASP | 119.D O | 2.975 | 2.063 | |
| GLN | 70.D NE2 | PRO | 192.C O | 3.183 | 2.218 | |
| SER | 149.D OG | GLN | 61.C OE1 | 2.823 | 1.911 | |

| | | | | | | |
|-------------|------------|-----|-----------|-------|-------|-----|
| TRP | 171.D NE1 | LEU | 141.C O | 2.797 | 1.795 | |
| DE (WT -WT) | 10 H-bonds | | | | | 123 |
| GLN | 61.D NE2 | SER | 149.E OG | 2.955 | 2.019 | |
| TRP | 77.D NE1 | GLU | 120.E OE1 | 2.748 | 1.864 | |
| TRP | 77.D NE1 | GLU | 120.E OE2 | 2.887 | 1.964 | |
| ARG | 101.D NH1 | LYS | 109.E O | 3.12 | 2.329 | |
| ARG | 101.D NH2 | ASP | 111.E OD1 | 2.66 | 1.749 | |
| ASN | 129.D ND2 | SER | 172.E OG | 2.876 | 1.855 | |
| GLN | 70.E NE2 | PRO | 192.D O | 2.982 | 1.977 | |
| SER | 117.E OG | GLY | 189.D O | 2.819 | 2.229 | |
| ARG | 121.E NH1 | PRO | 143.D O | 2.818 | 1.995 | |
| TRP | 171.E NE1 | LEU | 141.D O | 2.808 | 1.806 | |
| EA (WT-WT) | 13 H-bonds | | | | | 185 |
| ASN | 69.A ND2 | MET | 63.E O | 2.867 | 1.858 | |
| GLN | 70.A NE2 | PRO | 192.E O | 3.007 | 2.022 | |
| ARG | 121.A NE | ASN | 75.E OD1 | 2.826 | 2.012 | |
| ARG | 121.A NH1 | ASN | 75.E OD1 | 2.855 | 2.005 | |
| TRP | 171.A NE1 | LEU | 141.E O | 2.78 | 1.844 | |
| SER | 172.A OG | ASN | 129.E OD1 | 2.833 | 1.943 | |
| GLN | 61.E NE2 | SER | 149.A OG | 2.974 | 2.171 | |
| TRP | 77.E NE1 | ASP | 119.A O | 2.684 | 1.953 | |
| ARG | 101.E NE | TYR | 173.A OH | 2.872 | 1.87 | |
| ARG | 101.E NH1 | TYR | 173.A OH | 3.428 | 2.629 | |
| ARG | 101.E NH2 | PRO | 39.A O | 3.361 | 2.439 | |
| ARG | 101.E NH2 | LEU | 40.A O | 2.843 | 2.132 | |

| | | | | | | |
|-----|-----------|-----|-----------|-------|-------|--|
| ASN | 129.E ND2 | ASP | 111.A OD2 | 2.675 | 1.806 | |
|-----|-----------|-----|-----------|-------|-------|--|

Table S. 2: Full list of interfacial hydrogen bond contacts in A-dupa7 receptor before MD simulations.

| Donor | | Acceptor | | Donor-Acceptor distance (Å) | Donor hydrogen-Acceptor distance (Å) | Total number of contacts |
|--------------|-----------|----------|-----------|-----------------------------|--------------------------------------|--------------------------|
| AB (dup- WT) | 3 H-bonds | | | | | 132 |
| GLN | 70.B NE2 | PRO | 102.A O | 3.452 | 2.486 | |
| ARG | 121.B NH1 | PHE | 32.A O | 3.467 | 2.54 | |
| TRP | 171.B NE1 | LEU | 51.A O | 2.964 | 2.141 | |
| BC (WT-WT) | 4 H-bonds | | | | | 214 |
| ASN | 75.B ND2 | GLU | 120.C OE1 | 2.714 | 1.989 | |
| ARG | 101.B NH1 | SER | 172.C O | 3.492 | 2.74 | |
| ASN | 133.B ND2 | TYR | 173.C OH | 3.115 | 2.243 | |
| TRP | 171.C NE1 | LEU | 141.B O | 2.71 | 1.699 | |
| CD (WT-WT) | 3 H-bonds | | | | | 158 |
| TRP | 77.C NE1 | ASP | 119.D O | 2.909 | 2.017 | |
| GLN | 70.D NE2 | PRO | 192.C O | 3.233 | 2.255 | |
| TRP | 171.D NE1 | LEU | 141.C O | 2.787 | 1.78 | |
| DE (WT-WT) | 8 H-bonds | | | | | 131 |
| GLN | 61.D NE2 | SER | 149.E OG | 2.97 | 2.111 | |
| TRP | 77.D NE1 | GLU | 120.E OE1 | 2.675 | 1.821 | |
| TRP | 77.D NE1 | GLU | 120.E OE2 | 3.006 | 2.199 | |
| ARG | 101.D NH1 | LYS | 109.E O | 3.184 | 2.453 | |
| ARG | 101.D NH2 | ASP | 111.E OD1 | 2.62 | 1.879 | |
| ASN | 129.D ND2 | SER | 172.E OG | 2.9 | 1.886 | |
| GLN | 70.E NE2 | PRO | 192.D O | 2.984 | 1.974 | |
| TRP | 171.E NE1 | LEU | 141.D O | 2.757 | 1.751 | |

| | | | | | | |
|--------------|-----------|-----|-----------|-------|-------|----|
| EA (WT -dup) | 3 H-bonds | | | | | 73 |
| SER | 82.A OG | ASN | 129.E OD1 | 3.027 | 2.276 | |
| GLN | 61.E NE2 | SER | 59.A OG | 2.857 | 2.155 | |
| ARG | 101.E NE | TYR | 83.A OH | 2.509 | 1.595 | |

Table S. 3: Full list of interfacial hydrogen bond contacts in AB-dupa7 receptor before MD simulations.

| Donor | | Acceptor | | Donor-Acceptor distance (Å) | Donor hydrogen-Acceptor distance (Å) | Total number of contacts |
|--------------|------------|----------|-----------|-----------------------------|--------------------------------------|--------------------------|
| AB (dup-dup) | 3 H-bonds | | | | | 78 |
| ARG | 31.A NH1 | ASP | 29.B OD1 | 3.079 | 2.374 | |
| TRP | 81.B NE1 | LEU | 51.A O | 2.806 | 1.972 | |
| TYR | 83.B OH | ASN | 39.A OD1 | 2.717 | 1.816 | |
| BC (dup- WT) | 3 H-bonds | | | | | 134 |
| TRP | 108.C NE1 | GLN | 13.B OE1 | 2.979 | 2.092 | |
| SER | 117.C OG | GLY | 99.B O | 2.766 | 2.064 | |
| TRP | 171.C NE1 | LEU | 51.B O | 2.77 | 1.797 | |
| CD (WT-WT) | 5 H-bonds | | | | | 166 |
| LYS | 28.C N | GLU | 41.D OE2 | 3.474 | 2.652 | |
| TRP | 77.C NE1 | ASP | 119.D O | 2.976 | 2.066 | |
| GLN | 70.D NE2 | PRO | 192.C O | 3.166 | 2.201 | |
| SER | 149.D OG | GLN | 61.C OE1 | 2.818 | 1.893 | |
| TRP | 171.D NE1 | LEU | 141.C O | 2.794 | 1.792 | |
| DE (WT-WT) | 10 H-bonds | | | | | 122 |
| GLN | 61.D NE2 | SER | 149.E OG | 2.95 | 2.013 | |
| TRP | 77.D NE1 | GLU | 120.E OE1 | 2.756 | 1.878 | |
| TRP | 77.D NE1 | GLU | 120.E OE2 | 2.872 | 1.946 | |
| ARG | 101.D NH1 | LYS | 109.E O | 3.126 | 2.336 | |
| ARG | 101.D NH2 | ASP | 111.E OD1 | 2.665 | 1.754 | |
| ASN | 129.D ND2 | SER | 172.E OG | 2.882 | 1.861 | |
| GLN | 70.E NE2 | PRO | 192.D O | 2.981 | 1.977 | |

| | | | | | | |
|-------------|-----------|-----|-----------|-------|-------|----|
| SER | 117.E OG | GLY | 189.D O | 2.813 | 2.213 | |
| ARG | 121.E NH1 | PRO | 143.D O | 2.808 | 1.97 | |
| TRP | 171.E NE1 | LEU | 141.D O | 2.803 | 1.801 | |
| EA (WT -WT) | 3 H-bonds | | | | | 68 |
| TRP | 81.A NE1 | LEU | 141.E O | 3.308 | 2.435 | |
| SER | 82.A OG | ASN | 129.E OD1 | 3.017 | 2.103 | |
| ARG | 101.E NE | TYR | 83.A OH | 2.798 | 1.852 | |

Table S. 4: Full list of interfacial hydrogen bond contacts in AC-dupa7 receptor before MD simulations.

| Donor | | Acceptor | | Donor-Acceptor distance (Å) | Donor hydrogen-Acceptor distance (Å) | Total number of contacts |
|--------------|-----------|----------|-----------|-----------------------------|--------------------------------------|--------------------------|
| AB (dup- WT) | 3 H-bonds | | | | | 123 |
| GLN | 70.B NE2 | PRO | 102.A O | 3.454 | 2.485 | |
| ARG | 121.B NH1 | PHE | 32.A O | 3.199 | 2.306 | |
| TRP | 171.B NE1 | LEU | 51.A O | 2.786 | 1.897 | |
| BC (WT -dup) | 6 H-bonds | | | | | 219 |
| ASN | 75.B ND2 | GLU | 120.C OE1 | 2.686 | 1.819 | |
| ARG | 101.B NH1 | SER | 172.C O | 3.267 | 2.418 | |
| ASN | 133.B ND2 | TYR | 173.C OH | 3.196 | 2.335 | |
| TRP | 171.C NE1 | LEU | 141.B O | 2.77 | 1.764 | |
| SER | 172.C OG | LEU | 131.B O | 3.516 | 3.001 | |
| SER | 172.C OG | ASN | 133.B O | 2.884 | 2.01 | |
| CD (dup- WT) | 3 H-bonds | | | | | 70 |
| ARG | 101.C NH1 | SER | 82.D O | 2.862 | 2.092 | |
| SER | 59.D OG | GLN | 61.C OE1 | 3.189 | 2.417 | |
| TRP | 81.D NE1 | LEU | 141.C O | 2.683 | 1.698 | |
| DE (WT-WT) | 4 H-bonds | | | | | 177 |
| GLN | 70.E NE2 | PRO | 102.D O | 2.883 | 1.878 | |
| TRP | 108.E NE1 | GLN | 13.D OE1 | 3.107 | 2.124 | |
| SER | 117.E OG | GLY | 99.D O | 2.858 | 2.303 | |
| TRP | 171.E NE1 | LEU | 51.D O | 2.815 | 1.807 | |
| EA (WT -dup) | 3 H-bonds | | | | | 62 |
| TRP | 81.A NE1 | LEU | 141.E O | 3.267 | 2.39 | |

| | | | | | | |
|-----|----------|-----|-----------|-------|-------|--|
| SER | 82.A OG | ASN | 129.E OD1 | 3.007 | 2.087 | |
| ARG | 101.E NE | TYR | 83.A OH | 2.802 | 1.854 | |

Table S. 5: Full list of interfacial hydrogen bond contacts in ABC-dupa7 receptor before MD simulations.

| Donor | | Acceptor | | Donor-Acceptor distance (Å) | Donor hydrogen-Acceptor distance (Å) | Total number of contacts |
|--------------|------------|----------|-----------|-----------------------------|--------------------------------------|--------------------------|
| AB (dup-dup) | 3 H-bonds | | | | | 78 |
| ARG | 31.A NH1 | ASP | 29.B OD1 | 3.021 | 2.264 | |
| TRP | 81.B NE1 | LEU | 51.A O | 2.802 | 1.969 | |
| TYR | 83.B OH | ASN | 39.A OD1 | 2.711 | 1.803 | |
| BC (dup-dup) | 1 H-bonds | | | | | 86 |
| TRP | 81.C NE1 | LEU | 51.B O | 2.808 | 2.019 | |
| CD (dup- WT) | 4 H-bonds | | | | | 158 |
| GLN | 70.D NE2 | PRO | 102.C O | 3.244 | 2.276 | |
| TRP | 108.D NE1 | GLN | 13.C OE1 | 2.886 | 2.019 | |
| TRP | 171.D NE1 | LEU | 51.C O | 2.83 | 1.819 | |
| TYR | 173.D OH | GLN | 13.C OE1 | 2.956 | 2.251 | |
| DE (WT-WT) | 10 H-bonds | | | | | 120 |
| GLN | 61.D NE2 | SER | 149.E OG | 2.949 | 2.012 | |
| TRP | 77.D NE1 | GLU | 120.E OE1 | 2.763 | 1.887 | |
| TRP | 77.D NE1 | GLU | 120.E OE2 | 2.862 | 1.935 | |
| ARG | 101.D NH1 | LYS | 109.E O | 3.13 | 2.34 | |
| ARG | 101.D NH2 | ASP | 111.E OD1 | 2.667 | 1.756 | |
| ASN | 129.D ND2 | SER | 172.E OG | 2.887 | 1.866 | |
| GLN | 70.E NE2 | PRO | 192.D O | 2.979 | 1.974 | |
| SER | 117.E OG | GLY | 189.D O | 2.81 | 2.207 | |
| ARG | 121.E NH1 | PRO | 143.D O | 2.789 | 1.934 | |
| TRP | 171.E NE1 | LEU | 141.D O | 2.796 | 1.794 | |

| | | | | | | |
|--------------|-----------|-----|-----------|-------|-------|----|
| EA (WT -dup) | 3 H-bonds | | | | | 68 |
| TRP | 81.A NE1 | LEU | 141.E O | 3.289 | 2.415 | |
| SER | 82.A OG | ASN | 129.E OD1 | 3.014 | 2.1 | |
| ARG | 101.E NE | TYR | 83.A OH | 2.799 | 1.852 | |

Table S. 6: Full list of interfacial hydrogen bond contacts in ACD-dupa7 receptor before MD simulations.

| Donor | | Acceptor | | Donor-Acceptor distance (Å) | Donor hydrogen-Acceptor distance (Å) | Total number of contacts |
|--------------|-----------|----------|-----------|-----------------------------|--------------------------------------|--------------------------|
| AB (dup- WT) | 4 H-bonds | | | | | 178 |
| GLN | 70.E NE2 | PRO | 102.D O | 2.882 | 1.876 | |
| TRP | 108.E NE1 | GLN | 13.D OE1 | 3.111 | 2.128 | |
| SER | 117.E OG | GLY | 99.D O | 2.854 | 2.295 | |
| TRP | 171.E NE1 | LEU | 51.D O | 2.814 | 1.804 | |
| BC (WT -dup) | 3 H-bonds | | | | | 67 |
| TRP | 81.A NE1 | LEU | 141.E O | 3.252 | 2.375 | |
| SER | 82.A OG | ASN | 129.E OD1 | 3.007 | 2.089 | |
| ARG | 101.E NE | TYR | 83.A OH | 2.801 | 1.853 | |
| CD (dup-dup) | 3 H-bonds | | | | | 78 |
| ARG | 31.A NH1 | ASP | 29.B OD1 | 3.03 | 2.279 | |
| TRP | 81.B NE1 | LEU | 51.A O | 2.801 | 1.967 | |
| TYR | 83.B OH | ASN | 39.A OD1 | 2.709 | 1.799 | |
| DE (dup- WT) | 3 H-bonds | | | | | 121 |
| TRP | 108.C NE1 | GLN | 13.B OE1 | 2.971 | 2.082 | |
| SER | 117.C OG | GLY | 99.B O | 2.766 | 2.059 | |
| TRP | 171.C NE1 | LEU | 51.B O | 2.77 | 1.797 | |
| EA (WT -dup) | 3 H-bonds | | | | | 70 |
| ARG | 101.C NH1 | SER | 82.D O | 2.861 | 2.088 | |
| SER | 59.D OG | GLN | 61.C OE1 | 3.19 | 2.424 | |
| TRP | 81.D NE1 | LEU | 141.C O | 2.686 | 1.705 | |

Table S. 7: Full list of interfacial hydrogen bond contacts in $\alpha 7$ 4-dupa7 receptor before MD simulations.

| Donor | | Acceptor | | Donor-Acceptor distance (Å) | Donor hydrogen-Acceptor distance (Å) | Total number of contacts |
|--------------|-----------|----------|-----------|-----------------------------|--------------------------------------|--------------------------|
| AB (dup-dup) | 3 H-bonds | | | | | 133 |
| GLN | 70.B NE2 | PRO | 102.A O | 3.448 | 2.479 | |
| ARG | 121.B NH1 | PHE | 32.A O | 3.217 | 2.315 | |
| TRP | 171.B NE1 | LEU | 51.A O | 2.779 | 1.889 | |
| BC (dup-dup) | 5 H-bonds | | | | | 149 |
| ARG | 101.B NH1 | SER | 82.C O | 3.401 | 2.468 | |
| ARG | 101.B NH1 | SER | 87.C OG | 3.029 | 2.241 | |
| THR | 128.B OG1 | ILE | 27.C O | 3.459 | 2.578 | |
| TRP | 81.C NE1 | LEU | 141.B O | 2.742 | 1.846 | |
| TYR | 83.C OH | ASN | 133.B OD1 | 3.135 | 2.166 | |
| CD (dup-dup) | 3 H-bonds | | | | | 76 |
| TRP | 20.D NE1 | GLN | 13.C OE1 | 2.82 | 1.912 | |
| TRP | 81.D NE1 | LEU | 51.C O | 2.734 | 1.743 | |
| SER | 82.D OG | HIS | 37.C ND1 | 3.087 | 2.181 | |
| DE (dup- WT) | 3 H-bonds | | | | | 98 |
| ARG | 31.D NH1 | ASP | 29.E OD1 | 3.129 | 2.355 | |
| TRP | 20.E NE1 | GLN | 13.D OE1 | 3.037 | 2.076 | |
| TRP | 81.E NE1 | LEU | 51.D O | 2.776 | 1.766 | |
| EA (WT -dup) | 4 H-bonds | | | | | 89 |
| TRP | 81.A NE1 | LEU | 51.E O | 2.789 | 1.925 | |
| SER | 82.A OG | ASN | 39.E OD1 | 2.9 | 1.999 | |
| THR | 35.E OG1 | ASP | 29.A OD1 | 3.324 | 2.437 | |

| | | | | | | |
|-----|--------|-----|----------|-------|-------|--|
| PHE | 36.E N | ASP | 29.A OD1 | 3.274 | 2.411 | |
|-----|--------|-----|----------|-------|-------|--|

Table S. 8: Full list of interfacial hydrogen bond contacts in 5-dupa7 receptor before MD simulations.

| Donor | | Acceptor | | Donor-Acceptor distance (Å) | Donor hydrogen-Acceptor distance (Å) | Total number of contacts |
|--------------|-----------|----------|----------|-----------------------------|--------------------------------------|--------------------------|
| AB (dup-dup) | 3 H-bonds | | | | | 77 |
| ARG | 31.A NH1 | ASP | 29.B OD1 | 2.972 | 2.159 | |
| TRP | 81.B NE1 | LEU | 51.A O | 2.793 | 1.956 | |
| TYR | 83.B OH | ASN | 39.A OD1 | 2.699 | 1.779 | |
| BC (dup-dup) | 1 H-bonds | | | | | 86 |
| TRP | 81.C NE1 | LEU | 51.B O | 2.801 | 2.01 | |
| CD (dup-dup) | 3 H-bonds | | | | | 74 |
| TRP | 20.D NE1 | GLN | 13.C OE1 | 2.816 | 1.908 | |
| TRP | 81.D NE1 | LEU | 51.C O | 2.731 | 1.739 | |
| SER | 82.D OG | HIS | 37.C ND1 | 3.079 | 2.173 | |
| DE (dup-dup) | 3 H-bonds | | | | | 99 |
| ARG | 31.D NH1 | ASP | 29.E OD1 | 3.108 | 2.328 | |
| TRP | 20.E NE1 | GLN | 13.D OE1 | 3.035 | 2.075 | |
| TRP | 81.E NE1 | LEU | 51.D O | 2.774 | 1.764 | |
| EA (dup-dup) | 4 H-bonds | | | | | 88 |
| TRP | 81.A NE1 | LEU | 51.E O | 2.786 | 1.92 | |
| SER | 82.A OG | ASN | 39.E OD1 | 2.894 | 1.981 | |
| THR | 35.E OG1 | ASP | 29.A OD1 | 3.324 | 2.403 | |
| PHE | 36.E N | ASP | 29.A OD1 | 3.205 | 2.314 | |

Table S. 9: Full list of interfacial hydrogen bond contacts in the canonical a7 (WT) receptor after MD simulations.

| Donor | | Acceptor | | Donor-Acceptor distance (Å) | Donor hydrogen-Acceptor distance (Å) | Total number of contacts |
|------------|-----------|----------|-----------|-----------------------------|--------------------------------------|--------------------------|
| AB (WT-WT) | 8 H-bonds | | | | | 75 |
| ASN | 75.A ND2 | PHE | 122.B O | 2.862 | 1.948 | |
| ARG | 101.A NE | ASP | 111.B OD1 | 2.821 | 1.85 | |
| ARG | 101.A NE | ASP | 111.B OD2 | 2.844 | 1.85 | |
| ARG | 101.A NH1 | LYS | 109.B O | 2.966 | 2.229 | |
| ARG | 101.A NH1 | ASP | 111.B OD1 | 2.944 | 2.008 | |
| ARG | 101.A NH1 | ASP | 111.B OD2 | 3.1 | 2.191 | |
| ASN | 133.A ND2 | TYR | 173.B OH | 3.329 | 2.587 | |
| ARG | 155.B NH2 | LEU | 231.A O | 3.322 | 2.344 | |
| BC (WT-WT) | 2 H-bonds | | | | | 17 |
| ARG | 101.B NH2 | TRP | 171.C O | 2.617 | 1.85 | |
| THR | 128.B OG1 | GLU | 120.C OE2 | 3.355 | 2.531 | |
| CD (WT-WT) | 1 H-bonds | | | | | 105 |
| ARG | 121.D NH2 | GLN | 61.C OE1 | 3.069 | 2.119 | |
| DE (WT-WT) | 7 H-bonds | | | | | 79 |
| GLN | 61.D NE2 | ASN | 69.E O | 2.728 | 2.042 | |
| GLN | 61.D NE2 | GLN | 70.E O | 3.536 | 2.589 | |
| ASN | 75.D ND2 | GLU | 120.E O | 2.54 | 1.548 | |
| ARG | 101.D NH1 | ASP | 111.E OD1 | 3.162 | 2.199 | |
| ARG | 101.D NH1 | ASP | 111.E OD2 | 3.311 | 2.404 | |
| GLN | 70.E NE2 | ASN | 193.D O | 3.56 | 2.883 | |
| SER | 149.E OG | GLN | 61.D OE1 | 2.83 | 1.92 | |

| | | | | | | |
|------------|-----------|-----|----------|-------|-------|----|
| EA (WT-WT) | 3 H-bonds | | | | | 22 |
| ASN | 46.A ND2 | PRO | 95.E O | 3.212 | 2.35 | |
| ARG | 121.A NE | ASN | 75.E OD1 | 2.833 | 1.884 | |
| ARG | 121.A NH1 | ASN | 75.E OD1 | 2.998 | 2.162 | |

Table S. 10: Full list of interfacial hydrogen bond contacts in A-dupa7 receptor after MD simulations.

| Donor | | Acceptor | | Donor-Acceptor distance (Å) | Donor hydrogen-Acceptor distance (Å) | Total number of contacts |
|--------------|-----------|----------|----------|-----------------------------|--------------------------------------|--------------------------|
| AB (dup- WT) | 7 H-bonds | | | | | 83 |
| GLU | 105.A N | ASN | 69.B OD1 | 3.534 | 2.613 | |
| TRP | 108.B NE1 | GLN | 13.A OE1 | 2.994 | 2.136 | |
| ARG | 121.B NH1 | PHE | 32.A O | 3.17 | 2.472 | |
| ARG | 121.B NH1 | ASP | 33.A OD1 | 2.901 | 2.009 | |
| ARG | 121.B NH2 | ASP | 33.A OD1 | 2.742 | 1.768 | |
| ALA | 124.B N | PHE | 32.A O | 2.877 | 1.908 | |
| SER | 177.B OG | GLN | 11.A OE1 | 3.562 | 2.817 | |
| BC (WT-WT) | 5 H-bonds | | | | | 98 |
| LEU | 29.B N | GLU | 41.C OE1 | 3.177 | 2.334 | |
| LEU | 29.B N | GLU | 41.C OE2 | 3.461 | 2.52 | |
| TYR | 30.B N | GLU | 41.C OE1 | 3.523 | 2.519 | |
| ARG | 101.B NH1 | TRP | 171.C O | 3.519 | 2.664 | |
| SER | 117.C OG | GLY | 189.B O | 3.519 | 2.807 | |
| CD (WT-WT) | 9 H-bonds | | | | | 136 |
| LYS | 28.C N | GLU | 41.D OE1 | 2.748 | 1.742 | |
| LYS | 28.C N | GLU | 41.D OE2 | 3.358 | 2.561 | |
| LEU | 29.C N | GLU | 41.D OE1 | 2.884 | 1.893 | |
| LEU | 29.C N | GLU | 41.D OE2 | 3.23 | 2.475 | |
| TYR | 30.C N | GLU | 41.D OE2 | 3.221 | 2.229 | |
| ASN | 75.C ND2 | ASP | 119.D O | 3.246 | 2.304 | |
| TRP | 77.C NE1 | ASP | 119.D O | 2.874 | 1.874 | |

| | | | | | | |
|------------|-----------|-----|-----------|-------|-------|----|
| ARG | 101.C NE | SER | 172.D O | 2.826 | 2.165 | |
| ARG | 101.C NH1 | SER | 172.D O | 2.274 | 1.465 | |
| DE (WT-WT) | 5 H-bonds | | | | | 65 |
| LYS | 28.D N | GLU | 41.E OE1 | 3.065 | 2.218 | |
| LYS | 28.D N | GLU | 41.E OE2 | 2.995 | 2.078 | |
| LEU | 29.D N | GLU | 41.E OE1 | 3.038 | 2.273 | |
| ARG | 101.D NH1 | ASP | 111.E OD2 | 3.359 | 2.536 | |
| SER | 117.E OG | SER | 188.D O | 3.547 | 2.753 | |
| EA (WT-WT) | 6 H-bonds | | | | | 76 |
| ARG | 65.A NH1 | ASN | 193.E O | 2.482 | 1.589 | |
| ARG | 65.A NH2 | ASN | 193.E O | 3.193 | 2.52 | |
| LYS | 31.E NZ | GLN | 17.A O | 3.539 | 2.709 | |
| ARG | 101.E NE | SER | 82.A OG | 2.963 | 2.057 | |
| ARG | 101.E NH1 | ASP | 26.A OD1 | 2.642 | 1.92 | |
| ARG | 101.E NH1 | SER | 82.A OG | 2.997 | 2.123 | |

Table S. 11: Full list of interfacial hydrogen bond contacts in AB-dupa7 receptor after MD simulations.

| Donor | | Acceptor | | Donor-Acceptor distance (Å) | Donor hydrogen-Acceptor distance (Å) | Total number of contacts |
|--------------|-----------|----------|-----------|-----------------------------|--------------------------------------|--------------------------|
| AB (dup-dup) | 2 H-bonds | | | | | 25 |
| GLN | 13.A NE2 | ILE | 21.B O | 3.305 | 2.382 | |
| ALA | 23.B N | GLN | 13.A OE1 | 3.211 | 2.437 | |
| BC (dup- WT) | 3 H-bonds | | | | | 29 |
| GLN | 13.B NE2 | TYR | 173.C OH | 3.062 | 2.378 | |
| ARG | 42.C N | GLN | 13.B OE1 | 2.966 | 2.087 | |
| TRP | 171.C NE1 | PRO | 52.B O | 2.979 | 2.171 | |
| CD (WT-WT) | 6 H-bonds | | | | | 52 |
| LEU | 29.C N | GLU | 41.D OE1 | 3.102 | 2.404 | |
| LEU | 29.C N | GLU | 41.D OE2 | 2.689 | 1.978 | |
| TYR | 30.C N | GLU | 41.D OE2 | 3.091 | 2.213 | |
| GLN | 106.C NE2 | LEU | 40.D O | 3.423 | 2.645 | |
| TYR | 115.D OH | HIS | 127.C O | 2.845 | 2.229 | |
| ARG | 121.D NH1 | GLN | 61.C OE1 | 2.897 | 2.716 | |
| DE (WT-WT) | 5 H-bonds | | | | | 51 |
| LEU | 29.D N | GLU | 41.E OE1 | 2.858 | 1.948 | |
| TYR | 30.D N | GLU | 41.E OE1 | 3.129 | 2.142 | |
| ARG | 101.D NH1 | ASP | 111.E OD1 | 2.811 | 1.856 | |
| ARG | 101.D NH2 | ASP | 111.E OD1 | 3.468 | 2.721 | |
| ARG | 101.D NH2 | SER | 172.E OG | 3.025 | 2.177 | |
| EA (WT-dup) | 5 H-bonds | | | | | 84 |
| ARG | 65.A NH1 | GLU | 195.E OE1 | 2.918 | 2.007 | |

| | | | | | | |
|-----|-----------|-----|-----------|-------|-------|--|
| ARG | 65.A NH1 | GLU | 195.E OE2 | 2.894 | 2.02 | |
| ARG | 65.A NH2 | GLU | 195.E OE1 | 3.05 | 2.199 | |
| ARG | 65.A NH2 | GLU | 195.E OE2 | 2.944 | 2.092 | |
| GLN | 139.E NE2 | SER | 82.A O | 3.428 | 2.605 | |

Table S. 12: Full list of interfacial hydrogen bond contacts in AC-dupa7 receptor after MD simulations.

| Donor | | Acceptor | | Donor-Acceptor distance (Å) | Donor hydrogen-Acceptor distance (Å) | Total number of contacts |
|--------------|-----------|----------|-----------|-----------------------------|--------------------------------------|--------------------------|
| AB (dup- WT) | 6 H-bonds | | | | | 39 |
| GLN | 11.A N | SER | 172.B O | 3.205 | 2.205 | |
| ARG | 31.A NE | PHE | 122.B O | 3.052 | 2.131 | |
| ARG | 31.A NH1 | PHE | 122.B O | 2.893 | 1.961 | |
| PHE | 32.A N | ASP | 123.B OD1 | 3.532 | 2.527 | |
| GLN | 49.A NE2 | GLU | 211.B OE2 | 3.546 | 2.627 | |
| ALA | 124.B N | PHE | 32.A O | 2.953 | 2.011 | |
| BC (WT-dup) | 1 H-bonds | | | | | 117 |
| ASN | 193.B ND2 | GLN | 70.C OE1 | 3.278 | 2.618 | |
| CD (dup- WT) | 6 H-bonds | | | | | 34 |
| ARG | 101.C NE | SER | 82.D O | 3.051 | 2.288 | |
| ARG | 101.C NH1 | SER | 82.D O | 2.859 | 2.058 | |
| ARG | 121.C NH1 | ASP | 29.D OD2 | 3.104 | 2.106 | |
| ARG | 121.C NH2 | ASP | 29.D OD1 | 3.33 | 2.322 | |
| THR | 128.C OG1 | ILE | 21.D O | 3.521 | 2.642 | |
| TYR | 83.D OH | ASN | 129.C OD1 | 3.126 | 2.314 | |
| DE (WT-WT) | 1 H-bonds | | | | | 52 |
| ARG | 42.E NE | GLN | 11.D OE1 | 3.691 | 2.721 | |
| EA (WT -dup) | 1 H-bonds | | | | | 58 |
| ASN | 75.E ND2 | ASP | 26.A O | 3.475 | 2.476 | |

Table S. 13: Full list of interfacial hydrogen bond contacts in ABC-dupa7 receptor after MD simulations.

| Donor | | Acceptor | | Donor-Acceptor distance (Å) | Donor hydrogen-Acceptor distance (Å) | Total number of contacts |
|--------------|------------|----------|-----------|-----------------------------|--------------------------------------|--------------------------|
| AB (dup-dup) | 4 H-bonds | | | | | 61 |
| GLN | 11.A N | SER | 82.B O | 3.241 | 2.296 | |
| TYR | 100.A N | TYR | 61.B O | 3.137 | 2.169 | |
| TYR | 61.B N | SER | 98.A O | 3.108 | 2.114 | |
| ARG | 65.B NH2 | LEU | 141.A O | 3.354 | 2.411 | |
| BC (dup-dup) | 10 H-bonds | | | | | 130 |
| GLN | 13.B NE2 | SER | 82.C O | 2.8 | 1.83 | |
| THR | 35.B N | ASP | 29.C O | 2.865 | 1.875 | |
| ASN | 39.B ND2 | ILE | 21.C O | 3.699 | 2.919 | |
| ARG | 139.B NH1 | ASP | 63.C OD1 | 2.855 | 1.857 | |
| ARG | 139.B NH1 | ASP | 63.C OD2 | 2.94 | 1.95 | |
| ARG | 139.B NH2 | ASP | 63.C OD1 | 2.568 | 1.752 | |
| ARG | 139.B NH2 | ASP | 63.C OD2 | 2.662 | 1.844 | |
| ALA | 23.C N | ASN | 39.B OD1 | 2.95 | 1.968 | |
| ASP | 29.C N | THR | 35.B OG1 | 2.892 | 1.907 | |
| TYR | 83.C OH | ASN | 39.B OD1 | 2.847 | 1.892 | |
| CD (dup- WT) | 1 H-bonds | | | | | 49 |
| TYR | 115.D OH | ASP | 33.C O | 2.686 | 1.793 | |
| DE (WT-WT) | 3 H-bonds | | | | | 59 |
| ARG | 101.D NH2 | ASP | 111.E OD1 | 3.568 | 2.581 | |
| ARG | 101.D NH2 | ASP | 111.E OD2 | 3.544 | 2.569 | |
| ARG | 101.D NH2 | SER | 172.E OG | 2.872 | 1.941 | |

| | | | | | | |
|--------------|-----------|-----|---------|-------|-------|----|
| EA (WT -dup) | 3 H-bonds | | | | | 33 |
| TYR | 83.A OH | ARG | 101.E O | 3.274 | 2.322 | |
| LYS | 28.E N | HIS | 18.A O | 3.528 | 2.843 | |
| LYS | 31.E NZ | ILE | 16.A O | 3.286 | 2.304 | |

Table S. 14: Full list of interfacial hydrogen bond contacts in ACD-dupa7 receptor after MD simulations.

| Donor | | Acceptor | | Donor-Acceptor distance (Å) | Donor hydrogen-Acceptor distance (Å) | Total number of contacts |
|--------------|-----------|----------|-----------|-----------------------------|--------------------------------------|--------------------------|
| AB (WT-WT) | 5 H-bonds | | | | | 48 |
| GLN | 13.D N | LEU | 40.E O | 2.97 | 2.016 | |
| ARG | 42.E N | GLN | 11.D O | 3.518 | 2.521 | |
| LYS | 109.E N | GLN | 13.D OE1 | 3.223 | 2.261 | |
| ARG | 121.E NH1 | PHE | 32.D O | 2.707 | 2.026 | |
| ARG | 121.E NH2 | PHE | 32.D O | 3.225 | 2.245 | |
| BC (WT-dup) | 1 H-bonds | | | | | 64 |
| ARG | 101.E NH2 | SER | 87.A O | 2.884 | 2.013 | |
| CD (dup-dup) | 1 H-bonds | | | | | 26 |
| TYR | 61.B OH | GLY | 99.A O | 3.411 | 2.649 | |
| DE (dup- WT) | 8 H-bonds | | | | | 59 |
| ARG | 31.B NH1 | GLU | 120.C OE1 | 2.612 | 1.642 | |
| ARG | 31.B NH1 | GLU | 120.C OE2 | 2.662 | 1.658 | |
| ARG | 31.B NH2 | GLU | 120.C OE1 | 2.966 | 1.993 | |
| ARG | 31.B NH2 | GLU | 120.C OE2 | 2.657 | 1.649 | |
| ASN | 39.B ND2 | TRP | 171.C O | 3.123 | 2.491 | |
| ASN | 103.B ND2 | LYS | 68.C O | 3.102 | 2.536 | |
| TRP | 171.C NE1 | ASN | 39.B O | 3.011 | 2.169 | |
| LYS | 214.C NZ | GLN | 11.B OE1 | 3.012 | 2.202 | |
| EA (WT -dup) | 2 H-bonds | | | | | 49 |
| ASN | 129.C ND2 | TYR | 83.D OH | 2.788 | 1.992 | |
| CYS | 122.D SG | GLN | 139.C OE1 | 3.595 | 2.308 | |

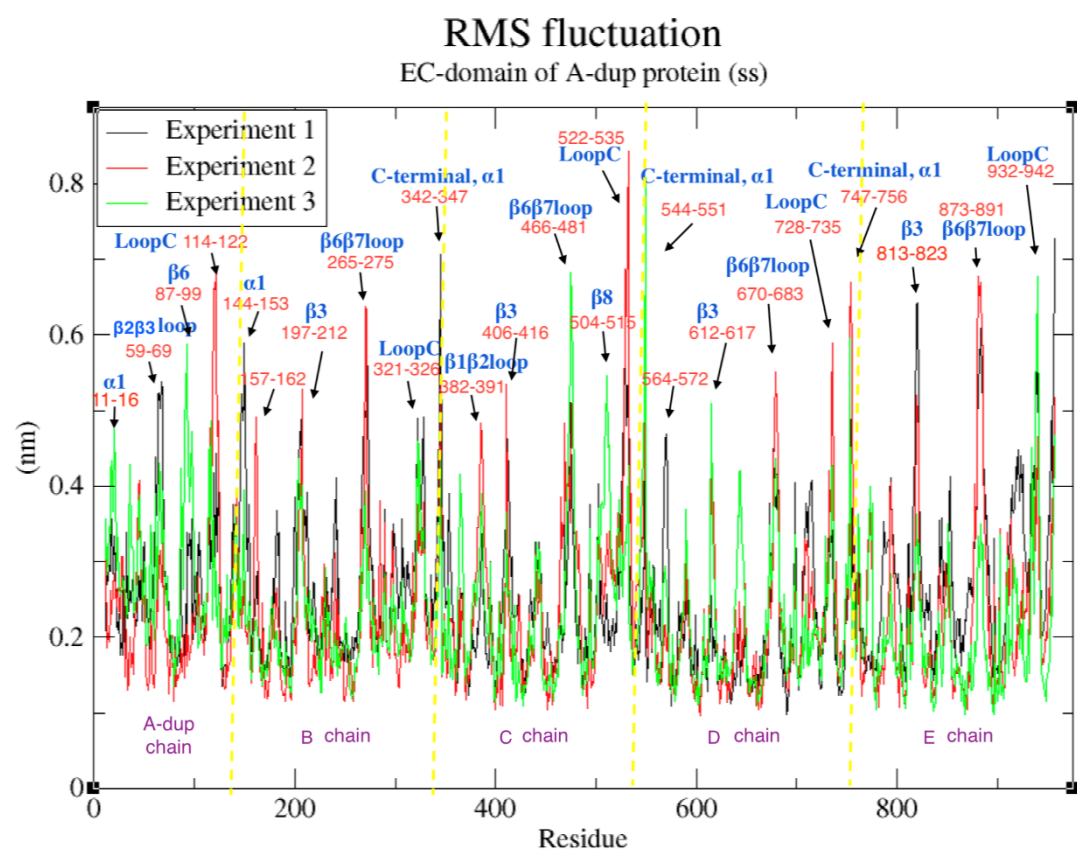
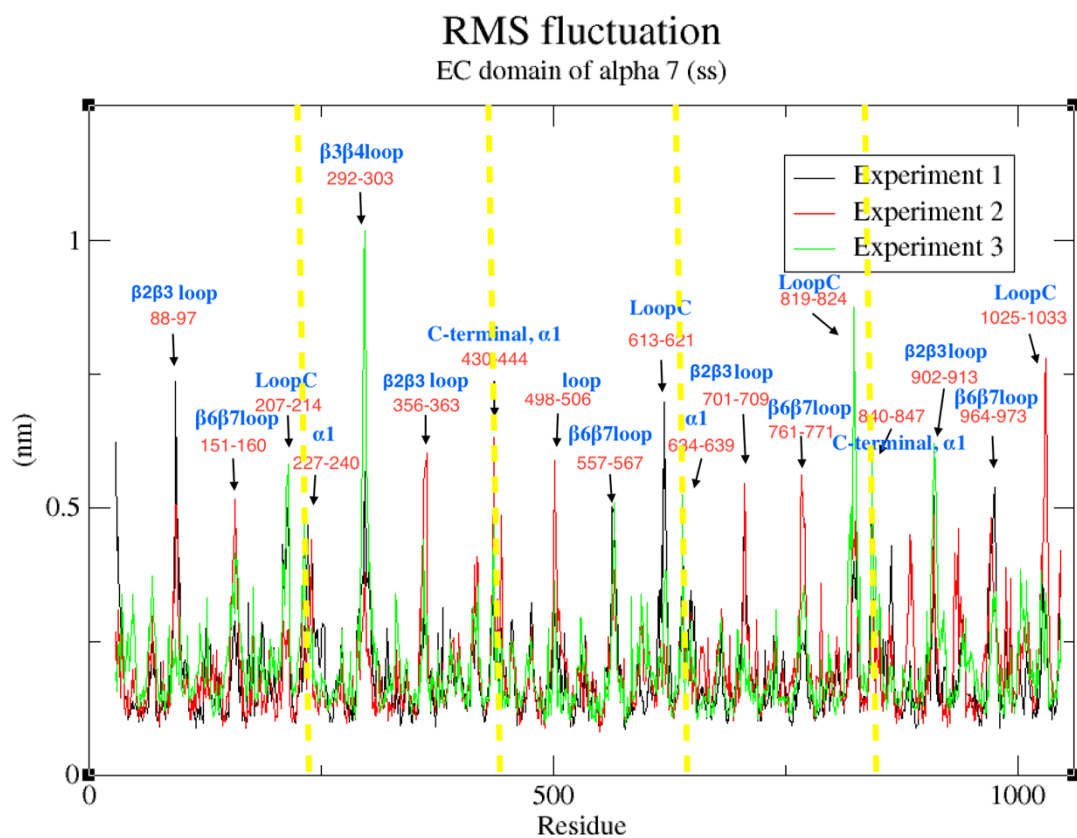
Table S. 15: Full list of interfacial hydrogen bond contacts in 4-dupa7 receptor after MD simulations.

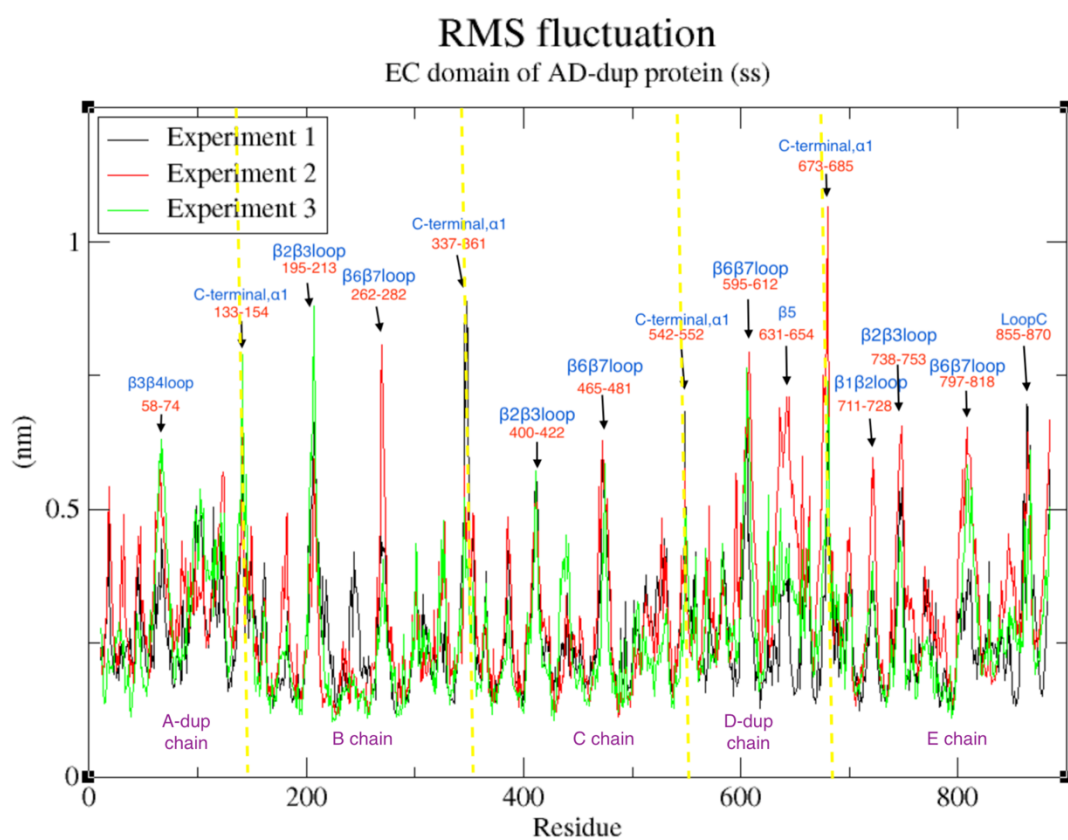
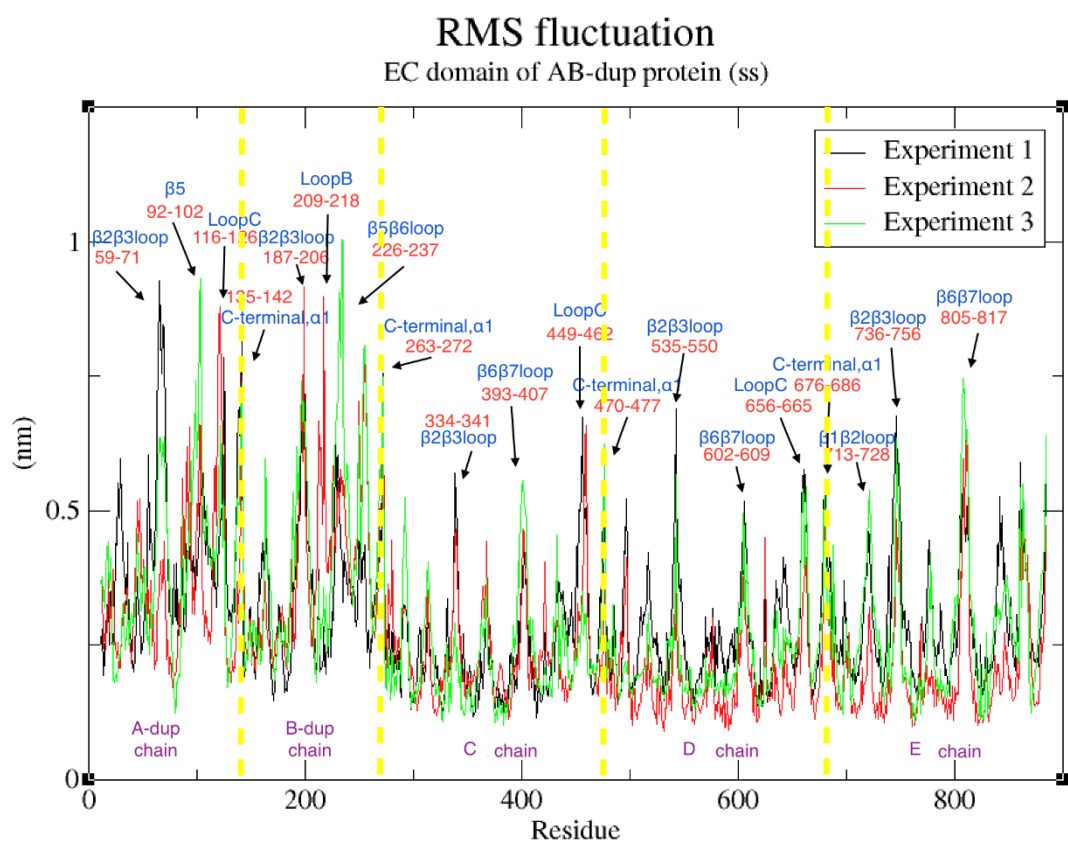
| Donor | | Acceptor | | Donor-Acceptor distance (Å) | Donor hydrogen-Acceptor distance (Å) | Total number of contacts |
|--------------|-----------|----------|-----------|-----------------------------|--------------------------------------|--------------------------|
| AB (dup-dup) | 4 H-bonds | | | | | 39 |
| GLN | 11.A N | GLU | 215.B OE1 | 2.86 | 1.995 | |
| GLN | 11.A N | GLU | 215.B OE2 | 2.817 | 1.922 | |
| GLN | 11.A N | TYR | 217.B OH | 3.318 | 2.494 | |
| GLY | 174.B N | GLN | 13.A OE1 | 3.039 | 2.23 | |
| BC (dup-dup) | 5 H-bonds | | | | | 62 |
| THR | 128.B OG1 | ASP | 29.C OD1 | 2.992 | 2.13 | |
| THR | 128.B OG1 | ASP | 29.C OD2 | 2.829 | 2.018 | |
| ASN | 129.B N | ASP | 29.C OD1 | 3.312 | 2.351 | |
| ASN | 129.B N | ASP | 29.C OD2 | 2.991 | 2.005 | |
| VAL | 130.B N | ASP | 29.C OD2 | 3.362 | 2.355 | |
| CD (dup-dup) | 6 H-bonds | | | | | 69 |
| ARG | 31.C NE | GLU | 30.D OE1 | 2.677 | 1.808 | |
| ARG | 31.C NE | GLU | 30.D OE2 | 2.689 | 1.822 | |
| ARG | 31.C NH1 | GLU | 30.D OE1 | 2.506 | 1.687 | |
| ARG | 31.C NH1 | GLU | 30.D OE2 | 2.531 | 1.726 | |
| ASN | 103.C N | CYS | 60.D O | 2.919 | 1.961 | |
| SER | 59.D OG | TYR | 100.C O | 3.182 | 2.29 | |
| DE (dup-WT) | 3 H-bonds | | | | | 167 |
| HIS | 37.D NE2 | SER | 82.E OG | 2.943 | 2.033 | |
| ASN | 39.D ND2 | ILE | 21.E O | 3.314 | 2.418 | |
| TRP | 81.E NE1 | LEU | 51.D O | 3.396 | 2.389 | |

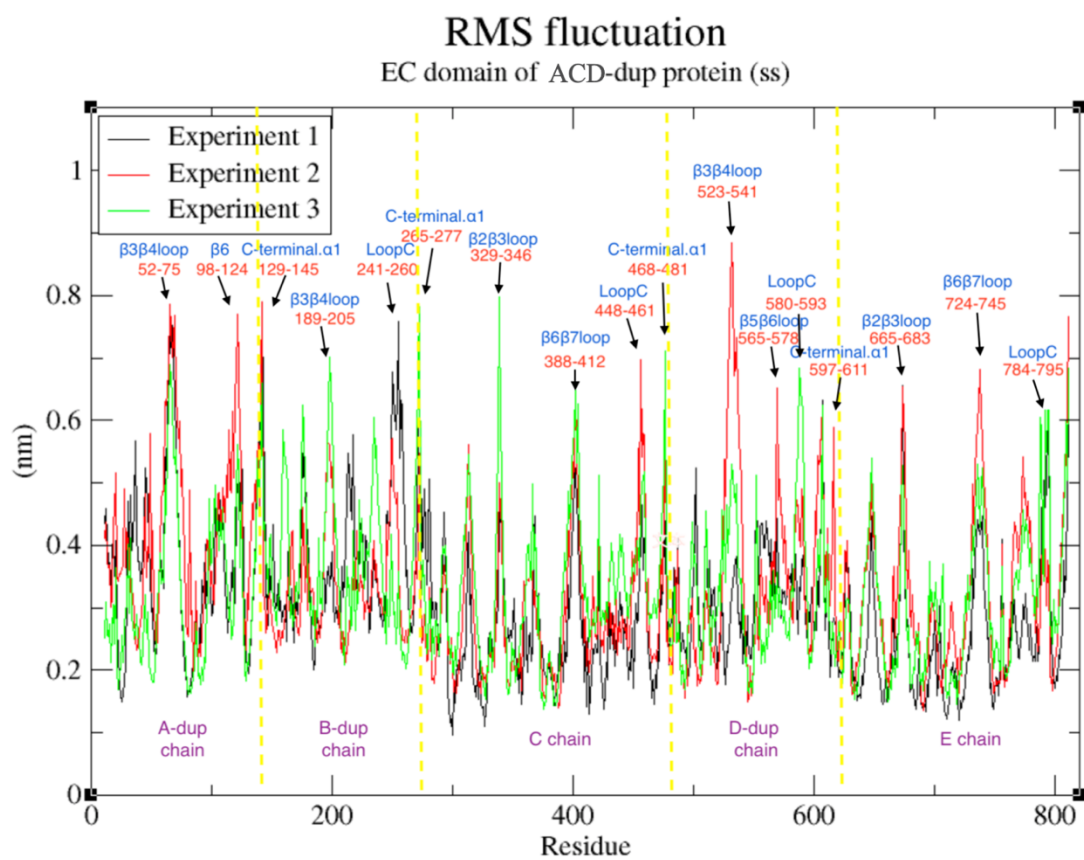
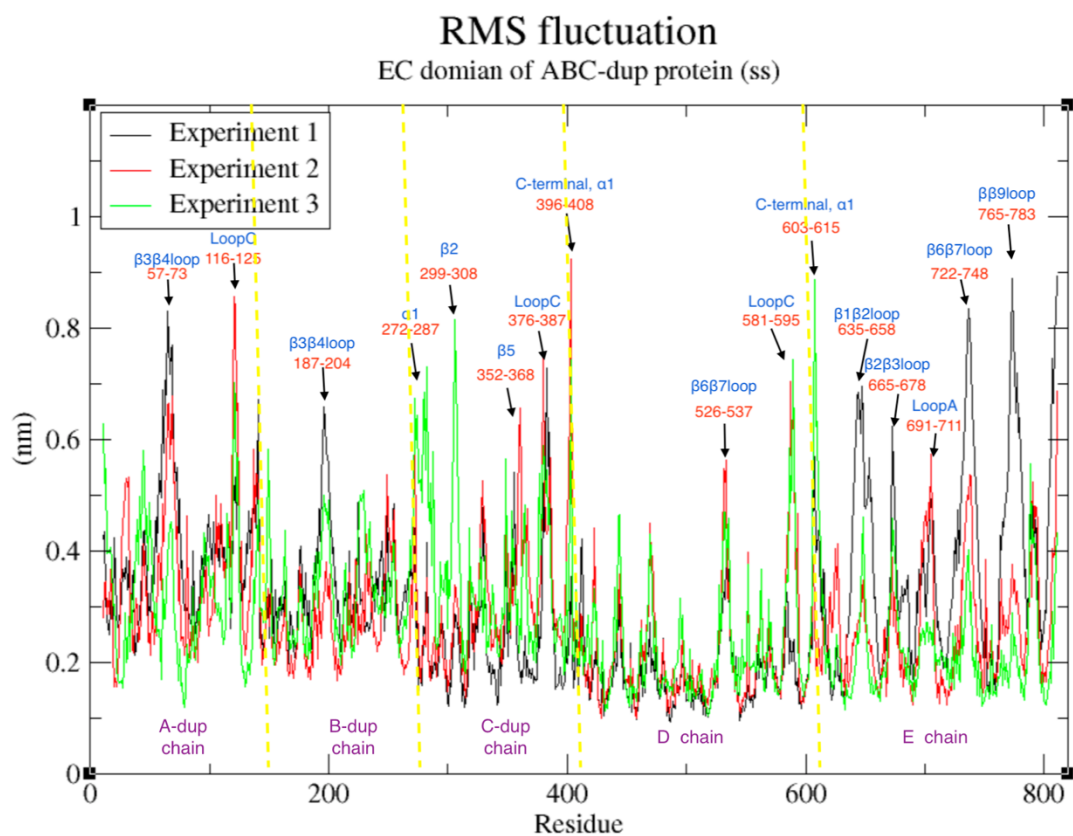
| | | | | | | |
|-------------|-----------|-----|----------|-------|-------|-----|
| EA (WT-dup) | 3 H-bonds | | | | | 138 |
| GLY | 84.A N | GLN | 13.E OE1 | 2.992 | 1.992 | |
| ALA | 34.E N | ILE | 27.A O | 2.876 | 1.991 | |
| ASN | 39.E ND2 | SER | 82.A O | 2.884 | 2.01 | |

Table S. 16: Full list of interfacial hydrogen bond contacts in 5-dupa7 receptor after MD simulations.

| Donor | | Acceptor | | Donor-Acceptor distance (Å) | Donor hydrogen-Acceptor distance (Å) | Total number of contacts |
|--------------|-----------|----------|--------|-----------------------------|--------------------------------------|--------------------------|
| AB (dup-dup) | 2 H-bonds | | | | | 88 |
| GLN | 13.A NE2 | TRP | 81.B O | 3.568 | 2.56 | |
| ALA | 34.A N | ILE | 27.B O | 3.24 | 2.258 | |
| BC (dup-dup) | 0 H-bonds | | | | | 103 |
| CD (dup-dup) | 1 H-bonds | | | | | 44 |
| TRP | 81.D NE1 | HIS | 37.C O | 2.981 | 2.005 | |
| DE (dup-dup) | 1 H-bonds | | | | | 106 |
| GLU | 30.E N | PHE | 32.D O | 2.933 | 1.939 | |
| EA (dup-dup) | 0 H-bonds | | | | | 39 |







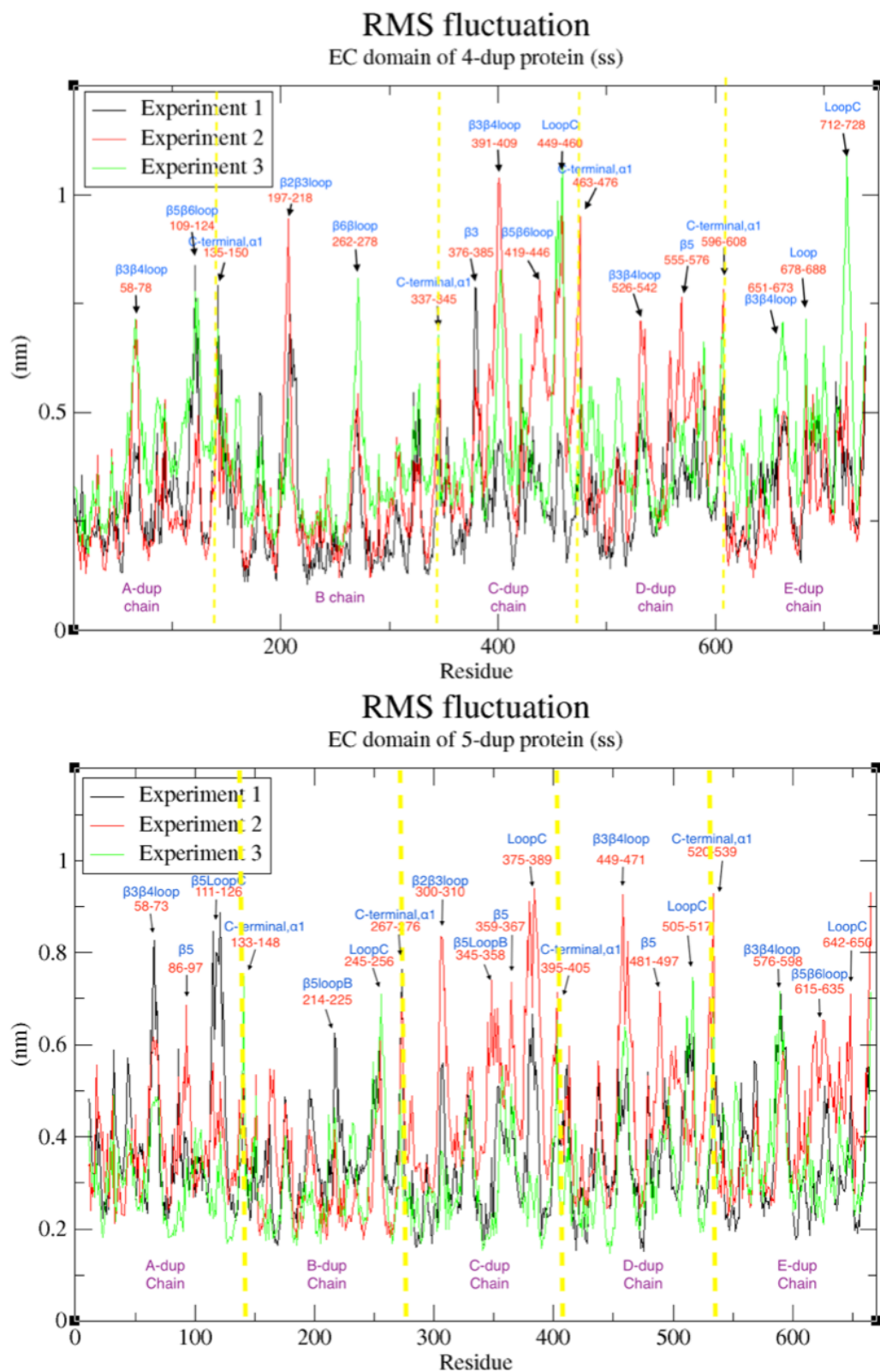


Figure S. 2: RMSF (averaged per-residue) obtained for extracellular (EC) domains of all receptor stoichiometries investigated in this study during 100 ns of all-atom MD simulation. Each simulation has been performed in triplicate. The data for all replicas are denoted as Experiment 1 - Experiment 3.

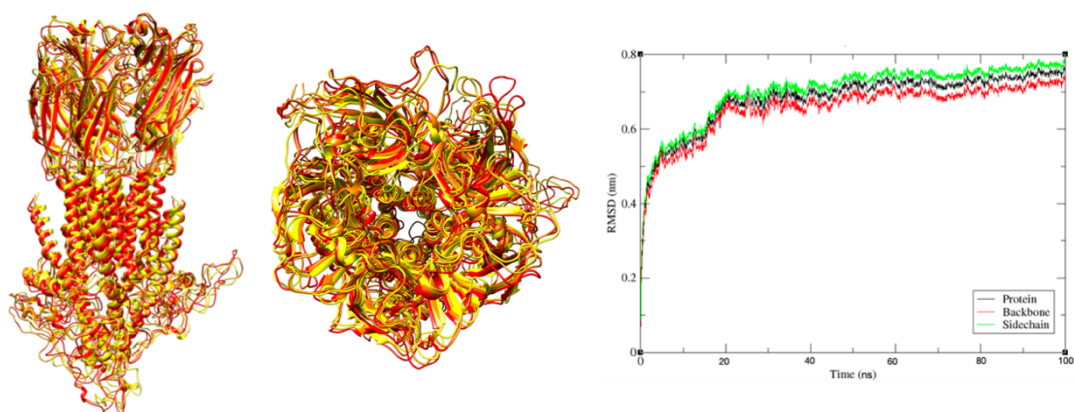


Figure S. 3: Left panel: The side view of the three conformations of the full-length $\alpha 7$ receptor. Middle panel: The top view of the three conformations of the full-length $\alpha 7$ receptor. Yellow, orange and red signify representative configurations around 15 ns, 50 ns and 85 ns of the all-atom simulation, respectively. Right panel: root-mean-square deviation (RMSD) calculated for $\alpha 7$ WT receptor during 100 ns of all-atom MD simulation. The black line shows data obtained for all atoms in the full-length receptor. The red line shows data for the protein backbone. The green line shows data for the side chain of the $\alpha 7$ WT receptor.

Table S. 17: Interaction between the ‘gatekeeping’ glutamate E254 (GLU) residue within WT- $\alpha 7$ and dup $\alpha 7$ subunits in different pentamer stoichiometries and passing Ca^{2+} cation (Ca), calculated from umbrella sampling simulations. The orange diamond represents the interaction between E254 and Ca^{2+} .

| | WT | A | AB | AD | | ABC | | ACD | | 4 | | 5 |
|--------------|----|----|-----|-----|----|-----|-----|-----|----|-----|-----|-----|
| Subunit | WT | WT | Dup | Dup | WT | Dup | Dup | Dup | WT | Dup | Dup | Dup |
| Ca-GLU (O) | ◆ | | | | | | | | | | | |
| Ca-GLU (OE1) | ◆ | ◆ | ◆ | ◆ | ◆ | ◆ | | ◆ | ◆ | ◆ | ◆ | ◆ |
| Ca-GLU (OE2) | ◆ | ◆ | ◆ | ◆ | ◆ | ◆ | ◆ | ◆ | ◆ | ◆ | ◆ | ◆ |
| Ca-GLU (CD) | ◆ | ◆ | ◆ | ◆ | ◆ | ◆ | ◆ | ◆ | ◆ | ◆ | ◆ | ◆ |
| Ca-GLU (C) | ◆ | | | | | | | | | | | |

Table S. 18: Hydrogen bonds and electrostatic interactions between WT $\alpha 7$ /dup $\alpha 7$ subunits in different receptor stoichiometries obtained from the umbrella sampling simulations. Red triangles represent the interactions between the canonical $\alpha 7$ (WT) subunit protein residues and Ca^{2+} cation. Blue diamonds represent the interactions between the protein residues in the dup $\alpha 7$ subunit and Ca^{2+} cation.

| H-bonds | WT | A | AB | AD | | ABC | | ACD | | 4 | | 5 |
|---------|----|----|-----|----|-----|-----|-----|-----|-----|-----|-----|-----|
| Subunit | WT | WT | Dup | TW | Dup | Dup | Dup | WT | Dup | Dup | Dup | Dup |
| Number | 6 | 5 | 5 | 6 | 5 | 3 | 5 | 9 | 8 | 6 | 6 | 4 |
| LEU-PHE | ▲ | ▲ | | ▲ | ◆ | | | ▲ | ◆ | ◆ | | ◆ |
| VAL-PHE | ▲ | ▲ | ◆ | ▲ | ◆ | | ◆ | | ◆ | | ◆ | |
| ALA-MET | ▲ | ▲ | | | | | ◆ | | ◆ | | ◆ | |
| GLU-LEU | ▲ | | | ▲ | | | ◆ | ▲ | ◆ | ◆ | ◆ | ◆ |
| ILE-LEU | ▲ | ▲ | | ▲ | | ◆ | | ▲ | | ◆ | ◆ | ◆ |
| MET-VAL | ▲ | | ◆ | ▲ | | | | | | | ◆ | |
| VAL-MET | | | ◆ | | | ◆ | | ▲ | | ◆ | | |
| TYR-ALA | | ▲ | | | | | | ▲ | | ◆ | | |
| LEU-VAL | | | ◆ | ▲ | ◆ | | ◆ | | ◆ | | ◆ | |
| ILE-VAL | | | ◆ | | ◆ | | ◆ | | | | | |
| ALA-LEU | | | | | ◆ | ◆ | | ▲ | | ◆ | | ◆ |
| LYS-GLU | | | | | | | | ▲ | | | | |
| ASN-ALA | | | | | | | | ▲ | | | | |
| ALA-VAL | | | | | | | | ▲ | | | | |
| TYR-ILE | | | | | | | | | ◆ | | | |

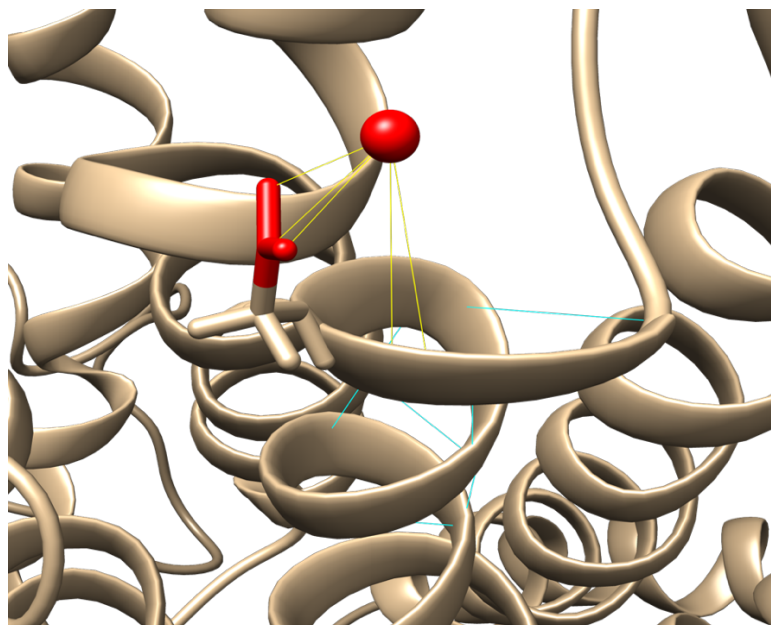


Figure S. 4: The Ca^{2+} interaction scheme within the WT receptor, at the entrance to the transmembrane (TM) region. Glutamate E254 (red) 'gatekeeper' residue interacts with the Ca^{2+} cation (yellow lines). The blue line represents the favourable electrostatic interaction between Ca^{2+} and the receptor TM region. The receptor backbone is rendered as secondary structure elements (all but E254 side chains are omitted for clarity) and coloured golden brown.

References

1. C. K. Jones, N. Byun and M. Bubser, *Neuropsychopharmacology*, 2012, **37**, 16-42.
2. W. Ockenga, S. Kuhne, S. Bocksberger, A. Banning and R. Tikkanen, *Genes (Basel)*, 2013, **4**, 171-197.
3. S. Mitra, S. N. Khatri, M. Maulik, A. Bult-Ito and M. Schulte, *Int J Mol Sci*, 2020, **21**.
4. C. C. Cafe-Mendes, H. M. Garay-Malpartida, M. B. Malta, L. de Sa Lima, C. Scavone, Z. S. Ferreira, R. P. Markus and T. Marcourakis, *Neurosci Lett*, 2017, **636**, 218-224.
5. V. Gatta, G. Mengod, M. Reale and A. M. Tata, *Biomedicines*, 2020, **8**.
6. H. Katsuki and K. Matsumoto, in *Nicotinic Acetylcholine Receptor Signaling in Neuroprotection*, eds. A. Akaike, S. Shimohama and Y. Misu, Singapore, 2018, DOI: 10.1007/978-981-10-8488-1_7, pp. 113-136.
7. M. Noda and A. I. Kobayashi, *J Physiol Sci*, 2017, **67**, 235-245.
8. Y. Yue, R. Liu, W. Cheng, Y. Hu, J. Li, X. Pan, J. Peng and P. Zhang, *Int Immunopharmacol*, 2015, **29**, 504-512.
9. R. R. Resende and A. Adhikari, *Cell Commun Signal*, 2009, **7**, 20.
10. D. L. Mann, *Circ Res*, 2002, **91**, 988-998.
11. S. V. Maurer and C. L. Williams, *Front Immunol*, 2017, **8**, 1489.
12. T. H. Ferreira-Vieira, I. M. Guimaraes, F. R. Silva and F. M. Ribeiro, *Curr Neuropsychopharmacol*, 2016, **14**, 101-115.
13. N. J. Woolf and L. L. Butcher, *Behav Brain Res*, 2011, **221**, 488-498.
14. H. Wang, M. Yu, M. Ochani, C. A. Amella, M. Tanovic, S. Susarla, J. H. Li, H. Wang, H. Yang, L. Ulloa, Y. Al-Abed, C. J. Czura and K. J. Tracey, *Nature*, 2003, **421**, 384-388.
15. S. Giunta, *Immun Ageing*, 2006, **3**, 12.
16. C. Franceschi, P. Garagnani, G. Vitale, M. Capri and S. Salvioli, *Trends Endocrinol Metab*, 2017, **28**, 199-212.
17. A. Panda, A. Arjona, E. Sapey, F. Bai, E. Fikrig, R. R. Montgomery, J. M. Lord and A. C. Shaw, *Trends Immunol*, 2009, **30**, 325-333.
18. C. Franceschi, M. Bonafe, S. Valensin, F. Olivieri, M. De Luca, E. Ottaviani and G. De Benedictis, *Ann N Y Acad Sci*, 2000, **908**, 244-254.
19. A. J. Pietrobon, F. M. E. Teixeira and M. N. Sato, *Front Immunol*, 2020, **11**, 579220.
20. B. Giunta, F. Fernandez, W. V. Nikolic, D. Obregon, E. Rrapo, T. Town and J. Tan, *J Neuroinflammation*, 2008, **5**, 51.
21. E. Boren and M. E. Gershwin, *Autoimmun Rev*, 2004, **3**, 401-406.

22. T. Fulop, G. Dupuis, J. M. Witkowski and A. Larbi, *Rev Invest Clin*, 2016, **68**, 84-91.
23. H. E. Gendelman, *J Neurovirol*, 2002, **8**, 474-479.
24. S. E. Ebert, P. Jensen, B. Ozenne, S. Armand, C. Svarer, D. S. Stenbaek, K. Moeller, A. Dyssegaard, G. Thomsen, J. Steinmetz, B. H. Forchhammer, G. M. Knudsen and L. H. Pinborg, *Eur J Neurol*, 2019, **26**, 1426-1432.
25. M. G. Fehlings and D. H. Nguyen, *J Clin Immunol*, 2010, **30 Suppl 1**, S109-112.
26. J. Gehrmann, Y. Matsumoto and G. W. Kreutzberg, *Brain Res Brain Res Rev*, 1995, **20**, 269-287.
27. B. A. t Hart and W. F. den Dunnen, *J Neuroimmune Pharmacol*, 2013, **8**, 757-759.
28. P. Troncoso-Escudero, A. Parra, M. Nassif and R. L. Vidal, *Front Neurol*, 2018, **9**, 860.
29. W. W. Chen, X. Zhang and W. J. Huang, *Mol Med Rep*, 2016, **13**, 3391-3396.
30. H. Gonzalez, J. Ottervald, K. C. Nilsson, N. Sjogren, T. Miliotis, H. Von Bahr, M. Khademi, B. Eriksson, S. Kjellstrom, A. Vegvari, R. Harris, G. Marko-Varga, K. Borg, J. Nilsson, T. Laurell, T. Olsson and B. Franzen, *J Proteomics*, 2009, **71**, 670-681.
31. S. Janelidze, N. Mattsson, E. Stomrud, O. Lindberg, S. Palmqvist, H. Zetterberg, K. Blennow and O. Hansson, *Neurology*, 2018, **91**, e867-e877.
32. S. Abu-Rumeileh, P. Steinacker, B. Polisch, A. Mammanna, A. Bartoletti-Stella, P. Oeckl, S. Baiardi, C. Zenesini, A. Huss, P. Cortelli, S. Capellari, M. Otto and P. Parchi, *Alzheimers Res Ther*, 2019, **12**, 2.
33. M. Fei, W. Jianghua, M. Rujuan, Z. Wei and W. Qian, *J Neurol Sci*, 2011, **305**, 92-96.
34. P. Lewczuk, J. Kornhuber, E. Vanmechelen, O. Peters, I. Heuser, W. Maier, F. Jessen, K. Burger, H. Hampel, L. Frolich, F. Henn, P. Falkai, E. Ruther, H. Jahn, C. Luckhaus, R. Perneczky, K. Schmidtke, J. Schroder, H. Kessler, J. Pantel, H. J. Gertz, H. Vanderstichele, G. de Meyer, F. Shapiro, S. Wolf, M. Bibl and J. Wiltfang, *Exp Neurol*, 2010, **223**, 366-370.
35. B. Borroni, M. Di Luca and A. Padovani, *Eur J Pharmacol*, 2006, **545**, 73-80.
36. R. Anand, S. A. Amici, G. Ponath, J. I. Robson, M. Nasir and S. B. McKay, *Autism-A Neurodevelopmental Journey from Genes to Behaviour*, 2011, 123-146.
37. P. Jain, A. M. Chaney, M. L. Carlson, I. M. Jackson, A. Rao and M. L. James, *J Nucl Med*, 2020, **61**, 1107-1112.
38. W. C. Kreisl, M. J. Kim, J. M. Coughlin, I. D. Henter, D. R. Owen and R. B. Innis, *Lancet Neurol*, 2020, **19**, 940-950.
39. R. Rupprecht, V. Papadopoulos, G. Rammes, T. C. Baghai, J. Fan, N. Akula, G. Groyer, D. Adams and M. Schumacher, *Nat Rev Drug Discov*, 2010, **9**, 971-988.
40. H. Hampel, M. M. Mesulam, A. C. Cuello, M. R. Farlow, E. Giacobini, G. T. Grossberg, A. S. Khachaturian, A. Vergallo, E. Cavedo, P. J. Snyder and Z. S. Khachaturian, *Brain*, 2018, **141**, 1917-1933.
41. M. R. Picciotto, M. J. Higley and Y. S. Mineur, *Neuron*, 2012, **76**, 116-129.

42. R. Gamage, I. Wagnon, I. Rossetti, R. Childs, G. Niedermayer, R. Chesworth and E. Gyengesi, *Front Cell Neurosci*, 2020, **14**, 577912.
43. H. O. Kalkman and D. Feuerbach, *Cell Mol Life Sci*, 2016, **73**, 2511-2530.
44. P. Edison, H. A. Archer, A. Gerhard, R. Hinz, N. Pavese, F. E. Turkheimer, A. Hammers, Y. F. Tai, N. Fox, A. Kennedy, M. Rossor and D. J. Brooks, *Neurobiol Dis*, 2008, **32**, 412-419.
45. E. Vowinckel, D. Reutens, B. Becher, G. Verge, A. Evans, T. Owens and J. P. Antel, *J Neurosci Res*, 1997, **50**, 345-353.
46. Y. Ouchi, E. Yoshikawa, Y. Sekine, M. Futatsubashi, T. Kanno, T. Ogusu and T. Torizuka, *Ann Neurol*, 2005, **57**, 168-175.
47. B. Ji, J. Maeda, M. Sawada, M. Ono, T. Okauchi, M. Inaji, M. R. Zhang, K. Suzuki, K. Ando, M. Staufenbiel, J. Q. Trojanowski, V. M. Lee, M. Higuchi and T. Suhara, *J Neurosci*, 2008, **28**, 12255-12267.
48. I. Muramatsu, T. Masuoka, J. Uwada, H. Yoshiki, T. Yazama, K. S. Lee, K. Sada, M. Nishio, T. Ishibashi and T. Taniguchi, in *Nicotinic Acetylcholine Receptor Signaling in Neuroprotection*, eds. A. Akaike, S. Shimohama and Y. Misu, Singapore, 2018, DOI: 10.1007/978-981-10-8488-1_3, pp. 45-58.
49. D. A. Brown, *Brain Neurosci Adv*, 2019, **3**, 2398212818820506.
50. M. Di Bari, G. Di Pinto, M. Reale, G. Mengod and A. M. Tata, *Cent Nerv Syst Agents Med Chem*, 2017, **17**, 109-115.
51. E. J. Kooi, M. Prins, N. Bajic, J. A. Belien, W. H. Gerritsen, J. van Horssen, E. Aronica, A. M. van Dam, J. J. Hoozemans, P. T. Francis, P. van der Valk and J. J. Geurts, *Acta Neuropathol*, 2011, **122**, 313-322.
52. M. Di Bari, M. Reale, M. Di Nicola, V. Orlando, S. Galizia, I. Porfilio, E. Costantini, C. D'Angelo, S. Ruggieri, S. Biagioni, C. Gasperini and A. M. Tata, *Int J Mol Sci*, 2016, **17**.
53. N. Tabet, *Age Ageing*, 2006, **35**, 336-338.
54. R. M. Eglen, *Auton Autacoid Pharmacol*, 2006, **26**, 219-233.
55. E. X. Albuquerque, E. F. Pereira, M. Alkondon and S. W. Rogers, *Physiol Rev*, 2009, **89**, 73-120.
56. K. J. Thompson and A. B. Tobin, *Cell Signal*, 2020, **70**, 109545.
57. A. M. Tata, *Recent Pat CNS Drug Discov*, 2008, **3**, 94-103.
58. K. Kawashima, T. Fujii, Y. Watanabe and H. Misawa, *Life Sci*, 1998, **62**, 1701-1705.
59. T. Fujii, Y. Watanabe, T. Inoue and K. Kawashima, *Neurochem Res*, 2003, **28**, 423-429.
60. T. B. Strom, A. Deisseroth, J. Morganroth, C. B. Carpenter and J. P. Merrill, *Proc Natl Acad Sci U S A*, 1972, **69**, 2995-2999.
61. S. G. Anagnostaras, G. G. Murphy, S. E. Hamilton, S. L. Mitchell, N. P. Rahnama, N. M. Nathanson and A. J. Silva, *Nat Neurosci*, 2003, **6**, 51-58.
62. A. Caccamo, S. Oddo, L. M. Billings, K. N. Green, H. Martinez-Coria, A. Fisher and F. M. LaFerla,

Neuron, 2006, **49**, 671-682.

63. L. Burghaus, U. Schutz, U. Krempel, R. A. de Vos, E. N. Jansen Steur, A. Wevers, J. Lindstrom and H. Schroder, *Brain Res Mol Brain Res*, 2000, **76**, 385-388.
64. J. M. Kulak, J. M. McIntosh and M. Quik, *Mol Pharmacol*, 2002, **61**, 230-238.
65. C. M. Martin-Ruiz, J. A. Court, E. Molnar, M. Lee, C. Gotti, A. Mamalaki, T. Tsouloufis, S. Tzartos, C. Ballard, R. H. Perry and E. K. Perry, *J Neurochem*, 1999, **73**, 1635-1640.
66. E. Nizri, M. Irony-Tur-Sinai, O. Lory, A. Orr-Urtreger, E. Lavi and T. Brenner, *J Immunol*, 2009, **183**, 6681-6688.
67. M. R. Farlow, *Clin Pharmacokinet*, 2003, **42**, 1383-1392.
68. B. Lu, K. Kwan, Y. A. Levine, P. S. Olofsson, H. Yang, J. Li, S. Joshi, H. Wang, U. Andersson, S. S. Chavan and K. J. Tracey, *Mol Med*, 2014, **20**, 350-358.
69. P. Ke, B. Z. Shao, Z. Q. Xu, X. W. Chen, W. Wei and C. Liu, *CNS Neurosci Ther*, 2017, **23**, 875-884.
70. D. Liu, G. Richardson, F. M. Benli, C. Park, J. V. de Souza, A. K. Bronowska and I. Spyridopoulos, *Clin Sci (Lond)*, 2020, **134**, 2243-2262.
71. B. G. Childs, M. Gluscevic, D. J. Baker, R. M. Laberge, D. Marquess, J. Dananberg and J. M. van Deursen, *Nat Rev Drug Discov*, 2017, **16**, 718-735.
72. F. Olivieri, R. Recchioni, F. Marcheselli, A. M. Abbatecola, G. Santini, G. Borghetti, R. Antonicelli and A. D. Procopio, *Curr Pharm Des*, 2013, **19**, 1710-1719.
73. Z. H. Lin, Y. C. Li, S. J. Wu, C. Zheng, Y. Z. Lin, H. Lian, W. Q. Lin and J. F. Lin, *J Cell Mol Med*, 2019, **23**, 4746-4758.
74. C. A. Baez-Pagan, M. Delgado-Velez and J. A. Lasalde-Dominicci, *J Neuroimmune Pharmacol*, 2015, **10**, 468-476.
75. M. Kucera and A. Hrabovska, *Ceska Slov Farm*, 2015, **64**, 254-263.
76. E. L. Saw, Y. Kakinuma, M. Fronius and R. Katare, *J Mol Cell Cardiol*, 2018, **125**, 129-139.
77. J. Lu and W. Wu, *Int Immunopharmacol*, 2021, **93**, 107391.
78. M. Rosas-Ballina, P. S. Olofsson, M. Ochani, S. I. Valdes-Ferrer, Y. A. Levine, C. Reardon, M. W. Tusche, V. A. Pavlov, U. Andersson, S. Chavan, T. W. Mak and K. J. Tracey, *Science*, 2011, **334**, 98-101.
79. F. Delbart, M. Brams, F. Gruss, S. Noppen, S. Peigneur, S. Boland, P. Chaltin, J. Brandao-Neto, F. von Delft, W. G. Touw, R. P. Joosten, S. Liekens, J. Tytgat and C. Ulens, *J Biol Chem*, 2018, **293**, 2534-2545.
80. A. Chatzidaki and N. S. Millar, *Biochem Pharmacol*, 2015, **97**, 408-417.
81. C. L. Morales-Perez, C. M. Noviello and R. E. Hibbs, *Nature*, 2016, **538**, 411-415.
82. S. D. Buckingham, A. K. Jones, L. A. Brown and D. B. Sattelle, *Pharmacol Rev*, 2009, **61**, 39-61.
83. P. Iturriaga-Vasquez, J. Alzate-Morales, I. Bermudez, R. Varas and M. Reyes-Parada, *Pharmacol Res*,

-
- 2015, **101**, 9-17.
84. M. Cecchini and J. P. Changeux, *Neuropharmacology*, 2015, **96**, 137-149.
85. P. N. Kao and A. Karlin, *J Biol Chem*, 1986, **261**, 8085-8088.
86. A. Karlin, *Nat Rev Neurosci*, 2002, **3**, 102-114.
87. A. Karlin and M. H. Akabas, *Neuron*, 1995, **15**, 1231-1244.
88. M. O. Ortells and G. G. Lunt, *Trends Neurosci*, 1995, **18**, 121-127.
89. E. X. Albuquerque, E. F. Pereira, N. G. Castro, M. Alkondon, S. Reinhardt, H. Schroder and A. Maelicke, *Ann N Y Acad Sci*, 1995, **757**, 48-72.
90. J. A. Dani and D. Bertrand, *Annu Rev Pharmacol Toxicol*, 2007, **47**, 699-729.
91. C. Gotti and F. Clementi, *Prog Neurobiol*, 2004, **74**, 363-396.
92. V. Itier and D. Bertrand, *FEBS Lett*, 2001, **504**, 118-125.
93. D. Kalamida, K. Poulas, V. Avramopoulou, E. Fostieri, G. Lagoumintzis, K. Lazaridis, A. Sideri, M. Zouridakis and S. J. Tzartos, *FEBS J*, 2007, **274**, 3799-3845.
94. J. A. Dani, *Int Rev Neurobiol*, 2015, **124**, 3-19.
95. L. M. Hendrickson, M. J. Guildford and A. R. Tapper, *Front Psychiatry*, 2013, **4**, 29.
96. C. Gotti, M. Moretti, A. Gaimarri, A. Zanardi, F. Clementi and M. Zoli, *Biochem Pharmacol*, 2007, **74**, 1102-1111.
97. A. R. Lew, T. R. Kellermayer, B. P. Sule and K. Szigeti, *Curr Genomics*, 2018, **19**, 420-430.
98. M. L. Sinkus, S. Graw, R. Freedman, R. G. Ross, H. A. Lester and S. Leonard, *Neuropharmacology*, 2015, **96**, 274-288.
99. Y. Jiang, H. Yuan, L. Huang, X. Hou, R. Zhou and X. Dang, *Gene*, 2019, **714**, 143996.
100. M. L. Sinkus, M. J. Lee, J. Gault, J. Logel, M. Short, R. Freedman, S. L. Christian, J. Lyon and S. Leonard, *Brain Res*, 2009, **1291**, 1-11.
101. E. Cooper, S. Couturier and M. Ballivet, *Nature*, 1991, **350**, 235-238.
102. N. Le Novere and J. P. Changeux, *J Mol Evol*, 1995, **40**, 155-172.
103. R. C. Hogg, M. Raggenbass and D. Bertrand, *Rev Physiol Biochem Pharmacol*, 2003, **147**, 1-46.
104. A. Graham, J. A. Court, C. M. Martin-Ruiz, E. Jaros, R. Perry, S. G. Volsen, S. Bose, N. Evans, P. Ince, A. Kuryatov, J. Lindstrom, C. Gotti and E. K. Perry, *Neuroscience*, 2002, **113**, 493-507.
105. E. D. Levin, F. J. McClernon and A. H. Rezvani, *Psychopharmacology (Berl)*, 2006, **184**, 523-539.
106. M. Alkondon and E. X. Albuquerque, *J Pharmacol Exp Ther*, 1993, **265**, 1455-1473.
107. J. Corradi and C. Bouzat, *Mol Pharmacol*, 2016, **90**, 288-299.

-
108. L. Chiodo, T. E. Malliavin, L. Maragliano, G. Cottone and G. Ciccotti, *PLoS One*, 2015, **10**, e0133011.
109. A. Gharpure, C. M. Noviello and R. E. Hibbs, *Neuropharmacology*, 2020, **171**, 108086.
110. K. Brejc, W. J. van Dijk, R. V. Klaassen, M. Schuurmans, J. van Der Oost, A. B. Smit and T. K. Sixma, *Nature*, 2001, **411**, 269-276.
111. P. Rucktooa, A. B. Smit and T. K. Sixma, *Biochem Pharmacol*, 2009, **78**, 777-787.
112. N. Unwin, *J Mol Biol*, 2005, **346**, 967-989.
113. N. Unwin and Y. Fujiyoshi, *J Mol Biol*, 2012, **422**, 617-634.
114. N. Unwin, *Q Rev Biophys*, 2013, **46**, 283-322.
115. R. J. Hilf and R. Dutzler, *Nature*, 2008, **452**, 375-379.
116. N. Bocquet, H. Nury, M. Baaden, C. Le Poupon, J. P. Changeux, M. Delarue and P. J. Corringer, *Nature*, 2009, **457**, 111-114.
117. N. Calimet, M. Simoes, J. P. Changeux, M. Karplus, A. Taly and M. Cecchini, *Proc Natl Acad Sci U S A*, 2013, **110**, E3987-3996.
118. A. Gharpure, J. Teng, Y. Zhuang, C. M. Noviello, R. M. Walsh, Jr., R. Cabuco, R. J. Howard, N. T. Zaveri, E. Lindahl and R. E. Hibbs, *Neuron*, 2019, **104**, 501-511 e506.
119. M. Zouridakis, A. Papakyriakou, I. A. Ivanov, I. E. Kasheverov, V. Tsetlin, S. Tzartos and P. Giastas, *Front Pharmacol*, 2019, **10**, 474.
120. R. M. Walsh, Jr., S. H. Roh, A. Gharpure, C. L. Morales-Perez, J. Teng and R. E. Hibbs, *Nature*, 2018, **557**, 261-265.
121. C. M. Henault, C. Govaerts, R. Spurny, M. Brams, A. Estrada-Mondragon, J. Lynch, D. Bertrand, E. Pardon, G. L. Evans, K. Woods, B. W. Elberson, L. G. Cuello, G. Brannigan, H. Nury, J. Steyaert, J. E. Baenziger and C. Ulens, *Nat Chem Biol*, 2019, **15**, 1156-1164.
122. M. Brams, C. Govaerts, K. Kambara, K. L. Price, R. Spurny, A. Gharpure, E. Pardon, G. L. Evans, D. Bertrand, S. C. Lummis, R. E. Hibbs, J. Steyaert and C. Ulens, *Elife*, 2020, **9**.
123. L. Polovinkin, G. Hassaine, J. Perot, E. Neumann, A. A. Jensen, S. N. Lefebvre, P. J. Corringer, J. Neyton, C. Chipot, F. Dehez, G. Schoehn and H. Nury, *Nature*, 2018, **563**, 275-279.
124. P. H. Celie, S. E. van Rossum-Fikkert, W. J. van Dijk, K. Brejc, A. B. Smit and T. K. Sixma, *Neuron*, 2004, **41**, 907-914.
125. A. Karlin, R. N. Cox, M. Dipaola, E. Holtzman, P. N. Kao, P. Lobel, L. Wang and N. Yodh, *Ann NY Acad Sci*, 1986, **463**, 53-69.
126. S. M. Sine and A. G. Engel, *Nature*, 2006, **440**, 448-455.
127. R. Spurny, S. Debaveye, A. Farinha, K. Veys, A. M. Vos, T. Gossas, J. Atack, S. Bertrand, D. Bertrand, U. H. Danielson, G. Tresadern and C. Ulens, *Proc Natl Acad Sci U S A*, 2015, **112**, E2543-2552.

-
128. D. K. Williams, J. Wang and R. L. Papke, *Biochem Pharmacol*, 2011, **82**, 915-930.
129. A. Taly, P. J. Corringer, D. Guedin, P. Lestage and J. P. Changeux, *Nat Rev Drug Discov*, 2009, **8**, 733-750.
130. T. S. Mohamed, S. S. Jayakar and A. K. Hamouda, *Front Mol Neurosci*, 2015, **8**, 71.
131. D. Gundisch and C. Eibl, *Expert Opin Ther Pat*, 2011, **21**, 1867-1896.
132. D. R. Macallan, G. G. Lunt, S. Wonnacott, K. L. Swanson, H. Rapoport and E. X. Albuquerque, *FEBS letters*, 1988, **226**, 357-363.
133. I. Wenningmann and J. P. Dilger, *Mol Pharmacol*, 2001, **60**, 790-796.
134. A. Miyazawa, Y. Fujiyoshi and N. Unwin, *Nature*, 2003, **423**, 949-955.
135. O. Beckstein, P. C. Biggin and M. S. Sansom, *The Journal of Physical Chemistry B*, 2001, **105**, 12902-12905.
136. M. H. Akabas, *Nat Chem Biol*, 2006, **2**, 128-130.
137. N. Unwin, *FEBS Lett*, 2003, **555**, 91-95.
138. N. Unwin, A. Miyazawa, J. Li and Y. Fujiyoshi, *J Mol Biol*, 2002, **319**, 1165-1176.
139. W. Xu, S. Gelber, A. Orr-Urtreger, D. Armstrong, R. A. Lewis, C. N. Ou, J. Patrick, L. Role, M. De Biasi and A. L. Beaudet, *Proc Natl Acad Sci U S A*, 1999, **96**, 5746-5751.
140. T. Booker, R. Allen, M. Marks, S. Grady, P. Whiteaker, J. Kolman, K. Smith and A. Collins, 1999.
141. J. P. Changeux, *Nat Rev Neurosci*, 2010, **11**, 389-401.
142. M. W. Holladay, M. J. Dart and J. K. Lynch, *J Med Chem*, 1997, **40**, 4169-4194.
143. N. Andersen, J. Corradi, S. M. Sine and C. Bouzat, *Proc Natl Acad Sci U S A*, 2013, **110**, 20819-20824.
144. J. Gault, M. Robinson, R. Berger, C. Drebing, J. Logel, J. Hopkins, T. Moore, S. Jacobs, J. Meriwether, M. J. Choi, E. J. Kim, K. Walton, K. Buiting, A. Davis, C. Breese, R. Freedman and S. Leonard, *Genomics*, 1998, **52**, 173-185.
145. C. Cameli, E. Bacchelli, M. De Paola, G. Giucastro, S. Cifiello, G. Collo, M. M. Cainazzo, L. A. Pini, E. Maestrini and M. Zoli, *Eur J Hum Genet*, 2018, **26**, 1824-1831.
146. A. M. de Lucas-Cerrillo, M. C. Maldifassi, F. Arnalich, J. Renart, G. Atienza, R. Serantes, J. Cruces, A. Sanchez-Pacheco, E. Andres-Mateos and C. Montiel, *J Biol Chem*, 2011, **286**, 594-606.
147. T. Araud, S. Graw, R. Berger, M. Lee, E. Neveu, D. Bertrand and S. Leonard, *Biochem Pharmacol*, 2011, **82**, 904-914.
148. K. Szigeti, I. Ihnatovych, B. Birkaya, Z. Chen, A. Ouf, D. C. Indurthi, J. E. Bard, J. Kann, A. Adams, L. Chaves, N. Sule, J. S. Reisch, V. Pavlik, R. H. B. Benedict, A. Auerbach and G. Wilding, *EBioMedicine*, 2020, **59**, 102892.
149. S. V. Kalmady, R. Agrawal, D. Venugopal, V. Shivakumar, A. C. Amaresha, S. M. Agarwal, M. Subbanna,

-
- A. Rajasekaran, J. C. Narayanaswamy, M. Debnath and G. Venkatasubramanian, *J Neural Transm (Vienna)*, 2018, **125**, 741-748.
150. T. W. Chan, S. Langness, O. Cohen, B. P. Eliceiri, A. Baird and T. W. Costantini, *Inflamm Res*, 2020, **69**, 631-633.
151. A. Maroli, S. Di Lascio, L. Druifuca, S. Cardani, E. Setten, M. Locati, D. Fornasari and R. Benfante, *J Neuroimmunol*, 2019, **332**, 155-166.
152. Y. Kunii, W. Zhang, Q. Xu, T. M. Hyde, W. McFadden, J. H. Shin, A. Deep-Soboslay, T. Ye, C. Li, J. E. Kleinman, K. H. Wang and B. K. Lipska, *Am J Psychiatry*, 2015, **172**, 1122-1130.
153. M. Lasala, C. Fabiani, J. Corradi, S. Antollini and C. Bouzat, *Front Cell Neurosci*, 2019, **13**, 37.
154. H. Y. Wang, D. H. Lee, M. R. D'Andrea, P. A. Peterson, R. P. Shank and A. B. Reitz, *J Biol Chem*, 2000, **275**, 5626-5632.
155. N. Maatuk and A. O. Samson, *Neurotoxicology*, 2013, **34**, 236-242.
156. M. Hassan, S. Shahzadi, H. Raza, M. A. Abbasi, H. Alashwal, N. Zaki, A. A. Moustafa and S. Y. Seo, *Int J Neurosci*, 2019, **129**, 666-680.
157. H. Y. Wang, C. Trocme-Thibierge, A. Stucky, S. M. Shah, J. Kvasic, A. Khan, P. Morain, I. Guignot, E. Bouguen, K. Deschet, M. Pueyo, E. Mocaer, P. J. Ousset, B. Vellas and V. Kiyasova, *Alzheimers Res Ther*, 2017, **9**, 54.
158. E. L. Goldwaser, N. K. Acharya, H. Wu, G. A. Godsey, A. Sarkar, C. A. DeMarshall, M. C. Kosciuk and R. G. Nagele, *J Alzheimers Dis*, 2020, **74**, 345-361.
159. E. P. Barykin, A. I. Garifulina, E. V. Kravkova, E. N. Spirova, A. A. Anashkina, A. A. Adzhubei, I. V. Shelukhina, I. E. Kasheverov, V. A. Mitkevich, S. A. Kozin, M. Hollmann, V. I. Tsetlin and A. A. Makarov, *Cells*, 2019, **8**.
160. H. R. Parri, C. M. Hernandez and K. T. Dineley, *Biochem Pharmacol*, 2011, **82**, 931-942.
161. M. Bencherif, *Acta Pharmacol Sin*, 2009, **30**, 702-714.
162. G. V. Mokrov, O. A. Deeva and T. A. Gudasheva, *Curr Pharm Des*, 2021, **27**, 217-237.
163. L. Veenman and M. Gavish, *Pharmacol Ther*, 2006, **110**, 503-524.
164. V. Papadopoulos, M. Baraldi, T. R. Guilarte, T. B. Knudsen, J. J. Lacapere, P. Lindemann, M. D. Norenberg, D. Nutt, A. Weizman, M. R. Zhang and M. Gavish, *Trends Pharmacol Sci*, 2006, **27**, 402-409.
165. L. Veenman, V. Papadopoulos and M. Gavish, *Curr Pharm Des*, 2007, **13**, 2385-2405.
166. J. M. Bernassau, J. L. Reversat, P. Ferrara, D. Caput and G. Lefur, *J Mol Graph*, 1993, **11**, 236-244, 235.
167. Y. Guo, R. C. Kalathur, Q. Liu, B. Kloss, R. Bruni, C. Ginter, E. Kloppmann, B. Rost and W. A. Hendrickson, *Science*, 2015, **347**, 551-555.
168. V. Papadopoulos, J. Fan and B. Zirkin, *J Neuroendocrinol*, 2018, **30**.

-
169. Z. Ilkan and F. G. Akar, *Front Physiol*, 2018, **9**, 1518.
170. M. Gavish, I. Bachman, R. Shoukrun, Y. Katz, L. Veenman, G. Weisinger and A. Weizman, *Pharmacol Rev*, 1999, **51**, 629-650.
171. S. Surinkaew, S. Chattipakorn and N. Chattipakorn, *Can J Cardiol*, 2011, **27**, 262 e263 -213.
172. L. Veenman, E. Levin, G. Weisinger, S. Leschiner, I. Spanier, S. H. Snyder, A. Weizman and M. Gavish, *Biochem Pharmacol*, 2004, **68**, 689-698.
173. N. Leducq, F. Bono, T. Sulpice, V. Vin, P. Janiak, G. L. Fur, S. E. O'Connor and J. M. Herbert, *J Pharmacol Exp Ther*, 2003, **306**, 828-837.
174. L. Veenman and M. Gavish, *Curr Mol Med*, 2012, **12**, 398-412.
175. L. Veenman, Y. Shandalov and M. Gavish, *J Bioenerg Biomembr*, 2008, **40**, 199-205.
176. J. J. Lacapere and V. Papadopoulos, *Steroids*, 2003, **68**, 569-585.
177. M. Garnier, N. Boujrad, S. O. Ogwuegbu, J. R. Hudson, Jr. and V. Papadopoulos, *J Biol Chem*, 1994, **269**, 22105-22112.
178. E. Levin, A. Premkumar, L. Veenman, W. Kugler, S. Leschiner, I. Spanier, G. Weisinger, M. Lakomek, A. Weizman, S. H. Snyder, G. W. Pasternak and M. Gavish, *Biochemistry*, 2005, **44**, 9924-9935.
179. T. R. Guilarte, M. K. Loth and S. R. Guariglia, *Trends Pharmacol Sci*, 2016, **37**, 334-343.
180. V. Papadopoulos, L. Lecanu, R. C. Brown, Z. Han and Z. X. Yao, *Neuroscience*, 2006, **138**, 749-756.
181. J. Versijpt, J. C. Debruyne, K. J. Van Laere, F. De Vos, J. Keppens, K. Strijckmans, E. Achten, G. Slegers, R. A. Dierckx, J. Korf and J. L. De Reuck, *Mult Scler*, 2005, **11**, 127-134.
182. K. B. Mihalak, F. I. Carroll and C. W. Luetje, *Mol Pharmacol*, 2006, **70**, 801-805.
183. I. Bacher, B. Wu, D. R. Shytle and T. P. George, *Expert Opin Pharmacother*, 2009, **10**, 2709-2721.
184. H. Kitagawa, T. Takenouchi, R. Azuma, K. A. Wesnes, W. G. Kramer, D. E. Clody and A. L. Burnett, *Neuropsychopharmacology*, 2003, **28**, 542-551.
185. T. E. Wilens, J. Biederman, T. J. Spencer, J. Bostic, J. Prince, M. C. Monuteaux, J. Soriano, C. Fine, A. Abrams, M. Rater and D. Polisner, *Am J Psychiatry*, 1999, **156**, 1931-1937.
186. F. Chauveau, N. Van Camp, F. Dolle, B. Kuhnast, F. Hinnen, A. Damont, H. Boutin, M. James, M. Kassiou and B. Tavitian, *J Nucl Med*, 2009, **50**, 468-476.
187. J. Wang, M. Sanchez-Rosello, J. L. Acena, C. del Pozo, A. E. Sorochinsky, S. Fustero, V. A. Soloshonok and H. Liu, *Chem Rev*, 2014, **114**, 2432-2506.
188. E. L. Werry, F. M. Bright, O. Piguet, L. M. Ittner, G. M. Halliday, J. R. Hodges, M. C. Kiernan, C. T. Loy, J. J. Kril and M. Kassiou, *Int J Mol Sci*, 2019, **20**.
189. J. Knowles and G. Gromo, *Nat Rev Drug Discov*, 2003, **2**, 63-69.
190. J. P. Hughes, S. Rees, S. B. Kalindjian and K. L. Philpott, *Br J Pharmacol*, 2011, **162**, 1239-1249.

-
191. O. J. Wouters, M. McKee and J. Luyten, *JAMA*, 2020, **323**, 844-853.
192. Y. Tang, W. Zhu, K. Chen and H. Jiang, *Drug Discov Today Technol*, 2006, **3**, 307-313.
193. W. Yu and A. D. MacKerell, Jr., *Methods Mol Biol*, 2017, **1520**, 85-106.
194. M. T. Muhammed and E. Aki-Yalcin, *Chem Biol Drug Des*, 2019, **93**, 12-20.
195. E. Krieger, S. B. Nabuurs and G. Vriend, *Methods Biochem Anal*, 2003, **44**, 509-523.
196. C. Chothia and A. M. Lesk, *EMBO J*, 1986, **5**, 823-826.
197. P. Majumder, in *Statistical Modelling and Machine Learning Principles for Bioinformatics Techniques, Tools, and Applications*, Springer, 2020, pp. 119-133.
198. H. M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov and P. E. Bourne, *Nucleic Acids Res*, 2000, **28**, 235-242.
199. C. UniProt, *Nucleic Acids Res*, 2021, **49**, D480-D489.
200. S. F. Altschul, W. Gish, W. Miller, E. W. Myers and D. J. Lipman, *J Mol Biol*, 1990, **215**, 403-410.
201. S. Henikoff and J. G. Henikoff, *Proc Natl Acad Sci U S A*, 1992, **89**, 10915-10919.
202. M. O. Dayhoff, *Atlas of protein sequence and structure*, 1972, **5**, 89-99.
203. A. Waterhouse, M. Bertoni, S. Bienert, G. Studer, G. Tauriello, R. Gumienny, F. T. Heer, T. A. P. de Beer, C. Rempfer, L. Bordoli, R. Lepore and T. Schwede, *Nucleic Acids Res*, 2018, **46**, W296-W303.
204. M. Bertoni, F. Kiefer, M. Biasini, L. Bordoli and T. Schwede, *Sci Rep*, 2017, **7**, 10480.
205. P. Benkert, M. Biasini and T. Schwede, *Bioinformatics*, 2011, **27**, 343-350.
206. P. Benkert, S. C. Tosatto and D. Schomburg, *Proteins*, 2008, **71**, 261-277.
207. G. Studer, C. Rempfer, A. M. Waterhouse, R. Gumienny, J. Haas and T. Schwede, *Bioinformatics*, 2020, **36**, 2647.
208. A. Fiser, R. K. Do and A. Sali, *Protein Sci*, 2000, **9**, 1753-1773.
209. T. I. Croll, *Acta Crystallogr D Struct Biol*, 2018, **74**, 519-530.
210. J. Jumper, R. Evans, A. Pritzel, T. Green, M. Figurnov, O. Ronneberger, K. Tunyasuvunakool, R. Bates, A. Zidek, A. Potapenko, A. Bridgland, C. Meyer, S. A. A. Kohl, A. J. Ballard, A. Cowie, B. Romera-Paredes, S. Nikolov, R. Jain, J. Adler, T. Back, S. Petersen, D. Reiman, E. Clancy, M. Zielinski, M. Steinegger, M. Pacholska, T. Berghammer, S. Bodenstein, D. Silver, O. Vinyals, A. W. Senior, K. Kavukcuoglu, P. Kohli and D. Hassabis, *Nature*, 2021, **596**, 583-589.
211. A. R. Leach and A. R. Leach, *Molecular modelling: principles and applications*, Pearson education, 2001.
212. F. Wang, Clemson University, 2007.
213. K. Vanommeslaeghe, O. Guvench and A. D. MacKerell, Jr., *Curr Pharm Des*, 2014, **20**, 3281-3292.
214. M. González, *École thématique de la Société Française de la Neutronique*, 2011, **12**, 169-200.

-
215. B. R. Brooks, C. L. Brooks, 3rd, A. D. Mackerell, Jr., L. Nilsson, R. J. Petrella, B. Roux, Y. Won, G. Archontis, C. Bartels, S. Boresch, A. Caflisch, L. Caves, Q. Cui, A. R. Dinner, M. Feig, S. Fischer, J. Gao, M. Hodoscek, W. Im, K. Kuczera, T. Lazaridis, J. Ma, V. Ovchinnikov, E. Paci, R. W. Pastor, C. B. Post, J. Z. Pu, M. Schaefer, B. Tidor, R. M. Venable, H. L. Woodcock, X. Wu, W. Yang, D. M. York and M. Karplus, *J Comput Chem*, 2009, **30**, 1545-1614.
216. L. D. Schuler, X. Daura and W. F. Van Gunsteren, *Journal of computational chemistry*, 2001, **22**, 1205-1218.
217. W. L. Jorgensen, D. S. Maxwell and J. Tirado-Rives, *Journal of the American Chemical Society*, 1996, **118**, 11225-11236.
218. K. Lindorff-Larsen, S. Piana, K. Palmo, P. Maragakis, J. L. Klepeis, R. O. Dror and D. E. Shaw, *Proteins*, 2010, **78**, 1950-1958.
219. M. P. Allen, *Computational soft matter: from synthetic polymers to proteins*, 2004, **23**, 1-28.
220. A. Hospital, J. R. Goni, M. Orozco and J. L. Gelpi, *Adv Appl Bioinform Chem*, 2015, **8**, 37-47.
221. M. De Vivo, M. Masetti, G. Bottegoni and A. Cavalli, *J Med Chem*, 2016, **59**, 4035-4061.
222. J. D. Durrant and J. A. McCammon, *BMC Biol*, 2011, **9**, 71.
223. W. C. Swope, H. C. Andersen, P. H. Berens and K. R. Wilson, *The Journal of chemical physics*, 1982, **76**, 637-649.
224. H. J. Berendsen, J. v. Postma, W. F. van Gunsteren, A. DiNola and J. R. Haak, *The Journal of chemical physics*, 1984, **81**, 3684-3690.
225. M. Parrinello and A. Rahman, *Journal of Applied physics*, 1981, **52**, 7182-7190.
226. S. Kmiecik, D. Gront, M. Kolinski, L. Wieteska, A. E. Dawid and A. Kolinski, *Chemical reviews*, 2016, **116**, 7898-7936.
227. S. J. Marrink, H. J. Risselada, S. Yefimov, D. P. Tieleman and A. H. De Vries, *The journal of physical chemistry B*, 2007, **111**, 7812-7824.
228. L. Monticelli, S. K. Kandasamy, X. Periole, R. G. Larson, D. P. Tieleman and S. J. Marrink, *J Chem Theory Comput*, 2008, **4**, 819-834.
229. T. Z. Sen, Y. Feng, J. V. Garcia, A. Kloczkowski and R. L. Jernigan, *Journal of chemical theory and computation*, 2006, **2**, 696-704.
230. J. R. López-Blanco and P. Chacón, *Current opinion in structural biology*, 2016, **37**, 46-53.
231. A. P. Heath, L. E. Kavragi and C. Clementi, *Proteins*, 2007, **68**, 646-661.
232. J. Kästner, *Wiley Interdisciplinary Reviews: Computational Molecular Science*, 2011, **1**, 932-942.
233. M. Mills and I. Andricioaei, *J Chem Phys*, 2008, **129**, 114101.
234. A. M. Ferrenberg and R. H. Swendsen, *Computers in Physics*, 1989, **3**, 101-104.

-
235. D. Kozakov, L. E. Grove, D. R. Hall, T. Bohnuud, S. E. Mottarella, L. Luo, B. Xia, D. Beglov and S. Vajda, *Nat Protoc*, 2015, **10**, 733-755.
236. C. H. Ngan, T. Bohnuud, S. E. Mottarella, D. Beglov, E. A. Villar, D. R. Hall, D. Kozakov and S. Vajda, *Nucleic Acids Res*, 2012, **40**, W271-275.
237. X. Y. Meng, H. X. Zhang, M. Mezei and M. Cui, *Curr Comput Aided Drug Des*, 2011, **7**, 146-157.
238. D. B. Kitchen, H. Decornez, J. R. Furr and J. Bajorath, *Nat Rev Drug Discov*, 2004, **3**, 935-949.
239. C. C. Huang, E. C. Meng, J. H. Morris, E. F. Pettersen and T. E. Ferrin, *Nucleic Acids Res*, 2014, **42**, W478-484.
240. S. R. Brozell, S. Mukherjee, T. E. Balius, D. R. Roe, D. A. Case and R. C. Rizzo, *Journal of computer-aided molecular design*, 2012, **26**, 749-773.
241. N. Schneider, G. Lange, S. Hindle, R. Klein and M. Rarey, *J Comput Aided Mol Des*, 2013, **27**, 15-29.
242. I. Reulecke, G. Lange, J. Albrecht, R. Klein and M. Rarey, *ChemMedChem*, 2008, **3**, 885-897.
243. SeeSAR version 10.3.3; BioSolveIT GmbH, Sankt Augustin, Germany, 2021, www.biosolveit.de/SeeSAR).
244. D. Kozakov, D. R. Hall, B. Xia, K. A. Porter, D. Padhorny, C. Yueh, D. Beglov and S. Vajda, *Nat Protoc*, 2017, **12**, 255-278.
245. S. R. Comeau, D. W. Gatchell, S. Vajda and C. J. Camacho, *Nucleic Acids Res*, 2004, **32**, W96-99.
246. X. Daura, K. Gademann, B. Jaun, D. Seebach, W. F. Van Gunsteren and A. E. Mark, *Angewandte Chemie International Edition*, 1999, **38**, 236-240.
247. T. W. Costantini, X. Dang, M. V. Yurchyshyna, R. Coimbra, B. P. Eliceiri and A. Baird, *Mol Med*, 2015, **21**, 323-336.
248. M. Lasala, J. Corradi, A. Bruzzone, M. D. C. Esandi and C. Bouzat, *J Biol Chem*, 2018, **293**, 10707-10717.
249. T. W. Costantini, T. W. Chan, O. Cohen, S. Langness, S. Treadwell, E. Williams, B. P. Eliceiri and A. Baird, *Proc Natl Acad Sci U S A*, 2019, **116**, 7932-7940.
250. N. Andersen, J. Corradi, M. Bartos, S. M. Sine and C. Bouzat, *J Neurosci*, 2011, **31**, 3662-3669.
251. E. F. Pettersen, T. D. Goddard, C. C. Huang, G. S. Couch, D. M. Greenblatt, E. C. Meng and T. E. Ferrin, *J Comput Chem*, 2004, **25**, 1605-1612.
252. C. C. Huang, E. C. Meng, J. H. Morris, E. F. Pettersen and T. E. Ferrin, *Nucleic Acids Research*, 2014, **42**, W478-W484.
253. Z. Yang, K. Lasker, D. Schneidman-Duhovny, B. Webb, C. C. Huang, E. F. Pettersen, T. D. Goddard, E. C. Meng, A. Sali and T. E. Ferrin, *J Struct Biol*, 2012, **179**, 269-278.
254. S. Kmiecik, D. Gront, M. Kolinski, L. Wieteska, A. E. Dawid and A. Kolinski, *Chem Rev*, 2016, **116**, 7898-7936.

-
255. S. J. Marrink, H. J. Risselada, S. Yefimov, D. P. Tieleman and A. H. de Vries, *J Phys Chem B*, 2007, **111**, 7812-7824.
256. D. Van Der Spoel, E. Lindahl, B. Hess, G. Groenhof, A. E. Mark and H. J. Berendsen, *J Comput Chem*, 2005, **26**, 1701-1718.
257. S. Jo, T. Kim, V. G. Iyer and W. Im, *J Comput Chem*, 2008, **29**, 1859-1865.
258. E. L. Wu, X. Cheng, S. Jo, H. Rui, K. C. Song, E. M. Davila-Contreras, Y. Qi, J. Lee, V. Monje-Galvan, R. M. Venable, J. B. Klauda and W. Im, *J Comput Chem*, 2014, **35**, 1997-2004.
259. S. Jo, T. Kim and W. Im, *PLoS One*, 2007, **2**, e880.
260. T. Darden, D. York and L. Pedersen, *The Journal of chemical physics*, 1993, **98**, 10089-10092.
261. B. Hess, H. Bekker, H. J. Berendsen and J. G. Fraaije, *Journal of computational chemistry*, 1997, **18**, 1463-1472.
262. W. J. Allen, T. E. Balius, S. Mukherjee, S. R. Brozell, D. T. Moustakas, P. T. Lang, D. A. Case, I. D. Kuntz and R. C. Rizzo, *J Comput Chem*, 2015, **36**, 1132-1156.
263. I. Ihnatovych, T. K. Nayak, A. Ouf, N. Sule, B. Birkaya, L. Chaves, A. Auerbach and K. Szigeti, *Transl Psychiatry*, 2019, **9**, 59.
264. J. P. Roberts, S. A. Stokoe, M. F. Sathler, R. A. Nichols and S. Kim, *J Biol Chem*, 2021, DOI: 10.1016/j.jbc.2021.100402, 100402.
265. T. Chan, E. Williams, O. Cohen, B. P. Eliceiri, A. Baird and T. W. Costantini, *Neurosci Lett*, 2019, **690**, 126-131.
266. J. L. Sun, S. A. Stokoe, J. P. Roberts, M. F. Sathler, K. A. Nip, J. Shou, K. Ko, S. Tsunoda and S. Kim, *Neurobiol Aging*, 2019, **84**, 166-177.
267. A. Jakalian, D. B. Jack and C. I. Bayly, *J Comput Chem*, 2002, **23**, 1623-1641.
268. I. T. Desta, K. A. Porter, B. Xia, D. Kozakov and S. Vajda, *Structure*, 2020, **28**, 1071-1081 e1073.
269. A. Lee, K. Lee and D. Kim, *Expert opinion on drug discovery*, 2016, **11**, 707-715.
270. S. Venneti, B. J. Lopresti and C. A. Wiley, *Glia*, 2013, **61**, 10-23.
271. M. Pannell, V. Economopoulos, T. C. Wilson, V. Kersemans, P. G. Isenegger, J. R. Larkin, S. Smart, S. Gilchrist, V. Gouverneur and N. R. Sibson, *Glia*, 2020, **68**, 280-297.
272. A. M. Scarf, L. M. Ittner and M. Kassiou, *J Med Chem*, 2009, **52**, 581-592.
273. P. Gut, M. Zweckstetter and R. B. Banati, *Trends in Endocrinology & Metabolism*, 2015, **26**, 349-356.
274. B. Largeau, A. C. Dupont, D. Guilloteau, M. J. Santiago-Ribeiro and N. Arlicot, *Contrast Media Mol Imaging*, 2017, **2017**, 6592139.
275. N. K. Ramakrishnan, M. Hird, S. Thompson, D. J. Williamson, L. Qiao, D. R. Owen, A. F. Brooks, P. J. H. Scott, S. Bacallado, J. T. O'Brien and F. I. Aigbirhio, *Eur J Nucl Med Mol Imaging*, 2021, DOI:

276. L. Zhang, K. Hu, T. Shao, L. Hou, S. Zhang, W. Ye, L. Josephson, J. H. Meyer, M. R. Zhang, N. Vasdev, J. Wang, H. Xu, L. Wang and S. H. Liang, *Acta Pharm Sin B*, 2021, **11**, 373-393.
277. A. Daina, O. Michielin and V. Zoete, *Sci Rep*, 2017, **7**, 42717.