Molecular Dynamics Simulations of Protein-Membrane Interactions Focusing on PI3Kα and Its Oncogenic Mutants

A thesis presented in fulfilment of the requirements for the degree of

Doctor of Philosophy

in

Computational Biochemistry

at Massey University, Albany, New Zealand.

William A. Irvine

2017

Abstract

The interactions between proteins and membranes are key to many aspects of biological function. Molecular dynamics simulations can provide insight into both atomic-level structural details and energetics of protein-membrane interactions. This thesis describes the development of a physiologically accurate brain lipid bilayer, and its use in molecular dynamics simulations to characterise how proteins that are important drug targets interact with the cell membrane. A method for rapidly identifying the orientation of a protein that interacts most favourably with a membrane was also developed and tested.

The first chapter provides an introduction to molecular dynamics and its role in the context of this research.

The second chapter details the development of a cellular membrane with a physiologically representative brain lipid composition. This was done through the testing of simple systems prior to the construction of two more complex lipid bilayers comprising phosphatidylethanolamine (PE), phosphatidylcholine (PC), phosphatidylserine (PS), phosphatidylinositide 4,5 bisphosphate (PIP_2), sphingomyelin, and cholesterol.

The third chapter implements the brain lipid bilayer in the development of a rotational interaction energy screening method designed to predict the most favourable orientation of a protein with respect to the cellular membrane. The functionality of the method was validated through application to two membrane proteins commonly implicated in cancer: the phosphatase and tensin homolog (PTEN), and the p110α-p85α phosphatidyl-inositol kinase (PI3Kα) complex.

The fourth chapter corresponds to the main focus of this research, the behaviour of wild type PI3Kα and two of its oncogenic mutants (E545K and H1047R) with regards to membrane and substrate interaction. It was primarily found that H1047R's increased membrane affinity allowed it to sample a catalytically competent orientation independently of *Ras*, unlike the wild type. Furthermore, it was also found that the position of the C terminal tail with regards to the substrate binding pocket was crucial in the achievement of a catalytically competent position against the cellular membrane.

The fifth and final chapter describes a cytochrome P450 system embedded in a cellular membrane. It was primarily found that the properties of its ingress and egress tunnels depended on the presence or absence of a substrate in the active site.

Acknowledgements

Firstly, I would like to thank my supervisor and mentor of the last four years, Dr Jane Allison without whom this would not have been possible. Thank you for granting me this opportunity and believing in me through it all.

To my fiancée, Patrice, you motivate me beyond belief. Thank you for all the love, the care, the understanding, and most of all the patience you afforded me in the completion of my thesis.

To my parents, Andrew and Sylvie, you are both truly inspiring. Thank you for developing me into the person I am today, and for instigating the drive I need to achieve my goals.

To my siblings, Jonathan and Melissa, being an only child would be far too much pressure. Thank you for relieving me of that burden through your own excellence, I'm proud of you both.

To my extended family and all my friends who have been there for me in person or in spirit, it means so much. Thank you for all the encouragement.

Secondly, to my collaborators, co-supervisors, and colleagues - Dr Jack Flanagan, Dr Joel Tyndall, Dr Thomas Piggot - thank you for your priceless guidance and advice as I stepped into each of your fields.

On that note, I would also like to thank my fantastic research group, present and past members - Tom, Lukas, Eli, Ivan, Ashar, and Shamim. All our lengthy discussions and your inputs were a fundamental part of this research.

Last but by no means least, I would like to thank the INMS, the CTCP, and by extension, Massey University, it was great to have the blessing to undertake my research in your amazing facilities. Of course, I would be remiss if I did not also express my gratitude for my selection as a recipient of a Massey University Doctoral Scholarship.

Table of Contents

List of Figures and Tables

Chapter 1

Chapter 2

Chapter 3

Chapter 4

Chapter 5

Dictionary of Common Abbreviations

ATP - Adenosine Triphosphate
APL - Area per Lipid
ABD - Adaptor Binding Domain
AKT - Protein Kinase B
BH - Bcl-2 Homology
CHOL - Cholesterol
CYP - Cytochrome P450
DNA - Deoxyribonucleic Acid
DSSP - Define Secondary Structure of Proteins
DP - Dipalmitoyl
DO - Dioleoyl
DOP - Deuterium Order Parameter
E545K - Mutation (E→K) at position 545
GUI - Graphical User Interface
GPCR - G-Protein Coupled Receptor
GTPase - Guanosine Triphosphatase
HDx - Hydrogen Deuterium Exchange
H1047R - Mutation (H→R) at position 1047
LINCS - Linear Constraint Solver Algorithm
LJ - Lennard-Jones
MBD - Membrane Binding Domain
MD - Molecular Dynamics
NMR - Nuclear Magnetic Resonance
NVT - Number, Volume, Temperature
NPT - Number, Pressure, Temperature
PI3K - Phosphatidylinositide 3-Kinase
PIK3CA - PI3K Catalytic Subunit Alpha
PC - Phosphatidylcholine
PE - Phosphatidylethanolamine
PS - Phosphatidylserine
PIP_2 - Phosphatidylinositde 4,5 Bisphosphate
PO - Palmitoyl Oleoyl
PMF - Potential of Mean Force
PME - Particle Mesh Ewald
RMSD - Root-Mean-Square Deviation
RMSF - Root-Mean-Square Fluctuation
RBD - Ras Binding Domain
SH2 - Src Homology 2
SO - Stearoyl Oleoyl
SA - Stearoyl Arachidonoyl
SPC - Simple Point Charge
SASA - Solvent Accessible Surface Area
SGML - Sphingomyelin
SRS - Substrate Recognition Site
VDW - Van der Waals
VPL - Volume per Lipid

Chapter 1 - Introduction

1.1 - Literature Review

1.1.1 - Characterisation of Proteins

Proteins are molecules that carry out many of the functions required for life, including but not limited to: maintenance of cell structure, transport of molecules, signal transduction and catalysis. Proteins are constructed from amino acid building blocks, the sequence of which is determined by the genes which encode them. Their shapes and functions are intrinsically linked to their amino acid sequences, which determine their three-dimensional structure. The function of a protein is generally highly dependent on its structure and dynamics, and disease often results from protein malfunction.

One way of classifying proteins is on the basis of their cellular function. Structural proteins such as keratin and elastin maintain the shape and size of specific cells around the body which require rigidity. They tend to be fibrous and insoluble in nature. Transport proteins, such as haemoglobin, move materials around the body, whether they are small ions or larger macromolecules. Defence proteins, commonly referred to as antibodies, belong to the immune system which identifies and neutralises foreign pathogens in the body.

When studying the link between protein malfunction and disease, the two most heavily implicated classes are enzymes and signalling proteins. Enzymes are highly specific proteins which catalyse various chemical reactions around the cell, ranging from metabolism to deoxyribonucleic acid (DNA) transcription. Signalling proteins generally collaborate with hormones and regulate physiological behaviour and biological pathways around the body. Both enzymes and signalling proteins commonly interact with cell membranes. In fact, as much as 30% of the human genome encodes membrane proteins [1], many of which act as drug targets [2]. As both enzymes and signalling proteins regulate cellular function and behaviour, any alteration in their structure will more often than not negatively impact a cascade of activities, resulting in disease.

There exist several experimental methods to characterise protein structure, such as x-ray and neutron diffraction from crystals or fibres, mass spectrometry, solid- and solution-state nuclear magnetic resonance (NMR), transmission electron microscopy, atomic-force microscopy, and most recently x-ray free electron lasers. All of these methods, with the exception of NMR, deal with characterising and visualising single structures at specific points in time. While NMR does take into account structural changes in that the measured observables are time- and ensemble-averages,

molecular modelling must be used to interpret these in terms of structural properties and their dynamics. Computational techniques such as molecular dynamics allow study of molecular structure in atomistic detail throughout simulations spanning hundreds of nanoseconds or more. This not only allows visualisation of the structure, but also gives insight on the dynamics that give way to the protein's function. While not a novel phenomenon, the use of molecular dynamics to explore biomolecular interactions has only recently increased in significance due to improved computer processing speeds, which now enables simulations of experimentally or physiologically relevant timescales to be run.

1.1.2 - Molecular Dynamics

Molecular dynamics (MD) involves the study of a system of interacting particles through time using computer simulation, according to the rules laid out in a force field. In the case of large biological molecules such as proteins, each particle typically represents an atom but coarse-grained representations also exist whereby multi-atom clusters are modelled as a single bead.

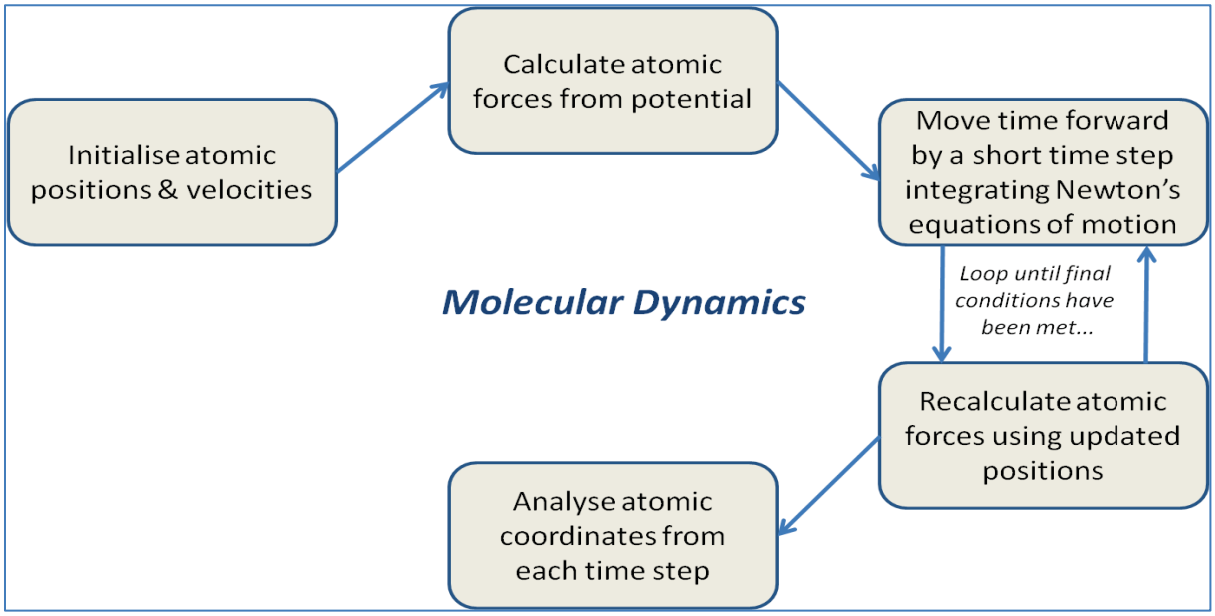


Figure 1i - *Highly simplified flow chart depicting the order of operations that take place during a molecular dynamics simulation. Final conditions could refer to system convergence or a point in the simulation at which the process of interest has been completed.*

Each step of a MD simulation depicted in Figure 1i will be described below along with the typical decisions and choices a user faces when applying the technique.

Initialise Atomic Positions and Velocities

Initial coordinates for a molecule of interest are typically obtained from the protein data bank [3]. In atomistic models, each atom of each molecule is assigned Cartesian coordinates describing its position in three-dimensional space, and initial velocities are drawn from a Maxwell-Boltzmann distribution. The system is typically energy minimised, solvated, and the charges optionally balanced through the introduction of counter-ions prior to the beginning of a MD simulation. Generally, MD simulations can either be run in an NVT (constant number of atoms, volume, and temperature) or more commonly, an NPT (constant number of atoms, pressure, and temperature) ensemble. Temperature and pressure are regulated via a user-defined thermostat and barostat, respectively.

Calculate Atomic Forces from Potential

The atoms comprising the molecules interact with each other in the simulation according to rules laid out in a force field. This contains parameters that define each atom type, in terms of their size, mass and charge. Additionally, it contains potential functions describing any pre-defined interactions between pairs (bond lengths - Eq. 1), triplets (bond angles - Eq. 2) or quartets (dihedral angles - Eq. 3, Eq. 4) of atoms joined by one or multiple bonds, and the so-called "van der Waals" (VDW) (Eq. 5) and electrostatic interactions (Eq. 6) between pairs of non-bonded atoms which are described in more detail in Section 3.2. The sum of all these terms gives rise to the potential energy of a system (Eq. 7), and is the derivative of the potential energy with respect to the atomic positions (Eq. 8).

Bonded Terms (Eq. 1 to Eq. 4)

$$V(b) = k_b(b - b_0)^2 \qquad \textbf{(Eq. 1)}$$

Where the potential energy (V) of a bond length (b) is a function of its deviation from the equilibrium bond length (b_0) regulated by a force constant (k_b).

$$V(\theta) = k_\theta(\theta - \theta_0)^2 \qquad \textbf{(Eq. 2)}$$

Where the potential energy (V) of a bond angle (ϑ) is a function of its deviation from the equilibrium bond angle (θ_0) regulated by a force constant (k_ϑ).

$$V(\varphi) = k_\varphi(1 + \cos(n\varphi - \varphi_0)) \qquad \textbf{(Eq. 3)}$$

Where the potential energy (V) of a dihedral (φ) is a cosine expansion of its deviation from the equilibrium dihedral (φ_0) regulated by a force constant (k_φ); the multiplicity (n) determines the symmetry of the rotor.

$$V(\psi) = k_\psi(\psi - \psi_0)^2 \qquad \textbf{(Eq. 4)}$$

Where the potential energy (V) of an improper dihedral (ψ) is a function of its deviation from the equilibrium dihedral (ψ_0) regulated by a force constant (k_ψ).

Non-Bonded Terms (Eq. 5 and Eq. 6)

$$V(LJ) = 4\varepsilon[\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6}] \qquad \textbf{(Eq. 5)}$$

Where the potential energy (V) of a VDW interaction is described by the Lennard-Jones (LJ) potential measured at all distances (r) less than a user-specified cut-off; the potential is attractive at all distances greater than σ, with a minimum (most attractive) energy (ε) at an optimal distance.

$$V(C) = \frac{1}{4\pi\varepsilon_0}\frac{q_1 q_2}{r} \qquad \textbf{(Eq. 6)}$$

Where the potential energy (V) of an electrostatic interaction is described by the Coulombic (C) potential between two charged atoms (q) at a distance (r) regulated by Coulomb's constant ($\frac{1}{4\pi\varepsilon_0}$).

$$\sum_{total} = \sum_{bonded} + \sum_{non-bonded} \qquad \textbf{(Eq. 7)}$$

Where the total potential energy of a system is the sum of the bonded and non-bonded potentials described in Eq. 1 through Eq. 6.

$$\boldsymbol{F_i} = -\frac{\delta V}{\delta r_i} \qquad \textbf{(Eq. 8)}$$

Where the force (F) on an atom (i) is the derivative of its potential interactions (V) at a position (r).

Move Time Forward by a Short Step Integrating Newton's Equations of Motion

A suitable timestep is chosen (typically 1 to 2 fs) which allows for the conservation of energy and the integration of the equations of motion at each step using for example, the leap-frog algorithm [4]. The movement of the atoms is simulated by numerically integrating Newton's equations of motion (Eq. 9) for all atoms of the system, giving rise to new positions.

$$\frac{\mathrm{d}r_i}{\mathrm{d}t} = v_i \,; \frac{\mathrm{d}v_i}{\mathrm{d}t} = \frac{F_i}{m_i} \qquad \textbf{(Eq. 9)}$$

*Where the rate of change (*dt*) of an atom position (r) is its velocity (v), and the rate of change of its velocity is a function of the force (F) exerted on its mass (m).*

Recalculate Atomic Forces Using Updated Positions

Atomic forces are recalculated as outlined above, with the previous two steps looped until final conditions defined by the user have been met.

Analyse Atomic Coordinates from Each Time Step

The atomic coordinates of the system are written to an output file as a function of time in the form of a trajectory. The behaviour of the system can then be analysed through characterisation of the atomic motions and dynamics, as well as the change in their energies.

1.1.3 - History of Molecular Dynamics

The application of MD methods was conceived several decades ago in the 1950s [5], with the earliest MD simulation involving atoms interacting only through perfect collisions. This model was then improved through the application of a smooth, continuous potential more representative of atomic interactions [6]. As the mathematical formulation describing atom interactions and the availability of computers improved, so did the complexity of the systems being simulated.

The first MD simulation of a biological macromolecule was published 40 years ago, involving the bovine pancreatic trypsin inhibitor [7]. This simulation lasted all of 9.2 ps with a very basic molecular mechanics potential, but achieved the goal in the field of establishing proteins as dynamic structures. This opened the door for a series of MD simulations exploring the motion of protein and nucleic acid structures over the upcoming years [8-16].

The trend has continued with the complexity of systems simulated increasing gradually over time, taking us to a point where the simulation of biological systems comprising several macromolecules or molecular aggregates such as portions of the cell membrane is now possible [17, 18]. This is thanks largely to the ever increasing processing power of the computers used to carry out MD simulations, but also the development of the theoretical and methodological aspects resulting in the simplification of the algorithm describing the behaviour of molecules. Another major improvement in the field is the accessibility and availability of a large number of macromolecule structures, with the protein data bank housing over 100,000 entries of protein and nucleic acid structures [3].

1.1.4 - Advantages of Molecular Dynamics

As explained in section 1.1.1, the cause of diseases is often protein malfunction which is borne of an alteration in its structure and dynamics. Fundamentally, the development of methods for preventing and treating these diseases depends upon characterising protein structure and motion.

MD simulations are invaluable in this aspect, as they are capable of not only characterising structural properties and motion but also doing so on a highly detailed atomistic level as a function of time [19]. As such, MD is applied in a wide range of fields from pharmaceutical chemistry to biophysics [20]. The important aspects of any given system are user-defined, allowing for variety in force field and system representation. Various MD simulation programs with their own force fields exist such as CHARMM [21], AMBER [22], and GROMOS [23], all of which are fine tuned for the behaviour of particular molecules. Other programs exist that are able to apply a range of different force fields. Whilst atomic representations generally lead to the most accurate reproduction of system behaviour, coarse-grained representations can be used where time scales and/or system sizes

exceed the reach of current computing capabilities when describing atomistic systems [24]. Other time-saving or resource-saving alterations include solvent representation, where solvent can be represented either implicitly [25, 26] or explicitly [27, 28] depending on the importance of the role it plays in the system [29]. The current generation of computers allow simulation codes such as AMBER [30], CHARMM [31], GROMACS [32], and NAMD [33], to utilise messaging passing interface (MPI) resulting in the use of several (anywhere from 2 to 1000) computer cores in parallel, thereby reducing computational time even more when used with high performance computing (HPC) clusters.

It is important to note that MD simulations in isolation provide limited insight unless properly evaluated in tandem with experimental data as a means of validating behaviour and predictions drawn from the motion of molecules through a trajectory. Improvement of analytical methods is equally as important as the improvement in speed and accuracy of the MD simulations. To do this, more and more methods for analysing simulations [34-37] are being developed which focus on the common ground between experimental and computational methods, many of which will be used in this research.

1.1.5 - Difficulties Facing Molecular Dynamics

Utilisation of these computing resources to explore biomolecular phenomena relies upon the development of force fields and algorithms which give an accurate representation and characterisation of the thermodynamic and kinetic properties of the simulated system. In principle, any system comprising many different molecules can be simulated, but only once they have been parameterised. The interaction functions and their parameters for common biomolecules have been carefully calibrated over the years by comparing quantities calculated from MD simulations and quantities measured experimentally or computed from quantum mechanics calculations. Parameters are seldom available for novel molecules such as drugs, however, or even for many of the myriad lipids found in the cell membrane. While automated parameterisation methods exist [38], these do not guarantee that the parameters will be compatible with existing force fields used to represent the remainder of the system.

However, the most glaring limitation of MD remains its inability to access timescales representative of biological processes which typically take place over milliseconds to seconds. The accessibility of even microsecond timescales in MD is heavily dictated by system size which is further limited due to computer memory and the time it would take to compute all the interactions.

When characterising the behaviour of molecules within a system, it is essential to determine the important aspects of the system, and where focusing on the accuracy of parameters and representation will yield the most biologically relevant outcome. In this research, protein-membrane interaction is the key process being evaluated by MD. As such, systems were built to be as biologically accurate as possible in terms of components present, and extra care was applied to the development and/or application of protein and membrane parameters.

The question which needs to be asked when performing molecular dynamics simulations is: is additional insight into the behaviour of the system being gained through simulation and subsequent analysis? Ideally, the simulation will both reproduce any available experimental data, thus confirming its relevance, but also provide new insight in terms of *e.g.* atomic-level detail of the conformational changes underlying a biochemical process, or the effect of changes that could not be made experimentally. Therefore, the choice of force field and program are of utmost importance. Will the force field accurately represent the system components, and, is the program able to generate MD simulations of the system on a time scale where the observations of the system components become relevant?

1.1.6 - Scope of This Research

GROMACS 4 [39] will primarily be used as it represented one of the fastest MD simulation packages at the time this research was undertaken, and was therefore used throughout for the sake of consistency. The GROMOS 54A7 force field [40] will be used with modifications and adaptations where stated as it represented a force field which has been calibrated against thermodynamic, as well as structural, data and thus makes a good choice for observing inter-molecular interactions.

Chapter 2 details the development of a biologically representative cellular membrane following the parameterisation and testing of its lipid components. Chapter 3 employs the technique of molecular dynamics in the development of a predictive rotational screening method to shed light on optimal protein-membrane interactions. Chapter 4 focuses on the similarities and differences in behaviour of phosphatidylinositide 3-kinase (PI3K) and its oncogenic mutants, particularly how the changes relate to its membrane and substrate interactions. Finally, Chapter 5 will describe the properties of a cytochrome P450 system embedded in a cellular membrane, and how it is affected by the presence or absence of a substrate in the active site.

1.2 - Bibliography

1. Krogh, A., B. Larsson, G. von Heijne, and E.L.L. Sonnhammer, *Predicting transmembrane protein topology with a hidden markov model: application to complete genomes1.* Journal of Molecular Biology, 2001. **305**(3): p. 567-580.
2. Overington, J.P., B. Al-Lazikani, and A.L. Hopkins, *How many drug targets are there?* Nat Rev Drug Discov, 2006. **5**(12): p. 993-996.
3. Berman, H.M., J. Westbrook, Z. Feng, G. Gilliland, T.N. Bhat, H. Weissig, I.N. Shindyalov, and P.E. Bourne, *The Protein Data Bank.* Nucleic Acids Research, 2000. **28**(1): p. 235-242.
4. Hockney, R.W., S.P. Goel, and J.W. Eastwood, *Quiet high-resolution computer models of a plasma.* Journal of Computational Physics, 1974. **14**(2): p. 148-158.
5. Alder, B.J. and T.E. Wainwright, *Phase Transition for a Hard Sphere System.* The Journal of Chemical Physics, 1957. **27**(5): p. 1208-1209.
6. Rahman, A., *Correlations in the Motion of Atoms in Liquid Argon.* Physical Review, 1964. **136**(2A): p. A405-A411.
7. McCammon, J.A., B.R. Gelin, and M. Karplus, *Dynamics of folded proteins.* Nature, 1977. **267**: p. 585-590.
8. Case, D.A. and M. Karplus, *Dynamics of ligand binding to heme proteins.* J. Mol. Biol., 1979. **132**: p. 343-368.
9. Ichiye, T. and M. Karplus, *Fluorescence depolarization of tryptophan residues in proteins: a molecular dynamics study.* Biochemistry, 1983. **22**: p. 2884-2893.
10. Harvey, S.C., M. Prabhakaran, B. Mao, and J.A. McCammon, *Phenylalanine transfer RNA: molecular dynamics simulation.* Science, 1984. **223**: p. 1189-1191.
11. Irikura, K.K., B. Tidor, B.R. Brooks, and M. Karplus, *Transition from B to Z DNA: contribution of internal fluctuations to the configurational entropy difference.* Science, 1985. **229**: p. 571-572.
12. Brunger, A.T., C.L. Brooks, and M. Karplus, *Active site dynamics of ribonuclease.* Proc. Natl. Acad. Sci. USA, 1985. **82**: p. 8458-8462.
13. Smith, J., S. Cusack, U. Pezzeca, B.R. Brooks, and M. Karplus, *Inelastic neutron scattering analysis of low frequency motion in proteins: A normal mode study of the bovine pancreatic trypsin inhibitor.* J. Chem. Phys., 1986. **85**: p. 3636-3654.
14. Dobson, C.M. and M. Karplus, *Internal motion of proteins: Nuclear magnetic resonance measurements and dynamic simulations.* Methods Enzymol., 1986. **131**: p. 362-389.
15. Nilsson, L., G.M. Clore, A.M. Gronenborn, A.T. Brunger, and M. Karplus, *Structure refinement of oligonucleotides by molecular dynamics with nuclear Overhauser effect interproton distance restraints: application to 5[prime] d(C-G-T-A-C-G)2.* J. Mol. Biol., 1986. **188**: p. 455-475.
16. Frauenfelder, H., *Thermal expansion of a protein.* Biochemistry, 1987. **26**: p. 254-261.
17. Brandman, R., Y. Brandman, and V.S. Pande, *A-Site Residues Move Independently from P-Site Residues in all-Atom Molecular Dynamics Simulations of the 70S Bacterial Ribosome.* PLoS ONE, 2012. **7**(1): p. e29377.
18. Khalid, S., Nils A. Berglund, Daniel A. Holdbrook, Yuk M. Leung, and J. Parkin, *The membranes of Gram-negative bacteria: progress in molecular modelling and simulation.* Biochemical Society Transactions, 2015. **43**(2): p. 162-167.
19. Karplus, M. and J.A. McCammon, *Molecular dynamics simulations of biomolecules.* Nat Struct Mol Biol, 2002. **9**(9): p. 646-652.
20. Adcock, S.A. and J.A. McCammon, *Molecular Dynamics: Survey of Methods for Simulating the Activity of Proteins.* Chemical Reviews, 2006. **106**(5): p. 1589-1615.
21. Brooks, B.R., *CHARMM: a program for macromolecular energy, minimization, and dynamics calculations.* J. Comput. Chem., 1983. **4**: p. 187-217.
22. Weiner, P.W. and P.A. Kollman, *AMBER: assisted model building with energy refinement.* J. Comput. Chem., 1981. **2**: p. 287-303.

23. Scott, W.R.P., *The GROMOS biomolecular simulation program package.* J. Phys. Chem. A, 1999. **103**: p. 3596-3607.
24. Orozco, M., L. Orellana, A. Hospital, A.N. Naganathan, A. Emperador, O. Carrillo, and J.L. Gelpí, *Coarse-grained Representation of Protein Flexibility. Foundations, Successes, and Shortcomings.* Advances in Protein Chemistry and Structural Biology, 2011. **85**: p. 183-215.
25. Roux, B.t. and T. Simonson, *Implicit solvent models.* Biophysical Chemistry, 1999. **78**(1): p. 1-20.
26. Vorobjev, Y.N., *Advances in Implicit Models of Water Solvent to Compute Conformational Free Energy and Molecular Dynamics of Proteins at Constant pH.* Advances in Protein Chemistry and Structural Biology, 2011. **85**: p. 281-322.
27. Berendsen, H.J.C., J.R. Grigera, and T.P. Straatsma, *The missing term in effective pair potentials.* The Journal of Physical Chemistry, 1987. **91**(24): p. 6269-6271.
28. Jorgensen, W.L., J. Chandrasekhar, J.D. Madura, R.W. Impey, and M.L. Klein, *Comparison of simple potential functions for simulating liquid water.* The Journal of Chemical Physics, 1983. **79**(2): p. 926-935.
29. Anandakrishnan, R., A. Drozdetski, Ross C. Walker, and Alexey V. Onufriev, *Speed of Conformational Change: Comparing Explicit and Implicit Solvent Molecular Dynamics Simulations.* Biophysical Journal, 2015. **108**(5): p. 1153-1164.
30. D.A. Case, D.S.C., T.E. Cheatham, III, T.A. Darden, R.E. Duke, T.J. Giese, H. Gohlke, A.W. Goetz, D. Greene, N. Homeyer, S. Izadi, A. Kovalenko, T.S. Lee, S. LeGrand, P. Li, C. Lin, J. Liu, T. Luchko, R. Luo, D. Mermelstein, K.M. Merz, G. Monard, H. Nguyen, I. Omelyan, A. Onufriev, F. Pan, R. Qi, D.R. Roe, A. Roitberg, C. Sagui, C.L. Simmerling, W.M. Botello-Smith, J. Swails, R.C. Walker, J. Wang, R.M. Wolf, X. Wu, L. Xiao, D.M. York and P.A. Kollman, *AMBER 2017.* 2017.
31. Brooks, B.R., C.L. Brooks, A.D. MacKerell, 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, *CHARMM: The Biomolecular Simulation Program.* Journal of Computational Chemistry, 2009. **30**(10): p. 1545-1614.
32. Abraham, M.J., T. Murtola, R. Schulz, S. Páll, J.C. Smith, B. Hess, and E. Lindahl, *GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers.* SoftwareX, 2015. **1–2**: p. 19-25.
33. Phillips, J.C., R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, E. Villa, C. Chipot, R.D. Skeel, L. Kalé, and K. Schulten, *Scalable molecular dynamics with NAMD.* Journal of Computational Chemistry, 2005. **26**(16): p. 1781-1802.
34. Koehl, P. and M. Levitt, *Structure-based conformational preferences of amino acids.* Proceedings of the National Academy of Sciences, 1999. **96**(22): p. 12524-12529.
35. Chau, P.L. and A.J. Hardwick, *A new order parameter for tetrahedral configurations.* Molecular Physics, 1998. **93**(3): p. 511-518.
36. Eisenhaber, F., P. Lijnzaad, P. Argos, C. Sander, and M. Scharf, *The double cubic lattice method: Efficient approaches to numerical integration of surface area and volume and to dot surface contouring of molecular assemblies.* Journal of Computational Chemistry, 1995. **16**(3): p. 273-284.
37. Lemkul, J.A. and D.R. Bevan, *Assessing the Stability of Alzheimer's Amyloid Protofibrils Using Molecular Dynamics.* The Journal of Physical Chemistry B, 2010. **114**(4): p. 1652-1660.
38. Malde, A.K., L. Zuo, M. Breeze, M. Stroet, D. Poger, P.C. Nair, C. Oostenbrink, and A.E. Mark, *An Automated Force Field Topology Builder (ATB) and Repository: Version 1.0.* Journal of Chemical Theory and Computation, 2011. **7**(12): p. 4026-4037.

39. Pronk, S., S. Pall, R. Schulz, P. Larsson, P. Bjelkmar, R. Apostolov, M.R. Shirts, J.C. Smith, P.M. Kasson, D. van der Spoel, B. Hess, and E. Lindahl, *GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit.* Bioinformatics, 2013. **29**(7): p. 845-54.
40. Schmid, N., A.P. Eichenberger, A. Choutko, S. Riniker, M. Winger, A.E. Mark, and W.F. van Gunsteren, *Definition and testing of the GROMOS force-field versions 54A7 and 54B7.* European Biophysics Journal, 2011. **40**(7): p. 843.

Chapter 2 - Lipid Parameterisation and Membrane Building

2.1 - Literature Review

2.1.1 - Overview

Biological membranes are lipid bilayers that coat cells and their organelles, establishing a physical boundary and a thus a micro-environment in which the cell or organelle can carry out its necessary biological processes [1]. The formation of organelles leads to compartmentalisation of the cell, where each compartment is associated with a specific function [2]. Cell membranes consist primarily of lipids, which are made up of hydrophobic fatty acid tails and a polar headgroup which differs from lipid to lipid. The types of lipids present in a membrane and their relative abundance are linked with bilayer function and properties [3]. For instance, membrane properties such as permeability and fluidity [4] are vital to bilayer functionality. It is also increasingly being realised that membrane-associated proteins are sensitive to the particular types of lipids present in the membrane [5-7].

In 1972, Singer and Nicolson proposed the fluid mosaic model for the substructure of cell membranes [8], in which the various components are scattered randomly in the lateral dimension of the membrane to form a "mosaic", with lateral diffusion of these components lending the membrane fluidity. Considerably more is now known about the substructure of cell membranes and the non-lipid components, such as proteins, that reside in and interact with membranes, as is explored in later chapters. Lipids commonly act as cellular targets for signalling proteins and are a vital part of the processes these proteins regulate [9]. As such, the focus of this chapter will be the development of bilayers of physiologically relevant lipid compositions for their use in molecular dynamics simulations with signalling proteins.

2.1.2 - Advantages of Computational Representation

Understanding the myriad biological functions associated with cellular and organellar membranes and the factors governing the localisation and function of membrane-associated proteins is a necessity. This can be achieved through thorough characterisation of the molecular composition of membranes, how this varies in space and between different cell types and biological states, and the effect that this has on cellular processes. The site of interaction with membrane-associated proteins is determined by the location and function of the membrane, which is defined by its lipid composition [10, 11].

While it is possible to obtain low-resolution images of intact cell membranes with techniques such as fluorescence microscopy [12], the nanometre scale of membrane domains - such as clusters of particular lipids such as PIP_2 - renders them too small for such studies to shed light on their molecular level composition and structure. Additionally, these techniques cannot report on the molecular-level details of membrane dynamics and how this translates to protein interaction. Therefore, to gain increased insight into the structure and behaviour of membranes, alternative methods such as computer simulations are required [13].

2.1.3 - Molecular Dynamics Studies of Membranes

Molecular dynamics is a tool that can be used to shed light on the dynamic nature of lipid membranes and how that relates to their function. However, most experimental and computational studies of membranes to date have been of simplified model systems comprising only one (usually dipalmitoylphosphatidylcholine, DPPC) [14-16] or two different types [17, 18] of phospholipids or other membrane-associated molecules such as cholesterol [19, 20]. Increasingly over time, research is reflecting the increased value placed on mixed membranes and the link between lipid composition and membrane function [21, 22]. Real biological membranes, however, are much more complex, comprising typically at least five to ten different types of lipids [23]. To gain a full understanding of the behaviour of cell membranes, it is necessary to simulate similarly complex systems. This understanding is limited not only by the requirement that each lipid be parameterised in a manner that is consistent with the remainder of the simulated system, but also the computing power required to simulate bilayers of a significant size for physiologically relevant timescales. To date, this has often been done using coarse-grained models [24, 25], which are useful for studying processes occurring over large spatial and temporal scales, but lack the detail accessible with atomic-level simulations. In recent times, Khalid *et al.* have advanced the field of atomistic lipid simulations, focusing on the development of bacterial membranes [26, 27]; however, little simulation work is available on mammalian membranes with eukaryotic membrane studies typically being for yeast [28, 29]. Advances in the development of supercomputers Anton and Anton2 [30, 31] have allowed for molecular dynamics simulations on millisecond timescales; however, for most researchers, access to hundreds of nanoseconds to microsecond timescales remains the limit on high performance computing clusters.

2.1.4 - Scope of This Research

Developing the capability for simulating membranes of realistic biological composition at an atomic level represents a crucial first step towards characterising their interactions with proteins and their

involvement in signal transduction, transport of drugs and other molecules through the bilayer, as well as the fundamental biophysics of cell membranes. As such, the goal was to develop parameters for lipid components most commonly found in eukaryotic cell membranes, ensuring they were functional in tandem, so as to allow for the construction of more realistic biological membranes through the variation in their composition. In particular, the focus was on the replication of a lipid bilayer typical of a brain membrane, so as to study its interaction with PI3K, as explored in Chapter 4. The lipids of choice were phosphatidylethanolamine (PE), phosphatidylcholine (PC), phosphatidylserine (PS), phosphatidylinositide 4,5 bisphosphate (PIP_2), sphingomyelin (SGML), and cholesterol (CHOL), which contained fatty acid tails varying in length and saturation, as these are the key components of mammalian brain cell membranes, as implemented in experimental studies involving PI3K [32]. Parameters were first tested by simulating single- and dual-lipid bilayers prior to the construction of two brain lipid bilayers, the second of which was simulated using both GROMOS and CHARMM parameters (Section 2.2). The first brain lipid bilayer only took headgroup variation into consideration whilst the second brain lipid bilayer also introduced tail variation.

2.1.5 - Parameter Validation

Typically, force field parameters are validated through their ability to reproduce experimental measurements. In the case of lipids, this manifests itself in the characteristics of the membrane the lipids aggregate to form. The most common and easily measurable characteristic of a membrane is the area per lipid across its lateral plane, despite its shortcomings as a benchmark for membrane properties [33]. Various other experimentally measurable characteristics include the volume per lipid, bilayer thickness, electron density profiles, and order parameters describing the phase of the bilayer [34-37]. These values are used in the majority of parameter set validations when it comes to the construction of lipid bilayers [38-41], and will therefore be used as a benchmark when evaluating the performance of lipid parameters used in this chapter.

2.2 - Methods

All simulations were prepared, performed, and analysed using the GROMACS 4.6 molecular dynamics package and its associated tools [42]. The lipid parameters were based on the GROMOS 54A7 and CHARMM36 force fields [43, 44], with minor additions as described below.

2.2.1 - Coordinates

Prior to the construction of single- and dual-lipid bilayers, all initial coordinates for individual phospholipids were taken with palmitoyl and oleoyl tails (PO). These coordinates were also later implemented in the first brain lipid bilayer.

Initial coordinates for POPE, POPC and CHOL were taken from the final coordinates of a 200 ns simulation of a bilayer of those components [45].

Initial coordinates for POPS were manually generated by combining the coordinates of a molecule of POPE with those of a molecule of (dioleoyl-) DOPS taken from a DOPS bilayer equilibrated for 40 ns at 303 K [46]. The coordinates of the common atoms of the POPE and the DOPS were superimposed and the third hydrogen on the second carbon of POPE replaced with the carboxyl group of DOPS.

Initial coordinates for SGML were taken from a bilayer of 1600 SGML molecules equilibrated for 3.8 ns at 300 K using the GROMOS 43A1-S3 force field [47].

Initial coordinates for PIP_2 were manually generated using a combination of the coordinates of a POPC molecule for the tails, and those of a (dipalmitoyl-) DPP_2 molecule [48] for the headgroup.

The coordinates for the second brain lipid bilayer introduced tail variation and included stearoyl (S) and arachidonoyl (A) tails. New coordinates for SOPE, POPC, SOPS, $SAPIP_2$, CHOL, and neuronal SGML as implemented in the second brain lipid bilayer were generated via the CHARMM GUI membrane builder [28] from which coordinates are available for a wide range of lipids.

2.2.2 - Parameters

The GROMOS 54A7 force field has only one lipid building block, DPPC. Parameters for POPC, which differs from DPPC only in terms of one of the acyl chains, were previously derived and tested by Poger *et al.* [38] and were used without adjustment here. For the remainder of the molecules, existing parameters from the GROMOS 54A7 force field were used wherever possible. Lists of the new interaction types introduced are given in Appendix A.

Molecular building blocks for POPE and POPS were mostly identical to that for POPC other than the parameters for interactions involving their differing head group atoms. For both POPE and POPS, the united-atom methyl groups of the PC head group moiety (atoms C1, C2, C3, Figure 2i-A) were replaced by hydrogen atoms, with all bonded (bond lengths, bond angles, dihedral angles and improper angles) and non-bonded methyl-nitrogen interactions replaced by standard 54A7 hydrogen-nitrogen interaction terms. Parameters for the additional carboxyl group in the POPS head group (atoms CF, OI, OJ, Figure 2i-C) were taken from existing bonded and non-bonded carbon-oxygen interactions included within the 54A7 force field.

The head group of SGML is identical to that of PC; thus, the same parameters were used. Beyond that, SGML differs in that a sphingosine group replaces the glycerol moiety that links the head group and fatty acid tails of the phospholipids (atoms N14, C13, C35, OA36, Figure 2i-D). All parameters for this region of SGML were taken from 43A1-S3 force field [49], an earlier GROMOS force field that has been specifically modified for simulating lipids. One SGML fatty acid chain is palmitate, for which parameters already existed. The other 16 carbon long tail features a double bond which was parameterised using the acyl chain double bond parameters from the oleate tail of POPC [38].

For PIP_2, head group parameters for the inositol ring (atoms C1-C6, Figure 2i-F) with the inclusion of phosphate groups (atoms O4x, P4 and O5x, P5, Figure 2i-F) were chosen on the basis of similarity to existing interactions included within the 54A7 force field.

For CHOL, new interaction types needed to be added to the 54A7 force field from the 43A1-S3 force field to describe the bonded and non-bonded interactions between carbon atoms in a non-aromatic ring. The ring carbons were assigned the existing CH2r (CH_2 in a ring) or CH0 (bare sp3 carbon with 4 bound heavy atoms) atom types with zero partial charge other than atom C2 (Figure 2i-E), which has a partial charge of +0.270 due to the attached hydroxyl group. Parameters for interaction terms describing the ideal length of a single carbon-carbon bond adjacent to a double carbon-carbon bond in a six-membered hydrocarbon ring and a single carbon-carbon bond in a five-membered hydrocarbon ring, bond angles at either end of a double bond in a six-membered hydrocarbon ring and in a five-membered hydrocarbon ring, and a weaker “improper” term describing the planarity of the five-membered hydrocarbon ring were also taken from the modified GROMOS 43A1-S3 force field [49]. For all other interactions, existing 54A7 parameters were used.

These parameters were repeated with the exception of those describing the tails when the second brain lipid bilayer was built, with double bond parameters (taken from the Poger *et al.* [38] oleoyl

chain) introduced where appropriate. CHARMM36 [44] parameters for the second brain lipid bilayer were obtained via the CHARMM GUI membrane builder [28].

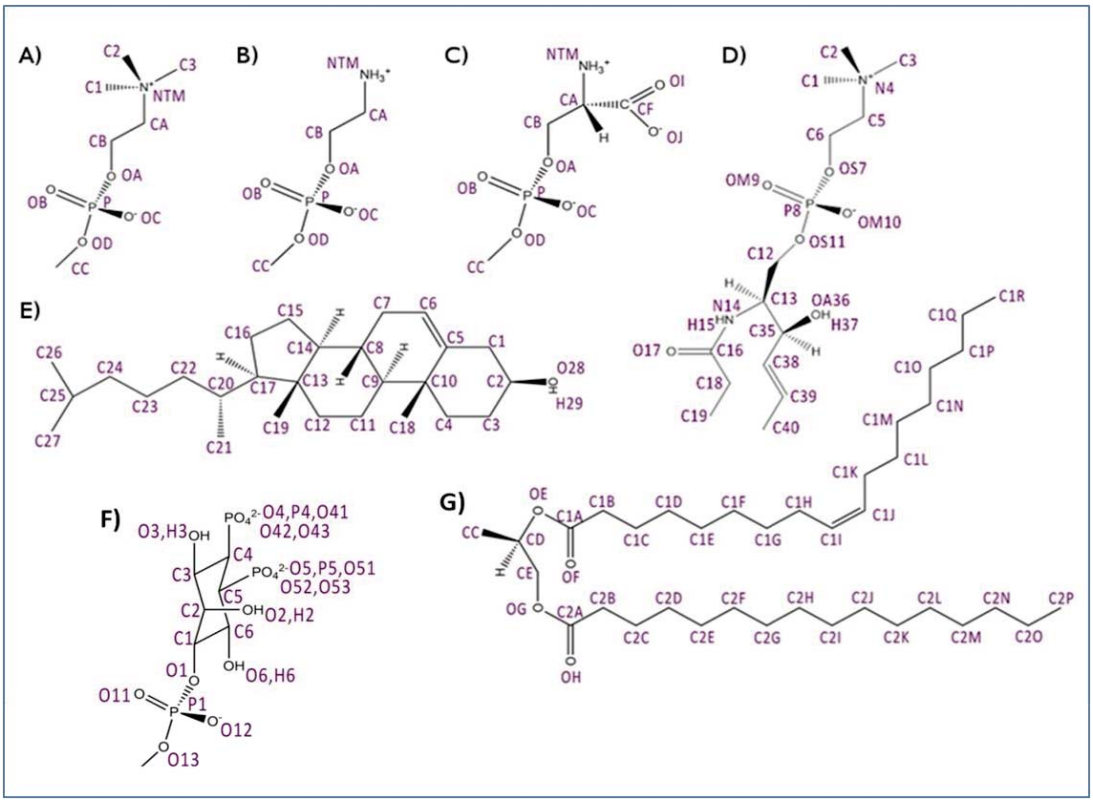


Figure 2i - *Chemical structures of the components of the brain lipid bilayer.* ***A)*** *phosphatidylcholine (PC);* ***B)*** *phosphatidylethanolamine (PE);* ***C)*** *phosphatidylserine (PS);* ***D)*** *sphingomyelin;* ***E)*** *cholesterol;* ***F)*** *phosphatidylinositide 4,5 bisphosphate (PIP2);* ***G)*** *palmitoyl-oleoyl (PO) tails. The atom names referred to in the text and used in the parameter files are shown in purple.*

2.2.3 - Bilayer Building Methods

Prior to incorporation into a bilayer, each individual lipid molecule was first subjected to 1000 steps of energy minimisation using the steepest descent algorithm in vacuum with the 54A7 force field modified as described above. The newly synthesised parameters were tested by simulating a series of single- and dual-lipid bilayer systems in solution: pure POPC, POPE, and POPS bilayers; as well as POPC/CHOL bilayers in varying cholesterol concentrations, ranging from 0% to 30%. In the case of POPS, positive counter-ions were also included to neutralise the negative charge from the carboxylate in the headgroup.

Each binary system was built by first randomly distributing lipids in their correct proportions into an ordered leaflet; the leaflet was then reflected and translated to form the other half of the bilayer. The POPC/CHOL bilayers all consisted of 128 lipids in total, with ratios of phospholipid to cholesterol of 16:0, 15:1, 14:2, 13:3, 12:4 and 11:5; whilst the pure bilayers also consisted of 128 lipids in total.

The first brain lipid bilayer system was constructed by first building a small leaflet with composition matching that of a neuronal cell membrane (45% POPE, 15% POPC, 20% POPS, 5% PIP_2, 10% CHOL, 5% SGML). This leaflet was mirrored to form a bilayer before being multiplied to form a final system with 480 lipids and dimensions of 13.73 nm by 12.06 nm by 6.93 nm. The second brain lipid bilayer system was constructed in the same way, altering only the tails to SOPE, POPC, SOPS, $SAPIP_2$, CHOL and neuronal SGML but maintaining the same relative composition giving a final system with 800 lipids and dimensions of 16.36 nm by 16.36 nm by 7.36 nm.

Each complete bilayer was subjected to 1000 steps of energy minimisation using the steepest descent algorithm in vacuum with the modified 54A7 force field, then solvated in a rectangular box using the simple point charge (SPC) [50] water model. In the case of the first brain lipid bilayer, the system was neutralised by addition of 216 cations (Na^+) to give a final composition of 216 POPE, 72 POPC, 96 POPS, 24 PIP_2, 48 CHOL, 24 SGML, 19639 SPC water and 216 Na^+ ions. In the case of the second brain lipid bilayer, the system was neutralised by the addition of 360 cations (Na^+) to give a final composition of 360 SOPE, 120 POPC, 160 SOPS, 40 $SAPIP_2$, 80 CHOL, 40 SGML, 33232 SPC water and 360 Na^+ ions.

The second brain lipid bilayer system was constructed to be directly comparable to the system constructed through the CHARMM membrane builder [28], which allowed coordinates for new lipids with the remodelled tails to be generated, as well as providing an opportunity to simulate an identical system using an alternative force field. A heterogenous symmetric bilayer of identical lipid composition was specified in the input, which generated an output with an identical number of lipids but instead used 23231 TIP3P water molecules [51] and 320 potassium ions.

2.2.4 - Simulation Methods

The MD simulations of each system using the modified GROMOS 54A7 force field were initiated with the following equilibration scheme. First, the initial velocities were randomly generated from a Maxwell-Boltzmann distribution at 50 K. The system was then heated to the ideal temperature (300 K for POPC, 308 K for POPE, 300 K for POPS, 300 K for the Brain Lipid Bilayer) over the course of 100 ps in the NVT ensemble, with temperature controlled using the Berendsen thermostat [52], with a temperature coupling constant (τ_T) of 0.1 ps, and periodic boundary conditions. This ideal

temperature was representative of a physiologically relevant temperature of at least 10 K above the phase transition point of each lipid [53]. The LINCS algorithm [54] was used with an order of 4 to constrain bond lengths and water bond angles, allowing for an integration time step of 2 fs. The centre of mass motion was removed every 100 ps. Non-bonded interactions were calculated using a grid cut-off scheme. The non-bonded Lennard-Jones interactions within a cut-off distance of 0.9 nm were calculated at every step from a pair list that was updated every fifth time step. At this point, interactions between atoms within 1.4 nm were also calculated and were kept constant between updates. Electrostatic interactions were calculated using particle mesh Ewald (PME) summation [55], with a cut-off distance of 0.9 nm. The system was then further equilibrated for 400 ps at the ideal temperatures outlined above in the NPT ensemble, with pressure controlled using the Berendsen barostat in a semi-isotropic environment, with two pressure coupling constants (τ_p) of 0.5 ps and two isothermal compressibilities of 4.5 x 10^{-5} bar^{-1}.

The final equilibrated coordinates were used as the starting configuration for 100 ns MD simulations at the ideal temperatures, with coordinates saved every 200 ps.

The MD simulation of the system built using the CHARMM GUI membrane builder, and therefore the CHARMM36 force field [44], was initiated with the equilibration scheme provided by the builder. The differences with the GROMOS 54A7 equilibration scheme are as follows. A Verlet cut-off scheme [56] was used to describe neighbouring atoms, with non-bonded interactions cut off at 1.2 nm and force-switched between 1.0 and 1.2 nm. Electrostatic interactions were also calculated using PME summation, instead with a cut-off distance of 1.2 nm. When simulating in the NPT ensemble, the temperature of 303.15 K was controlled using the Nose-Hoover [57, 58] thermostat, with τ_T = 1 ps. The pressure was instead controlled using the Parrinello-Rahman [59] barostat in a semi-isotropic environment, with two pressure coupling constants of 5.0 ps and two isothermal compressibilities of 4.5 x 10^{-5} bar^{-1}.

2.2.5 - Analysis Methods

Area per molecule for single lipid systems was calculated as a function of the box area divided by the number of lipids in each leaflet, averaged across the two leaflets.

Area per molecule for the binary systems was calculated using the following relationships [19]:

$$A_{lipid} = \frac{2A_{box}}{(1-x)N_{lipids}}\left[1 - \frac{xN_{lipids}V_{chol}}{V_{box} - N_{water}V_{water}}\right], \quad \textbf{(Eq. 1)}$$

$$A_{chol} = \frac{2A_{box}V_{chol}}{V_{box} - N_{water}V_{water}}, \quad \textbf{(Eq. 2)}$$

where *A* is the area of the bilayer or of a single molecule in the plane of the bilayer (A_{box} is the area of the solvent-facing surface of the bilayer assuming it is planar), *V* is the volume of the complete, solvated system (V_{box}) or of an individual molecule, *N* is the total number of lipids of all types in both sides of the bilayer, x is the mole fraction of cholesterol. The value of V_{water} (0.030 nm^3) was obtained from a box of equilibrated SPC water, while the value of V_{CHOL} (0.593 nm^3) was derived from crystal structures by Hofsaβ *et al.* [19].

For a system of more than two components, such as the brain lipid bilayers, relationships of the type given in Eq. 1 and Eq. 2 are no longer applicable, thus an alternative method of calculating the area of each component that utilises Voronoi tessellation as implemented in the program APL@voro [60] was also used. This method involves projecting the centre of mass of each molecule onto a plane and tessellating the area of the plane into individual molecule sections by bisecting the lines that join the centres of mass [61].

The thickness of the bilayer was computed from the electron density profile calculated using the Gromacs program *g_density* as the distance between the highest points of the peaks corresponding to high electron density that occur at either side of the bilayer. In this way, the average length of specific lipid types as well as that of the entire bilayer could be obtained.

The average volume per lipid was calculated by multiplying the average area per lipid by half the average thickness of the membrane. More specific values for each lipid type were calculated from the average area per lipid type and the thickness calculated for the section of the bilayer where that lipid type was found as detailed above.

The deuterium order parameters of only the saturated lipid tails within each bilayer were calculated using the Gromacs program *g_order,* which calculates how ordered the carbon atoms within each acyl chain of the fatty acid tails are over the course of the simulation. Deuterium order parameters of unsaturated lipid tails were avoided due to known issues with their calculation by *g_order* [62].

2.3 - Results and Discussion

In order to validate the parameters newly introduced to the GROMOS 54A7 force field, simple bilayers containing one or two components were first constructed so as to test key physical properties previously outlined. These were compared, where possible, to experimental and/or computational values calculated from similar systems.

2.3.1 - Overview of Physical Properties

Area per Lipid

Area per lipid (APL) is one of the key parameters used to assess the accuracy of simulations of lipid bilayers. While calculation of the average area per lipid molecule for single component bilayers is simply the total area of the leaflet divided by the number of lipids, assuming negligible curvature, it is less trivial for more complex mixtures. One possibility is to use relationships of the type given in Eq. 1 and 2. However, these rely on having reasonable estimates for the volume of an individual molecule of all but one of the system components. Even if these are accurate within their own system, they may not necessarily reflect the actual average volume of that type of molecule at each time point in the simulation of every other system. Moreover, this method of calculating the average area per lipid masks the variation in area between different molecules of the same type of lipid at a given time point: *i.e.* only the variation in the average area of each type of lipid over the course of the simulation is accessible.

An alternative means of calculating area per lipid that does not suffer from these drawbacks is Voronoi tessellation. The centre of mass of each molecule is projected onto a plane parallel to the surface of the bilayer, which is then tessellated into sections corresponding to each molecule by bisecting the lines that join the molecular centres of mass [61]. Voronoi tessellation has, however, been shown to sometimes overestimate the degree of variation in area between different lipids [63].

Bilayer Thickness

Another key physical property is the bilayer thickness. This was estimated as the distance between the peaks in the electron density profile of each lipid, *i.e.* assuming that the greatest electron density lies at, or close to, the solvent-facing end of each molecule. The appropriateness of this assumption will depend on the nature of the lipid head group. In particular, CHOL sits much lower in the bilayer than the phospholipids, so that its most electron-dense moiety, the hydroxyl group, lies some way from the surface of the bilayer. In comparison, the phosphate group in phospholipids is highly electron dense, and lies close to the bilayer surface. Thus the effective thickness of the bilayer as a whole was estimated from the combined electron density profile for all lipid types.

Volume per Lipid

Volume per lipid (VPL) was simply a function of the area per lipid calculated as outlined above, and half the average thickness of the lipid(s) across the bilayer taken from the electron density profiles.

Deuterium Order Parameters

The deuterium order parameters (DOP) describe how ordered the acyl chains of the fatty acid tails are within each lipid, which describes the membrane fluidity and in which phase it resides.

2.3.2 - Single Lipid Systems

For the three single lipid systems consisting of POPC, POPE, and POPS, the area per lipid (Table 2A, Figure 2ii-B) was calculated using Eq. 1 from section 2.2.5 for every time point during the simulation, and then averaged over all time points following initial equilibration which occurred at approximately 25 ns (Figure 2ii-B). Any variation in the timeseries plot therefore represents a variation in average area per phospholipid over the course of the simulation, as opposed to variation across all lipids at any given time point.

The average area per lipid of POPC was determined to be approximately 0.63 nm^2 in a relatively thin bilayer (3.25 nm) (Figure 2ii-C) giving an average volume per lipid of approximately 1.03 nm^3 (Table 2A). This is in good agreement with experiments [64-66] and previous simulations of PC [38, 67, 68] which gave an area per lipid in the range of 0.60 to 0.65 nm^2. The volume is lower than reported, but this is undoubtedly due to the method of volume calculation which does not incorporate areas of the lipid beyond the phosphate in the headgroup (Figure 2ii-C). When calculated as a difference between the box volume and the volume occupied by water using 0.03 nm^3 as in Eq. 1, the volume per lipid of the POPC system is instead 1.24 nm^3 which is in good agreement with the aforementioned experimental data.

The average area per lipid of POPE was determined to be approximately 0.50 nm^2 in a relatively thick bilayer (3.98 nm) (Figure 2ii-C) giving a volume of 1.00 nm^3, or 1.17 nm^3 when calculated as a function of the box volume not occupied by water (Table 2A). The area per lipid is lower than the 0.56 nm^2 measured experimentally and in previous simulations [69, 70]; however, the volume is in better agreement with the 1.15 nm^3 measured. With the volume being representative but the area being lower than expected, one possible explanation is due to the POPE lipids being more rigid and therefore not the optimal representation of the fluid phase, as explained in the order parameter discussion below. As outlined in section 2.2.4, simulations of POPE in this research were conducted at 308 K, in parallel with other simulations being at least 10 K higher than the phase transition point

of each lipid. The phase transition point of POPE (298 K) is closer to its simulation temperature than that of POPS (287 K) and POPC (271 K) [53] and is therefore expected to form a more viscous bilayer.

The average area per lipid of POPS was determined to be approximately 0.55 nm^2 in a moderately thick bilayer (3.82 nm) (Figure 2ii-C) giving a volume of 1.06 nm^3, or 1.20 nm^3 when calculated as a function of the box volume not occupied by water (Table 2A), allowing for any error due to the presence of unmeasured sodium counter-ions. This is in good agreement with previous simulation work [71, 72] which yielded area per lipid at 0.55 nm^2 and volume per lipid at 1.19 nm^3.

At any given temperature, a lipid bilayer can either exist in a liquid or a solid gel phase [73]. Lipids will be free to diffuse when in a liquid phase, but are less free to do so when in a gel phase due to a lack of kinetic energy. *In vivo*, cellular membranes exist in the liquid phase, where the temperature of their environment is in excess of their phase transition point. The temperature of the simulation will therefore control the phase of a bilayer, and as such was varied across each single lipid system. The POPE system had a temperature of 308 K, whilst the other two had a temperature of 300 K. Deuterium order parameters are a good indication of the fluidity of the membrane; if the acyl chain of the fatty acid has low order parameters, it is therefore more mobile and "disordered", describing a system in the liquid phase. The deuterium order parameters were calculated for the palmitoyl (P) fatty acid tail, as it allowed for calculation on a fully saturated tail negating any complications which may arise from the double bond in the oleoyl (O) fatty acid tail. As the deuterium order parameter gets further from zero, the more ordered the acyl chain would be. In all three cases of single lipid systems analysed, the P tail of POPC, POPE, and POPS all reflected deuterium order parameters expected from a fluid phase bilayer [14, 74, 75], with the magnitude of the parameters ranging from approximately 0.25 to 0.12 for POPC and POPS, and a slightly higher, more ordered, 0.34 to 0.18 for POPE which explains the thicker though not yet gel like nature of the bilayer (Table 2A, Figure 2ii).

Table 2A

Physical properties (Section 2.3.1) of the lipids in the single lipid and binary cholesterol systems

Physical Properties	Single Lipid Systems			POPC + CHOL %				
	POPC	POPE	POPS	6.25%	12.5%	18.75%	25%	31.25%
APL (nm^2)	0.631	0.501	0.554	0.591	0.569	0.558	0.545	0.532
Thickness (nm)	3.25	3.98	3.82	3.55	3.68	3.72	3.83	3.95
VPL (nm^3)	1.03	1.00	1.06	1.05	1.05	1.04	1.04	1.05
DOP Range	0.23 - 0.13	0.34 - 0.18	0.25 - 0.12	0.25 - 0.11	0.28 - 0.14	0.30 - 0.16	0.32 - 0.16	0.37 - 0.19

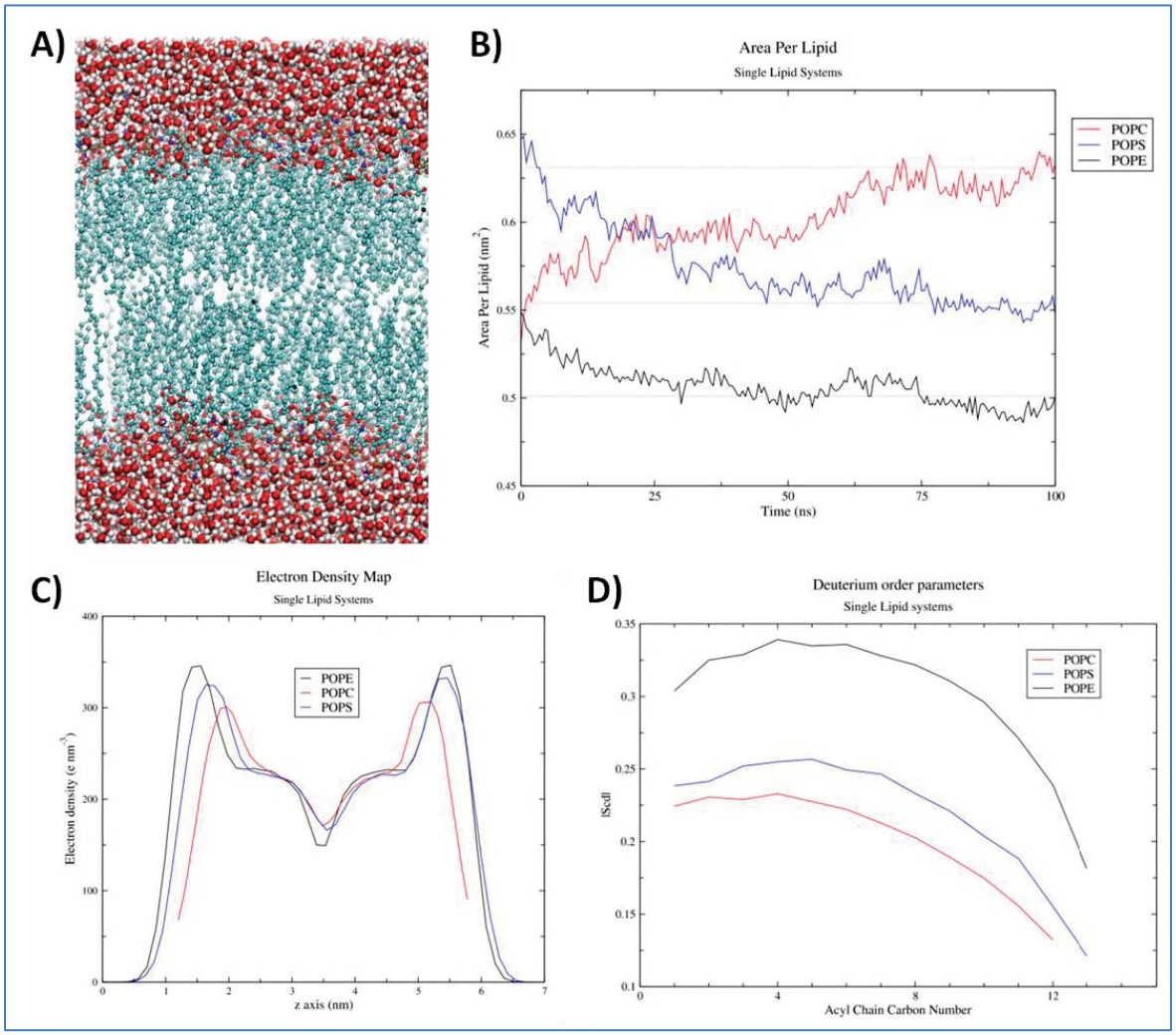


Figure 2ii - ***A)*** *Snapshot of the POPE system following 100 ns of simulation time with the atoms coloured by name - C (cyan), N (blue), O (red), H (white).* ***B)*** *Area per lipid for the three single phospholipid systems - POPE (black), POPC (red), POPS (blue).* ***C)*** *Electron density map for the three single phospholipid systems. The z axis represents the length of the normal to the bilayer.* ***D)*** *Deuterium order parameters of the palmitoyl (P) tails for the three single phospholipid systems.*

2.3.3 - Effect of Cholesterol

As the properties of cholesterol are not best observed in pure bilayers, an ideal way of observing their behaviour is through their effect on phospholipid bilayers as cholesterol concentration is increased. A cholesterol molecule is too short in length to span the leaflet of a typical lipid bilayer, that is where the tails comprise of 16 or more carbon atoms, and as such, it sits buried amongst the phospholipid molecules, with the hydrophobic majority of cholesterol packed around the phospholipid tails (Figure 2iii-A). Its hydroxyl group instead nestles in amongst the carbonyl groups in the POPC (in this case) molecules allowing the POPC head groups to act as a molecular umbrella. The

area per cholesterol remains relatively stable in comparison to that of POPC, ranging from 0.284 nm^2 at 6.25% to 0.257 nm^2 at 31.25% (Figure 2iii-B), whilst the overall box area of the system decreases. This ultimately means that with a reduction in the number of phospholipids as the concentration of cholesterol increases, the larger relative decrease in area per lipid of POPC (0.631 nm^2 at 0% cholesterol to 0.532 nm^2 at 31.25% cholesterol) is primarily responsible for this change (Table 2A, Figure 2iii-C). This decrease occurs at a greater rate than expected when compared with ideal mixing [76] where the area per molecule is simply a weighted average of the area of the pure components. This is known as the condensation effect and was first observed almost 100 years ago [77].

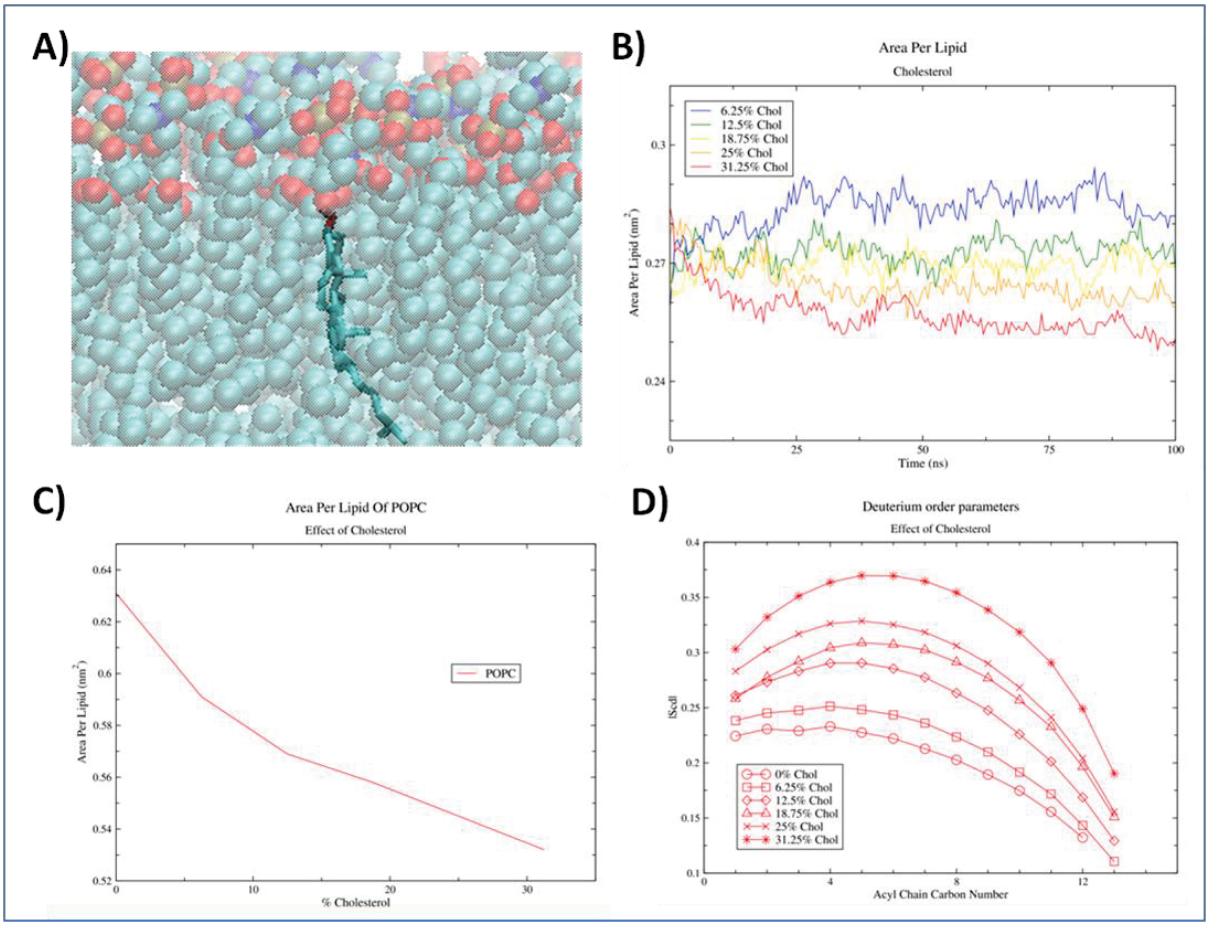


Figure 2iii - ***A)*** *Snapshot of the POPC/25% cholesterol system following 100 ns of simulation time, highlighting the relative position of the cholesterol molecule with the atoms coloured by name - C (cyan), N (blue), O (red), H (white).* ***B)*** *Area per cholesterol molecule in each of the binary POPC/cholesterol systems coloured from cold (6.25%) to hot (31.25%) as the concentration increases.* ***C)*** *Change in the average area per POPC molecule in each of the binary POPC/cholesterol systems as cholesterol concentration increases.* ***D)*** *Deuterium order parameters for the palmitoyl tail of the POPC lipids in each of the binary POPC/cholesterol systems.*

Cholesterol is seen to also thicken the bilayer (Table 2A), suggesting that the POPC tails are increasing in length and therefore becoming more ordered. This is also reflected in the deuterium order parameter ranges (Table 2A, Figure2iii-D), which become more and more ordered particularly towards the glycerol end of the acyl chain as cholesterol concentration increases. This is to be expected due to cholesterol packing itself in between neighbouring POPC fatty acid tails, and therefore results in a less fluid membrane.

The overall effects of cholesterol as outlined above on a lipid bilayer are in good agreement with previous studies done exploring various cholesterol concentrations [19, 78].

2.3.4 - Brain Lipid Bilayer

As the aim was to create a lipid bilayer representative of a brain cellular membrane, the properties of all combined lipid components also needed to be tested. Due to the complex nature of the brain lipid bilayer, comprising six different components, area per lipid was calculated using APL@voro (Section 2.2.5). Average area per lipid was calculated over the course of the 100 ns simulation with a tessellation being drawn every 0.5 ns. The average areas of each component were taken from the 200 tessellations and are estimated as follows: POPE 0.610 nm^2, POPC 0.595 nm^2, POPS 0.615 nm^2, PIP_2 0.645 nm^2, SGML 0.633 nm^2, and CHOL 0.283 nm^2 (Table 2B, Figure 2iv-A). The primary phospholipids tested in the single component systems (POPC, POPE, and POPS) all had similar areas in the brain lipid bilayer, approximately 0.6 nm^2, which is not reflective of their variance when tested in isolation – ranging from 0.50 nm^2 (POPE) to 0.63 nm^2 (POPC) (Table 2A). CHOL's area is effectively identical to that observed in the binary POPC/CHOL when cholesterol had a low concentration (Figure 2iii-B), whilst SGML's area of 0.633 nm^2 is considerably higher than the 0.55 nm^2 reported in the literature [47, 79, 80]. Little data exist in the way of area per lipid for PIP_2 but considering the nature of its headgroup, it can be expected to be larger than its phospholipid counterparts.

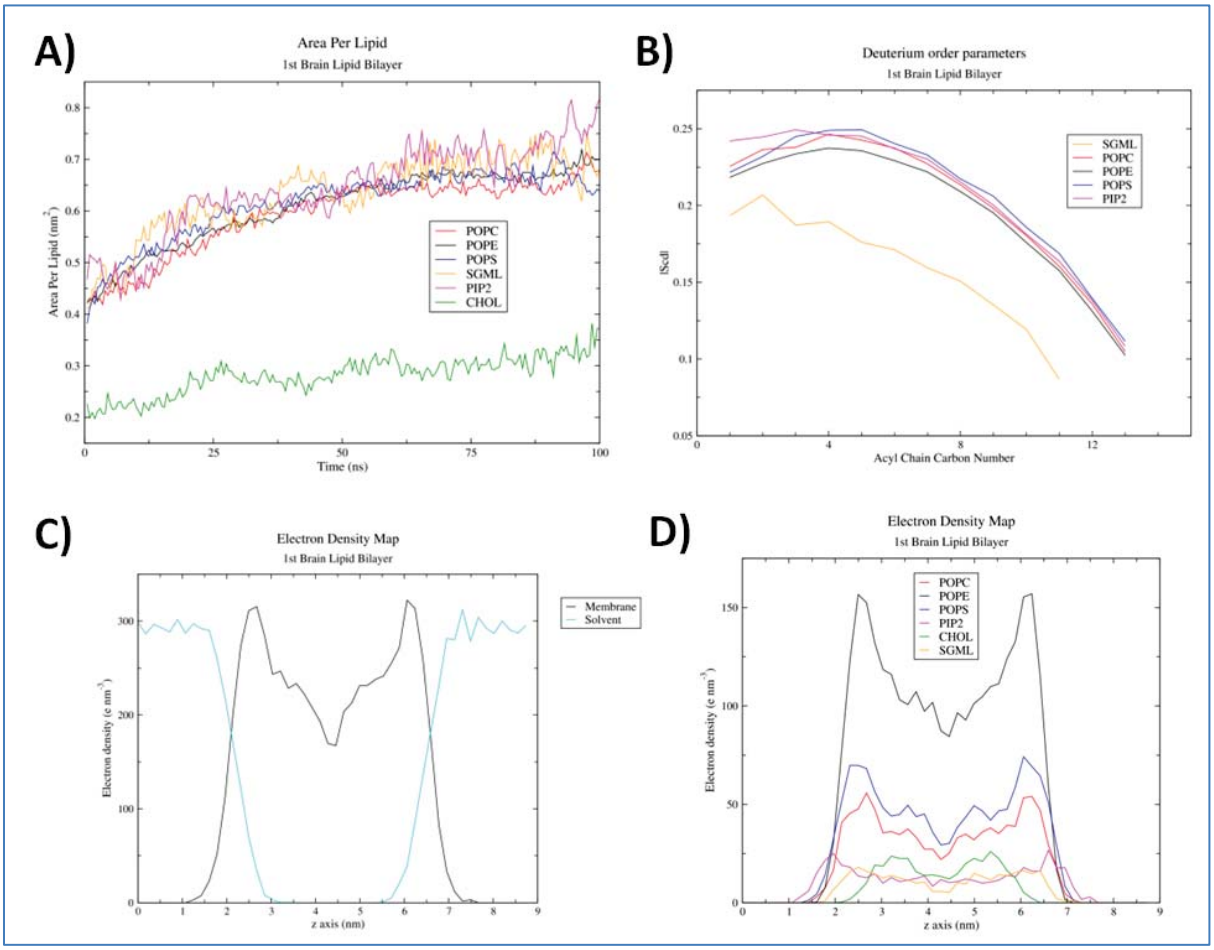


Figure 2iv - ***A)*** *Area per lipid for each component of the brain lipid bilayer - POPE (black), POPC (red), POPS (blue), PIP_2 (pink), SGML (orange), and CHOL (green).* ***B)*** *Deuterium order parameters of the saturated tails for the relevant components of the brain lipid bilayer.* ***C)*** *Electron density map for the complete brain lipid bilayer system including solvent.* ***D)*** *Electron density map for each component of the brain lipid bilayer. The z axis represents the length of the normal to the bilayer in both cases.*

The area per lipid values for individual lipids included in the APL@voro calculations were averaged across each time point and plotted as histograms showing their variation within each lipid type (Figure 2v). This allowed not only an observation of the area per lipid variance, but also the potential error range in the Voronoi tessellation's estimation of the areas. Considering the composition of the brain lipid bilayer, these errors will have the most effect on the area per lipid estimations of the most sparse lipid components, whilst the area per lipid estimation of POPE (the dominant component) will, and did, exhibit the most normal distribution. Despite the normality of most distributions, it is evident that the lipid areas, or APL@voro's estimate of those areas, fluctuate greatly. The smallest and largest lipids within each distribution were approximately 50% smaller and larger than the calculated means respectively.

In the case of all components but SGML, APL@voro overestimated the area of at least one molecule displayed as a clear outlier on each histogram. The mean was not always representative of the mode which provides two possibilities for the accepted areas of each component. This is particularly the case of SGML, which exhibited a bimodal distribution, suggesting its area may be calculated differently on the basis of its neighbouring lipids.

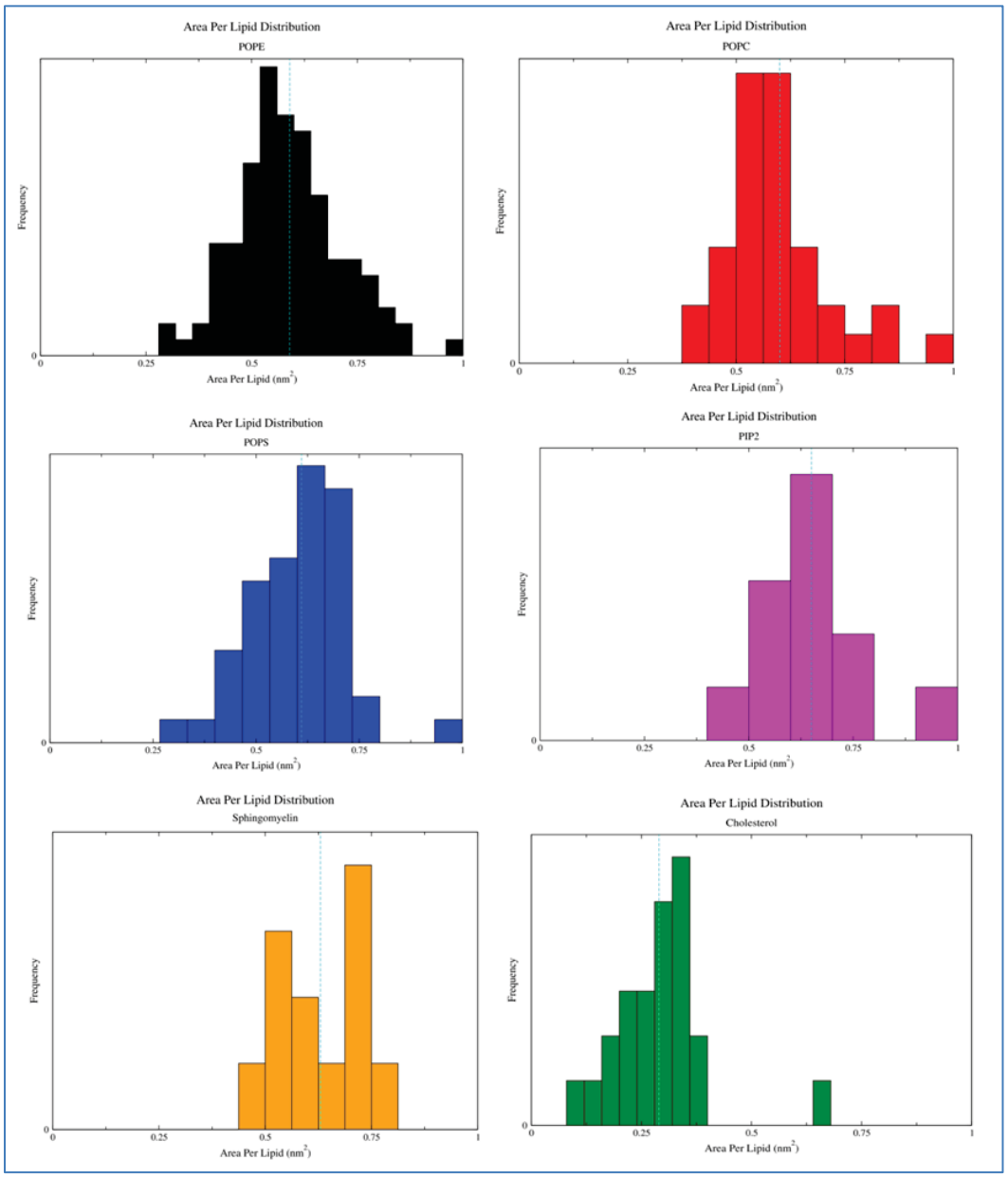


Figure 2v - *Distributions of the individual area per lipid values for each component chosen for calculation by APL@voro - POPE (black), POPC (red), POPS (blue), PIP_2 (pink), SGML (orange), and CHOL (green). In each case, the mean is indicated by a cyan line.*

A few things need to be taken into consideration when evaluating Voronoi tessellation areas. Firstly, it is known to overestimate the area of larger components in the system [63]. Secondly, it is dependent on the user's choice of key atom(s) around which to tessellate the plane. In the case of the area reported, the phosphate group was used for all lipids, whilst the hydroxyl group was used in the case of cholesterol. For the sake of comparison, when all atoms are selected across all lipids, the areas per lipid are estimated as follows: POPE 0.609 nm^2, POPC 0.612 nm^2, POPS 0.610 nm^2, PIP_2 0.680 nm^2, SGML 0.591 nm^2, and CHOL 0.269 nm^2. These values are closer to their experimentally determined areas, but are not so substantially different than the values obtained when a few key atoms are selected; with the latter method providing a considerable boost in calculation speed. Finally, the estimated areas depend on the location or "slice" of the membrane plane to be tessellated, with key atoms selected at the hydrophobic/hydrophilic interface of the bilayer doing the best job of not skewing the calculated areas of larger and smaller components of the membrane [17, 61, 63, 81].

With such a large system, areas per lipid need to be considered with a grain of salt as they are not the best indication of the lipid parameters reproducing experimental behaviour, due to their high sensitivity to simulation input parameters, especially those governing electrostatics [33]. It is also unknown how area per lipid might vary for such a complex mixture, as area properties varied even with the addition of cholesterol (Figure 2iii-C).

Volume per lipid is a more useful benchmark and also converges far more quickly than area per lipid [38]. While volume per lipid can be calculated in the same way as previously (as a function of area and half the thickness), the bilayer thickness of a mixed component membrane proves a trickier task than simply calculating its electron density. Due to each component having a different thickness, and therefore a different density profile, the overall membrane thickness is heavily biased towards whichever the most dominant component of the bilayer happens to be – in this case POPE. Membrane thickness as a whole was instead calculated as the difference between the two points on either side at which the electron density of the solvent exceeded that of the lipids.

The overall membrane thickness was found to be approximately 4.5 nm (Figure 2iv-C). Volume per lipid for each component was calculated using the thicknesses of the lipids from their individual electron density maps (Table 2B, Figure 2iv-D), which were found to be as follows: POPE 3.75 nm, POPC 3.59 nm, POPS 3.73 nm, PIP_2 4.63 nm, SGML 3.93 nm, and CHOL 2.13 nm. This, in turn, gives a volume per lipid for each component as follows: POPE 1.14 nm^3, POPC 1.07 nm^3, POPS 1.15 nm^3, PIP_2 1.50 nm^3, SGML 1.24 nm^3, and CHOL 0.30 nm^3. When looking back at the single lipid systems, it can now be seen that the exaggeration in phospholipid area has led to a larger than expected volume,

while the underestimation of the cholesterol area has led to a volume smaller than the 0.593 nm^3 calculated from the crystal structure.

While all calculations stemming from APL@voro must be considered cautiously, interpretation of the deuterium order parameters can be done with more confidence. It can be seen that all the palmitoyl tails of the phospholipid components exhibit disorder parameters in line with a bilayer in the liquid phase (Table 2B, Figure 2iv). The SGML tail is even more disordered, which is interesting considering suggestions it is expected to be found in lipid rafts along with cholesterol [12, 20, 82] - a molecule which was observed to increase order and is linked to the regulation of membrane fluidity [83, 84].

Table 2B

Physical properties (Section 2.3.1) of the lipids in the brain lipid bilayer systems

Physical Properties	PC	PE	PS	PIP_2	SGML	CHOL
APL (nm^2)						
a[1]	0.595	0.610	0.615	0.645	0.633	0.283
b[2]	0.586	0.543	0.558	0.577	0.558	0.260
c[3]	0.498	0.494	0.530	0.630	0.631	0.258
Thickness (nm)						
a	3.59	3.75	3.73	4.63	3.93	2.13
b	4.35	4.03	4.51	4.97	4.37	1.72
c	3.93	4.25	4.27	5.33	4.55	2.31
VPL (nm^3)						
a	1.07	1.14	1.15	1.50	1.24	0.30
b	1.27	1.09	1.26	1.43	1.22	0.22
c	0.98	1.05	1.13	1.68	1.44	0.30
DOP Range						
a	0.24 - 0.11	0.23 - 0.10	0.25 - 0.12	0.25 - 0.11	0.20 - 0.08	nc[4]
b	0.31 - 0.15	0.30 - 0.11	0.30 - 0.12	0.30 - 0.12	nc	nc
c	0.34 - 0.20	0.34 - 0.15	0.34 - 0.15	0.34 - 0.16	nc	nc

[1]- *The first brain lipid bilayer system is denoted by 'a'.*

[2]- *The CHARMM brain lipid bilayer system is denoted by 'b'.*

[3]- *The second brain lipid bilayer system is denoted by 'c'.*

[4]- *Uncalculated values (for lipids without saturated tails) are denoted by 'nc'.*

2.3.5 - CHARMM Lipid Bilayer

The same calculations were carried out within APL@voro for the CHARMM membrane builder bilayer, using the same key atoms for all components. The areas per lipid appear to have converged better in this simulation (Figure 2vi-A), with average estimates found to be as follows: SOPE 0.543 nm^2, POPC 0.586 nm^2, SOPS 0.558 nm^2, $SAPIP_2$ 0.577 nm^2, Neuronal SGML (NSGML) 0.558 nm^2, and

CHOL 0.260 nm^2 (Table 2B). These areas are in good agreement with previously reported experimental values across the board, with the exception of POPC which was roughly 6% smaller here.

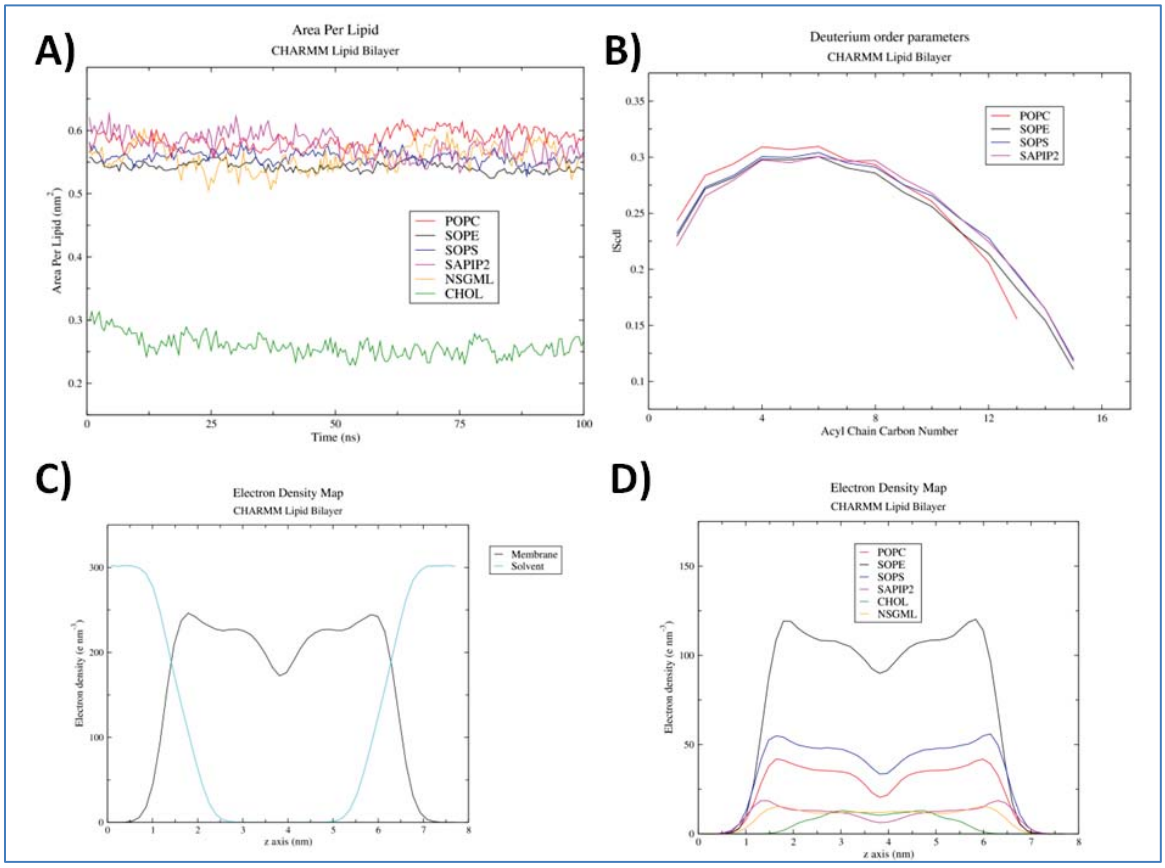


Figure 2vi - ***A)*** *Area per lipid for each component of the CHARMM lipid bilayer - SOPE (black), POPC (red), SOPS (blue), $SAPIP_2$ (pink), NSGML (orange), and CHOL (green).* ***B)*** *Deuterium order parameters of the saturated tails for the relevant components of the CHARMM lipid bilayer.* ***C)*** *Electron density map for the complete CHARMM lipid bilayer system including solvent.* ***D)*** *Electron density map for each component of the CHARMM lipid bilayer. The z axis represents the length of the normal to the bilayer in both cases.*

With an overall thickness of 4.86 nm and more ordered saturated tails (stearoyl and palmitoyl), it appeared to be less fluid than the first brain lipid bilayer, but remained within the range of DOP values expected for a fluid phase (Table 2B, Figure 2vi-B). Individual thicknesses for the components of SOPE 4.03 nm, POPC 4.35 nm, SOPS 4.51 nm, $SAPIP_2$ 4.97 nm, NSGML 4.37 nm, and CHOL 1.72 nm, gave rise to volumes of 1.09 nm^3, 1.27 nm^3, 1.26 nm^3, 1.43 nm^3, 1.22 nm^3, and 0.22 nm^3 respectively (Table 2B, Figure 2vi-D). The calculated volume for cholesterol is strikingly low which may once again be a function of APL@voro underestimating its area. As no part was taken in the parameterisation of these lipids, no conclusions will be made beyond acknowledging its good

agreement with experimental data, which makes this system ideal as a benchmark for comparison with the second brain lipid bilayer which incorporated the same tail configurations for each lipid component.

2.3.6 - Second Brain Lipid Bilayer

As the second brain lipid bilayer was also simulated using GROMOS style parameters, the analysis procedure was repeated allowing comparison with the CHARMM parameters. Areas per lipid were calculated as: SOPE 0.494 nm^2, POPC 0.498 nm^2, SOPS 0.530 nm^2, $SAPIP_2$ 0.630 nm^2, NSGML 0.631 nm^2, and CHOL 0.258 nm^2 (Table 2B, Figure 2vii-A). The values for SOPE and POPC were smaller than seen in the CHARMM lipid bilayer, while the values for $SAPIP_2$ and NSGML were larger.

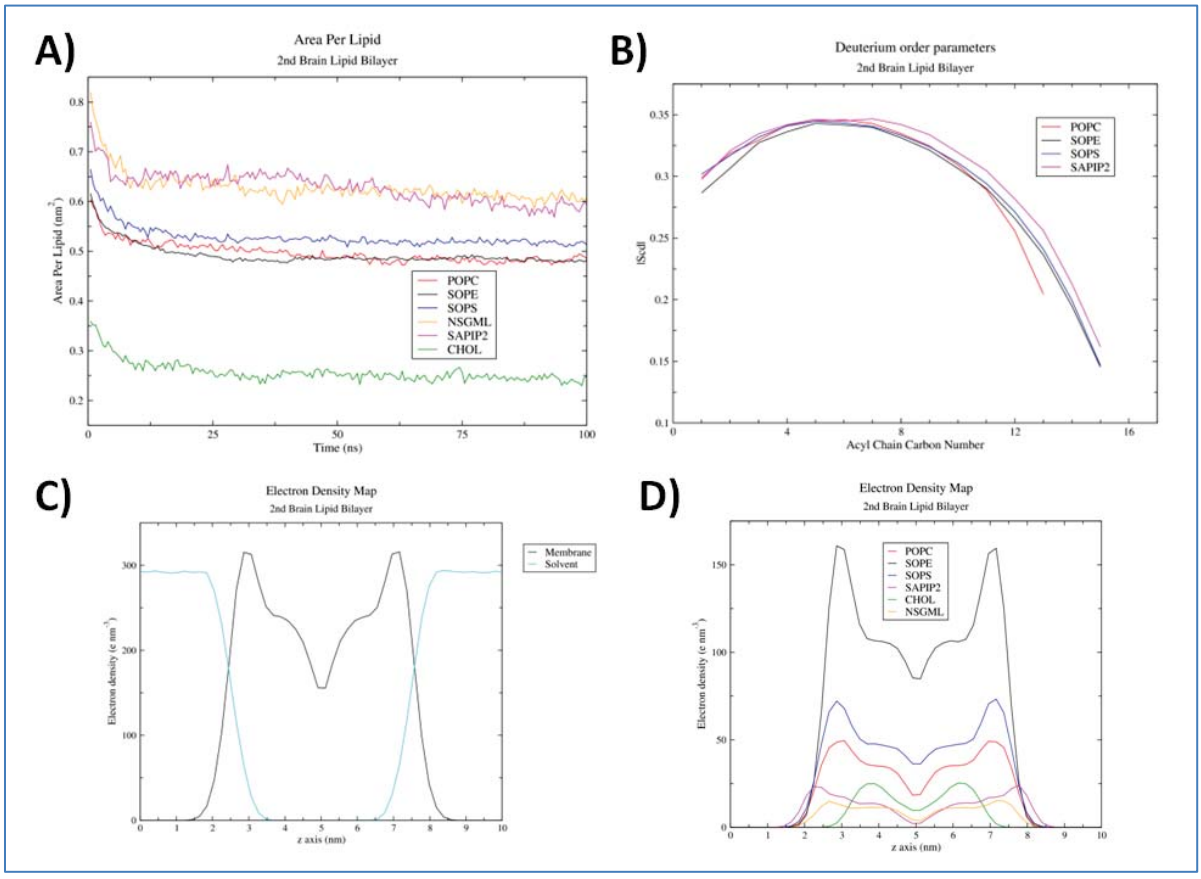


Figure 2vii - ***A)*** *Area per lipid for each component of the second brain lipid bilayer - SOPE (black), POPC (red), SOPS (blue), $SAPIP_2$ (pink), NSGML (orange), and CHOL (green).* ***B)*** *Deuterium order parameters of the saturated tails for the relevant components of the second brain lipid bilayer.* ***C)*** *Electron density map for the complete second brain lipid bilayer system including solvent.* ***D)*** *Electron density map for each component of the second brain lipid bilayer. The z axis represents the length of the normal to the bilayer in both cases.*

A larger membrane thickness of 5.11 nm and the slightly more ordered nature of the tails would suggest that this new brain lipid bilayer is once again less fluid than the CHARMM lipid bilayer (Table 2B, Figure 2vii-B). Individual thicknesses of: SOPE 4.25 nm, POPC 3.93 nm, SOPS 4.27 nm, $SAPIP_2$ 5.33, NSGML 4.55 nm, and CHOL 2.31 nm, gave rise to volumes per lipid of: SOPE 1.05 nm^3, POPC 0.98 nm^3, SOPS 1.13 nm^3, $SAPIP_2$ 1.68 nm^3, NSGML 1.44 nm^3, and CHOL 0.30 nm^3 (Table 2B, Figure 2vii-D).

Once again, values for SOPE and POPC were smaller than seen in the CHARMM bilayer, considerably in the case of POPC, and values for $SAPIP_2$ and NSGML were larger. However, considerably different simulation input parameters (Section 2.2.4) were used in the generation of both systems to reflect the force fields with which the lipids were parameterised. Along with the nature of the assembled parameters (Section 2.2.2) from different iterations of the GROMOS force fields, this could have had a significant effect on bilayer properties dependent on the version of Gromacs in which they were modelled [85]. Also, considering the parameters for POPC were untouched between the first brain lipid bilayer and this one, the contrast in results specific to this lipid are stark and underline how evaluation of single lipid properties within mixed lipid bilayers can affect results.

2.4 - Conclusion

The focus of this research was to construct a biologically realistic membrane, in this case a lipid bilayer representative of a brain membrane with a physiologically relevant lipid composition. This required the assimilation of coordinates and parameters for the various components - PC, PE, PS, PIP_2, SGML, and CHOL - compatible with the GROMOS 54A7 force field. The parameters were tested through simulation and analysis of small single lipid and binary systems, and they were validated through their ability to reproduce the structural and dynamic properties of previously analysed lipid systems, both computationally and experimentally.

Simulations of single lipid and binary systems showed the assimilated GROMOS 54A7 parameters to result in good reproduction of bilayer properties, lending confidence to their suitability for building a brain lipid bilayer. The construction of the first brain lipid bilayer was followed by the testing of its behaviour to observe that the lipid components could also reproduce structural and dynamic properties when simulated in a large mixed composition. Using the CHARMM GUI membrane builder, the realistic nature of the membrane was further improved in the construction of a second brain lipid bilayer, following modification of the lipid tails to those expected in such a composition.

Simulations of the mixed lipid systems also showed the assimilated GROMOS 54A7 parameters to result in good reproduction of bilayer properties. This allows for the use of this brain lipid bilayer in any molecular dynamics simulation where the system being observed occurs in neuronal cells at the membrane interface. Furthermore, these parameters can also be extended to the construction of lipid bilayers found in a range of cells through variation of their composition.

The importance of lipid composition has long been highlighted in membrane behaviour and plays a role in cellular signalling and lipid recognition in enzymatic studies. This ensures that any molecular dynamics study on a human system in which a cellular membrane plays a part can be as accurately represented as possible so as to derive new insight into the behaviour of the membrane itself and the proteins which interact with it.

2.5 - Bibliography

1. Korn, E.D., *Structure and Function of the Plasma Membrane : A biochemical perspective.* The Journal of General Physiology, 1968. **52**(1): p. 257-278.
2. Laude, A.J. and I.A. Prior, *Plasma membrane microdomains: organisation, function and trafficking.* Molecular membrane biology, 2004. **21**(3): p. 193-205.
3. Spector, A.A. and M.A. Yorek, *Membrane lipid composition and cellular function.* J Lipid Res, 1985. **26**(9): p. 1015-35.
4. Lande, M.B., J.M. Donovan, and M.L. Zeidel, *The relationship between membrane fluidity and permeabilities to water, solutes, ammonia, and protons.* J Gen Physiol, 1995. **106**(1): p. 67-84.
5. Lee, A.G., *How lipids affect the activities of integral membrane proteins.* Biochimica et Biophysica Acta (BBA) - Biomembranes, 2004. **1666**(1–2): p. 62-87.
6. McIntosh, T.J. and S.A. Simon, *Bilayers as Protein Solvents: Role of Bilayer Structure and Elastic Properties.* The Journal of General Physiology, 2007. **130**(2): p. 225-227.
7. Phillips, R., T. Ursell, P. Wiggins, and P. Sens, *Emerging roles for lipids in shaping membrane-protein function.* Nature, 2009. **459**(7245): p. 379-385.
8. Singer, S.J. and G.L. Nicolson, *The fluid mosaic model of the structure of cell membranes.* Science, 1972. **175**(4023): p. 720-31.
9. Lemmon, M.A., *Membrane recognition by phospholipid-binding domains.* Nat Rev Mol Cell Biol, 2008. **9**(2): p. 99-111.
10. Spector, A.A. and M.A. Yorek, *Membrane lipid composition and cellular function.* Journal of Lipid Research, 1985. **26**(9): p. 1015-35.
11. Lande, M.B., J. M. Donovan and M. L. Zeidel, *The relationship between membrane fluidity and permeabilities to water, solutes, ammonia, and protons.* The Journal of General Physiology, 1995. **106**(1): p. 67-84.
12. Stöckl, M.T. and A. Herrmann, *Detection of lipid domains in model and cell membranes by fluorescence lifetime imaging microscopy.* Biochimica et Biophysica Acta (BBA) - Biomembranes, 2010. **1798**(7): p. 1444-1456.
13. Ingólfsson, H.I., C. Arnarez, X. Periole, and S.J. Marrink, *Computational 'microscopy' of cellular membranes.* Journal of Cell Science, 2016.
14. Berger, O., O. Edholm, and F. Jähnig, *Molecular dynamics simulations of a fluid bilayer of dipalmitoylphosphatidylcholine at full hydration, constant pressure, and constant temperature.* Biophysical Journal, 1997. **72**(5): p. 2002-2013.
15. Moore, P.B., C.F. Lopez, and M.L. Klein, *Dynamical Properties of a Hydrated Lipid Bilayer from a Multinanosecond Molecular Dynamics Simulation.* Biophysical Journal. **81**(5): p. 2484-2494.
16. Feller, S.E., R.M. Venable, and R.W. Pastor, *Computer Simulation of a DPPC Phospholipid Bilayer: Structural Changes as a Function of Molecular Surface Area.* Langmuir, 1997. **13**(24): p. 6555-6561.
17. Pandit, S.A., D. Bostick, and M.L. Berkowitz, *Mixed Bilayer Containing Dipalmitoylphosphatidylcholine and Dipalmitoylphosphatidylserine: Lipid Complexation, Ion Binding, and Electrostatics.* Biophysical Journal, 2003. **85**(5): p. 3120-3131.
18. Janosi, L. and A.A. Gorfe, *Simulating POPC and POPC/POPG Bilayers: Conserved Packing and Altered Surface Reactivity.* Journal of Chemical Theory and Computation, 2010. **6**(10): p. 3267-3273.
19. Hofsäß, C., E. Lindahl, and O. Edholm, *Molecular Dynamics Simulations of Phospholipid Bilayers with Cholesterol.* Biophysical Journal, 2003. **84**(4): p. 2192-2206.
20. Berkowitz, M.L., *Detailed molecular dynamics simulations of model biological membranes containing cholesterol.* Biochimica et Biophysica Acta (BBA) - Biomembranes, 2009. **1788**(1): p. 86-96.

21. Wu, Emilia L., Patrick J. Fleming, Min S. Yeom, G. Widmalm, Jeffery B. Klauda, Karen G. Fleming, and W. Im, *E. coli Outer Membrane and Interactions with OmpLA.* Biophysical Journal. **106**(11): p. 2493-2502.
22. Bera, I. and J.B. Klauda, *Molecular Simulations of Mixed Lipid Bilayers with Sphingomyelin, Glycerophospholipids and Cholesterol.* The Journal of Physical Chemistry B, 2017.
23. van Meer, G., D.R. Voelker, and G.W. Feigenson, *Membrane lipids: where they are and how they behave.* Nature reviews. Molecular cell biology, 2008. **9**(2): p. 112-124.
24. Berglund, N.A., T.J. Piggot, D. Jefferies, R.B. Sessions, P.J. Bond, and S. Khalid, *Interaction of the Antimicrobial Peptide Polymyxin B1 with Both Membranes of E. coli: A Molecular Dynamics Study.* PLOS Computational Biology, 2015. **11**(4): p. e1004180.
25. Stansfeld, Phillip J. and Mark S.P. Sansom, *Molecular Simulation Approaches to Membrane Proteins.* Structure. **19**(11): p. 1562-1572.
26. Ortiz-Suarez, Maite L., F. Samsudin, Thomas J. Piggot, Peter J. Bond, and S. Khalid, *Full-Length OmpA: Structure, Function, and Membrane Interactions Predicted by Molecular Dynamics Simulations.* Biophysical Journal, 2016. **111**(8): p. 1692-1702.
27. Khalid, S., Nils A. Berglund, Daniel A. Holdbrook, Yuk M. Leung, and J. Parkin, *The membranes of Gram-negative bacteria: progress in molecular modelling and simulation.* Biochemical Society Transactions, 2015. **43**(2): p. 162-167.
28. Jo, S., J.B. Lim, J.B. Klauda, and W. Im, *CHARMM-GUI Membrane Builder for Mixed Bilayers and Its Application to Yeast Membranes.* Biophysical Journal, 2009. **97**(1): p. 50-58.
29. Monje-Galvan, V. and J.B. Klauda, *Modeling Yeast Organelle Membranes and How Lipid Diversity Influences Bilayer Properties.* Biochemistry, 2015. **54**(45): p. 6852-6861.
30. Shaw, D.E., R.O. Dror, J.K. Salmon, J.P. Grossman, K.M. Mackenzie, J.A. Bank, C. Young, M.M. Deneroff, B. Batson, K.J. Bowers, E. Chow, M.P. Eastwood, D.J. Ierardi, J.L. Klepeis, J.S. Kuskin, R.H. Larson, K. Lindorff-Larsen, P. Maragakis, M.A. Moraes, S. Piana, Y. Shan, and B. Towles, *Millisecond-scale molecular dynamics simulations on Anton*, in *Proceedings of the Conference on High Performance Computing Networking, Storage and Analysis*. 2009, ACM: Portland, Oregon. p. 1-11.
31. Shaw, D.E., J.P. Grossman, J.A. Bank, B. Batson, J.A. Butts, J.C. Chao, M.M. Deneroff, R.O. Dror, A. Even, C.H. Fenton, A. Forte, J. Gagliardo, G. Gill, B. Greskamp, C.R. Ho, D.J. Ierardi, L. Iserovich, J.S. Kuskin, R.H. Larson, T. Layman, L.S. Lee, A.K. Lerer, C. Li, D. Killebrew, K.M. Mackenzie, S.Y.H. Mok, M.A. Moraes, R. Mueller, L.J. Nociolo, J.L. Peticolas, T. Quan, D. Ramot, J.K. Salmon, D.P. Scarpazza, U.B. Schafer, N. Siddique, C.W. Snyder, J. Spengler, P.T.P. Tang, M. Theobald, H. Toma, B. Towles, B. Vitale, S.C. Wang, and C. Young. *Anton 2: Raising the Bar for Performance and Programmability in a Special-Purpose Molecular Dynamics Supercomputer*. in *SC14: International Conference for High Performance Computing, Networking, Storage and Analysis*. 2014.
32. Burke, J.E., O. Perisic, G.R. Masson, O. Vadas, and R.L. Williams, *Oncogenic mutations mimic and enhance dynamic events in the natural activation of phosphoinositide 3-kinase p110alpha (PIK3CA).* Proc Natl Acad Sci U S A, 2012. **109**(38): p. 15259-64.
33. Anézo, C., A.H. de Vries, H.-D. Höltje, D.P. Tieleman, and S.-J. Marrink, *Methodological Issues in Lipid Bilayer Simulations.* The Journal of Physical Chemistry B, 2003. **107**(35): p. 9424-9433.
34. Petrache, H.I., S. Tristram-Nagle, K. Gawrisch, D. Harries, V.A. Parsegian, and J.F. Nagle, *Structure and Fluctuations of Charged Phosphatidylserine Bilayers in the Absence of Salt.* Biophysical Journal, 2004. **86**(3): p. 1574-1586.
35. Stidder, B., G. Fragneto, and S.J. Roser, *Structure and stability of DPPE planar bilayers.* Soft Matter, 2007. **3**(2): p. 214-222.

36. Pan, J., F.A. Heberle, S. Tristram-Nagle, M. Szymanski, M. Koepfinger, J. Katsaras, and N. Kučerka, *Molecular structures of fluid phase phosphatidylglycerol bilayers as determined by small angle neutron and X-ray scattering.* Biochimica et biophysica acta, 2012. **1818**(9): p. 2135-2148.
37. Kučerka, N., B.W. Holland, C.G. Gray, B. Tomberli, and J. Katsaras, *Scattering Density Profile Model of POPG Bilayers As Determined by Molecular Dynamics Simulations and Small-Angle Neutron and X-ray Scattering Experiments.* The Journal of Physical Chemistry B, 2012. **116**(1): p. 232-239.
38. Poger, D., W.F. Van Gunsteren, and A.E. Mark, *A new force field for simulating phosphatidylcholine bilayers.* Journal of Computational Chemistry, 2010. **31**(6): p. 1117-1125.
39. Klauda, J.B., R.M. Venable, J.A. Freites, J.W. O'Connor, D.J. Tobias, C. Mondragon-Ramirez, I. Vorobyov, A.D. MacKerell, and R.W. Pastor, *Update of the CHARMM all-atom additive force field for lipids: Validation on six lipid types.* The journal of physical chemistry. B, 2010. **114**(23): p. 7830-7843.
40. Jämbeck, J.P.M. and A.P. Lyubartsev, *Derivation and Systematic Validation of a Refined All-Atom Force Field for Phosphatidylcholine Lipids.* The Journal of Physical Chemistry B, 2012. **116**(10): p. 3164-3179.
41. Dickson, C.J., B.D. Madej, Å.A. Skjevik, R.M. Betz, K. Teigen, I.R. Gould, and R.C. Walker, *Lipid14: The Amber Lipid Force Field.* Journal of Chemical Theory and Computation, 2014. **10**(2): p. 865-879.
42. Pronk, S., S. Pall, R. Schulz, P. Larsson, P. Bjelkmar, R. Apostolov, M.R. Shirts, J.C. Smith, P.M. Kasson, D. van der Spoel, B. Hess, and E. Lindahl, *GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit.* Bioinformatics, 2013. **29**(7): p. 845-54.
43. Schmid, N., A.P. Eichenberger, A. Choutko, S. Riniker, M. Winger, A.E. Mark, and W.F. van Gunsteren, *Definition and testing of the GROMOS force-field versions 54A7 and 54B7.* Eur Biophys J, 2011. **40**(7): p. 843-56.
44. Huang, J. and A.D. MacKerell, *CHARMM36 all-atom additive protein force field: Validation based on comparison to NMR data.* Journal of Computational Chemistry, 2013. **34**(25): p. 2135-2145.
45. Wennberg, C.L., D. van der Spoel, and J.S. Hub, *Large Influence of Cholesterol on Solute Partitioning into Lipid Membranes.* Journal of the American Chemical Society, 2012. **134**(11): p. 5351-5361.
46. Jämbeck, J.P.M. and A.P. Lyubartsev, *Another Piece of the Membrane Puzzle: Extending Slipids Further.* Journal of Chemical Theory and Computation, 2013. **9**(1): p. 774-784.
47. Chiu, S.W., S. Vasudevan, E. Jakobsson, R.J. Mashl, and H.L. Scott, *Structure of Sphingomyelin Bilayers: A Simulation Study.* Biophysical Journal, 2003. **85**(6): p. 3624-3636.
48. Holdbrook, D.A., Y.M. Leung, T.J. Piggot, P. Marius, P.T.F. Williamson, and S. Khalid, *Stability and Membrane Orientation of the Fukutin Transmembrane Domain: A Combined Multiscale Molecular Dynamics and Circular Dichroism Study.* Biochemistry, 2010. **49**(51): p. 10796-10802.
49. Chiu, S.-W., S.A. Pandit, H.L. Scott, and E. Jakobsson, *An Improved United Atom Force Field for Simulation of Mixed Lipid Bilayers.* The Journal of Physical Chemistry B, 2009. **113**(9): p. 2748-2763.
50. Berendsen, H.J.C., J.P.M. Postma, W.F. van Gunsteren, and J. Hermans, *Interaction Models for Water in Relation to Protein Hydration*, in *Intermolecular Forces*, B. Pullman, Editor. 1981, Springer Netherlands. p. 331-342.
51. Jorgensen, W.L., J. Chandrasekhar, J.D. Madura, R.W. Impey, and M.L. Klein, *Comparison of simple potential functions for simulating liquid water.* The Journal of Chemical Physics, 1983. **79**(2): p. 926-935.

52. Berendsen, H.J.C., J.P.M. Postma, W.F. van Gunsteren, A. DiNola, and J.R. Haak, *Molecular dynamics with coupling to an external bath.* The Journal of Chemical Physics, 1984. **81**(8): p. 3684-3690.
53. Silvius, J.R. *Thermotropic Phase Transitions of Pure Lipids in Model Membranes and Their Modifications by Membrane Proteins*. Lipid-Protein Interactions 1982; Available from: https://avantilipids.com/tech-support/physical-properties/phase-transition-temps/.
54. Hess, B., H. Bekker, H.J.C. Berendsen, and J.G.E.M. Fraaije, *LINCS: A linear constraint solver for molecular simulations.* Journal of Computational Chemistry, 1997. **18**(12): p. 1463-1472.
55. Essmann, U., L. Perera, M.L. Berkowitz, T. Darden, H. Lee, and L.G. Pedersen, *A smooth particle mesh Ewald method.* The Journal of Chemical Physics, 1995. **103**(19): p. 8577-8593.
56. Páll, S. and B. Hess, *A flexible algorithm for calculating pair interactions on SIMD architectures.* Computer Physics Communications, 2013. **184**(12): p. 2641-2650.
57. Nose, S., *A unified formulation of the constant temperature molecular dynamics methods.* The Journal of Chemical Physics, 1984. **81**(1): p. 511-519.
58. Hoover, W.G., *Canonical dynamics: Equilibrium phase-space distributions.* Physical Review A, 1985. **31**(3): p. 1695-1697.
59. Parrinello, M. and A. Rahman, *Strain fluctuations and elastic constants.* The Journal of Chemical Physics, 1982. **76**(5): p. 2662-2666.
60. Lukat, G., J. Krüger, and B. Sommer, *APL@Voro: A Voronoi-Based Membrane Analysis Tool for GROMACS Trajectories.* Journal of Chemical Information and Modeling, 2013. **53**(11): p. 2908-2925.
61. Edholm, O. and J.F. Nagle, *Areas of Molecules in Membranes Consisting of Mixtures.* Biophysical Journal, 2005. **89**(3): p. 1827-1832.
62. Piggot, T.J., Á. Piñeiro, and S. Khalid, *Correction to Molecular Dynamics Simulations of Phosphatidylcholine Membranes: A Comparative Force Field Study.* Journal of Chemical Theory and Computation, 2017. **13**(4): p. 1862-1865.
63. Pandit, S.A., S. Vasudevan, S.W. Chiu, R.J. Mashl, E. Jakobsson, and H.L. Scott, *Sphingomyelin-Cholesterol Domains in Phospholipid Membranes: Atomistic Simulation.* Biophysical Journal, 2004. **87**(2): p. 1092-1100.
64. Nagle, J.F. and M.C. Wiener, *Structure of fully hydrated bilayer dispersions.* Biochimica et Biophysica Acta (BBA) - Biomembranes, 1988. **942**(1): p. 1-10.
65. Nagle, J.F., R. Zhang, S. Tristram-Nagle, W. Sun, H.I. Petrache, and R.M. Suter, *X-ray structure determination of fully hydrated L alpha phase dipalmitoylphosphatidylcholine bilayers.* Biophysical Journal, 1996. **70**(3): p. 1419-1431.
66. Kučerka, N., J.F. Nagle, J.N. Sachs, S.E. Feller, J. Pencer, A. Jackson, and J. Katsaras, *Lipid Bilayer Structure Determined by the Simultaneous Analysis of Neutron and X-Ray Scattering Data.* Biophysical Journal, 2008. **95**(5): p. 2356-2367.
67. Heller, H., M. Schaefer, and K. Schulten, *Molecular dynamics simulation of a bilayer of 200 lipids in the gel and in the liquid crystal phase.* The Journal of Physical Chemistry, 1993. **97**(31): p. 8343-8360.
68. Poger, D. and A.E. Mark, *On the Validation of Molecular Dynamics Simulations of Saturated and cis-Monounsaturated Phosphatidylcholine Lipid Bilayers: A Comparison with Experiment.* Journal of Chemical Theory and Computation, 2010. **6**(1): p. 325-336.
69. Rand, R.P. and V.A. Parsegian, *Hydration forces between phospholipid bilayers.* Biochimica et Biophysica Acta (BBA) - Reviews on Biomembranes, 1989. **988**(3): p. 351-376.
70. Jämbeck, J.P.M. and A.P. Lyubartsev, *An Extension and Further Validation of an All-Atomistic Force Field for Biological Membranes.* Journal of Chemical Theory and Computation, 2012. **8**(8): p. 2938-2948.
71. Mukhopadhyay, P., L. Monticelli, and D.P. Tieleman, *Molecular Dynamics Simulation of a Palmitoyl-Oleoyl Phosphatidylserine Bilayer with Na(+) Counterions and NaCl.* Biophysical Journal, 2004. **86**(3): p. 1601-1609.

72. Jurkiewicz, P., L. Cwiklik, A. Vojtíšková, P. Jungwirth, and M. Hof, *Structure, dynamics, and hydration of POPC/POPS bilayers suspended in NaCl, KCl, and CsCl solutions.* Biochimica et Biophysica Acta (BBA) - Biomembranes, 2012. **1818**(3): p. 609-616.
73. Kaiser, H.-J., D. Lingwood, I. Levental, J.L. Sampaio, L. Kalvodova, L. Rajendran, and K. Simons, *Order of lipid phases in model and plasma membranes.* Proc Natl Acad Sci U S A, 2009. **106**(39): p. 16645-16650.
74. Douliez, J.P., A. Léonard, and E.J. Dufourc, *Restatement of order parameters in biomembranes: calculation of C-C bond order parameters from C-D quadrupolar splittings.* Biophysical Journal, 1995. **68**(5): p. 1727-1739.
75. Petrache, H.I., S.W. Dodd, and M.F. Brown, *Area per lipid and acyl length distributions in fluid phosphatidylcholines determined by (2)H NMR spectroscopy.* Biophysical Journal, 2000. **79**(6): p. 3172-3192.
76. McConnell, H.M. and A. Radhakrishnan, *Condensed complexes of cholesterol and phospholipids.* Biochimica et Biophysica Acta (BBA) - Biomembranes, 2003. **1610**(2): p. 159-173.
77. Leathes, J.B., *Croonian Lectures ON THE RÔLE OF FATS IN VITAL PHENOMENA.* The Lancet, 1925. **205**(5303): p. 803-807.
78. de Meyer, F. and B. Smit, *Effect of cholesterol on the structure of a phospholipid bilayer.* Proceedings of the National Academy of Sciences, 2009. **106**(10): p. 3654-3658.
79. Maulik, P.R. and G.G. Shipley, *N-Palmitoyl Sphingomyelin Bilayers: Structure and Interactions with Cholesterol and Dipalmitoylphosphatidylcholine.* Biochemistry, 1996. **35**(24): p. 8025-8034.
80. Li, X.M., J.M. Smaby, M.M. Momsen, H.L. Brockman, and R.E. Brown, *Sphingomyelin interfacial behavior: the impact of changing acyl chain composition.* Biophysical Journal, 2000. **78**(4): p. 1921-1931.
81. Falck, E., M. Patra, M. Karttunen, M.T. Hyvonen, and I. Vattulainen, *Lessons of slicing membranes: interplay of packing, free area, and lateral diffusion in phospholipid/cholesterol bilayers.* Biophys J, 2004. **87**(2): p. 1076-91.
82. Silvius, J.R., *Role of cholesterol in lipid raft formation: lessons from lipid model systems.* Biochimica et Biophysica Acta (BBA) - Biomembranes, 2003. **1610**(2): p. 174-183.
83. Cooper, R.A., *Influence of increased membrane cholesterol on membrane fluidity and cell function in human red blood cells.* Journal of Supramolecular Structure, 1978. **8**(4): p. 413-430.
84. Krause, M.R. and S.L. Regen, *The Structural Role of Cholesterol in Cell Membranes: From Condensed Bilayers to Lipid Rafts.* Accounts of Chemical Research, 2014. **47**(12): p. 3512-3521.
85. Reisser, S., D. Poger, M. Stroet, and A.E. Mark, *The Real Cost of Speed: The Effect of a Time-Saving Multiple-Time-Stepping Algorithm on the Accuracy of Molecular Dynamics Simulations.* Journal of Chemical Theory and Computation, 2017.

Chapter 3 - Protein-Membrane Interactions

3.1 - Literature Review

3.1.1 - Overview

As outlined in Chapter 1, a protein's function is linked to its structure, which in turn is determined by its sequence. Mutation of this sequence that leads to a structural change can cause an alteration in the protein's function. Enzymes and signalling proteins are of utmost importance in the body when it comes to regulation of physiological behaviour through various biological pathways. As such, when a mutation in these proteins alters their function, this will more often than not negatively impact a cascade of activities, resulting in disease. Both enzymes and signalling proteins commonly interact with cell membranes, with this sub group appropriately being named membrane proteins. Simply put, membrane proteins are proteins which interact with cell lipid bilayers found in the body. These interactions can either be integral, involving membrane proteins which are permanently anchored within the bilayer, or peripheral, involving membrane proteins which are only temporarily bound to the bilayer or even to an integral membrane protein. This chapter focuses on the development of a rotational interaction energy screening method able to rapidly predict the orientation in which peripheral membrane proteins will exhibit transient complexes with cell lipid bilayers. The method incorporates non-bonded interactions described in molecular dynamics (Section 3.2) to determine the most favourable interface between a membrane protein and a bilayer of physiologically relevant lipid composition developed in Chapter 2.

3.1.2 - Peripheral Membrane Proteins

Peripheral membrane proteins commonly fall under two classes of proteins – enzymes and signalling proteins, but play a part in many other biological processes such as vesicle formation. Enzymes are highly specific proteins which catalyse various chemical reactions around the cell, ranging from metabolism to deoxyribonucleic acid (DNA) transcription. Signalling proteins generally collaborate with hormones and regulate physiological behaviour and biological pathways around the body. Considering the role of peripheral membrane proteins in the human body, it should come as no surprise that a large number of them are drug targets [1, 2], with half of all known small molecule drugs binding to membrane proteins [3]. It is important therefore to characterise how they interact with the membrane, in terms of orientation, a mode of action and relative location of drug binding sites, as well as how these factors differ in wild type functional proteins compared to any disease-linked mutated variants.

3.1.3 - Protein-Membrane Binding

Peripheral membrane proteins bind to the cell lipid bilayer in different ways. There can be loops or helices, short, long or bent, embedded at various depths into the membrane or even lying flat on its surface [4, 5]. The residues which interact with the cell lipid bilayer can be characterised as membrane binding or targeting domains, and these differ across families of membrane proteins [6, 7]. Some examples of membrane binding or targeting domains include C1 (conserved region 1) domains [8], C2 (conserved region 2) domains [9], PX (phox homology) domains [10] and PH (pleckstrin homology) domains [11].

Transient protein membrane interactions can broadly be broken down into VDW interactions and electrostatic interactions, with the former describing membrane anchors embedded into the hydrophobic region of the membrane, and the latter describing surface to surface interactions involving the hydrophilic region of the membrane [12]. Electrostatic interactions can either be described as non-specific, generally long range, or specific, generally short range; where the favourability of these interactions determine the likelihood of membrane binding occurring at a particular site [13, 14].

For these interactions to lead to protein function there are two groups which must work in tandem - amino acids from the protein and lipids from the membrane. Just as the protein sequence defines its structure and function, so will the lipid composition of the membrane correlate with its location in the body and thus its function [15, 16]. Most signalling proteins bind site specifically to cellular membrane regions which are the points of recognition to initiate the pathways these proteins regulate. The acidic phospholipids introduced in Chapter 2, phosphatidylserine (PS) and phosphatidylinositide (PI) are in almost all cases the binding targets of signalling proteins; however, the domains and mechanisms which bind them vary greatly [17].

PIs are one of the least common phospholipids found in human cell membranes, but are one of the major cell signalling recognition sites [18]. They have a further level of differentiation on the basis of phosphorylation at the 3-, 4-, and 5- positions of the inositol ring, and ratios between these phosphorylated lipids within a cellular compartment act as another control mechanism for signalling pathways [19]. PIP_2 in particular serves as an important target in processes such as exocytosis, endocytosis, membrane trafficking, and enzyme activation [20]. Despite the low level of PIP_2 in cellular membranes, approximately 0.05% [17], these phosphoinositides are localised in clusters within the cellular membranes around the cell, allowing for the proteins which target them to carry out their functions at specific sites [21].

Due to the small percentage of PIs within the cell, it can be expected that signalling proteins which target them have a high specificity, while signalling proteins which target PS and other anionic lipids have a rather low specificity due to the more abundant nature of the lipid (8.5%) [17]. Other mechanisms are involved in the regulation of PS targeting proteins, such as those which include C2 domains. These domains commonly rely on calcium levels to determine when membrane binding occurs [22-24]; however, not all proteins that include C2 domains rely on calcium ions to regulate their binding [9].

3.1.4 - Protein-Membrane Association

There are two parts to membrane protein interaction - membrane association, and membrane binding. Membrane association is usually achieved through non-specific long range interactions [25], whereas membrane binding is achieved through specific short range interactions, for instance, through signalling targets in the membrane bound by motifs within the protein sequence [26]. Research suggests that electrostatic interactions guide the protein towards its initial membrane docking configuration, at which point hydrophobic embedding stabilises the membrane interaction [17, 24, 27]. The membrane association is a key part of not only translocation of the protein to the correct region of the cellular membrane, but also allowing for membrane binding in the correct orientation to achieve the protein's function. Even proteins within the same family often use different methods for membrane binding [28, 29], further increasing specificity.

When considering the behaviour of disease-linked malfunctioning membrane proteins, inhibition of their function can therefore be achieved in a multitude of ways - disrupting the favourable electrostatic protein-membrane interface, removing or altering the target molecule found within the membrane, or disrupting the specific sequence motifs or recognition sites within the protein which allow it to bind. In drug design, the first two methods could have dire implications on a multitude of protein functions beyond the target, and as such, the third option stands out as the most target-specific and logistically feasible method of inhibition.

3.1.5 - Importance of Lipid Composition

Considering the targets that signalling proteins rely on for their function, the lipid composition of cellular membranes is of high importance, and can influence the local structure, dynamics and activity of membrane proteins [30-32]. As outlined in Chapter 2, in times past, computational studies commonly profiled membranes as monolipid bilayers involving only DPPC, for example, evolving to bilayers with two kinds of lipid [33-35]. However, if the aim is to observe the membrane targeting of a protein which binds to a particular phospholipid, then it's only logical that the phospholipid be

included in the system. Studies with simplified membranes will be unable to follow the membrane targeting of signalling proteins which rely on specific lipid compositions to be functional.

Another shortcoming of many existing computational studies is the use of coarse-grained simulations to characterise system components [36-38]. These simulations eliminate the atomistic detail of the components within the system; most importantly, those interactions with each other and with the protein residue side chains and membrane lipid headgroups.

As explained in Chapter 1, there are two major limitations to molecular dynamics, namely a lack of parameters and a lack of computing power, both of which restrict the potential for longer time scale simulations of complex systems. For a biologically representative cellular membrane to be simulated, not only do the individual lipids need to be parameterised within a force field, but they also need to be compatible with each other. While there will always be limitations, this study seeks to exploit novel lipid parameters previously assembled [39] and tested in Chapter 2, and the advent of more powerful computing resources to partially overcome these limitations.

3.1.6 - Limitations of Experimental Assays

In vivo studies have shown the importance of both hydrophobic embedment into the bilayer and basic residues binding to the lipid headgroups as crucial to the function of membrane proteins [24, 27, 40]. Equally important is the lipid composition, with fluctuations in acidic phospholipid levels resulting in drastic differences in membrane binding due to the electrostatic favourability of the interactions with basic residue clusters on the protein surface [41]. Disruption of this protein-membrane interaction is essentially a complete neutralisation of its function, and therefore it is important to be able to accurately measure any effect on protein-membrane interaction pre- and post-disruption. To do this experimentally, a membrane binding assay has to be developed with a means of detecting interaction, which incorporates the native protein capable of its function as well as a realistic composition of the membrane it is expected to bind to, the site of its target [42].

A range of assays exists including cosedimentation and coflotation assays, surface plasmon resonance, and various fluorescence spectroscopy methods.

In cosedimentation assays, the concentration of one of the components needs to be varied to excess in an effort to calculate the dissociation constant between the protein and the membrane [43]. It makes more sense to vary the lipid concentration due to the cost-efficiency of obtaining it in abundance in comparison to a purified protein. However, only an approximate value for the dissociation constant can be obtained due to the multivalent nature of protein-membrane binding [42].

Surface plasmon resonance can be used to measure lipid specificity of protein-membrane interactions through the alteration of lipid composition [44]; however, signals are approximate and cannot always be pinpointed to the protein's interaction with a specific type of lipid.

Fluorescence resonance energy transfer assays are commonly used to look at protein-membrane binding as they provide additional insight into the dynamic nature of the interactions. Tryptophan fluorescence, in the case of the protein, can be used to determine the region of the protein undergoing binding due to the fluorescence of the tryptophan changing as it moves into a more hydrophobic environment in the membrane than in the solvent [45]. Additionally, lipid labelling in either the headgroup or the tails [46, 47] can be used to determine the extent and the depth to which the protein inserts itself into the membrane [48]. However, these fluorescence assays are heavily reliant on the tagged amino acid and lipid being close enough for fluorescence transfer to take place, and on their emission and excitation spectra overlapping [49]. As such, they are not always guaranteed to work, and fail to give an accurate distance or quenching measurement for more than one specific pair of sites per assay.

3.1.7 - Computational Evaluation of Protein-Membrane Interactions

To achieve maximum insight into the behaviour of membrane proteins as it relates to their function, protein-membrane interactions need to be characterised on a residue level and lipid level. As outlined above, experimental methods to test these interactions not only require successful purification of the protein in question, but also development of a sensitive and specific lipid binding assay. Alterations in the protein sequence or lipid composition are money and time consuming, and are more easily achieved computationally, with the only limitation being initial parameterisation. Due to the atomistic level of detail in computational representation, the characterisation of membrane-protein affinity on a residue and lipid specific level becomes a realistic goal.

In the past, membrane and solvent properties in protein-membrane screening methods have been described in very basic and broad terms due to the restraints of computing power and parameterisation of single components within them. Membranes have evolved from smeared-charge models which permeate a negative charge through a slab representative of a bilayer in an attempt to describe the effect of the distribution of acidic phospholipids across the membrane [50]. In combination with a smeared-charge model of a peptide with a distribution of a positive charge across the surface due to the effect of the distribution of basic residues, this gives a computationally cheap method of evaluating protein membrane interactions. However, this fails to describe clustering of lipids and/or structural motifs within membranes and proteins, allowing for specific recognition and binding sites.

Computational studies have further evolved to pinpoint the necessity for observation of proteins on a molecular (residue) level to determine their effect on a charged membrane [22]. Methods such as the Poisson-Boltzmann equation [51] were developed to describe the electrostatic properties of biological macromolecules using a finite difference approximation for a static structure [52, 53]. Compromises for computational efficiency are also made elsewhere, describing water as a dielectric constant allowing for the description of long-range electrostatics but sacrificing molecular detail and protein dynamics in the process. Whilst able to determine a putative protein-membrane interface, this method negates the possibility for any short-range specific interactions between residues and lipids to be characterised. Lipid composition was not given nearly as much focus as the protein, choosing instead to describe membranes using atomistic models of one or two lipids [54].

3.1.8 - Scope of This Research

As discussed in Chapter 2, if the hope remains to visualise and characterise molecular interactions accurately, then the system which incorporates these interactions must also have accurately described components. Phospholipid composition and concentration, and membrane properties help determine the binding site and function of membrane proteins. Using an atomistic representation of a lipid bilayer with a physiologically relevant composition, as well as atomistic representations of protein and water, this chapter presents a method of rapidly evaluating and characterising protein-membrane interactions. This method seeks to determine the correct orientation of peripheral membrane proteins with respect to the membrane on the basis of the electrostatic and VDW energies between them. The atomistic detail further allows identification of individual residues and lipids important to the process, leading to the development of experimentally testable hypotheses behind protein-membrane interactions.

3.2 - Non-Bonded Interaction Energy Terms

The goal of this work was to develop a method for rapidly evaluating how the interaction potential between a protein and a membrane varies with respect to the orientation of the protein relative to the membrane, and the membrane composition. While there exist several approximate methods for doing so, as outlined above, the goal here was to replicate the protein-membrane interactions that exist during a molecular dynamics simulation; therefore the protein-membrane interactions are evaluated using the non-bonded interaction energy terms from the same force field (GROMOS 54A7 [55]) as was used in the simulations.

As outlined in Chapter 1, in an MD simulation, the protein-membrane interactions are described by the non-bonded terms of the force field, namely the Lennard-Jones (van der Waals) and Coulombic (electrostatic) potentials. These terms are evaluated for pairs of atoms in the protein and membrane within a cut-off distance. For the GROMOS force fields and simulation setup used here, the Lennard-Jones interactions are set to zero outside the cut-off distance, and the Coulombic interactions are treated with the particle-mesh Ewald (PME) [56] approximation.

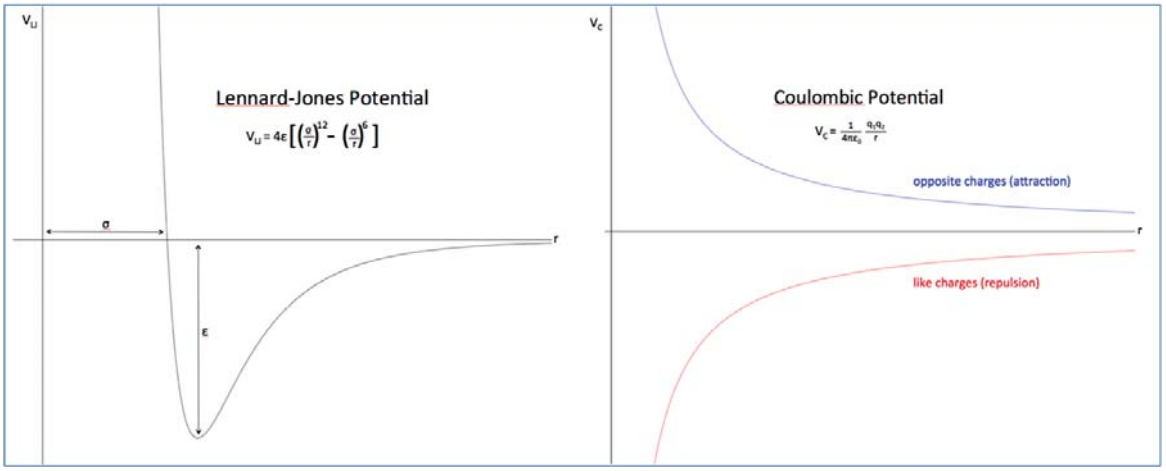


Figure 3i - *The Lennard-Jones (V_{LJ}) and Coulombic (V_C) potentials. The majority of Lennard-Jones energy terms presented in the interaction energy screening are at a distance (r) greater than the minimum and as such will always be in the favourable region (below zero) of the potential. The nature of the Coulombic energies depends on whether the signs of the partial charges of the atoms involved are the same or opposite.*

At distances greater than σ and less than the cut-off, all van der Waals (VDW) interactions, described by the Lennard-Jones potential (Figure 3i), are attractive to some degree, with a minimum (most attractive) interaction energy of ε at an optimal atom-atom distance. As the protein and membrane get closer together, the Lennard-Jones potential ascends steeply, making interactions repulsive for the most part at very short atom-atom distances and therefore preventing steric overlap of atoms.

Coulomb's constant ($\frac{1}{4\pi\varepsilon_0}$) describes the electrostatic interaction energy between two charged atoms (q_1 and q_2). These include polar atoms, which are assigned partial charges in atomic-level force fields. These interactions are favourable when unlike charges come close together, and vice versa. Coulombic interactions are in general of higher magnitude than VDW interactions, and remain significant over larger atom-atom distances due to the 1/*r* scaling (Figure 3i). At very short atom-atom distances, however, the repulsive part of the Lennard-Jones term is much greater than the Coulombic attraction between unlike charges, thus preventing steric overlap. The Coulombic interactions are therefore expected to dominate the protein-membrane interaction at large distances, whereas at shorter distances, a favourable Lennard-Jones interaction could also result if there is a large protein-membrane interaction surface.

3.3 - Method Development

3.3.1 - Test System Methods

A test system was used to understand the balance between the relative contributions to the non-bonded interaction energy and how they vary with the distance between the interacting components. The GROMACS 4 simulation [57] software separates the short- and long-range contributions to the Coulombic and Lennard-Jones (LJ) interactions; the long-range Coulombic contribution comes from the PME approximation, so was not included in this testing.

A series of systems described by the GROMOS 54A7 force field [55] was constructed including a slab of 817 decane molecules and a single methane molecule translated to minimum distances of 0.5 nm, 1 nm, 1.5 nm, and 2 nm away from the decane (Figure 3ii). The systems were separated into two groups, where the methane molecule was either given a small negative charge of -0.2, or left uncharged. The decane slab was given a positive charge of 1 spread evenly across the first and tenth carbons, or left uncharged, respectively. These two groups of systems were further separated on the basis of their input files which either had single range cut-offs for both Coulombic and LJ calculations, or twin range cut-offs for the LJ calculations but not the Coulombic calculations. In a twin-range cut-off system, energies within the first cut-off are calculated every step, whilst energies between the first and second cut-off are evaluated less often. The GROMOS force fields are designed to be used with a twin-range cut-off, but are often used with only a single-range cut-off. The two possibilities were therefore tested to observe how GROMACS described LJ terms with a twin range cut-off and how it affected the overall LJ value.

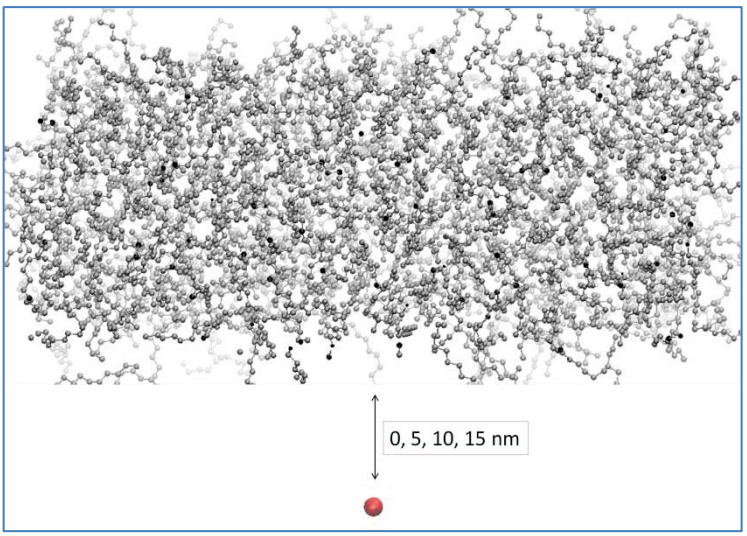


Figure 3ii - *Illustration of the test system developed to evaluate the calculation of non-bonded interactions in GROMACS 4. The system features a decane slab (gray) and a methane molecule (red) placed at varying distances away from each other, with varying charges placed on each component.*

All systems were solvated with single point charge (SPC) water [58], with the 0.5 nm system being duplicated with no solvent to explore the effect a polar solvent had on the Coulombic calculations. Harmonic position restraints with a force constant of 1000 $kJ.mol^{-1}$ were applied to both the methane molecule and the decane slab. Each system was subjected to 1000 steps of steepest descent energy minimisation before undergoing a short 50 fs MD simulation from which the energies averaged over the course of the simulation were calculated.

The simulations were initiated with the following scheme. Firstly, the initial velocities were randomly generated from a Maxwell-Boltzmann distribution at 50 K in the NPT ensemble. The temperature was controlled using the Berendsen thermostat [59] with a temperature coupling constant (τ_T) of 0.1 ps. The pressure was controlled using the Berendsen barostat with a pressure coupling constant (τ_p) of 0.5 ps and an isothermal compressibility of 4.5×10^{-5} bar^{-1} in isotropic conditions. Structures were saved every 10 fs with the LINCS algorithm [60] used with an order of 4 to constrain bond lengths and water bond angles, allowing for an integration time step of 2 fs. Non-bonded interactions were calculated using a grid cut-off scheme. For the single range cut-off scheme, LJ interactions were calculated within a cut-off distance of 1 nm at every step from a pair list that was updated every fifth time step. For the twin range cut-off scheme, LJ interactions within a cut-off distance of 0.9 nm were calculated at every step from a pair list that was updated every fifth time step, at which point, interactions between atoms within 1.4 nm were also calculated and were kept constant between updates. Electrostatic interactions were calculated using particle mesh Ewald (PME) summation [56], with a cut-off distance of 1 nm for the single range cut-off LJ systems, and 0.9 nm for the twin range cut-off LJ systems. The LJ energies in the twin range systems were separated into short-range (S-R) and long-range (L-R) interactions.

Following the short MD simulation, the minimum distance between the methane molecule and the decane slab was recalculated to ensure the energies were being evaluated at the expected distance.

3.3.2 - Test System Results

Table 3A

Results for uncharged system featuring a twin range LJ cut-off scheme

rlist = 0.9 ; rcoulomb = 0.9 ; rvdw = 1.4 (nm) Uncharged Decane Slab + Uncharged Methane Molecule			
Distance After MD (nm)	**Coul. Energy (kJ.mol^{-1})**	**S-R LJ Energy (kJ.mol^{-1})**	**L-R LJ Energy (kJ.mol^{-1})**
0.511	0	-2.37	-0.507
0.971	0	0	-0.144
1.43	0	0	0
1.86	0	0	0
0.511 (no solvent)	0	-2.37	-0.507

Table 3B

Results for charged system featuring a twin range LJ cut-off scheme

rlist = 0.9 ; rcoulomb = 0.9 ; rvdw = 1.4 (nm) Charged Decane Slab + Charged Methane Molecule			
Distance After MD (nm)	**Coul. Energy (kJ.mol^{-1})**	**S-R LJ Energy (kJ.mol^{-1})**	**L-R LJ Energy (kJ.mol^{-1})**
0.511	-0.295	-2.37	-0.507
0.971	0	0	-0.144
1.43	0	0	0
1.86	0	0	0
0.511 (no solvent)	-0.295	-2.37	-0.507

Table 3C

Results for uncharged system featuring a single range LJ cut-off scheme

rlist = 1.0 ; rcoulomb = 1.0 ; rvdw = 1.0 (nm) Uncharged Decane Slab + Uncharged Methane Molecule			
Distance After MD (nm)	**Coul. Energy (kJ.mol^{-1})**	**S-R LJ Energy (kJ.mol^{-1})**	**L-R LJ Energy (kJ.mol^{-1})**
0.511	0	-2.65	0
0.971	0	0	0
1.43	0	0	0
1.86	0	0	0
0.511 (no solvent)	0	-2.65	0

Table 3D

Results for charged system featuring a single range LJ cut-off scheme

rlist = 1.0 ; rcoulomb = 1.0 ; rvdw = 1.0 (nm) Charged Decane Slab + Charged Methane Molecule			
Distance After MD (nm)	**Coul. Energy (kJ.mol^{-1})**	**S-R LJ Energy (kJ.mol^{-1})**	**L-R LJ Energy (kJ.mol^{-1})**
0.511	-0.680	-2.65	0
0.971	0	0	0
1.43	0	0	0
1.86	0	0	0
0.511 (no solvent)	-0.680	-2.65	0

The behaviour of the methane decane test system was as expected in all cases, and interaction energies between the methane molecule and the decane slab were unaffected by the presence or absence of solvent under all conditions tested.

For the twin range VDW cut-off scheme, short range LJ energies (within 0.9 nm) were always zero when the methane molecule was at a minimum distance in excess of the 0.9 nm cut-off, and favourable (-2.37 $kJ.mol^{-1}$) when within the cut-off (Table 3A, Table 3B). Long range LJ energies (between 0.9 nm and 1.4 nm) decreased in magnitude as the methane molecule move further away from the decane slab, returning to zero once outside the 1.4 nm cut-off range. Coulombic energies were mildly favourable (-0.295 $kJ.mol^{-1}$) when the methane molecule was within the 0.9 nm cut-off, and were never calculated beyond the cut-off. Long-range Coulombic energies (in excess of 0.9 nm) are approximated by PME and are unreported here as they are not atom-specific contributions to the interaction energy.

For the single range LJ cut-off scheme, all LJ energies within the 1 nm cut-off are considered as short range, so that no long range VDW energies exist. As such, long range LJ energies were evaluated as zero across the conditions tested (Table 3C, Table 3D). The increased cut-off distance of 1 nm compared to 0.9 nm resulted in an increase in the magnitude of both the short range LJ energies (-2.65 $kJ.mol^{-1}$) and Coulombic energies (-0.680 $kJ.mol^{-1}$). Once again, both energies were evaluated as zero outside the cut-off range as seen in the twin range LJ cut-off schemes.

When extrapolating the results to a protein-membrane system, the size of the protein must be taken into account. At any given minimum distance within the LJ and Coulombic cut-offs, the protein will exist both inside and outside the cut-offs, whilst exhibiting a far greater charge and having considerably more contact points. The majority of those contact points will not exist in the short range LJ calculations. As the VDW contribution to the overall interaction energy is heavily dependent on the number of contact points, the long range LJ energy contribution is a better representation of the interaction energy between two large elements within a system, such as a protein and a membrane.

Moving onto the rotational interaction energy screening of bigger systems, the use of a twin range LJ cut-off scheme was implemented in line with all other MD simulations carried out in this thesis, as there were no shortcomings in their evaluation when compared to a single range cut-off scheme. It was also decided that Coulombic and long range LJ interaction energies between the protein and membrane would be considered as they were the most ideal evaluation of the electrostatic and VDW contributions to the protein's approach to the membrane.

3.4 - Application to Protein-Membrane Systems

3.4.1 - Script Development

The overall aim of this chapter was to determine the most energetically favourable orientation of a protein relative to a membrane in terms of Coulombic and VDW interactions. This would allow the determination of the optimal protein-membrane interface from which to start a simulation, and also the likely membrane composition preferences of the protein. To do so, a python script (Appendix B) was developed to use a range of GROMACS tools to rapidly and effectively screen all orientations of a protein in the presence of a lipid bilayer. This script was initially tested and debugged on a desktop computer featuring 8 cores to ensure it was correctly following the rotation scheme. This initial rotation scheme featured two angles, representing rotation around the x and z axes, with 10° increments at a series of protein membrane distances (Section 3.5.2). Following this, the script was parallelised and optimised with the help of Dr Ben Roberts for the NeSI PAN Cluster [61] using a more complex rotation scheme described below. An array submission script was used where all iterations were run sequentially on 4 cores, grouped by their rotation around the x axis. The steps carried out by the script are as follows:

Firstly, the protein coordinates and membrane coordinates were separated into unique files, allowing any changes in orientation to be made solely to the protein. The protein was rotated using pitch, roll, and yaw variables around the x, y, and z axes respectively, in chosen increments (30° for the purposes of this thesis) ranging from 0-360°, as a means of uniformly sampling the protein surface in all orientations. Each rotated configuration was recombined with the initial coordinates of the membrane. To ensure each configuration was situated at an identical distance away from the membrane, the minimum distance between the protein and the membrane was calculated, and the protein was translated along the y axis (normal to the membrane surface) to give a chosen final minimum protein-membrane distance (0.5 nm for the purposes of this thesis).

Next, each protein-membrane system was solvated using the SPC water model and energy minimised and a short MD simulation applied as per the test system protocol (Section 3.3.1). The short-range electrostatic and short- and long-range VDW interaction energies between the protein and the membrane were calculated and averaged over the course of the simulation. These energy values were plotted against pitch, roll, and yaw rotation angles in 3D heat maps to identify patterns in which orientations of the protein are most favourable in its approach to the cell membrane.

The outputs of this script include a data file listing the rotation angles and energies calculated, the system coordinates of each rotated protein orientation, and the GROMACS MD run input file (tpr)

which incorporates the topology, coordinates and simulation parameters of the system. This allows for further analysis such as characterising individual residue contributions and the like.

3.4.2 - Results of Initial Testing of Protein-Membrane Interaction Screening

Phosphatidylinositide 3-kinase (PI3K) is a signalling enzyme which participates in pathways regulating cellular outcomes such as growth, proliferation, and survival [62]. It activates these pathways via the phosphorylation of its substrate PIP_2, which resides in the inner leaflet of the cell's plasma membrane. Mutations of PI3Kα, such as the H1047R (histidine to arginine at position 1047) variant, lead to hyper-activation of the pathway and are implicated in cancer development [63]. As such, characterisation of the PI3Kα complex requires a clear understanding of its interaction with the membrane, and study of this interaction led to the development of this method. The characterisation of the PI3Kα-membrane interface is the primary focus of the research described in Chapter 4 - a detailed background and discussion on PI3Kα is given there. However, for the purposes of the results shown in this chapter, analysis will be centered on testing the protein-membrane interaction screening method.

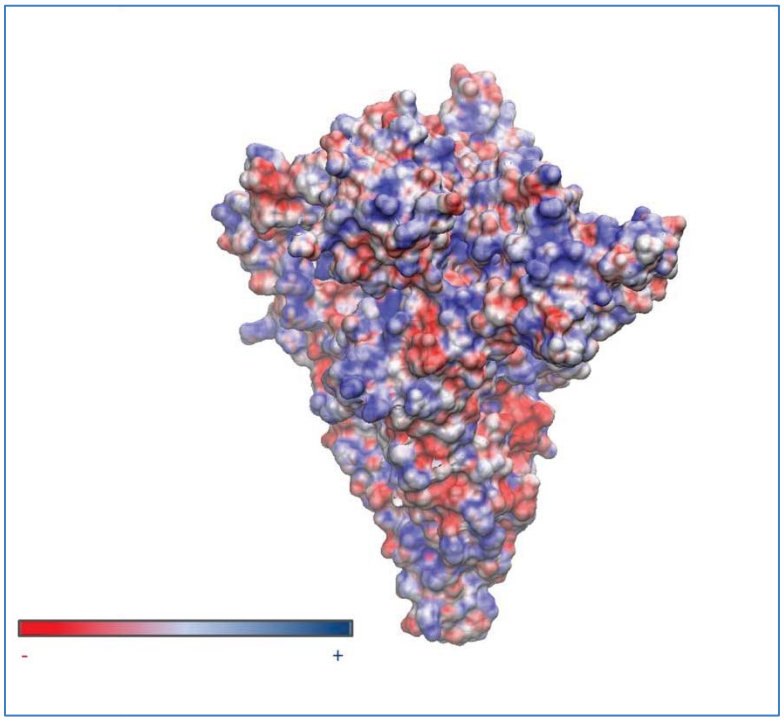


Figure 3iii - *Electrostatic potential of PI3Kα's surface (PDB ID: 4A55 [40]), ranging from negative (red) to positive (blue) on the basis of the nature of the residues and their side chains. This image was generated using the Adaptive Poisson-Boltzmann Solver (APBS) [64, 65] implemented in VMD [66].*

As seen in Figure 3iii, PI3Kα is a complex structure whose surface has a wide distribution of positive and negative charge. However, despite this complexity, the optimal orientation of PI3Kα to the membrane must be consistent for it to be functional. The favourability of a PI3Kα-membrane interface would be difficult to determine on the basis of just its surface electrostatic potential, and as such, the rotational interaction energy screening presents itself as a rapid alternative for the prediction of the most favourable interface.

3.4.2.1 - Effect of Protein-Membrane Distance on Interaction Energy

The results of the initial testing of the script are presented in 2D heat maps (Figure 3iv), as the rotation scheme revolved only around two axes, where each square is representative of a particular combination of rotation angles and therefore a particular orientation of the protein with respect to the membrane. The PI3Kα H1047R mutant in the presence of a brain lipid bilayer inclusive of its PIP_2 substrate (as developed in Chapter 2) was used as an initial test system. An equilibrated structure of H1047R-PI3Kα (see Chapter 4 for details) was used for the screening. This oncogenic mutant of PI3Kα increases its activity by increasing its membrane affinity, known from experimental data [40, 67], making this a feasible test system to see whether the most favourable results of the protein-membrane interaction energy screening were in keeping with what is known about how this protein interacts with the membrane.

As previously discussed (Section 3.1.4), membrane interaction is driven by both non-specific long range and specific short-range interactions giving way to membrane association and binding. As such, both Lennard-Jones and Coulombic interactions were analysed at varying distances to evaluate their relative strengths and overall contribution to the protein's membrane binding orientation. The VDW and Coulombic energies and how they related to optimal protein orientation were first evaluated at the closest distance of 0.5 nm in detail (Figure 3iv) before being compared to larger distances of 1 nm and 1.4 nm (Figure 3v). As seen in Figure 3iv, Coulombic interaction energies dominate the VDW energies by a factor of over tenfold, with a favourable VDW profile not necessarily corresponding to a favourable Coulombic profile. Region A represents an orientation with favourable VDW, but unfavourable Coulombic energy, while Region B represents the most favourable orientation when considering both VDW and Coulombic energies.

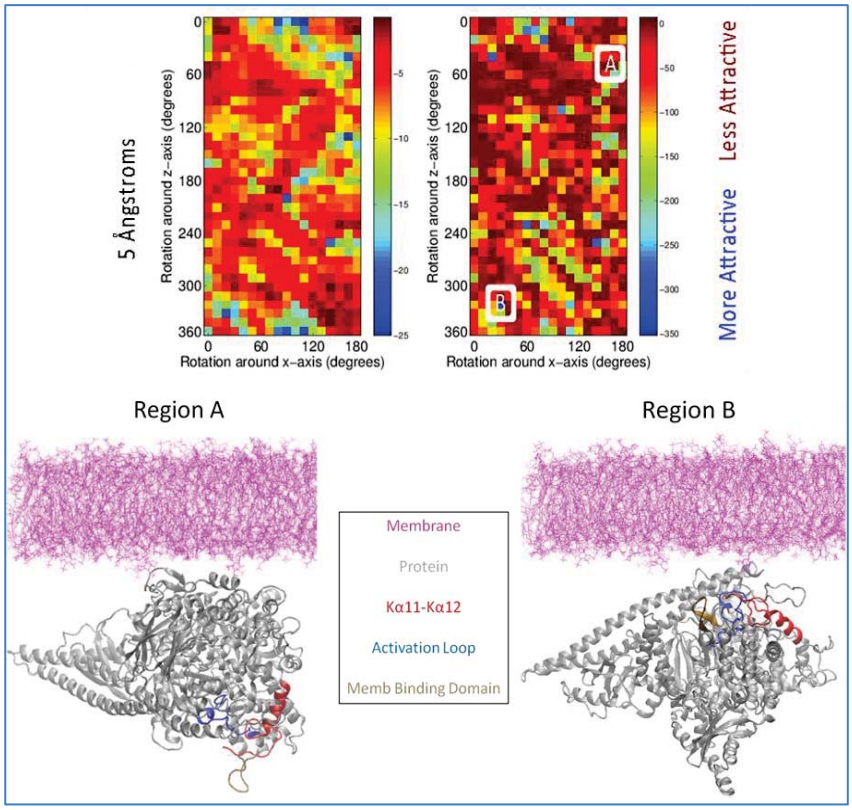


Figure 3iv - *Rotational interaction energy screening results for the H1047R mutant system at a minimum distance of 0.5 nm (5 Å) from the cell membrane surface.* ***Top panel****: heat maps of the (from left to right) VDW and Coulombic energy terms. The colouring goes from warm (red) to cold (blue) on the basis of the favourability of the protein-membrane interaction at that particular orientation; however, the values measured in kJ.mol^{-1} on each scale differ. Two regions have been selected on the rightmost plot - A represents an example of an orientation with favourable VDW, but unfavourable Coulombic energy; B represents the orientation with the most favourable VDW and Coulombic energies.* ***Bottom panel****: protein-membrane systems corresponding to A and B, with key parts of the protein highlighted.*

Focusing on the coloured domains (Figure 3iv) essential for substrate recognition and membrane binding, the protein membrane orientation sampled in Region B is similar to the orientation expected based on the experimental Hydrogen Deuterium Exchange (HDx) data [67], as these domains are all in close proximity and facing the membrane surface (see section 4.3.2.5 for more details). This suggests the biochemically relevant PI3Kα-membrane interaction is dominated by Coulombic interactions, as the protein membrane orientation sampled in Region A is inconsistent with the HDx data. The PI3Kα interface presented to the membrane in Region B features a high percentage of basic residues (arginine and lysine) which interact favourably with the anionic lipids (POPS and PIP_2) present in the brain lipid membrane (see section 4.3.3.3 for more details).

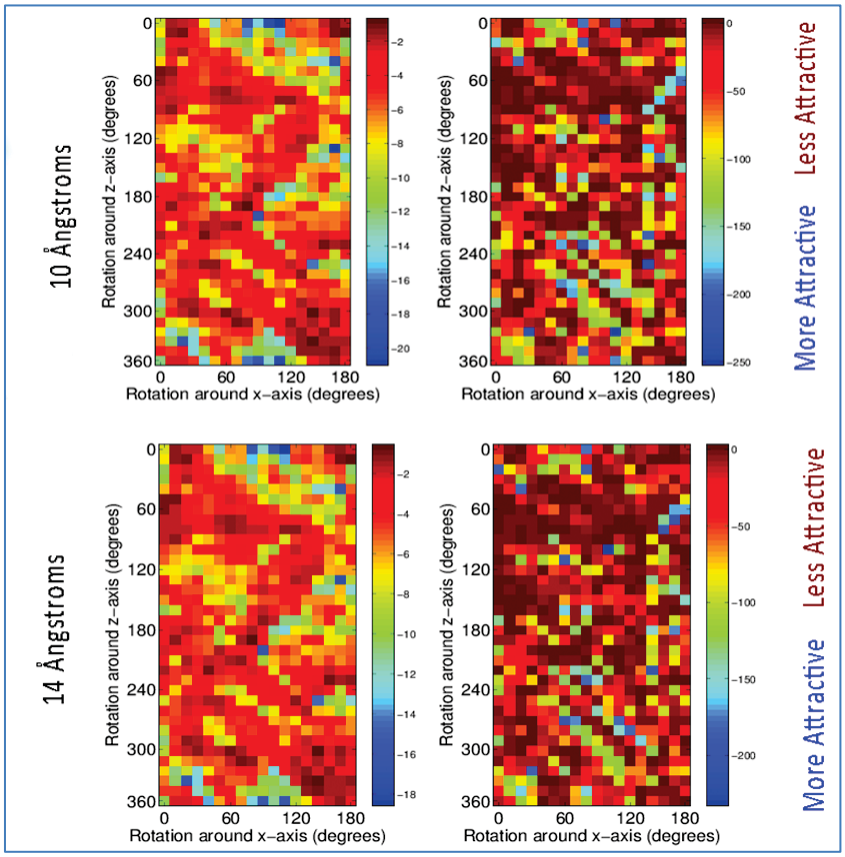


Figure 3v - *Rotational interaction energy screening results for the H1047R system at a minimum distance of 1 nm (10 Å) (top panel) and 1.4 nm (14 Å) (bottom panel) from the cell membrane. Heat maps of the (from left to right) VDW and Coulombic energy terms. The colouring goes from warm (red) to cold (blue) on the basis of the favourability of the protein-membrane interaction at that particular orientation; however, the values measured in kJ.mol^{-1} on each scale differ.*

As the minimum distance between the PI3Kα H1047R mutant and the membrane increases from 0.5 nm to 1 nm and 1.4 nm, the favourability profile of the energies remains similar across all orientations (Figure 3v). As expected, the only difference is the strength of the interaction in both the VDW and Coulombic energy profiles. As this minimum distance increases, the most favourable VDW profile energy decreases from -25 kJ.mol^{-1} at 0.5 nm to -18 kJ.mol^{-1} at 1.4 nm, representing a 30% reduction due to the Lennard-Jones potential tending more and more towards zero. However, the most significant change is seen in the Coulombic energy profile, which sees a decrease from -350 kJ.mol^{-1} in the most favourable orientation at 0.5 nm to -220 kJ.mol^{-1} at 1.4 nm, representing a 40% reduction due to the decrease in magnitude of the (favourable) Coulombic interaction energy as the distance increases.

As the favourability profile of the energies remained similar across all distances, scaled only by relative magnitude, it was decided that the closest protein-membrane distance of 0.5 nm was most suitable for the evaluation of all future protein-membrane interactions.

3.4.2.2 - Effect of Complex Rotational Scheme on Optimal Protein-Membrane Orientation

The parallelisation of the script for use on high performance computing (HPC) platforms allowed a more comprehensive surface sampling rotation scheme to be used. The protein was subjected to rotation around three axes (pitch, roll, and yaw) and therefore the results will be presented as 3D heat maps projected on cubes (Figure 3vi), where each mini-cube is representative of one combination of rotation angles. The results from the interaction energy screening of wild type PI3Kα in the presence of a brain lipid bilayer following both the initial and more complex rotation schemes were used for comparison. These results will never be perfectly reproducible, due to the coordinates following each rotation undergoing a short energy minimisation and molecular dynamics simulation prior to the calculation of interaction energies. In the case of the 2D plots, with a sampling range of 180° around the x axis and 360° around the z axis using increments of 10°, there were roughly 700 configurations. In the case of the 3D plots, with a sampling range of 360° in pitch, roll, and yaw, using increments of 30°, there were roughly 1700 configurations. This difference in sampling allowed for a wider range of energy values in the case of the HPC testing with VDW energies going from 0 to -25 $kJ.mol^{-1}$ in comparison to 0 to -18 $kJ.mol^{-1}$ for the initial testing (Figure 3vi). The same was evident for the Coulombic energies, with ranges being 0 to -560 $kJ.mol^{-1}$ and 0 to -350 $kJ.mol^{-1}$ respectively. This suggests that the more complex rotation scheme allows for the evaluation of a wider range of favourable orientations, including those not sampled by the simpler rotation scheme.

However, despite the difference in overall magnitude, the most favourable orientations of PI3Kα with respect to the membrane identified by both versions of the rotation script were similar, with some of the top-ranking configurations in the more complex rotation scheme not being sampled beforehand in the two-dimensional rotation scheme. While ideally sampling would be done in smaller, *e.g.* 1°, increments across all axes, this is not computationally feasible within the time restraints of the project. Nonetheless, this method does a good job of identifying the PI3Kα orientation relative to the membrane that might be expected based on the experimental HDx data (see section 4.3.2.5 for more details) [67].

The simple and complex rotation scheme methods of surface sampling produced a series of configurations where the equivalent orientation had its energies evaluated. This occurred whenever the simple rotation scheme featured a rotational increment as a multiple of 30° and the complex rotation scheme featured a rotational increment of 0° around the y axis (roll). At these points, accounting for slight changes in orientation and protein-membrane distance (Figure 3vii), the results are directly comparable. A series of configurations were selected at regular intervals of roughly 120° rotations around the z axis when the rotation around the x axis was 0°; and then intervals of 60°

rotations around the x axis for one randomly selected rotation around the z axis. They were compared in tabular format (Table 3E). While the Coulombic energy values in the complex rotation scheme do not perfectly reproduce those in the simple rotation scheme, both methods give a similar profile of Coulombic interaction energies across the different orientation angles selected, *i.e.* relative favourability. These differences can easily be explained by small-scale changes in amino acid side chain and lipid head group orientation (Figure 3vii) that occur during the energy minimisation and MD steps, as these are the two key charged elements in the system that contribute most to the Coulombic interaction energy.

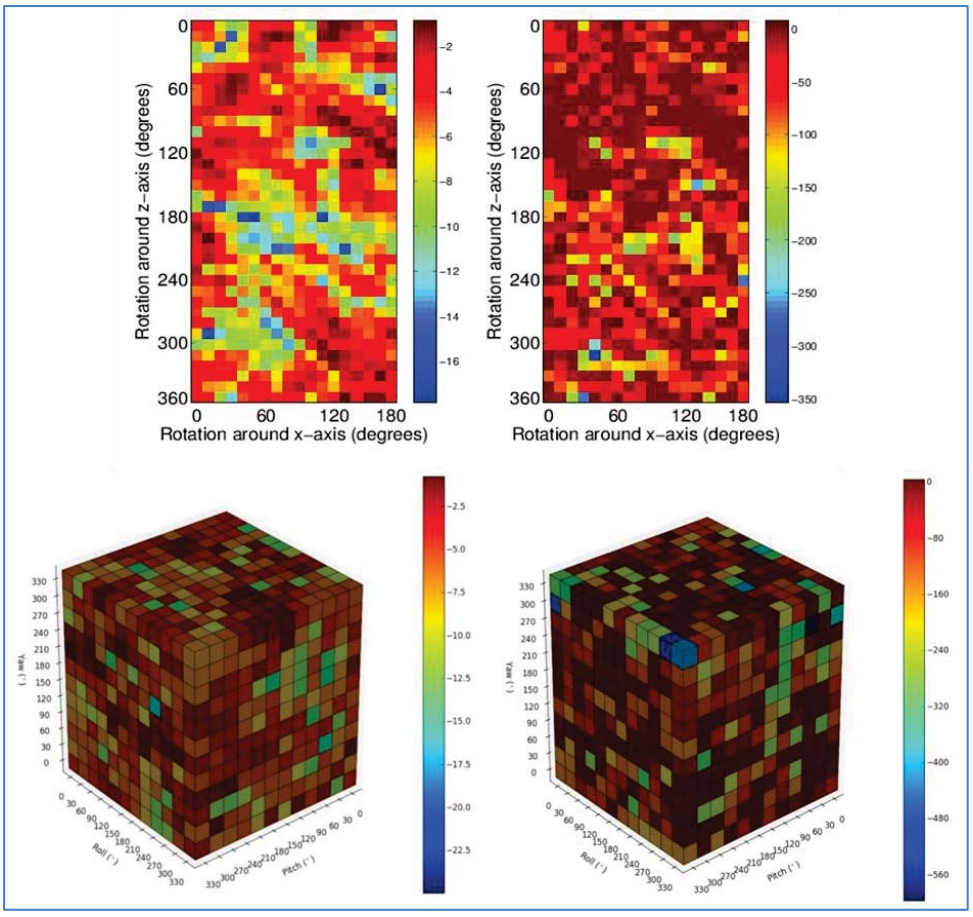


Figure 3vi - *Interaction energy screening results for the wild type system at a minimum distance of 0.5 nm from the cell membrane. Top panel (2D Screening): heat maps of the (from left to right) VDW and Coulombic energy terms. Bottom panel (3D Screening): heat cubes of the (from left to right) VDW and Coulombic energy terms. The colouring goes from warm (red) to cold (blue) on the basis of the favourability of the protein-membrane interaction at that particular orientation; however, the values measured in kJ.mol^{-1} on each scale differ.*

Table 3E

Comparison of Coulombic energies across both rotation schemes

Simple Rotation Scheme		Complex Rotation Scheme	
Rotation around x,z (°)	Coul. Energy (kJ.mol^{-1})	Rotation around x,y,z (°)	Coul. Energy (kJ.mol^{-1})
0,0	-70.5	0,0,0	-86.2
0,120	0.00	0,0,120	0.01
0,210[1]	**-5.10**	**0,0,210[1]**	**-7.19**
0,330	-82.8	0,0,330	-86.6
60,90	-15.4	60,0,90	-3.85
120,300	0.38	120,0,300	-1.72

[1] *- The orientation used in the structural alignment pictured in Figure 3vii.*

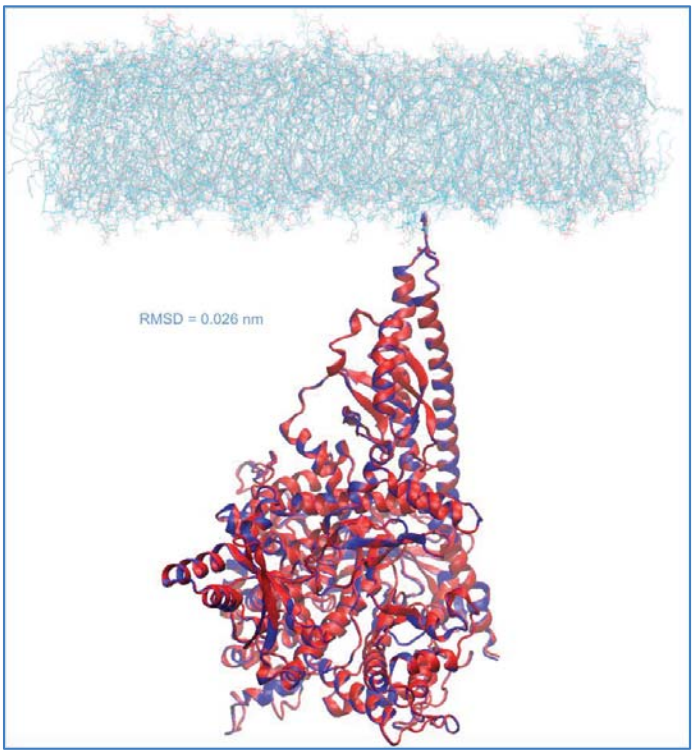


Figure 3vii - *Structural alignment of cartoon represented PI3Kα at the highlighted orientations in Table 3E. Simple rotation scheme: PI3Kα coordinates shown in blue, membrane coordinates shown in cyan. Complex rotation scheme: PI3Kα coordinates shown in red, membrane coordinates shown in pink. The atom-positional root-mean-square deviation between the two structures is also presented.*

3.5 - Conclusion

The initial testing of the script used a simple rotation scheme around two axes to identify the most favourable PI3Kα orientation matching those expected based on experimental data and the relative positioning of its membrane binding domains. This further confirmed the belief that the concept of protein-membrane orientation being determined by Coulombic interaction energy was a reasonable one. Following improvement of the sampling using a more complex rotation scheme, and parallelisation of the script, evaluation of the results led to the conclusion that not only were the results valid, but a greater range in interaction energy values showed that the simple rotation scheme missed out on some of the most favourable orientations.

It was henceforth decided that the parallelised and more complex method would be used for the evaluation of protein-membrane interactions, and was later used to obtain a more comprehensive insight into PI3Kα's interaction with the membrane, results for which can be found in Chapter 4. Initial testing was done using PI3Kα as it was a protein that was well understood due to it being the primary focus of this thesis; however, more rigorous testing and validation was also undertaken exploring further applications of this method. This is described in the attached manuscript (Appendix C) submitted for publication in the ***Journal of Molecular Recognition***.

3.6 - Bibliography

1. Krogh, A., B. Larsson, G. von Heijne, and E.L.L. Sonnhammer, *Predicting transmembrane protein topology with a hidden markov model: application to complete genomes1.* Journal of Molecular Biology, 2001. **305**(3): p. 567-580.
2. Overington, J.P., B. Al-Lazikani, and A.L. Hopkins, *How many drug targets are there?* Nat Rev Drug Discov, 2006. **5**(12): p. 993-996.
3. Tautermann, C.S., *GPCR structures in drug design, emerging opportunities with new structures.* Bioorganic & Medicinal Chemistry Letters, 2014. **24**(17): p. 4073-4079.
4. Sheetz, M.P., J.E. Sable, and H.G. Dobereiner, *Continuous membrane-cytoskeleton adhesion requires continuous accommodation to lipid and cytoskeleton dynamics.* Annu. Rev. Biophys. Biomol. Struct., 2006. **35**: p. 417-434.
5. Di Paolo, G. and P. De Camilli, *Phosphoinositides in cell regulation and membrane dynamics.* Nature, 2006. **443**: p. 651-657.
6. Bhattacharyya, R.P., A. Remenyi, B.J. Yeh, and W.A. Lim, *Domains, motifs, and scaffolds: the role of modular interactions in the evolution and wiring of cell signaling circuits.* Annu. Rev. Biochem., 2006. **75**: p. 655-680.
7. Perisic, O., H.F. Paterson, G. Mosedale, S. Lara-González, and R.L. Williams, *Mapping the Phospholipid-binding Surface and Translocation Determinants of the C2 Domain from Cytosolic Phospholipase A2.* Journal of Biological Chemistry, 1999. **274**(21): p. 14979-14987.
8. Colon-Gonzalez, F. and M.G. Kazanietz, *C1 domains exposed: from diacylglycerol binding to protein-protein interactions.* Biochim. Biophys. Acta, 2006. **1761**: p. 827-837.
9. Cho, W. and R.V. Stahelin, *Membrane binding and subcellular targeting of C2 domains.* Biochim. Biophys. Acta, 2006. **1761**: p. 838-849.
10. Stahelin, R.V., D. Karathanassis, K.S. Bruzik, M.D. Waterfield, J. Bravo, R.L. Williams, and W. Cho, *Structural and Membrane Binding Analysis of the Phox Homology Domain of Phosphoinositide 3-Kinase-C2α.* Journal of Biological Chemistry, 2006. **281**(51): p. 39396-39406.
11. Lemmon, M.A. and K.M. Ferguson, *Signal-dependent membrane targeting by pleckstrin homology (PH) domains.* Biochem. J., 2000. **350**: p. 1-18.
12. Murray, D., N. Ben-Tal, B. Honig, and S. McLaughlin, *Electrostatic interaction of myristoylated proteins with membranes: simple physics, complicated biology.* Structure, 1997. **5**(8): p. 985-989.
13. Cho, W., *Membrane Targeting by C1 and C2 Domains.* Journal of Biological Chemistry, 2001. **276**(35): p. 32407-32410.
14. Stahelin, R.V. and W. Cho, *Differential Roles of Ionic, Aliphatic, and Aromatic Residues in Membrane−Protein Interactions: A Surface Plasmon Resonance Study on Phospholipases A2.* Biochemistry, 2001. **40**(15): p. 4672-4678.
15. Spector, A.A. and M.A. Yorek, *Membrane lipid composition and cellular function.* Journal of Lipid Research, 1985. **26**(9): p. 1015-35.
16. Lande, M.B., J.M. Donovan, and M.L. Zeidel, *The relationship between membrane fluidity and permeabilities to water, solutes, ammonia, and protons.* J Gen Physiol, 1995. **106**(1): p. 67-84.
17. Lemmon, M.A., *Membrane recognition by phospholipid-binding domains.* Nat Rev Mol Cell Biol, 2008. **9**(2): p. 99-111.
18. Stephens, L., A. McGregor, and P. Hawkins, *Biology of Phosphoinositides*. 2000, Nature Publishing Group.
19. McLaughlin, S. and D. Murray, *Plasma membrane phosphoinositide organization by protein electrostatics.* Nature, 2005. **438**: p. 605-611.
20. Czech, M.P., *PIP2 and PIP3: Complex Roles at the Cell Surface.* Cell, 2000. **100**(6): p. 603-606.

21. McLaughlin, S., J. Wang, A. Gambhir, and D. Murray, *PIP2 and proteins: interactions, organization, and information flow.* Annu. Rev. Biophys. Biomol. Struct., 2002. **31**: p. 151-175.
22. Murray, D. and B. Honig, *Electrostatic Control of the Membrane Targeting of C2 Domains.* Molecular Cell, 2002. **9**(1): p. 145-154.
23. Verdaguer, N., S. Corbalan-Garcia, W.F. Ochoa, I. Fita, and J.C. Gomez-Fernandez, *Ca2+ bridges the C2 membrane-binding domain of protein kinase C[alpha] directly to phosphatidylserine.* EMBO J., 1999. **18**: p. 6329-6338.
24. Davletov, B., O. Perisic, and R.L. Williams, *Calcium-dependent membrane penetration is a hallmark of the C2 domain of cytosolic phospholipase A2 whereas the C2A domain of synaptotagmin binds membranes electrostatically.* J. Biol. Chem., 1998. **273**: p. 19093-19096.
25. Stahelin, R.V. and W. Cho, *Roles of calcium ions in the membrane binding of C2 domains.* Biochemical Journal, 2001. **359**(Pt 3): p. 679-685.
26. Cho, W. and R.V. Stahelin, *Membrane-protein interactions in cell signaling and membrane trafficking.* Annu. Rev. Biophys. Biomol. Struct., 2005. **34**: p. 119-151.
27. Kufareva, I., M. Lenoir, F. Dancea, P. Sridhar, E. Raush, C. Bissig, J. Gruenberg, R. Abagyan, and M. Overduin, *Discovery of novel membrane binding structures and functions.* Biochemistry and cell biology = Biochimie et biologie cellulaire, 2014. **92**(6): p. 555-563.
28. Nalefski, E.A. and J.J. Falke, *The C2 domain calcium-binding motif: structural and functional diversity.* Protein Science : A Publication of the Protein Society, 1996. **5**(12): p. 2375-2390.
29. Blatner, N.R., R.V. Stahelin, K. Diraviyam, P.T. Hawkins, W. Hong, D. Murray, and W. Cho, *The Molecular Basis of the Differential Subcellular Localization of FYVE Domains.* Journal of Biological Chemistry, 2004. **279**(51): p. 53818-53827.
30. Lee, A.G., *How lipids affect the activities of integral membrane proteins.* Biochimica et Biophysica Acta (BBA) - Biomembranes, 2004. **1666**(1–2): p. 62-87.
31. McIntosh, T.J. and S.A. Simon, *Bilayers as Protein Solvents: Role of Bilayer Structure and Elastic Properties.* The Journal of General Physiology, 2007. **130**(2): p. 225-227.
32. Phillips, R., T. Ursell, P. Wiggins, and P. Sens, *Emerging roles for lipids in shaping membrane-protein function.* Nature, 2009. **459**(7245): p. 379-385.
33. Berger, O., O. Edholm, and F. Jähnig, *Molecular dynamics simulations of a fluid bilayer of dipalmitoylphosphatidylcholine at full hydration, constant pressure, and constant temperature.* Biophysical Journal, 1997. **72**(5): p. 2002-2013.
34. Pandit, S.A., D. Bostick, and M.L. Berkowitz, *Mixed Bilayer Containing Dipalmitoylphosphatidylcholine and Dipalmitoylphosphatidylserine: Lipid Complexation, Ion Binding, and Electrostatics.* Biophysical Journal, 2003. **85**(5): p. 3120-3131.
35. Hofsäß, C., E. Lindahl, and O. Edholm, *Molecular Dynamics Simulations of Phospholipid Bilayers with Cholesterol.* Biophysical Journal, 2003. **84**(4): p. 2192-2206.
36. Berglund, N.A., T.J. Piggot, D. Jefferies, R.B. Sessions, P.J. Bond, and S. Khalid, *Interaction of the Antimicrobial Peptide Polymyxin B1 with Both Membranes of E. coli: A Molecular Dynamics Study.* PLOS Computational Biology, 2015. **11**(4): p. e1004180.
37. Merchant, B.A. and J.D. Madura, *A Review of Coarse-Grained Molecular Dynamics Techniques to Access Extended Spatial and Temporal Scales in Biomolecular Simulations.* Annual Reports in Computational Chemistry, 2011. **7**: p. 67-87.
38. Kmiecik, S., D. Gront, M. Kolinski, L. Wieteska, A.E. Dawid, and A. Kolinski, *Coarse-Grained Protein Models and Their Applications.* Chemical Reviews, 2016. **116**(14): p. 7898-7936.
39. Irvine, W.A., *Testing of parameters for a biologically accurate brain membrane and molecular dynamics simulations exploration of membrane interactions and conformational changes exhibited by p110a and its oncogenic mutants*, in *Institute of Natural and Mathematical Sciences*. 2014, Massey University, Albany, New Zealand: Massey University.
40. Hon, W.-C., A. Berndt, and R.L. Williams, *Regulation of lipid binding underlies the activation mechanism of class IA PI3-kinases.* Oncogene, 2012. **31**(32): p. 3655-3666.

41. Murray, D., A. Arbuzova, B. Honig, and S. McLaughlint, *The role of electrostatic and nonpolar interactions in the association of peripheral proteins with membranes*, in *Current Topics in Membranes*. 2002, Academic Press. p. 277-307.
42. Zhao, H. and P. Lappalainen, *A simple guide to biochemical approaches for analyzing protein–lipid interactions.* Molecular Biology of the Cell, 2012. **23**(15): p. 2823-2830.
43. Pollard, T.D., *A Guide to Simple and Informative Binding Assays.* Molecular Biology of the Cell, 2010. **21**(23): p. 4061-4067.
44. Beseničar, M., P. Maček, J.H. Lakey, and G. Anderluh, *Surface plasmon resonance in protein–membrane interactions.* Chemistry and Physics of Lipids, 2006. **141**(1–2): p. 169-178.
45. Kraft, C.A., J.L. Garrido, L. Leiva-Vega, and G. Romero, *Quantitative Analysis of Protein-Lipid Interactions Using Tryptophan Fluorescence.* Science Signaling, 2009. **2**(99): p. pl4-pl4.
46. Gambhir, A., *Electrostatic sequestration of PIP2 on phospholipid membranes by basic/aromatic regions of proteins.* Biophys. J., 2004. **86**: p. 2188-2207.
47. Rusu, L., A. Gambhir, S. McLaughlin, and J. Radler, *Fluorescence correlation spectroscopy studies of peptide and protein binding to phospholipid vesicles.* Biophys. J., 2004. **87**: p. 1044-1053.
48. Saarikangas, J., H. Zhao, A. Pykäläinen, P. Laurinmäki, P.K. Mattila, P.K.J. Kinnunen, S.J. Butcher, and P. Lappalainen, *Molecular Mechanisms of Membrane Deformation by I-BAR Domain Proteins.* Current Biology. **19**(2): p. 95-107.
49. Qian, J.I.E., B. Yao, and C. Wu, *Fluorescence resonance energy transfer detection methods: Sensitized emission and acceptor bleaching.* Experimental and Therapeutic Medicine, 2014. **8**(5): p. 1375-1380.
50. McLaughlin, S., *The electrostatic properties of membranes.* Annu. Rev. Biophys. Biophys. Chem., 1989. **18**: p. 113-136.
51. Andelman, D., *CHAPTER 12 - Electrostatic Properties of Membranes: The Poisson-Boltzmann Theory*, in *Handbook of Biological Physics*, R. Lipowsky and E. Sackmann, Editors. 1995, North-Holland. p. 603-642.
52. Davis, M.E. and J.A. McCammon, *Electrostatics in biomolecular structure and dynamics.* Chemical Reviews, 1990. **90**(3): p. 509-521.
53. Honig, B. and A. Nicholls, *Classical electrostatics in biology and chemistry.* Science, 1995. **268**(5214): p. 1144-1149.
54. Ben-Tal, N., B. Honig, R.M. Peitzsch, G. Denisov, and S. McLaughlin, *Binding of small basic peptides to membranes containing acidic lipids: theoretical models and experimental results.* Biophys. J., 1996. **71**: p. 561-575.
55. Schmid, N., A.P. Eichenberger, A. Choutko, S. Riniker, M. Winger, A.E. Mark, and W.F. van Gunsteren, *Definition and testing of the GROMOS force-field versions 54A7 and 54B7.* European Biophysics Journal, 2011. **40**(7): p. 843.
56. Essmann, U., L. Perera, M.L. Berkowitz, T. Darden, H. Lee, and L.G. Pedersen, *A smooth particle mesh Ewald method.* The Journal of Chemical Physics, 1995. **103**(19): p. 8577-8593.
57. Pronk, S., S. Pall, R. Schulz, P. Larsson, P. Bjelkmar, R. Apostolov, M.R. Shirts, J.C. Smith, P.M. Kasson, D. van der Spoel, B. Hess, and E. Lindahl, *GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit.* Bioinformatics, 2013. **29**(7): p. 845-54.
58. Berendsen, H.J.C., J.R. Grigera, and T.P. Straatsma, *The missing term in effective pair potentials.* The Journal of Physical Chemistry, 1987. **91**(24): p. 6269-6271.
59. Berendsen, H.J.C., J.P.M. Postma, W.F. van Gunsteren, A. DiNola, and J.R. Haak, *Molecular dynamics with coupling to an external bath.* The Journal of Chemical Physics, 1984. **81**(8): p. 3684-3690.
60. Hess, B., H. Bekker, H.J.C. Berendsen, and J.G.E.M. Fraaije, *LINCS: A linear constraint solver for molecular simulations.* Journal of Computational Chemistry, 1997. **18**(12): p. 1463-1472.

61. Masoud-Ansari, S. *NeSI Pan Cluster*. 2015; Available from: https://wiki.auckland.ac.nz/display/CER/NeSI+Pan+Cluster.
62. Foster, F.M., C.J. Traer, S.M. Abraham, and M.J. Fry, *The phosphoinositide (PI) 3-kinase family.* J Cell Sci, 2003. **116**(Pt 15): p. 3037-40.
63. Cleary, J.M. and G.I. Shapiro, *Development of phosphoinositide-3 kinase pathway inhibitors for advanced cancer.* Curr Oncol Rep, 2010. **12**(2): p. 87-94.
64. Dolinsky, T.J., J.E. Nielsen, J.A. McCammon, and N.A. Baker, *PDB2PQR: an automated pipeline for the setup of Poisson–Boltzmann electrostatics calculations.* Nucleic Acids Research, 2004. **32**(suppl 2): p. W665-W667.
65. Baker, N.A., D. Sept, S. Joseph, M.J. Holst, and J.A. McCammon, *Electrostatics of nanosystems: Application to microtubules and the ribosome.* Proceedings of the National Academy of Sciences, 2001. **98**(18): p. 10037-10041.
66. Humphrey, W., A. Dalke, and K. Schulten, *VMD: visual molecular dynamics.* J Mol Graph, 1996. **14**(1): p. 33-8, 27-8.
67. Burke, J.E., O. Perisic, G.R. Masson, O. Vadas, and R.L. Williams, *Oncogenic mutations mimic and enhance dynamic events in the natural activation of phosphoinositide 3-kinase p110alpha (PIK3CA).* Proc Natl Acad Sci U S A, 2012. **109**(38): p. 15259-64.

Chapter 4 - Phosphatidylinositide 3-Kinase

4.1 - Literature Review

4.1.1 - Overview

Phosphatidylinositide 3-kinases (PI3Ks) are proteins that participate in signalling pathways regulating factors such as cell growth, proliferation, and survival (Figure 4i) [1]. These signalling pathways are triggered by the phosphorylation of the hydroxyl group at the 3-position of the inositol ring of phosphatidylinositols (PIs), converting them from phosphatidylinositol 4,5-bisphosphate (PIP_2) to phosphatidyl 3,4,5-trisphosphate (PIP_3), hence the name given to the enzymes involved [2].

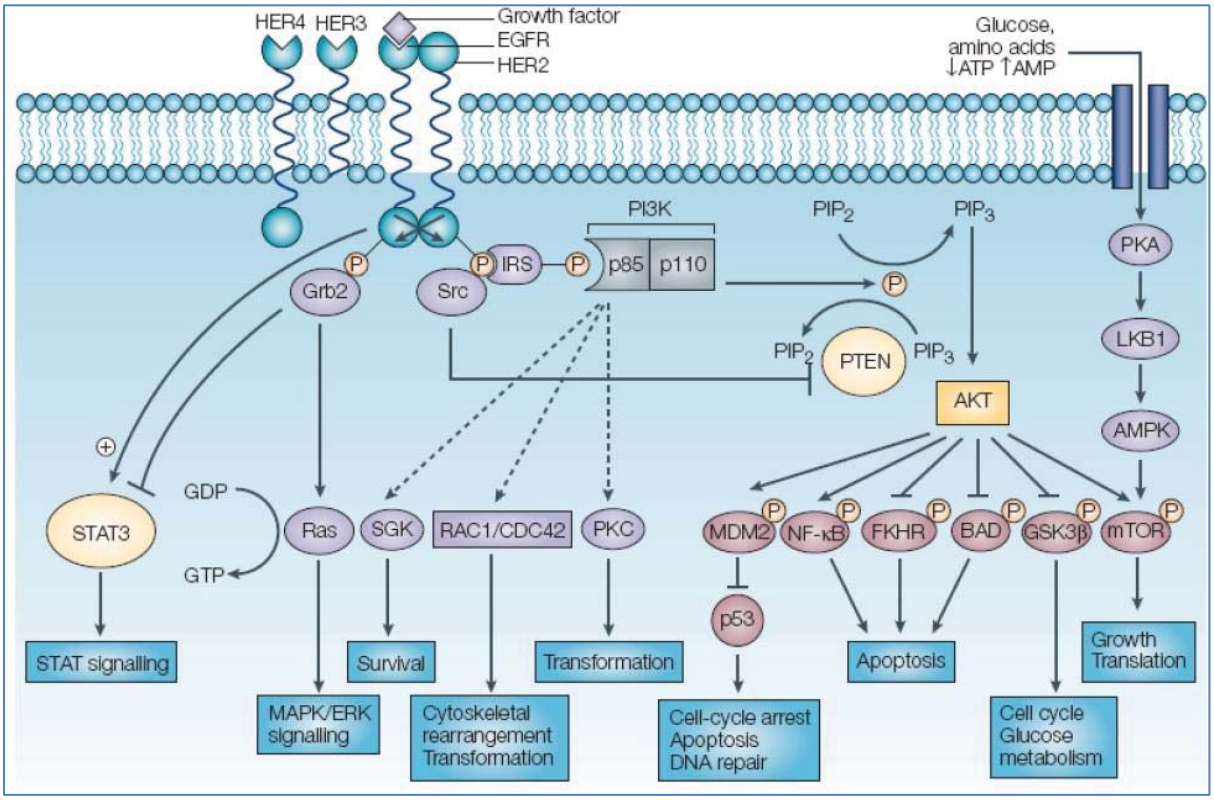


Figure 4i - *Complete Class I PI3K signalling pathway taken unchanged from [3], with PI3Kα comprised of two subunits (p85 and p110) shown in gray. A cascade of functions is triggered by the act of PI3K phosphorylating PIP_2 to PIP_3, including the activation of AKT (orange box) which inhibits a series of transcription factors (pink boxes) primarily mediating cellular functions such as apoptosis and cell-cycle arrest (cyan boxes on the right). This indirectly results in an increase in cell survival (cyan box on the left) through downstream activation of SGK and other proteins (purple boxes).*

An increase in PI3K activity may sometimes result in increased cell proliferation and cell survival, implicating PI3Ks in cancer development [4], with some of the genes encoding these proteins having

been identified as oncogenes [5]. One of these genes, *PIK3CA*, encodes the catalytic subunit (p110) of the α isoform of PI3K [6], one of the three Class IA PI3Ks - the other two being the β and δ isoforms. There also exists a γ isoform belonging to Class IB, with the two classes differentiated only by their regulatory subunits (p85 for class IA, p84 or p101 for class IB). Importantly, the catalytic subunit maintains sequence similarity across all isoforms [7, 8]. Looking at these subunits in more detail, they can be broken down into various domains each of which carries out a specific function of the protein (Figure 4ii).

PI3Ks typically consist of a catalytic subunit as well as a regulatory or inhibitory subunit. The p110α subunit consists of an N terminal adaptor binding domain (ABD), a *Ras* binding domain (RBD), a C2 domain, a helical domain, and finally the all important C terminal kinase domain [9]. This subunit normally works in tandem with the regulatory p85 subunit, controlling the level of activity of the protein, as well as its connectivity to other proteins [10].

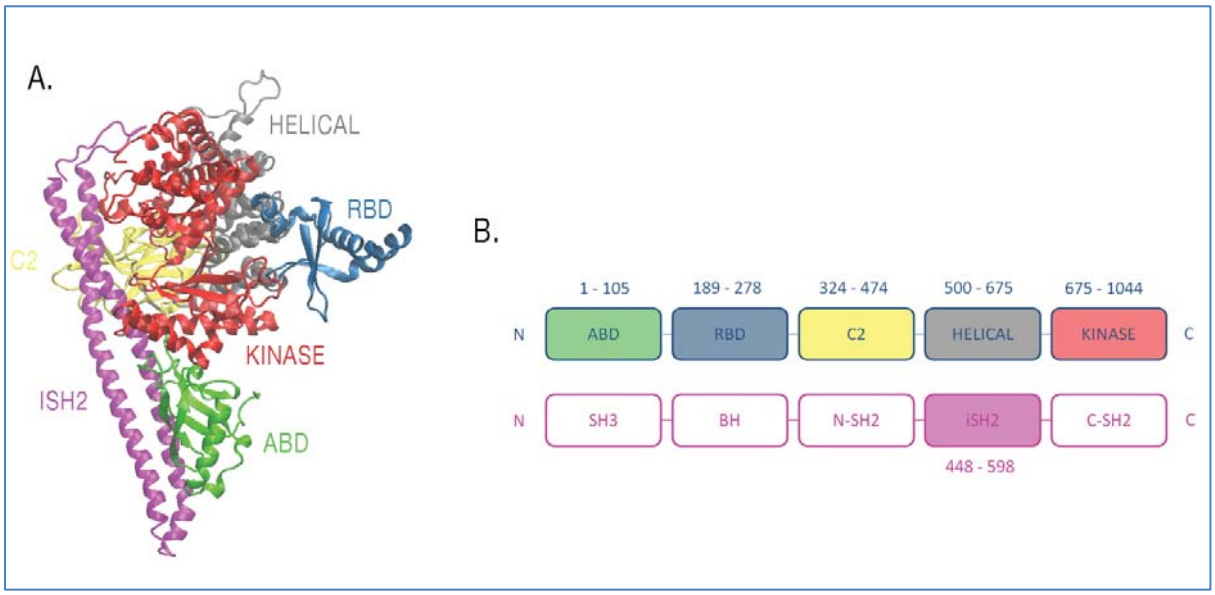


Figure 4ii - ***A)*** *Cartoon structure of the p110α catalytic subunit in complex with the iSH2 domain of the p85α regulatory subunit (PDB ID: 4A55 [11]), coloured according to domain.* ***B)*** *The domains present in (A) (shaded boxes) and their respective residue numbers comprising the p110α (top) and p85α (bottom) subunits.*

There are several variants of the regulatory subunit - p85α, p55α, p50α, p85β and p55γ, all of which contain two Src homology 2 (SH2) domains, one found on the N terminus side and the other on the C terminus side. These two SH2 domains are then separated by a coiled coil referred to as the inter or iSH2 domain, which is the primary point of interaction between the two subunits [12]. The alpha form of the p85 subunit is also characterised by an Src homology 3 (SH3) domain and a Bcl-2

homology (BH) domain straddled by two proline rich regions [13]. Without activating signals from receptor tyrosine kinase (RTK), *Ras*, or G protein-coupled receptors (GPCR), the contacts formed between the regulatory subunits and catalytic subunits are inhibitory to their function [14]. In the case of the Class IA PI3Ks, these activating signals result in the decoupling of inhibitory contacts and the recruitment of the protein to the plasma membrane, where it carries out its function - phosphorylation of PIP_2 [15, 16]. The presence of the product, PIP_3, then activates AKT-dependent and AKT-independent pathways, whilst PTEN regulates the presence of PIP_3 in the membrane by dephosphorylating it back to PIP_2 (Figure 4i).

4.1.2 - Domain Function

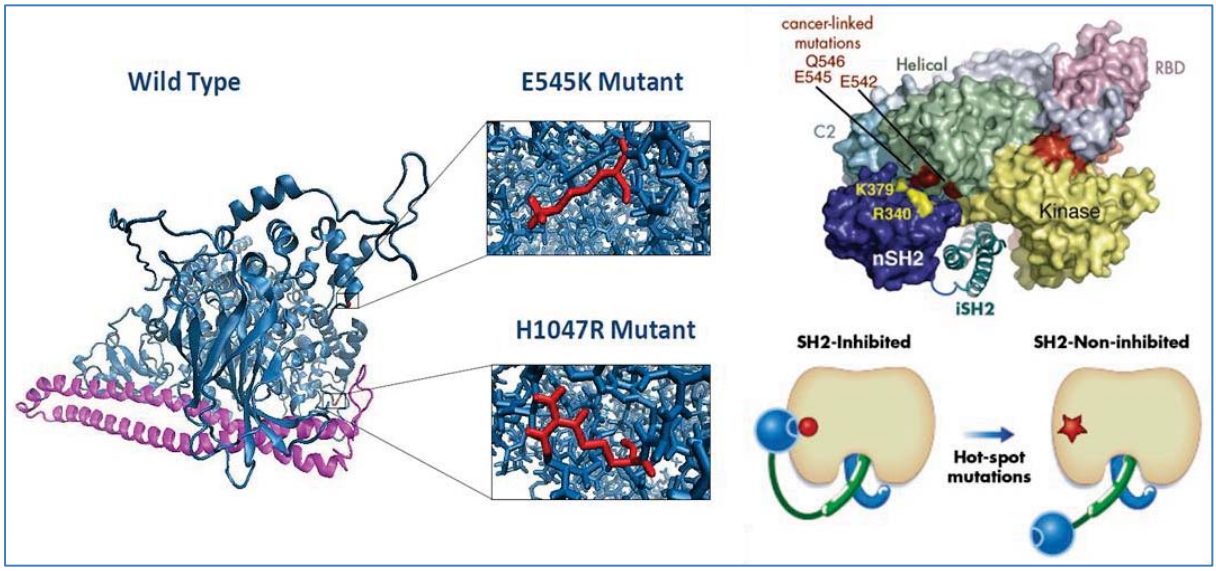
While there have been several proposed functions of each domain within PI3Kα, some are yet to be confirmed. In the p110α subunit, the ABD is simply proposed to bind to the p85α subunit, while the RBD is known to bind to a GTPase from the *Ras* family - a source of induced activation of the PI3Kα enzyme [17, 18]. The C2 domain has been found to bind to the membrane, independent of calcium, as in other proteins including a conserved C2 region such as phosphatase and tensin homolog (PTEN) [19, 20]. The purpose of the helical domain is still a mystery; however it is involved in binding to other proteins such as protein kinase C in the p110γ isoform [21]. Finally, the kinase domain is the source of the catalytic activity of PI3Kα - the phosphorylation of PIP_2 [22-24].

In the p85α subunit, the SH2 domains come to the fore, with the nSH2 domain proposed to form an inhibitory contact with the p110α subunit (Figure 4iii) [25], directly mediating its activity through the binding and dissociation of phosphopeptides [26]. The iSH2 domain appears to serve only as a link between the nSH2 and cSH2 domains, facilitating their contact with the p110α subunit, as the presence of the coiled coil in isolation does not inhibit enzyme activity [27]. The cSH2 domain is proposed to contact the RBD of the p110α subunit, preventing its stimulation by *Ras* [17], but seems to require all other domains to be an effective inhibitor. The functions of the remaining domains in p85α are not clearly understood beyond them binding a series of small proteins to assist in regulation, while the nSH2 contact has been proposed as the primary source of inhibition [27].

4.1.3 - Mutations

Over 1000 mutations of *PIK3CA* have been identified [28], the majority occurring at two hotspots within the p110α subunit - E545K and H1047R (Figure 4iii)[29, 30]. These two point mutations both increase the catalytic activity of p110α via different methods resulting in enhanced downstream signalling of the aforementioned AKT pathway [31], making these enzymes a major target for cancer therapy. Experimental studies have been undertaken to comprehend and characterise the processes

behind the increase in catalytic activity, all of which point to conformational changes in the p110α subunit of the enzyme being responsible. The helical domain mutation E545K releases the inhibition imposed by p85α by displacing the nSH2 domain, and is sensitive to further activation by *Ras*; while the H1047R has improved membrane binding and is not sensitive to activation by *Ras*, but is still subject to inhibition by p85α [32-34]. These two mutations therefore exhibit synergy in their effect



on the increased activity of PI3Kα, suggesting they operate by different mechanisms [8, 13, 34].

Figure 4iii - ***On the left****: a cartoon representation of the p110α subunit (blue) in complex with the iSH2 domain of p85α (purple), highlighting the location of the two major point mutations (red).* ***On the right****: a space filling representation of the p110α subunit in complex with the iSH2 and nSH2 domains of the p85α subunit, highlighting the locations of major oncogenic point mutations in p110α (red) and p85α (yellow) at the point of inhibitory contact between the nSH2 domain and p110α [25].*

Complete crystal structures of p110α in complex with p85α do not exist, however they have been determined for p110α in complex with the iSH2 and nSH2 domains of p85α [11, 13, 33, 35-37]. Analysis of these structures revealed the interaction points between the subunits, as well as the relative positions of residues 545 and 1047. It is understood that while the contact between the iSH2 domain (p85α) and ABD (p110α) is the primary source of subunit interaction, the nSH2 domain also contacts p110α at various points in the helical, C2 and kinase domains [13]. Interestingly enough, the p85α contact points in the helical domain occur in and around residue 545, while the contact points in the kinase domain occur around residue 1047 (Figure 4iii) [33].

Hydrogen deuterium exchange mass spectroscopy was used to identify the regions of PI3Kα involved in membrane binding, and to identify any differences resulting from the oncogenic mutations [38, 39]. It was found that the contact between the nSH2 domain and p110α not only inhibited its activity

through interaction, but also through forming a scaffold restricting any interdomain movements required for membrane binding. Once this interface was disrupted, conformational changes in the ABD allowed the kinase domain to better interact with lipids. It was also found that H1047R had an increased affinity for membranes due to a conformational change in the lipid binding surface, whereas E545K did not, further supporting their difference in mechanism.

4.1.4 - Advantages of Computational Representation

Experimentally, these are the current heights of understanding regarding how these mutations affect the activity of PI3Kα; however, more needs to be understood about the specific residues and conformational changes giving way to membrane binding. Contemporary experimental methods mostly sample structures at only one point in time in limited residue-specific detail [40, 41], but molecular dynamics (MD) simulations provide a means of transcending these heights to obtain a more detailed and dynamic view of the structural changes involved in molecular recognition. In the case of PI3Kα, these include how the structure and dynamics are affected by the oncogenic mutants, and how wild type and oncogenic mutants of PI3Kα are affected by interaction with the cell membrane and, potentially, other known interaction partner proteins. This information could prove to be vital for improving drug specificity in the future.

To date, there have been several molecular dynamics simulations of PI3Kα. They have primarily involved either the wild type p110α or its mutants complexed to an inhibitor in aqueous solution [42-45]. Binding analysis of the ligands aside, it was found that the major structural differences between the bound and unbound forms of the protein lay in the membrane binding domains (residues 721-727, 863-873 and 966-974). There also appears to be clustering of these domains and the WIF motif (residues 1057-1059), a loop previously found to be a key determinant for lipid binding along with the activation loop (residues 940 to 960) [11].

Other developments in the PI3K field involving MD simulations explored the effects of five p110α mutations (R38C, R88Q, E542K, E545K, N345K) on the dynamics of the p110α-p85α complex [46]. Following MD simulations and the tracking of fluctuations exhibited by the various PI3Kα domains, it was found that all these tumour-associated mutations which occur at the interface between p110α and p85α increase activity by weakening the interactions between the catalytic and inhibitory subunits.

4.1.5 - Scope of This Research

While these simulations allowed new insight to be made into the behaviour of the protein, they did not address one of the most important aspects of its function and environment - the membrane interaction. As the purpose of Class I PI3Ks is the phosphorylation of lipids $PI(4,5)P_2$ to $PI(3,4,5)P_3$, a substantial leap in our understanding of the enzyme would be to characterise its behaviour in the presence of its substrate in an effort to understand the structural changes giving way to its function at the atomic level. Importantly, HDx has provided some information on the dynamics of the protein-membrane interaction, and MD provides the perfect opportunity to further develop the resolution of this data.

To do this, MD simulations of PI3K would be better carried out in a system more closely reflecting its realistic biological environment. The main components of this system would therefore be PI3Kα in its most active form (in the absence of the nSH2 and cSH2 domains of p85α) and a membrane including the lipid substrate PIP_2. Physiologically, contacts with these two SH2 domains would be released through phosphotyrosine binding activating PI3Kα [38, 39]; however, as phosphotyrosine will not be included in the system, a crystal structure (PDB ID: 4A55) [11] not inclusive of the nSH2 and cSH2 domains will be used. In Part I, the wild type PI3Kα and the two most common oncogenic mutants (H1047R and E545K) were simulated in the presence of the brain lipid bilayer developed in Chapter 2. In Part II, PI3Kα was remodelled to incorporate its complete sequence inclusive of the C terminal tail. Further alterations were also introduced to test a series of hypotheses developed in Part I. All systems were explored using molecular dynamics and other computational tools, with the aim of understanding the roles of each domain in PI3Kα's function, as well as the mechanism by which the mutations increase PI3Kα's activity.

4.2 - Methods

4.2.1 - Simulation Methods

The simulations of the p110α-p85α complex in the presence of a brain lipid bilayer developed in Chapter 2 were prepared and carried out as detailed below.

Coordinates for wild type p110α in complex with the iSH2 subdomain of wild type p85α (PDB ID: 4A55) [11] with the ligand coordinates removed were obtained from Jack Flanagan (University of Auckland), who modelled missing residues from the crystal structure (310 to 322, 410 to 419, 501 to 525, 863 to 870) without any structural restraints. At the time, this was the only structure available with electron density describing the activation loop. Point mutations were introduced manually at positions 545 (E→K) and 1047 (H→R) using VMD's Mutate Residue extension. A bilayer in the correct proportions and of appropriate size was constructed so the x (13.25 nm) and y (12.42 nm) dimensions approximately mirrored the size of the box of water required to adequately solvate the protein as outlined in Chapter 2. The dimensions of the box containing the membrane were then elongated along the z (16.72 nm) axis, allowing room for the protein coordinates to be implemented 2 nm away. The complete system was then subjected to 1000 steps of steepest descent energy minimisation before being solvated with SPC water [47] and neutralised with the addition of 216 cations (Na^+).

Each simulation was initiated with the following equilibration scheme. Firstly, the initial velocities were randomly generated from a Maxwell-Boltzmann distribution at 50 K. The system was then heated to 300 K over the course of 100 ps in the NVT ensemble. The temperature was controlled using the Berendsen thermostat [48] with a temperature coupling constant (τ_T) of 0.1 ps. The system was further equilibrated for 400 ps at 300 K in the NPT ensemble. The pressure was controlled using the Berendsen barostat with a pressure coupling constant (τ_p) of 0.5 ps and an isothermal compressibility of 4.5 x 10^{-5} bar^{-1} in semi-isotropic conditions.

The final structure was used as the starting configuration for an initial 50 ns production run at 300 K, with structures saved every 200 ps which was run as part of a previously assessed body of work [49]. In this research, the simulations were extended by 100 ns to 150 ns of run time. The LINCS algorithm [50] was used with an order of 4 to constrain bond lengths and water bond angles, allowing for an integration time step of 2 fs. The centre of mass motion was removed every 100 ps. Non-bonded interactions were calculated using a grid cut-off scheme. The non-bonded interactions within a cut-off distance of 0.9 nm were calculated at every step from a pair list that was updated every fifth time step. At this point, interactions between atoms within 1.4 nm were also calculated and were kept

constant between updates. Electrostatic interactions were calculated using particle mesh Ewald (PME) summation [51], with a cut-off distance of 0.9 nm. Duplicate runs (150 ns each) have been generated for each simulation, as well as new simulations (150 ns each) starting from configurations generated following the rotational screening of protein-membrane interaction energies (Chapter 3).

4.2.2 - Coordinate Building Methods

New human p110α-p85α model structures were built to test the importance of various residues and domains of the protein to its interaction with the membrane. Figure 4iv details the changes made in producing each new construct. As crystal structures of human p110α with the complete Kα12 helix in either the open or closed conformation have not yet been determined, this helix had to be modelled using structural information from crystal structures of other related enzymes. Firstly, the crystal structure of the Drosophila Class III PI3K, Vps34 (PDB ID: 2X6H) [52], was used as it exhibits the Kα12 helix in an open conformation (Figure 4v) allowing substrate easier access to the binding pocket. Secondly, the crystal structure of pig p110γ (PDB ID: 1E7U) [53] was used as it exhibits the Kα12 helix in a closed conformation (Figure 4v), inhibiting substrate access.

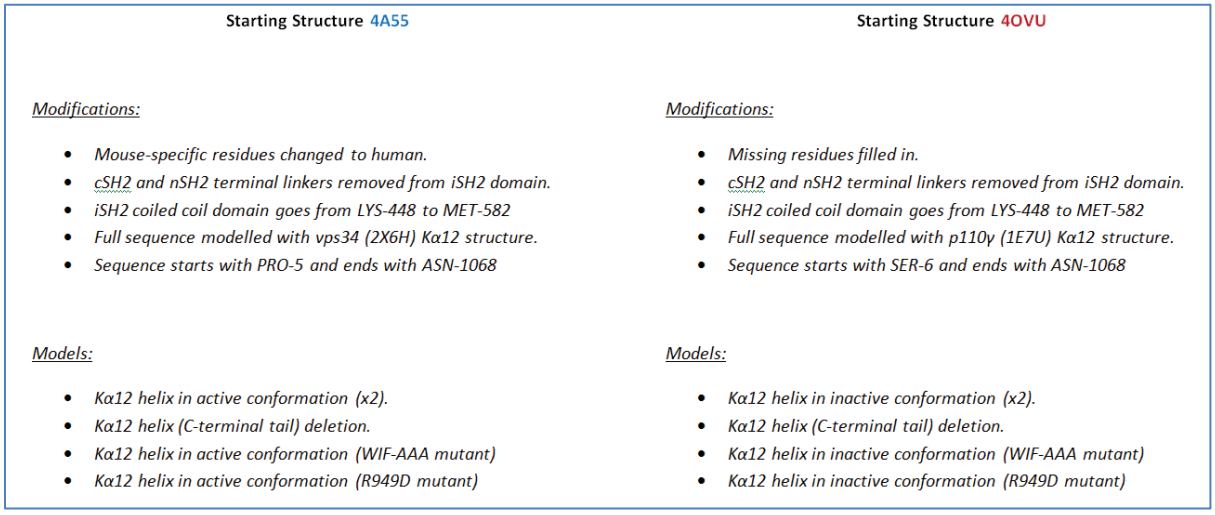


Figure 4iv - *Crystal structures used of mouse and human p110α-p85α, the modifications made to each prior to the commencement of model building, and a list of new constructs built for each. On the left, the Kα12 helix in its active conformation was added to PDB ID 4A55. On the right, the Kα12 helix in its inactive conformation was added to PDB ID 4OVU.*

To first ensure that the Kα12 helical structure was being modelled in an acceptable conformation, the complete (*i.e.* after the modifications described in Figure 4iv, other than addition of the Kα12 residues) 4A55 and 4OVU structures of p110α were aligned with each target structure (2X6H and 1E7U, respectively) using the RMSD and Stamp alignment tools in VMD. Following that, two

secondary structure prediction programs, Jpred and PredictProtein [54, 55], were used to ascertain whether the Kα12 residues from the human sequence were expected to be in a helix, and if so, the point at which the helix began and ended (Figure 4vi). Using this information, the *automodel* and *loopmodel* scripts provided with the Modeller program [56] were tailored and applied to map the coordinates of the residues of the Kα12 helix of human p110α onto those of the 2X6H or 1E7U structural templates, applying restraints to residues 1033 to 1047 (Kα11) and residues 1051 to 1064 (Kα12) (Figure 4vi).

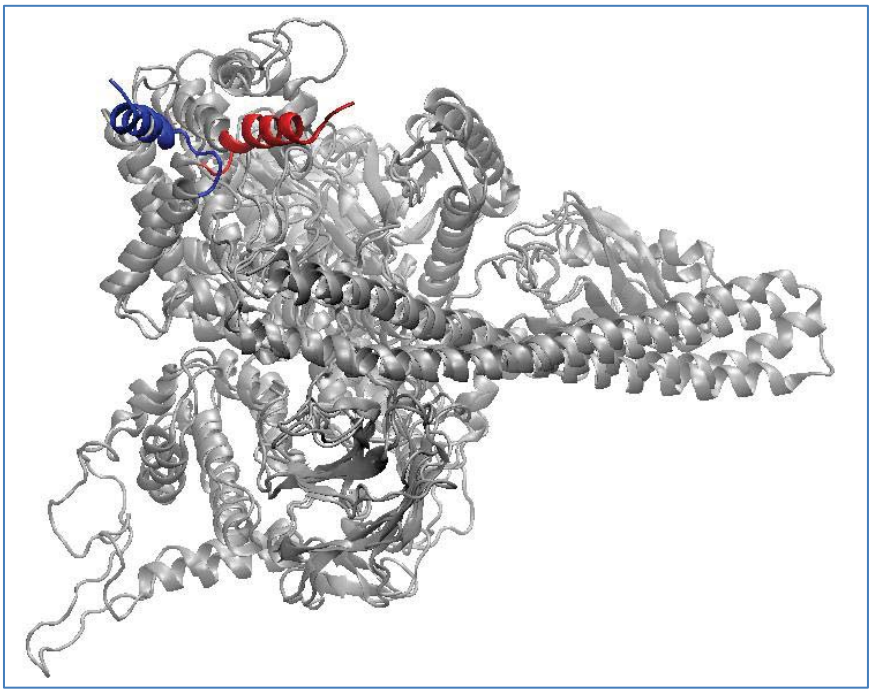


Figure 4v - *Alignment of the two complete human p110α-p85α structures (gray) with the modelled Kα12 helix shown in its open (blue) and closed (red) conformations.*

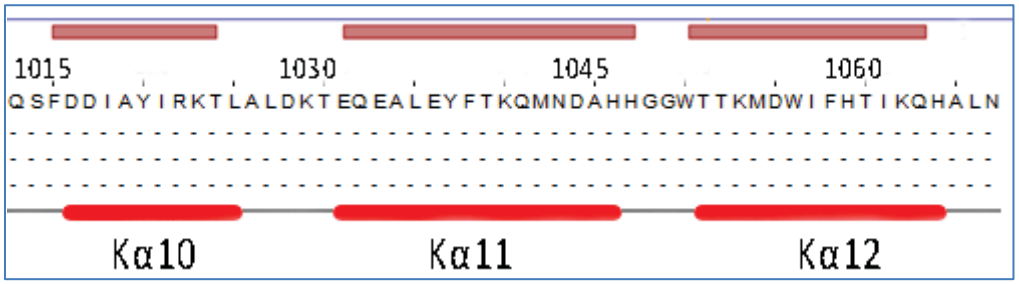


Figure 4vi - *The combined results of the secondary structure prediction programs, with PredictProtein [55] at the top and JPred [54] at the bottom. In each case, residues 1033 to 1047 and residues 1051 to 1064 were predicted to be in a helical conformation (PredictProtein represented by a mauve bar, JPred represented by a red bar). For reference, the residues numbered from 1013 to 1067 and their corresponding helices in the p110α-p85α structure have also been shown.*

Five models were built for each construct listed in Figure 4iv before Molprobity [57] was used to check for any steric overlap. Any errors due to atom overlap or bond irregularities were eliminated by substituting that block of residues with the same block from another model until a model with no errors remained. Each new construct then underwent energy minimisation and a 50 ns equilibration in solution following the simulation procedure detailed above.

4.2.3 - Interaction Energy Screening Methods

Rotational interaction energy screening between PI3Kα, and both a brain lipid bilayer and a neutral DPPC bilayer were carried out as detailed in Chapter 3.

4.2.4 - Analysis Methods

This section serves to give some background to each of the analytical tools used to characterise the behaviour of PI3Kα in the project, and the relevant inputs and parameters opted for in their use. Analysis of the simulations was primarily carried out using GROMACS 4.6 [58] tools, and simulations were visualised using VMD [59]. The methods listed here will be applicable throughout Chapters 4 and 5 unless otherwise stated.

Root-Mean-Square Deviation

The program *g_rms* was used to calculate the root-mean-square deviation (RMSD) of all atoms in a protein over the course of a simulation, using the initial coordinates as a reference structure. This deviation is then averaged over the entire protein at each time point.

Cluster Analysis

The program *g_cluster* was used to cluster protein structures over the course of a simulation using an all-by-all matrix of the RMSD of atom-pair distances to define the distance between structures. The gromos algorithm [60] was used to count the number of neighbours using 0.25 nm as a cut-off, with a cluster being defined as containing 50 neighbouring structures.

Eigenvector Calculation

The program *g_covar* was used to calculate and diagonalise a mass-weighted covariance matrix of the protein structure fitted to the initial coordinates as a reference structure. The primary eigenvector taken was representative of the most dominant conformational shift over the specified time frame before being analysed with *g_anaeig*. This subsequent program was used to plot the root-mean-square fluctuation of the Cα carbons in the protein structure over the specified time frame.

Root-Mean-Square Fluctuation

The fluctuation in residue position and its overall motion was determined using the program *g_rmsf*. This program computed the root-mean-square fluctuation (RMSF) of the protein atomic positions over the course of the simulation, which was then averaged for each residue.

Solvent Accessible Surface Area

The program *g_sas* was used to compute the total solvent accessible surface area (SASA) of each protein residue over the course of the simulation in two 15 ns phases, one before and one after membrane interaction. The solvent probe radius was specified at 0.14 nm which represents the radius of a water molecule.

Secondary Structure

The program *do_dssp* was used to compute the secondary structure of the protein for each time frame over the course of the simulation. This assigns the most likely class of secondary structure on the basis of the hydrogen bonding pattern, using the algorithm *Define Secondary Structure of Proteins* (DSSP) [61].

Distance Calculation

The program *g_mindist* was used to determine the minimum distance between two atoms or groups of atoms, for example to track protein-membrane or residue-lipid distances over the course of the simulation.

PIP_2 Substrate Alignment

The H1047R mutant system was broken down into coordinate files every 25 ns over the course of the 150 ns simulation. For each output frame, the protein was structurally aligned with that of the 4OVV crystal structure of PI3Kα with PIP_2 bound [35], with the PIP_2 coordinates also included in the aligned coordinate file. The RMSD between atoms in the headgroups of the PIP_2 substrate from the 4OVV structure and the PIP_2 substrate in the simulation was then calculated for each aligned output frame.

PIP_2 Hydrogen Bonding Network

The program *g_hbond* was used to compute all hydrogen bonds exhibited by the PIP_2 substrate for the 60 ns snapshot of the H1047R mutant simulation, which represented the point at which the substrate was closest to residue R949. The hydrogen bonds for R949 were also computed at this point. The water molecules to which they hydrogen bonded were determined, and the process

repeated one final time to compute the hydrogen bonds exhibited by each of these water molecules. This water-mediated hydrogen bonding network of PIP_2-H_2O-H_2O-R949 was visualised and represented in VMD using the Hydrogen Bonds analysis tool.

Per Residue Decomposition of Protein-Membrane Interaction Energy

The coordinates of the wild type and H1047R mutant systems were saved following 100 ns of simulation time. All residues within the 1.4 nm non-bonded cut-off distance (Section 4.2.1) of the membrane were recorded, and then listed as energy groups along with the membrane in a new simulation input file. A short unrestrained 25-step simulation was run using this input file and each set of coordinates, before the interaction energies between each residue and the membrane were calculated using the program *g_energy*. The total interaction energy between the protein and the membrane was calculated as a sum of the individual residue energies, and compared to the value obtained using the entire protein and the membrane as energy groups.

Potential of Mean Force

Equilibrated starting structures for the pulling simulation were obtained from the final snapshots of the wild type and H1047R mutant simulations in the presence of the brain lipid bilayer. The starting structures were then pulled away from the positionally restrained membrane along the z-axis over the course of 2 ns using a spring constant of 1000 $kJ.mol^{-1}.nm^{-2}$ and a pull rate of 0.01 $nm.ps^{-1}$ in NPT conditions identical to those described in section 4.2.1 with the exception of coupling algorithms and non-bonded cut-offs. The temperature was controlled using the Nose-Hoover thermostat [62, 63], while the pressure was regulated using the Parrinello-Rahman barostat [64]. The non-bonded cut-offs used were 1.4 nm for both Coulombic and Lennard-Jones interactions. The final center-of-mass distance between the protein and the membrane following the pulling simulation was 3.4 nm.

From this simulation, coordinates were generated at a protein-membrane distance every 0.2 nm to be used as the umbrella sampling windows. Each window was simulated in NPT conditions for 20 ns with a harmonic restraint enforcing the distance between the protein and membrane centre of mass with a force constant of 1000 $kJ.mol^{-1}.nm^{-2}$. The results were analysed using the weighted histogram analysis method as implemented in the program *g_wham*.

The free energy of binding (ΔG°) is related to the dissociation constant (K_d) between the protein and the membrane via the equation: $\Delta G^\circ = RT \ln \frac{K_d}{c}$, where *R* is the ideal gas constant (8.31 $J.K^{-1}.mol^{-1}$), *T* is the temperature at which the simulations were carried out, and *c*° is the standard reference concentration (1 $mol.L^{-1}$) .

ConAn

The program ConAn (Csaba Daday and Davide Mercadante) was used to implement the program *g_mdmat* to characterise intra-protein interactions within PI3Kα. These interactions are defined from a distance matrix consisting of the smallest distance between residue pairs. Computation of residue contacts was truncated at 1 nm.

Hydrogen Bonding Occupancy

A python script developed by Ricardo Soares was used to extract hydrogen bonds between residues generated by *g_hbond* describing their occupancy over the course of the simulation. A threshold of 15% occupancy was chosen. For every time block of 10 ns, the occupancy of donor-acceptor pairs involving the specified residues was averaged and plotted as a single point.

4.3 - Results and Discussion (Part I)

The results of this chapter will be split into two parts. The first part will detail observations and analysis of the effect of the mutations H1047R and E545K on the behaviour of the PI3Kα heterodimer (p110α-p85α) in the presence of a brain lipid bilayer. The second part on the other hand will focus more on the behaviour of the C terminal tail, and how any alterations to that tail affect the dynamics and function of p110α-p85α.

4.3.1 - Effect of Mutation

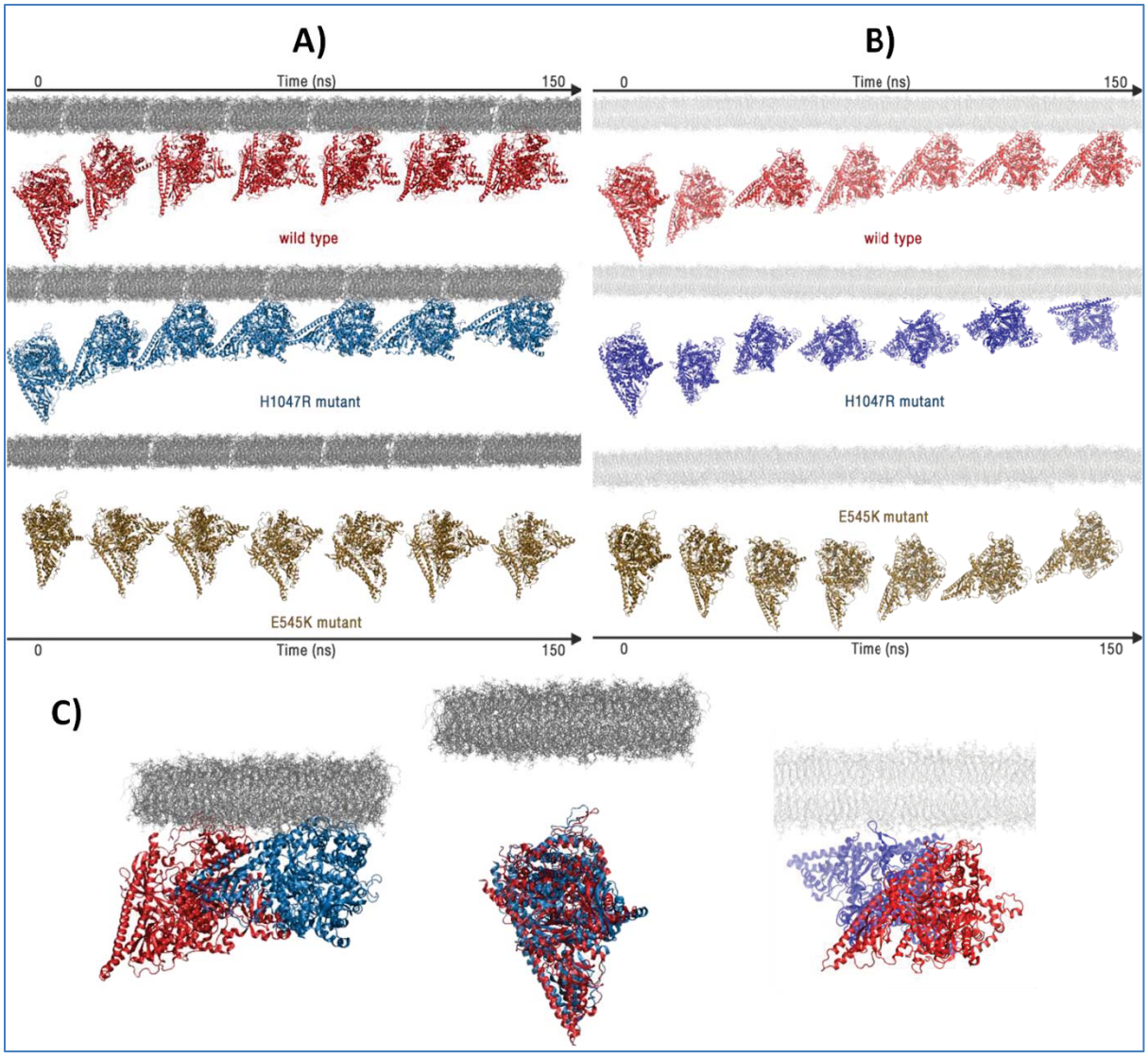


Figure 4vii - ***A,B)*** *Snapshots from the first (A) and second (B) runs of the MD simulations over the course of 150 ns shown at 25 ns intervals, revealing the position of the protein relative to the cell membrane (gray). The wild type is shown in red, the H1047R mutant in blue and the E545K mutant in gold. Water molecules have been removed for clarity.* ***C)*** *Snapshots of the initial (middle) and final (first run left, second run right) conformations of the wild type (red) and the H1047R mutant (blue) relative to the cell membrane are superimposed, highlighting their difference in orientation.*

The primary source of results and analysis in this section are the three systems (wild type, H1047R, and E545K) simulated for 150 ns along with a brain lipid bilayer in a solvated box, as described in the methods. These systems have been duplicated, and results will be displayed for both molecular dynamics runs unless otherwise stated. In Part I, analysis will be categorised as either structural or to do with membrane interaction. The latter section foregoes analysis of the E545K mutant, as p110α-p85α never came into contact with the brain lipid bilayer in these systems (Figure 4vii). The E545K mutant instead remained in the solvent, even adopting a more disordered state in the first run, with the p110α subunit and iSH2 domain of the p85α subunit drifting apart (Figure 4vii). However, in the second run, there is some movement towards the membrane as well as a change in orientation closer in similarity to the wild type and H1047R systems (Figure 4vii). Potential reasons for the lack of interaction with the membrane will be discussed in the following section.

With regards to the wild type and H1047R systems, both p110α-p85α complexes eventually approached and interacted with the brain lipid bilayer over the course of the simulations. This initial interaction occurred much earlier in the first run (approximately 30 ns), instead taking closer to 100 ns in the second run. However, that is the extent of the similarity between the systems, as both prior to and after initial contact with the membrane, each protein reoriented itself over the remainder of the simulations, settling on rather different final orientations in each case (Figure 4vii-C). To gain further insight into the reasons behind the difference in orientation, the structural and energetic properties of the system were calculated.

4.3.2 - Structural Analysis

4.3.2.1 - Root Mean Square Deviation

To determine how much the protein structure changes during the course of the simulation, the atom-positional root-mean-square deviation (RMSD) of all atoms from the initial conformation of the p110α-p85α complex at each point in time was calculated. RMSD gives a single value quantifying the structural deviation of the entire protein from its initial structure at each time point in the simulation, averaged over all residues of the protein. This calculation also sheds light on the point at which the protein structure attained equilibrium within the context of the system.

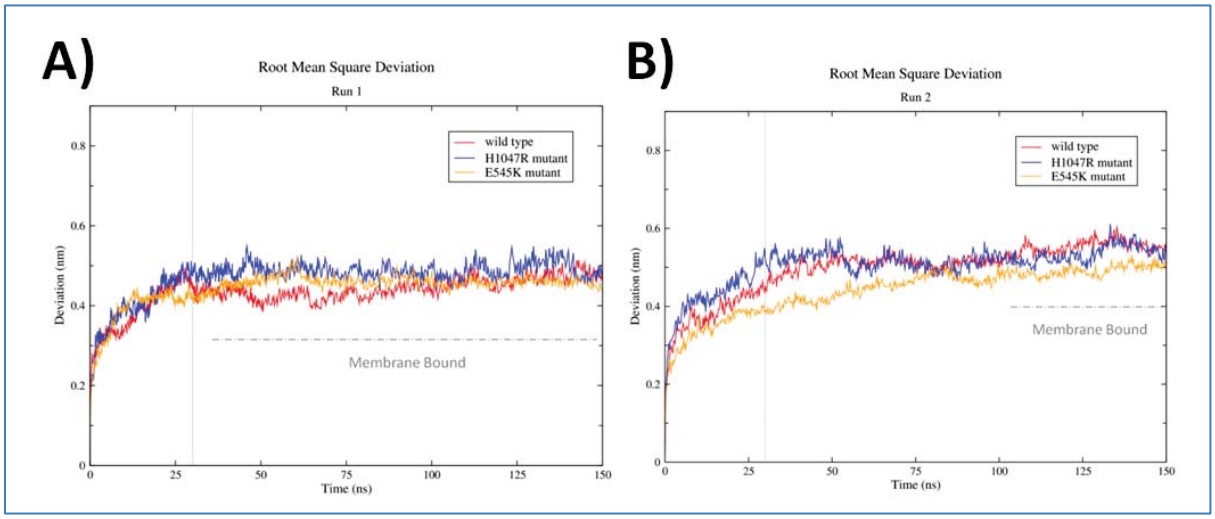


Figure 4viii - *A,B)* *Atom-positional RMSD of the p110α-p85α structures from their initial conformations over the course of the first (A) and second (B) 150 ns simulation runs in the presence of a cell membrane. The wild type is shown in red, the H1047R mutant in blue and the E545K mutant in gold. The dotted vertical line indicates the time at which the RMSD reaches a plateau for all three systems. The dashed horizontal line indicates the time in which the protein was membrane bound.*

Each system appears to have equilibrated at approximately 30 ns, shortly before interaction with the membrane in the cases of the wild type and the H1047R mutant in the first run. Despite similar equilibration times for the duplicate runs, wild type and H1047R take considerably longer to initiate contact with the membrane. While there is fluctuation in the RMSD of all systems after this convergence point, there is little overall upwards or downwards trend, indicating that the structure does not change markedly if and when the protein interacts with the membrane.

4.3.2.2 - Cluster and Eigenvector Analysis

Clustering of the protein structures exhibited across all time points in the simulation allows the determination of whether any major conformational changes were undertaken, and at what points these changes occurred in the simulation. This is done using an all by all structural comparison with the data represented as a matrix of all atom-positional RMSD values (upper left of each plot, Figure 4ix), and then sorted into distinct clusters of neighbouring structures with an all atom RMSD cut-off of 0.25 nm (lower right of each plot, Figure 4ix). Eigenvectors represent the trajectory of the major conformational changes, and in tandem with the cluster analysis, enable the pinpointing of which domains in the protein underwent the conformational change exhibited at a specific point in time.

Looking at the cluster maps (Figure 4ix), all simulations which approached the membrane can be broken down into three main segments, where the p110α-p85α complex adopts one or two distinct

structures within each of these. When linked back to the timeseries plot of each simulation (Figure 4vii), these segments can broadly be categorised as “in solvent” (S), “approach or interaction with membrane” (A), and “membrane bound” (M). Note that a cluster being assigned the same colour in two different simulations does not mean that the structures are similar, *i.e.* the clusters and their corresponding colours are assigned independently to each simulation.

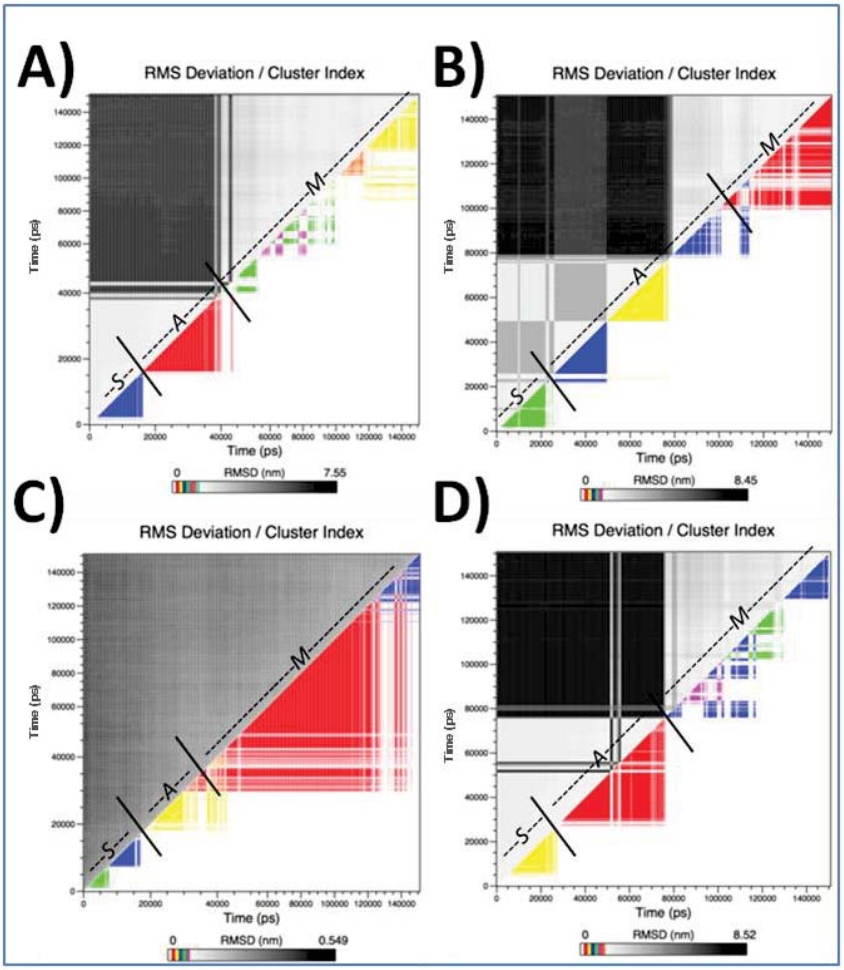


Figure 4ix - *Matrix map of protein structures coloured on the basis of their clustering where neighbouring structures are defined as having atom-positional RMSD values within 0.25 nm. Results from the first (A) and second (B) runs of the wild type system are shown at the top, while results from the first (C) and second (D) runs of the H1047R system are shown at the bottom. The segments representing solvent (S), membrane approach (A), and membrane bound (M) states are also shown.*

For the first wild type system, the S phase is denoted by a blue triangle, the A phase by a red triangle, and the M phase initially by green and purple triangles (Figure 4ix-A). In the duplicate system, the S phase is denoted by a green triangle, the A phase by blue and yellow triangles, and the M phase by a red triangle (Figure 4ix-B). For the first H1047R mutant system, the S phase is denoted by green and blue triangles, the first of which potentially highlights equilibration due to the

introduction of the mutation. The A phase is denoted by a yellow triangle, while the M phase initially by a red triangle (Figure 4ix-C). In the duplicate system, the S phase is denoted by a yellow triangle, the A phase by a red triangle, and the M phase by blue, pink, and green triangles (Figure 4ix-D).

Keeping in mind that the duplicates (approximately 100 ns) interact with the membrane considerably later than the original simulations (approximately 35 ns) (Figure 4viii), this essentially means that two major conformational changes take place in all four simulations. The first occurs as the protein recognises the membrane and approaches it, eventually making contact; and the second occurs once contact has been established, at which point the protein potentially adopts a conformation suitable for lipid binding and its catalytic function.

As the system duplicates adopted similar phases for wild type and H1047R, the major structural changes for one of the duplicates were calculated in the form of eigenvectors, encompassing the time frame across the period in which each conformational shift was undertaken. For example, the transition from the S phase to the A phase in the duplicate run of the H1047R mutant system occurred after 30 ns, and so the main eigenvector for the time frame between 15 ns and 45 ns was calculated. The RMSF (Section 4.3.2.3) of the Cα carbons in the protein structure over the course of this time frame was calculated to shed insight onto the primary influence behind the eigenvector describing the structural change.

Prior to membrane interaction, both the wild type and H1047R mutant underwent conformational changes in similar regions of the protein (Figure 4x). The first of these was in the RBD, which is understandable considering the *Ras* protein plays a part in PI3Kα's membrane recruitment, and as such PI3Kα potentially prepares for *Ras* binding when in the presence of a membrane. Continuing along the residue sequence, the following two conformational changes were in very mobile loops preceding and following the C2 domain, potentially reorienting it in preparation for membrane interaction. The membrane binding domain between residues 863 and 873 also exhibited structural changes during membrane approach, along with the activation loop between residues 940 and 960, and the C terminal tail between residues 1050 and 1064. These last three blocks of residues play some role in membrane interaction and catalytic function [11, 65], and will be explored in further detail in section 4.3.3. Interestingly, whilst the membrane binding domain and C terminal tail continued their conformational changes to a lesser degree following first contact with the membrane, the activation loop did not. The activation loop is home to arginine 949 amongst other residues, whose primary function is as a PIP_2 substrate recognition site [11]. Once a PIP_2 molecule is located in the initial membrane interaction, it is plausible that the activation loop needs to stay relatively static to allow for PI3Kα's catalytic function.

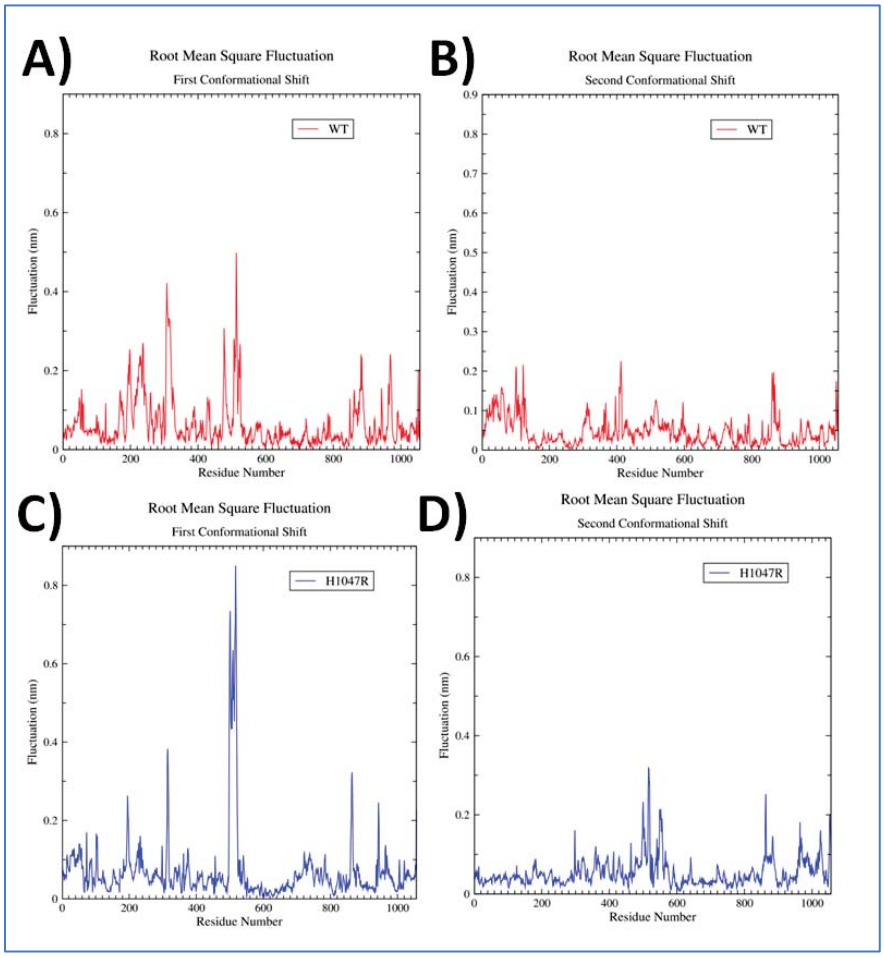


Figure 4x - ***A,B)*** *RMSF of the Cα carbons in the wild type protein over the course of 15 ns during the first (A) and second (B) conformational changes seen in Figure 4ix.* ***C,D)*** *RMSF of the Cα carbons in the H1047R protein over the course of 15 ns during the first (C) and second (D) conformational changes seen in Figure 4ix.*

Differences between the wild type and H1047R mutant are mostly seen following membrane interaction (Figure 4x-B,D). Whilst both continue to exhibit conformational changes in and around the C2 domain (residues 324 - 474), the H1047R mutant does so to a greater degree which may play a part in increasing the protein's membrane affinity - the proposed effect of the mutation. The wild type undergoes a conformational change in the ABD (residues 1 - 105), which is unseen in the H1047R mutant. The ABD is expected to bind to regulatory SH2 domains found in the p85α subunit and potentially plays a part in restricting the level of membrane binding in the functional wild type protein [13]. Membrane interactions seen in these systems will be explored in greater depth in section 4.3.3.

The cluster analysis was also repeated for the wild type and H1047R systems, this time using crystal structures as a reference template; the E545K mutant structure has yet to be crystallised. In the case

of the wild type structure, PDB ID 4OVU [35] was used as the best representation of the wild type despite the presence of the nSH2 domain; and in the case of the H1047R structure, PDB ID 3HIZ [33] was used as the best representation of the H1047R mutant and also included the nSH2 domain. The aim was to see if any conformations sampled throughout the simulation adopted a similar structure to that of the crystallised protein once the nSH2 domain coordinates were omitted.

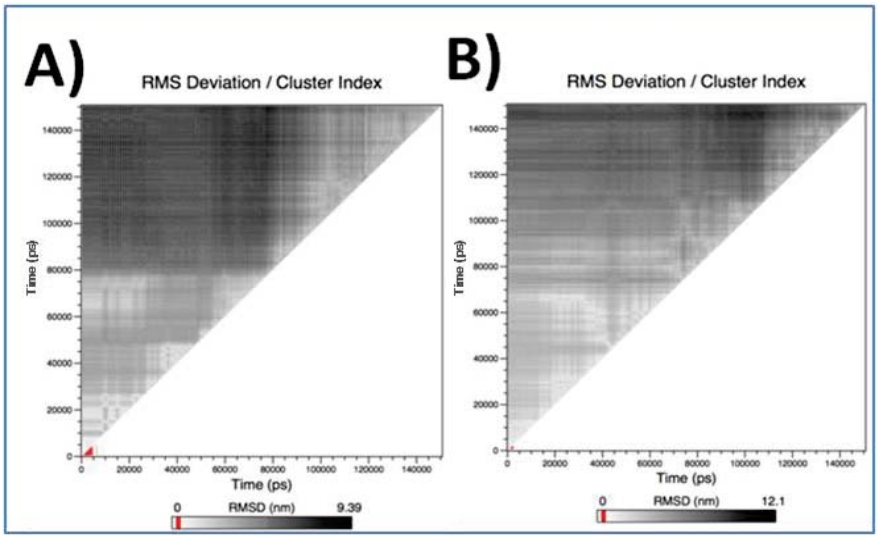


Figure 4xi - *Matrix maps of the wild type (A) and H1047R protein structures coloured on the basis of its clustering with the 4OVU crystal structure [35] and 3HIZ crystal structure [33] respectively, over the course of the 150 ns wild type (A) and H1047R (B) simulations. Neighbouring structures are defined as having atom-positional RMSD values within 0.25 nm.*

It was found in both cases that the static structures of the p110α-p85α complex captured through crystallography were only represented in the early stages of the simulation prior to full equilibration (Figure 4xi). There was no similarity detected as the protein structure approached, interacted with, and eventually bound to the membrane. This would suggest that these crystal structures of the p110α-p85α complex are not representative of the conformations adopted by the protein in its membrane-bound state. Furthermore, any intra-protein interactions seen in the static crystallised structure are not guaranteed to exist *in vivo* when PI3Kα is catalytically active, and these simulations may prove more efficient at mapping the nature of the PI3Kα-membrane interaction.

4.3.2.3 - Root Mean Square Fluctuation

While the RMSD quantifies the structural deviation of the entire protein averaged over all residues, the RMSF is the RMSD of each atom (or the group of atoms making up a residue) from its average position during the course of the simulation, averaged over time (and, often, summed over atoms in a residue). Large RMSF values will therefore refer to those residues whose positions deviate most

relative to the average structure of the protein during its approach and/or interaction with the membrane.

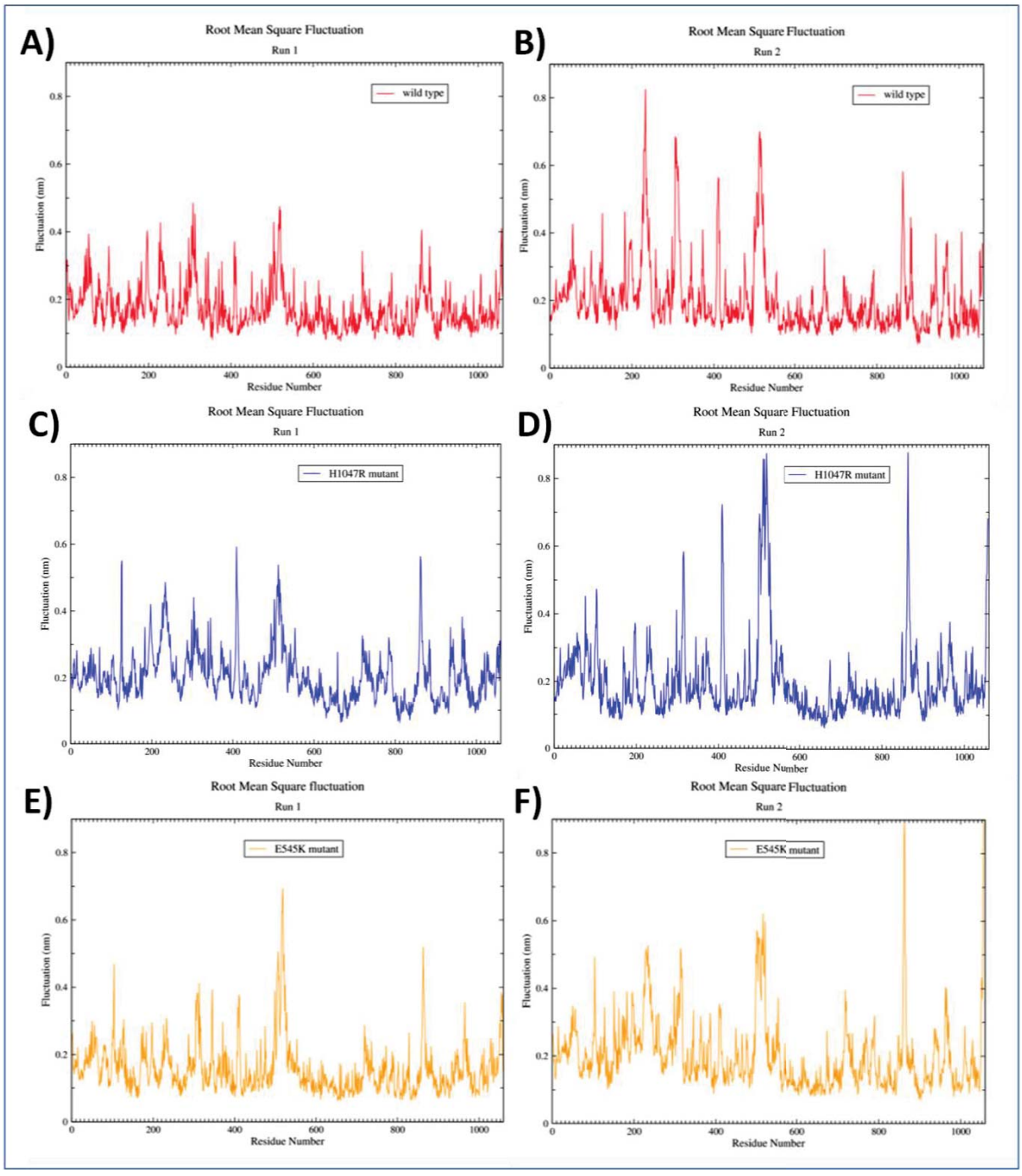


Figure 4xii - *Atom-positional RMSF of each p110α-p85α residue over the course of the 150 ns simulations in the presence of a cell membrane. The first (A) and second (B) runs of the wild type system simulations are shown in red; the first (C) and second (D) runs of the H1047R system simulations are shown in blue; the first (E) and second (F) runs of the E545K system simulations are shown in gold.*

The RMSF of each residue revealed which residues exhibit the most conformational changes during the simulation (Figure 4xii). Across duplicate runs for each isoform, the residues of the p110α-p85α complex which fluctuated most were essentially identical, although they exhibited considerably more fluctuation in the second run. However, it is important to note that the RMSF merely indicates the extent of conformational change and not the nature of it, so it fails to differentiate between functionally linked conformational change, and a mobile loop in the structure. A value of 0.4 nm was chosen as the cut-off point for a residue having exhibited considerable and significant fluctuation as it represented a twofold increase in the modal fluctuation of the majority of the residues.

Across all systems, there were distinct blocks of residues belonging to the p110α subunit which underwent considerably more fluctuation than the others (Figure 4xii). These are residues 210-225, 300-310, 500-515, 860-875, and 1050-1065. There was also mutant specific fluctuation in residues 405-415 in the H1047R mutant system; and residues 100-110 in the E545K mutant system. Each residue block will be addressed separately.

The first set of residues is found in the *Ras* Binding Domain, which as the name suggests is charged with *Ras* protein recruitment. As the p110α-p85α complex recognises the membrane, it can be expected to undergo conformational changes in this region to allow for *Ras* binding [18]; however, *Ras* was not included in the system and as such this cannot be further tested. This structural change is also seen in the cluster analysis.

The next two sets of residues also found in the cluster analysis are loops found on either side of the C2 domain, which is involved in membrane binding across other protein families [19, 20]. While the loops themselves are isolated in solution in the simulations, the C2 domain makes contact with the membrane. This contact could have been facilitated by the dynamics of these loops on either side.

The final two residue blocks which exhibit motion in all systems are found in two regions associated with membrane interaction. Whilst the E545K system does not approach the membrane, it can be seen in the cluster analysis that conformational changes in these regions occur mainly prior to interaction, and potentially as positioning for optimal membrane contact. The block of residues 860-875 is situated between the N lobe and C lobe of the p110α kinase domain and has previously been proposed as a membrane binding domain of PI3Kα [10]. This area of the protein is a part of the catalytic site and, according to experimental results [39], is in the region which contacts the membrane and will be explored in section 4.3.3. The block of residues 1050-1065 is situated in the C terminal tail of the p110α kinase domain. This area of the protein is a part of the catalytic site and is in close contact with the bilayer. This loop is also preceded by the mutation hotspot H1047R which is

understood to affect the activity of PI3Kα by facilitating its interaction with the membrane. It is also the location of the highly conserved and important WIF (Trp, Ile, Phe) motif found at residues 1057-1059, proposed to insert itself into the membrane and undergo hydrophobic interactions [11, 39]. Any membrane interactions exhibited by these loops will be explored in section 4.3.3.

4.3.2.4 - Solvent Accessible Surface Area

Solvent accessible surface area (SASA) represents the extent to which each atom or residue in the protein is accessible to the solvent. When a change in SASA is observed in terms of interaction with the membrane, it can help identify residues which have undergone either a conformational change to become more exposed to or shielded from solvent, or the residues which have made contact with the lipids and potentially buried themselves into the membrane. This makes the analysis here most relevant for the wild type and H1047R systems, but not the E545K system. As such, SASA of each residue for the first two systems was calculated before and after interaction with the membrane for each run, and plotted on the same graph (Figure 4xiii). The maximum solvent accessible surface area of any residue across all systems was approximately 2 nm^2. As such, a cut-off of 0.50 nm^2 representing 25% of that maximum was used as a threshold to find significant differences in area before and after membrane interaction. Residue blocks which exhibited increased solvent exposure were denoted by a red arrow, whilst those with decreased exposure were denoted by a blue arrow.

Following membrane interaction, the wild type and H1047R systems primarily exhibited increased solvent exposure in three similar regions - the linker between the ABD and RBD, the linker between the C2 and helical domains, and the linker between the iSH2 and nSH2 domains (Figure 4xiii). Considering the conformational changes previously observed in sections 4.3.2.2 and 4.3.2.3, the first two linkers may have become more exposed to solvent in the reorientation of the RBD and C2 domains. As the nSH2 domain is not present in this system, no functional insight can be gleaned from the SH2 linker, as its increased mobility is most likely due to the absence of the nSH2 domain.

Decreased solvent exposure on the other hand was more substantial in the H1047R system in comparison to the wild type system (Figure 4xiii). This can be expected, as from the time series (Figure 4vii) it could be seen that H1047R presented more of its surface to the membrane, lying almost parallel, whilst the wild type system sat more perpendicular to the membrane surface. In the wild type system, there was decreased solvent exposure in the C2 and kinase domains, which were located at the protein-membrane interface. This was also seen to a greater degree in the H1047R systems, with decreased solvent exposure also observed down the length of the protein incorporating the iSH2 domain, and considerably more residues in the kinase domain.

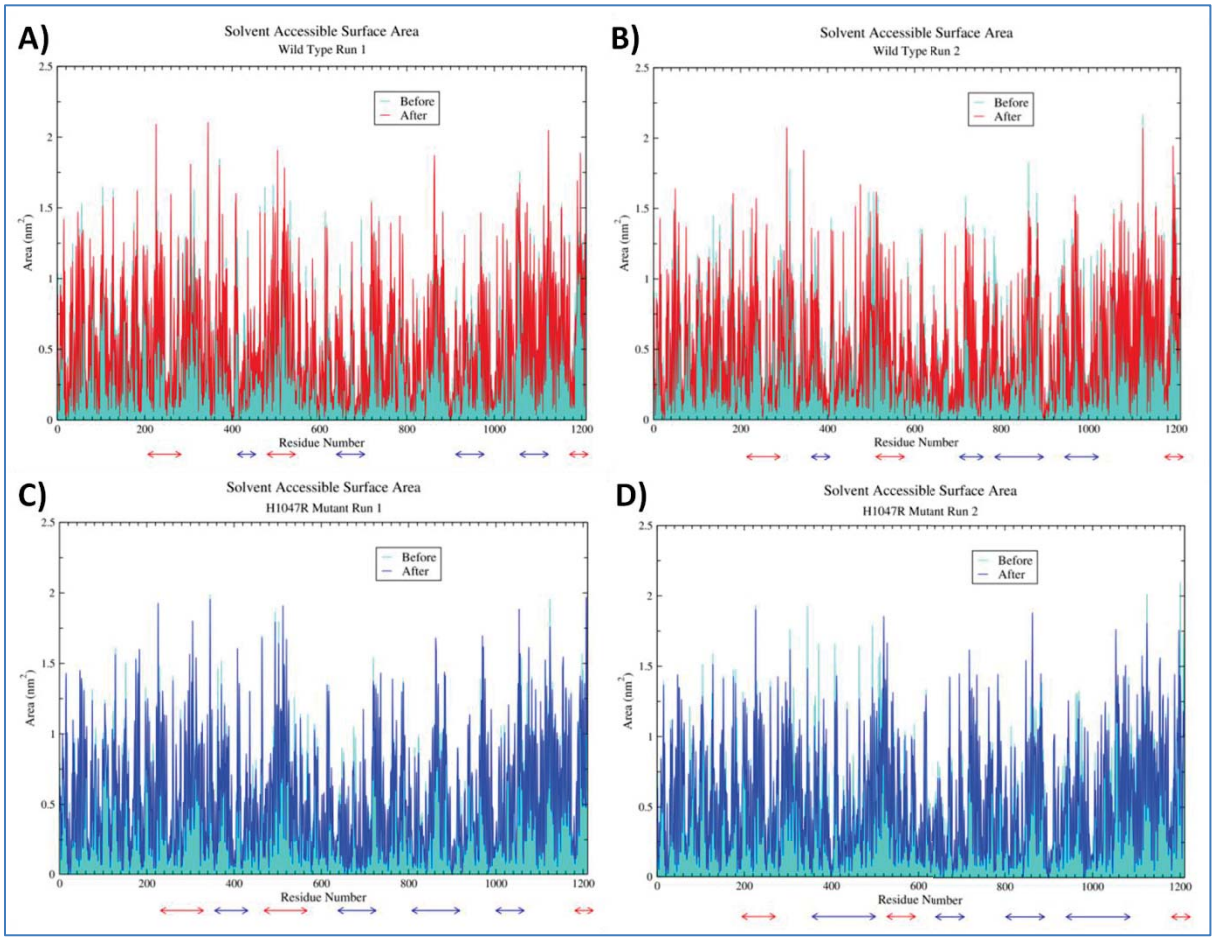


Figure 4xiii - *SASA of each protein in 15 ns phases before and after coming into contact with the cell membrane. In each case the SASA before membrane contact is shown in cyan. The SASA of the protein after membrane contact in the first (A) and second (B) runs of the wild type system simulations are shown in red; the SASA of the protein after membrane contact in the first (C) and second (D) runs of the H1047R system simulations are shown in blue. Blocks of residues which exhibit a significant decrease (blue arrow) and increase (red arrow) in SASA are denoted below the x axis.*

4.3.2.5 - Comparison to HDx-MS Data

The hydrogen deuterium exchange mass spectrometry (HDx-MS) data measured experimentally for p110α-p85α in the presence and absence of a cell membrane [38, 39] provide a means of validating the MD simulations. These data were available in the form of the percentage change in HDx between different conditions (and between wild type and mutants) of short (5-10 residues) partially overlapping peptides obtained by protease digestion of p110α-p85α prior to MS. While the amount of hydrogen-deuterium exchange depends at least in part upon the degree of solvent exposure, the nature of the HDx-MS data made it difficult to directly compare to the solvent accessible surface area (SASA) calculated from the MD simulations. Instead, the final snapshot from the MD simulations of the wild type and the H1047R mutant were coloured according to the HDx-MS data, and compared to the similar analysis carried out by Burke *et al.* [39] on the crystal structures and their proposed model of p110α-p85α membrane interaction (Figure 4xiv).

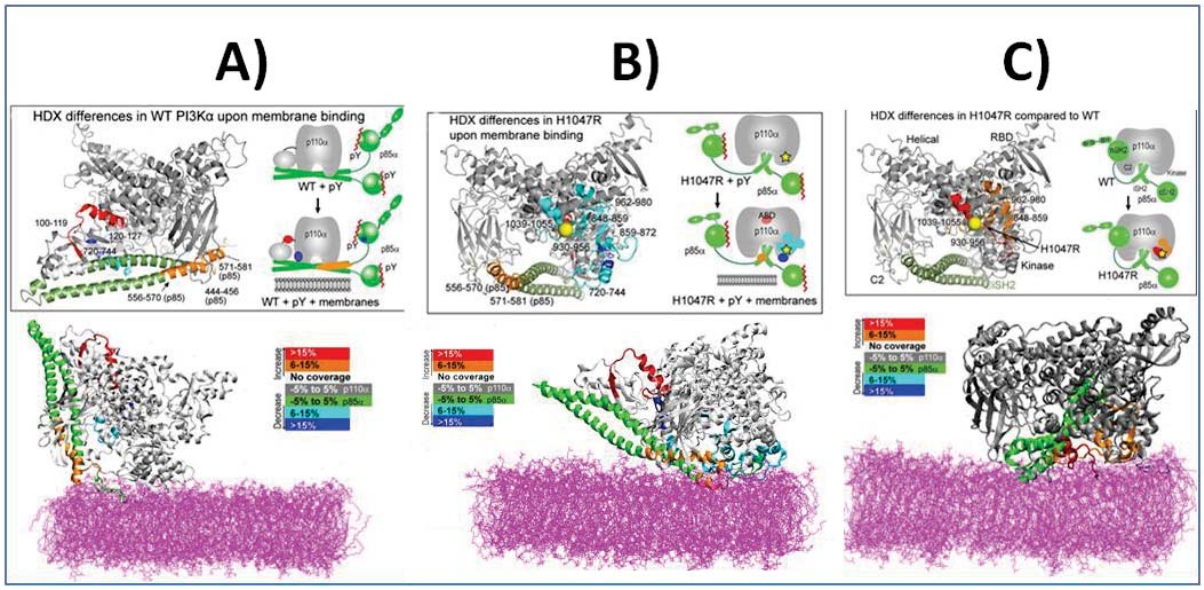


Figure 4xiv - *Changes in HDx of activated p110α-p85α upon interaction with the cell membrane. The upper panels are taken from Burke* et al. *[39], showing the change in HDx upon activation by phosphotyrosine (pY) and membrane binding of wild type (A) and H1047R (B) mapped onto the crystal structure of p110α-p85α and onto a cartoon representation of their proposed model of protein-membrane interaction. The lower panels show the final snapshot from the first runs of the 150 ns simulations of wild type (A) and H1047R (B) systems coloured and oriented similarly, with the membrane shown in pink. The scale of relative deuteration is shown in the middle. The differences between the changes in HDx observed in the two systems are identically represented in (C) on the final snapshot of the H1047R system.*

Firstly, the two systems which approached and interacted with the membrane (wild type and H1047R) were each coloured on the basis of the blocks of residues which exhibited either an

increase or decrease in relative deuteration upon interaction with the membrane in the HDx-MS data for that protein (Figure 4xiv-A,B). Following this, the difference in the change in relative deuteration upon interacting with the membrane between wild type and H1047R was calculated and mapped onto the H1047R system to show the change in membrane interaction due to the presence of the mutation (Figure 4xiv-C). The H1047R mutant exhibits more significant decrease in deuteration along the kinase domain surface upon interaction with the membrane, indicating that this is the part of the protein that contacts the membrane in the HDx assays (Figure 4xiv-B). This is replicated by the simulation data, where the final snapshots show the H1047R mutant system almost parallel to the membrane, with the same interface presented. On the other hand, the wild type system presents itself to the membrane in an orientation less parallel than that of H1047R, which results in a less substantial change in deuteration along the kinase domain (Figure 4xiv-A,C).

4.3.2.6 - Secondary Structure Analysis

Secondary structure analysis was carried out on the C terminal tail of PI3Kα in each system to monitor its behaviour in solution, as well as in the presence of a membrane. The C terminal tail consists of the Kα11 (residues 1031 to 1046) and Kα12 (residues 1051 to 1062) helices, although the latter has not been crystallised in a helical form in the p110α-p85α complex. The lack of electron density describing the PI3Kα Kα12 structure indicates that it is either too disordered to be helical or is highly mobile. Considering the helical nature of Kα12 in analogous proteins VPS34 and p110γ (PDB IDs: 2X6H and 1E7U, respectively - section 4.2.2), the latter explanation is more likely. Kα12 is also seen as a helix by secondary structure prediction programs (Figure 4vi), and has been proposed to be present in either a closed conformation as seen in 1E7U or open conformation as seen in 2X6H (Figure 4v). Kα12 is expected to be in a closed form in the cytosol but an open form on the membrane, allowing interaction and subsequent insertion [52].

As these simulations were run from the 4A55 crystal structure as the starting point, Kα12 started off in a non-helical form, whilst Kα11 was in a well defined helix. Considering Kα12's importance in wild type PI3Kα membrane binding, and the relative position of the H1047R mutation between the Kα11 and Kα12 helices, it is possible the H1047R mutant increases membrane affinity through induced conformational changes in and around the C terminal tail. Looking at the predicted secondary structure analysis of these simulations, Kα11 remains in a well-defined α-helix in all three systems (Figure 4xv). Kα12 on the other hand is only seen as an α-helix briefly in the E545K system (Figure 4xv-C) which remains in solution. It is interesting to note that the two systems which do interact with the membrane (wild type and H1047R) exhibit a less defined structure in Kα12 in comparison to the E545K mutant (Figure 4xv-A,B), especially following membrane interaction (at approximately 100

ns). The behaviour of the C terminal tail and its effect on membrane binding will be explored in further detail in section 4.4.

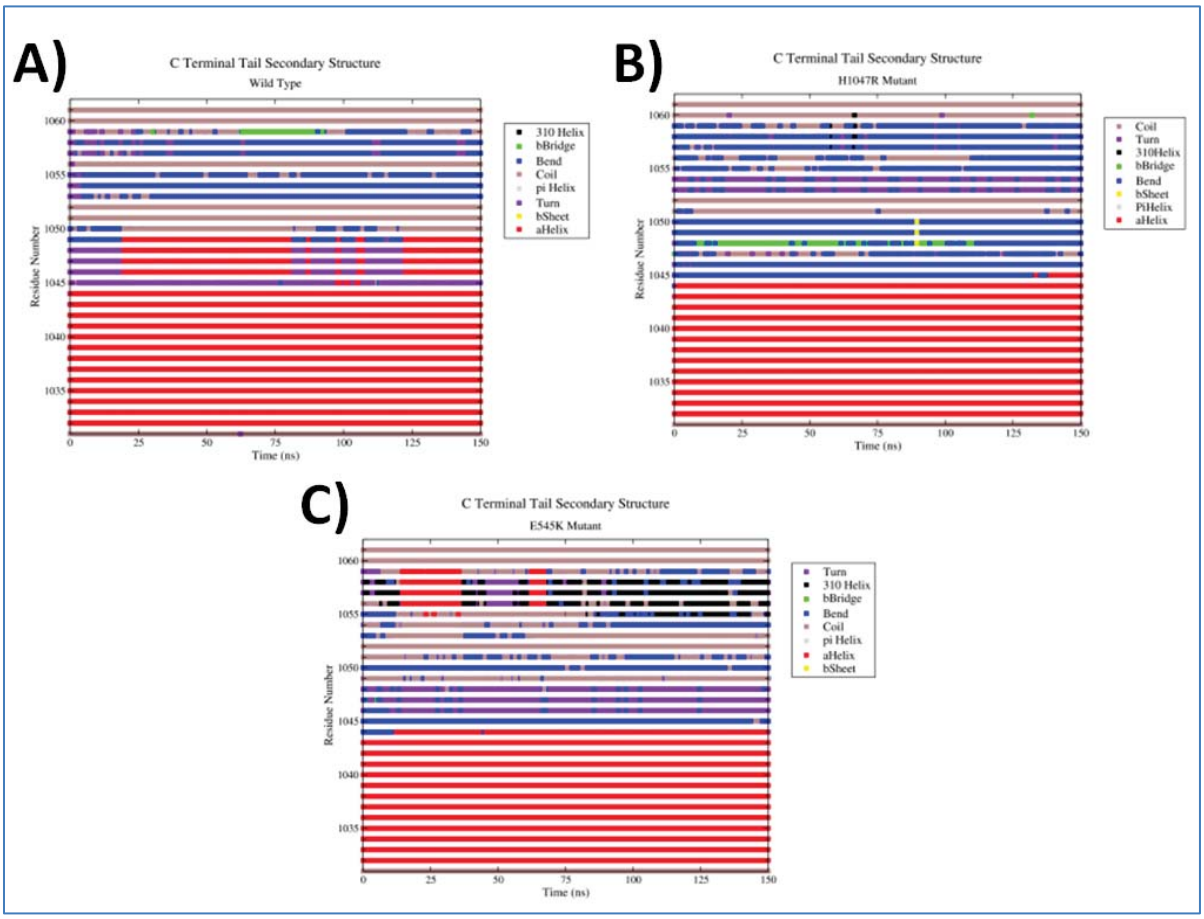


Figure 4xv - *Secondary structure analysis of the C terminal tail in the p110α subunit comprising of Kα11 (residues 1031 to 1046) and Kα12 (residues 1051 to 1062) over the course of the 150 ns simulations of the wild type (A), H1047R (B), and E545K (C) systems. Colouring is done on the basis of DSSP definitions of coils (brown), turns (purple), bends (blue), β-bridges (green), β-sheets (yellow), 3_{10}-helices (black), π-helices (gray), and most importantly α-helices (red).*

4.3.3 - Membrane Interaction

The structural analysis of all three systems and their duplicate simulations were taken into consideration before the protein-membrane interactions were analysed.

Firstly, the E545K mutant exhibits no membrane interaction throughout 150 ns of simulation, and as such will be exempt from this section. A repeat of this behaviour gives more confidence to the possibility of the E545K p110α-p85α complex not being viable to carry out membrane binding as quickly as the wild type and H1047R systems. A change in orientation towards the latter segment of the second run suggests that the E545K system may yet interact with the lipid bilayer, but on a longer timescale than used in this research. It is interesting to note that the E545K mutant's activity

has been found to be uninhibited by p85α but is further activated by *Ras* [32-34], a component not present in the systems analysed, which may explain the slower membrane interaction. The opposite is true for the H1047R mutant, whose increased activity is largely dependent on increased membrane affinity [32-34], a feature which will be explored in this section.

Secondly, in the case of the wild type and H1047R systems, the proteins adopt identical orientations with respect to the membrane across the duplicate simulations. As seen in section 4.3.2.3, the residues and domains which present themselves to the brain lipid bilayer are identical across duplicate simulations, with the most notable difference between the duplicate simulations being how quickly they approached the membrane. Both H1047R simulations but neither wild type simulation replicated the HDx data (section 4.3.2.5), and therefore no particular system could be chosen for further analysis of the membrane interaction based on this criteria. As such, analysis centered on the membrane interaction of each system will primarily incorporate the first run of the systems, as they spent the majority of the 150 ns simulations in contact with the brain lipid bilayer. Results from the second run of each simulation are referenced by figures found in Appendix D.

4.3.3.1 - Distance Analysis

The minimum distance between each of the identified membrane binding domains (MBDs) [65] and the anionic lipids found in the membrane were calculated over the course of the wild type and H1047R mutant simulations. These were residues 721-727 (MBD1), 863-873 (MBD2), and 966-974 (MBD3). As the WIF motif (residues 1057-1059) is expected to make a hydrophobic interaction, the minimum distance to all lipids was calculated instead for these residues (Figure 4xvi, Figure Di).

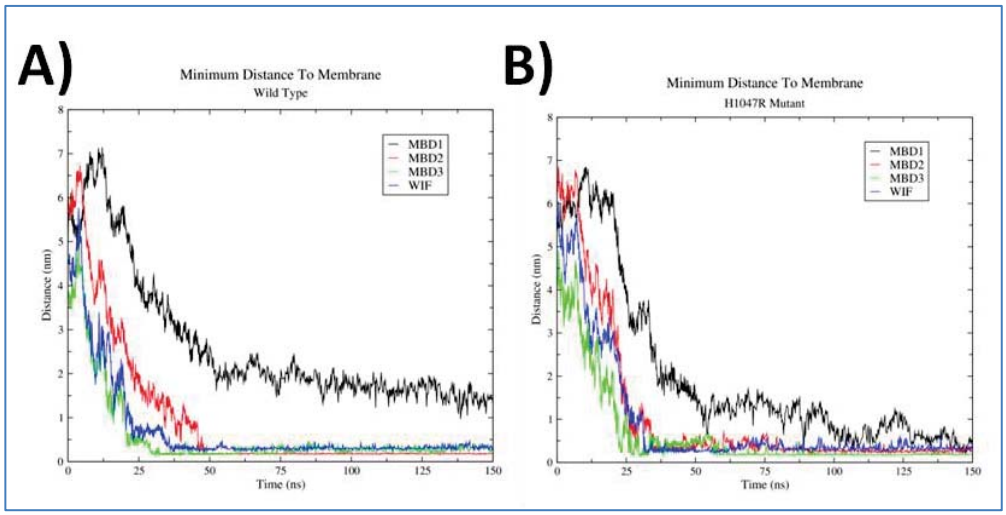


Figure 4xvi - *The minimum distance between MBDs (1-black, 2-red, 3-green, WIF-blue) and the cell membrane over the course of the first run of the 150 ns simulation of the wild type (A) and H1047R (B) systems.*

These protein-lipid distances were also broken down on a per-residue basis, looking at which of the residues which constitute the MBDs made first contact, and which residues embedded the furthest into the membrane making the closest contact. To do so, snapshots of each simulation were taken of the MBDs at time points where the protein-membrane distance was closest, focusing on the residues within each MBD that the lipids were interacting with. A common feature in all MBDs is the presence of lysines, typically flanked by hydrophobic residues. Lysine, a basic residue, would interact favourably with the anionic lipids, whilst the hydrophobic residues could potentially embed themselves beyond the level of the charged headgroups (Figure 4xvii, Figure Dii).

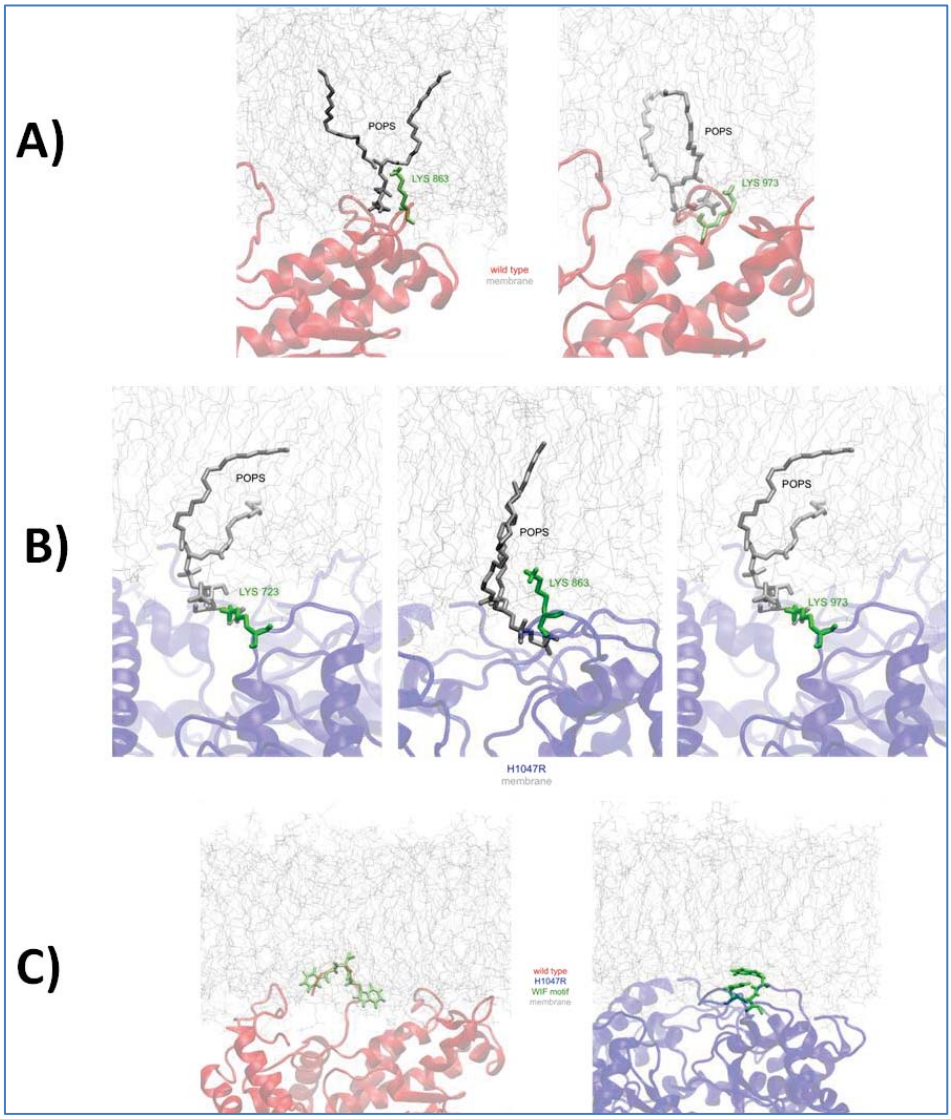


Figure 4xvii - *Final snapshots of the wild type (red) and H1047R (blue) proteins following the first run of 150 ns simulations in the presence of a cell membrane (gray). Key residues (green) and key lipids (dark gray) are highlighted.* ***A)*** *From left to right: LYS 863 (MBD2) and LYS 973 (MBD3) seen interacting with a POPS lipid.* ***B)*** *From left to right: LYS 723 (MBD1), LYS 863 (MBD2) and LYS 973 (MBD3) seen interacting with a POPS lipid.* ***C)*** *Embedment of the C terminal tail into the cell membrane showing the position of the WIF motif in the wild type (left) and H1047R (right) systems.*

In the case of the wild type system, MBD1 does not approach the membrane closely enough to interact, but does so slowly in the H1047R mutant system albeit not closely enough to embed itself (Figure 4xvi, Figure Dii). As seen in Figure 4vii, this agrees with the larger PI3Kα-membrane interface established by the mutant, and also plays a part in H1047R having increased membrane affinity.

For both the wild type and H1047R mutant systems, the same basic residues are found to be interacting with anionic lipids. In the case of MBD1 (only in H1047R), LYS 723 is the closest to a POPS lipid; for MBD2, it's LYS 863; and for MBD3 it's LYS 973 (Figure 4xvii-A,B). In both systems, the WIF motif in the C terminal tail is the most deeply embedded set of residues in the protein (Figure 4xvii-C) which solidifies theories behind its anchoring of wild type PI3Kα to the membrane [11].

4.3.3.2 - Substrate Analysis

PIP_2 is a substrate of PI3Kα, and a PIP_2 binding pocket has been observed by Miller *et al.* [35] based on the 4OVV crystal structure, in which a PIP_2 molecule is bound to PI3Kα. The specific PIP_2 lipid molecule closest to this binding pocket over the course of the simulation was identified in the first run of the H1047R system. The distance between each residue of the protein and this potential PIP_2 substrate was tracked over the course of this simulation. The residue closest to the PIP_2 substrate at any point in time was the arginine at position 949 (R949), which has been previously identified as a PIP_2 recognition point [66]. The distance between relevant regions of the protein (the binding pocket identified by Miller *et al.* [35] and R949) and the nearest PIP_2 molecule are shown in Figure 4xviii and Figure Diii. The PIP_2 interaction was substantially later in the second run of simulations (Appendix D).

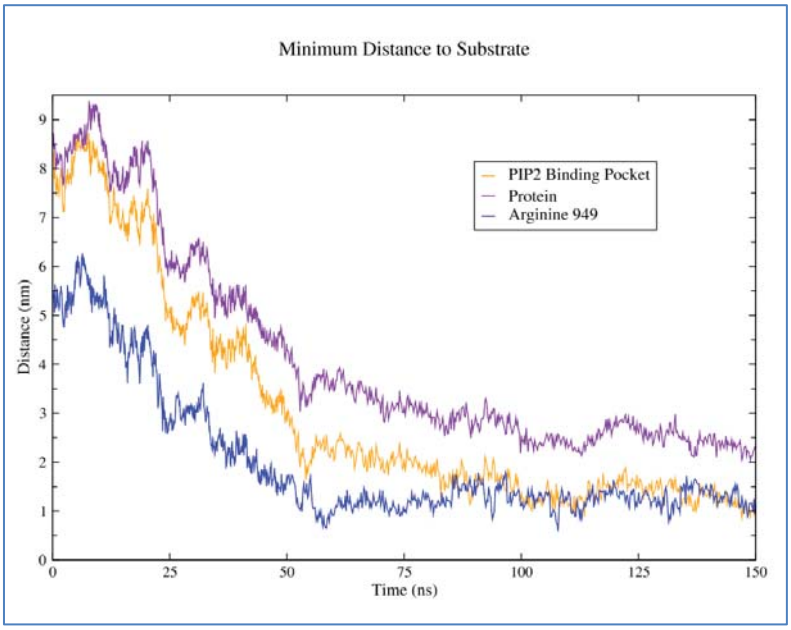


Figure 4xviii - *Centre of mass distance between p110α-p85α (purple) and the identified PIP_2 substrate over the course of the first run of the H1047R system simulation. The minimum distances of the PIP_2 binding pocket (orange) and R949 (blue) to the same PIP_2 substrate are also shown.*

At approximately 60 ns and 110 ns into the simulation, the R949 residue comes within 0.7 nm of the PIP_2 substrate. Interestingly, in between these two time points or interactions, the protein appears to "fall over" onto the membrane (Figure 4vii) as the PIP_2 substrate slowly moves out of the lipid bilayer. This is highlighted by the relatively rapid reduction in the distance between the binding pocket and the substrate (Figure 4xviii). To better understand the potential interaction between the R949 residue and the PIP_2 substrate, a detailed structural and hydrogen bonding analysis of the system was carried out at approximately 60 ns, the time point of the first interaction and initial substrate recognition (Figure 4xviii). While 0.7 nm is not close enough for direct contact between the R949 residue and the PIP_2 substrate, there remained the possibility of a water-mediated contact.

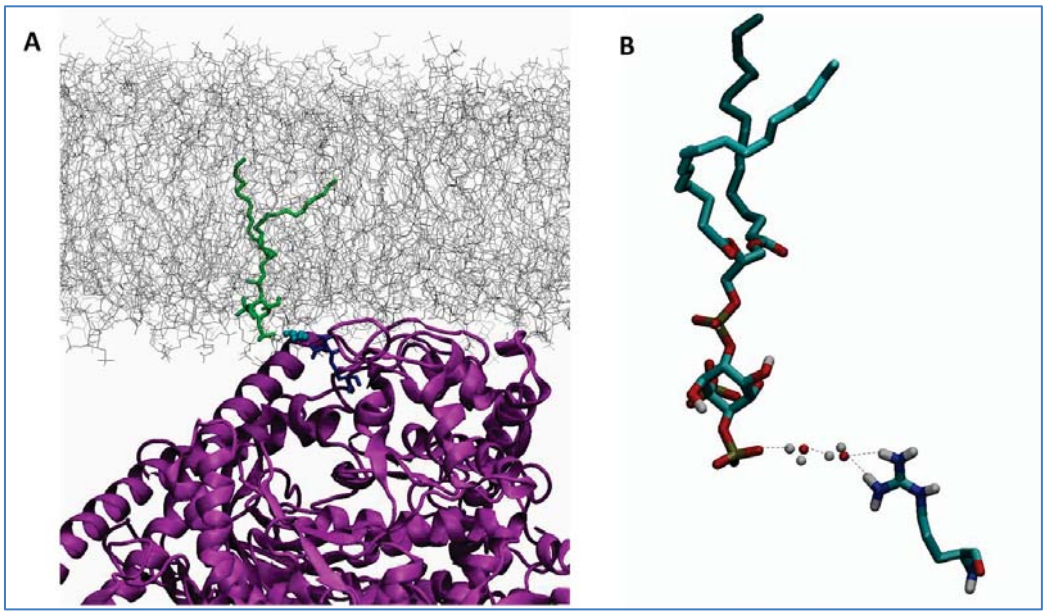


Figure 4xix - ***A)*** *Snapshot of p110α-p85α (purple) in contact with the cell membrane (gray) at approximately 60 ns in the first run of the H1047R system simulation representing the distance between R949 (blue) and the PIP_2 substrate (green). The water molecules (cyan) mediating the contact between R949 and PIP_2 are also shown.* ***B)*** *Hydrogen bonding network exhibited by PIP_2 (left) and R949 (right) with the two water molecules mediating the contact. Molecules have been coloured on the basis of atom type - C (cyan), O (red), H (white), N (blue), and P (gold).*

Using the Gromacs program *g_hbond*, a hydrogen bond network was mapped out in the 60 ns snapshot (Figure 4xix), which comprised hydrogen bonds between the oxygen atom of the phosphate group in the PIP_2 headgroup and the hydrogen atom of the closest water molecule, the oxygen atom of this water molecule and the hydrogen atom of the adjacent water molecule, and finally two hydrogen bonds between the oxygen atom of this water molecule and the hydrogen atoms of the two amino groups of the R949 side chain. The nature of the interaction between R949 and the PIP_2 substrate is a transient one as they do not remain permanently connected via these

hydrogen bonds. Rather, this hydrogen bonding network is dynamic as the water molecules move around the system and the recognition point is relatively fleeting.

Following this, the H1047R p110α-p85α structure at 25 ns timepoints through the 150 ns simulation was aligned with that of the 4OVV crystal structure with PIP_2 bound (Figure 4xx-A). Once aligned, the RMSD between the head group of the prospective PIP_2 substrate identified as approaching the protein in the H1047R simulation and that of the PIP_2 substrate bound in the pocket of the 4OVV crystal structure was calculated (Figure 4xx-B). The RMSD between the PIP_2 substrates was approximately 8 nm at the beginning of the simulation. At the point in the simulation (75 ns) at which the H1047R p110α-p85α presented its full face to the membrane, which places the binding pocket adjacent to the membrane (Figure 4vii, Figure 4xvii), the RMSD between the PIP_2 substrates was approximately 2 nm (Figure 4xx-B). From this point onwards, PIP_2 continues its movement into the binding pocket as the RMSD between the substrates slowly reduces to just less than 1 nm following 150 ns of simulation time.

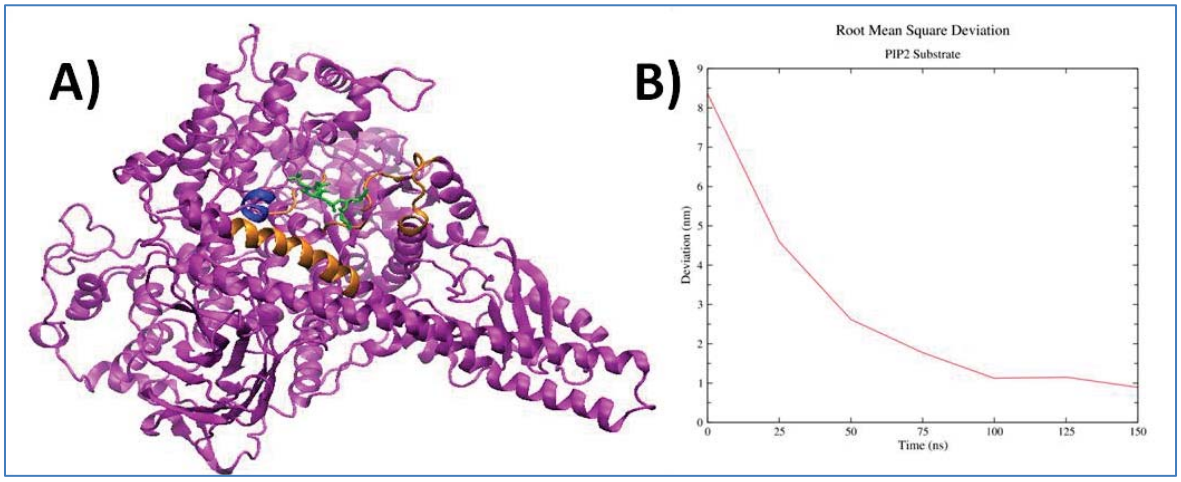


Figure 4xx - ***A)*** *Cartoon representation of p110α-p85α (purple) following 150 ns of the first run of the H1047R system simulation aligned with the 4OVV PDB structure (not shown). The position of the 4OVV PIP_2 substrate (green) in the binding pocket (orange) has also been shown, highlighting the closeness of the PIP_2 headgroup to the LYS-LYS-LYS-LYS miniature helix (blue).* ***B)*** *Atom-positional RMSD between the headgroups of the PIP_2 substrate molecule identified in the first run of the simulated H1047R p110α-p85α system and that in the binding pocket of the 4OVV crystal structure.*

Placing the 4OVV PIP_2 substrate into the more functionally representative H1047R membrane-bound structure shows its proposed position in the binding pocket [35] at a time when the protein is carrying out its kinase activity. The PIP_2 headgroup is seen in close proximity to a loop consisting of four consecutive lysine residues (941 - 944) which may play a role in either further substrate recognition or catalysis (Figure 4xx-A). This suggests that as the protein initiates first contact with

the membrane, and recognises the substrate through R949, it "falls over" onto the membrane into this functionally representative state. The relative position on the membrane of this "falling over" is potentially dictated by a further recognition of the PIP_2 substrate through K941/K942/K943/K944 along the way as it enters the binding pocket. As catalytic activity is unable to be characterised by MD simulation, this final position of the protein parallel to the membrane surface with a PIP_2 substrate entering the binding pocket will be used as a proxy for a representation of a catalytically competent structure, a precursor to ATP binding and subsequent catalytic activity.

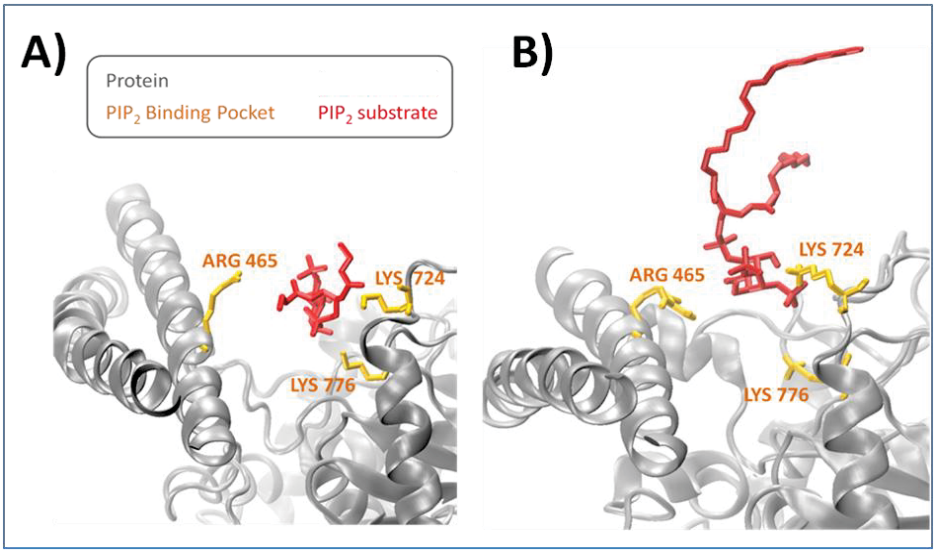


Figure 4xxi - ***A)*** *Snapshot of p110α-p85α (gray) from the 4OVV crystal structure focused on the binding pocket highlighting the residues (orange) interacting with the PIP_2 substrate (red).* ***B)*** *Snapshot of p110α-p85α (gray) following 150 ns of the first run of the H1047R system focused similarly highlighting the same residues (orange) interacting with the identified PIP_2 substrate (red).*

Returning to the simulation results, the final snapshot of H1047R after 150 ns of simulation revealed the final position of the PIP_2 substrate (Figure 4xxi-B). At this RMSD of less than 1 nm from the PIP_2 in the 4OVV structure (Figure 4xx-B), the substrate is seen to be entering the binding pocket. The RMSD is, however, not zero, explained by the difference in orientation of the substrate in the 4OVV structure (Figure 4xxi-A). This might be due to the absence of lipid tails in the 4OVV PIP_2 substrate, and an absence of a lipid bilayer in the 4OVV crystal structure allowing for more space in the pocket. In contrast, the PIP_2 used in the simulations features 16-carbon tails. Despite this, the PIP_2 substrate is seen in close proximity to the same residues in both cases, namely K724 and K776 of the kinase domain (p110α), and R465 of the iSH2 domain (p85α) (Figure 4xxi). The residue which stands out here is K776 which is located in the phosphate binding loop, conserved in the PI3K family [67]. A very recent publication using kinetic analyses to provide mechanistic insight into PI3Kα catalysis has also found K776 to be crucial in not only the recognition of the PIP_2 substrate, but also that of ATP [68].

4.3.3.3 - Electrostatic Analysis

As seen in section 4.3.3.1, a key feature of the PI3Kα-membrane interaction is the interaction between basic residues found in the membrane binding domains and anionic (POPS) lipids found in the lipid bilayer. In fact, the interface presented to the membrane by PI3Kα involves three domains - the C2 domain, the iSH2 domain, and the kinase domain - and consists of 10 to 15 residue long clusters (inclusive of membrane binding domains) which feature upwards of 40% basic residues (arginine or lysine) (Figure 4xxii-B). The distance between this interface and the membrane decreases over the course of the simulation (Figure 4xxii-A) and the number of hydrogen bonds formed between these three domains and the lipid bilayer increases over the same time frame (Figure 4xxi-A). This further serves to confirm that these 10 to 15 residue clusters, in particular the membrane binding domains, approach and interact with the membrane in this simulation.

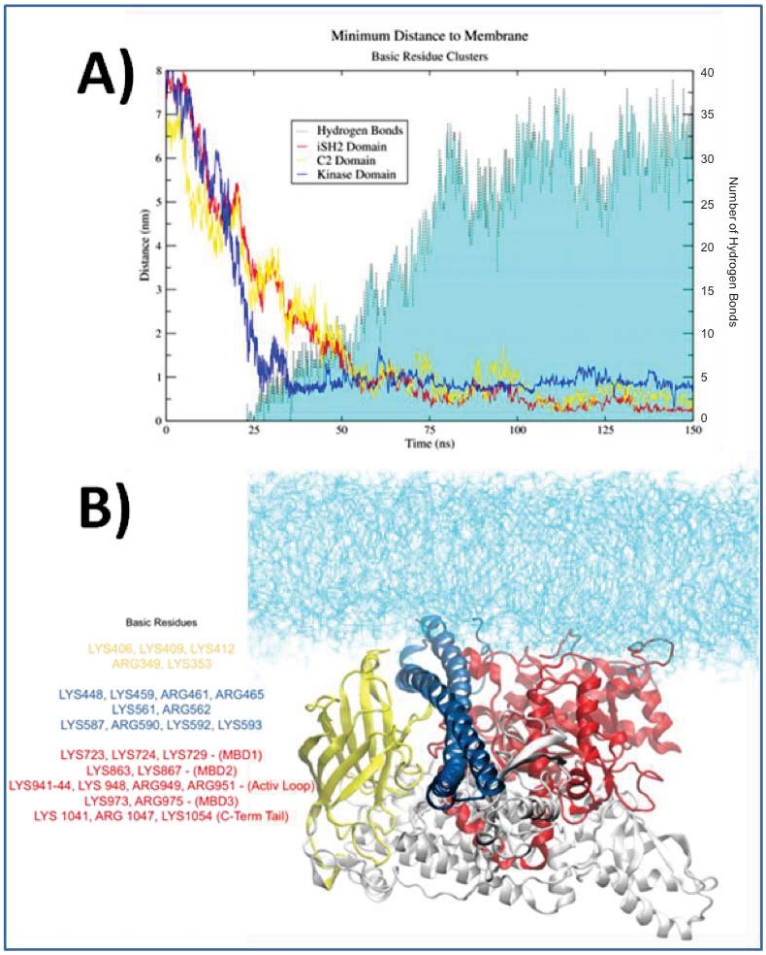


Figure 4xxii - ***A)*** *The minimum distance between the C2 (yellow), kinase (red) and iSH2 (purple) domains and the cell membrane throughout the first run of the H1047R system simulation. The number of hydrogen bonds (cyan) formed between p110α-85α and the cell membrane over the same time frame is also represented.* ***B)*** *Right: Final snapshot of p110α-p85α (white) interacting with the cell membrane (cyan) following 150 ns of the first run of the H1047R system simulation. The C2, kinase, and iSH2 domains are coloured as in (A). Left: A list of the basic residues grouped into clusters within each domain, coloured as in (A), found within 1 nm of the cell membrane.*

The basis of this recognition, approach, and interaction is the favourable electrostatic interactions between the basic residue side chains and the anionic lipid head groups, which was later confirmed in section 4.3.3.4. Using the interaction energy screening method developed in Chapter 3, the most Coulombically favourable orientation of both the wild type and H1047R systems with respect to the brain lipid bilayer was determined (Figure 4xxiii). In each case, the same orientation was observed, but only in the presence of the brain lipid bilayer confirming the importance of the membrane's lipid composition, as this orientation was unseen when the process was repeated in the presence of a neutral DPPC bilayer (Figure 4xxiii-A). The orientation identified here (Figure 4xxiii-B,C) is the same as that in which H1047R binds to the membrane in both simulations (Figure 4vii). However, this is not the case in the two wild type simulations (Figure 4vii); as from an electrostatic standpoint, the H1047R mutant explores an orientation allowing for maximum membrane affinity, yet the wild type does not and is potentially restricted from doing so without the assistance of the *Ras* protein.

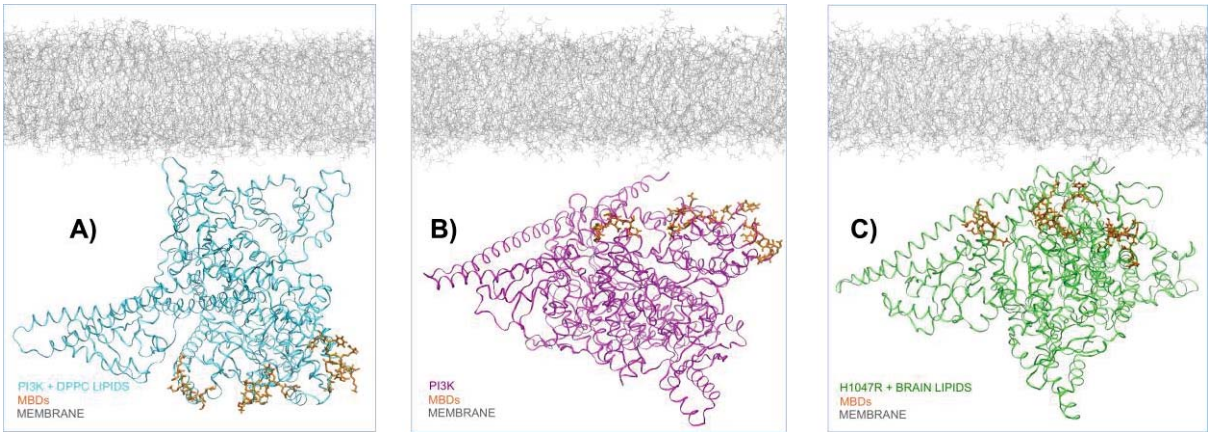


Figure 4xxiii - *Snapshots of the most Coulombically favourable orientation of p110α-p85α with respect to the cell membrane (gray) as identified by rotational interaction energy screening. The MBDs (MBD1, MBD2, MBD3, WIF) are highlighted in orange.* ***A)*** *Wild type (cyan) in the presence of a pure DPPC bilayer.* ***B)*** *Wild type (magenta) in the presence of a brain lipid bilayer.* ***C)*** *H1047R (green) in the presence of a brain lipid bilayer.*

4.3.3.4 - Per-Residue Decomposition of Interaction Energies

The functional orientation of PI3Kα relative to the membrane is driven by Coulombic interactions; however, it was unclear as to which residues in particular made the highest contribution to this attraction. To investigate this, the coordinates from the wild type and H1047R systems were obtained after 100 ns, reflecting a state where the protein was bound to the membrane via all possible MBDs (Figure 4xvi). All residues within the system were assigned their own index group along with the membrane, prior to a 25 step simulation being run from this state as described in the interaction energy screening method (Section 3.3.1). The per-residue interaction energies obtained

in this way would therefore be representative of the individual contribution of each residue in the protein to the total interaction energy between the protein and the membrane at a point in time where they were bound. The difference in the key residues behind the interaction between the wild type and H1047R systems indicate which residues enable the membrane interaction in either protein, and also which residues are fundamental to lipid binding regardless of mutation.

As a control, the interaction energy between the entire protein and the membrane was also calculated, ensuring that the sum of the individual residue contributions was equivalent to the total. It was found that of the 114 wild type residues and 137 H1047R residues within the Coulombic cut-off distance from the membrane, more than 75% of the energy contribution came from a mere few residues in each case (Figure 4xxiv, Table 4A). The wild type p110α-p85α complex also had a lower total protein-membrane interaction energy (-1484 kJ.mol^{-1}) compared to the mutant H1047R complex (-1742 kJ.mol^{-1}) which is in keeping with H1047R's proposed mechanism of increasing activity through improved membrane affinity [8, 34].

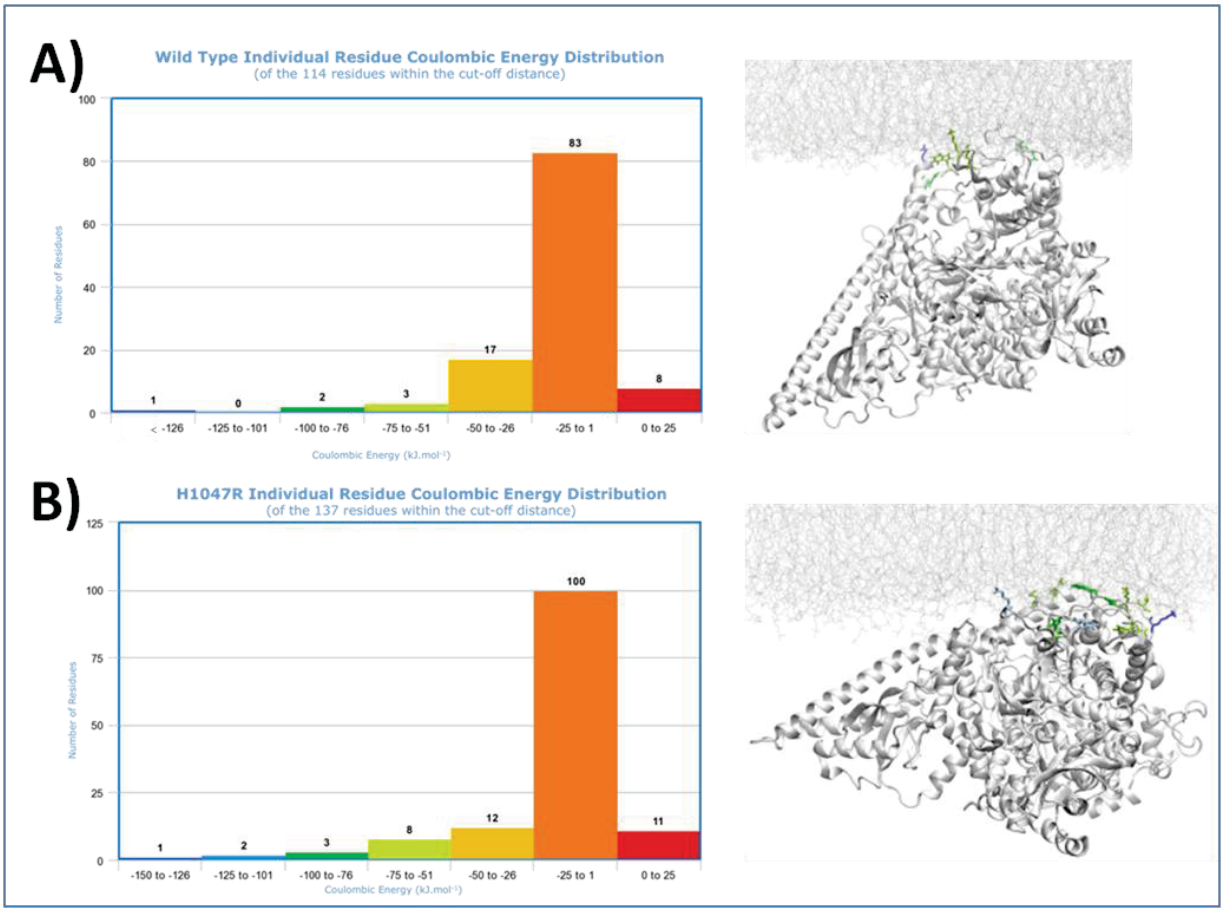


Figure 4xxiv - *Left: Histograms of how the Coulombic interaction energies between p110α-p85α and the cell membrane are distributed amongst individual residues in the wild type (A) and H1047R (B) systems. Right: Final snapshots of p110α-p85α (white) interacting with the cell membrane (gray) following 100 ns of the first runs of the wild type (A) and H1047R (B) system simulations. The corresponding residues from the three most favourable histogram bins are coloured accordingly.*

The wild type interaction energy contribution was primarily made up from residues in the SH2 linkers (Table 4A), floppy loops which in the full length protein would be attached to either the nSH2 or cSH2 inhibitory domains. Returning to the timeseries snapshots (Figure 4vii), these linkers are seen to disrupt the region between the p85α iSH2 domain and p110α kinase domain (PIP_2 binding pocket) by driving them apart, which may have skewed the final orientation of the wild type protein onto the membrane by altering the aforementioned interface incorporating the C2, iSH2, and kinase domains (Section 4.3.3.3). In reality, there would be domains bound to these linkers, and they would not be free to insert themselves into the PIP_2 binding pocket. This issue was later addressed in section 4.4.1. The H1047R interaction energy contribution was primarily made up from residues in the kinase domain (Table 4A), as expected given the parallel nature of its membrane-binding orientation during the simulations.

Table 4A

Top ten interaction energy contributing residues from wild type and H1047R

Wild Type Residues	Coul. Energy ($kJ.mol^{-1}$)	H1047R Mutant Residues	Coul. Energy ($kJ.mol^{-1}$)
LYS 448 (p85α)	-320	LYS 587 (p85α)	-154
LYS 587 (p85α)	-90.5	LYS 867 (p110α)	-124
TYR 453 (p85α)	-79.8	HIS 875 (p110α)	-122
LYS 867 (p110α)	-75.3	ASP 1056 (p110α)	-79.8
GLN 871 (p110α)	-67.9	TRP 1057 (p110α)	-77.2
SER 874 (p110α)	-51.2	HIS 917 (p110α)	-76.6
LYS 973 (p110α)	-47.0	GLU 458 (p85α)	-59.4
ASN 595 (p85α)	-46.4	GLN 455 (p85α)	-58.0
THR 972 (p110α)	-41.9	LYS 593 (p85α)	-55.9
GLY 868 (p110α)	-41.7	THR 585 (p85α)	-53.7

To allow for the unnatural SH2 linker positions in the wild type system, the top five contributing residues from the p110α subunit only were determined and ranked for each system. In the case of wild type, the interaction energy was dominated by the membrane binding domain between residues 860 to 875, as well as those in the Kα7 helix following the activation loop. In the case of H1047R, interaction energy was also dominated by residues in and around the same membrane binding domain, as well as HIS 917 in the catalytic loop, and the Kα12 C terminal tail helix, which includes the WIF motif found to insert itself into the membrane. It seems that the membrane binding domain of residues 863-873 is essential to the membrane affinity of the wild type and H1047R p110α-p85α complexes, especially LYS 867 - a basic residue found to be the highest ranking p110α residue in each case. The more parallel nature of H1047R's interaction with the membrane is also represented in its results which bring the catalytic loop into play, whilst the effect of wild type's more perpendicular interaction and disruption due to the SH2 linkers can also be seen in its results.

4.3.3.5 - Potential of Mean Force

The binding free energy of two bodies, in this case the p110α-p85α complex and the membrane, can be approximated through derivation from the potential of mean force (PMF). The PMF of the two system components is extracted from a series of umbrella sampling simulations at increasing distances from each other. In each window or simulation, the protein was harmonically restrained at a fixed centre of mass distance from the membrane using an umbrella biasing potential. These restraints allow the protein to sample configurational space at each distance. Overlapping windows of each simulation across the series of distances enable the construction of a PMF curve, through which the binding free energy between p110α-p85α and the brain lipid bilayer can be determined.

The starting configurations for the wild type and H1047R systems were obtained from the endpoints of the first runs of each simulation before they underwent solvation, minimisation, and equilibration as outlined in section 4.2.4. The new configurations were generated by position restraining the membrane whilst pulling the protein along the normal to the membrane at a fixed rate of 0.01 nm per ps. The distance between p110α-p85α and the membrane was analysed at each time point of the simulation, before configurations every 0.2 nm away were selected. This resulted in 18 overlapping configurations initially (Figure 4xxv-A,B), which were simulated in NPT conditions for 20 ns with a harmonic restraint enforcing the distance between the protein and membrane centre of mass with a force constant of 1000 $kJ.mol^{-1}.nm^{-2}$.

Orientational sampling was disregarded as it was unfeasible to run a sufficiently long simulation to capture all possible rotations by the protein. This aspect of sampling was also already done by the screening method albeit at only one distance of 0.5 nm, and as such the PMF here instead describes the motion of PI3Kα along a direct path to the membrane (along the z axis), in a fixed orientation defined by the simulation. The PMF was extracted using the Weighted Histogram Analysis Method as implemented in *g_wham* and the binding free energy profile was constructed for each system between p110α-p85α and the membrane, whereby the change in binding free energy (ΔG) represents the difference between the maximum and minimum values on the curve. Where there were discontinuities in the PMF, or where the distributions of the relative position of p110α-p85α to the membrane at neighbouring distances did not overlap, additional umbrella sampling runs positioned at intermediate distances were run - one for wild type, two for H1047R.

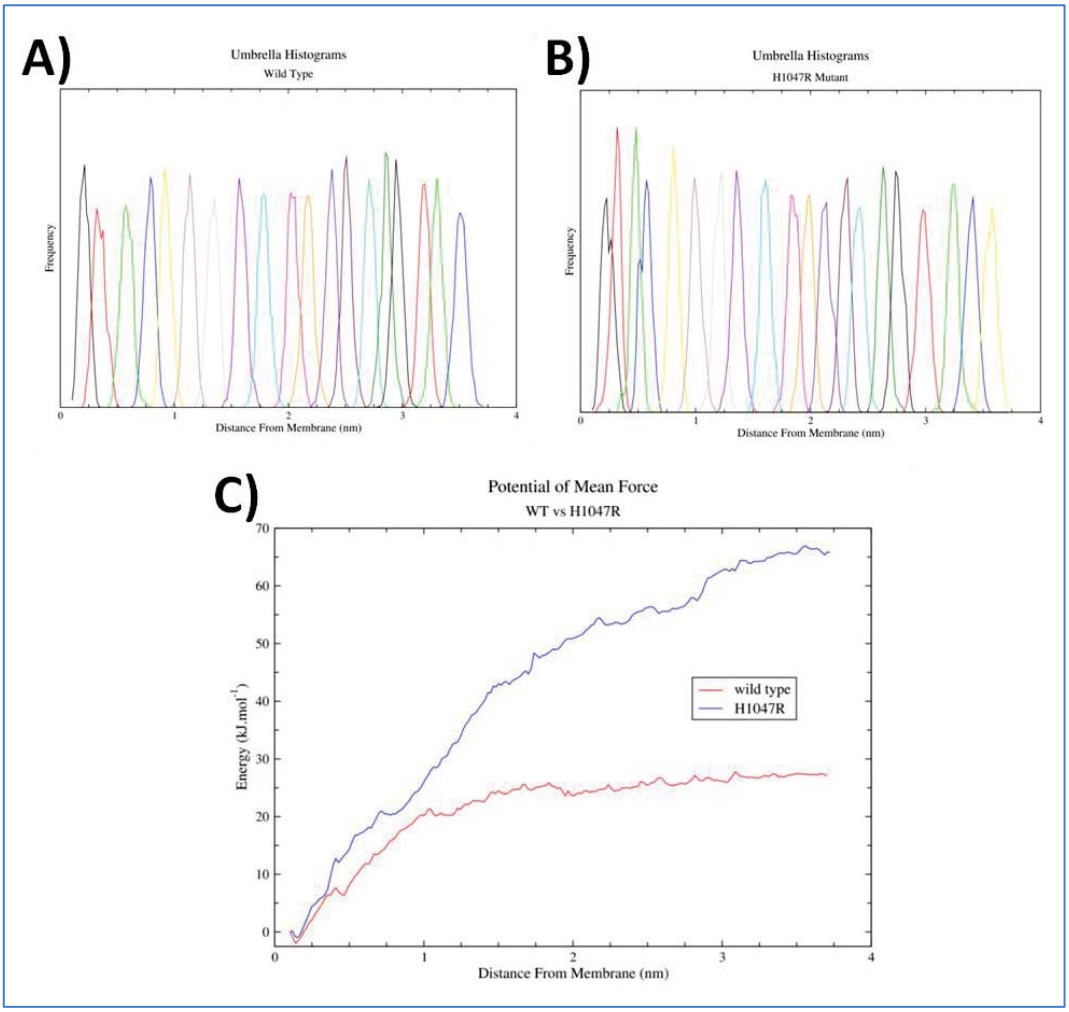


Figure 4xxv - ***A,B)*** *Histogram plots of wild type (A) and H1047R (B) umbrella sampling highlighting the overlap in configurational space sampled by each window at an increasing distance of 0.2 nm away from the cell membrane surface.* ***C)*** *PMF curve calculated on the basis of the umbrella sampling of wild type (red) and H1047R (blue) at increasing distances from the cell membrane surface.*

In the case of the wild type system, the change in binding free energy (maximum to minimum) between p110α-p85α and the brain lipid bilayer was found to be -29.5 kJ.mol^{-1}, while that of the H1047R system was found to be -67.8 kJ.mol^{-1} (Figure 4xxv-C). Once again, it can be seen that the H1047R mutant exhibits a much higher membrane affinity than wild type as discussed in section 4.3.3.4, consistent with lipid binding assays by Burke *et al.* [39]. The curve shows that the PMF between H1047R and the membrane remains high to much greater distances than for wild type suggesting that H1047R is much happier bound to the membrane than the wild type. This is most likely due to the more negative Coulombic interaction energy between H1047R and the membrane in the binding orientation observed in the simulation, and from which the PMF calculations were initiated. These interactions remain strong at distances beyond the 1.4 nm Coulombic cut-off, at which point the curves diverge most (Figure 4xxv-C).

4.3.4 - Intra-Protein Interactions

The program ConAn was used to characterise intra-protein interactions within PI3Kα across the first simulations of wild type and H1047R as they spent the majority of the 150 ns in contact with the brain lipid bilayer, making for a potentially more catalytically relevant representation of their structures. The goal was to determine whether particular contacts within the protein structures, particularly the catalytic domain, explained the difference in their orientations when interacting with the membrane.

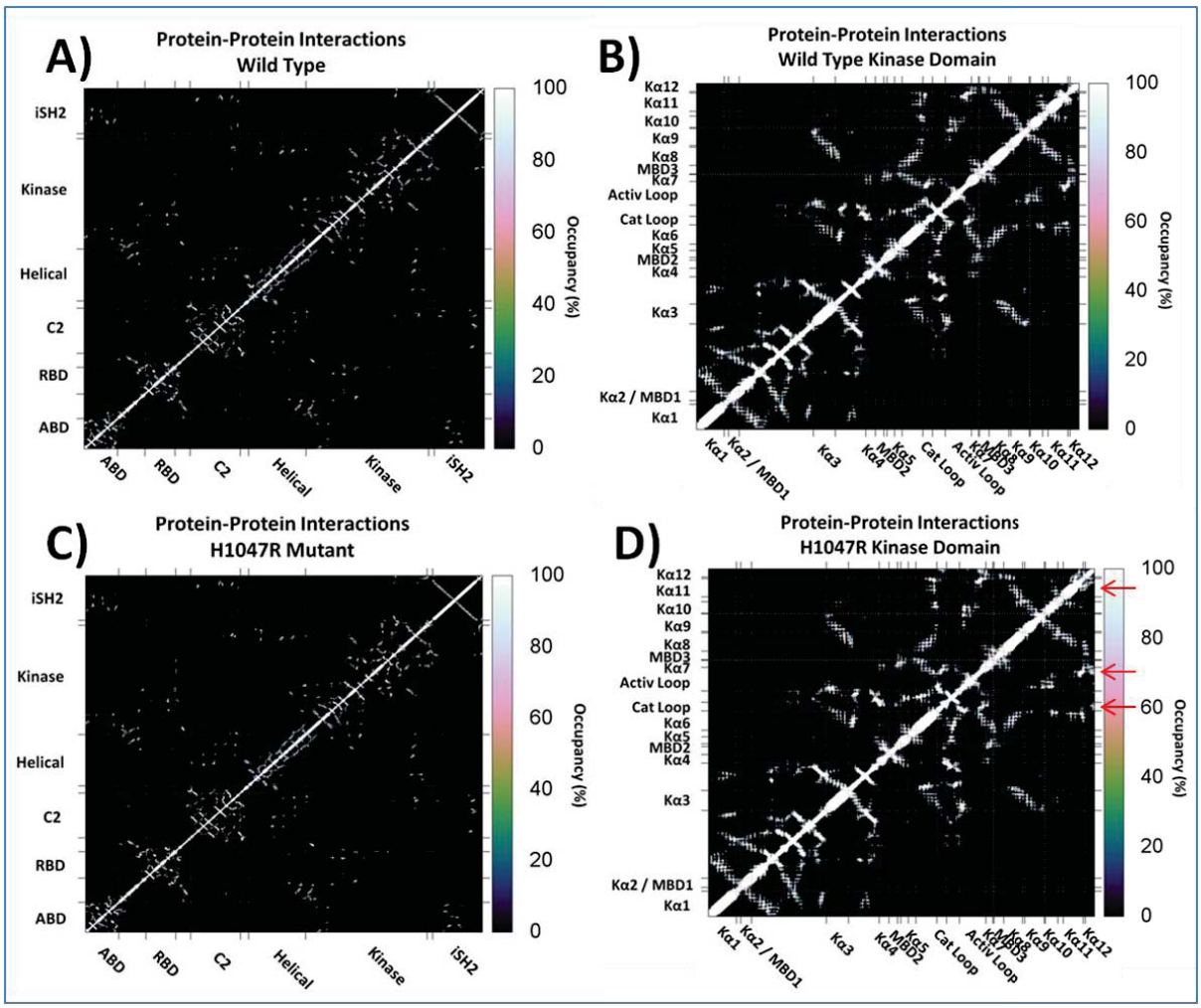


Figure 4xxvi - ***A,C)*** *Map of the intra-protein interactions exhibited by the entire wild type (A) and H1047R (C) p110α-p85α complexes and their occupancies, with gridlines defined by each domain.* ***B,D)*** *The same interactions in (A) and (C) focused within the kinase domains of the wild type (B) and H1047R (D) p110α-p85α complexes, with gridlines defined by important kinase domain segments. Additional contacts made by the H1047R Kα12 tail are indicated by red arrows in (D).*

The wild type and H1047R mutant systems exhibit similar interaction networks within the p110α-p85α complex at first glance (Figure 4xxvi-A,C). However, focusing on the kinase domain (Figure

4xxvi-B,D) reveals that conformational differences between the wild type and H1047R mutant have resulted in the Kα12 C terminal tail of the H1047R mutant making more contacts with the remainder of the kinase domain - including the Kα11 helix, the activation loop, and the catalytic loop. However, in the second run of simulations, both systems exhibited these contacts (Figure Div).

Focusing on the source of these changes, *g_hbond* was used to characterise hydrogen bonds made by residue 1047 (Figure 4xxvii, Figure Dv) which is a histidine in wild type and an arginine in the H1047R mutant. Arginine (R1047), thanks to the charge delocalisation through its double bond and nitrogen lone pairs, is able to form more hydrogen bonds than histidine (H1047). Over the course of the simulation, R1047 forms 39 hydrogen bonds in comparison to H1047's 11. Once multiple donors and acceptors from the same residues are merged, the picture becomes a little clearer.

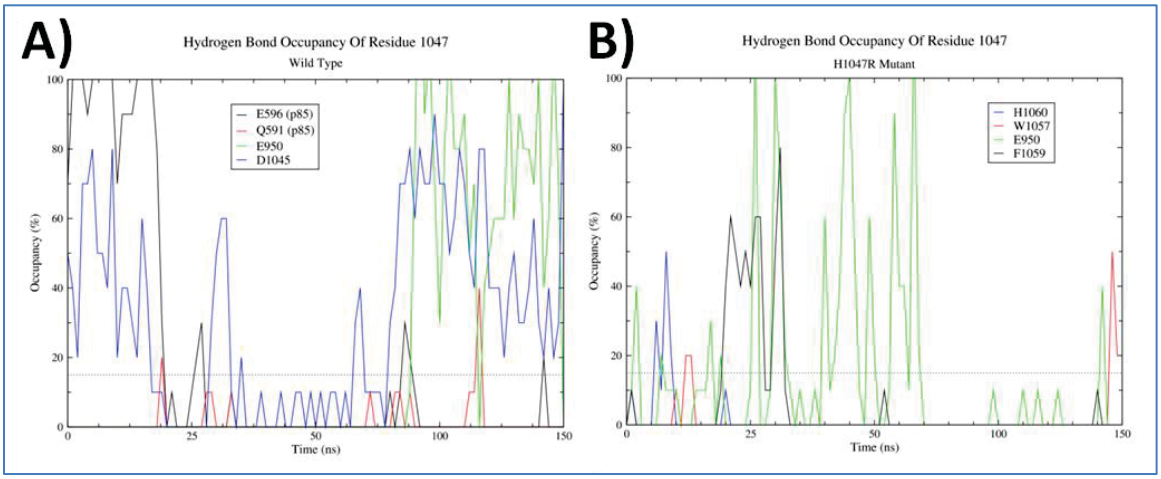


Figure 4xxvii - *Relative occupancy of the hydrogen bonds formed by residue 1047 over the course of the first run of the wild type (A) and H1047R (B) system simulations. Hydrogen bonds with E596 (black), Q591 (red), and D1045 (blue) exist only in the wild type system (A). Hydrogen bonds with W1057 (red), F1059 (black), and H1060 (blue) exist only in the H1047R system. Hydrogen bonds with E950 exist in both cases and are represented by a green line.*

The hydrogen bonding similarities between R1047 and H1047 in the first set of simulations begin and end with glutamic acid at position 950 (E950) found in the activation loop (Figure 4xxvii). This contact is only exhibited by R1047 in the second set of simulations (Figure Dv). Considering R949 has been seen to be fundamental to substrate recognition [11], it's interesting to see its neighbour implicated in hydrogen bonding with a mutation hotspot located in the C terminal tail. For H1047R to increase catalytic activity through improved membrane affinity, it's understandable why the mutation would occur in a membrane anchoring domain, at a specific residue which makes contact with the activation loop responsible for recognition of its lipid substrate. Whilst R1047 forms hydrogen bonds with E950 from the moment membrane interaction occurs (Figure 4xxvii-B), H1047

takes considerably longer to establish this contact (Figure 4xxvii-A). This potentially means the H1047R mutant is able to recognise the substrate sooner than wild type due to improved interactions between the residue at position 1047 and the activation loop. These interactions between the C terminal tail and the activation loop are potentially able to keep the substrate recognition site in position from the moment membrane interaction occurs in the H1047R mutant, thus accelerating its reorientation into a catalytically competent position. The comparatively slower formation of this contact in the wild type could postpone substrate recognition following membrane interaction. However, considering the H1047L mutant also exhibits improved membrane binding despite leucine's differing chemical properties [11], this interaction with the activation loop achieved by R1047 cannot be the primary mechanism behind PI3Kα's increased activity.

The hydrogen bonding differences between H1047 and R1047 are even more interesting. H1047 interacts with residues 591 and 596 of the p85α regulatory subunit in the SH2 linker region, prior to wild type's interaction with the membrane at approximately 30 ns in the first run (Figure 4xxvii-A) and at 100 ns in the second run (Figure Dv). This contact is absent in the H1047R mutant system and could be inhibitory to the wild type's membrane affinity. It might also be the cause of the disrupted nature of the region between the iSH2 and kinase domains previously discussed in section 4.3.3.4.

Finally, R1047 is seen to interact with the Kα12 C terminal tail, particularly the WIF motif. Hydrogen bonding exists between R1047 and phenylalanine at position 1059 (F1059) of the WIF motif, but only prior to interaction with the membrane at approximately 30 ns in the first run (Figure 4xxvii-B). Contacts with Kα12 are even more prevalent in the second run (Figure Dv). These contacts are absent in the wild type systems, and could be helping the C terminal tail maintain a position to insert itself into the membrane more efficiently and thus increasing the membrane binding of H1047R.

4.3.5 - Conclusion

Using the experimental and computational results presented above, a hypothesis regarding the factors governing the approach of PI3Kα to the membrane in an optimal orientation for interaction with the membrane was developed, and how this differs between wild type and the H1047R mutant. Across the various analyses carried out targeted at membrane interaction and the differences between wild type and the H1047R mutant, one thing is clear. H1047R's membrane affinity exceeds that of the wild type p110α-p85α complex, which has been established as the mutant's method of increasing PI3Kα's catalytic activity in this research and in previous studies [34, 39].

The increased activity of H1047R through improved membrane affinity seems to be influenced by a wide range of interactions, both specific short-range and non-specific long-range. Attractive

Coulombic or electrostatic interactions between basic residues in the protein and anionic lipids in the membrane determine the orientation at which PI3Kα approaches the bilayer. These basic residues, which are conserved across the wild type and H1047R enzymes, are clustered in the membrane binding regions of the kinase domain, as well as the C2 and iSH2 domains creating a positively charged surface along this interface. However, this interface is more presented in the H1047R mutant resulting in a flattened orientation with respect to the membrane, compared to the more offset orientation of the wild type. The result is an improved protein-membrane interaction which places the substrate recognition site (R949) at a location where the PIP_2 substrate is available, and places the PIP_2 binding pocket adjacent to the membrane. The final position of the p110α-p85α complex is maintained through the C terminal tail anchoring.

On the basis of the rotational protein-membrane interaction energy screening, this final position should be identical in the wild type and H1047R mutant; however, the simulation results point to this only being possible in the H1047R system. This difference in orientation results in a higher membrane affinity in the H1047R mutant and appears to be driven by intra-protein interactions, which also potentially leads to accelerated substrate recognition and improved C terminal tail anchoring. The H1047R mutant's activity has been found to be independent of the *Ras* protein, responsible for membrane recruitment [32-34], which suggests that wild type PI3Kα may potentially depend on the *Ras* protein, or a longer timescale to achieve this catalytically functional orientation.

The intra-protein interactions involving residue 1047 have marked similarities and differences between the wild type and H1047R. In both cases, residue 1047 makes contact with residue E950 within the activation loop, which is the neighbour of the substrate recognition residue R949. However, H1047 in the wild type tends towards contacts with the SH2 domain linkers which are possibly more inhibitory, whilst R1047 in the H1047R mutant tends towards contacts with the C terminal tail which are more conducive to membrane binding.

In Part II of the results, completely modelled p110α-p85α constructs without the potentially inhibitory SH2 linkers were built, along with a series of C terminal tail mutants. This will allow an improved characterisation of the importance of the contacts made by residue 1047. Essentially, this will suggest whether the interactions between H1047 and the regulatory domain are key to regulating membrane affinity, and if the interactions between R1047 and the WIF motif are necessary to facilitate membrane binding. The effect of the relative position of the C terminal tail will also be explored, seeing whether modelling it in either an open or closed conformation (Section 4.2.2) has any effect on its interactions with both the membrane and the protein itself.

4.4 - Results and Discussion (Part II)

4.4.1 - Importance of C Terminal Tail

As outlined in section 4.1.4, Hon *et al.* proposed the importance of the kinase activation loop and the C terminal tail in the interaction of p110α-p85α with the cellular membrane through mutagenesis studies [11]. To provide further insight into these findings, a series of new p110α-p85α constructs were modelled to replicate the mutants used by Hon *et al*.

It was found that both the polybasic activation loop (electrostatic) and the C terminal WIF motif (hydrophobic) played a part in lipid binding which was also explored in Part I. They therefore created a series of catalytically active (p110α in complex with the iSH2 domain of p85α) PI3Kα mutants for lipid binding experiments. These involved an activation loop mutant (R949D), a Kα12 (C terminal tail) deletion mutant, and a Kα12 triple mutant of the WIF motif (WIF-AAA) reducing its hydrophobicity.

These constructs were replicated allowing the proposed hypothesis (Section 4.3.5) regarding the forces that govern membrane binding to be tested, most importantly those surrounding the behaviour of the C terminal tail.

4.4.2 - Construct Building

A total of ten structures were modelled, five representative of PI3Kα with the Kα12 helix modelled outwards in an open conformation, and five with the Kα12 helix modelled inwards in a closed conformation (Section 4.2.2). In each case, three of the five models were replicates of Hon *et al.*'s mutants - an R949D mutant (R949D), a Kα12 deletion mutant (Kα12del), and a WIF-AAA mutant (WIF-AAA) [11]. All three of the mutants were modelled onto the structures which featured the Kα12 tail modelled as a helix, previously uncharacterised in any PI3Kα crystal structure. The final two models involved the wild type Kα12 tail, which was either modelled as a restrained helix (Kα12helix) or in its originally unrestrained ribbon state (Original) (Figure 4iv, Figure 4v).

However, before commencing any model building, template PI3Kα structures had to be selected. 4A55 was chosen for models set to include Kα12 in an open conformation as the crystal structure was in a proposed "active" state due to the absence of the cSH2 and nSH2 domains of p85α, and the presence of the activation loop structure defined by the electron density. However, as 4A55 is the crystal structure of mouse p110α, any residues which differed from those found in human p110α were amended. 4OVU was chosen for models set to include Kα12 in a closed conformation as the crystal structure was in a proposed "inactive" state due to the presence of the inhibitory nSH2 domain of p85α. A complete list of the modifications made to each structure is described in Figure

4iv. Importantly, all remnants of the cSH2 and nSH2 domains, and their linkers to the iSH2 domain were removed. This would theoretically result in a maximally uninhibited state of the protein as outlined in section 4.1.2.

All ten structures were prepared for simulation and equilibrated for 50 ns as described in section 4.2.1, before being submitted for rotational interaction energy screening in the presence of the brain lipid bilayer.

4.4.3 - Structural Analysis

To determine which mutation had the largest effect on the p110α-p85α structure, the RMSD of all atoms during the equilibration simulation of each construct from the atoms present in the crystal structure template on which it was built was calculated. All constructs equilibrated fairly quickly with any structural changes triggered by the various changes that were made taking place inside the first 10 ns. This is indicated by little upwards or downwards trend being observed in the fluctuations beyond this point (Figure 4xxviii).

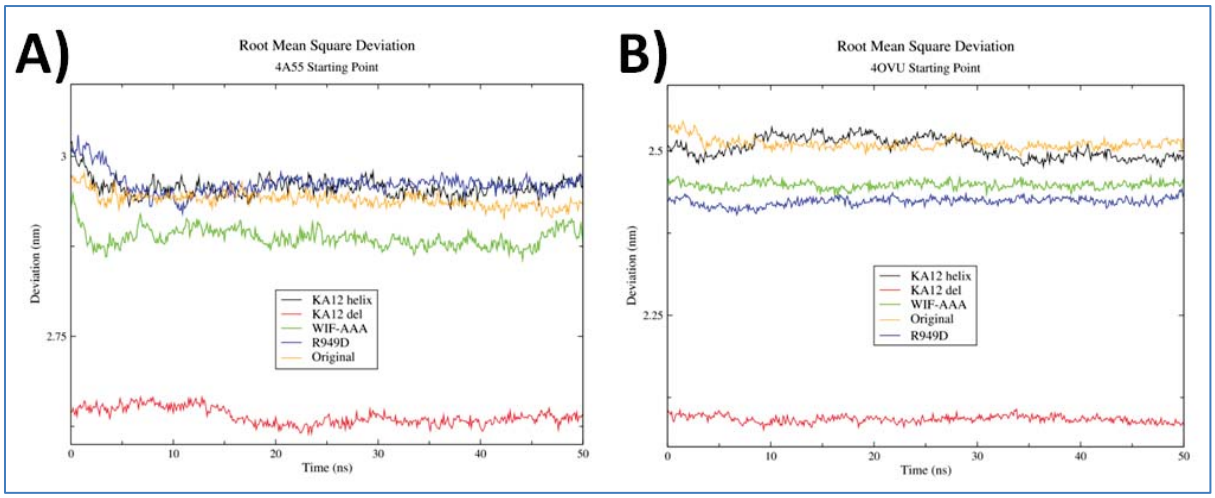


Figure 4xxviii - *Atom-positional RMSD of the new p110α-p85α constructs from the 4A55 (A) and 4OVU (B) crystal structure templates over the course of the 50 ns equilibration runs in solvent. Kα12helix is shown in black, Kα12del in red, WIF-AAA in green, R949D in blue, and Original in orange.*

In the case of both the 4A55 crystal structure and 4OVU crystal structure starting points, Kα12del exhibited fewer conformational changes than the other novel constructs. This is to be expected as the protein structure will not have been overly disrupted due to the removal of terminal residues. On the other hand, the constructs which exhibited the most conformational changes were Kα12helix and Original. To explore the source of these conformational changes, the RMSF of the Kα12helix and Original mutants was calculated over the course of the equilibration.

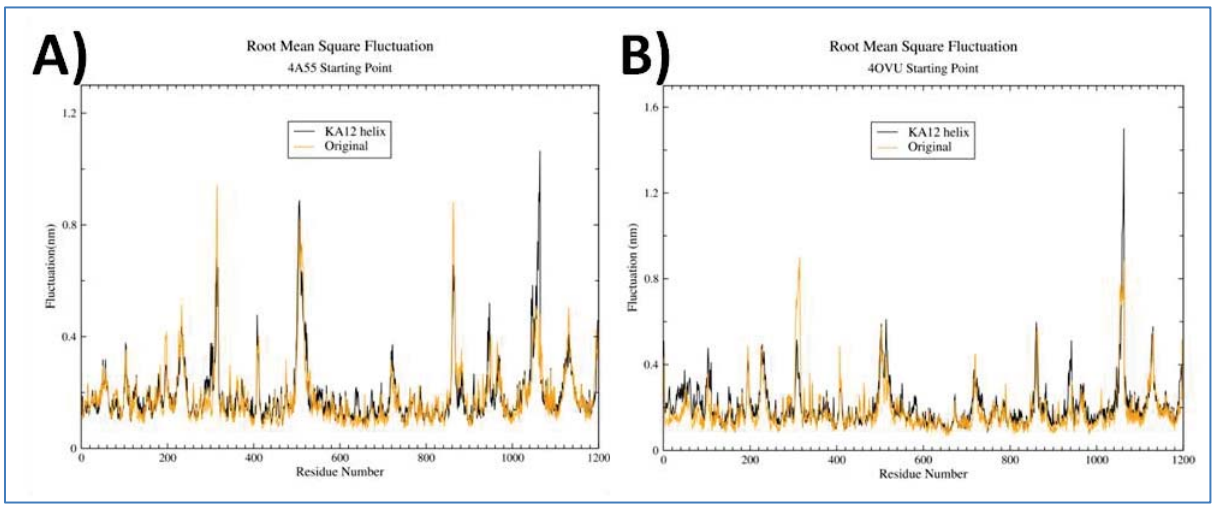


Figure 4xxix - *Atom-positional RMSF of each p110α-p85α residue over the course of the 50 ns equilibration runs in solvent. The Kα12helix (black) and Original (orange) constructs represent those which had the highest deviation from the 4A55 (A) and 4OVU (B) crystal structures (Figure 4xxix).*

As also seen in the RMSD, the 4A55 starting point exhibits a higher degree of conformational change than the 4OVU starting point, with twice as many residues fluctuating highly (Figure 4xxix). This could be explained by the 4A55 structure originally being crystallised with a ligand inside the C terminal lobe active site, as well as a number of residues being changed from the mouse sequence to the human one. However in both cases, the Kα12 tail (residues 1050 - 1064) and the activation loop (residues 940 - 960) underwent more fluctuation in Kα12hel than in Original. The remainder of the residues with the highest RMSF values were similar to those observed in Figure 4xii, with Original exhibiting a higher degree of change in the loop preceding the C2 domain (residues 300 - 310).

Another structural aspect of interest was the secondary structure of the C terminal tail. As it was modelled with restraints as a helix in an inwards and outwards pointing state, the nature of the tail following a simulation with no restraints was of interest. The aim was to determine the stability of the helical structure, and whether or not this was affected by its relative position.

In both cases, the final two or three residues of the Kα12 tail are disordered (Figure 4xxx) which is in keeping with the secondary structure prediction (Figure 4vi). However, when the Kα12 tail was modelled pointing outwards, the α-helix started unfolding further after approximately 20 ns, with the final six or seven residues in the tail becoming more disordered (Figure 4xxx-A). On the other hand, when modelled pointing inwards, the α-helix remained intact in this region throughout the equilibration process (Figure 4xxx-B). The secondary structure prediction also suggested a non-helical nature of the kink between the Kα11 and Kα12 helices (Figure 4vi) which is seen to a much greater degree when the Kα12 tail was modelled pointing outwards (Figure 4xxx-A).

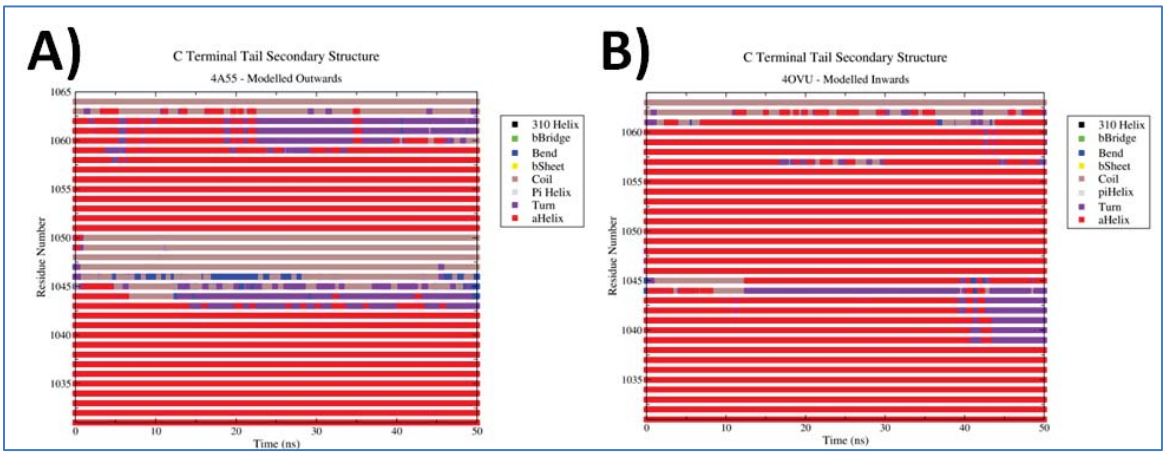


Figure 4xxx - *Secondary structure analysis of the C terminal tail in the p110α subunit comprising Kα11 (residues 1031 to 1046) and Kα12 (residues 1051 to 1062) over the course of the 50 ns equilibration runs when modelled pointing outwards (A) and inwards (B). Colouring is done on the basis of DSSP definitions as listed in Figure 4xv.*

The unordered nature of the end of the Kα12 tail modelled pointing outwards can be potentially explained by its increased exposure to solvent. When the tail is modelled inwards, it is in the vicinity of other residues within the kinase domain discussed in section 4.4.5.2 which may assist in maintaining its helical nature. The environment of the Kα12 tail is the most likely explanation for its unfolding when it is modelled pointing outwards, seeing as the tail has the same sequence in both cases.

4.4.4 - Interaction Energy Screening of Constructs

Rotational interaction energy screening was carried out between each of the constructs and the membrane at a distance of 0.5 nm as described in Chapter 3. The goal was to determine the most favourable orientation of each construct relative to the membrane on the basis of their Coulombic interactions, and whether any of the changes described in section 4.4.2 resulted in a different orientation than the one observed in section 4.3.3.3.

The majority of the newly modelled constructs were found to have a most favourable orientation consistent with the previously identified membrane binding interface presented to the membrane in each case for all of the top five results (Figure 4xxxi). On first thought, this is to be expected, as this interface will have the most positively charged residues regardless of a few mutations. This is especially considering that the majority of these constructs had alterations done to the C terminal tail, whose function in membrane binding is not of an electrostatic nature, but rather a hydrophobic one. However, it is important to note that these constructs will have undergone some level of conformational change following the mutations, and as such may not outwardly present the same residues on their surface. There are two constructs that did not have orientations consistent with

the known membrane binding interface in their top five results, and they both occurred from the 4A55 starting point.

The first result is R949D, whose equilibrated structure presented its membrane binding domains away from the membrane, with the C2 domain being the closest (Figure 4xxxi-D). This is reminiscent of the result where the p110α-p85α complex was screened against a neutral DPPC bilayer (Figure 4xxiii). Considering the function of residue R949 in PIP_2 recognition, it is plausible that were it not for this residue, PI3Kα may indeed function as any other C2 domain protein. However, this result was not reproduced in the 4OVU structure (Figure 4xxxi-H), and may be starting conformation specific.

The second result is WIF-AAA, which produced an orientation identical to that seen in the wild type simulations in Part I (Figure 4xxxi-C). This is a remarkable result which may suggest that interactions between the WIF motif and the kinase domain may also play a role in PI3Kα's membrane approach, assisting in the protein being flat against the membrane. This outcome was not observed with the 4OVU structure (Figure 4xxxi-H). These results led to more questions than answers, discussed below.

Firstly, the helix being modelled pointing inwards or outwards had little overall effect on the most favourable orientations identified in the rotational interaction energy screening results, which focused primarily on Coulombic medium-range interactions. However, due to the protein-membrane distance of 0.5 nm, it remained to be seen whether the C terminal tail position had an effect on the more short-range characteristics such as membrane insertion and substrate interaction. Also, it was unknown whether the introduction of the H1047R mutation to the constructs would initiate any conformational differences in both cases with regards to intra-protein interactions.

Secondly, as WIF-AAA produced a result identical to the wild type simulations in Part I, it would be interesting to explore whether the introduction of an H1047R mutation would influence the orientation of a WIF-AAA p110α-p85α construct with respect to the membrane.

To address these questions, three constructs were chosen - Kα12helix modelled pointing outwards from a 4A55 starting point, Kα12helix modelled pointing inwards from a 4OVU starting point, and WIF-AAA from a 4A55 starting point. The H1047R mutation was introduced in each case, resulting in six new systems: three wild type and three H1047R. These systems each underwent 150 ns molecular dynamics simulations in the presence of the brain lipid bilayer with the previously described protocol (Section 4.2.1) from a starting point taken directly from the screening results. This means the p110α-p85α complex was only 0.5 nm away from the membrane to start with and as such was more biased in its approach following identification of the most likely binding orientation observed in the screening.

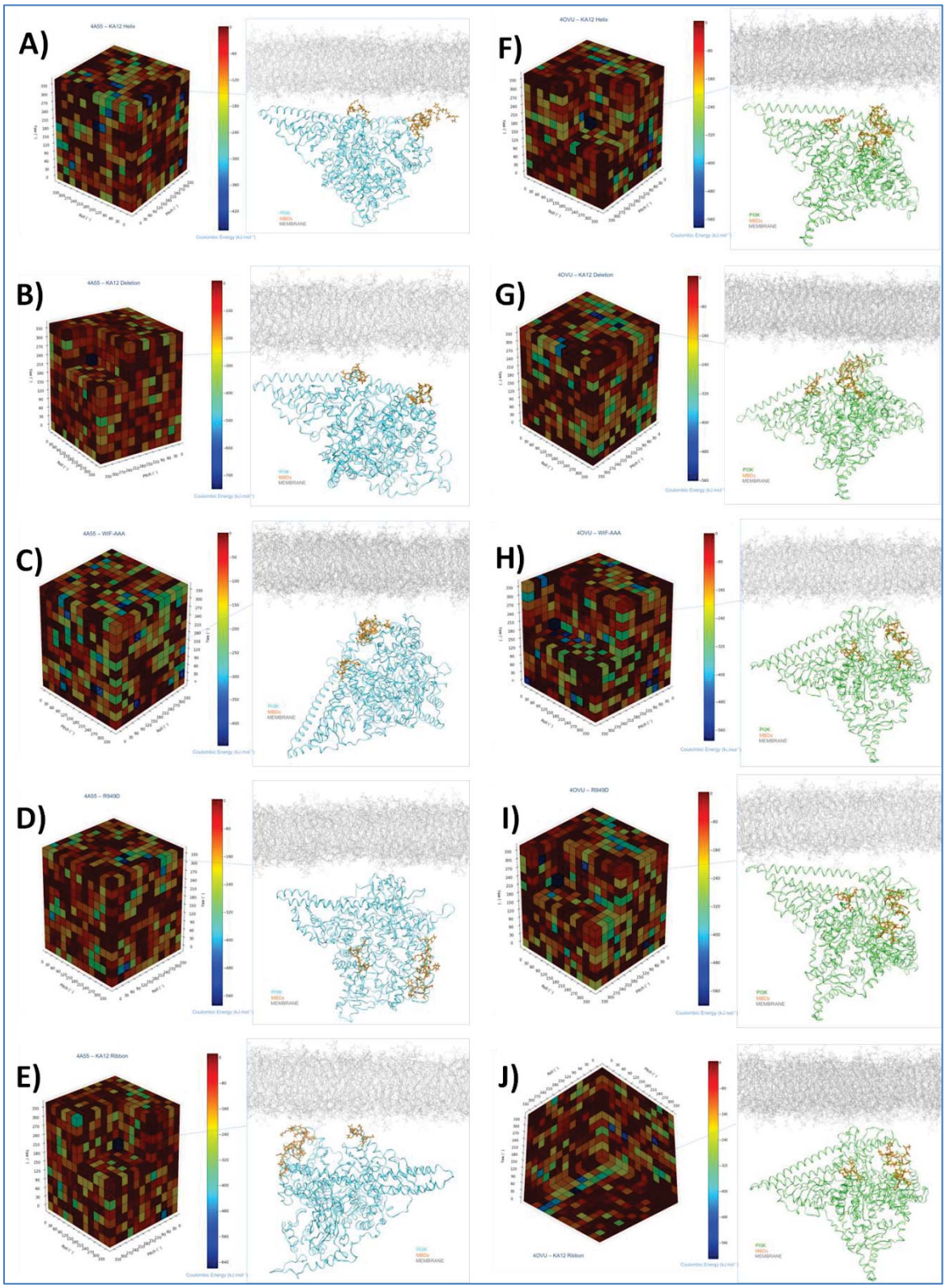


Figure 4xxxi - ***Left**: Heat cubes of the Coulombic energy terms. The colouring goes from warm (red) to cold (blue) on the basis of the favourability of the protein-membrane interaction at that particular orientation, measured in kJ.mol^{-1}. **Right**: Snapshots of the most Coulombically favourable orientations of p110α-p85α constructs built from the 4A55 starting point (A-E, cyan) and 4OVU starting point (F-J, green) with respect to the cell membrane (gray) as identified by rotational interaction energy screening. The MBDs (MBD1, MBD2, MBD3, WIF) are highlighted in orange. From top to bottom: Kα12helix, Kα12del, WIF-AAA, R949D, and Original.*

4.4.5 - Simulations of Chosen Constructs

4.4.5.1 - Effect of C Terminal Tail Modifications

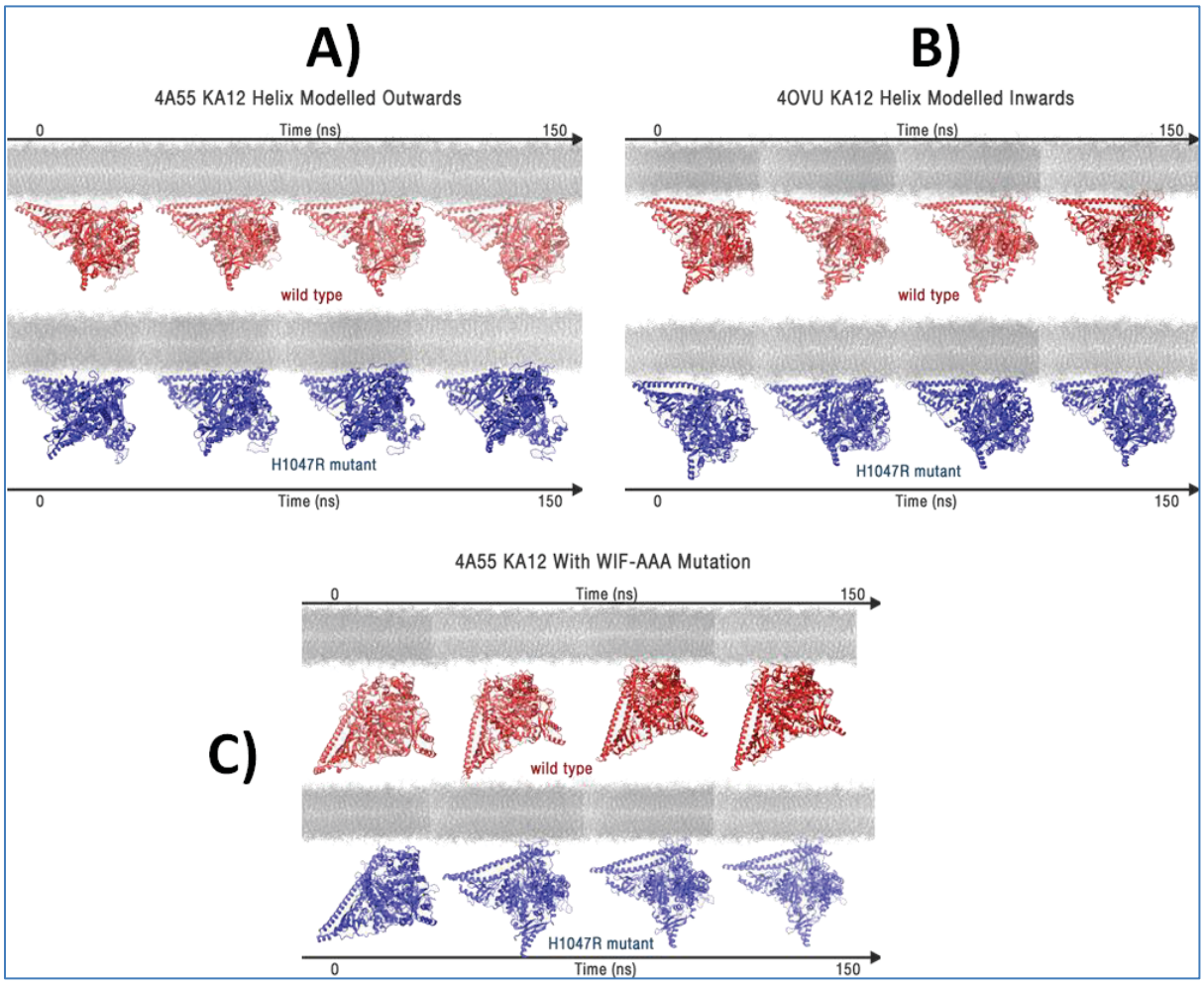


Figure 4xxxii - *Snapshots from the simulations of chosen constructs, from a starting point identified by rotational interaction energy screening, over the course of 150 ns shown at 50 ns intervals, revealing the position of the protein relative to the cell membrane (gray). The wild type is shown in red, the H1047R mutant in blue in all cases. Water molecules have been removed for clarity.* ***A)*** *Kα12helix modelled pointing outwards (4A55 crystal structure template);* ***B)*** *Kα12helix modelled pointing inwards (4OVU crystal structure template);* ***C)*** *WIF-AAA (4A55 crystal structure template).*

For the most part, the p110α-p85α complex approached the membrane in its starting orientation determined by the rotational interaction energy screening (Figure 4xxxii). The previously identified membrane binding interface was maintained, with the PIP_2 binding pocket presented towards the membrane. However, there was a significant change in the orientation of the WIF-AAA mutant proteins with respect to the membrane during the simulations (Figure 4xxxii-C). The starting orientation determined by the screening was more or less identical to the final orientation of the

wild type simulations in part I (Figure 4vii). The wild type p110α-p85α with the WIF-AAA mutation approached the membrane more slowly than its H1047R mutant counterpart, and also at a more upright angle, mimicking the wild type simulations in part I. However, despite the identical biased starting orientation, the H1047R mutant introduced to the same system resulted in the p110α-p85α complex interacting with the membrane sooner and also at an orientation in line with not only the H1047R mutant simulations in part I (Figure 4vii), but also the orientations identified by the screening (Figure 4xxiii, Figure 4xxxi). This suggests that the introduction of the H1047R mutation is what is essential to this more parallel orientation against the membrane, which results in increased membrane binding and a more accessible PIP_2 binding pocket, based on Part I. It can be concluded that H1047R increases membrane binding and activity by altering the preferred orientation of p110α-p85α with respect to the membrane surface, the mechanism behind which was previously touched on in section 4.3.4, and will be explored below.

It has been previously determined that H1047R's activity is independent of activation by *Ras*, whose function is membrane recruitment in wild type PI3K [32-34]. Based on the simulation results in Parts I and II (Figure 4vii, Figure 4xxxii), it would appear that the wild type p110α-p85α complex is unable to flatten itself parallel against the membrane in the simulations, whilst H1047R is able to achieve this orientation on its own. This removes a level of control in the PI3K pathway, causing PI3K to be in a constantly hyperactive state without *Ras* GTPase activation, and as such the signal for cell proliferation remains switched on leading to the development of tumours.

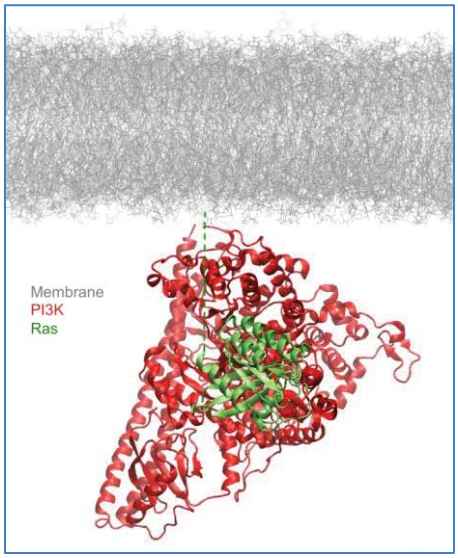


Figure 4xxxiii - *Final snapshot of the WIF-AAA wild type p110α-p85α complex (red) modelled from the 4A55 crystal structure template following 150 ns of simulation in the presence of a cell membrane (gray). The proposed position of the* Ras *protein (green) has been modelled from a structural alignment with the 1HE8 crystal structure [69], with a putative C terminal tail incorporating the final 20 residues represented by a green dashed line.*

In order to investigate the notion that the catalytically viable orientation of wild type p110α-p85α with respect to the membrane interface is assisted by *Ras* recruitment, a structural and sequence alignment was done with the final orientation of p110α-p85α in the WIF-AAA wild type simulation and p110γ in complex with RAS G12V (PDB ID: 1HE8) [69]. The proposed position of *Ras* alongside the RBD highlights its relative orientation to the membrane (Figure 4xxxiii). The 20 residue long C terminal tail of *Ras*, absent from the crystal structure, is proposed to span the gap between the RBD-bound *Ras* and the membrane surface, and a similar putative C terminal tail was drawn by Pacold *et al.* [69]. Considering the relative positions of p110α-p85α, *Ras*, and the membrane, it is quite possible that if the *Ras* recruitment to the membrane was in play, it would assist in flattening p110α-p85α's relative orientation to be more parallel to the membrane surface, as it could effectively pull the lower end of the complex upwards by increasing PI3Kα's overall membrane binding affinity. However, the H1047R mutation is able to overcome this limitation either more quickly or on its own.

4.4.5.2 - Intra-Protein Interactions

The program CoNaN was used to map the intra-protein interactions of each of the six proteins over the course of the 150 ns simulation in the presence of a brain lipid bilayer. The goal was to determine whether the changes introduced to the p110α-p85α complex had any effect on the residue contacts between the Kα12 tail and the remainder of the kinase domain observed in section 4.3.4. As both the changes and these contacts were located in the kinase domain of the protein, the intra-protein interactions evaluated were restricted to this region of the protein.

There are two things to consider when evaluating these results (Figure 4xxxiv). Firstly, each starting p110α-p85α structure had already undergone conformational changes, triggered by the C terminal tail mutations or alterations introduced in the modelling phase, in the 50 ns equilibration period. Furthermore, the 150 ns simulation also allowed for equilibration following the introduction of the H1047R mutation. Secondly, all interactions shown here equate to those exhibited by the p110α-p85α structure in a membrane interacting form, while those in Part I also incorporated various interactions made in solution throughout its membrane approach.

The only system where a clear difference is seen between the wild type and H1047R mutant is the Kα12 helix modelled pointing inwards (Figure 4xxxiv-B,E), where the wild type C terminal tail makes more interactions with the remainder of the kinase domain than the H1047R mutant. This is due to a shift in the position of the C terminal tail in the H1047R mutant, resulting in its interactions being more in line with that of the structures where the Kα12 helix was modelled pointing outwards. This suggests that this shift in position may be representative of an active p110-p85α structure and is

further explored in section 4.4.5.3. The remaining systems (Kα12helix modelled pointing outwards and WIF-AAA) exhibited similar interactions across the wild type and H1047R (Figure 4xxxiv).

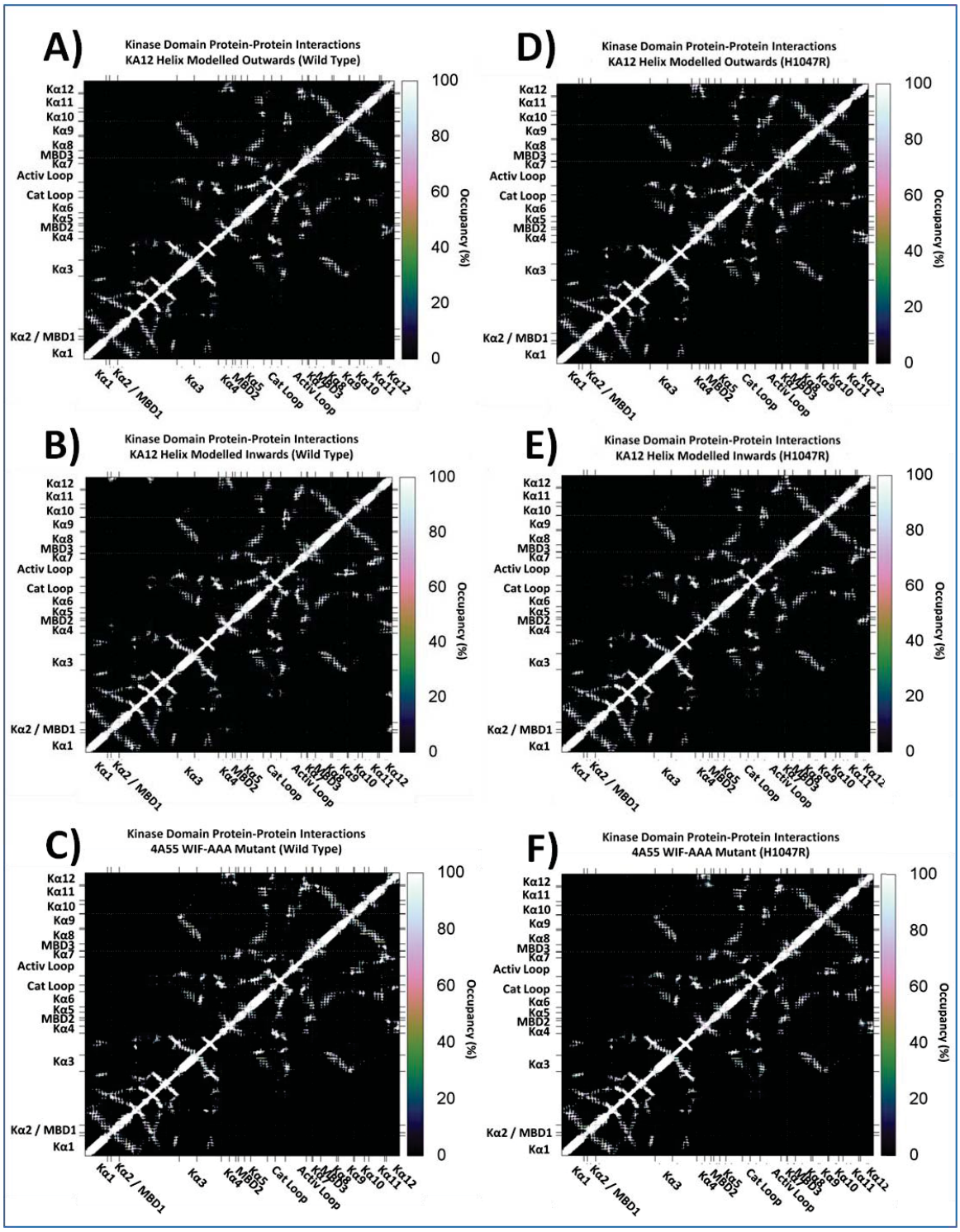


Figure 4xxxiv - *Map of the intra-protein interactions exhibited by the kinase domains of the chosen wild type (A-C) and H1047R (D-F) p110α-p85α constructs and their occupancies, with gridlines defined by important kinase domain segments. From top to bottom: Kα12helix modelled pointing outwards (4A55 crystal structure template), Kα12helix modelled pointing inwards (4OVU crystal structure template), WIF-AAA (4A55 crystal structure template).*

Once again, the hydrogen bonds made by residue 1047 were brought into focus to shed some light onto any differences in behaviour between the wild type and H1047R mutant constructs.

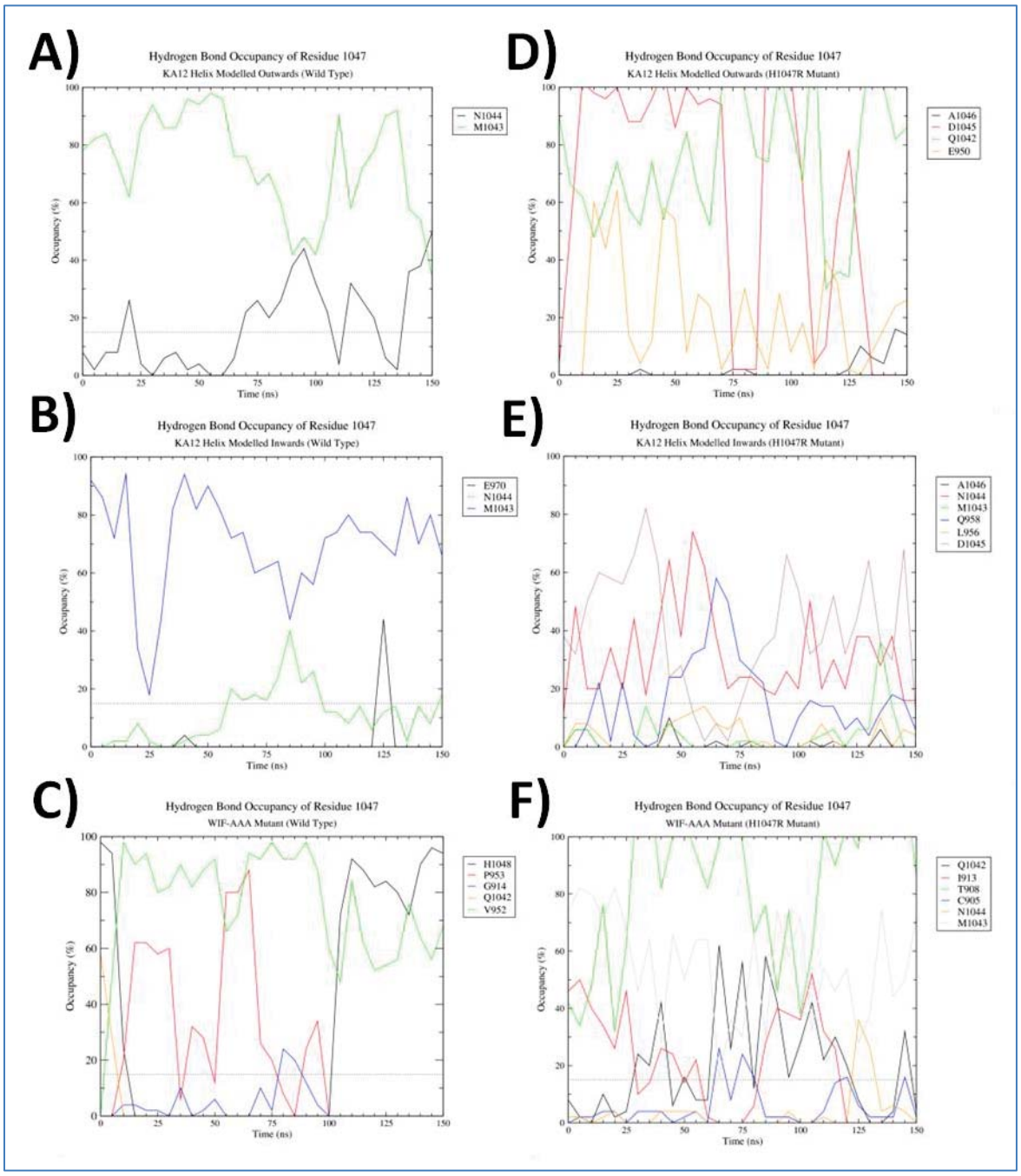


Figure 4xxxv - *Relative occupancy of the hydrogen bonds formed by residue 1047 over the course of the chosen wild type (A-C) and H1047R (D-F) construct simulations in the presence of a cell membrane. From top to bottom: Kα12helix modelled pointing outwards (4A55 crystal structure template), Kα12helix modelled pointing inwards (4OVU crystal structure template), WIF-AAA (4A55 crystal structure template).*

It can firstly be seen that there are many more interactions with nearby residues (within five residues of residue 1047) (Figure 4xxxv) in comparison with the systems explored in Part I (Figure 4xxvii). This is due to the C terminal tail having now been modelled as an α-helix and as such, hydrogen bonds exist between the backbones of residues along this helix. Once again, the arginine at position 1047 (H1047R mutant) forms considerably more hydrogen bonds than the wild type histidine (Figure 4xxxv). In the case of the Kα12 helices modelled pointing inwards and outwards, R1047 forms hydrogen bonds with the activation loop as seen in the Part I system (Figure 4xxxv-D,E), whilst H1047 does not regardless of the position of the C terminal tail, unlike the Part I system (Figure 4xxxv-A,B). This would suggest that despite the chemical similarity in the histidine and arginine side chains (amino groups), the sheer length of the arginine side chain becomes a key factor in the formation of hydrogen bonds with residues that histidine is unable to contact. The lengthier nature of the arginine side chain could also be the reason behind the shift in C terminal tail position when the Kα12 helix was initially modelled inwards (Section 4.4.5.3), resulting in a potentially activated p110α-p85α complex.

On the other hand, the WIF-AAA systems are unique cases due to the differing protein orientations relative to the membrane (Figure 4xxxii), whereby a conformational change in the kinase region has resulted in hydrogen bond formation unseen in the other systems (Figure 4xxxv). Interestingly, whilst the H1047 continues to exhibit hydrogen bonds with the activation loop as seen in the Part I system (Figure 4xxxv-C), the R1047 instead predominantly forms hydrogen bonds with residues either within or directly preceding the catalytic loop (residues 912-920) (Figure 4xxxv-F). This alteration of the kinase domain structure may have been a driving force behind the WIF-AAA/H1047R mutant's final orientation on the membrane despite its starting orientation, which the wild type was unable to achieve (Figure 4xxxii). As the interaction with the catalytic loop was unseen in Part I (Figure 4xxvii), it is possible that the H1047R mutation is also able to overcome the conformational changes instigated by the WIF-AAA mutation, through the creation of new contacts involving R1047.

4.4.5.3 - Membrane Interaction and Substrate Analysis

While the orientation of the Kα12 helix does not affect p110α-p85α's interaction with the membrane during the simulations, changes in the helix orientation might be expected to have more of an effect on the insertion of the C terminal tail itself into the membrane. As such, the relative position of the C terminal tail to the membrane was determined for all four cases and whether this was indicative of any insertion into or interaction with the membrane.

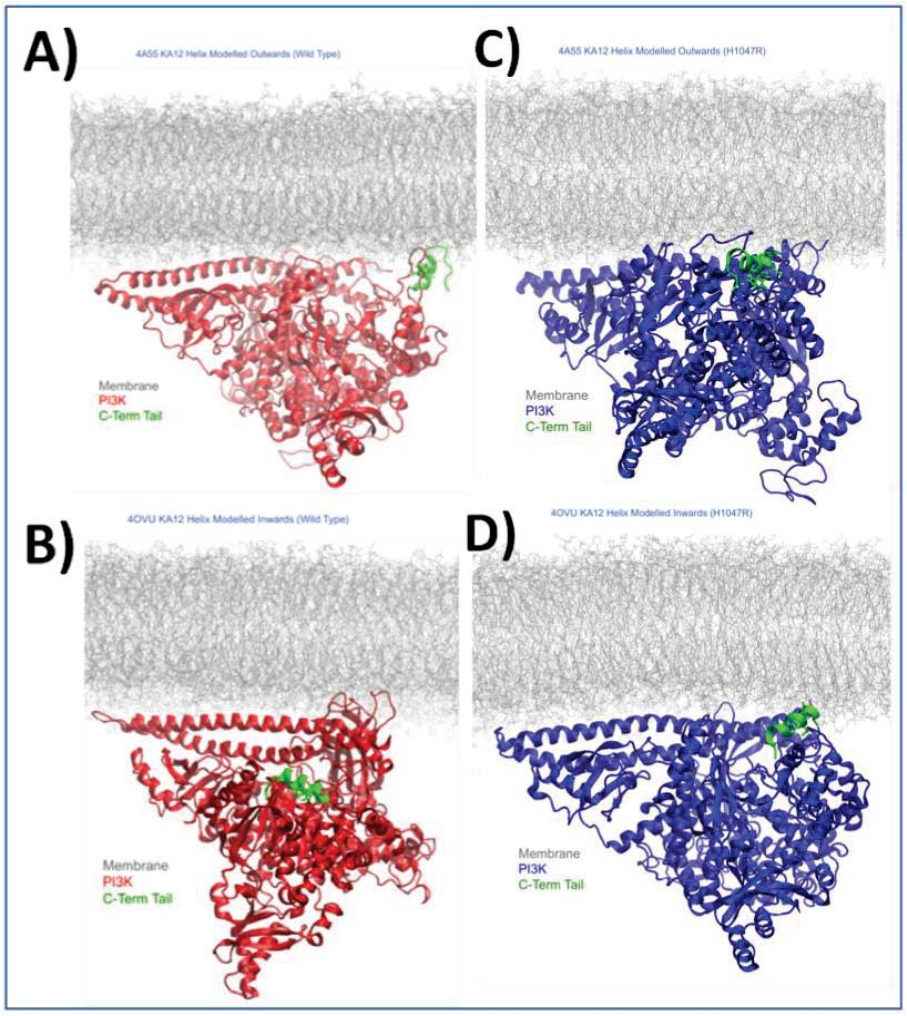


Figure 4xxxvi - *Final snapshots of the chosen wild type (red) and H1047R (blue) constructs following 150 ns of simulation in the presence of a cell membrane (gray), highlighting the relative position of the C terminal tail (green).* ***A,C)*** *Kα12helix modelled pointing outwards (4A55 crystal structure template).* ***B,D)*** *Kα12helix modelled pointing inwards (4OVU crystal structure template).*

When the Kα12 helix was modelled pointing outwards, both the wild type and H1047R mutant had a final orientation relative to the membrane allowing for the C terminal tail to come into contact with the lipids (Figure 4xxxvi-A,C), as observed in Part I (Figure 4xvii). However, when the Kα12 helix was modelled pointing inwards, only the H1047R mutant ended in an orientation which allowed C terminal tail and membrane interaction (Figure 4xxxvi-D). Furthermore, the H1047R mutant presented the PIP_2 binding pocket and the full C2-iSH2-kinase domain interface to the membrane whilst the wild type p110α-p85α construct kinase domain was in a more offset position. This would suggest that Kα12 helix in an inwards position prevents the wild type from binding to the membrane effectively, but this barrier is overcome by the H1047R mutant possibly giving it an additional mechanism for increasing membrane affinity and catalytic activity. This issue was non-existent when the Kα12 helix was modelled in an outwards position, which essentially clears the PIP_2 binding

pocket of any interference. The starting C terminal tail position was identical when modelled pointing inwards for the wild type and H1047R systems, yet throughout the simulation, gravitated towards different positions (Figure 4xxxvii). While the Kα12 helix in the wild type system remained in the same lateral position relative to the binding pocket, the H1047R system exhibited a more open conformation, with it shifting away from the binding pocket. This could have been the key factor in its binding orientation and subsequent substrate interaction, explored below.

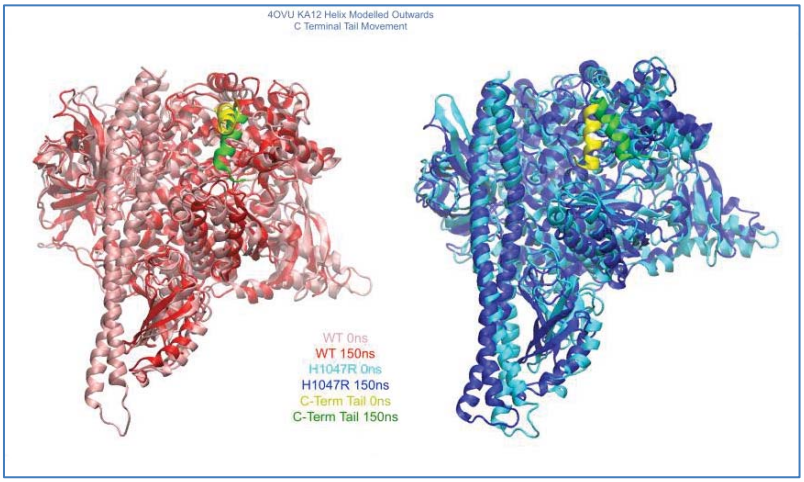


Figure 4xxxvii - ***Left****: Snapshots of the chosen wild type construct with the Kα12helix modelled pointing inwards after 0 ns (pink) and 150 ns (red) of simulation in the presence of a cell membrane.* ***Right****: Snapshots of the chosen H1047R construct with the Kα12helix modelled pointing inwards after 0 ns (cyan) and 150 ns (blue) of simulation in the presence of a cell membrane. The relative positions of the C terminal tail before (yellow) and after (green) 150 ns of simulation are highlighted.*

As expected from the above results, when the Kα12 helix was modelled pointing outwards, both the wild type and H1047R mutant positioned themselves so as to interact with a PIP_2 lipid substrate (Figure 4xxxviii-A,B). Following 150 ns of simulation, the binding pocket in each case was adjacent to a PIP_2 lipid that sits slightly lower than the membrane surface. This would suggest that once the Kα12 helix is in an open configuration, the p110α-p85α complex is catalytically active; however, disregarding the bias in initial orientation in these simulations, the wild type may either require more time or some assistance from *Ras* in getting to that position as previously discussed in section 4.4.5.1. The H1047R seemingly needs no such help, and is also unperturbed by the Kα12 helix being modelled pointing inwards. A PIP_2 substrate is also seen adjacent to the binding pocket in this system (Figure 4xxxviii-C), something that is not the case in the wild type system, potentially due to the inhibitory final position of its C terminal tail (Figure 4xxxvii). This is further supported by the catalytically competent position of the wild type simulations in Part I, where the Kα12 tail was neither helical nor pointing in any particular direction.

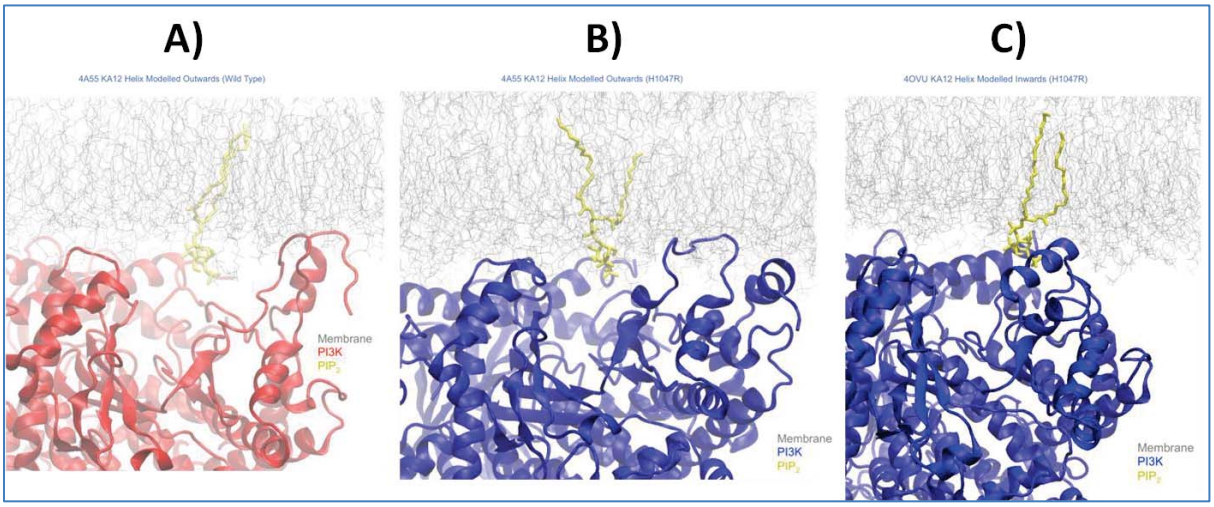


Figure xxxviii - *Final snapshots of the chosen wild type (red) and H1047R (blue) constructs following 150 ns of simulation in the presence of a cell membrane (gray). A potential PIP_2 substrate (yellow) is seen entering the binding pocket in each case.* ***A,B)*** *Kα12helix modelled pointing outwards (4A55 crystal structure template),* ***C)*** *Kα12helix modelled pointing inwards (4OVU crystal structure template).*

4.4.6 - Conclusion

The aim of these simulations was to explore the structure and dynamics of PI3Kα in its most active form (in the absence of the nSH2 and cSH2 domains of p85α) in the presence of a cell membrane including its PIP_2 substrate. In Part I, the effect of mutation hotspot H1047R on PI3Kα's membrane affinity was explored. It was concluded that the fundamental difference between the p110α-p85α wild type and H1047R mutant is the ability of the H1047R mutant to get into a catalytically competent position through increased membrane affinity. It does so without the proposed assistance of *Ras* unlike the wild type which instead was found to bind to the membrane at an offset angle which did not present the PIP_2 binding pocket to the membrane. Also important in this transition from an offset angle to a more parallel orientation is the necessity of placing residue R949 in a position to carry out its role of PIP_2 recognition.

In Part II, the aim was further refined to focus on the role of the C terminal tail, and how changes introduced to the Kα12 helix altered the behaviour of PI3Kα. Whilst in Part I the role of the C terminal tail was restricted to that of a membrane anchor, it was also seen in Part II to be linked to the accessibility of the PIP_2 binding pocket. When positioned away from the pocket, it allowed p110α-p85α to be catalytically competent, with both wild type and H1047R observed to be interacting with a PIP_2 substrate. However, when it was modelled pointing inwards, wild type is unable to achieve a catalytically competent position at the membrane surface, while the H1047R mutation initiated a conformational change that moved the Kα12 helix away from the binding

pocket, allowing catalytically competent interaction with the membrane surface and concomitant PIP_2 interaction.

In summary, PI3Kα's membrane interaction is complex and multi-faceted involving a basic residue interface governing its general orientation through Coulombic attraction to anionic lipids, followed by substrate recognition by R949 and C terminal tail membrane interaction ensuring a catalytically competent position. PI3Kα's catalytic activity has many levels of regulation both intrinsically, in the form of the C terminal tail and inhibitory p85α subunit, and extrinsically, in the form of recruitment and activating proteins such as *Ras*. Oncogenic mutants which increase PI3Kα activity do so by overcoming any one or more of these levels of regulation, with H1047R achieving a catalytically competent state without the assistance of *Ras*. The E545K mutant whose membrane interaction was unable to be determined by this study has been proposed to be independent of the regulatory p85α subunit which suggests it disengages interactions between the inhibitory nSH2 domains and the catalytic p110α subunit. Overall, these simulations have provided mechanistic insight into the role of the C terminal tail and the effect of the H1047R mutation on PI3Kα behaviour.

4.5 - Bibliography

1. Williams, R., A. Berndt, S. Miller, W.C. Hon, and X. Zhang, *Form and flexibility in phosphoinositide 3-kinases.* Biochem Soc Trans, 2009. **37**(Pt 4): p. 615-26.
2. Foster, F.M., C.J. Traer, S.M. Abraham, and M.J. Fry, *The phosphoinositide (PI) 3-kinase family.* J Cell Sci, 2003. **116**(Pt 15): p. 3037-40.
3. Cully, M., H. You, A.J. Levine, and T.W. Mak, *Beyond PTEN mutations: the PI3K pathway as an integrator of multiple inputs during tumorigenesis.* Nat Rev Cancer, 2006. **6**(3): p. 184-92.
4. Cleary, J.M. and G.I. Shapiro, *Development of phosphoinositide-3 kinase pathway inhibitors for advanced cancer.* Curr Oncol Rep, 2010. **12**(2): p. 87-94.
5. Arcaro, A. and A.S. Guerreiro, *The phosphoinositide 3-kinase pathway in human cancer: genetic alterations and therapeutic implications.* Curr Genomics, 2007. **8**(5): p. 271-306.
6. Chomczyk, M.A. and P. Czajka, *[PIK3CA mutations in the most common types of cancer].* Postepy Biochem, 2013. **59**(3): p. 280-4.
7. Vadas, O., J.E. Burke, X. Zhang, A. Berndt, and R.L. Williams, *Structural basis for activation and inhibition of class I phosphoinositide 3-kinases.* Sci Signal, 2011. **4**(195): p. re2.
8. Burke, J.E. and R.L. Williams, *Synergy in activating class I PI3Ks.* Trends in Biochemical Sciences. **40**(2): p. 88-100.
9. Amzel, L.M., C.-H. Huang, D. Mandelker, C. Lengauer, S.B. Gabelli, and B. Vogelstein, *Structural comparisons of class I phosphoinositide 3-kinases.* Nat Rev Cancer, 2008. **8**(9): p. 665-669.
10. Gabelli, S., C.H. Huang, D. Mandelker, O. Schmidt-Kittler, B. Vogelstein, and L.M. Amzel, *Structural Effects of Oncogenic PI3Kα Mutations.* Current topics in microbiology and immunology, 2010. **347**: p. 43-53.
11. Hon, W.-C., A. Berndt, and R.L. Williams, *Regulation of lipid binding underlies the activation mechanism of class IA PI3-kinases.* Oncogene, 2012. **31**(32): p. 3655-3666.
12. Backer, J.M., *The regulation of class IA PI 3-kinases by inter-subunit interactions.* Curr Top Microbiol Immunol, 2010. **346**: p. 87-114.
13. Huang, C.-H., D. Mandelker, O. Schmidt-Kittler, Y. Samuels, V.E. Velculescu, K.W. Kinzler, B. Vogelstein, S.B. Gabelli, and L.M. Amzel, *The Structure of a Human p110α/p85α Complex Elucidates the Effects of Oncogenic PI3Kα Mutations.* Science, 2007. **318**(5857): p. 1744-1748.
14. Zhao, L. and P.K. Vogt, *Class I PI3K in oncogenic cellular transformation.* Oncogene, 0000. **27**(41): p. 5486-5496.
15. Stephens, L.R., K.T. Hughes, and R.F. Irvine, *Pathway of phosphatidylinositol(3,4,5)-trisphosphate synthesis in activated neutrophils.* Nature, 1991. **351**(6321): p. 33-39.
16. Vanhaesebroeck, B., J. Guillermet-Guibert, M. Graupera, and B. Bilanges, *The emerging mechanisms of isoform-specific PI3K signalling.* Nature Reviews Molecular Cell Biology, 2010. **11**(5): p. 329-341.
17. Jiménez, C., C. Hernández, B. Pimentel, and A.C. Carrera, *The p85 Regulatory Subunit Controls Sequential Activation of Phosphoinositide 3-Kinase by Tyr Kinases and Ras.* Journal of Biological Chemistry, 2002. **277**(44): p. 41556-41562.
18. Siempelkamp, B.D., M.K. Rathinaswamy, M.L. Jenkins, and J.E. Burke, *Molecular mechanism of activation of class IA phosphoinositide 3-kinases (PI3Ks) by membrane-localized HRas.* Journal of Biological Chemistry, 2017.
19. Zhang, D. and L. Aravind, *Identification of novel families and classification of the C2 domain superfamily elucidate the origin and evolution of membrane targeting activities in eukaryotes.* Gene, 2010. **469**(1–2): p. 18-30.
20. Steck, P.A., M.A. Pershouse, S.A. Jasser, W.K.A. Yung, H. Lin, A.H. Ligon, L.A. Langford, M.L. Baumgard, T. Hattier, T. Davis, C. Frye, R. Hu, B. Swedlund, D.H.R. Teng, and S.V. Tavtigian, *Identification of a candidate tumour suppressor gene, MMAC1, at chromosome 10q23.3 that is mutated in multiple advanced cancers.* Nat Genet, 1997. **15**(4): p. 356-362.

21. Walser, R., J.E. Burke, E. Gogvadze, T. Bohnacker, X. Zhang, D. Hess, P. Küenzi, M. Leitges, E. Hirsch, R.L. Williams, M. Laffargue, and M.P. Wymann, *PKCβ Phosphorylates PI3Kγ to Activate It and Release It from GPCR Control.* PLOS Biology, 2013. **11**(6): p. e1001587.
22. Cantley, L.C., *The phosphoinositide 3-kinase pathway.* Science, 2002. **296**(5573): p. 1655-7.
23. Vanhaesebroeck, B. and M.D. Waterfield, *Signaling by distinct classes of phosphoinositide 3-kinases.* Exp Cell Res, 1999. **253**(1): p. 239-54.
24. Vivanco, I. and C.L. Sawyers, *The phosphatidylinositol 3-Kinase AKT pathway in human cancer.* Nat Rev Cancer, 2002. **2**(7): p. 489-501.
25. Miled, N., Y. Yan, W.-C. Hon, O. Perisic, M. Zvelebil, Y. Inbar, D. Schneidman-Duhovny, H.J. Wolfson, J.M. Backer, and R.L. Williams, *Mechanism of Two Classes of Cancer Mutations in the Phosphoinositide 3-Kinase Catalytic Subunit.* Science, 2007. **317**(5835): p. 239-242.
26. Carpenter, C.L., K.R. Auger, M. Chanudhuri, M. Yoakim, B. Schaffhausen, S. Shoelson, and L.C. Cantley, *Phosphoinositide 3-kinase is activated by phosphopeptides that bind to the SH2 domains of the 85-kDa subunit.* Journal of Biological Chemistry, 1993. **268**(13): p. 9478-83.
27. Yu, J., C. Wjasow, and J.M. Backer, *Regulation of the p85/p110α Phosphatidylinositol 3'-Kinase: DISTINCT ROLES FOR THE N-TERMINAL AND C-TERMINAL SH2 DOMAINS.* Journal of Biological Chemistry, 1998. **273**(46): p. 30199-30203.
28. Forbes, S.A., D. Beare, P. Gunasekaran, K. Leung, N. Bindal, H. Boutselakis, M. Ding, S. Bamford, C. Cole, S. Ward, C.Y. Kok, M. Jia, T. De, J.W. Teague, M.R. Stratton, U. McDermott, and P.J. Campbell, *COSMIC: exploring the world's knowledge of somatic mutations in human cancer.* Nucleic Acids Research, 2015. **43**(D1): p. D805-D811.
29. Kang, S., A.G. Bader, and P.K. Vogt, *Phosphatidylinositol 3-kinase mutations identified in human cancer are oncogenic.* Proc Natl Acad Sci U S A, 2005. **102**(3): p. 802-7.
30. Meyer, D.S., S. Koren, C. Leroy, H. Brinkhaus, U. Muller, I. Klebba, M. Muller, R.D. Cardiff, and M. Bentires-Alj, *Expression of PIK3CA mutant E545K in the mammary gland induces heterogeneous tumors but is less potent than mutant H1047R.* Oncogenesis, 2013. **2**: p. e74.
31. Bader, A.G., S. Kang, and P.K. Vogt, *Cancer-specific mutations in PIK3CA are oncogenic in vivo.* Proc Natl Acad Sci U S A, 2006. **103**(5): p. 1475-9.
32. Hao, Y., C. Wang, B. Cao, B.M. Hirsch, J. Song, S.D. Markowitz, R.M. Ewing, D. Sedwick, L. Liu, W. Zheng, and Z. Wang, *Gain of interaction with IRS1 by p110alpha-helical domain mutants is crucial for their oncogenic functions.* Cancer Cell, 2013. **23**(5): p. 583-93.
33. Mandelker, D., S.B. Gabelli, O. Schmidt-Kittler, J. Zhu, I. Cheong, C.-H. Huang, K.W. Kinzler, B. Vogelstein, and L.M. Amzel, *A frequent kinase domain mutation that changes the interaction between PI3Kα and the membrane.* Proceedings of the National Academy of Sciences, 2009. **106**(40): p. 16996-17001.
34. Zhao, L. and P.K. Vogt, *Helical domain and kinase domain mutations in p110α of phosphatidylinositol 3-kinase induce gain of function by different mechanisms.* Proceedings of the National Academy of Sciences, 2008. **105**(7): p. 2652-2657.
35. Miller, M.S., O. Schmidt-Kittler, D.M. Bolduc, E.T. Brower, D. Chaves-Moreira, M. Allaire, K.W. Kinzler, I.G. Jennings, P.E. Thompson, P.A. Cole, L.M. Amzel, B. Vogelstein, and S.B. Gabelli, *Structural basis of nSH2 regulation and lipid binding in PI3Kα.* Oncotarget, 2014. **5**(14): p. 5198-5208.
36. Furet, P., V. Guagnano, R.A. Fairhurst, P. Imbach-Weese, I. Bruce, M. Knapp, C. Fritsch, F. Blasco, J. Blanz, R. Aichholz, J. Hamon, D. Fabbro, and G. Caravatti, *Discovery of NVP-BYL719 a potent and selective phosphatidylinositol-3 kinase alpha inhibitor selected for clinical evaluation.* Bioorganic & Medicinal Chemistry Letters, 2013. **23**(13): p. 3741-3748.
37. Zhao, Y., X. Zhang, Y. Chen, S. Lu, Y. Peng, X. Wang, C. Guo, A. Zhou, J. Zhang, Y. Luo, Q. Shen, J. Ding, L. Meng, and J. Zhang, *Crystal Structures of PI3Kα Complexed with PI103 and Its Derivatives: New Directions for Inhibitors Design.* ACS Medicinal Chemistry Letters, 2014. **5**(2): p. 138-142.

38. Burke, J.E. and R.L. Williams, *Dynamic steps in receptor tyrosine kinase mediated activation of class IA phosphoinositide 3-kinases (PI3K) captured by H/D exchange (HDX-MS).* Adv Biol Regul, 2013. **53**(1): p. 97-110.
39. Burke, J.E., O. Perisic, G.R. Masson, O. Vadas, and R.L. Williams, *Oncogenic mutations mimic and enhance dynamic events in the natural activation of phosphoinositide 3-kinase p110alpha (PIK3CA).* Proc Natl Acad Sci U S A, 2012. **109**(38): p. 15259-64.
40. Durrant, J. and J.A. McCammon, *Molecular dynamics simulations and drug discovery.* BMC Biology, 2011. **9**(1): p. 71.
41. Adcock, S. and J. McCammon, *Molecular dynamics: survey of methods for simulating the activity of proteins.* Chem Rev, 2006. **106**: p. 1589 - 1615.
42. Gkeka, P., T. Evangelidis, M. Pavlaki, V. Lazani, S. Christoforidis, B. Agianian, and Z. Cournia, *Investigating the structure and dynamics of the PIK3CA wild-type and H1047R oncogenic mutant.* PLoS Comput Biol, 2014. **10**(10): p. e1003895.
43. Gkeka, P., A. Papafotika, S. Christoforidis, and Z. Cournia, *Exploring a Non-ATP Pocket for Potential Allosteric Modulation of PI3Kα.* The Journal of Physical Chemistry B, 2015. **119**(3): p. 1002-1016.
44. Gao, Y., Y. Ma, G. Yang, and Y. Li, *Molecular Dynamics Simulations to Investigate the Binding Mode of the Natural Product Liphagal with Phosphoinositide 3-Kinase α.* Molecules, 2016. **21**(7): p. 857.
45. Han, M. and J.Z.H. Zhang, *Class I Phospho-inositide-3-kinases (PI3Ks) Isoform-Specific Inhibition Study by the Combination of Docking and Molecular Dynamics Simulation.* Journal of Chemical Information and Modeling, 2010. **50**(1): p. 136-145.
46. Echeverria, I., Y. Liu, S.B. Gabelli, and L.M. Amzel, *Oncogenic mutations weaken the interactions that stabilize the p110α-p85α heterodimer in phosphatidylinositol 3-kinase α.* FEBS Journal, 2015. **282**(18): p. 3528-3542.
47. Berendsen, H.J.C., J.R. Grigera, and T.P. Straatsma, *The missing term in effective pair potentials.* The Journal of Physical Chemistry, 1987. **91**(24): p. 6269-6271.
48. Berendsen, H.J.C., J.P.M. Postma, W.F. van Gunsteren, A. DiNola, and J.R. Haak, *Molecular dynamics with coupling to an external bath.* The Journal of Chemical Physics, 1984. **81**(8): p. 3684-3690.
49. Irvine, W.A., *Testing of parameters for a biologically accurate brain membrane and molecular dynamics simulations exploration of membrane interactions and conformational changes exhibited by p110a and its oncogenic mutants*, in *Institute of Natural and Mathematical Sciences*. 2014, Massey University, Albany, New Zealand: Massey University.
50. Hess, B., H. Bekker, H.J.C. Berendsen, and J.G.E.M. Fraaije, *LINCS: A linear constraint solver for molecular simulations.* Journal of Computational Chemistry, 1997. **18**(12): p. 1463-1472.
51. Essmann, U., L. Perera, M.L. Berkowitz, T. Darden, H. Lee, and L.G. Pedersen, *A smooth particle mesh Ewald method.* The Journal of Chemical Physics, 1995. **103**(19): p. 8577-8593.
52. Miller, S., B. Tavshanjian, A. Oleksy, O. Perisic, B.T. Houseman, K.M. Shokat, and R.L. Williams, *Shaping Development of Autophagy Inhibitors with the Structure of the Lipid Kinase Vps34.* Science, 2010. **327**(5973): p. 1638-1642.
53. Walker, E.H., M.E. Pacold, O. Perisic, L. Stephens, P.T. Hawkins, M.P. Wymann, and R.L. Williams, *Structural determinants of phosphoinositide 3-kinase inhibition by wortmannin, LY294002, quercetin, myricetin, and staurosporine.* Mol Cell, 2000. **6**(4): p. 909-19.
54. Drozdetskiy, A., C. Cole, J. Procter, and G.J. Barton, *JPred4: a protein secondary structure prediction server.* Nucleic Acids Research, 2015.
55. Yachdav, G., E. Kloppmann, L. Kajan, M. Hecht, T. Goldberg, T. Hamp, P. Hönigschmid, A. Schafferhans, M. Roos, M. Bernhofer, L. Richter, H. Ashkenazy, M. Punta, A. Schlessinger, Y. Bromberg, R. Schneider, G. Vriend, C. Sander, N. Ben-Tal, Rost, and Burkhard, *PredictProtein—an open resource for online prediction of protein structural and functional features.* Nucleic Acids Research, 2014.

56. Eswar, N., B. Webb, M.A. Marti-Renom, M.S. Madhusudhan, D. Eramian, M.-y. Shen, U. Pieper, and A. Sali, *Comparative Protein Structure Modeling Using MODELLER*, in *Current Protocols in Protein Science*. 2001, John Wiley & Sons, Inc.
57. Chen, V.B., W.B. Arendall, J.J. Headd, D.A. Keedy, R.M. Immormino, G.J. Kapral, L.W. Murray, J.S. Richardson, and D.C. Richardson, *MolProbity: all-atom structure validation for macromolecular crystallography.* Acta Crystallographica Section D: Biological Crystallography, 2010. **66**(Pt 1): p. 12-21.
58. Pronk, S., S. Pall, R. Schulz, P. Larsson, P. Bjelkmar, R. Apostolov, M.R. Shirts, J.C. Smith, P.M. Kasson, D. van der Spoel, B. Hess, and E. Lindahl, *GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit.* Bioinformatics, 2013. **29**(7): p. 845-54.
59. Humphrey, W., A. Dalke, and K. Schulten, *VMD: visual molecular dynamics.* J Mol Graph, 1996. **14**(1): p. 33-8, 27-8.
60. Daura, X., K. Gademann, B. Jaun, D. Seebach, W.F. van Gunsteren, and A.E. Mark, *Peptide Folding: When Simulation Meets Experiment.* Angewandte Chemie International Edition, 1999. **38**(1-2): p. 236-240.
61. Kabsch, W. and C. Sander, *Dictionary of protein secondary structure: Pattern recognition of hydrogen-bonded and geometrical features.* Biopolymers, 1983. **22**(12): p. 2577-2637.
62. Nose, S., *A unified formulation of the constant temperature molecular dynamics methods.* The Journal of Chemical Physics, 1984. **81**(1): p. 511-519.
63. Hoover, W.G., *Canonical dynamics: Equilibrium phase-space distributions.* Physical Review A, 1985. **31**(3): p. 1695-1697.
64. Parrinello, M. and A. Rahman, *Strain fluctuations and elastic constants.* The Journal of Chemical Physics, 1982. **76**(5): p. 2662-2666.
65. Gkeka, P., T. Evangelidis, M. Pavlaki, V. Lazani, S. Christoforidis, B. Agianian, and Z. Cournia, *Investigating the Structure and Dynamics of the PIK3CA Wild-Type and H1047R Oncogenic Mutant.* PLOS Computational Biology, 2014. **10**(10): p. e1003895.
66. Pirola, L., M.J. Zvelebil, G. Bulgarelli-Leva, E. Van Obberghen, M.D. Waterfield, and M.P. Wymann, *Activation Loop Sequences Confer Substrate Specificity to Phosphoinositide 3-Kinase α (PI3Kα): FUNCTIONS OF LIPID KINASE-DEFICIENT PI3Kα IN SIGNALING.* Journal of Biological Chemistry, 2001. **276**(24): p. 21544-21554.
67. Walker, E.H., O. Perisic, C. Ried, L. Stephens, and R.L. Williams, *Structural insights into phosphoinositide 3-kinase catalysis and signalling.* Nature, 1999. **402**(6759): p. 313-320.
68. Maheshwari, S., M.S. Miller, R. O'Meally, R.N. Cole, L.M. Amzel, and S.B. Gabelli, *Kinetic and structural analyses reveal residues in phosphoinositide 3-kinase α that are critical for catalysis and substrate recognition.* Journal of Biological Chemistry, 2017.
69. Pacold, M.E., S. Suire, O. Perisic, S. Lara-Gonzalez, C.T. Davis, E.H. Walker, P.T. Hawkins, L. Stephens, J.F. Eccleston, and R.L. Williams, *Crystal Structure and Functional Analysis of Ras Binding to Its Effector Phosphoinositide 3-Kinase Gamma.* Cell, 2000. **103**(6): p. 931-944.

Chapter 5 - Cytochrome P450

5.1 - Literature Review

5.1.1 - Overview

Cytochromes P450 (CYPs) are haemoproteins, belonging to a superfamily of proteins which contain a haem group as a cofactor. To date, roughly 60 active genes have been discovered to code for CYPs in the human genome [1], but they are found to exist in all kingdoms of life including bacteria and viruses [2]. CYP enzymes are oxidases which catalyse a variety of metabolic reactions, and play a role in the synthesis of molecules such as steroid hormones, bile acids, cholesterols and other fatty acids. They also play a major role in the metabolism of toxins and foreign substances introduced to the body, such as pharmaceutical drugs [3, 4].

The name cytochrome P450 comes from its location in microsomal vesicles (cyto), and its coloured nature (chrome), pigmented (P) so as to have a 450 nm wavelength absorption peak upon binding carbon monoxide [5]. They are further divided on the basis of sequence similarity, with letters and numbers following CYP [6] denoting its family (number), sub-family (letter), and gene (number); for example CYP3A4. The focus of this chapter will be on the simulation of CYP51A1, also known as lanosterol 14α-demethylase, embedded in a mixed lipid bilayer with and without its lanosterol substrate bound to the active site.

5.1.2 - Clinical Role of CYP

The elimination of foreign substances (xenobiotics) from the body commences with detoxification of compounds occurring in the liver prior to their distribution through the circulatory system [7]. Considering the role of CYPs, they are unsurprisingly primarily located in the liver, but are also found throughout the digestive tract, and have been recently discovered even in the brain [8, 9]. Within the cell, CYP enzymes are either located in the endoplasmic reticulum where they typically metabolise xenobiotics, or in the mitochondria where they are primarily involved in the synthesis of endogenous substances such as cholesterol [10].

Within the liver, there are three phases of metabolism. Phase I essentially renders compounds more hydrophilic through processes such as hydrolysis and oxidation, so as to make them more easily eliminated by the kidneys. CYP enzymes are responsible for the majority of phase I reactions, with as many as half of the drugs introduced to the human body primarily cleared by CYP enzymes [11]. The increased solubility of drug molecules following phase I metabolism facilitates their distribution into

the bloodstream [12]. Phase II is a conjugation reaction involving the addition of a polar group to further increase solubility, and does not involve CYP enzymes, but rather transferase enzymes [13]. Phase III involves the absorption and distribution of drug molecules, leading to their eventual elimination from the body, facilitated by drug transporters which can move the molecules across cellular barriers [14].

Of all the enzymes involved in drug metabolism within the human body, over 70% of them are accounted for by CYP enzymes [15]. The rate of metabolism of a particular drug compound by CYP enzymes typically determines the dose of the drug administered, with the slowly metabolised drugs requiring a low dose and vice versa [16]. Mutations in CYP enzymes tend to lead to a build-up of toxic substances in the body, and also a reduction in the synthesis of important molecules such as cholesterol [17].

5.1.3 - Biosynthesis of Sterols

As many as 14 CYP enzymes are known to be involved in the metabolism of cholesterol and its derivatives, with all but one of those having a role in cholesterol degradation. This involves the formation of bile acids and steroid hormones contributing to cholesterol homeostasis in the human body [18]. The final CYP enzyme is involved in the biosynthesis of cholesterol, belonging to the CYP51 family [19]. However, it also exists in fungi where it is involved in the biosynthesis of ergosterol, a steroid which alters the permeability and fluidity of fungal membranes, much like cholesterol with human cellular membranes [20]. In all cases, lanosterol serves as the substrate and precursor of the product sterol whereby it is demethylated at the 14α position through three consecutive oxidation steps [21], giving CYP51 its alternative name of lanosterol 14α-demethylase. CYP51 is a common target for inhibition in antifungal medication due to the importance of ergosterol in fungal membranes [22, 23]. A fundamental basis of CYP51 inhibition in antifungal medication has relied on retaining the function of the human variant whilst disrupting the activity of the fungal variant [24-26], and as such is mainly applied topically to avoid side effects on cholesterol homeostasis [27, 28].

5.1.4 - CYP51 Structure and Function

Considering the role of CYP51 enzymes, it is necessary for the lanosterol substrate to be oriented correctly so as to successfully undergo demethylation [19]. This substrate orientation has been the subject of controversy [29]. CYP51 function is highlighted in the conservation of specific residues clustered in regions known as substrate recognition sites (SRS) which are located around the

similarly conserved haem binding region [30] (Figure 5i). Substitution mutagenesis studies centered on SRS-1 and SRS-4 have revealed significant reductions in or complete loss of CYP51 activity [31-35].

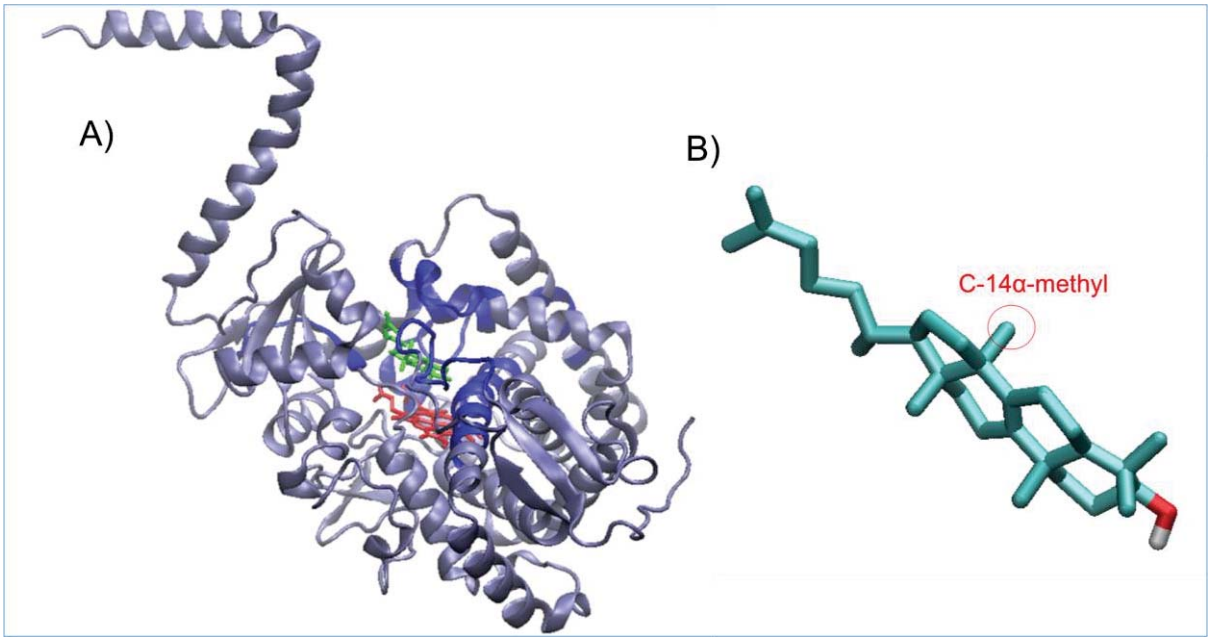


Figure 5i - ***A)*** *Cartoon structure of* S. cerevisiae *CYP51 with lanosterol bound (PDB ID: 4LXJ). CYP51 is shown in light blue, with the lanosterol (green) and haem cofactor (red) present in the binding cavity. All six of the substrate recognition sites are coloured in dark blue - SRS1 (A122 to C142), SRS2 (Y229 to F236), SRS3 (R252 to K259), SRS4 (E303 to A320), SRS5 (H377 to K386), and SRS6 (T507 to G515).* ***B)*** *Structure of lanosterol coloured according to atom type - C (cyan), O (red), H (white). The carbon from the methyl group at the 14α position has been circled in red.*

An x-ray diffraction crystal structure of CYP51 (from *S. cerevisiae*) has been solved with its substrate lanosterol bound (PDB ID: 4LXJ) which allowed insight into its potential orientation relative to the cellular membrane, as well as identifying potential substrate and product channels [36]. CYP51 orientation relative to the cellular membrane plays a key role in ensuring the substrate enters and exits the active site successfully [37], whilst maintaining stability in its position through an N terminal amphipathic transmembrane helical domain [38-40]. This anchoring is proposed to result in the mouth of the substrate channel being slightly embedded into the cellular membrane within the hydrophobic environment [41, 42]. This proposed orientation supports the hypothesis that the lanosterol substrate is directly taken from the cellular membrane into the active site where it undergoes demethylation prior to exiting the active site through a separate channel [43-46].

5.1.5 - Insight from Molecular Dynamics

The positioning of CYP enzymes in the cellular membrane has been studied using computer modelling and simulations in phospholipid bilayers, using a range of approaches from coarse-grained to multi-scale representations [47-49]. However, these have not involved the CYP51 family until very recently, where *T. brucei* CYP51 (PDB ID: 3G1Q) [45] was modelled within a POPC bilayer, along with various other CYP enzymes [50]. This was initially done using a fully coarse-grained model to determine membrane orientation before being converted to an atomistic model to explore the nature of the ligand tunnels. Two major tunnels were identified on the basis of their radius, where one potentially serves as an entrance tunnel allowing access to the active site from the membrane, whilst the other serves as an exit tunnel allowing access to the solvent, and potentially to a subsequent enzyme for further oxidation. After undergoing CYP51's function of sterol biosynthesis, the product can exit through either tunnel dependent on its properties, potentially reinserting into the cellular membrane. It was deduced that due to the rigid nature of the CYP51 binding site, inhibition would require compounds similar in shape and chemical properties to the lanosterol substrate.

5.1.6 - Scope of This Research

In collaboration with Assoc. Prof. Joel Tyndall (University of Otago), atomistic MD simulations of a *S. cerevisiae* CYP51 enzyme (PDB ID: 4LXJ) [36] embedded in a lanosterol and POPC membrane were performed, with and without the lanosterol substrate bound in the active site. The aim was to characterise the nature of CYP51's interaction with the membrane and its effect on the active site, highlighting the differences when the substrate was present and absent. Furthermore, a lipid bilayer consisting of a lanosterol substrate allowed for new insight into the behaviour of the substrate when both inside and outside of the active site. Whilst CYP51's demethylation function cannot currently be observed through MD simulation, the dynamics of its substrate can be observed as it relates to the proposed ingress and egress tunnels identified in the CYP51 structure.

5.2 - Methods

5.2.1 - System Construction

Coordinates for POPC were obtained as detailed in Chapter 2, whilst coordinates for lanosterol were taken from the 4LXJ crystal structure [36], which was also the source of the CYP51 coordinates. The CYP51 structure was obtained in yeast, which was used as a surrogate fungus. Lanosterol parameters were required prior to any MD simulations being performed; they were generated by the Automated Force Field Topology Builder (ATB) [51]. A lipid bilayer with a composition featuring an 11:5 ratio of POPC to lanosterol molecules was constructed as outlined in Chapter 2. POPC was used as the primary phospholipid as it features predominantly in yeast plasma membranes, whilst the phospholipid to sterol ratio is also representative of yeast plasma membranes [52, 53].

5.2.2 - Simulation Methods

The completed lipid bilayer was subjected to 1000 steps of energy minimisation using the steepest descent algorithm in vacuum with the 54A7 force field [54] modified to include lanosterol parameters, then solvated in a rectangular box using the simple point charge (SPC) [55] water model. The MD simulation was initiated with the following equilibration scheme. First, the initial velocities were randomly generated from a Maxwell-Boltzmann distribution at 50 K. The system was then heated to 300 K over the course of 100 ps in the NVT ensemble, with temperature controlled using the Berendsen thermostat [56], with a temperature coupling constant (τ_T) of 0.1 ps, and periodic boundary conditions. The LINCS algorithm [57] was used with an order of 4 to constrain bond lengths and water bond angles, allowing for an integration time step of 2 fs. The centre of mass motion was removed every 100 ps. Non-bonded interactions were calculated using a grid cut-off scheme. The non-bonded interactions within a cut-off distance of 0.9 nm were calculated at every step from a pair list that was updated every fifth time step. At this point, Lennard-Jones interactions between atoms within 1.4 nm were also calculated and were kept constant between updates. Electrostatic interactions were calculated using particle mesh Ewald (PME) summation [58] outside a cut-off distance of 0.9 nm. The system was then further equilibrated for 400 ps at the ideal temperatures outlined above in the NPT ensemble, with pressure controlled using the Berendsen barostat in a semi-isotropic environment, with two pressure coupling constants (τ_p) of 0.5 ps and two isothermal compressibilities of 4.5×10^{-5} bar^{-1}. The final coordinates were used as the starting configuration for a 100 ns MD simulation at 300 K.

Once the lipid bilayer was fully equilibrated, the CYP51 protein was embedded into the membrane using the *g_membed* program as per the protocol outlined by Wolf *et al.* [59]. The orientation for embedment was determined through analogy to other CYP systems previously mentioned (section 5.1.5). The system was then reminimised, resolvated, and neutralised by addition of 3 anions (Cl^-) prior to being re-equilibrated as per the scheme outlined above. The final coordinates were used as the starting configuration for a 150 ns MD simulation at 300 K, with coordinates saved every 50 ps. This process was done twice for the CYP51 protein - once with lanosterol bound in the active site as per the crystal structure, and once with it removed from the active site.

5.2.3 - Analysis Methods

Analysis of the simulations was carried out using GROMACS 4.6 [60] tools for which further detail can be found in section 4.2.4, and simulations were visualised using VMD [61]. Ligand tunnel analysis was carried out using the program CAVER3.0 [62].

5.3 - Results and Discussion

The focus of this research was on the behaviour of CYP51 when in its native membrane-bound state within a lipid bilayer including its lanosterol substrate. The effect of the presence or absence of lanosterol inside the substrate binding cavity was also observed, taking the dynamic and structural changes of the proposed ingress and egress tunnels into consideration. The system consisting of CYP51 with no lanosterol bound will simply be referred to as CYP51, whilst the system featuring a lanosterol substrate in the active site will be referred to as CYP51(L) from this point forwards. Tunnel nomenclature (2f - membrane access, S - solvent access) is taken from Cojocaru *et al.* [63].

5.3.1 - Overview of Conformational Changes

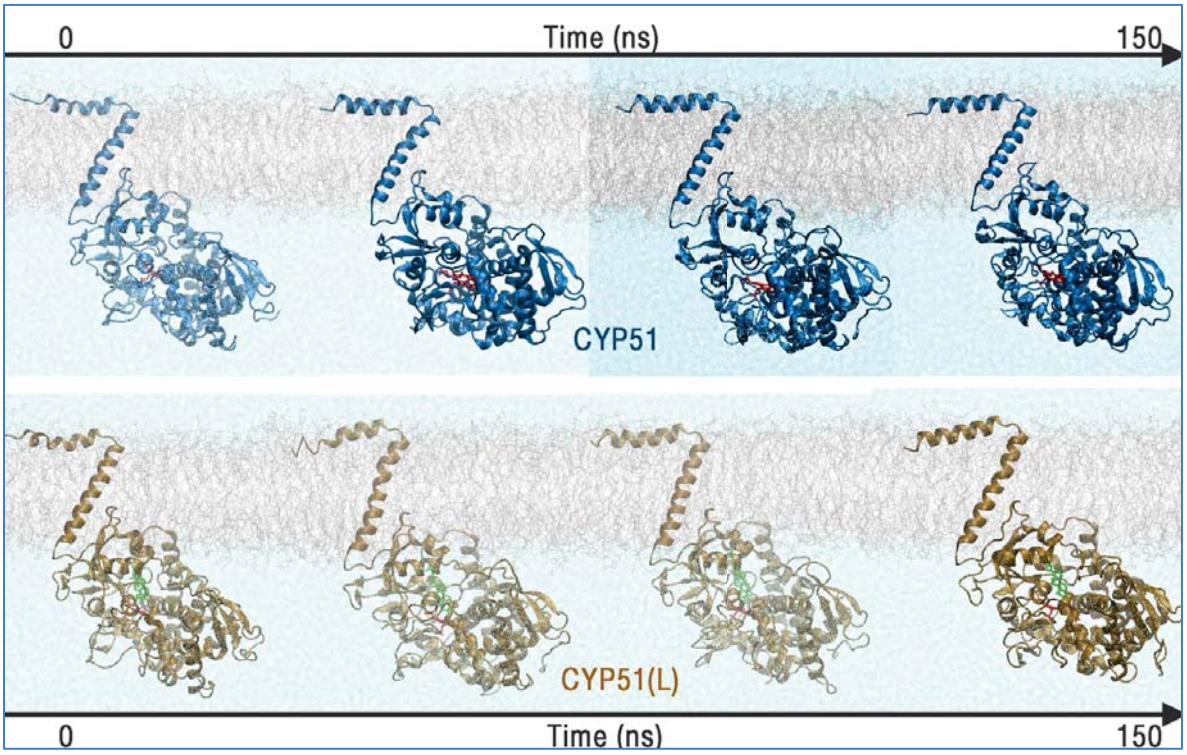


Figure 5ii - *Snapshots from the MD simulations over the course of 150 ns shown at 50 ns intervals, revealing the position of CYP51 embedded in the cell membrane (gray). The CYP51 protein is shown in a blue cartoon structure, the CYP51(L) protein in a gold cartoon structure, the lanosterol substrate in green, and the haem cofactor in red. Water molecules are represented by miniscule cyan dots.*

The CYP51 and CYP51(L) systems had identical starting conformations but differed slightly in their trajectories (Figure 5ii), with CYP51 adopting a looser conformation around the A (residues 76 to 87) and A' (residues 70 to 74) helices situated at the protein-membrane interface. This resulted in CYP51(L) having a final conformation packed more tightly against the membrane, which may have had an effect on the openness of the 2f tunnel whose mouth incorporates residues in the A' helix.

Otherwise, the systems were fundamentally similar, with the amphipathic helix retaining its position spanning the lipid bilayer. The haem group remained planar in nature, adopting an angle of approximately 60° against the normal of the membrane plane, which is in agreement with experimental [64] and computational data [50].

5.3.2 - Structural Analysis

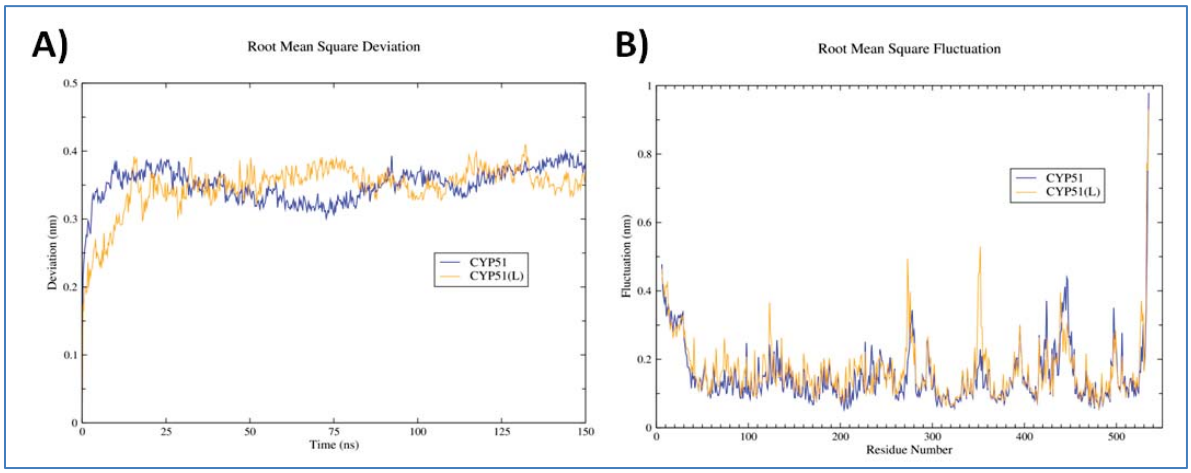


Figure 5iii - ***A)*** *Atom-positional RMSD of the CYP51 (blue) and CYP51(L) (gold) protein structures from their initial conformations over the course of the 150 ns simulation embedded in a cell membrane.* ***B)*** *Atom-positional RMSF averaged over each residue of the CYP51 (blue) and CYP51(L) (gold) proteins averaged over the course of the 150 ns simulation embedded in a cell membrane.*

The root mean square deviation (RMSD) of the CYP51 and CYP51(L) systems revealed that the simulations had equilibrated after approximately 20 ns (Figure 5iii-A). A maximum deviation at any point of 0.2 nm from the original structure suggests that both systems are stable within the lipid bilayer. This essentially means that the behaviour of CYP51 in the membrane over the last 125 ns of the simulation will be indicative of its functional behaviour within the limitations of the parameters and components present in the system.

The root mean square fluctuation (RMSF) highlights residues exhibiting the most conformational change over the course of the simulation (Figure 5iii-B). Both CYP51 and CYP51(L) exhibited relatively high fluctuation in the following regions: the GH loop (residues 274-284), the general region between residues 400-460 which features β1 and β2 (fungal specific loop), and finally the β4 hairpin (residues 506-511) located within the substrate recognition site 6 (SRS6). This is in good agreement with the results of previous simulations of various CYP enzymes in solution and membrane-bound [50]. However, whilst the region between residues 350-365 exhibits low levels of fluctuation in the CYP51 system as seen in Yu *et al.*'s simulation results, the mobility in this region is increased more

than twofold in the previously uncharacterised CYP51(L) system. This would indicate a conformational difference in this region, corresponding to the JK loop, which is influenced by the presence of lanosterol within the active site. This region lies between SRS4 and SRS5, and does not exist within the infrastructure of either the ingress or egress tunnels, so the cause and effect of this conformational change is not due to direct interaction with lanosterol. It remains a possibility, however, that this alteration could have consequently led to the collapse or opening of one of these tunnels, which will be explored in section 5.3.5. SRS1 (residues 122-142), SRS2 (residues 229-236), SRS3 (residues 252-259), SRS4 (residues 303-320), and SRS5 (residues 377-386) all exhibited minimal fluctuation; suggesting that the integrity of the substrate binding cavity is well maintained.

5.3.3 - Membrane Interaction

As previously mentioned, the transmembrane helix remains in a position spanning the lipid bilayer throughout the course of the 150 ns simulation in both the case of CYP51 and CYP51(L). This can be seen in the maintenance of upwards of 20 hydrogen bonds between backbones of the helix residues and the membrane following equilibration. This equilibration process can therefore be characterised by an increase in hydrogen bond formation, as the transmembrane helix adapts to being in the membrane. The vast majority of the 50 residues encompassing this transmembrane helix are hydrophobic in nature, with the one notable exception being arginine at position 30 (R30). R30 is positioned at the start of the second helix with its side chain situated at the level of the lipid headgroups, which suggests interaction with the polar region of the membrane could play a role in determining the point at which the N terminal tail turns to be perpendicular to the membrane normal after having spanned the lipid bilayer along the normal.

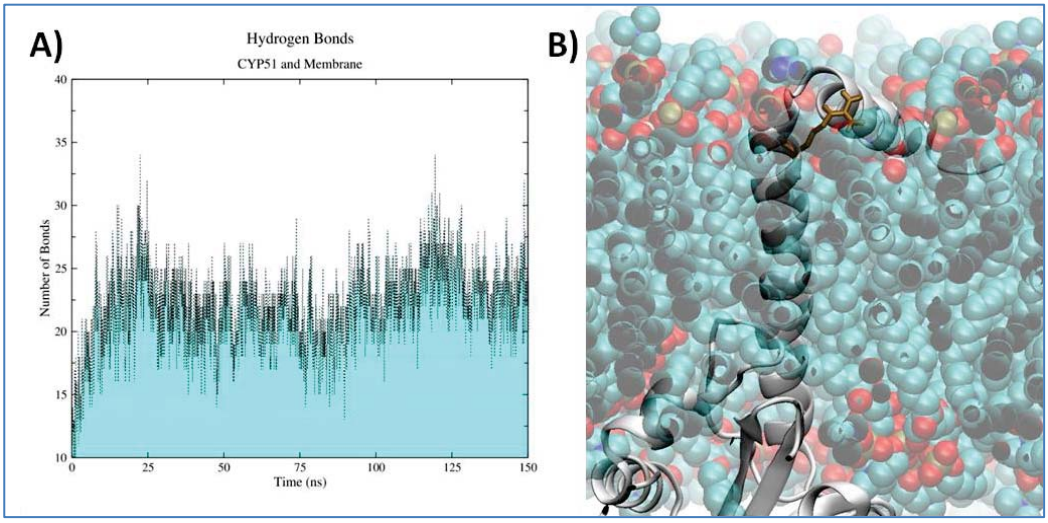


Figure 5iv - ***A)*** *The number of hydrogen bonds formed between CYP51 and the membrane over the course of the 150 ns simulation.* ***B)*** *The position of the transmembrane helix (white) anchored to the membrane coloured by atom type - C (cyan), N (blue), O (red), H (white), P (gold), showing the relative position of R30 (gold) at the level of the lipid headgroups.*

The FG loop (residues 234-247) has been found to be embedded in the membrane in both experimental [65] and computational [50] studies. This is also the case in these simulations, with both the CYP51 and CYP51(L) systems having this feature (Figure 5v). The positioning of this loop within the membrane undoubtedly plays a part in substrate recruitment, as it also contributes residues to the mouth of the 2f ingress tunnel, along with the A' helix (residues 70-74) and β4 hairpin (residues 506-511).

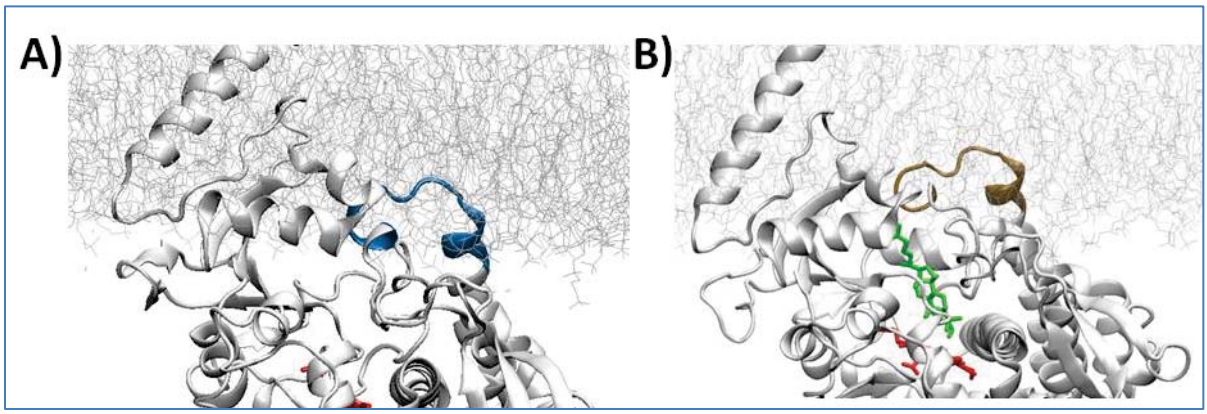


Figure 5v - ***A)*** *Snapshot of the CYP51 protein (white) after 150 ns of simulation embedded in the cell membrane (gray) highlighting the position of the FG loop (blue). The haem cofactor (red) is also shown.* ***B)*** *Snapshot of the CYP51(L) protein (white) after 150 ns of simulation embedded in the cell membrane (gray) highlighting the position of the FG loop (gold). The haem cofactor (red) and lanosterol substrate (green) are also shown.*

5.3.4 - Substrate Behaviour

In the CYP51(L) system, lanosterol was situated in the substrate binding cavity at the beginning of the simulation. The minimum distance between the lanosterol substrate and the lipid bilayer was monitored over the course of the simulation to see whether its position altered. The lanosterol substrate starts out approximately 1 nm away from the lipids in the cell membrane, and attains a minimum distance of 0.4 nm closely following equilibration (25 ns). It does not, however, completely exit the binding cavity (Figure 5vi).

It has been previously suggested that, once in the binding cavity, the substrate can exit through the S tunnel or equivalently back through the 2f ingress tunnel into the membrane depending on the substrate properties [43-46]. This would suggest that the chemical change effected by demethylation causes the ergosterol/cholesterol precursor to exit CYP51 to be further oxidised to its ergosterol/cholesterol product [21], whereas the unmodified lanosterol may instead return to the cellular membrane. The retention of lanosterol in the binding cavity during the CYP51(L) simulation may simply be due to the relatively short timescale of the simulation compared to physiological

timescales. As well as increasing the length of the simulation, it would be of interest in the future to also carry out simulations with an ergosterol/cholesterol precursor bound.

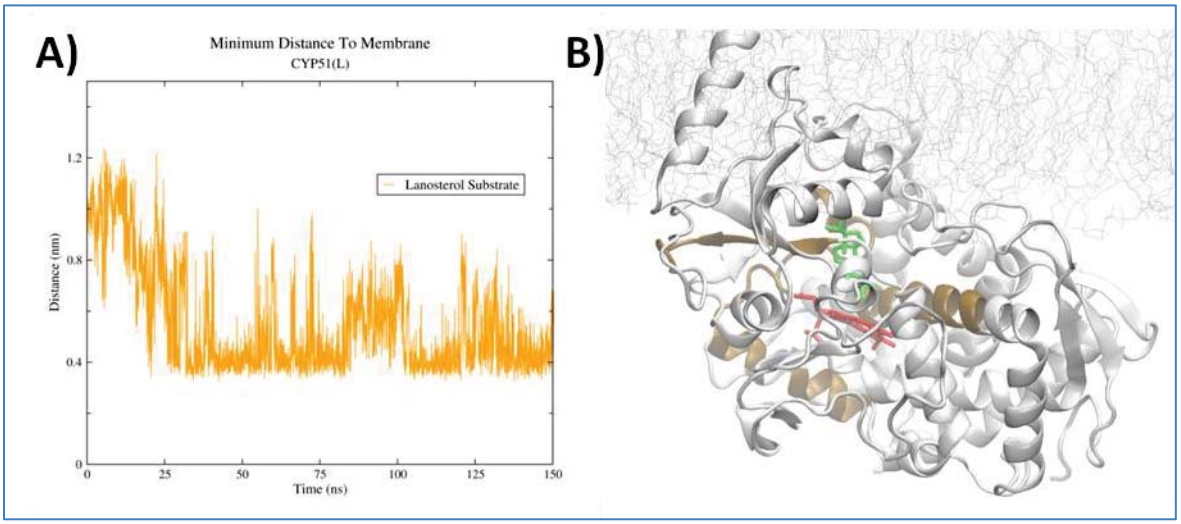


Figure 5vi - ***A)*** *The minimum distance between the lanosterol substrate and the cell membrane in the CYP51(L) system.* ***B)*** *Snapshot of the CYP51(L) protein (white) after 150 ns of simulation embedded in the cell membrane (gray) highlighting the position of the substrate binding cavity (gold). The haem cofactor (red) and lanosterol substrate (green) are also shown.*

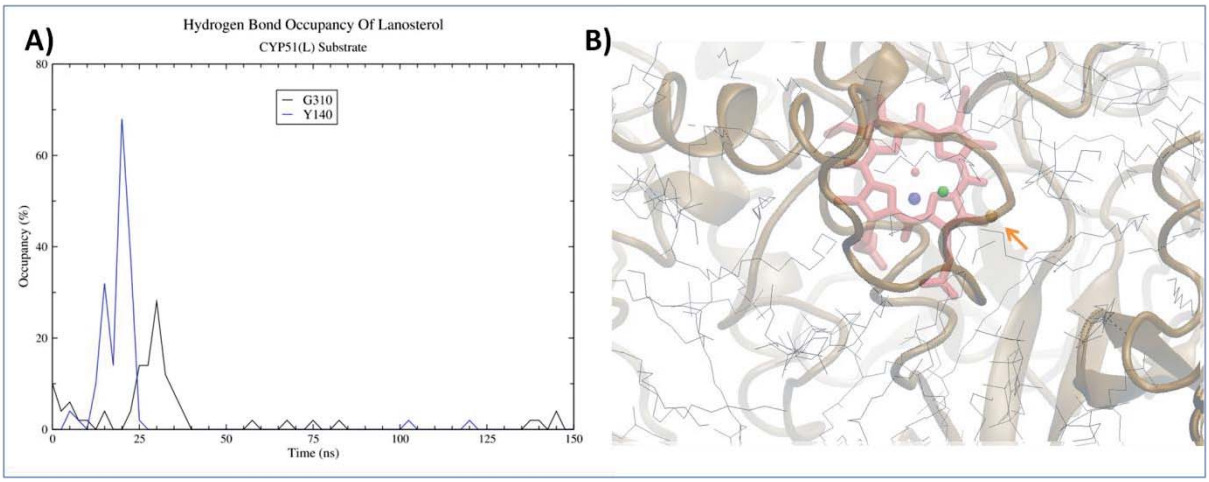


Figure 5vii - ***A)*** *Relative occupancy of the hydrogen bonds formed by the lanosterol substrate in the CYP51(L) system over the course of the 150 ns simulation. Hydrogen bonds with G310 are shown by a black line and hydrogen bonds with Y140 are shown by a blue line.* ***B)*** *Looking down through the membrane (gray) into the active site: relative positions of the 14-methyl group to the haem cofactor (red) in the aligned 3P99 structure (blue), 4LXJ structure (orange), and the CYP51(L) system (green) following 150 ns of simulation. The CYP51(L) protein is also shown in gold.*

The simulated CYP51 protein exhibits two hydrogen bonds with the lanosterol substrate, one with the tyrosine at residue position 140 (Y140), and the other with glycine at residue position 310 (G310) (Figure 5vii-A). These two residues are situated in SRS1 and SRS4 respectively, belonging to the two most highly conserved motifs of the CYP51 family [19]. These 2 bonds exist primarily in the first 40 ns of the simulation, at which point they disappear due to the lanosterol substrate closing the distance to the haem cofactor over the course of the simulation (Figure 5vii-B). The aforementioned controversy in lanosterol orientation (Section 5.1.4) [29] surrounds the position of the 14-methyl group targeted for oxidation relative to the haem cofactor, whereby crystal structure 3P99 [45] presents 14α-methylenecyclopropyl-Δ7-24,25-dihydrolanosterol (MCP) in a flipped orientation when compared to the lanosterol substrate of the 4LXJ crystal structure [36]. Following substrate recognition by SRS1 and SRS4, it is necessary for the lanosterol substrate to approach the haem for the functional oxidation activity of CYP51 to occur. It is plausible that the orientation of lanosterol in the 4LXJ crystal structure is representative of a pre-catalytic state, whereby contacts with SRS1 and SRS4 occur; alternatively, the orientation of MCP in crystal structure 3P99 is representative of a catalytic state, whereby the 14-methyl group points towards the iron in the functional haem cofactor. The simulation of the CYP51(L) simulation has thus provided insight into a potential pathway describing the substrate movement required for catalysis, so as to bring the 14-methyl group from a position described by the 4LXJ structure closer to a position described by the 3P99 structure (Figure 5vii-B).

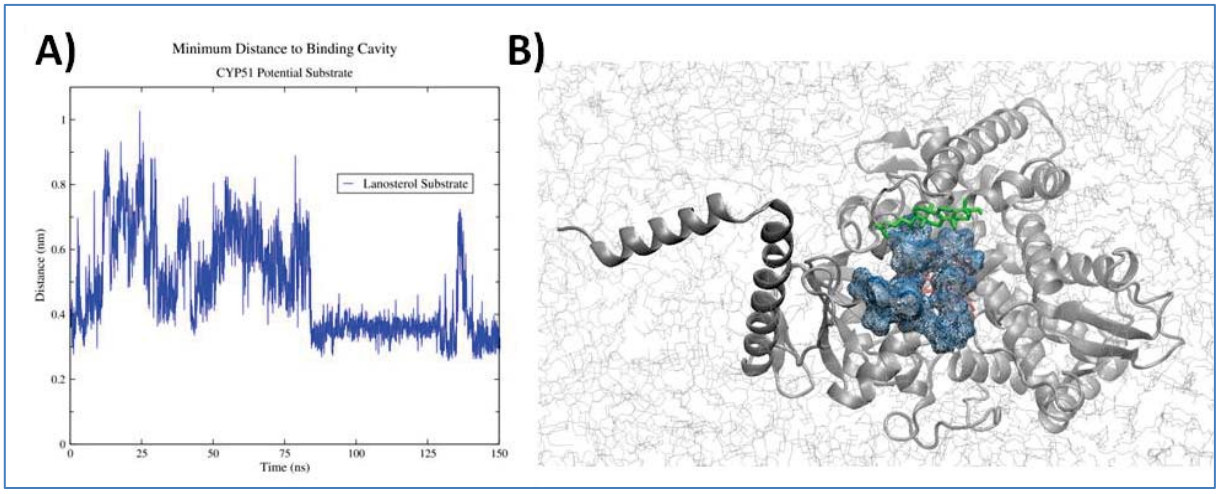


Figure 5viii - ***A)*** *The minimum distance between the potential lanosterol substrate and the binding cavity in the CYP51 system.* ***B)*** *Snapshot of the CYP51 protein (white) after 150 ns of simulation looking down through the cell membrane (gray) highlighting the relative position of the potential lanosterol substrate (green) to the tunnel entry to the substrate binding cavity (blue). The haem cofactor (red) is also shown.*

On the other hand, the CYP51 system featured no lanosterol substrate within the binding cavity. The behaviour of all lanosterols in the lipid bilayer were monitored over the course of the simulation, before one was chosen as the most likely potential substrate based on its proximity to the ingress tunnel. The CYP51 enzyme diffused across the membrane before settling in a position adjacent to the potential substrate, resulting in a minimum distance of 0.3 nm between the binding cavity and the substrate for the final half of the simulation (Figure 5viii). It also went from an expectedly upright position parallel to the remainder of the lipids to a rather surprising perpendicular position. As the entrance to the 2f tunnel is embedded beyond the level of the lipid headgroups, it is necessary for the lanosterol substrate to alter its orientation for either its polar end or tail end to be accessible to the tunnel entrance for it to end up in the orientation seen in the 4LXJ crystal structure [36] or 3P99 crystal structure [45] respectively. Thus, this change in orientation may be a precursor to entry into the tunnel.

5.3.5 - Behaviour of Ingress/Egress Tunnels

The goal of this section was to follow the behaviour of the proposed ingress and egress tunnels over the course of the simulation, to see whether they became or remained accessible on the basis of the presence or absence of a substrate in the binding cavity. To do so initially, the final protein structure in the CYP51(L) system was analysed, with the focus on residues at the mouth of the 2f (membrane entry) and S (solvent exit) tunnels (Figure 5ix).

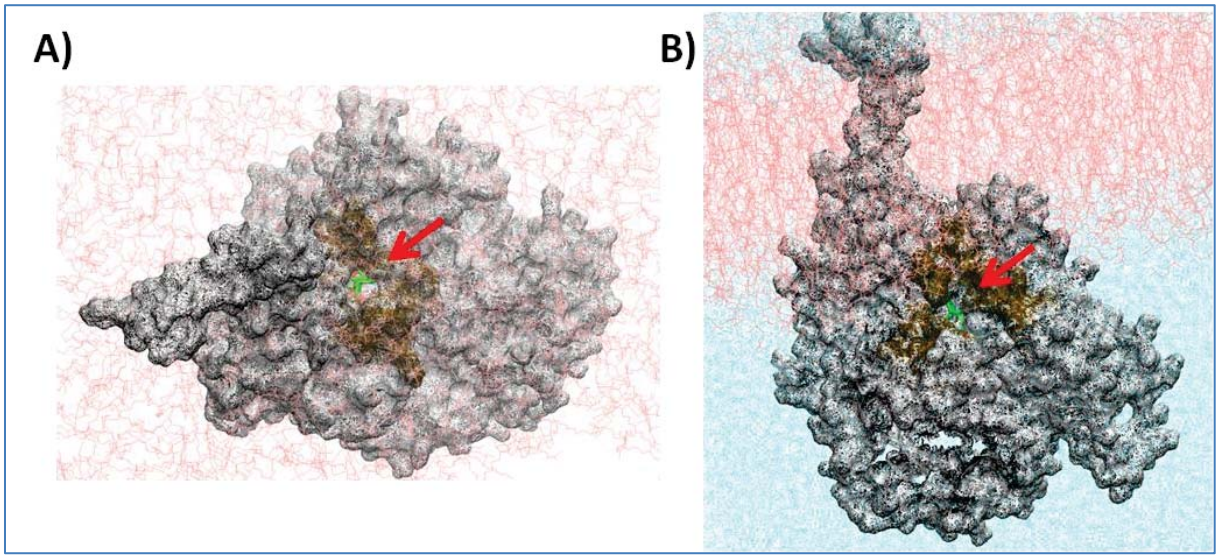


Figure 5ix - ***A)*** *Snapshot of the CYP51(L) protein (silver) depicted in a wireframe surface representation looking down through the cell membrane (pink). The 2f tunnel (red arrow) and its lining residues (gold) are highlighted with the lanosterol substrate (green) visible through the tunnel.* ***B)*** *Snapshot of the CYP51(L) protein (silver) depicted in a wireframe surface embedded in the cell membrane (pink). The S tunnel (red arrow) and its lining residues (gold) are highlighted with the lanosterol substrate (green) visible through the tunnel. The water molecules are also shown in cyan.*

A clear path was shown from the membrane into the binding cavity via the 2f tunnel, as well as from the binding cavity into the solvent via the S tunnel. The lanosterol substrate situated within the binding cavity is easily visible in both cases; however, its orientation would suggest that exit via the 2f tunnel back into the membrane is the most likely outcome were the simulation extended indefinitely.

The program CAVER3.0 [62] was used to analyse the behaviour of the two tunnels over the course of the simulation, in both the CYP51 and CYP51(L) systems. This program determines the most favourable pathway to access the protein surface from a user selected starting point within a cavity. The trajectories were broken down into individual frames every 50 ns, at which point the coordinates were evaluated for the presence of tunnels. The starting point of the analysis was determined for each coordinate file individually as being approximately 0.4 nm higher than the iron in the haem group as the best representation of the active site [63]. The tunnels were determined to have either been open or closed at each particular time point, based on a bottleneck radius of 0.14 nm, with a greater radius signifying an open nature [62].

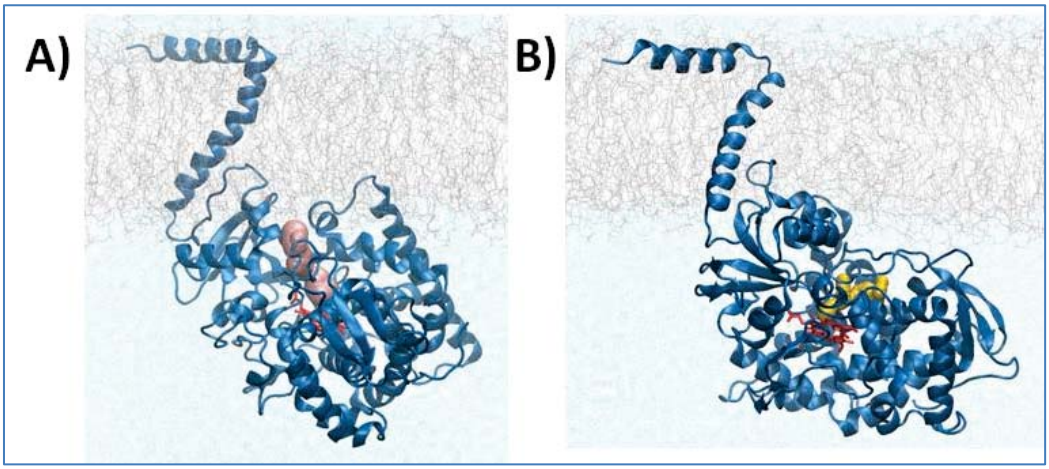


Figure 5x - ***A)*** *Snapshot of the CYP51 protein (blue) embedded in the cell membrane (gray) after 0 ns of simulation with the 2f tunnel (pink) shown as analysed in CAVER3.0. The haem cofactor is also seen in red, with the water molecules depicted in cyan.* ***B)*** *Snapshot of the CYP51 protein (blue) embedded in the cell membrane (gray) after 50 ns of simulation with the S tunnel (yellow) shown as analysed in CAVER3.0. The haem cofactor is shown in red, with the water molecules depicted in cyan.*

Following the use of a probe radius of 0.12 nm as recommended by Cojocaru *et al.* [63], only two tunnels were ever detected over the course of either simulation (Figure 5x); these were, as surmised, tunnels 2f and S. The 2f tunnel was lined by the A' helix (residues 70-73), SRS1 (residues 126-140), F" helix (residues 236-242), SRS4 (residues 311-318), SRS5 (residues 379-385), and the β4

hairpin in SRS6 (residues 506-510). The S tunnel was lined by the F helix (residues 228-237), SRS4 (residues 312-321), SRS5 (residues 379-380), and the β4 hairpin in SRS6 (residues 508-512). These tunnel lining residues agree with Yu *et al.*'s simulation findings [50]

In the CYP51 system where the lanosterol substrate was removed from the active site, the 2f tunnel was only found to be open at the beginning of the simulation (Figure 5x-A) at which point the S tunnel was closed. As the simulation progressed, the S tunnel was found to be open after 50 ns (Figure 5x-B), 100 ns, and 150 ns, while the 2f tunnel was closed at these points.

In the CYP51(L) system where the lanosterol substrate was left in the active site, both the 2f and S tunnels were found to be open at the beginning of the simulation (Figure 5xi). However, contrary to the CYP51 system, the 2f tunnel remained open for the entire duration of the simulation, while the S tunnel was found to be closed at all other time points analysed.

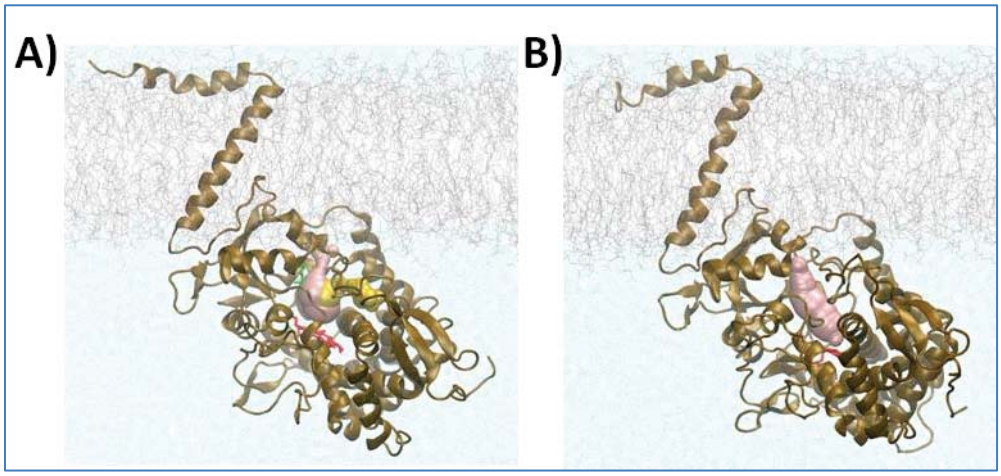


Figure 5xi - ***A)*** *Snapshot of the CYP51(L) protein (gold) embedded in the cell membrane (gray) after 0 ns of simulation with the 2f tunnel (pink) and S tunnel (yellow) shown as analysed in CAVER3.0. The haem cofactor is also seen in red, the lanosterol substrate in green, and the water molecules depicted in cyan.* ***B)*** *Snapshot of the CYP51(L) protein (gold) embedded in the cell membrane (gray) after 50 ns of simulation with the 2f tunnel (pink) shown as analysed in CAVER3.0. The haem cofactor is also seen in red, the lanosterol substrate in green, and the water molecules depicted in cyan.*

When the binding cavity is occupied, the 2f tunnel remains open whilst the S tunnel remains closed (Figure 5xi-B) suggesting one of two possibilities - either the eventual product will enter the membrane, or the S tunnel will only open once the CYP51 enzyme has carried out its function. When the binding cavity is unoccupied, the 2f tunnel remains closed whilst the S tunnel remains open (Figure 5x-B) also suggesting one of two possibilities - either the eventual substrate is recruited from the cytosol, or more likely, the 2f tunnel will only open once a potential substrate has been recognised.

This is not in keeping with Yu *et al.*'s work which found that both ligand-bound and ligand-free CYP51 maintained both tunnels open when membrane-bound [50]. These differences can be explained by the nature of the substrate, which was lanosterol in this study (solved by crystal structure) and the MCP inhibitor in Yu *et al.*'s work modelled through analogy following conversion from a coarse-grained representation to an all-atom model. Yu *et al.*'s work also used a *T. bruicei* CYP51 enzyme (29% sequence identity [66]) with the amphipathic helix modelled, while the *S. cerevisiae* CYP51 enzyme structure used in this case was solved with a much longer amphipathic helix included.

5.4 - Conclusion

The CYP51 protein systems appear to be structurally stable when embedded inside a lipid bilayer, regardless of the presence or absence of a lanosterol substrate in its binding cavity. The stability of the membrane-bound protein is undoubtedly enhanced by the presence of a transmembrane helix spanning the length of the membrane normal, making an almost 90° turn at the level of the opposite headgroups potentially determined by the presence of a lone polar residue (R30) in the helix. The residues surrounding the binding cavity exhibit little conformational change suggesting the integrity of the binding cavity is kept through CYP51's stable membrane interaction.

When the lanosterol substrate is present in the binding cavity, it initially interacts with the highly conserved SRS1 and SRS4 regions prior to its repositioning to be closer to the haem cofactor for the remainder of the simulation. Considering the oxidation function of CYP51, this final orientation of the lanosterol substrate is plausibly indicative of a catalytically competent position. This position is unseen in the 4LXJ crystal structure as the lanosterol substrate may not advance from a position allowing substrate recognition until CYP51 is in a membrane-bound state.

It has been proposed previously that the substrate has two pathways in and out of the binding cavity, namely the 2f (membrane access) and S (solvent access tunnels). The simulations show that the open or closed nature of these tunnels appears to be regulated by the presence or absence of a substrate inside the binding cavity. The behaviour of the tunnels over the course of the simulation suggests that when the active site is occupied by a substrate, the 2f tunnel is open whilst the S tunnel is closed. However, when the active site is devoid of a substrate, the opposite is true. This could be a result of the source of substrate recruitment; however, observation of the recruitment of a substrate from the membrane or the cytosol would require considerably more simulation time. The nature of the product may also determine the preferred exit pathway of the enclosed sterol, which could highlight differences between the behaviour of human CYP51 (cholesterol product) and fungal CYP51 (ergosterol product) for example.

For additional information to be gleaned on the effect the nature of the substrate or product has on CYP51's behaviour, additional simulations would have to be run with a range of ligands, including ergosterol and its precursors, contained in the active site with potentially a range of cell membranes consisting of varying lipid compositions reflective of the host organism.

5.5 - Bibliography

1. Nelson, D.R., *Introductory remarks on human CYPs.* Drug Metabolism Reviews, 2002. **34**(1-2): p. 1-5.
2. Lamb, D.C., L. Lei, A.G.S. Warrilow, G.I. Lepesheva, J.G.L. Mullins, M.R. Waterman, and S.L. Kelly, *The First Virally Encoded Cytochrome P450.* Journal of Virology, 2009. **83**(16): p. 8266-8269.
3. Hannemann, F., A. Bichet, K.M. Ewen, and R. Bernhardt, *Cytochrome P450 systems—biological variations of electron transport chains.* Biochimica et Biophysica Acta (BBA) - General Subjects, 2007. **1770**(3): p. 330-344.
4. Wijnen, P.A.H.M., R.A.M. Op Den Buijsch, M. Drent, P.M.J.C. Kuipers, C. Neef, A. Bast, O. Bekers, and G.H. Koek, *Review article: the prevalence and clinical relevance of cytochrome P450 polymorphisms.* Alimentary Pharmacology & Therapeutics, 2007. **26**: p. 211-219.
5. Ortiz de Montellano, P.R., *The 1994 Bernard B. Brodie Award Lecture. Structure, mechanism, and inhibition of cytochrome P450.* Drug Metabolism and Disposition, 1995. **23**(11): p. 1181-1187.
6. Nelson, D.R., L. Koymans, T. Kamataki, J.J. Stegeman, R. Feyereisen, D.J. Waxman, M.R. Waterman, O. Gotoh, M.J. Coon, R.W. Estabrook, I.C. Gunsalus, and D.W. Nebert, *P450 superfamily: update on new sequences, gene mapping, accession numbers and nomenclature.* Pharmacogenetics, 1996. **6**(1): p. 1-42.
7. Oinonen, T. and K.O. Lindros, *Zonation of hepatic cytochrome P-450 expression and regulation.* Biochem J, 1998. **329 (Pt 1)**: p. 17-35.
8. Krishna, D.R. and U. Klotz, *Extrahepatic metabolism of drugs in humans.* Clin Pharmacokinet, 1994. **26**(2): p. 144-60.
9. Meyer, R.P., M. Gehlhaus, R. Knoth, and B. Volk, *Expression and function of cytochrome p450 in brain drug metabolism.* Curr Drug Metab, 2007. **8**(4): p. 297-306.
10. Estabrook, R.W., *A passion for P450s (remembrances of the early history of research on cytochrome P450).* Drug Metabolism and Disposition, 2003. **31**(12): p. 1461-1473.
11. Bertz, R.J. and G.R. Granneman, *Use of in vitro and in vivo data to estimate the likelihood of metabolic pharmacokinetic interactions.* Clin Pharmacokinet, 1997. **32**(3): p. 210-58.
12. Guengerich, F.P., *Common and Uncommon Cytochrome P450 Reactions Related to Metabolism and Chemical Toxicity.* Chemical Research in Toxicology, 2001. **14**(6): p. 611-650.
13. Jakoby, W.B. and D.M. Ziegler, *The enzymes of detoxication.* J Biol Chem, 1990. **265**(34): p. 20715-8.
14. Xu, C., C.Y. Li, and A.N. Kong, *Induction of phase I, II and III drug metabolism/transport by xenobiotics.* Arch Pharm Res, 2005. **28**(3): p. 249-68.
15. Guengerich, F.P., *Cytochrome P450 and Chemical Toxicology.* Chemical Research in Toxicology, 2008. **21**(1): p. 70-83.
16. Lynch, T. and A. Price, *The effect of cytochrome P450 metabolism on drug response, interactions, and adverse effects.* Am Fam Physician, 2007. **76**(3): p. 391-6.
17. Nebert, D.W. and D.W. Russell, *Clinical importance of the cytochromes P450.* The Lancet. **360**(9340): p. 1155-1162.
18. Pikuleva, I.A., *Cytochrome P450s and cholesterol homeostasis.* Pharmacology & Therapeutics, 2006. **112**(3): p. 761-773.
19. Lepesheva, G.I. and M.R. Waterman, *Sterol 14α-Demethylase Cytochrome P450 (CYP51), a P450 in all Biological Kingdoms.* Biochimica et biophysica acta, 2007. **1770**(3): p. 467-477.
20. Becher, R. and S.G. Wirsel, *Fungal cytochrome P450 sterol 14alpha-demethylase (CYP51) and azole resistance in plant and human pathogens.* Appl Microbiol Biotechnol, 2012. **95**(4): p. 825-40.
21. Waterman, M.R. and G.I. Lepesheva, *Sterol 14 alpha-demethylase, an abundant and essential mixed-function oxidase.* Biochem Biophys Res Commun, 2005. **338**(1): p. 418-22.

22. Kritchevsky, D., *Biochemistry of Steroids and Other Isopentenoids. William R. Nes , Margaret Lee McKean Cholesterol. John R. Sabine.* The Quarterly Review of Biology, 1978. **53**(3): p. 298-299.
23. Lepesheva, G.I., T.Y. Hargrove, Y. Kleshchenko, W.D. Nes, F. Villalta, and M.R. Waterman, *CYP51: A major drug target in the cytochrome P450 superfamily.* Lipids, 2008. **43**(12): p. 1117-25.
24. Lamb, D.C., D.E. Kelly, M.R. Waterman, M. Stromstedt, D. Rozman, and S.L. Kelly, *Characteristics of the heterologously expressed human lanosterol 14α-demethylase (other names: P45014DM, CYP51, P45051) and inhibition of the purified human and Candida albicans CYP51 with azole antifungal agents.* Yeast, 1999. **15**(9): p. 755-763.
25. Kelly, S.L., D.C. Lamb, M. Cannieux, D. Greetham, C.J. Jackson, T. Marczylo, C. Ugochukwu, and D.E. Kelly, *An old activity in the cytochrome P450 superfamily (CYP51) and a new story of drugs and resistance.* Biochem Soc Trans, 2001. **29**(Pt 2): p. 122-8.
26. Lepesheva, G.I., W.D. Nes, W. Zhou, G.C. Hill, and M.R. Waterman, *CYP51 from Trypanosoma brucei is obtusifoliol-specific.* Biochemistry, 2004. **43**(33): p. 10789-99.
27. Miettinen, T.A., *Cholesterol metabolism during ketoconazole treatment in man.* J Lipid Res, 1988. **29**(1): p. 43-51.
28. Debeljak, N., M. Fink, and D. Rozman, *Many facets of mammalian lanosterol 14α-demethylase from the evolutionarily conserved cytochrome P450 family CYP51.* Archives of Biochemistry and Biophysics, 2003. **409**(1): p. 159-171.
29. Choi, J.Y., L.M. Podust, and W.R. Roush, *Drug Strategies Targeting CYP51 in Neglected Tropical Diseases.* Chemical Reviews, 2014. **114**(22): p. 11242-11271.
30. Yoshida, Y., Y. Aoyama, M. Noshiro, and O. Gotoh, *Sterol 14-demethylase P450 (CYP51) provides a breakthrough for the discussion on the evolution of cytochrome P450 gene superfamily.* Biochem Biophys Res Commun, 2000. **273**(3): p. 799-804.
31. Lepesheva, G.I., C. Virus, and M.R. Waterman, *Conservation in the CYP51 family. Role of the B' helix/BC loop and helices F and G in enzymatic function.* Biochemistry, 2003. **42**(30): p. 9091-101.
32. Nitahara, Y., K. Kishimoto, Y. Yabusaki, O. Gotoh, Y. Yoshida, T. Horiuchi, and Y. Aoyama, *The amino acid residues affecting the activity and azole susceptibility of rat CYP51 (sterol 14-demethylase P450).* J Biochem, 2001. **129**(5): p. 761-8.
33. Bellamine, A., G.I. Lepesheva, and M.R. Waterman, *Fluconazole binding and sterol demethylation in three CYP51 isoforms indicate differences in active site topology.* J Lipid Res, 2004. **45**(11): p. 2000-7.
34. Aoyama, Y., *Recent progress in the CYP51 research focusing on its unique evolutionary and functional characteristics as a diversozyme P450.* Front Biosci, 2005. **10**: p. 1546-57.
35. Vidakovic, M., S.G. Sligar, H. Li, and T.L. Poulos, *Understanding the role of the essential Asp251 in cytochrome p450cam using site-directed mutagenesis, crystallography, and kinetic solvent isotope effect.* Biochemistry, 1998. **37**(26): p. 9211-9.
36. Monk, B.C., T.M. Tomasiak, M.V. Keniya, F.U. Huschmann, J.D.A. Tyndall, J.D. O'Connell, R.D. Cannon, J.G. McDonald, A. Rodriguez, J.S. Finer-Moore, and R.M. Stroud, *Architecture of a single membrane spanning cytochrome P450 suggests constraints that orient the catalytic domain relative to a bilayer.* Proceedings of the National Academy of Sciences of the United States of America, 2014. **111**(10): p. 3865-3870.
37. White, S.H., A.S. Ladokhin, S. Jayasinghe, and K. Hristova, *How membranes shape protein structure.* J Biol Chem, 2001. **276**(35): p. 32395-8.
38. Williams, P.A., J. Cosme, V. Sridhar, E.F. Johnson, and D.E. McRee, *Mammalian microsomal cytochrome P450 monooxygenase: structural adaptations for membrane binding and functional diversity.* Mol Cell, 2000. **5**(1): p. 121-31.

39. Wade, R.C., D. Motiejunas, K. Schleinkofer, Sudarko, P.J. Winn, A. Banerjee, A. Kariakin, and C. Jung, *Multiple molecular recognition mechanisms. Cytochrome P450--a case study.* Biochim Biophys Acta, 2005. **1754**(1-2): p. 239-44.
40. Zhao, Y., M.A. White, B.K. Muralidhara, L. Sun, J.R. Halpert, and C.D. Stout, *Structure of microsomal cytochrome P450 2B4 complexed with the antifungal drug bifonazole: insight into P450 conformational plasticity and membrane interaction.* J Biol Chem, 2006. **281**(9): p. 5973-81.
41. Black, S.D., *Membrane topology of the mammalian P450 cytochromes.* FASEB J, 1992. **6**(2): p. 680-5.
42. Berka, K., T. Hendrychova, P. Anzenbacher, and M. Otyepka, *Membrane position of ibuprofen agrees with suggested access path entrance to cytochrome P450 2C9 active site.* J Phys Chem A, 2011. **115**(41): p. 11248-55.
43. Podust, L.M., T.L. Poulos, and M.R. Waterman, *Crystal structure of cytochrome P450 14alpha -sterol demethylase (CYP51) from Mycobacterium tuberculosis in complex with azole inhibitors.* Proc Natl Acad Sci U S A, 2001. **98**(6): p. 3068-73.
44. Strushkevich, N., S.A. Usanov, and H.W. Park, *Structural basis of human CYP51 inhibition by antifungal azoles.* J Mol Biol, 2010. **397**(4): p. 1067-78.
45. Lepesheva, G.I., H.W. Park, T.Y. Hargrove, B. Vanhollebeke, Z. Wawrzak, J.M. Harp, M. Sundaramoorthy, W.D. Nes, E. Pays, M. Chaudhuri, F. Villalta, and M.R. Waterman, *Crystal structures of Trypanosoma brucei sterol 14alpha-demethylase and implications for selective treatment of human infections.* J Biol Chem, 2010. **285**(3): p. 1773-80.
46. Magistrato, A., J. Sgrignani, R. Krause, and A. Cavalli, *Single or Multiple Access Channels to the CYP450s Active Site? An Answer from Free Energy Simulations of the Human Aromatase Enzyme.* J Phys Chem Lett, 2017. **8**(9): p. 2036-2042.
47. Cojocaru, V., K. Balali-Mood, M.S. Sansom, and R.C. Wade, *Structure and dynamics of the membrane-bound cytochrome P450 2C9.* PLoS Comput Biol, 2011. **7**(8): p. e1002152.
48. Denisov, I.G., A.Y. Shih, and S.G. Sligar, *Structural differences between soluble and membrane bound cytochrome P450s.* J Inorg Biochem, 2012. **108**: p. 150-8.
49. Berka, K., M. Paloncyova, P. Anzenbacher, and M. Otyepka, *Behavior of human cytochromes P450 on lipid membranes.* J Phys Chem B, 2013. **117**(39): p. 11556-64.
50. Yu, X., V. Cojocaru, G. Mustafa, O.M.H. Salo-Ahen, G.I. Lepesheva, and R.C. Wade, *Dynamics of CYP51: implications for function and inhibitor design.* Journal of molecular recognition : JMR, 2015. **28**(2): p. 59-73.
51. Malde, A.K., L. Zuo, M. Breeze, M. Stroet, D. Poger, P.C. Nair, C. Oostenbrink, and A.E. Mark, *An Automated Force Field Topology Builder (ATB) and Repository: Version 1.0.* Journal of Chemical Theory and Computation, 2011. **7**(12): p. 4026-4037.
52. Schneiter, R., B. Brügger, R. Sandhoff, G. Zellnig, A. Leber, M. Lampl, K. Athenstaedt, C. Hrastnik, S. Eder, G. Daum, F. Paltauf, F.T. Wieland, and S.D. Kohlwein, *Electrospray Ionization Tandem Mass Spectrometry (Esi-Ms/Ms) Analysis of the Lipid Molecular Species Composition of Yeast Subcellular Membranes Reveals Acyl Chain-Based Sorting/Remodeling of Distinct Molecular Species En Route to the Plasma Membrane.* The Journal of Cell Biology, 1999. **146**(4): p. 741-754.
53. Alvarez, F.J., L.M. Douglas, and J.B. Konopka, *Sterol-Rich Plasma Membrane Domains in Fungi.* Eukaryotic Cell, 2007. **6**(5): p. 755-763.
54. Schmid, N., A.P. Eichenberger, A. Choutko, S. Riniker, M. Winger, A.E. Mark, and W.F. van Gunsteren, *Definition and testing of the GROMOS force-field versions 54A7 and 54B7.* European Biophysics Journal, 2011. **40**(7): p. 843.
55. Berendsen, H.J.C., J.P.M. Postma, W.F. van Gunsteren, and J. Hermans, *Interaction Models for Water in Relation to Protein Hydration*, in *Intermolecular Forces*, B. Pullman, Editor. 1981, Springer Netherlands. p. 331-342.

56. Berendsen, H.J.C., J.P.M. Postma, W.F. van Gunsteren, A. DiNola, and J.R. Haak, *Molecular dynamics with coupling to an external bath.* The Journal of Chemical Physics, 1984. **81**(8): p. 3684-3690.
57. Hess, B., H. Bekker, H.J.C. Berendsen, and J.G.E.M. Fraaije, *LINCS: A linear constraint solver for molecular simulations.* Journal of Computational Chemistry, 1997. **18**(12): p. 1463-1472.
58. Essmann, U., L. Perera, M.L. Berkowitz, T. Darden, H. Lee, and L.G. Pedersen, *A smooth particle mesh Ewald method.* The Journal of Chemical Physics, 1995. **103**(19): p. 8577-8593.
59. Wolf, M.G., M. Hoefling, C. Aponte-Santamaría, H. Grubmüller, and G. Groenhof, *g_membed: Efficient insertion of a membrane protein into an equilibrated lipid bilayer with minimal perturbation.* Journal of Computational Chemistry, 2010. **31**(11): p. 2169-2174.
60. Pronk, S., S. Pall, R. Schulz, P. Larsson, P. Bjelkmar, R. Apostolov, M.R. Shirts, J.C. Smith, P.M. Kasson, D. van der Spoel, B. Hess, and E. Lindahl, *GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit.* Bioinformatics, 2013. **29**(7): p. 845-54.
61. Humphrey, W., A. Dalke, and K. Schulten, *VMD: visual molecular dynamics.* J Mol Graph, 1996. **14**(1): p. 33-8, 27-8.
62. Chovancova, E., A. Pavelka, P. Benes, O. Strnad, J. Brezovsky, B. Kozlikova, A. Gora, V. Sustr, M. Klvana, P. Medek, L. Biedermannova, J. Sochor, and J. Damborsky, *CAVER 3.0: A Tool for the Analysis of Transport Pathways in Dynamic Protein Structures.* PLOS Computational Biology, 2012. **8**(10): p. e1002708.
63. Cojocaru, V., P.J. Winn, and R.C. Wade, *The ins and outs of cytochrome P450s.* Biochimica et Biophysica Acta (BBA) - General Subjects, 2007. **1770**(3): p. 390-401.
64. Baylon, J.L., I.L. Lenov, S.G. Sligar, and E. Tajkhorshid, *Characterizing the Membrane-Bound State of Cytochrome P450 3A4: Structure, Depth of Insertion, and Orientation.* Journal of the American Chemical Society, 2013. **135**(23): p. 8542-8551.
65. De Lemos-Chiarandini, C., A.B. Frey, D.D. Sabatini, and G. Kreibich, *Determination of the membrane topology of the phenobarbital-inducible rat liver cytochrome P-450 isoenzyme PB-4 using site-specific antibodies.* J Cell Biol, 1987. **104**(2): p. 209-19.
66. Rice, P., I. Longden, and A. Bleasby, *EMBOSS: the European Molecular Biology Open Software Suite.* Trends Genet, 2000. **16**(6): p. 276-7.

Chapter 6 - Conclusions and Future Perspectives

This research sought to apply the techniques of molecular dynamics and computational analyses to characterise protein-membrane interactions on an atomistic level. Membrane proteins typically regulate cellular function and behaviour; thus, alteration in their structure will more often than not negatively impact a cascade of activities, resulting in disease. To make these proteins effective drug targets in the treatment or cure of such diseases, their behaviour must be well understood in terms of not only structure and dynamics, but also their interactions with the cellular membrane interface. Molecular dynamics presents itself as a valuable tool in the study and observation of protein-membrane systems, allowing detailed insight into specific and non-specific interactions.

In the field of molecular dynamics, proteins have been carefully parameterised whereas less importance has been placed on the cell membrane and its lipid components. Lipid composition is intrinsically linked to both the location and function of the cell membrane; and so, to gain a full understanding of their role in protein-membrane interactions, complex systems reflective of the physiological composition of lipid bilayers must also be studied. Chapter 2 outlines the creation of parameters for novel lipids of varying headgroups and fatty acid tails for the GROMOS 54a7 force field. These parameters were then validated through the construction of single, dual, and multi-lipid bilayers leading to the construction of brain lipid bilayers featuring biologically relevant composition. This opens the door for the use of these novel lipids to build a wide array of lipid bilayers with compositions reflective of the environment being modelled. Depending on the bilayer, further parameterisation may need to be undertaken to introduce additional lipid types, such as phosphatidylglycerol and cardiolipin. Ideally, the current brain lipid bilayer simulations should be extended towards a 500 ns timescale, so as to observe membrane features such as lipid diffusion and clustering.

The parameterisation of these lipids and the construction of the brain lipid bilayer allowed for a more accurate representation of protein-membrane systems. Protein-membrane interactions can be broken down into two stages: membrane association and membrane binding. Membrane binding typically features specific short-range interactions and often sees embedment of the protein into the lipid bilayer. On the other hand, membrane association describes the approach of the protein towards the membrane in an orientation necessary for its function, via non-specific long-range interactions. These interactions are often electrostatic, commonly involving positively charged basic residues and negatively charged anionic lipids.

Using this theoretical knowledge, a rotational sampling method was developed to rapidly screen protein orientations with respect to the membrane so as to identify the most favourable protein-membrane interface. Furthermore, atomistic representations of all system components allowed for per-residue decomposition of the protein-membrane interaction energies facilitating the identification of key individual residues and lipids important to the membrane association process. In tandem with the construction of lipid bilayers representative of a particular membrane protein's environment, this method allows for the prediction of the functional orientation of the protein with respect to the membrane, ranking the individual contribution of residues in achieving this orientation. This provides new and essential insight into potential drug targets for the disruption of the protein-membrane interface fundamental to the protein's function. Additionally, this allows for the initiation of molecular dynamics simulations of protein-membrane systems from an advanced stage in terms of their association, potentially saving both time and resources.

Whereas much of the work has focused on the behaviour of the protein, it would be useful and interesting to observe the significance of the role other elements play within the system. The effect of the variation of lipid composition, particularly anionic lipids, on the most favourable orientation would be a fundamental step in solidifying the nature of protein-membrane association. This screening method also outputs potential energies on the basis of interatomic distances, failing to take into account the effect of the surrounding solvent and ions. Whilst PMF is a better estimation of membrane affinity, the substantial difference in calculation timescale becomes a hindrance. It would be intriguing to determine the feasibility of introducing a method incorporating aspects of both techniques, in an effort to produce a more realistic evaluation of protein-membrane interactions.

Molecular dynamics simulations were done of PI3Kα, a signalling protein regulating cellular functions such as proliferation and survival, via the phosphorylation of PIP_2 to PIP_3 occurring at the membrane interface. This study found that an oncogenic mutant of PI3Kα, H1047R, increases activity through the improvement of its membrane affinity. It achieves this through conformational changes initiated due to the nature of the arginine side chain, resulting in the presentation of more basic residues along the membrane-binding interface. This allows for faster membrane association, earlier substrate recognition, and the achievement of a catalytically competent orientation independent of other activating and recruitment proteins. A theory was put forward that the wild type PI3Kα's achievement of this orientation is potentially dependent on the *Ras* activating protein. Eventually, it would be ideal to carry out molecular dynamics simulations of a *Ras*-bound PI3Kα in the presence of a lipid bilayer. As such a crystal structure has not yet been resolved, this system would need to be modelled analogously to the PI3Kγ structure.

A comprehensive study into the nature of PI3Kα's C terminal tail was also carried out, using predictive modelling to determine the effect of its position relative to the substrate binding pocket. When modelled pointing outwards, away from the binding pocket, PI3Kα was able to achieve a catalytically competent orientation whereby the tail was embedded into the lipid bilayer. However, when modelled pointing inwards, into the binding pocket, the wild type PI3Kα was unable to achieve this orientation. Interestingly, the H1047R mutation initiated a shift of the tail away from the binding pocket, allowing the catalytically competent orientation to be achieved. This finding adds another facet to the mechanism by which this mutation increases PI3Kα activity. Experimental procedures must be carried out to validate the predictions made by these molecular dynamics simulations; however, this research has potentially led to a more fundamental understanding of PI3Kα's membrane interaction and how that leads to its function.

Further molecular dynamics studies featuring changes in membrane composition and the nature of the PIP_2 substrate are already being carried out in an attempt to provide further context to the function of the many membrane binding domains found in PI3Kα. These new simulations look at the effect of the naturally variable length and saturation of the PIP_2 fatty acid tails on PI3Kα's ability to recognise the substrate. Extension of the current simulations would also be helpful in determining whether the observed PIP_2 substrate would continue its path into the binding pocket, and also whether the E545K system would eventually demonstrate membrane interaction. Further testing of the multi-faceted inhibition of PI3Kα's activity and how each mutation overcomes it could be carried out. In addition to Ras-bound PI3Kα, other potential examples of this include: the E545K mutation introduced to a PI3Kα system with the nSH2 domain bound to determine the mechanism of decoupling; the H1047L mutation introduced to a PI3Kα system to observe the parallels with H1047R's conformational changes leading to an improved membrane affinity.

Another protein-membrane system was studied, featuring a cytochrome P450 enzyme embedded into a lipid bilayer. The stable nature of this arrangement during the molecular dynamics simulations of this system was comforting, and allowed for a more confident analysis of CYP51's behaviour. While simulations are unable to demonstrate catalytic activity, they can potentially predict catalytic pathways highlighting interactions between the enzyme and the substrate. In the case of the CYP51-membrane system, this was manifested in the lanosterol substrate interacting with previously identified substrate recognition sites en route to a repositioning alongside the haem cofactor present in the protein.

The volumes of the ingress and egress tunnels situated on the membrane and solvent interfaces of CYP51 were also characterised, leading to a theory of their regulation by the presence or absence of a substrate within the binding cavity. It was found that when the binding cavity was occupied, the membrane access tunnel remained open, whereas the solvent access tunnel remained closed. However, when the binding cavity was unoccupied, the opposite was true. The nature of the substrate may also have an effect on its preferred egress pathway. Physiologically, CYP51 works alongside other membrane proteins allowing for substrates to be passed laterally to adjacent enzymes. In the future, it would be interesting to be able to simulate these proteins embedded in the membrane in more complex systems with a range of substrates and their precursors. This would allow for a more complete characterisation of their effect on the nature of the binding cavity, as well as shed more light onto the functional behaviour of CYP51.

In summary, molecular dynamics is a uniquely useful tool when it comes to the characterisation of protein-membrane interactions, doing so in a dynamic nature with an atomistic level of detail. Complex systems can be built featuring the protein in an environment reflective of its physiological state, leading to accurate predictions of its behaviour and the identification of components essential to its function. When applied to both a transiently bound peripheral membrane protein (PI3Kα) and an integral transmembrane protein (CYP51), novel insight was gained into their function. With as many as half of medicinal drugs targeting membrane proteins, it is of utmost importance that the function of these proteins is well understood, primarily involving the characterisation of their interactions with the cellular membrane which is often the source of their substrate. This research will hopefully introduce and establish novel methods in the field for the study of protein-membrane interactions, particularly in their application to an understanding of PI3Kα which was the main focus of this thesis.

Appendix A - Alterations to Force Field Files

New bond length for carbon-carbon bonds in 5-membered and 6-membered hydrocarbon rings.

```
#define gb_53           0.150       0.750E+07
; C,CHn - C,CHn         5-6 mem rings     800
```

New bond angle for carbon-carbon single bonds in 5-membered and 6-membered hydrocarbon rings.

```
#define ga_55           108.0       535.00
; CHn - CHn - CHn 5-6 mem rings
```

New bond angle for carbon-carbon double bonds in 6-membered hydrocarbon rings.

```
#define ga_56           124.0       615.00
; CHn - CH = CH - CHn          6 mem ring
```

New improper dihedral angle for 5-membered planar cholesterol ring.

```
#define gi_1a           0.0         33.5
; planar group          5 mem ring
```

All GROMOS54A7 parameter files for lipids used in this research are available upon request.

Appendix B - Rotational Interaction Energy Screening Script

```
#!/usr/bin/python

# script to iterate over different orientations and distances of the
protein
# the membrane and call gromacs for a short MD to calculate the
interaction energy

import sys, os
import getopt
import shutil
from string import replace
from subprocess import Popen, check_call, PIPE, CalledProcessError

def usage():
    print("""usage: {0} -h
       {0} <pitch> <roll> <yaw> -x <centre_x> -y <centre_y> -z
<centre_z>
           [ -s <exe_suffix> ]""".format(progname))
    return

exitcode = { 'ok' : 0, 'warning' : 1, 'critical' : 2, 'unknown' : 3 }

progname = os.path.basename(__file__)

try:
    opts, args = getopt.gnu_getopt(sys.argv[1:], "hs:x:y:z:", ["help",
"suffix="])
except getopt.GetoptError as err:
    # print help information and exit:
    print str(err) # will print something like "option -a not
recognized"
    usage()
    sys.exit(exitcode['critical'])

centre = [None, None, None]
suffix = ""
try:
    for opt, arg in opts:
        if opt in ("-h", "--help"):
            usage()
            sys.exit(exitcode['ok'])
        elif opt in ("-s", "--suffix"):
            suffix = arg
        elif opt == "-x":
            centre[0] = float(arg)
        elif opt == "-y":
            centre[1] = float(arg)
        elif opt == "-z":
            centre[2] = float(arg)
        else:
            assert False, "unhandled option"
except ValueError:
```

```
    sys.stderr.write("""Centre coordinates must be floating-point
numbers (or convertible to such)
""")
    sys.exit(exitcode['critical'])

if len(args) != 3 or len(centre) != 3:
    usage()
    sys.exit(exitcode['critical'])

# Angle of rotation from starting position (degrees)
try:
    pitch = int(args[0]) # Angle of rotation about X
    roll = int(args[1])  # Angle of rotation about Y
    yaw = int(args[2])   # Angle of rotation about Z
except ValueError:
    sys.stderr.write("""Angles of rotation must be integers (or
convertible to integers)
""")
    sys.exit(exitcode['critical'])

[pitch, roll, yaw] = [angle % 360 for angle in [pitch, roll, yaw]]
rotstring = "_{0:03d}_{1:03d}_{2:03d}".format(pitch, roll, yaw)

# Get the location of GROMACS
try:
    gromacsbindir = os.path.join(os.environ['EBROOTGROMACS'], 'bin')
except KeyError:
    sys.stderr.write("""The environment variable EBROOTGROMACS is not
defined.
Please load a GROMACS module, then try again.
""")
    sys.exit(exitcode['critical'])

# Make sure the suffix is correct
editconf_exe = os.path.join(gromacsbindir,
"editconf{0}".format(suffix))
if not os.path.isfile(editconf_exe):
    suggested_suffix = None
    for gromacsprogram in os.listdir(gromacsbindir):
        # Assume only one "editconf"
        if gromacsprogram.startswith("editconf"):
            suggested_suffix = replace(gromacsprogram, "editconf", "",
1)
            break
    if suggested_suffix is None:
        sys.stderr.write("Error: editconf is missing -- something is
badly wrong!\n")
    else:
        sys.stderr.write("{0}: no such file -- perhaps suffix should be
{1}?\n".format(editconf_exe, suggested_suffix))
    sys.exit(exitcode['critical'])

# protein name?
mut = 'pi3k'
# output directory for putting all the files that we generate into
outdir = mut+'_rot-ene'
try:
    os.mkdir(outdir)
```

```
except:
    pass

# centre of box
# TODO: Consider whether to supply as an argument
# protein-membrane coordinates (solvated)
#protmb = mut+'_mb.gro'

# open a log file to see what's happening
log_filename = rotstring + ".log"
log = open(log_filename,'w')

# -- first cut out protein and membrane coordinates
#f = open(protmb,'r')
#l = f.readlines()
# membrane coordinates only (in big box)
mb = 'mb.gro'
#fm = open(mb,'w')
#init = 'Membrane in Box;\n 22788\n'
#fm.write(init)
#for i in range(12869,35657):
#  fm.write(l[i])
end = '  13.27970  26.50000  12.44970\n'
#fm.write(end)
#fm.close()
# protein coordinates only
prot = mut+'.gro'
#fp = open(prot,'w')
#init = 'Protein to Rotate;\n 12867\n'
#fp.write(init)
#for i in range(2,12869):
#  fp.write(l[i])
#fp.write(end)
#fp.close()

stolen_data = []
# new name for rotated protein coordinates according to current
iteration
rotprot = os.path.join(outdir, replace(prot,'.gro',rotstring+'.gro'))
# rotate and re-centre protein
# note that all arguments to check_call must be strings, not ints or
floats
rotate_cmd = [ "{0}{1}".format(os.path.join(gromacsbindir, 'editconf'),
suffix) ]
rotate_cmd.extend(["-f", prot])
rotate_cmd.extend(["-rotate", str(pitch), str(roll), str(yaw)])
rotate_cmd.extend(["-center"])
rotate_cmd.extend([str(coord) for coord in centre])
rotate_cmd.extend(["-o", rotprot ])
log.write("Running rotate command: {0}\n".format(rotate_cmd))
check_call(rotate_cmd)

# new name for rotated protein coordinates + membrane
rotprotmb = os.path.join(outdir,
replace(prot,'.gro',rotstring+'mb.gro'))
# read in rotated protein coordinates
fi = open(rotprot,'r')
li = fi.readlines()
```

```
fi.close()
fm = open(mb,'r')
lm = fm.readlines()
fm.close()
# open output file for combined protein + membrane coordinates
fo = open(rotprotmb,'w')
# write header with total number of atoms
init = 'Rotated protein + membrane;\n35480\n'
fo.write(init)
# write protein coordinates (without box)
for i in range(2,len(li)-1):
    fo.write(li[i])
# write membrane coordinates (still known from initial extraction of
membrane coords)
for i in range(2,len(lm)-1):
    fo.write(lm[i])
# write box size (still known from initial extraction of coords)
fo.write(end)
fo.close()

# output file for mindist
mdfile =
os.path.join(outdir,"{0}_mindist{1}.xvg".format(mut,rotstring))
# calculate mininum distance between protein and membrane
# it asks for groups:
#Select a group: 1
#Selected 1: 'Protein'
#Select a group: 12
#Selected 32: 'Other'
# have automated group selection:
while True:
    dist_cmd = [ "{0}{1}".format(os.path.join(gromacsbindir,
'g_mindist'), suffix) ]
    dist_cmd.extend(['-f', rotprotmb])
    dist_cmd.extend(['-n', "{0}.ndx".format(mut)])
    #dist_cmd.extend(['-s', "{0}_inter.tpr".format(mut)])
    dist_cmd.extend(['-od', mdfile])

    log.write("Running distance command: {0}\n".format(dist_cmd))
    proc = Popen(dist_cmd, stdin=PIPE)
    proc.communicate('1\n12\n')
    retcode = proc.wait()
    if retcode != 0:
        raise CalledProcessError(returncode=retcode, cmd=dist_cmd)

# extract mindist from file
    fd = open(mdfile,'r')
    mindist = float(fd.readlines()[19].split()[1])
    fd.close()
    log.write('Mindist = '+str(mindist)+'\n\n')

    # Let's pretend for now that we want the edge of the protein to be
exactly 0.5 nm from the membrane surface
    # this should move it away if it overlaps, away if it's too close,
and closer if it's too far away
    if mindist > 0.05 : transdist = round(mindist - 0.05,3)
    # NOTE: may not round as expected so may not quite shift enough
    else : transdist = -1*round(0.05 - mindist,3)
```

```
    log.write('Transdist = '+str(transdist)+'\n\n')

    # translate to current working distance (assuming we only translate
in z)
    transprot = "{0}_translated.gro".format(rotstring)
    translate_cmd = [ "{0}{1}".format(os.path.join(gromacsbindir,
'editconf'), suffix) ]
    translate_cmd.extend(["-f", rotprot])
    translate_cmd.extend(["-translate", "0", "0", str(transdist)])
    translate_cmd.extend(["-o", transprot ])
    log.write("Running translate command: {0}\n".format(translate_cmd))
    check_call(translate_cmd)

    # replace rotated protein coordinates with translated protein
coordinates
    os.rename(transprot, rotprot)

    # combine coordinates (overwriting previous)
    # read in translated protein coordinates
    fi = open(rotprot,'r')
    li = fi.readlines()
    fi.close()
    fm = open(mb,'r')
    lm = fm.readlines()
    fm.close()
    # open output file for combined protein + membrane coordinates
(overwriting previous)
    fo = open(rotprotmb,'w')
    # write header with total number of atoms
    init = 'Rotated protein + membrane;\n35480\n'
    fo.write(init)
    # write protein coordinates
    for i in range(2,len(li)-1):
        fo.write(li[i])
    # write membrane coordinates (still known from initial extraction
of membrane coords)
    for i in range(2,len(lm)-1):
        fo.write(lm[i])
    # write box size (still known from initial extraction of coords)
    fo.write(end)
    fo.close()

    # re-calculate protein-membrane distance
    log.write("Running distance command: {0}\n".format(dist_cmd))
    proc = Popen(dist_cmd, stdin=PIPE)
    proc.communicate('1\n12\n')
    retcode = proc.wait()
    if retcode != 0:
        raise CalledProcessError(returncode=retcode, cmd=dist_cmd)
    #check_call(dist_cmd)

    # extract new mindist from file
    fd = open(mdfile,'r')
    mindist = float(fd.readlines()[19].split()[1])
    fd.close()
    log.write('New mindist = '+str(mindist)+'\n\n')

    if mindist > 0.05 or -0.01 < transdist < 0.01 : break
```

```
shutil.copy("{0}_beta.top".format(mut),
"{0}_beta{1}.top".format(mut,rotstring))

# Solvate each rotated protein + membrane system
solvate_cmd = [ "{0}{1}".format(os.path.join(gromacsbindir, "genbox"),
suffix) ]
solvate_cmd.extend([ '-cp', rotprotmb ])
solvate_cmd.extend([ '-cs', 'spc216.gro' ])
solvate_cmd.extend([ '-p', "{0}_beta{1}.top".format(mut,rotstring) ])
solvate_cmd.extend([ '-o', "{0}_solv{1}.gro".format(mut,rotstring) ])
log.write("Running solvate command: {0}\n".format(solvate_cmd))
check_call(solvate_cmd)

# Energy minimise solvated system
min_grompp_cmd = [ "{0}{1}".format(os.path.join(gromacsbindir,
'grompp'), suffix) ]
min_grompp_cmd.extend([ '-f', "memb_minim.mdp" ])
min_grompp_cmd.extend([ '-c', "{0}_solv{1}.gro".format(mut,rotstring)
])
min_grompp_cmd.extend([ '-p', "{0}_beta{1}.top".format(mut,rotstring)
])
min_grompp_cmd.extend([ '-o', "{0}_em{1}.tpr".format(mut,rotstring) ])

log.write("Preprocessing energy minimisation:
{0}\n".format(min_grompp_cmd))
check_call(min_grompp_cmd)

# Run energy minimisation
min_mdrun_cmd = [ "srun" ]
min_mdrun_cmd.extend([ "{0}{1}".format(os.path.join(gromacsbindir,
'mdrun'), suffix) ])
min_mdrun_cmd.extend([ '-deffnm', "{0}_em{1}".format(mut,rotstring) ])
log.write("Running energy minimisation: {0}\n".format(min_mdrun_cmd))
check_call(min_mdrun_cmd)

# prepare for 25-step MD with necessary energy groups included
md_grompp_cmd = [ "{0}{1}".format(os.path.join(gromacsbindir,
'grompp'), suffix) ]
md_grompp_cmd.extend([ '-f', "final.mdp" ])
md_grompp_cmd.extend([ '-c', "{0}_em{1}.gro".format(mut,rotstring) ])
md_grompp_cmd.extend([ '-p', "{0}_beta{1}.top".format(mut,rotstring) ])
md_grompp_cmd.extend([ '-o', "{0}{1}_energy".format(mut,rotstring) ])
log.write("Preprocessing MD simulation: {0}".format(md_grompp_cmd))
check_call(md_grompp_cmd)

# run 25-step MD
md_mdrun_cmd = [ "srun" ]
md_mdrun_cmd.extend([ "{0}{1}".format(os.path.join(gromacsbindir,
'mdrun'), suffix) ])
md_mdrun_cmd.extend([ '-deffnm', "{0}{1}_energy".format(mut,rotstring)
])
log.write("Running MD simulation: {0}".format(md_mdrun_cmd))
check_call(md_mdrun_cmd)

# calculate interaction energy (vdw as well as electrostatic)
# 47  Coul-SR:Protein-Other
# 48  LJ-SR:Protein-Other
```

```
# 49  LJ-LR:Protein-Other
md_energy_cmd = [ "{0}{1}".format(os.path.join(gromacsbindir,
"g_energy"), suffix) ]
md_energy_cmd.extend([ '-f', "{0}{1}_energy.edr".format(mut, rotstring)
])
md_energy_cmd.extend([ '-s', "{0}{1}_energy.tpr".format(mut, rotstring)
])
md_energy_cmd.extend([ '-o', "{0}{1}_energy.xvg".format(mut, rotstring)
])
log.write("Computing energies: {0}".format(md_energy_cmd))
proc = Popen(md_energy_cmd, stdin=PIPE)
proc.communicate('47\n48\n49\n')
retcode = proc.wait()
if retcode != 0:
    raise CalledProcessError(returncode=retcode, cmd=md_energy_cmd)

#select_cmd = [ "{0}{1}".format(os.path.join(gromacsbindir,
"g_select"), suffix) ]
#select_cmd.extend([ '-f', "{0}{1}_energy.gro".format(mut, rotstring)
])
#select_cmd.extend([ '-s', "{0}{1}_energy.tpr".format(mut, rotstring)
])
#select_cmd.extend([ '-select', "'\"reslect\" residues within 1.4 of
group other'" ])
#select_cmd.extend([ '-oi', "{0}{1}.dat".format(mut, rotstring) ])
#log.write("Selecting residues: {0}".format(select_cmd))
#check_call(select_cmd)

with open("{0}{1}_energy.xvg".format(mut, rotstring),'r') as fh:
    important_line = fh.readlines()[-1].split()
stolen_data.append('%d,%d,%d\t%s\t%s\t%s'%(pitch,roll,yaw,important_lin
e[1],important_line[2],important_line[3]))

with open('outputfile_'+str(rotstring)+'.txt','w') as fh:
    for i in stolen_data: fh.write(i+'\n')

# Throw away unneeded files
to_delete = [f for f in os.listdir(".") if
"{0}_energy.trr".format(rotstring) in f]
for file in to_delete:
    os.remove(file)

to_delete = [f for f in os.listdir(".") if "_em{0}".format(rotstring)
in f]
for file in to_delete:
    os.remove(file)

to_delete = [f for f in os.listdir(".") if "_solv{0}".format(rotstring)
in f]
for file in to_delete:
    os.remove(file)

to_delete = [f for f in os.listdir(".") if "mdout" in f]
for file in to_delete:
    os.remove(file)

sys.exit(exitcode['ok'])
```

Appendix C - Rotational Interaction Energy Screening Manuscript

Computational prediction of protein-membrane interactions

William A. Irvine[1,2], Jack U. Flanagan[3,5] and Jane R Allison[1,2,4,5]

1. Centre for Theoretical Chemistry and Physics, Massey University Auckland, Private Bag 102904, 0632 Auckland, New Zealand
2. Institute of Natural and Mathematical Sciences, Massey University Auckland, Private Bag 102904, 0632 Auckland, New Zealand
3. Auckland Cancer Society Research Centre, University of Auckland, Private Bag 92019, Auckland, New Zealand
4. Biomolecular Interaction Centre, University of Canterbury, Private Bag 4800, Christchurch 8140, New Zealand
5. Maurice Wilkins Centre for Molecular Biodiscovery, University of Auckland, Private Bag 92019, Auckland, New Zealand

Introduction

In carrying out their physiological roles, proteins must interact with other cellular components. A large proportion of proteins interact with the cell membrane, with most cellular functions occurring in or around membranes, and 60% of drug targets are located at the cell surface[1]. As such, it is important to understand how the characteristics of both proteins and membranes affect their interaction and the impact of any changes in these characteristics.

Proteins that interact with the cell membrane may be integral or peripheral. The latter class are soluble and interact reversibly with the membrane surface. While some peripheral membrane proteins utilise post-translational modifications to govern their interaction with the membrane, others interact with the membrane directly, most commonly via a membrane binding or targeting domain[2], such as the C1 (conserved region 1)[3,4], C2 (conserved region 2)[4,5], and PH (pleckstrin homology) domains[6]. However not all proteins that contain these domains will have the same membrane binding mode[7].

Protein-membrane interaction requires first, membrane association, and second, membrane binding. A major contribution to the approach of peripheral membrane proteins to the membrane surface is non-specific electrostatic interactions[8]. These typically involve anionic lipid species and polycationic protein domains, but the opposite can also occur[9,10]. These long-range interactions align the protein into a binding-competent orientation relative to the lipid bilayer [11-13].

The subsequent membrane binding typically involves specific short-range interactions[14]. These may involve van der Waals interactions from both the close association of the protein with the membrane surface and from the embedding of hydrophobic moieties – proteinaceous or post-translationally added – into the membrane core[15,16], as well as electrostatic interactions

between specific amino acids or lipid recognition domains and particular lipid head groups such as phosphatidylinositol [4,17-20]. It is these specific short-range electrostatic interactions that govern the localisation of peripheral membrane proteins at a particular site on the membrane surface [11, 12]. Additionally, there is increasing evidence of reciprocal influence between membrane lipid composition and the local structure, dynamics and activity of membrane proteins[9,21-29].

The association and binding of peripheral membrane proteins to the membrane can be studied computationally using a number of different approaches, which can be loosely organised into three categories: molecular dynamics (MD) simulation (using atomic-level or coarse-grained representations); evaluation of the Coulombic electrostatic interaction energy using approximate methods; and evaluation of the preference for a hydrophobic versus hydrophilic environment. The latter two methods typically involve approximate descriptions, such as determining the electrostatic energy of a protein at the charged planar membrane surface based on the finite difference Poisson-Boltzmann method[18,30-33], determining the transfer energy of moving a protein from water to a hydrophobic planar slab[34], or weighting the solvent-accessible surface area of amino acids[35]. While electrostatic interactions are an important aspect of the membrane association of most peripheral membranes, methods involving hydrophobic/hydrophilic preferences are better suited to integral membrane proteins. Neither method is well-suited to evaluating the effect of specific interactions. Only MD simulations can incorporate all the different contributions to protein-membrane association, and additionally allow the entire process to be followed. MD simulations are particularly useful for testing hypotheses regarding the influence of lipid composition or changes to the protein such as mutation or truncation on membrane association and binding. However the size of a protein-membrane system and the time-scales involved make it unfeasible to use MD simulations in a predictive or even statistically meaningful manner.

We describe here the rotational interaction energy profiling (RIEP) method for evaluating the optimal orientation of a protein for membrane association and binding that uses the detailed molecular description and associated energy functions used in MD simulations, without the need to run lengthy simulations. Our focus is on the forces that guide peripheral membrane proteins towards the membrane such that they approach the membrane in a binding-competent orientation. The ability to use lipid bilayers of varying complexity allows the effect of lipid composition on the protein-membrane interaction to be tested. Key residues involved in the protein-membrane interaction can also be identified. We utilise explicit solvent to correctly capture solvation effects, and so that the interaction potential energies are directly comparable to those sampled during molecular dynamics simulations, allowing the optimal orientation to be used to efficiently seed simulations to follow the process of membrane association, as well as identify the most promising protein variants and membrane compositions to study experimentally. We validate our method on two cytosolic proteins that associate with the cell membrane, phosphatase and tensin homolog (PTEN) and the p110α/p85 phosphatidyl-inositol kinase (PI3Kα) complex.

Methods

Rotational interaction energy profiling (RIEP)

The Coulombic and van der Waals contribution to the protein-membrane potential energy over different orientations of the protein with respect to the membrane surface were computed by using a python wrapper script to call GROMACS (v4.6.7) tools[36]. The protein uniformly sampled different orientations (pitch-roll-yaw) with respect to the membrane at a granularity of 30°. Here, a single minimum protein-bilayer distance of 0.5 nm was used to ensure that a large proportion of the protein was within the cut-off distance so that the interactions are computed exactly, while still probing the forces that govern the process of membrane association. At each orientation, the system was solvated with explicit water, energy minimised for 1000 steps of steepest descent, and subjected to a short (25-step) molecular dynamics (MD) simulation to allow relaxation of the water molecules prior to evaluating the van der Waals and Coulombic interaction potential energies between the protein and the membrane. The contributions of individual residues a particular rotation were determined by re-computing the Coulombic and van der Waals interaction energy for each residue within the relevant cut-off distance. RIEP calculations were run in parallel across multiple cores.

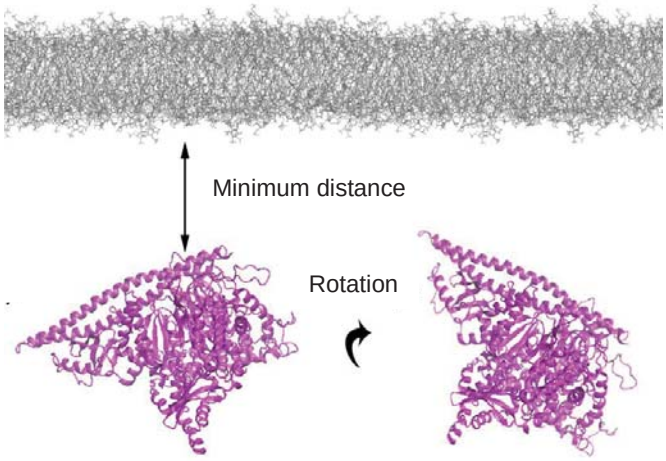


Figure 1. Rotational interaction energy profiling (RIEP) setup and procedure. The Coulombic and van der Waals contributions to the protein-membrane interaction potential energy are calculated for each orientation of the protein (magenta) with respect to the membrane (grey) at a specified minimum protein-membrane distance.

System

Proteins and lipids were modelled using the GROMOS 54A7 force field[37], with parameters for the lipids developed in-house[38] with the exception of dipalmotoyl-sn-glycero-3-phosphocholine (DPPC)[39], and the single point charge (SPC) water model[40]. The neuronal cell membrane contained 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine (POPC), 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphoethanolamine (POPE), 1-palmitoyl-2-oleoylphosphatidylserine (POPS), cholesterol (CHOL), sphingomyelin (SGML) and phosphatidylinositol 4,5-bisphosphate (PIP_2) in the ratio 3:9:4:2:1:1 (240 lipid molecules per leaflet). The simulation box dimensions (and therefore also the xy dimensions of the lipid bilayer) were chosen such that there was a minimum of 2 nm between the protein and the sides of the simulation box in all orientations, and simulations were run under periodic boundary conditions.

Simulations

Each lipid bilayer was pre-equilibrated for 50 ns. During both the pre-equilibration and the interaction energy calculations, the following simulation procedure was followed: temperature was maintained using the Berendsen thermostat[41], with a coupling time constant of 0.1 ps, and pressure was maintained using weak coupling[41] with a time constant of 0.5 ps and an isothermal compressibility of 4.5×10^{-5} bar^{-1}. Bond lengths were constrained using the LINCS algorithm[42] with an order of 4 to constrain bond lengths and water bond angles, allowing for an integration time step of 2 fs. The Coulombic interactions were cut-off at 0.9 nm, with long-range Coulombic effects approximated using PME[43], and the van der Waals interactions were cut-off at 1.4 nm with no long-range corrections.

Analysis

The protein-membrane interaction energies were plotted using matplotlib[44]. Molecular structures were visualised using VMD[45].

Results and Discussion

RIEP was validated using two peripheral membrane proteins: phosphatase and tensin homologue (PTEN) and the p110α/p85 phosphatidyl-inositol kinase (PI3Kα) complex; and two different membranes: a pure DPPC bilayer, and a model cell membrane containing the type and proportion of lipids typically found in a mammalian brain cell membrane. Both of our test proteins contain C2 domains, but bind to the membrane surface in different ways and have different lipid targets – PS, in the case of PTEN, and PIP_2 in the case of PI3Kα – both of which are included in the cell membrane.

PTEN

PTEN is a protein and lipid phosphatase. It has tumour suppressor function via its dephosphorylation of phosphatidylinositol-3,4,5-trisphosphate (PIP_3) to phosphatidylinositol 4,5-bisphosphate (PIP_2), which reduces intracellular levels of PIP_3 and so antagonises the phosphoinositol-3-kinase/AKT signaling pathway, suppressing cell survival and proliferation[46,47]. PTEN comprises two major domains, an N-terminal phosphatase domain and a tensin-like C-terminal domain[48] that in turn contains a C2 domain and a tail domain. The C2 domain is implicated in its recruitment to phospholipid membranes[49] which is critical in order for PTEN to carry out its function.

Here, RIEP was carried out for PTEN positioned 0.5 nm from a representative cell membrane containing POPC, POPE, POPS, CHOL, SGML and PIP_2. At this distance, the Coulombic contribution to the protein-membrane interaction energy dominated the van der Waals contribution by a factor of 18 (Supporting Information Figure S1), thus the former is the driving force for attracting the protein to the membrane. The Coulombic interaction energies ranged from –453 to 3.52 $kJ.mol^{-1}$ (Figure 2A), with only 5 orientations exhibiting energies less than –401 $kJ.mol^{-1}$ (Supporting Information Figure S2). While there is some clustering of the orientations with less favourable interaction energies visible in Figure 2A, it should be noted that adjacent squares represent rotation of the protein by 30°, which is clearly sufficient to produce a marked change in the interaction energy. Additionally, orientations that appear far apart in the cube may correspond to essentially identical orientations of the protein with respect to the membrane, for instance via rotation of the protein in the xy (membrane surface) plane, which results in similar but not identical protein-membrane interactions as the different types of lipids are distributed randomly across the membrane surface.

The orientation of PTEN with respect to the membrane with the most favourable Coulombic interaction energy is shown in Figure 2B. This orientation is in agreement with that identified as optimal for membrane binding by Das et al. [49] in that the membrane binding domain (MBD) faces the membrane surface, along with the C2 domain and the phosphatase domain. Based on alanine mutation of cationic patches, the phosphatase domain is thought to be essential for electrostatically-driven membrane binding[48,50], while the C2 domain plays an indirect role, with its C-terminal tail promoting stability[51,52].

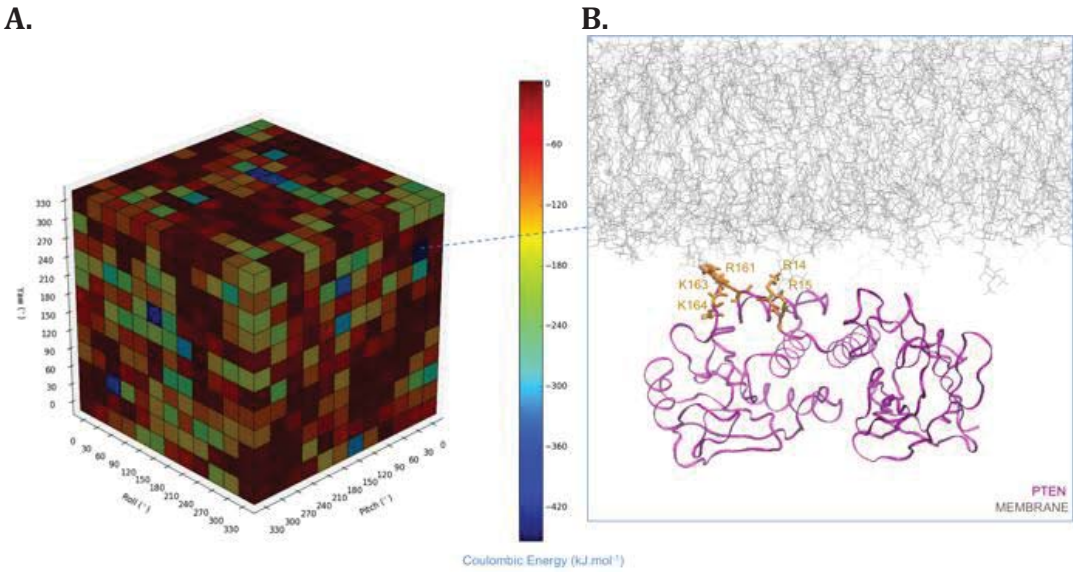


Figure 2. The most favourable orientation of PTEN with respect to the model cell membrane. A. Coulombic contribution to the protein-membrane interaction potential energy at a protein-membrane distance of 0.5 nm as a function of the orientation of the protein with respect to the membrane normal. B. the most favourable orientation (E^{pot}_{Coul} = –453 kJ.mol^{-1}) of PTEN (magenta ribbon) with respect to the membrane (grey lines), with the five residues that contribute most to the Coulombic protein-membrane interaction potential energy labelled (orange sticks).

A per-residue decomposition of the energy for this orientation revealed that 100% of the total Coulombic interaction energy could be assigned to just four basic amino acids (Supporting Information Figure S3). Three of these four residues correspond to residues found in the cationic patch in the phosphatase domain identified by Das et al.[49] as being critical to non-specific electrostatic affinity to a POPC/POPS membrane, and which were also found to be involved in membrane binding in other studies [53,54]. RIEP is therefore capable of identifying not only the correct binding orientation of the protein with respect to a representative cell membrane, but the key residues involved.

PI3Kα

Phosphatidylinositide 3-kinases (PI3K) are proteins that phosphorylate PIP_2 to PIP_3, triggering signalling pathways that stimulate cell growth, proliferation, and survival[55]. They have been implicated in cancer development[56] and the genes encoding these proteins identified as oncogenes[57].

The PI3Kα isoform consists of a catalytic subunit (p110α) and a regulatory subunit (p85α), which work in tandem to control protein activity[58]. Like PTEN, the p110α subunit contains a C2 domain, but the catalytic kinase domain is in the C-terminal region, and there are also an N-terminal adaptor binding domain (ABD), a *Ras* binding domain (RBD), and a helical domain[59]. While PI3Kα activation requires binding to a GTPase from the *Ras* family via the RBD – a source of induced activation of the PI3K enzyme[60], it is the kinase domain that mediates membrane binding through multiple contacts. These membrane

binding domains (MBDs) comprise residues 721-727, 863-873, 966-974, and feature a strong positive charge due to the relatively high percentage of basic residues [61]. However in order for the active site to be presented to the membrane such that phosphorylation of PIP_2 occurs, the C2, iSH2 and kinase domains must all face the membrane [62]. All three of these domains also contain a number of basic residues, which are situated along the membrane binding interface. Additionally, the hydrophobic WIF motif (residues 1057-1059) anchors PI3Kα to the membrane surface by inserting into the hydrophobic core of the membrane [63].

RIEP of PI3Kα positioned 0.5 nm from the representative cell membrane again revealed that Coulombic interactions make the largest contribution to the protein-membrane interaction energy at this distance (Supporting Information Figure S4), and so are primarily responsible for governing the approach of the protein to the membrane.

The orientation of PI3Kα with respect to the cell membrane with the most favourable Coulombic interaction energy has the membrane binding domains (MBDs) as well as the WIF motif facing the membrane surface (Figure 3), in agreement with both lipid binding assays [63] and identification of the membrane binding residues using hydrogen deuterium exchange mass spectrometry [62]. Importantly, RIEP is able to identify the correct membrane-binding orientation even when other domains in addition to the common membrane-binding C2 domain are involved.

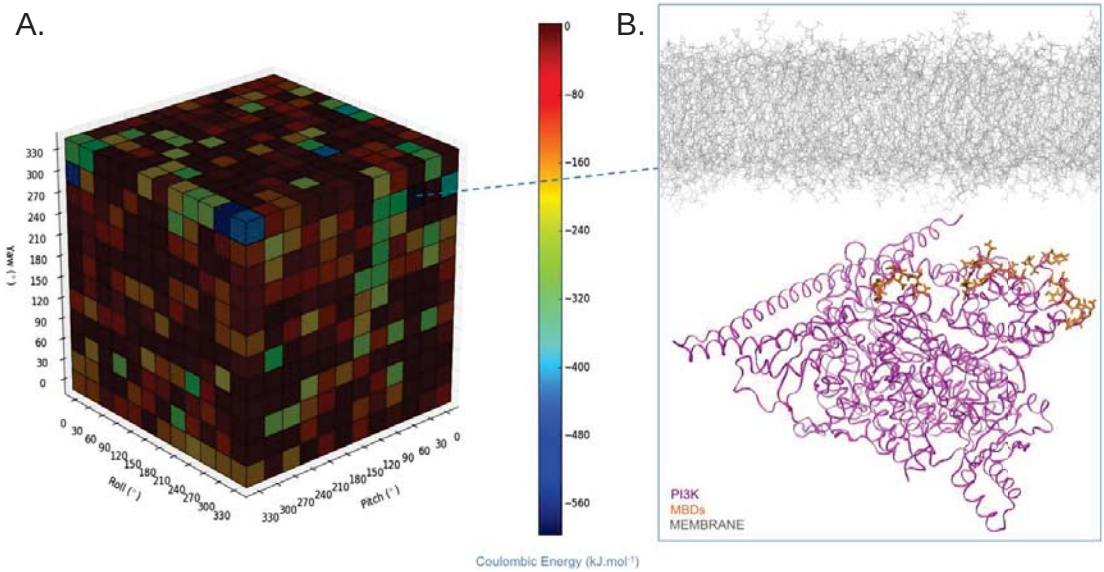


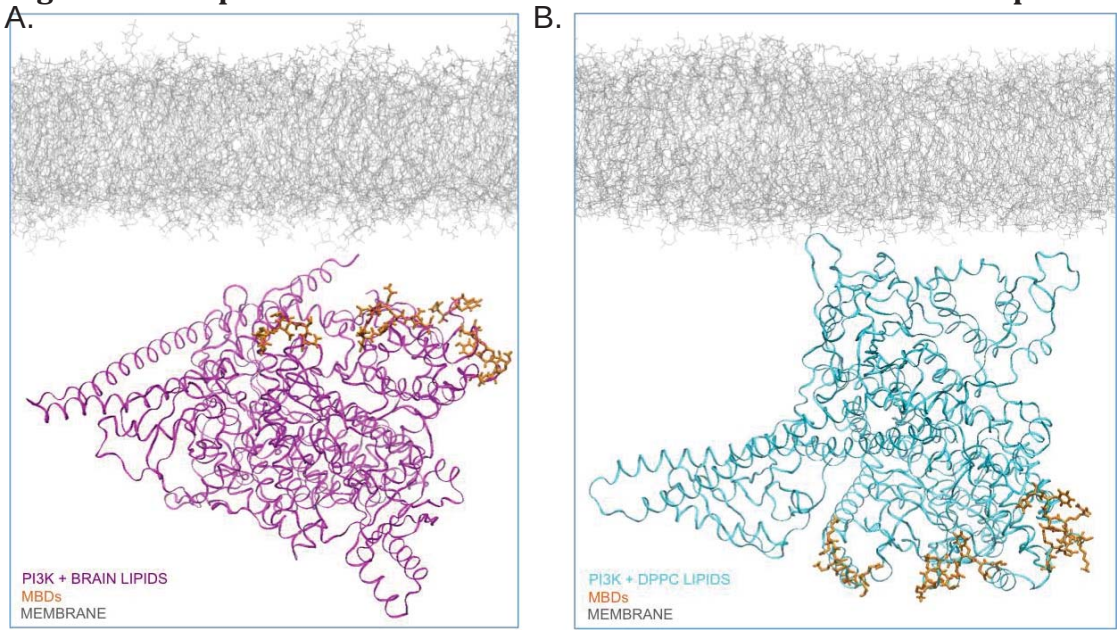
Figure 3. The most favourable orientation of PI3Kα with respect to the model cell membrane. A. Coulombic contribution to the protein-membrane interaction potential energy at a protein-membrane distance of 0.5 nm as a function of the orientation of the protein with respect to the membrane normal. B. The most favourable orientation (E^{pot}_{Coul} = –596 kJ.mol^{-1}) of PI3Kα (magenta ribbon) with respect to the membrane (grey lines), with the MBDs and WIF motif drawn in orange.

To test the sensitivity of the method to removal of the electrostatic interactions between negatively charged lipid head groups and basic residues in the protein, RIEP was carried out for PI3Kα positioned 0.5 nm from a pure DPPC bilayer. Unlike the representative cell membrane, where the inclusion of POPS creates an overall negative charge on the membrane surface, DPPC has a net charge of zero. As expected, the overall protein-membrane interaction potential energy in the presence of DPPC was greatly reduced compared to that in the presence of the model cell membrane due to substantial reduction in the magnitude of the Coulombic contribution to the protein-membrane interaction potential energy (Supporting Information Figures S4 – S7).

The RIEP of PI3Kα with the cell membrane clearly identified a set of six highly similar optimal membrane binding orientations (Figure S8), all of which present the MBDs to the membrane surface, in keeping with experimental observations [62,64]. In contrast, while the most favourable orientations detected by conducting RIEP with the DPPC bilayer are similar to one another, none matched the membrane binding orientation identified from RIEP with the cell membrane and predicted experimentally; rather, the MBDs and the C2 and iSH2 domains face away from the membrane surface (Figure 4, S9).

The six-fold decrease in the Coulombic protein-membrane interaction potential energy between the model cell membrane and the pure DPPC bilayer can be attributed largely to the removal of anionic lipids, highlighting the importance of membrane composition in capturing the non-specific electrostatic interactions that attract many membrane-associated proteins to the membrane surface.

Figure 4. Comparison of favourable orientations of PI3Kα with respect to the model cell membrane and the DPPC bilayer. A. The most favourable orientation of PI3Kα with respect to the model cell membrane. B. A representative example of the orientations identified as favourable when conducting RIEP with a pure DPPC bilayer. The residues drawn explicitly in orange correspond to the MBDs and the WIF motif.



Conclusion

The RIEP method rapidly evaluates the most favourable protein-membrane interactions at the level of both overall protein orientation with respect to the membrane, and key residues involved in membrane binding. The method was validated by reproduction of known membrane binding interactions and membrane composition preferences for the proteins PTEN and PI3Kα. The identification of the most likely binding orientation of a protein with respect to the membrane provides an efficient means of selecting initial coordinates for simulation of the membrane binding process, minimising the time spent sampling unproductive orientations of the protein with respect to the membrane. The key residues involved in membrane binding can also be identified and their contribution quantified, providing hypotheses about the drivers of the protein-membrane interaction that can be tested experimentally by site-directed mutagenesis. Together, these two facets also allow prediction and testing of the response of protein-membrane interactions to different membrane lipid compositions.

Acknowledgements

The authors acknowledge Benjamin Roberts for his help in developing the energy screening code, and the New Zealand eScience Infrastructure (NeSI) for high performance computing time. J.R.A. is supported by a Marsden Fast Start grant (MAU-13-039), a Rutherford Discovery Fellowship (15-MAU-001), and a Marsden grant (15-UOA-105), J.U.F. is supported by a Marsden grant (15-UOA-105), and W.A.I. is supported by a Massey University Doctoral Scholarship.

References

References

1. Overington JP, Al-Lazikani B, Hopkins AL: **How many drug targets are there?** *Nat Rev Drug Discov* (2006) **5**(12):993-996.

2. Lemmon MA: **Membrane recognition by phospholipid-binding domains.** *Nat Rev Mol Cell Biol* (2008) **9**(2):99-111.

3. Colon-Gonzalez F, Kazanietz MG: **C1 domains exposed: From diacylglycerol binding to protein-protein interactions.** *Biochim Biophys Acta* (2006) **1761**(827-837.

4. Cho W: **Membrane targeting by c1 and c2 domains.** *Journal of Biological Chemistry* (2001) **276**(35):32407-32410.

5. Cho W, Stahelin RV: **Membrane binding and subcellular targeting of c2 domains.** *Biochim Biophys Acta* (2006) **1761**(838-849.

6. Lemmon MA, Ferguson KM: **Signal-dependent membrane targeting by pleckstrin homology (ph) domains.** *Biochem J* (2000) **350**(1-18.

7. Zhang D, Aravind L: **Identification of novel families and classification of the c2 domain superfamily elucidate the origin and evolution of membrane targeting activities in eukaryotes.** *Gene* (2010) **469**(1-2):18-30.

8. Goldenberg NM, Steinberg BE: **Surface charge: A key determinant of protein localization and function.** *Cancer Res* (2010) **70**(4):1277-1280.

9. Escriba PV, Gonzalez-Ros JM, Goni FM, Kinnunen PK, Vigh L, Sanchez-Magraner L, Fernandez AM, Busquets X, Horvath I, Barcelo-Coblijn G: **Membranes: A meeting point for lipids, proteins and therapies.** *Journal of cellular and molecular medicine* (2008) **12**(3):829-875.

10. Mustonen P, Lehtonen J, Koiv A, Kinnunen PK: **Effects of sphingosine on peripheral membrane interactions: Comparison of adriamycin, cytochrome c, and phospholipase a2.** *Biochemistry* (1993) **32**(20):5373-5380.

11. Heimburg T, Marsh D: **Protein surface-distribution and protein-protein interactions in the binding of peripheral proteins to charged lipid membranes.** *Biophys J* (1995) **68**(2):536-546.

12. Johnson JE, Cornell RB: **Amphitropic proteins: Regulation by reversible membrane interactions (review).** *Molecular membrane biology* (1999) **16**(3):217-235.

13. Murray D, Arbuzova A, Hangyas-Mihalyne G, Gambhir A, Ben-Tal N, Honig B, McLaughlin S: **Electrostatic properties of membranes containing acidic lipids and adsorbed basic peptides: Theory and experiment.** *Biophys J* (1999) **77**(6):3176-3188.

14. Cho W, Stahelin RV: **Membrane-protein interactions in cell signaling and membrane trafficking.** *Annu Rev Biophys Biomol Struct* (2005) **34**(119-151.

15. Murray D, Arbuzova A, Honig B, McLaughlin S: **The role of electrostatic and nonpolar interactions in the association of peripheral proteins with membranes.** *Curr Top Membr* (2002) **52**(271-302.

16. Murray D, Arbuzova A, Honig B, McLaughlint S: **The role of electrostatic and nonpolar interactions in the association of peripheral proteins with membranes.** In: *Current topics in membranes.* Volume 52*.* Academic Press, (2002):277-307.

17. Khan HM, He T, Fuglebakk E, Grauffel C, Yang B, Roberts MF, Gershenson A, Reuter N: **A role for weak electrostatic interactions in peripheral membrane protein binding.** *Biophys J* (2016) **110**(6):1367-1378.

18. Murray D, Ben-Tal N, Honig B, McLaughlin S: **Electrostatic interaction of myristoylated proteins with membranes: Simple physics, complicated biology.** *Structure* (1997) **5**(8):985-989.

19. Stahelin RV, Cho W: **Differential roles of ionic, aliphatic, and aromatic residues in membrane–protein interactions: A surface plasmon resonance study on phospholipases a2.** *Biochemistry* (2001) **40**(15):4672-4678.

20. Wrabl JO, Shortle D, Woolf TB: **Correlation between changes in nuclear magnetic resonance order parameters and conformational entropy: Molecular dynamics simulations of native and denatured staphylococcal nuclease.** *Proteins Struct Funct Genet* (2000) **38**(123-133.

21. Lee AG: **How lipids affect the activities of integral membrane proteins.** *Biochimica et Biophysica Acta (BBA) - Biomembranes* (2004) **1666**(1–2):62-87.

22. McIntosh TJ, Simon SA: **Bilayers as protein solvents: Role of bilayer structure and elastic properties.** *The Journal of General Physiology* (2007) **130**(2):225-227.

23. Phillips R, Ursell T, Wiggins P, Sens P: **Emerging roles for lipids in shaping membrane-protein function.** *Nature* (2009) **459**(7245):379-385.

24. Hunte C, Richers S: **Lipids and membrane protein structures.** *Current opinion in structural biology* (2008) **18**(4):406-411.

25. Lee AG: **Biological membranes: The importance of molecular detail.** *Trends in biochemical sciences* (2011) **36**(9):493-500.

26. Laganowsky A, Reading E, Allison TM, Ulmschneider MB, Degiacomi MT, Baldwin AJ, Robinson CV: **Membrane proteins bind lipids selectively to modulate their structure and function.** *Nature* (2014) **510**(7503):172-175.

27. Barrera NP, Zhou M, Robinson CV: **The role of lipids in defining membrane protein interactions: Insights from mass spectrometry.** *Trends in cell biology* (2013) **23**(1):1-8.

28. Contreras FX, Ernst AM, Wieland F, Brugger B: **Specificity of intramembrane protein-lipid interactions.** *Cold Spring Harbor perspectives in biology* (2011) **3**(6).

29. Saliba AE, Vonkova I, Gavin AC: **The systematic analysis of protein-lipid interactions comes of age.** *Nat Rev Mol Cell Biol* (2015) **16**(12):753-761.

30. Murray D, Honig B: **Electrostatic control of the membrane targeting of c2 domains.** *Molecular Cell* (2002) **9**(1):145-154.

31. Singh SM, Murray D: **Molecular modeling of the membrane targeting of phospholipase c pleckstrin homology domains.** *Protein Sci* (2003) **12**(1934-1953.

32. Ben-Tal N, Honig B, Miller C, McLaughlin S: **Electrostatic binding of proteins to membranes. Theoretical predictions and experimental results with charybdotoxin and phospholipid vesicles.** *Biophys J* (1997) **73**(4):1717-1727.

33. Diraviyam K, Stahelin RV, Cho W, Murray D: **Computer modeling of the membrane interaction of fyve domains.** *J Mol Biol* (2003) **328**(3):721-736.

34. Lomize AL, Pogozheva ID, Lomize MA, Mosberg HI: **The role of hydrophobic interactions in positioning of peripheral proteins in membranes.** *BMC Struct Biol* (2007) **7**(44.

35. Kufareva I, Lenoir M, Dancea F, Sridhar P, Raush E, Bissig C, Gruenberg J, Abagyan R, Overduin M: **Discovery of novel membrane binding structures and functions.** *Biochemistry and cell biology = Biochimie et biologie cellulaire* (2014) **92**(6):555-563.

36. Pronk S, Pall S, Schulz R, Larsson P, Bjelkmar P, Apostolov R, Shirts MR, Smith JC, Kasson PM, van der Spoel D, Hess B *et al*: **Gromacs 4.5: A high-throughput and highly parallel open source molecular simulation toolkit.** *Bioinformatics (Oxford, England)* (2013) **29**(7):845-854.

37. Schmid N, Eichenberger AP, Choutko A, Riniker S, Winger M, Mark AE, van Gunsteren WF: **Definition and testing of the gromos force-field versions 54a7 and 54b7.** *European Biophysics Journal* (2011) **40**(7):843.

38. Irvine WA: **Testing of parameters for a biologically accurate brain membrane and molecular dynamics simulations exploration of membrane interactions and conformational changes exhibited by p110a and its oncogenic mutants.** In: *Institute of Natural and Mathematical Sciences.* Masters of Science*.* Massey University, Albany, New Zealand, Massey University (2014).

39. Poger D, Van Gunsteren WF, Mark AE: **A new force field for simulating phosphatidylcholine bilayers.** *Journal of Computational Chemistry* (2010) **31**(6):1117-1125.

40. Berendsen HJC, Grigera JR, Straatsma TP: **The missing term in effective pair potentials.** *The Journal of Physical Chemistry* (1987) **91**(24):6269-6271.

41. Berendsen HJC, Postma JPM, van Gunsteren WF, DiNola A, Haak JR: **Molecular dynamics with coupling to an external bath.** *The Journal of Chemical Physics* (1984) **81**(8):3684-3690.

42. Hess B, Bekker H, Berendsen HJC, Fraaije JGEM: **Lincs: A linear constraint solver for molecular simulations.** *Journal of Computational Chemistry* (1997) **18**(12):1463-1472.

43. Essmann U, Perera L, Berkowitz ML, Darden T, Lee H, Pedersen LG: **A smooth particle mesh ewald method.** *The Journal of Chemical Physics* (1995) **103**(19):8577-8593.

44. Hunter JD: **Matplotlib: A 2d graphics environment.** *Computing In Science & Engineering* (2007) **9**(3):90-95.

45. Humphrey W, Dalke A, Schulten K: **Vmd: Visual molecular dynamics.** *Journal of molecular graphics* (1996) **14**(1):33-38, 27-38.

46. Song MS, Salmena L, Pandolfi PP: **The functions and regulation of the pten tumour suppressor.** *Nat Rev Mol Cell Biol* (2012) **13**(5):283-296.

47. Yin Y, Shen WH: **Pten: A new guardian of the genome.** *Oncogene* (2008) **27**(41):5443-5453.

48. Lee JO, Yang H, Georgescu MM, Di Cristofano A, Maehama T, Shi Y, Dixon JE, Pandolfi P, Pavletich NP: **Crystal structure of the pten tumor suppressor: Implications for its phosphoinositide phosphatase activity and membrane association.** *Cell* (1999) **99**(3):323-334.

49. Das S, Dixon JE, Cho W: **Membrane-binding and activation mechanism of pten.** *Proceedings of the National Academy of Sciences of the United States of America* (2003) **100**(13):7491-7496.

50. Vazquez F, Ramaswamy S, Nakamura N, Sellers WR: **Phosphorylation of the pten tail regulates protein stability and function.** *Molecular and cellular biology* (2000) **20**(14):5010-5018.

51. Georgescu MM, Kirsch KH, Kaloudis P, Yang H, Pavletich NP, Hanafusa H: **Stabilization and productive positioning roles of the c2 domain of pten tumor suppressor.** *Cancer Res* (2000) **60**(24):7033-7038.

52. Georgescu MM, Kirsch KH, Akagi T, Shishido T, Hanafusa H: **The tumor-suppressor activity of pten is regulated by its carboxyl-terminal region.** *Proceedings of the National Academy of Sciences of the United States of America* (1999) **96**(18):10182-10187.

53. Shenoy S, Shekhar P, Heinrich F, Daou M-C, Gericke A, Ross AH, Lösche M: **Membrane association of the pten tumor suppressor: Molecular details of the protein-membrane complex from spr binding studies and neutron reflection.** *PLOS ONE* (2012) **7**(4):e32591.

54. Lumb Craig N, Sansom Mark SP: **Defining the membrane-associated state of the pten tumor suppressor protein.** *Biophysical Journal* (2013) **104**(3):613-621.

55. Williams R, Berndt A, Miller S, Hon WC, Zhang X: **Form and flexibility in phosphoinositide 3-kinases.** *Biochem Soc Trans* (2009) **37**(Pt 4):615-626.

56. Cleary JM, Shapiro GI: **Development of phosphoinositide-3 kinase pathway inhibitors for advanced cancer.** *Curr Oncol Rep* (2010) **12**(2):87-94.

57. Arcaro A, Guerreiro AS: **The phosphoinositide 3-kinase pathway in human cancer: Genetic alterations and therapeutic implications.** *Curr Genomics* (2007) **8**(5):271-306.

58. Gabelli SB, Huang CH, Mandelker D, Schmidt-Kittler O, Vogelstein B, Amzel LM: **Structural effects of oncogenic pi3kalpha mutations.** *Curr Top Microbiol Immunol* (2010) **347**(43-53.

59. Amzel LM, Huang CH, Mandelker D, Lengauer C, Gabelli SB, Vogelstein B: **Structural comparisons of class i phosphoinositide 3-kinases.** *Nat Rev Cancer* (2008) **8**(9):665-669.

60. Jiménez C, Hernández C, Pimentel B, Carrera AC: **The p85 regulatory subunit controls sequential activation of phosphoinositide 3-kinase by tyr kinases and ras.** *Journal of Biological Chemistry* (2002) **277**(44):41556-41562.

61. Gkeka P, Evangelidis T, Pavlaki M, Lazani V, Christoforidis S, Agianian B, Cournia Z: **Investigating the structure and dynamics of the pik3ca wild-type and h1047r oncogenic mutant.** *PLOS Computational Biology* (2014) **10**(10):e1003895.

62. Burke JE, Perisic O, Masson GR, Vadas O, Williams RL: **Oncogenic mutations mimic and enhance dynamic events in the natural activation of phosphoinositide 3-kinase p110alpha (pik3ca).** *Proceedings of the National Academy of Sciences of the United States of America* (2012) **109**(38):15259-15264.

63. Hon W-C, Berndt A, Williams RL: **Regulation of lipid binding underlies the activation mechanism of class ia pi3-kinases.** *Oncogene* (2012) **31**(32):3655-3666.

64. Hon WC, Berndt A, Williams RL: **Regulation of lipid binding underlies the activation mechanism of class ia pi3-kinases.** *Oncogene* (2012) **31**(32):3655-3666.

Supporting Information for

Computational prediction of protein-membrane interactions

William A. Irvine[1,2], Jack U. Flanagan[3,5] and Jane R Allison[1,2,4,5]

1. Centre for Theoretical Chemistry and Physics, Massey University Auckland, Private Bag 102904, 0632 Auckland, New Zealand
2. Institute of Natural and Mathematical Sciences, Massey University Auckland, Private Bag 102904, 0632 Auckland, New Zealand
3. Auckland Cancer Society Research Centre, University of Auckland, Private Bag 92019, Auckland, New Zealand
4. Biomolecular Interaction Centre, University of Canterbury, Private Bag 4800, Christchurch 8140, New Zealand
5. Maurice Wilkins Centre for Molecular Biodiscovery, University of Auckland, Private Bag 92019, Auckland, New Zealand

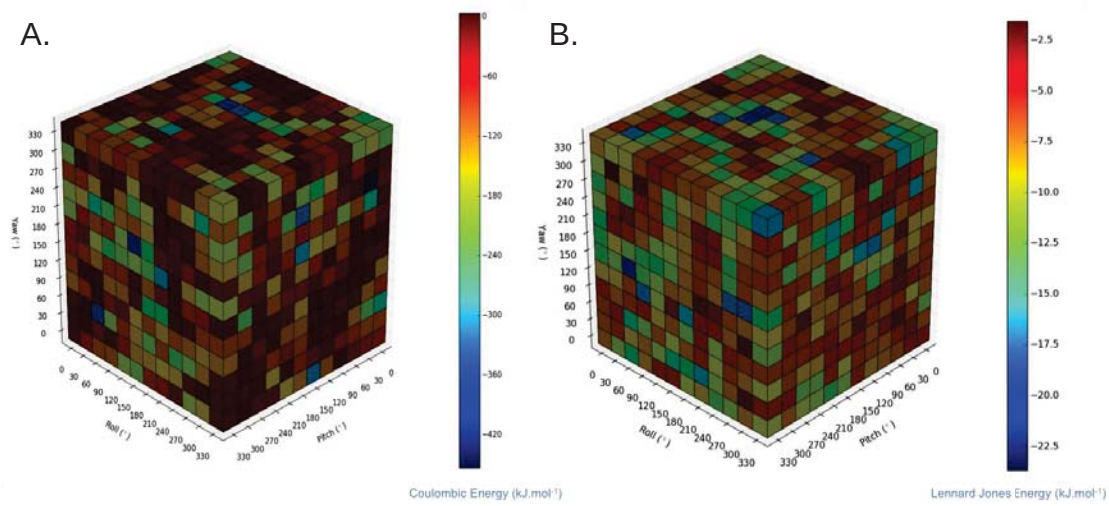


Figure S1. Electrostatic interactions dominate the PTEN-membrane interaction energy. A. Coulombic and B. Lennard-Jones contribution to the interaction potential energy between PTEN and the model cell membrane at a protein-membrane distance of 0.5 nm as a function of the orientation of the protein with respect to the membrane normal.

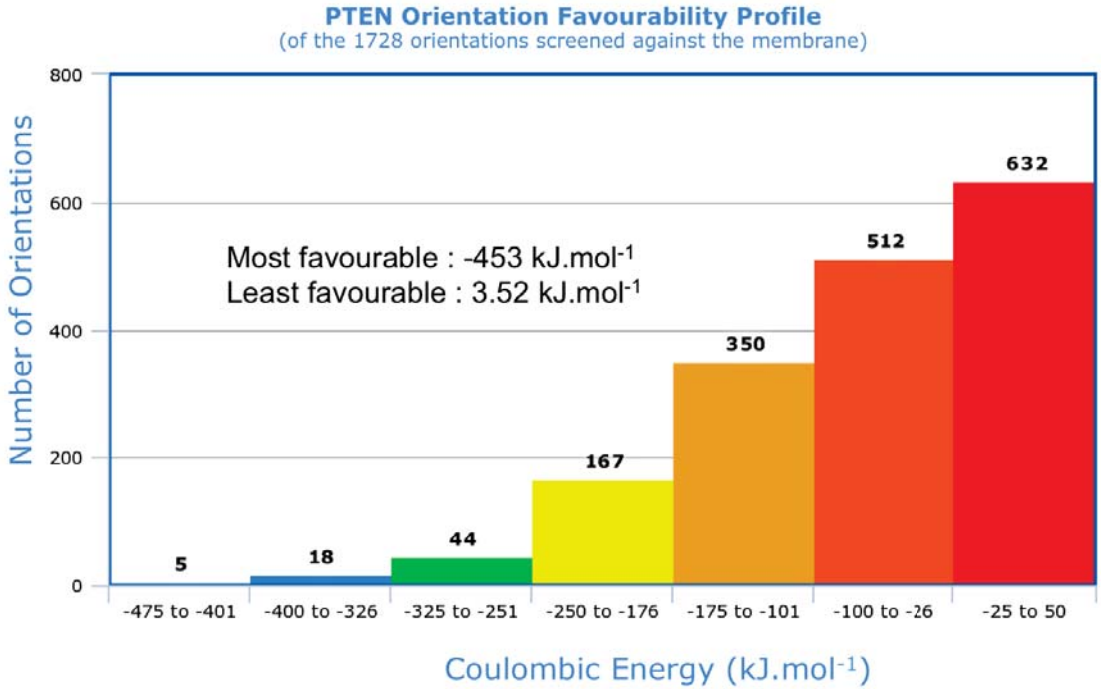


Figure S2. Only a small proportion of orientations exhibit highly favourable PTEN-membrane Coulombic interaction energies. Histogram of the distribution of values of the Coulombic interaction potential energy between PTEN and the model cell membrane at a protein-membrane distance of 0.5 nm over all orientations of the protein with respect to the membrane normal.

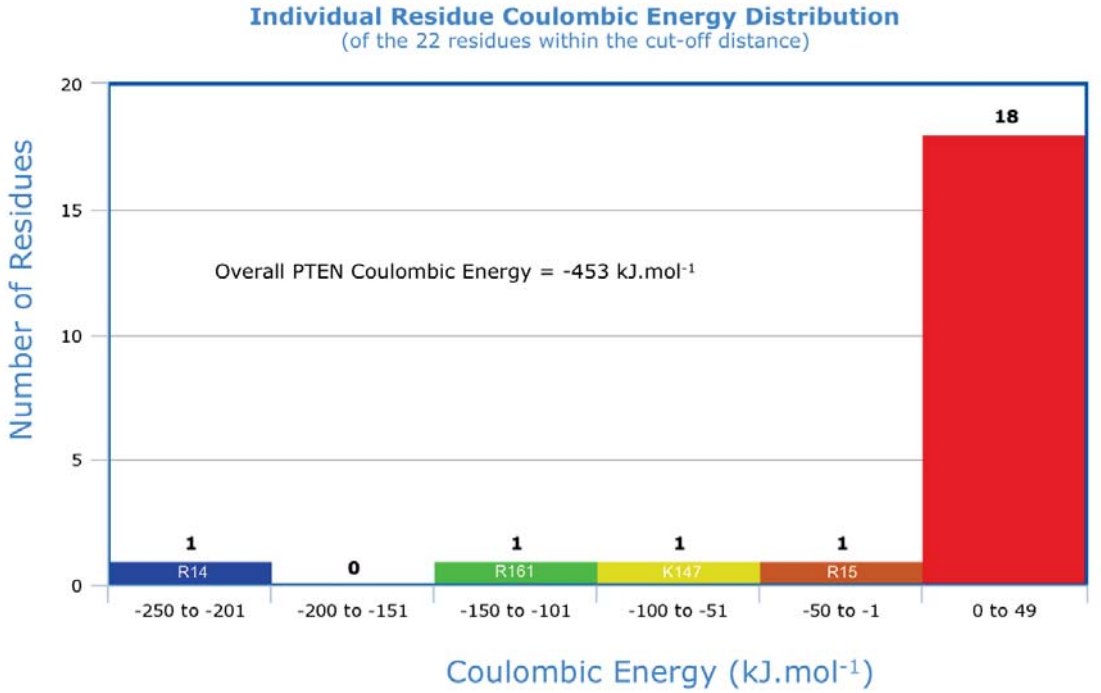


Figure S3. A few residues dominate the PTEN-membrane Coulombic interaction energy. Histogram of the distribution of values of the per-residue Coulombic interaction potential energies between PTEN and the model cell membrane at a protein-membrane distance of 0.5 nm at the most favourable orientation of the protein with respect to the membrane normal.

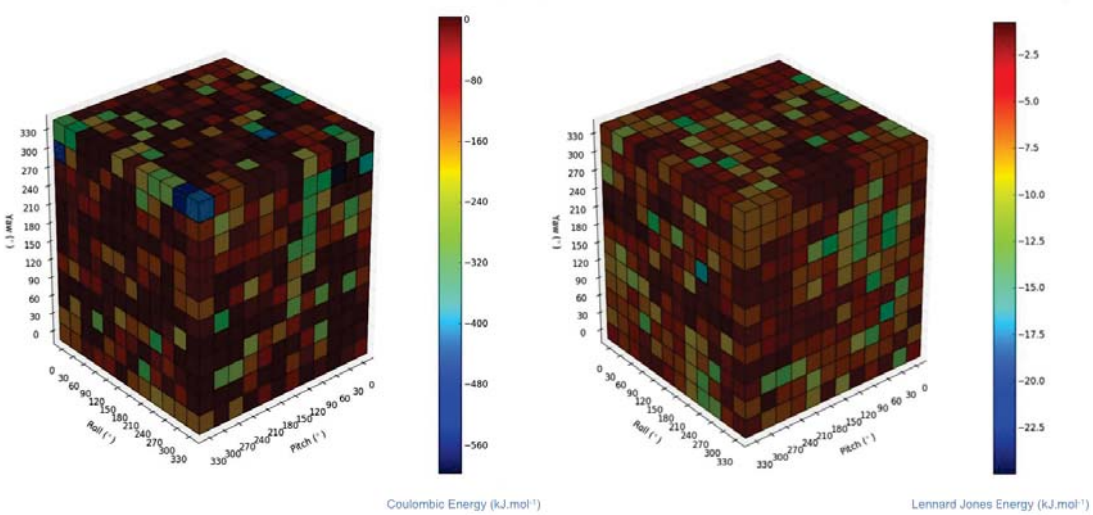


Figure S4. Electrostatic interactions dominate the PI3Kα-membrane interaction energy. A. Coulombic and B. Lennard-Jones contribution to the interaction potential energy between PI3Kα and the model cell membrane at a protein-membrane distance of 0.5 nm as a function of the orientation of the protein with respect to the membrane normal.

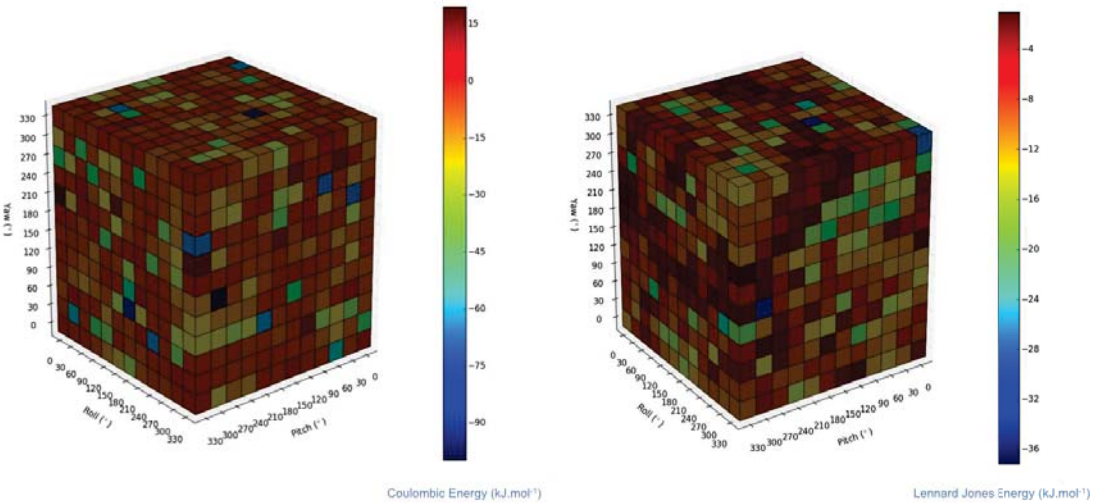


Figure S5. Electrostatic interactions do not dominate the PI3Kα-DPPC interaction energy. A. Coulombic and B. Lennard-Jones contribution to the interaction potential energy between PI3Kα and a DPPC bilayer at a protein-bilayer distance of 0.5 nm as a function of the orientation of the protein with respect to the bilayer normal.

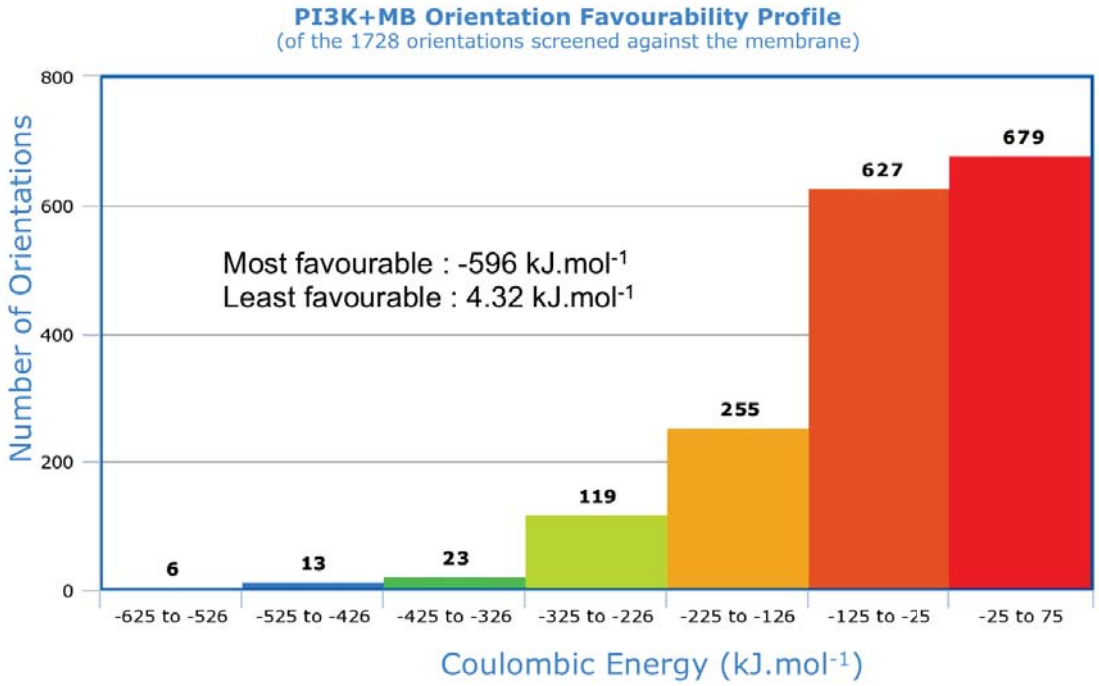


Figure S6. Only a small proportion of orientations exhibit highly favourable PI3Kα-membrane Coulombic interaction energies. Histogram of the distribution of values of the Coulombic interaction potential energy between PI3Kα and the model cell membrane at a protein-membrane distance of 0.5 nm over all orientations of the protein with respect to the membrane normal.

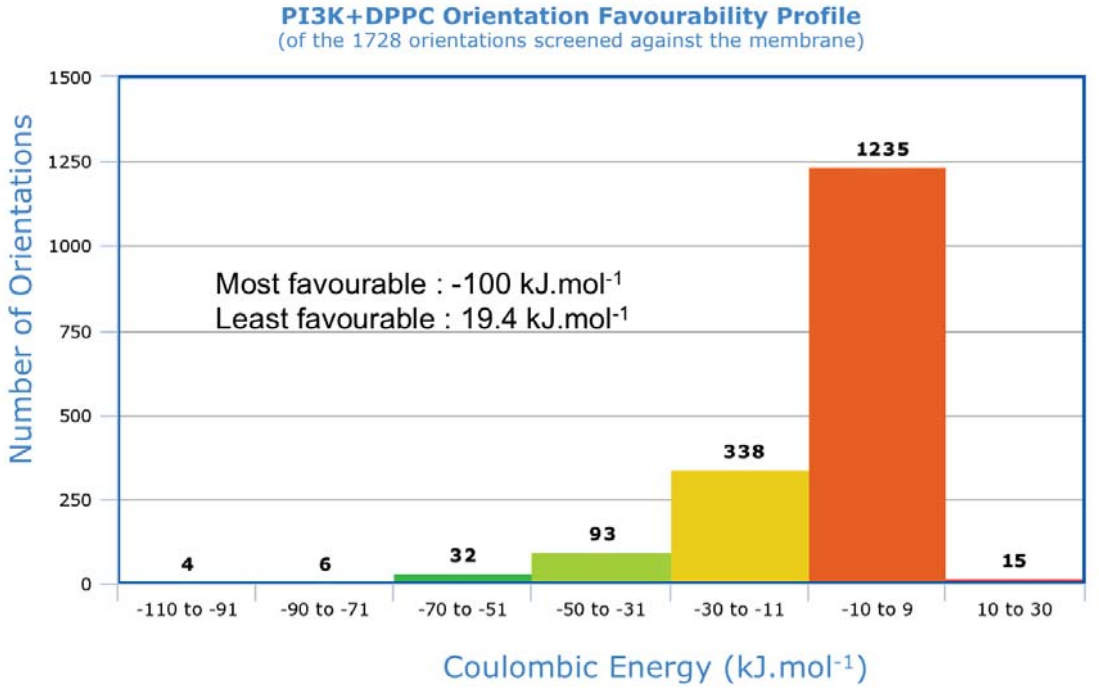


Figure S7. Only a small proportion of orientations exhibit highly favourable PI3Kα-DPPC Coulombic interaction energies. Histogram of the distribution of values of the Coulombic interaction potential energy between PI3Kα and a DPPC bilayer at a protein-bilayer distance of 0.5 nm over all orientations of the protein with respect to the bilayer normal.

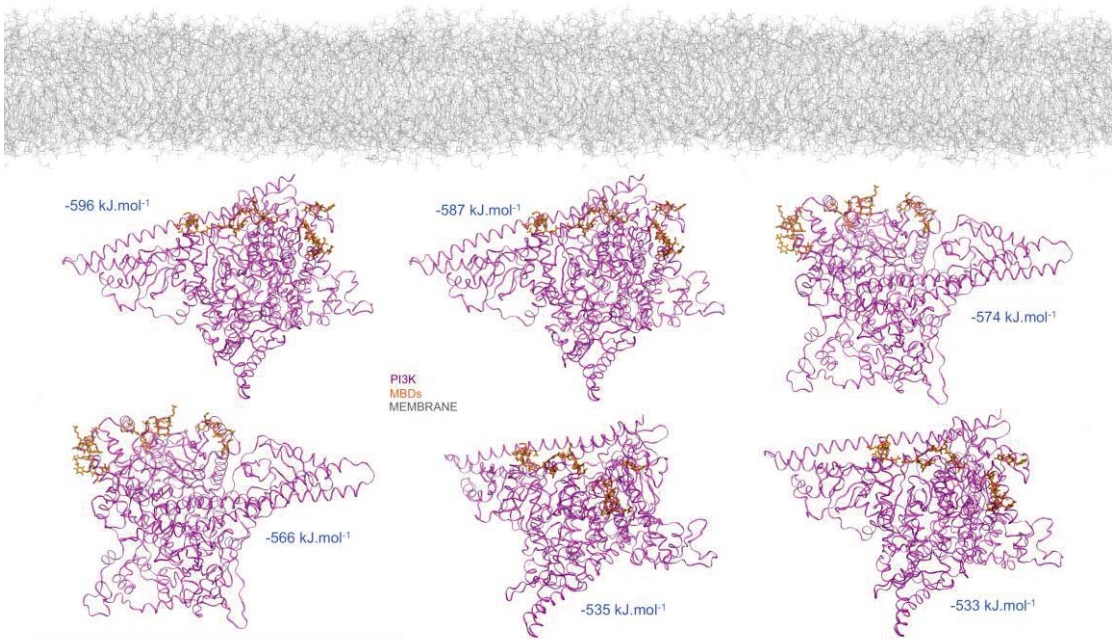


Figure S8. The most favourable PI3Kα-membrane orientations all present the MBDs to the membrane. Snapshots of the six orientations of PI3Kα with the most favourable Coulombic interaction potential energies (as labelled) with the model cell membrane at a protein-membrane distance of 0.5 nm. PI3Kα is shown in pink, the MBDs and WIF motif in orange, and the membrane in grey.

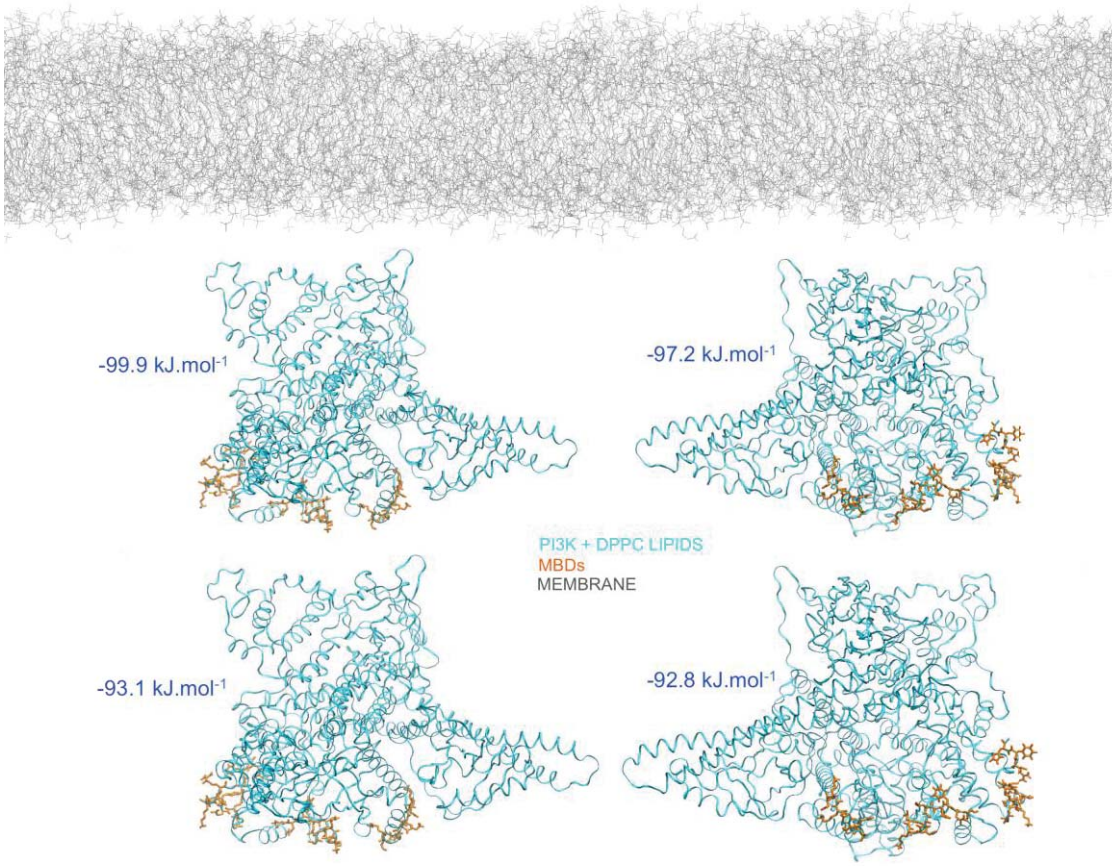


Figure S9. The most favourable PI3Kα-DPPC orientations do not present the MBDs to the DPPC bilayer. Snapshots of the four orientations of PI3Kα with the most favourable Coulombic interaction potential energies (as labelled) with the DPPC bilayer at a protein-bilayer distance of 0.5 nm. PI3Kα is shown in cyan, the MBDs and WIF motif in orange, and the bilayer in grey.

Appendix D - Supplementary PI3K Data

Supplementary Information for Section 4.3.3.1

Protein-Membrane Distance Analysis

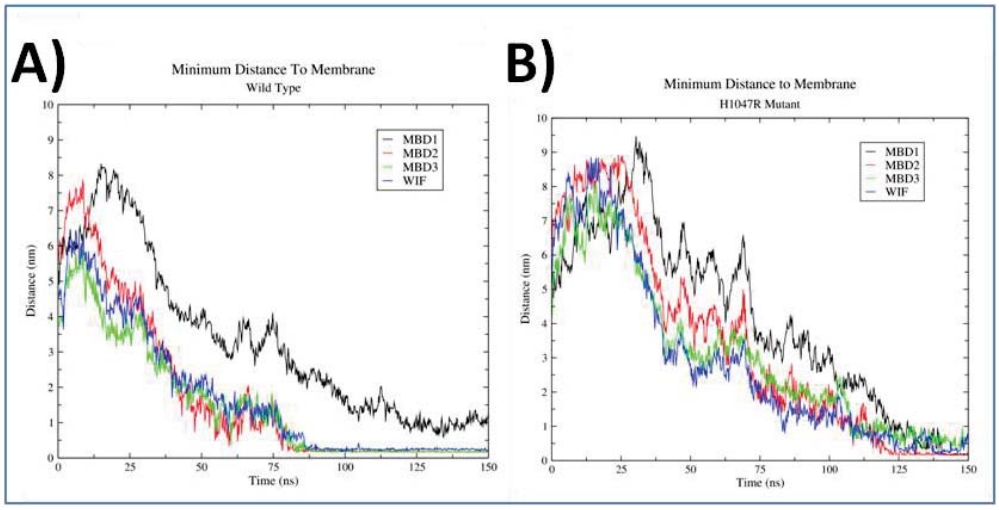


Figure Di - *The minimum distance between MBDs (1-black, 2-red, 3-green, WIF-blue) and the cell membrane over the course of the second run of the 150 ns simulation of the wild type (A) and H1047R (B) systems.*

The second run of the wild type simulation took 50 ns longer to establish a protein-membrane interaction than the first run (Figure Di-A, Figure 4xvi-A), while the second run of the H1047R simulation took almost 100 ns longer to establish a protein-membrane interaction than in the first run (Figure Di-B, Figure 4xvi-B). This resulted in the PI3Kα system spending considerably less time in a catalytically competent position, where the membrane binding domains were all interacting with the membrane resulting in the C2-iSH2-kinase domain interface being presented. However, as in the first set of simulations, MBD1 only interacts with the membrane in the H1047R system and not the wild type system due to the difference in orientation observed in Figure 4vi and explored in further detail in Chapter 4.

For both the wild type and H1047R mutant systems, the same basic residues are also found to be interacting with anionic lipids in the second run of simulations. In the case of MBD1 (only in H1047R), LYS 723 is the closest to a POPS lipid; for MBD2, it's LYS 863; and for MBD3 it's LYS 973 (Figure Dii-A,B). In both systems, the WIF motif in the C terminal tail is the most deeply embedded set of residues in the protein (Figure Dii-C) albeit not to the extent seen in the first run of simulations

(Figure 4xvii-C). In the case of MBD3 in the H1047R system, LYS 973 is also substantially further away from a POPS lipid in the second run of simulations than in the first (Figure 4xvii-B, Figure Dii-B).

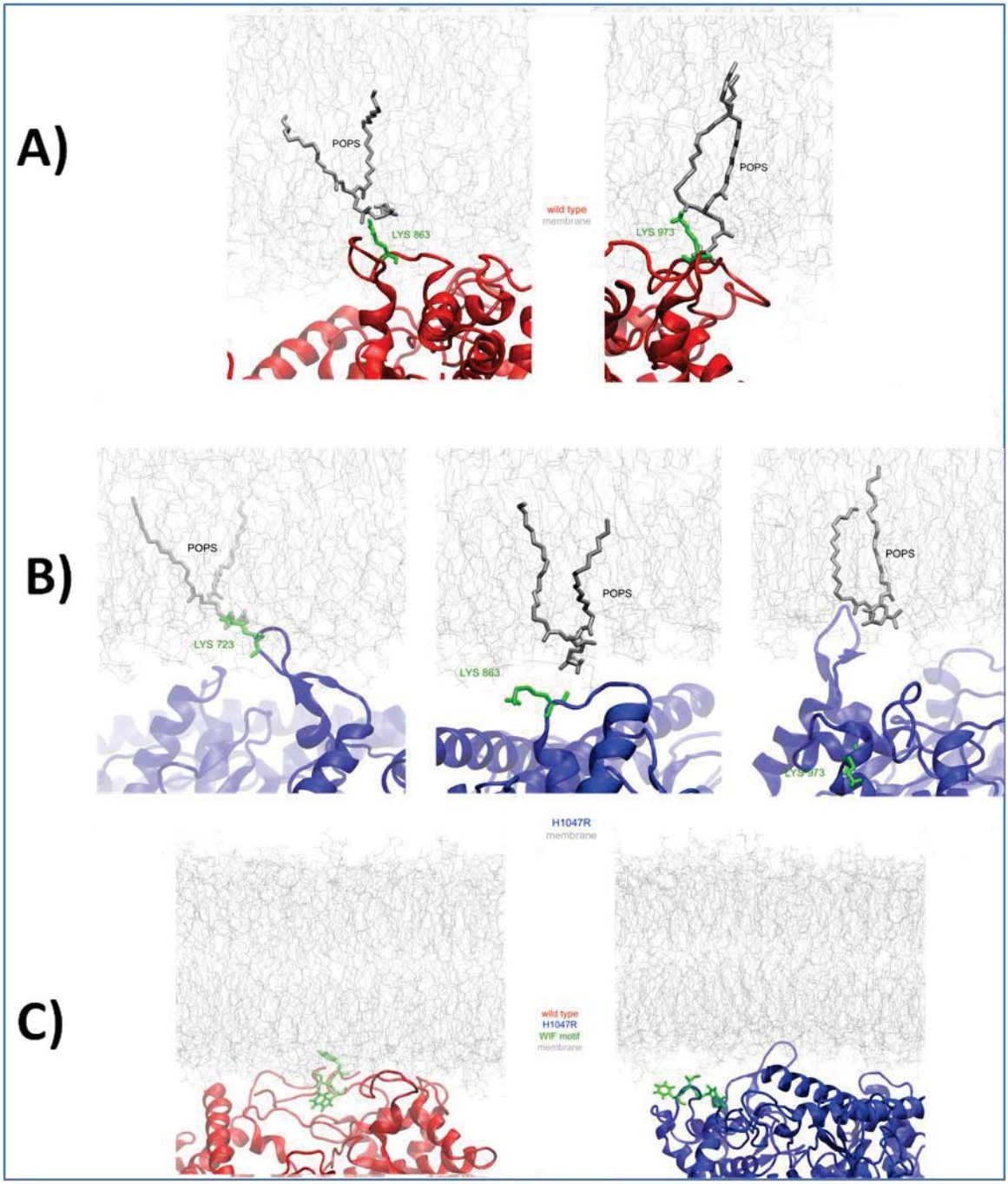


Figure Dii - *Final snapshots of the wild type (red) and H1047R (blue) proteins following the second run of 150 ns simulations in the presence of a cell membrane (gray). Key residues (green) and key lipids (dark gray) are highlighted.* ***A)*** *From left to right: LYS 863 (MBD2) and LYS 973 (MBD3) seen interacting with a POPS lipid.* ***B)*** *From left to right: LYS 723 (MBD1), LYS 863 (MBD2) and LYS 973 (MBD3) seen interacting with a POPS lipid.* ***C)*** *Embedment of the C terminal tail into the cell membrane showing the position of the WIF motif in the wild type (left) and H1047R (right) systems. The WIF motif does not embed to the same extent as in the first run of both systems (Figure 4xvii-C).*

Supplementary Information for Section 4.3.3.2

Substrate Analysis

As in the first run of simulations, only the H1047R system explored a catalytically competent position in the second run of simulations, with the PIP_2 binding pocket presented to the lipid bilayer. As such, the distance between the potential substrate (closest PIP_2 lipid molecule to the binding pocket) and relevant regions of the protein (the binding pocket and R949) was calculated (Figure Diii-A).

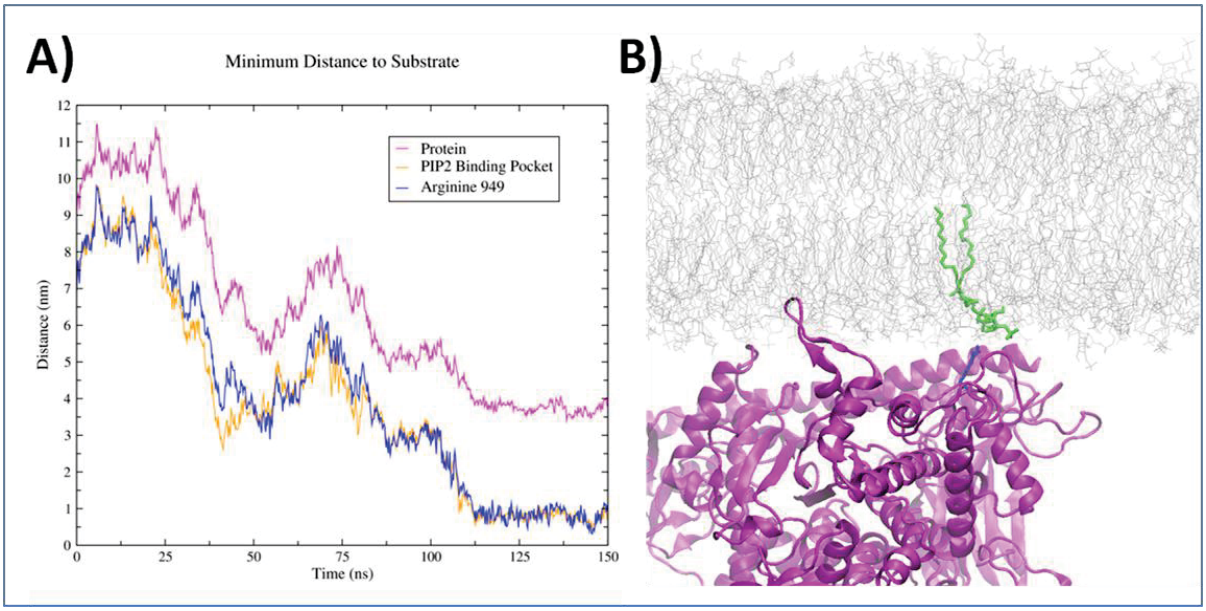


Figure Diii - ***A)*** *Centre of mass distance between p110α-p85α (purple) and the identified PIP_2 substrate over the course of the second run of the H1047R system simulation. The minimum distances of the PIP_2 binding pocket (orange) and R949 (blue) to the same PIP_2 substrate are also been shown.* ***B)*** *Snapshot of p110α-p85α (purple) in contact with the cell membrane (gray) at approximately 145 ns in the second run of the H1047R system simulation representing the distance between R949 (blue) and the PIP_2 substrate (green).*

In the case of the second run of the H1047R simulation, R949 got within 1 nm of the PIP_2 substrate following 110 ns of simulation, with the minimum distance occuring at the 145 ns timepoint (Figure Diii). Hoeever, unlike in the first run of the H1047R simulation, no hydrogen bond network was found to exist between R949 and the PIP_2 substrate, water mediated or otherwise. In the first run of simulations, first contact between H1047R and the membrane occurred around 35 ns prior to R949 exhibiting a hydrogen bonding network with PIP_2 (Figure 4xix). As the second run of simulations only established first contact with the membrane at approximately 120 ns (Figure Di), it was unlikely that R949 had the opportunity to develop a hydrogen bonding network with PIP_2 prior to the completion of the 150 ns simulation.

Supplementary Information for Section 4.3.4

Intra-Protein Interactions

Using ConAn to establish intra-protein contacts within PI3Kα revealed H1047R to form more contacts between Kα12 and the remainder of the kinase domain when compared to wild type in the first run of simulations (Figure 4xxvi). While this remains the case in the second run of simulations, it is not to the level observed in the first run, with the wild type enzyme also exhibiting contacts between Kα12 and the Kα11 helix, activation loop, and catalytic loop. These contacts were only observed in the H1047R system in the first run of simulations, but are seen in both systems in the second run.

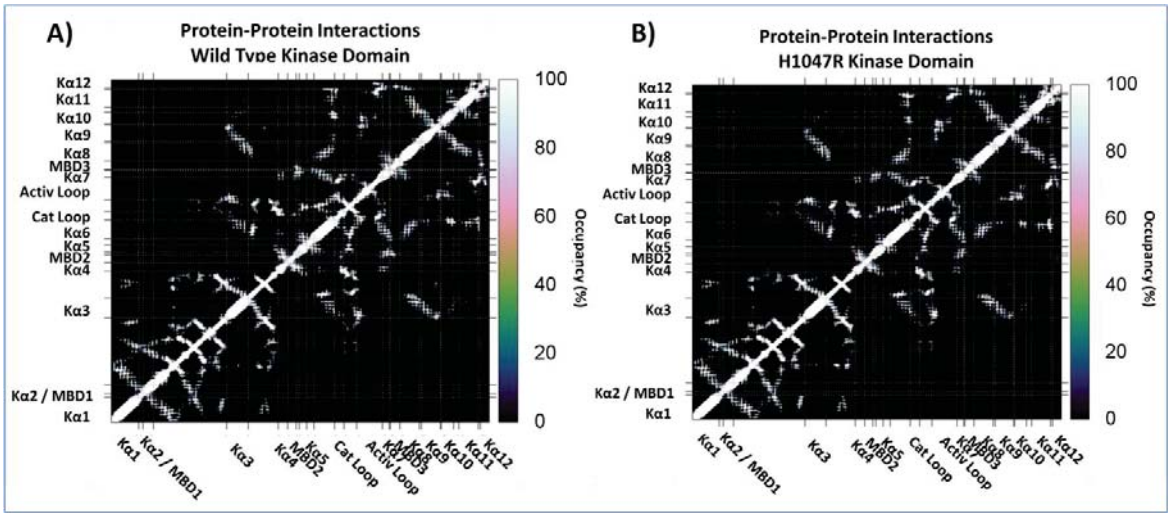


Figure Div - *Map of the intra-protein interactions exhibited by the kinase domain of the wild type (A) and H1047R (B) p110α-p85α complexes and their occupancies, with gridlines defined by important kinase domain segments.*

To gain additional insight, the hydrogen bonds made by residue 1047 were characterised for the second run of simulations (Figure Dv). In the first run, histidine (H1047) was found to make contacts which were deemed more inhibitory in nature (with the SH2 domains), whilst arginine (R1047) was found to make contacts deemed more conducive to PI3Kα's function (with the C terminal tail) (Figure 4xxvii). They were both found to interact with E950 in the first run of simulations, however that is not the case in the second run of simulations. This contact with the activation loop only exists in the second run of the H1047R simulation, seemingly as it got closer to the membrane (100 ns) (Figure Dv-B). H1047's contacts with the SH2 linker prior to membrane interaction is repeated in the second run of the wild type simulations (Figure Dv-A), whilst R1047's contacts with the C terminal tail prior to membrane interaction is once again seen in the second run of the H1047R simulations (Figure Dv-B).

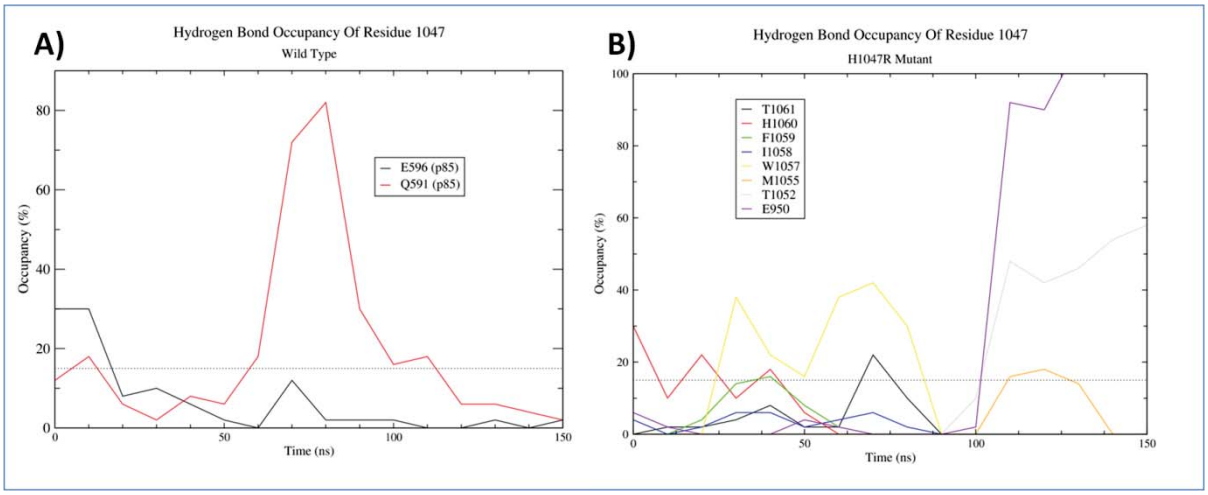


Figure Dv - *Relative occupancy of the hydrogen bonds formed by residue 1047 over the course of the first run of the wild type (A) and H1047R (B) system simulations. Hydrogen bonds with E596 (black) and Q591 (red) exist only in the wild type system (A). Hydrogen bonds with the WIF motif (residues 1057-1059) and remainder of the C terminal tail (residues 1050-1064) exist only in the H1047R system. Hydrogen bonds with E950 exist only in the H1047R second simulation run.*