Studies of the MuSK System Using Molecular Dynamics Simulations

Amanda C. Buyan

Wadham College

University of Oxford

Submitted in partial fulfillment
of the requirements for the degree of

Doctor of Philosophy

Trinity Term 2015

Abstract

Studies of the MuSK System Using Molecular Dynamics Simulations

Amanda C. Buyan
Wadham College
University of Oxford
Doctor of Philosophy
Trinity Term 2015

Formation and maintenance of neuromuscular junctions, as well as signalling from the nerve to the muscle, depends on the concerted action of a number of membrane proteins. Muscle-Specific Kinase (MuSK) and Downstream-of-Kinase 7 (Dok7) are key players in these processes. MuSK acts as the central component in the complexes responsible for nerve-muscle communication. Dok7 then regulates full activation of MuSK. Despite available structural data for these proteins, the structure of the overall Dok7/MuSK complex in a cell membrane remains elusive. In this thesis, a multi scale molecular simulation approach (combining coarse-grained and atomistic simulations) was used to study the formation of the Dok7/MuSK/membrane complex. Three major insights into this process were obtained. Firstly, dimerisation of the transmembrane helix of MuSK is mediated by a small-x-x-x-small amino acid motif, and it is not affected significantly by the addition of anionic lipids or unstructured extensions on each end of the helices. Secondly, Dok7 interacts with PIP-containing bilayers via a positive patch on its pleckstrin-homology (PH) domain. Thirdly, it is shown that MuSK's behaviour in a complex model membrane is potentially influenced by the presence of Dok7. Taken together, these results allow formulation of a defined model of the structure and mechanism of MuSK in a model membrane.

Acknowledgements

First, I would like to thank my supervisor, Mark Sansom, who first recommended the Wellcome Trust programme, and also decided to take me on as a D.Phil student. He has provided important guidance in my projects, and has taught me that genuinely nice and caring people can indeed get ahead in life.

I would also like to thank the members of the lab, especially David, Joe, Anna, Maria, Tere, Matthieu, Mickaël, Tom, Nick, Fiona, Nathalie, and Jerome. A huge thanks to Laura and Julia, for continued friendship and support for the past two years. A special thanks to Heidi Koldsø, who, aside from being a fabulous and tolerant deskmate for three years, provided me with a valuable "rendering intervention" as well as the inspiration to be a better scientist. Another special thanks to Antreas Kalli, who provided most of the scripts and insight in Chapters 4 and 5, and who was always willing to help me with both of these projects, all while maintaining an impressively happy disposition, even in the worst of times.

Many thanks to my friends here outside SBCB, including Annie, Marta and Diana. Thanks to Lucy, who dealt with living with me for a year and is still my friend. Hannah, Radhika, and Jan - thanks for all your help and support in kickboxing. A thanks as well to my fellow coursemates Ellie, Alex, Marko and Santiago, for being there these past four years.

A huge thanks to many of my friends"across the pond," especially Connie, Kathryn, and Jess, for their continuing friendship beyond our tenure at 1301 Spring Street. A special thanks to Rachel, for continuing to be my "oboe buddy" long after our musical relationship concluded. Another special thanks to "my friend Sadie," who has been the sounding board for many frustrations, and whose conversations via Skype have provided many reliefs for difficult times.

And finally, I would like to thank my family, especially my mother, father, and brother, whose unwavering support in many ways throughout these 4 years have been invaluable. A special thanks to my grandmother's support, who passed away during this D.Phil. I know she would have been proud to have seen this thesis come to fruition.

Amanda Buyan
September 2015, University of Oxford

List of Publications Arising From This Thesis

Buyan, A., Kalli, A.C., and Sansom, M.S.P. "Multiscale simulations suggest a mechanism for the association of Dok7's PH Domain with PIP-containing Bilayers." (Manuscript in Preparation for PLoS Comp. Biol.)

Buyan, A., Kalli, A.C., and Sansom, M.S.P. "Elucidating a Potential Mechanism for MuSK Activation via its Interactions with Dok7 in a complex membrane." (Manuscript in Preparation for Nature Comms.)

Contents

List of Figures

List of Tables

List of Abbreviations, Acronyms and Symbols

AChR	Acetylcholine Receptor
AT-MD	Atomistic Molecular Dynamics
CG	Coarse-Grained
CG-MD	Coarse Grain-Molecular Dynamics
COM	Centre of Mass
Dok7	Downstream-of-Kinase 7
DPPC	dipalmitoylphosphatidylcholine
EGFR	Epidermal Growth Factor Receptor
ENM	Elastic Network Model
G3	Laminin G-like
GDP	Guanosine Diphosphate
GEF	Guanidine Nucleotide Exchange Factor
GpA	Glycophorin A
GPCR	G-Protein Coupled Receptor
GROMACS	GROningen MAchine for Chemical Simulations
GTP	Guanisine Triphosphate
HFIP	Hexaflouro-2-propanol
IRS	Insulin Receptor Substrate
JM	Juxtamembrane
L_D	Lipid Disordered
L_O	Lipid Ordered
LH	Left-Handed

LMPG	1-myristoyl-2-hydroxy-sn-glycero-3-[phospho-rac-(1-glycerol)]
Lrp4	Low-Density Lipoprotein Receptor-Related 4
MD	Molecular Dynamics
MuSK	Muscle-Specific Kinase
NMJ	Neuromuscular Junction
PBC	Periodic Boundary Conditions
PC	Phosphatidylcholine
PDB	Protein Data Bank
PH	Pleckstrin Homology
PIP	phosphatidylinositol phosphates
PIP_2	1,2-Diacyl-sn-glycero-3-phospho-(1-D-myo-inositol 4,5-bisphosphate)
PM	Plasma Membrane
PME	Particle Mesh Ewald
POPC	1-palmitoyl, 2-oleoyl-sn-glycero-3-phosphocholine
POPS	1-palmitoyl-2-oleoyl-sn-glycerol-3-phosphoethanolserine
PS	Phosphatidylserine
PTB	Phosphotyrosine Binding
QM	Quantum Mechanics
RDF	Radial Distribution Function
RH	Right-Handed
RMSD	Root Mean Square Deviation
RTK	Receptor Tyrosine Kinase
SDS	sodium dodecyl-sulfate
SPP	Single Protein Production
TM	Transmembrane
TOXCAT	ToxR-mediated activation of a reporter gene encoding Chloramphenicol Acetyltransferase (CAT)
VDW	van der Waals

1

Introduction

1.1 Membrane Proteins and Transmembrane Signalling

Membrane proteins make up a considerable part of the human genome (∼30%)[1], and are important in transmitting signals from the outside to the inside of the cell. This makes them liable to be drug targets[2]. They are also involved in many cellular processes, from signalling[3] to cell-cell recognition[4], and can be divided into two categories: integral membrane proteins, which span across the membrane (i.e. trans-

porters, channels and signalling receptors) and peripheral membrane proteins, which interact with the membrane surface (i.e. pleckstrin homology (PH) domains). (Figure 1.1). Peripheral membrane proteins can also attach to integral membrane proteins, and these interactions, as well as their interactions with the membrane, act to carry signals inside the cell and regulate activity of integral membrane proteins. The majority of membrane proteins are α-helical, with their membrane portion consisting of a single spanning transmembrane (TM) α-helix, or a bundle of α-helices. At the time of writing this chapter (August-September 2015), the number of known structures of membrane proteins was 2350, 713 of them being unique structures, according to MemProtMD (http://sbcb.bioch.ox.ac.uk/memprotmd/beta/)[5].

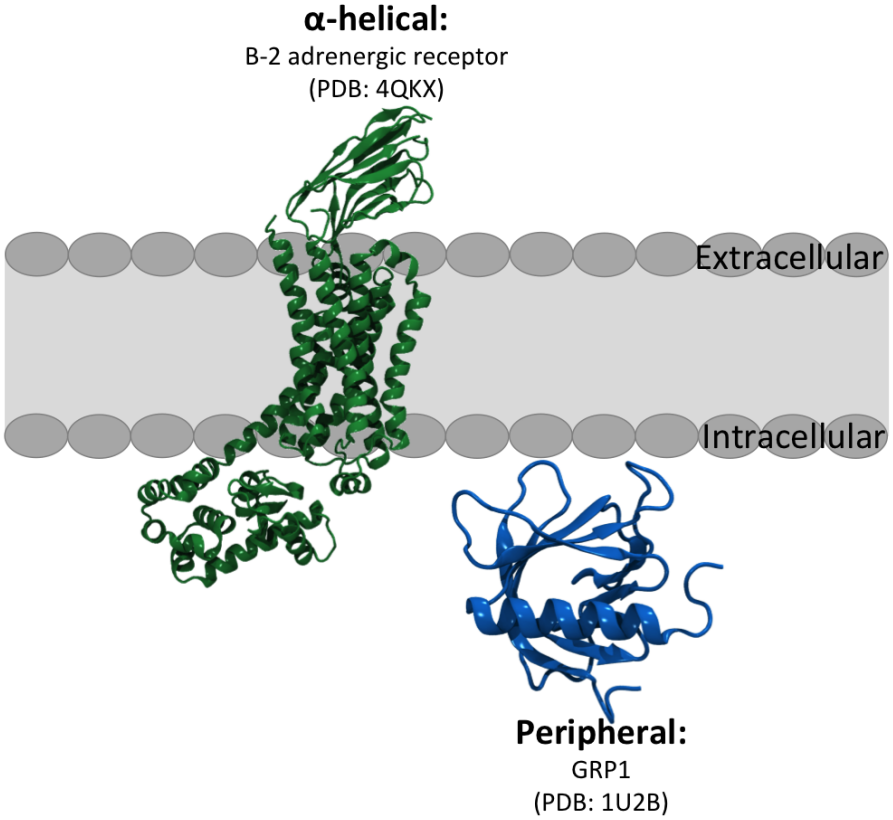

Figure 1.1: Examples of membrane proteins. The first (left, green) is an α-helical integral membrane proteins, and the second (right, blue) is a peripheral membrane protein. All proteins are shown in cartoon representation. The membrane is represented by a grey box and grey ovals.

There are a number of large families of membrane proteins, such as G-Protein Coupled

Receptors (GPCRs) and Receptor Tyrosine Kinases (RTKs)[6]. GPCRs are proteins with seven membrane-spanning α-helices which bind a variety of ligands, such as hormones, drugs and ions. The binding of these ligands induces a conformational change within the GPCR itself. Following the conformational change, the GEF domain of the GPCR allosterically activates the heterotrimeric G protein bound to the intracellular side of the GPCR. This facilitates exchange of a GDP for a GTP, causing dissociation of the heterotrimeric G protein from the GPCR. This allows the G protein to modulate activity of other proteins and facilitates further extracellular signalling. GPCRs are involved in many signalling processes, from neurotransmission to hormonal pathways[7].

The GPCR signalling process is very similar to other signalling proteins. The process starts when a ligand binds to the ectodomain of the membrane protein in question. This binding induces conformational changes both in the ectodomain and membrane portion, carrying the message across the membrane and stimulating the signalling inside the cell.

Human RTKs are a superfamily of 58 signalling proteins which are grouped into 20 subfamilies[8]. Structurally, they consist of an extracellular domain, a single-spanning TM helix, a juxtamembrane (JM) region, and a kinase domain (Figure 1.2). The signal transduction pathway of RTKs involve the ligand first binding to the extracellular domain. This triggers dimerization of the receptor, followed by autophosphorylation of important tyrosine residues in either the JM region, or most often, the kinase domain[4;6]. This phosphorylation allows other signalling proteins to assemble around it, resulting in many signalling cascades responsible for many biological processes and diseases, such as achondroplasia[9] to cell differentiation[10] to nerve/muscle cell communication[11;12].

Specific to RTKs, it is known that a small-x-x-x-small amino acid motif is common

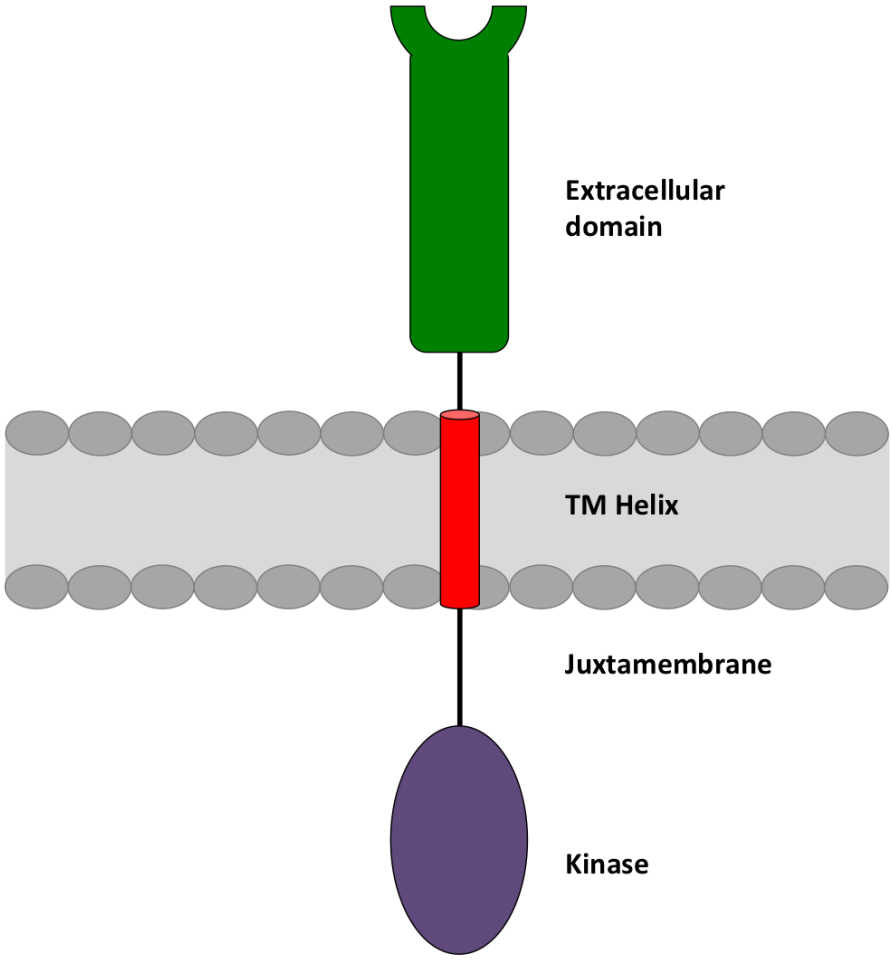

Figure 1.2: Schematic of a receptor tyrosine kinase. The extracellular domain is green rectangle, the transmembrane helix is a red cylinder, and the kinase domain is a purple oval. Unstructured regions are denoted by a black line. The lipid bilayer is represented by grey rectangles and ovals.

amongst their TM domains, and is thought to mediate dimerisation of this family of receptors[13]. An alignment of selected RTKs shows the prevalence of a small-x-x-x-small amino acid motif (Figure 1.3A), which is shown to be involved in dimerisation in six solved structures of TM helix dimers (Figure 1.3B). The first structure of a TM helix dimer was solved in 1997 by MacKenzie et al.[14] of the glycophorin A (GpA) dimer. Since then, Senes et al.[15] determined there was a prevalence of small-x-x-x-small motifs, especially when the 'small' amino acid was glycine. This corroborated the findings of the GpA dimer, and since then, there have been numerous structures

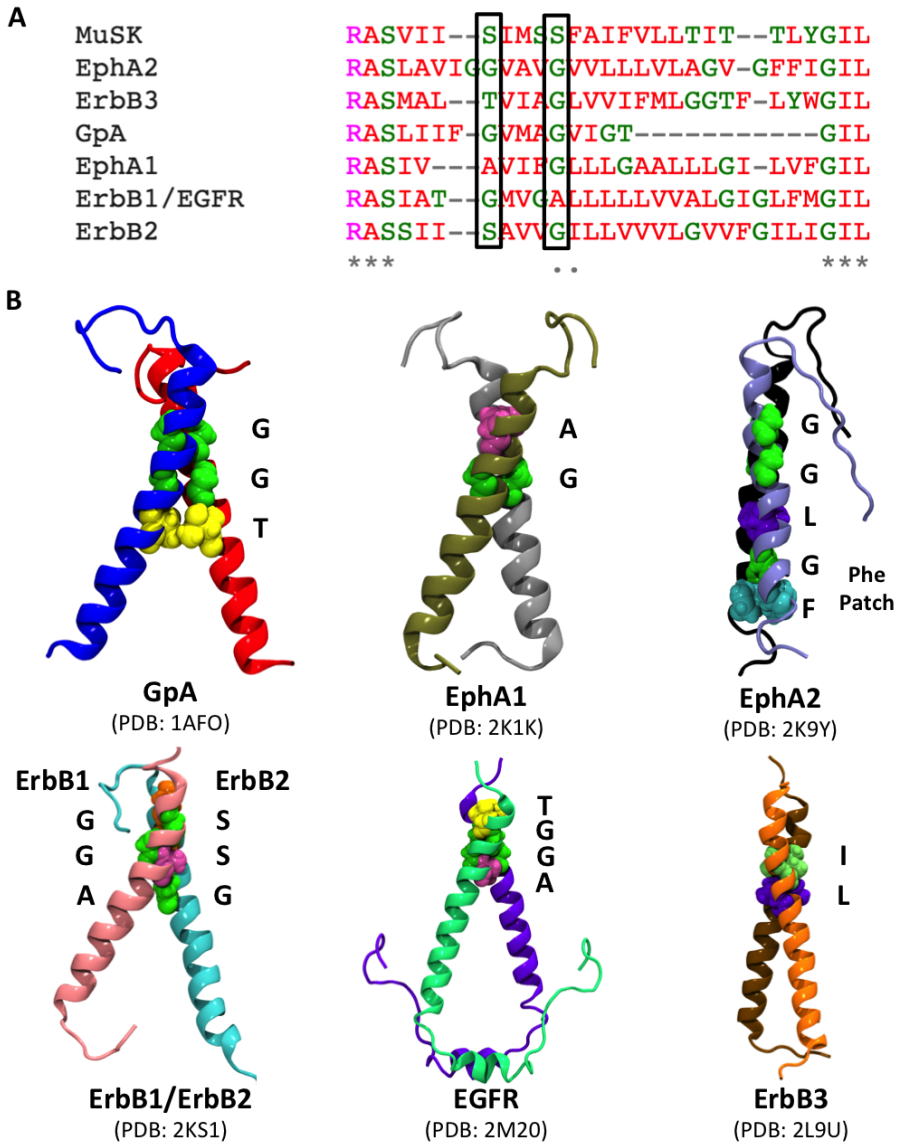

Figure 1.3: A. Sequence alignment of solved structures of TM helix dimers with Muscle-Specific Kinase (MuSK). The small-x-x-x-small amino acid motifs are highlighted by black boxes. All sequences shown except GpA are RTKs. The alignment was made using the EBI clustalW tool muscle (http://www.ebi.ac.uk/Tools/msa/muscle/). B. Solved structures of TM helix dimers. All are in new cartoon representation. Small-x-x-x-small amino acid motifs are in van der Waals (VDW) representation, with various colours for different types of amino acid residues.

of TM helix dimers solved, including EphA2[16], EGFR[17], and others[18–20]. These structures provide insight into how single-spanning transmembrane helices associate with each other, and potentially carry signals across the bilayer.

The structure of the TM helix + JM region of Epidermal Growth Factor Receptor

(EGFR) was solved in 2013[17], and a companion paper containing simulations of this structure[21] was also published, and attempted to answer some of the questions raised in the experimental paper, such as the role of lipids and the aforementioned small-x-x-x-small amino acid motif. Together, these papers have expanded on the activation mechanism of EGFR, and highlighted the complicated nature of RTK activation mechanisms. Both of these papers suggest that the TM helix plays a vital role in receptor activation, as evidenced by immunoflourescence experiments[17]. Upon mutation of the TM region interfacial residues, EGFR's activation is significantly inhibited, which is both verified experimentally and computationally[17;21]. An additional contributing factor of EGFR activation is the presence of anionic lipids in the membrane. Anionic lipids, specifically PIP_2[22], function to stabilise the active conformation. This observation is potentially because PIP_2 stabilises the antiparallel JM helix. The JM portion is also interacting with the membrane, which may explain the presence of so many positive charges in the JM region. These positive charges may provide relief through auto inhibition via electrostatic interactions with PIP_2, and are integral in EGFR's activation. This work highlights the importance of protein-lipid interaction, and the fact that they must also be taken into account when describing the signal transduction process.[23–26]

1.2 Role of Lipids

The plasma membrane (PM) is a complicated entity which contains a diverse range of proteins and lipids, and has also been found to be asymmetric[27] (Figure 1.4A). The outer leaflet of the plasma membrane contains mostly phosphotidylcholines (PC) and sphingolipids, whereas the inner leaflet contains mostly phosphatidylethanolamines (PE), phosphatidylserines (PS), and phosphoinositides (PI), with cholesterol filling in the rest of the composition in both leaflets[28;29]. Even though the composition and

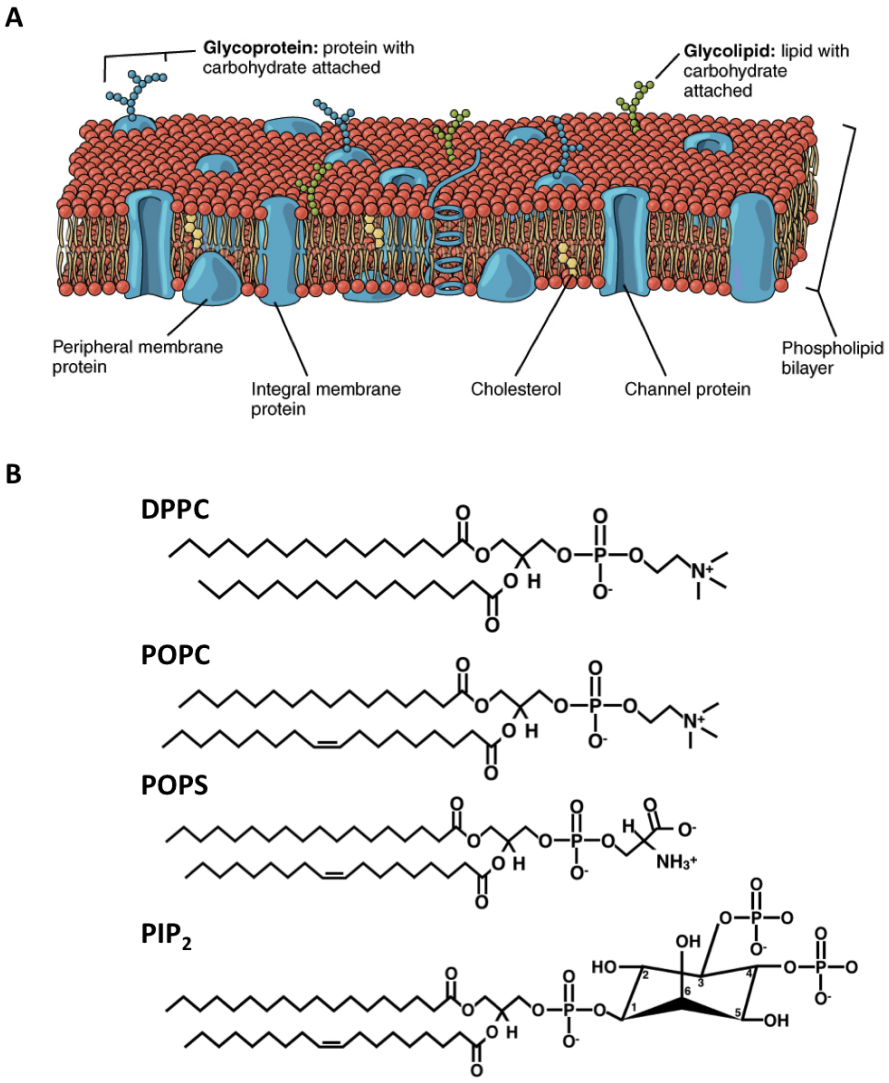

Figure 1.4: A. The phospholipid bilayer. It contains many different components, including integral membrane proteins, peripheral membrane proteins, lipids, and sugar moieties on specific components. (Figure courtesy of http://cnx.org/content/m46021/1.5/). B. Chemical structures of lipids used in the simulations in this thesis: DPPC (dipalmitoylphosphatidylcholine), POPC (1-palmitoyl, 2-oleoyl-sn-glycero-3-phosphocholine), POPS (1-palmitoyl-2-oleoyl-sn-glycerol-3-phosphoethanolserine), and PIP_2 (1,2-Diacyl-sn-glycero-3-phospho-(1-D-myo-inositol 4,5-bisphosphate)).

asymmetry vary between different cell types, the types of lipids in each membrane fall into three categories: phospholipids, sphingolipids and sterols. Phospholipids are lipids that have hydrophobic tails that are connected to a hydrophilic head by a glycerol group. This phosphate-nitrogen group makes that portion of the molecule hydrophilic. Phospholipids make up the bulk of the lipid bilayers *in vivo*. Exam-

ples of phospholipids are POPC or POPS (Figure 1.4B). Sphingolipids are fatty acid derivatives of sphingosine, and are involved in everything from apoptosis to differentiation, often in the form of sphingomyelin. Sterols are a subgroup of the steroids, characterised by ring structures. An example of a sterol is cholesterol. All lipids used in this thesis are in Figure 1.4B.

Due to the asymmetry of the plasma membrane, especially with the presence of PS, PIPs and cholesterol in the membrane[28;29], there is potential for the membrane to form nano- or micro-domains, or "lipid rafts."[29;30] This presents a more structured view of the plasma membrane, in which the lipids and proteins partition into two phases: the lipid disordered (L_D) and lipid ordered (L_O), adding a layer of complexity to the previously proposed "fluid mosaic model."[29–32] Experiments looking at micro-domains in isolated plasma membrane vesicles derived from mammalian cells, and also in model membranes, suggest how rafts may form in cells[33]. However, the precise nature and size of lipid rafts remains unknown[26;34;35], as any data gathered about lipid rafts have previously been limited by the diffraction limit of optical microscopy (∼200-300nm). Recently, new advances in super-resolution microscopy have been able to visualise lipid rafts between 5nm and 200nm in size[36]. It is believed that these nano-domains, or "lipid rafts," are fluid structures enriched in cholesterol, sphingomyelin and gangliosides[30], which would function as signalling hubs for the cell[28;37;38]. One of the places these signalling hubs are thought to be found in the neuromuscular junction[39].

1.3 MuSK's Origin, Role in NMJ and Functional Data

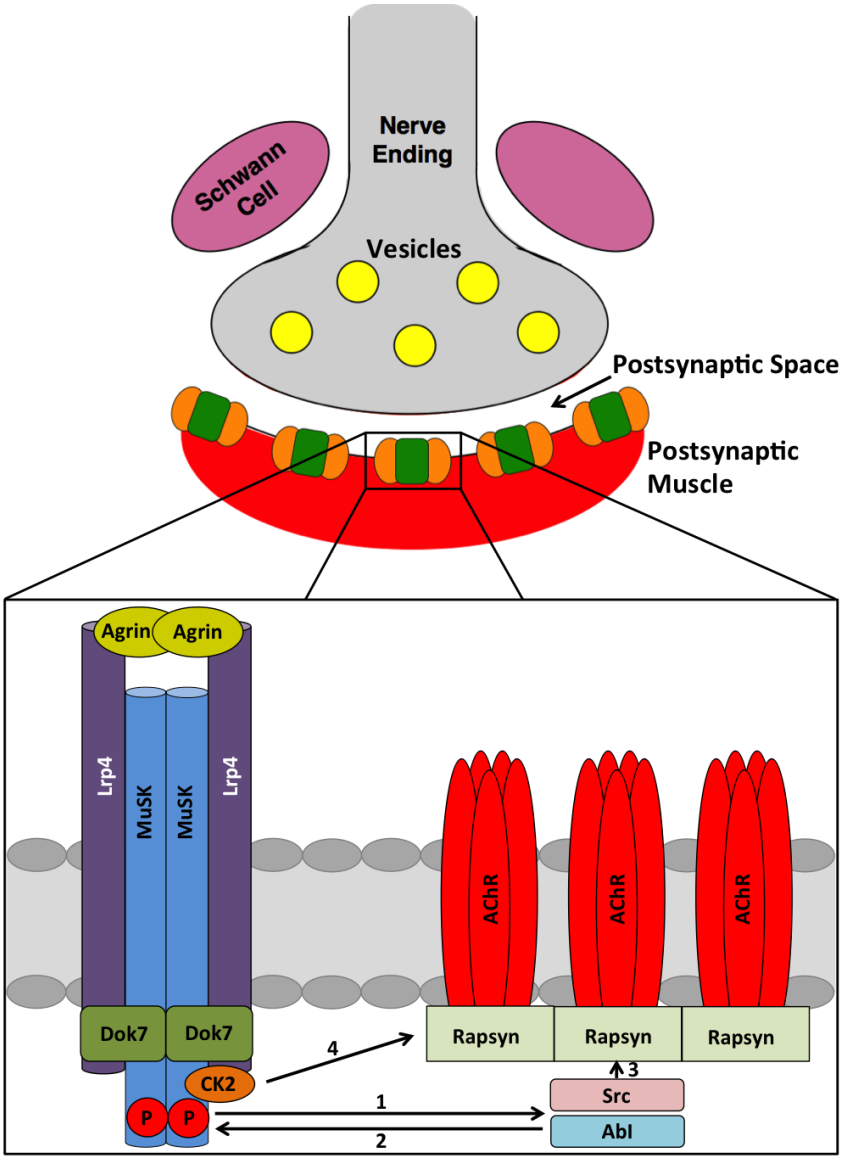

Figure 1.5: Schematic of the neuromuscular junction. The nerve ending is shown in grey on the top, Schwann cells are shown in pink ovals, vesicles carrying neurotransmitters are yellow circles, and the MuSK complex is shown in green boxes and orange ovals. The postsynaptic muscle is shown in a red line. MuSK (blue cylinders) is in complex with the neurotransmitter Agrin (dark yellow ovals), Lrp4 (purple cylinders) and Dok7 (green rounded rectangles). Dok7 recruits casein kinase 2 (CK2), dark orange oval), and continues a signalling cascade which ends by assembling the complex including Rapsyn(green rectangles), Src (pink rectangle), Abl (light blue rectangle), and acetylcholine receptors (red ovals). Figure adapted from Ghazanfari et al.[11]

The neuromuscular junction (NMJ) is a complicated entity, which functions to signal from the nerve ending to the muscle ending, resulting in muscle contraction (Figure 1.5). Vesicles carrying various neurotransmitters proceed to the cell wall of the nerve ending, whereby the vesicles fuse with the cell wall, causing neurotransmitters to be secreted across the postsynaptic space. These neurotransmitters bind to receptors residing in the membrane of the postsynaptic muscle, one of which is the MuSK complex (Figure 1.5, in box). It was previously suspected that there was a protein kinase involved in formation and maintenance of the neuromuscular junction. In 1993, a homologue of human Muscle-Specific Kinase (MuSK) was first discovered in the electric ray *Torpedo californa*, where it was enriched in electric organ and skeletal muscle[40]. This tissue in electric rays is homologous to human muscle tissue, and it was concluded that this RTK was also expressed in human muscle cells, though only the kinase domain was conserved between different species[40]. MuSK has since been implicated in mammalian NMJ formation and maintenance[41–43], as mice without MuSK die at birth from improperly formed neuromuscular junctions[41;44;45]. MuSK was hypothesised to be activated by a neurotransmitter, and glycoprotein, called agrin[46]. Interestingly, mice lacking agrin also die at birth, in a similar manner to mice lacking MuSK[47], which confirmed the hypothesis that MuSK is activated by agrin. Indeed, without MuSK, clustering of acetylcholine receptors (AChRs) is unable to take place[48], as agrin-dependent signalling is lost[43]. In addition, MuSK has been shown to co-localise with AChRs[49], which are critical for NMJ formation and maintenance. This suggested that MuSK actually may function to cluster AChRs, though the mechanism by which this happens was unknown at the time.

Even though it was known that MuSK and agrin are both necessary for AChR clustering, and that agrin activates MuSK, they, in fact, do not interact directly[48]. It was proposed that a hypothetical protein, deemed a myotube-associated specificity component (MASC), was the actual receptor for agrin, not MuSK. Later genetic studies

implicated a protein called low-density lipoprotein receptor-related 4 (Lrp4) in NMJ formation[50–52]. Mice without Lrp4 exhibited the same phenotypes as mice without MuSK[52]. Two years later, two independent studies confirmed that Lrp4 is indeed necessary for clustering of AChRs and formation of NMJs[53;54], as this is the sought-after component that acts as a bridge between neural agrin and MuSK. Interestingly, the full mechanism of how agrin and Lrp4 activate MuSK is still unknown[43].

Another important component of this pathway is a protein called Downstream-of-kinase 7 (Dok7). In 2006, this protein was found to be a unique member of the Dok family, and has been implicated in the activation of MuSK[55]. Mice without Dok7 also give similar phenotypes to Lrp4 and MuSK mutants[55]. Later, in 2010, a structure of the Dok7 dimer was solved, and was co-crystallised with some of the JM region of human MuSK, which will be discussed in section 1.6[56]. Together, these four proteins are the core of MuSK signalling to cluster acetylcholine receptors. All of this information can be consolidated into one pathway.

1.4 Activation pathway of MuSK and *myasthenia gravis*

After it is secreted across the postsynaptic space, Agrin binds to Lrp4. Lrp4, meanwhile, is sitting in an inactive heterodimer with MuSK. Upon binding of Agrin, two MuSK:Lrp4:Agrin trimers come together to form a 2:2:2 MuSK:Lrp4:Agrin complex. This renders MuSK partially active via a rearrangement or conformational change, leading to an autophosphorylation event at $pTyr_{553}$. This phosphorylation recruits the peripheral protein Downstream-of-Kinase 7 (Dok7), which induces further conformational changes in MuSK and results in autophosphorylation of three tyrosines on MuSK: Tyr_{750}, Tyr_{754} and Tyr_{755}(Figure 1.5, bottom, and Figure 1.6). This renders

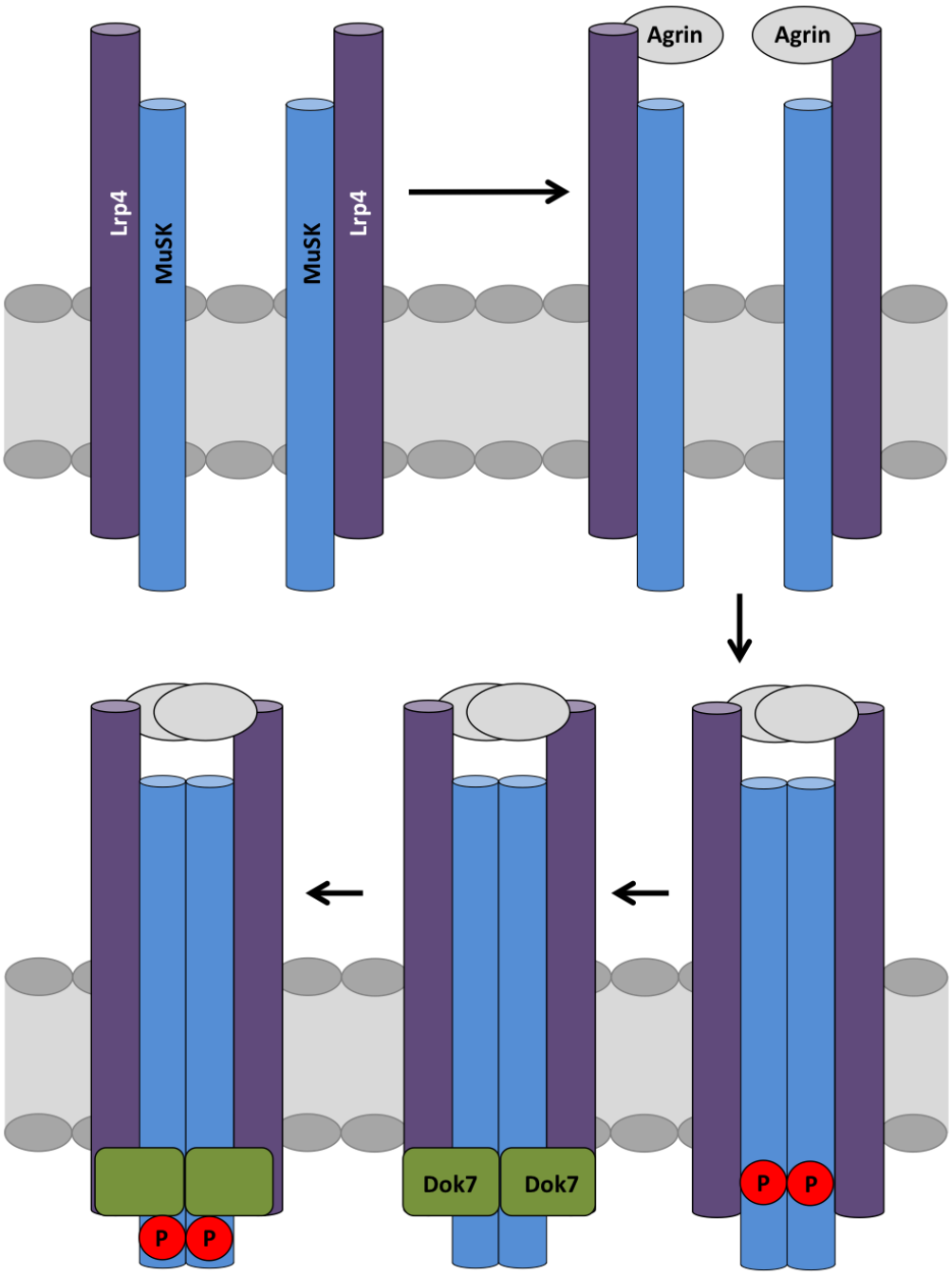

Figure 1.6: The assembly process of the MuSK complex. MuSK is represented by blue cylinders, Lrp4 by purple cylinders, agrin as grey ovals, Dok7 as green rounded rectangles, and phosphorylated tyrosines by red circles. The lipid bilayer is represented by grey ovals and rectangle.

MuSK fully activated (Figure 1.6) by disinhibiting the catalytic domain, and activates Abl and Src kinases. These kinases provide positive feedback for MuSK activation,

and are important in stabilising AChR clusters. This happens when Src and Abl phosphorylate residues on the AChR β-subunit, which recruits rapsyn for clustering AChRs. Casein Kinase 2 (CK2) phosphorylates serine residues on MuSK, which promotes cluster stability (Figure 1.5, bottom)[11;12]. Any disruption in this pathway leads to a disease called *myasthenia gravis*[57–63]. Symptoms of *myasthenia gravis* include muscle weakness, eye drooping, and fatigue[64]. Mutations in this disease mainly affect the Agrin:Lrp4:MuSK association[59;65] and the Dok7:MuSK association[66;67]. No mutations occurring in the general population are known to affect the association of Dok7 with the bilayer or MuSK with itself.

1.5 Structural Biology of MuSK

Three crystal structures exist for domains of the MuSK receptor: two Immunoglobulin-like (IG) domains, a frizzled cysteine-rich domain (Fz-CRD), and a kinase domain (Figure 1.7). However, these will only be discussed briefly, as they are not used in this thesis. There are no known structures for the MuSK TM Helix.

Ig1-Ig2

The first two immunoglobulin-like domains of rat MuSK were solved in 2006 by Stiegler et al[68]. They were determined to belong to the intermediate set (I-set) of the immunoglobulin superfamily, and were crystallised as a dimer. Their interface is proposed from the resulting crystal structure, and is similar to FGFR-1[69]. This interface is thought to be the dimerisation interface involved in MuSK signalling.

Fz-CRD

The Frizzled-Cysteine Rich Domain (Fz-CRD) was solved in 2009 by Stiegler et al.[70] The Cysteine Rich Domain was solved as a dimer, and was found to be similar to Fz8-CRDs and secreted FRP3-CRD proteins. The main difference between each monomer

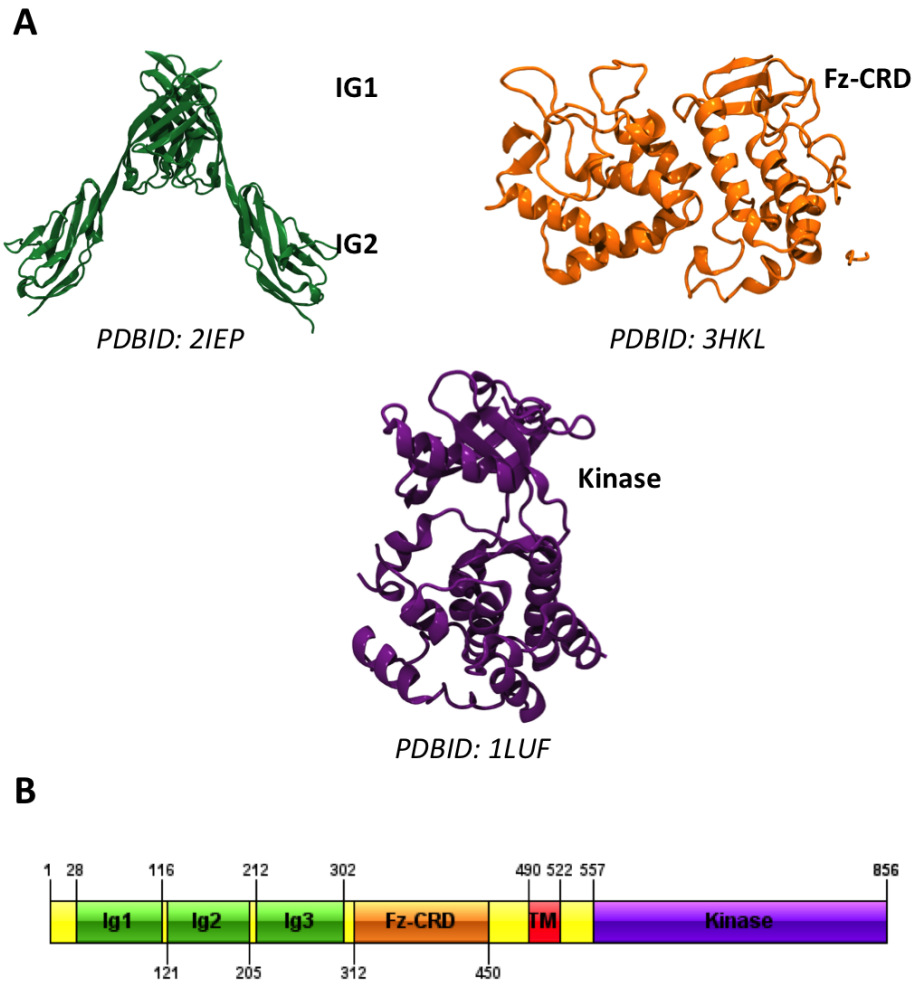

Figure 1.7: A. Crystal structures of different domains of MuSK. The Ig1-Ig2 domains (in green cartoon), the Fz-CRD (in orange cartoon), and the kinase domain (in purple cartoon) together total ∼70% of the receptor. B. Domain map of MuSK with residue numbers. Immunoglobulin-like domains (Ig) are in green, the Frizzled-Cysteine Rich domain (CRD) is in orange, the TM helix is in red, and the kinase domain is in purple.

in the structure is that one part of one monomer's structure is uncoiled, whereas the same part of the structure in the other monomer is α-helical. Whether this happens upon dimerisation is unknown.

Kinase Domain

The kinase domain was solved in 2002 by Till et al[71]. The crystal structure of the kinase domain was solved without its autophosphorylation loop (A loop), but the

mechanism proposed by the crystallographers is that the activation is through the A loop, as well as JM inhibition, most likely by $pTyr_{553}$ on the JM region.

1.6 Structural Biology of other proteins in MuSK complex

Three other proteins in the complex have had parts of their structures solved - Lrp4's β-propeller domain in complex with the third laminin G-like domain of Agrin, and the first 210 residues of Dok7.

Lrp4 in complex with Agrin

In 2008, it was determined that Lrp4 was the mediating factor in Agrin-mediated MuSK activation[53;54]. Not long after this, a crystal structure (Figure 1.8A) was determined of Lrp4 and Agrin in complex (Protein Data Bank (PDB) ID: 3V64 and 3V65)[72]. The first β propeller domain of Lrp4 was solved in complex with the G3 domains of Agrin. It was crystallised in a 2:2 Lrp4:Agrin complex, and the following mechanism was suggested: Agrin binds the Lrp4:MuSK heterodimer first via the z8 loop of Lrp4. Then, the Agrin binding induces this complex to dimerise, leading to the pathway of activation described in sections 1.3 and 1.4.

Dok7

As mentioned in section 1.3, Dok7 was found to be the protein that was responsible for fully activating MuSK by binding to $pTyr_{553}$ and inducing conformational changes that leads to further autophosphorylation of MuSK and activation[55]. It is a member of a family of adaptor proteins Dok1-7, and IRS1-4[73]. It also directly interacts with the cytoplasmic portion of MuSK and both domains are essential for MuSK activation in cells, especially the PH domain[55;74]. The first 210 residues of Dok7 have

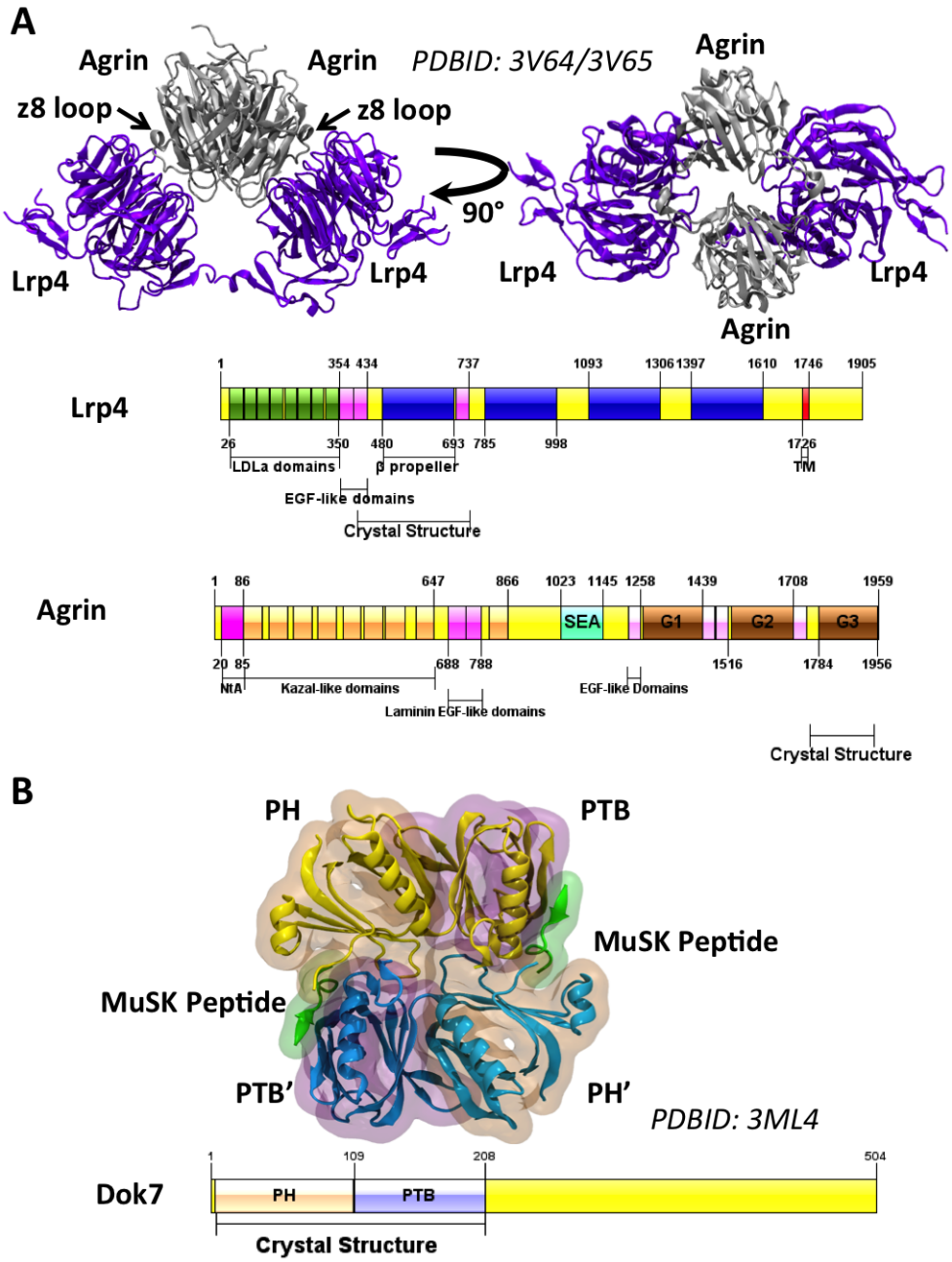

Figure 1.8: A. Crystal structure of Lrp4 bound to Agrin (pdb ids: 3V64 and 3V65). Agrin is in grey cartoon, while Lrp4 is in purple cartoon representation. In the domain maps, LDLa (LDL-receptor class A) domains are in green, EGF-like domains are in pink, β propeller domains are in blue, TM helices are in red, Nta domains are in dark pink, Kazal-like domains are in orange, SEA domains are in turquoise, and laminin G-like domains are in brown. Unstructured regions are in yellow. B. Crystal structure of Dok7 (pdb id: 3ML4). Each monomer is in blue or yellow cartoon representation, with each PH domain in orange surface, and each PTB domain in purple surface. Each MuSK peptide is in green cartoon and surface representation. The domain map has the PH domain as a light orange, the PTB domain as a light purple, and unstructured regions in yellow.

been crystallised (PDB ID 3ML4, Figure 1.8B)[56], with the remaining 300 residues remaining uncrystallised due to those domains being disordered. In Bergamin et al, mutations were done to determine that Dok7 associated with the unstructured JM region of MuSK via both its PH and Phosphotyrosine Binding (PTB) domain. Key mutations in these areas, including some that lead to the disease *myasthenia gravis*, affected association between Dok7 and MuSK, decreasing phosphorylation levels of MuSK. In addition, it was known that Dok7 interacts with lipids via its PH domain, though there is not much, if any, literature on this subject following Bergamin et al. The mechanism of association to the bilayer remained unknown, and will be explored in Chapter 5.

1.7 PH Domains

PH domains are a structurally conserved family of protein domains which have been shown to associate with phosphatidylinositol phosphates (PIPs)[75;76]. They are found in peripheral membrane proteins, which associate with the bilayer and often are involved in signalling cascades. A PH domain's structure consists of four anti-parallel β-strands, connected with varying lengths of loops, followed by a conserved fifth stretch of secondary structure, often a C-terminal helix[77;78]. In between the first two β-strands resides a positive loop that is responsible for binding PIPs[79]. Dok7's PH domain was shown to associate with PIPs via a mutation of a residue close to the aforementioned positive loop (R54)[56], and is comparable regions that have been shown to be important for the formation of the PH/PIP complexes for other PH domains[80;81]. Since the mutational data only gives the residues that interact with PIPs, but not the orientation on the bilayer, simulations are needed to figure out the mechanism of association of the protein to the membrane.

1.8 MD simulations of membrane proteins, including RTKs and other signalling complexes

Recently, the number of known membrane protein structures has increased as the techniques for solving crystal structures have improved and changed to accommodate membrane proteins (http://sbcb.bioch.ox.ac.uk/memprotmd/beta/). However, these structures, while necessary and informative, are snapshots of a protein, and in most cases are unable to capture the full range of dynamics that membrane proteins are capable of. In addition, the membrane protein structures that are available are often in unrealistic membrane environments[82;83]. Computational methods, specifically molecular dynamics (MD), are often ideal methods to bridge the gap between static structures and conformational changes. However, large conformational changes are usually not seen, as they happen on timescales that are not very attainable with classical atomistic molecular dynamics (AT-MD) simulations, even with advanced computational resources such as ANTON[21].

Despite this limitation, MD simulations have been able to provide some mechanistic insights that experiments would otherwise be unable to definitively answer. Simulations have been run on ANTON to describe a potential mechanism, as well as the importance of lipid composition, of EGFR activation. Recently, Arkhipov et al. have determined that the transmembrane domain has a crucial role in signalling, in addition to the lipids[21]. They have also simulated a model of the the full-length receptor with anionic lipids, to determine the effect of lipids on the receptor's behaviour. This has been extended by Kaszuba et al., who have done a similar thing with a more complex bilayer, and also with sugar moieties added onto the protein[84]. Both works have provided valuable insight into EGFR dynamics.

Coarse-grained Molecular Dynamics (CG-MD) simulations have been successfully

used on many systems, such as simulations on self-association of TM helices in the bilayer[85–88]. These simulations have provided insights into potential structures, both known and unknown, of these helices, as well as switching between states[89] as a potential mechanism of signalling. Free energy calculations can also be performed on these systems to determine the free energy of association of two helices[90]. Expanding on this, computational studies also have been done on full-length receptors in complex membranes[91] to determine the behaviour of the protein and the lipids. There have also been studies done of proteins, specifically PH domains, associating with the bilayer[92–96], as well as assembly of macromolecular complexes to determine the proposed mechanism of activation of a signalling system[93].

1.9 Thesis Overview

As mentioned in sections 1.3-1.7, various structural and functional data provide important insights into the MuSK pathway, both in forming NMJs and also maintaining them. The pathway of activation has just recently been determined, though despite the functional data available, the precise mechanism of MuSK activation is still unknown. For example, it is unknown how MuSK is fully activated by Dok7, or how the TM region homodimerises with itself, and also how the TM and JM region behaves in the presence of a complex bilayer. Therefore, this thesis aims to address these questions by providing insights into MuSK and its potential activation pathway. The structural data previously mentioned makes MuSK and Dok7 good targets for MD simulations. MD simulations are the main method used in this thesis to explore MuSK and Dok7's dynamics in a more biologically relevant environment.

In Chapter 2, the theory behind molecular dynamics simulation techniques is discussed. In Chapter 3, MuSK's dimerisation behaviour in bilayers which vary in lipid

composition is explored. In addition, the effect of various unstructured extensions on MuSK's helix has in lipid bilayers is also explored. These results demonstrate that the lipid composition of the bilayer has very little effect on the association of the MuSK TM helices, nor do the added unstructured extensions, suggesting something else is necessary for the activation mechanism. Additionally, Chapter 3 presents NMR data gathered of the MuSK TM Helix, as well as the jackknife method applied to TM helix dimer simulations to determine how many individual simulations are actually needed for convergence of selected properties of a system.

In Chapter 4, Dok7's association with a PIP-containing bilayer is explored. Chapter 4 demonstrates that the PH domain is the main driving force behind association of Dok7 with the bilayer, which is consistent with previous experimental and simulation-based evidence of PH domains. This remains consistent upon mutation of the key residues binding PIP lipids, and suggests that the PH domain drives Dok7 to associate with the membrane, while the PTB domain is responsible for another part of the signalling process, mainly interacting with MuSK's JM region for further downstream signalling.

In Chapter 5, the Dok7/MuSK complex is built and simulated in a complex bilayer. Together with Chapters 3 and 4, a potential structure of Dok7 with the MuSK TM+JM region is determined, and a potential mechanism is provided for the (in)activation of MuSK by Dok7 with insights from the MuSK TM helix dimer dynamics.

The outcome of this thesis is important from a biological perspective, in that it is possible to suggest a model and mechanism of signalling from individual crystal structures. Overall, this thesis provides information about a unique member of the RTK family, which may give important insight into other activation mechanisms of other members of the RTK family.

2

Simulation Theory and Methods

This chapter aims to discuss the methods used in this thesis. The primary method used is Molecular Dynamics (MD) simulations, for which a system of particles is prepared, and dynamics are applied to the system to evolve the positions and energies of the system over a timestep t until the target time is reached. Also included are summaries of other modelling and computational techniques employed in subsequent chapters.

2.1 Equations of Motion and Computational Implementation

It is important to consider the level of atomic detail needed when simulating biological systems. Ideally, systems could be simulated with a very high level of detail, down to the quantum level. Quantum Mechanics (QM) calculations have the ability to attain this detail, as they take into account all of the atoms' movements, most importantly the movement of electrons. However, these calculations are computationally very expensive, and can really only be used for ligands with a small amount of amino acids (5-6, totalling to ∼1,500 atoms)[97] surrounding the ligand. To overcome this barrier, classical MD simulations are employed. Classical MD simulations use the Born-Oppenheimer approximation, which allows all atoms to be represented as a set of Cartesian coordinates, as it assumes that the position of the electrons surrounding the nucleus of the atom will react instantaneously to any change in the nucleus[98], saving computational time.

To describe the movement of atoms in a classical system, Newton's equation of motion are used to evolve the system over time. In the system's initial configuration, there are initial masses (m_i) for every atom and positions of all the atoms (r_i). What is needed is the force acting on a particular atom, as the forces evolve the system. To do this, Newton's first equation of motion, which relates the force F_i acting on atom i to the atoms' mass m_i and acceleration $a_i(t)$:

$$F_i = m_i a_i(t) \tag{2.1}$$

This can be rearranged to obtain

$$\frac{\partial^2 r_i}{\partial t^2} = \frac{F_{x_i}}{m_i} \tag{2.2}$$

which relates the positions of each atoms to its force and mass. r_i is the position of atom i, and the motion of an atom with mass m_i along a coordinate x_i is given by the force F_{x_i} acting in the x direction. The force F_i on atom i is calculated in the following manner:

$$F_i = -\frac{\partial V}{\partial r_i} \tag{2.3}$$

where r_i is the position of atom i, and V is a potential energy function.

As it is not possible to solve these equations analytically, numerical methods have been developed to solve differential equations with a good degree of accuracy and a small amount of error. GROMACS[99] uses the Verlet method, which is also known as the 'Leapfrog Algorithm.'[100] Rather than calculating velocities at full-time intervals, it calculates velocities at half-time intervals, which is used to calculate the positions of each atom at half-time intervals (Equation 2.4 and 2.5). Both of these quantities use the forces on the atoms previously described in Equation 2.3.

$$r_i(t + \frac{\Delta t}{2}) = r_i(t) + v_i(t + \frac{\Delta t}{2})\Delta t \tag{2.4}$$

$$v_i(t + \frac{\Delta t}{2}) = v_i(t - \frac{\Delta t}{2} + \frac{F_i(t)}{m_i})\Delta t \tag{2.5}$$

Δt is the timestep given for the simulations.

2.2 Force Fields

For classical Molecular Dynamics, the forcefield is a set of parameters for and interactions between all atoms present in a particular system[98]. These parameters contribute to the overall energy of the system($U(r^N)$), which can be described in both kinetic (U^{bonded}) and potential ($U^{nonbonded}$) energy terms:

$$U(r^N) = \sum U^{bonded} + \sum U^{nonbonded} \tag{2.6}$$

where r is the position of all of the particles in the system, and N is the number of particles in the system.

2.2.1 Bonded Interactions

$$\sum U^{bonded} = \sum U^{bonds} + \sum U^{angles} + \sum U^{dihedrals} + \sum U^{impropers} \tag{2.7}$$

The bonded interactions in a system describe the kinetic energy of a system, in that there is energy shared between two atoms. The energy contribution of all bonded interactions in the system is the sum of all of the energy terms associated with the bond, angle, dihedral and improper dihedral angles (Equation 2.7 and Figure 2.1).

All of the terms in Equation 2.7 are described by Equations 2.8-2.11. Equation 2.8 and Figure 2.1A describes the energy contribution for all of the bonds in the system:

$$\sum U^{bonds} = \sum_{bonds} \frac{1}{2} k_b (b - b_0)^2 \tag{2.8}$$

where k_b is the force constant, b is the radius of the bond, and b_0 is the equilibrium bond distance (usually obtained from a crystal structure or an idealised model). k_b

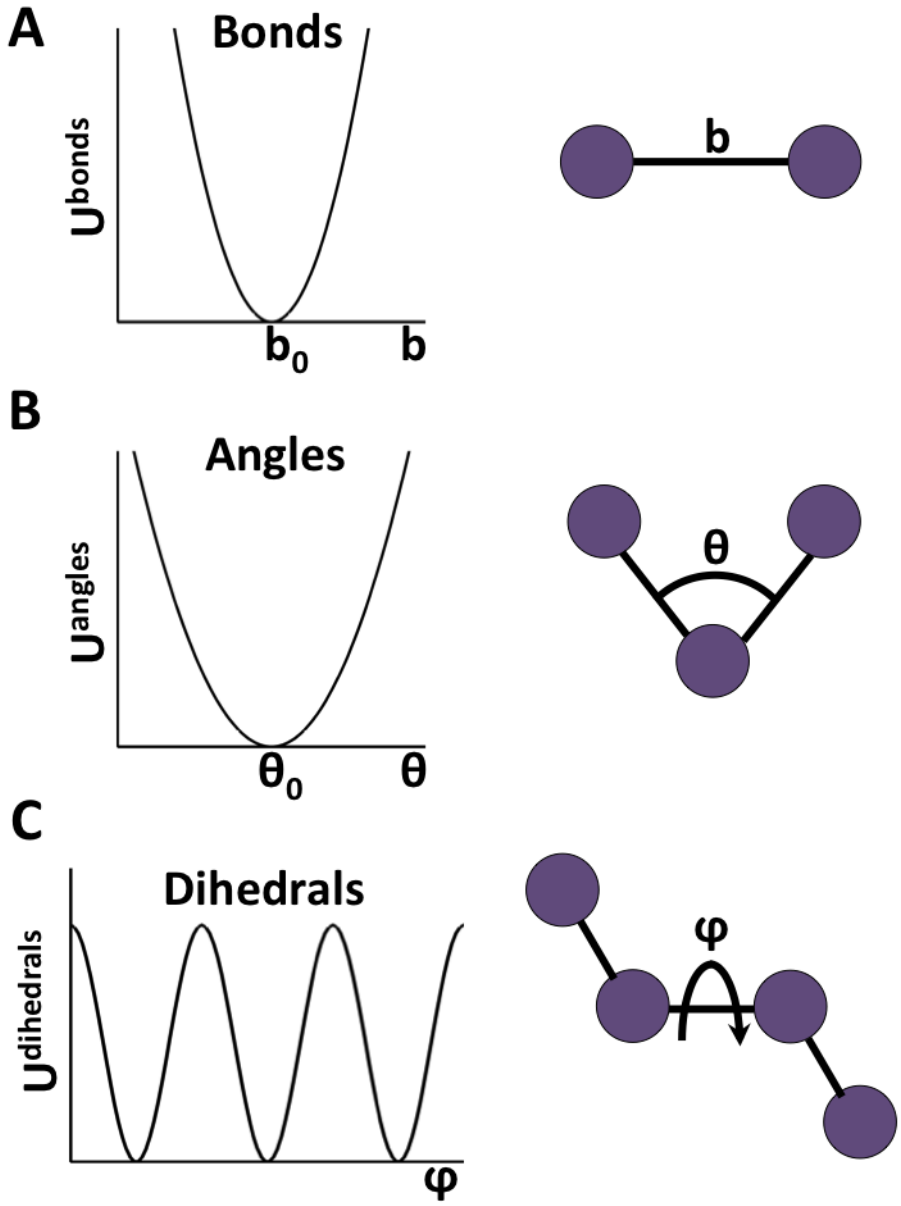

Figure 2.1: The different types of bonded interactions from Equation 2.7. The first interaction is the bond between two atoms (A), then the angles between three atoms (B), and the dihedral angles formed between four atoms (C). The individual atoms are purple spheres, and the bonds are solid black lines.

is determined from IR spectroscopy[98].

The energy contribution for all angles is described by Equation 2.9 and Figure 2.1B:

$$\sum U^{angles} = \sum_{angles} \frac{1}{2} K_\theta [\theta - \theta_0]^2 \tag{2.9}$$

where K_θ is the force constant of the angle, θ is the bond angle, and θ_0 is the equilibrium bond angle. Finally, the energy contribution for both proper and improper dihedrals are given by Equations 2.10 and 2.11, with Equation 2.10 illustrated in Figure 2.1C:

$$\sum U^{dihedrals} = \sum_{dihedrals} K_\varphi(1 + cos(n - n_0)) \tag{2.10}$$

$$\sum U^{impropers} = \sum_{impropers} \frac{1}{2} K_\omega(\omega - \omega_0)^2 \tag{2.11}$$

where K_φ is the force constant, n is the current angle, n_0 is the minimum value of $U^{dihedrals}$, K_ω is the value of the force constant, ω is the value of the angle, and ω_0 is the most favourable value of the angle.

2.2.2 Non-bonded interactions

The non-bonded interactions describe the potential energy in a system, as two particles that are not bonded contain energy that would otherwise be shared between them via a bond. These interactions consist of both Lennard-Jones and electrostatic potentials (Equation 2.12 and Figure 2.2):

$$U^{nonbonded} = U^{LJ} + U^{electrostatics} \tag{2.12}$$

The Lennard-Jones potential, is defined by Equation 2.13 and Figure 2.2A:

$$U^{LJ} = \sum_{i,j} 4\epsilon_{i,j} \left[\left(\frac{c_{ij}}{r_{ij}} \right)^{12} - \left(\frac{c_{ij}}{r_{ij}} \right)^{6} \right] \tag{2.13}$$

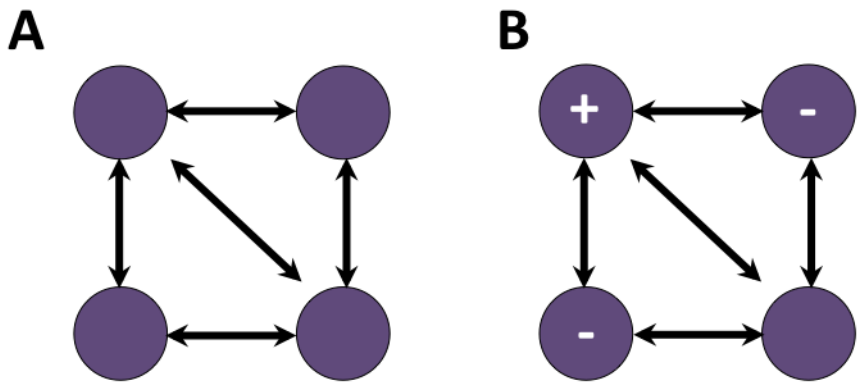

Figure 2.2: The two types of interactions that contribute to nonbonded interactions: Lennard-Jones (A) and electrostatic (B). Individual atoms are purple spheres, interactions are shown by solid black arrows, and charges are white text.

where $\epsilon_{i,j}$ is the measure of the attraction between atoms i and j, c_{ij} is the 'collision distance,' or the distance at which the energy between two particles is equal to zero. r_{ij} is the distance between particles i and j. The $(\frac{c_{ij}}{r_{ij}})^{12}$ term represents the short-range repulsion forces inherent in the system, such as forces described by the Pauli Exclusion Principle, and the $(\frac{c_{ij}}{r_{ij}})^{6}$ term represents the attractive forces inherent in the system, such as London dispersion forces.

The electrostatics term is Coulombs law (Figure 2.2B and Equation 2.14):

$$U^{electrostatics} = \sum_{i,j} \frac{q_i q_j}{4\pi\epsilon r_{ij}} \tag{2.14}$$

where q_i and q_j represent the partial charges of particles i and j, ϵ is the permittivity of free space, and r_{ij} is the distance between the two particles. q_i and q_j are determined from quantum mechanics calculations[98].

2.2.3 All-Atom Forcefields

In MD simulations, atoms can be explicitly represented and simulated. This is computationally expensive and time consuming, but is advantageous due to the system inherently being high resolution. Two examples of all-atom forcefields are OPLS[101;102] and CHARMM[103]. Conversely, united-atom forcefields are more commonly used, as they do not explicitly take into account the non-polar hydrogen atoms, which are more computationally expensive. This reduces the amount of calculations further, enabling an even longer timescale to be reached. An example of a united-atom forcefield is GROMOS96, which has been implemented in GROMACS[99].

2.2.4 Coarse-Grained Forcefields

Coarse-grained (CG) forcefields are used in addition to all-atom forcefields for molecular dynamics simulations. Simulations run with coarse-grained forcefields are able to achieve larger, even experimental-length scales, as well as much longer timescales[104]. This is due to simplification of the system, in that approximately four non-hydrogen atomistic particles are combined into one large coarse-grained particle (Figure 2.3). This 4:1 mapping reduces the number of particles in the system by a factor of 4, which in turn simplifies the calculations for each timestep and increases the speed of calculation of a CG-MD simulation by up to a factor of three compared to an AT simulation[105]. Most simulations run in this thesis are in the MARTINI (v2.1 and 2.2) forcefield, and some are in the BOND[86] forcefield, which is a local modified version of the MARTINI force field.

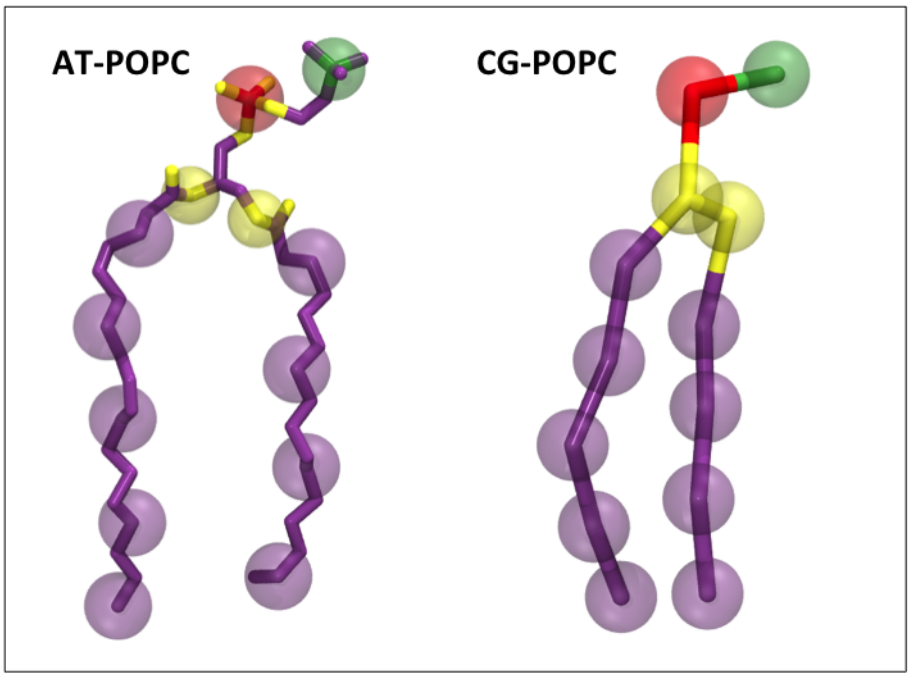

Figure 2.3: An atomistic (left) and coarse-grained (right) representation of a POPC lipid. Nonpolar groups are in green, charged groups are in red, polar groups are in yellow, and apolar groups are in purple. The bonds between the atoms are licorice representation, and coarse-grained particles are spheres.

2.2.5 BOND vs. MARTINI forcefields

There are two coarse-grained forcefields used in this thesis: the MARTINI forcefield, and the BOND forcefield. The MARTINI forcefield was originally developed by Marrink et al. for lipid and surfactant studies[106], but has been extended to biomolecules to study dynamics of biological systems. The BOND forcefield, is a local modification of the original MARTINI parameters[86]. Both force fields have four main particles:

- Polar (P) - Polar particles are neutral, water-soluble groups.
- Nonpolar (N) - Nonpolar particles represent neutral groups of atoms which are partially polar and partially apolar.
- Apolar (C) - Apolar particles are used for hydrophobic groups.
- Charged (Q) - Charged particles are for ionised groups.

In addition, charged (Q) and non polar (N) groups have four different subtypes, depending if they are neutral (0), are either a hydrogen donor or acceptor (d/a), or are both a hydrogen donor and acceptor (da). Thus, a non-polar group of atoms with no hydrogen bonding capabilities, would have the bead type N0. MARTINI further expands on this principle by dividing the P (polar) and C (apolar) into different subtypes according to their polarity, with I being the lowest polarity and V being the highest degree of polarity.

Table 2.1: CG Particle Types for both MARTINI and BOND forcefields.

Residue	Side chain particles	
	MARTINI	BOND
Hydrophobic		
Ala	C	C
Ile	C1	C
Leu	C1	C
Pro	C2	C
Val	C2	C
Cys	C5	N0
Met	C5	N0
Charged		
Glu	Qa, P1 (uncharged)	Qa (-1)
Asp	Qa, P3 (uncharged)	Qa (-1)
Lys	C3-Qd, C3-P1 (uncharged)	C+Qd
Arg	N0-Qd, N0-P4 (uncharged)	Nd+Qd
Aromatic		
Phe	SC4-SC4-SC4	C+C
Trp	SC4-SP1-SC4-SC4	Nd+C
His	SC4-SP1-SP1	C+Nda
Tyr	SC4-SC4-SP1	C+P
Polar uncharged		
Asn	P5	P
Gln	P4	P
Ser	P1	Nda
Thr	P1	Nda

Backbone atoms in both forcefields are denoted by a single CG particle, but the side chains are treated differently (Figure 2.4). In MARTINI, between zero and four CG particles are used for the side chains, whilst in the BOND forcefield, a maximum of

two CG beads are used per side chain. This mainly affects side chains with rings, like phenylalanine. The difference between each amino acid particle types and subtypes are in Table 2.1.

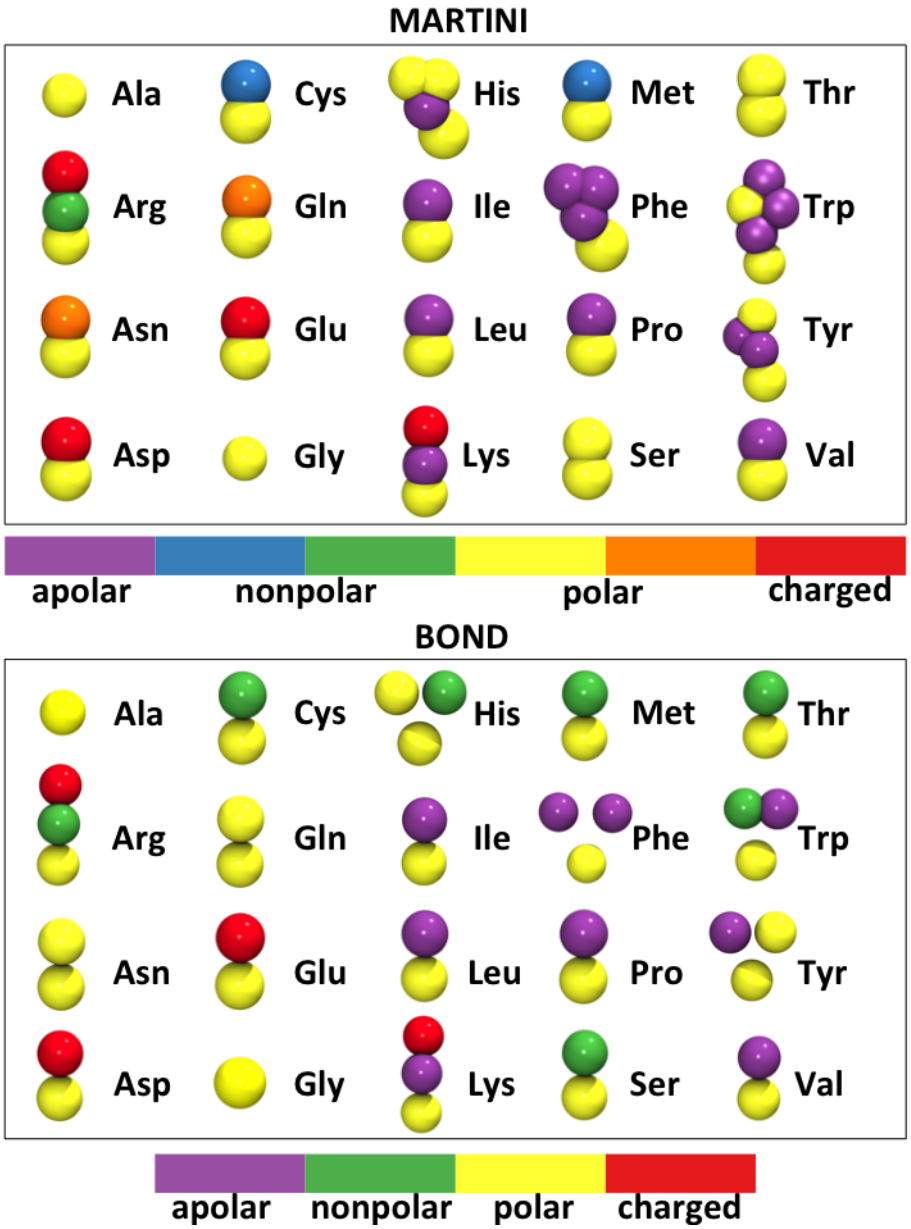

Figure 2.4: Coarse-grained representations of all 20 amino acids in the MARTINI (top) and BOND (bottom) CG forcefields. All of the amino acid particles are spheres, and coloured to match their bead type. Bead types are specified by the colour bar at the bottom of each box, with the MARTINI particles having two colours each for nonpolar and polar particles (blue and green for nonpolar, yellow and orange for polar). The top half of figure is based on Monticelli et al.[107], and the bottom half of the figure was created as a comparison to the above.

Non-bonded interactions

Non-bonded interactions are implemented in a similar manner in both forcefields. Electrostatics are represented as a coulomb potential, and all of the non-Coulombic interactions in the system are combined into a term similar to a Lennard-Jones-like potential, which is parameterised against atomistic simulations. BOND uses five different levels of strength of the non-bonded interactions (I=attractive, II=semi-attractive, III=intermediate, IV=semi-repulsive, V=repulsive). MARTINI uses a similar scale, except instead of five different levels of strength, it goes from O (no interactions) to IX (super repulsive) (Table 2.2).

Table 2.2: Table of MARTINI Particle Interactions

		Q				P					N				C				
	sub	da	d	a	0	5	4	3	2	1	da	d	a	0	5	4	3	2	1
Q	da	O	O	O	II	O	O	O	I	I	I	I	I	IV	V	VI	VII	IX	IX
	d	O	I	O	II	O	O	O	I	I	I	III	I	IV	V	VI	VII	IX	IX
	a	O	O	I	II	O	O	O	I	I	I	I	III	IV	V	VI	VII	IX	IX
	0	II	II	II	IV	I	O	I	II	III	III	III	III	IV	V	VI	VII	IX	IX
P	5	O	O	O	I	O	O	O	O	O	I	I	I	IV	V	VI	VI	VII	VIII
	4	O	O	O	O	O	I	I	II	II	III	III	III	IV	V	VI	VI	VII	VIII
	3	O	O	O	I	O	I	I	II	II	II	II	II	IV	IV	V	V	VI	VII
	2	I	I	I	II	O	II	II	II	II	II	II	II	III	IV	IV	V	VI	VII
	1	I	I	I	III	O	II	II	II	II	II	II	II	III	IV	IV	IV	V	VI
N	da	I	I	I	III	I	III	II	II	II	II	II	II	IV	IV	V	VI	VI	VI
	d	I	III	I	III	I	III	II	II	II	II	IIII	II	IV	IV	V	VI	VI	VI
	a	I	I	III	III	I	III	II	II	II	II	II	III	IV	IV	V	VI	VI	VI
	0	IV	IV	IV	IV	IV	IV	IV	III	III	IV	IV	IV	IV	IV	IV	IV	V	VI
C	5	V	V	V	V	V	V	IV	IV	IV	IV	IV	IV	IV	IV	IV	IV	V	V
	4	VI	VI	VI	VI	VI	VI	V	IV	IV	V	V	V	IV	IV	IV	IV	V	V
	3	VII	VII	VII	VII	VI	VI	V	V	IV	VI	VI	VI	IV	IV	IV	IV	IV	IV
	2	IX	IX	IX	IX	VII	VII	VI	VI	V	VI	VI	VI	V	V	V	IV	IV	IV
	1	IX	IX	IX	IX	VIII	VIII	VII	VI	VI	VI	VI	VI	VI	V	V	IV	IV	IV

*Level of interaction indicates the well depth in the LJ potential: O, ϵ=5.6 kJ/mol; I, ϵ=5.0 kJ/mol; II, ϵ=4.5 kJ/mol; III, ϵ=4.0 kJ/mol; IV, ϵ=3.5 kJ/mol; V, ϵ=3.1 kJ/mol; VI, ϵ=2.7 kJ/mol; VII, ϵ=2.3 kJ/mol; VIII, ϵ=2.0 kJ/mol; IX, ϵ=2.0 kJ/mol. The LH parameter σ=0.47nm for all interaction levels except level IX for which σ=0.62nm. Four different CG sites are considered: charged (Q), polar (P), non polar (N) and apolar (C). Subscripts are used to further distinguish groups with different chemical nature: 0, no hydrogen-bonding capabilities are present; d, groups acting as hydrogen bond donor; a, groups acting as hydrogen bond acceptor; da, groups with both donor and acceptor options; 1-5, indicating increasing polar affinity.

Elastic Network Model

Elastic Network Models (ENM) were introduced to describe motion within a protein, specifically conformational changes, as an alternative to normal mode analysis[108–110]ĖNMs describe a structure as a network of point masses connected to one another with springs, only when the distance between the particles is less than a specified cutoff distance, R_C. For most systems, the force of the spring, K_{spring}, along with the cutoff distance R_C, remains the same for all particles over the network of points.

Elastic Network Model As Applied to Proteins in MARTINI

After the initial MARTINI model was released, protein structure parameters were also released. The initial forcefield extension to proteins is MARTINIv2.1[107]. This uses an elastic network model[111] to restrain protein structure as described in the above paragraph, so there is little deviation of the protein's native structure throughout the simulation. This curtails any large conformational changes or unwanted folding events.

MARTINIv.2.2

A new version of the MARTINI forcefield (MARTINIv2.2)[112] was recently released, with improved parameters for the protein part of the forcefield. It aims to correct some inaccuracies of the previous forcefield, in that numerous side chain parameters were changed to more closely match atomistic simulations. The most dramatic changes were in side chains containing rings, along with backbone-backbone bonds. These changes to the forcefield resulted in a model that is more closely related to and more applicable to membrane proteins, as compared to atomistic simulations.

2.2.6 CG2AT Conversion and AT Simulation

CG simulations are a powerful tool used to capture macroscopic properties of a system, such as oligomerisation and association of a protein to the bilayer. However, due to the simplified representation, they are not able to accurately capture the microscopic properties of a system, such as the movement of side chains and conformational changes. To assess these microscopic properties, interesting phenomena that are observed in coarse-grain can also be observed in atomistic simulations using coarse-grain to atomistic conversion approaches. The CG2AT program (written in perl and developed by Dr. Phillip Stansfield)[113] was used in all results chapters (Chapters 3-5) to convert CG generated models to an atomistic representation. The CG particles in the models were converted to atomistic using a fragment-based approach.

Steps in the CG2AT protocol:

1. Rename all CG lipids and protein particles to their atomistic counterpart
2. Proteins - reconstruction of the AT coordinates from the CG model using Pulchra[114] and MODELLER[115]
3. Lipids - alignment of AT lipid fragments to the CG particle of each lipid molecule
4. Energy minimisation of protein and lipids individually
5. Protein and lipid systems are combined and energy minimised further
6. System is then solvated, and ions are added to neutralise the system
7. Energy minimisation of solvated system
8. System is equilibrated for ins with the protein C_α atoms restrained (force constant = 10 kJ/mol/Å^2)
9. Restraints are released for production run

Energy minimisation and equilibration are necessary to prevent dramatic movements of regions in the system that could potentially cause instability. Production runs were performed on different initial configurations and different starting velocities.

2.3 Simulation Techniques

2.3.1 Timesteps and Constraints

Time steps and constraints are important factors in a simulation, as having the right time step and constraints present prevents a system from becoming unstable. The generally suggested method of choosing a timestep is selecting one 10 times smaller than the fastest movement inherent in the system. For example, in a fully atomistic system, the vibration of a C-H bond is generally the quickest motion, which is on the order of 1 to 2fs, so the ideal timestep for atomistic simulations is 0.1-0.2fs. In a coarse-grained system, however, since the fastest motion is slower, bigger time steps can be employed, resulting the ability to reach much longer timescales. This helps prevent crashing and accummulation of errors. Constraints are used during an MD simulation to constrain bond lengths such that they satisfy a given reference length. Examples of constraint algorithms are the LINCS[116] and SHAKE[117] algorithms. In this thesis, the LINCS algorithm was used to constrain bond length in atomistic simulations. The LINCS algorithm eliminates bond vibration, fixing bonds at a particular length, so as to reduce time for calculating flexibility between molecules.

2.3.2 Periodicity and Periodic Boundary Conditions

Periodicity must be taken into account in most biological simulations. At a certain size, a box of particles becomes too large for efficient calculations. This means there

must be a cutoff at the edge of the box to compensate. Simulations of a finite system will be prone to artefacts due to interactions of particles with the boundary (box) of the system. There are a number of ways to overcome these artefacts. One is the imposition of Periodic Boundary Conditions (PBCs) on the system, such that the system is treated as a lattice, with infinite copies of itself spreading out in all directions. Periodic Boundary Conditions can be used with numerous box shapes, with the most common being a cubic system, shown with periodicity in Figure 2.5.

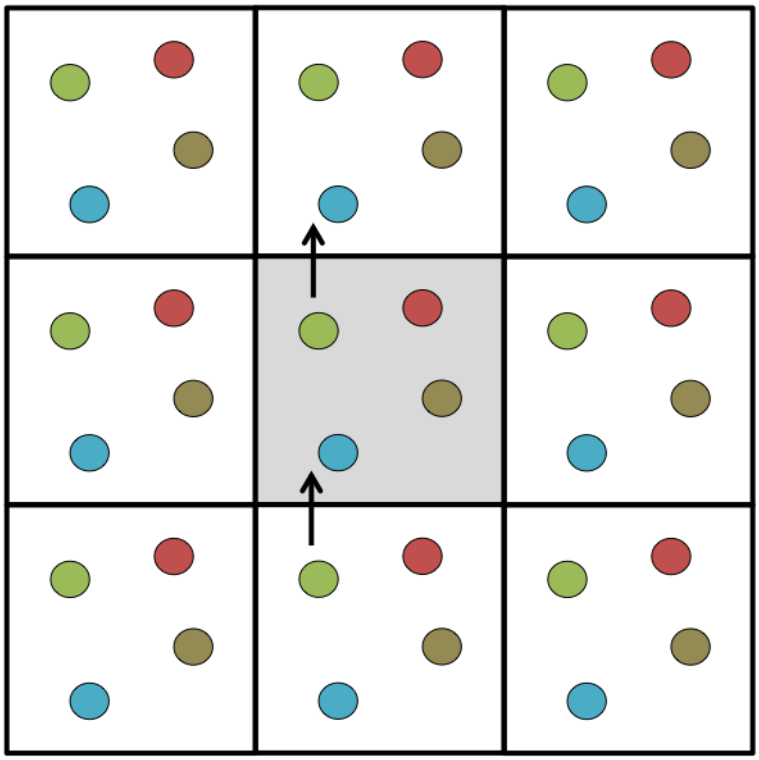

Figure 2.5: Periodic Boundary Conditions. The red, green, yellow and blue circles are particles, the grey box is the primary image, and the white boxes are periodic images of the primary image.

In a system with PBCs, particles are able to pass through the periodic boundaries and will appear on the opposite side of the simulation box. Interactions also can take place across the periodic boundaries. Since the system is stretched infinitely, having suitable cut-offs and schemes such as the Particle Mesh Ewald (PME) method is crucial.

2.3.3 Particle Mesh Ewald

The Particle Mesh Ewald method (PME) is an electrostatic summation method, which aims to speed up calculations of long range electrostatic interactions between particles[118]. Long-range electrostatics require lots of computational power, due to the sheer number of calculations required, and the number of calculations increases exponentially (on the order of $O(N^2)$) as the number of particles in the system increases. This is due to the electrostatic calculations. To prevent loads of unnecessary calculations, an interaction cutoff can be introduced to prevent too many long-range electrostatics calculations. However, due to the small rate of decay of the potential, introducing a cut-off would introduce discontinuity in the system, affecting its accuracy. The PME method aims to reduce the number of electrostatic calculations by splitting the calculation into two parts: a short-range potential and a long-range potential. The short-range potential converges very quickly in real space, and the long-range potential converges slowly. However, if the long-range potential is transformed into Fourier space, it converges very quickly. Because the two components converge very quickly, albeit in their respective spaces, then truncation of the calculations is possible, increasing the efficiency with only a small decrease in accuracy.

2.3.4 Temperature and Pressure Coupling

All simulations in this thesis were conducted under NPT conditions, in which the number of particles (N), the pressure (P), and the temperature (T) were fixed. During NPT simulations, the pressure and temperature are fixed using a barostat and a thermostat, respectively, to prevent any large fluctuations in pressure and temperature. By fixing the temperature and pressure in the simulation, this gives simulations the ability to more closely match experimental conditions. Numerous barostat and thermostat algorithms are available, such as stochastic collisions[119], where a particle

is randomly chosen for velocity reassignment (a similar principle to the Berendsen thermostat and barostat), and an extended system[120;121], which couples the system to a thermal reservoir, so there is flow between the system and the reservoir.

All simulations in this thesis used the Berendsen thermostat and barostat.

2.3.5 The Berendsen Thermostat

The Berendsen thermostat[122] relates the temperature of the system to its particles' energy by coupling the system to an external heat bath where the temperature is fixed at the required value. The velocities are rescaled at each timestep by an amount proportional to the difference between the reference temperature, T_{ref}, and the temperature of the system, T:

$$\frac{dT(t)}{dt} = \frac{1}{\tau_T}(T_{ref} - T) \tag{2.15}$$

τ_T represents the coupling parameter between temperature and pressure for a particular component of a system. Typically, in a simulation, proteins are coupled separately from each lipid type and the surrounding solvent. This is to prevent simulations where the overall temperature of the system is correct, but each component would vary wildly; for example, if the protein would increase in temperature, another component, the solvent, would decrease in temperature to compensate.

2.3.6 The Berendsen Barostat

The Berendsen barostat[122] works in a similar way to the Berendsen thermostat by having an external "pressure bath," as a sort of a reference pressure. This way, the box size can be rescaled to reflect the pressure in the system. The rate of change is

essentially the same for pressure coupling as with the temperature coupling:

$$\frac{dP(t)}{dt} = \frac{1}{\tau_P}(P_{ref} - P) \tag{2.16}$$

where P_{ref} is the reference pressure, P is the pressure of the system, and τ_P is the coupling parameter.

2.3.7 Energy Minimisation

Energy minimisation is an essential step in the building of simulations. This step is important for relaxing the system, which would reduce and even remove unfavourable interactions prior to the production runs[98]. This is especially useful on protein structures to also reduce the steric clashes between side chains and waters, or side chains and lipids, in a system. The energy minimisation algorithm used for all energy minimisations in this thesis is the steepest descent algorithm, which searches for minima by proceeding down the potential energy landscape until a minima is reached.

Another energy minimisation algorithm is the conjugate gradient algorithm. This algorithm is similar to the previous algorithm, but focuses on reducing the energy of the system to a very low energy. The direction of movement in the potential energy landscape is similar to the steepest descent method.

2.4 Convergence Analysis

With the advent of molecular dynamics simulations, as well as the increased efficiency of computational resources, techniques must be developed to assess how stable a system's properties are, or how particular properties of a simulation converges. Many methods are available, such as calculating the free energies of a system[123] to determine

energetic differences between states, running simulations for a longer timescale[124] to determine the stability of a system, and running many simulations coupled with a bootstrapping technique[125] to assess various parameters for transmembrane helix dimerisation.

2.5 Homology Modelling

The models and mutants used in the simulations presented in Chapter 4 and 5 were generated using MODELLER[115]. In Chapter 4, MODELLER was used to build unresolved loops in the Dok7 crystal structure, as well as mutate specific residues for *in silico* mutational studies. In Chapter 5, MODELLER was used to both build the unresolved loops in the crystal structure, as well as build the unstructured juxtamembrane region that connects Dok7 and MuSK. The homology modelling process contains the following:

1. Selection of template(s)
2. Alignment of template(s) with manual editing of sequences if necessary
3. Build model using MODELLER functions
4. Model evaluation
5. Use in simulations

The models were then visually assessed for any non-physiological characteristics, and used in the simulations.

2.6 Software and Analysis

All simulations in this thesis were performed using the GROMACS (http://www.gromacs.org/) bimolecular simulation package. Analyses were performed using various GROMACS analysis tools and in-house scripts written in bash, Python and tcl. Visualisation of the trajectories and other graphics were performed in VMD[126].

3

MuSK's Interaction in Bilayers and a Statistical Sidenote

3.1 Background - MuSK

Muscle-Specific Kinase (MuSK) is an 800-residue member of the receptor tyrosine kinase (RTK) family (Figure 3.1A, left)[12]. Mutations and defects in MuSK disrupt the signalling from the nerve cell to the muscle cell[12], which lead to the disease *myasthenia gravis*[57–63]. There are crystal structures of four different domains of

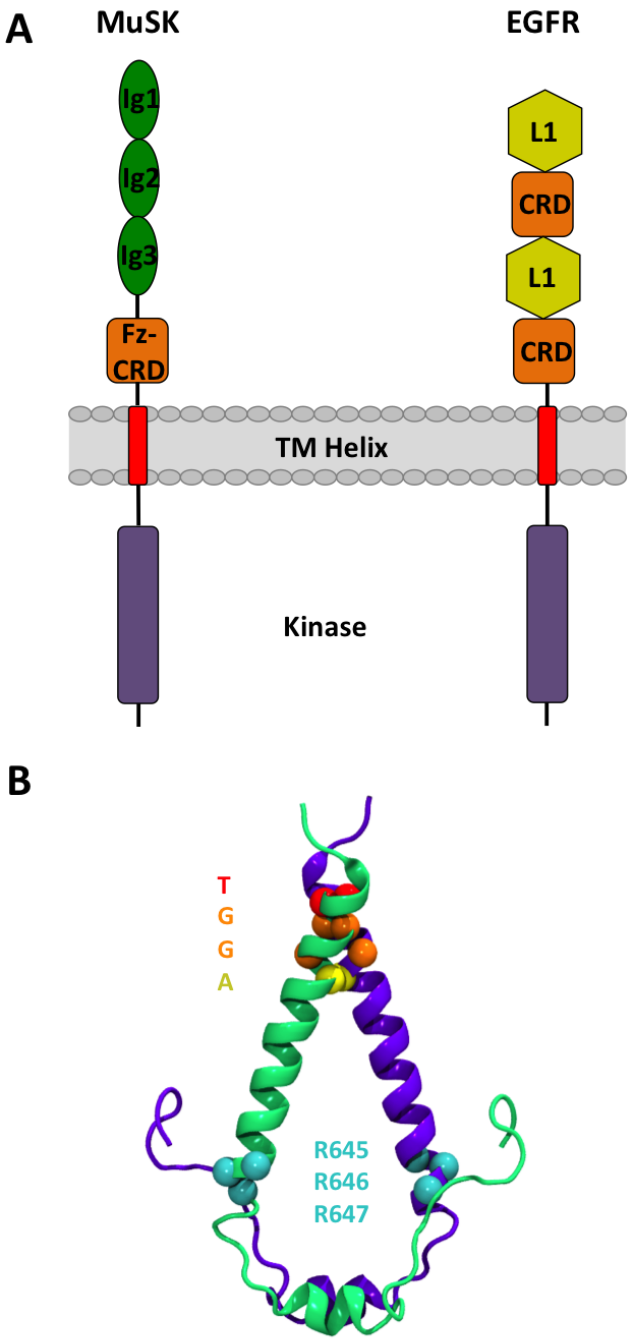

Figure 3.1: Cartoon schematic of MuSK (left) and EGFR (right). Immunoglobulin-like domains are green ovals, cysteine-rich domains are in orange squares, receptor L domains are in yellow hexagons, and the TM and kinase domains are in red and purple cylinders, respectively. Unstructured regions are represented by black lines. The bilayer is represented by a grey rectangle and grey ovals. B. NMR and MD structure of the EGFR TM Helix (pdb id 2M20) plus some residues of the juxtamembrane region. The backbone is represented as green and purple cartoon, and the interfacial residues are in red, orange and yellow VDW representation. R645, R646 and R647 are represented by blue VDW.

MuSK[68;70;71], which structurally characterise ∼70% of the receptor (see Chapter 1, Section 1.5). However, the structure of the transmembrane (TM) helix of MuSK has not been determined, and consequently the structure of the dimer is uncertain.

Other TM helix structures have been published, as discussed in Section 1.1 (Figure 1.3). One of these key structures is of the TM and JM region of EGFR[17]. Epidermal Growth Factor Receptor (EGFR) (Figure 3.1A, right) is another member of the RTK family, which is implicated in cell growth, survival, proliferation and differentiation[4;127]. A structure of the TM+JM portion of EGFR has been published[17](Figure 3.1B), so the structure of the right-handed dimer is known. There is a small-x-x-x-small amino acid motif that dominates the interactions between the transmembrane helices, which is seen in other RTKs. The structure shows that the helices intersect at an angle of -44°, and is dominated by a TGxxGA motif. There are also mutations on EGFR that correspond with lipid interactions needed for activation, specifically R645, R646 and R647 in the JM region[22].

Lipids are becoming an increasingly important component of biological systems. There is an increasing amount of evidence that lipids play a role in signalling, as evidenced by EGFR. The plasma membrane model contains a diverse range of lipids, especially anionic lipids like POPS and PIP_2[128]. Since these anionic lipids have been shown to interact with the positively charged JM region of RTKs[129], it follows that both lipids and the addition of unstructured regions of the TM helices may affect dimerisation. Thus, the aim of this chapter is to determine the effect of lipid composition and unstructured extensions of the protein on MuSK's dimerisation behaviour.

It is known that MuSK interacts with itself, from a study which tested the propensity of the RTKs to homodimerise via the TOXCAT assay[13]. All RTKs dimerise to an extent, but with varying strengths, as compared to GpA. NMR experiments were performed to determine whether or not the MuSK TM region could be expressed, if

it is in fact helical, and whether it is folded in a membrane-like environment.

3.2 Experimental Procedures

3.2.1 The Single Protein Production System

To express the transmembrane portion of MuSK, the Single-Protein Production System (SPP)[130] was chosen because it has the potential to facilitate inexpensive isotope labelling approaches (Figure 3.2). Because of a condensation step right before induction, it is possible to grow a large amount of cells in unlabelled media to log phase, at which point they are condensed into a small amount of labelled media. This saves on labelled materials, which is particularly important for labelling schemes that require expensive precursors, such as deuteration and methyl protonation. In addition to being economical, it is also selective in its expression of protein, due to the presence of MazF. MazF is a sequence-specific endoribonuclease which selectively cleaves single-stranded RNAs containing ACA codons. Mutating or removing all of the ACA codons in the DNA allows for exclusive expression of the protein of interest after the addition of IPTG. This has the added advantage of selectively labelling the protein of interest, since IPTG is added directly after condensation, minimising the usage of labelled media by the cells for other cellular proteins and processes. It may also be possible to carry out structural studies on unpurified cell extracts, since only the protein of interest is isotopically enriched.

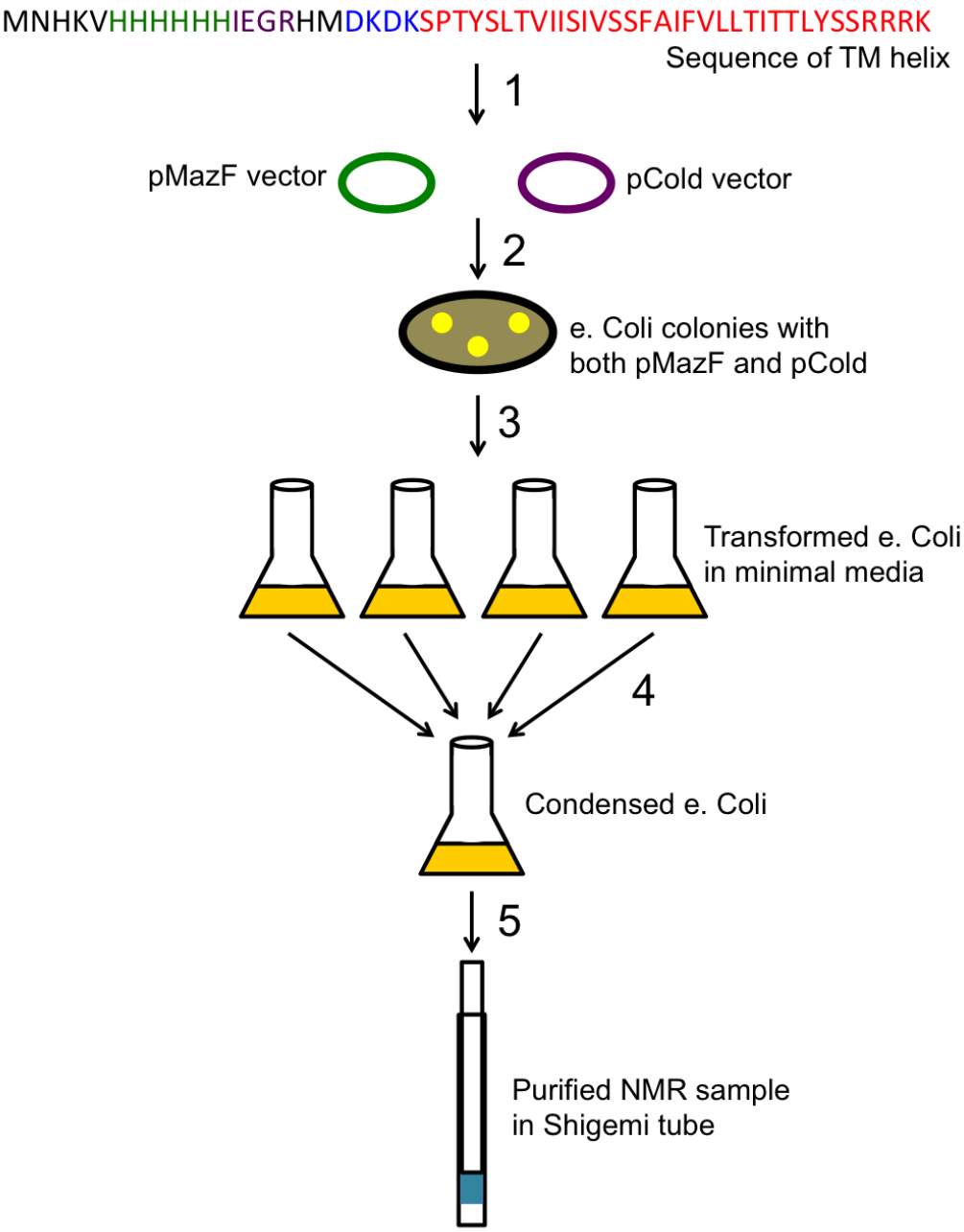

Figure 3.2: A diagram of the Single Protein Production (SPP) system. The sequence of interest is at the top, with a histidine tag highlighted in green, a factor Xa cleavage site in purple, residues to aid solubilisation in blue, and the sequence of MuSK in red. DNA vectors are either green (pMazF) or purple (pCold) ovals. *e. coli* colonies are represented as yellow circles, and the petri dish is represented as a brown oval outlined in black. Flasks are in white triangles and cylinders, outlined in black, with the media coloured yellow inside. The Shigemi NMR tube is represented by white rectangles outlined in black, with the sample coloured in blue.

3.2.2 Cloning of MuSK into pCOLD and Expression in the SPP System

Residues 489-522 of human MuSK, along with a his tag and extra residues to aid solubilisation were cloned into the pCold vector (Figure 3.2, 1). These were then transformed into BL21 cells competent with the pMazF plasmid. These cells were grown on plates with ampicillin and chloramphenicol (Figure 3.2, 2). pMazF-competent cells that had taken up the pCold vector were picked from the cell plates, and put into minimal media with casamino acids overnight at 37°C . The cells were then spun down for 15 minutes at 3300 g at 25°C and were resuspended in 2L unlabelled M9 media, with no casamino acids (Figure 3.2, 3). They were then grown until they reach an OD(600) of 0.5, were subsequently cold-shocked for 5 minutes in an ice bath, then incubated at 15°C for 45 minutes. The cells were centrifuged at 3300 g for 15 minutes at 15°C, and resuspended in M9 1x wash buffer (per litre: 1.28g $NaH_2PO_4 \bullet 7H_2O$ (6.8g NaH_2PO_4), 3.0g KH_2PO_4, 0.5g NaCl, 1.0g NH_4Cl, pH 7.4). The cells were centrifuged at 3300 g for 15 minutes at 15°C, then resuspended in labelled media, and induced with IPTG, at a final concentration of 10mM IPTG. The cells were grown overnight at 15°C (Figure 3.2, 4). Through an expression test, it was determined that condensing by a factor of 4, and inducing with a final concentration of 10mM ITPG resulted in the highest protein expression.

3.2.3 Purification

A SoFast HMQC spectrum on MuSK in the unpurified *E. coli* cell membrane environment was collected, but it was determined that purification of the protein was needed to measure peaks with resolution suitable for resonance assignment. The cells were spun down for 15 minutes at 4000 rpm at 4°C. They were then resuspended in 100mL

lysis buffer (50mM Tris, 200mM NaCl, pH 8.0), and 100μL of a protease inhibitor cocktail was added, as well as 10mg of lysozyme from chicken egg white. This was stirred for a half hour at room temperature, then the solution was sonicated twice for five minutes each, with the sonicator on for 20 seconds, off for 40 seconds. The solution was then ultracentrifuged for 1 hour at 200,000 g at 4°C. The supernatant from this was poured off, and the pellet was resuspended in Extraction Buffer (20mM Tris, 200mM NaCl, 8M urea, pH 8.0) using a Dounce homogeniser. 0.6mL of Triton X-100 was then added to this solution, and it was stirred for one hour at 4°C. The solution was then ultracentrifuged for 1 hour at 200,000 g at 4°C, and the pellet was kept for preparation of NMR samples (Figure 3.2, 5).

3.2.4 Preparation of Samples for NMR

0.002g protein was dissolved in 1mL of 100% formic acid. This solution was diluted to 10% formic acid, then frozen and lyophilised. One of the pellets was dissolved in 1mL of HFIP with 100mM LMPG for resolubilisation. Three consecutive thin films were made from the pellet, each using 1mL of HFIP to aid solubilisation of both protein and detergent. The sample was then lyophilised and taken up in 500μL of KH_2PO_4. This sample was put into a round-bottom NMR tube for spectra collection at 600MHz. Another pellet was dissolved in 40mM SDS and 8M urea, then dialysed twice against 20mM KH_2PO_4, then 20mM NaH_2PO_4 with 8mM SDS. This was subsequently put into a Shigemi tube for spectra collection at 950MHz.

3.3 Simulation Procedures

For all helices used in this study, the sequences used were obtained from UNIPROT (http://www.uniprot.org). The TM domain for each sequence was predicted using a

Figure 3.3: Sequences of MuSK used in the simulation studies. The small-x-x-x-small amino acid motif is highlighted in red, and the unstructured extensions are highlighted in green and cyan. The numbering of the residues is taken from the sequence of human MuSK.

consensus TM predictor[131]. They were then built using PyMOL. Both the EGFR and MuSK test sets were run on SIDEKICK by Khairul Abd Halim using the BOND[86] forcefield. These were run using GROMACS3.2.2, with the Berendsen thermostat and barostat used for temperature (T=323K) and pressure coupling[122]. The LINCS algorithm was used to constrain bond length[116]. The timestep used was 20fs. Two helices were placed in a preformed DPPC bilayer 60Å apart. The box measured approximately 9nm x 9nm x 15nm.

All three models used in this study were built using PyMOL, with the sequences used given in Figure 3.3. The other CG simulations were performed using the MARTINI2.2 forcefield[107;112] using GROMACS4.5.5. Two helices were placed 60Å apart, and then lipids self-assembled around the helices with restraints placed on the helices. Once the bilayer was formed, the restraints on the helices were released, and the helices were then allowed to explore the bilayer environment via a random walk (Figure 3.4A). For the bilayers with POPC and POPS, exchange lipids[128] was used to exchange some of the POPC molecules for POPS. The three bilayers that were used were 100% POPC, 100% DPPC, and 80%PC + 20%PS, totalling five different MuSK+bilayer systems used: MuSK in 100% POPC, MuSK in 100% DPPC, MuSK in 80%PC + 20%PS, MuSK+4 in 80%PC + 20%PS, and MuSK+8 in 80%PC+20%PS. Anionic lipids were added into the inner leaflet to act as a simple model of the more diverse anionic lipids of the inner leaflet of the plasma membrane. The Berendsen thermostat and barostat were used for temperature (T=323K) and pressure coupling[122], and the LINCS algorithm was used to constrain bond length[99]. The time step used was 20fs

at a temperature of 323K. There were approximately 5300 water particles in the box, which measured approximately 12nm x 6.5nm x 12nm.

The atomistic simulations were generated by taking a coarse-grain snapshot of the system, which corresponds to either the left-handed or right-handed modes of the crossing angle, and converting it to all-atom representation using a fragment-based approach[113]. These AT-MD simulations were run using the GROMOS96 53a6 forcefield[132]. The LINCS algorithm[99] was used to constrain bond lengths, while the Parinello-Rahman barostat[133] and Berendsen[122] thermostat was used (T=310K). The box length was approximately 12nm x 6.5nm x 17nm, with Particle Mesh Ewald (PME)[118] used for long-range electrostatics. Equilibration was run on these systems for 1ns with the Cα atoms restrained. This was followed by 0.1μs of unrestrained simulation. Analysis was performed using GROMACS[116], VMD[126], and locally written scripts. A summary of the simulations run for this study is in Table 1.

Table 3.1: Simulations Performed in MuSK Study

Simulation	Duration	Forcefield	Simulator
Coarse-Grained Simulations			
EGFR TM Test Set	50 x 0.5μs	BOND	Khairul Abd-Halim
MuSK TM Test Set	77 x 0.5μs	BOND	Khairul Abd-Halim
MuSK TM DPPC	100 x 1μs	MARTINI2.2	Amanda Buyan
MuSK TM POPC	100 x 1μs	MARTINI2.2	Amanda Buyan
MuSK TM PC/PS	50 x 1μs	MARTINI2.2	Amanda Buyan
MuSK TM+4 PC/PS	50 x 1μs	MARTINI2.2	Amanda Buyan
MuSK TM+8 PC/PS	50 x 1μs	MARTINI2.2	Amanda Buyan
Atomistic Simulations			
LH MuSK TM+8 PC/PS	3 x 0.1μs	GROMOS53a6	Amanda Buyan
RH MuSK TM+8 PC/PS	3 x 0.1μs	GROMOS53a6	Amanda Buyan

*All PC/PS/bilayers contained 80%PC + 20%PS.

3.4 NMR Spectra

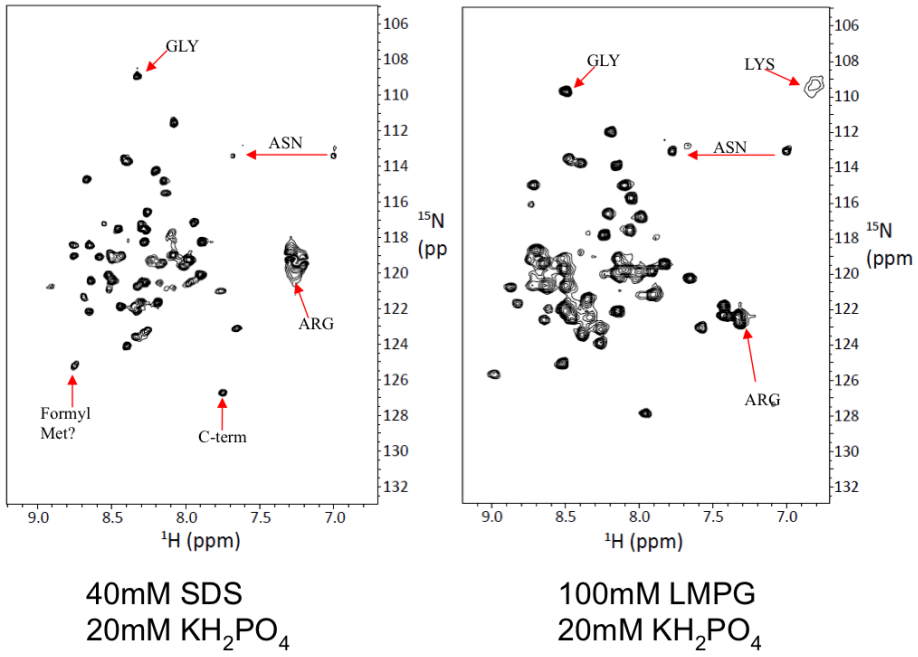

Figure 3.4: NMR spectra of MuSK in SDS micelles (left) and LMPG micelles (right). The SDS spectra was collected at 950MHz, and the LMPG spectra was collected at 600MHz. Peaks of interest are highlighted by red arrows.

Two NMR spectra were obtained of the MuSK TM Helix - one in 1-myristoyl-2-hydroxy-sn-glycero-3-[phospho-rac-(1-glycerol)] (LMPG) micelles and one in sodium dodecyl-sulfate (SDS) micelles (Figure 3.4). The moderate peak dispersion indicates that the protein is indeed folded in micelles, and that it is folded in an α-helical manner. It is unclear whether or not there are helix dimers forming, but the cross peaks are well-defined in both spectra, indicating that the protein adopts a single conformation in both SDS and LMPG detergent micelles.

3.5 Simulation Results

Next, we sought to determine the dimerisation interface of MuSK via MD simulations. To test this, simulations are run where the helices are placed at least 60Å apart from

each other (Figure 3.5), and allowed to move around in the bilayer and freely assemble into a dimer. However, many simulation are needed to accurately determine the dimerisation interface, due to the approximate nature of coarse-grained forcefields (see Section 2.2.4 for an extended discussion). To determine the interactions between two helices in a bilayer, the SIDEKICK method[85] was developed to run many simulations of systems to attain sufficient sampling. However, the next question is how to assess the convergence of the data to determine if a system has been properly sampled.

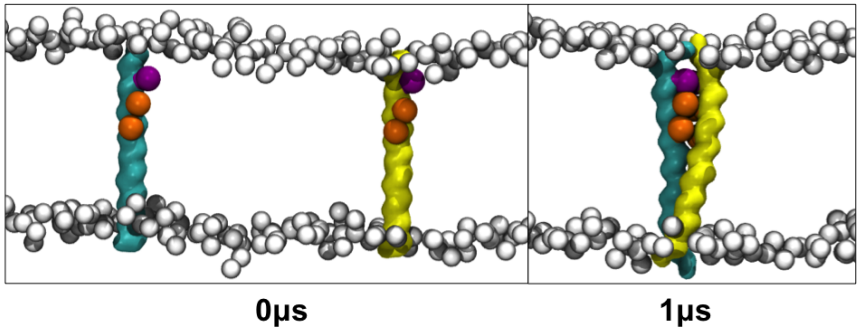

Figure 3.5: Start and end of a dimerisation simulation with duration 1μs. Helices are in cyan and yellow surface, and threonines and serines are in purple or orange VDW.

3.6 Issue of Sampling and Convergence

As discussed in Section 2.4, assessing convergence of particular properties of a system is important to determine if your sampling is sufficient. For these simulations, we are interested to know when specific properties of a system have converged, namely the crossing angle and the interfacial residues, such that there is little deviation from the mean and modes of the crossing angle and contacts when additional simulations are added to the mix. For this reason, the jackknife method was chosen as a way to assess convergence of transmembrane helix dimer systems. However, while the jackknife principle was chosen, it was traditionally used for one sample size, whereas we wanted to sample over all the possible subsets (each subset is referred to as M) of the population size (denoted N).

3.6.1 Jackknife Theory

The jackknife method was invented in 1949 by Quenouille[134], and later by Tukey[135] in 1958. Initially, it was determined that if you had a sample size, denoted as N, and removed one of the observations (denoted θ), you would get an estimate of θ as

$$\hat{\theta}_{J(1)} = N\hat{\theta}_N - (N-1)\bar{\theta}_{(i)} \tag{3.1}$$

where $\hat{\theta}_{J(1)}$ is the estimate of θ based on removing 1 observation, $\hat{\theta}_N$ is the actual theta, and

$$\bar{\theta}_{(i)} = \frac{1}{n}\sum_{i=1}^{N}\hat{\theta}_{(i)} \tag{3.2}$$

The variance is denoted as

$$\sigma^2 = \frac{N-1}{N}\sum_{i=1}^{N}(\hat{\theta}_{(i)} - \bar{\theta}_{(i)})^2 \tag{3.3}$$

$\hat{\theta}_{J(1)}$ is the estimate of the overall statistic with 1 observation removed, $\bar{\theta}_{(i)}$ is the average value with N-1 samples, and $\bar{\theta}_{(i)}$ is the value of theta with sample i removed from the population. Later, this idea was extended to allow for removing a sample size larger than 1, denoted by M, which gives the formula

$$\hat{\theta}_{J(M)} = n\hat{\theta}_N - (N-M)\bar{\theta}_{(i)} \tag{3.4}$$

where $\hat{\theta}_{J(1)}$ is the estimate of θ based on removing M observations, $\hat{\theta}_N$ is the actual theta, and

$$\bar{\theta}_{(i)} = \frac{M}{N} \sum_{i=1}^{N} \hat{\theta}_{(i)} \tag{3.5}$$

The variance of removing M observations is

$$\sigma^2 = \frac{N-M}{N} \sum_{i=1}^{N} (\hat{\theta}_{(i)} - \bar{\theta}_{(i)})^2 \tag{3.6}$$

where $\bar{\theta}_{(i)}$ is the average value with N-M samples, and $\bar{\theta}_{(i)}$ is the value of theta with sample i removed from the population. However, to extend this idea so convergence is assessed, then an approach is needed where it would sample all possible subsets, or iterate, over all possible Ms, not just one particular M.

3.6.2 The Iterative Jackknife Method

Later, in 2012, Wilke[136] expanded on the jackknife principle and used an iterative approach. In this technique, instead of using one subset, all subset sizes of 1 to N-1 were iterated over to determine the minimum sample size needed to detect 'hotspots' on fMRI scans of patients' brains, providing a starting point for convergence analysis. However, sampling all possible combinations for each subset properly proved difficult computationally. For example, if N is 50, and 1 simulation is removed from the population, there are 50 different subsets of 49. However, as more and more simulations are removed, the number of possible combinations of subsets increases according to Equation 3.7:

$$C_{max} = \binom{N}{M} = \frac{N!}{(M! * (N-M)!)} \tag{3.7}$$

where C_{max} is the total possible number of unique combinations sampled without

replacement, N is the total number of simulations, and M is the number of simulations randomly chosen from the population N. This quickly becomes infeasible to sample all possible combinations of each subset size, as removing even 5 simulations from the population would result in $2.1*10^6$ possible combinations. However, Wilke determined that if a random sample of 100 subsets is taken from all the possible combinations, it is sufficient sampling to represent the population of all possible combinations.

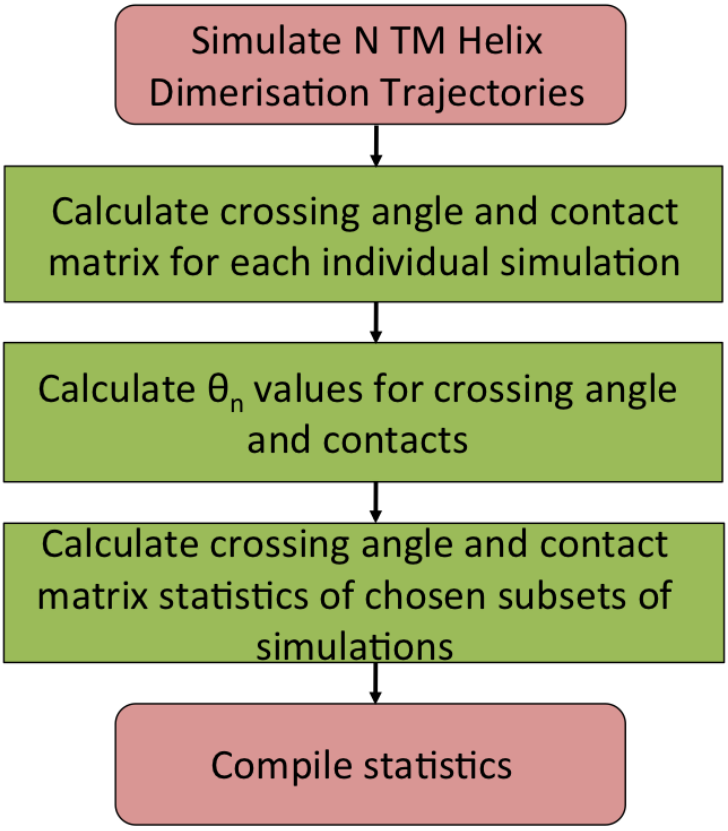

Figure 3.6: Workflow for jackknife algorithm. Start/stop of the algorithm are red rounded rectangles, and actions are green rectangles.

This principle was adapted to assess the crossing angle and the contacts of transmembrane helix dimer simulations (Figure 3.6). Crossing angles and contact matrices were calculated for all dimerised frames of all dimerisation simulations. Calculations of the overall mean and mode of the crossing angle, as well as the most frequently occurring contacts (θ_Ns) were run. Then, for each subset numbering 1 to N-1 (denoted M), the means, modes and probability for both right-handed (RH) and left-handed

(LH) dimers were determined for each randomly chosen subset of size M. The data generated for the crossing angles is represented as box plots, with the black line representing the median of the data, the boxes representing the first and third quartiles of the data, and the end of the lines represent the minimum and maximum values found from 1000 randomly drawn combinations. 1000 subsets were chosen instead of 100, as previously suggested, due to the different type of data used, and due to better convergence with 1000 subsets for this set of data.

To find the most frequently occurring interfacial contacts, the average contacts overall, the RH and the LH dimer were calculated for each dimerised frame. Then, for each combination of simulations chosen for a particular subset M, all of these frames were iterated through to find the ten smallest C_α distances for each frame. Frames designated one of the ten 'closest contacts' were given 1s, while the other contacts were given 0s. These contacts were determined for all 1000 combinations. Once these frequency matrices were calculated, they were added together to make a 'frequency matrix,' where all contacts would have a value between 0 and 1000, depending on how frequently the contacts were designated the 'closest contacts.' This was done for both right- and left-handed dimers, as well as an overall picture.

3.6.3 Results - EGFR test set

A training set was needed to test the accuracy of the program, as well as the precision. The training set chosen was 50 independent dimerisation simulations performed using the SIDEKICK method[85]. These simulations only included the transmembrane region of EGFR. These simulations were chosen because it was previously shown qualitatively (Figure 3.7A) that the crossing angle converged for one set of randomly chosen subsets. In addition, there is a known structure of the transmembrane domain of EGFR, so a direct comparison can be made of the calculated mean crossing angle

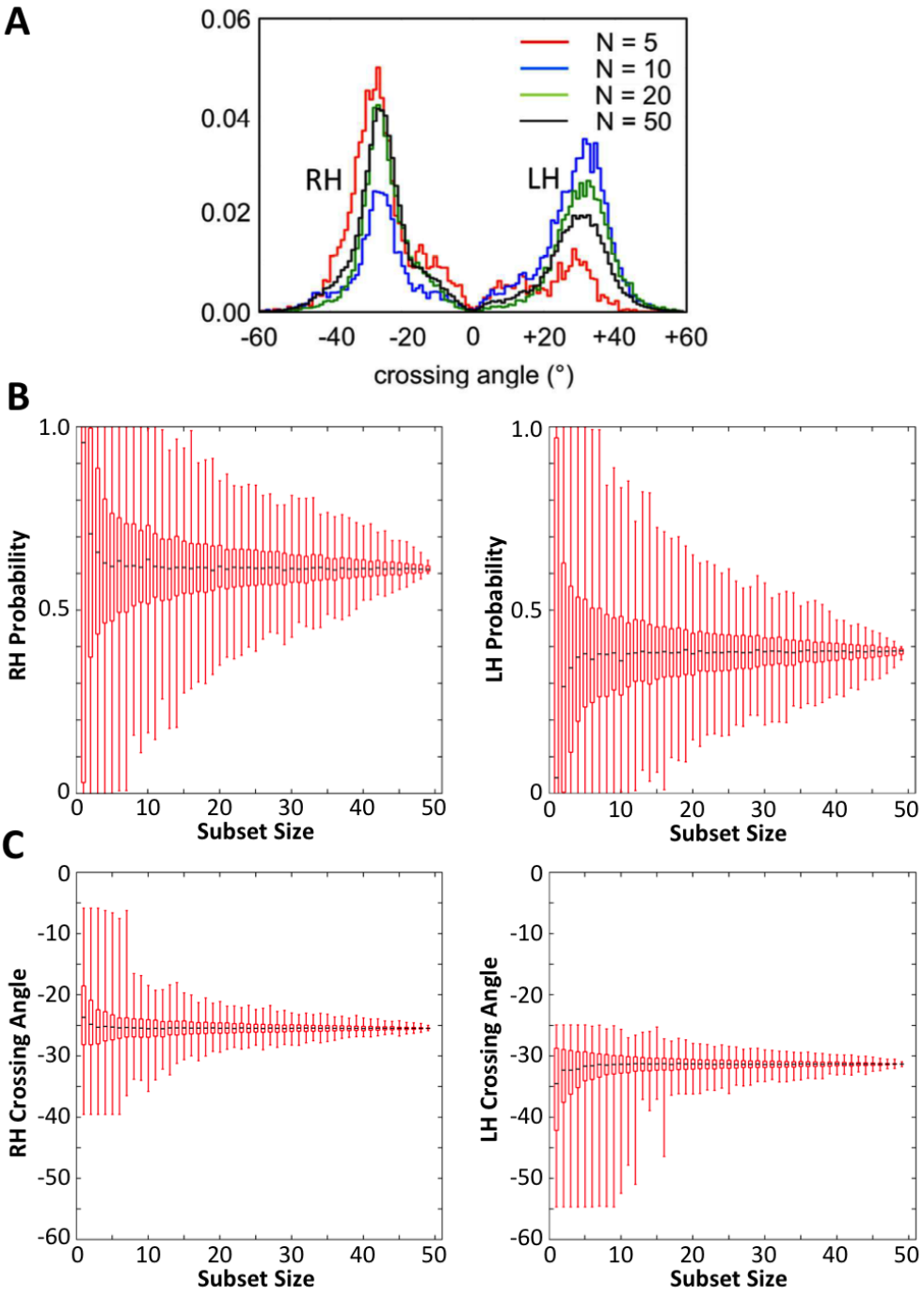

Figure 3.7: Crossing Angle Analysis for the EGFR test set. A. Qualitative crossing angle analysis performed previously by Khairul Abd Halim. The random 5 simulations are in red, 10 in blue, 20 in green, and 50 in black. B. Boxplot analysis of crossing angle probabilities for RH (left) and LH (right) crossing angles. Black lines are the median of the data, the first and third quartiles are the bottom and the top of the box, respectively, and the ends of the bars represent either the minimum or maximum of the data sampled. C. Boxplot analysis of the mean crossing angle for RH (left) and LH (right) crossing angles.

and the most frequently occurring contacts.

First, the crossing angle was analysed and compared with the qualitative results. Previously, after 20 simulations, the probability as well as the mode of the left-handed and right-handed dimers remained relatively unchanged (Figure 3.7A). This was then compared to the results from the jackknife script. The overall median for the probability of RH (Figure 3.7B, left) and LH (Figure 3.7B, right) crossing angles remained stable after approximately 15-20 simulations. This can also be observed when looking at the size of the boxes, as the change in the height of the boxes decreases as the subset size increases. As the subset size increases, the height of the box decreases, indicating that the majority of the data is relatively close to the mode. The extremes of the data (the minima and maxima sampled) are still relatively far apart, but the distance between the extremes and the quartiles decreases as the subset size increases. This is also true of the median of the mean crossing angles (Figure 3.7C, with RH crossing angles on the left and LH crossing angles on the right). Together, these suggest a sort of convergence of the data, with regards to the crossing angle.

Next, the contacts made between the residues of each helix were analysed. When M=5, the main contacts are apparent; however, there is background noise in the top right corner of the frequency matrix (Figure 3.8A). However, when the subset size increases, even after M=20, the most frequently occurring contacts begin to become more apparent, and the noise of other contacts begins to decrease, all the way to becoming relatively non-existent in M=45. These top contacts coincide with the interfacial contacts determined by the recently solved NMR structure. These observations agree with previous data, that a small-x-x-x-small amino acid motif is responsible for tight, favourable binding between two transmembrane helices. When comparing a representative right-handed dimer structure from the simulation to the

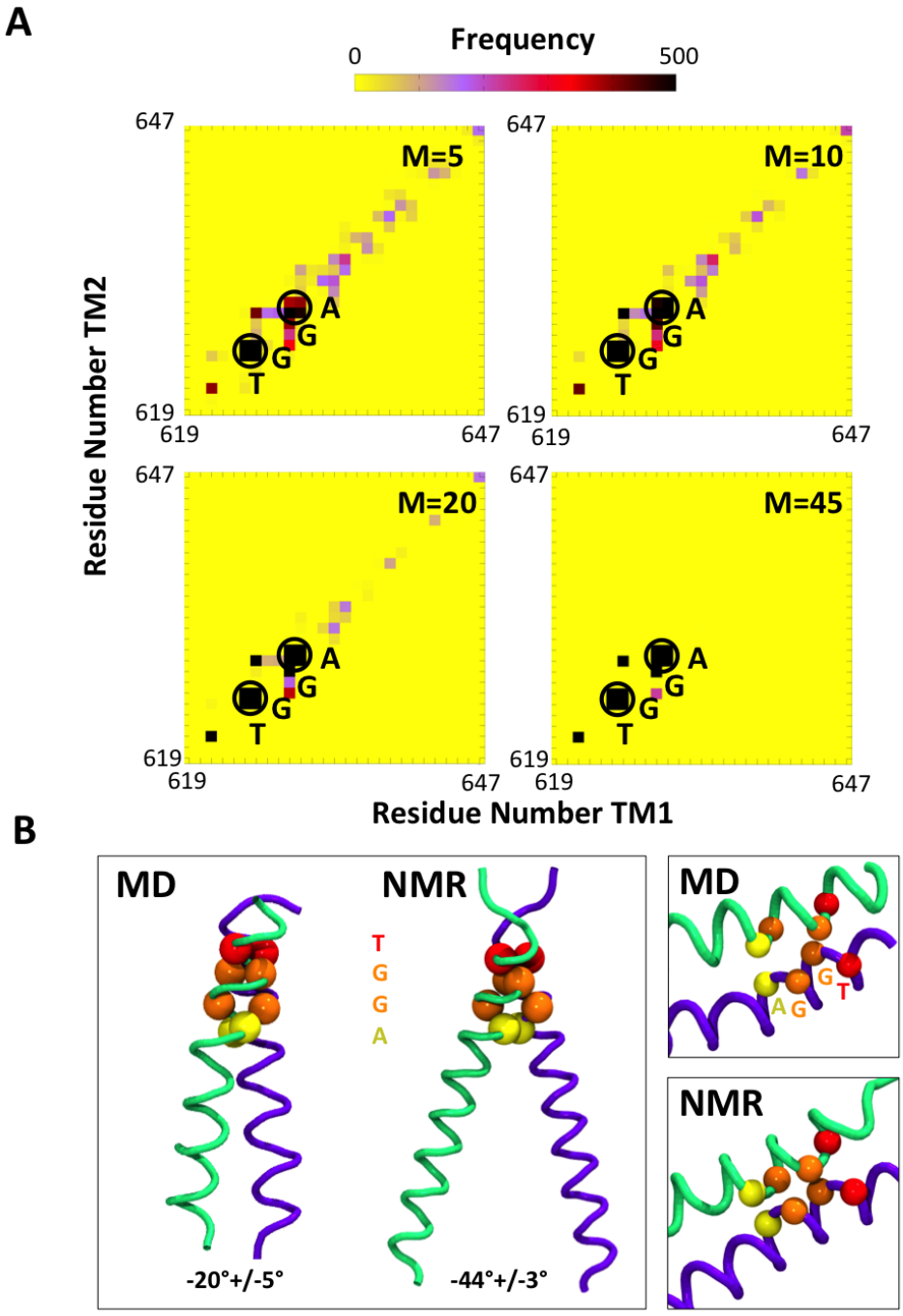

Figure 3.8: Contact Analysis for the EGFR test set. A. Frequency matrices for subset sizes 5, 10, 20, and 45, denoted by the letter M. The contacts responsible for dimerisation are highlighted in black circles. B. Contacts shown on MD structure vs. NMR structure. The contacting residues are in VDW representation, and the C_α atoms are in either green or purple tube representation.

NMR structure (Figure 3.8B), they look remarkably similar each other, with the contacting residues in similar positions in both structures. The large difference between

the two is the crossing angle at which they are represented. The crossing angle of the MD structure is ∼-25°+/-5°, whilst the NMR structure has a crossing angle of ∼-44°+/-3°. This can potentially be explained by the MD simulations only being performed on the TM helix, while the NMR structure contained both the TM helix and the JM region, which contains short α-helices bound together in an antiparallel fashion, pushing the two helices apart.

3.6.4 Results - MuSK test set

A test set was needed to determine if the number of simulations needed was system independent or system dependent, and if the crossing angle and interfacial contacts were able to converge on an unknown system. Simulations run on the MuSK transmembrane helix by Khairul Abd Halim were chosen, as there is no known structure of the TM region. There were 77 independent dimerisation simulations run using SIDEKICK[85], which only includes the TM region of MuSK (Figure 3.9). These crossing angle and contact matrices are different from EGFR, in that there is no clear preference for a RH versus a LH dimer. There are also no significant differences between the right-handed and the left-handed dimer contacts, in that they both showed a preference for the TxxxSxxxS motif. The crossing angle of MuSK also converges slower than the crossing angle of EGFR, as evidenced by the fact it takes at least 35 simulations to start to see convergence of both the crossing angle and the contacts.

When a more detailed look was done for the crossing angle, it was found that there were four relatively distinct modes (Figure 3.10). When these were unpacked, it revealed two distinct right-handed and left-handed crossing angle modes (Figure 3.10B), but two 'intermediate' states, suggesting a potential mechanism for switching between RH and LH dimers. While a potential mechanism of switching between states is interesting, without a method like free energy calculations, or running longer simulations,

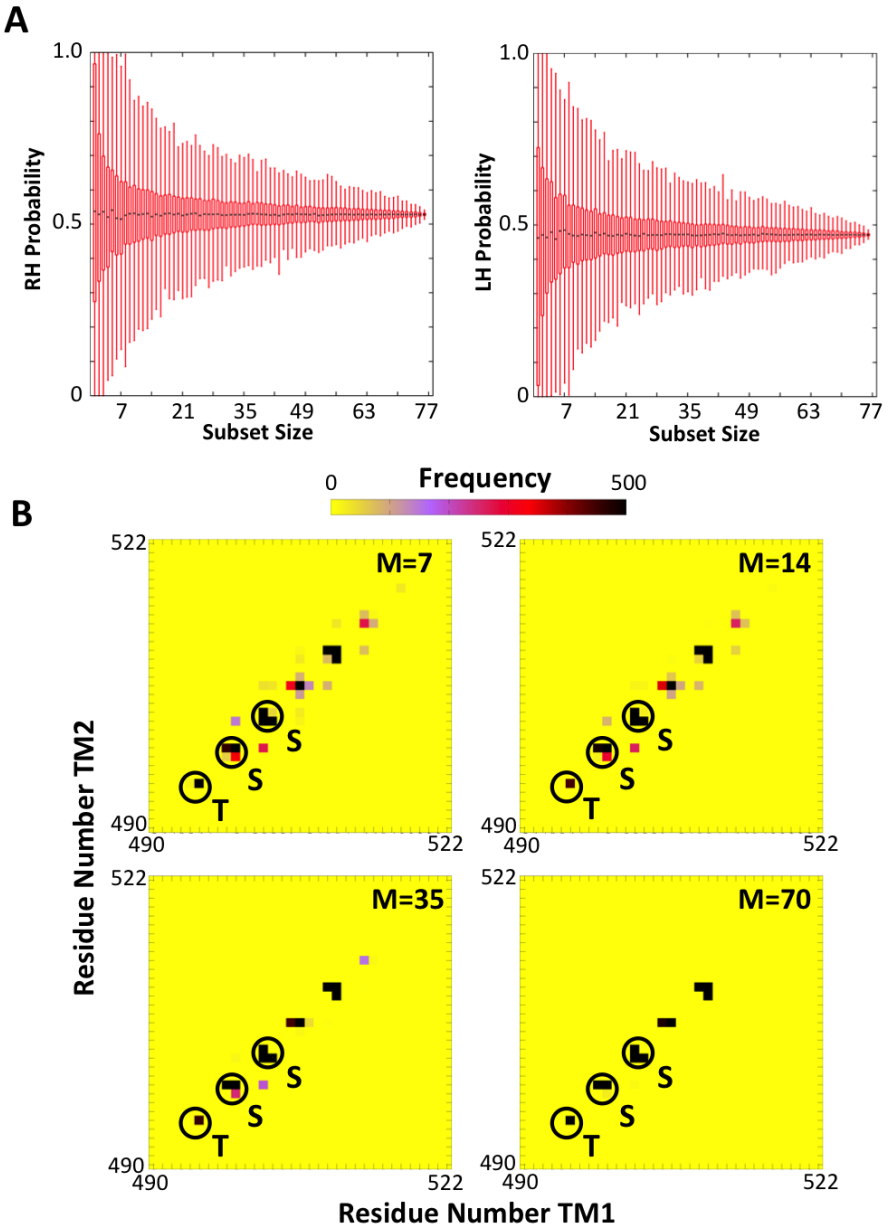

Figure 3.9: Jackknife analysis for crossing angle (A) and contacting residues (B) of MuSK test set. Contacts are highlighted by black circles.

the mechanism as well as direction of switching between RH and LH dimers will remain unclear. As this was not seen in the jackknife analysis, this points to the fact that, while the jackknife method works for determining macroscopic properties, it might not be able to distinguish subtleties in the activation mechanisms.

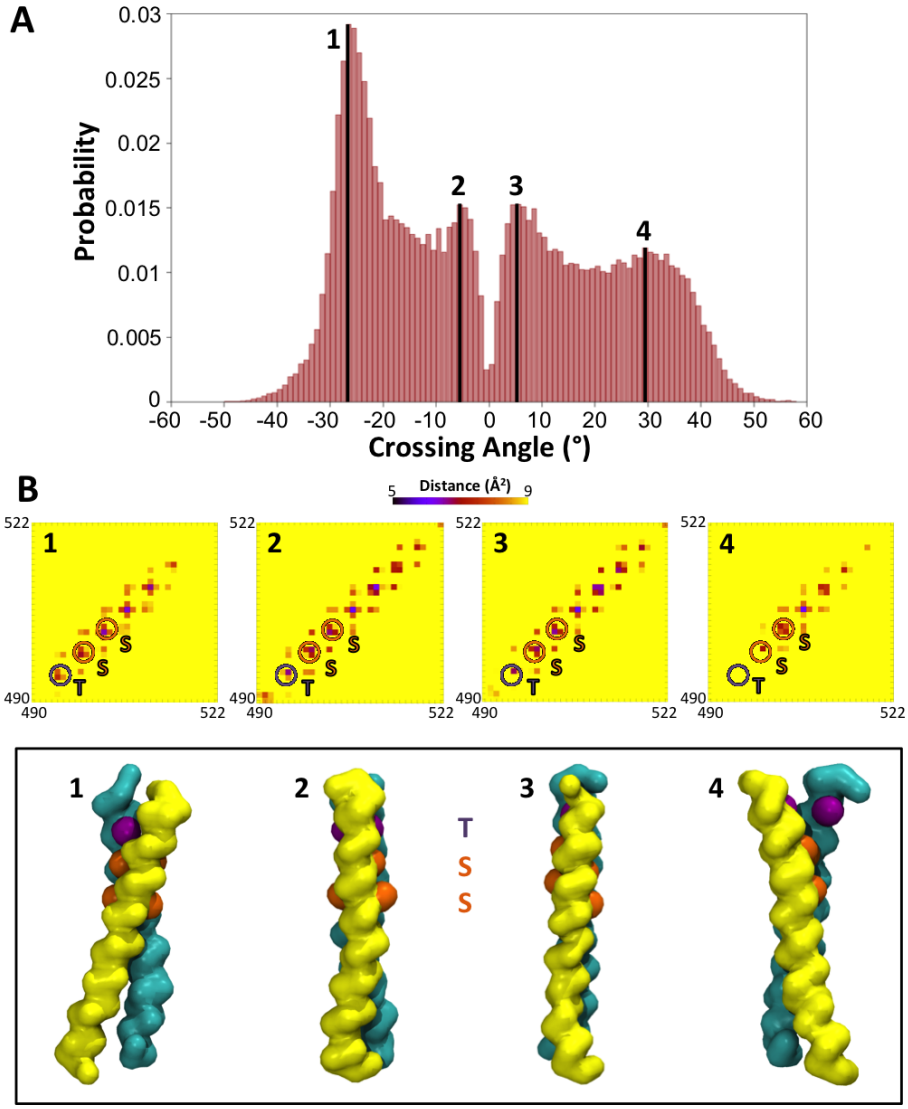

Figure 3.10: MuSK crossing angle distribution breakdown. A. Crossing angle distribution for test set of MuSK simulations. Four 'modes' are highlighted by black lines. The number of simulations included in each histogram, denoted by N, are as follows: N=100 for DPPC and POPC, N=5 for 80%PC+20%PS. B. Contact matrices for each of the four modes. The distances are calculated from the average of all of the structures that are contained in the bin denoted mode 1. The number of structures for mode 1 is 10902, mode 2 is 5905, mode 3 is 4810, and mode 4 is 5718. The threonine and serine contacts are highlighted by purple and orange circles, respectively. C. Representative images of the modes highlighted in A. Each TM is in either yellow or cyan surface, and threonines and serines are in purple and orange spheres, respectively.

3.7 Dimerisation of MuSK and The Influence of Lipid Bilayer Composition

After determining that the number of simulations needed for convergence of the crossing angle and the contacting residues for the MuSK TM dimer system was at least 35 simulations, it was decided to perturb two aspects of the system. Changing the lipid composition in the bilayer and the length of unstructured regions added to the helix were chosen to see the effect on the crossing angle and interfacial residues. These characteristics of the system were chosen because of the diverse range of lipids inherent in the plasma membrane, as well as the fact unstructured extensions connect the TM helix to the intra- and extracellular portions of the protein, and may have roles in the mechanism of signal transduction. It was decided to use 50 simulations for all systems, as the previous work has shown that convergence is reached after 35 simulations, plus extra to account for differences in forcefield and bilayer lipid composition.

Both the crossing angle and the contacting residues were examined for perturbations. First, the crossing angle was examined for MuSK TM DPPC, MuSK TM POPC, and MuSK TM PC/PS (Figure 3.11A). The crossing angle modes did not change significantly as the composition of the bilayer was varied, as they remained around -30° for right-handed helices (Figure 3.12A), and around +20° for left-handed helices. The probability also remained very similar across the different bilayer compositions. The crossing angle modes, as well as the probability, also did not change significantly when the unstructured extensions extensions were added (Figure 3.11B).

This analysis was expanded by plotting the crossing angle versus distance between the Centres of Mass (COMs) for all 6 systems, to determine if there were any differences in the distances between the COMs of the helices. The modes are still around -30°

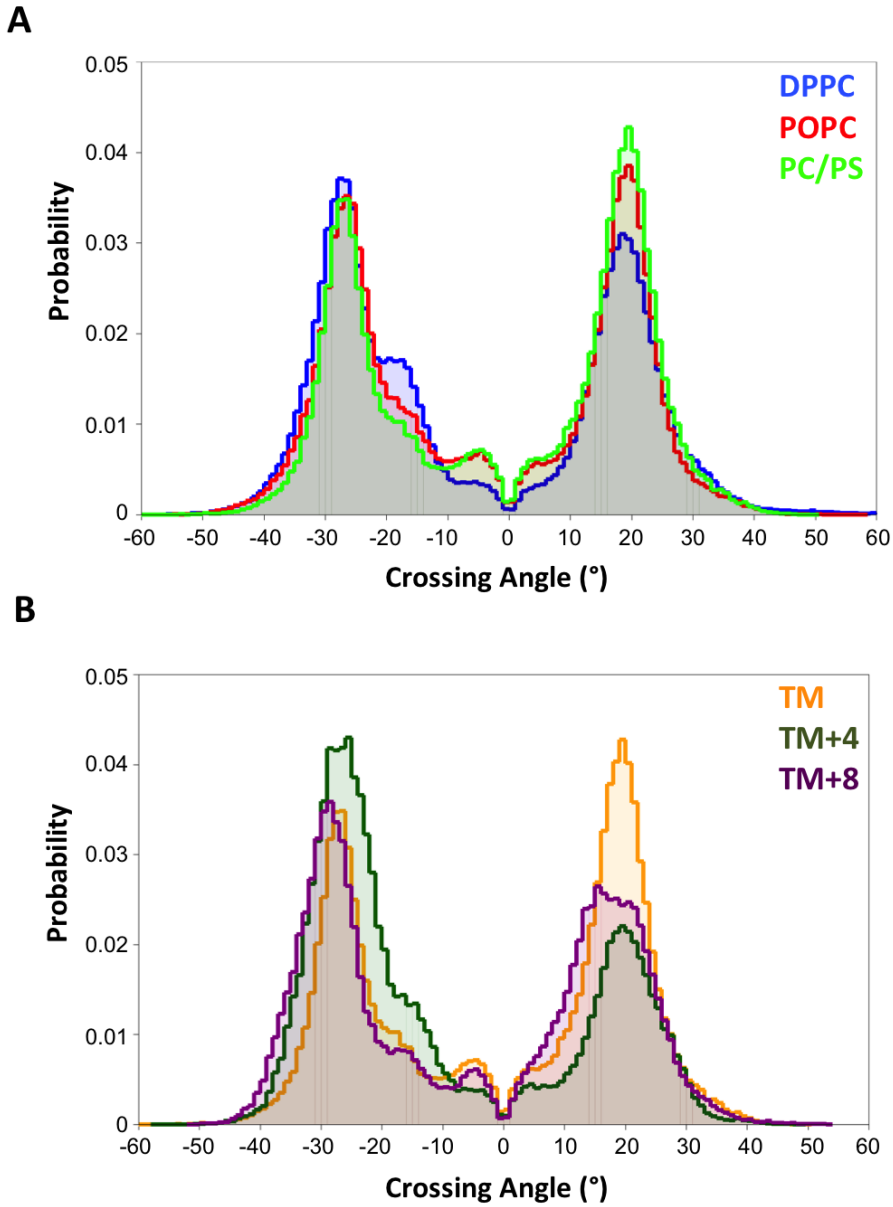

Figure 3.11: Crossing Angle of MuSK models. A. Crossing angle distribution for the TM helix models in three different bilayers: pure POPC (red), pure DPPC (blue), and 80%POPC + 20%POPS (green). B. Crossing angle distribution for TM (orange), TM+4 (green), and TM+8 (purple) systems in a 80%POPC + 20%POPS bilayer.

for right-handed helices, and around +20° for left-handed helices, but the distance between the COM differs (Figure 3.12). For the majority of the right-handed dimers, they reside at a distance of 6Ås between the helices, while the left-handed modes reside between 8-10Ås, indicating that the RH dimers are more tightly packed together than the LH dimers.

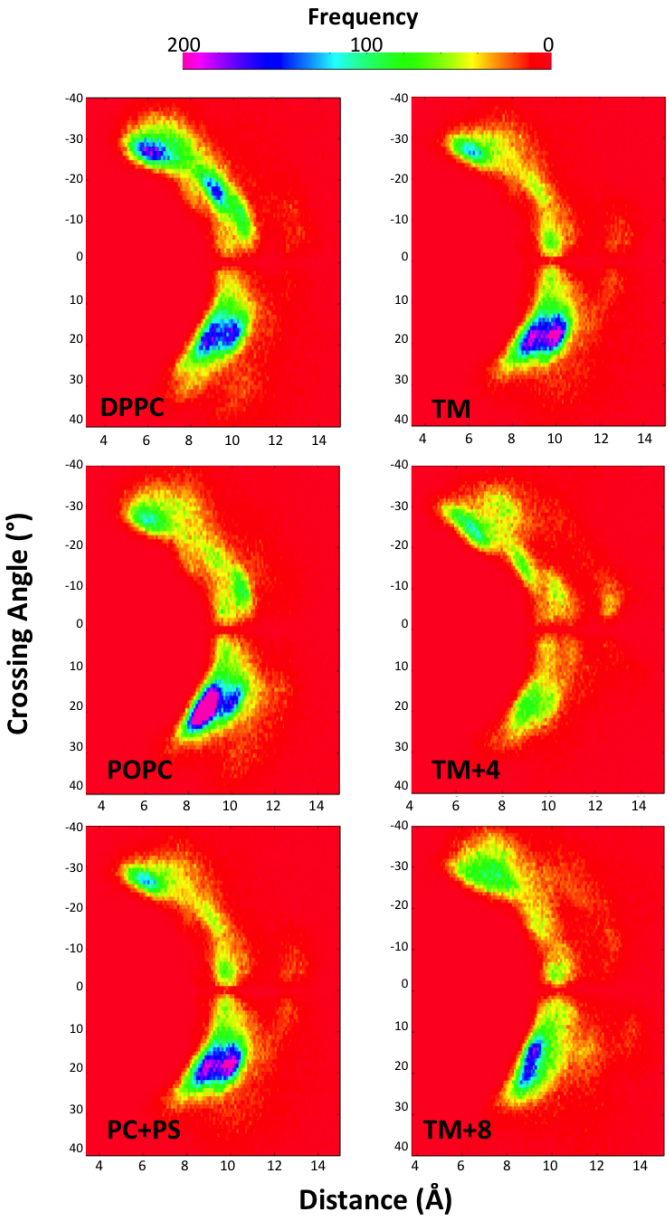

Figure 3.12: Crossing Angle versus Distance Between COMs of the helices for all 6 systems. Frequency is shown from red (0) to magenta (200). The bilayers that are tested are in the left column, and the addition of extensions (in PC/PS bilayer) are in the right column.

The contacts of each RH and LH dimer were analysed for each model (Figure 3.13). To do this, the LH structures were averaged into one structure, and the distances of this average dimer were calculated. The closest 10 contacts were then chosen, and given a 1. All other contacts were given a 0. This was also done for the RH dimer. As extensions are added to the helix, the main contacts are maintained (highlighted by

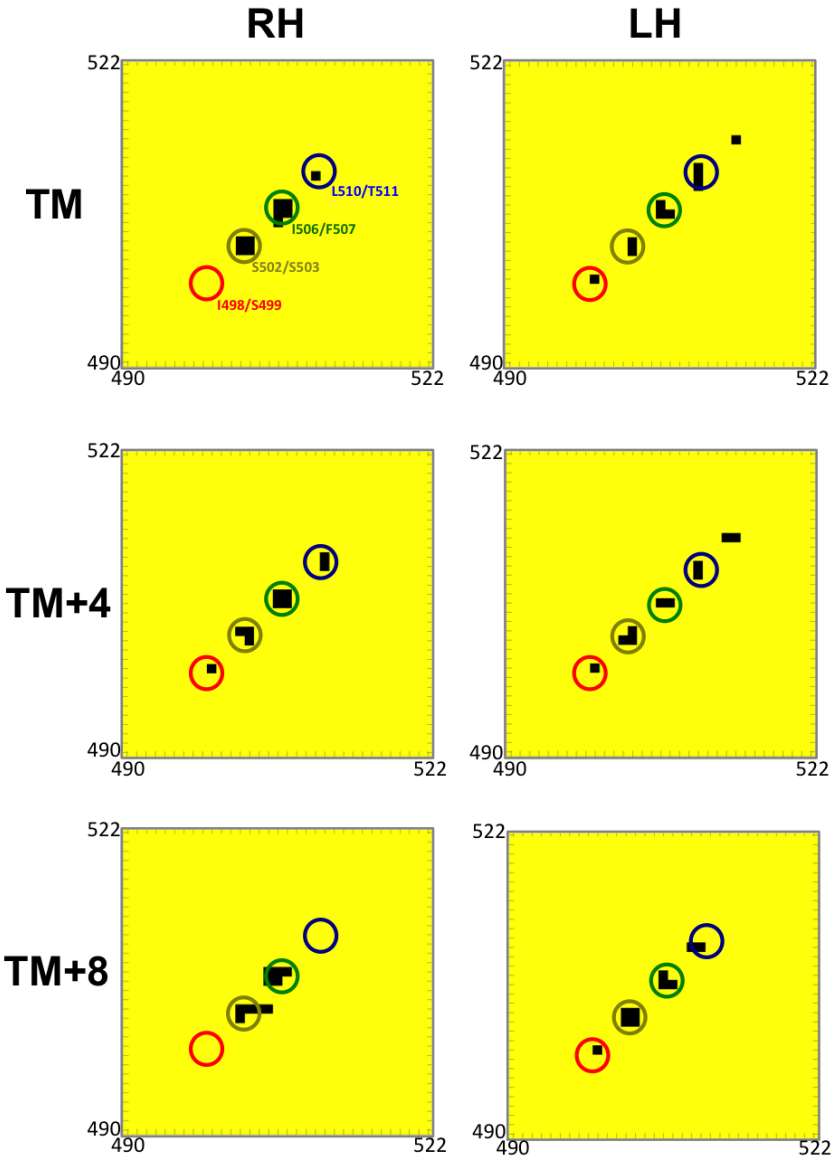

Figure 3.13: Right-handed (RH) versus left-handed (LH) contact matrices for TM, TM+4 and TM+8 in POPC+POPS. Black squares represent the ten closest contacts found of all RH and LH dimers. The circles surround the closest contacts made by MuSK: red for I498/S499, dark yellow for S502/S503, green for I506/F507, and blue for L510/T511.

the red, yellow, green and blue circles), but contacts vary in the right-handed more than the left-handed. It is seen that, as the unstructured extensions are added to the TM, the LH dimer remains more or less stable in the contacts it makes. However, the contacts in the RH dimer change slightly as the red and blue contacts are lost, resulting in more of a "smearing" in the middle two contacts. This could indicate

that the dimer is packing together more closely, and there are more contacts made around the central residues involved in dimerisation. Taking the top 10 contacts may not be enough to sufficiently sample this effect.

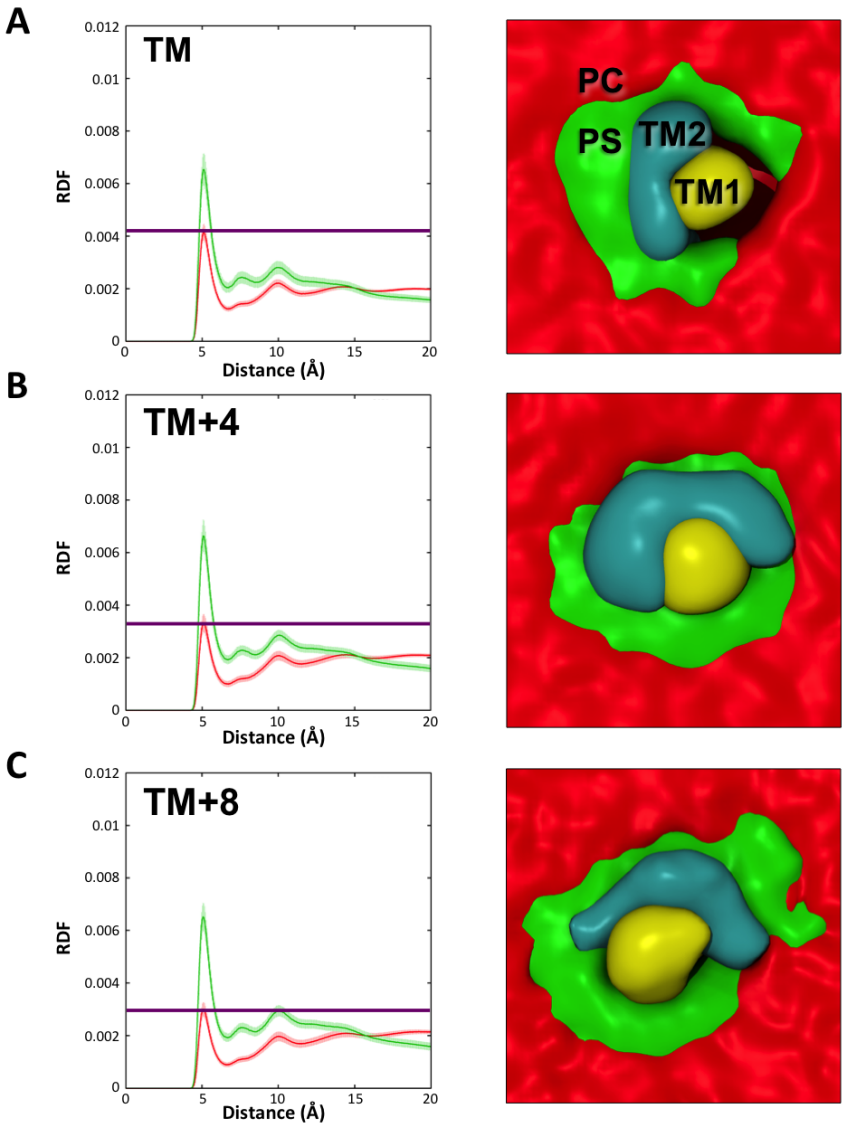

Figure 3.14: Radial Distribution Functions (RDFs) and Volmap plots of the MuSK TM (A), TM+4 (B), and TM+8 (C) systems. POPS is represented in green on the RDF and a green surface in the Volmap plots. POPC is represented as a red line in the RDFs, and red surface in the Volmap plots. The centred transmembrane helix is represented as a yellow surface in the Volmap plots, and the second transmembrane helix is represented as a cyan surface. The height of the POPC curve is indicated by the purple line on all of the RDF plots.

Next, the behaviour of the lipids were analysed. Radial Distribution Functions (RDFs) were calculated for each system, and Volmap plots were generated. As the

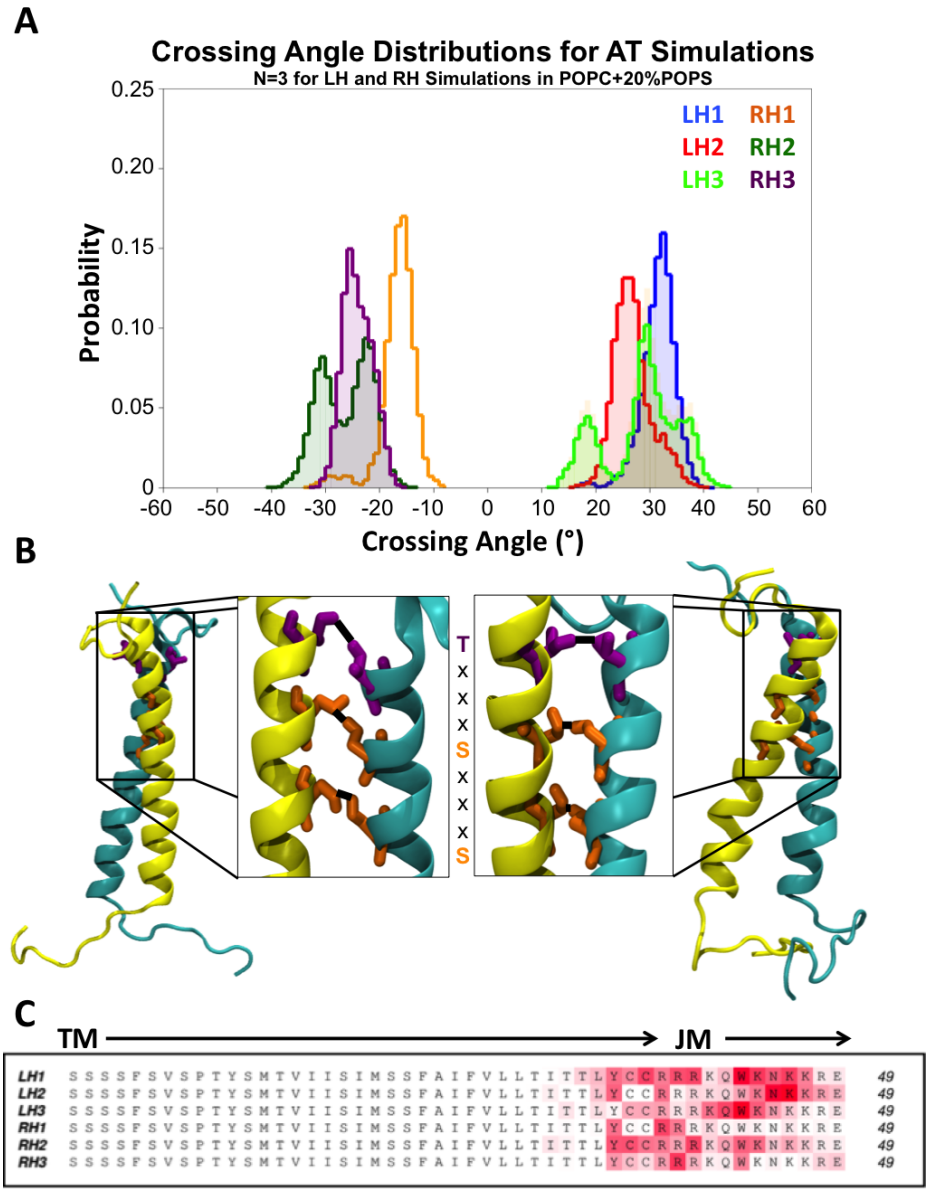

Figure 3.15: Atomistic simulations of right-handed and left-handed dimers in the PC/PS systems. A. Crossing angle histogram of all of the replicates. Each simulation is in a different colour. B. Representative examples of RH and LH dimers, as well as their respective interfaces. Each TM is in either cyan or yellow cartoon representation, the threonine in the interface is violet licorice, and the serines in the interface are represented by orange licorice. C. The sequence of MuSK+8 for each replicate. The red shading indicates the level of interaction POPS lipids have with a particular residue.

unstructured extensions increase in number of residues, from 0 (Figure 3.14A) to 4 (Figure 3.14B) to 8 (Figure 3.14C), the propensity of POPS to cluster around the dimer increases. While the maximum height of the POPS line does not appear to

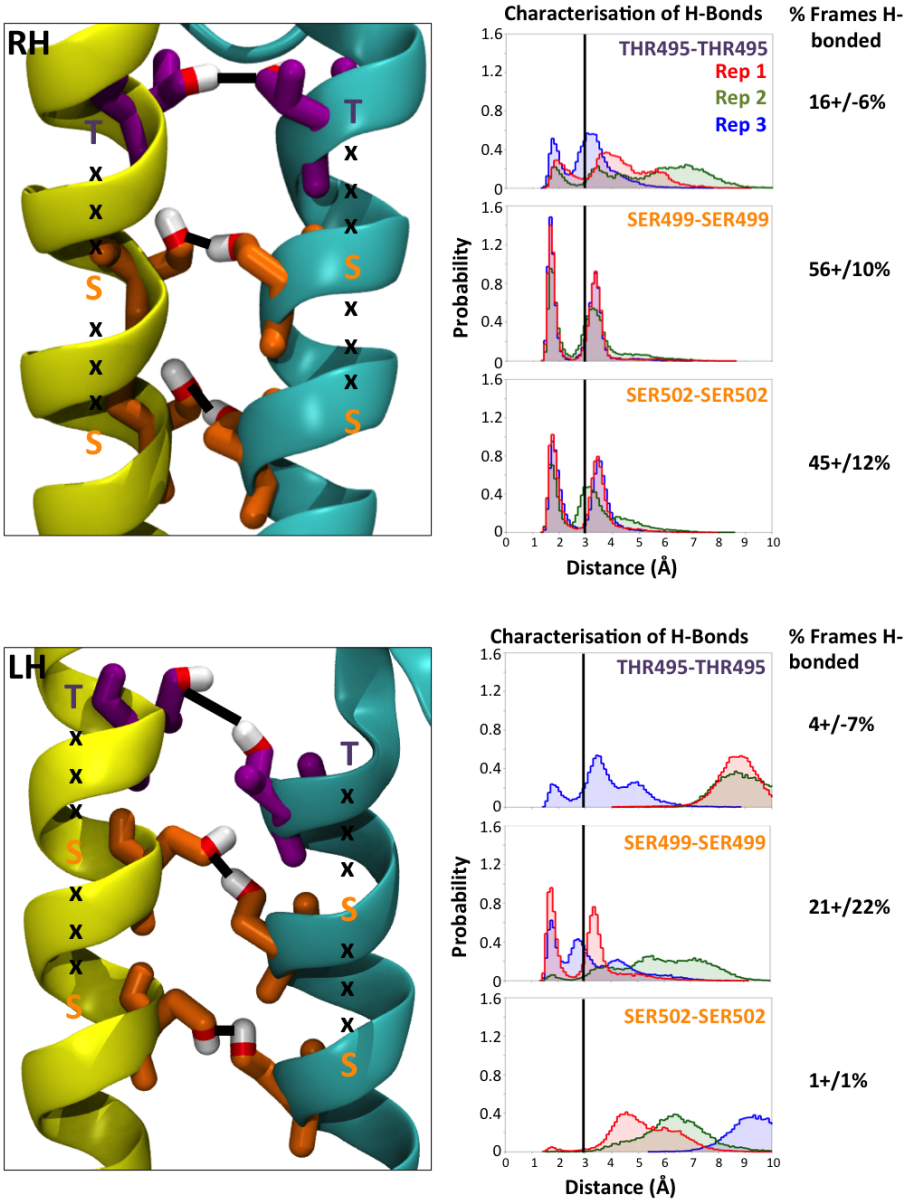

Figure 3.16: Hydrogen Bonding Analysis of the interface in RH (A) and LH (B) dimers, with the distances calculated shown in black lines. Each TM is in either cyan or yellow cartoon representation, the threonine in the interface is violet licorice, and the serines in the interface are represented by orange licorice. For the H-bond plots, replicates are separated and represented by either blue, red or green histograms. The 3Åmark is shown as a black line.

change on the RDF plots, the maximum height of the POPC curve decreases as the extension is increased, indicated by the purple line in Figure 13. POPS, therefore, has more propensity to interact with the dimer. This is further evidenced by the Volmap plots (Figure 3.14A, B, and C). As the propensity for the POPC to interact

with the protein decreases, the volume around MuSK occupied by POPS increases.

Three representative frames of MuSK simulations with a crossing angle that corresponded to both the LH and the RH modes were then chosen, converted to atomistic representation, and simulated for 100ns. Each simulation maintained its initial handedness, with none of the simulations switching between a left- or right-handed dimer (Figure 3.15A). The contacts for each simulation were then calculated, and 4 replicates (3 RH, 1LH) maintained the TxxxSxxxS motif previously seen in the coarse-grained simulations (Figure 3.15B). Lipid interactions were also calculated (Figure 3.15C), and mapped onto the sequence according to the likelihood of interaction with POPS. Overall, it appears that the preferred lipid contacts made are with cationic residues, preferentially the arginines and lysines in the beginning of the TM, and the first few residues of the JM sequence.

The hydrogen bonds for both the RH and LH interfaces were calculated, as shown in Figure 16. For each of the three simulations, it was determined that the SxxxS motif previously predicted and seen in simulations is more tightly packed in right-handed dimers than in left-handed dimers, as there are more hydrogen bonds formed between the serines in the right-handed dimers than the left-handed dimers. The threonine contact was overall less stable than the serine contacts, but it was still more prominent in the RH dimer than in the LH dimer, indicating that the RH dimer is more tightly packed than the LH dimer.

3.8 Conclusions and Relation to MuSK Biology

Determining packing motifs of TM helices is an important step in determining the full activation mechanism of a receptor. This dimerisation is crucial to carrying a signal from the outside to the inside of a cell. Even though the structure and interacting

residues were not able to be determined by NMR, an alternative to this is running MD simulations. Running many simulations in which the two helices self-assemble was the method chosen for this problem. However, with an increased number of simulations, determining macroscopic properties and convergence remain a difficult task. Using the iterative jackknife method, we are able to see that the macroscopic properties of both systems are not affected that much, and the jackknife is able, even with 50-77 simulations, to determine differences between two RTK systems. It could also determine the most prominent contacts for both RH and LH dimers, though as it stands, it is not a sensitive enough technique to detect a potential mechanism for helix dimers switching between handedness.

Convergence is an important thing to assess in a system, as it can determine whether or not there is sufficient sampling. There has recently been another simulation study looking at helix dimerisation and convergence in detail[125], which also aims to determine the optimal number of simulations to accurately describe properties of TM helix dimer systems. Wassenaar et al. have determined that the number of simulations needed for sufficient sampling is significantly different than those presented in this thesis (N=500 simulations needed at a minimum). The potential reasons for these differences is the system they have chosen is different than the ones presented in this thesis (GpA dimerisation, trimers, and GPCR dimers), as well as the parameters which they chose to assess convergence. This difference is quite significant, however, and it could be because Wassenaar et al. chose to look at energies and Euler angles, rather than solely at the crossing angles and the contact matrices. Another reason could be that the bootstrap was used to sample with replacement, rather than using a special case of the bootstrap, the jackknife, and sampling without replacement.

Receptor Tyrosine Kinases are known to interact with lipids, and even be regulated by lipids[22]. As it has been shown that all RTKs recruit anionic lipids, it follows to

ask the question of whether or not the lipids have any effect on the modulation of the dimerisation of MuSK, or whether they have a different role entirely. We have examined three simple lipid systems, and found that there was not very much difference on whether or not MuSK indeed dimerises, nor the contacts maintained when the system dimerises. However, lipids are preferentially clustering around the positively charged juxtamembrane region, as when there is more charged added onto the juxtamembrane region, this attracts more anionic lipids. This could be recruitment of lipids for signalling, or localisation to lipid rafts. In any case, both the interacting residues and the lipids present around the protein have their distinct roles in signalling transduction.

4

Dok7's Interaction With A PIP-Containing Membrane

4.1 Role of Dok7 in MuSK Pathway

Downstream-of-Kinase 7 (Dok7) is a 500-residue cytoplasmic protein[55] that is known to interact with MuSK. It is the cytoplasmic activator of MuSK[12]. Defects in this protein lead to myasthenia gravis[57–63]. A crystal structure of the first 210 residues of Dok7 has been published[56] (Figure 4.1), which excludes the (unstructured) C-

terminal half of the protein. Dok7 was crystallised as an asymmetric homodimer (Figure 4.1A). Each monomer containing a pleckstrin-homology (PH) domain and a phosphotyrosine binding (PTB) domain. It was co-crystallised with a 13-residue peptide containing part of the juxtamembrane (JM) sequence of human MuSK, with a phosphorylated tyrosine residue ($pTyr_{553}$), which is known to bind to Dok7.

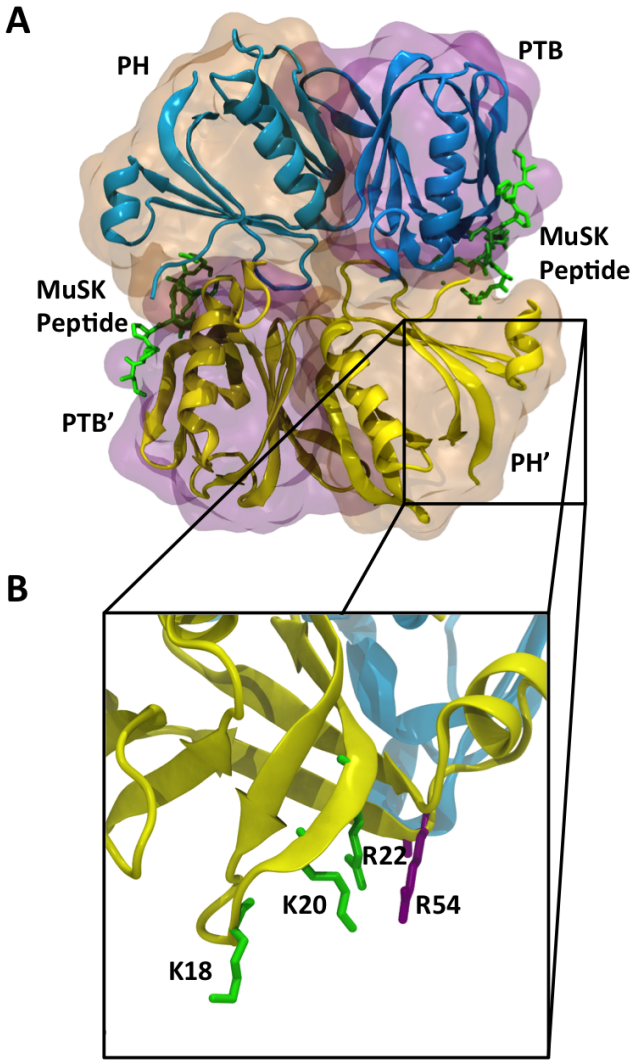

Figure 4.1: Crystal structure of the first 210 residues of Dok7 (PDB ID 3ML4). Each monomer is either yellow or blue cartoon representation, with each PH domain highlighted in surface orange and each PTB domain in in surface purple. The MuSK peptide is shown in green stick representation. Representations adapted from Bergamin, et al.[56] B. The positive patch that is proposed to bind PIP lipids. The backbone of the protein is in yellow cartoon representation. The residue that was mutated to alanine (R54A) and tested experimentally is in purple stick representation, whereas the residues postulated to interact with PIPs (K18, K20, R22) are in green stick representation.[56]

As covered in Chapter 1, PH domains are known to bind phosphatidyl inositol phosphate (PIP) lipids[75;76]. Bergamin, et al.[56] then tested the proclivity of Dok7 to bind PIP lipids via fluorescence polarisation, and it was found to bind to a variety of derivatised PIP lipids, namely PI-(4,5)P_2, PI-(3,4,5)P_3 and PI-(3,4)P_3 lipids. Upon mutation of a key residue (R54) to alanine, the authors found there was a loss in PIP binding compared to the wild-type. They therefore determined that this residue was of importance in the localisation of Dok7 to the bilayer and binding to PIP lipids[56].

The authors also postulated that three residues (K18, K20 and R22) were also involved in PIP binding, as they were close in proximity to R54 in the crystal structure and formed a highly basic region, which would be ideal for a highly negative lipid to bind (Figure 4.1B). However, the authors were unable to determine the mechanism by which Dok7 associates with the bilayer. As this was yet unknown, the aim of this chapter is to determine the mechanism of association of Dok7 to PIP-containing bilayers via molecular dynamics (MD) simulations.

4.2 Multiscale Approach for Peripheral Membrane Proteins

To determine how Dok7 associates with the bilayer, a multiscale approach was employed. First, the protein of interest is converted to a coarse-grained representation, with an elastic network applied to the protein. Then, the protein is placed in a box 7 nm away from a preformed model membrane containing 356 lipids. The box is then solvated with CG water, ions are added at a concentration of 150mM, and energy minimised. This process is repeated at least 20 times, with the protein being placed in a random orientation relative to the bilayer in each respective box to ensure independence of the replicates. The replicates are then equilibrated for 2.5ns and

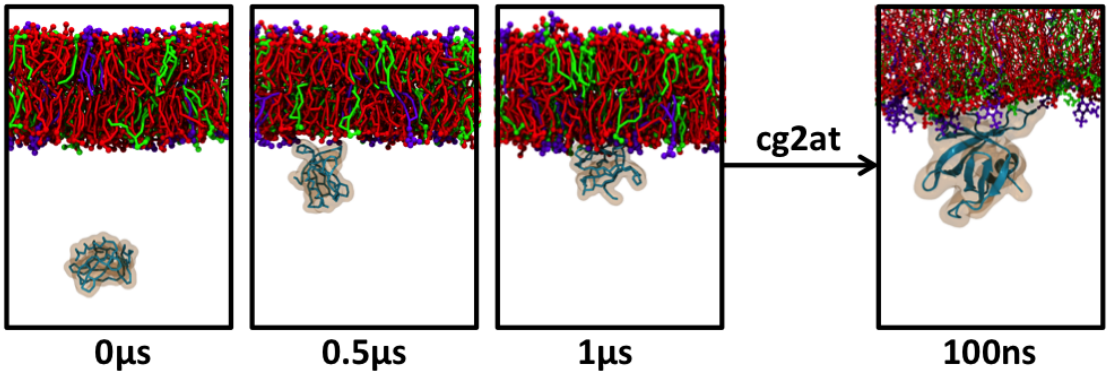

Figure 4.2: A multiscale approach. The protein is placed away from a model bilayer, and allowed to diffuse toward the bilayer for 1μs. A representative frame is chosen, converted to atomistic representation using a fragment-based approach, and simulated for 100ns. POPC lipids are in red, POPS lipids are in green, and PIP_2 lipids are in purple. The protein is in either blue DynamicBonds or cartoon representation, with orange surface on the outside of it.

submitted to a local server, where the simulations run for 1μs. The production runs are then downloaded, stored, and analysed. After analysis is completed, representative frames are then converted to atomistic representation using a fragment-based approach[113] and simulated for 100ns (Figure 4.2). All models and mutations were made using MODELLER[115].

All CG-MD simulations were run with GROMACS 4.5.5, the Berendsen thermostat and barostat were used for temperature (T=323K) and pressure coupling[122]. The LINCS algorithm was used to constrain bond length[99]. The MARTINI2.1[107;112] forcefield was used. The timestep used was 20fs at a temperature of 310K. There were approximately 15,000 waters in each CG box, which measured approximately 11nm x 11nm x 23.5nm.

All AT-MD simulations were run using the GROMOS 53a6 forcefield[132]. The LINCS algorithm[99] was used to constrain bond lengths, while the Parinello-Rahman barostat[133] and Berendsen[122] thermostat was used (T=310K). The box length was approximately 10.5nm x 10.5nm x 13.5nm, with Particle Mesh Ewald (PME)[118] used for long-range electrostatics. Equilibration was run on these systems for 1 ns with the

Table 4.1: Simulations Performed in the Dok7 Study

Simulation	Protein	Duration
	Coarse-Grained Simulations	
CG-PH-pip2	Dok7 PH	20 x1 μs
CG-PH-pip3	Dok7 PH	20 x 1μs
CG-PH-pc	Dok7 PH	20 x 1μs
CG-PH-pcps	Dok7 PH	20 x 1μs
CG-PTB-pip2	Dok7 PTB	20 x 1μs
CG-PTB-pip3	Dok7 PTB	20 x 1μs
CG-Dok7m-pip2	Dok7 (PH-PTB)	20 x 1μs
CG-Dok7m-pip3	Dok7 (PH-PTB)	20 x 1μs
CG-Dok7m-pc	Dok7 (PH-PTB)	20 x 1μs
CG-Dok7m-pcps	Dok7 (PH-PTB)	20 x 1μs
CG-Dok7R54E-pip2	Dok7 (PH-PTB;R54E)	20 x 1μs
CG-Dok75mut-pip2	Dok7 (PH-PTB;5mut)	20 x 1μs
CG-Dok7R54A-pip2	Dok7 (PH-PTB;R54A)	20 x 1μs
CG-Dok7d-pip2	Dok7$(\text{PH-PTB})_2$	50 x 1μs
	Atomistic Simulations	
AT-PH-pip2	Dok7 PH	3 x 0.1μs
AT-m1-pip2	Dok7 (PH-PTB) M1	3 x 0.1μs
AT-m2-pip2	Dok7 (PH-PTB) M2	3 x 0.1μs
AT-d1a-pip2	Dok7 $(\text{PH-PTB})_2$ M1A	3 x 0.1μs
AT-d1b-pip2	Dok7 $(\text{PH-PTB})_2$ M1B	3 x 0.1μs
AT-d2-pip2	Dok7 $(\text{PH-PTB})_2$ M2	3 x 0.1μs
AT-d3-pip2	Dok7 $(\text{PH-PTB})_2$ M3	3 x 0.1μs

*In all simulations a POPC/POPS/PI(4,5)P_2 (75%/20%/5%) was used, with the exception of CG-PH-pc and CG-Dok7m-pc for which a POPC (100%) was used, and CG-Dok7m-pcps for which a POPC/POPS (80%/20%). Additionally in the CG-PH-pcpspip3, CG-PTB-pcpspip3 and CG-Dok7m-pcpspip3 a POPC/POPS/PI(3,4,5)P_3 (75%/20%/5%) was used.

Cα atoms restrained (force constant is 10 kJ/mol/Å^2). This was followed by 0.1μs of unrestrained simulation. Analysis was performed using GROMACS[116], VMD[126], Bendix[137], and locally written scripts.

4.3 PIP-mediated association of Dok7 to the bilayer

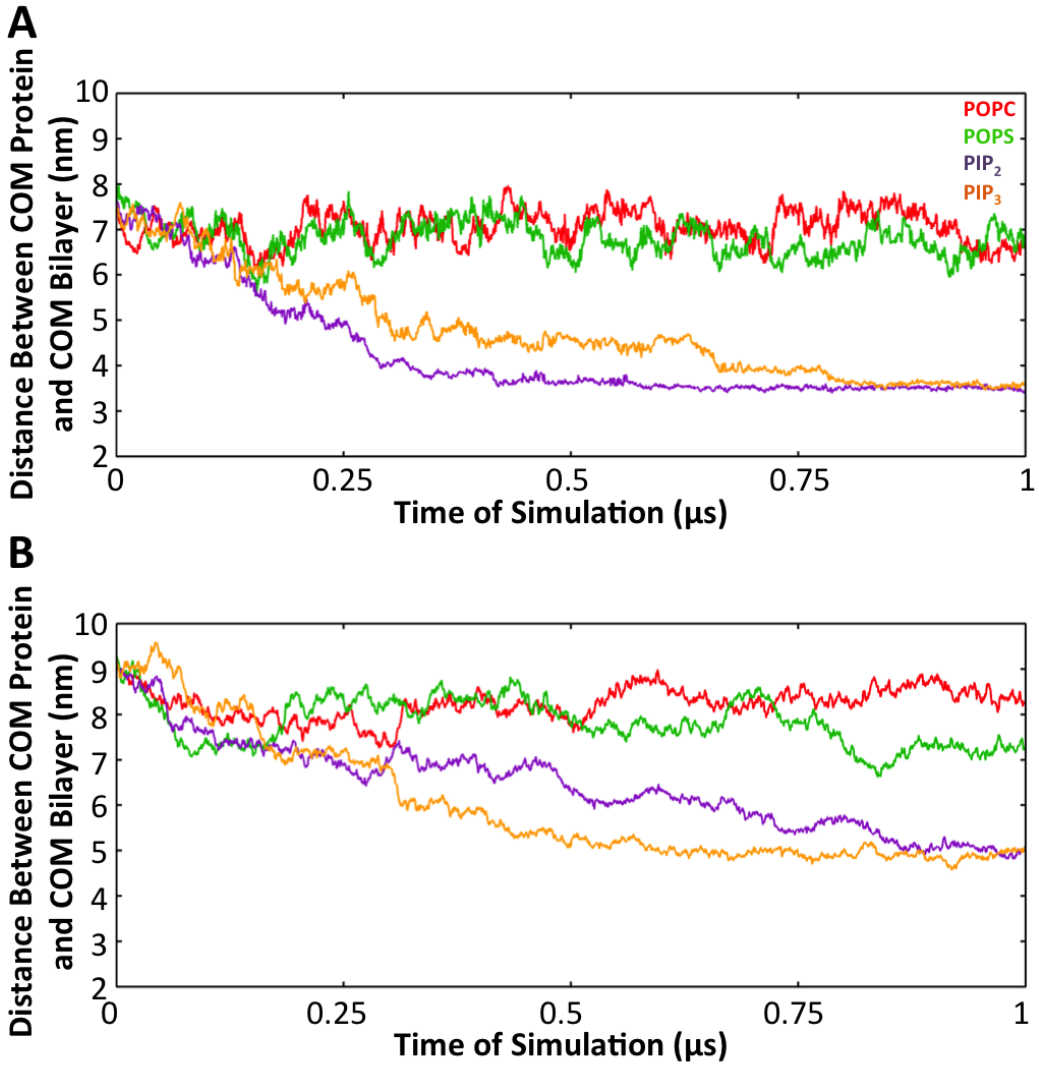

Figure 4.3: Average distance over time for Dok7(PH) (A) and Dok7(PH-PTB) (B) for four different lipid bilayer compositions. COM is the Centre of Mass of either the protein or the bilayer. The red lines corresponds to 100% POPC, the green lines correspond to a bilayer with 80% POPC + 20% POPS, the purple lines are a bilayer with 75% POPC + 20% POPS + 5% PIP_2, and the orange lines correspond to a bilayer with 75% POPC + 20% POPS + 5% PIP_3. All 20 simulations were averaged at each time step in the simulation.

Control simulations were first run with both Dok7(PH) and Dok7(PH-PTB) to see whether or not the presence of PIP lipids was necessary for association of Dok7 to the membrane, and to check if it would bind in a canonical form similar to other

PH domains. The bilayers used were 100% POPC, 80% POPC+20%POPS, and 75% POPC+20% POPS+5% PIP_2 or PIP_3 lipids. When both systems were finished and analysed, it was found that the PIP lipids were indeed necessary for association of Dok7 to the bilayer (Figure 4.3), with no individual dissociation events seen in 1μs of simulation. Since Dok7 is more likely to interact with PIP_2 lipids rather than PIP_3, due to the higher concentration present in the plasma membrane[128], it was decided to continue with all PIP bilayers containing PIP_2.

4.4 Association of Dok7 constructs to PIP_2-containing bilayers

Three different substructures of the crystal were built using the crystal structure of Dok7: Dok7(PH), Dok7(PH-PTB), and Dok7(PH-PTB)$_2$. The Dok7(PH) construct was built to determine the behaviour of the PH domain of Dok7 on its own, to determine if it behaves like the PH domains of other proteins. The Dok7(PH-PTB) construct was built to determine how the addition of a PTB domain would alter the binding behaviour of the protein. Finally, the Dok7(PH-PTB)$_2$ was built to determine the behaviour of the dimer in the presence of the membrane, as it is in a dimer when it activates MuSK, as described in Section 1.3.

4.4.1 Calculation of Energy Landscapes and Convergence Analysis

Energy landscapes were calculated using all 20 simulations of each system. First, a template structure of the preferred orientation of the protein on the membrane was chosen. This template is verified against the canonical binding mode of most PH

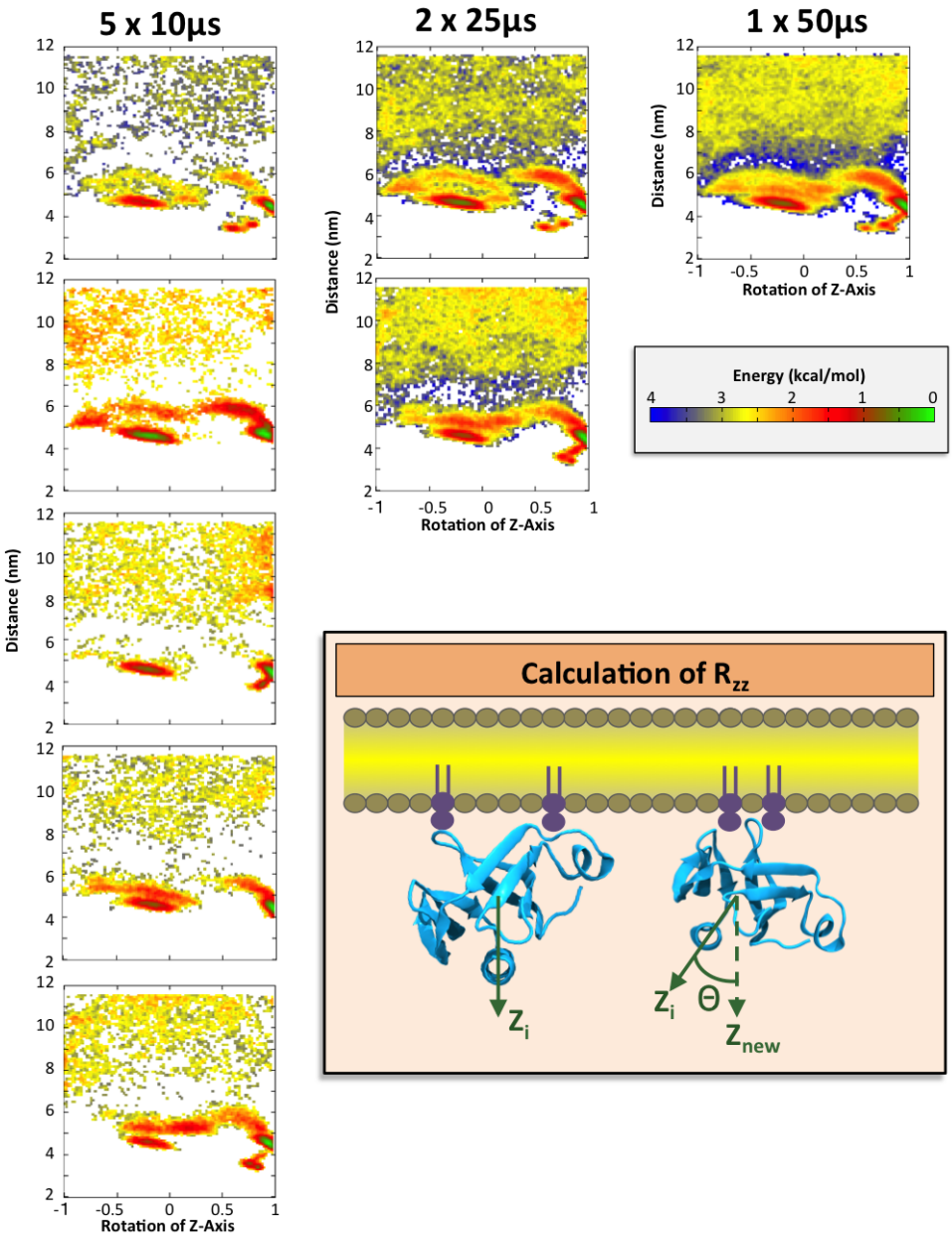

Figure 4.4: Calculation of energy landscapes and convergence analysis for Dok7(PH-PTB)$_2$. The first column of energy landscapes is 5 groups of 10 simulations, which are each 1μs in duration. The second column is 2 groups of 25 simulations, again each 1μs each. The final column is all 50 simulations run for Dok7(PH-PTB)$_2$ in one energy landscape. The energy bar for all simulations is below the energy landscape in the third column. A diagram showing the calculation of the energy landscapes is inset. The protein is aligned to a template structure (an example of which is on the left), and the change in angle in the zz vector is determined (right). This is plotted against distance from the COM of the bilayer. The proteins are in blue cartoon representation, the zz vectors are in green arrows, the head groups are in orange ovals, and PIP_2 is represented by purple ovals and lines.

domains, which is where a structurally conserved helix lies parallel to the membrane, as described in Chapter 2 (Figure 4.4). After the selection of the template, the trajectories were pre-processed to centre the trajectory on the protein, as well as remove rotation about x and y, to just observe rotation about the z axis. After this step, the protein of each frame was aligned with the given template structure. The change in the zz vector upon alignment (Figure 4.4) was then tabulated, as well as the distance of the centre of mass (COM) of the protein and the COM of the bilayer. This was done for all 20 simulations. The data was then plotted with the distance between the COM of the protein and the COM of the bilayer versus the rotation of the zz vector, giving a frequency plot. Then, a reverse Boltzmann transformation was taken to get approximate energies. Finally, the global minima was shifted to zero.

Next, convergence was assessed by taking the most complicated system (Dok7(PH-PTB)$_2$) and running a convergence analysis with 50 simulations (Figure 4.4). The simulations were grouped in the following way: 5 groups of 10 simulations, each 1μs, 2 groups of 25 simulations, each 1μs, and 1 group of all 50 simulations, each 1μs. For 10 simulations, the basic details of the energy landscapes are present in all simulations, but all features are not represented in all five energy landscapes in the first column of Figure 4.4. In the second column, all parts of the energy landscape are present, as well as in the third column. In addition, convergence has been reached since the depth of the wells has not changed, no matter how many simulations are added. The only change observed is the amount of sampling of the protein in solution.

4.4.2 Dok7(PH)

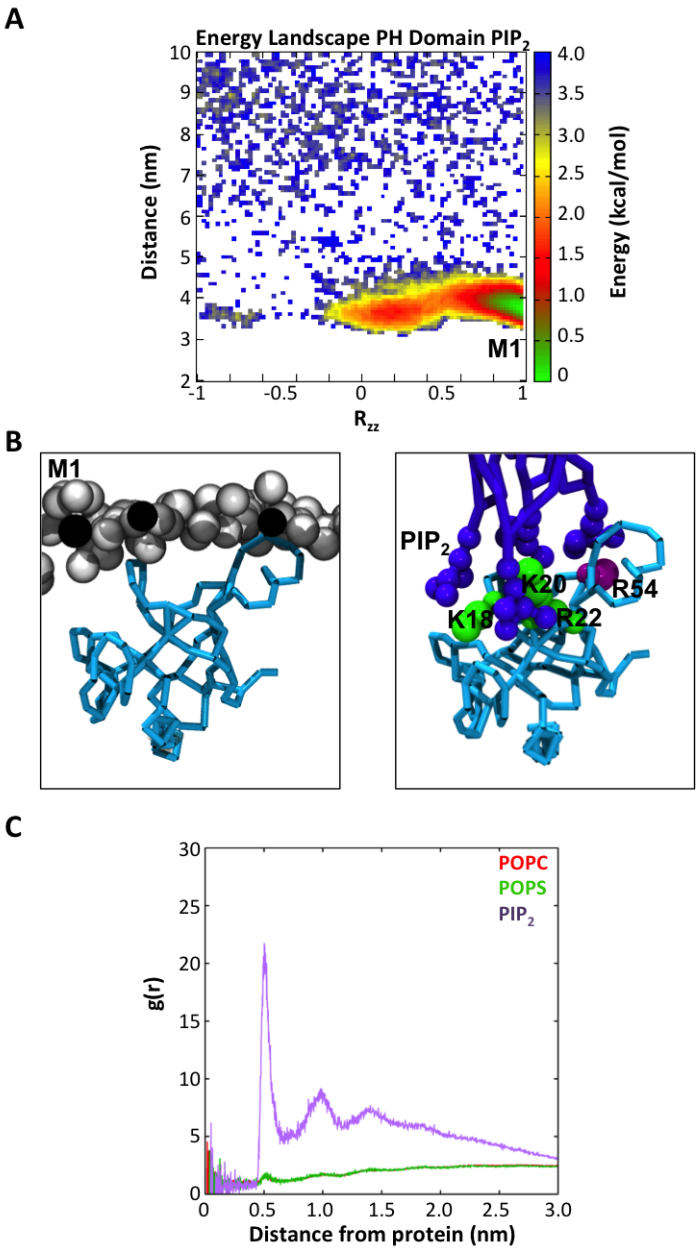

Figure 4.5: A. Energy landscape of the PH domain. B. Orientation of the PH domain in M1 (left). The protein is in blue DynamicBonds representation, and the lipid head groups are in white spheres representation. This orientation is also binding PIP_2 lipids (right), with K18, K20, R22 (in green spheres) and R54 (purple spheres) highlighted. PIP_2 lipids are in purple licorice and spheres representation. C. A representative RDF of the leaflet where the PH domain is bound. PIP_2 is the purple line, POPS is the green line, and POPC is the red line.

First, Dok7(PH) was analysed for preferred binding modes and lipid interactions. Only one binding mode (M1) is seen on the energy landscape (Figure 4.5A).

Upon inspection, Dok7(PH) is found to bind in a canonical form similar to most PH domains[79–81], with the conserved structural helix oriented parallel to the bilayer, and with the PIP_2 head group hugged in between a short, conserved loop between β strands 1 and 2, and the helix between β strands 3 and 4 (Figure 4.5B).

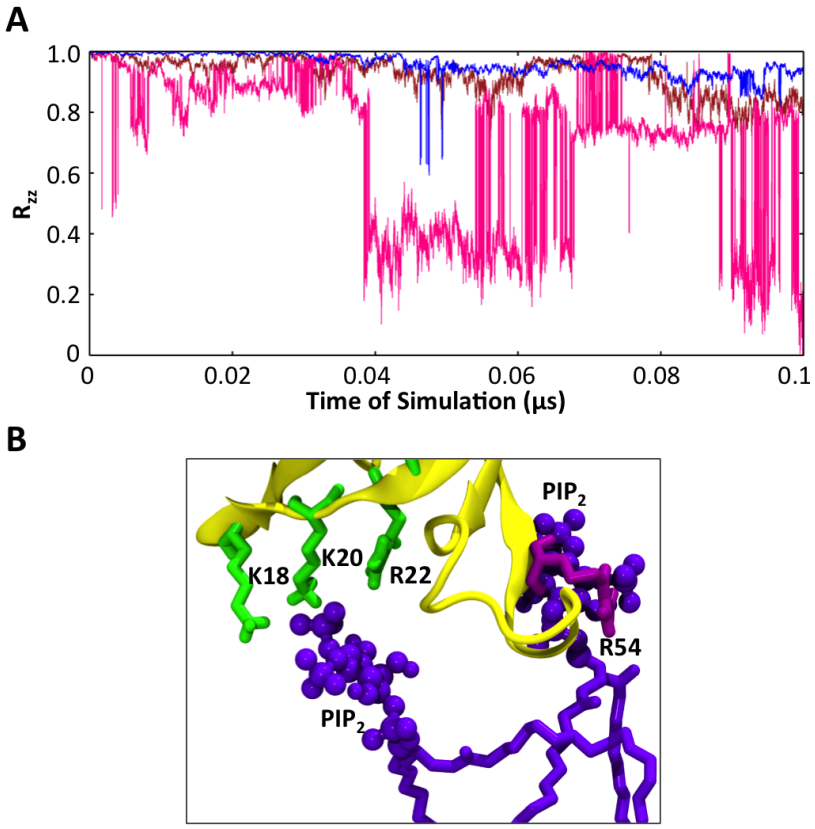

Figure 4.6: Atomistic simulations of Dok7(PH). A. Change in R_{zz} relative to time of simulation, with R_{zz} being the rotation of the frame relative to the provided template. Each replicate is shown separately in either a pink, blue or rust line. B. Example of atomistic interactions with lipids. PIP_2 is in purple CPK and licorice representation, and R54 is in purple licorice representation. K18, K20, and R22 are in green licorice representation, and the PH domain is in yellow cartoon representation.

This is the aforementioned positive pocket present in the PH domain, with K18, K20, R22 and R54 preferentially interacting with the PIP_2 head group. This was further quantified by calculation of the radial distribution functions (RDFs) of the

different lipid species around Dok7(PH) (a representative RDF in Figure 4.5C). The RDFs show a large peak at 0.5nm, for PIP_2, with a second peak at 1.0nm, indicating that PIP_2 preferentially clusters around the PH domain. This clustering was not observed for POPC or POPS, and only the leaflet to which Dok7(PH) was bound was considered in this calculation.

Three frames of the preferred orientation of Dok7(PH) were chosen, converted to an atomistic representation, and simulated for 0.1μs. During these simulations, the protein maintained its orientation in 2 of the 3 replicates (Figure 4.6A) throughout the simulation, along with the preferred lipid contacts identified from the CG simulations (residues K18, K20, R22 and R54, Figure 4.6B).

4.4.3 Dok7(PH-PTB)

Next, the energy landscape of Dok7(PH-PTB) was calculated and compared to Dok7(PH) to determine what effect the presence of another domain would have on the energy landscape, as well as the mechanism of which Dok7 would associate with the bilayer.

When the PTB domain was added to the model, the energy landscape becomes slightly more complicated, revealing two binding modes: one corresponding to Dok7(PH-PTB) only associating with the membrane via its PH domain (M1), and the other corresponding to Dok7(PH-PTB) lying flat on the membrane, with both domains binding to the bilayer (M2). (Figure 4.7)

It was found that Dok7(PH-PTB) first associates with the bilayer via its PH domain (M1), and then the PTB domain interacts transiently with the bilayer before eventually finding a favourable orientation and settling onto the bilayer (M2).

To determine whether or not this interaction was transient over longer timescales, three simulations that finished in M1 were extended to 5μs. When the distance

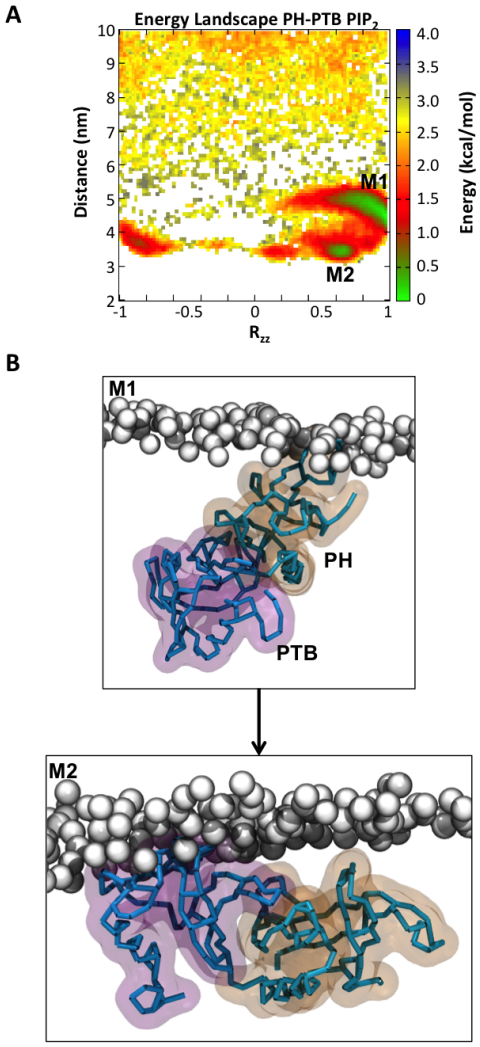

Figure 4.7: A. Energy landscape of Dok7(PH-PTB). Two binding modes are labelled (M1 and M2). B. Visualisation of M1 and M2 binding modes from the energy landscapes. The monomer is in blue DynamicBonds representation, the PH domain is highlighted in surface orange, the PTB domain is highlighted in surface purple, and the lipid head groups are in white spheres representation.

between the protein and the bilayer over the time of the simulation was plotted (Figure 4.8A), it was found that M2 was indeed transient, as the distance between the protein and the bilayer switched between 3 nm (M2) and 5 nm (M1) during the 5μs. Since the PH domain anchors itself to the bilayer first, and does not dissociate (Figure 4.8B), then the distance shift seen in Figure 4.8A must be due to the PTB domain dissociating. This is indeed the case, as the distance between the PTB domain

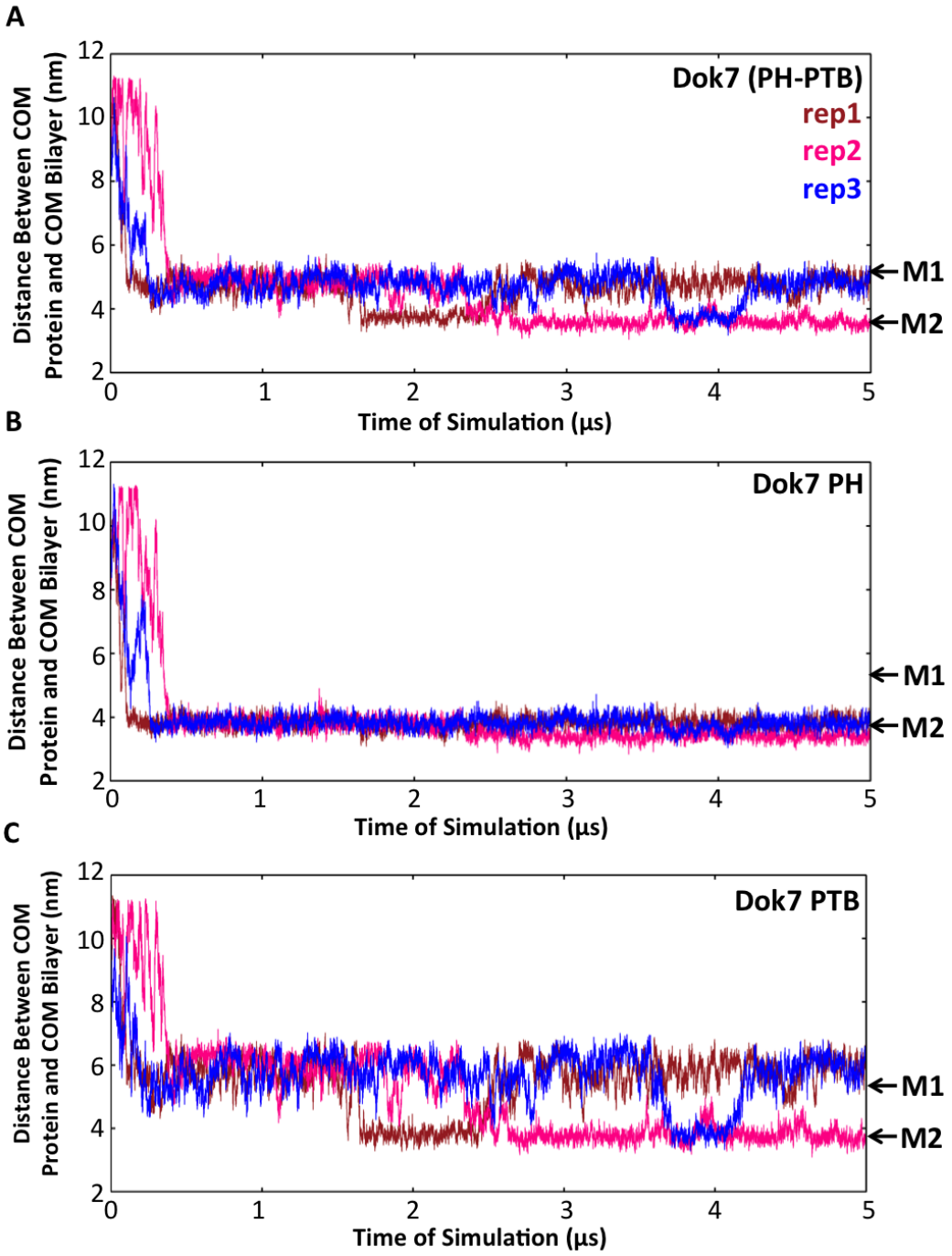

Figure 4.8: Distance between COM protein and COM of the centre of the bilayer vs. overall time of simulation. The distances shown are for Dok7(PH-PTB) (A), Dok7 PH (B), and Dok7 PTB (C). M1 and M2 are highlighted on the graphs with arrows. Each replicate is shown separately - rep1 is a pink line, rep2 is a blue line, and rep3 is a rust line.

and the bilayer fluctuates between 4nm and 6nm, indicating that the PTB domain's transient interactions with the bilayer are the reason for the switch between M1 and

M2.

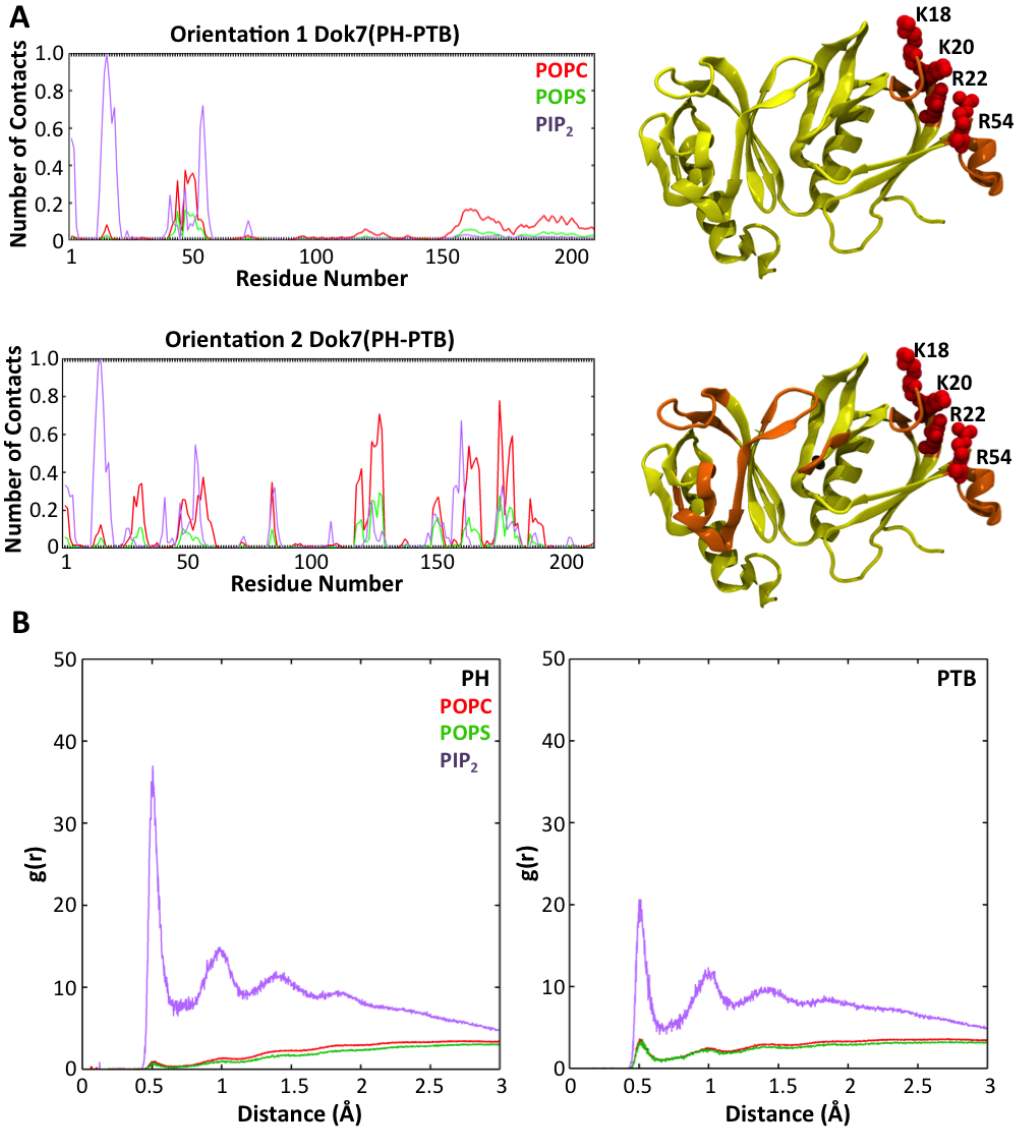

Figure 4.9: Lipid analysis for Dok7(PH-PTB). A. Lipid interactions for each orientation on the energy landscape in Figure 4.8. Normalised lipid contacts (divided by the number of contacts the residue with the most contacts makes) are shown on the left side. All red lines are POPC, all green lines are POPS, and all purple lines are PIP_2. The most prominent lipid interactions from PIP_2 are highlighted in orange cartoon representation on top of the yellow cartoon Dok7(PH-PTB) structure to the right of each graph. K18, K20, R22 and R54 are shown in red spheres. B. Representative RDFs for each domain of Dok7(PH-PTB). PIP_2 is the purple line, POPS is the green line, and POPC is the red line. The RDF of the PH domain is on the left, and the RDF of the PTB domain is on the right.

When the lipid interactions were compared between M1 and M2, the only significant difference between the two modes was the presence of lipid contacts in the PTB domain (Figure 4.9A). When compared to the lipid contacts in Dok7(PH), again, the same positive patch in the PH domain was responsible for interactions with PIP lipids in both M1 and M2 (Figure 4.9A). Interestingly, when the RDFs for each individual domain were calculated for Dok7(PH-PTB), the peak at 0.5nm for the PH domain is larger than the peak at 0.5nm for the PTB domain, indicating that, although there is still clustering of PIP lipids around the PTB domain, PIP_2 preferentially clusters around the PH domain (Figure 4.9B).

Again, three frames of each preferred orientation of Dok7(PH-PTB) were chosen, converted to an atomistic representation, and simulated for 0.1μs. Each orientation was maintained throughout each simulation (Figure 4.10A), as were the lipid contacts seen in the CG simulations (Figure 4.10B). No significant conformational changes were observed for either M1 or M2 in an atomistic simulation.

To test the requirement of this positive patch to promote association, two *in silico* mutants of Dok7(PH-PTB) were made: R54E (harking back to the experimental mutation performed), and K17E, K18E, R20E, R22E, and R54E (hereafter referred to as 5mut). These two models were simulated in pipeline previously described with a 75% POPC+20% POPS+5% PIP_2 bilayer, similar to the WT. Reversing the charge of these basic residues did not affect whether or not Dok7(PH-PTB) associated with the bilayer (Figure 4.11), but its orientation on the bilayer.

Mutating one residue (R54E) did not significantly change the energy landscape, but resulted in more of a "smearing" of M1 (Figure 4.11, middle), presumably because the initial encounter between Dok7 and the membrane was weakened. However, when the 5mut was placed in the simulation box with the same model bilayer, it completely changed how the monomer associated with the bilayer. The 5mut did not fully retain

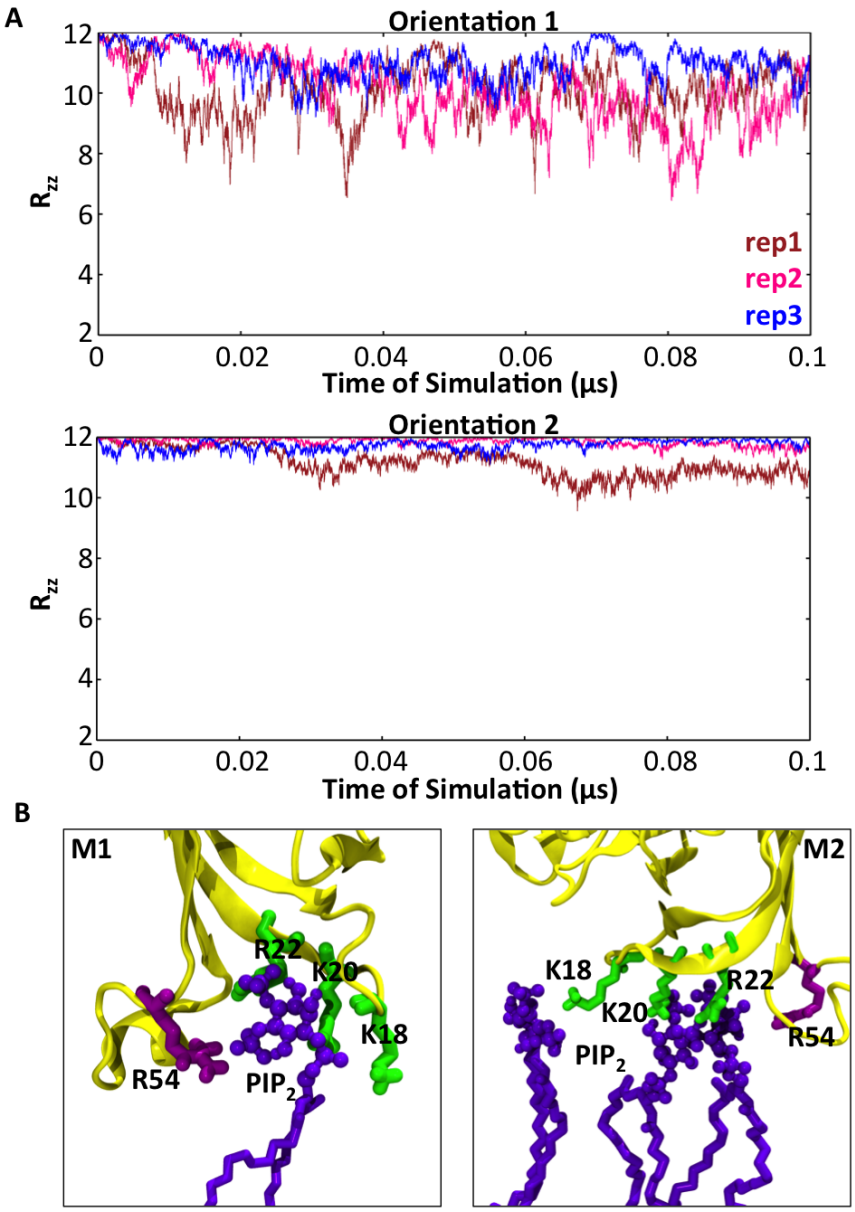

Figure 4.10: Atomistic simulations of Dok7(PH-PTB) orientations M1 and M2. A. R_{zz} (rotation of the zz vector relative to a given template) over the time of simulation. Each replicate is a separate line, with rep1 in pink, rep2 in blue, and rep3 in rust. B. Representative lipid interactions of the atomistic simulations. K18, K20 and R22 are in green licorice representation, while R54 is in purple licorice representation. PIP_2 is in purple licorice and CPK representation, and the PH domain of Dok7 is in yellow new cartoon.

the orientation of M1 compared to the wild type, as the conserved helix is not away from yet parallel to the bilayer; instead , it was in the opposite orientation, where the helix was still parallel, but facing the bilayer. M2 was entirely different, for the protein was lying on its 'back,' as it were, instead of on its 'feet,' resulting in an

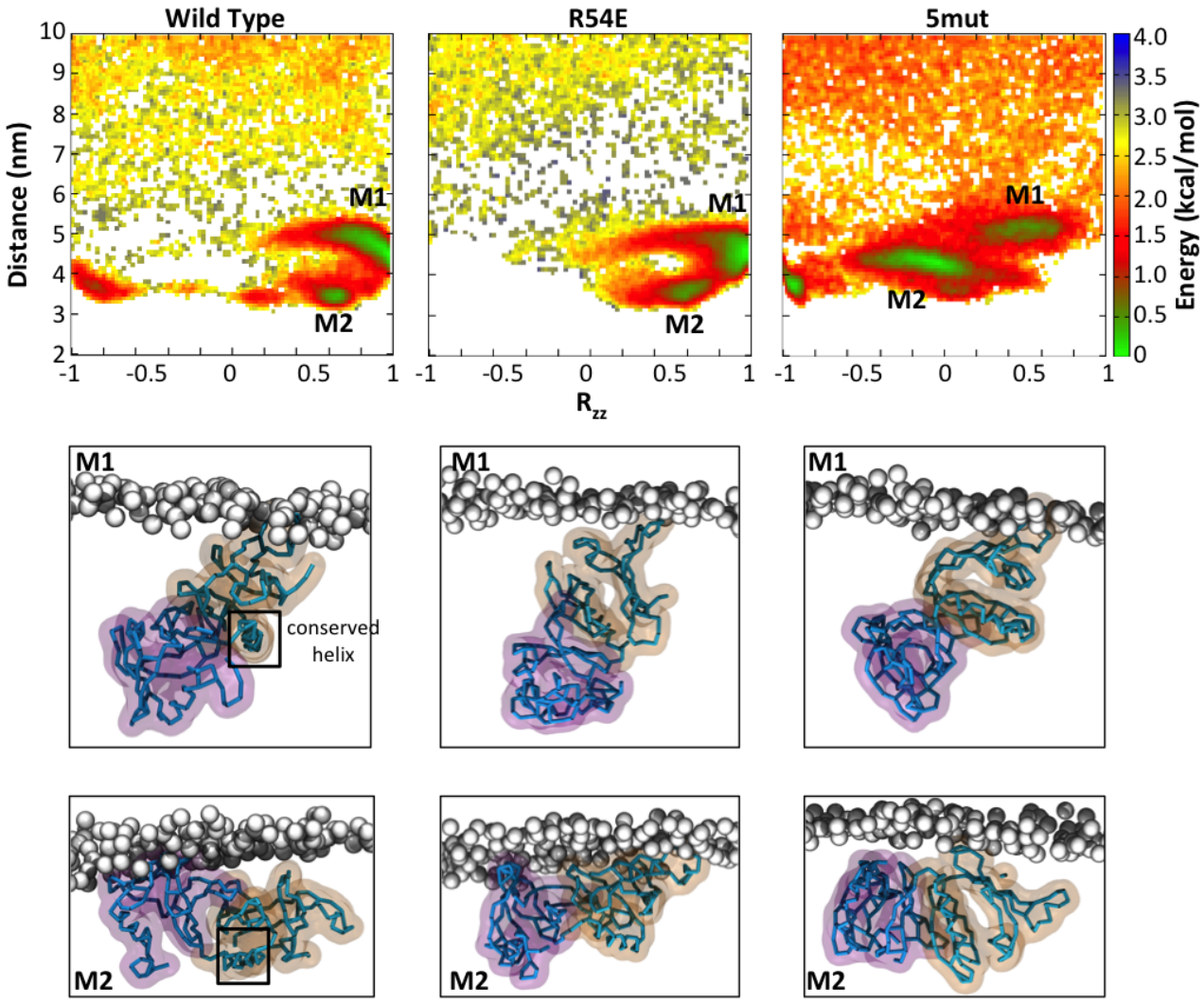

Figure 4.11: Energy landscapes and binding modes of the WT Dok7(PH-PTB), R54E Dok7(PH-PTB) and the 5mut Dok7(PH-PTB). All binding modes are under their respective energy landscape. The backbone of the protein is represented by blue DynamicBonds, the PH domain is in surface orange, and the PTB domain is in surface purple. The lipid head groups are in white spheres representation. The conserved helix is highlighted by a black box.

unproductive orientation for MuSK activation.

4.4.4 Dok7(PH-PTB)$_2$

To determine whether or not other interactions were at play in stabilising the interaction of Dok7 with a PIP-containing bilayer, simulations of the dimeric form of Dok7

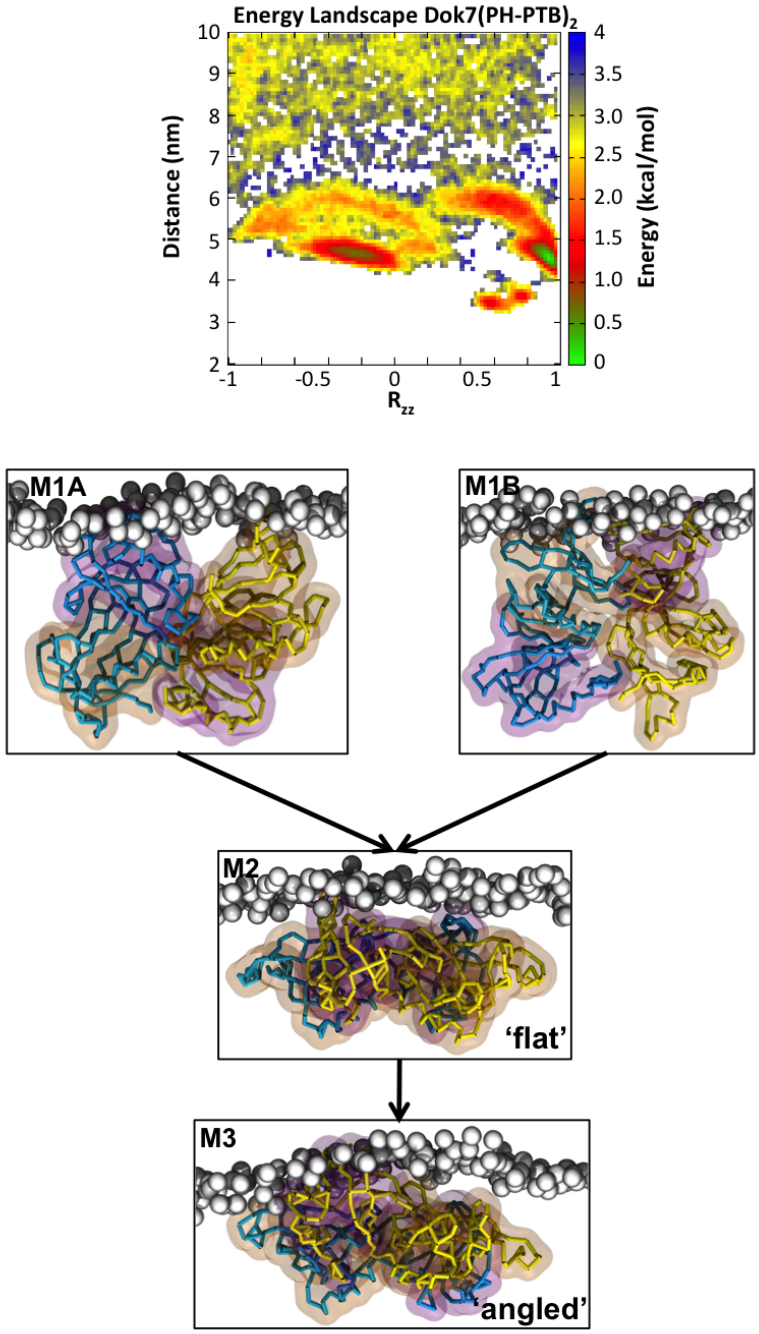

Figure 4.12: Energy landscape of Dok7(PH-PTB)$_2$, with three distinct binding modes (M1A, M1B, M2, and M3). Example images of each binding mode is shown below the energy landscape. Each monomer is in either yellow or blue DynamicBonds representation, with the PH domains in surface orange, and the PTB domains in surface purple. Lipid head groups are highlighted in white spheres representation.

from the crystal structure (omitting the MuSK-derived peptides) were performed. This was done to determine if the PH domain is still the driving force in the associ-

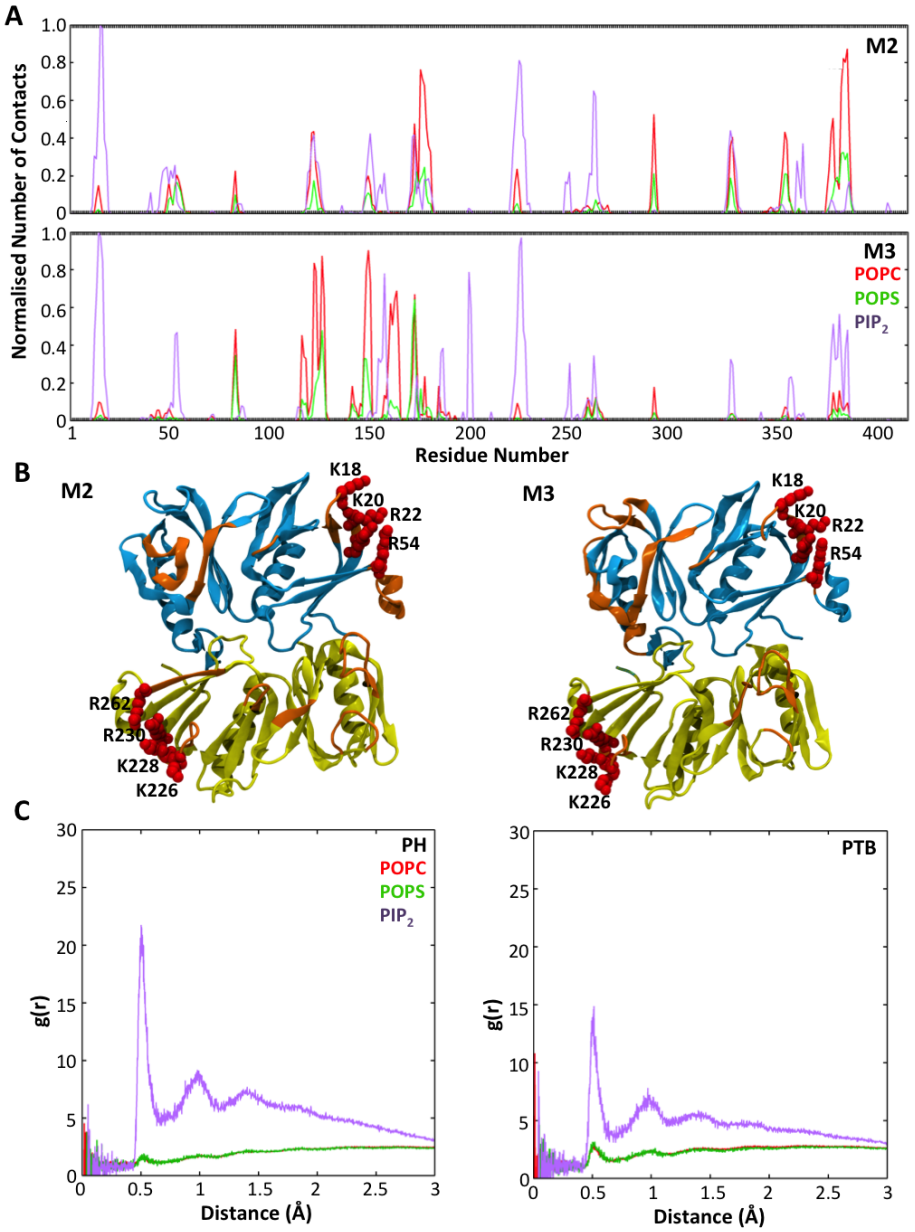

Figure 4.13: Interactions of Dok7(PH-PTB)$_2$ with lipids. A. Normalised number of lipid interactions of M2 and M3 (top two graphs). Red lines are POPC interactions, green lines are POPS interactions, and purple lines are PIP_2 interactions. B. Lipid interactions mapped onto dimer structure (bottom two structures). Each Dok7 monomer is either in blue or yellow cartoon representation, with the prominent PIP_2 lipid interactions mapped onto the structure in orange cartoon representation. The proposed interacting residues (K18, K20, R22 and R54) are in red spheres. C. Representative RDFs of one PH domain and one PTB domain on Dok7(PH-PTB)$_2$. Red lines are POPC, green lines are POPS, and purple lines are PIP_2.

ation of the bilayer of Dok7, as well as to determine what the most likely orientation of Dok7(PH-PTB)$_2$ on the bilayer.

In this case, for 20 x 1μs simulations, the energy landscape contains three distinct binding modes (Figure 4.12). In M1A and M1B. the PH domain of one monomer is bound to the bilayer, along with the PTB domain of the other monomer. These are not each distinct binding modes, as they are of a homodimer and only differ by a 180° rotation. In M2 and M3, the Dok7 dimer is found to lie 'flat' on the membrane, such that the plane of the dimer is approximately parallel to that of the bilayer. In M2, it lies 'flat' on the membrane, such that both PH domains are interacting with the bilayer. In M3, the dimer adopts an 'angled' orientation, where one PH domain is forced away from the bilayer as its partner PTB domain interacts with the bilayer.

The interaction of Dok7(PH-PTB)$_2$ with PIP_2 was quantified for all binding modes. For M1A and M1B, the main lipid interactions were with one PH domain and the opposite PTB domain, with the aforementioned four residues (K18, K20, R22 and R54) playing a significant part in lipid binding. For M2, there were interactions with both PH and PTB domain, but the predominant interactions were with the PH domain (Figure 4.13A). For M3, the interactions with lipids were largely similar to M2, but with the exception of more interactions with one of the PTB domains in the dimer (Figure 4.13B).

Finally, 0.1μs atomistic simulations were performed for the M2 and M3 binding modes, with the frames selected having the same distance and R_{zz} component as the corresponding well of interest on the energy landscape. For the M2 simulations, the Dok7(PH-PTB)$_2$ interaction with the bilayer remained stable throughout the simulation (Figure 4.14A). In the M3 simulations, the PH domain forced away from the bilayer preferentially curls back towards the bilayer to interact with PIP lipids within the first few nanoseconds. There was some conformational change within the protein, between each monomer, with the most significant initial change occurring in M3, rather than M2 (Figure 4.14A). Atomistic simulations were also run for M1A

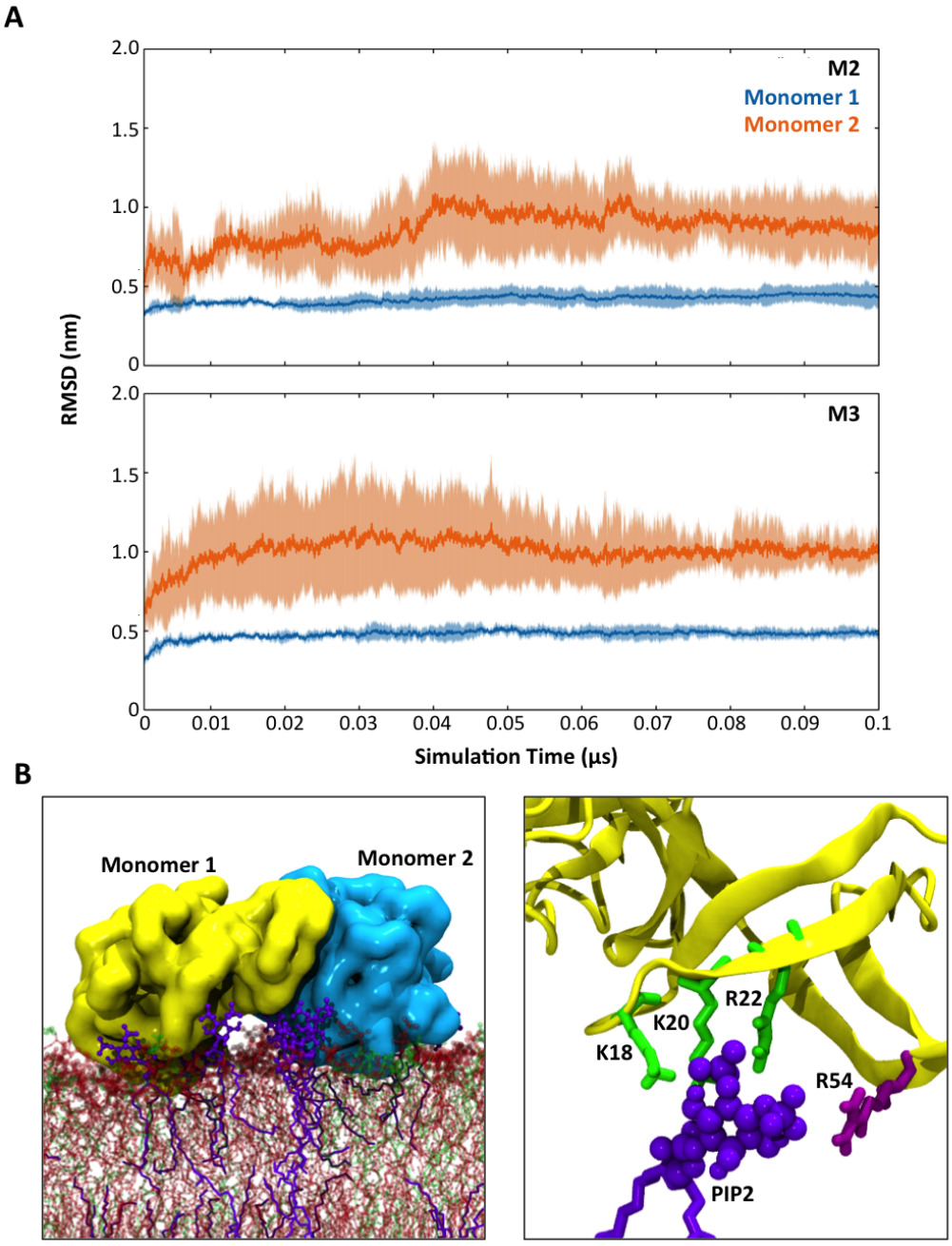

Figure 4.14: A. Average RMSD of three replicates of M2 and M3. The RMSD is calculated for each monomer separately, with the dimer being aligned to monomer 1. Error bars are shown in the surface around the line. B. Example lipid interactions of M2 and M3. K18, K20, R22 (in green stick representation) and R54 (shown in purple stick representation) are still responsible as the main contacts for the interactions with PIP_2.

and M1B, but no significant conformational changes occurred. The aforementioned lipid contacts were prominent throughout the simulation (Figure 4.14B). These ob-

servations, along with the likelihood that the PH domain would, in fact, preferentially interact with the membrane, led us to combine the results from Chapter 3 with this chapter to propose a potential structure in Figure 4.15.

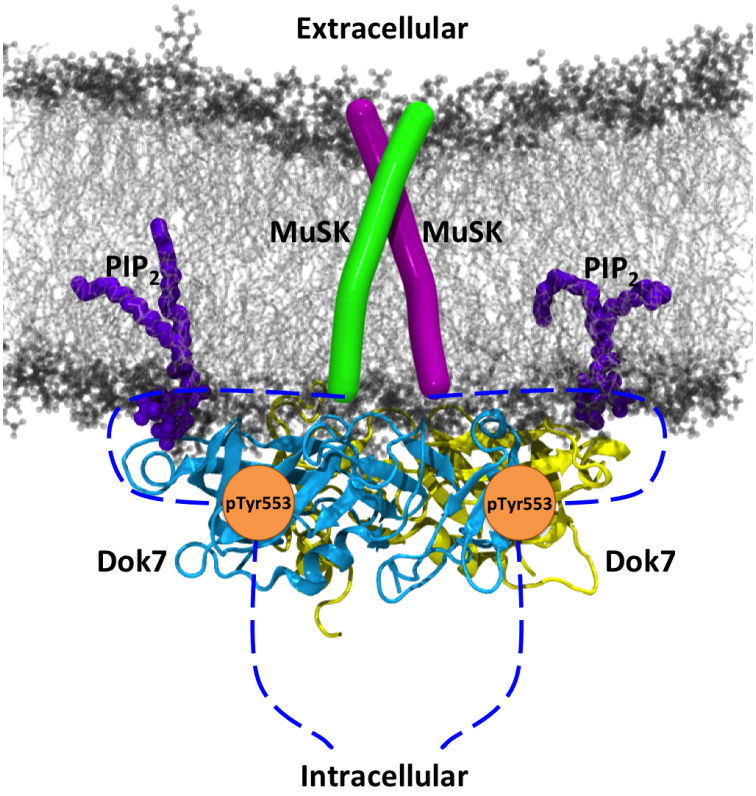

Figure 4.15: The theoretical complex of Dok7 and MuSK. PIP_2 is shown in purple liquorice, MuSK is shown as green and purple bendices, and Dok7 is shown as blue and yellow cartoon representation. The postulated location of the JM region is in a blue dashed line. The location of $pTyr_{553}$ is shown as an orange circle.

4.5 Discussion

From the simulations performed in this study, we can conclude that Dok7 interacts with the membrane first via its PH domain, in a manner consistent with other PH domains. This orientation is where the β1-β2 loop interfaces with PIP lipids. In addition to this loop, a conserved helix, discussed above and in Chapter 1, is parallel to the bilayer. The positively charged residues that were hypothesised to interact with

PIP lipids *in vitro* were shown to interact with PIP_2 *in silico*, even driving association with the bilayer in the Dok7(PH) simulations. These extended to the Dok7(PH-PTB) simulations, where the energy landscape gets more complicated. However, the monomeric form still associated with the bilayer first via its PH domain, followed by the PTB domain transiently associating with the bilayer Dok7(PH-PTB) still interacted with PIP_2 via K18, K20, R22 and R54, which is consistent with the results from the PH domain.

Following on from this, the manner in which Dok7 interacts with the membrane becomes increasingly complicated from Dok7(PH-PTB) to Dok7(PH-PTB)$_2$. As Dok7 is in its dimeric form with MuSK in the membrane, we had to postulate what would be the most favourable orientation. Even though M1 is the most sampled orientation to the membrane, we reasoned that this would not be the most productive orientation, as it places one of Dok7(PH-PTB)$_2$'s PH domains away from the bilayer. This is unlikely to happen *in vivo*, as PH domains are known to associate and prefer to interact with PIP lipids, both in the simulations performed here and previous studies. In addition, both PTB domains would need to be in close proximity to the membrane to bind to $pTyr_{553}$ on MuSK. M1 has one $pTyr_{553}$ binding site away from the membrane, so the MuSK receptor would not be able to fully activate without both $pTyr_{553}$ binding sites nearby to facilitate activation. Having one PH and one PTB domain away from the bilayer would prevent full activation of MuSK, which suggests that M1 is not physiologically relevant. This leaves either M2 or M3 as the most physiologically relevant binding mode. Due to the increased conformational changes in M3 compared to M2, as well as the proclivity of the PH domain to associate and interact with lipid on the membrane, we conclude that the most likely orientation of Dok7 on the bilayer is in M2. This would allow the PTB domain to be available to bind to $pTyr_{553}$.

5

Putting It All Together: the Dok7/MuSK/Membrane Complex

5.1 Background: Dok7/MuSK/membrane interaction

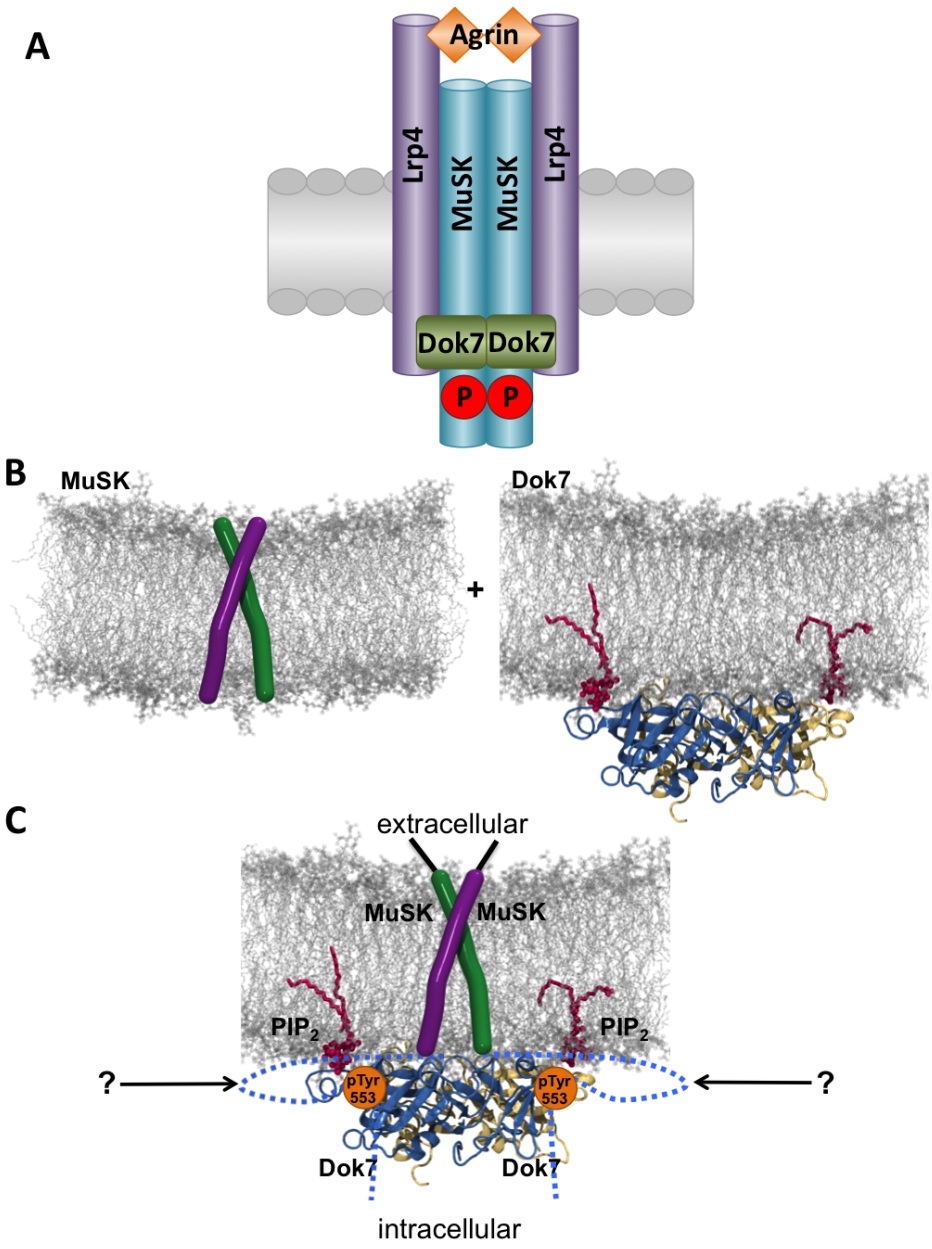

Figure 5.1: A. Schematic of Dok7/MuSK/Lrp4/Agrin complex before starting to activate acetylcholine receptors. MuSK is represented by blue cylinders, Lrp4 by purple cylinders, Dok7 by green rounded rectangles, and Agrin by orange diamonds. Phosphorylated tyrosines are represented by red spheres, and the bilayer is represented both by grey ovals and a grey rectangle. B. Results from Chapter 3 (left) and Chapter 4 (right). The MuSK dimer (left) is shown as green and purple bendices, the Dok7 dimer (right) is shown in blue or yellow cartoon representation, and the PIP_2 molecules are shown in pink CPK and licorice representation. The other lipids in the system are shown as white licorice representation. C. Proposed structure of Dok7/MuSK/membrane complex. MuSK and Dok7 are in the same representations as in B, the proposed juxtamembrane region is in a blue dashed line, and the position of the phosphotyrosine from the crystal structure is shown in an orange circle.

Dok7 is a 500-residue cytoplasmic protein[55] that interacts with and functions to fully activate MuSKwhen Dok7 binds to a phosphorylated tyrosine (pTyr553) on MuSK's JM region (Figure 5.1A). This process furthers a signalling cascade to cluster acetylcholine receptors (AChRs) so the nerve cell can signal to the muscle cell[12]. The crystal structure of Dok7 was solved in 2010 by Bergamin, et al.[56] The first 210 residues were structurally determined, and the dimer of Dok7 was co-crystallised with a 13-residue phosphopeptide corresponding to 13 residues of the juxtamembrane (JM) domain of human MuSK. SEC-MALS data from Bergamin et al. suggests that Dok7 and MuSK interact to form a 1:1 or a 2:2 complex, suggesting that Dok7 binds to MuSK before dimerising[56]. Autophosphorylation data from Bergamin et al. also suggests that Dok7 is needed for full activation of MuSK, as the level of phosphorylation on MuSK increases with the presence of Dok7[56].

In Chapter 3, the most likely MuSK transmembrane (TM) helix dimer structure in a lipid bilayer was determined to be right-handed, due to the extra stability imposed by a threonine residue above the SxxxS motif (Figure 5.1B - left). In Chapter 4, the most likely orientation of Dok7 on the membrane was determined to be a 'flat' orientation, as the shielding of the $pTyr_{553}$ binding site would not be physiologically relevant (Figure 5.1B - right). However, it is unknown whether both MuSK and Dok7 would behave as they have previously when present together in the system. It is also unknown what role the unstructured juxtamembrane (JM) region of MuSK plays when present with both proteins (Figure 5.1C). Thus, this chapter aims to explore what happens when complexes with varying stoichiometry of Dok7 and MuSK are built and simulated with a complex membrane.

5.2 Methods

5.2.1 Building the Models

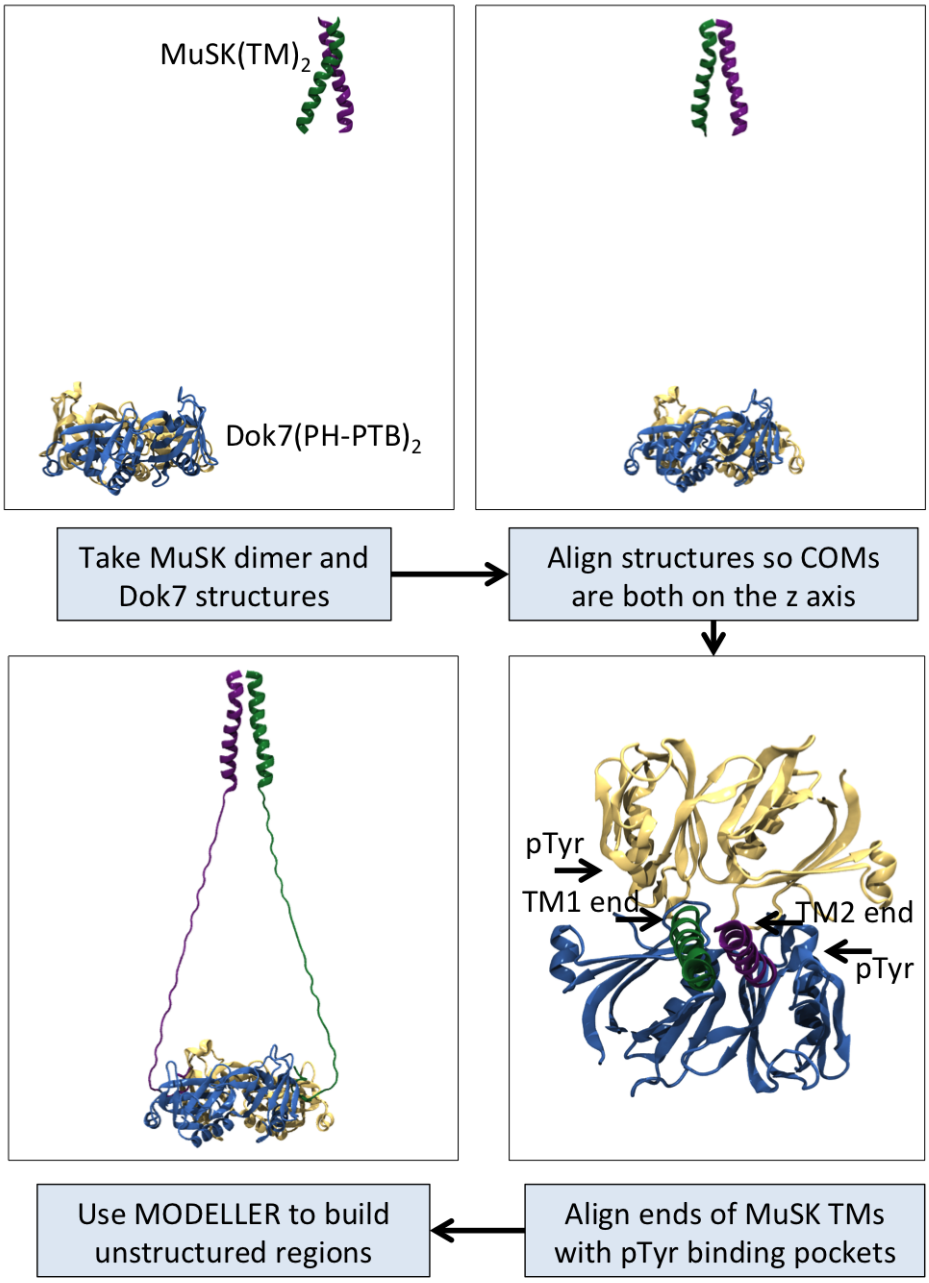

Figure 5.2: Workflow for building a Dok7(PH-PTB)$_N$MuSK(TM+JM)$_N$ model. The unstructured juxtamembrane region of MuSK that connects the two proteins was built fully extended. MuSK(TM) and MuSK(TM+JM) is shown in either green or purple cartoon representation. The Dok7(PH-PTB) is shown in either yellow or blue cartoon representation. The end of the TMs and the pTyrr$_{553}$ binding site is highlighted in the third panel.

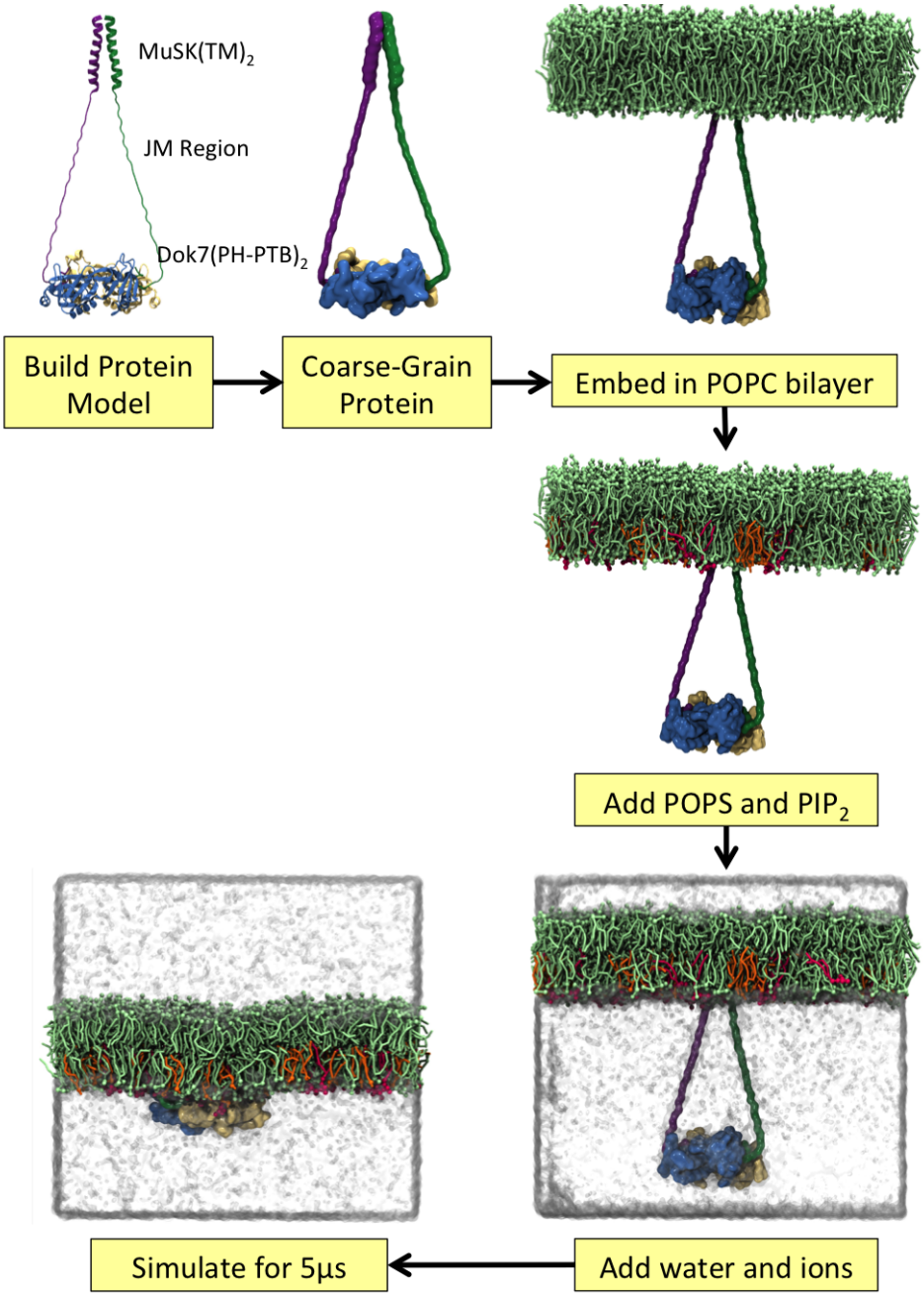

Figure 5.3: Workflow for simulation setup. The atomistic model is shown in cartoon representation, with each MuSK(TM+JM) either in green or purple, and each Dok7(PH-PTB) in either blue or yellow. The coarse-grained model is represented in QuickSurf, with the same colouring scene as the atomistic model. Bilayers are represented in licorice and spheres representation, with POPC in green, POPS in orange, and PIP_2 in pink. Water is represented as white surface representation. Ions are omitted for clarity.

First, the models for the TM helices and the crystal structure of Dok7 (either the monomer or dimer) were arranged so the the COM of the TM helices of MuSK

and the COM of the crystal structure of Dok7 were aligned to the z-axis. Then, the MuSK TM dimer was rotated so the ends of the TM helix facing the Dok7 dimer were in line with the $pTyr_{553}$ binding site, to facilitate building of the juxtamembrane region. After that, the unstructured juxtamembrane region of MuSK was built as a link between the TM helix of MuSK and the PTB domain of Dok7, with the MuSK phosphopeptide in the crystal structure used as a guide for where part of the MuSK juxtamembrane region interacts with Dok7. (Figure 5.2)

Once the model is built using MODELLER[115], with the membrane binding face of Dok7 oriented towards MuSK (and also the bilayer), the protein is coarse-grained using the MARTINI2.1[107] forcefield. The model is then embedded in a POPC bilayer. Using exchange_lipids[128], POPS and PIP_2 are added to the inner leaflet at a ratio of 73%:20%:7% POPC:POPS:PIP_2. Water particles and ions (at a concentration of 150mM) are then added to the simulation box, and then the system is simulated for 5μs (Figure 5.3).

5.2.2 Simulation Parameters

All CG-MD simulations were run with GROMACS 4.5.5, the Berendsen thermostat and barostat were used for temperature (T=323K) and pressure coupling[122]. The LINCS algorithm was used to constrain bond length[99]. The time step used was 20fs at a temperature of 310K. There were approximately 80,000 water particles in each CG box, which measured approximately 21nm x 21nm x 20nm.

The atomistic simulation was generated by taking a coarse-grain snapshot of the system, which corresponds to the distance and R_{zz} value of the energy minima corresponding to the 'flat' orientation of Dok7 on the bilayer, and converting it to all-atom representation using a fragment-based approach[113]. This AT-MD simulation were run using the GROMOS96 53a6 forcefield[132]. The LINCS algorithm[99] was

used to constrain bond lengths, while the Parinello-Rahman barostat[133] and Berendsen[122] thermostat was used (T=310K). The box length was approximately 20.5nm x 20.5nm x12.5nm, with Particle Mesh Ewald (PME)[118] used for long-range electrostatics. Equilibration was run on these systems for 2ns with the Cα atoms restrained. This was followed by 0.5μs of unrestrained simulation. Analysis was performed using GROMACS[116], VMD[126], Bendix[137], and locally written scripts. A summary of the simulations run for this study is in Table 1.

Table 5.1: Simulations Performed in The Dok7/MuSK/Membrane Study

Simulation	Duration
Coarse-Grained Simulations	
MuSK(TM+JM)	5 x 5 μs
Dok7(PH-PTB)MuSK(TM+JM)	4 x 5 μs
Dok7(PH-PTB)$_2$MuSK(TM+JM)	3 x 5 μs
Dok7(PH-PTB)$_2$MuSK(TM+JM)$_2$	3 x 5 μs
Atomistic Simulations	
Dok7(PH-PTB)$_2$MuSK(TM+JM)$_2$	1 x 0.5 μs

*In all simulations a POPC/POPS/PI(4,5)P$_2$ (73%/20%/7%) was used.

5.2.3 Parameterisation of Phosphotyrosine

Parameters were designed for the GROMOS53a6[132] forcefield for the phosphotyrosine molecule. The initial structure of the phosphotyrosine was obtained from the crystal structure of Dok7 and the MuSK phosphopeptide. Hydrogens were then added in the appropriate places using the VMD plugin Molefacture, and the parameterisation file was built using tyrosine parameters, and adjusting accordingly for the phosphate group present on the end of the tyrosine. The phosphotyrosine was simulated first on its own in water, then in a peptide, followed with the PTB domain of Dok7 and the peptide from the crystal structure. There were 3 x 5ns simulations run for each system. In each system, the RMSD was calculated relative to the template structure, and it was found that the RMSD was very low, with the main motion found being

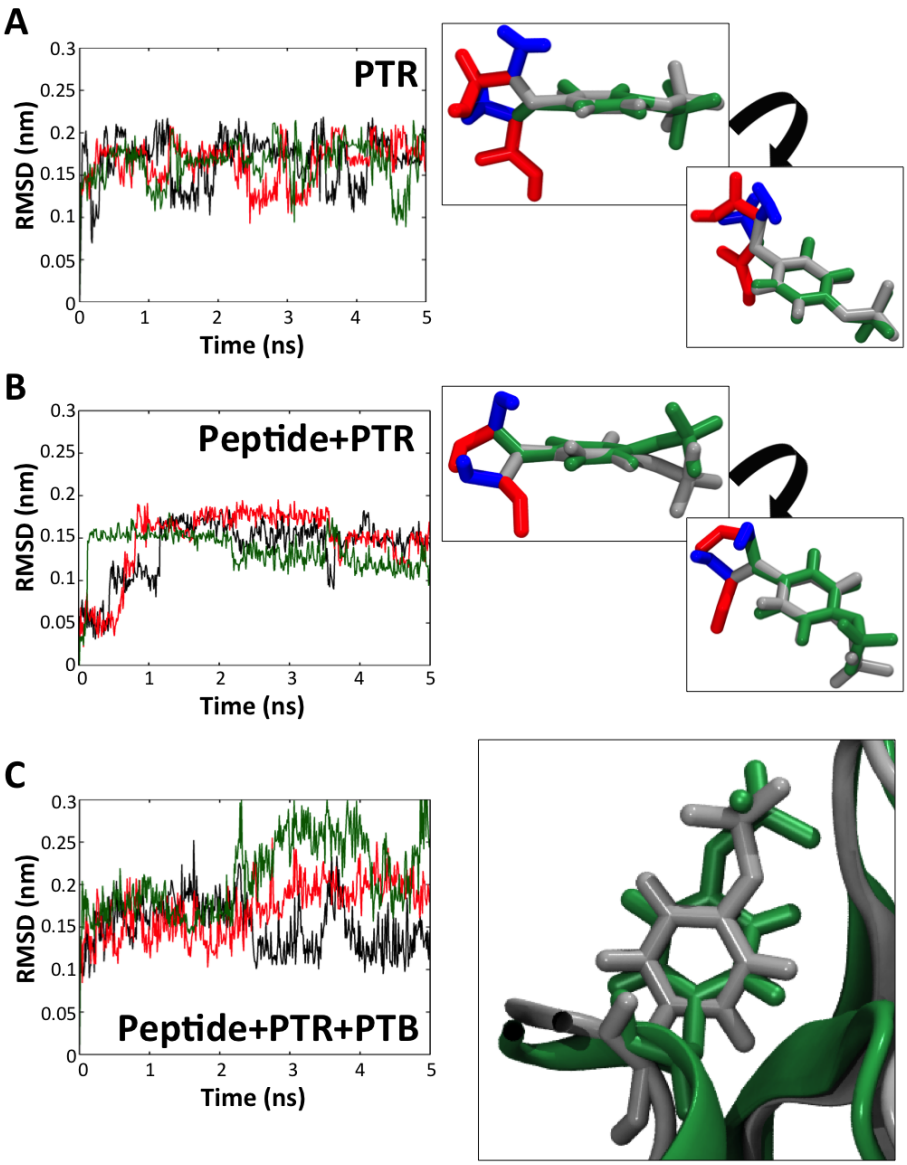

Figure 5.4: Simulations of phosphotyrosine in water as just PTR (A), PTR in MuSK peptide (B), or PTR in MuSK peptide bound to the PTB domain of Dok7 (C). Each individual replicate's RMSD is either in black, green or red line. For A and B, the C terminus is in red, the N terminus is in blue, the first frame's sidechain is silver, and the last frame is in green. All phosphotyrosines in A and B are shown in licorice representation. For C, the first frame of the simulation shows the PTB domain in silver cartoon and the phosphotyrosine in silver stick representation. The last frame is the PTB domain in green cartoon and the phosphotyrosine in green stick representation.

rotation around the carbon in between the tyrosine ring and the backbone, and also the oxygen between the tyrosine ring and the phosphate group (Figure 5.4). In addition, when aligning the first and second frame of the phosphotyrosine+PTB+peptide

in water, the phosphotyrosine remained in a similar position throughout the 5ns simulation.

5.3 Initial Simulations of MuSK(TM+JM)

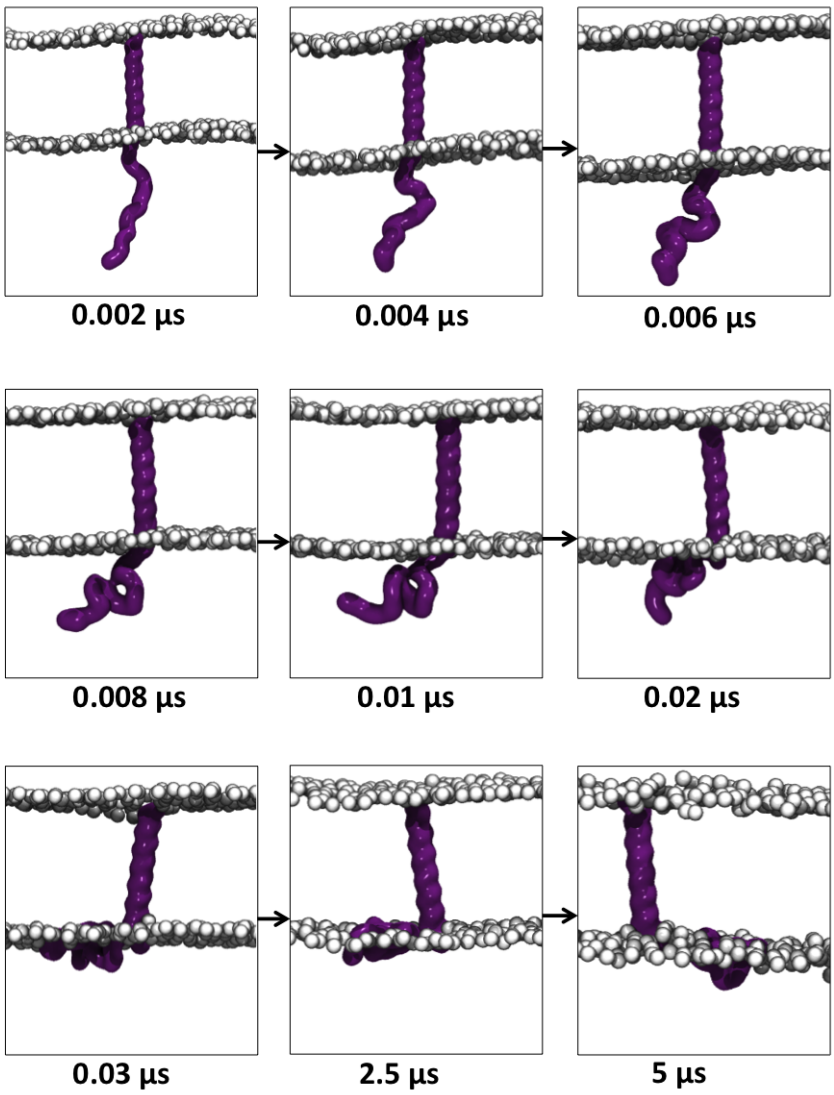

Figure 5.5: Snapshots of MuSK(TM+JM) system over a 5μs simulation. MuSK(TM+JM) is in purple QuickSurf, and the lipids are in white VDW representation.

First, an initial system with the MuSK TM helix plus the JM region (MuSK(TM+JM)) (Figure 5.6A) was simulated to determine the behaviour of the juxtamembrane re-

gion, as well as the lipids around the protein. The juxtamembrane region collapsed onto the membrane within the first few nanoseconds of the simulation (Figure 5.5) and remained there throughout the entire 5μs, which meant the would-be phosphorylated tyrosine wasn't exposed to the solvent, which was expected. However, since the JM region collapses onto the membrane, it would follow that the juxtamembrane region would interact with the lipids in the membrane.

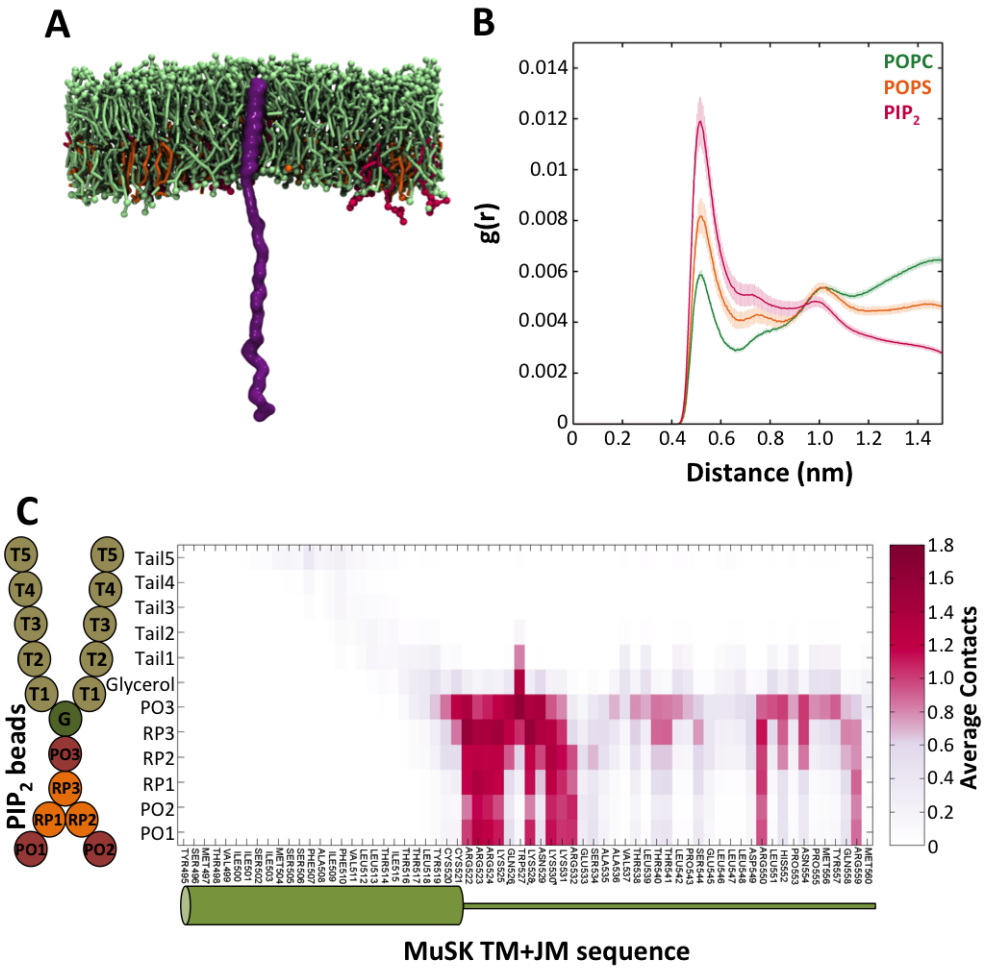

Figure 5.6: Simulations of MuSK(TM+JM) with POPC/POPS/PIP_2 bilayer. A. Example image of the MuSK(TM+JM) system. MuSK(TM+JM) is in purple QuickSurf, POPC is in green, POPS is in orange, and PIP_2 is in pink. All lipids are in licorice and spheres representation. B. RDF of MuSK(TM+JM). POPC is in green, POPS is in orange, and PIP_2 is in pink. Error bars are calculated by taking the standard deviation from all five simulations run. C. Fractional lipid interaction of MuSK(TM+JM) with PIP_2 by bead type. PIP_2 is represented by goldenrod, green, red and orange spheres. MuSK(TM+JM) is represented by a green cylinder and line.

Thus, the next step was to determine where the lipids were in relation to the protein. To do this, an average radial distribution function (RDF) (Figure 5.6B) was calculated. The RDF reveals that the PIP_2 lipids have a higher propensity to cluster around the highly positively charged JM region of the protein, out-competing POPS and POPC molecules, as seen in previous work on juxtamembrane regions of RTKs[129]. This was further analysed by calculating the fractional lipid interactions (Figure 5.6C) of each residue of MuSK with each bead of PIP_2 lipids. Overall, it was found that the head groups had a very high propensity for the first 10 residues of the juxtamembrane region, as these were the most positively charged. It also had a propensity to cluster around the whole JM region, but it is mainly centred on the first 10 residues of the juxtamembrane region.

5.4 MuSK+Dok7 Simulations

As the tyrosine was not going to be exposed enough to attempt free assembly of the complex, it was decided to build complexes with varying numbers of Dok7(PH-PTB)s and MuSK(TM+JM)s. The crystal structure was used as a guide to where the juxtamembrane region of MuSK(TM+JM) would be bound to Dok7(PH-PTB).

5.4.1 Dok7(PH-PTB)MuSK(TM+JM) Complex

First, we built the Dok7(PH-PTB)MuSK(TM+JM) model which had Dok7(PH-PTB) connected to the JM region of MuSK, but Dok7(PH-PTB) was placed away from the bilayer. Within the first few nanoseconds Dok7(PH-PTB) associated with the bilayer, as the JM region of MuSK(TM+JM) collapsed to the bilayer, which took with it Dok7(PH-PTB). Interestingly, in the presence of MuSK(TM+JM), Dok7(PH-PTB) still begins to associate with the bilayer via its PH domain (Figure 5.7B), in a sim-

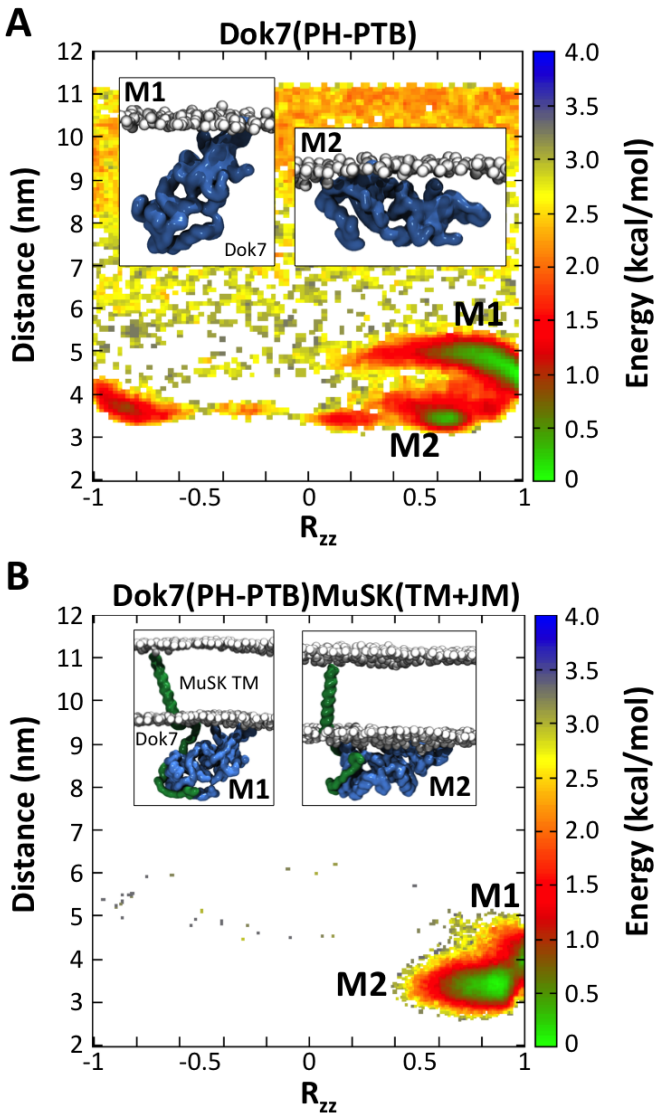

Figure 5.7: Energy Landscapes of Dok7(PH-PTB) without (A) and with (B) MuSK(TM+JM) present. Each energy landscape has 2 productive orientations (labelled 1 and 2). Dok7(PH-PTB) is in blue surface representation, MuSK(TM+JM) is in green surface representation, and the lipid head groups are in white spheres representation.

ilar fashion to the Dok7(PH-PTB) alone (Figure 5.7A). However, in the presence of MuSK, Dok7(PH-PTB) prefers the 'flat' binding mode (M2), rather than the 'angled' binding mode (M1). This preference can be attributed to the JM region of MuSK(TM+JM) pulling the PTB domain towards the bilayer in a flat orientation, and that Dok7(PH-PTB) remains in this orientation for the remainder of the simulation.

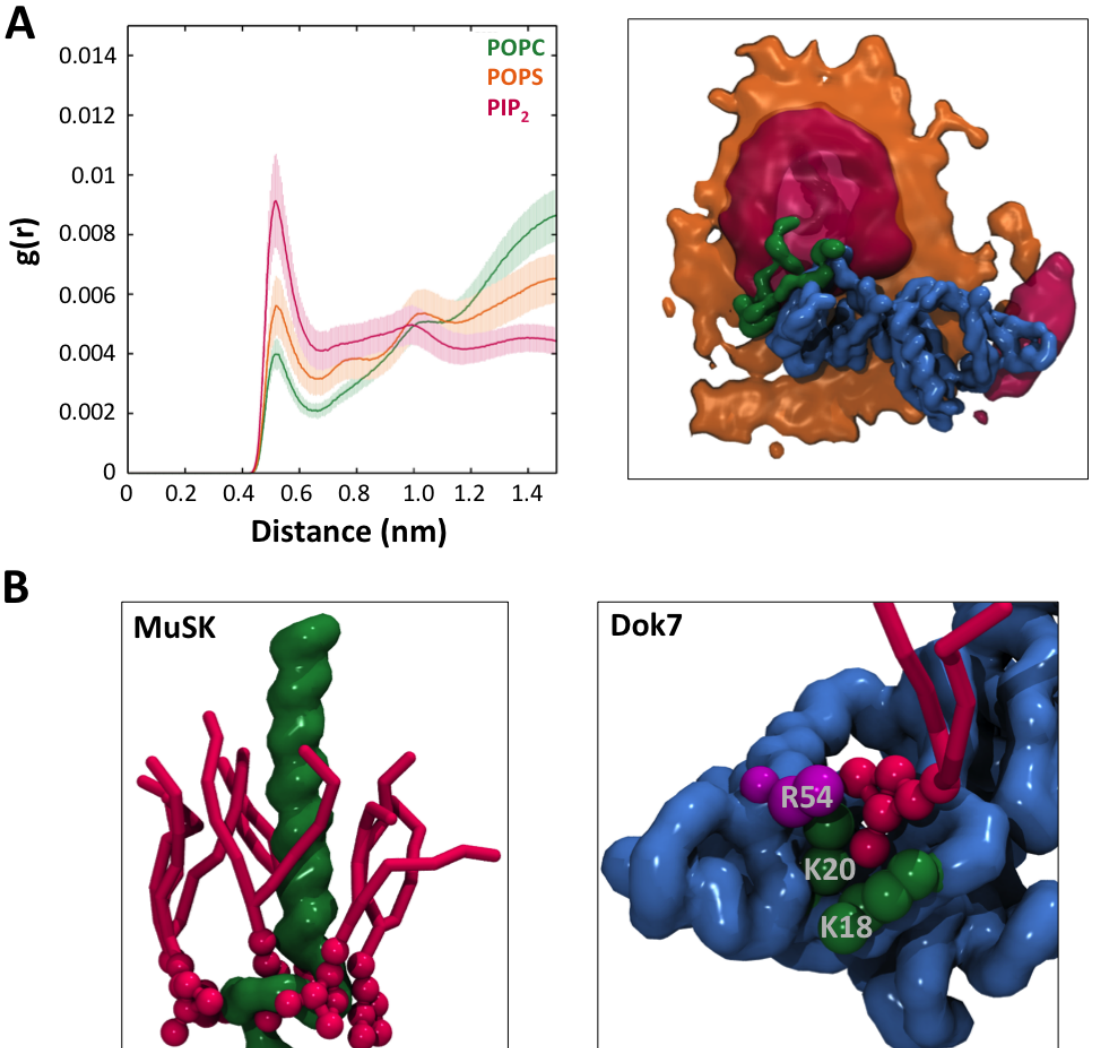

Figure 5.8: Lipid interactions of the Dok7(PH-PTB)MuSK(TM+JM) complex. A. RDF of Dok7(PH-PTB)MuSK(TM+JM) complex (left), as well as a representative Volmap plot (right). PIP_2 is in pink surface representation, POPS is in orange surface representation, MuSK is in green surface representation, and Dok7 is in blue surface representation. B. Representative lipid interactions for both MuSK(TM+JM) (left) and Dok7(PH-PTB) (right). PIP_2 is in pink, MuSK(TM+JM) is in green surface representation, and Dok7(PH-PTB) is in blue representation. Individual residues are in spheres representation, with the lysines in green and arginine in purple.

The lipid interactions for the Dok7(PH-PTB)MuSK(TM+JM) complex were examined for all simulations. The average RDF was calculated for all simulations (Figure 8A - left), and it was still found that PIP_2 has a propensity for clustering around the complex and out-competing POPC and POPS for interaction with the proteins. In the RDF, the peak at 0.5nm corresponds to the first shell of lipids around the

protein. This is further evidenced by a representative Volmap plot (Figure 5.8A - right), which shows PIP_2 clustering both around the positively charged JM region of MuSK(TM+JM) and the positively charged PH domain of Dok7. This is also seen when snapshots are taken of MuSK(TM+JM) and Dok7(PH-PTB) (Figure 5.8B). Interestingly, when looking at the residues in Dok7(PH-PTB) that bind to PIP_2, they are the same residues that were previously implicated in PIP binding in the previous chapter (K18, K20, R22, and R54).

5.4.2 Dok7(PH-PTB)$_2$MuSK(TM+JM) and Dok7(PH-PTB)$_2$ MuSK(TM+JM)$_2$ Complexes

Next, the Dok7(PH-PTB)$_2$MuSK(TM+JM) system was built, with the PH domain of one monomer facing the bilayer, as well as the PTB domain of the other monomer. This was compared to previous simulations run with just Dok7(PH-PTB)$_2$ (Figure 5.9A). with one simulation preferring binding mode M1A (Figure 5.9B), and the other two preferring somewhere in between binding modes M2 and M3. When the Dok7(PH-PTB)$_2$(MuSK)$_2$ complex was run, one of the simulations preferred to reside in the M1B orientation, while the remainder of the simulations preferred binding mode M2 (Figure 5.9C). This could be explained by the 2 MuSK(TM+JM)s restricting movement of the PTB domain of either monomer towards the membrane, acting simultaneously as an anchor and a barrier. Dok7(PH-PTB)$_2$ is tethered to the bilayer via MuSK, but acts as a barrier for the PTB domain to insert itself into the bilayer, like orientations seen in Section 4.4.4.

The interactions between the lipids and the protein in the system was analysed for both MuSK(TM+JM) and Dok7(PH-PTB)$_2$ separately (Figure 5.10A). As before, PIP_2 preferentially bids to both of the proteins, as evidenced by a strong peak at 0.5nm for both MuSK(TM+JM) and Dok7(PH-PTB)$_2$, indicating the presence of

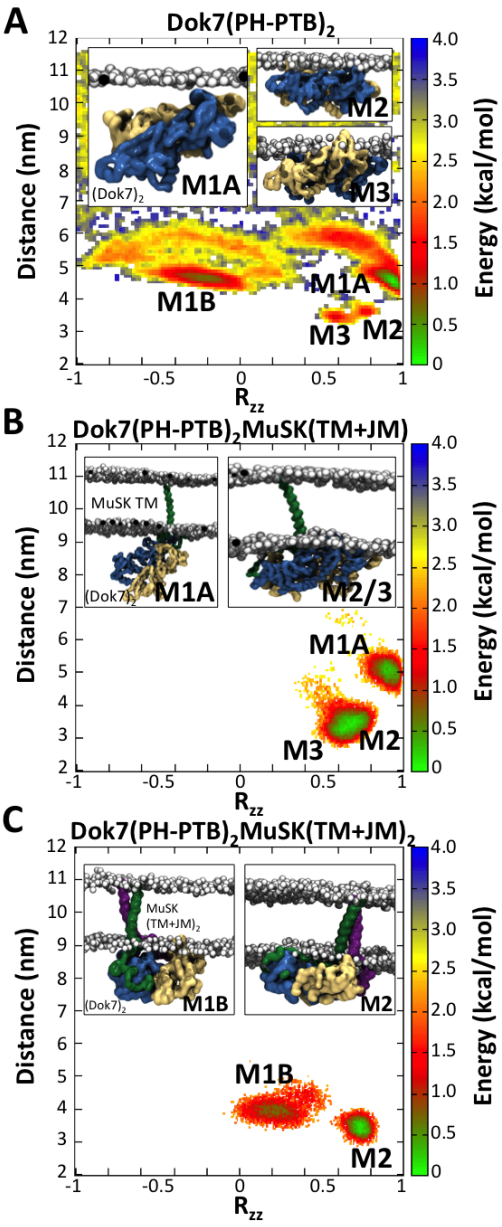

Figure 5.9: Energy Landscapes of the Dok7(PH-PTB)$_2$ (A), Dok7(PH-PTB)$_2$MuSK(TM+JM) (B), and Dok7(PH-PTB)$_2$MuSK(TM+JM)$_2$ Complex (C). Each Dok7(PH-PTB) is either blue or yellow surface representation, while each MuSK(TM+JM) is either green or purple surface representation. The lipid head groups are in white spheres representation. Each productive binding mode is labelled with either M1A, M1B, M2, or M3.

PIP$_2$ very close to the origin, which is the centred MuSK(TM). However, as the preferred lipid interactions for MuSK(TM+JM) taper off as the 5nm mark is approached,the preferential lipid interactions of Dok7(PH-PTB)$_2$ increase as the distance from the origin increases, indicating that the edges of the protein is where the

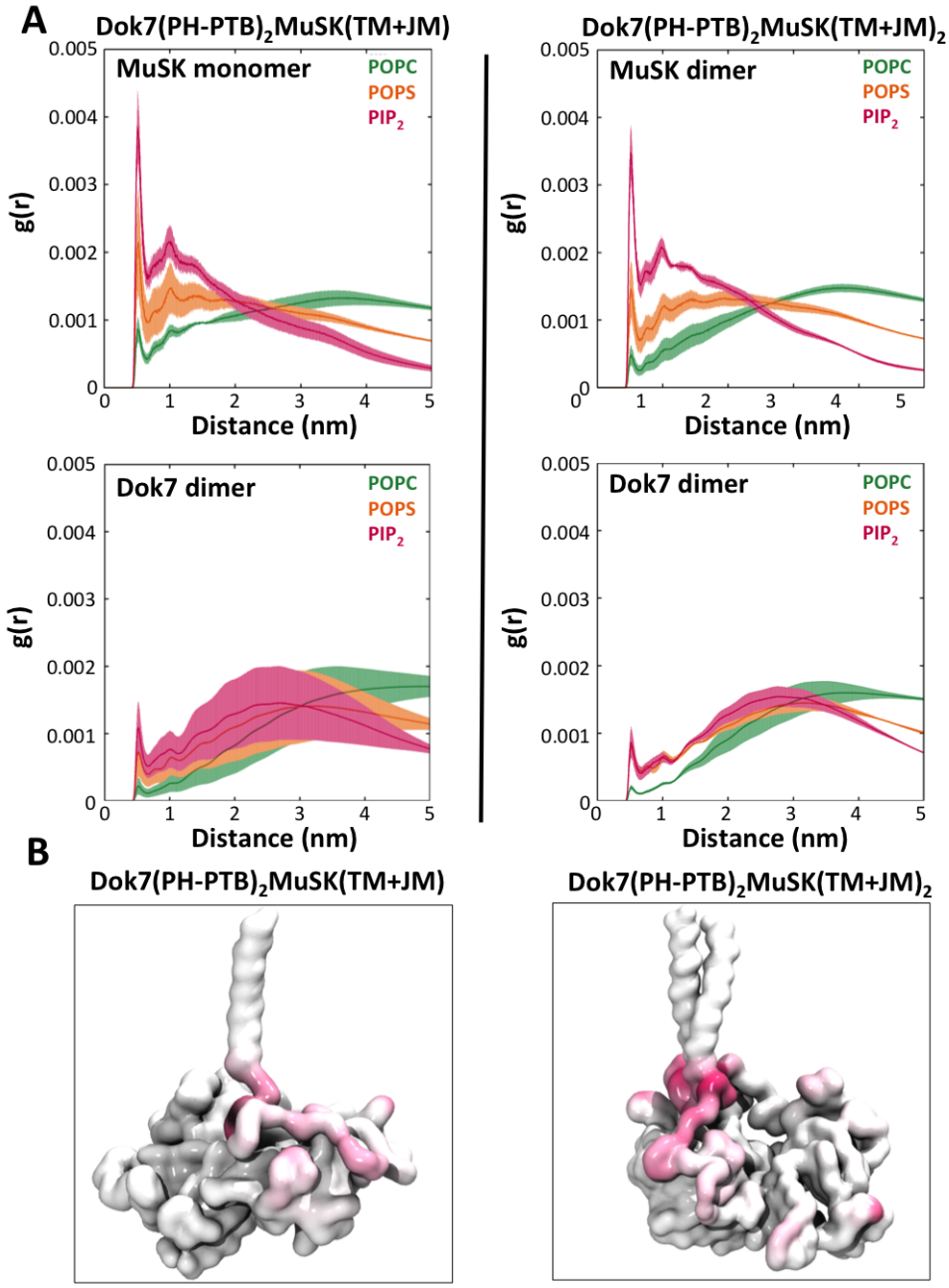

Figure 5.10: Lipid Analysis of both the Dok7(PH-PTB)$_2$MuSK(TM+JM) Complex and Dok7(PH-PTB)$_2$MuSK(TM+JM)$_2$ Complex. A. RDFs for MuSK(TM+JM) (left) and Dok7(PH-PTB)$_2$ (right). RDFs were calculated by separating PIP$_2$ is in pink, POPS is in orange, and POPC is in green. B. Fractional lipid interactions plotted onto the structure of the protein. The protein is in surface representation, with no lipid interactions in white, and the most preferential lipid interactions in pink.

PIP$_2$s preferentially bind to Dok7(PH-PTB)$_2$. This is further indicated by having the propensity of PIP lipids to bind to particular residues onto the structure (Figure

5.10B). The particular parts of the protein that are highlighted are, again, the first 10 residues of the juxtamembrane region of MuSK(TM+JM), and also the PH domains of (PH-PTB)$_2$, which confirms what has been seen in previous simulations.

5.5 The Final Dok7(PH-PTB)$_2$MuSK(TM+JM)$_2$ Atomistic Simulation and Structure

Next, a snapshot from the Dok7(PH-PTB)$_2$(MuSK)$_2$ simulations, which had the distance and R_{zz} corresponding to M2 (Dok7 in a 'flat' orientation), was chosen. It was then converted to atomistic representation using a fragment-based approach[113] and simulated for 0.5μs to determine the behaviour of MuSK(TM+JM), Dok7(PH-PTB)$_2$ and the lipids present in the membrane (Figure 5.11A).

5.5.1 Lipid Interactions of Dok7(PH-PTB)$_2$(MuSK)$_2$ Atomistic Complex

The propensity of lipid species to bind to different parts of the complex was again first examined by calculating an RDF for MuSK(TM+JM)$_2$ and Dok7(PH-PTB)$_2$ separately (Figure 5.11B). As it has been seen before in the CG simulations, MuSK(TM+JM)$_2$ has a higher propensity to bind PIP_2 rather than Dok7(PH-PTB)$_2$ (Figure 5.11C - left), but Dok7 also binds PIP_2 when the distance of the RDF is about 3 nm, which correlates with PIPs clustering around one or both of the PH domains of Dok7(PH-PTB)$_2$ (Figure 5.11C - right). To quantitatively determine whether Dok7(PH-PTB)$_2$ or MuSK(TM+JM)$_2$ has more PIP_2 bound, the lipid interactions over time was calculated separately for the MuSK(TM+JM) and Dok7(PH-PTB)$_2$ present in the system (Figure 5.11D). Indeed, MuSK(TM+JM)$_2$ maintained more contacts over time than

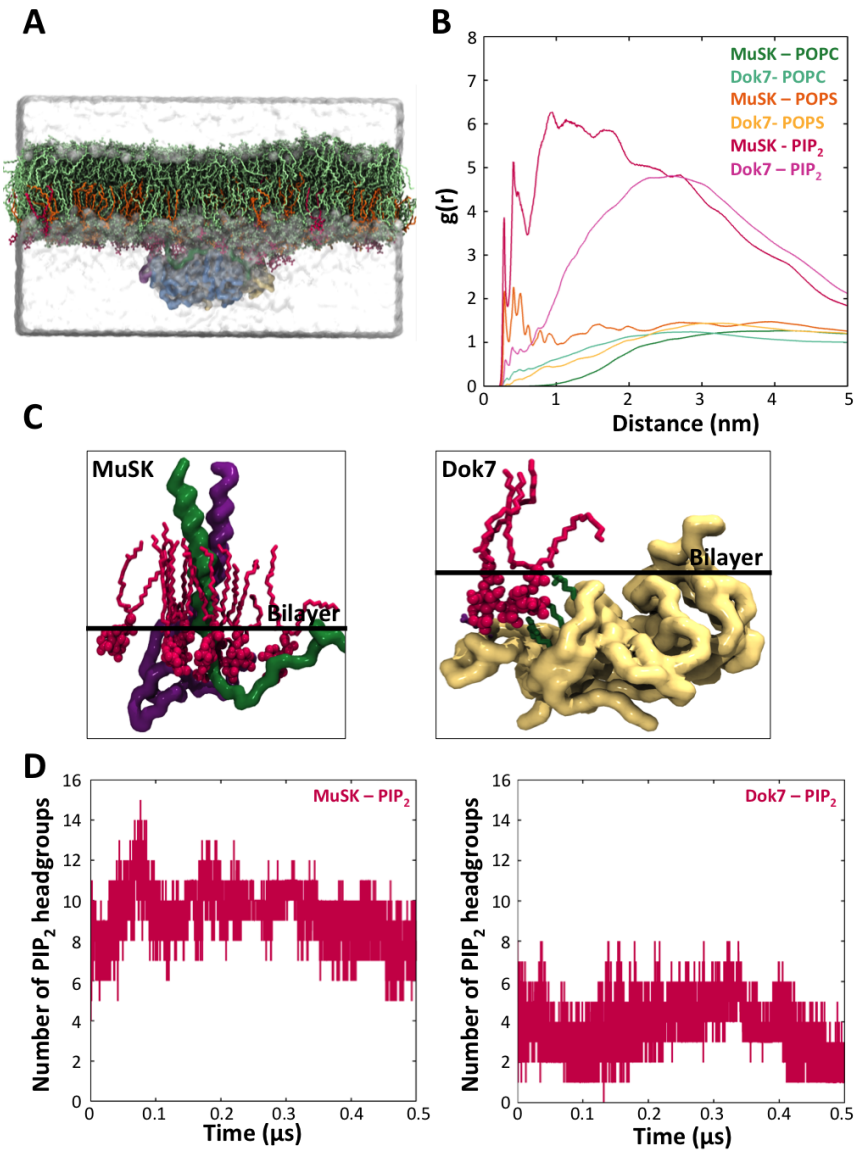

Figure 5.11: Lipid interactions of atomistic simulation. A. Image of atomistic simulation. Lipids are in licorice and spheres representation, with POPC in green, POPS in orange, and PIP_2 in pink. Each Dok7(PH-PTB) is either in blue or yellow surface representation, and each MuSK(TM+JM) is either in green or purple surface representation. Water is shown as a white surface. Ions are omitted for clarity. B. RDF of both MuSK(TM+JM)$_2$ and Dok7(PH-PTB)$_2$. POPC lines are in green, POPS are in orange, and PIP_2 are in pink. C. Examples of lipid interactions for both MuSK(TM+JM)$_2$ and Dok7(PH-PTB)$_2$. PIP_2 is in pink licorice and spheres representation, with each MuSK(TM+JM) in either green or purple surface, and each Dok7(PH-PTB) in yellow surface. D. Number of lipid interactions over time for both MuSK(TM+JM) and Dok7(PH-PTB)$_2$. Both lines are coloured pink.

Dok7(PH-PTB)$_2$, indicating that, while Dok7(PH-PTB)$_2$ binds PIP_2, PIP_2 preferentially binds to MuSK(TM+JM)$_2$, which confirms what was observed in the CG simulations.

5.5.2 Analysis of MuSK Dimer

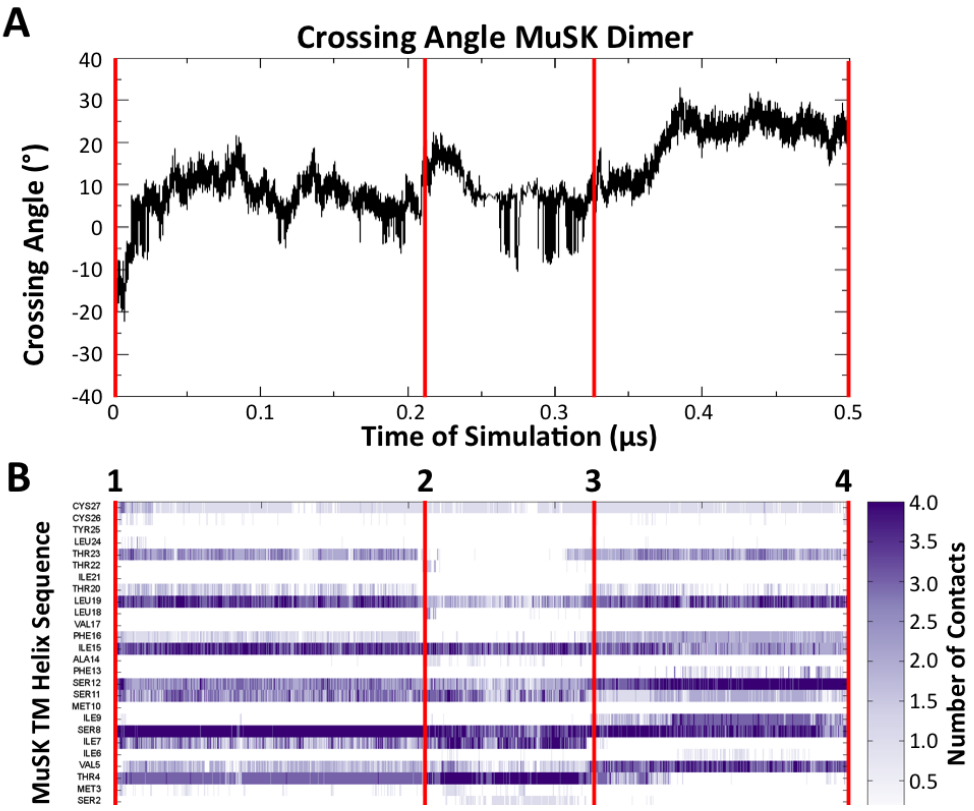

Figure 5.12: Crossing Angle and first part of contact analysis of MuSK TM. A. Crossing angle of the MuSK dimer over the simulation. The black line is the crossing angle over time, while the red lines are select frames that are expanded upon in Figure 11. B. Contacts of one helix of the MuSK dimer. The color scale ranges from white (0 contacts) to purple (4 contacts).

Further analysis on the MuSK TM helix dimer was run to determine its behaviour in the simulation in the presence of Dok7(PH-PTB)$_2$. Upon inspection of the crossing angle, the MuSK TM helix dimer started out as a right-handed dimer, but over the 0.5μs simulation, it switched to a left-handed dimer (Figure 5.12A), with some contacting residues changing between point 1 and point 4 on the simulation, specifically on points 2 and 3 (Figure 5.12B). Time points 1 and 4 were chosen as they were the start and end of the simulation, respectively. Time point 2 was selected because contacts on the C terminus of the helix began to shift away from each other, and time point 3 was chosen due to a shift in the threonine contact on the N terminus of the helix.

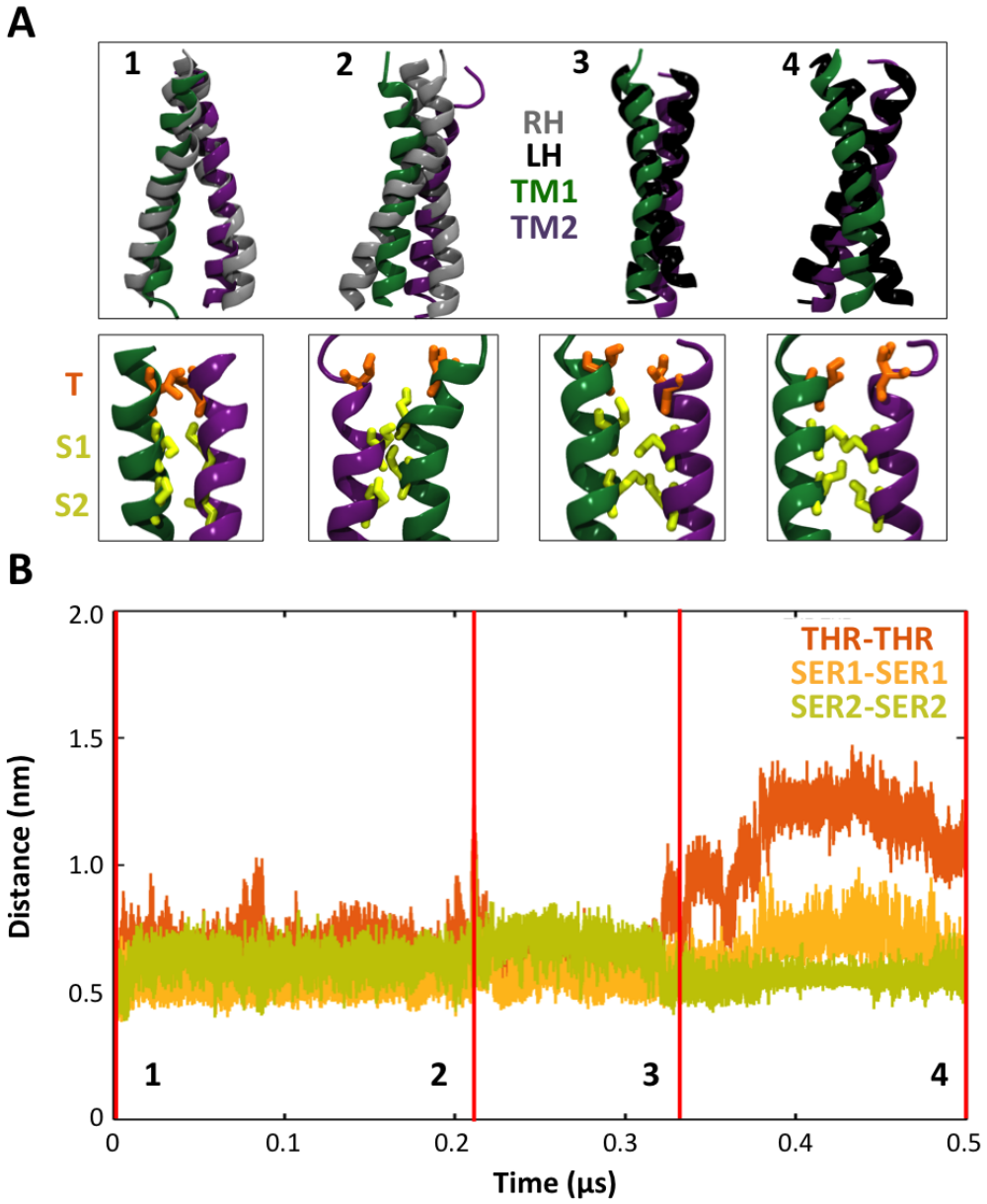

Figure 5.13: Analysis of MuSK dimer. A. Snapshots of dimers at four different times of the simulation (see Figure 10 for time points with corresponding crossing angle). All TMs are shown in cartoon representation, with TM1 in green, TM2 in purple, the right-handed template in grey, and the left-handed template in black. Threonines are shown in orange stick representation, and serines are shown in yellow representation. B. Distance between the Cα atoms of the threonines and series shown above. The threonine distance is in orange, and the serine distances in are different shades of yellow. The corresponding time points are shown as red lines.

To determine how this was happening, snapshots from time points 1-4 were taken and aligned with previously determined templates (Figure 5.13A, top). It is seen that at time point 1, the dimer is in a right-handed state, and at time point 2, it

still maintains some right-handed proclivity, but is in transition. This is even more apparent in time point 3, where it is beginning to look like a left-handed dimer. The transition is complete in time point 4, where the dimer resembles the left-handed template. The main contacts are between the aforementioned threonine and two serines. At the different timepoints, it appears that the threonine maintains contact only in the right-handed and potentially the transition dimers, but does not maintain contacts in the left-handed dimers (Figure 13A, bottom). As these were looked at more quantitatively, by checking the distance between the Cα atoms of each pair of residues (Figure 5.13B), it was clear that the two serines were maintaining contact with each other, while near the end of the simulation the threonines were coming apart during the simulation, and not interacting with each other as tightly as the core serines, which also indicates a transition from RH to LH dimer.

5.5.3 Behaviour of Phosphotyrosine

The behaviour of the phosphotyrosines between the first and last frame of the simulation was determined. Overall, the phosphotyrosines remained in the general area of the binding pocket, with the phosphotyrosine on the opposite side of the MuSK dimer (Figure 5.14 - left) moving away from the original binding pocket, but still maintained interactions with Dok7, albeit on a different interface. The other phosphotyrosine (Figure 5.14 - right) was closer to maintaining its position than the first phosphotyrosine, but still remained interacting with Dok7, and may have started interacting with MuSK. It is unclear what influence MuSK's TM helix dimer has on the behaviour of the phosphotyrosines.

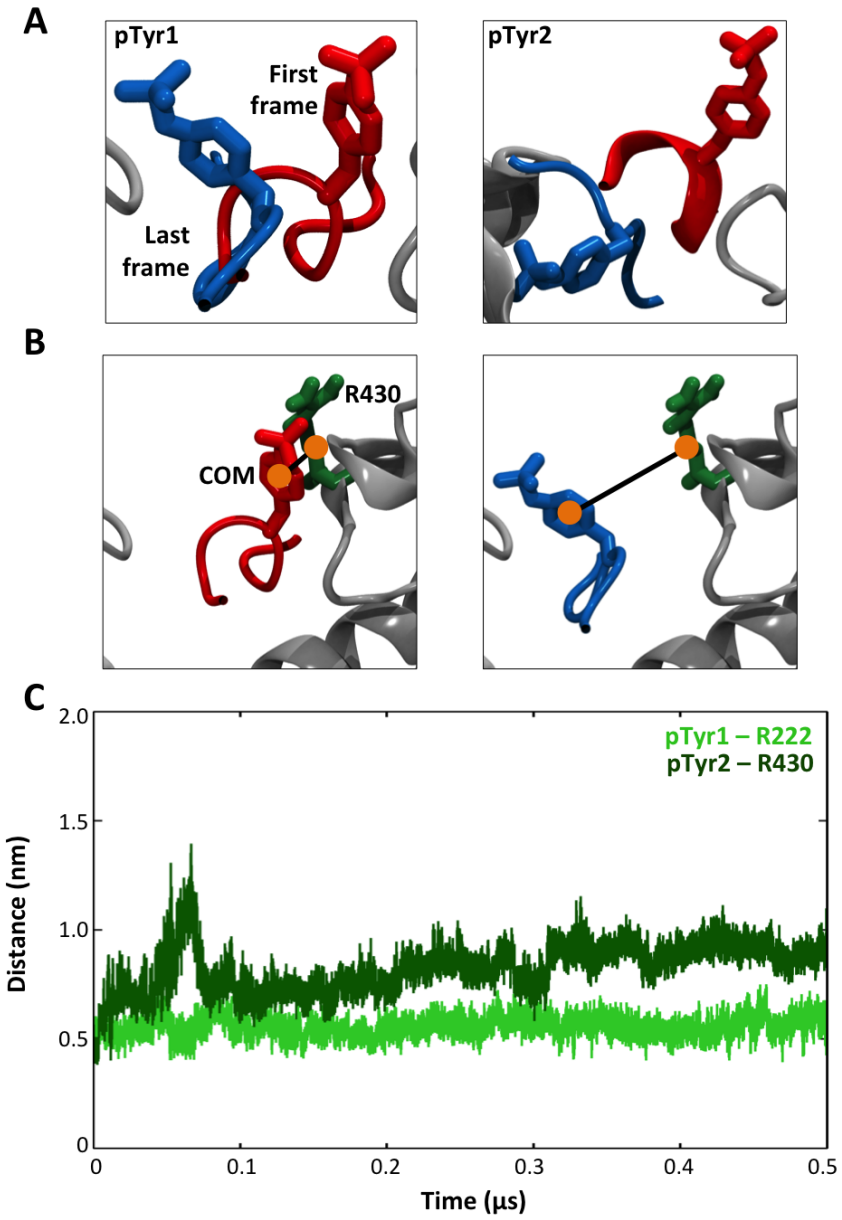

Figure 5.14: Behaviour of Phosphotyrosines during the simulation. A. The first and last frame of the phosphotyrosines in simulation. The first frame backbone is shown entirely in silver, the first frame phosphotyrosine is in red licorice, and the last frame phosphotyrosine is in blue licorice. B. Diagram of the calculation of distances between the pTyrs and a residue close to the backbone. pTyr is in either red or blue licorice, the MuSK peptide is in either red or blue cartoon representation, the backbone of the protein is in silver cartoon representation, and R430 is in green licorice presentation. The COM of each residue is represented by an orange circle, and the distance is shown as a black line. C. Distance between the COMs of the phosphotyrosine in question and the residue closest to it on Dok7 in the beginning of the simulation. The first phosphotyrosine is in light green, and the second phosphotyrosine is in dark green.

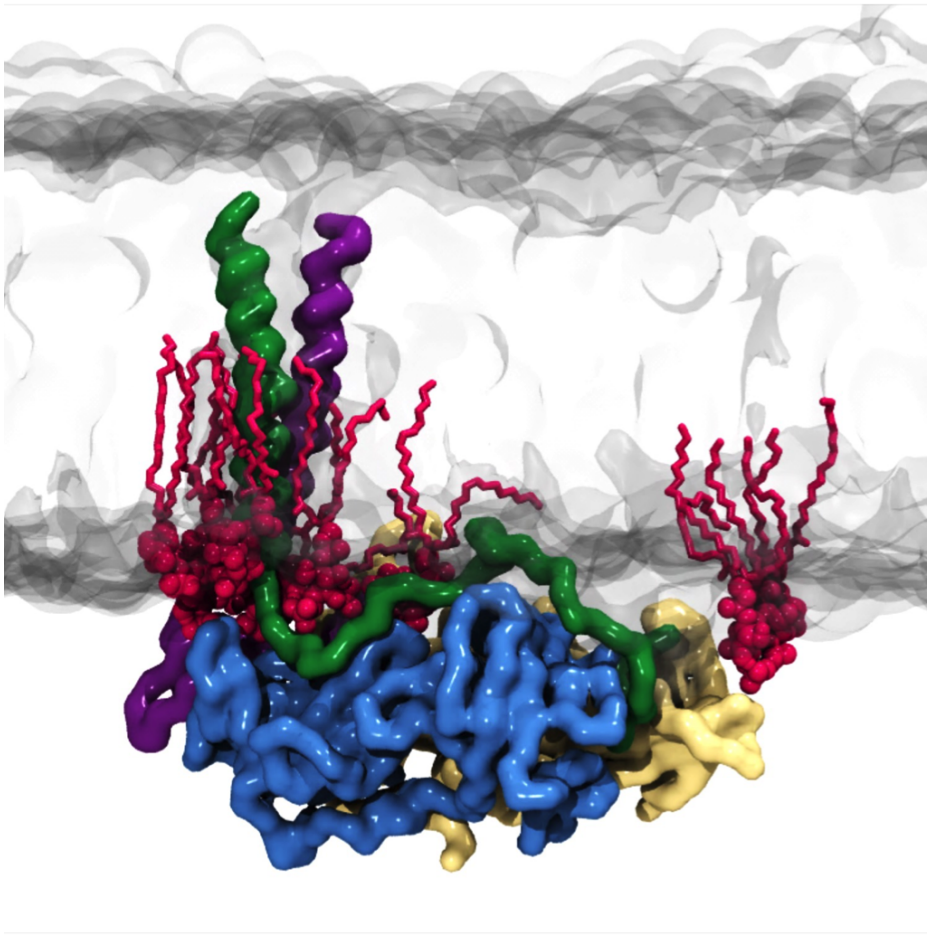

Figure 5.15: Final snapshot of the atomistic simulation. Dok7(PH-PTB) is shown in either blue or yellow surface representation, MuSK(TM+JM) is shown in either green or purple surface representation, the bilayer is shown in transparent white surface, and PIP_2 is shown in pink licorice and CPK representation.

5.5.4 Final Model and Bigger Picture

After putting all of this analysis together, the final snapshot was taken of the atomistic simulation, which gives a prospective model of the Dok7/MuSK/membrane complex (Figure 5.15). The final snapshot is of a LH dimer, and it is unknown whether this is the active or inactive form of the Dok7/MuSK/membrane complex.

The model can then be put into the bigger context in Figure 5.16. The proposed structure of the Dok7/MuSK/membrane complex is shown with the full-length MuSK

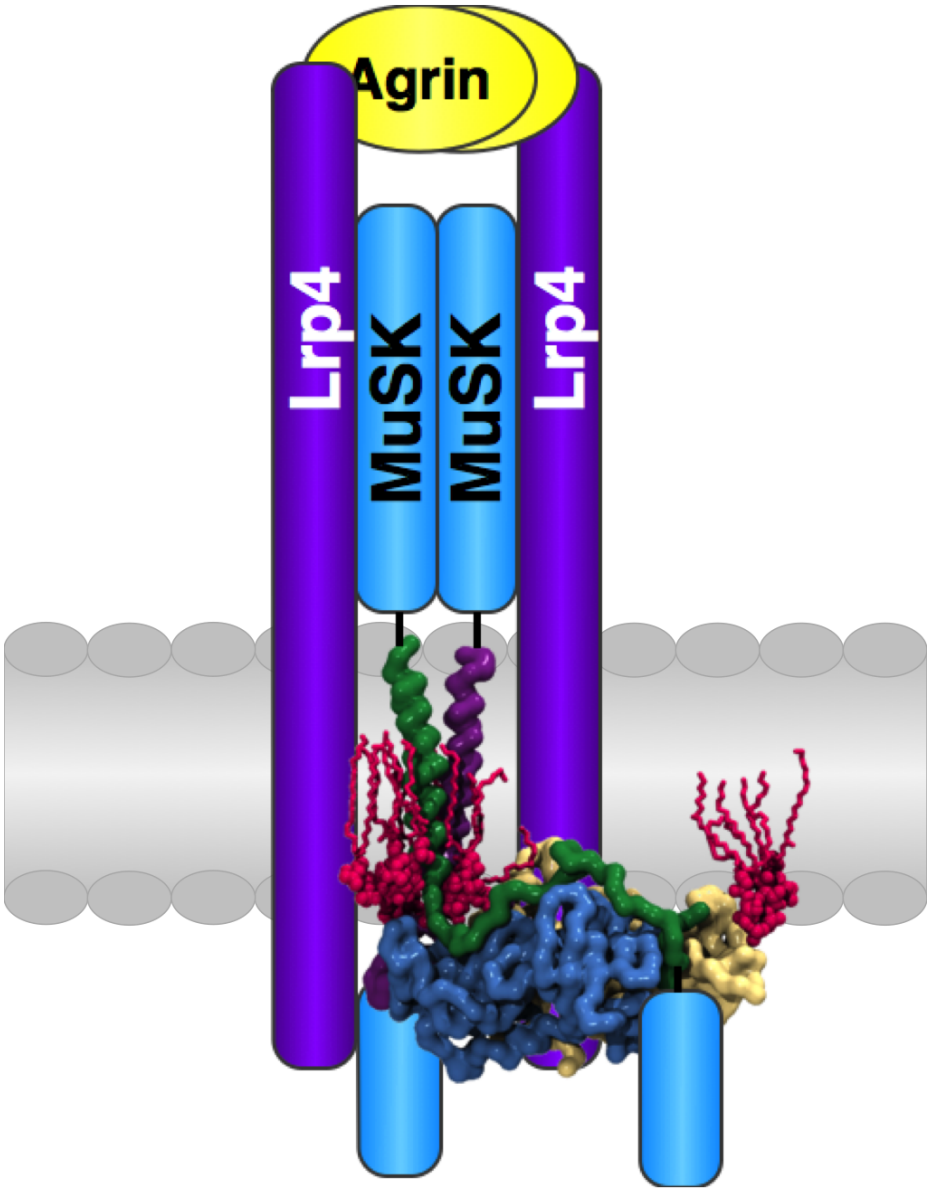

Figure 5.16: Final snapshot of the atomistic simulation in context with the overall complex. The bilayer is represented by grey ovals and rectangles, with Lrp4 in purple rectangles and MuSK in cyan rectangles and black lines.

receptor, as well as the full-length Lrp4 receptor. Even though it is a small part of the complex, it is an important one, and can be expanded upon in the future.

5.6 Conclusions and Speculation of Dok7/MuSK/ Membrane Mechanism

From the simulations in this study, previous behaviour of the MuSK TM helix dimer in the membrane, as well as Dok7(PH-PTB)$_2$ in the periphery, have been confirmed. PIP_2 preferentially clusters around both MuSK and Dok7, as seen previously in each proteins' independent simulation, but now it is known that MuSK(TM+JM) preferentially interacts with PIP_2 over Dok7(PH-PTB). This is true no matter the stoichiometry of Dok7(PH-PTB) to MuSK(TM+JM). It was also found that the presence of MuSK(TM+JM) also influences the energy landscapes of Dok7(PH-PTB), first by forcing Dok7(PH-PTB) to prefer the flat binding mode in the Dok7(PH-PTB)MuSK(TM+JM) simulations. However, Dok7(PH-PTB) still first associates with the bilayer via its PH domain. Then, in the Dok7(PH-PTB)$_2$MuSK(TM+JM) simulations, only two orientations are seen, one with the PH domain of one monomer of Dok7(PH-PTB)$_2$ and the PTB domain of the other monomer of Dok7(PH-PTB)$_2$ (M1A or M1B) binding with the bilayer, and one well which combines two orientations previously seen in just the Dok7(PH-PTB)$_2$ simulations. Finally, the Dok7(PH-PTB)$_2$MuSK(TM+JM)$_2$ simulations only have two orientations as well, but there is a clear preference for the 'flat' orientation (M2) previously seen in Dok7(PH-PTB)$_2$ simulations, rather than the M1A or M1B orientation previously preferred. This orientation was previously implicated in Section 4.4.4 as the most likely and potentially biologically relevant orientation for Dok7. These simulations are a nice confirmation that the 'flat' orientation is the preferred orientation of Dok7(PH-PTB) on the bilayer.

One representative frame for the Dok7(PH-PTB)$_2$MuSK(TM+JM)$_2$ complex was converted to atomistic representation and simulated for 0.5μs. This simulation contained

some interesting mechanistic implications. First, it was confirmed that Dok7 and MuSK maintained contacts with PIP_2, with MuSK still out-competing Dok7 for the PIP_2 lipids in the system. Combined with previous simulation studies on RTKs and lipids[129], it points to the probability that, even though some RTKs recruit peripheral proteins that most likely associate with anionic lipids, the RTKs will still out-compete these peripheral proteins for anionic lipids, specifically PIPs. However, far more interesting is the fact that, with the presence of Dok7, MuSK switched from being a right-handed dimer to a left-handed dimer throughout the 0.5μs simulation, with the last frame of the simulation as a representative picture of the Dok7/MuSK/membrane complex as Figure 15. This could potentially have implications for how Dok7 either activates MuSK for further phosphorylation, in that, when the handedness switches from right-handed to left-handed, conformational changes are induced such that MuSK is able to auto-phosphorylate itself. Conversely, this could be how the system deactivates before Dok7 unbinds and releases MuSK to go back to its resting state. To determine the full mechanism, however, more (and/or longer) simulations are needed of the atomistic system.

6

Conclusions and Future Directions

6.1 Summary and Conclusions

Prior to this thesis, the identification and the role of MuSK has been elucidated by mainly functional assays[40–43;48], as well as some structural studies, which include the kinase domain[71], the first two Immunoglobulin-like domains Ig1-Ig2[68], and the Frizzled-Cysteine Rich Domain (Fz-CRD)[70] of MuSK, as well as the first 210 residues of Dok7[56]. MuSK activation is a dynamic process, and even though the steps in the process are known, the details of the mechanism by which MuSK transmits the signal

from the outside to the inside of the cell via its TM helix to recruit Dok7 remains unknown. However, with the structure of Dok7 and TM helix prediction software, this allowed the use of simulations and preliminary NMR data to study this system.

First, a short experimental study was carried out on the MuSK TM helix, and though it was determined that it was folded in micelles, as well as helical, it was unclear whether the helix was present as a monomer or dimer in micelles. High-throughput simulation approaches[85;125] have been developed to run lots of small dimerisation simulations to describe the overall behaviour and modulating residues of TM helix dimerisation, giving the ability to study these systems in lieu of experimental data. There was a preliminary study done on the SIDEKICK[85] server of the MuSK TM helix previously, and it was shown that its behaviour was modulated by a small-x-x-x-small amino acid motif that is common among RTKs[15]. This study was extended to different bilayer compositions, as well as the addition of unstructured extensions on either end of the predicted helix, to determine if these changes to the system would cause any perturbations of the crossing angle and the small-x-x-x-small amino acid motif modulating dimerisation. These changes did not significantly affect the dimerisation propensity of the system, and demonstrates the robustness of the RTK systems to maintain their dimerisation motif.

The crystal structure of the first 210 residues of Dok7 was solved in 2010[56], providing the starting point of a simulations study of Dok7/membrane association. Computational and experimental studies of proteins with PH domains have been carried out[75;76;93;95;96], providing the groundwork and background for this study. These results demonstrate that Dok7's association with membranes is modulated by the PIP lipid species present in the bilayer, and a positive patch in the PH domain previously identified by the crystallographers is responsible for the association of Dok7 to the membrane. Mutations which disrupt the nature of this positive patch result in

perturbed binding of Dok7 to the bilayer. The manner in which Dok7 binds to the bilayer becomes increasingly complex as domains are added, culminating in a complex binding mechanism of the Dok7(PH-PTB)$_2$. This allowed us to determine that a 'flat' orientation of the dimer was the most likely of the binding modes, but without the presence of MuSK, it could not be confirmed.

The Dok7/MuSK/membrane complex was then built and simulated, as the behaviour of its individual components could be compared to the overall system's behaviour. Various ratios of Dok7/MuSK (1:1, 2:1 and 2:2) were simulated to determine how Dok7 would associate with the bilayer in the presence of MuSK's TM+JM region. Dok7's binding was influenced greatly, maintaining a 'flat' orientation in both the 1:1 and 2:2 complexes, validating the claim in the previous chapter that Dok7 is most likely going to be found in a 'flat' orientation on the bilayer in the presence of MuSK. Since this was all done in coarse-grain, it was decided to convert a frame which was representative of the 'flat' orientation of Dok7 to atomistic to further explore the lipid interactions, as well as the protein-protein interactions and potential conformational changes. In the simulation, it was seen that the MuSK dimer, which began in a right-handed conformation, switched to the left-handed dimer, potentially mediated by Dok7, and evidenced by a loss of a threonine contact above the small-x-x-x-small amino acid motif in MuSK.

6.2 Future Directions

One atomistic simulation, while interesting in its potential implications, is not enough to determine a definitive mechanism of switching. At least 1-2 more atomistic simulations of the 2:2 complex would have to be run, as well as simulations of the MuSK dimer on its own without MuSK present, would have to be run to start to answer

these questions. Addition of domains, such as the kinase domain, would provide a valuable insight into how the intracellular signalling would progress. Once this is understood, the addition of extracellular domains could also provide valuable insights into the behaviour of the whole Dok7/MuSK/membrane complex.

More generally, this thesis has focused mainly on the MuSK receptor, but could potentially be extended to RTKs, especially those with peripheral membrane proteins needed for signal transduction, like the FGFR system. FGFR has a peripheral protein, FRS2, that is activated upon phosphorylation of FGFR, and activates a signalling pathway, ending in p38 and ERK[9]. Techniques used in this thesis could be used in conjunction with known structural data of FGFR to shed light on one of FGFR's signalling pathways inside the cell.

Bibliography

[1] M. S. Almén, K. J. V. Nordström, R. Fredriksson, and H. B. Schiöth, "Mapping the human membrane proteome: a majority of the human membrane proteins can be classified according to function and evolutionary origin.," *BMC Biology*, vol. 7, 2009.

[2] Y. Arinaminpathy, E. Khurana, D. M. Engelman, and M. B. Gerstein, "Computational analysis of membrane proteins: the largest class of drug targets," *Drug Discovery Today*, vol. 14, pp. 1130–1135, 2009.

[3] G. Grigoryan, D. T. Moore, and W. F. DeGrado, "Transmembrane Communication: General Principles and Lessons from the Structure and Function of the M2 Proton Channel, K+ Channels, and Integrin Receptors," *Annu. Rev. Biochem.*, vol. 27, no. 5, pp. 417–428, 2009.

[4] M. A. Lemmon and J. Schlessinger, "Cell signaling by receptor tyrosine kinases.," *Cell*, vol. 141, no. 7, pp. 1117–1134, 2010.

[5] P. Stansfeld, J. Goose, M. Caffrey, E. Carpenter, J. Parker, S. Newstead, and M. Sansom, "MemProtMD: Automated Insertion of Membrane Protein Structures into Explicit Lipid Membranes," *Structure*, vol. 23, pp. 1350–1361, 2015.

[6] K. A. Holland and I. B. Holland, *Transmembrane Signalling*. John Wiley & Sons, Ltd., 2001.

[7] H. W. Choe, J. H. Park, Y. J. Kim, and O. P. Ernst, "Transmembrane signaling by GPCRs: Insight from rhodopsin and opsin structures," *Neuropharmacology*, vol. 60, pp. 52–57, 2011.

[8] D. R. Robinson, Y. M. Wu, and S. F. Lin, "The protein tyrosine kinase family of the human genome.," *Oncogene*, vol. 19, no. 49, pp. 5548–5557, 2000.

[9] J. Narayana and W. A. Horton, "FGFR3 biology and skeletal disease," *Connective Tissue Research*, pp. 1–18, 2015.

[10] I. N. Maruyama, "Activation of transmembrane cell-surface receptors via a common mechanism? The "rotation model"," *BioEssays*, pp. 1–9, 2015.

[11] N. Ghazanfari, K. J. Fernandez, Y. Murata, M. Morsch, S. T. Ngo, S. W. Reddel, P. G. Noakes, and W. D. Phillips, "Muscle specific kinase: organiser of synaptic membrane domains.," *Int. J. Biochem. Cell B.*, vol. 43, no. 3, pp. 295–298, 2011.

[12] S. R. Hubbard and K. Gnanasambandan, "Structure and Activation of MuSK, a Receptor Tyrosine Kinase Central to Neuromuscular Junction Formation," *Biochimica et Biophysica Acta*, vol. 1834, no. 10, pp. 2166–2169, 2013.

[13] C. Finger, C. Escher, and D. Schneider, "The single transmembrane domains of human receptor tyrosine kinases encode self-interactions.," *Science*, vol. 2, no. 89, p. ra56, 2009.

[14] K. R. MacKenzie, J. H. Prestegard, and D. M. Engelman, "A transmembrane helix dimer: structure and implications.," *Science*, vol. 276, pp. 131–133, 1997.

[15] A. Senes, M. Gerstein, and D. M. Engelman, "Statistical analysis of amino acid patterns in transmembrane helices: the GxxxG motif occurs frequently and in association with beta-branched residues at neighboring positions.," *J. Mol. Biol.*, vol. 296, pp. 921–936, 2000.

[16] E. V. Bocharov, M. L. Mayzel, P. E. Volynsky, K. S. Mineev, E. N. Tkach, Y. S. Ermolyuk, A. A. Schulga, R. G. Efremov, and A. S. Arseniev, "Left-handed dimer of EphA2 transmembrane domain: Helix packing diversity among receptor tyrosine kinases.," *Biophys. J.*, vol. 98, no. 5, pp. 881–889, 2010.

[17] N. Endres *et al.*, "Conformational Coupling across the Plasma Membrane in Activation of the EGF Receptor," *Cell*, vol. 152, pp. 543–556, 2013.

[18] E. V. Bocharov *et al.*, "Spatial structure and pH-dependent conformational diversity of dimeric transmembrane domain of the receptor tyrosine kinase EphA1," *J. Biol. Chem.*, vol. 283, no. 43, pp. 29385–29395, 2008.

[19] K. S. Mineev, E. V. Bocharov, Y. E. Pustovalova, O. V. Bocharova, V. V. Chupin, and A. S. Arseniev, "Spatial Structure of the Transmembrane Domain Heterodimer of ErbB1 and ErbB2 Receptor Tyrosine Kinases," *J. Mol. Biol.*, vol. 400, pp. 231–243, 2010.

[20] K. S. Mineev, N. F. Khabibullina, E. N. Lyukmanova, D. A. Dolgikh, M. P. Kirpichnikov, and A. S. Arseniev, "Spatial structure and dimer-monomer equilibrium of the ErbB3 transmembrane domain in DPC micelles," *Biochimica et Biophysica Acta*, vol. 1808, pp. 2081–2088, 2011.

[21] A. Arkhipov, Y. Shan, R. Das, N. F. Endres, M. P. Eastwood, D. E. Wemmer, J. Kuriyan, and D. E. Shaw, "Architecture and membrane interactions of the EGF receptor," *Cell*, vol. 152, pp. 557–569, 2013.

[22] I. E. Michailidis, R. Rusinova, A. Georgakopoulos, Y. Chen, R. Iyengar, N. K. Robakis, D. E. Logothetis, and L. Baki, "Phosphatidylinositol-4,5-bisphosphate regulates epidermal growth factor receptor activation," vol. 461, pp. 387–397, 2011.

[23] R. G. W. Anderson and K. Jacobson, "A role for lipid shells in targeting proteins to caveolae, rafts, and other lipid domains.," *Science*, vol. 296, pp. 1821–1825, 2002.

[24] Y. Gambin, R. Lopez-Esparza, M. Reffay, E. Sierecki, N. S. Gov, M. Genest, R. S. Hodges, and W. Urbach, "Lateral mobility of proteins in liquid membranes revisited.," *PNAS*, vol. 103, pp. 2098–2102, 2006.

[25] T. Apajalahti, P. Niemela, P. N. Govindan, M. S. Miettinen, E. Salonen, S. J. Marrink, and I. Vattulainen, "Concerted diffusion of lipids in raft-like membranes," *Faraday discussions*, vol. 144, pp. 411–430, 2010.

[26] K. Simons and J. l. Sampaio, "Membrane organization and lipid rafts," *Cold Spring Harbor Perspect. Biol.*, pp. 1–18, 2011.

[27] A. Zachowski, "Phospholipids in animal eukaryotic membranes: transverse asymmetry and movement.," *Biochem. J.*, vol. 294, pp. 1–14, 1993.

[28] A. Shevchenko and K. Simons, "Lipidomics: coming to grips with lipid diversity," *Nat. Rev. Mol. Cell Biol.*, pp. 593–598, 2010.

[29] U. Coskun and K. Simons, "Cell membranes: The lipid perspective," *Structure*, vol. 19, pp. 1543–1548, 2011.

[30] G. van Meer, D. R. Voelker, and G. W. Feigenson, "Membrane lipids: where they are and how they behave.," *Nat. Rev. Mol. Biol.*, vol. 9, pp. 112–124, 2008.

[31] E. Ikonen, "Cellular cholesterol trafficking and compartmentalization.," *Nat. Rev. Mol. Biol.*, vol. 9, pp. 125–138, 2008.

[32] S. J. Singer and G. L. Nicolson, "The Fluid Mosaic Model of the Structure of Cell Membranes," *Science*, vol. 175, pp. 720–731, 1972.

[33] I. Levental, M. Grzybek, and K. Simons, "Raft domains of variable properties and compositions in plasma membrane vesicles.," *PNAS*, vol. 108, pp. 11411–11416, 2011.

[34] L. Rajendran and K. Simons, "Lipid rafts and membrane dynamics.," *J. Cell. Sci.*, vol. 118, pp. 1099–1102, 2005.

[35] S. Munro, "Lipid Rafts: Elusive or Illusive?," *Cell*, vol. 115, pp. 377–388, 2003.

[36] D. M. Owen, A. Magenau, D. Williamson, and K. Gaus, "The lipid raft hypothesis revisited - new insights on raft composition and function from super-resolution fluorescence microscopy," *BioEssays*, vol. 34, pp. 739–747, 2012.

[37] K. Simons and D. Toomre, "Lipid rafts and signal transduction.," *Nat. Rev. Mol. Biol.*, vol. 1, pp. 31–39, 2000.

[38] J. B. Helms and C. Zurzolo, "Lipids as targeting signals: Lipid rafts and intracellular trafficking," *Traffic*, vol. 5, pp. 247–254, 2004.

[39] D. Zhu, W. C. Xiong, and L. Mei, "Lipid rafts serve as a signaling platform for nicotinic acetylcholine receptor clustering.," *Neuroscience*, vol. 26, pp. 4841–4851, 2006.

[40] C. G. Jennings, S. M. Dyer, and S. J. Burden, "Muscle-specific trk-related receptor with a kringle domain defines a distinct class of receptor tyrosine kinases.," *PNAS*, vol. 90, pp. 2895–2899, 1993.

[41] DeChiara *et al.*, "The receptor tyrosine kinase MuSK is required for neuromuscular junction formation in vivo.," *Cell*, vol. 85, pp. 501–512, 1996.

[42] T. T. Kummer, T. Misgeld, and J. R. Sanes, "Assembly of the postsynaptic membrane at the neuromuscular junction: Paradigm lost," *Curr. Op. Neuro.*, vol. 16, pp. 74–82, 2006.

[43] H. Wu, W. C. Xiong, and L. Mei, "To build a synapse: signaling pathways in neuromuscular junction assembly.," *Development*, vol. 137, pp. 1017–1033, 2010.

[44] W. Lin, R. W. Burgess, B. Dominguez, S. L. Pfaff, J. R. Sanes, and K. Lee, "Distinct roles of nerve and muscle in postsynaptic differentiation of the neuromuscular synapse," *Nature*, vol. 410, pp. 1057–1064, 2001.

[45] X. Yang, S. Arber, C. William, L. Li, Y. Tanabe, T. M. Jessell, C. Birchmeier, and S. J. Burden, "Patterning of muscle acetylcholine receptor gene expression in the absence of motor innervation," *Neuron*, vol. 30, pp. 399–410, 2001.

[46] U. McMahon, "The Agrin Hypothesis," *Cold spring Harb. Symp. Quant. Biol.*, vol. 55, pp. 573–584, 1990.

[47] M. Gautam, P. G. Noakes, L. Moscoso, F. Rupp, R. H. Scheller, J. P. Merlie, and J. R. Sanes, "Defective neuromuscular synaptogenesis in agrin-deficient mutant mice," *Cell*, vol. 85, pp. 525–535, 1996.

[48] D. J. Glass *et al.*, "Agrin acts via a MuSK receptor complex," *Cell*, vol. 85, pp. 513–523, 1996.

[49] D. M. Valenzuela, T. N. Stitt, P. S. DiStefano, E. Rojas, K. Mattsson, D. L. Compton, L. Nuñez, J. S. Park, J. L. Stark, and D. R. Gies, "Receptor tyrosine kinase specific for the skeletal muscle lineage: expression in embryonic muscle, at the neuromuscular junction, and after injury.," *Neuron*, vol. 15, pp. 573–584, 1995.

[50] E. B. Johnson, R. E. Hammer, and J. Herz, "Abnormal development of the apical ectodermal ridge and polysyndactyly in Megf7-deficient mice," *Human Molecular Genetics*, vol. 14, pp. 3523–3538, 2005.

[51] D. Simon-Chazottes, S. Tutois, M. Kuehn, M. Evans, F. Bourgade, S. Cook, M. T. Davisson, and J. L. Guénet, "Mutations in the gene encoding the low-density lipoprotein receptor LRP4 cause abnormal limb development in the mouse," *Genomics*, vol. 87, pp. 673–677, 2006.

[52] S. D. Weatherbee, K. V. Anderson, and L. A. Niswander, "LDL-receptor-related protein 4 is crucial for formation of the neuromuscular junction.," *Development*, vol. 133, no. 24, pp. 4993–5000, 2006.

[53] B. Zhang, S. Luo, Q. Wang, T. Suzuki, W. C. Xiong, and L. Mei, "LRP4 serves as a coreceptor of agrin.," *Neuron*, vol. 60, no. 2, pp. 285–297, 2008.

[54] N. Kim, A. L. Stiegler, T. O. Cameron, P. T. Hallock, A. M. Gomez, J. H. Huang, S. R. Hubbard, M. L. Dustin, and S. J. Burden, "Lrp4 is a receptor for Agrin and forms a complex with MuSK.," *Cell*, vol. 135, no. 2, pp. 334–342, 2008.

[55] K. Okada *et al.*, "The muscle protein Dok-7 is essential for neuromuscular synaptogenesis.," *Science*, vol. 312, pp. 1802–1805, 2006.

[56] E. Bergamin, P. T. Hallock, S. J. Burden, and S. R. Hubbard, "The cytoplasmic adaptor protein Dok7 activates the receptor tyrosine kinase MuSK via dimerization.," *Mol. Cell*, vol. 39, pp. 100–109, 2010.

[57] W. Hoch, J. McConville, S. Helms, J. Newsom-Davis, A. Melms, and A. Vincent, "Auto-antibodies to the receptor tyrosine kinase MuSK in patients with myasthenia gravis without acetylcholine receptor antibodies.," *Nature Medicine*, vol. 7, pp. 365–368, 2001.

[58] J. McConville and A. Vincent, "Diseases of the neuromuscular junction.," *Curr. Op. Pharm.*, vol. 2, pp. 296–301, 2002.

[59] C. Huzé *et al.*, "Identification of an agrin mutation that causes congenital myasthenia and affects synapse function.," *Am. J. Hum. Gen.*, vol. 85, pp. 155–167, 2009.

[60] J. Spillane, D. J. Beeson, and D. M. Kullmann, "Myasthenia and related disorders of the neuromuscular junction.," *J. Neurol. Neurosurg. Psychiatry*, vol. 81, pp. 850–857, 2010.

[61] K. Shigemoto *et al.*, "Myasthenia gravis experimentally induced with muscle-specific kinase.," *Ann. N. Y. Acad. Sci.*, vol. 1132, pp. 93–98, 2008.

[62] E. H. Niks, J. B. M. Kuks, J. H. J. Wokke, H. Veldman, E. Bakker, J. J. G. M. Verschuuren, and J. J. Plomp, "Pre- and postsynaptic neuromuscular junction abnormalities in musk myasthenia.," *Muscle & Nerve*, vol. 42, pp. 283–288, 2010.

[63] A. Ben Ammar *et al.*, "A mutation causes MuSK reduced sensitivity to agrin and congenital myasthenia.," *PloS one*, vol. 8, p. e53826, 2013.

[64] V. Kumar and H. J. Kaminski, "Treatment of myasthenia gravis.," *Curr. Neurol. Neurosci. Rep.*, vol. 11, pp. 89–96, 2011.

[65] B. Ohkawara *et al.*, "LRP4 third β-propeller domain mutations cause novel congenital myasthenia by compromising agrin-mediated MuSK signaling in a position-specific manner.," *Hum. Mol. Gen.*, pp. ddt578–, 2013.

[66] J. S. Müller *et al.*, "Phenotypical spectrum of DOK7 mutations in congenital myasthenic syndromes.," *Brainy*, vol. 130, pp. 1497–1506, 2007.

[67] R. A. Maselli *et al.*, "Mutations in MUSK causing congenital myasthenic syndrome impair MuSK-Dok-7 interaction.," *Hum. Mol. Gen.*, vol. 19, pp. 2370–2379, 2010.

[68] A. L. Stiegler, S. J. Burden, and S. R. Hubbard, "Crystal structure of the agrin-responsive immunoglobulin-like domains 1 and 2 of the receptor tyrosine kinase MuSK.," *J. Mol. Biol.*, vol. 364, no. 3, pp. 424–433, 2006.

[69] A. N. Plotnikov, J. Schlessinger, S. R. Hubbard, and M. Mohammadi, "Structural basis for FGF receptor dimerization and activation," *Cell*, vol. 98, pp. 641–650, 1999.

[70] A. L. Stiegler, S. J. Burden, and S. R. Hubbard, "Crystal structure of the frizzled-like cysteine-rich domain of the receptor tyrosine kinase MuSK.," *J. Mol. Biol.*, vol. 393, pp. 1–9, 2009.

[71] J. H. Till, M. Becerra, A. Watty, Y. Lu, Y. Ma, T. A. Neubert, S. J. Burden, and S. R. Hubbard, "Crystal structure of the MuSK tyrosine kinase: insights into receptor autoregulation.," *Structure*, vol. 10, no. 9, pp. 1187–1196, 2002.

[72] Y. Zong, B. Zhang, S. Gu, K. Lee, J. Zhou, G. Yao, D. Figueiredo, K. Perry, L. Mei, and R. Jin, "Structural basis of agrin-LRP4-MuSK signaling.," *Genes & Dev.*, vol. 26, no. 3, pp. 247–258, 2012.

[73] D. Cai, S. Dhe-Paganon, P. A. Melendez, J. Lee, and S. E. Shoelson, "Two new substrates in insulin signaling, IRS5/DOK4 and IRS6/DOK5," *J. Biol. Chem.*, vol. 278, no. 28, pp. 25323–25330, 2003.

[74] A. Inoue, K. Setoguchi, Y. Matsubara, K. Okada, N. Sato, Y. Iwakura, O. Higuchi, and Y. Yamanashi, "Dok-7 activates the muscle receptor kinase MuSK and shapes synapse formation.," *Science signaling*, vol. 2, no. 59, p. ra7, 2009.

[75] S. E. Lietzke, S. Bose, T. Cronin, J. Klarlund, A. Chawla, M. P. Czech, and D. G. Lambright, "Structural basis of 3-phosphoinositide recognition by pleckstrin homology domains.," *Mol. Cell.*, vol. 6, pp. 385–394, 2000.

[76] K. M. Ferguson, J. M. Kavran, V. G. Sankaran, E. Fournier, S. J. Isakoff, E. Y. Skolnik, and M. A. Lemmon, "Structural basis for discrimination of 3-phosphoinositides by pleckstrin homology domains.," *Mol. Cell*, vol. 6, pp. 373–384, 2000.

[77] R. J. Haslam, H. B. Koide, and B. A. Hemmings, "Pleckstrin domain homology," vol. 363, pp. 309–310, 1993.

[78] B. J. Mayer, R. Ren, K. L. Clark, and D. Baltimore, "A putative modular domain present in diverse signaling proteins.," *Cell*, vol. 73, no. 4, pp. 629–630, 1993.

[79] J. P. DiNitto and D. G. Lambright, "Membrane and juxtamembrane targeting by PH and PTB domains," *Biochimica et Biophysica Acta*, vol. 1761, no. 8, pp. 850–867, 2006.

[80] C. C. Milburn, M. Deak, S. M. Kelly, N. C. Price, D. R. Alessi, and D. M. F. van Aalten, "Binding of phosphatidylinositol 3 , 4 , 5-trisphosphate to the pleckstrin homology domain of protein kinase B induces a conformational change," vol. 538, pp. 531–538, 2003.

[81] J. Liu, K. Fukuda, Z. Xu, Y. Q. Ma, J. Hirbawi, X. Mao, C. Wu, E. F. Plow, and J. Qin, "Structural basis of phosphoinositide binding to kindlin-2 protein pleckstrin homology domain in regulating integrin activation.," *J. Biol. Chem.*, vol. 286, pp. 43334–43342, 2011.

[82] H. Luecke, B. Schobert, H. T. Richter, J. P. Cartailler, and J. K. Lanyi, "Structure of bacteriorhodopsin at 1.55 A resolution.," *JMB*, vol. 291, no. 4, pp. 899–911, 1999.

[83] T. Gonen, Y. Cheng, P. Sliz, Y. Hiroaki, Y. Fujiyoshi, S. C. Harrison, and T. Walz, "Lipid-protein interactions in double-layered two-dimensional AQP0 crystals.," *Nature*, vol. 438, pp. 633–638, 2005.

[84] K. Kaszuba, M. Grzybek, A. Orowski, R. Danne, T. Róg, K. Simons, U. Coskun, and I. Vattulainen, "N-Glycosylation as determinant of epidermal growth factor receptor conformation in membranes," *PNAS*, vol. 112, p. 201503262, 2015.

[85] B. A. Hall, K. A. Halim, A. Buyan, B. Emmanouil, and M. S. P. Sansom, "A Sidekick for Membrane Simulations: Automated Ensemble Molecular Dynamics Simulations of Transmembrane Helices," *JCTC*, p. 140407160657002, 2014.

[86] P. J. Bond and M. S. P. Sansom, "Insertion and assembly of membrane proteins via simulation.," *JACS*, vol. 128, pp. 2697–2704, 2006.

[87] E. Psachoulia, D. P. Marshall, and M. S. P. Sansom, "Molecular Dynamics Simulations of the Dimerization of Transmembrane Alpha-Helices," *Acc. Chem. Res.*, vol. 43, no. 3, 2010.

[88] E. Psachoulia, P. W. Fowler, P. J. Bond, and M. S. P. Sansom, "Helix - Helix Interactions in Membrane Proteins: Coarse-Grained Simulations of Glycophorin A Helix Dimerisation," *Biochemistry*, vol. 47, no. 40, 2008.

[89] M. Chavent, A. P. Chetwynd, P. J. Stansfeld, and M. S. P. Sansom, "Dimerization of the EphA1 Receptor Tyrosine Kinase Transmembrane Domain: Insights into the Mechanism of Receptor Activation," *Biochemistry*, pp. 6641–6652, 2014.

[90] D. Sengupta and S. J. Marrink, "Lipid-mediated interactions tune the association of glycophorin A helix and its disruptive mutants in membranes.," *PCCP*, vol. 12, pp. 12987–12996, 2010.

[91] K. B. Abd Halim, H. Koldsø, and M. S. P. Sansom, "Interactions of the EGFR juxtamembrane domain with PIP_2-containing lipid bilayers: Insights from multiscale molecular dynamics simulations.," *Biochimica et Biophysica Acta*, vol. 1850, pp. 1017–1025, 2015.

[92] E. Psachoulia and M. S. P. Sansom, "Interactions of the pleckstrin homology domain with phosphatidylinositol phosphate and membranes: characterization via molecular dynamics simulations.," *Biochemistry*, vol. 47, no. 14, pp. 4211–4220, 2008.

[93] A. C. Kalli, I. D. Campbell, and M. S. P. Sansom, "Multiscale simulations suggest a mechanism for integrin inside-out activation.," *PNAS*, vol. 108, no. 29, pp. 11890–11895, 2011.

[94] A. C. Kalli, I. Devaney, and M. S. P. Sansom, "Interactions of phosphatase and tensin homologue (PTEN) proteins with phosphatidylinositol phosphates: insights from molecular dynamics simulations of PTEN and voltage sensitive phosphatase.," *Biochemistry*, vol. 53, no. 11, pp. 1724–1732, 2014.

[95] A. C. Kalli, G. Morgan, and M. S. P. Sansom, "Interactions of the auxilin-1 PTEN-like domain with model membranes result in nanoclustering of phosphatidyl inositol phosphates.," *Biophys. J.*, vol. 105, pp. 137–145, 2013.

[96] C. N. Lumb and M. S. P. Sansom, "Finding a needle in a haystack: the role of electrostatics in target lipid recognition by PH domains.," *PLoS Comp. Biol.*, vol. 8, p. e1002617, 2012.

[97] J. Khandogin and D. M. York, "Quantum descriptors for biological macromolecules from linear-scaling electronic structure methods," *Proteins*, vol. 56, no. 4, pp. 724–737, 2004.

[98] A. Leach, *Molecular Modelling: Principles and Applications (2nd Edition)*. Prentice Hall, 2001.

[99] B. Hess, S. Uppsala, and E. Lindahl, "GROMACS 4 : Algorithms for Highly Efficient , Load-Balanced , and Scalable Molecular Simulation," *JCTC*, pp. 435–447, 2008.

[100] R. W. Hockney, S. P. Goel, and J. W. Eastwood, "Quiet high-resolution computer models of a plasma," *J. Comput. Physics*, vol. 14, pp. 148–158, 1974.

[101] W. L. Jorgensen, D. S. Maxwell, and J. Tirado-Rives, "Development and Testing of the OLPS All-Atom Force Field on Conformational Energetics and Properties of Organic Liquids," *J. Am. Chem. Soc.*, vol. 118, pp. 11225–11236, 1996.

[102] W. L. Jorgensen and J. Tirado-Rives, "The OPLS Potential Functions for Proteins. Energy Minimizations for Crystals of Cyclic Peptides and Crambin," *J. Am. Chem. Soc.*, vol. 110, pp. 1657–1666, 1988.

[103] A. D. MacKerell, N. Banavali, and N. Foloppe, "Development and current status of the CHARMM force field for nucleic acids.," *Biopolymers*, vol. 56, pp. 257–265, 2001.

[104] S. J. Marrink and D. P. Tieleman, "Perspective on the Martini model.," *Chem. Soc. Rev.*, vol. 42, pp. 6801–6822, 2013.

[105] S. J. Marrink, H. J. Risselada, S. Yefimov, D. P. Tieleman, and A. H. D. Vries, "The MARTINI Force Field : Coarse Grained Model for Biomolecular Simulations The MARTINI Force Field : Coarse Grained Model for Biomolecular Simulations," *J. Phys. Chem. B.*, vol. 111, pp. 7812–7824, 2007.

[106] S. J. Marrink, A. D. Vries, and A. E. Mark, "Coarse grained model for semiquantitative lipid simulations," *J. Phys. Chem. B.*, vol. 108, pp. 750–760, 2004.

[107] L. Monticelli, S. K. Kandasamy, X. Periole, R. G. Larson, D. P. Tieleman, and S. J. Marrink, "The MARTINI Coarse-Grained Force Field: Extension to Proteins," *JCTC*, vol. 4, pp. 819–834, 2008.

[108] I. Bahar and A. J. Rader, "Coarse-grained normal mode analysis in structural biology," *Curr. Op. Str. Biol.*, vol. 15, no. 5, pp. 586–592, 2005.

[109] J. Ma, "Usefulness and limitations of normal mode analysis in modeling dynamics of biomolecular complexes," *Structure*, vol. 13, no. 3, pp. 373–380, 2005.

[110] F. Tama and C. L. Brooks *Ann. Rev. Biophys. Biomol. Str.*, vol. 35, no. 1, pp. 115–133, 2006.

[111] X. Periole, M. Cavalli, S. J. Marrink, and M. A. Ceruso, "Combining an elastic network with a coarse-grained molecular force field: Structure, dynamics, and intermolecular recognition," *JCTC*, vol. 5, pp. 2531–2543, 2009.

[112] D. H. de Jong, G. Singh, W. F. D. Bennett, C. Arnarez, T. A. Wassenaar, L. V. Schäfer, X. Periole, D. P. Tieleman, and S. J. Marrink, "Improved Parameters for the Martini Coarse-Grained Protein Force Field," *JCTC*, vol. 9, pp. 687–697, 2013.

[113] P. J. Stansfeld and M. S. Sansom, "From Coarse Grained to Atomistic: A Serial Multiscale Approach to Membrane Protein Simulations," *JCTC*, vol. 7, pp. 1157–1166, 2011.

[114] P. Rotkiewicz and J. Skolnick, "Fast Procedure for Reconstruction of Full-Atom Protein Models from Reduced Representations," *J. Comput. Chem.*, pp. 2967–2970, 2008.

[115] A. Fiser and A. Sali, "Modeller: Generation and Refinement of Homology-Based Protein Structure Models," *Method. Enzym.*, vol. 374, pp. 461–491, 2003.

[116] B. Hess, "P-LINCS: A Parallel Linear Constraint Solver for Molecular Simulation," *JCTC*, vol. 4, pp. 116–122, 2008.

[117] J. P. Ryckaert, G. Ciccotti, and H. J. C. Berendsen, "Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes," *J. Comput. Phys.*, vol. 23, pp. 327–341, 1977.

[118] U. Essmann, L. Perera, M. L. Berkowitz, T. Darden, H. Lee, and L. G. Pedersen, "A smooth particle mesh Ewald method," *J. Chem. Phys.*, vol. 103, pp. 31–34, 1995.

[119] H. C. Anderson, "Molecular dynamics simulations at constant temperature and/or pressure," *J. Chem. Phys.*, vol. 72, pp. 2384–2393, 1980.

[120] S. Nosé, "A molecular dynamics method for simulations in the canonical ensemble," *Molecular Physics*, vol. 52, pp. 255–268, 1984.

[121] W. G. Hoover, "Canonical dynamics: Equilibrium phase-space distributions," *Phys. Rev. A*, vol. 31, pp. 1695–1697, 1985.

[122] H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, a. DiNola, and J. R. Haak, "Molecular dynamics with coupling to an external bath," *J. Chem. Phys.*, vol. 81, p. 3684, 1984.

[123] T. Hansson, C. Oostenbrink, and W. van Gunsteren, "Molecular dynamics simulations," *Curr. Op. Struct. Biol.*, vol. 12, pp. 190–196, 2002.

[124] E. Lindahl and M. S. P. Sansom, "Membrane proteins: molecular dynamics simulations.," *Curr. Op. Struct. Biol.*, vol. 18, pp. 425–431, 2008.

[125] T. A. Wassenaar, K. Pluhackova, A. Moussatova, D. Sengupta, S. J. Marrink, D. P. Tieleman, and R. A. Böckmann, "High-Throughput Simulations of Dimer and Trimer Assembly of Membrane Proteins. The DAFT Approach.," *JCTC*, p. 150317170838005, 2015.

[126] W. Humphrey, A. Dalke, and K. Schulten, "VMD: Visual Molecular Dynamics," *J. Mol. Graph.*, vol. 14, pp. 33–38, 1996.

[127] N. J. Bessman, D. M. Freed, and M. A. Lemmon, "Putting together structures of epidermal growth factor receptors," *Curr. Op. Struct. Biol.*, vol. 29, pp. 95–101, 2014.

[128] H. Koldsø, D. Shorthouse, J. Helie, and M. S. Sansom, "Lipid Clustering Correlates with Membrane Curvature as Revealed by Molecular Simulations of Complex Lipid Bilayers," *PLOS Comp. Biol.*, vol. 10, 2014.

[129] G. Hedger, M. S. P. Sansom, and H. Koldsø, "The juxtamembrane regions of human receptor tyrosine kinases exhibit conserved interaction sites with anionic lipids," *Scientific Reports*, vol. 5, p. 9198, 2015.

[130] M. Suzuki, L. Mao, and M. Inouye, "Single protein production (SPP) system in Escherichia coli.," *Nat. Protoc.*, vol. 2, pp. 1802–1810, 2007.

[131] J. M. Cuthbertson, D. A. Doyle, and M. S. P. Sansom, "Transmembrane helix prediction: a comparative evaluation and analysis.," *PEDS*, vol. 18, pp. 295–308, 2005.

[132] C. Oostenbrink, A. Villa, A. E. Mark, and W. F. van Gunsteren, "A biomolecular force field based on the free enthalpy of hydration and solvation: the GROMOS force-field parameter sets 53A5 and 53A6.," *J. Comp. Chem.*, vol. 25, pp. 1656–1676, 2004.

[133] M. Parrinello, "Polymorphic transitions in single crystals: A new molecular dynamics method," *J. Appl. Phys.*, vol. 52, p. 7182, 1981.

[134] M. H. Quenouille, ""Problems in Plane Sampling"," *Ann. Math. Stat.*, pp. 355–375, September 1949.

[135] J. W. Tukey, "Bias and confidence in not quite large samples," *Ann. Math. Stat.*, 1958.

[136] M. Wilke, "An Iterative Jackknife Approach for Assessing Reliability and Power of fMRI Group Analyses," *PLoS ONE*, vol. 7, p. e35578, 2012.

[137] A. C. E. Dahl, M. Chavent, and M. S. P. Sansom, "Bendix: intuitive helix geometry analysis and abstraction.," *Bioinformatics*, vol. 28, pp. 2193–2194, Aug. 2012.