MOLECULAR DYNAMICS SIMULATION STUDIES OF TRANSMEMBRANE SIGNALLING PROTEINS

Khairul Bariyyah Abd Halim



Structural Bioinformatics and Computational Biochemistry Unit

Department of Biochemistry

St Hilda's College, Oxford

Hilary, 2014

Thesis submitted in partial fulfilment of the requirement

for the degree of Doctor of Philosophy at the University of Oxford

ABSTRACT

Molecular dynamics simulation studies of transmembrane signalling proteins

Receptor tyrosine kinases (RTKs) are a major class of cell surface receptors, important in cell signalling events associated with a variety of functions. High-throughput (HTP), coarse-grained molecular dynamics (CG-MD) simulations have been used to investigate the dimerization of the transmembrane (TM) domain of selected RTKs, including epidermal growth factor receptor (EGFR) and muscle-specific kinase (MuSK). EGFR activation requires not only a specific TM dimer interface, but also a proper orientation of its juxtamembrane (JM) domain. Phosphatidylinositol 4,5-bisphosphate (PIP_2) is known to abolish EGFR phosphorylation through interaction with basic residues within the JM domain. Here, a multiscale approach was used to investigate anionic lipid clustering around the TM-JM junction and how such clustering is modulated by the mutation of basic residues. The simulations demonstrated that PIP_2 may help stabilize the JM-A antiparallel dimer, which may in turn help stabilize TM domain helix packing of the N-terminal dimerization motif. A proximal TM domain residue has been implicated in the inhibition of ganglioside GM3 in phase-separated membranes. Here, CG simulations were used to explore the dynamic behaviour of the EGFR TM domain dimer in GM3-containing and GM3-depleted bilayers designed to resemble lipid-disordered (Ld) and phase-separated (Ld/Lo) membranes. The simulations suggest that the presence of GM3 in Ld/Lo bilayers can disrupt and destabilize the TM dimer, which helps to explain why GM3 may favour monomeric EGFR in vivo. To gain insights into the dynamic nature of the intact EGFR, a nearly complete EGFR dimer was modelled using available structural data and embedded in an asymmetric compositional complex bilayer, which resembles the mammalian plasma membrane. The results demonstrated the dynamic nature of the EGFR ectodomain and its predicted interactions with lipids in the local bilayer. Strong protein-lipid interactions, as well as lipid-lipid interactions, affect the local clustering of lipids and the diffusion of lipids in the vicinity of the protein on both leaflets.

Acknowledgements

First and foremost, I am enormously grateful to my supervisor, Prof Mark Sansom, for giving me the opportunity to be a part of SBCB, and for all his help, support, encouragement and understanding that kept up my enthusiasm for completing my study.

I owe a great debt to previous and current members of SBCB; firstly to Dr Ben Hall and Dr Beatrice Emmanouil, who introduced me to MD simulations and Dr Phillip Stansfeld for his help with CG2AT and Modeller during my early days in the lab. Secondly, to other colleagues in the SBCB, especially Antreas, who has always been very helpful and happy for discussion, Amanda and Heidi for their help, comments and suggestions, and Pragya Chohan for being such a great friend. I also would like to thank everybody else at the SBCB who have made it such a wonderful lab to work. Outside of the SBCB, I would like to thank Dr Sohail, who helped me a lot when I first arrived in Oxford. I feel very fortunate to have amazing friends, my best friends Aliah Hawari and Siti Aminah Setu and other Malaysian friends (especially Syarif, Nurhayati, Ridza, Farhana, Shikin, Oyie, Juliana), thank you for all the good times and the moral support throughout the years.

Yayasan Khazanah, St Hilda's College and the Biochemistry Department are gratefully acknowledged for generous funding. I have been incredibly fortunate to be a part of the Yayasan Khazanah-Oxford Centre for Islamic Studies (OCIS) Merdeka Scholarship program. Many special events and regular meetings organized by OCIS for OCIS scholars have made me feel very exceptionally welcomed, especially when I first arrived in Oxford. Also, I would like to thank my advisor at OCIS, Dr Basil Mustafa, for his advice and support during my DPhil study.

Also, I would like to thank my parents, Abd Halim and Rabiyah, my family and friends in Malaysia for their love and support in everything I do. I am truly grateful and thankful to my husband, Ramlee Ab Hamid, for his unconditional love and support throughout my life and my studies. I could never have done this without him. This thesis is dedicated to him and my precious little boys Amir and Arfan.

Khairul Bariyyah Abd Halim
University of Oxford, April 2014

List of publications arising from this thesis

Hall, B.A, Halim, K.B.A., Buyan, A., Emmanouil, B. and Sansom, M.S.P. (2014) A Sidekick for Membrane Simulations: Automated Ensemble Molecular Dynamics Simulations of Transmembrane Helices, Journal of Chemical Theory and Computation 10(5), pp2165-2175. DOI:10.1021/ct500003g

Abd Halim,K.B., Koldsø, H., and Sansom, M.S.P. (2014) Multiscale MD Simulations of the Interactions of the EGFR Transmembrane Domain with Complex Lipid Bilayers (manuscript submitted for BBA)

Abd Halim,K.B., Koldsø, H., Goose, J., Kalli, A., and Sansom, M.S.P. EGFR in a Biologically Realistic Bilayer: A Coarse-grained Molecular Dynamics Study (manuscript in preparation for Biochemistry).

Table of Contents

Abstract
Acknowledgements
List of Publications
List of Abbreviations

CHAPTER 1 1
1.1 BIOLOGICAL MEMBRANES 1
1.2 MEMBRANE LIPIDS 1
1.3 MEMBRANE PROTEINS 4
1.4 MEMBRANE PROTEIN FOLDING 5
1.5 SIGNALLING ACROSS MEMBRANES 8
1.5.1 Epidermal growth factor receptor (EGFR) 9
1.5.2 Involvement of the C-terminal tail in EGFR activation 15
1.5.3 Regulation of EGFR by lipids 15
1.6 MOLECULAR DYNAMICS AND RELATED COMPUTATIONAL APPROACHES TO STUDY MEMBRANE PROTEINS, ESPECIALLY RTKS 16
1.7 THESIS OVERVIEW 17

CHAPTER 2 19
2.1 MOLECULAR DYNAMICS SIMULATIONS 19
2.2 EQUATION OF MOTION 19
2.3 THE FORCE FIELD 20
2.3.1 Bonded interactions 21
2.3.2 Non-bonded interactions 22
2.3.3 Integration Algorithm 23
2.3.4 All-atom force field 24
2.3.5 CG force field 25
2.4 HIGH THROUGHPUT (HTP) CG-MD 29
2.5 CG2AT CONVERSION AND AT SIMULATION 30
2.6 CONSTRAINTS 31
2.7 PERIODIC BOUNDARY CONDITION (PBC) 31
2.8 TEMPERATURE AND PRESSURE CONTROL 32
2.9 ENERGY MINIMIZATION 32
2.10 HOMOLOGY MODELING 32
2.11 SOFTWARE AND ANALYSIS 33

CHAPTER 3 34

3.1 INTRODUCTION 34
3.1.1 EGFR and MuSK *35*
3.2 SIMULATION SETUP 36
3.3 RESULTS 40
3.3.1 CG-MD simulation of EGFR *40*
3.3.2 CG-MD simulation of the EGFR mutant *43*
3.4 CONVERGENCE ANALYSIS 46
3.5 MULTISCALE MD SIMULATIONS OF MUSK 47
3.5.1 Conversion to atomistic and production run *50*
3.6 SIMULATIONS OF MUSK MUTANTS 57
3.7 CONCLUSIONS AND FUTURE STUDIES 59
CHAPTER 4 61
4.1 INTRODUCTION 61
4.1.1 Regulation of EGFR by PIP_2 ..... *62*
4.1.2 Regulation of EGFR by GM3 ..... *63*
4.2 SYSTEM SETUP 64
4.3 CG-MD SIMULATIONS OF TM-JM IN PIP2-CONTAINING BILAYERS 66
4.3.1 AT-MD simulations of TM-JM in PC:PS:PIP_2 bilayer *74*
4.3.2 Control simulations in POPC and POPC:POPS bilayers *78*
4.4 CG-MD SIMULATIONS OF EGFR TM DOMAIN IN GM3 CONTAINING BILAYERS 80
4.4.1 Protein-lipid interactions *85*
4.4.2 Comparison of WT EGFR and K618G dimers in an Ld/Lo bilayer *87*
4.4.3 Control simulations in POPC and POPC:GM3 bilayers *87*
4.5 DISCUSSION AND CONCLUSIONS 88
CHAPTER 5 90
5.1 INTRODUCTION 90
5.2 HOMOLOGY MODELLING OF THE INTACT EGFR 91
5.3 SIMULATION SETUP 93
5.4 BEHAVIOUR OF INTACT EGFR IN A COMPLEX LIPID BILAYER 96
5.5 PROTEIN-LIPID INTERACTIONS 102
5.6 LIPID MOBILITY 105
5.7 DISCUSSION 107
CHAPTER 6 110
BIBLIOGRAPHY 113

List of abbreviations

AT atomistic
AT-MD atomistic molecular dynamics simulations
Cer ceramide
CG coarse-grained
CG2AT coarse-grained to atomistic representation
CG-MD coarse-grained molecular dynamics simulations
CHOL cholesterol
DAG diacylglycerol
DMPC 1,2-dimyristoyl-sn-glycero-3-phosphocholine
DPPC 1,2-dipalmitoyl-sn-glycero-3-phosphocholine
DUPC 1,2-(cis-cis-9,12-octadecadienoyl)-sn-glycero-3-phosphatidylcholine
EGF epidermal growth factor
EGFR epidermal growth factor receptor
ENM elastic network model
ER endoplasmic reticulum
FRET fluorescent resonance energy transfer
GM3 αNeu5Ac-(2-3)-βGal-(1-4)-βGlc-Ceramide
GSL glycosphigolipid
GPCR G protein-coupled receptor
GROMACS GROningen Machine for Chemical Simulations
HA high affinity
HTP high-throughput
JM juxtamembrane
LA low affinity
L_d lipid disordered
LH left-handed
L_o lipid ordered
MC monte carlo
MD molecular dynamics
MM molecular mechanics
MuSK muscle-specific kinase
NMR nuclear magnetic resonance
PDB protein databank
PDGFR platelet derived growth factor receptor
PIP_2 phosphatidylinositol-4,5-biphosphate
PA phosphatidic acid
PBC periodic boundary condition
PC phosphatidylcholine
PE phosphatidylethanolamine
PG phosphatidylglycerol
PI phosphatidylinositol
PM plasma membrane
PME particle mesh Ewald
POPC 1-palmitoyl,2-oleoyl-sn-glycero-3-phosphocholine
POPE 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphoethanolamine
POPS 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphoethanolserine
PS phosphatidylserine

QM quantum mechanics
RDF radial distribution function
RH right-handed
RMSD root mean square deviation
RMSF root mean square fluctuation
RTK receptor tyrosine kinase
SM sphingomyelin
TM transmembrane
VDW van der Waals
WT wild type

CHAPTER 1

Introduction

1.1 BIOLOGICAL MEMBRANES

Biological membranes encapsulate living cells, separating and protecting their inner compartment from external environments. Membranes are made of lipids and proteins; these two components act together as a dynamic barrier to regulate the passage of molecules across the membranes. The transport of molecules can occur via simple diffusion across the membrane or via active (e.g., pumps) and passive transport (e.g., channels, carrier proteins) systems, which are facilitated by membrane-associated proteins (*1*). Membrane proteins are responsible for various specialized membrane functions; some act as receptors that are involved in cell-cell recognition and signalling across membrane (e.g., receptor tyrosine kinases (RTKs) (*2*), integrins (*3*, *4*) and G-protein coupled receptors (GPCRs)) (*5*), while others participate in energy transduction (e.g., ATP synthase for oxidative phosphorylation in mitochondria). The signalling across membranes that is mediated by RTKs is of primary interest in this thesis and is discussed in more detail in the following sections and throughout the thesis chapters.

1.2 MEMBRANE LIPIDS

Approximately 5% of eukaryotic genomes encode lipids, which number in the thousands (*6*). Structurally, lipids comprise a hydrophilic head group and hydrophobic carbon tails. The former tend to associate with aqueous environments and with each other, while the latter self-associate to spontaneously form the membrane bilayer (Figure 1-1A) (*7*). Lipids can be classified into three main classes: glycerophospholipids, sphingolipids and sterols. Further modification of the head group by the addition of various chemical moieties to the sn-3 position of the diacylglycerol (DAG) backbone generates a variety of phosphatidyl lipids or, if left unmodified, phosphatidic acid (PA). For example, the attachment of choline, ethanolamine, serine, glycerol or inositol groups to the sn-3 position yields phosphatidylcholine (PC), phosphatidylethanolamine (PE), phosphatidylserine (PS), phosphatidylglycerol (PG)

and phosphatidylinositol (PI), respectively. The chemical structures of all lipids used in this thesis are shown in Figure 1-1B.

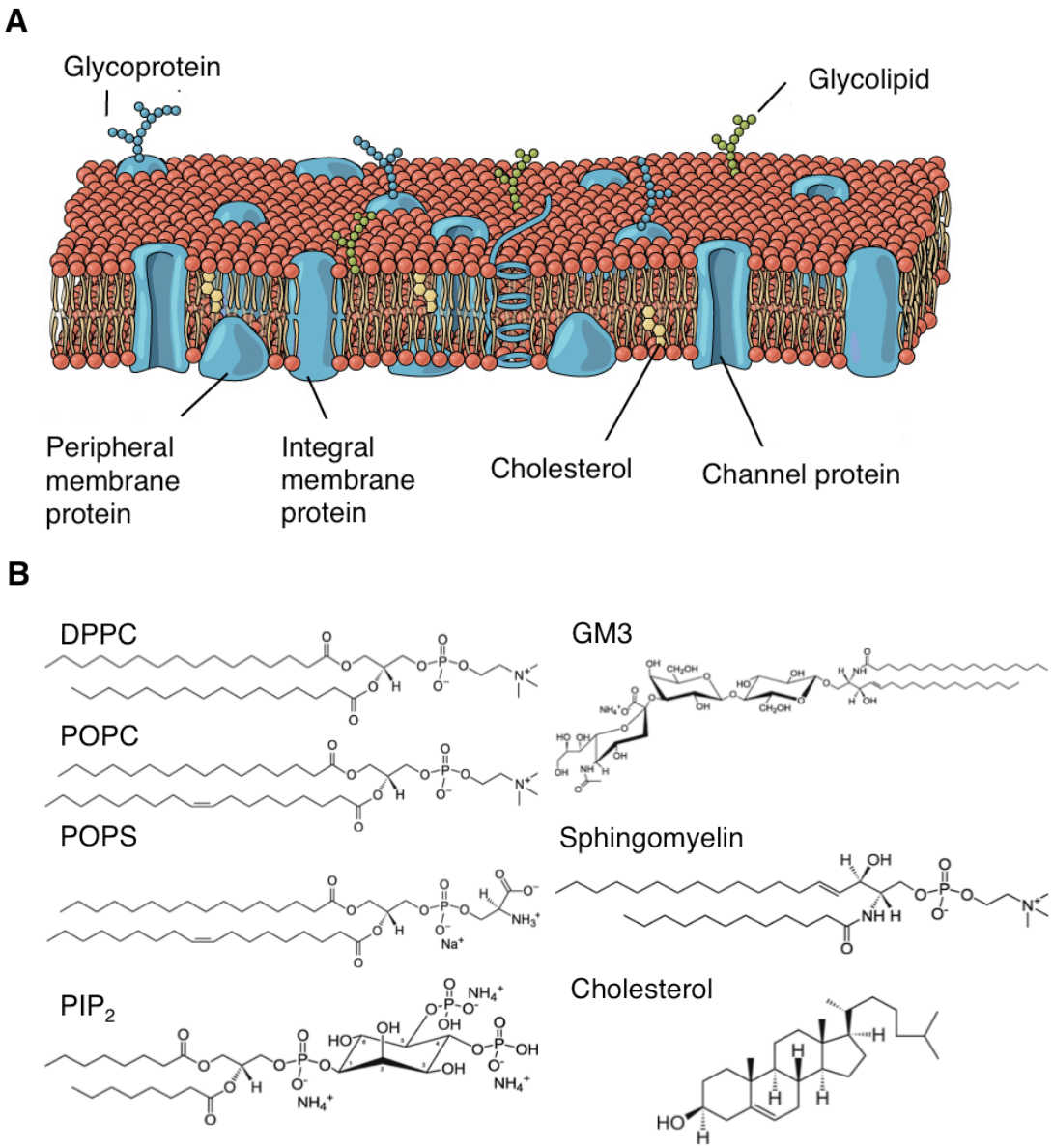


Figure 1-1. A) The cell membrane is a phospholipid bilayer containing many different molecular components, including proteins and cholesterol, some with carbohydrate groups attached. Image from Connexions, http://cnx.org/content/m46021/1.5/. B) The chemical structures of all lipids used in the MD simulations presented in the thesis. POPC (1-palmitoyl,2-oleoyl-sn-glycero-3-phosphocholine), POPE (1-palmitoyl-2-oleoyl-sn-glycero-3-phosphoethanolamine), POPS (1-palmitoyl-2-oleoyl-sn-glycero-3-phosphoethanolserine).

Meanwhile, the sn-1 and sn-2 positions of the glycerol backbone are connected to fatty acid chains of varying lengths (16–18 carbons atoms) and differing degrees of saturation via ester, alkyl ether or alkenyl ether bonds (*8*). PS, PI and PG have a net negative charge at physiological PH, while PC and PE are neutral (zwitterionic). Sphingolipids, in contrast, have a ceramide (Cer) backbone made of sphingosine amid-linked to a fatty acid. Cer can be connected to the phosphocholine moiety of sphingomyelin (SM) or to glycans (carbohydrate head groups) to form glycosphingolipids (GSLs). SM and GSLs are the major sphingolipids in mammalian cells. Sterols are the major non-polar lipids of cell membranes; cholesterol (CHOL) is a particularly important component of eukaryotic membranes (*7–9*).

Lipids are primarily synthesised in the endoplasmic reticulum (ER), although some are synthesised in the Golgi (e.g., sphingolipids) prior to their transport to other organelles. Simple, pure, phospholipid bilayers can adopt a gel phase below their melting temperature (T_m), a fluid or liquid disordered phase (L_d) above their T_m, and a liquid ordered phase (L_o) in the presence of CHOL (*7*). The different lipid phases (L_d and L_o) co-exist as microdomains, more commonly known as rafts. The new conception of lipid rafts adds another level of complexity to the fluid mosaic model of membranes proposed by Singer and Nicholson several decades ago (*10*). Lipid rafts are suggested to be dynamic, nano-scale structures (<200 nm) enriched in CHOL, SM and gangliosides (*7*), and they act as a platform for cellular signalling (*11*, *12*). Lipid raft serves as a platform for membrane protein sorting and signalling protein complexes (*11*, *13*). Some membrane proteins are associated with raft domains, while some are excluded from the raft domains (*14*), (*15*). The observation of microdomains in isolated plasma membrane vesicles derived from mammalian cells and in model membranes suggests how rafts may form in cells (*16*). Studies of lipid rafts in giant unilamellar vesicles (GUVs) composed of cholesterol and two phospholipids (saturated and unsaturated acyl chains) showed the formation of micrometre-sized Lo domains that coexist with Ld regions of the membrane. However, the precise nature of rafts (i.e., their size, lifetime and composition) and their functions in living cells remain elusive and contentious (*15*, *17*, *18*). This is because direct observations of the raft domains in vivo have been hampered by the diffraction limit of optical microscopy (~200–300 nm). However, recent advances in super-resolution microscopy have enabled the visualization of raft domains of ~5 to 200 nm in size (*19*). New advances in the field of super-resolution

microscopy are likely to provide new insights into raft structure and function, which will lead to further refinement of the lipid raft hypothesis (*19*).

The lipid composition of different membranes varies throughout the cell and the distribution of various lipids between the two membrane leaflets is either symmetric (e.g., the ER) or asymmetric (e.g., Golgi, endosomal and plasma membranes) (*20*). The outer leaflet of the plasma membrane contains mainly PC, SM and GSLs (e.g., GM3), while PS, PE and PI are generally enriched in the inner leaflet of the bilayer. In addition to phospholipids, the plasma membrane is also enriched in CHOL (about 20–25%) (*9*) compared with only 1% in the ER. Varying compositions of CHOL in various membranes, from the site of lipid synthesis in the ER towards the Golgi and PM, are linked to protein sorting, trafficking and signalling across membrane (*9*).

1.3 MEMBRANE PROTEINS

More than 20-30% of the human genome encodes membrane proteins (*21*, *22*). Membrane proteins found on the surface of the cell, called peripheral proteins, are covalently attached to the lipids on the inner or outer membrane surface. Membrane proteins that span the membrane one or more times are classified as integral membrane proteins. The majority of integral membrane proteins are made of a single spanning TM α-helix or a bundle of α-helices, while a minority of membrane proteins have a beta-barrel structure (Figure 1-2).

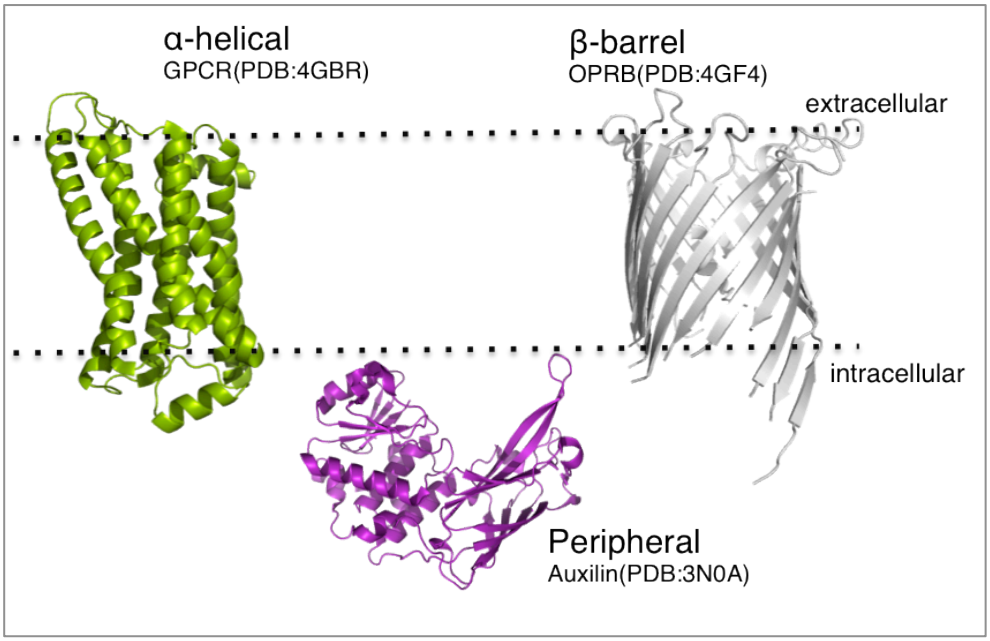


Figure 1-2. The integral (α-helical and β-barrel) and peripheral membrane proteins shown in cartoon representation.

The total number of 3D structures of known membrane proteins in the database of membrane protein of known structures at the time of the writing of this chapter (March 2014) was 1,442, 454 of these being unique. Several other databases, such as the Protein Databank of Transmembrane Proteins (PDBTM), the Membrane Protein Databank, the database of Orientations of Proteins in Membranes (OPM), also gather structures, data and studies and link them to relevant information about membrane proteins solved by various structure determination methods (e.g., X-ray crystallography, nuclear magnetic resonance (NMR) and cryo-electron microscopy (EM)).

1.4 MEMBRANE PROTEIN FOLDING

Folding of α-helical membrane proteins in cell membranes occurs via lateral association of the TM α-helices, which proceeds in two steps, insertion and assembly into a stable tertiary conformation, as described by the Popot and Engleman two-stage model (*23*). One of the driving forces of the lateral association of TM α-helices with membranes is the strong tendency of the TM helices to adopt a more energetically favourable structure via dimerization or oligomerization. Association of the TM α-helices is also driven by close contact between the interacting helices, which is mediated by polar

contacts between weakly polar resides such as Gly, Ala, Ser and Thr. These residues exist in a specific sequence pattern, or motif, collectively known as small-xxx-small or GG4 motifs, such as the GxxxG motif that was first recognized and studied extensively in glycophorin A (GpA) (*24*). Similar motifs are also found in the TM region of the major coat protein (MCP) from bacteriophages, in α_{IIb} integrin (*25*), in syndecan-2 (*26*), and also in the pro-apoptotic protein BNIP3 (*27*). Finger et al. also showed that 99% of RTKs formed homo- or heterodimers and that GxxxG-like motifs are also common in this receptor family (*28*).

The small-xxx-small motif allows the close approach of α-helices due to the surface complementarity between interacting helices, often described as "knob-into-holes" and "ridge-into-groove" packing, which promotes hydrogen bonding between weakly polar residues and between CαH and carbonyl groups across the dimer interface (*29*). These interactions can be further stabilized by van der Waals (VDW) contacts between non-polar residues, and π-π stacking, involving aromatic residues, across the packing interface. In addition, the presence of a highly polar residue, such as Asp or Glu, within a TM domain could also drive the association via hydrogen bonding and salt bridge formation. Meanwhile, positively charged residues, such as Lys and Arg, can also modulate the strength of helix-helix interactions via electrostatic interactions with negatively charged membranes at the membrane-water interface. These positively charged residues, which are frequently found at the end of the TM domain and also in the juxtamembrane (JM) domain, have been shown to influence the affinity of TM helix associations in the membrane (*30*).

In addition, the packing and stability of TM helices is influenced in part by the lipid composition of membranes, as these helices need to match the hydrophobic thickness of the bilayer with the hydrophobic length of their TM domain. This influences the tilting of the helices with respect to the bilayer plane, which may also play a role in stabilizing and sorting membrane proteins into signalling platforms (i.e., lipid rafts) (*31*). Perturbations (i.e., mutations) of stable TM helix-helix interactions can result in various diseases (*32*). Thus, TM domains have become an interesting and important target for understanding the principles that govern the TM helix-helix interactions. A better understanding of the helix-helix association could potentially aid in the design of specific peptides or small molecule inhibitors that could interfere with the contact sites

involved in the conformational changes of the receptor within the bilayer that lead to downstream signalling (*33*).

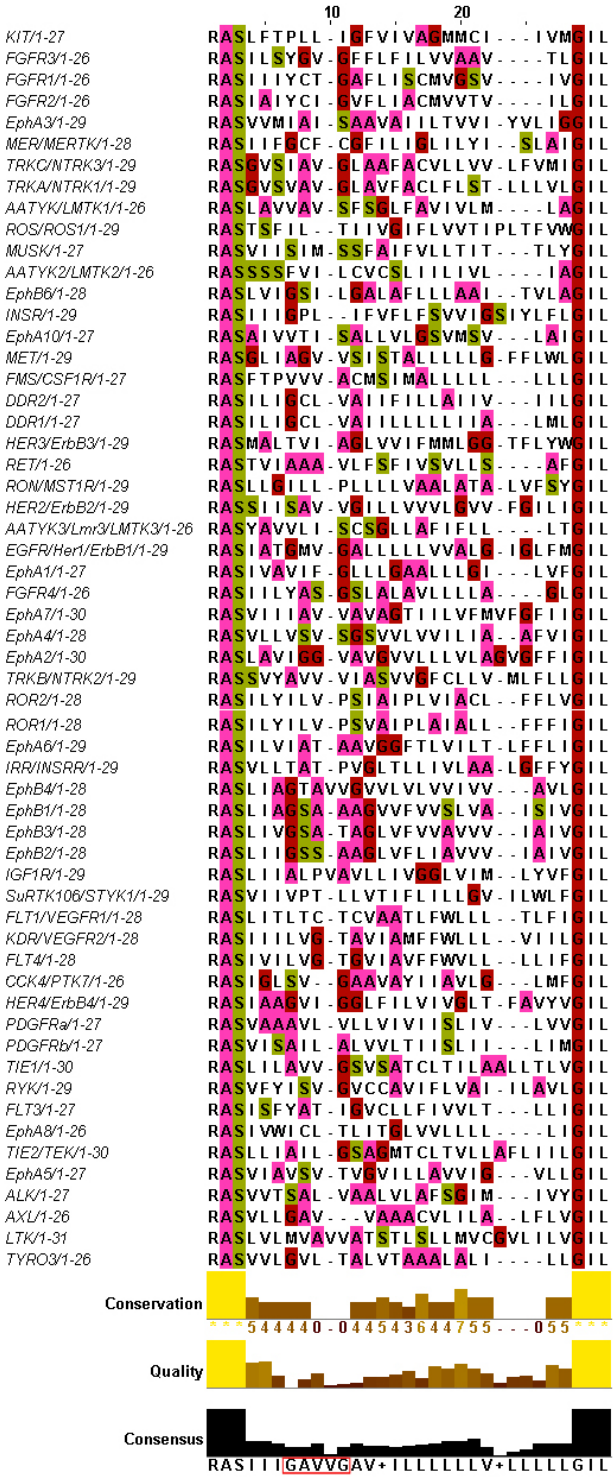


Figure 1-3. A multiple sequence alignment of 58 transmembrane domains of RTKs using EBI Muscle ClustalW http://www.ebi.ac.uk/Tools/muscle/ showing the consensus of GxxxG motifs in RTKs. The TM sequences were obtained from a recent TOXCAT study of 58 RTKs dimerization domains by Finger et al. (*28*).

1.5 SIGNALLING ACROSS MEMBRANES

External signalling molecules are detected by receptors that are embedded in the plasma membrane, which will then transmit the signals across the membrane. Signalling by single-span TM proteins, such as RTKs, involves ligand-dependent or ligand-independent dimerization, which is the first critical step in their activation mechanism (*34*). RTKs play an important role in signal transduction (*2*, *35*). Structurally, RTKs comprise an extracellular domain, a single-span TM domain, a JM region and an intracellular kinase domain. There are 58 RTKs, which are grouped into 20 subfamilies (*36*). In general, the activation of RTKs involves the binding of a ligand to the extracellular domain, which will trigger dimerization, followed by the autophosphorylation of key tyrosine residues. The phosphorylated tyrosine residues serve as sites for the binding and assembly of other signalling proteins associated with multiple cytoplasmic signalling pathways, which are linked to variety of biological processes, including cell growth, migration, differentiation and apoptosis (*2*). Epidermal growth factor receptor (EGFR) is the most well studied RTK and is the focus of this thesis. A more detailed description of EGFR is given in the following sections.

The small-xxx-small residue motif is common in RTKs and has been suggested to mediate homo- or heterodimerization of these receptors (see Figure 1-3) (*28*). However, despite the significant involvement of the TM domains in folding and biological activity, only a few structures of single-spanning TM domains have been determined thus far. TM domains of RTKs are often studied in membrane environments separately from the extracellular and intracellular domains due to the difficulty in obtaining high-resolution structures of intact, full-length proteins using available structural biology methods (*37*). A recent mutagenesis study of the FGFR3 receptor demonstrated that mutations within the TM domains persists in an isolated TM to the same extent as in the full-length protein (*38*). This study provides evidences that the study of RTKs' TM domains in isolation is relevant to full-length receptors. An investigation of helix-helix interactions mediated by the small-xxx-small residue motif is discussed in Chapter 3.

1.5.1 Epidermal growth factor receptor (EGFR)

EGFR is a member of the ErbB family of receptors, a distinct subfamily of RTKs involved in regulating key events associated with cell growth, differentiation and migration (*35*). EGFR, also known as ErbB1 or HER1, has three homologs in humans, namely ErbB2/HER2, ErbB3/HER3 and ErbB4/HER4. EGFR is activated by peptide growth factors (e.g., EGF), while ErbB3 and ErbB4 are activated by multiple variants of neuregulins (*39*) and there is no known soluble ligand for ErbB2 (see full list in Table 1-1). Members of the ErbB family associate to form either homodimers or heterodimers, thereby providing a high degree of signalling diversity (*40*). Activation of EGFR initiates a variety of signalling cascades, including the Ras/MAPK, PI3K/Akt, PLC/PKC and STAT pathways, associated with diverse cellular responses, including proliferation, differentiation, migration and apoptosis (Figure 1-4).

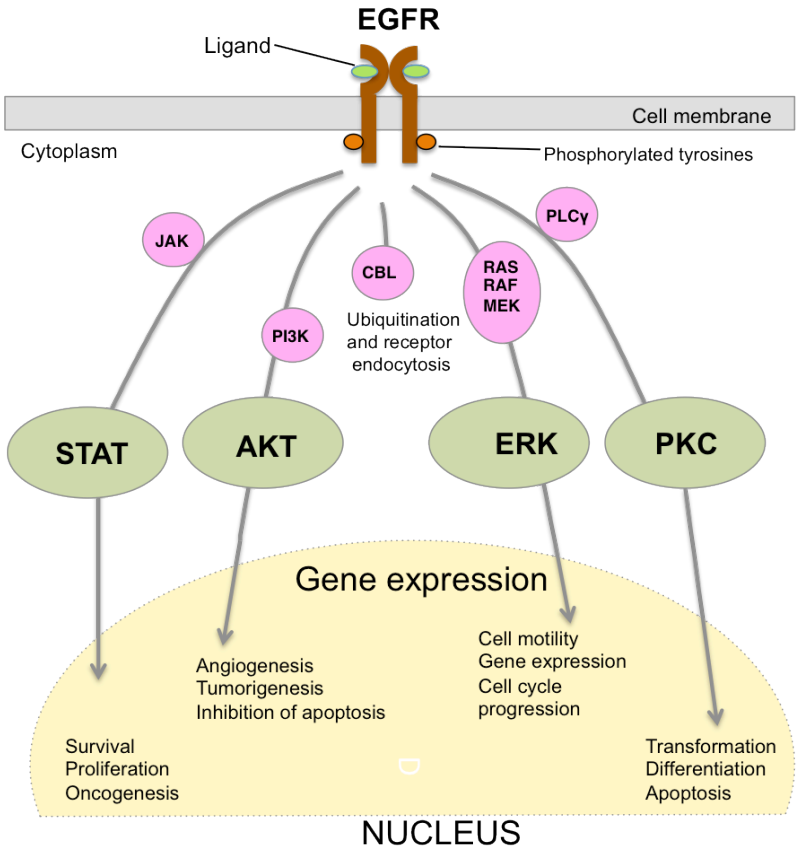


Figure 1-4. Epidermal growth factor receptor (EGFR) activation of signal transduction pathways associated with cancers. The activation of EGFR on tumour cells leads to phosphorylation of tyrosine residues in the kinase domain of the receptor and subsequent activation of the Ras/Raf/MEK/ERK, STAT, or PI3K/Akt pathways. Figure adapted from (*41*).

Activated EGFR and other RTKs are internalized within minutes of ligand stimulation, primarily via clathrin-mediated endocytosis or clathrin-independent endocytosis. After internalization, the activated receptors are sorted for recycling or degradation (*2*). Several studies suggest that following internalization, activated RTKs, including EGFR, continue to activate intracellular signalling pathways (*2*). Mutation and overexpression of these receptors has been implicated in various forms of cancers and, thus, it they have been a prime target for drug/inhibitor design for therapeutic applications (*42*). A number of tyrosine kinase inhibitors (TKIs) targeting EGFR, such as erlotinib and gefitinib, are used for treatment of non-small-cell lung cancer. For example, lapatinib, which targets EGFR and ErbB2, is currently used to treat breast cancer patients. Monoclonal antibodies, such Trastuzumab and cetuximab, are also used as cancer treatments; a detailed review is provided elsewhere (*43*).

Table 1-1. A short summary of ErbB ligands and kinase activity.

Receptor	EGFR	ErbB2	ErbB3	ErbB4
Ligand	EGF, TGF-α, Amphiregulin, HB-EGF, Betacellulin, Epiregulin Epigen	No known ligand	Heregulin, Neuregulin	Heregulin, Neuregulin, HB-EGF, Betacellulin, Epiregulin
Kinase	Active	Active	Little or no kinase activity	Active

EGFR is a 170 kDa transmembrane glycoprotein. Its 1,186 amino acids subdivided into a large extracellular region (621 amino acids), a single-spanning transmembrane domain (23 amino acids), an intracellular juxtamembrane (JM) region (38 amino acids), a tyrosine kinase domain (269 amino acids) and a C-terminal regulatory region (233 amino acids) (*44*). The extracellular region of EGFR consists of two homologous ligand binding domains (I and III) and two Cys-rich domains (II and IV), as shown in Figure 1-5A (reviewed by Ferguson (*45*)).

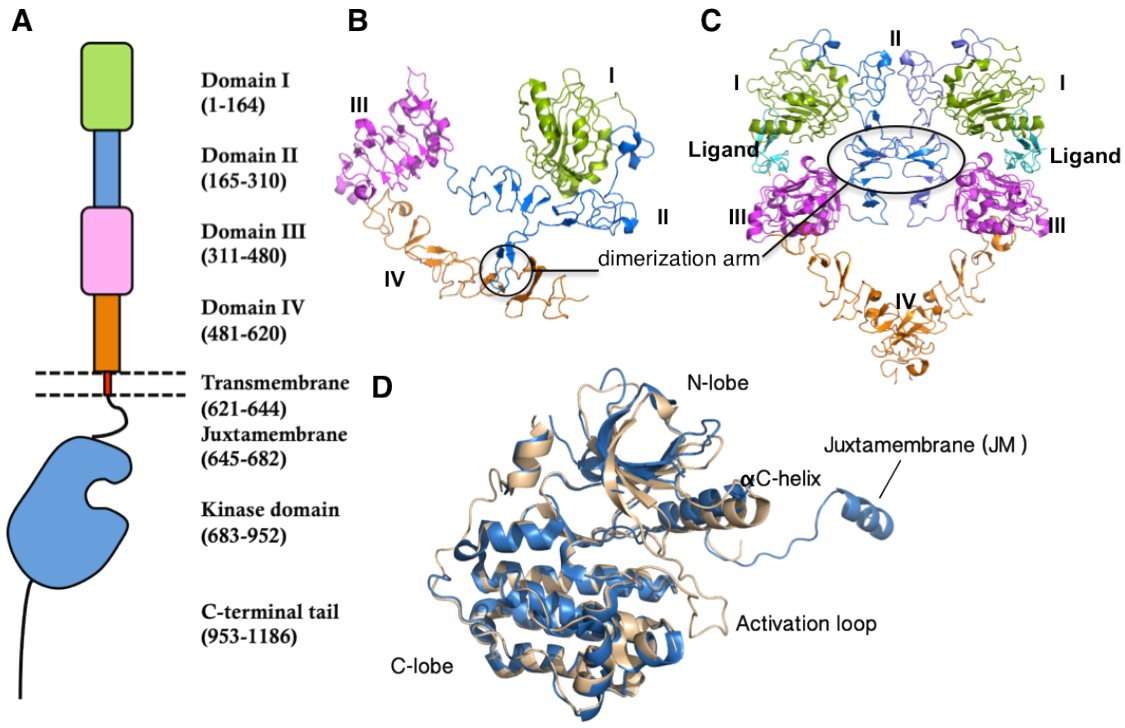


Figure 1-5. A) The schematic representation of monomeric EGFR domains. Residue numbering corresponds to the human EGFR , excluding the 24 residue signal sequence. B) Inactive EGFR monomer in a tethered conformation is shown in cartoon representation. Domains I, II, III and IV are labelled and shown in green, blue, magenta and orange, respectively. The dimerization arm on domain II is circled. C) The extended dimer with two ligands (cyan) bound to the ligand binding sites on each dimer subunit (domains I and III). D) Crystal structure of the EGFR JM-kinase domain (PDB:3GOP (blue)) superimposed on another crystal structure (PDB:2JIU(light orange)). The figure is adapted from (*45*).

The ectodomain of EGFR can exist as a monomer (tethered or extended) or as a preformed dimer. Ligand binding induces conformational rearrangements of the preformed dimer or conformational changes in the tethered monomer, which result in an extended structure (*46–52*). Ligand binding to domain I and III exposes a hairpin

loop region in domain II, also known as the dimerization arm, which promotes the formation of a dimeric EGFR complex via interactions between the dimerization arms (see Figure 1-5B,C) (*2*). In the absence of ligand, the conversion of a tethered to an extended conformation is insufficient for activation (*53*), while the preformed dimer remains inactive.

In an inactive state, the EGFR kinase domain dimer adopts a symmetric conformation (*54*). In the presence of ligand, EGFR is activated in an allosteric manner through the interaction of the C-lobe of one monomer (activator) with the N-lobe of the other monomer (receiver), forming an asymmetric kinase domain dimer (*54–56*) (see Figure 1-6B). The N-lobe and C-lobe of the kinase domain have an activation loop and an αC-helix important for activation, but which are not involved in phosphorylation (phosphorylation of the activation loop can occur in some RTKs). The asymmetric kinase dimer is stabilized by the participation of the JM domain in the asymmetric contact (details on the JM domain are provided in the introduction to Chapter 4) by forming an antiparallel dimer, promoting the formation of the asymmetric kinase domain dimer (*54*). Indeed, a recent NMR study revealed that the TM-JM domain adopts a specific TM interface at its N-terminal motif, not at its C-terminus, which is compatible with the JM antiparallel dimer, as proposed by Jura et al. (*57*). Furthermore, the JM region has also been suggested to play an important role in other RTKs, such as Eph receptors (in the binding of EPhB2/Nuk to SH2 domains) (*58*)), MuSK, and PDGFR-beta (an alanine-scanning mutagenesis study of PDGFR-beta has shown that the Ala substitution in the JM region of this receptor results in the constitutive activation of the receptor (*59*), which occurs via the disruption of the inhibitory WW domain in the JM region). Other activation mechanisms have been proposed, including the electrostatic switch triggered by the binding of calmodulin to the calmodulin binding site on the JM-A segment (*60*).

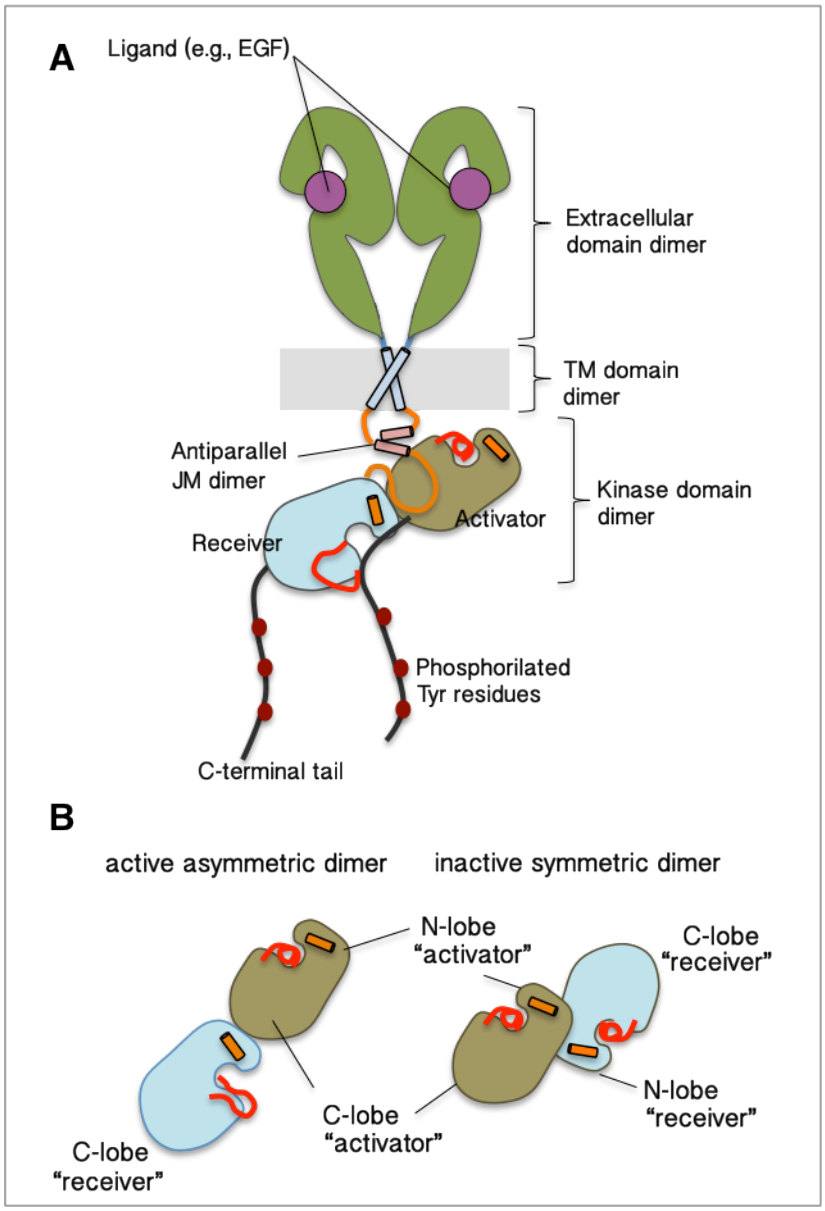


Figure 1-6. Schematic representation of EGFR domain organization of an active EGFR. A) In the presence of ligand, the TM domain associates at a specific interface that favours the proper orientation of the JM antiparallel helix dimer in the cytoplasmic site. This will in turn stabilize an asymmetric association of the kinase domain. B) Schematic diagram of an asymmetric kinase domain dimer forming a contact interface at the N-lobe of the receiver and C-lobe of the activator (left). Schematic diagram of a symmetric kinase domain dimer conformation (right). This figure is adapted from (*54*).

In addition, EGFR exhibits two distinct classes of binding affinities: low affinity (LA), K_D = 2 nM (~90%), and high affinity (HA), K_D = 300 pM (~10%) based on a Scatchard plot analysis (*61*). This was initially thought to correspond to the existence of two distinct populations of receptors. However, McDonald and Pike demonstrated that the concave nature of the Scatchard plot is indicative of negative cooperativity (*62*), which means that ligand binding to the first binding site on the ectodomain dimer reduces the binding affinity of the second binding site for EGF (*63*, *64*). However, this phenomenon cannot be explained by the available crystal structures of the human ectodomain that display a symmetric structure (PDB:3NJP) (*65*). In 2010, a structural study of the isolated *Drosophila* ectodomain by Alvarado et al. provided the basis of the negative cooperativity (PDB: 3LTF and 3LTG) (*64*). This study suggests that high affinity ligand binding to the first binding site and the singly liganded dimer imposed some restraint that reduced the binding affinity of the second site. The existence of singly liganded, active dimers of human EGFR was provided by cell-based assays of EGFR and ErbB4 (*66*). It has also been suggested that the different affinities are linked to different localizations of the receptor on the plasma membrane or oligomerization in the form of a back-to-back dimer that lies almost flat on the membrane surface (*67*, *68*). The existence of HA and LA EGFR has also been shown to activate distinct signalling pathways (*69*).

Additionally, in a recent study, He and Hristova showed that replacing the JM with an unstructured sequence only affects the transition from an inactive to active dimer. The stability of the inactive dimer and ligand binding were not affected by the perturbation (*70*). This study provides further insight into the role of the JM domain in the EGFR activation mechanism. A recent negative-stain EM study of the nearly complete EGFR (lacking the TM domain and the residues 999–1186 at the C-term tail) provided the simultaneous visualization of the extracellular domain and the cytoplasmic domain (*71*). This study revealed that there is only one ectodomain conformation of an activated receptor that is coupled to multiple kinase domain arrangements. Thus, the JM domain plays an important role in modulating the association of the “activator” and “receiver” in the kinase domain. Structural studies of the kinase domains of other ErbB receptors suggest that the mechanism of activation adopted by EGFR is shared by other family members, including ErbB3 (*72*). Nonetheless, the mechanistic detail of the

conformational coupling between the different domains of EGFR has not been fully elucidated.

1.5.2 Involvement of the C-terminal tail in EGFR activation

Recent structural and biochemical studies have shown that the activation of EGFR kinase by dimerization also involves the C-terminal tail. The C-terminal tail is rich in proline residues and contains five phosphorylation sites that are important, but not essential, for EGFR activity, as EGFR is allosterically activated by asymmetric dimerization of its kinase domain (*73*, *74*). A study of the V824R mutation at the asymmetric dimer interface suggests that the C-terminal region plays an inhibitory role by occluding the JM docking site on the kinase activator (*54*). Another study of an EGFR mutant that lacked the C-terminal tail showed ligand-independent constitutive activation (*75*). However, despite extensive characterizations of the active kinase domain, the conformational state of the C-terminal tail is still uncertain (*76*) and no structural data is available for this region. Note that the C-terminal domain is not included in the modelling and simulation of the intact EGFR presented in Chapter 5.

1.5.3 Regulation of EGFR by lipids

Apart from the conformational changes in the EGFR domains that provide a tight control of EGFR activity, the lipid environment is thought to play a role in the regulation of EGFR activity (*77–79*). EGFR is localized in CHOL-rich membrane domains, such as caveolae and lipid rafts. It has been suggested that EGFR activity is suppressed in lipid rafts and that EGFR is activated outside of raft domains (*80*). CHOL depletion has been shown to increase EGFR activity, decrease internalization and down-regulate the receptor (*81*). EGFR was identified as the target of the inhibitory action of GM3; other gangliosides exerted much smaller effects on EGFR, indicating the involvement of highly specific interactions (*82*). A recent study by Coskun et al. further showed that EGFR activity is inhibited by ganglioside GM3 only in CHOL-rich Ld/Lo domains and not in CHOL-depleted Ld domains (*77*). They proposed that GM3 binding to EGFR could perturb the intrinsic capacity of EGFR TM-JM dimerization, thus further stabilizing the inactive conformation or preventing the activation of the receptor via asymmetric kinase dimerization (*77*).

EGFR is also regulated by PIP_2 via its JM domain (*78*). Michailidis et al. demonstrated the PIP_2 binding to the JM region increased EGFR activity, as EGFR phosphorylation increased upon interaction with PIP_2. Binding of PIP_2 to the receptor is abolished through the neutralization of key residues in the JM region (*78*). A recent study of the wild-type (WT) and oncogenic Neu TM-JM domains using solid state NMR and fluorescence resonance energy transfer (FRET) provided evidences that the TM-JM domain of WT Neu is released from the membrane upon addition of PIP_2, and that the oncogenic/active conformation is not bound to the negatively charged membrane (*30*). This study provides evidence that when the TM-JM domain is in a proper orientation that reflects the active state of the receptor, the JM domain is released from membrane. However, the precise mechanism of EGFR regulation by PIP_2 and GM3 is still not completely understood.

1.6 MOLECULAR DYNAMICS AND RELATED COMPUTATIONAL APPROACHES TO STUDY MEMBRANE PROTEINS, ESPECIALLY RTKS

Computational modelling and simulation techniques have been extensively used to study membrane proteins (*83*). For RTKs, computational techniques, such as modelling and simulations, have been employed to study the dynamic association of isolated TM domains embedded in membrane bilayers, and also to study the dynamics and mechanistic aspects of their activation and signalling (e.g., the conformational transition of the kinase domain from an inactive to active state) (*84–87*).

In general, MD simulations are one of the most effective and informative approaches for detailed investigations of various interactions and processes that occur in biomolecular systems (e.g., protein-protein and protein-lipid interactions, conformational changes, interactions with water, ions, ligands, etc.) (*88*). However, this method is sensitive to the initial starting configurations, and MD calculations require expensive computational resources. Thus, the time-scales that can be reached are restricted to the system size and available computational resources. Despite its limitations, this method has been extensively used to probe the dynamics and energetic properties of TM helix associations of RTKs and other TM signalling proteins (*89–93*).

MD simulations have provided a useful tool to probe and to characterize atomic-level details of conformational changes (e.g., the transition of the kinase domain from an active to inactive state) of RTKs to gain insights into the sequence of molecular events that occur upon ligand binding, which induce conformational changes that eventually lead to the activation of the receptor. EGFR is one of the most studied RTKs experimentally and computationally; the following paragraph provides a short review of MD simulation studies of EGFR.

MD simulations have been used to study clinically identified mutants of ErbB kinase at nanosecond to microsecond timescales. This study examines the molecular mechanism underlying the aberrant signalling of its oncogenic mutants (*86*, *87*). Additionally, docking, combined with MD simulations of inhibitor binding to active/inactive EGFRs, provides insights into EGFR acquired resistance and selectivity towards different inhibitors currently used in cancer treatment (*94*). A more recent application of MD simulations has been extended by Arkhipov et al. to the modelling and simulation of EGFR using a divide-and-conquer approach, in which they performed microsecond timescale atomistic MD simulations of the monomeric (both tethered and extended) and dimeric forms of EGFR, in both active-ligand bound and inactive ligand-free dimers, as well as nearly complete/full-length EGFR (*95*). This study shed some light on the conformational coupling across EGFR domains that leads to its activation. MD simulations have also been employed to study the possible oligomerization structure of EGFR in a POPC membrane (*68*, *96*). The free energy landscape of TM helix association can be explored using potential mean force (PMF) calculations (*89*, *92*) or using an enhanced sampling method, called metadynamics (*97*). The energetics of EGFR TM domain association has been explored by PMF (*92*) and metadynamics (*98*).

1.7 THESIS OVERVIEW

This thesis describes the use of CG-MD and AT-MD simulations to gain insights into the dimerization and protein-lipid interaction of selected RTKs. The theoretical background of the methods used in the thesis is explained in Chapter 2. Chapter 3 discusses the use of CG-MD HTP methodology to predict the dimerization of TM signalling proteins EGFR and MuSK, which contain small residue motifs. The CG predicted dimer showed a good correlation with experimental data, indicating the reliability of this method, despite its limitations, for use in investigating other TM

signalling proteins. The effect of mutations in the small residue motif on TM packing was also investigated.

Chapter 4 focuses on more compositionally complex bilayer, in contrast to the simple, pure, DPPC bilayer studied in Chapter 3. Chapter 4 is divided into two parts. The first part examines the TM-JM domain of EGFR in bilayers containing anionic lipids. CG-MD and AT-MD simulations were performed using the TM-JM NMR structure in PIP_2-containing and PIP_2-depleted membranes. The JM domain plays an important role in stabilizing the asymmetric kinase domain of the intact receptor. It has also been implicated in the interaction with membranes, and, thus, this region was included to investigate its interactions with anionic lipids. The effects of mutations within the JM region on the JM-JM association and JM-membrane interaction were also examined. In the second part, I present the CG-MD simulations of an EGFR TM dimer in GM3-containing bilayers. The effect of mutation of the proximal TM residue that has been implicated in GM3-EGFR inhibition was also investigated.

Chapter 5 considers the intact EGFR in a complex, physiologically relevant bilayer using CG-MD simulations. The dynamics of protein-lipid interactions and lateral diffusion of the different lipids in the complex, asymmetric bilayer, which mimics the mammalian plasma membrane, is investigated and discussed in this chapter. The final chapter, Chapter 6, provides a summary of the thesis and future research directions.

CHAPTER 2

Theory and Methods

2.1 MOLECULAR DYNAMICS SIMULATIONS

MD simulations were first used in the 1950s to study the behaviour of atoms using a hard-sphere model (*99*), and they have been widely used ever since. The first application in the field of protein chemistry was pioneered by McCammon and co-workers in 1977 (*100*). Currently, MD simulations have been extensively used and have become an important tool to study biological systems, for example, to study protein dynamics using experimentally determined structures or homology models. MD simulations have been used alongside experimental techniques, as the 3D structure alone cannot fully describe and characterize the structural dynamic properties of proteins (*101*, *102*). MD simulations provide a picture of the time-dependent evolution of protein structural dynamics on nanosecond to microsecond timescales (*102*). The development of coarse-graining extends the accessible timescale further. Modelling of complex biological systems using a multiscale approach has becoming increasingly popular. The recent Nobel Prize in Chemistry was awarded to Martin Karplus, Michael Levitt and Arieh Warshel for "Development of Multiscale Models for Complex Chemical Systems", in recognition of the importance of molecular modelling and simulations, thereby providing inspiration for current and on-going research in this field (*103*).

This chapter briefly describes the theoretical background of MD simulations employed in the subsequent chapters. Additionally, it includes some descriptions of other methods used in modelling and analysis of the MD simulations.

2.2 EQUATION OF MOTION

The underlying principles behind MD simulations is Molecular Mechanics (MM), which refers to the use of simple potential energy functions to model molecular systems (*104*). MD is one of the many useful applications of MM, other examples are Monte Carlo (MC) and ligand docking simulations (*105*). In MM, each atom is treated as a point mass, based only on its nuclear position, while the positions of its electrons are

ignored (the Born-Oppenheimer approximation). This is in contrast with Quantum Mechanics (QM), which takes into account the movement of electrons and, thus, is very computationally demanding and limited to relatively small systems.

During MD simulations, all particles (an atom or a group of atoms) in a system are allowed to interact over time and their consequential movements are calculated by solving Newton's equation of motions (*104*). Thus, the force F on an atom of mass m and acceleration $\boldsymbol{a}$ is:

$$\boldsymbol{F} = m\boldsymbol{a} \tag{2.1}$$

During MD simulations, the force F acting on a particle or atom i is calculated using the negative derivative of the potential energy function:

$$F = -\frac{d}{dr} V\left(\mathrm{r}_1, \mathrm{r}_2, \ldots\ldots\ldots \mathrm{r}_\mathrm{N}\right) \tag{2.2}$$

A Hamiltonian description of a system combines the kinetic energy and potential energy terms. The potential energy is the sum of the contributions of the bonded and non-bonded interactions between atoms or particles in a system.

$$V(\boldsymbol{r}^{\boldsymbol{N}}) = \sum V^{bonded} + \sum V^{non-bonded} \tag{2.3}$$

$V(\boldsymbol{r}^{\boldsymbol{N}})$ denotes potential energy as a function of the positions, r, of N number of atoms or particles in the system.

2.3 THE FORCE FIELD

A critical aspect of MD is the mathematical model used to approximate the atomic-level force action on the simulated system. This is done based on the force field that has been set up to perform the calculation. The force field comprises a set of parameters and mathematical functions for calculating the bonded and non-bonded interactions of atoms. There are many force fields in existence, eight of them have been systematically compared by (*106*). A few examples of widely used force fields are CHARMM (*107*, *108*) and OPLS (*109*, *110*). The force field must be sufficiently accurate, which means

that it should perform consistently well in reproducing experimental data (e.g., NMR, kinetics or thermodynamics data).

2.3.1 Bonded interactions

The contribution of bonded interactions is given by the sum of the bond, angle, improper dihedral angle and torsional angle terms, as presented in equation 2.4.

$$\sum V^{bonded} = \sum V^{bonds} + \sum V^{angles} + \sum V^{impropers} + \sum V^{torsionals} \quad 2.4$$

Equations 2.5 to 2.8 describe the energetic penalties associated with the deviation of bonds and angles away from their reference values. The dependencies of the energy on the variations in bonds, angles and torsions are illustrated in Figure 2-1.

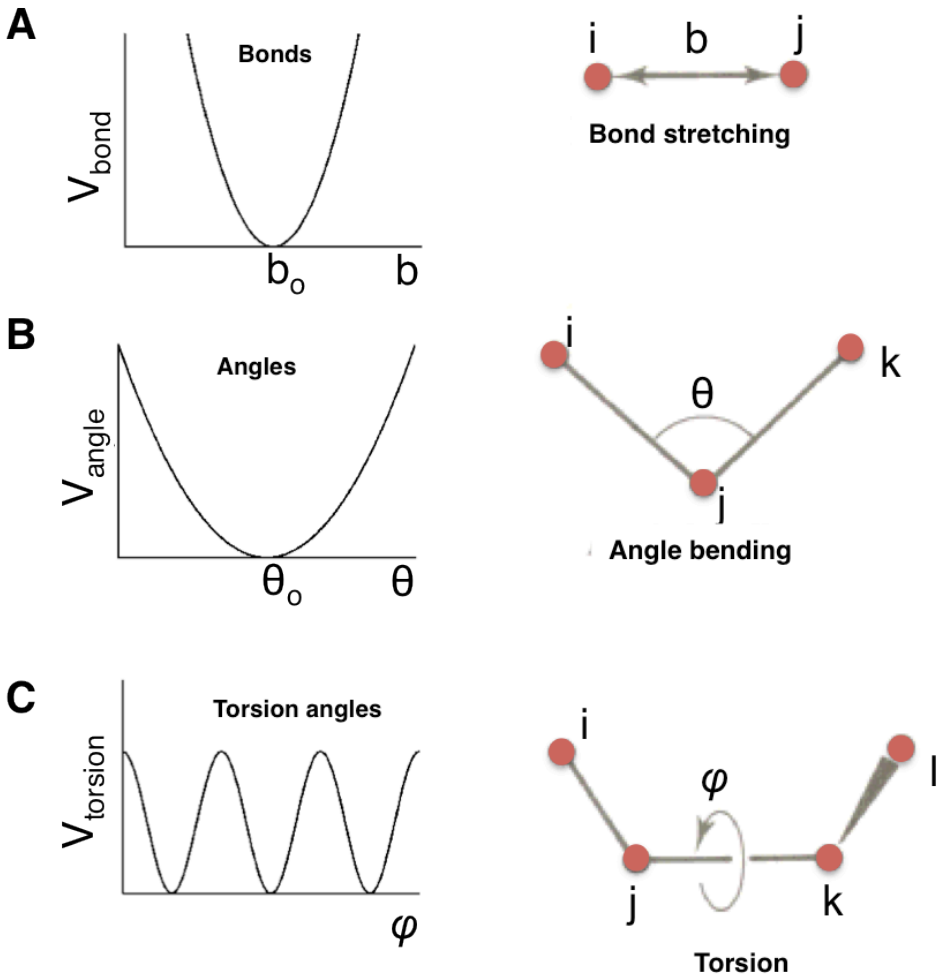


Figure 2-1. Dependencies of the potential energy on the variations in bonds (A), angles (B) and torsions (C). Schematics of the types of bonded interactions are shown on the right.

$$\sum V^{bonds} = \sum_{bonds} \frac{K_b}{2} \left[b - b_0 \right]^2 \quad 2.5$$

Equation 2.5 models the interactions between bonded atoms using harmonic potential; b_0 is the reference value for bond length derived (e.g., from crystal structures or model compounds) and K_b is the force constant. Any deviation from the reference value results in an increase in energy. Equation 2.6 models the changes in the bond angles between three atoms by harmonic potential, where θ_0 is the equilibrium bond angle and K_θ is the force constant.

$$\sum V^{angles} = \sum_{angles} \frac{K_\theta}{2} \left[\theta - \theta_0 \right]^2 \quad 2.6$$

The improper dihedral angle term is given by equation 2.7. The torsional term between four atoms is given in equation 2.8. The equation models how the energy changes when a bond rotates. K_ω and K_φ are the force constants.

$$\sum V^{impropers} = \sum_{impropers} \frac{K_\omega}{2} \left[\omega - \omega_0 \right]^2 \quad 2.7$$

$$\sum V^{torsionals} = \sum_{torsional} \frac{K_\varphi}{2} \left[1 + \cos(n - \gamma) \right] \quad 2.8$$

φ in equation 2.8 is the torsion angle. The value of n gives the number of minimum points in the function as the bond is rotated through 360 degrees. γ determines where the torsion angle passes through its minimum value.

2.3.2 Non-bonded interactions

Non-bonded interactions are interactions between all pairs of atoms that are in different molecules or that are in the same molecules, but separated by at least three bonds (equation 2.9). The van der Waals (equation 2.10) and electrostatic interactions (equation 2.11) are modelled using Lennard-Jones and Coulomb potentials, respectively (*104*).

$$V^{non-bonded} = V^{VDW} + V^{electrostatics} \quad 2.9$$

The VDW interactions between a pair of atoms (i, j) are shown below:

$$V^{VDW} = \sum_{pairs\ i,j} 4\varepsilon_{ij} \left[\left(\frac{C_{ij}}{r_{ij}} \right)^{12} - \left(\frac{C_{ij}}{r_{ij}} \right)^{6} \right] \quad 2.10$$

where ε_{ij} represents the well depth at the energy minimum and the r_{ij} distance between the two atoms, and C_{ij} is the distance at which the VDW interaction is zero.

$$V^{electrostatics} = \sum_{pairs\ i,j} \frac{q_i q_j}{4\pi\varepsilon_0} \frac{1}{r_{ij}} \quad 2.11$$

q_i and q_j represent the charges of atom i and j, respectively, ε_0 is the permittivity of free space, and r_{ij} is the separation between the two atoms.

The calculation of the non-bonded interactions are computationally expensive due to periodicity (see description for the periodic boundary condition (PBC)), which means that the non-bonded interactions should be calculated for every pair of particles in the central box and in its periodic images. To overcome this, a non-bonded cut-off can be used for short range van der Waals interactions. However, such a cut-off cannot be used for long-range electrostatic interactions because it can lead to artefacts. Thus, the particle mesh Ewald (PME) method (*111*, *112*), which is based on the Ewald method, is used to calculate electrostatic interactions. The Ewald method separates the calculation of the electrostatic interactions into short- and long-range terms that converge quickly in their respective space, real and reciprocal space, respectively. Cut-off distances are used in the real space and Fourier transformations are used in the reciprocal space. In contrast, PME uses fast Fourier transformations for both the short- and long-range terms, which allows for a much faster convergence of both components. PME was used for all the AT simulations, whereas the cut-off method was used in CG-MD simulations.

2.3.3 Integration Algorithm

In MD simulations, the initial conformation of a biological macromolecule is typically obtained from an NMR or crystallographic structure. During MD simulations, due to fact that the motions of all particles are coupled together, it is impossible to solve the equation of motion analytically. Thus, the equations of motion are integrated using a finite difference method. This method essentially breaks down the integration into many

small stages, with a fixed time interval, δt. The total force acted on each particle at time t is calculated as the vector sum of its interactions with other particles. We can determine the acceleration of the particles from the force and calculate the new positions and velocities at time $t + \delta t$. This process is repeated to calculate the next position and velocity at every time step during an MD simulation (*104*).

In GROMACS, a leap-frog algorithm is used to integrate Newton's equation of motion (*113*). The leapfrog algorithm is a variant of the most commonly used Verlet algorithm and is better at handling velocities compared with the original Verlet algorithm. It uses the following terms:

$$v\left(t + \frac{1}{2}\Delta t\right) = v\left(t - \frac{1}{2}\Delta t\right) + \frac{\Delta t}{m}F(t) \qquad 2.12$$

$$r(t + \Delta t) = r(t) + \Delta t v(t + \frac{1}{2}\Delta t) \qquad 2.13$$

The leapfrog algorithm uses the positions $\mathbf{r}(t)$ and velocities v at time $t - \frac{1}{2}\Delta t$ to update the positions and determine the velocities at time t using the force, $\boldsymbol{F}$(t), calculated for $\mathbf{r}(t)$. The equations are solved simultaneously, with temperature and pressure kept within required values, and the coordinates are output at given intervals. The coordinates as a function of time represent the trajectory of a system.

2.3.4 All-atom force field

Individual atoms and interactions between them can be explicitly represented and simulated during MD simulations, although this is computationally expensive and time consuming. OPLS (*109*, *110*) and CHARMM (*108*) are two examples of all atom force fields. All-atom MD simulations are computationally demanding because all atoms in the system are treated explicitly and calculations will involve every single pair of atoms in the system. Conversely, united-atom force fields take into account all atoms except the non-polar hydrogen. An example of a united-atom force field is GROMOS96, which has been implemented in GROMACS (*114*).

2.3.5 CG force field

A coarse-grained force field allows MD simulations to reach timescales, system sizes and complexities that are inaccessible by all-atom force fields. This is achieved through simplification of the atomistic representation by grouping four atoms into a single interaction site of a CG particle (Figure 2-2). The 4:1 mapping reduces the atomistic details, degrees of freedom and interaction sites and, thus, simplifies calculations and increases the speed of an MD simulation by up to three orders of magnitude compared with an atomistic model (*115*).

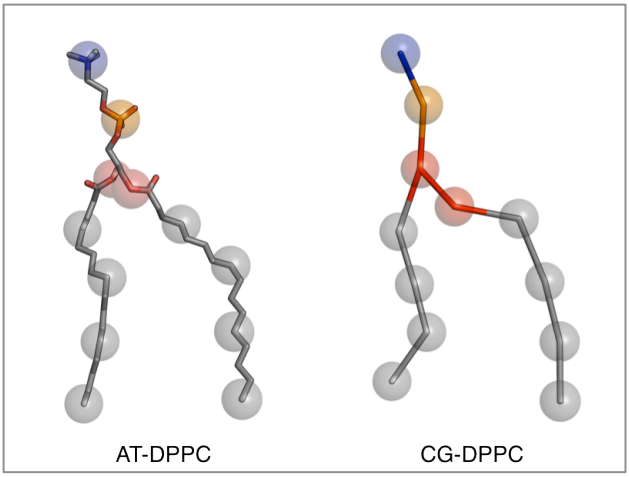
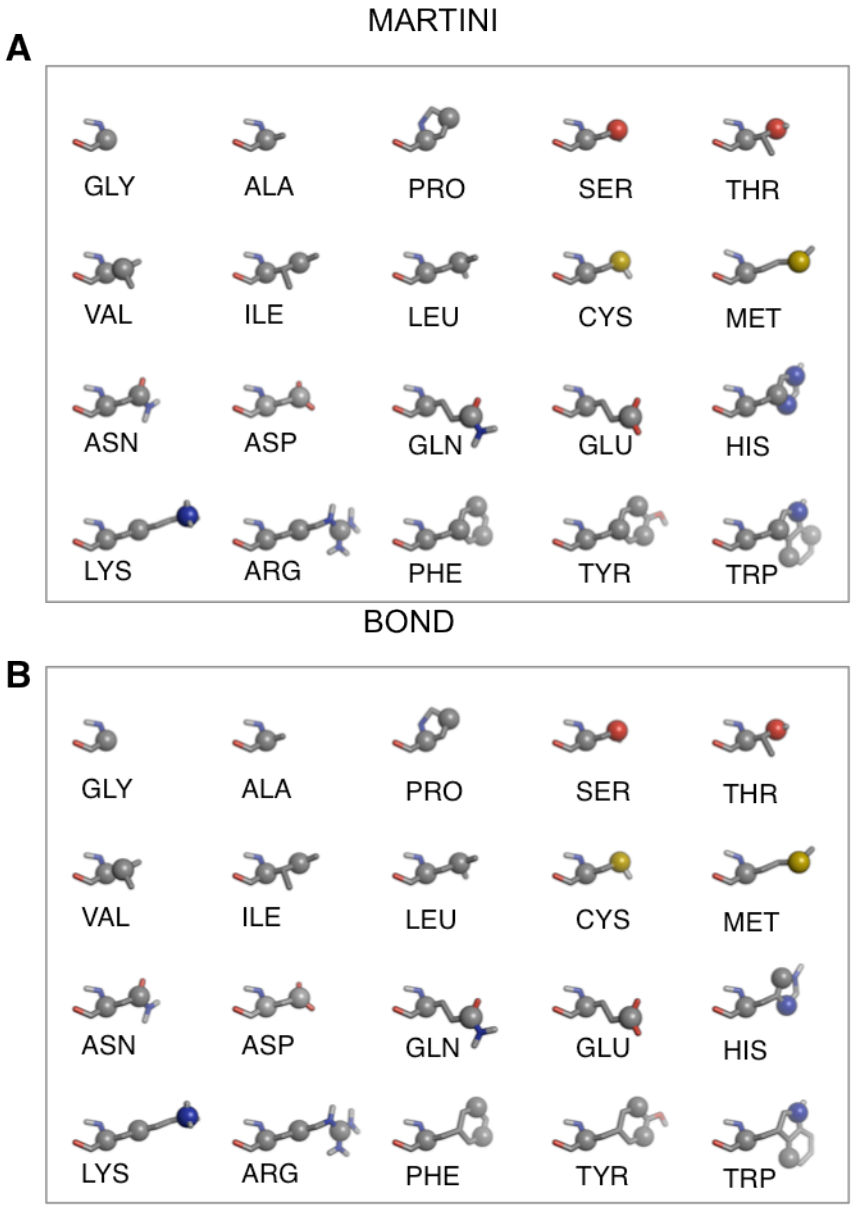


Figure 2-2. AT and CG representations of a DPPC lipid. In the CG model a 4:1 mapping of heavy atoms is made. In the CG model, the CG particles are coloured as: red for glycerol, grey for acyl chain, yellow for phosphate and blue for choline. Figure courtesy of Dr Phillip Stansfeld.

A locally modified version of the MARTINI force field (BOND force field) and MARTINI (v2.1) were employed in MD simulations presented in this thesis. The MARTINI force field was developed by Marrink et al. for semi-quantitative lipid and surfactant simulations, which they later extended to model biomolecules to study protein dynamics (*116*). Meanwhile, the BOND force field, which is based on the original MARTINI CG parameters for lipids, was developed by the Sansom group (*117*). In a CG model, the centre of mass of a group of atoms is used to determine the centre of mass of the CG particle. Both force fields have four main particles, namely P (polar), N (nonpolar), C (apolar) and Q (charged). P is for neutral and water-soluble groups of atoms, and C is for neutral groups of hydrophobic atoms. Meanwhile, N and Q are for mixed (partly polar and apolar) and charged groups of atoms, respectively, with four subtypes (0 = none, d = donor, a = acceptor, da = both) assigned to reflect

hydrogen bonding capacities. MARTINI further divided the P and C particles into five subtypes to distinguish their degree of polarity, with 1 being lower polarity and 5 being higher polarity. In MARTINI, the amino acid backbone is represented by a single particle. For the side chains, zero to four particles are used by MARTINI, whilst BOND uses a maximum of two particles for the side chain (refer to Figure 2-3). Thus, all amino acids are represented by the same number of particles in MARTINI and BOND, except for those with aromatic side chains, such Phe, Tyr, Trp, and His, which have extra numbers of particles in MARTINI (2-3B). For instance, the benzene group of Phe is represented by two CG particles in BOND and three CG particles in MARTINI (Figure 2-3). The different assignment of CG main particle types and subtypes for backbone and side chains of all amino acids are summarized in Table 2-1. BOND uses five different levels of strength of the non-bonded interactions between CG particles (I = attractive, II = semi-attractive, III = intermediate, IV = semi-repulsive, V = repulsive). Meanwhile, the strength of interactions in the MARTINI CG model are expanded to ten (*116*).



42-3. The CG particles for protein sidechains for MARTINI (A) and BOND (B). The principle difference is found in the number of particles used for the aromatic residues. Figure provided by Dr Phillip Stansfeld.

Table 2-1. CG main particle types and subtypes for MARTINI and Bond force fields.

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

Both force fields model the non-bonded interactions using standard Lennard-Jones potentials, and interactions between charged particles are modelled using Columbic potentials. Meanwhile, bonds and angles, collectively referred to as bonded parameters, are treated with a harmonic potential, with its respective force constants matching the experimental or atomistic MD data. In the BOND force field, for example, bond potential uses a force constant of 1,250 $kJmol^{-1}nm^{-2}$, with an equilibrium bond length of 0.38 nm. BOND angle potentials use a force constant of 25 $kJmol^{-1}nm^{-2}$ for secondary structure elements and 35 $kJmol^{-1}nm^{-2}$ for random coils. The equilibrium bond angles are 90°, 130° and 120° for an alpha-helices, beta-strand segments and random coils, respectively. BOND uses a harmonic distance restraint with a force constant of 1,000 kJ mol^{-1} nm^{-2} to represent the secondary structure of H-bonds between the protein backbone particles (*118*). Conversely, the MARTINI model uses proper dihedrals to impose secondary structural elements on the peptide backbone (*116*).

For protein tertiary structure, to prevent the protein structure from unfolding or becoming severely distorted during CG MD simulations, an elastic network model (ENM) is used (*119*). An ENM uses harmonic distance restraints between backbone particles, with a cut-off distance of 7 Å and a force constant of 10 kJ $mol^{-1}nm^{-2}$ to maintain the secondary structural elements of a protein or a peptide.

2.4 HIGH THROUGHPUT (HTP) CG-MD

Coarse-graining reduces the computational cost and allows CG-MD simulations to be run on a large scale, in a high throughput manner, enabling a significantly improved sampling for statistical analysis of a system and its properties of interest. A high-throughput (HTP) CG-MD pipeline, called Sidekick, was developed by Dr Benjamin Hall and has been used to study helix insertion and dimerization (*120*). Sidekick requires input sequence in FASTA format to generate an atomistic helix that is then converted to a CG representation for use in the simulation setup and runs. The Sidekick pipeline, from input sequence to retrieval of results, is fully automated and provides an efficient and systematic way of running a large ensemble of CG-MD. This is because the simulations are performed in duplicate (usually 100 or more) across a dedicated cluster and workstations, while individual simulations and analyses are performed on a

single processor. Thus, multiple simulations of a given ensemble with different initial configurations can be performed in parallel. Sidekick has been used to study the behaviour of helices in bilayers using TM helices and model peptides (*91*, *121–123*), a bacterial signalling system (*124*) and fusion peptides (e.g., simian immunodeficiency virus (*125*, *126*). An application of the HTP CG-MD methodology is presented in Chapter 3.

2.5 CG2AT CONVERSION AND AT SIMULATION

The CG model is a powerful tool to capture certain properties, such as peptide aggregation and protein lipid interactions; however, due to its simplified representation of atoms, details of interactions, such as hydrogen bonding and conformational transitions that might occur during the simulations, cannot be captured. However, interesting phenomena that occur during CG simulations can be investigated in detail at the atomistic level using reverse coarse-graining or coarse-grained to atomistic conversion strategies. The CG2AT program was employed to convert the CG generated models (Chapter 3) and the final state of the corresponding CG system to an atomistic representation (see the simulation of the TM-JM domain in Chapter 4). The CG to AT conversions in this thesis were performed using the CG2AT program developed (written in Perl) by Dr Phillip Stansfeld (*127*). The CG particles are converted to atomistic representations using a fragment-based fitting approach (*128*). The first step in the CG2AT conversion protocol is renaming all CG lipids and protein particles to their atomistic counterpart. Lipids and proteins are converted separately. For proteins, the conversion involves the reconstruction of the AT coordinates from the CG model using Pulchra (*126*) and MODELER (*130*). Lipids are converted by alignment of AT lipid fragments to the CG particle of each lipid molecule. This step is repeated for all the lipids in the system, followed by a subsequent energy minimization. The energy minimized proteins and lipids are combined and energy minimized further. The system is then solvated, and counterions are added to neutralize the system (*125*). The initial atomistic systems were energy minimized and equilibrated for 1 ns with the protein Cα atoms restrained (force constant = 10 kJ/mol/ $Å^2$), and subsequently followed by unrestrained production runs. Energy minimization and equilibration were necessary to prevent dramatic movement of some regions in the simulation systems that will cause

instability in the system. Production runs were performed in duplicate using the same initial configuration, but with different initial velocities.

2.6 CONSTRAINTS

Constraint algorithms, such as LINCS (*129*) and SHAKE (*130*), are used during an MD simulation to constrain bond length to a reference value that satisfies the equilibrium bond geometry. The LINCS algorithm was used in this thesis. The constraint is implemented by eliminating the fastest motion in a system, the bond vibration, which has little effect overall, but slows down the speed of the simulation. Thus, larger time steps could be implemented.

2.7 PERIODIC BOUNDARY CONDITION (PBC)

A PBC is used in molecular simulations to mimic the presence of an infinite bulk surrounding the model system. The system is treated as a lattice made up of infinite copies of itself in all direction (Figure 2-4). GROMACS uses PBC in combination with minimum image convention, which is implemented by taking into account only one of the nearest image of a particle to calculate short-range and non-bonded interactions (*113*). The use of a PBC allows a better approximation of the physical behaviour of a system; it also has some drawbacks, as computing the potential energy for all interactions now requires an infinite sum. Thus, suitable cut-offs are used to calculate non-bonded interactions between pairs of particles that would otherwise become infinite with the implementation of a PBC.

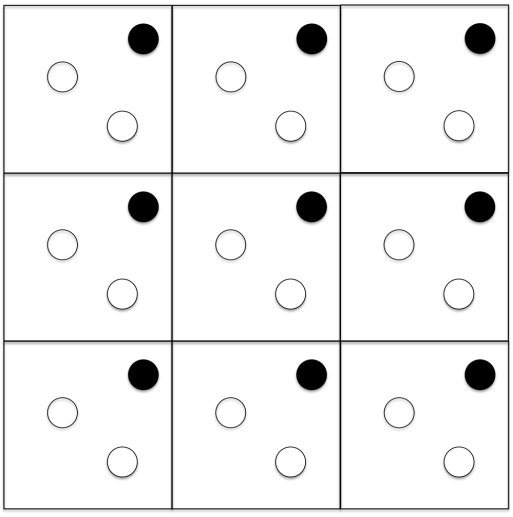


Figure 2-4. Schematic of periodic boundary conditions in two dimensions.

2.8 TEMPERATURE AND PRESSURE CONTROL

All simulations in this thesis were conducted under NPT conditions, in which the number of particles (N), pressure (P) and temperature (T) were fixed. In an NPT ensemble, temperature and pressure are regulated during the simulation using a thermostat and barostat. During a CG-MD simulation, the temperature was maintained using the Berendsen algorithm, in which the temperature is weakly coupled to an external bath (Berendsen thermostat) (*131*). In this algorithm, the instantaneous temperature is maintained at a desired temperature by scaling the velocities at each time step. This is implemented by correcting the deviations of an instantaneous temperature, T, from the desired temperature, T_O, by multiplying the velocities by a certain factor, λ, to ensure that the average temperature of the system is maintained at the desired temperature. Simulations in pure DPPC and POPC bilayers were performed at 323 K and 300 K, respectively. Different temperatures were used for different lipid mixtures, as described in Chapters 4 and 5.

The Berendsen barostat, with a coupling constant, τ_p = 10 ps, and a compressibility of 3 x 10^{-5} bar^{-1} (Chapter 3) or 5 x 10^{-5} bar^{-1} (Chapter 4), was used for pressure coupling in the CG-MD simulations. The Parrinello-Rahman barostat, with τ_p = 1.0 and a compressibility of 4.6 x 10^{-5} bar^{-1}, was used in all AT-MD simulations.

2.9 ENERGY MINIMIZATION

Energy minimization is an important step in the preparation of a system for an MD simulation. It is important to energy minimize a protein structure to relax the system, thereby releasing any unfavourable interaction prior to the MD simulations (*104*). The energy minimizations in all the MD simulations performed in this thesis were performed using the steepest descent algorithm; the search for minima proceeded down the potential energy landscape until a minimum was reached. Another commonly used energy minimization algorithm is the conjugate gradient, which works in similar way, although the direction of movement is determined by the previous step.

2.10 HOMOLOGY MODELING

The mutants used in the simulations presented in Chapter 4 were generated using MODELLER (*132*). This program was also used to construct the nearly complete

EGFR (Chapter 5) using the MODELLER multiple templates function. Multiple crystallographic structures were used to plug the missing fragments in any of the templates used. The simultaneous use of multiple templates in comparative modelling helps to minimize errors in the missing region constructed with MODELLER. The homology modelling process involves: (i) selection of suitable templates, (ii) alignment of the chosen templates with manual editing when necessary, (iii) construction of the model using MODELLER functions, and (iv) model evaluation. The models were assessed using MODELLER using the DOPE potential to select the best model for energy minimization prior to conversion to CG representation and its subsequent use in simulations.

2.11 SOFTWARE AND ANALYSIS

All simulations in this thesis were performed using the GROMACS biomolecular simulation package and analyses were performed using the GROMACS analysis tool or in-house scripts. Visualizations of the trajectories and graphics were performed in VMD (*133*) and PyMOL.

CHAPTER 3

RTK transmembrane helix dimerization: An investigation using a multiscale molecular dynamics simulation approach

3.1 INTRODUCTION

Lateral association of TM alpha helices has been associated with a defined interaction between specific sequence motifs. The motifs are commonly known as small-xxx-small or GxxxG motifs, and are responsible for close packing and stable TM helix-helix interactions (*134–136*), as exemplified in GpA (*24*). Many experimental and computational studies have examined the principles that govern the association of TM domain alpha helices association to understand the behaviour and properties of membrane proteins.

Dimerization is an important step in receptor-mediated cell signalling (*137–139*). CG-MD simulations have been used in combination with AT-MD simulations (multiscale approach) in the study of TM helix dimerization (*93*) (*140*). A HTP CG-MD approach has also been used to study the dynamics of single helix insertion and assembly, as well as for TM dimerization, with better sampling, thereby generating a good starting structure for further AT-MD refinement (*140*). Thus, it is of interest to extend this approach to explore TM helix associations of many other TM signalling proteins containing similar motifs. Here, two member of the RTK family with small-xxx-small motifs, the EGFR and MUSK were used as test cases to study TM helix dimerization using the HTP CG-MD approach. EGFR was chosen on the basis of its importance in various human cancers (*42*, *141–143*), as well as the abundance of experimental data from computational and structural perspectives. This receptor is the most well studied member of the RTK family, and the structure of its TM domain has been recently solved by NMR (PDB:2M20) (*57*). MuSK was chosen because of its importance in neuromuscular junction (NMJ) formation and also because of a lack of computational and structural information on its TM domain.

3.1.1 EGFR and MuSK

The TM region of EGFR contains one small-xxx-small motif at its C and N-terminal ends (highlighted in green in Figure 3-1). In its active conformation, the dimer contact interface is formed at the N-terminal motif ($G^{625}MVGA^{629}$), with a crossing angle of ~−44° ± 3°, as revealed by the NMR structure (*57*). During receptor activation, the helix-helix association alter the orientation of the JM region to promote a proper asymmetric association of the kinase domain. In addition, mutation of the four N-terminal residue motif to Ile (marked as asterisks in Figure 3-1) resulted in significant inhibition of the receptor (*57*).

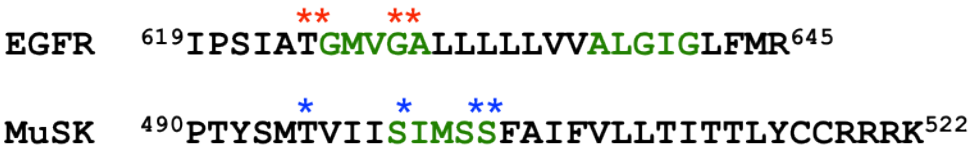


Figure 3-1. The TM sequence of EGFR and MuSK used in this study. The small-xxx-small motif is highlighted in green. The asterisks mark the residues mutated to Ile (red) and Ala (blue) to generate mutants for both candidates.

The MuSK TM domain contains a small-xxx-small ($S^{499}IMSS^{503}$). To date, there is no NMR structure available for the MuSK TM domain. This receptor plays an important role in the formation and maintenance of the NMJ (see schematic diagram in Figure 3-2) (*144*). Dysfunction in MuSK or other synaptic proteins in the pathway results in severely impaired NMJs, and is one of the causes of a family of diseases, called congenital myasthenic syndromes (CMS) (*145*). Like other RTKs (detailed discussion in Chapter 1), MuSK consists of three immunoglobulin-like (Ig) domains and a Frizzled cysteine-rich domain (C6) in its extracellular domain, a single-spanning TM helix, and the cytoplasmic domain comprises a JM domain and a tyrosine kinase catalytic domain. MuSK is indirectly activated by agrin, a heparin sulphate proteoglycan secreted from motor nerve terminals. This is because L-agrin binds another protein, called Lrp4, which exists in a heterodimer with MuSK and mediates MuSK activity (*146*). Although the molecular mechanisms and the key players in MuSK signalling pathway have been identified, details on the MuSK TM domain are still lacking, from both experimental

and computational perspectives. Thus, both EGFR and MuSK provide a good test case to explore the dimerization of TM alpha helices containing small residue motifs.

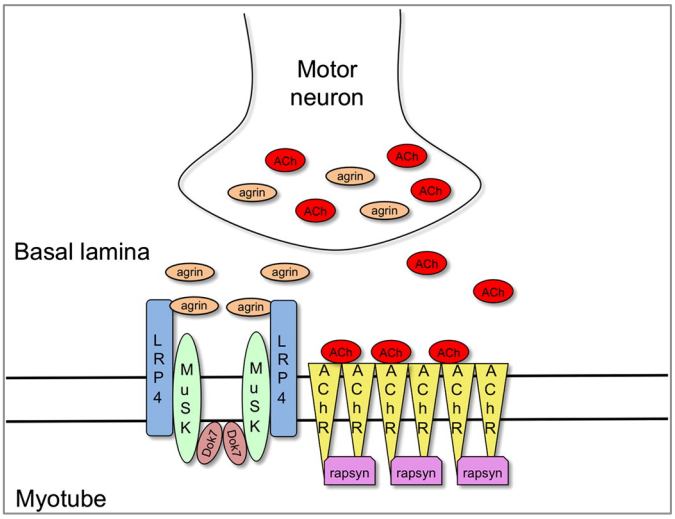


Figure 3-2. Schematic representation of key players in NMJ formation. MuSK is required for acetylcholine receptors (AchR) clustering, a critical step in NMJ formation. MuSK is indirectly activated by agrin, which is secreted from the motor neuron to the basal lamina, because Agrin binds to Lrp4, which in turn activates MuSK. Activated MuSK phosphosphorylates dowstream signalling proteins, which results in AchR clustering and muscle contraction. Picture adapted from (*147*).

3.2 SIMULATION SETUP

The sequences used in this study were obtained from UNIPROT (http://www.uniprot.org/), and the TM domain for each sequence was predicted using a consensus TM predictor (*148*). The CG-MD self-assembly simulations were performed with two helices positioned parallel to each other in a preformed dipalmitoyl phosphatidylcholine (DPPC) bilayer and separated from each other by a distance of 60 Å, a separation distance sufficiently large to prevent any influence of van der Waals or electrostatic interactions that might introduce any bias towards dimer formation (see Figure 3-3). A CG-MD HTP methodology was used in this study to enhance sampling of the association and dynamics of the EGFR TM helices in lipid bilayers (*120*). This approach enables automatic running and analysis of a large ensemble of CG-MD simulations. A list of all CG-MD and AT-MD simulations performed are detailed in Table 3-1.

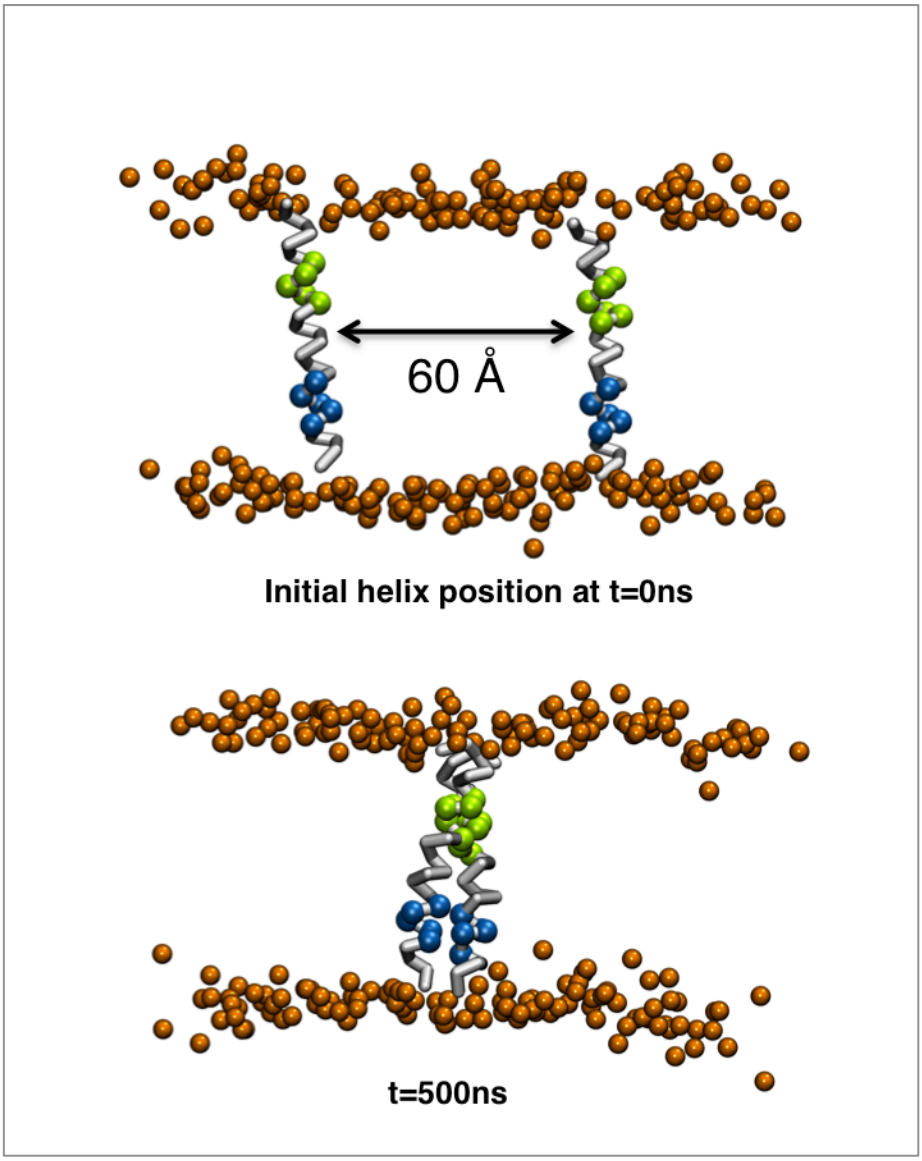


Figure 3-3. The initial configuration of completely separated EGFR TM helices in a DPPC bilayer at t = 0 ns and of dimerized helices at t = 500 ns. The N-terminal and C-terminal small-xxx-small residue motifs are shown in green and blue spheres, respectively. The snapshots were taken from one of the EGFR CG-MD simulations. The phosphate group particles of the DPPC lipids are shown as orange spheres; water particles are not shown for clarity.

Table 3-1 Summary of TM α-helices simulations

Simulation	No. of simulations x Length
CG-MD	
EGFR WT	50 x 500 ns
EGFR 4I (T624I/G625I/G628I/A629I)	25 x 500 ns
MuSK WT	77 x 500 ns
MuSK4A (T495A/S499A/S502A/S503A)	50 x 500 ns
MuSK 3A (S499A/S502A/S503A)	50 x 500 ns
MuSK T495A	50 x 500 ns
AT-MD	
MUSK WT_parallel AT	3 x 100 ns
MUSK WT_RH AT	3 x 100 ns
MUSK WT_LH AT	3 x 100 ns

The system was solvated with CG water particles and, subsequently, an appropriate number of counter ions (NA+ or CL-) were added to neutralize the system. Lennard-Jones functions and van der Waals interactions were shifted to zero between 0.9 and 1.2 nm. Meanwhile, electrostatics interactions were shifted to zero between 0 and 1.2 nm, with a relative dielectric constant of 20. The non-bonded neighbour list was updated every 10 steps. All simulations were conducted at constant temperature, pressure and particle numbers. Temperature and pressure were coupled using the Berendsen algorithm at 323 K ((τ_T = 1 ps and τ_P = 10 ps). The CG-MD simulations were performed in GROMACS3.2.2 (www.gromacs.org) (*149*, *150*).

Cluster analysis using GROMACS g_cluster on the Cα particle was performed to identify dimers that are more frequently found (*151*). On the basis of distances between structures, this method clusters the CG predicted structures by counting the number of neighbours using a RMSD cut-off of 0.48. This value was chosen because it produced a reasonable number of clusters for a number of test cases. During the clustering, the structures with the largest number of neighbours were eliminated from the rest of the

structures in the pool of clusters. This procedure was repeated for the remaining structures in the pool. Different cut-off values were used for each cluster to obtain a reasonable number of clusters. The representative structures from the cluster analysis were converted back to an atomistic model using locally written script (*127*).

TM helix packing was evaluated by its crossing angle value, characterized as left-handed (LH), parallel or right-handed (RH), as illustrated in Figure 3-4. A locally written script developed by Amanda Buyan was used to calculate the crossing angle of a pair of interacting helices at a cut-off distance of 10 Å. At greater distances, the helices were not regarded as dimers and, thus, the frames were discarded from the crossing angle calculations. The crossing angle was calculated by drawing a line through the center of mass (COM) of the first and last four atoms of each helix; the angle between the two lines gives the crossing angle of a given TM helix dimer.

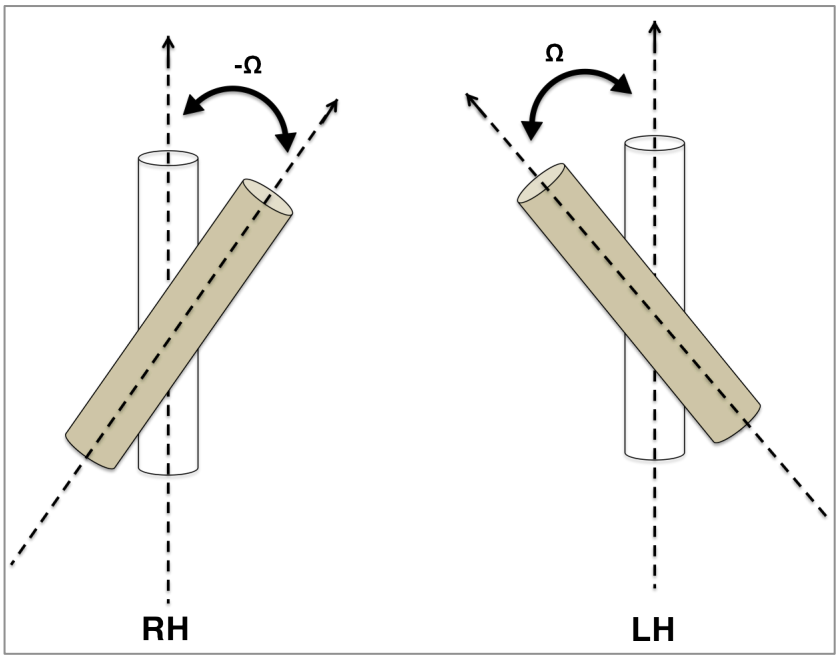


Figure 3-4. Schematic of two helices packing in right-handed (RH) and left-handed (LH) fashions; the crossing angle value (Ω) is negative for RH helices and positive for LH helices.

3.3 RESULTS

3.3.1 CG-MD simulation of EGFR

The CG-MD simulations predicted stable dimers for the EGFR TM domains, with some variability in terms of the time taken by these helices to form a dimer. Dimerization of the helices occurred mostly within the first few hundred nanoseconds, which led to the formation of a stable dimer; dissociation was not observed after dimer formation. The evolution of the inter-helix distance throughout the simulations for all 50 repeats is depicted in Figure 3-5A. To evaluate helix-helix packing, a crossing angle analysis was performed on the combined trajectories of the complete set of 50 CG-MD simulations. The crossing angle distribution for EGFR TM dimers revealed that the CG-MD predicted dimers were predominantly right handed, with a modal crossing angle of approximately -25º, as shown in Figure 3-5B.

To classify and rank the dimers structures, cluster analysis was performed on the CG-MD ensemble. The two top clusters obtained from cluster analysis corresponded to RH and LH structures, with the RH dimer being the most populated structures. Inspection of the predicted dimers obtained from the CG-MD ensemble revealed the close proximity of both the N-and C-terminal motifs at the dimer interface. However, the predicted dimers packed more closely at the N-terminal motif than the C-terminal motif, as evidenced by the mean contact distance calculated from the combined trajectories of the last 100 ns of the CG-MD simulations (Figure 3-5C). The contact matrix also revealed the key residues in close contact at the dimer interface, namely Gly625, Gly628 and Ala629 of the $G^{625}MVG^{628}A^{629}$ motif at the N-terminal region, which were oriented towards each other at the dimer interface, as seen in the NMR structure (Figure 3-6B). Meanwhile, for the less populated LH structures, Gly625 and Ala629 were oriented away from each other. Representative RH and LH structures generated from cluster analysis of the CG-MD simulations are shown in (Figure 3-6 (A, C).

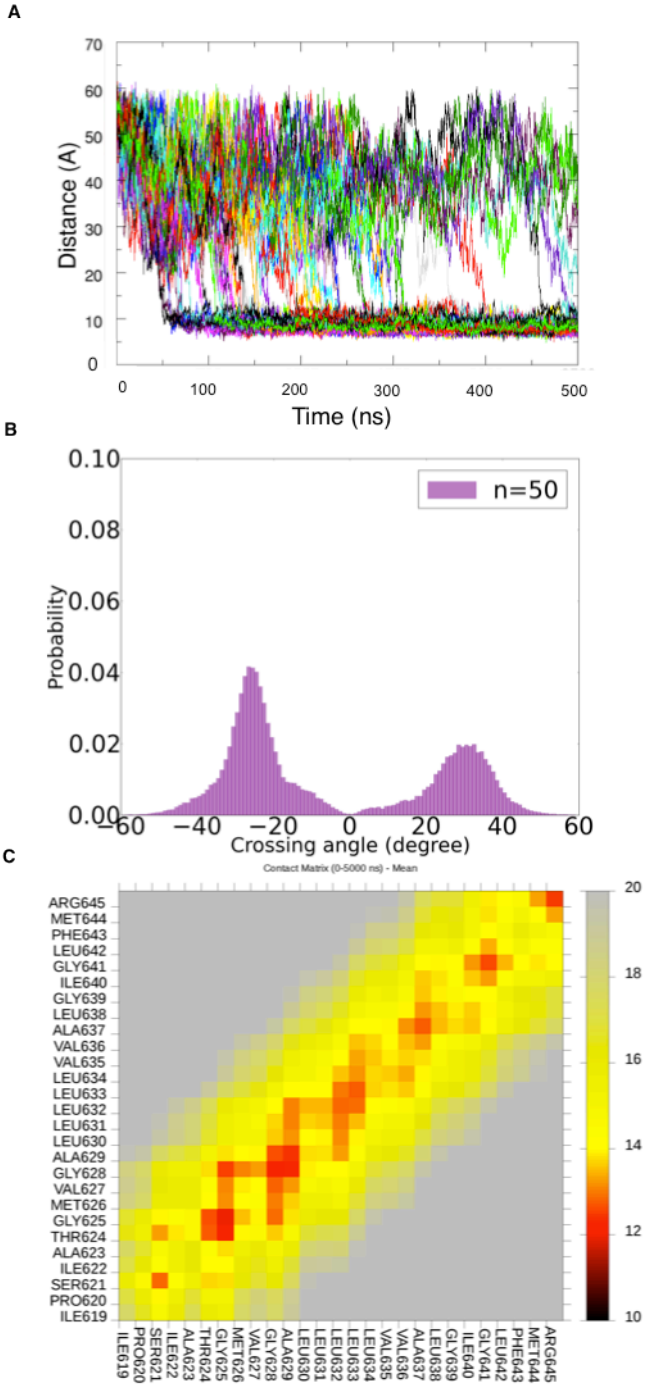


Figure 3-5. A) Evolution of interhelix distance over time and the crossing angle distribution for the combined trajectories of N=50. Note that in seven out of the 50 simulations, the helices remained separated throughout the simulations. B) The crossing angle distribution calculated for all 50 CG-MD simulations. C) The mean contact matrix for combined trajectories of the last 100 ns of CG-MD high throughput ensemble, showing a closer contact at the N-terminal motif compared with the C-terminal motif.

Overall, it is clear that the CG-MD HTP methodology used here predicted the correct packing and contact interface of EGFR, as evidenced by their agreement with the NMR structure, although the crossing angle value was smaller than that of the NMR structure (~−44° ± 3°) (*57*). This could potentially be due to the soft nature of the coarse-grained particles, as well as the presence of an α-helix dimer in the JM region, which could potentially push the two helices away from each other, resulting in a bigger crossing angle. Bilayer thickness can also influence the packing of the dimer, as the NMR structure was solved in DMPC/DHPC micelles, whilst a DPPC bilayer was used in the current simulation study.

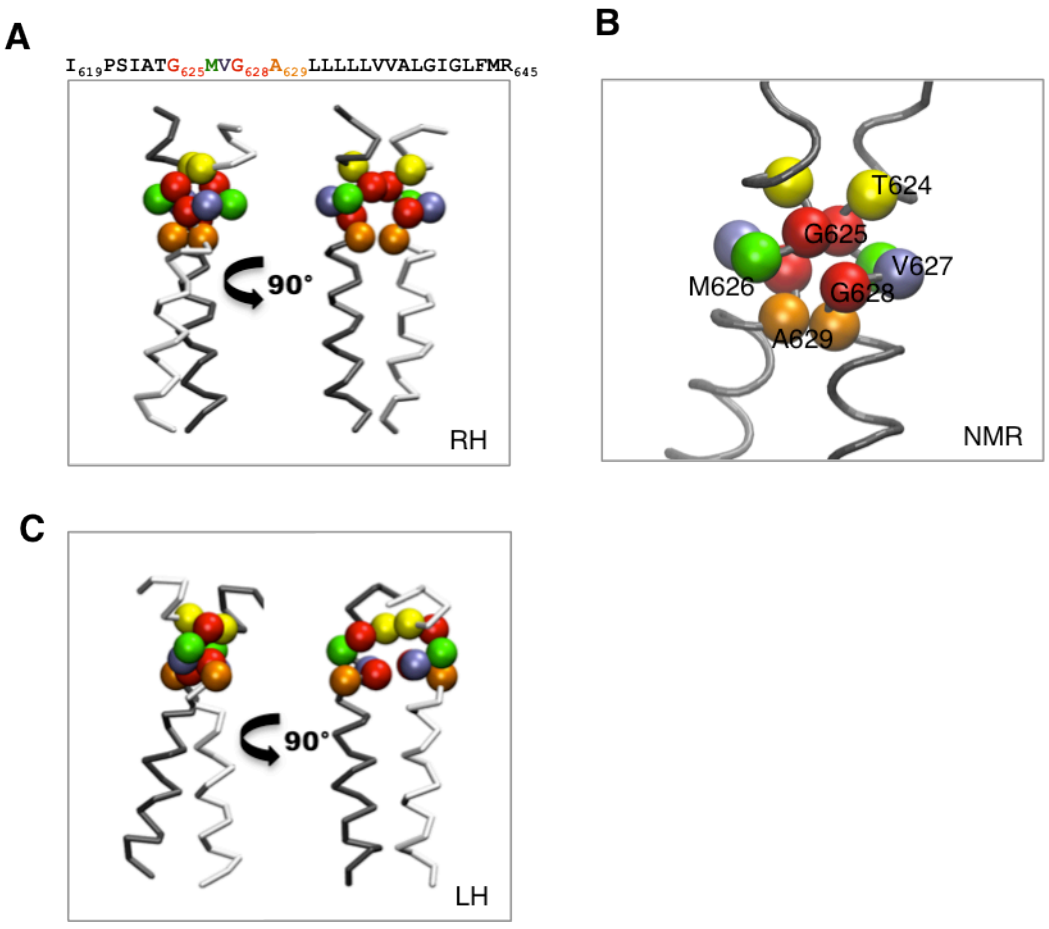


Figure 3-6. The representative or middle structures obtained from cluster analysis of the CG-MD ensemble. A) The most populated RH dimer structure. B) The NMR structure (PDB: 2M20) is shown in the center for comparison. C) The LH dimer. The C-alpha atoms of the key residues are shown as spheres (Thr = yellow, Gly = red, Met = green, Val = grey and Ala = orange).

3.3.2 CG-MD simulation of the EGFR mutant

A recent experimental study by Endres et al. demonstrated that a quadruple mutation T624I/G625I/G628I/A629I (4I mutant) led to a significant inhibition of the receptor (*57*). Earlier, another mutagenesis study by Lu et al. showed that mutations in the N-terminal motif (T624L, G625L, G628L and A629L) do not disrupt EGFR activation. To test the sensitivity of CG simulations to changes in the TM sequence motif, 25 CG-MD simulations were conducted for the more potent 4I mutant using two different force fields, BOND and MARTINI, for comparison. The association of the helices in each individual simulation was monitored based on the evolution of the inter-helix distances throughout the simulations, as shown in Figure 3-7(A, B).

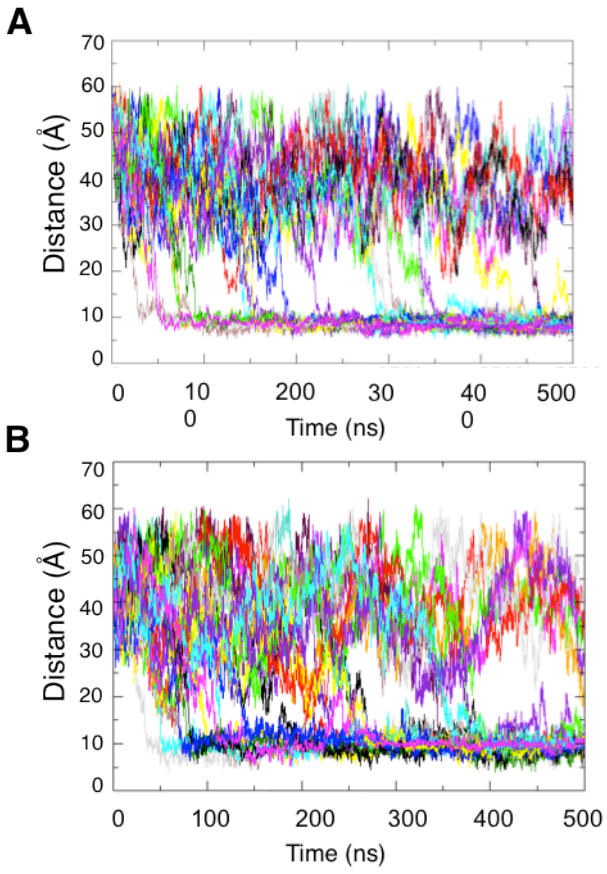


Figure 3-7. The evolution of inter-helix distance as a function of time for each repeat of both set of CG-MD simulations using BOND (A) and MARTINI force fields (B).

CG-MD simulations of the 4I mutant resulted in dimerization in both data sets, despite the mutation of four key residues at the dimer interface. This was not surprising, as a disruptive mutant, such as G83L in GpA, can still form a dimer in CG-MD simulations (*90*). In the AT simulations by Arkhipov et al., the quadruple mutation disrupted the dimer interface and the dimer dissociated (*95*). However, no dissociation was observed in the CG-MD simulations. The crossing angle distributions of the 4I mutant from both sets do not significantly differ from the WT receptor (Figure 3-8), and both force fields predicted the 4I mutant, CG-MD generated dimers to be mainly RH.

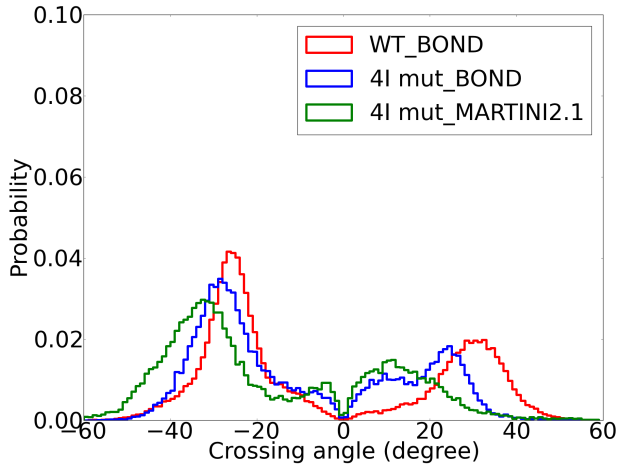


Figure 3-8. Crossing angle distributions for WT and 4I mutant EGFRs. The crossing angle distributions for the 4I mutant (N=25), performed with BOND and MARTINI force fields, are shown in blue and green, respectively. Crossing angle distributions for the WT simulations (N=50) with the BOND force field are shown in red for comparison.

The CG-MD simulations of the 4I mutant predicted dimers with crossing angle distributions similar to that of the WT EGFR, which suggests that the 4I mutant did not completely alter the crossing angle of the dimer. Thus, it is useful to investigate whether the mutation altered the orientation of key residues at the dimer interface. To investigate the 4I mutant helix-helix packing interface, the top two clusters obtained from the cluster analysis of the 4I mutant CG-MD ensemble were examined. The most frequently observed RH dimers of the 4I mutant were different from the WT RH dimer, as the G625I residues are directed towards each other and the G628I residues are oriented away from each other (Figure 3-9B). This is in contrast to the WT packing, where both G625

and G628 residues are facing each other and making close contact at the dimer interface (Figure 3-9A). Thus, the helix-packing interface is perturbed relative to the WT RH dimer packing interface. Meanwhile, for the less populated LH structure the G625I is pointing away and G628I is pointing toward each other as depicted in Figure 3-9C.

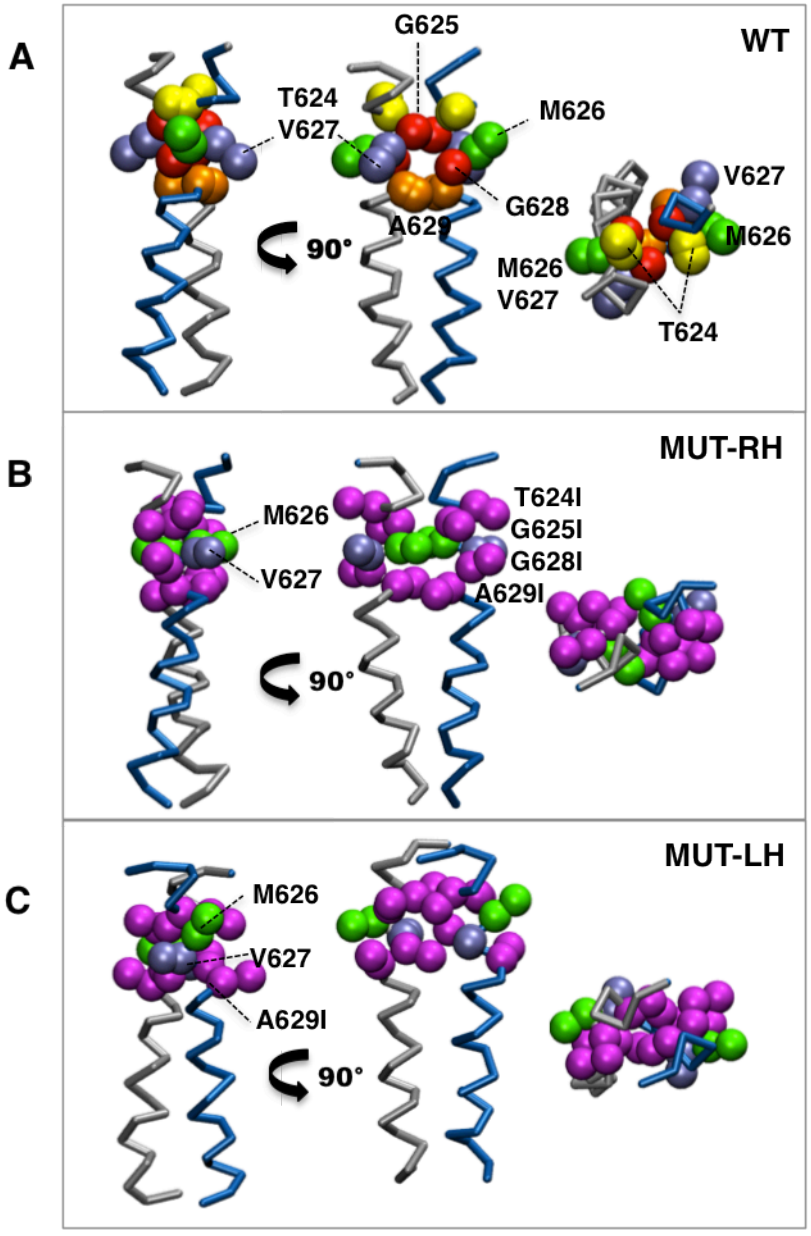


Figure 3-9. Representative structures obtained from cluster analysis of the WT and 4I mutant CG-MD simulations. A) The predicted RH dimer for the WT receptor, with the Thr624 and GMVGA motif residues shown as spheres (Thr = yellow, Gly = red, Met = green, Val = grey and Ala = orange). B) The most populated RH dimer predicted by the CG-MD simulations for the 4I mutant. C) A representative structure of the predicted LH dimer of the 4I mutant. The T624I/G625I/G628I/A629I mutations are shown as purple spheres.

3.4 CONVERGENCE ANALYSIS

A convergence analysis was performed on MuSK and EGFR CG-MD ensembles consisting of 77 and 50 dimerization simulations, respectively. The convergence was analysed in terms of the helix crossing angle distributions and the RH/LH ratio. For EGFR, the ensemble of 50 dimerization simulations was analysed in terms of the helix crossing angle distributions, as these allow for ready identification of the frequency of formation of right-handed (i.e., negative crossing angle) helix dimers. The full ensemble of 50 simulations showed a strong preference for right handed packing, in agreement with the NMR structures. Therefore, this ensemble was used to explore the convergence of the simulations to this characteristic structure. The first pass at such an analysis consisted of randomly selecting a single subset of 5, 10, or 20 simulations from the full ensemble of 50 simulations and analysing the crossing angle distributions for each subset. This suggested that running smaller ensembles could result in a biased view of the helix-packing model. For example, the randomly selected frequency of right handed crossing for the single $N = 10$ subset was slightly less than the frequency of left handed packing of the helices (Figure 3-10A). The RH/LH ratio was obtained by dividing the total number of structures with RH crossing angles by the total number of structures with LH crossing angles. This calculation was repeated for each of the subsets and the complete set. The convergence analysis was repeated for MuSK. Based on this analysis, one can judge that reasonable convergence was attained after a sub-ensemble size of ca. 20 or above, for both EGFR and MuSK. This is a valuable indication of what sizes of ensembles to generate if HTP methodologies are used to explore different families of related TM helix dimers via CG simulations (Figure 3-10). This conclusion was supported by a proper statistical analysis of the EGFR ensemble using the Jackknife method, which was performed by Amanda Buyan and is discussed in (*152*).

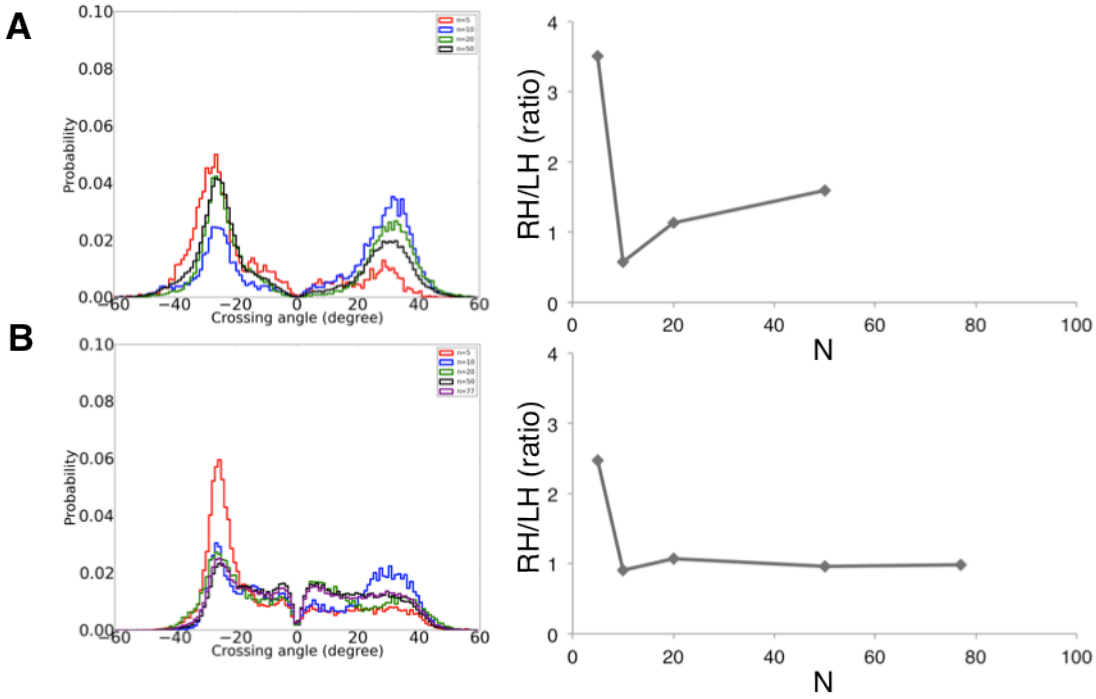


Figure 3-10. Convergence analysis of EGFR ensembles. Crossing angle distributions for the complete set and subsets of N = 5,N = 10, N = 20, N = 50 drawn from the complete set are shown. The crossing angle distributions of EGFR (A) and MuSK (B) CG-MD simulations reached convergence at N >= 20 (green line). On the right are the RH/LH ratio plots, as a function of N, for each of the test cases, which again suggest a convergence at N >= 20.

3.5 MULTISCALE MD SIMULATIONS OF MUSK

In the CG-MD simulations, dimers spontaneously formed mostly within the first hundreds of nanoseconds and remained associated for the duration of the simulations. This suggests that stable dimers were formed, as no dissociation was observed (Figure 3-11A). Based on the crossing angle distribution calculated for the MuSK CG-MD ensemble of 77 trajectories, it appears that the MuSK dimer sampled a wider range of crossing angles compared with EGFR, with a peak at ~ -25º, which corresponded to RH packing (Figure 3-11B). The crossing angles, calculated every 200 ps, were used to calculate the percentage of different crossing angles sampled by the MuSK dimer in the CG-MD simulations. It was found that MuSK adopted a RH conformation in 49% of the frames while the remaining 51% corresponded to a LH conformation. In addition, 27% of the frames corresponded to MuSK dimers that had crossing angle values greater than 20°, and 24% of the frames belonged to dimers with crossing angle less than -20°. Thus, 49% of the dimers had crossing angle values between -20° and 20°, which corresponded

to parallel conformations (Figure 3-11B). The fact that the four packing modes in the distribution were approximately equally populated suggests that the MuSK TM dimer is loosely packed and can adopt multiple conformations. Cluster analyses performed on the CG-MD ensemble, however, suggest that the most populated structures were dimers with crossing angles of less than 20º, with both helices oriented almost parallel to each other in the bilayer. Snapshots of the representative structures extracted from the trajectories based on cluster analysis are displayed in Figure 3-11C.

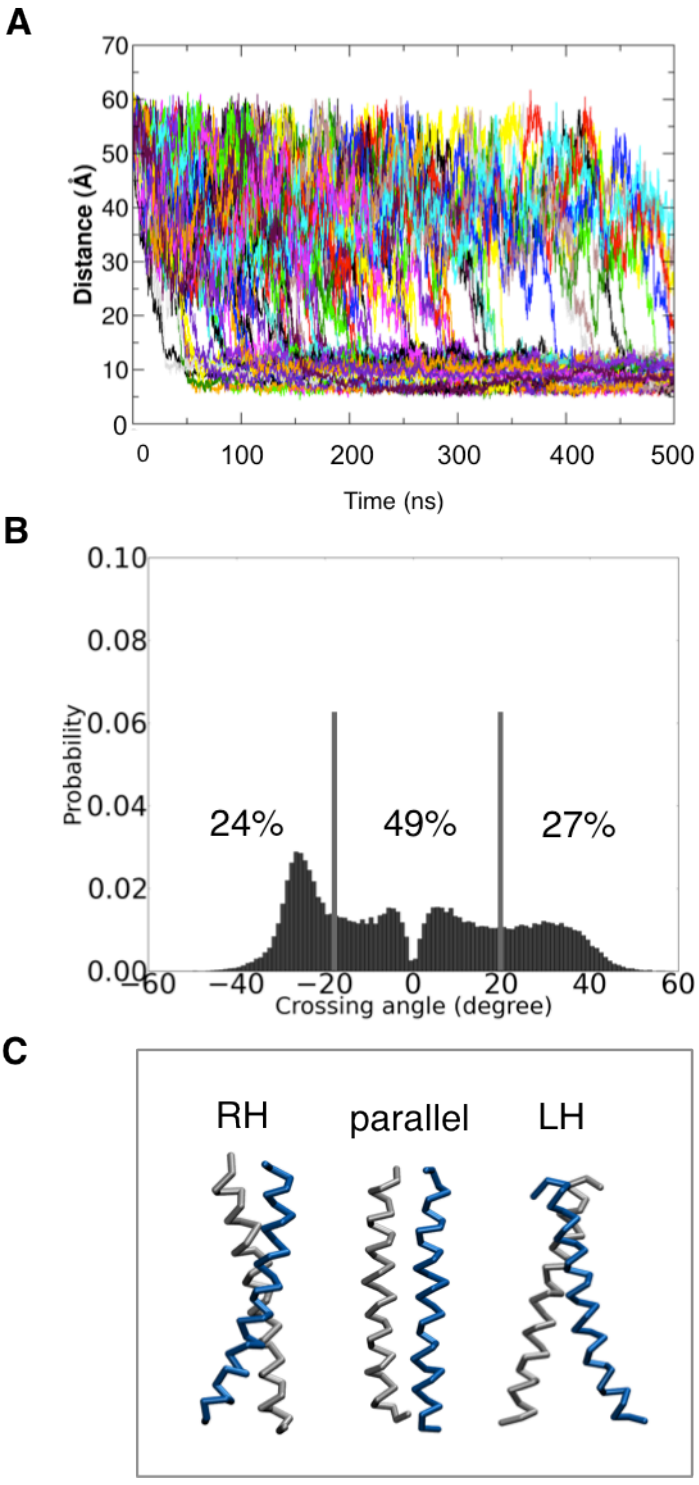


Figure 3-11. A) The evolution of interhelix distance during the CG-MD simulations (N=77). B) The crossing angle distribution of the CG-MD simulations. The histogram shows that the crossing angle values for the MuSK dimer are widely distributed. C) The representative structures obtained from cluster analysis of the MuSK CG-MD ensemble.

Visual inspection of the representative parallel structures extracted from CG-MD trajectories demonstrated that T495, S499, S502, I506, L510, T513 and T517 are positioned at the dimer interface. Meanwhile, the dimer interface is formed mainly by the residues at the N-terminal end of the TM domain, T495, S499, S502 and S503 for the RH dimer, and S499, S502 and S503 for the LH dimer. For a more detailed investigation of the helix-helix interaction predicted in the CG-MD simulation, three structures representing parallel, LH and RH dimers from the CG-MD simulations were selected for refinement by all-atom simulations.

3.5.1 Conversion to atomistic and production run

To investigate in more detail the packing of the MuSK TM dimers suggested by the CG-MD simulations, the three representative structures from the CG-MD simulations were converted to AT representations using a fragment-based approach (*125*) (details of the method are given in subsection 2.5). The initial atomistic systems were energy minimized and equilibrated for 1 ns, with the protein Cα restrained, followed by unrestrained production runs for 100 ns. The Cα RMSDs and RMSFs of the three different dimers conformation during the AT simulations were calculated to assess the stability of the different conformations. A stable RMSD was observed in two of the three repeats of RH simulations relative to the parallel and LH simulations. This suggests that the RH dimer is relatively more stable than the LH and parallel dimers, though the average RMSD of the LH conformation is relatively lower than the average RMSDs of the RH and parallel conformations. The RMSD values for the LH dimer increased throughout the simulations (Figure 3-12A), in contrast to the RH dimer, which began to plateau after 60 ns.

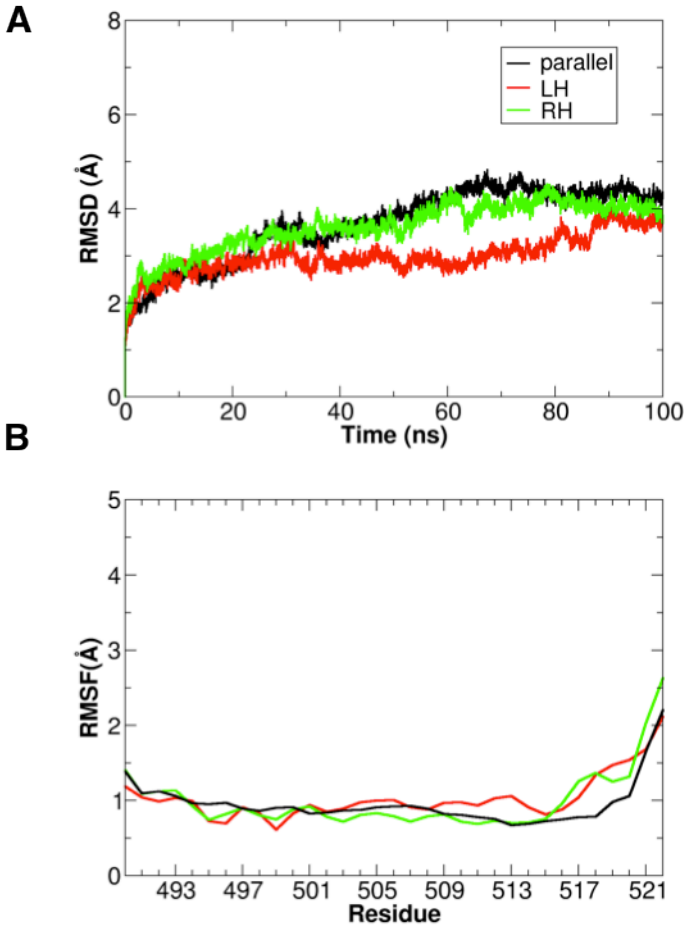


Figure 3-12. A) The C∝ RMSDs for parallel, RH and LH MuSK structures averaged over three repeats. B) The RMSFs of the C∝ atoms showing the lowest values for the Thr495 and Ser499 residues for both the RH and LH dimers. The RH dimers have a relatively low C∝ RMSF compared with the parallel and LH dimers.

It was observed that the tilt angle of all dimer structures increased during the AT simulations. The average tilt of all three conformations increased during the simulations, from 10° to approximately 25° and 35° for the parallel and RH structure, respectively. Meanwhile, the tilt angle for the LH structure increased slightly, from 20° to 30°. The tilting of the 33-residue helix was expected to minimize the hydrophobic mismatch between the helix and lipid bilayer (Figure 3-13A). Snapshots of the initial and final structures for the different conformations are depicted in Figure 3-14B.

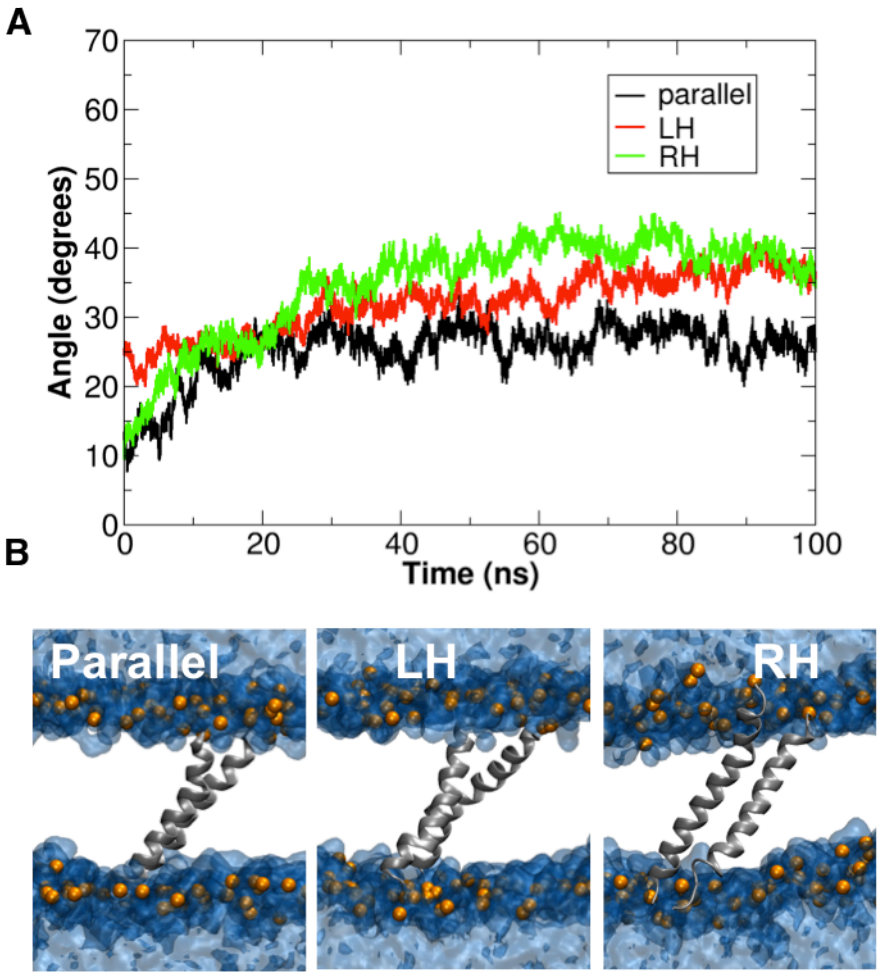


Figure 3-13. A) Tilt angle of parallel, LH and RH dimers of MuSK averaged over three repeated AT-MD simulations. The tilt angles of the RH and parallel dimer increase rapidly during the first 20 ns and become stable after 40 ns. B) Snapshots of one of the repeats for the parallel, LH and RH dimers at t = 100 ns during the AT-MD simulations. The lipid phosphate atoms are represented as orange spheres and water molecules are shown in surface representation in blue.

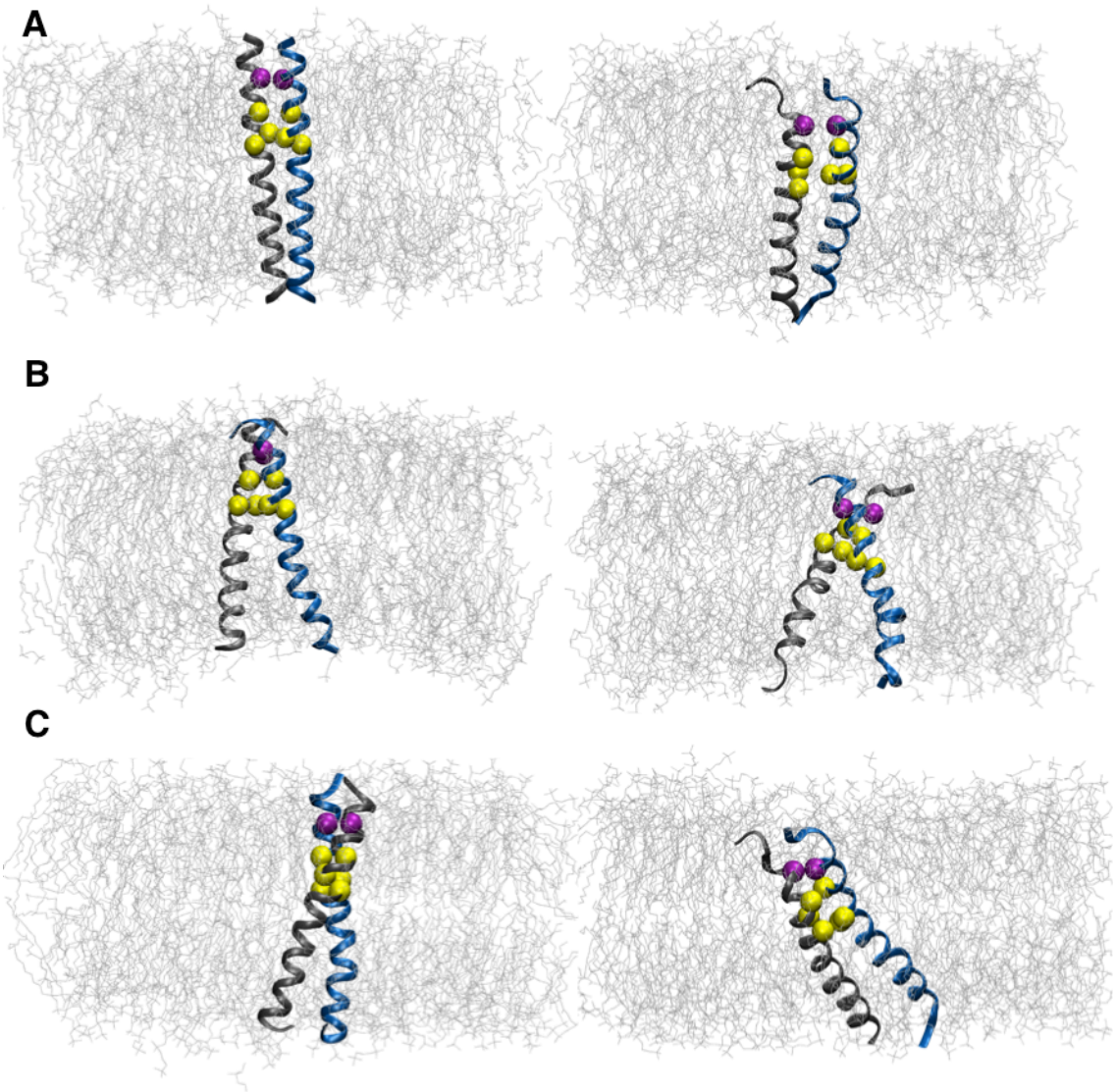


Figure 3-14. Cartoon representation of the MuSK dimer in AT-MD simulation. Snapshots of parallel (A), LH (B) and RH (C) conformations taken from one of the three repeats of the AT simulations. On the left is a snapshot at t = 0 ns, and a snapshot at t = 100 ns is shown on the right. The C∝ atoms of Ser and Thr residues are represented as yellow and purple spheres, respectively. The DPPC lipids are shown as lines in grey.

To further analyse the MuSK dimerization interface, a hydrogen bond analysis was performed to determine the number of interactions between the MuSK dimer subunits during the AT-MD simulations. A comparison of the hydrogen bonds formed between the small-xxx-small (Ser-xxx-Ser) motif of MuSK among these three different conformations revealed that more hydrogen bonds were formed between the residues of

the Ser-xxx-Ser motif of the RH dimer than for those of the parallel or LH dimers. Two hydrogen bonds were formed by the Ser-xxx-Ser motif in the parallel dimer, while up to four and six hydrogen bonds were formed in the LH and RH dimers, respectively (Figure 3-15).

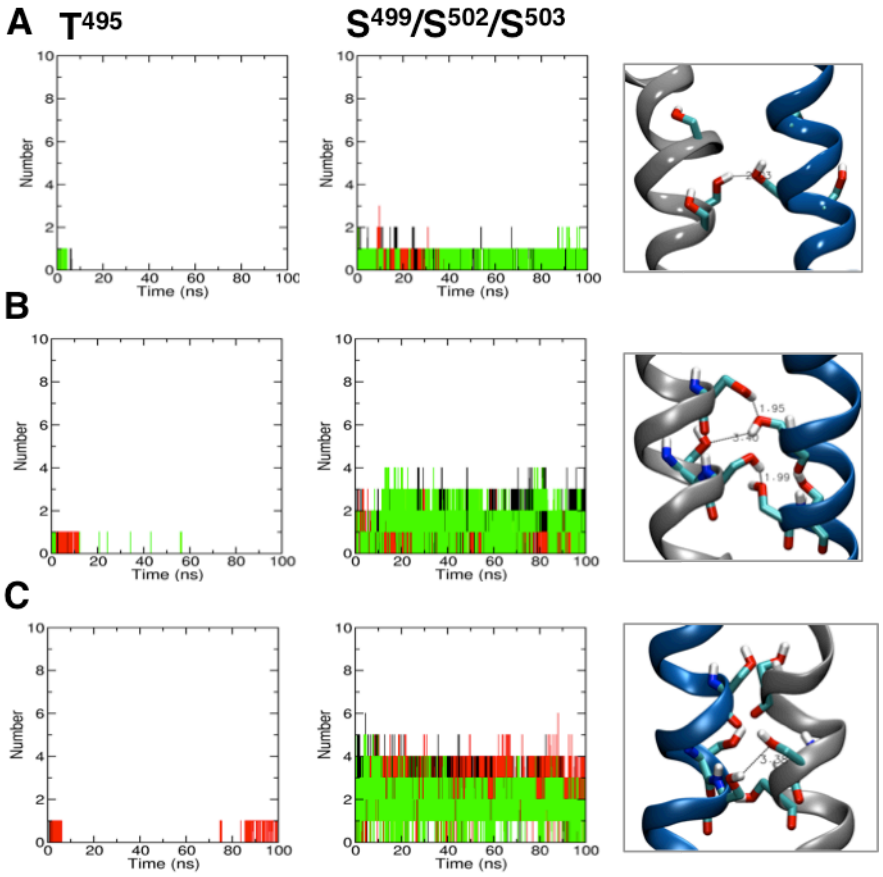


Figure 3-15. Hydrogen bonding between Thr495 and the SxxxS motif throughout the 100-ns simulations. Individual simulations are shown in black, red and green lines, which represent three independent repeats for each of the different starting conformations in A (parallel), B (LH), and C (RH).

Overall, the RH dimer appears to be stabilized by more hydrogen bonds formed by the Ser-xxx-Ser motif. Relatively more residues, namely T495, I498, S499, S502, S503, I506, I507, L510, and L511, are in close proximity at the dimer interface in the RH dimer, thus allowing close contact and hydrogen bonding between the dimer subunit compared with the LH dimers (Figure 3-16). S499, S502 and S503 of the RH dimer participate in a hydrogen bond network at the dimer interface. For the LH dimer, T495, S499, S502 and I506 are the interfacial residues in both MuSK dimers; however, Ser503 is oriented slightly away from the dimer interface (see Figure 3-15B) and not oriented at

the interface as observed in the WT. Meanwhile, for the parallel dimer, despite having more residues at the interface, only Ser502 is involved in stable hydrogen bonding throughout the simulation, as shown earlier in Figure 3-15A. Another stable hydrogen bond was observed between T513 at the C-terminal end. Thus, the parallel dimer is relatively less tightly packed at the Ser-xxx-Ser motif compared with the RH or LH dimers. The main inter-helix contact averaged from the last 50 ns trajectory from three repeats for each conformation were gathered and merged, as depicted in Figure 3-17. As can be seen from the contact matrix, T495 and residues in the Ser-xxx-Ser motif are in close contact at the dimer interface, particularly in the RH dimer. For the parallel dimer, the contact matrix revealed that residues at the C-terminus of the parallel dimer are also in close contact, possibly playing a role in stabilizing the dimer in a parallel conformation.

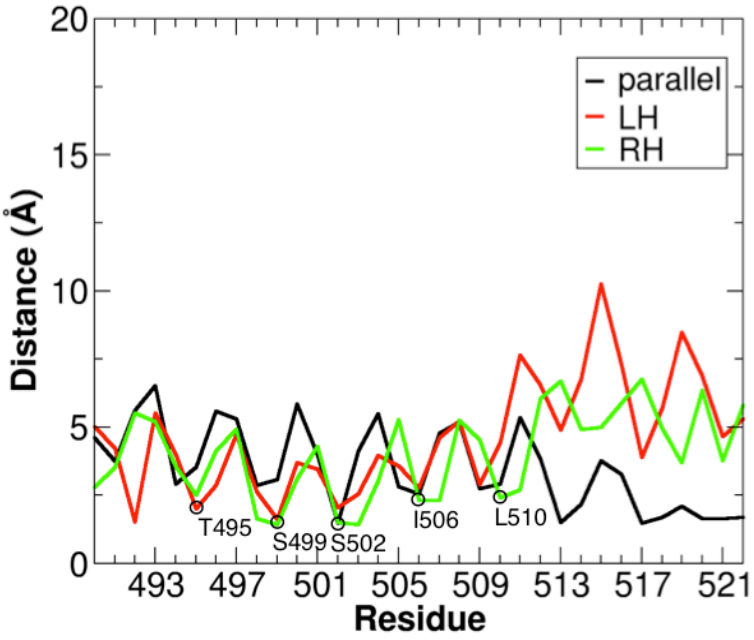


Figure 3-16. The residues at the dimer contact interface as observed in AT simulations. The minimum distance between residues averaged over the last 50 ns of the three repeats of each type of AT simulation.

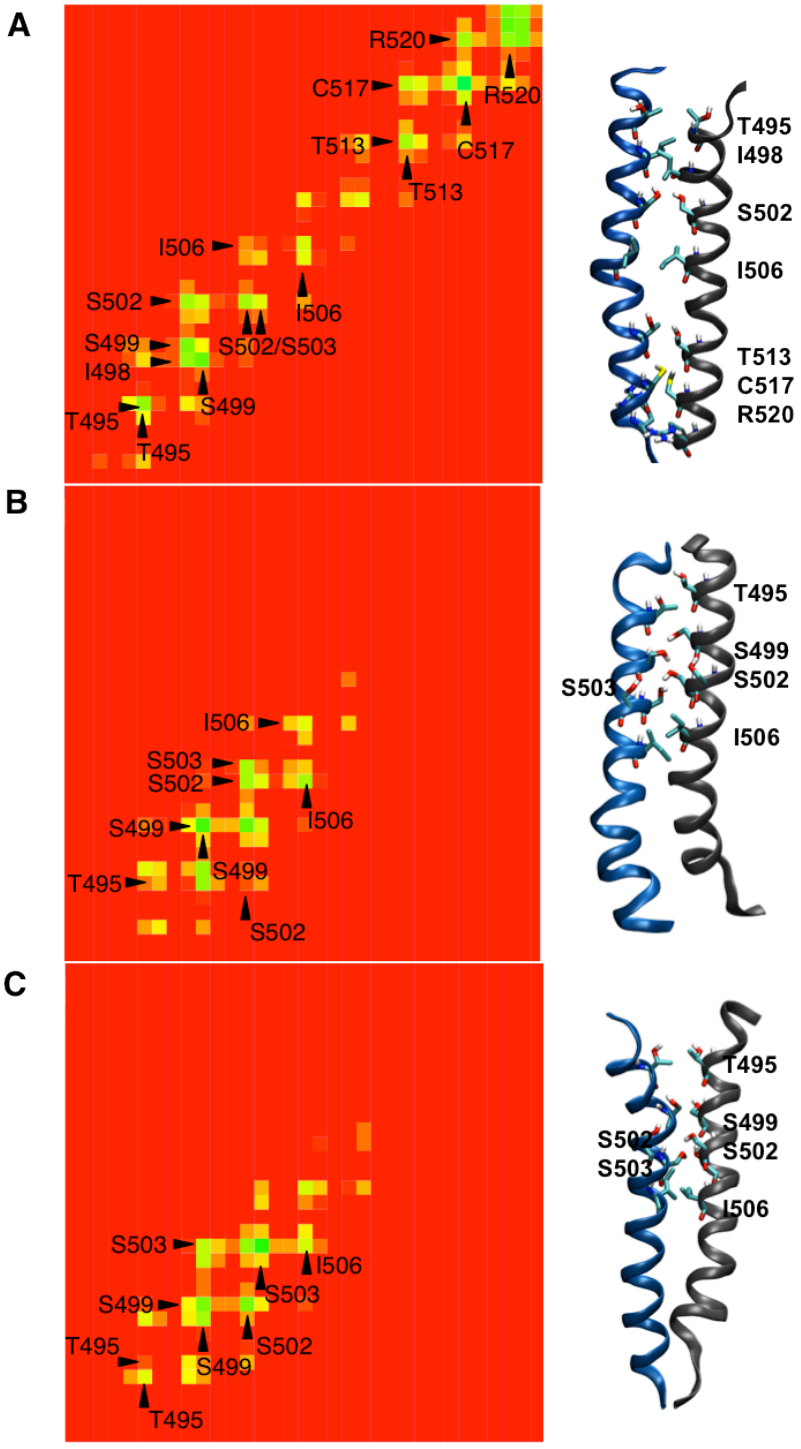


Figure 3-17. The mean smallest contact between C∝ atoms averaged over the last 50 ns (from all three simulations) of each type of conformation. The residues involved at the dimer interface of the parallel, LH and RH dimers are shown in A, B and C, respectively.

3.6 SIMULATIONS OF MUSK MUTANTS

The MuSK WT CG and AT simulations suggested that T495, S499, SS502 and S503 are important in MuSK dimer packing. Because there is no structural data on the mutations of the MuSK TM domain, CG-MD is a useful tool to explore the possible effect of mutation of residues thought to be important for the dimer contact interface. Three mutants were simulated here; the sequences used are shown in Figure 3-18.

```
MuSK WT  490      PTYSMTVIISIMSSFAIFVLLTITTLYCCRRRK 522
4A mutant         PTYSMAVIIAIMAAFAIFVLLTITTLYCCRRRK
3A mutant         PTYSMTVIIAIMAAFAIFVLLTITTLYCCRRRK
T495A mutant      PTYSMAVIISIMSSFAIFVLLTITTLYCCRRRK
```

Figure 3-18. Comparison of the TM sequences of WT MuSK and mutants used in this study. Thr495, Ser499, Ser502 and Ser503 were mutated to Ala, as highlighted in red.

It was observed in the CG-MD simulations that when all four residues were mutated to Ala (4A), the dimer was predicted to mainly adopt a RH conformation. The crossing angle distributions of all the mutants were analysed and compared with those of the WT TM, as shown in Figure 3-19.

Comparison of the crossing angle distributions revealed that the 3A mutant was very similar to the WT (Figure 3-19C). However, for the single Thr495 to Ala mutation, the predicted dimers appeared to favour either RH or LH, rather than parallel, conformations, as judged by the bimodal crossing angle distribution. The crossing angle distribution shifted toward more positive values due to the T495A mutation, which may disrupt WT-like parallel packing (Figure 3-19B). Taken together, these results suggest that Thr495 may play an important role in modulating helix packing in the MuSK dimer.

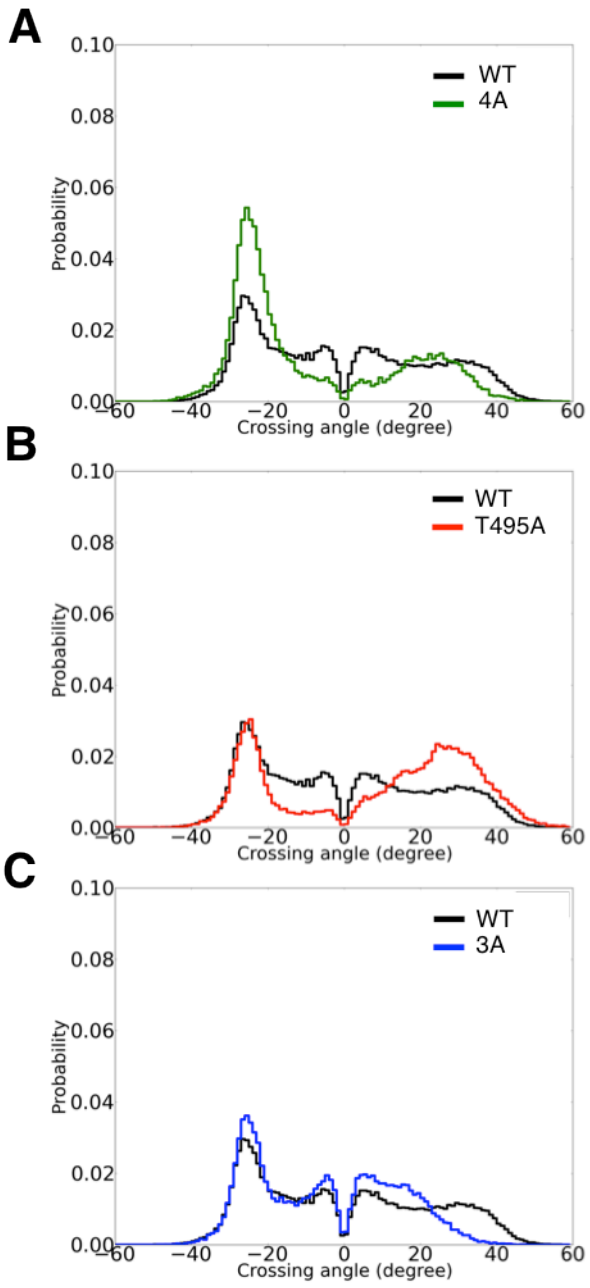


Figure 3-19. The crossing angle distribution for the MuSK mutants compared with the WT TM, as predicted by CG-MD simulations. The crossing angle distribution for MuSK 4A, Thr495Ala, and 3A mutants are shown in A, B, and C, respectively. The crossing angle of the WT is shown as a black line.

3.7 CONCLUSIONS AND FUTURE STUDIES

In this study, the CG-MD simulations predicted the correct packing of EGFR, with the dimer interface formed by the N-terminal motif, as observed in the NMR structure. Meanwhile, the CG-MD simulations of the EGFR 4I mutant predicted that the EGFR TM helices still form dimers, despite the quadruple mutation of its N-terminal motif to Ile. This is not entirely unexpected, as the most disruptive GpA mutant, (G83L), still forms a dimer in CG-MD simulations. In the MARTINI CG model, Leu and Ile are very similar, thus, it is not surprising that the helices can still form a dimer. The CG-MD simulations suggest that the 4I mutant adopts RH packing, with a similar crossing angle as the WT TM, but with different residues at the dimer interface, thereby indicating that the dimer structure is perturbed relative to that of the WT TM.

A multiscale approach was employed to predict the dimerization and the association of MuSK TM helices in DPPC lipid bilayers. In the CG-MD simulations, the MuSK dimer was observed to adopt a flexible and looser packing interface, as it sampled various possible conformations, from right handed to parallel to left handed, with crossing angles ranging from -40 to 40°, with a peak of approximately -22° for the RH conformation, which suggests that MuSK has a slight preference for RH packing. Additional AT simulations of the CG-predicted dimers suggest that T495, S499xxS503S405, I506, L510 and T513 form the contact interface of MuSK. The subsequent AT simulations suggest that the Ser-xxx-Ser motif (S499xxS503S405) may play an important role in MuSK dimerization, as this motif is involved in hydrogen bonds networks at the contact interface. In addition, T495 may play an important role in promoting some degree of flexibility in MuSK packing, as this residue underwent transient hydrogen bonding in the simulations. It is possible that MuSK has a looser packing interface and can adopt various conformations mediated by the interactions between T495 on the first helix and the T495 on the second helix. This is not unique to MuSK, as such loose dimer packing has previously been observed in other RTKs (*153*), such as the FGFR3 receptor (*154*). It was suggested that such flexibility is important for receptor signalling and allows for better adaptation to heterogeneous membrane environments (*154*).

From the CG-MD simulations, it was observed that a single T495A mutation did not significantly disrupt the WT packing mode or change the dimer interface, although it reduced the populations of parallel dimers. Meanwhile, the S499A/S503A/S503A (3A)

mutant displayed a crossing distribution that closely resembles that of the WT TM domain. This suggests that T495 may have the ability to modulate MuSK TM helix packing. It is useful to compare the results obtained from this study with experimental data once it becomes available. However, to reliably predict the dimerization of TM proteins without known structures using the CG-MD approach, a large number of simulations are required to obtain a better sampling of dimer conformations. Previously, CG simulations were performed in small sets of manual simulations (ca. up to 10 simulations of several microseconds), or up to 100 or more CG-MD simulations using HTP methodology, which was employed in this study. In this study, a simple convergence analysis conducted on the three data sets suggests that the CG-MD simulations began to converge after 20 simulations, and fully converged after 50 simulations.

It is also important to note that not only mutations of the important residues can affect dimerization; the bilayer composition and lipid environment can also affect the dimerization strength and energetics of the TM helix association, which is beyond the scope of this study. It would be interesting to further investigate the free energy association and the energetics of the various conformations adopted by the MuSK WT TM domain, as well as the effect of mutations on the dimerization and stability of the MuSK TM domain. Additionally, to obtain a better understanding of MuSK TM helix association, it would be interesting to gain some insight into the behaviour of the helices by varying the thickness of the bilayer, using a more physiologically relevant bilayer, and extending this study to the micellar environment.

CHAPTER 4

MD simulations of the EGFR TM-JM interactions with lipids

4.1 INTRODUCTION

In this chapter, the simulations were performed using a more biologically relevant system, including the juxtamembrane (JM) domain embedded in lipid bilayers containing anionic phospholipids and ganglioside. As described earlier in Chapter 1, the JM domain plays a crucial role in the activation of EGFR (*57*), although the mechanistic details remain elusive. The JM domain begins at the C-terminus of the TM domain, extending from residues 645 to 682, referred to as JM-A (645–663) and JM-B (664–682) (*54*). The JM-A and JM-B domains are critical for the dimerization and activation of EGFR, as a significant reduction in activity (~10-fold) was observed when JM-A was removed, whilst an ~70 fold increase in EGFR activity was observed when both JM-A and JM-B are present. It has been shown that the receptor loses its activity completely when JM-AB is removed (*54*). The JM-A segment has three clusters of positively charged residues (Arg and Lys) that have been suggested to bind to negatively charged lipids (*78*). These three clusters of positively charged residues are R645/R646/R647, R651/K652/R653 and R656/R657 (see Figure 4-1). The central portion of the JM-A segment also has hydrophobic residues in an i, i+3, i+4 pattern in an alpha-helical conformation within the L_{655}RRLL_{659} motif (as previously revealed by a solution NMR study (*54*) and the recent NMR structure (*57*)). The interface of the JM-A dimer is composed of residues L655, L658 and L659, and JM-A association can occur outside membrane, as judged by the NMR-derived water accessibility of the JM residues (*57*). The NMR structure of the TM-JM domain revealed a specific contact interface at the N-terminal motif of the TM domain that is compatible with the proper orientation of the intracellular JM domain (*57*). Thus, both the TM and JM domains play a critical role in facilitating and stabilizing the asymmetric kinase dimer formation crucial for EGFR activation (*54*, *55*).

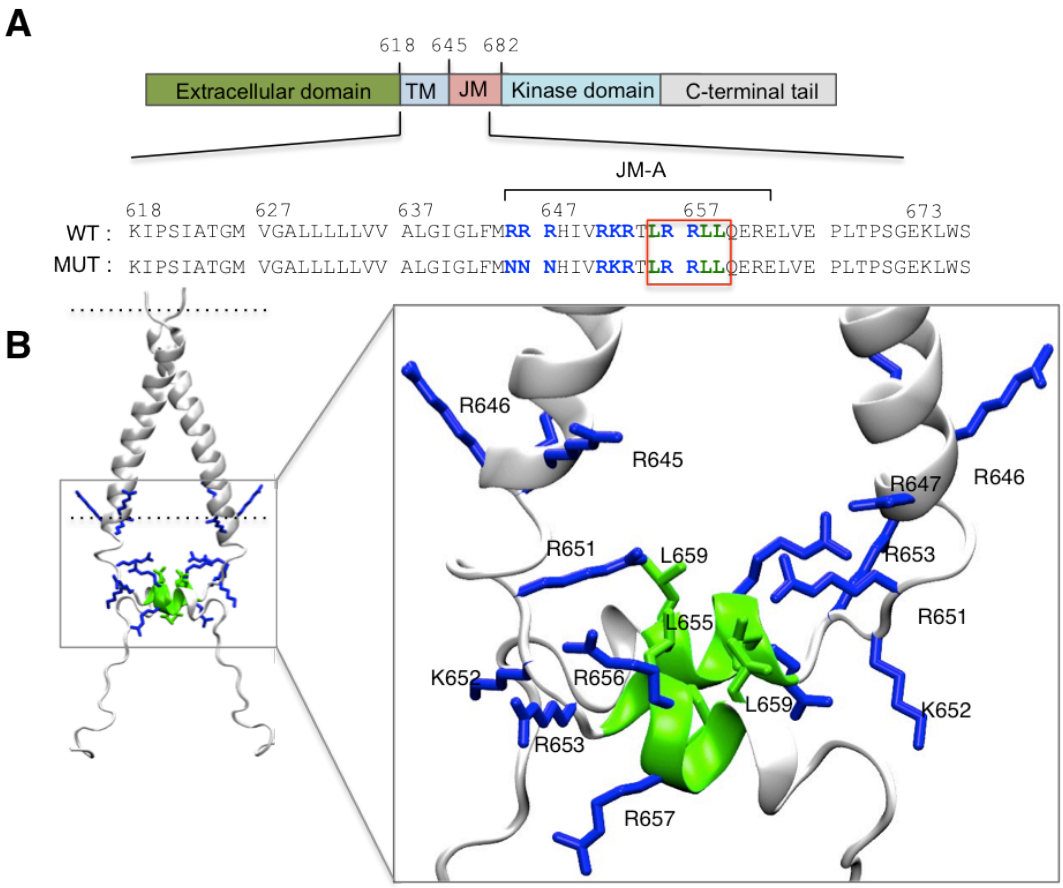


Figure 4-1 . A) Schematic diagram of the TM-JM domain. The amino acid sequence of the TM-JM domain. JM-A and JM-B are shown in red and black boxes, respectively. B) The NMR structure of the TM-JM domain (PDB : 2M20) is shown, with clusters of negatively charged residues highlighted. The Arg residues are shown as stick representations in blue and the Leu residues within the LRRLL motif are shown in green. The LRRLL motif is highlighted in green.

4.1.1 Regulation of EGFR by PIP_2

PIP_2, which is the most abundant PI derivative, constitutes ~1% of the phospholipids in the plasma membrane (*155*). The JM domain of EGFR has been implicated to interact with PIP_2 (*60*) and also Ca^{2+}/calmodulin binding (*156*). In a recent study, Michailidis et al. reported that EGFR is regulated by PIP_2 via the positively charged residue clusters in the JM domain. They also demonstrated that mutation of the first Arg/Lys to Asn abolished PIP_2 binding and reduced phosphorylation of the receptor (*78*). A recent study by Matsushita et al. of the closely related Neu TM-JM domain using solid-state NMR and fluorescence microscopy demonstrated that the TM-JM domain formed tight interactions with anionic lipids and the JM domain was released from the membrane upon ligand-induced activation (*30*). Additionally, the JM interaction with the membrane

has been investigated using long time scale AT-MD in a POPC/POPS bilayer (*95*). Although the TM-JM interaction with anionic lipids is important, the behaviour of the TM-JM domain in a PIP_2-containing bilayer has not been extensively studied. In this chapter, the dynamic behaviour of the TM-JM domain was investigated in PIP_2-containing and PIP_2-depleted membranes.

4.1.2 Regulation of EGFR by GM3

GM3 is a ganglioside derivative composed of an oligosaccharide head group structure containing one sialic acid residue. GM3 is known to inhibit the functions of fibroblast growth factor receptor (FGFR) (*157*), platelet-derived growth factor receptor (PDGFR) (*158*), the insulin receptor (*159*), urokinase-like plasminogen activator (*160*), integrin and EGFR (*161*). It has been shown that GM3 can affect kinase activity (*162*), and both dimeric/monomeric forms of EGFR in A431 cells were inhibited only by GM3 (*163*). It was also demonstrated that an N-linked glycan, with GlcNac termini, linked to interacted with GM3, thereby causing the inhibition (*164*). More recently, reconstitution experiments by Coskun et al. provided the specificity of EGFR inhibition, i.e., the inhibitory action occurred with GM3, and not other ganglioside species, and the inhibition of EGFR by GM3 only occurred in a phase-separated, Ld/Lo membrane (*77*). The inhibition of EGFR activity by GM3 can be released by removing neuraminic acid in the GM3 head group or by mutating the membrane-proximal residue K618 to a Gly. This study also suggests the potential regulatory role of GM3 on the allosteric structural transition from an inactive to active state in response to ligand binding (*77*). Apart from the inhibition of EGFR phosphorylation, GM3 was also reported to influence the ErbB receptor (ErbB2) distribution in lipid raft-containing membranes (*165*).

Here, a combination of CG-MD and AT simulations was used to investigate the dynamic behaviour of the TM and TM-JM domains in bilayers of different lipid compositions. This chapter has been divided into two parts. The first part will focus on the simulation of the TM-JM domain in PIP_2-containing and PIP_2-depleted membranes. The second part of the chapter will discuss the CG-MD simulations performed with the aim of investigating the behaviour of the TM dimer in GM3-containing bilayers.

4.2 SYSTEM SETUP

The CG-MD simulations were performed using the NMR structure of the EGFR TM-JM domain (PDB:2M20) (*166*). The NMR structure was converted to a CG representation suitable for the MARTINI force field (v2.1) (*116*). The NMR structure was simulated with ENM restraints to preserve its secondary structure. The simulations were performed at 300 K for a pure POPC bilayer and at 310 K for POPC:POPS and POPC:POPS:PIP_2 ; both these temperatures are well above the phase transition temperature of these lipids (Figure 4-2) bilayers. An initial 100 ns CG-MD simulation was performed to self-assemble the phospholipids (POPC) around the dimeric protein. POPS and PIP_2 molecules were subsequently incorporated into the equilibrated POPC bilayer by exchanging them for POPC molecules; this procedure was implemented using an in-house lipid exchange script (*167*). The same protocol was used to set the GM3-containing bilayers. After lipid exchange, the system was energy minimized, followed by a 2–5 ns equilibration with position restraints on the protein and 3 μs production runs. Prior to performing unrestrained production run simulations, the systems were energy minimized using the steepest descent method for 5,000 steps. The systems consist of 247 lipids in a simulation box of 100 x 100 x 120 $Å^3$. To further investigate the TM-JM interactions with PIP_2, three repeats of a 100 ns AT simulation were performed using snapshots obtained from POPC:POPS:PIP_2 CG-MD simulations. All the CG-MD and AT simulations were performed using GROMACS 4.5.4.

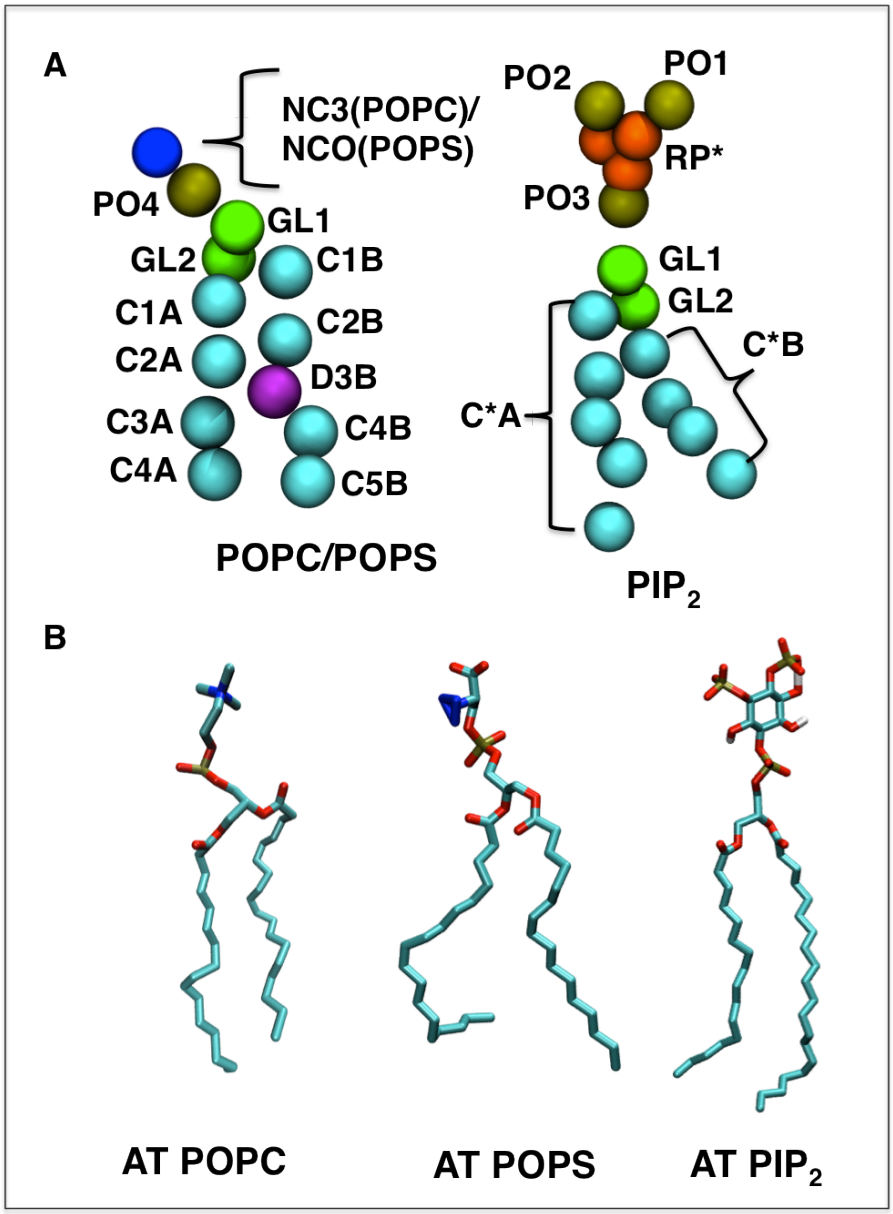


Figure 4-2. A) The CG structures of POPC, POPS and PIP_2 used in the simulations. B) The atomistic structures of POPC, POPS and PIP_2.

4.3 CG-MD SIMULATIONS OF TM-JM IN PIP2-CONTAINING BILAYERS

To explore the interactions between PIP_2 and the JM segment, the NMR structure of the EGFR TM-JM domain was simulated in a POPC:POPS:PIP_2 bilayer, as summarized in Table 4-1.

Table 4-1. Multiscale simulations of the TM-JM domains (WT and mutant) in PIP_2-containing and PIP_2-depleted membranes.

Simulation	System	Duration
CG-MD		
TM-JM (WT)	POPC	3 x 3 μs
TM-JM (WT)	POPC, POPS	3 x 3 μs
TM-JM (WT)	POPC, POPS, PIP_2	3 x 5 μs
TM-JM_mut[a]	POPC, POPS, PIP_2	3 x 3 μs
AT SIMULATIONS		
TM-JM (WT)	POPC, POPS, PIP_2	3 x 100 ns

The TM-JM domain used the NMR structure (PDB:2M20)

[a] R645N/R646N/R647N triple mutant.

In the initial configuration, the POPS and PIP_2 lipids were randomly distributed in the bilayer, as illustrated in Figure 4-3A. To monitor the interaction of the PIP_2 molecules with the protein, the evolution of the minimum distance between the protein and six PIP_2 molecules was calculated over the course of the CG-MD simulation. During the simulations, the initial contacts between the PIP_2 molecules and the protein were established mostly within the first 500 ns (see Figure 4-3B). It can be seen from the plot that out of six PIP_2 molecules, three to four PIP_2 molecules formed long-lived interactions with the proteins, whilst the rest of the PIP_2 molecules were weakly bound and dissociated multiple times during the simulations.

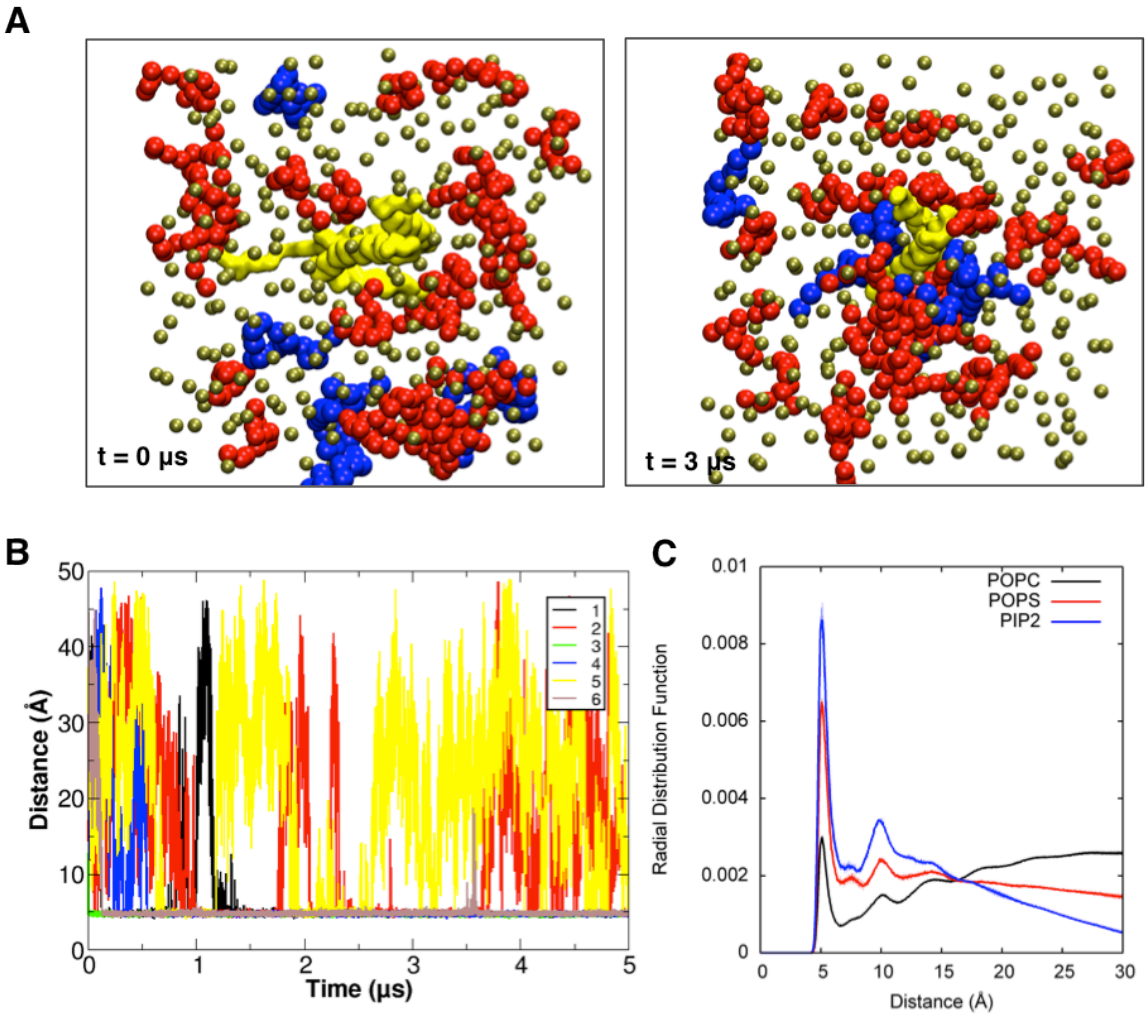


Figure 4-3. A) Snapshot of the starting configurations at t = 0 ns, in which the PIP_2 molecules are farther away from the protein and at t = 3 µs, in which PIP_2 interacts with the protein. The Cα trace of the protein is shown in surface represenatation in yellow, POPS and PIP_2 lipids are shown as red and blue spheres, respectively. The phosphate groups of the POPC are shown as brown spheres. The rest of the components of the system (water, ions, POPC lipid tails) are not shown for clarity. B) The time evolution of the minimum distance between PIP_2 and protein as a function of time for a representative CG-MD simulation of the TM-JM (WT) domain in a POPC:POPS:PIP_2 bilayer. A different colour is used for each individual PIP_2 molecule. C) CG Radial distribution function (RDF) between the head group of the lipids surrounding the protein for the different lipid species in the POPC:POPS:PIP_2 bilayer.

It is of interest to investigate the behaviour of the different lipids surrounding the EGFR TM-JM domain. To do so, the radial distribution functions of the centre of mass of the lipid head group particle (PO4) relative to the centre of mass of the protein were calculated (Figure 4-3C). It is apparent from the RDFs that the PIP_2 lipids were observed at the highest probability within ~5 Å distance from the protein. A clustering of POPS lipids was also observed around the protein, but with a much lower probability than PIP_2, suggesting that the interaction between the EGFR TM-JM domain and anionic lipids is dominated by PIP_2. The lipid clustering around the protein can be visualized using the density map, as shown in Figure 4-4(A-C). The density maps illustrate the clustering of anionic lipids in a ring like pattern around the protein, with an obvious segregation of PIP_2 to the membrane region in close proximity to the protein (Figure 4-4C).

To identify the key residues involved in the protein-lipid interactions during the CG-MD simulations, the average protein-lipid contacts within a 6-Å cut-off were calculated. The contact analysis revealed that the positively charged amino acids from residues 645–657 in the JM domain accounted for more than 50% of the total contacts with PIP_2 (see Figure 4-5A). The high percentage of contacts was observed mainly with the first three clusters of positively charged residues, and no significant contact with lipids was observed beyond L658, except for Q660 (~40%), suggesting that the C-terminal ends of both monomers were completely detached from the bilayer (see snapshots in Figure 4-4D).

.

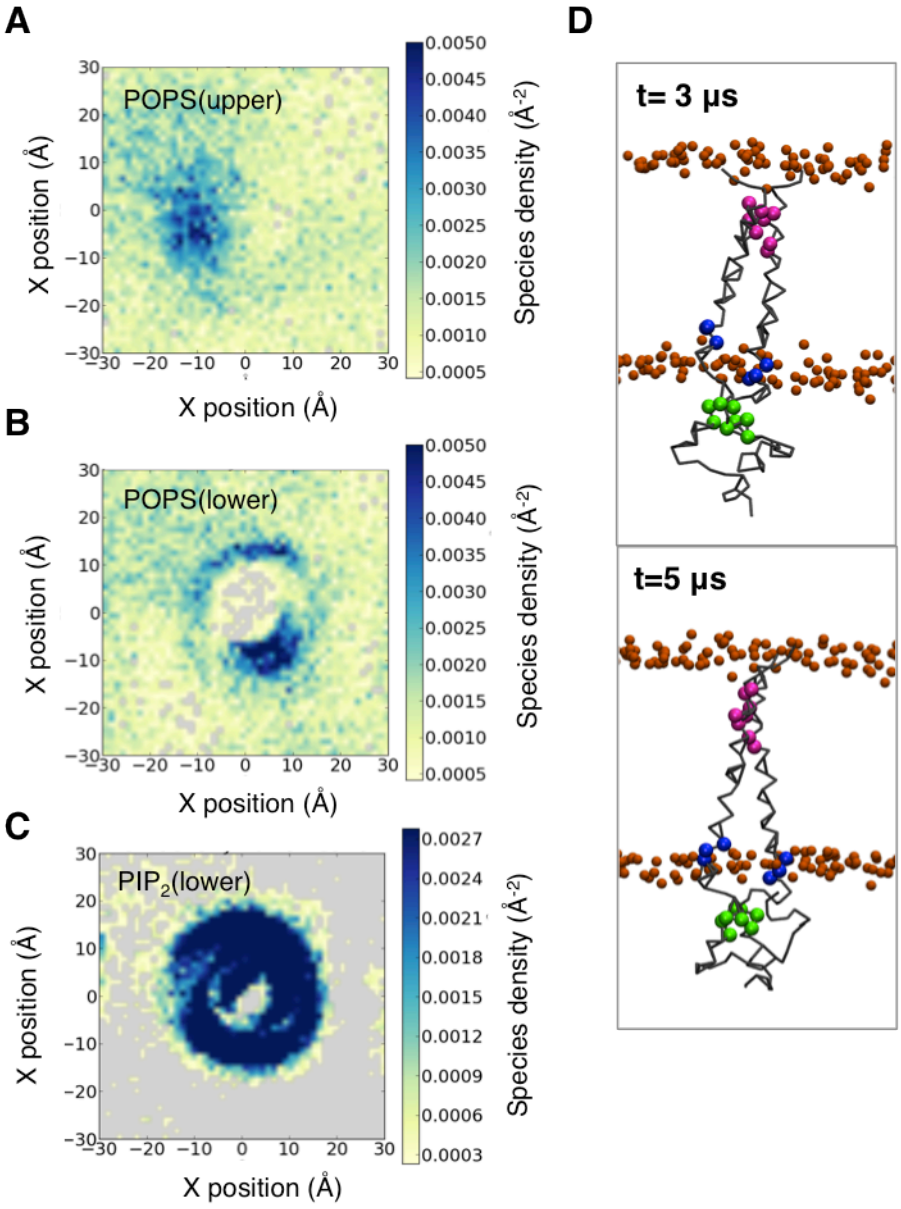


Figure 4-4. Time-averaged 2D particle densities around the protein were used to illustrate the clustering of anionic lipids around the wild-type EGFR TM-JM domain. A) Clustering of POPS molecules on the upper leaflet of the POPC:POPS: PIP_2 bilayer. B) Clustering of POPS on the lower leaflet of the POPC:POPS: PIP_2 bilayer. C) Clustering of PIP_2 on the lower leaflet. D) Snapshots of the EGFR TM-JM domain at t = 3 µs and t = 5 µs; the N-terminal motif is shown as pink spheres, the R645/R646/R647 and LRRLL motifs are shown as blue and green spheres, respectively. The Cα trace is shown in black, and the phosphate groups of the lipids are shown as orange spheres.

To assess the preferential interaction between the JM domain and lipids, the fractional contact between the protein and lipid head groups, relative to the number of lipids within the 6-Å cut-off, were calculated over the course of the simulations. It is also evident from the plot, as indicated previously in the RDFs, that PIP_2 dominates the interaction with the

protein throughout the simulation, with the more persistence contacts made by positively charged residues, particularly R645, R647 and K652 (Figure 4-5D).

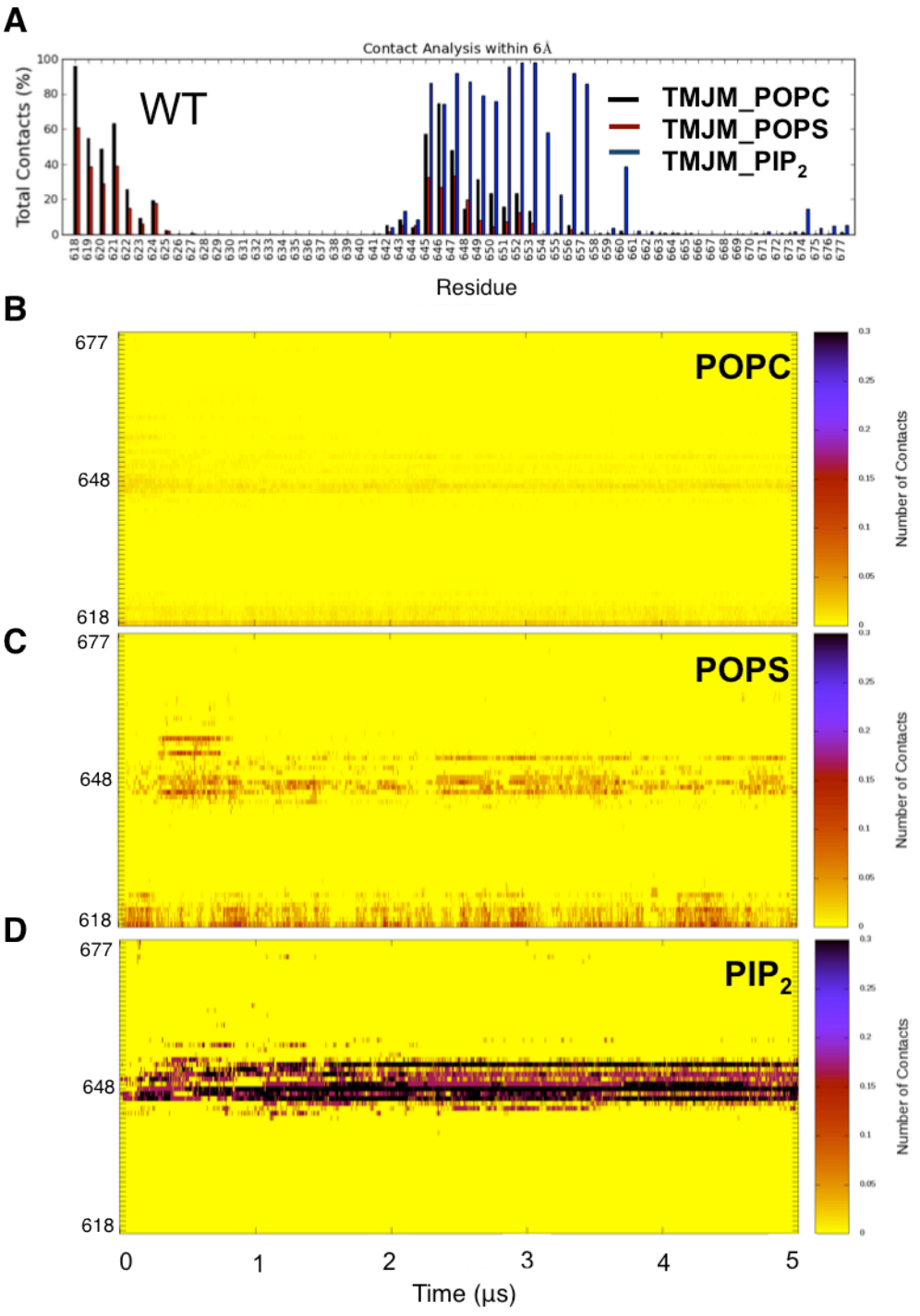


Figure 4-5. A) Normalized average number of contacts between the phosphate group of the lipids and protein within the 6 Å cut-off. B-D) Interactions between the head groups of the lipids and protein within the 6 Å cut-off were calculated for a representative simulation of the TM-JM domain in a POPC:POPS:PIP_2 bilayer. Interactions between the POPC head groups and protein (B), POPS head groups and protein (C) and PIP_2 head groups and protein (D).

To test the effect of anionic lipid clustering around the JM domain, the first three positively charged residues, R645, R646, and R647, within the JM-A domain were mutated to Asn. This mutation was selected based on a previous study by Michalaidis et al., in which they demonstrated that neutralization of R645 to R647 by mutation to Asn abolished EGFR phosphorylation (*78*). As expected, the PIP_2 molecules were no longer in close contacts with the N645/N646/N647 residues, although high percentages of total contacts (>50%) were observed between H648, R651, K652, R653, R656 and R657. The interaction between the lipids and proteins were further examined by analysing the average number of interactions between all proteins and the head groups of the different lipid species. The average number of interactions of PIP_2 with the WT and mutant proteins were mapped onto the CG structure (Figure 4-6) using a blue, red and green (BRG) colour scale.

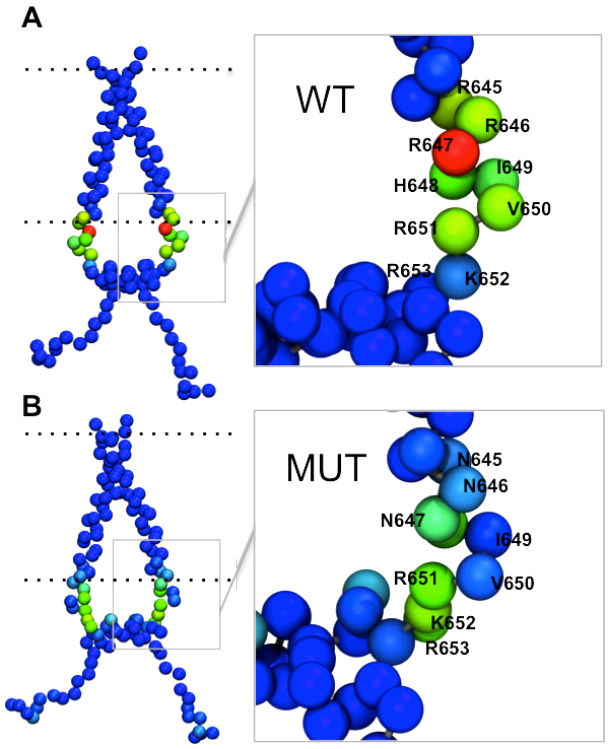


Figure 4-6. The average number of contacts of a representative simulation showing the interactions between the EGFR TM-JM domain and PIP_2. The contacts are mapped onto the initial CG model of the TM-JM (t = 0 ns). A) The average number of contacts between PIP_2 and the WT sequence. B) The average number of contacts between PIP_2 and the R645N/R646N/R647N mutant. A blue to red colour scale is used to indicate no interaction (blue) and strong interaction (red). The Cα particle are shown as spheres. The lipid bilayer is represented by dotted black lines.

To visualize how anionic lipid clustering is affected by the mutation, lipid density maps were computed for the anionic lipids over the course of the CG-MD simulations. Figure

4-7(A-C) shows how the clustering of PIP_2 and POPS was affected by the mutation. It was observed in the simulation with the mutated form of the protein, that either one or both JM-A monomers made contacts with the lipid bilayer, as exemplified in the snapshots of the final frames of the CG-MD simulations (Figure 4-7D).

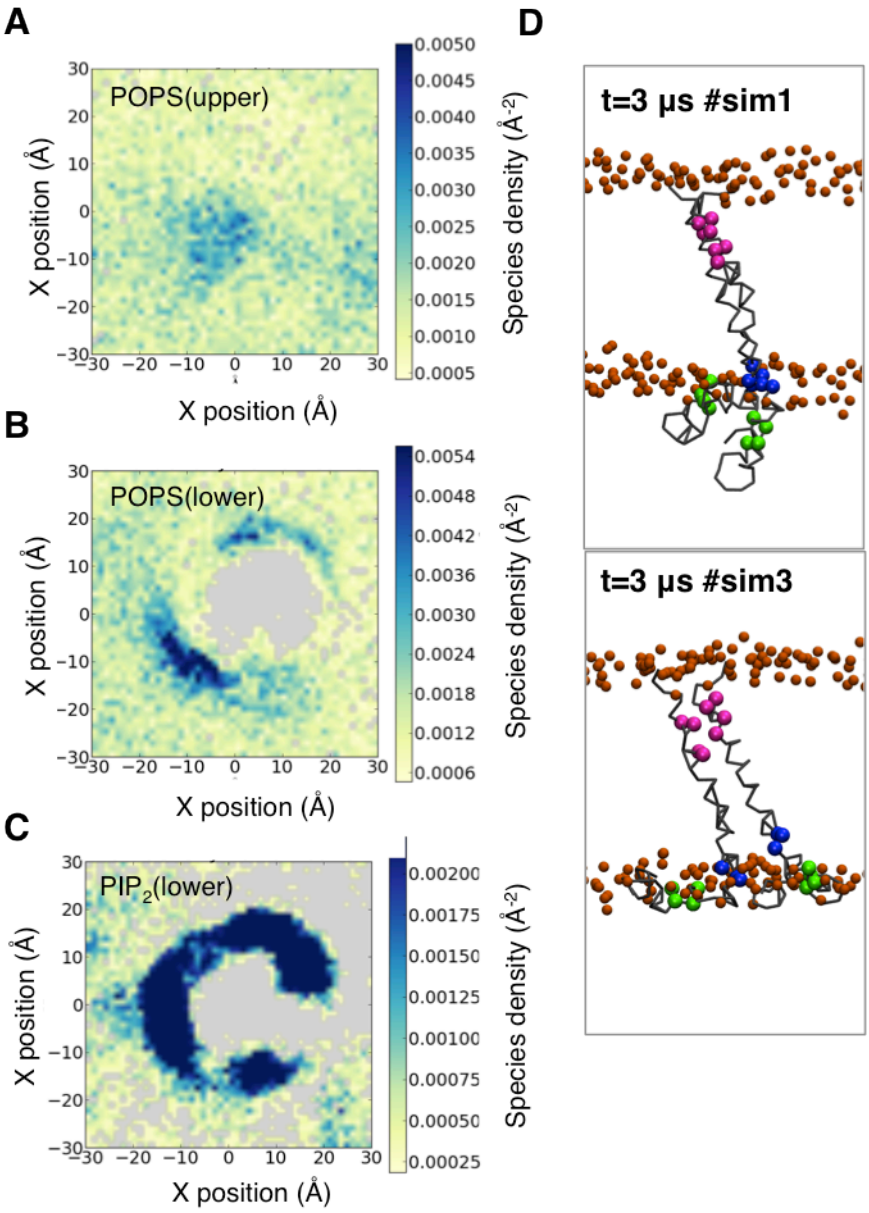


Figure 4-7. Time-averaged 2D particle densities around the protein were used to illustrate the clustering of anionic lipids around the mutant protein. A) Clustering of POPS on the upper leaflet of the POPC:POPS:PIP_2 bilayer. B) Clustering of POPS on the lower leaflet. C) Clustering of PIP_2 on the lower leaflet. D) Snapshots of the EGFR TM-JM domain at t = 3 µs and t = 5 µs; the N-terminal motif is shown as pink spheres. R645N/R646N/R647N and the LRRLL motif are shown as blue and green spheres, respectively. The C-alpha trace is shown in black, and the phosphate groups of the lipids are shown as orange spheres.

The CG-MD simulations revealed that PIP_2 and POPS clustering was slightly different from that of the WT protein, as the high density clusters of these anionic lipids were located farther away from the protein as a result of the interactions between the lipids and the R651/K652/R653 and R656/R657 residues, as well as other residues along the JM segment (Figure 4-8A). This was reflected in the protein-lipid contact matrices calculated for every 200 ps, as shown in Figure 4-8B-D.

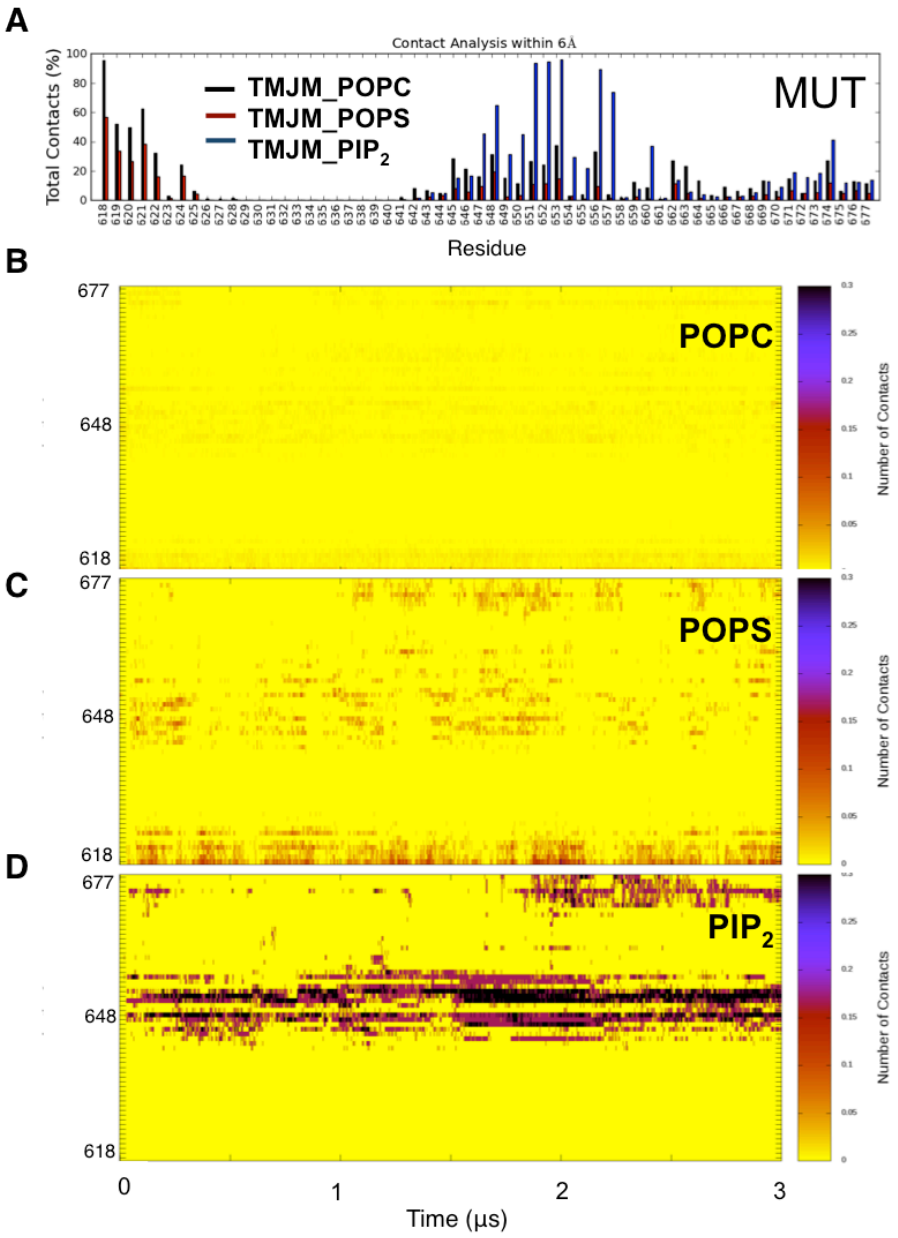


Figure 4-8. A) Normalized average number of contacts between the phosphate groups of the lipids and protein within the 6 Å cut-off. Interactions between the head groups of the lipids and protein within the 6 Å cut-off were calculated for a representative simulation of the mutated TM-JM domain for interactions between POPC (A), POPS (B) and PIP_2 (C) and the protein.

4.3.1 AT-MD simulations of TM-JM in PC:PS:PIP_2 bilayer

A final snapshot from each of the three repeats of the CG-MD simulation in the PIP_2-containing bilayer was converted to an atomistic representation, followed by 100-ns atomistic simulations. The AT-MD simulations were performed using a GROMOS96 43a1 force field. A simple point charge (SPC) water model was used because SPC is recommended for use with the GROMOS96 43a1 force field. Conformational drift of the TM-JM domain during the AT simulation was assessed by measuring the Cα RMSD values relative to those of the initial structure. In one of the AT simulations in which the JM was in an antiparallel orientation, the TM-JM domain was stable throughout the simulation, as indicated by the RMSD values, which begin to plateau after 40 ns (Figure 4-9A). The LRRLL motif within the JM-A region remained helical and was not attached to the bilayer. The flexibility of the JM domain can be seen in the Cα RMSF values; note the low RMSF values for the LRRLL motif due to antiparallel dimer formation, and high RMSF values for the C-terminal end of the JM domain, which was observed to move freely in the solvent (Figure 4-9B).

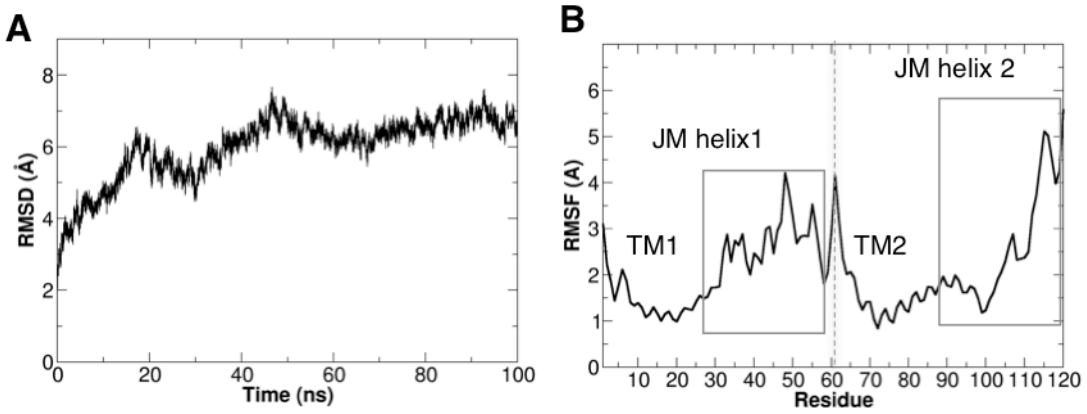


Figure 4-9. A) The Cα RMSD relative to that of the starting structure as a funtion of time. B) The RMSF profiles for the dimer subunit calculated for one of the AT simulations, in which the TM-JM domain was observed to be very stable and adopted the same conformation as the NMR structure.

In the same simulation, the JM domain was initially in an antiparallel orientation, although the LRRLL motif from one of the monomers was parallel to the membrane, and another LRRLL motif from the second monomer was perpendicular to the membrane. It was observed that the LRRLL motifs adopted an antiparallel dimer, as seen in the NMR, structure at the end of the simulation; see the final snapshot in Figure 4-10A. In all the AT

simulations, the JM region remained unbound from the membrane, and this can be clearly depicted by the density profiles of the entire JM domain, R645-647 and the LRRLL motif, as analysed previously for the CG-MD simulations, see Figure 4-10B.

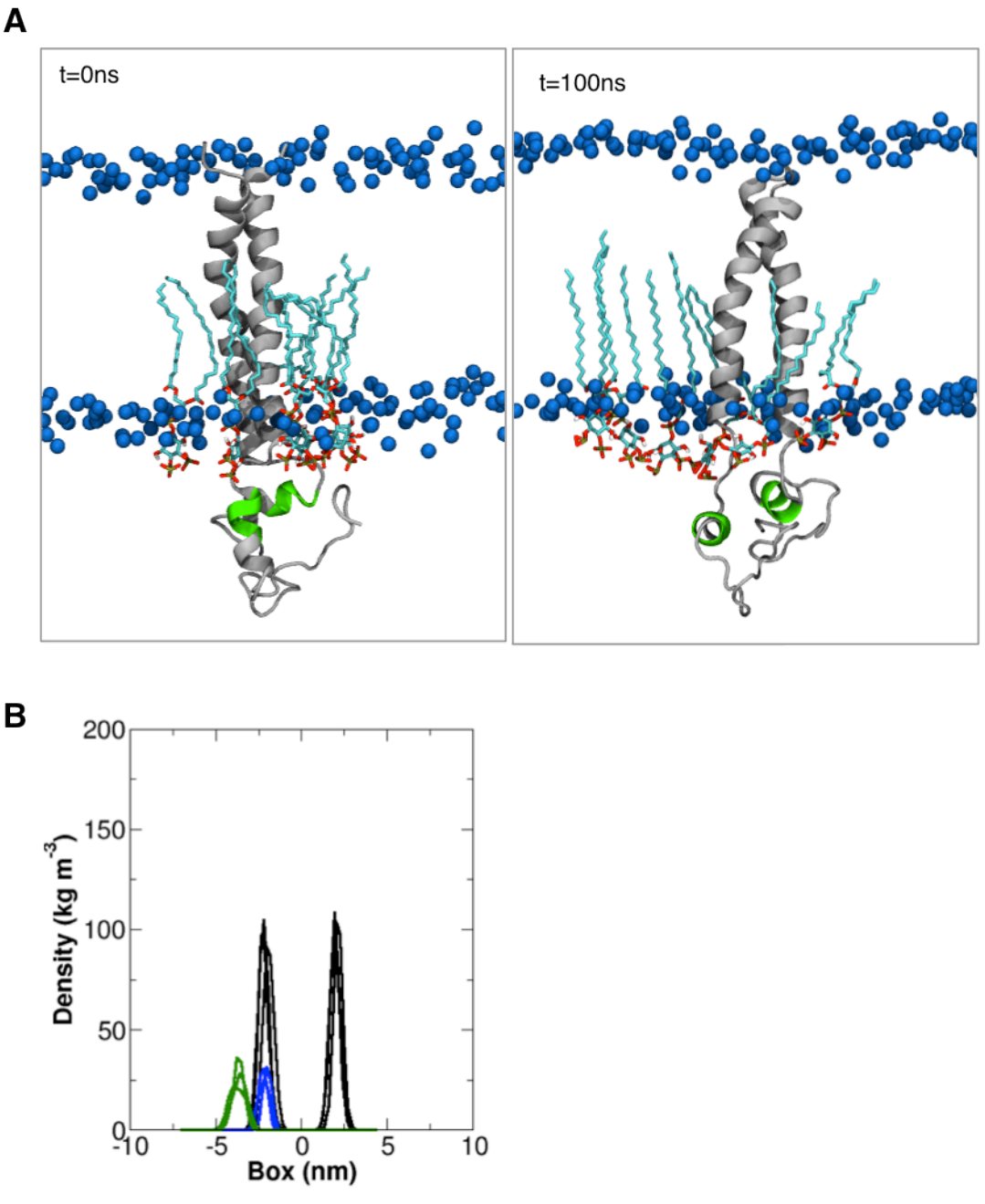


Figure 4-10. A) Snapshots of the EGFR TM-JM domain at t = 0 ns and t = 100 ns. The protein is shown as a cartoon representation, with the LRRLL motif coloured green. Lipid phosphate groups are shown as blue spheres and PIP_2 lipids are shown as stick representations. The rest of the elements in the system are not shown for clarity. B) Density profiles, relative to the positions of the lipid phosphate groups (black), along the normal of the membrane for R645 (blue), the LRRLL motif (green) and the entire JM segment (purple) for all three repeats of the AT simulation.

An overall average of approximately seven hydrogen bonds were formed between the JM domain and the POPC lipids (despite POPS constituting 90% of the lipids), while approximately six and three hydrogen bonds were formed between the JM domain and PIP_2 and POPS, respectively. It is clear from this comparison that PIP_2 interacted favourably with the JM domain, particularly with residues 645–647, as reflected by the minimum distance between the JM-A residues and PIP_2 (see Figure 4-11A-D). In all repeats of the AT-MD simulations, the JM domain remained detached from the membrane, which suggests that PIP_2 stabilizes the JM domain in an antiparallel orientation away from the bilayer, as depicted in the snapshot of the final frame in Figure 4-11E.

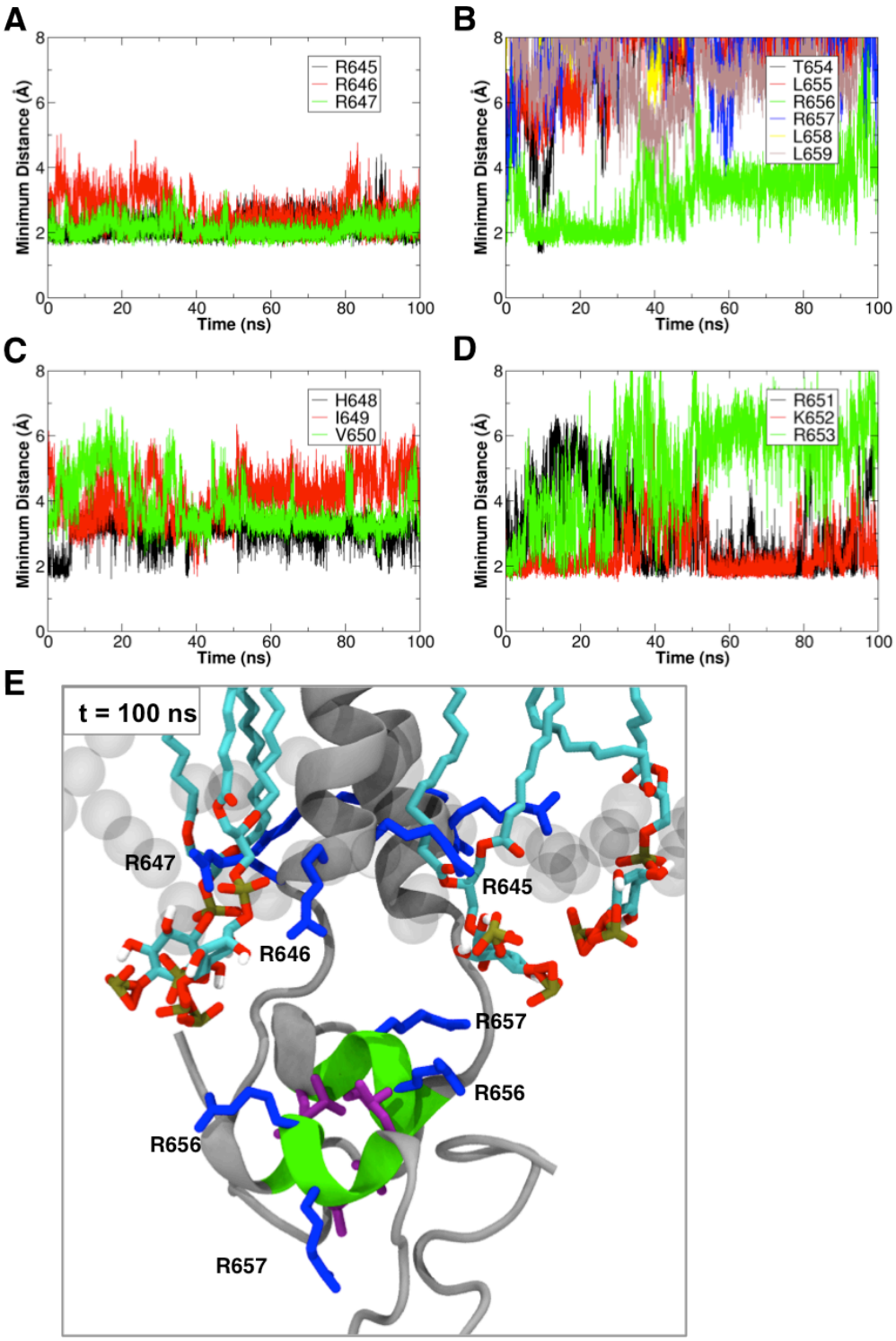


Figure 4-11. Time evolution of minimum distance as a function of time between amino acid residues of the JM-A domain and PIP_2 lipids. The minimum distance between residues R645/R646/R647 (A), residue T654 and the LRRLL motif (B), residues 648–650 (C), and residues R651/K652/R653 with PIP_2 (D). E) A snapshot of the AT simulation at 100 ns illustrates the position of the antiparallel JM-A domain and PIP_2. The LRRLL motif is shown in green, with Arg and Leu residues shown as licorice representations in blue and purple, respectively.

4.3.2 Control simulations in POPC and POPC:POPS bilayers

CG-MD simulations were conducted in POPC and POPC:POPS bilayers without PIP_2 using the same protein structures as in the previous PIP_2-containing simulations. These simulations were performed to compare the behaviour of the TM-JM domain in PIP_2-containing and PIP_2-depleted bilayers. In the POPC:POPS bilayer, a degree of clustering of POPS molecules was observed around the protein, forming either the ring-like pattern observed for PIP_2, or two or three discrete patches around the protein in the lower leaflet. POPS lipids also clustered in the outer leaflet of the bilayer around the proximal T residues 618 to 624, with more than 50% of the total contacts with K618. In simulations in POPC and POPC:POPS bilayers, the JM region made some contacts with the membrane, forming an average of 10–20% of the total contacts with the bilayer (Figure 4-12A,B).

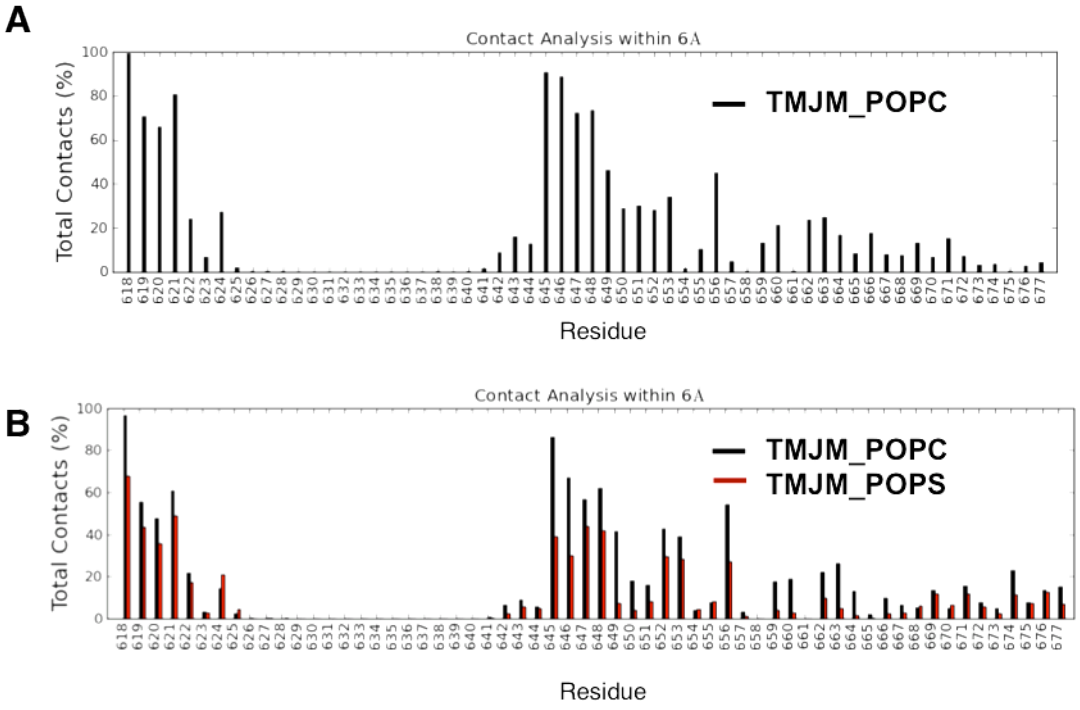


Figure 4-12. The percentage of the average number of contacts between phosphate groups of POPC and POPS (using a cut-off distance of 6 Å) averaged over three repeats of the CG-MD simulations in A) a pure POPC bilayer and B) a POPC:POPS bilayer.

The interactions between the JM domain and the different bilayers can also be illustrated by comparing the distances between the COM of the membrane-proximal residues R645–648 and the COM of the JM domain, with respect to the COM of the phosphate groups of

the bilayer (Figure 4-13). Note that only the WT JM domain remained away from membrane when the simulations were extended to 5 µs (see inset in Figure 4-13C). For other PIP_2-depleted membranes and the mutant JM domain, the JM-A antiparallel dimer was disrupted and interacted with the bilayer.

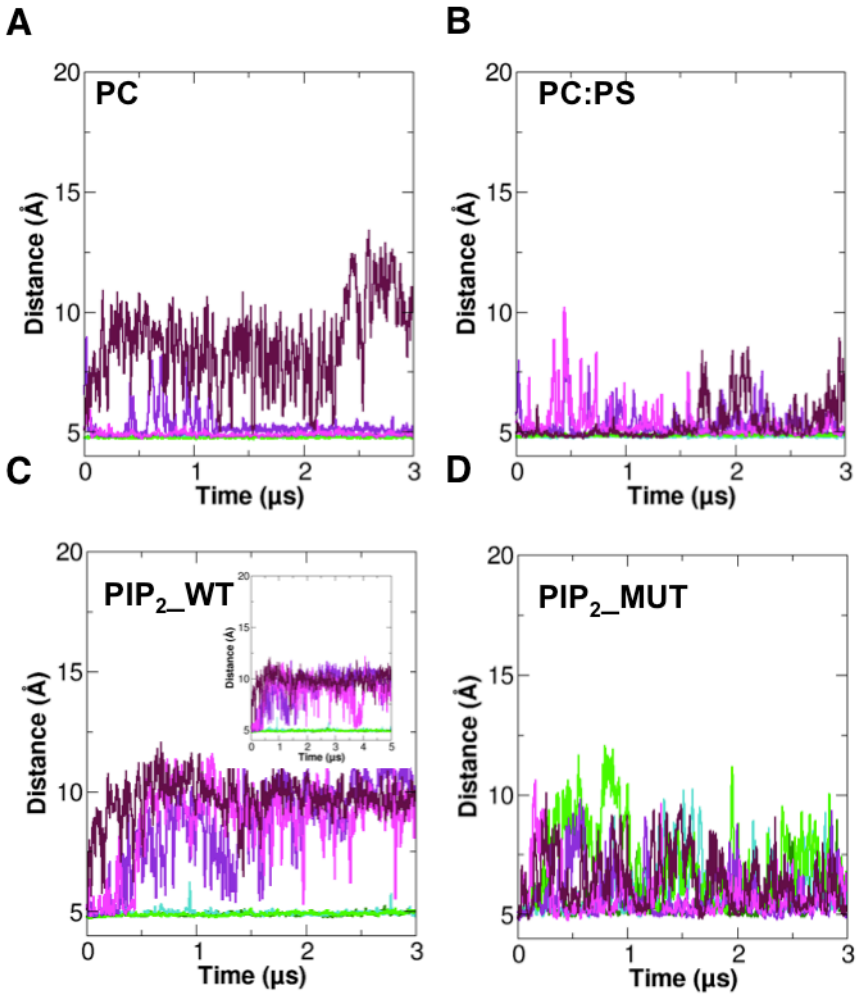


Figure 4-13. . Comparison of the evolution of minimum distances between the COM of residues R645–647, the COM of the LRRLL motif and the COM of the membrane (purple). The distances between R645–647 and the COM of the bilayer (green) for A) a pure POPC bilayer, B) a POPC:POPS bilayer and C-D) POPC:POPS:PIP_2 bilayers with wild-type and mutant proteins. Individual simulations of the three repeats in all of the bilayer types are shown in different shades of green and purple for R645–647 and the LRRLL region, respectively.

The association of the JM domain and the membrane was also reflected in the density profiles of residues R645–647, the LRRLL motif, the JM domain and the phosphate groups of the lipids. In the PIP_2-depleted bilayer, the LRRLL density profile almost overlapped with the density profile of R645–647, thereby providing additional evidence of JM-membrane interactions in PIP_2-depleted membranes (Figure 4-14).

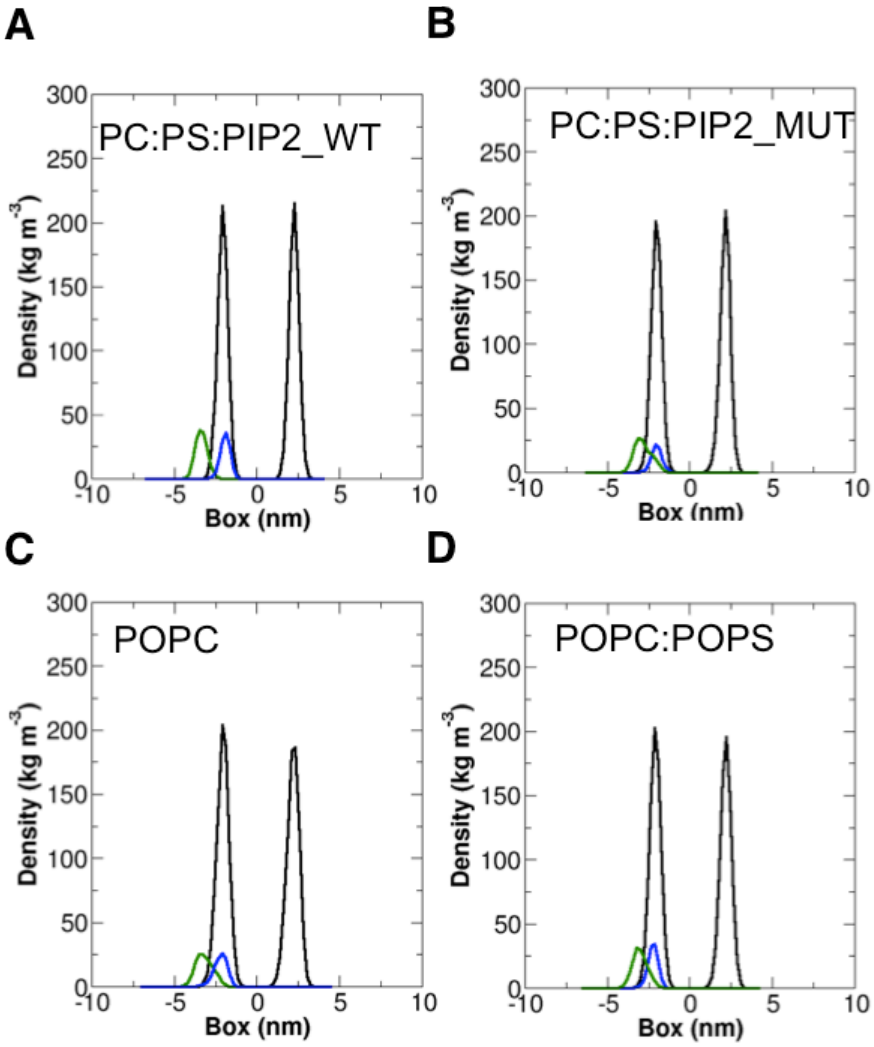


Figure 4-14. Density profiles along the normal of the membrane calculated for the WT and mutant proteins. Residues R645/R646/R647 at the membrane-proximal are shown in blue, and the LRRLL motif within the N-terminal region of JM-A region is shown in green. The density of the phosphate groups of the bilayer is shown in black.

4.4 CG-MD SIMULATIONS OF EGFR TM DOMAIN IN GM3 CONTAINING BILAYERS

This section constitutes the second part of Chapter 4. Note that only the TM domain was considered in this set of simulations because the aim was to test how GM3 modulates TM helix packing using the CG-MD approach. The EGFR TM domain dimer was modelled based on the NMR structure of the TM-JM domain (*57*). The CG-MD simulations were performed in GM-containing bilayers to investigate the effects of GM3 on the packing of the EGFR TM helices. The CG-MD simulations are detailed in Table 4-2. DPPC, DUPC

(1,2-(cis-cis-9,12-octadecadienoyl)-sn-glycerol-3-phosphatidylcholine), CHOL and GM3 mixture were assumed to phase separate. This assumption was made based on a previous study by Risselada et al., who showed that a ternary lipid mixture, DPPC:DUPC:CHOL, phase separated at 295 K (*168*). Thus, the simulations for the WT and mutant proteins in the DPPC:DUPC:CHOL:GM3 bilayer (50:28:20:2) were performed at 295 K (equivalent to the temperature used in the original MARTINI study of domain formation (*166*)). A self-assembly simulation was performed for 100 ns to generate a pure POPC bilayer consisting of 250 POPC lipids. GM3 and other lipids species were incorporated in the pure POPC bilayer using a similar protocol to that described for PIP_2-containing bilayers. The simulated system contained 250 lipids, ~5,000 CG waters and ~170 counter ions in a simulation box of dimensions 100 x 100 x 120 $Å^3$. Control simulations were performed for the WT sequence at 323 K in a DPPC:DUPC:CHOL:GM3 bilayer, because it is known from previous simulations in the lab using the ternary mixture (DPPC:DUPC:CHOL) formed an Ld membrane at this temperature (*169*). Other control simulations in simpler bilayers, consisting of pure POPC and POPC:GM3, were performed at 310 K.

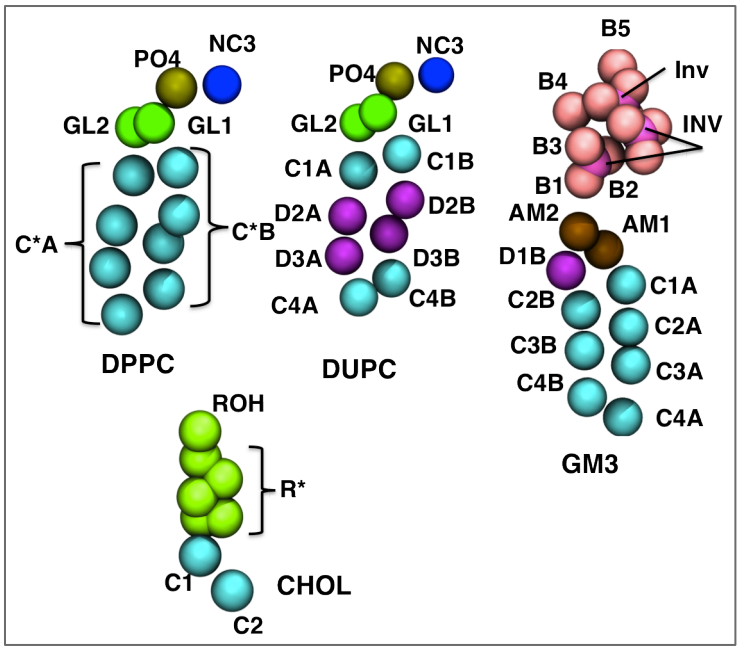


Figure 4-15. The CG structures for DPPC, DUPC, CHOL and GM3 used in the simulations.

Table 4-2. Summary of CG-MD simulations of TM in GM3-containing bilayers and control simulations.

System	Composition	N x Length
GM3-containing bilayer		
TM_WT	DPPC:DUPC:CHOL:GM3 (T = 295 K)	4 x 3 μs
TM_mut(K618G)	DPPC:DUPC:CHOL:GM3 (T = 295 K)	3 x 3 μs
Control simulations		
TM_WT	DPPC:DUPC:CHOL:GM3 (T = 323 K)	3 x 3 μs
	POPC	3 x 5 μs
	POPC:GM3	3 x 3 μs
TM (618 to 645)		

One of the primary interests in this study was to examine the regime of GM3 interactions with the TM dimer during the simulations. Thus, the minimum distances between GM3 and protein over the course of the simulations were calculated for the four TM_WT simulations, as presented in Figure 4-16A. It can be seen from the plot that not all the GM3 lipids interacted with the protein simultaneously. Most of the time, only two to three GM3 lipids interacted with the protein, except in one of the simulations, in which five GM3 lipids clustered around the protein, as depicted in Figure 4-16A(sim4). Overall, it seems that GM3 interacted with the TM dimer, but did not permanently bind to it, although it did form long-lived interactions with the TM dimer in one of the simulations (sim4). In general, the GM3 lipids associated and dissociated from the protein multiple times, and interacted favourably with other GM3 lipids to form small GM3 clusters.

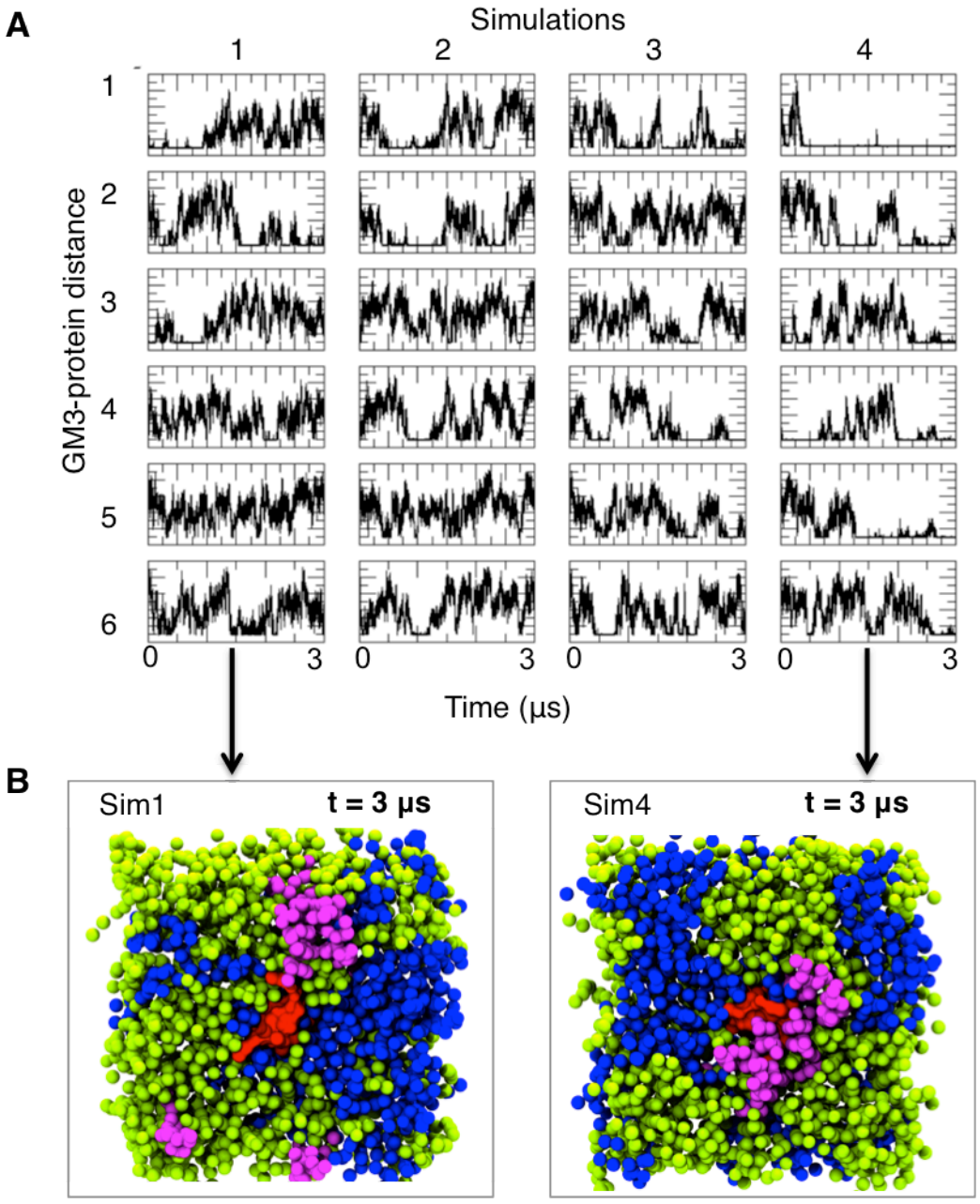


Figure 4-16. Time evolution of minimum distances between the six GM3 lipids and protein for the TM_WT CG-MD simulations in a DPPC:DUPC:CHOL:GM3 bilayer. B) Two representative snapshots of the final frame taken from two of the TM_WT simulations (sim 1 and sim4). The DPPC lipids are shown as green spheres, DUPC lipids as blue spheres, and GM3 as magenta spheres. The EGFR TM helix dimer is shown as a surface representation in red.

To further investigate the dynamic behaviour of the EGFR TM dimer in the Ld/Lo bilayer, dimer association was monitored by plotting the inter-helix distance as a function of time. The plot indicates that in two simulations, the dimer dissociated for several hundred nanoseconds for up to 1.5 μs (Figure 4-17A, sim1 and sim3). However, in one of the simulations, despite long-lived interactions between GM3 and the protein, the dimer was found to be stable (Figure 4-16A, sim4), as reflected in the inter-helix contact matrix (Figure 4-17B, sim4).

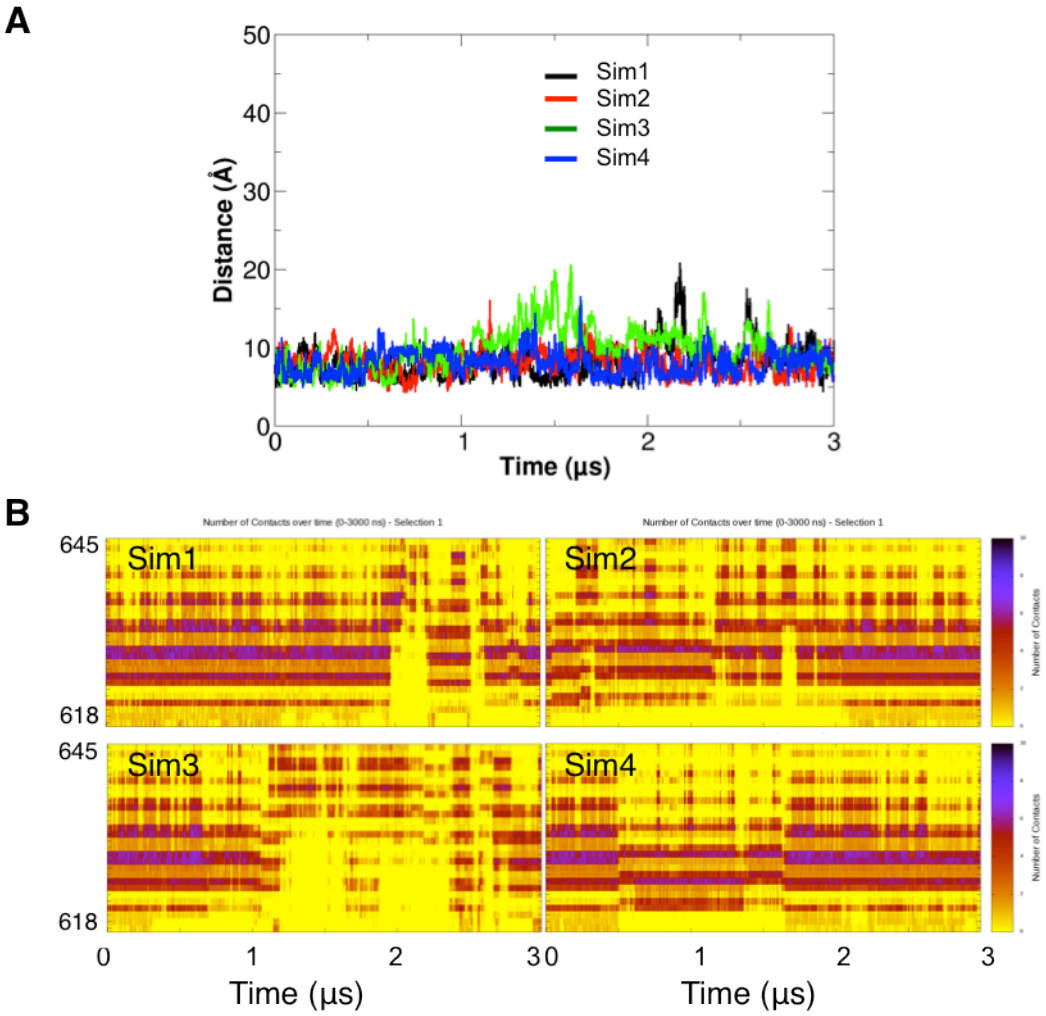


Figure 4-17. A) Time evolution of inter-helix distances for the TM_WT CG-MD simulations. Each individual simulation is shown in a different colour. B) Time evolution contact matrix calculated for each of the individual simulations of TM_WT protein in a DPPC:DUPC:CHOL:GM3 bilayer (T = 295 K), as labelled.

To explore how the interactions may modulate EGFR TM dimer packing, the contact matrices were calculated every 200 ps. The contact matrices revealed that the contact

interface of the dimer, formed by residues T624/G625/G628/A629, was completely lost when the dimer dissociated. Although the dimer reformed at the end of the simulation, the contact interfaces were variable, as reflected in the contact matrix between the interfacing residues, as shown in Figure 4-18A. The result suggests that the protein-lipid interaction has the potential to destabilize helix-helix packing, but did not lead to permanent dissociation; only the dimer interface was slightly perturbed relative to the initial packing (Figure 4-18B, t = 3μs).

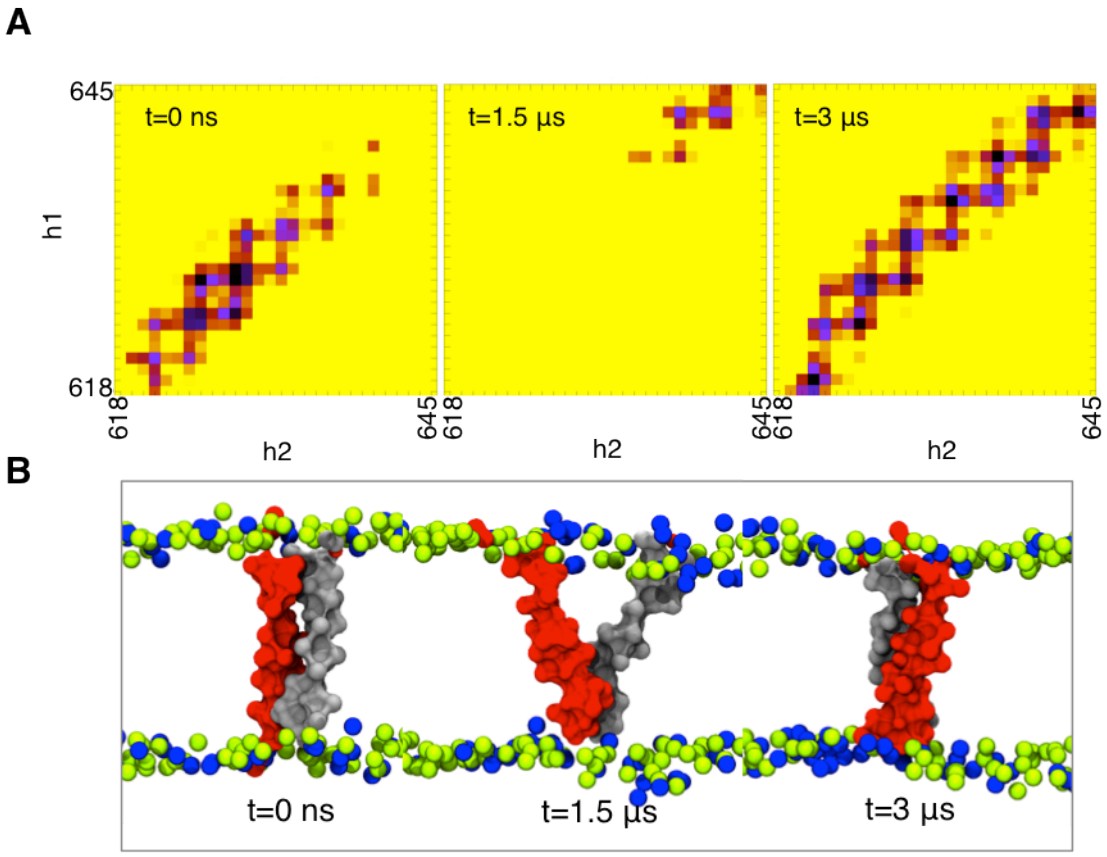


Figure 4-18. A) The helix-helix contact matrix calculated for a representative simulation at 0 ns, 1.5 μs and 3 μs and showing the initial contact interface of the TM helix dimer changed significantly, even after the dimer reformed after dissociation (t = 1.5 μs). B) Snapshots of the matching CG-MD trajectory showing the dissociation and reassociation of the EGFR TM dimer.

4.4.1 Protein-lipid interactions

The interaction between protein and GM3, which was of primary interest, was assessed by analysing the fractional interactions of the protein and the GM3 per frame per lipid. The fractional contact matrix between the head group of GM3 and the protein within a 6 Å cut-off revealed that GM3 interacted with five to six residues on the N-terminal ends of the EGFR TM domain, with a relatively higher number of contacts with residue K618 (see

Figure 4-19A). This is in agreement with an experimental study by Coskun et al. (*77*). Additionally, it was evident from the time evolution contact analysis that a stable interaction was made between the GM3 head group and K618 throughout the simulation (see Figure 4-19B). GM3 also interacted with other residues at the N-terminal end, namely I619, P620, S621, I622, albeit less strongly, and these interactions were with the GM3 head group. The highest number of contacts was observed between the sialic acid particles of GM3 (both B4 and B5) and the membrane-proximal residue K618 (see Figure 4-19A).

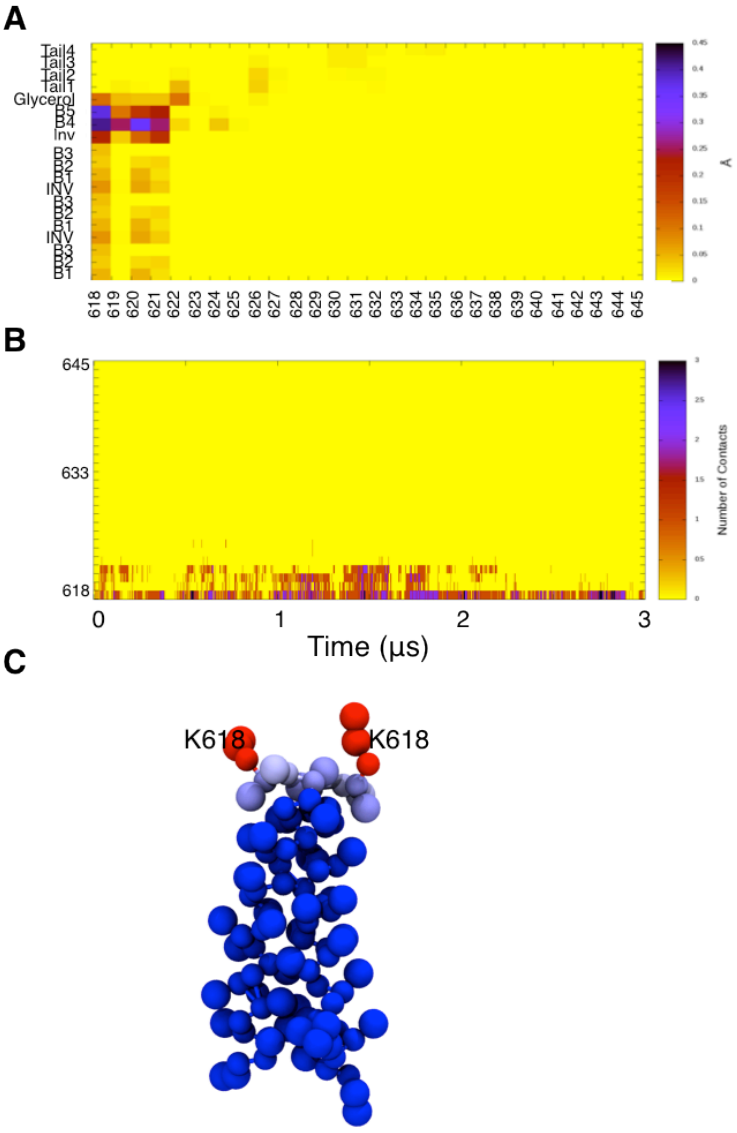


Figure 4-19. A) The fractional contact matrix of interactions between GM3 and protein for four representative CG-MD simulations (all other simulations had the same pattern of protein-GM3 interactions). B) The time-evolution contact matrix between the GM3 head group and protein throughout the simulation. C) Interaction between the GM3 head group and the EGFR TM domain dimer within a 6 Å cut-off was mapped onto the initial CG dimer structure. The blue to red color scale indicates no interaction (blue) and strong interactions (red).

4.4.2 Comparison of WT EGFR and K618G dimers in an Ld/Lo bilayer

It has been shown experimentally that GM3 inhibition can be released by removing the neuraminic acid component of GM3 or by mutating residue K618 (or K642 according to an alternative EGFR numbering scheme that includes the 24 residue signal sequence) to Gly (*77*). The K618G mutant has been shown experimentally to abolish the inhibition of EGFR by GM3 (*77*). To test whether the experimentally observed release of GM3 inhibition by the mutation of K618 at membrane-proximal sites could be detected by CG-MD, four simulations of the K618G mutant in an Ld/Lo bilayer were performed at 295 K. During the simulations, the EGFR TM dimer remained stable and tightly packed, as indicated by the close contact between residues of the dimer, and all four repeats converged on the same contact pattern (data not shown).

4.4.3 Control simulations in POPC and POPC:GM3 bilayers

Control simulations in pure POPC and POPC:GM3 bilayers were also performed for comparison. As expected, the EGFR TM dimer remained relatively tightly packed in POPC, as well as in POPC:GM3, bilayers compared with the Ld/Lo bilayer. However, in the presence of POPC:GM3, there was an occasion in one of the three simulations in which the TM dimer dissociated by up to 20 Å for several hundred nanoseconds, although the dimer reformed and gained its initial packing. Overall, it appears that the presence of GM3 in the Ld bilayer did not cause any significant changes to the TM dimer.

It is also interest to test the effect of temperature changes on TM packing. In a previous study of phase separation by Risellada & Marrink, a DPPC:DUPC:CHOL bilayer clearly underwent domain formation at 295K (*168*). In this study, the simulations were repeated at 323 K for both WT and mutant proteins, assuming that the DPPC:DUPC:CHOL bilayer was in the Ld phase at 323 K. Interestingly, a more pronounced dissociation of the dimer subunit was observed for both WT (Figure 4-20B) and mutant (data not shown). As demonstrated by the inter-helix distances plot, there were several occasions in which the dimer subunits were spaced farther apart (by up to 60 Å), and a similar behaviour was observed for the mutant.

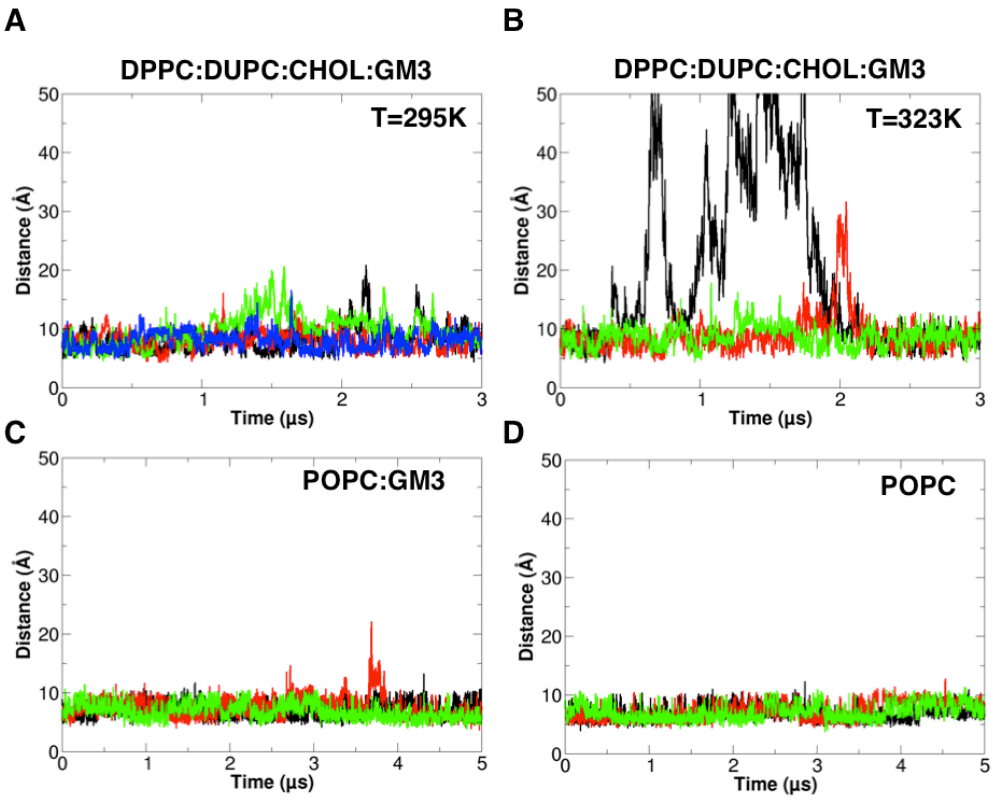


Figure 4-20. The evolution of inter-helix minimum distances as a function of time for the CG-MD simulations in a DPPC:DUPC:CHOL:GM3 bilayer at T = 295 K (A) and T = 323 K (B). Control simulation in POPC:GM3 (C) and pure POPC (D) bilayers revealed a slight perturbation of EGFR dimer packing in the presence of GM3 lipids.

4.5 DISCUSSION AND CONCLUSIONS

It has been proposed that electrostatic interaction between the negatively charged residues of the JM region and anionic lipids play an important role in the regulation of the EGFR receptor. Here, CG and AT MD simulations have been used to study the behaviour of the EGFR TM-JM domain in the presence of POPS and PIP_2 lipids. The CG-MD simulations revealed that the positively charged residues within the JM-A domain are able to attract and sequester PIP_2, resulting in clustering of PIP_2 around the TM-JM region. In PIP_2 containing membranes, the most prominent contacts are made between the first cluster of positively charged residues (R645/R646/R647), which is in agreement with a previous experimental study by Michailidis et al. 2011 (*78*). A comparison of the interactions between the JM domain and PIP_2 in WT and mutant proteins during the CG-MD simulations suggests that the mutation not only affected PIP_2 binding, but also disrupted the orientation of the JM domain, resulting in dissociation of the antiparallel JM dimer.

Meanwhile, in PIP_2 depleted membranes, the JM domain is bound to the membrane and the JM-A dimer is either dissociated completely or partially deformed.

Further refinement with AT simulations revealed that the TM-JM dimer is stable and the JM-A domain oriented itself in an antiparallel manner that corresponds to the active state of the receptor, which is in agreement with the NMR structure (*57*). This is also in agreement with the behaviour of the Neu TM-JM domain in a recent study by Matsushita et al. using solid-state NMR and fluorescent microscopy, who demonstrated that the JM domain was flexible and detached from the negatively charged membrane in the presence of PIP_2 (*30*). The interaction of the EGFR TM-JM domain with anionic lipids has also been investigated using long timescale, all-atom MD simulations by Arkhipov et al. (*95*), who demonstrated that increasing amounts of anionic lipids (POPS) stabilized the association of the TM domain at its N-terminal motif, which favoured an antiparallel association of the JM-A segment. Their simulations also showed that the JM-A region of the monomeric TM-JM domain was embedded in the membrane. Taken together, these simulations support the notion that the negatively charged residues in the JM domain are able to sequester PIP_2 and that the JM-A antiparallel helix association occurs outside the membrane. Moreover, these simulations identified the effect of anionic lipid clustering around the protein, as well as the orientation of the JM domain with respect to the membrane (i.e., mutation or PIP_2-depletion). It appears that the presence of PIP_2 helps stabilize the TM-JM N-terminal association and the JM-A helix dimer formed by the LRRLL motif.

Meanwhile, the CG-MD simulations of the EGFR TM domain in an Ld/Lo bilayer revealed a preferential interaction of the sialic acid components of GM3 lipids with residue K618, which is in agreement with experimental observations (*77*). In a DPPC:DUPC:CHOL:GM3 bilayer, several short-term dissociations of the WT TM dimer were observed, whereas in POPC and POPC:GM3 bilayers and the K618G mutant, the TM dimer was relatively stable and remained intact throughout the simulation. This suggests that the strength of interaction of the TM helices is sensitive to the bilayer composition and properties (phase-separation). It is interesting that the CG-MD simulation is able to capture the preferential interaction between protein and lipid. This study is very preliminary and merits further investigation via AT simulations.

CHAPTER 5

Simulation of the intact EGFR in a physiologically relevant bilayer

5.1 INTRODUCTION

Recent advances in lipidomics have revealed the remarkable diversity of lipids and have contributed to an increased understanding of their diverse biological roles (*170*). The composition of lipids in a cell varies depending on species and intracellular membrane type (*7*). The PM bilayer is also known to be asymmetric, with mostly PC and SM in the upper leaflet and PE, PS and PIs in the lower leaflet, whilst cholesterol (CHOL) occupies both leaflets (*20*). Lipids not only form the matrix of the cell membrane in which proteins reside, they are also directly involved in membrane trafficking, regulating membrane proteins and in membrane subcompartmentalization (e.g., lipid rafts) (*7*, *13*, *15*, *171*). These microdomains form distinct Lo phases in the lipid bilayer and co-exist with the Ld (unsaturated glycerolipids) (*14*). EGFR is one of many signalling proteins known to localize in lipid rafts. Its association with lipid rafts is CHOL dependent, thus CHOL depletion from the PM triggers ligand-independent activation of EGFR (*172*, *173*), and raft association sequesters EGFR in a state inaccessible for ligand binding without altering its ligand affinity (*172*). It is also known that inhibition of EGFR by GM3 only occurs in a raft environment (*8*, *77*).

Most published computational studies on EGFR have been limited to isolated EGFR domains, mainly focusing on the kinase domain and its conformational transition from an inactive to active state, as well as the effect of oncogenic mutations on kinase inhibitor binding (*87*, *141*, *174*), (*84*). Recently, MD studies of intact EGFR have begun to emerge. Koland performed MD simulations of the complete EGFR to investigate the mechanism of self-phosphorylation of the P-sites on the C-terminal domain using a CG approach (*175*). However, in this study, the membrane bilayer was modelled in a highly simplified fashion as two planar barriers 33 Å normal to the z-axis between in which the TM domain was oriented. Long timescale, all-atom simulations of a nearly complete EGFR (the C-terminal tail was not included) were performed by Arkhipov et al. to shed some light on the interaction of the JM and kinase domains with membranes consisting of POPC and POPS

(*95*). Simulations of intact EGFR or EGFR domains in isolation (monomeric/dimeric form) have thus far been conducted in simple membranes. Therefore, no MD simulations of intact EGFR in a more complex, physiologically relevant bilayer have yet been reported. This chapter describes the construction of an intact, nearly complete model of the EGFR, and the use of CG-MD to probe the dynamics of protein-lipid interactions and lipid reorganization in a physiologically relevant (i.e., compositionally complex) bilayer.

5.2 HOMOLOGY MODELLING OF THE INTACT EGFR

The full-length EGFR comprises 1,186 amino acid residues subdivided into a: (i) an ectodomain (residues 1–620), (ii) a TM domain (621–644), (iii) a JM domain (645–682), (iv) a kinase domain (683–952) and (v) a C-terminal tail (953–1186) (*45*). The structure of the intact EGFR has not been determined, but there is a wealth of available structural information for different regions of EGFR from a number of X-ray crystallography and NMR studies. Here, a doubly liganded, intact EGFR (996 amino acids) in an active dimeric form was generated using MODELLER (automodel function with multiple templates) (*176*). Firstly, the TM and intracellular kinase domains were modelled separately from the ectodomain. The asymmetric kinase domain dimer was modelled using multiple templates (PDB: 2GS6, 3GOP, 2JIU in Figure 5-1A,B) (*55*, *56*, *177*). The TM domain, the antiparallel LRRLL dimer within the JM-A region, and the rest of the JM region were modelled based on the NMR structure (PDB: 2M20) (*57*). The JM-B region, which is known to form a "juxtamembrane latch", was modelled by superimposing the 3GOP structure onto the NMR structure of the TM-JM domain. The TM-JM-kinase domain was then connected to the ectodomain using the crystal structure of the extracellular domain (PDB: 3NJP) (*65*) with the bound ligands to obtain a model of the full-length EGFR receptor (Figure 5-2). The resulting homology models were evaluated using the MODELLER molpdf and DOPE scoring methods, and the best model out of 20 models was selected for further refinement and characterization using MolProbity analysis (*178*). Energy minimizations using the steepest descent method in a vacuum were performed on the homology model prior to its subsequent use in the MD simulations.

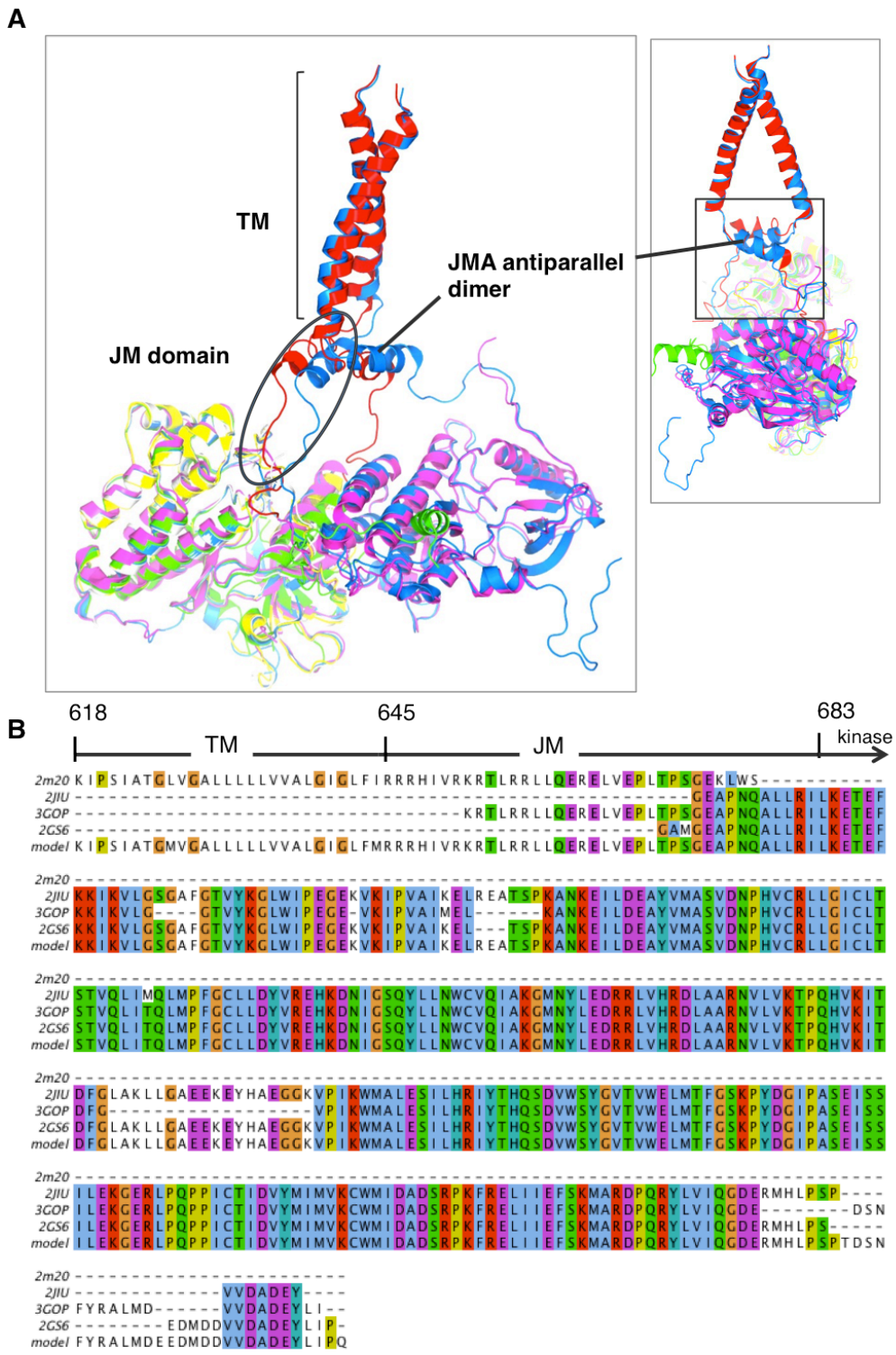


Figure 5-1. A) Superimposition of multiple structures used to construct the TM-kinase domain (PDB: 2M20 = red, 3GOP = green, 2JIU = magenta, 2GS6 = yellow, TM-kinase model = blue). On the right is the same picture rotated 90°. B) Alignment of the multiple sequences used to model the TM-kinase domain. The alignment was performed in Jalview with manual editing.

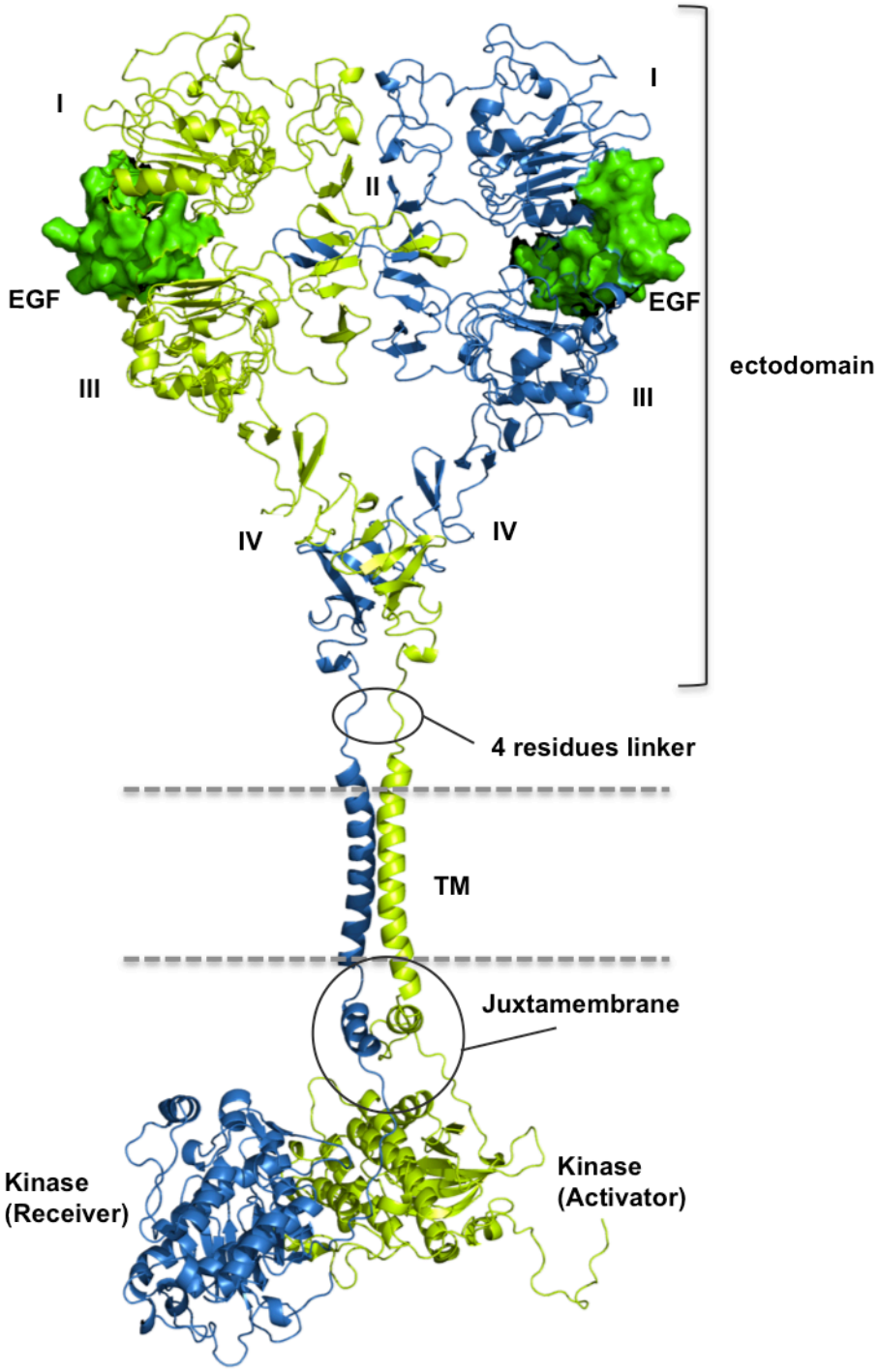


Figure 5-2. Cartoon model of the intact EGFR complexed with two EGF ligands, as shown in surface representation (green). Extracellular domains I, II, III and IV are labelled. Chains A and B are coloured blue and green, respectively. The approximate location of the membrane bilayer is marked by dotted grey lines.

5.3 SIMULATION SETUP

The model of the EGFR active asymmetric dimer was converted to a CG representation suitable for the MARTINI force field (v2.1) (*115*), and a elastic network model (ENM)

was employed to maintain its secondary structure. To mimic the lipid composition in a typical plasma membrane, simulations were performed in an asymmetric bilayer comprising POPC, POPE, POPS, sphingomyelin (SM), PIP_2, GM3 and CHOL which was constructed from a pure POPC bilayer using a lipid exchange protocol (*167*).. The upper leaflet consisted of POPC:POPE:SM:CHOL:GM3 (40:10:15:25:10) and the lower leaflet consisted of POPC:POPE:POPS:PIP_2:CHOL (10:40:15:25:10). The overall composition of the seven lipid types (POPC:POPE:POPS:SM:PIP_2:GM3:CHOL) was 25:25:7.5:5:7.5:5:25. The pure POPC bilayer was provided by Heidi Koldsoe (*167*). The protein was inserted manually into the pure POPC bilayer (Figure 5-3A). After protein insertion, other lipid species were incorporated into the bilayer at random positions (Figure 5-3B). The position of the TM domain in the bilayer was based on the position of the TM-JM domain in the bilayer in previous simulations (Chapter 4). Energy minimizations and equilibrations were performed to allow the lipids to relax around the protein. The equilibrations were carried for 2 to 5 ns, and the backbone particles of the protein were restrained to allow lipids to equilibrate around protein. Then, the restraints were removed to allow the protein to adopt a stable position within the membrane for 10 ns (Figure 5-3C). The simulations were performed with only one EGFR molecule in a bilayer of approximate dimensions 165 x 165 $Å^2$ containing a total of 1,219 lipids (POPC = 355, POPE = 267, POPS = 80, SM = 138, GM3 = 55, PIP_2 = 53, CHOL = 271). MDanalysis (*179*) and in-house scripts were used for the analysis (*167*, *180*).

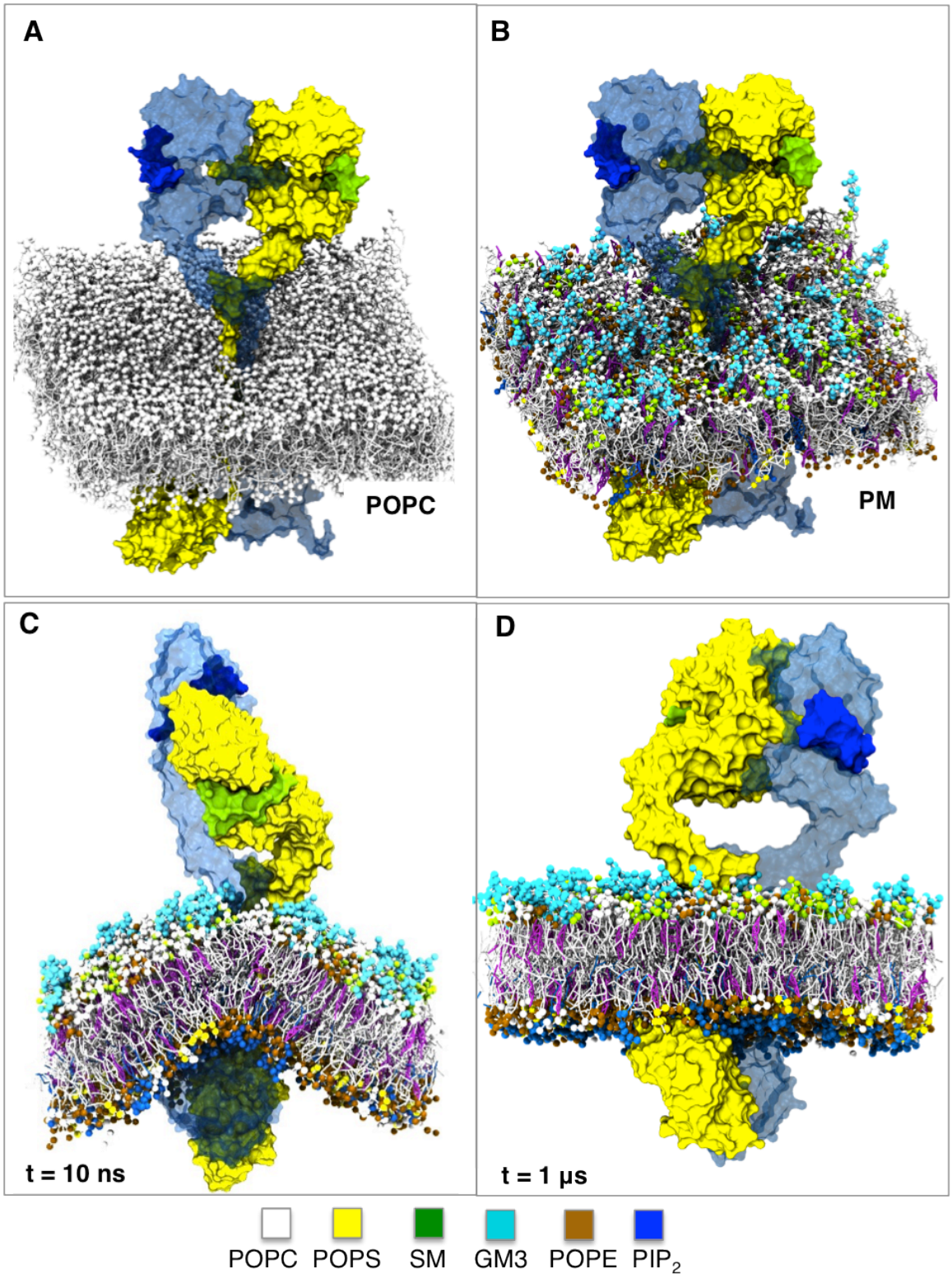


Figure 5-3. A) Surface representation of the CG model of EGFR. The protein was manually inserted into a pure POPC bilayer. The TM domain position in the bilayer was approximated by aligning it with the TM domain from previous simulations using VMD. B) A snapshot of the asymmetric bilayer after other lipid species were exchanged with POPC. C) A snapshot of the EGFR in the asymmetric bilayer during bilayer equilibration. D) A snapshot of the system at 1 µs during the production run. The figure was produced with VMD (*133*). The colour scheme for the lipids is shown at the bottom of the figure.

5.4 BEHAVIOUR OF INTACT EGFR IN A COMPLEX LIPID BILAYER

Analysis of the CG-MD trajectories of the intact EGFR receptor revealed that the ectodomain approached and formed strong, i.e., long-lived interactions, with the bilayer. Ectodomain motions were expected due to the presence of a flexible linker connecting the ectodomain and TM domain (*63*). The movement of the ectodomain relative to the membrane can be depicted by the evolution of the distances between centre of mass of the N-terminal residue of EGF on each dimer subunit and the centre of mass of the phosphate groups of the bilayer. Figure 5-4 illustrates the movement of the ectodomain in one of the simulations (sim1), in which it approached the membrane at 2 μs and remained close to the bilayer during the remainder of the simulation. In the second (sim2) simulation, the ectodomain moved towards the bilayer quickly, at around 500 ns, and remained in close contact with the bilayer until the end of the simulation. The ligand-membrane distances observed during the simulation are in agreement with FLIM-FRET microscopy experiments performed by Tynan et al., in which they showed that the closest approaches of the ligand and membrane ranged from 3.0 to 4.0 nm and 6.9 to 7.0 nm for short and long distances, respectively (*66*). They also showed a high affinity, ligand-binding human EGFR conformation, which is consistent with a flat alignment of the ectodomain region on the plasma membrane, with a ligand membrane distance < 4 nm. The ligand-membrane distance of the ectodomain dimer standing upright, observed in crystal structure is about 10 nm. However, in the present study, the ectodomain only tilted and lied on one side of the dimer, and did not lie flat on the membrane (*67*). This is perhaps due to the implementation of the CG with ENM constraint (to maintain the overall structure) that may restrict conformational changes within the ectodomain (formation of an asymmetric ectodomain dimer), thereby allowing it to lie flat on membrane. The behaviour of the ectodomain is also consistent with previous AT-MD simulations by Arkhipov et al., in which the ectodomain was found to be flexible and often rested on the membrane (*95*). Overall, the behaviour of the ectodomain is in agreement with FRET measurements and previous MD studies (*67*, *68*, *95*).

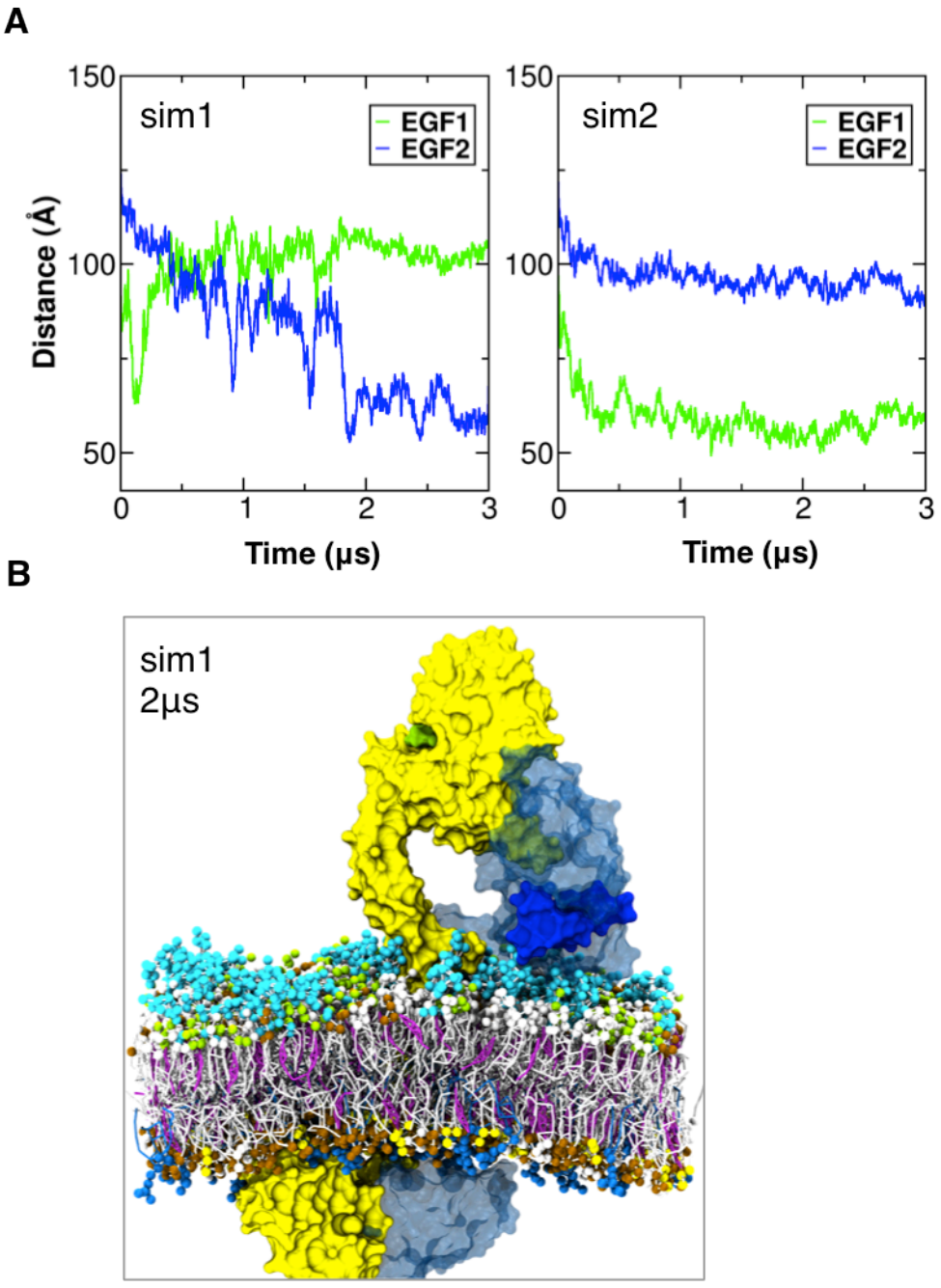


Figure 5-4. A) Evolution of distances between the COM of EGF on each monomer during the 3 μs simulation showing that the ectodomain moved towards bilayer, tilted, and rested on the bilayer. B) A snapshot of the receptor showing its resting position on the membrane surface at 2 μs in sim1. The protein and the ligands bound to their ligand binding sites are shown in a surface representation, and the lipids are shown in stick representations.

It is of interest to investigate the dynamics of the EGFR and lipids in a compositionally complex bilayer. In addition, the long-lived interaction between the tilted ectodomain and the bilayer signifies the presence of protein-lipid interactions that are likely to affect the behaviour of lipids around the protein. Thus, it is of interest to quantify the behaviour of the lipids by calculating the RDF of each lipid species around the TM domain. The RDF profiles revealed that three lipid species, namely PIP_2, GM3 and CHOL, formed favourable interactions with EGFR. In the upper leaflet, the interactions between EGFR and lipids were dominated by CHOL, followed by GM3 and POPE, as shown in the RDFs calculated for all lipids in each leaflet (see Figure 5-5A,B). SM and POPC in the upper leaflet had similar RDF profiles. Meanwhile, in the lower leaflet, the highest peak was observed for PIP_2 within an ~5 Å confinement from the TM domain. This was followed by a second peak at ~10 Å and a third peak at ~15 Å, suggesting that PIP_2 not only forms favourable interactions with the TM domain, but also interacts with the kinase domain. The peaks observed for CHOL were of the same height in the lower and upper leaflets. Meanwhile, the density distribution for POPS was much lower than that of PIP_2 and CHOL, although it was higher than that of POPE and POPC (Figure 5-5B). The average coordination numbers for POPC, POPS, SM, GM3, PIP_2 and CHOL within 10 Å cut-off were 0.28, 0.08, 0.09, 0.152, 0.42, 0.17 and 1.01, respectively.

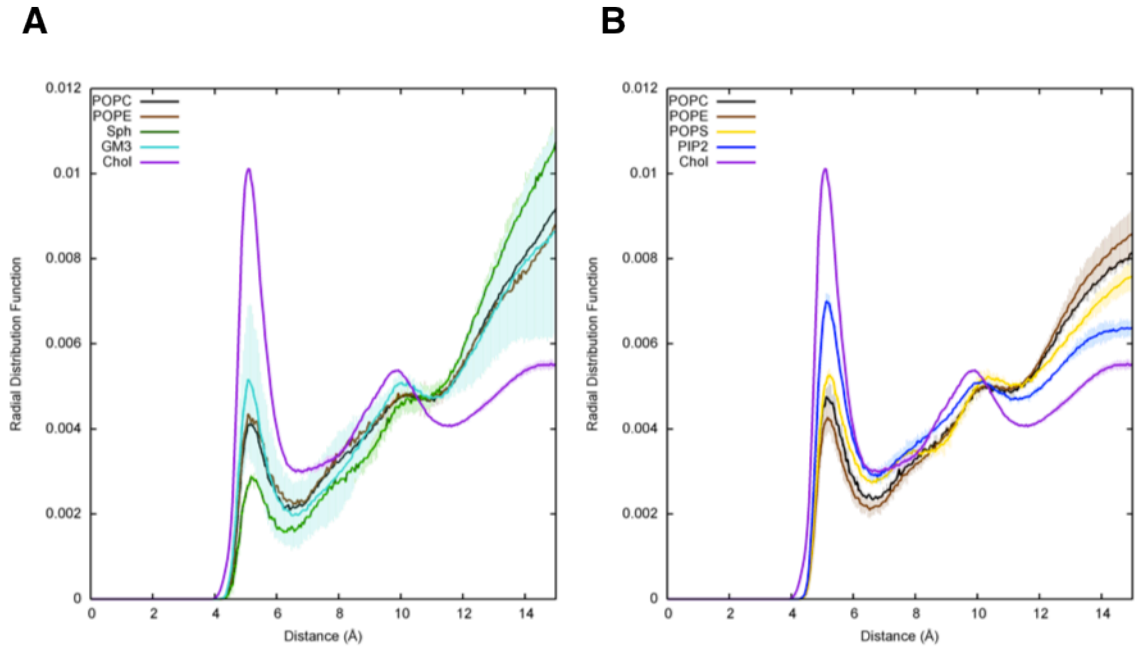


Figure 5-5. The calculated RDFs of the centre of mass of the lipid head groups of all lipid species in the system. B) The RDFs of lipids in the upper leaflet. C) The RDFs of lipids in the lower leaflet.

To investigate the organization of the different lipids around EGFR during the simulation, the time-averaged lipid densities around the protein were calculated. In the upper leaflet, several large, high-density clusters of GM3 were observed, suggesting the formation of GM3 nano-domains (Figure 5-6). The formation of high-density clusters also indicates strong interactions between the ectodomain and TM domain with membrane. This is an interesting observation, as GM3 has been shown to be involved in the regulation of EGFR through its inhibition of EGFR activity in phase-separated membrane enriched in CHOL and SM (*77*).

The density maps also illustrated a clustering of negatively charged lipids in the lower leaflet, particularly high-density and large patches of PIP_2 clusters formed around the protein, reflecting a strong interaction between PIP_2 and the EGFR intracellular domains. Several smaller, lower-density patches of POPS lipids were also found around the protein, as depicted in Figure 5-7. The clustering of these two anionic lipids was expected due to the presence of charged residues on the intracellular JM and kinase domains that have been suggested to be responsible for membrane interactions (*78*). Some clustering of POPE was also observed in both leaflets, with much higher-density clusters formed in the upper leaflet compared to the lower leaflet, which is consistent with its RDF profile. For POPC lipids, no significant clustering was observed. A clustering of CHOL around the TM domain was also observed in the lower leaflet near the TM domain, while the clustering was less prominent and more dispersed in the upper leaflet.

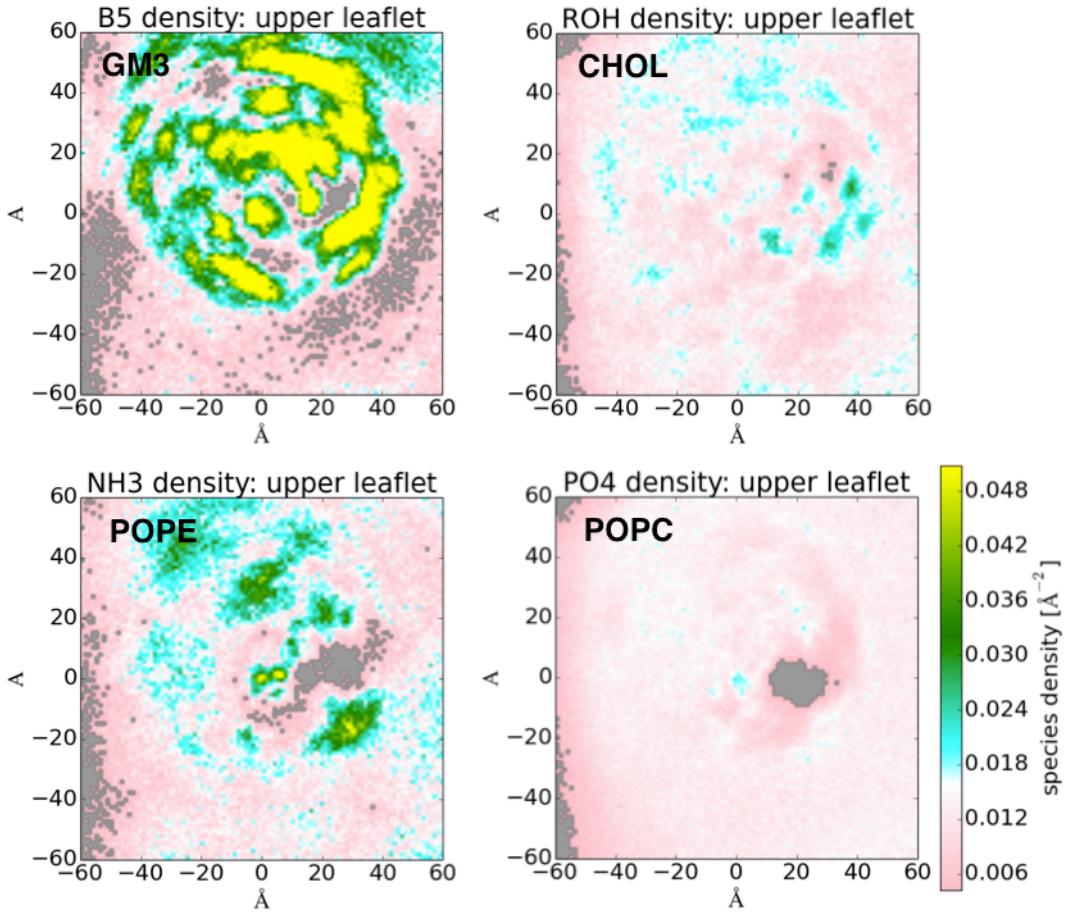


Figure 5-6. Time-averaged, two-dimentional head group particle densities around the TM domain of EGFR in the upper leaflet were obtained from concatenated trajectories of the last 1 μs of the two simulations.

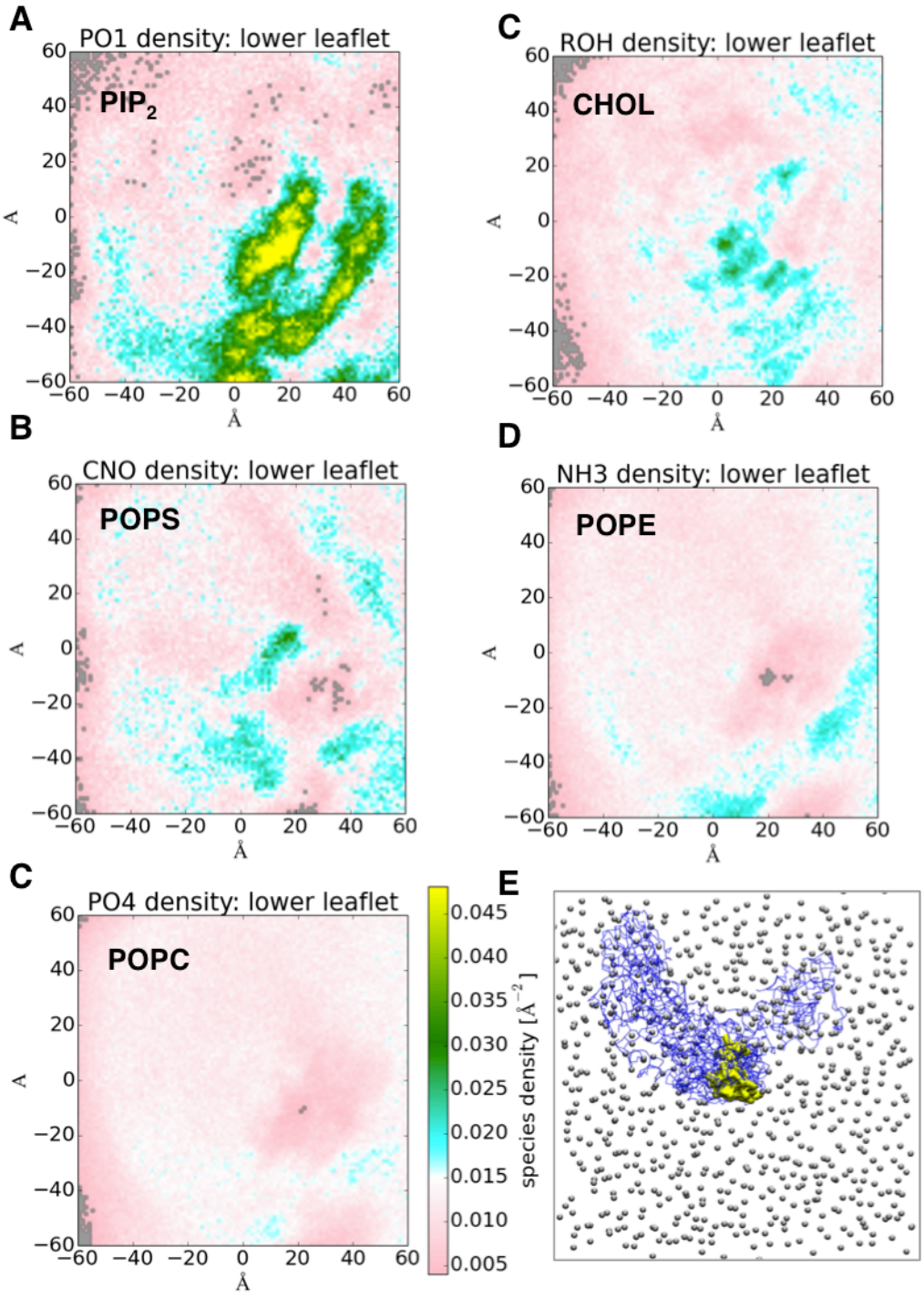


Figure 5-7. (A-D) Time-averaged, two-dimentional head group particle densities around the TM domain of EGFR in the lower leaflet were obtained from concatenated trajectories of the last 1 μs of the two simulations for all lipids in the system, as labelled. E) A snapshot of the protein from the top, with the Cα trace shown in blue and the TM shown as a surface representation in yellow.

5.5 PROTEIN-LIPID INTERACTIONS

To analyse the protein lipid-interactions, fractional protein-lipid contacts analysis within a 6-Å cut-off was performed for each of the different lipid species. The analyses revealed that GM3 lipids made extensive contacts with many residues of the ectodomain, as well as with residues at the N-terminal end of the TM domain. A snapshot of the intact EGFR receptor illustrating the clustering of GM3 near the TM domain and extracellular domains III and IV is shown in Figure 5-8. Previously, in a CG-MD simulation of the TM domain in a GM3-containing bilayer (Chapter 4), the GM3 head group components interacted strongly with residue K618. In the intact receptor, the GM3 lipids also interacted strongly with K618, and also with several other residues adjacent to K618, namely G606, P607, N615, G616, P617, I619 and T624 of the TM domain. Many amino acid residues of the ectodomain were also in contact with the head groups of GM3. This can be illustrated by mapping the fractional contacts between EGFR and GM3 to the EGFR structure (Figure 5-9).

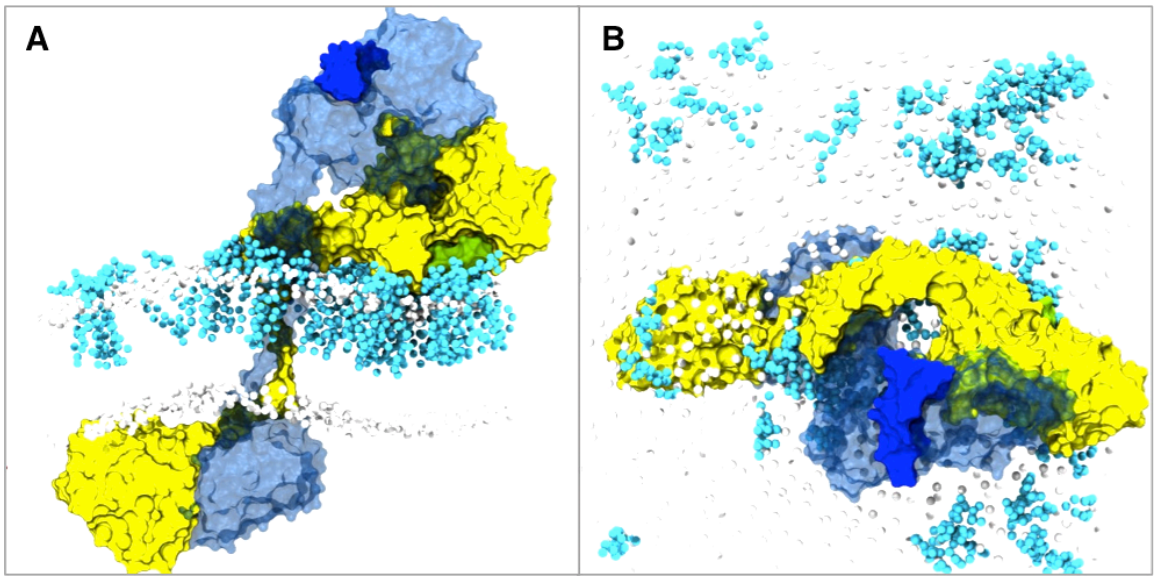


Figure 5-8. Sideview (A) and topview (B) of the trajectories showing the clustering of GM3 (shown in cyan) around the protein during the simulations. The phosphate groups of the other lipids are shown as white spheres.

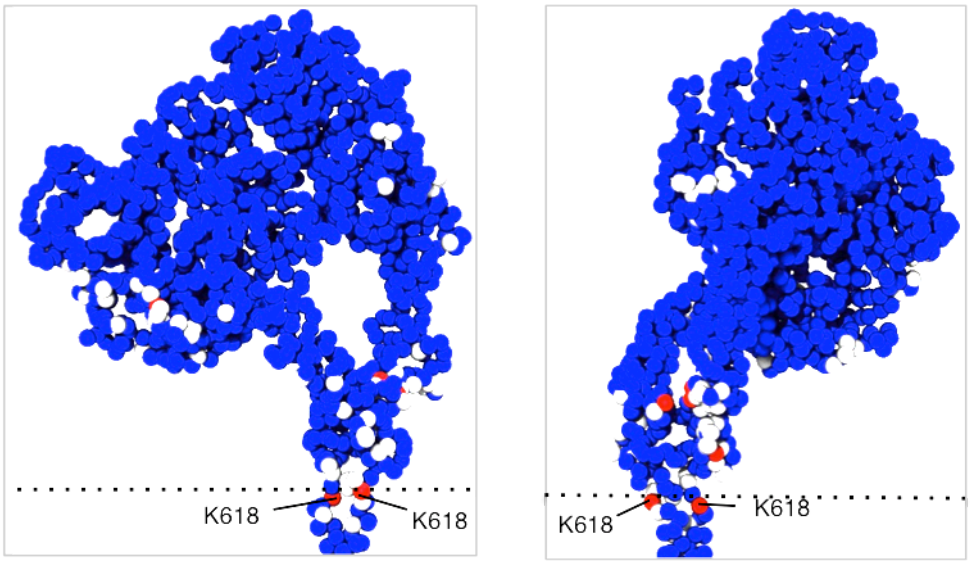


Figure 5-9. Interactions between GM3 head groups and EGFR within a 6-Å cut-off were mapped onto the initial structure. The Blue to red colour scale indicates no interaction (blue) or strong interactions (red). Note the strong interaction between the residues in the TM domain and domain IV with GM3 head groups. The figure on the right is rotated 90° rotation relative to the one the left. The bilayer is indicated by the dotted black line.

In the lower leaflet, high-density clustering of PIP_2 lipids was observed, suggesting strong interactions between PIP_2 and EGFR, which is consistent with the PIP_2 RDF profile presented in Figure 5-5. PIP_2 interacted with residues 645–650, 653–654, 656–657 and Q660 of the TM-JM simulation, with the highest number of contacts with the first two Arg residues of the JM-domain, R645 and R646 (Figure 5-10). This is in agreement with the study of Michailidis et al. and the previous observation from the CG-MD simulation of the TM-JM domain in the POPC:POPS:PIP_2 bilayer (Chapter 4). PIP_2 also interacted with residues 863–865, K905, 907–911, P935, S938, K925, Y954, R953 and K946 of the intracellular kinase domain. Clustering of POPS was also observed in the lower leaflet, albeit with lower probability than PIP_2. Clustering of POPE in the upper leaflet was due to lipid-lipid interactions, as this was not reflected in the contact analysis between POPE and the protein, whilst in the lower leaflet, the contact analysis revealed some interactions between POPE and TM, as well as with the kinase domains. CHOL interacted with the TM domain.

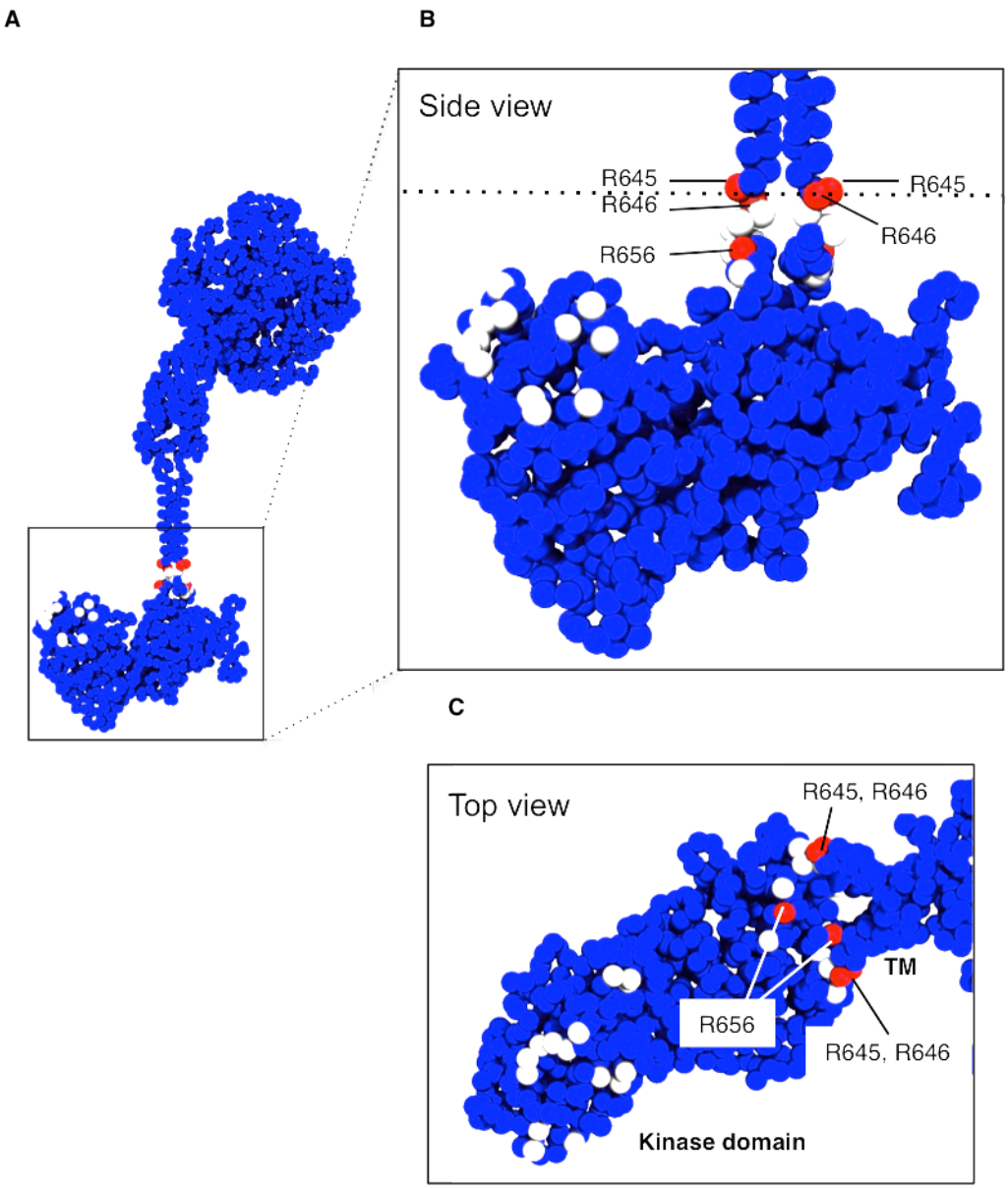


Figure 5-10. Interactions between PIP_2 head groups and EGFR within a 6-Å cut-off were mapped onto the CG structure of EGFR, shown as a sphere representation. The blue to red colour scale indicates no interaction (blue) or strong interactions (red). Note the strong interactions between PIP_2 and the basic residues in the JMA domain. (B,C) Sideview and topview of the PIP_2 head group-protein interactions. The bilayer is indicated by the dotted black line.

5.6 LIPID MOBILITY

To investigate the effect of the interactions with EGFR on the mobility of the different lipids in the complex membrane, the diffusion coefficients of all lipid species were calculated over the final 1 μs trajectories of the 2 x 3 μs trajectories. In general, the diffusion of lipids that were in close proximity to the protein was much lower than those that were far from the protein. Those lipids that formed strong contacts with the protein had much lower diffusion coefficients than freely moving lipids. In the simulation, the strongest interactions with EGFR were observed for GM3, PIP_2 and CHOL. The strong interactions between GM3 and EGFR were reflected by the diffusion coefficient of GM3 (2.13×10^{-7} cm^2/s), the lowest among all lipid species considered. The diffusion coefficient of PIP_2 was 2.28×10^{-7} cm^2/s, slightly higher than that of GM3, but lower than that of POPS lipids (2.87×10^{-7} cm^2/s). The diffusion coefficients for POPE, POPC, CHOL and SM were, in the following order, POPC<POPE<CHOL<SM. SM was the fastest moving lipid, with a diffusion coefficient of 3.82×10^{-7} cm^2/s, which is consistent with its lowest density distribution around EGFR, as revealed by the RDF, thereby suggesting that it did not interact strongly with EGFR.

The mobility of annular lipids showed a dependence on the distance from the protein. For example, the diffusion coefficient of POPC was 1.73×10^{-7} cm^2/s at a 0–10 Å annulus from the protein, and this value increased at 2030 Å to ~2.2×10^{-7} cm^2/s and began to plateau at 30 Å or more. The diffusion coefficients of GM3 and PIP_2 below 30 Å were much reduced compared with other lipid species (see Figure 5-11). Comparing the diffusion coefficients of lipid species found in both upper and lower leaflets revealed that the diffusion coefficients differed in different leaflets. This is possibly due to lipid-lipid (lipid nano-domain formation) or protein-lipid interactions adjusted to accommodate the shape and nature (i.e., the distribution of charged residues) of the protein domains. The mobility of POPC was slightly higher in the upper leaflet ($D = 2.94 \times 10^{-7}$ cm^2/s) than in the lower leaflet ($D = 2.82 \times 10^{-7}$ cm^2/s). Similarly, the diffusion coefficient of POPE was much higher in the upper leaflet than in the lower leaflet, with the value of 3.86×10^{-7} cm^2/s for the former and 2.88×10^{-7} cm^2/s for the latter. The difference in the diffusion coefficient of POPE may be due to a sampling issue, as there was more POPE in the lower leaflet (40%) compared with the upper leaflet (10%). The diffusion coefficients converged more quickly in leaflets highly populated with POPE or POPC lipids (see Figure 5-12).

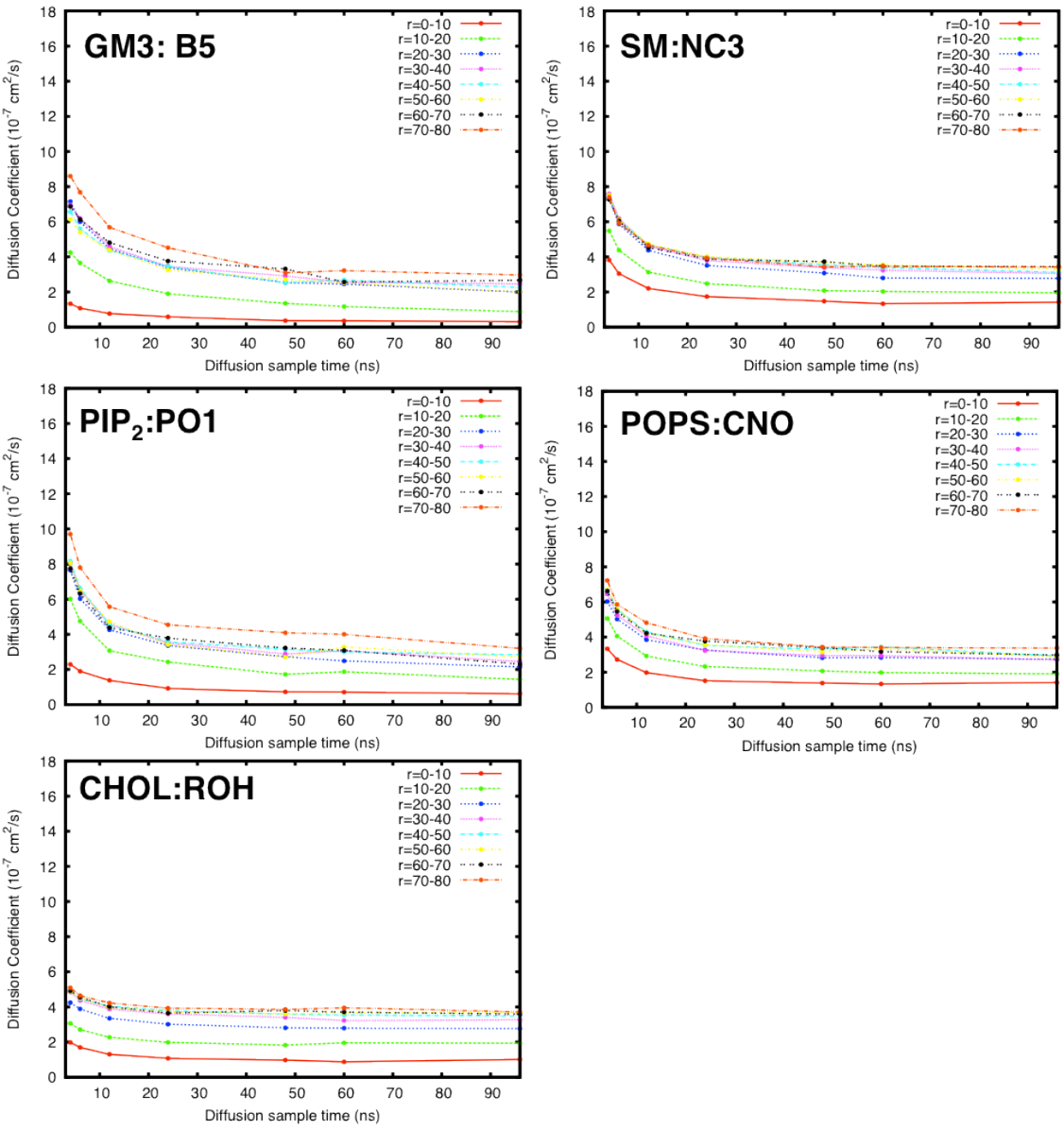


Figure 5-11. The lipid diffusion coefficients of the lipid head groups as a function of distance from the protein and of observation times for GM3, SM, PIP_2, POPS and CHOL.

CHOL, conversely, is able to flip-flop between leaflets and, despite having a high-density distribution close to the protein, as judge by the RDF, its mobility was much higher than those of PIP_2 and GM3. The diffusion coefficient for CHOL was more or less similar in both leaflets, suggesting that it mainly interacts with the TM region and its mobility was not significantly affected by the presence of the extra- or intracellular domains.

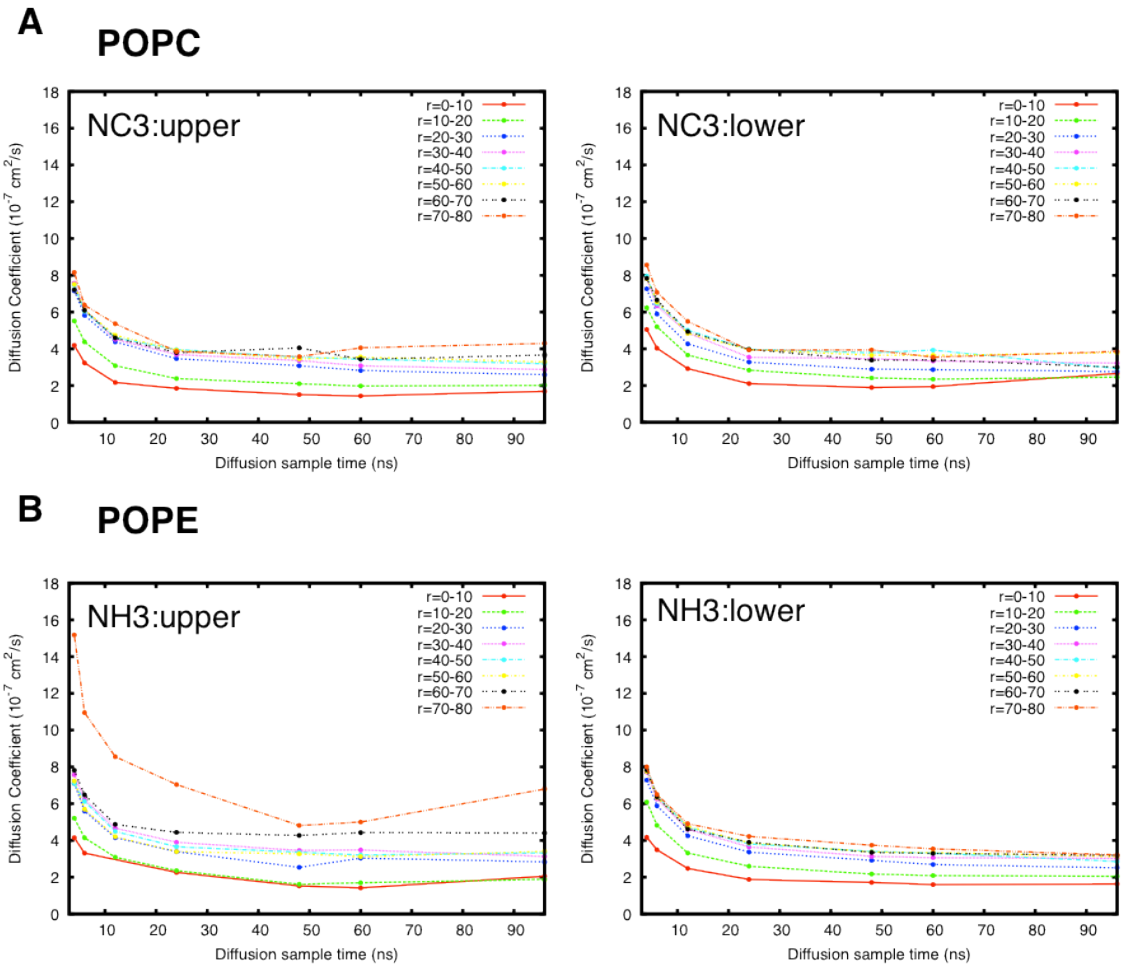


Figure 5-12. The diffusion coefficients for the upper and lower leaflets of the bilayer as a function of distance and observation times for POPC (A) and POPE head groups (B).

5.7 DISCUSSION

A doubly-liganded, active asymmetric dimer model of EGFR was simulated in an asymmetric bilayer that mimicked the lipid composition of the mammalian cell plasma membrane. Note that the receptor was held in an active conformation due to the implementation of a CG model plus ENM to maintain its secondary structural elements. However, the aim here was to investigate the dynamic behaviour of intact EGFR and protein-lipid interactions in a physiologically relevant bilayer. The lipid environment has been proposed to play an important role in the regulation of EGFR activity, and EGFR is known to localize to lipid raft, although it is also activated outside lipid rafts (*181*). However, the mechanistic details regarding EGFR activation are still not clear.

The most important finding of the simulation of intact EGFR presented in this chapter is that GM3, PIP_2 and CHOL formed strong interactions with EGFR. This was reflected by clustering of the GM3 and PIP_2 lipids, which is also indicative of the formation of nano-clusters. However, phase-separation was not observed in the membrane, and this was expected, as modification of the lipids is required to induce phase-separation, as reported in the MD studies of model membranes (*168*, *182*, *183*). Additionally, it probably also reflects the size of the system.

Comparison of the human EGFR structure with its *Drosophila* counterpart has revealed an asymmetric structure of the *Drosophila* ectodomain that has never been observed in the crystal structure of the isolated human EGFR (*48*, *64*). This asymmetry of the ectodomain also exists in the human EGFR, and this was observed during AT-MD simulations (*67*). Although, in this study, the ectodomain did not lie flat on the membrane surface as shown earlier in the MD simulations of Kastner et al., strong interactions were observed between one of the dimer subunits and the membrane due to interactions with GM3. The CG-MD simulations of the intact receptor in the present study also captured strong interactions between the membrane-proximal residues, including one (K618) that was shown to be involved in direct interaction with GM3. Previously, Lingwood et al. suggested that inactive, tethered monomers of EGFR could preferentially associate in a dynamic fashion with nanoscale lipid rafts (*184*).

Another interesting result from this study is that extensive interactions between anionic lipids, particularly between PIP_2, and the JM-A dimer confirm the important role of this lipid in the regulation of EGFR activity by stabilizing the antiparallel orientation of the JM-A, as previously suggested by Jura et al. (*54*). This is also in accordance with previous observations of TM-JM simulations (Chapter 4). These strong interactions in the contextt of full-length receptor may help stabilize the JM-A antiparallel dimer, and this, in turn, may help to stabilize the asymmetric kinase dimer. This also helps explain why EGFR activity (phosphorylation) is completely abolished in the absence of PIP_2 (*78*). In the present simulation, it was observed that PIP_2 also interacted with some residues of the kinase domain. This is not entirely surprising, as the active asymmetric dimer also interacted, albeit less extensively, with membranes containing POPS in AT-MD simulations by Arkhipov et al. (*95*). Because only the active asymmetric conformation was used in this study, the extent of membrane interactions between the asymmetric active and symmetric inactive kinase cannot be compared here.

EGFR regulation by membrane lipids occurs in a CHOL-dependent manner, as the depletion of CHOL stimulates EGFR activation (*81*). Thus, the high CHOL density in the vicinity of the TM domain that was observed in the simulations reflects its importance in EGFR function. In addition to the clustering of lipids, the mobility of these lipids was also affected by lipid clustering, as well as by protein-lipid and lipid-lipid interactions, as judged by the reduction in their diffusion rates. This effect was most obvious in the diffusion rates of GM3 and PIP_2, which were in agreement with those of a previous study that demonstrated a reduced mobility of lipids in crowded membranes (*180*).

In summary, the CG-MD simulations of the intact receptor presented here appear to capture protein-lipid interactions, particularly the interaction between the membrane-proximal residues of the ectodomain and intracellular domain that are known to be important in the regulation of EGFR activity. The CG-MD simulations, however, are unlikely to capture any conformational transitions and,, thus AT-MD simulations are required to investigate how the interaction with lipids contributes to the conformational coupling between different EGFR domains. Additionally, in future investigations, the C-terminal tail could also be included to investigate the behaviour of the full-length EGFR in complex membranes.

CHAPTER 6

Summary and conclusions

A preliminary study was conducted to explore the dimerization of the TM domains of selected RTKs in simple membranes consisting of neutral, pure DPPC bilayers using a CG-MD high-throughput approach in combination with AT-MD. The results from the preliminary study sparked an interest for a more detailed study of one of the candidates from the preliminary study, EGFR, a relatively well-studied member of ErbB family protein. When the simulations were conducted in 2010, the structure of the TM domain dimer of EGFR was not yet available; the only structure of the TM domain available was that of the EGFR/ErbB2 heterodimer (PDB: 2KS1). Thus, the results presented in Chapter 3 showed that CG-MD is a valuable tool for predicting dimerization, as evidenced by the good agreement of the data with the NMR structure, which was solved in 2013. Other studies have shown that this method is sufficient to capture the nature of protein-protein associations (e.g., dimerization) in simplified systems, and is sensitive to small perturbations (e.g., mutations and deletions) (*91*, *93*). Subsequent atomistic simulations allow further refinement of the CG generated models.

In 2011, several papers highlighted the role of lipids in the regulation of EGFR, further supporting previous studies regarding the importance and significant involvement of membrane-proximal extracellular residues (e.g., K642-GM3, JM-PIP_2) in the activation of the receptor via interactions with the membrane (*8*, *77*, *78*). These studies motivated further exploration of the behaviour of the EGFR TM domain in GM3- and PIP_2-containing membranes, and these results were discussed in Chapter 4. The investigation of the GM3-EGFR interaction was performed using CG approach in GM3-containing bilayers. The simulation was designed to mimic the experimental data in which the author observed a direct interaction between GM3 and the membrane-proximal residue K618 only in phase-separated membrane of giant unilamellar vesicles (*77*). The CG-MD simulations revealed that GM3 interacted favourably with K618, and has the potential to destabilize the helix-helix association. Meanwhile, in Ld bilayers (POPC and POPC:GM3) or in the simulation of the K618G mutant, the TM dimer packing remained stable throughout the simulation. Thus, the CG-MD simulations provide an insight into the effect of the lipid environment on the packing of the EGFR TM domain dimer.

The NMR structure of the TM-JM domain, which became available in 2013 (*57*), provides a good starting point to investigate the JM domain interaction using CG-MD, followed by AT-MD. The results demonstrate that the JM domain indeed interacted strongly with PIP_2 in a PIP_2-containing membrane. The strong interaction between the JM region and anionic lipids (PIP_2, POPS) was shown to stabilize the TM-JM-A domain, in which the JM-A domain interacted in an antiparallel manner throughout the 5 μs CG-MD simulation, and was further stabilized throughout the 100 ns AT-MD simulation. Mutations that disrupt the electrostatic interactions between the JM domain and the membrane were shown to alter the lipid organization around the EGFR TM domain, which resulted in the dissociation of the antiparallel JM-A dimer, thereby causing it to interact with the membrane. Interestingly, depletion of PIP_2 (POPC:POPS membrane) destabilized the TM-JM-A dimer and resulted in the partial or full deforming of the JM-A-antiparallel orientation, which resulted in its subsequent embedment in the membrane. This explains why the depletion of PIP_2 in vivo abolishes EGFR phosphorylation and, hence, reduces receptor activity.

More recent studies have indicated that the activation of EGFR is not as simple as previously thought. It involves coupling of the different domains and the specific association and orientation of the TM-JM and kinase domains (*54*, *57*, *71*). How the movements of these different domains are coupled is not completely understood. It was suggested that the lipid environment also has a significant impact on EGFR activity. Thus, to fully understand the mechanism of activation of the receptor, it is crucial to consider the intact receptor in a more physiologically relevant bilayer. This motivated another investigation, and added another level of complexity (i.e., bilayer asymmetry and complex lipid mixtures). Additionally, an intact EGFR was modelled to gain more insights into EGFR protein-lipid and lipid-lipid interactions.

Prior to the publication of the TM-JM domain structure in 2013, there were abundant crystallographic structures of the intracellular domain in active/inactive conformations (*43*, *52–54*, *173*). The structures of the ectodomain in monomeric tethered and extended forms and of the two-liganded dimer are also available (*65*). Prior to the study presented in Chapter 5, a preliminary study was conducted using a homology model of the intact EGFR, with the TM domain modelled using the EGFR/ErbB2 heterodimer structure (data not shown in this thesis) in a simple membrane. The initial simulations gave an idea of the optimal system configuration and setup. Fortunately, the TM-JM NMR structure was

published in good timing with this effort, which allowed the construction of the nearly intact EGFR receptor.

The CG-MD simulations of the intact EGFR in asymmetric bilayer comprising POPC, POPE, POPS, PIP_2, GM3, SM and CHOL revealed favourable interactions with GM3, PIP_2 and CHOL. On the extracellular side, strong interactions were observed between GM3 and the ectodomain throughout the simulation. Previously, Kastner et al. showed experimentally (using FRET microscopy) that the ectodomain aligns flat on the membrane surface, which corresponded to a high-affinity EGFR, whilst the ectodomain in an upright conformation corresponded to a low affinity EGFR (*67*, *68*). They further suggested the ectodomain-membrane interaction would help stabilize the high-affinity liganded dimer on the membrane surface. The results presented in Chapter 5 revealed favourable interactions between the ectodomain and GM3, suggesting that it may play a role in stabilizing EGFR in inactive state within a raft environment through protein-lipid interactions (*67*, *68*). Note that glycosylation of the ectodomain has not been considered in this study, despite its importance in EGFR function and interaction with the ganglioside GM3. The intracellular kinase domain also preferentially associates with PIP_2 and POPS. The strong interactions between anionic lipids and the intact EGFR support the important role of the intracellular kinase-membrane interactions on EGFR function. This study, however, was unable to capture any conformational changes or large movements of these domains, with respect to each other, due to the implementation of CG with ENM to preserve secondary structure.

To study the protein-lipid interactions in more detail, lipid-lipid interactions, as well as the conformational dynamics of the different EGFR domains in an intact receptor, an all-atom MD simulation was required. The intact EGFR model in a complex bilayer presented in Chapter 5 could be converted back to an AT representation to further investigate the protein-lipid and lipid-lipid interactions. Of course, converting a CG system into an AT system consisting of ~600,000 particles will be computationally demanding and time-consuming. In addition, simulation of the intact EGFR in monomeric or higher order oligomers could also be performed to gain more insight into protein-lipid and protein-protein interactions in complex membranes.

This thesis has focused on EGFR, the most well studied of the ErbB receptors. This study could be extended to other ErbB family members that are suggested to adopt similar activation mechanisms, as well as to other RTKs (e.g., EphA).

Bibliography

1. Luckey, M. (2008) Membrane structural biology:with biochemical and biophysical foundation, Cambridge University Press.

2. Lemmon, M. A, and Schlessinger, J. (2010) Cell signaling by receptor tyrosine kinases., *Cell 141*, 1117–1134.

3. Harburger, D. S., and Calderwood, D. A. (2009) Integrin signalling at a glance., *Journal of Cell Science 122*, 159–163.

4. Ye, F., Kim, C., and Ginsberg, M. H. (2011) Molecular mechanism of inside-out integrin regulation., *Journal of Thrombosis and Haemostasis JTH*, *9 Suppl 1*, 20–25.

5. Choe, H. W., Park, J. H., Kim, Y. J., and Ernst, O. P. (2011) Transmembrane signaling by GPCRs: Insight from rhodopsin and opsin structures., *Neuropharmacology 60, 52-57*.

6. Sud, M., Fahy, E., Cotter, D., Brown, A., Dennis, E. A., Glass, C. K., Merrill, A. H., Murphy, R. C., Raetz, C. R. H., Russell, D. W., and Subramaniam, S. (2007) LMSD: LIPID MAPS structure database., *Nucleic Acids Research 35*, D527–D532.

7. Van Meer, G., Voelker, D. R., and Feigenson, G. W. (2008) Membrane lipids: where they are and how they behave., *Nature Reviews Molecular Cell Biology 9*, 112–124.

8. Coskun, Ü., and Simons, K. (2011) Cell membranes: the lipid perspective., *Structure 19*, 1543–1548.

9. Ikonen, E. (2008) Cellular cholesterol trafficking and compartmentalization., *Nature Reviews Molecular Cell Biology 9*, 125–138.

10. Singer, S. J., and Nicolson, G. L. (1972) The fluid mosaic model of the structure of cell membranes., *Science 175*, 720–731.

11. Simons, K., and Toomre, D. (2000) Lipid rafts and signal transduction., *Nature Reviews. Molecular Cell Biology 1*, 31–39.

12. Shevchenko, A., and Simons, K. (2010) Lipidomics: coming to grips with lipid diversity., *Nature Reviews Molecular Cell Biology 11*, 593–598.

13. Helms, J. B., and Zurzolo, C. (2004) Lipids as targeting signals: lipid rafts and intracellular trafficking., *Traffic 5*, 247–254.

14. Simons, K., and Ikonen, E. (1997) Functional rafts in cell membranes., *Nature 387*, 569–572.

15. Rajendran, L., and Simons, K. (2005) Lipid rafts and membrane dynamics., *Journal of Cell Science 118*, 1099–1102.

16. Levental, I., Grzybek, M., and Simons, K. (2011) Raft domains of variable properties and compositions in plasma membrane vesicles., *Proceedings of the National Academy of Sciences of the United States of America 108*, 11411–11416.

17. Munro, S. (2003) Lipid rafts: elusive or illusive?., *Cell 115*, 377–388.

18. Simons, K., and Sampaio, J. L. (2011) Membrane organization and lipid rafts., *Cold Spring Harbor perspectives in biology 3*, a004697.

19. Owen, D. M., Magenau, A., Williamson, D., and Gaus, K. (2012) The lipid raft hypothesis revisited--new insights on raft composition and function from super-resolution fluorescence microscopy., *BioEssays : news and reviews in molecular, cellular and developmental biology 34*, 739–747.

20. Zachowski, A. (1993) Phospholipids in animal eukaryotic membranes: transverse asymmetry and movement., *Biochemical Journal 294*, 1–14.

21. Krogh, A., Larsson, B., von Heijne, G., and Sonnhammer, E. L. (2001) Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes., *Journal of Molecular Biology 305*, 567–580.

22. Wallin, E., and von Heijne, G. (1998) Genome-wide analysis of integral membrane proteins from eubacterial, archaean, and eukaryotic organisms., *Protein Science 7*, 1029–1038.

23. Popot, J. L., and Engelman, D. M. (2000) Helical membrane protein folding, stability, and evolution., *Annual Review of Biochemistry 69*, 881–922.

24. MacKenzie, K. R., Prestegard, J. H., and Engelman, D. M. (1997) A transmembrane helix dimer: structure and implications., *Science 276*, 131–133.

25. Li, R., Gorelik, R., Nanda, V., Law, P. B., Lear, J. D., DeGrado, W. F., and Bennett, J. S. (2004) Dimerization of the transmembrane domain of Integrin alphaIIb subunit in cell membranes., *Journal of Biological Chemistry 279*, 26666–26673.

26. Dews, I. C., and Mackenzie, K. R. (2007) Transmembrane domains of the syndecan family of growth factor coreceptors display a hierarchy of homotypic and heterotypic interactions., *Proceedings of the National Academy of Sciences of the United States of America 104*, 20782–20787.

27. Lawrie, C. M., Sulistijo, E. S., and Mackenzie, K. R. (2010) Intermonomer hydrogen bonds enhance GxxxG-driven dimerization of the BNIP3 transmembrane domain : roles for sequence context in helix – helix association in membranes., *Journal of Molecular Biology 396*, 924–936.

28. Finger, C., Escher, C., and Schneider, D. (2009) The single transmembrane domains of human receptor tyrosine kinases encode self-interactions., *Science Signaling 2*, ra56.

29. Senes, A., Ubarretxena-Belandia, I., and Engelman, D. M. (2001) The calpha --- H...O hydrogen bond: a determinant of stability and specificity in transmembrane helix interactions., *Proceedings of the National Academy of Sciences of the United States of America 98*, 9056–9061.

30. Matsushita, C., Tamagaki, H., Miyazawa, Y., Aimoto, S., Smith, S. O., and Sato, T. (2013) Transmembrane helix orientation influences membrane binding of the intracellular juxtamembrane domain in Neu receptor peptides., *Proceedings of the National Academy of Sciences of the United States of America 110*, 1646–1651.

31. Muhle-Goll, C., Hoffmann, S., Afonin, S., Grage, S. L., Polyansky, A. A., Windisch, D., Zeitler, M., Bürck, J., and Ulrich, A. S. (2012) Hydrophobic matching controls the tilt and stability of the dimeric platelet-derived growth factor receptor (PDGFR) β transmembrane segment., *Journal of Biological Chemistry 287*, 26178–26186.

32. Li, E., You, M., and Hristova, K. (2006) FGFR3 dimer stabilization due to a single amino acid pathogenic mutation., *Journal of Molecular Biology 356*, 600–612.

33. Bennasroune, A., Fickova, M., Gardin, A., Dirrig-Grosch, S., Aunis, D., Crémel, G., and Hubert, P. (2004) Transmembrane peptides as inhibitors of ErbB receptor signaling., *Molecular Biology of The Cell 15*, 3464–3474.

34. Klemm, J. D., Schreiber, S. L., and Crabtree, G. R. (1998) Dimerization as a regulatory mechanism in signal transduction., *Annual Review of Immunology 16*, 569–592.

35. Schlessinger, J. (2002) Dimerization and activation of EGF receptor., *Receptor 110*, 669–672.

36. Robinson, D. R., Wu, Y. M., and Lin, S. F. (2000) The protein tyrosine kinase family of the human genome., *Oncogene 19*, 5548–5557.

37. Bordag, N., and Keller, S. (2010) Alpha-helical transmembrane peptides: a “divide and conquer” approach to membrane proteins., *Chemistry and Physics of Lipids*, *163*, 1–26.

38. Sarabipour, S., and Hristova, K. (2013) FGFR3 transmembrane domain interactions persist in the presence of its extracellular domain., *Biophysical Journal 105*, 165–171.

39. Schneider, M. R., and Wolf, E. (2009) The epidermal growth factor receptor ligands at a glance., *Journal of Cellular Physiology 218*, 460–466.

40. Zaczek, A., Brandt, B., and Bielawski, K. P. (2005) The diverse signaling network of EGFR, HER2, HER3 and HER4 tyrosine kinase receptors and the consequences for therapeutic approaches., *Histology and Histopathology 20*, 1005–1015.

41. Nyati, M. K., Morgan, M. A., Feng, F. Y., and Lawrence, T. S. (2006) Integration of EGFR inhibitors with radiochemotherapy., *Nature Reviews. Cancer 6*, 876–885.

42. Normanno, N., De Luca, A., Bianco, C., Strizzi, L., Mancino, M., Maiello, M. R., Carotenuto, A., De Feo, G., Caponigro, F., and Salomon, D. S. (2006) Epidermal growth factor receptor (EGFR) signaling in cancer., *Gene 366*, 2–16.

43. Yarden, Y., and Pines, G. (2012) The ERBB network: at last, cancer therapy meets systems biology., *Nature Reviews. Cancer 12*, 553–563.

44. Gullick, W. J., Downward, J., Parker, P. J., Whittle, N., Kris, R., Schlessinger, J., Ullrich, A., and Waterfield, M. D. (1985) The structure and function of the epidermal growth factor receptor studied by using antisynthetic peptide antibodies., *Proceedings of the Royal Society of London B 226*, 127–134.

45. Ferguson, K. M. (2008) Structure-based view of epidermal growth factor receptor regulation., *Annual Review of Biophysics 37*, 353–373.

46. Garrett, T. P. J., McKern, N. M., Lou, M., Elleman, T. C., Adams, T. E., Lovrecz, G. O., Kofler, M., Jorissen, R. N., Nice, E. C., Burgess, A. W., and Ward, C. W. (2003) The crystal structure of a truncated ErbB2 ectodomain reveals an active conformation, poised to interact with other ErbB receptors., *Molecular Cell 11*, 495–505.

47. Garrett, T. P. J., McKern, N. M., Lou, M., Elleman, T. C., Adams, T. E., Lovrecz, G. O., Zhu, H.-J., Walker, F., Frenkel, M. J., Hoyne, P. A., Jorissen, R. N., Nice, E. C., Burgess, A. W., and Ward, C. W. (2002) Crystal structure of a truncated epidermal growth factor receptor extracellular domain bound to Transforming Growth Factor α., *Cell 110*, 763–773.

48. Ogiso, H., Ishitani, R., Nureki, O., Fukai, S., Yamanaka, M., Kim, J.-H., Saito, K., Sakamoto, A., Inoue, M., Shirouzu, M., and Yokoyama, S. (2002) Crystal structure of the complex of human epidermal growth factor and receptor extracellular domains., *Cell 110*, 775–787.

49. Ferguson, K. M., Berger, M. B., Mendrola, J. M., Cho, H.-S., Leahy, D. J., and Lemmon, M. A. (2003) EGF activates its receptor by removing interactions that autoinhibit ectodomain dimerization., *Molecular Cell 11*, 507–517.

50. Cho, H.-S., Mason, K., Ramyar, K. X., Stanley, A. M., Gabelli, S. B., Denney, D. W., and Leahy, D. J. (2003) Structure of the extracellular region of HER2 alone and in complex with the Herceptin Fab., *Nature 421*, 756–760.

51. Greenfield, C., Hiles, I., Waterfield, M. D., Federwisch, M., Wollmer, A., Blundell, T. L., and McDonald, N. (1989) Epidermal growth factor binding induces a conformational change in the external domain of its receptor., *The EMBO Journal 8*, 4115–4123.

52. Bouyain, S., Longo, P. A., Li, S., Ferguson, K. M., and Leahy, D. J. (2005) The extracellular region of ErbB4 adopts a tethered conformation in the absence of ligand., *Proceedings of the National Academy of Sciences of the United States of America 102*, 15024–15029.

53. Elleman, T. C., Domagala, T., McKern, N. M., Nerrie, M., Lönnqvist, B., Adams, T. E., Lewis, J., Lovrecz, G. O., Hoyne, P. a, Richards, K. M., Howlett, G. J., Rothacker, J., Jorissen, R. N., Lou, M., Garrett, T. P., Burgess, a W., Nice, E. C., and Ward, C. W. (2001) Identification of a determinant of epidermal growth factor receptor ligand-binding specificity using a truncated, high-affinity form of the ectodomain., *Biochemistry 40*, 8930–8939.

54. Jura, N., Endres, N. F., Engel, K., Deindl, S., Das, R., Lamers, M. H., Wemmer, D. E., Zhang, X., and Kuriyan, J. (2009) Mechanism for activation of the EGF receptor catalytic domain by the juxtamembrane segment., *Cell 137*, 1293–1307.

55. Zhang, X., Gureasko, J., Shen, K., Cole, P. a, and Kuriyan, J. (2006) An allosteric mechanism for activation of the kinase domain of epidermal growth factor receptor., *Cell 125*, 1137–1149.

56. Red Brewer, M., Choi, S. H., Alvarado, D., Moravcevic, K., Pozzi, A., Lemmon, M. A., and Carpenter, G. (2009) The juxtamembrane region of the EGF receptor functions as an activation domain., *Molecular Cell 34*, 641–651.

57. Endres, N. F., Das, R., Smith, A. W., Arkhipov, A., Kovacs, E., Huang, Y., Pelton, J. G., Shan, Y., Shaw, D. E., Wemmer, D. E., Groves, J. T., and Kuriyan, J. (2013) Conformational coupling across the plasma membrane in activation of the EGF receptor., *Cell 152*, 543–556.

58. Holland, S. J., Gale, N. W., Gish, G. D., Roth, R. a, Songyang, Z., Cantley, L. C., Henkemeyer, M., Yancopoulos, G. D., and Pawson, T. (1997) Juxtamembrane tyrosine residues couple the Eph family receptor EphB2/Nuk to specific SH2 domain proteins in neuronal cells., *The EMBO Journal 16*, 3877–3888.

59. Irusta, P. M., Luo, Y., Bakht, O., Lai, C.-C., Smith, S. O., and DiMaio, D. (2002) Definition of an inhibitory juxtamembrane WW-like domain in the platelet-derived growth factor beta receptor., *Journal of Biological Chemistry 277*, 38627–38634.

60. McLaughlin, S., Smith, S. O., Hayman, M. J., and Murray, D. (2005) An electrostatic engine model for autoinhibition and activation of the epidermal growth factor receptor (EGFR/ErbB) family., *Journal of General Physiology 126*, 41–53.

61. Lemmon, M. A. (2009) Ligand-induced ErbB receptor dimerization., *Experimental Cell Research 315*, 638–648.

62. Macdonald, J. L., and Pike, L. J. (2008) Heterogeneity in EGF-binding affinities arises from negative cooperativity in an aggregating system., *Proceedings of the National Academy of Sciences of the United States of America 105*, 112–117.

63. Mayawala, K., Vlachos, D. G., and Edwards, J. S. (2005) Heterogeneities in EGF receptor density at the cell surface can lead to concave up scatchard plot of EGF binding., *FEBS Letters 579*, 3043–3047.

64. Alvarado, D., Klein, D. E., and Lemmon, M. A. (2010) Structural basis for negative cooperativity in growth factor binding to an EGF receptor., *Cell 142*, 568–579.

65. Lu, C., Mi, L.-Z., Grey, M. J., Zhu, J., Graef, E., Yokoyama, S., and Springer, T. a. (2010) Structural evidence for loose linkage between ligand binding and kinase activation in the epidermal growth factor receptor., *Molecular and Cellular Biology 30*, 5432–5443.

66. Liu, P., Cleveland, T. E., Bouyain, S., Byrne, P. O., Longo, P. A., and Leahy, D. J. (2012) A single ligand is sufficient to activate EGFR dimers., *Proceedings of the National Academy of Sciences of the United States of America 109*, 10861–10866.

67. Kästner, J., Loeffler, H. H., Roberts, S. K., Martin-Fernandez, M. L., and Winn, M. D. (2009) Ectodomain orientation, conformational plasticity and oligomerization of ErbB1 receptors investigated by molecular dynamics., *Journal of Structural Biology 167*, 117–128.

68. Tynan, C. J., Roberts, S. K., Rolfe, D. J., Clarke, D. T., Loeffler, H. H., Kästner, J., Winn, M. D., Parker, P. J., and Martin-Fernandez, M. L. (2011) Human epidermal growth factor receptor (EGFR) aligned on the plasma membrane adopts key features of Drosophila EGFR asymmetry., *Molecular and Cellular Biology 31*, 2241–2252.

69. Krall, J. A., Beyer, E. M., and MacBeath, G. (2011) High- and low-affinity epidermal growth factor receptor-ligand interactions activate distinct signaling pathways., *PLoS One 6*, e15945.

70. He, L., and Hristova, K. (2012) Consequences of replacing EGFR juxtamembrane domain with an unstructured sequence., *Scientific Reports 2*, 1–8.

71. Mi, L.-Z., Lu, C., Li, Z., Nishida, N., Walz, T., and Springer, T. A. (2011) Simultaneous visualization of the extracellular and cytoplasmic domains of the epidermal growth factor receptor., *Nature Structural & Molecular Biology 18*, 984–989.

72. Jura, N., Shan, Y., Cao, X., Shaw, D. E., and Kuriyan, J. (2009) Structural analysis of the catalytically inactive kinase domain of the human EGF receptor 3., *Proceedings of the National Academy of Sciences of the United States of America 106*, 21608–21613.

73. Gotoh, N., Tojo, A., Hino, M., Yazaki, Y., and Shibuya, M. (1992) A highly conserved tyrosine residue at codon 845 within the kinase domain is not required for the transforming activity of human epidermal growth factor receptor., *Biochemical and Biophysical Research Communications 186*, 768–774.

74. Kim, Y., Li, Z., Apetri, M., Luo, B., Settleman, J. E., and Anderson, K. S. (2012) Temporal resolution of autophosphorylation for normal and oncogenic forms of EGFR and differential effects of gefitinib., *Biochemistry 51*, 5212–5222.

75. Pines, G., Huang, P. H., Zwang, Y., White, F. M., and Yarden, Y. (2010) EGFRvIV: a previously uncharacterized oncogenic mutant reveals a kinase autoinhibitory mechanism., *Oncogene 29*, 5850–5860.

76. Gajiwala, K. S. (2013) EGFR: tale of the C-terminal tail., *Protein science : a publication of the Protein Society 22*, 995–999.

77. Coskun, Ü., Grzybek, M., Drechsel, D., and Simons, K. (2011) Regulation of human EGF receptor by lipids., *Proceedings of the National Academy of Sciences of the United States of America 108*, 9044–9048.

78. Michailidis, I. E., Rusinova, R., Georgakopoulos, A., Chen, Y., Iyengar, R., Robakis, N. K., Logothetis, D. E., and Baki, L. (2011) Phosphatidylinositol-4,5-bisphosphate regulates epidermal growth factor receptor activation., *Pflugers Archiv European Journal of Physiology 461*, 387–397.

79. Prinetti, A., Loberto, N., Chigorno, V., and Sonnino, S. (2009) Glycosphingolipid behaviour in complex membranes., *Biochimica et Biophysica Acta 1788*, 184–193.

80. Lu, Y.-C., and Chen, H.-C. (2011) Involvement of lipid rafts in adhesion-induced activation of Met and EGFR., *Journal of Biomedical Science 18*, 78.

81. Pike, L. J., and Casey, L. (2002) Cholesterol levels modulate EGF receptor-mediated signaling by altering receptor function and trafficking., *Biochemistry 41*, 10315–10322.

82. Bremer, E. G., Schlessinger, J., and Hakomori, S. (1986) Ganglioside-mediated modulation of cell growth. Specific effects of GM3 on tyrosine phosphorylation of the epidermal growth factor receptor., *Journal of Biological Chemistry 261*, 2434–2340.

83. Ash, W. L., Zlomislic, M. R., Oloo, E. O., and Tieleman, D. P. (2004) Computer simulations of membrane proteins., *Biochimica et Biophysica Acta 1666*, 158–189.

84. Shan, Y., Eastwood, M. P., Zhang, X., Kim, E. T., Arkhipov, A., Dror, R. O., Jumper, J., Kuriyan, J., and Shaw, D. E. (2012) Oncogenic mutations counteract intrinsic disorder in the EGFR kinase and promote receptor dimerization., *Cell 149*, 860–870.

85. Mirza, A., Mustafa, M., Talevich, E., and Kannan, N. (2010) Co-conserved features associated with cis regulation of ErbB tyrosine kinases., *PLoS One 5*, e14310.

86. Shih, A. J., Telesco, S. E., and Radhakrishnan, R. (2011) Analysis of somatic mutations in cancer: molecular mechanisms of activation in the ErbB family of receptor tyrosine kinases., *Cancers 3*, 1195–1231.

87. Shih, A. J., Telesco, S. E., Choi, S.-H., Lemmon, M. A., and Radhakrishnan, R. (2011) Molecular dynamics analysis of conserved hydrophobic and hydrophilic bond-interaction networks in ErbB family kinases., *The Biochemical Journal 436*, 241–251.

88. Perez-Aguilar, J. M., and Saven, J. G. (2012) Computational design of membrane proteins., *Structure 20*, 5–14.

89. Chetwynd, A., Wee, C. L., Hall, B. A., and Sansom, M. S. P. (2010) The energetics of transmembrane helix insertion into a lipid bilayer., *Biophysical Journal 99*, 2534–2540.

90. Psachoulia, E., Fowler, P. W., Bond, P. J., and Sansom, M. S. P. (2008) Helix-helix interactions in membrane proteins: coarse-grained simulations of glycophorin A helix dimerization., *Biochemistry 47*, 10503–10512.

91. Kalli, A. C., Hall, B. A., Campbell, I. D., and Sansom, M. S. P. (2011) A helix heterodimer in a lipid bilayer: prediction of the structure of an integrin transmembrane domain via multiscale simulations., *Structure 19*, 1477–1484.

92. Prakash, A., Janosi, L., and Doxastakis, M. (2010) Self-association of models of transmembrane domains of ErbB receptors in a lipid bilayer., *Biophysical Journal 99*, 3657–3665.

93. Psachoulia, E., Marshall, D. P., and Sansom, M. S. P. (2010) Molecular dynamics simulations of the dimerization of transmembrane alpha-helices., *Accounts of Chemical Research 43*, 388–396.

94. Park, J. H., Liu, Y., Lemmon, M. A., and Radhakrishnan, R. (2012) Erlotinib binds both inactive and active conformations of the EGFR tyrosine kinase domain., *The Biochemical Journal 448*, 417–423.

95. Arkhipov, A., Shan, Y., Das, R., Endres, N. F., Eastwood, M. P., Wemmer, D. E., Kuriyan, J., and Shaw, D. E. (2013) Architecture and membrane interactions of the EGF receptor., *Cell 152*, 557–569.

96. Webb, S. E. D., Roberts, S. K., Needham, S. R., Tynan, C. J., Rolfe, D. J., Winn, M. D., Clarke, D. T., Barraclough, R., and Martin-Fernandez, M. L. (2008) Single-molecule imaging and fluorescence lifetime imaging microscopy show different structures for high- and low-affinity epidermal growth factor receptors in A431 cells., *Biophysical Journal 94*, 803–819.

97. Limongelli, V., Bonomi, M., and Parrinello, M. (2013) Funnel metadynamics as accurate binding free-energy method., *Proceedings of the National Academy of Sciences of the United States of America 110*, 6358–6363.

98. Lelimousin, M., Limongelli, V., and Sansom, M. S. P. (2013) Transmembrane helix association of ErbB receptors investigated by coarse-grained metadynamics simulations., *Biophysical Journal 104*, 32a.

99. Alder, B. J., and Wainwright, T. E. (1957) Phase transition for a hard sphere system., *Journal of Chemical Physics 27*, 1208-1209.

100. McCammon, J. A., Gelin, B. R., and Karplus, M. (1977) Dynamics of folded proteins., *Nature 267*, 585–590.

101. Dodson, G. G., Lane, D. P., and Verma, C. S. (2008) Molecular simulations of protein dynamics: new windows on mechanisms in biology., *EMBO Reports 9*, 144–150.

102. Henzler-Wildman, K., and Kern, D. (2007) Dynamic personalities of proteins., *Nature 450*, 964–972.

103. Zhou, H.-X. (2014) Theoretical frameworks for multiscale modeling and simulation., *Current Opinion in Structural Biology 25C*, 67–76.

104. Leach, A. (2001) Molecular modelling: principles and applications (2nd Edition), p 784, Prentice Hall.

105. Adcock, S. A., and McCammon, J. A. (2006) Molecular dynamics: survey of methods for simulating the activity of proteins., *Chemical Reviews 106*, 1589–1615.

106. Lindorff-Larsen, K., Maragakis, P., Piana, S., Eastwood, M. P., Dror, R. O., and Shaw, D. E. (2012) Systematic validation of protein force fields against experimental data., *PLoS One 7*, e32131.

107. Brooks, B., and Bruccoleri, R. (1983) CHARMM: A program for macromolecular energy, minimization, and dynamics calculations., *Journal of Computational Chemistry 4*, 187–217.

108. MacKerell, a D., Banavali, N., and Foloppe, N. (2001) Development and current status of the CHARMM force field for nucleic acids., *Biopolymers 56*, 257–265.

109. Jorgensen, W. L., Maxwell, D. S., and Tirado-Rives, J. (1996) Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids., *Journal of the American Chemical Society 118*, 11225–11236.

110. Jorgensen, W. L., and Tirado-Rives, J. (1988) The OPLS potential functions for proteins. Energy minimizations for crystals of cyclic peptides and crambin., *Journal of the American Chemical Society 110*, 1657–1666.

111. Darden, T., York, D., and Pedersen, L. (1993) Particle mesh Ewald: An N(N) method for Ewald sums in large systems., *Journal of Chemical Physics*, AIP *98*, 10089–10092.

112. Essman, U., Perera, L., Berkowitz, M., Darden, T., Lee, H., and Pedersen, L. (1995) A smooth particle mesh Ewald method., *Journal of Chemical Physics 103*, 8577.

113. Apol, E., Apostolov, R., Berendsen, H. J. ., and Buuren, A. van. (2010) GROMACS user manual version 4.5.4.

114. Schuler, L. D., Daura, X., and Van Gunsteren, W. F. (2001) An improved GROMOS96 force field for aliphatic hydrocarbons in the condensed phase., *Journal of Computational Chemistry 22*, 1205–1218.

115. Marrink, S. J., Risselada, H. J., Yefimov, S., Tieleman, D. P., and de Vries, A. H. (2007) The MARTINI force field: coarse grained model for biomolecular simulations., *Journal of Physical Chemistry B 111*, 7812–7824.

116. Monticelli, L., Kandasamy, S. K., Periole, X., Larson, R. G., Tieleman, D. P., and Marrink, S.-J. (2008) The MARTINI coarse-grained force field: extension to proteins., *Journal of Chemical Theory and Computation 4*, 819–834.

117. Marrink, S. J., de Vries, A. H., and Mark, A. E. (2004) Coarse grained model for semiquantitative lipid simulations., *Journal of Physical Chemistry B 108*, 750–760.

118. Bond, P. J., and Sansom, M. S. P. (2006) Insertion and assembly of membrane proteins via simulation., *Journal of the American Chemical Society 128*, 2697–2704.

119. Atilgan, a R., Durell, S. R., Jernigan, R. L., Demirel, M. C., Keskin, O., and Bahar, I. (2001) Anisotropy of fluctuation dynamics of proteins with an elastic network model., *Biophysical Journal 80*, 505–515.

120. Hall, B., Halim, K., Buyan, A., and Sansom, M. A Sidekick for Membrane Simulations: Automated ensemble molecular dynamics simulations of transmembrane helices., *Journal of Chemical Theory and Computation 10*, 2165-2175

121. Vostrikov, V. V, Hall, B. A., Greathouse, D. V, Koeppe, R. E., and Sansom, M. S. P. (2010) Changes in transmembrane helix alignment by arginine residues revealed by solid-state NMR experiments and coarse-grained MD simulations., *Journal of the American Chemical Society 132*, 5803–5811.

122. Vostrikov, V. V, Hall, B. A., Sansom, M. S. P., and Koeppe, R. E. (2012) Accommodation of a central arginine in a transmembrane peptide by changing the placement of anchor residues., *Journal of Physical Chemistry B 116*, 12980–12990.

123. Hall, B. A., Chetwynd, A. P., and Sansom, M. S. P. (2011) Exploring peptide-membrane interactions with coarse-grained MD simulations., *Biophysical Journal*, *100*, 1940–1948.

124. Hall, B. A., Armitage, J. P., and Sansom, M. S. P. (2011) Transmembrane helix dynamics of bacterial chemoreceptors supports a piston model of signalling., *PLoS computational biology 7*, e1002204.

125. Lindau, M., Hall, B. A., Chetwynd, A., Beckstein, O., and Sansom, M. S. P. (2012) Coarse-grain simulations reveal movement of the synaptobrevin C-terminus in response to piconewton forces., *Biophysical Journal 103*, 959-969

126. Crowet, J.-M., Parton, D. L., Hall, B. A., Steinhauer, S., Brasseur, R., Lins, L., and Sansom, M. S. P. (2012) Multi-scale simulation of the simian immunodeficiency virus fusion peptide., *Journal of Physical Chemistry B 116*, 13713–13721.

127. Stansfeld, P. J., and Sansom, M. S. P. (2011) From coarse grained to atomistic: A serial multiscale approach to membrane protein simulations., *Journal of Chemical Theory and Computation* 7, 1157-1166

128. Rotkiewicz, P., and Skolnick, J. (2008) Fast procedure for reconstruction of full-atom protein models from reduced representations., *Journal of Computational Chemistry 29*, 1460–1465.

129. Hess, B., Bekker, H., Berendsen, H. J. C., and Fraaije, J. G. E. M. (1997) LINCS: A linear constraint solver for molecular simulations., *Journal of Computational Chemistry 18*, 1463–1472.

130. Ryckaert, J.-P., Ciccotti, G., and Berendsen, H. J. . (1977) Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes., *Journal of Computational Physics 23*, 327–341.

131. Berendsen, H. J. C., Postma, J. P. M., van Gunsteren, W. F., DiNola, A., and Haak, J. R. (1984) Molecular dynamics with coupling to an external bath., *Journal of Chemical Physics 81*, 3684-3690.

132. Eswar, N., Webb, B., Marti-renom, M. A., Madhusudhan, M. S., David, E., Shen, M., Pieper, U., and Sali, A. (2007) Comparative Protein structure modeling using MODELLER., *Current Protocols in Protein Science Chapter 2*.Unit 2.9

133. Humphrey, W., Dalke, A., and Schulten, K. (1996) VMD: visual molecular dynamics., *Journal of Molecular Graphics 14*, 33–8, 27–28.

134. Curran, A. R., and Engelman, D. M. (2003) Sequence motifs, polar interactions and conformational changes in helical membrane proteins., *Current Opinion in Structural Biology 13*, 412–417.

135. Russ, W. P., and Engelman, D. M. (2000) The GxxxG motif: a framework for transmembrane helix-helix association., *Journal of molecular biology 296*, 911–919.

136. Senes, A., Gerstein, M., and Engelman, D. M. (2000) Statistical analysis of amino acid patterns in transmembrane helices: the GxxxG motif occurs frequently and in association with beta-branched residues at neighboring positions., *Journal of Molecular Biology 296*, 921–936.

137. Lemmon, M. A. (2007) Activation and inhibition of the EGF receptor., *FASEB Journal 21*, A46–A46.

138. Ranganathan, R. (2007) Signaling across the cell membrane., *Science 318*, 1253–1254.

139. Li, E., and Hristova, K. (2006) Role of receptor tyrosine kinase transmembrane domains in cell signaling and human pathologies., *Biochemistry 45*, 6241–6251.

140. Kalli, A. C., Wegener, K. L., Goult, B. T., Anthis, N. J., Campbell, I. D., and Sansom, M. S. P. (2010) The structure of the talin/integrin complex at a lipid bilayer: an NMR and MD simulation study., *Structure 18*, 1280–1288.

141. Dixit, A., and Verkhivker, G. M. (2011) The energy landscape analysis of cancer mutations in protein kinases., *PLoS One 6*, e26071.

142. Fry, W. H. D., Kotelawala, L., Sweeney, C., and Carraway, K. L. (2009) Mechanisms of ErbB receptor negative regulation and relevance in cancer., *Experimental Cell Research 315*, 697–706.

143. Surawska, H., Ma, P. C., and Salgia, R. (2004) The role of ephrins and Eph receptors in cancer., *Cytokine & Growth Factor Reviews 15*, 419–433.

144. Glass, D. J., Apel, E. D., Shah, S., Bowen, D. C., DeChiara, T. M., Stitt, T. N., Sanes, J. R., and Yancopoulos, G. D. (1997) Kinase domain of the muscle-specific receptor tyrosine kinase (MuSK) is sufficient for phosphorylation but not clustering of acetylcholine receptors: Required role for the MuSK ectodomain?., *Proceedings of the National Academy of Sciences of the United States of America 94*, 8848–8853.

145. Beeson, D., Webster, R., Cossins, J., Lashley, D., Spearman, H., Maxwell, S., Slater, C. R., Newsom-Davis, J., Palace, J., and Vincent, A. (2008) Congenital myasthenic syndromes and the formation of the neuromuscular junction., *Annals Of The New York Academy Of Sciences 1132*, 99–103.

146. Kim, N., Stiegler, A. L., Cameron, T. O., Hallock, P. T., Gomez, A. M., Huang, J. H., Hubbard, S. R., Dustin, M. L., and Burden, S. J. (2008) Lrp4 is a receptor for agrin and forms a complex with MuSK., *Cell 135*, 334–342.

147. Hubbard, S. R., and Gnanasambandan, K. (2013) Structure and activation of MuSK, a receptor tyrosine kinase central to neuromuscular junction formation., *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics 1834*, 2166–2169.

148. Cuthbertson, J. M., Doyle, D. A., and Sansom, M. S. P. (2005) Transmembrane helix prediction: a comparative evaluation and analysis., *Protein Engineering, Design & Selection : PEDS 18*, 295–308.

149. Van Der Spoel, D., Lindahl, E., Hess, B., Groenhof, G., Mark, A. E., and Berendsen, H. J. C. (2005) GROMACS: Fast, flexible, and free., *Journal of Computational Chemistry 26*, 1701–1718.

150. Lindahl, E., Hess, B., and Van Der Spoel, D. (2001) GROMACS 3.0: a package for molecular simulation and trajectory analysis., *Journal of Molecular Modeling 7*, 306–317.

151. Daura, X., Gademann, K., Jaun, B., Seebach, D., van Gunsteren, W. F., and Mark, A. E. (1999) Peptide folding: When simulation meets experiment., *Angewandte Chemie International Edition 38*, 236–240.

152. Hall, B. A., Halim, K. A., Buyan, A., Emmanouil, B., and Sansom, M. S. P. (2014) A sidekick for membrane simulations: automated ensemble molecular dynamics simulations of transmembrane helices., *Journal of Chemical Theory and Computation*, 140407160657002.

153. Polyansky, A. A., Volynsky, P. E., and Efremov, R. G. (2012) Multistate organization of transmembrane helical protein dimers governed by the host membrane., *Journal of the American Chemical Society 134*, 14390-14400.

154. Volynsky, P. E., Polyansky, A. A., Fakhrutdinova, G. N., Bocharov, E. V., and Efremov, R. G. (2013) Role of dimerization efficiency of transmembrane domains in activation of fibroblast growth factor receptor 3 (FGFR3)., *Journal of the American Chemical Society 135*, 8105–8108.

155. McLaughlin, S., Wang, J., Gambhir, A., and Murray, D. (2002) PIP(2) and proteins: interactions, organization, and information flow., *Annual Review of Biophysics and Biomolecular Structure 31*, 151–175.

156. Sato, T., Pallavi, P., Golebiewska, U., McLaughlin, S., and Smith, S. O. (2006) Structure of the membrane reconstituted transmembrane-juxtamembrane peptide EGFR(622-660) and its interaction with Ca2+/calmodulin., *Biochemistry 45*, 12704–12714.

157. Toledo, M. S., Suzuki, E., Handa, K., and Hakomori, S. (2005) Effect of ganglioside and tetraspanins in microdomains on interaction of integrins with fibroblast growth factor receptor., *Journal of Biological Chemistry 280*, 16227–34.

158. Hynds, D. L., Summers, M., Van Brocklyn, J., O'Dorisio, M. S., and Yates, A. J. (1995) Gangliosides inhibit platelet-derived growth factor-stimulated growth, receptor phosphorylation, and dimerization in neuroblastoma SH-SY5Y cells., *Journal of Neurochemistry 65*, 2251–2258.

159. Langeveld, M., Ghauharali, K. J. M., Sauerwein, H. P., Ackermans, M. T., Groener, J. E. M., Hollak, C. E. M., Aerts, J. M., and Serlie, M. J. (2008) Type I Gaucher disease, a glycosphingolipid storage disorder, is associated with insulin resistance., *Journal of Clinical Endocrinology and Metabolism 93*, 845–851.

160. Wang, X.-Q., Sun, P., and Paller, A. S. (2005) Gangliosides inhibit urokinase-type plasminogen activator (uPA)-dependent squamous carcinoma cell migration by preventing uPA receptor/alphabeta integrin/epidermal growth factor receptor interactions., *Journal of Investigative Dermatology 124*, 839–848.

161. Miljan, E. A., and Bremer, E. G. (2002) Regulation of growth factor receptors by gangliosides., *Sciences STKE Signal Transduction Knowledge Environment 2002*, re15.

162. Song, W. X., Vacca, M. F., Welti, R., and Rintoul, D. A. (1991) Effects of gangliosides GM3 and De-N-acetyl GM3 on epidermal growth factor receptor kinase activity and cell growth., *Journal of Biological Chemistry 266*, 10174–10181.

163. Zhou, Q., Hakomori, S., Kitamura, K., and Igarashi, Y. (1994) GM3 directly inhibits tyrosine phosphorylation and de-N-acetyl-GM3 directly enhances serine phosphorylation of epidermal growth factor receptor, independently of receptor-receptor interaction., *Journal of Biological Chemistry 269*, 1959–1965.

164. Yoon, S.-J., Nakayama, K., Hikita, T., Handa, K., and Hakomori, S. (2006) Epidermal growth factor receptor tyrosine kinase is modulated by GM3 interaction with N-linked GlcNAc termini of the receptor., *Proceedings of the National Academy of Sciences of the United States of America 103*, 18987–18991.

165. Sottocornola, E., Misasi, R., Mattei, V., Ciarlo, L., Gradini, R., Garofalo, T., Berra, B., Colombo, I., and Sorice, M. (2006) Role of gangliosides in the association of ErbB2 with lipid rafts in mammary epithelial HC11 cells., *The FEBS journal 273*, 1821–1830.

166. Endres, N. F., Engel, K., Das, R., Kovacs, E., and Kuriyan, J. (2011) Regulation of the catalytic activity of the EGF receptor., *Current Opinion in Structural Biology 21*, 777–784.

167. Koldsø, H., Shorthouse, D., Helie, J., and Sansom, S. . M. (2014) Lipid clustering correlates with membrane curvature as revealed by molecular simulations of complex lipid bilayers.(revised manuscript submitted)

168. Risselada, H. J., and Marrink, S. J. (2008) The molecular face of lipid rafts in model membranes., *Proceedings of the National Academy of Sciences of the United States of America 105*, 17367–17372.

169. Parton, D. L., Tek, A., Baaden, M., and Sansom, M. S. P. (2013) Formation of Raft-Like Assemblies within Clusters of Influenza Hemagglutinin Observed by MD Simulations, *PLoS Computational Biology 9*, e1003034.

170. Wenk, M. R. (2010) Lipidomics: new tools and applications., *Cell 143*, 888–895.

171. Lingwood, D., Kaiser, H.-J., Levental, I., and Simons, K. (2009) Lipid rafts as functional heterogeneity in cell membranes., *Biochemical Society Transactions 37*, 955–960.

172. Roepstorff, K., Thomsen, P., Sandvig, K., and van Deurs, B. (2002) Sequestration of epidermal growth factor receptors in non-caveolar lipid rafts inhibits ligand binding., *Journal of Biological Chemistry 277*, 18954–18960.

173. Chen, X., and Resh, M. D. (2002) Cholesterol depletion from the plasma membrane triggers ligand-independent activation of the epidermal growth factor receptor., *Journal of Biological Chemistry 277*, 49631–49637.

174. Mustafa, M., Mirza, A., and Kannan, N. (2011) Conformational regulation of the EGFR kinase core by the juxtamembrane and C-terminal tail: a molecular dynamics study., *Proteins 79*, 99–114.

175. Koland, J. G. (2014) Coarse-grained molecular simulation of epidermal growth factor receptor protein tyrosine kinase multi-site self-phosphorylation., *PLoS Computational Biology 10*, e1003435.

176. Sali, A., and Blundell, T. L. (1993) Comparative protein modelling by satisfaction of spatial restraints., *Journal of Molecular Biology 234*, 779–815.

177. Yun, C.-H., Mengwasser, K. E., Toms, A. V, Woo, M. S., Greulich, H., Wong, K.-K., Meyerson, M., and Eck, M. J. (2008) The T790M mutation in EGFR kinase causes drug resistance by increasing the affinity for ATP., *Proceedings of the National Academy of Sciences of the United States of America 105*, 2070–2075.

178. Davis, I. W., Murray, L. W., Richardson, J. S., and Richardson, D. C. (2004) MOLPROBITY: structure validation and all-atom contact analysis for nucleic acids and their complexes., *Nucleic Acids Research 32*, W615–619.

179. Michaud-Agrawal, N., Denning, E. J., Woolf, T. B., and Beckstein, O. (2011) MDAnalysis: A toolkit for the analysis of molecular dynamics simulations., *Journal of Computational Chemistry 32*, 2319–2327.

180. Goose, J. E., and Sansom, M. S. P. (2013) Reduced lateral mobility of lipids and proteins in crowded membranes., *PLoS Computational Biology 9*, e1003033.

181. Pike, L. J., Han, X., and Gross, R. W. (2005) Epidermal growth factor receptors are localized to lipid rafts that contain a balance of inner and outer leaflet lipids: a shotgun lipidomics study., *Journal of Biological Chemistry 280*, 26796–26804.

182. De Jong, D. H., Lopez, C. A., and Marrink, S. J. (2013) Molecular view on protein sorting into liquid-ordered membrane domains mediated by gangliosides and lipid anchors., *Faraday Discussions 161*, 347-363.

183. Baoukina, S., Mendez-Villuendas, E., Bennett, W. F. D., and Tieleman, D. P. (2013) Computer simulations of the phase separation in model membranes., *Faraday Discussions 161*, 63-75.

184. Lingwood, D., and Simons, K. (2010) Lipid rafts as a membrane-organizing principle., *Science 327*, 46–50.