Energetics of protein interactions in the membrane via computer simulations



Jan Domański

Merton College

University of Oxford

A thesis submitted for the degree of

Doctor of Philosophy

Trinity 2017

Publications

The contents of this thesis are based on the following publications:

- J. Domański et al. "Balancing Force Field Protein-Lipid Interactions To Capture Transmembrane Helix-Helix Association". In: *J. Chem. Theory Comput.* (Feb. 2018)
- J. Domański et al. "Convergence and Sampling in Determining Free Energy Landscapes for Membrane Protein Association". In: *J. Phys. Chem. B* 121.15 (2017). PMID: 27807980, pp. 3364–3375
- G. Hedger et al. "Lipid loving ANTs: cardiolipin interactions and the organization of the adenine nucleotide translocase in model mitochondrial membranes". In: *Biochemistry* 55.45 (Nov. 2016), pp. 6238–6249

This thesis is dedicated to my sister and my parents:

two lovely Maria's and Maciej Domański

who's love and support helped me throughout the duration

of this PhD "experience"...

Thank you with all of my heart / Dziękuje Wam z całego serca :)

Acknowledgements

My sincere thanks go to Dr. Robert Best, Prof. Mark Sansom and Prof. Phillip Stansfeld for their patience and guidance in helping me form the ideas and refine the work presented in this thesis. A multitude of thanks and my gratitude to you guys for putting up with me, and with the small figure legends, for all those years.

I also wanted to thank the members of the Sansom, Stansfeld and Best labs. In particular, Wenwei Zheng, Pengfei Tian who made Washington feel like home. Also, Sarah-Beth Amos, George Hedger and Eiji Yamamoto for doing the same for Oxford. To all you, for your technical and scientific help, and for listening to my moaning about the academic life, my sincere thanks.

Countless people who shaped my scientific development: Dr. Pavlos Maragakis at D. E. Shaw Research, Profs. Lars Schäfer and Siewert-Jan Marrink from Groningen. Prof. Oliver Beckstein, now at ASU, for getting me hooked on scientific research, git and python. Prof. George Ratcliffe, my New College tutor for teaching me the foundations of biochemistry, and the scientific process... and that I should stay the hell away from experimental work. Marek Sawicki and Ewa Jastrzębska, my high-school chemistry and biology teachers, for showing me the fascinating word at the intersection of chemistry and biology. Bożena Grondziel, my primary school biology teacher, who was particularly fond of molecular biology and the implications it had for medicine.

Abstract

Protein-lipid interactions underpin the biological activities in cell membranes. However, the energetic aspect of these interactions is notoriously difficult to study. Here, I performed a number of enhanced sampling simulations on biologically-interesting systems using different collective variables to provide converged estimates of protein binding affinities in the membrane.

Replica exchange umbrella sampling (REUS) simulations were performed using a distance-based collective variable on two biologically relevant interactions: the ANT-Cardiolipin (CL) and Kir2.2-PIP_2. Converged PMFs were obtained for both systems; for the Kir2.2 R186A channel mutant, the potential of mean force (PMF) well-depth is reduced, in qualitative agreement with experimental data, where the mutation reduces conductance. These results demonstrate that REUS methodology can be robustly used to estimate protein-lipid binding interactions with the coarse-grained model, on different binding pockets and lipid types.

The REUS methodology was then extended to study prototypical transmembrane (TM) helix-helix interactions. A series of model helix-helix dimers was studied, experimentally known to span a range of affinities, including the Glycophorin A (GpA) TM fragment and a number of different Receptor Tyrosine Kinase (RTK) TM fragments. The coarse-grained MARTINI model correctly predicts the bound state to be most stable, for GpA. Unexpectedly the non-dimerizing GpA mutant was just as stable as the wild-type. Similarly, for RTKs, the calculated affinities were similar, in contrast to semi-quantitative TOXCAT data that suggests a range of affinities.

Using the collective variable developed to study the coarse-grained helix-helix systems, an all-atom PMF was calculated for wild-type GpA in a 1-palmitoyl,2-oleoyl-sn-glycero-3-phosphocholine (POPC) bilayer. Unexpectedly, the dimer state was found to be metastable, with the CHARMM36 and AMBER force fields, with the unbound state being the global energy minimum. Two simple corrections were proposed for the CHARMM36 force field, which stabilize the bound state making the model consistent with experiment.

Contents

1 Introduction **1**

1.1 Cells 1

1.2 Cell Membranes 2

1.3 Proteins in and around membranes 4

1.4 Experiments 7

1.4.1 Protein-protein interaction energetics in the membrane 8

1.4.1.1 Helix-helix interacitons 9

1.4.2 Protein-lipid interaction energetics in the membrane 11

1.5 Simulations 14

2 Methods **17**

2.1 Molecular dynamics simulations 19

2.1.1 Equations of motion 19

2.1.2 Numerical techniques 22

2.1.2.1 Constraints 22

2.1.2.2 Periodic Boundary Conditions 22

2.1.2.3 Van der Waals interaction 23

2.1.2.4 Charge-charge (Coulombic) interactions 25

2.1.2.5 Ensembles, temperature and pressure control 25

2.1.3 Force fields 26

2.1.3.1 All-atom force fields 28
2.1.3.2 Coarse-grained force fields 29
2.1.4 Detailed settings 29
2.2 Definitions 30
2.2.1 Free energy surface 30
2.2.2 Convergence 30
2.2.3 Collective variables 31
2.2.4 Definition of an inter-helical CV 32
2.3 Sampling methods 33
2.3.1 Replica exchange 33
2.3.2 Umbrella sampling 37
2.3.3 Metadynamics 38
2.3.4 Generating initial conditions for umbrella sampling 40
2.3.5 Recovering the unbiased free energy surface 42
2.3.6 Re-weighting along different collective variables 43
2.4 Comparison to experiment 44
2.4.1 Experimental K_d for helix-helix dimerization 44

3 Protein-lipid interactions in the membrane 46
3.1 Introduction to ANT 48
3.2 Introduction to Kir2.2 50
3.3 ANT–Cardiolipin interaction 53
3.3.1 Material and methods 53
3.3.2 Results and discussion 54
3.3.3 Conclusion and future directions 57
3.4 Kir2.2–PIP_2 interaction 60
3.4.1 Material and methods 60
3.4.2 Results and discussion 61

3.4.3 Conclusions and future work 64

4 Coarse-grained helix-helix dimerization 66

4.1 Introduction 66

4.1.1 Glycophorin A 66

4.1.2 Receptor Tyrosine kinases 68

4.2 Glycophorin A 72

4.2.1 Material and methods 72

4.2.2 Results 72

4.2.3 Conclusions and future work 83

4.3 Receptor Tyrosine Kinases 86

4.3.1 Material and methods 86

4.3.2 Results 87

4.3.3 Conclusions and future work 89

5 All-atom GpA dimerization 90

5.1 Glycophorin A 91

5.1.1 Introduction 91

5.1.2 Methods 94

5.1.2.1 Simulation methods 94

5.1.2.2 System preparation 95

5.1.2.3 Potential of mean force calculations 95

5.1.3 Analysis and visualization 97

5.1.4 Results and discussion 98

5.1.4.1 Potentials of mean force 98

5.1.4.2 Additional simulations 102

5.1.4.3 Checking lipid behaviour in simulation 104

5.1.4.4 Proposed force field fixes 106

5.1.4.5 Comparison to a coarse-grained model 110
5.1.5 Conclusions and future work 113

6 Conclusions **115**

6.1 Coarse-grained protein-lipid interaction 115
6.2 Coarse-grained helix-helix interactions 116
6.3 All-atom helix-helix interactions 117
6.4 Membrane simulations using models of different resolution 118

Bibliography **119**

Chapter 1

Introduction

1.1 Cells

The building block of all living organisms is the cell. Organisms can be made of a single cell, such as bacteria, and are then referred to as unicellular. Other organisms are made of multiple cells working together – those organisms are called multicellular. The cell boundary is the lipid bilayer that separates the external environment form the cellular environment, the cytosol. Based on genetic information, three cell types have been delineated: eukaryotic, prokaryotic (eubacteria) and archaeal. Eukaryotic cells are $10 - 100\,\mu m$ in diameter and can contain numerous organelles along with a nucleus. The nucleus stores the genetic information of the cell; organelles are intracellular compartments, serving specialized functions. Prokaryotic and archaeal cells are smaller, $1 - 10\,\mu m$ in diameter, and have no organelles and no nucleus.

Organelles are (in some instances) surrounded by lipid membranes of their own, separating them from the cell contents. These organelles are typically involved in specialized functions for instance, mitochondria are involved in metabolism and the endoplasmatic reticulum participates in protein synthesis . See Figure 1.1 for a diagram showing an eukaryotic cell containing a number of organelles of different type.

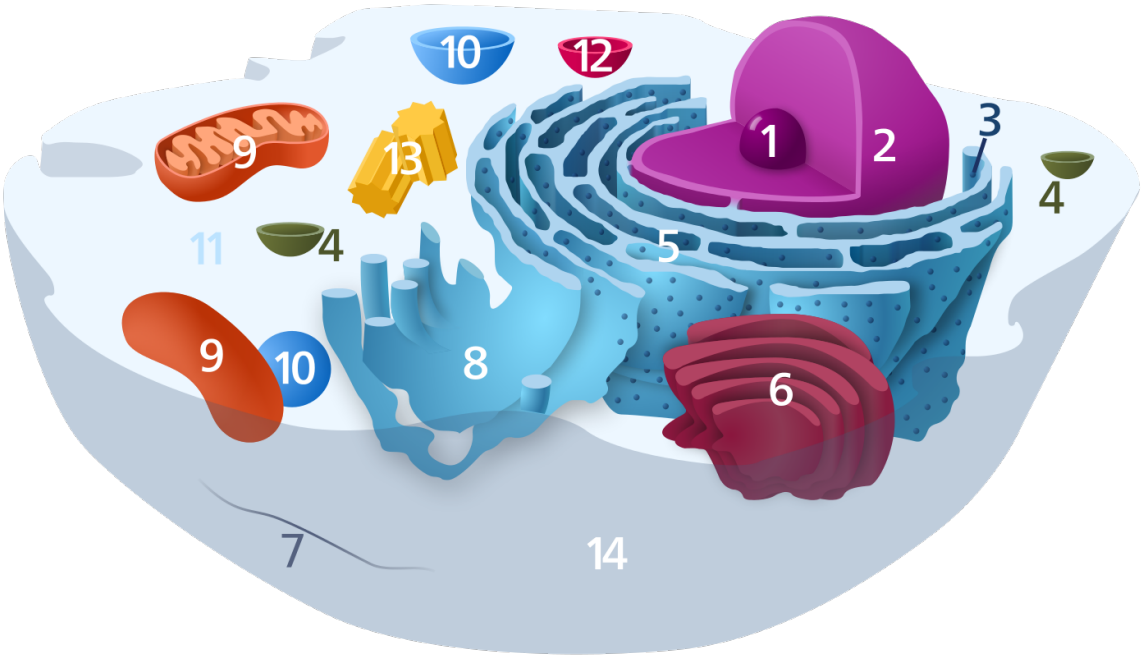


Figure 1.1: Diagram of id=jdan eukaryotic cell showing various organelle types. 1. Nucleolus 2. Nucleus 3. Ribosomes (dots on rough reticulum walls) 4. Vesicle 5. Rough endoplasmic reticulum 6. Golgi apparatus (or "Golgi body") 7. Cytoskeleton 8. Smooth endoplasmic reticulum 9. Mitochondrion 10. Vacuole 11. Cytosol 12. Lysosome 13. Centriole 14. Cell membrane Taken from `https://commons.wikimedia.org/wiki/File:Animal_Cell.svg`, accessed on 17. September 2017.

This thesis will be focused around the cell membrane, which holds the other organelles together in a confined space of a cell.

1.2 Cell Membranes

A structure called the cell membrane separates the cells from their external environment. Without this boundary contents of the cell would spill out to the exterior, diffusing into the surrounding environment . . Due to their amphipathic nature, lipids can adopt a stable bilayer structure. The hydrophilic heads of the lipids face away from the centre of the bilayer, while the hydrophobic lipid tails form the "greasy" interior. This arrangement results in the physicochemial properties of the memberne: its surface is charged and polar while its core remains apolar and hydrophobic. Due

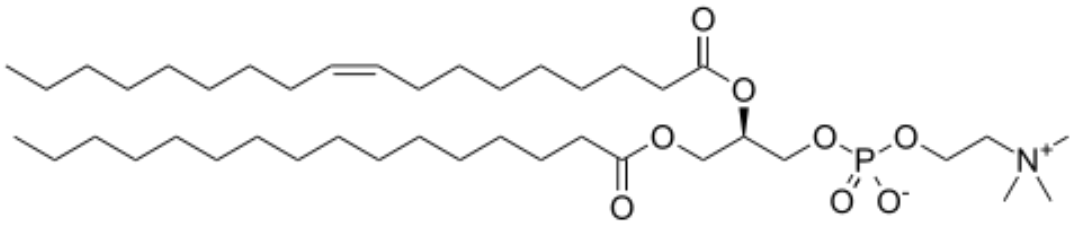


Figure 1.2: Chemical structure of 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine (POPC) lipid. Taken from `https://en.wikipedia.org/wiki/POPC#/media/File:1-palmitoyl-2-oleoylphosphatidylcholine.svg`, accessed on 14. Feb 2018.

to these physicochemical properties, the membrane does not allow many molecules to permeate across (such as sugars, ions or amino-acids) as they cannot move across the hydrophobic core. Three types of lipids build the bilayer: phospholipids, sphingolipids and sterols. The third lipid type, sterols does not form stable bilayers, but serves other functions: for example cholesterol is involved in regulating cell membrane fluidity.

The heterogeneity of biological membranes is also due to their lipid composition: the membranes are formed from multiple lipid types, often with covalent modifications such as glycans.

A particular group of lipids called phosphatidylcholines (PCs) is abundant in the animal cell membranes [132]. An example lipid from this group is 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine (POPC) (see Fig. 1.2 for its chemical structure). In order to reduce the complexity of the experimental setup setup, often a single lipid-type is chosen to mimic a biological membrane. POPC is the main lipid used in this thesis, as a "workhorse" lipid for the protein-protein interaction study. Other lipids are also used in this work to study protein-lipid interactions (these lipids will be introduced later).

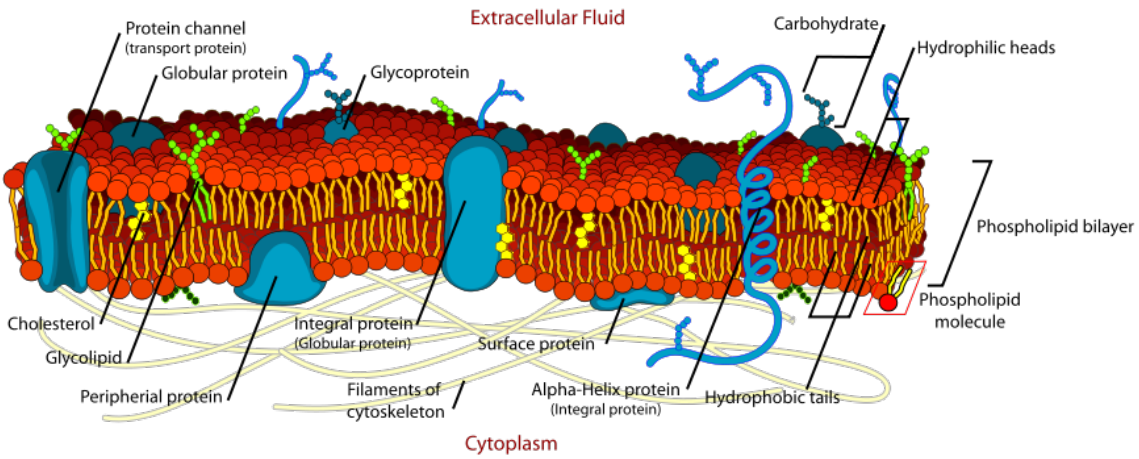


Figure 1.3: Diagram of the biological membrane structure, showing the double bilayer with embedded proteins and glycans, as well as the supporting cytoskeleton on the cytoplasmic side. Taken from `https://commons.wikimedia.org/wiki/File:Cell_membrane_detailed_diagram_en.svg`, accessed on 17. September 2017.

1.3 Proteins in and around membranes

Proteins and other mechanisms exist that permit regulated passage and uptake of substances that do not dissolve in the membrane core. Biological membranes contain numerous proteins embedded in the membrane (see Figure 1.3).

Membrane proteins are unique and distinct from their soluble counterparts. Firstly, their folding is more complicated and involves numerous associated proteins, as often the insertion into the membrane has to be catalysed. A number of protein complexes exists to facilitate insertion and folding into the membrane. The insertion can be coupled to the pH gradient (facilitated by the Hcf106 in bacteria), or mediated by the Tat complex (in which case, the transported protein is moved in a folded state), or threaded through the Sec YEG complex (powered by ATP hydrolysis, protein is unfolded during transit) [36]. How does the protein machinery that mediates membrane protein insertion gets itself into the membrane in the first place is unclear. Secondly, the surface properties of membrane proteins are different from soluble proteins. Membrane proteins have to interact both with the polar hydrophobic lipid head-groups and the hydrophobic lipid tails, while the soluble proteins have their hydrophobic core

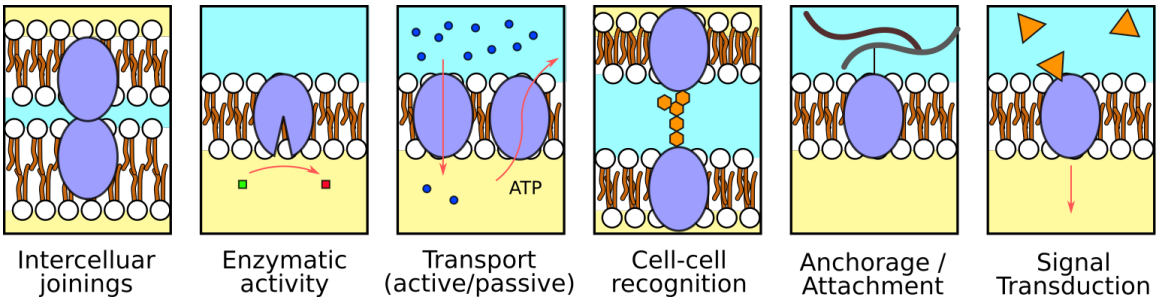


Figure 1.4: Classification of membrane proteins by their biological function. Adapted from `http://ib.bioninja.com.au/standard-level/topic-1-cell-biology/13-membrane-structure/membrane-proteins.html`, accessed on 17. September 2017.

buried and surface residues exposed to water. Two main types of membrane proteins are "membrane associated", which bind the bilayer surface, and "membrane integral" that are, at least partially, inserted into the hydrophobic bilayer. The integral membrane proteins can then be further subdivided, basing on their architecture as either helical (with one or more transmembrane helices) or β-barrel proteins.

Proteins associated with biological membranes play a variety of roles, outlined in Fig. 1.4. Examples of membrane biological function include:

1. Metabolism, for example in mitochondria the adenine nucleotide translocator (ANT) exchanges adenosine diphosphate (ADP) for adenosine triphosphate (ATP) molecules.

2. Ion permeation, for instance the Kir2.2 inward-rectifying ion channel involved in signal propagation across membranes.

3. Signalling and development, for instance the Pleckstrin homology (PH) domains selectively bind to phosphoinositides on the membrane surface.

4. Identification, for example the Glycophorin A and B are major sialoglycoproteins providing the basis for MNS blood type [163], defined by two antigen variants (M vs N, S vs s), on the Glycophorin gene, resulting in 4 possible

comibnations seen in the population. Glycophorin is also the target proteins recognized by the *Plasmodium falciparum* malaria parasite [128].

The functional importance of membrane proteins is linked to their location at the interface of the cell and the environment. Approximately a third of human genes are predicted to be membrane proteins [50, 2]. Around 50% of known drugs are thought to act via a membrane protein in their mechanism of action [82]. This underscores the role played by the membrane proteins, and their biomedical relevance. The medical importance of folding in the context of membrane proteins is best illustrated by the cystic fibrosis transmembrane conductance regulator (CFTR), where mutations induce misfolding and lead to a disease phenotype [62].

Protein-protein and protein-lipid interactions occur in the context of the membrane. The mechanism and energetics of these interactions affects biological function. Experimentally it is challenging to study these membrane systems: the amphipathic nature of the membrane makes it difficult to handle in a wet-lab, as components have to be carefully solubilized. Nonetheless, studies on isolated membrane species have yielded thermodynamic insight into their interactions. The primary challenge, computationally, remains the high viscosity of the bilayer which is 2-3 fold higher than for surrounding water [156].

How the protein and lipid species interact in the membrane, affects how proteins fold inside of the membrane environment. Experimentally, reversible folding experiments for membrane proteins are rarely attempted (but are a common-place for soluble counter-parts). Numerous chaperones and additional proteins mediate the insertion and folding of membrane proteins, further adding to the complexity [33]. The prototypical problem in membrane protein folding is a protein-protein association between two TM helices within a lipid bilayer. This simple interaction likely underlines folding of more complex, multi-pass transmembrane proteins. Multi-pass helical proteins represent a major class of membrane proteins, which is often impli-

cated in disease [214]. Interactions between individual TM helices are important for folding [158], stability [51] and the biological role [46] of membrane proteins.

The model describing the sequential folding and then the assembly of secondary structure elements in membrane protein folding was proposed by Popot and Engelman [158, 210]. In this model, the insertion of and folding of the helices is the first step, followed by the association of the folded helices in the second step. The membrane plays a critical role constraining the available configuration space to be explored by the interacting secondary structure elements. That view has changed over time, recognizing that a number of membrane mimicking environments do not seem to affect the final structure and function of proteins [159]. It is likely that at least conceptual similarities exist between the Popot and Engelman model and multi-pass TM helix insertion performed by the translocon. In simulation, spontaneous insertion and folding of small peptides can be observed at extremely high temperatures [197]. However, the latter step of helix-helix association is harder to study due to the high viscosity of the membrane and has not been, to my knowledge, observed to occur spontaneously in unbiased all-atom simulations.

1.4 Experiments

Experimental work on the thermodynamics of protein-lipid and protein-protein interactions are discussed in this section. Model systems for protein-protein interaction in the membrane include single-pass transmembrane (TM) helix homodimers. Examples include the Glycophorin A TM region and the TM regions of various receptor tyrosine kinases (RTKs), which have been studied by methods such as the TOXCAT assay, analytical ultracentrifugation, and optical tweezers [56, 60, 81].

In this section I will summarize the experimental work on the energetics of protein-protein and protein-lipid interactions in the membrane. These experimental studies

will serve as a reference point to compare simulation results; additionally, the limitations of experimental and computation techniques will be discussed.

1.4.1 Protein-protein interaction energetics in the membrane

Protein-lipid interaction have been characterized using a number of approaches. Starting with simplest building blocks and a membrane mimic, Radzicka [162] used analogues of individual amino-acids to measure their partitioning coefficient between water and cyclohexane. However, cyclohexane being an apolar solvent is a poor proxy for lipids present in biological membranes, it lacks the amphiphatic features present in membrane lipids, does not form bilayers and is at best a proxy for the greasy membrane core.. Side-chain analogues are also not a perfect approximation for amino-acids, because they lack the backbone contribution to the energetics, and cannot be used to probe any long-term interactions between amino-acids in a peptide sequence.

Recognizing these shortcomings, Wimley proposed a hydrophobicity scale based on POPC dialysis experiments with a series of Ac-WLxLL pentapeptides, where the residue "X" was replaced by one of the twenty natural amino-acids [211]. The advantage of this method is that it uses a complete peptide, albeit a short one, in a biological-like membrane. The disadvantage of this method is that it is impossible to distinguish if the pentapeptide is localized to the membrane interface, or if it partitions into the membrane. It is also limited in that the pentapeptide may not be represent the energetics of longer protein sequences.

Pentapeptides are closer to the *in vivo* situation than amino-acid analogues but still dissimilar to typical membrane proteins. To address the above issue, Fleming and co-workers systematically mutated a bilayer-facing residue on the side of a β-barrel outer membrane phospholipase A (OmpLA) to derive a "biological" hydrophobicity scale for twenty amino-acids [137]. The yield of the OmpLA folding reaction is was then indicative of how favourable is the insertion energy for each mutant; the the

location of the varied resuidue in the center of the barrel ensures that the residue is sidechaininserted into the bilayer.

Lastly, von Heijne and co-workers used a Sec translocon-based to measure, in a cell-free *in vitro* assay, how strongly each of the amino acids can drive a model helix insertion into the endoplasmic reticulum membrane. The Sec translocon is thought to insert folded helices, by opening to the side of the membrane. The translocation is stopped using an arrest sequence, which folds inside the Sec-bound ribosome and stalls it. [78, 77]. The authors interpreted their results as free energies, where the peptides stoppage is permanent. However, there is likely a kinetic component to the measurement: the arrest peptide can unfold and the translation can resume – distorting the measurement that assumes helix insertion is the only force pushing the TM helix into the membrane.

Experiments probing the protein-protein energetics in the membrane will be discussed, focusing on the helix-helix dimerization , as those are most relevant to the simulations presented in this thesis. Obtaining experimental values for individual amino-acids, or even side-chain analogues, in the membrane has not been possible, with the exception of a peptide backbone analogue [9]. The amino-acid dimerization affinities have been only estimated from their occurrence in protein structures, in a knowledge-based potential [135].

1.4.1.1 Helix-helix interacitons

Measurements of helix-helix dimerization propensity have been obtained by a range of experimental methods. Binding affinities for transmembrane helix pairs have been measured using analytical centrifugation, a work pioneered by Karen Fleming for the Glycophorin A homodimer [60]. Alanine scanning was then used to identify interaction hot-spots on the dimer that correspond to a region close to the GXXXG motif, mediating the helix interaction in structural models [61]. Analytical centrifugation

was further used to determine that Erythropoietin Receptor TM region, in detergent micelles, interacts more weakly than Glycophorin A [45].

The main limitation of the analytical centrifugation experiments is the detergent micelle environment, substituting for a lipid bilayer. This means that to estimate the K_d it is typically necessary to extrapolate detergent concentration to zero. For Glycophorin A (GpA), a common model system for TM-helix dimerization, there is a seven order-of-magnitude change between the extrapolated sedimentation equilibrium analytical ultracentrifugation, and the fluorescence resonance energy transfer (FRET) method, with FRET predicting much stronger GpA association in POPC lipids. The impact of this extrapolation has been reviewed in [121] and has been labelled by the authors as insignificant.

Bowie and Hong pioneered a FRET-based affinity measurement for TM helix dimerization [81]. Using a Glycophorin A construct with a biotin label, in either a lipid bilayer or in a detergent micelle, binding of monovalent streptavidin to the biotin labels will disrupt the dimer. A fluorescent pyrene probe is attached to Cys-72 allowing for the dimer-dimer distance to be measured [80]. The method was used to obtain Glycophorin A binding affinity in pure POPC bilayer as well as bacterial membrane; a number of dimerization-disrupting GpA mutants were also tested and their $\Delta\Delta G$ of dimerization relative to wild-type measured [81]. For Glycophorin A, the interaction energetics may further vary with membrane composition and crowding, particularly in cell membranes, as indicated by quantitative-imaging FRET experiments [22].

Estimates of relative binding energy have been obtained using a TOXCAT assay for Glycophorin A TM fragment and a series of Receptor Tyrosine Kinase TM regions [169, 55] TOXCAT is a bioassay probing the affinity of a TM helix-helix dimer of interest (see Figure 1.5). The principle behind the TOXCAT assay is that if the probed helix-helix dimer has high dimerization affinity, it will drive the dimerization of ToxR domains fused to it. That, in turn, leads to chloramphenicol (CAM) gene

expression (ToxR is a dimerization-dependent transcription factor) and makes the expressing cells CAM-resistant. The expression is done in bacterial cells on an antibiotic (selecting) medium.

In the TOXCAT method, dimer affinity is inferred, indirectly, by measuring enzymatic activity. This is in contrast to biophysical methods such as FRET or equilibrium-sedimentation ultracentrifugation, which have an advantage by directly measuring the interaction of interest. The advantage of TOXCAT is throughput: the Finger *et al.* study measures affinity of approx. 50 helix pairs, a feat that would be difficult to accomplish with other methods. While the sheer number of measurements is impressive, the quality of the readout is questionable because the error bars on each measurement are quite high.

Because the expression is done in bacterial cells, and the RTKs used are mammalian in origin it is not clear if the measured ranking translates to the native, biologically-relevant membranes. Furthermore, it is entirely possible in a bioassay that secondary effects occur: insertion of a foreign RTK TM fragment distorts the membrane fluidity, leading to whole-cell gene expression changes. It is challenging to control for these effects in any bioassay, not just TOXCAT.

For larger proteins, such as β-barrels, the only available data for the phospholipase OmpLA system, in detergent micelles, estimating the barrel-barrel dimerization energy to be between -35 and -25 kJ/mol [187]. The oligomeric state of GPCRs has been investigated as well, with earlier studies suggesting rhodopsin (a model GPCR) is freely diffusing [26, 111], and subsequent atomic force-microscopy (AFM) imaging suggesting extensive oligomerization [21].

1.4.2 Protein-lipid interaction energetics in the membrane

The lipid bilayer is a viscous environment, with an oil-like, hydrophobic interior formed by lipid tails. Indeed, the diffusion time within the plane of the bilayer is

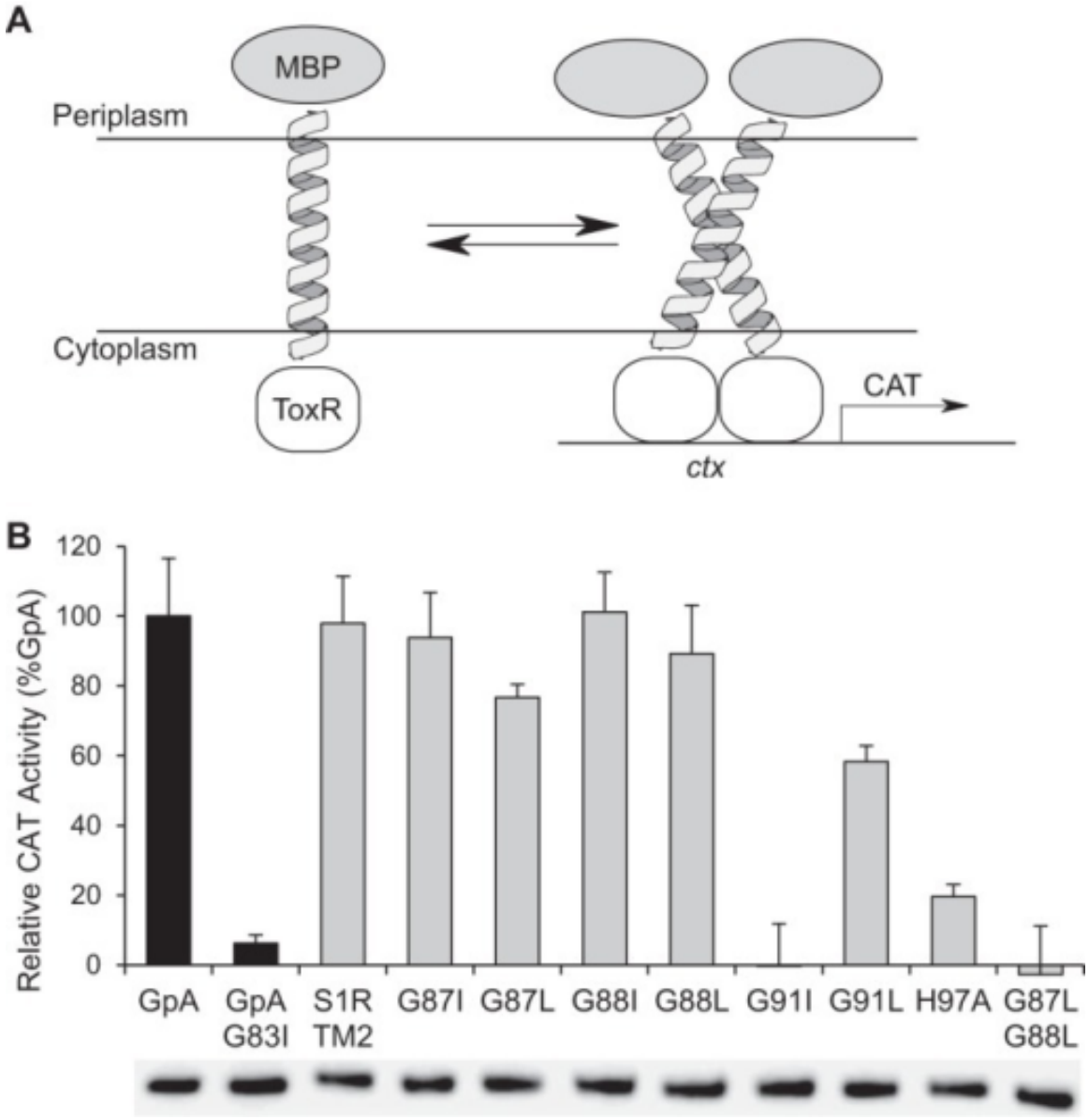


Figure 1.5: A diagrammatic explanation of the TOXCAT assay. Binding of two TM helical regions, brings together ToxR domains, a dimerization-dependent transcription factor. ToxR dimerization triggers expression of *ctx* gene encoding chloramphenicol acetyltransferase (CAT) enzyme. Expression of this protein confers resistance to chloramphenicol antibiotic to the bacteria expressing it, and the level of CAT expression is related to the strength of TM region dimerization. The figure was taken from [65] and shows the assay results for various dimerization disrupting and non-disrupting Glycophorin A mutants.

three orders of magnitude slower than for water [156]. Interactions between proteins and lipids that occur in this environment, underpin the functioning of the cell membrane. Without the proteins, the membrane would be a bilayer of heterogeneous lipids unable to perform many of its biological functions. It is therefore important to understand the protein-lipid interactions and evaluate computational models available, by comparing selected system to available experimental data. As thermodynamics data on binding is relatively sparse, I will rely also on X-ray structures and mutational work along with biochemical experiments to see how well simulations match up with experiment.

Interactions between proteins and individual lipids (rather than just bulk lipid membrane) have been of particular interest due to their medical relevance (this is discussed in more detail in Chapter 3). Despite numerous structural and biochemical studies demonstrating a specific lipid-protein binding, obtaining an experimental K_d measurements within the membrane bilayer has been challenging. Inwardly rectifying Kir channels have been crystallized with a phosphatidylinositol 4,5-bisphosphate (PIP_2) lipid bound, which lead to speculation about a possible lipid role in affecting the channel function[69].

G-Protein Coupled Receptors (GPCRs) are a major class of membrane receptors and frequent drug targets. Purified GPCRs were systematically reconstituted in lipidoparticles with different lipid-group types: it was found that phosphatidylgycerol promoted receptor activation and agonist binding, while phosphatidylethanolamine had the opposite effect [38]. A similar study was performed by a group working on neurotensin (NT) receptor 1 (NTS1). The receptor was reconstituted in zwitterionic (POPC), anionic (1-palmitoyl-2-oleoyl-sn-glycero-3-phosphoglycero, POPG) or mixed (POPC and POPG) bilayers. Here it was found that the agonist NT binding was unaffected by lipid composition. However, the affinity of the activated NTS1 receptor to its G-proteins increased with increasing POPG content. G-protein binding to the

activated receptor is needed for downstream signal transduction [87].

For Receptor Tyrosine Kinases (RTKs), such as the Epidermal Growth Factor Receptor (EGFR), the lipid modulation of receptor activity has been know from biochemical work [133, 28]. Phosphatidylinositol-4,5-bisphosphate (PIP_2) was found to directly affect EGFR activation by binding to juxtamembrane region of the receptor. Levels of PIP_2 were decreased by inhibiting the PI kinase pharmacologically, which lead to a decrease in EGFR tyrosine phosphorylation. However, the opposite was true (EGFR tyrosine phosphorylation was increased) when the levels of PIP_2 were increased by over expressing PIP5K1α [133]. Reconstituting EGFR into proteoliposomes of defined composition, a ganglioside monosialodihexosylganglioside (GM3) was found to promote EGFR activation by suppressing autophosphorylation of the intra-cellular domain, a key inhibitory signal, none of the closely related GM3-like lipids had this effect, and the lipid composition was found to have no effect on ligand binding [28].

1.5 Simulations

Progress in studying membrane proteins has been made despite the challenges in experimental handling. The number of X-ray and NMR structures for membrane proteins steadily increased over the years, providing inputs for simulations [208]. At the same time, improvements in force fields and computational power allowed for greater simulation accuracy and longer timescales, respectively. Simulations of membrane proteins have unique challenges of their own: due to the viscous nature of the lipid bilayer, lateral diffusion within the bilayer plane is slow, making sampling a much greater challenge than in aqueous solution.

A large number of simulations have focused on model systems for which experimental data is available. These simulations were often exploratory, providing intu-

itions about how the protein and lipid species interact in the membrane and how the various protein-protein and protein-lipid interactions might mechanistically drive processes observed in experiments. Another category of simulations focused on quantitatively probing the strength of interactions in the membrane. The aim here was to explicitly compare simulation to experimental data, focusing on simple systems where membrane composition and interacting species were controlled.

It is often challenging to compare computation results with experiment. If the computed result is poorly converged, any conclusions drawn may be due to poor sampling rather than due to the computational model. Simulation convergence is a challenge in the membrane environment, and that is particularly the case for all-atom models. While capturing all the atomistic degrees of freedom, the accessible timescales are approximately $0.1 - 1\ \mu s$ due to computational limitations.

To circumvent the sampling problem , two strategies can be adopted. Firstly, enhanced sampling methodologies can be used that allow exploration of the free energy surface at a faster pace than equilibrium simulations. Secondly, coarse-grained models were developed, which are characterized by much smoother energy landscape and thus can be used to sample much longer timescales, typically $1-10\mu s$. By coarse-graining away the atomistic degrees of freedom, resolution and accuracy is lost but sampling is much faster. This is particularly beneficial for some of the larger systems described in this thesis, for example the Kir channel. The coarse-grained model used is MARTINI as it has been used and validated for protein-lipid simulations at a variety of scales.

The MARTINI model is one of the more popular coarse-grained force fields available, used to model a variety of biomolecules. However, many other CG models exist, some of which are domain specific. Concerning this thesis, perhaps the most relevant model is PACE developed by Schulten and Han [204]. PACE uses a united-atom protein description with a coarse-grained description for the solvent and the lipid bilayer

(both of which use the MARTINI parameter set). However a further detailed description of this force-field is beyond the scope of this thesis and it is only mentioned here for completeness.

To map out if the limitations (if any) of the coarse-grained force fields, I will also perform simulations at all-atom resolution. Inclusion of all the relevant interactions should increase simulation accuracy at the expense of performance. The model chosen here will be CHARMM36, again a popular choice in the field for simulating mixed protein-lipid systems.

In this thesis, I will use both coarse-grained and all-atom models , together with a range of enhanced-sampling methods, to study protein-lipid interactions in model bilayers. Paying particular attention to simulation convergence, I will attempt to quantify the interaction energetics of model helix-helix dimers and compare to experiment when possible. A question that I hope to answer is how reliable are the available models for studying these phenomena and, in case shortcomings are identified, how can the available models be improved to be more realistic.

Chapter 2

Methods

In this chapter, I will introduce the fundamentals of molecular dynamics (MD) simulations. MD simulations have historically been used to study relatively simple systems, hundreds of atoms in size: first MD simulation of liquid water was done in 1974[191], followed shortly by the first *in vaccuo* protein simulation of the bovine pancreatic trypsin inhibitor (BPTI) performed by Karplus and co-workers [130]. Over time larger systems of thousands to million particles became accessible. Those large systems are of relevance because many biologically and medically relevant problems tend to be modelled in that size range.

Chemical systems of interest often complex polymers (such as protein), with multiple degrees of freedom resulting in complex free energy surfaces describing the state-space available to these polymers. These surfaces have two challenging characteristics: firstly, they can contain high energy barriers that are difficult to traverse via equilibrium sampling at physiological temperatures; secondly, their free energy surfaces are rough, slowing down diffusion, posing a challenge given the limited computational timescales available. In this chapter, I will introduce selected methods aimed at enhancing the sampling in molecular simulations. The enhanced sampling methods will rely on defining a suitable, 1D collective variable that describes the reaction of inter-

est, such as protein-lipid binding or helix-helix dimerization. Detailed introduction of an intermolecular D_{RMS} collective variable will follow, as it was used in this thesis to study transmembrane (TM) helix dimerization. I will also describe how to recover an unbiased energy surface from a biased enhanced sampling simulation. Unbiased free energy surfaces, recovered from simulation can be used to estimate population ratios. For example bound/unbound dimer to monomer ratio can be estimated and compared, directly, with the ratio obtained experimentally..

Comparison of simulation to experiment is meaningful when simulations are "converged". Convergence means that the sampling is sufficient that the simulation results do not change with increased simulation time (see Section 2.2.2 for a complete definition and discussion). Converged simulations are also insensitive to initalization from different starting conditions. .

2.1 Molecular dynamics simulations

Computers have dominated the field of molecular dynamics (MD) today. Before that the field was reliant on simple physical experiments: a box of plastic balls was shaken by the investigators (or, more likely, graduate students), who then annotated how the positions of plastic spheres (atoms) changed over time [12]. While primitive, this "shaking" example gives a good intuition to present day simulations: today the program performs the shaking of virtual spheres stored the computer's memory. In this section, I will describe representative numerical techniques used to accomplish that.

2.1.1 Equations of motion

The simulation system is typically composed of particles that model atoms, or groups of atoms. The particles have a point-charge and mass, which does not change over time. The way different particles interact is described by a potential energy function V (or just a 'potential function', 'energy function'). The point-charge particles interact with an energy that is dependent on their position r. In classical MD, Newtonian mechanics is used to calculate the forces on all the particles and their displacement over a short time (called a simulation "timestep"). The force $\vec{F}$ acting on each particle can be calculated by taking the 1st derivative of the energy function with respect to particle position:

$$\vec{F} = -\frac{dV(\vec{r})}{d\vec{r}} \tag{2.1}$$

The acceleration acting on each particle can obtained by rearranging the 2nd equation of motion:

$$\vec{F} = m\vec{a} \tag{2.2}$$

Where m is the mass of particles, and $\vec{a}$ is the acceleration. This calculation can be repeated multiple times over; every repetition is called a timestep. A typical simulation has billions of timesteps. From the forces acting on each particle, acceleration is calculated and used to get the velocities. Over the short duration of the timestep, velocity is used to calculate the particle displacement. More formally, this can be written as:

$$\vec{a} = \frac{d^2\vec{r}}{dt^2} = \frac{\vec{F}}{m} \tag{2.3}$$

Further the velocities and positions at the next timestep can be calculated by:

$$\begin{aligned} \vec{v}(t+dt) &\simeq \vec{a}dt + \vec{v}(t) \\ \vec{r}(t+dt) &= \vec{v}(t)dt + \frac{1}{2}\vec{a}(t)dt^2 + \vec{r}(t) \end{aligned} \tag{2.4}$$

From the equations above, the accelerations $\vec{a}$ and $\vec{v}$ at time t, can be used to calculate the velocities at the next timestep $\vec{v}(t+dt)$; meanwhile, the positions at the next timestep $\vec{r}(t+dt)$, can be calculated from the velocities $\vec{v}$, accelerations $\vec{a}$ and positions $\vec{r}$ at the previous timestep t.

Initial coordinates $\vec{r}$ are obtained from an experimental structure or otherwise derived (for example, by manual model building, or by a homology model). The initial velocities $\vec{v}_0$ are sampled from a Boltzmann distribution at the desired temperature T.

$$v_x = \sqrt{\frac{k_B T}{m}}\nu(0,1) \tag{2.5}$$

where v_x is the velocity along a single dimension, k_B is the Boltzmann constant, m is the particle masses and $\nu(0,1)$ is a normal distribution with mean of 0 and variance

of 1.

To sum up, by using the potential energy function to calculate the forces $\vec{F}$ we can calculate acceleration $\vec{a}$, for a given time t. The acceleration, velocities and positions at time t, denoted as $\vec{a}(t)$, $\vec{v}(t)$ and $\vec{r}(t)$ can be used to calculate the positions after Δt, a small increment of time, and give positions at time $t + \Delta t$.

Two popular algorithms for performing the integration of the equations of motion are Verlet [201] and Leapfrog integrators [79]. The Leapfrog is the default used by GROMACS [160] in this thesis. The integrator splits a single timestep into two half-steps, with alternating updates to positions and velocities :

$$\begin{aligned} \vec{r}(t + \Delta t) &= \vec{r}(t) + \Delta t \vec{v}(t + \Delta t/2) \\ \vec{v}(t + \Delta t/2) &= \vec{v}(t - \Delta t/2) + \Delta t \vec{a}(t) \\ \vec{v}(t) &= \frac{\vec{v}(t - \Delta t/2) + \vec{v}(t + \Delta t/2)}{2} \end{aligned} \tag{2.6}$$

Velocities at time t, denoted as $\vec{v}(t)$, are an arithmetic mean of velocities at previous and next half-timestep; this is only needed because the velocities at time t are not available directly from the integration. The integration updates velocities and positions at alternating half-timesteps, called "position" and "velocity" timesteps. The first two equations are the used to update the positions and velocities at each timestep. The third equation is used to approximate the velocity at the position timestep, but is not part of the integration.

The critical quantity for simulation accuracy and performance is the timestep duration Δt, in units of time, which has to be chosen suitably short to avoid numerical artefacts. This is typically determined empirically, to ensure stable integration of the equations of motions. The key quantity to consider in choosing the timestep is energy drift of the system (computed in the macrocanonical ensemble, NVE, introduced later). The timestep is typically taken to be around 1/10 the oscillation time of the

fastest motion [106], which is related to the mass of the particles and spring constants used to describe the interactions – i.e. the stronger the force constant, or the lighter the particles, the faster the motion and the shorter timestep has to be used, and thus the shorter the timescales accessible.

2.1.2 Numerical techniques

This subsection describes some commonly used techniques that are employed either to speed up the simulations or to ensure that simulations can be compared to experimental systems at infinite dilution.

2.1.2.1 Constraints

Bonds between atoms are typically approximated by a harmonic potential in the energy function (more on this in the following section "Force fields"). When one of the atoms involved in the oscillation has a low mass, this can lead to fairly fast oscillation time for the bond. An example here could be a hydrogen atom attached to either an oxygen atom in a water molecule or a carbon molecule in a peptide chain.

Ideally, to increase sampling time the longest possible timestep is desirable. It is therefore advantageous to eliminate some of the fast oscillations in the system by replacing the harmonic bond term with a constraint term, assuming it does not perturb some important feature of the system. Two popular algorithms for constraining bond lengths and angle are LINCS [76] and SHAKE[170]. Both of these algorithms will be used in this thesis.

2.1.2.2 Periodic Boundary Conditions

For all the simulations described in this thesis, periodic boundary condition is used: only a single copy of the system is simulated, but the particles can freely pass via the box walls to find themselves on the "other side" of the box . The method is known

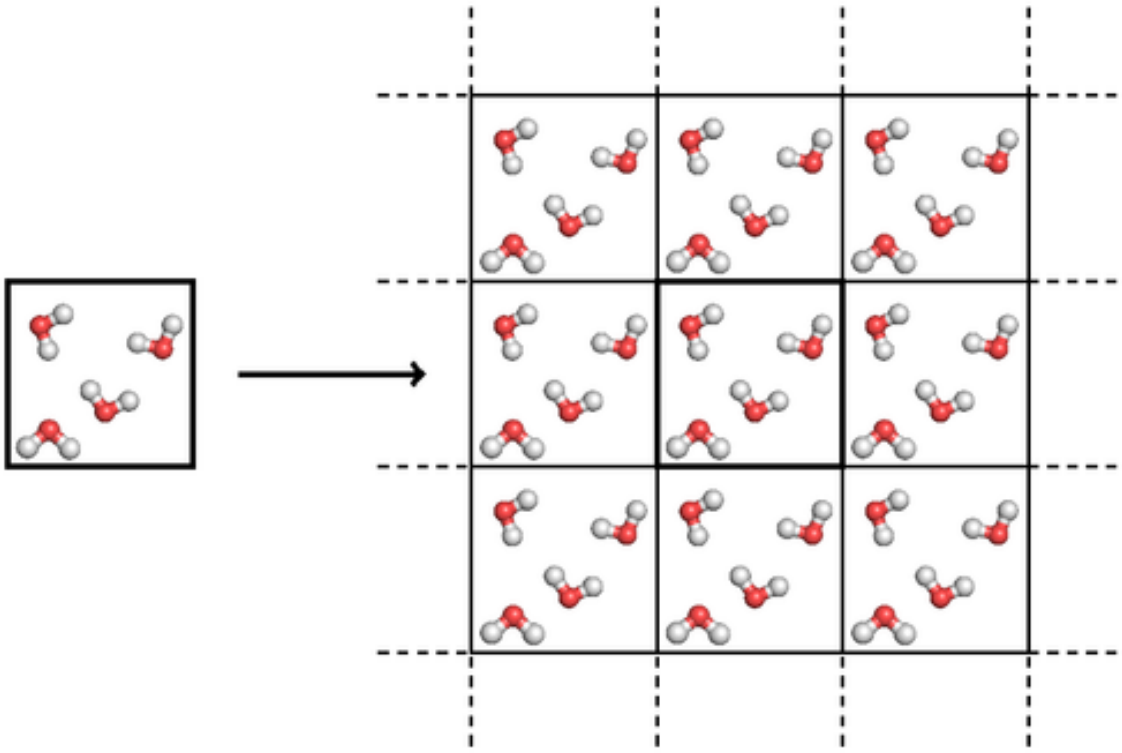


Figure 2.1: Diagram of a simulation box with 4 waters and periodic copies of the system. Taken from http://texample.net/tikz/examples/periodic-boundaries-conditions/, accessed on 2017-08-12.

as periodic boundary conditions. In essence, what is simulated is an infinite lattice of the molecular system, allowing comparison with experiment, as the simulation box is equivalent to a physical system with an equivalent particle concentration. See the Figure 2.1 for a visual aid.

2.1.2.3 Van der Waals interaction

Van der Waals forces are a relatively weak, non-covalent class of interactions between atoms. Due to their weak nature, these are only relevant between neighbouring particles. This interaction is typically modelled using the Lennard-Jones potential function, shown below in one commonly used form:

$$V_{LJ}(r) = 4\epsilon[(\frac{\sigma}{r})^{12} - (\frac{\sigma}{r})^{6}] \tag{2.7}$$

where ϵ is the potential well-depth, r is the distance between the two particles and σ is the distance at which the potential becomes zero. The r^{-6} term is a form of an attractive dispersion interactions, while the r^{-12} repulsion term attempts to model the Pauli exclusion principle, and is also referred to as the exchange interaction. The form of this potential is given here to highlight the r^{-6} distance-dependence and it is later covered in the "Force fields" section (further down).

The primary advantage of the Lennard-Jones (LJ) potential is the computational simplicity, and more accurate (but less performanent) potentials representing the van der Waals (vdW) interaction exist. Beyond a certain cutoff distance, the vdW interaction between two particles is so small it is often neglected. However, to understand which particles are neighbouring, an all-to-all distance calculation has to be performed at every time step, in a naive implementation. The $O(N^2)$ complexity makes this step computationally expensive. Typically, the all-to-all distance calculation can be performed only N steps (typically 10 or so) and used to construct a neighbour-list. Then, only the neighbour particles are included in the calculation. Further, to avoid cutoff effects, the potential energy of the force of the LJ interaction is switched smoothly to zero, between r_{switch} and r_{cutoff}.

Even after the above correction the potential energy would be non-zero at r_{cutoff} had the potential energy not been switched. The value of the Lennard-Jones potential at r_{cutoff} would be very small, 2.5 σ. The effect of this truncation can be quite significant for anisotropic systems (such as large macromolecules or liquid/gas interfaces). This long-term truncation effect can be addressed by applying a "tail correction" of the form [63]:

$$V_{LJ}^{\text{tail}} = \frac{8}{3}\pi\rho\epsilon\sigma^3[\frac{1}{3}(\frac{\sigma}{r_c})^9 - (\frac{\sigma}{r_c})^3] \tag{2.8}$$

where r_c is the cutoff distance, ϵ is the dielectric constant, and ρ is the density when $r > r_c$ and is assumed to be equal to the average number density. σ is the Lennard-

Jones parameter. This correction however assumes isotropic systems, which is not the case for most mixed protein-water systems or systems in phase transitions. Truncation of non-bonded interactions, irrespective of truncation scheme chosen (simple cutoff or force shift/switch) is known introduce nonphysical behaviour at interfaces [58]. A treatment analogous to the Particle Mesh Ewald (discussed subsection 2.1.2.4) has been proposed, using lattice sum Lennard-Jones that avoids the isotropic assumption, and implemented in recent GROMACS version [207].

2.1.2.4 Charge-charge (Coulombic) interactions

Coulombic interactions occur between point charges assigned to the simulated particles. In contrast to the LJ interaction potential which is relatively short-range, the Coulombic interactions decay with r^{-1}, making this interaction relevant at nanoscale distances. The short-range Coulombic component is calculated with a cutoff in real space, while the long-range Coulombic interactions are estimated using the Ewald sum in Fourier space [49]. The particle mesh Ewald (PME) algorithm is used in this thesis to calculate the long-range part of the Coulombic interaction [37, 48].

Avoiding electrostatic interaction artefacts is important in lipid simulations. Use of cutoffs, and truncating the long-range electrostatics, instead of the PME method, can induce a phase transition into an ordered, gel-like phase in DPPC lipid at the temperature a little above its experimental transition temperature [150].

2.1.2.5 Ensembles, temperature and pressure control

A molecular dynamics simulation has be run in a certain ensemble. A popular choice is the grand canonical ensemble, which presupposes that the number of particles N stays constant throughout the simulation. The energy of the particles in the box and the box volume can be subject to additional control. Possible thermodynamics ensembles include:

- the microcanonical NVE ensemble, where the number of particles, box volume and energy are kept constant,
- the canonical NVT ensemble, where the number of particles, box volume and temperature are kept constant,
- the isothermal-isobaric NPT ensemble, where the number of particles, box pressure and temperature are kept contestant.

In this thesis, nearly all simulations were performed under the NPT ensemble. Only control simulations were performed in different ensembles and these are clearly marked in this thesis. Temperature coupling using velocity rescaling with a stochastic term was used [18]. This thermostat is similar to the Berendsen thermostat in spirit [11] but the stochastic term ensures that correct canonical ensemble is generated. This feature is important, as the correct ensemble is needed to properly perform replica exchange simulations, used heavily in this thesis (introduced further in this Chapter, in the "Sampling Methods" section). The pressure was controlled using the Parrinello-Rahman barostat [147]; for lipid simulations, semi-isotropic coupling in XY and Z was used.

2.1.3 Force fields

A force field is composed of a potential energy function and a set of parameters. The potential energy function is a sum of the terms contributing to the energy function. The particular terms are related to the distance between pairs of atoms, or angles between triplets/quintuplets of particles. The parameter set contains all the various coefficients and spring constants that describe these interactions.

For every configuration of the system $\vec{r}$, the potential energy of that configuration $V(\vec{r})$ is a sum of bonded and nonbonded interaction contributions. Bonded terms

include the bond, angle and dihedral energies; nonbonded terms include the Lennard-Jones and Coulomb interaction energies. A simplified potential energy function is shown below:

$$
\begin{aligned}
V &= V_B + V_A + V_P + V_{LJ} + V_C \\
V_B &= \sum_{b=0}^{bonds} \frac{1}{2} k_b (d_b - d_{b,0})^2 \\
V_A &= \sum_{a=0}^{angles} \frac{1}{2} k_\theta (\theta_a - \theta_{a,0})^2 \\
V_P &= \sum_{d=0}^{dihedrals} k_\Phi (1 + \cos(n\Phi_d - \Phi_{d,0})) \\
V_{LJ} &= \sum_i \sum_j 4\epsilon [(\frac{\sigma}{r_{ij}})^{12} - (\frac{\sigma}{r_{ij}})^6] \\
V_C &= \sum_i \sum_j \frac{q_i q_j}{\epsilon r_{ij}}
\end{aligned}
\tag{2.9}
$$

Where V is the potential energy of the system, V_B, V_A, V_P are the the energies of bonds, angles and dihedrals, respectively. V_{LJ} is the Lennard-Jones van der Waals energy, V_C is the electrostatic interaction energy. Where k_b, k_θ and k_Φ are spring constants for bonds, angles and dihedrals respectively. $d_{b,0}$, $\theta_{a,0}$ and $\Phi_{d,0}$ describe the position of harmonic potential minimum and n describes the periodicity of the dihedral term. d_b and θ_a are the distance between atoms and angle respectively, and Φ_d is the dihedral value. The ϵ term describes the depth of V_{LJ} potential, while the ω is radius at which the potential has the lowest energy. For the electrostatic interactions, q_i and q_j are the charges on particles i and j, ϵ is the dielectric constant. r is the distance between particles i and j.

So far, the term "particles" was used to describe interacting species, rather than "atoms". The ambiguity arises because the particles in the simulation may model atoms or groups of atoms. Three major types of force field exist: coarse-grained and

all-atom.

- In all-atom force fields, every particle corresponds to a single atom.
- In coarse-grained force fields, e.g. MARTINI [125], a particle corresponds to at least two atoms.
- In united atom force fields, aliphatic hydrogen atoms attached to a carbon atom are typically represented by a single particle.

Both the coarse-grained and all-atom force fields are described further in the following two sections. Depending on the resolution, a single particle may represent an entire amino-acid residue in peptide chain, in even coarser models.

Force fields are "empirical", meaning they are developed to reproduce experimentally observed properties of molecular systems of varying size. The experimental observables typically include melting points, partitioning coefficients, diffusion coefficients, order parameters, radii of gyration. Overall, a mix of structural, thermodynamic/energetic and kinetic measures is to be reproduced via the numerical model. The force fields are then adjusted to best fit the desired properties.

2.1.3.1 All-atom force fields

A number of force field families exist in this category. One of the main differences between force fields lies in the chemical matter than can be described with them: certain force fields are protein-only, others include the lipid and carbohydrate parameters, as well as descriptions of small-molecules. The long-standing goal in the field of computational chemistry is to provide a general model that can be used in all macro-molecular simulations of biomolecules, often referred to as a "transferable" model. In classical molecular dynamics, these force fields are considered most accurate (due to 1-to-1 particle-atom mapping, matching the physical system). They have been used in simulations of lipids, proteins, and carbohydrates. Notable examples include

studies of small protein folding mechanisms [112, 14]. The main families of all-atom force field include GROMOS [66], OPLS [93], CHARMM [83] and AMBER [27]. In this thesis, recent versions of CHARMM and AMBER force fields are used.

2.1.3.2 Coarse-grained force fields

The coarse-grained force field used in this thesis is the MARTINI force field [125]. MARTINI force field was first developed for simple mixtures of lipids and surfactants [124]. It was then extended to lipid bilayers [125], proteins [136] and carbohydrates [116]. Originally developed to reproduce partitioning coefficients of various solutes in water, it typically maps 4 heavy atoms to one coarse-grained particle. Reduction of atomistic degrees of freedom has a marked effect on the free energy surface, which becomes much smoother. Smoother energy landscape allows for a much larger timestep to be used. Thus, much longer simulation times can be obtained, up to 3 orders of magnitude faster than for all-atom force fields. Additionally, much larger systems can be simulated. The increase in simulation speed comes at the cost of loss of atomic detail.

2.1.4 Detailed settings

The aim of this section was to provide an overview of the theory of molecular simulations, as they relate to the contents of this particular thesis. Detailed simulation settings used are described in the Methods section of each chapter.

2.2 Definitions

2.2.1 Free energy surface

Molecular simulation of polymers (including proteins) is governed by a potential energy function. The potential energy function describes interactions between atoms. The quantities of interest cannot be directly obtained from the potential energy. In other words, the states the system can adopt are not informed by the parameters used in the potential energy. For instance, for a protein simulation the stability of the folded state is only indirectly related to the parameters used by the energy function. Sampling is done along the potential energy function, with the hope to recover the free energy surface (FES), which describes the relative probability of the states the system can adopt.

The free energy surfaces has two characteristics that make it a challenge for computer simulation to obtain FES estimates:

- the energy barriers separating the states of interest, a thermodynamic challenge,
- roughness of the free energy surface, a kinetic challenge.

Obtaining a reliable free energy surface estimate via computer simulations is related to the concept of convergence discussed in the next paragraph.

2.2.2 Convergence

Molecular simulations are often run to characterize the free energy surface of a system under study. It is therefore important to understand how accurate is the estimate computed. A definition of converge might be that the simulation was run for long enough to obtain a representative equilibrium sampling. The closer the estimate to the underlying equilibrium distribution, the more reliable it is. However, the equilibrium

distribution is often the quantity pursued (it is not known). One way to assess convergence is whether the results depend on the chosen initial condition. After sufficient time, simulations started from different initial conditions should converge towards the same answer, a necessary although not sufficient condition for equilibrium sampling to be achieved.

2.2.3 Collective variables

In studying free energy surfaces of molecular systems, it is often useful to define a collective variable (also called a "reaction coordinate" in chemistry). A collective variable (CV) is a projection of the system coordinates onto a lower-dimensional (often 1-dimensional) space. Collective variables are often defined based on pre-existing knowledge of the system or chemical intuition about the process studied. Example of a CV include a centre-of-mass distance between two groups of atoms, or a radius of gyration of a group of atoms or, an RMSD to a reference structure.

Defining a "good" CV is a problem in itself. At minimum, a good CV should identify the kinetically relevant stables state and the free energy barriers separating them. A good CV should limited the overlap between distinct states, as far as possible. Incorrect choice of a CV, or a "bad CV" can make simulations difficult to converge. A collective variable can be used to compute projections of the free energy surface, or coupled with a biasing method in an enhanced sampling simulation (discussed in the "Sampling methods" section below).

Quality of the CVs can be gauged using the committor analysis: for a theoretical two-well landscape, transition trajectories started from any configuration selected at the barrier top should have a fifty-fifty probability of falling into an energy well on either side of the barrier; any deviations indicate a CV imperfection [126].

2.2.4 Definition of an inter-helical CV

In this thesis, I will study helix-helix interactions in membranes via simulation. This section discusses a collective variable called inter-helical D_{RMS}, which I have used in biasing simulations at different levels of resolution.

Two configurations X^A and X^B can be compared using a distance root-mean-square displacement D_{RMS}, as defined below:

$$D_{\mathrm{RMS}}(X^A, X^B) = \sqrt{\frac{1}{N_{cont}} \sum_{i \neq j \in contacts} [d(x_i^a, x_j^a) - d(x_i^b, x_j^b)]^2} \tag{2.10}$$

Where $d(x_i^a, x_j^a)$ is the distance between coordinates x_i and x_j of atoms i and j in configuration A; analogously for configuration B with $d(x_i^b, x_j^b)$. The sum runs for all atom pairs (i,j) in a defined contact list (the length of this list is denoted as N_{cont}).

The all-to-all distance calculation has to be performed at every simulation step, which becomes computationally expensive. Thus the contact list only includes atom pairs within some cutoff distance of each other, defined relative to native distance set. To further limit the set of atoms involved in the calculation, we only include the inter-helix distances in the calculation: contact pair (i,j) is only included in the contact list if i and j are not in the same molecule (hence the name, "inter-helical D_{RMS}").

2.3 Sampling methods

A challenge of molecular simulations is that, for chemical systems of interest, free energy landscapes are compled and the states of interest are usually transient, low probability (high in energy). In an unbiased molecular dynamics simulation, the population of states is determined by the Boltzmann distributions. Suppose a simple system described by a two-well potential, separated by a barrier of known height. The barrier-crossing probability is related to the height of the barrier and to the roughness of the underyling free energy surface. The wait times for observing these transitions (or rare events) increase exponentially with increasing barrier height. One way to address this problem is by running longer simulations. Longer simulations require more powerful computer chips: transistor power has been steadily doubling as per Moore's law [138] to provide ever-faster hardware. Special-purpose built chips such as ANTON [181] allow for further speedups. However, due to physical limits in transistor spacing between lanes, the speed up anticipated from Moore's law has plateaued [120].

A methodological solution to the timescale problem in molecular simulations has been to employ so-called "enhanced sampling" methods. These methods modify the underlying energy surface, by introducing an external bias. The effect of the bias can then be removed after the simulation, in post-processing, allowing the free energy surface of the system to be recovered.

2.3.1 Replica exchange

In replica exchange, a number of copies (often termed "windows") of the system is simulated. The copies have the same number of atoms and contain the same chemical system (same number of molecules of identical type). What is different between the replicas is the Hamiltonian (the energy function) or the temperature. The form of

the energy function is the same across the copies but different parameters are used. In replica exchange simulations, every N molecular dynamics steps, an exchange between replicas is attempted. The exchange attempt is a separate Monte-Carlo step. Exchanges are typically performed between neighbouring replicas (here labelled as '1' and '2').

Let us first consider the case of Temperature Replica Exchange MD (T-REMD), where the only thing different between replicas is their temperature (the Hamiltonian is the same). The exchange is accepted or rejected according to a Metropolis criterion:

$$P_{acc}(1 \leftrightarrow 2) = \min(1, \exp[(\frac{1}{k_B T_1} - \frac{1}{k_B T_2})(U_1 - U_2)]) \tag{2.11}$$

Where $P_{acc}(1 \leftrightarrow 2)$ is the probability of an exchange occurring between replicas 1 and 2, U_1 and U_2 are the potential energies of replicas 1 and 2, respectively; k_B is the Boltzmann constant, T_1 and T_2 are the temperatures of replica 1 and 2, respectively.

The simulations presented in this thesis were typically run at the NPT ensemble. The NPT replicas run at higher temperature, which lead to increase in volume of the simulation box[178]. The exchange criterion has to the be corrected for a volume/pressure component for simulations run at isobaric-isothermal ensemble:

$$P_{acc}(1 \leftrightarrow 2) = \min(1, \exp[(\frac{1}{k_B T_1} - \frac{1}{k_B T_2})(U_1 - U_2) + (\frac{P_1}{k_B T_1} - \frac{P_2}{k_B T_2})(V_1 - V_2)]) \tag{2.12}$$

Where P_1 and P_2 are the target/reference pressures for the replicas and V_1 and V_2 are their volumes at the time of exchange.

When the energy difference between the replicas is very high, the probability of exchange becomes near zero; when the energy difference is small, the exchange is near certain.

In the case of T-REMD, the only difference between replicas is the temperature

at which they are run at. The high-temperature replicas are thought to sample faster. The temperature of the last replica has to be then chosen suitably, typically by trial and error, to smooth the energy surface sufficiently for the transition to occur at reasonable timescales. Then the overlap between replicas has to be adequate to allow the "hot" replicas diffuse down the temperature ladder to the 1st replica, usually corresponding to the conditions of interest.

The performance benefit of T-REMD scales poorly with increasing system size. As the number of particles increases in the system, the potential energy grows larger. The overlap between replica energies becomes smaller and smaller, with increasing system size, for the same temperature ladder. To address this limitation Berne and co-workers have introduced the REST1 and REST2 methods [113, 205].

In the REST-type methods, instead of changing the system temperature the Hamiltonian is locally deformed: a scaling factor λ is used to construct a ladder, scaling down interaction terms and increasing sampling.

For REST1, the potential energy E of a configuration X in replica m can be expressed as:

$$E_m^{\mathrm{REST1}}(X) = E_{pp}(X) + \sqrt{\frac{\beta_0 + \beta_m}{2\beta_m}} E_{pw}(X) + \frac{\beta_0}{\beta_m} E_{ww}(X) \tag{2.13}$$

Where E_{pp}, E_{pw} and E_{ww} are the protein-protein, protein-water and water-water interaction energies, respectively. β_m and β_0 are the $\frac{1}{k_B T}$ for replica m and 0 respectively.

For REST1, the acceptance probability for an exchange between replicas m and n can be calculated using:

$$\begin{aligned} \Delta_{mn}(\mathrm{REST1}) = (\beta_m - \beta_n)[(E_{pp}(X_n) + \frac{1}{2}E_{pp}(X_n)) - \\ (E_{pp}(X_m) + \frac{1}{2}E_{pp}(X_m))] \end{aligned} \tag{2.14}$$

where X_m and X_n are the configurations of replicas m and n respectively, and β_m and β_n are the scaling factors applied to replicas m and n respectively.

For REST2, a slight change was made, and the potential energy of a replica is defined as:

$$E_m^{\mathrm{REST2}}(X) = \frac{\beta_m}{\beta_0} E_{pp}(X) + \frac{\beta_m}{\beta_0} E_{pw}(X) + E_{ww}(X) \tag{2.15}$$

Where are the terms are as per REST1 above.

For REST2, exchange between replicas n and m, the acceptance ratio can then be calculated using only the scaled components of the Hamiltonian:

$$\begin{aligned} \Delta_{mn}(\mathrm{REST2}) = (\beta_m - \beta_n)[(E_{pp}(X_n) - E_{pp}(X_m)) + \\ \frac{\sqrt{\beta_0}}{\sqrt{\beta_m} + \sqrt{\beta_n}} (E_{pw}(X_n) - E_{pw}(X_m))] \end{aligned} \tag{2.16}$$

Where all the terms are as per REST1, above.

The scaling is only performed on the selected fragment of the system, for example the protein, or in the case of this thesis on the bilayer POPC lipids. Energy overlap between neighbouring replicas improves, as only a small perturbation is done to the potential energy (used to calculate exchange probability).

In the first version of the algorithm, the REST1, both the temperature and the potential energy of each replica are modified. In REST2, on the other hand, the modification is done only to the potential energy function (individual force field terms are modified).

Berne and co-workers have successfully used the REST2 method to study small protein folding, using Trp-cage and GB1 β -hairpin as model systems.

2.3.2 Umbrella sampling

Another popular enhanced sampling method is umbrella sampling (US). US requires prior information about the chemical system to define a collective variable: a reduced representation of the system along which a bias can be applied to drive transitions across the energy landscape.

In T-REMD, the potential energy is the equivalent of a collective variable (CV). In the context of this thesis, where protein interactions within the membrane are studied, a frequently used coordinate is the centre-of-mass distance between two species - for example, two helices dimerizing in the membrane. Once a CV has been chosen it can be used to add a harmonic biasing potential to the Hamiltonian of the simulation:

$$V'(\vec{x}) = V(\vec{x}) + U(r(\vec{x})) \tag{2.17}$$

Where V is the original potential energy, a function of all the coordinates, represented by $\vec{x}$, U is the umbrella potential, a function of the collective variable $r(\vec{x})$, and V' is the total potential energy.

In the absence of an umbrella potential, the free energy can be obtained by defining the effective 1D potential on r, run with an unbiased MD simulation with the potential $V(\vec{x})$. The free energy $F(r)$ along r is then:

$$F(r) = -\log(P(r)) + const \tag{2.18}$$

where P(r) is the distribution of r from an infinitely long unbiased run.

Usually multiple windows are simulated with exchanges allowed between neighbouring copies of the system. The functional form of the perturbation potential $U(r)$ is commonly a harmonic function:

$$U(r) = \frac{1}{2}k(r - r_0)^2 \tag{2.19}$$

Where k is the spring constant, r is the position of the umbrella window in the reduced, 1D coordinate at any given time and r_0 is the minimum position of the bias. Multiple umbrella windows are typically run at the same time. The umbrellas are usually linearly spaced along the CV, with a suitable spring constant. The spring constant is empirically determined to be sufficiently rigid to ensure that copy stays close to the umbrella center for a given window, but also soft enough to allow for energy overlap (and thus exchanges) with neighbouring windows.

Umbrella sampling can be combined with replica exchange: it is then the difference in $U(r)$ bias energy between neighbouring replicas that is used to calculate the exchange probability (see Subsection 2.3.1). Replica exchange umbrella sampling (REUS) is particularly advantageous if the collective variable is imperfect, which is often the case in practice. The exchanges between neighbouring window allow the system to escape local minima, orthogonal to the biasing variable.

The umbrella sampling potential is history-independent i.e., it does not change with simulation time. This is in contrast to methods such as metadynamics, described in the following section.

2.3.3 Metadynamics

Metadynamics is an enhanced sampling method that requires a collective variable, much like umbrella sampling. Metadynamics is a generalization of umbrella sampling: instead of applying a fixed series of harmonic biases, the energy landscape is accumulated at simulation run-time by depositing "hills" of energy [104]. The energy "hills" are Gaussian of a certain height (in units of energy) and certain width (in units of the CV) deposited as the simulation progresses. The accumulation of these energy hills at a particular value of the CV will then bias the system to move away from that region. The free energy surface is estimated by taking the negative of the sum of the Gaussian hills. Thus, the simulation is run with an additional bias potential of the

form:

$$V(\vec{s},t) = \sum_{j=0}^{\tau<t} W \exp(-\sum_{i=1}^{d} \frac{(\vec{s}-\vec{s}_j)^2}{2\sigma_i^2}). \tag{2.20}$$

Where $V(\vec{s},t)$ is the energy of the system at time t as a function of collective variable vector $\vec{s}$. The sum runs overall accumulated Gaussians: W is the Gaussian height, τ is the number of Gaussians deposited prior to t, σ is the width of deposited Gaussian and d is the total number of CVs along which the Gaussians are deposited.

The free energy can then be recovered, assuming an infinitely long simulation, which build up the $V(\vec{s},t)$ potential:

$$-\lim_{t\to\infty} V(\vec{s},t) + const \simeq F(\vec{s}) \tag{2.21}$$

where $F(\vec{s})$ is the free energy along a collective variable $\vec{s}$.

The variant of metadynamics described above is the classical approach. One problem that is apparent with this approach is that the hills never stop being accumulated. Taking the infinite limit of simulation lenght, an infinite amount of energy will be deposited as Gaussians. This problem is referred to as "over-filling". Additionally, under specific conditions classical metadynamics can converge to the wrong free energy surface.

A variant of the method was proposed called well-tempered metadynamics where the height of the Gaussians decays as a function of simulation time [7]. This results in correct asymptotic behaviour: as the simulation is run "forever" the free energy surface estimation will converge.

$$V(\vec{s},t) = \sum_{j=0}^{\tau<t} W e^{-V(\vec{s},t)/\Delta T} \exp(-\sum_{i=1}^{d} \frac{(\vec{s}-\vec{s}_j)^2}{2\sigma_i^2}) \tag{2.22}$$

where all the symbols the same as in Eq. 2.20 but the height of the Gaussian is no longer fixed but instead decays smoothly: $-V(\vec{s},t)$ is the estimate for the bias energy

for the current CV position at time t, tuned down by a so-called bias-factor ΔT. The bias factor ensures that hills will not be excessively accumulated at a single position in CV space. Metadynamics has been proven to be converge to the correct equilibrium distribution at infinite run time. A number of conditions has to be met for this to be true: (1) the metadynamics system does not "jam" in response to external bias, (2) the application of the bias does not make the system unstable, (3) the bias applied has to be properly bound over the course of the simulation. All of the above conditions are easily satisfied in existing implementations of the method. However, the convergence is not guaranteed to be "quick", in practical terms. This is true particularly with a poorly chosen collective variable.

Metadynamics leads to a build-up of a history dependent potential, accumulated at run-time. However, a time-independent procedure has been proposed by Tiwary, to complement well-tempered metadynamics [193]. Over the course of a long metadynamics simulation, the bias accumulated is equal to the true free energy surface plus a time-dependent constant. The authors determine the relationship between the time-dependent term and simulation time, and use that relationship to estimate the local FES convergence.

2.3.4 Generating initial conditions for umbrella sampling

This thesis presents a number of replica exchange umbrella sampling (REUS) simulations. The simulations probe various protein-protein and protein-lipid complexes. Explanation below is done using an example of a ligand in a binding site but applies to other cases presented in this thesis. Diagrammatic explanation of how the initial condition are generated is shown in Fig. 2.2. What is referred to as the "bound" or "together" configuration comes usually from an experimental structure or otherwise a structural model. To generate the unbound state, a collective variable is defined such as the distance between the binding site and the ligand. A single umbrella potential

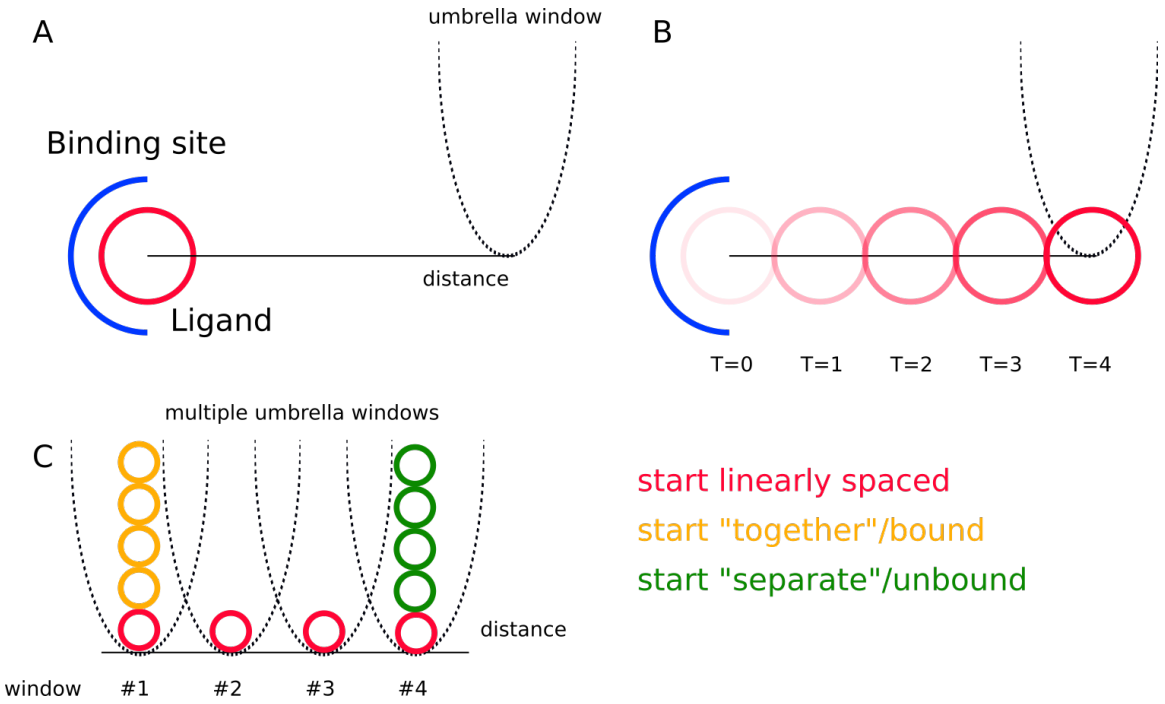


Figure 2.2: Generation of initial configurations for umbrella sampling simulations, as used in this thesis. (A) Starting from the "bound" state a single umbrella is positioned along a CV away from the binding site. (B) Over the course of the simulation, the ligand is moved towards the target position of the umbrella. Intermediate configurations along the CV are saved. (C) Three initial configurations are used in this thesis. All three share the same umbrella ladder, the only difference is the set of particle coordinates used to initialize the windows. In the "start bound" all windows start in the bound configuration, vice-versa for the "start unbound". In the linear/"default" case, the coordinates of the system are chosen to be close to each windows' target umbrella minimum.

is applied, centered at a distance far away from the binding site (see Fig. 2.2A). The distance is chosen by trial and error to ensure ligand is fully unbound. A single copy simulation is then run: the applied potential leads the ligand to move towards the target umbrella distance (away from the binding site). As the ligand is moving towards the target distance, intermediate configurations are saved (Fig. 2.2B, T=0 through to T=4). Lastly, this generates the "unbound" or "separate" configuration.

Alternatively, to generate an unbound state a constant force-simulation can be performed: the difference to the harmonic umbrella being that the force is fixed rather than distance-dependent.

Having described the procedure for generating a range of configurations, spanning the bound and unbound states, lets discuss how they relate to initializing the umbrella sampling simulation. For an umbrella ladder used in REUS simulations, all the potential window are spaced uniformly along the collective variable, with an identical spring constant. However, it is important to draw a distinction between the set of coordinates used to initialize the system and the center of the potential applied to each replica. The most popular approach in the field is for the window coordinates to be close to the target position of the umbrella potential; in this thesis this is referred to as the "linear" or "default" initial condition (Fig. 2.2C). The addition to that, in this thesis, is to use the same umbrella ladder with coordinates of all replicas in either "bound" or "unbound" configurations. This causes the edge widows to be far away from their target distance, and initially in the REUS simulation a large force is excreted to move these closer to the target window distance.

All this trouble is taken to ensure that the simulation result, in case of REUS simulations the PMF, is insensitive to initial conditions. Care must be taken to ensure that the initial pulling of edge windows, in "start together" and "start separate" simulations does not induce non-equilibrium changes, in the case of all-atom helix-helix system that would be (for example) helix unfolding.

2.3.5 Recovering the unbiased free energy surface

Enhanced sampling methods rely on modifying the Hamiltonian by adding a biasing potential along a defined CV. What is obtained over the course of those simulations is the apparent free energy surface $V'(r)$. Given that the size of added bias is known it should be possible to account for the bias and recover the free energy surface, in the absence of bias $V(r)$. For metadynamics this is reasonably straightforward: the inverse of the sum of accumulated hills is the free energy surface. The process is a bit more complicated for unbiasing umbrella sampling simulations. Conceptually, for

each window of the US simulation, the free energy curvature along the biasing CV is known. Assuming an adequate overlap of neighbouring windows, their ends can be aligned in a self-consistent way. Such "stitching together" can be performed via weighted histogram analysis method WHAM, see [100] for derivation. An alternative estimator, derived from WHAM exists and is called MBAR: it follows from the original WHAM paper[100], assuming zero bin-width histograms, a limiting case in WHAM [182].

2.3.6 Re-weighting along different collective variables

Once the free energy surface has been recovered via a WHAM calculation, or otherwise, the weight of each frame can be calculated in the equilibrium ensemble. This weight w is given by, for each frame $\vec{x}_i$:

$$w(\vec{x}_i) = \exp[-\frac{f_w - U_{bias}(r(\vec{x}_i))}{k_B T}] \tag{2.23}$$

Where f_w is the free energy of that window returned by WHAM, $U_{bias}(r)$ is the umbrella energy of configuration $\vec{x}_i$.

The simulation frames can then be histogrammed using the weights $P(r)$ along the original biasing coordinate. An identical PMF should be obtained as from the WHAM calculation. Using the weights obtained via the WHAM procedure, other collective variables can then also be projected along and plotted, either separately or in conjunction, with the original collective variable.

2.4 Comparison to experiment

Enhanced sampling tools are used in molecular simulations to provide an improved estimate of the equilibrium distribution of the system The enhanced sampling can be either done along a collective variable, or by a CV-free methods such as T-REMD. Assuming that estimate of the distribution is well-converged (for example by computing a PMF that is independent of initial conditions), it becomes useful in comparing the simulation-derived estimate to experimental measurement. A number of considerations apply, including: how does the experimental measurement relate to simulation (is the same thing being computed and measured experimentally), are the conditions comparable? In the context of this thesis, I was interested in how reliable the computational models are for studying the energetics of protein interactions within a membrane. In particular, how the computed dimerization constants compare with experimental estimates.

2.4.1 Experimental K_d for helix-helix dimerization

The dissociation constant K_d of a homodimeric species in a membrane can be defined as:

$$K_d = \frac{(1-y)[A][A]}{y[AA]} \tag{2.24}$$

Where $[A]$ and $[AA]$ are the concentrations of the monomeric and dimeric form, respectively, per lipid area. Fraction of time when the dimer is bound is defined as y. Eq. 2.24 can be derived by requiring that either:

1. chemical potential (time- or ensemble-averaged) of the monomer and of the dimer form should be equal at equilibrium (and assuming no effect on pressure of the free energy changes), or

2. forward- and reverse-fluxes (time averaged) for the dimerization reaction to be equal at equilibrium.

In this case, $[AA] = \frac{[A]}{2} = \frac{1}{\sigma}$, where σ is the lipid area explored in simulation.

This lipid area is the area of a circle with radius L, defined as mean helix-helix centre-of-mass distance in the last window of the umbrella sampling simulation.

The dissociation constant K_d can then be defined as:

$$K_d = 2[A]\frac{1-y}{y} = \frac{4}{\sigma}\frac{1-y}{y} \tag{2.25}$$

The reference concentration of 1 molecule per nm^2 is used. The standard free energy of association can the be calculated straightforwardly using $\Delta G = -RT \ln K_d$. A similar derivation was obtained by De Jong et al. also for heterodimers [39] (see also [75]).

The fraction bound can be calculated by suitably integrating the population distribution, obtained by Boltzmann inversion of the PMF:

$$y = \frac{\int_0^{d_c} e^{-\beta F(D_{\mathrm{RMS}})}\, dD_{\mathrm{RMS}}}{\int_0^{L} e^{-\beta F(D_{\mathrm{RMS}})}\, dD_{\mathrm{RMS}}} \tag{2.26}$$

Where $\beta = \frac{1}{k_B T}$, k_B is the Boltzmann constant and T is the absolute temperature. $F(D_{\mathrm{RMS}})$ is the PMF calculated along D_{RMS}. d_c is a cutoff value for what constitutes a bound state: everything below this value is considered bound.. d_c is chosen suitably to match the end part of the PMF where further separation does not change the free energy value. The value of y should ideally not be sensitive to the choice of d_c.

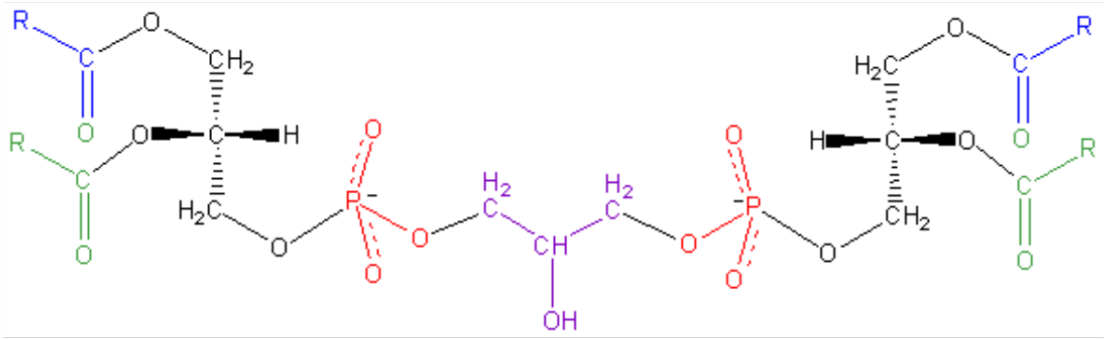
A related method of K_d estimation is shown in the methods section of Chapter 5.

Chapter 3

Protein-lipid interactions in the membrane

The two systems I want to focus on in this chapter are the adenine nucleotide translocase (ANT), a metabolite transporter found in the inner membrane of animal mitochondria, and the inwardly rectifying Kir2.2 channel involved in signal transduction. For both of these membrane proteins, lipid interactions have a crucial impact on biological function. The ANT translocase interacts with cardiolipin (CL), a lipid with four acyl chains instead of the usual two (see Fig. 3.1A for its chemical structure), which is thought to regulate the activity of a number of mitochondrial proteins, such as the membrane-embedded respiratory complexes (I, III, IV, and V) and other carrier family members (phosphate carrier, uncoupling protein); in addition CL is thought to interact with cytochrome c and creatine kinase [174]. . The Kir2.2 channel binds a negatively charged phosphatidylinositol 4,5-bisphosphate (PIP_2, see Fig. 3.1B for its chemical structure), which affects the channel conductance. The PIP_2 lipid contains a poly-anionic headgroup that can be present at variable protonation state under physiological conditions (pH of about 7), with the phosphatidylinositol headgroup either fully deprotonated (charge -4) or singly protonated (charge -3). Kir2.2 mutants are

A



B

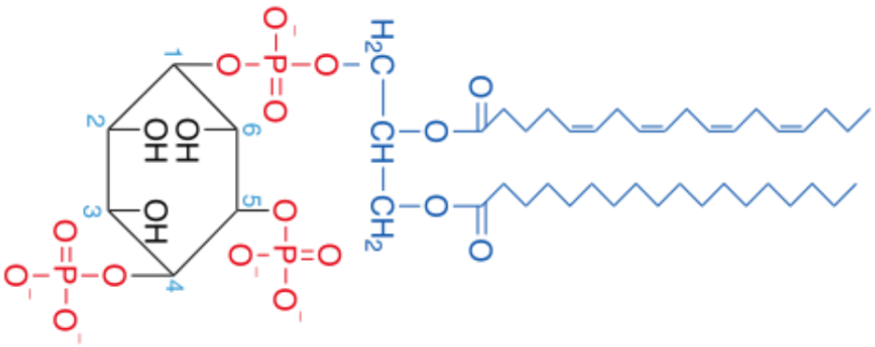


Figure 3.1: Chemical structure diagrams of lipids studied in this chapter (A) Cardiolipin (CL), with R-group indicating an acyl chain, (B) Phosphatidylinositol 4,5-bisphosphate (PIP_2). Taken from `https://en.wikipedia.org/wiki/Cardiolipin` and `https://en.wikipedia.org/wiki/Phosphatidylinositol_4,5-bisphosphate`, respectively, accession date 1st March 2018.

known to have a diminished sensitivity to PIP_2 [115, 215]. For both of these systems, I will estimate the binding affinity from a PMF calculation using a coarse-grained MARTINI model and compare with experimental data to see if there is, at least qualitative, agreement between simulation and experiment.

3.1 Introduction to ANT

Mitochondria are often called the 'powerhouse' of the cell [184]. Proteins in the electron transport chain drive the formation of electrochemical gradient across the mitochondrial inner membrane. That gradient is then harnessed to generate ATP , a molecule often referred to as the "energy currency" of the cell [97]. An essential part of that process is the exchange of ADP for ATP across the inner mitochondrial membrane [96] (see Fig. 3.2B). This exchange is performed by mitochondrial adenine nucleotide translocase (ANT, known as the ADP/ATP carrier, AAC). ANT is the most abundant member of a mitochondrial carrier (MC) family [101], which transport various metabolites across the mitochondrial membrane [146].

ANT is known to be functionally dependent on an interaction with cardiolipin, a lipid abundant in the mitochondrial membrane[95, 25]. CL is of broad medical interest: impairment of its synthesis is known as the Sengers syndrome, symptoms of which include muscle weakness and hypertrophic cardiomyopathy [67]. The molecular biology of the disease is an apparent depletion of the ANT protein from the mitochondrial membrane [129, 67], further highlighting the connection between ANT and CL.

At least two different lines of evidence point to a direct ANT-CL interaction. A NMR ^{31}P binding assay showed that CL interacts with bovine heart translocase [15]. All available structural models of the ANT and related translocases required CL to be present during crystallization and purification. The structural models themseleves were found with the lipid bound to sites on the periphery of the proteins [151, 145]. Available structures include crystallographic models of yeast and bovine ANT, both in a cytoplasm-facing conformation [151, 145]. An NMR model is also available for uncoupling protein 2 (UCP2) [10], a related protein from the mitochondrial carrier family.

The ANT protein-family architecture has a topology containing 6 TM helices

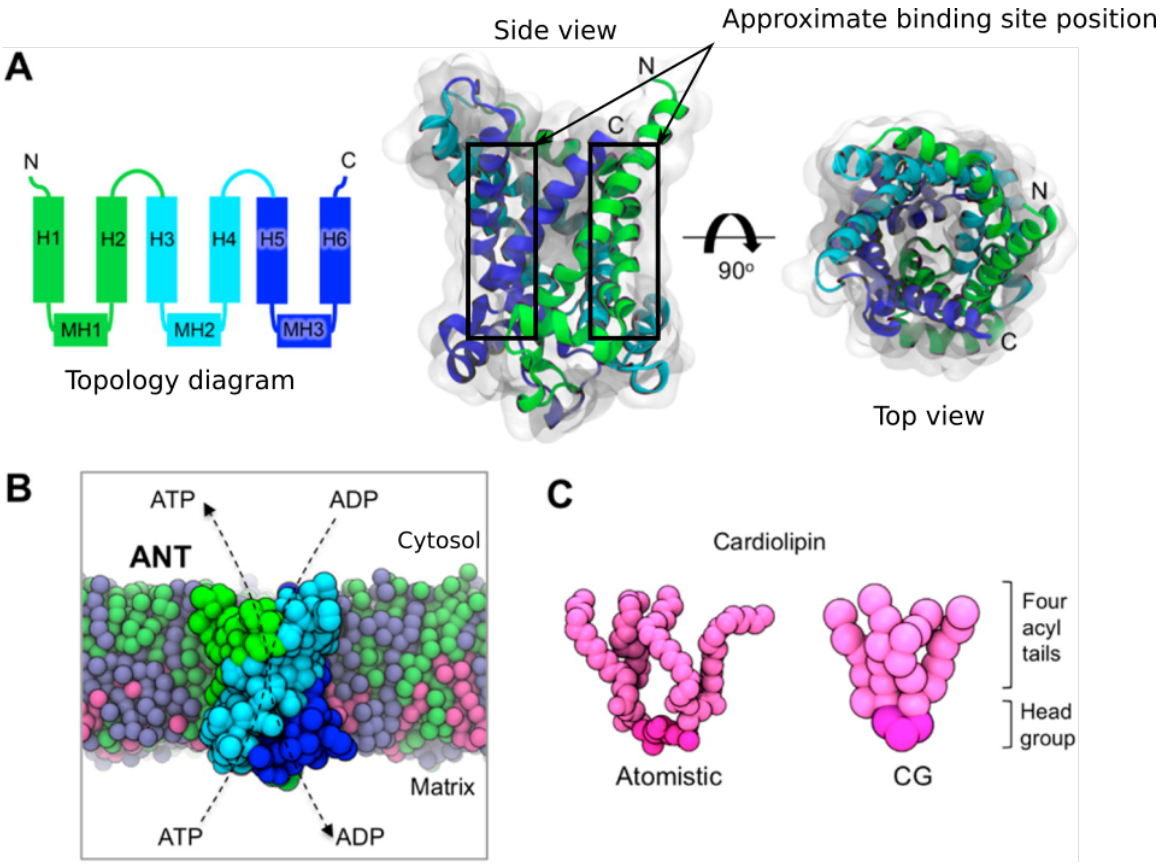


Figure 3.2: (A) Diagram of the topology of ANT with 6 TM and three amphipathic helices (MH1-3); rendering showing side- and top-views of the protein. Approximate location of two binding sites visible is indicated with black rectangles. (B) ANT embedded in a coarse-grained lipid bilayer, protein structure was taken from the cytoplasmic-facing configuration (Protein Data Bank entry 1OKC, [151]); Arrows indicate ATP/ADP exchange reaction. (C) Diagram showing the all-atom and coarse-grained representations of cardiolipin. Reproduced with permission from [73], CC-BY 4.0 license.

(H1–H6) with a pseudo 3-fold symmetry; 3 amphipathic helices (MH1–MH3) form the substrate binding sites (see Fig. 3.2A).

Despite the available biochemical and structural evidence suggesting a link between CL and mitochondrial carrier family protein, characterizing this interaction remains difficult. Computer simulations can be used to study protein-lipid interaction in model membranes [75]. Of particular usefulness are the coarse-grained simulations using a simplified representation of the atomistic system. Through the reduced number of particles, coarse-grained simulations are advantageous for large biologically-

relevant systems (see Fig. 3.2B for an example of a large-scale system). Simulations have been employed to help and identify lipid binding sites [71], as well as probing energetics [140, 72, 5] and kinetics of protein-lipid interactions in the membrane[43]. Other prominent uses of coarse-grained simulations include prediction of the PIP_2 binding site on the Kir channel [190], which was later confirmed via crystal structures [175, 69]. All known Aquaporin (Aqp) structures have also been characterized by simulations, suggesting a conserved lipid binding patterns on Aqp-family protein surface exposed to the membrane [188].

Here I apply replica exchange umbrella sampling (REUS) simulations, using a coarse-grained MARTINI force field, to obtain a PMF for CL binding to a site on ANT in the lipid bilayer. The binding energy was estimated for a single representative binding site on ANT. This result will likely be similar across all the sites, because of a high degree of homology between them (ANT is pseudo 3-fold symmetric); additionally, the crystal structure of bovine ANT1 (PDB entry 1OKC, used in this chapter) shows all three sites occupied by density interpreted as bound CL molecules. Having said that, it is possible that subtle differences exist between the CL-binding sites.

The experimental evidence from NMR binding measurements and crystallographic structures (described above) is qualitative but points to a stable binding site. I compare the energetic and structural characteristic based on the free energy profile estimate to available experimental data, as well as, investigate simulation convergence and reliability.

3.2 Introduction to Kir2.2

Protein channels allow for a selective or non-selective flow of solutes across the membrane. The passage of solutes occurs without the use of energy, down a concentration gradient. Ion channels, in particular, are an important group of proteins involved in

nerve transmission and muscle contraction. Ion channels can open and close in response to various stimuli [13], such as ligand binding, membrane stretching or heat [20]. Ion channels are also sensitive to their lipid environment: for instance, the bacterial KcsA potassium channel requires negatively-charged lipids types for its activity [74, 198].

Inwardly rectifying potassium (Kir) channels are capable of passing the ionic current more easily into the cell than out of the cell. Kirs are potassium-selective channels that play biologically important roles in the kidneys, insulin secretion and muscle contraction [42, 142, 164].

Kirs are homotetramers, with each subunit containing 2 TM helices, M1 and M2, with crystal structures available [102, 144]. The two helices are connected by a loop that forms part of the channel selectivity-filter, a region of the pore that mediates which types of ions are allowed to pass through. Both C-terminal and N-termial domains are intracellular; N-terminal is connected to M1 via an amphipathic M0 helix. N- and C-terminus form the intracellular domain [143], which regulates the gating of the channel and is affected by range of factors such as pH, ligand binding and interaction with regulatory proteins.

Phosphatidylinositol 4,5-bisphosphate (PIP_2) is thought to affect all members of this family [52, 114]. The effects of PIP_2 range from inhibition of the prokaryotic channels KirBac1.1 and 1.2 [47, 23] to activation of the mammalian Kir channels [53].

A structural model of PIP_2 binding to Kir has been difficult to obtain, as the lipid could not initially be co-crystallized with the protein. However, mutational and functional investigations of the channel have revealed that residues from both C- and N-terminal are involved in PIP_2 binding [183, 31, 30, 176, 91, 166]. These residues are often positively charged and thus likely to interact with the negative headgroup of PIP_2 lipids. Direct evidence from crystallographic models has however been lacking.

Given the key role played by PIP_2 in regulating membrane phenomena [131], including regulation of ion channel function [68], I performed PMF calculations of PIP_2 binding to the Kir2.2 ion channel TM region. Coarse-grained simulations have been used previously to characterize the binding of PIP_2 to the Kir-type channel [190, 175]. Additionally, experimental data exists for mutations that reduce PIP_2 binding to the channel [68, 115, 215]. Together with the greater complexity of PIP_2 headgroup, relative to CL, this interaction makes for an interesting benchmark of the coarse-grained force field and sampling methods available.

3.3 ANT–Cardiolipin interaction

3.3.1 Material and methods

Simulation parameters used in this chapter are shown in Table 3.1. The MARTINI 2.1 coarse-grained force field was used in this chapter [136, 125, 124], with a 4-tail bead POPC model. Cardiolipin parameters were taken from [35].

The system was built with the protein, centered and position restrained (using a force constant of $1000\,\mathrm{kJ/mol/nm^2}$), with a POPC bilayer around as described in [73]. The system contained approximately 4200 coarse-grained water molecules with 280 POPC lipids, a single POPC molecule in the upper-leaflet was replaced with a CL lipid. To neutralize any net charged in the system, randomly selected water molecules were converted into sodium and chloride ions. A collective variable was defined as the distance between CL glycerol bead (GL0) and a proline backbone bead from the Px[D/E]xx[K/R] motif found in the binding site (see Fig. 3.3A). Steered molecular dynamics (SMD) simulations were performed to generate the initial configurations of the system , with the CL molecule pulled away from the protein at the rate of 0.1 nm/ns (with a spring constant of $1000\,\mathrm{kJ/mol/nm^2}$, to a distance of about $3-4$ nm (as described in [73]). The biasing was done along the distance CV, roughly perpendicular to the binding site, in order to prevent sampling around any non-native binding sites. Coordinate system was kept fixed by placing an absolute position restraint was on three Proline residues found in the equivalent binding sites with a $400\,\mathrm{kJ/mol/nm^2}$ spring constant (as per [73]). To prevent the CL lipid from diffusing away from the 1D biasing path, a Y-component harmonic restraint was added with a spring constant of $1000\,\mathrm{kJ/mol/nm^2}$. Replica exchange umbrella sampling was performed with windows distributed uniformly along the distance coordinate between 1.6 and 4.5 nm with a $1000\,\mathrm{kJ/mol/nm^2}$ spring constant. Exchanges were attempted using the Metropolis criterion every 10000 steps.

Table 3.1: Summary of methods used in coarse-grained simulations

GROMACS version	4.6.7
Time-step	20 fs
Pressure control	Semi-isotropic in XY and Z, reference pressure of 1 bar, 12 ps coupling constant
Pressure control (equilibration)	Berendsen
Pressure control (production)	Parrinello-Rahman
Temperature control	Stochastic velocity rescaling temperature control (V-rescale), target temperature of 310 K, characteristic time of 1 ps
Lennard-Jones (LJ) interactions	Cut-off at 1.2 nm, switched-off smoothly between 0.9 and 1.2 nm
Coulombic interactions	Particle Mesh Ewald (PME) method with real-space cutoff of 1.2 nm and grid spacing of 1.2 nm

3.3.2 Results and discussion

The ADP/ATP carrier (AAC1/ANT) is known to interact with CL in the inner mitochondrial membrane. Protein-lipid interactions are important for the stability and function of biological membranes, and in this particular example, the energetics of the interaction can be characterized by PMF calculation. CL is a complex lipid with four acyl tails and thus should pose a sampling challenge relative to "regular" 2-tailed lipids where the conformational space for the tails to explore is smaller.

Replica exchange umbrella sampling was used to estimate the binding energy. The collective variable was chosen as a 1D path orthogonal to the binding site, using the distance (d) between the CL glycerol bead (GL0) and the Proline backbone bead in the Px[D/E]xx[K/R] of the ANT binding site . Simulations were started from different initial configurations to ensure that the PMF is insensitive to their choice (see the Methods chapter, for how the REUS starting configurations were generated). A "bound" state was defined with CL at site 1 of ANT ($d = 1.7$ nm); an "unbound" was defined at ($d = 4.5$ nm). The bound configuration was generated by running a $1 - 10\,\mu s$, equilibrium MD trajectory, during which CL diffused into and binds to site

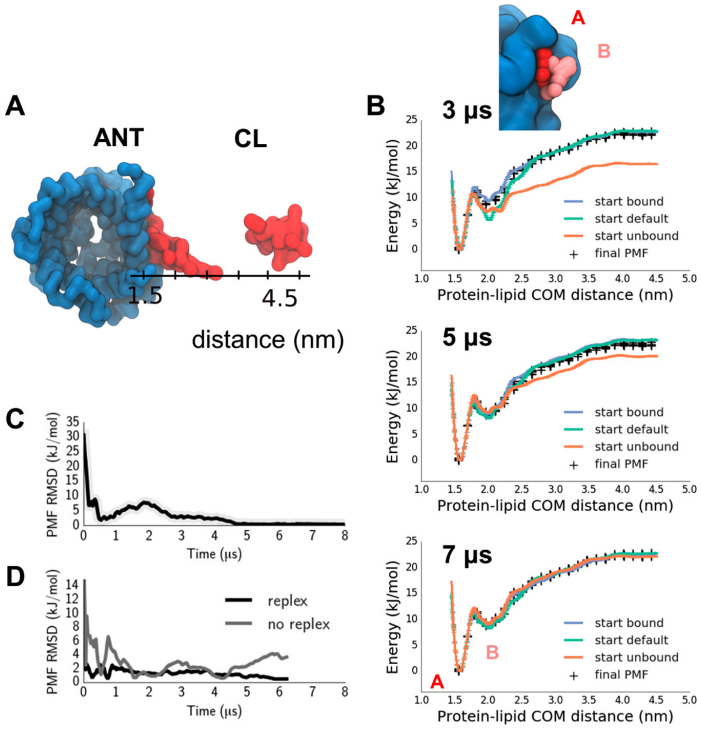


Figure 3.3: (A) REUS simulation setup and choice of collective variable, showing an approximate 1D axis along which the CL lipid (red) is moved relative to the ANT protein (in blue). The simulation was run with a single CL molecule, here two CL molecules are shown to highlight the difference between bound and unbound states. (B) PMF overlap at different time points of the simulations, calculated after truncating all REUS windows at 3, 5 and 7 μs of simulation time for the 3 independent initial conditions. Inset above the PMF contains the representative structures showing the position of CL headgroup relative to ANT binding site for the two minima labelled A and B. (C) Root mean square difference between PMFs obtained from bound and unbound initial configuration as a function of simulation length. (D) Effect of preventing replica exchanges ("replex") as a function of simulation length, show by comparing two simulations started with the default initial condition, with ANT and CL linearly spaced along the d CV. Reproduced with permission from [41], CC-BY 4.0 license.

1; a most representative structure from the bound cluster was then chosen, visually similar to the conformation seen crystallographically [73]. Biased molecular dynamics was used to generate the unbound initial structure, using a constant force on the bound CL to push it out of the binding site on ANT. An additional initial condition was also used, with the CL lipid positioned linearly further and further away from the binding site (see Fig. 3.3A for details).

Two minima were recovered by unbiasing the umbrella sampling simulations using WHAM to recover the PMF [186, 85]. The first minimum (labelled "A") at 1.5 nm had a well-depth of 20 kJ/mol, second minimum (labelled "B") at 2.0 had a-well depth of 10 kJ/mol (see Fig. 3.3B). PMF convergence was monitored as a function of simulation time by taking a difference of the PMFs started from bound and unbound initial conditions. After 5 μs, the two PMFs are within less than 1 kJ/mol of each other (see Fig. 3.3C).

Convergence is helped by allowing for exchanges between neighbouring umbrella sampling windows: the windows are allowed to exchange every 10000 simulation steps. When an independent simulation was performed, with exchanges disabled, after 7 μs the PMFs from bound and unbound initial conditions still fluctuate considerably (see Fig. 3.3D).

An essential but not sufficient condition for PMF convergence is an overlap of CV histograms. For ANT-CL binding along d, the overlap is nearly perfect and appears independent of initial configuration (see Fig. 3.4).

A good CV should be diffusive over the free energy surface. By reversing the effect of exchanges between neighbouring replicas ("demuxing") we can plot the time series of the CV d continously through the replica space (see Fig. 3.5). Indeed it appears that the system is diffusing well along the CV, while the replicas are getting occasionally trapped in the bound or unbound state. This is likely because the d describes the unbinding reaction well but re-binding requires some structural re-arrangement

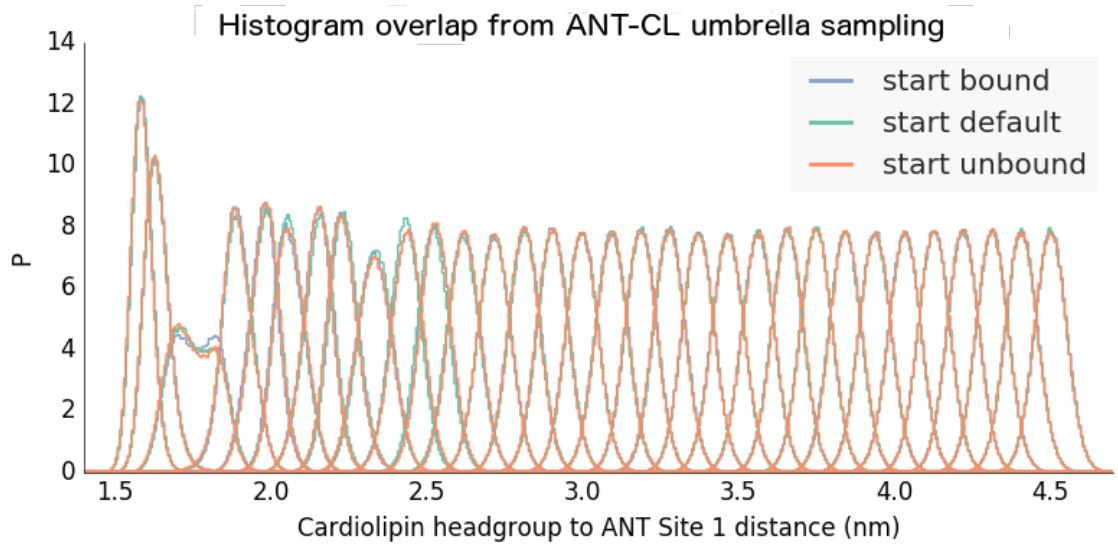


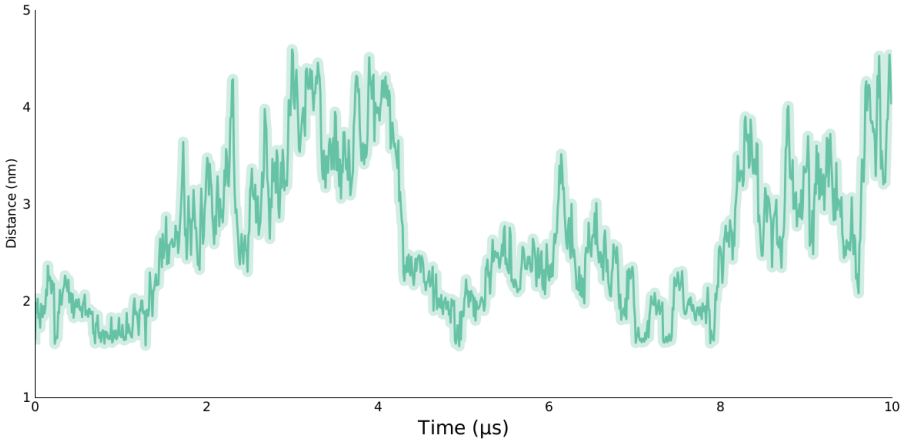
Figure 3.4: Overlap of CV d histograms from REUS simulations, plotted from 3 independent simulations started with different initial conditions. The ANT-CL complex was positioned either bound or unbound in all windows, or uniformly spaced along the d CV. Reproduced with permission from [41], CC-BY 4.0 license.

of the lipid, not just proximity to the protein binding site by centre-of-mass distance, d coordinate.

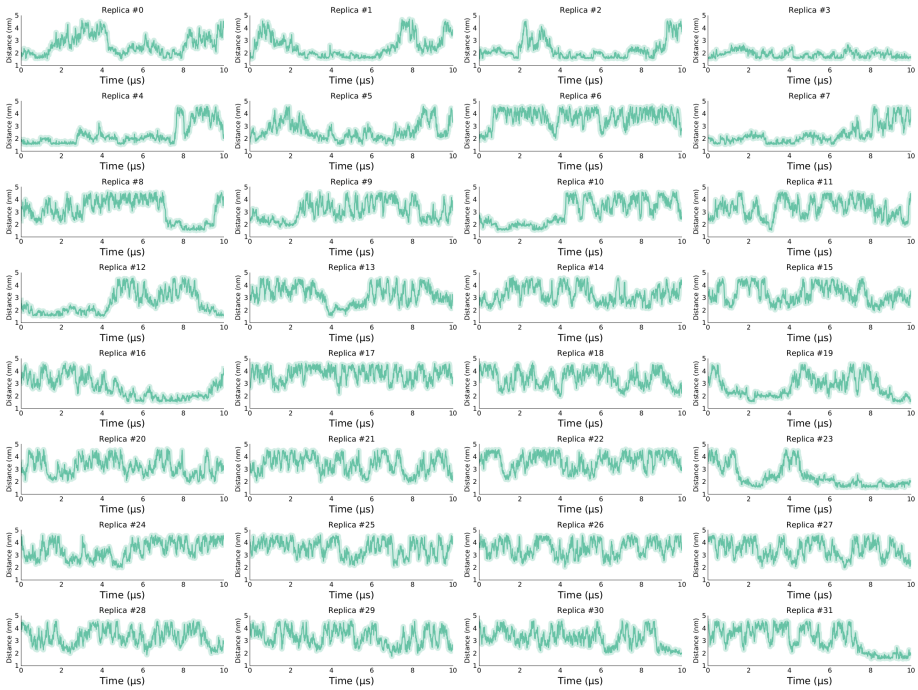
Being confident in the convergence of our simulations, we wanted to understand the structural interpretation of the PMF obtained. We selected representative frames from the two minima: it seems that the lipid head-group is displaced by 0.5 nm from the native minimum A to minimum B (see Fig. 3.3B, inset). The minima are separated by a steep 10 kJ/mol barrier that would be challenging to access using unbiased MD simulations.

3.3.3 Conclusion and future directions

I have presented converged binding PMFs of CL to a site of the ANT transporter. These are PMFs along a 1D reaction path, obtained using REUS, which should be comparable to binding estimates from free energy methods, such as free energy per-



(a) An example replica – replica number 1.



(b) Continuous (with respect to ensemble) CV timeseries for all replicas obtained by reverting the effect of exchanges.

Figure 3.5: Diffusion of the collective variable d between different umbrella windows as a function of time for continuous trajectories throught the umbrella space. Zoom-in of replica number 1, showing ANT-CL binding and unbinding events (upper panel). Time series for all other 32 replicas in REUS simulation, starting from the default initial condition (lower panel). Reproduced with permission from [41], CC-BY 4.0 license.

turbation (FEP), or thermodynamic integration (TI). In the free energy methods, the distance-based collective variable is absent but the ligand is sampled at a number of λ values. A more appropriate collective variable would also allow evaluation of possible non-native binding sites on the surface of ANT. To achieve convergence ,using a simple center-of-mass collectiva variable, the simulations took longer than typically-used 1 μs [72] in calculating protein-lipid binding PMFs. Here there is a difficulty defining a simple, 1D collective variable that can drive binding/unbinding along all possible sites on the protein surface exposed to the lipid environment. The entropy correction used can only account for the lost conformational freedom to diffuse in the plane of the membrane, it assumes no non-native binding sites exist. However, assuming that the native sites on ANT are dominant in affinity (over any other non-native sites) the estimation done here can be used to compare different binding sites on mitochondrial transporter family proteins.

The calculation was performed with wild-type ANT construct, site 1. The binding of CL to ANT mutants or other members of this family, such as UCP2 and Aac2p, could also be of interest in the future. Conversely, lipids related to CL could be investigated: is it the four acyl chains of the cardiolipin that make the interaction possible, or are there any 2-tailed lipids that interact with ANT?

Finally, it is thought that CL mediates ANT interactions with other proteins. This effect was addressed by Hedger et al. in [73]; however, a more detailed free energy calculation of 2 ANT proteins interacting with and without CL in the bilayer could shed some light on how the lipid mediates protein-protein interactions.

3.4 Kir2.2–PIP_2 interaction

3.4.1 Material and methods

Simulation parameters used in this chapter are shown in Table 3.1. Coarse-grained MARTINI 2.2 model [92] was for Kir2.2/PIP_2 simulations with a 5-bead tail model of POPC [206]. PIP_2 parameters were taken from [190].

The protein structure was prepared starting from the PDB 3SPI entry [69], a model for the wild-type Kir2.2. Each of the monomers in the tetrameter was truncated within the Asp-61 and Thr-191 region. This preserves the Arg-186 in the binding site that coordinates the 5' phosphate in the crystal structure. Two forms of the protein were simulated: the wild-type channel and a experimentally-characterized mutant which reduced affinity for PIP_2. In the mutant simulations, Arg-186 was mutated into Ala.

The protein was then embedded in a POPC bilayer and equilibrated using the MemProtMD pipeline [189]. To prevent the protein from rotating in the bilayer about the Z-axis, an absolute position restraint of 100 kJ/mol/nm was applied on the backbone beads in X- and Y-dimensions. Short-tail PIP_2 molecules found in the crystal structure were used as models for the initial configuration of the full-length PIP_2 lipids. To mimic experimentally used buffer conditions, randomly selected solvent molecules were replaced with sodium and chloride ions, up to a concentration of 0.15 M. The final system contained 12000 coarse-grained particles: 4 PIP_2 molecules, approximately 360 POPC lipids and around 6000 waters.

To sample the protein-lipid binding, a collective variable was defined between the PIP_2 headgroup particles and the Kir2.2 monomer centre of mass. The biasing was performed on the X-component of this distance, which resulted in a path roughly orthogonal to binding surface. To prevent lipid diffusion in the Y-direction, a harmonic restraint was used with a 100 kJ/mol spring constant. To prevent protein

rotation within the bilayer, backbone beads were position restrained in X and Y with a 100 kJ/mol/nm^2 spring constant.

Umbrella sampling was done with 16 windows, linearly spaced between 2.6 and 4.6 nm, with a 400 kJ/mol/nm^2 spring constant. Exchanges were attempted every 10000 steps. The initial structure for the bound state was taken from the PDB 3SPI structure, the unbound structure was taken from a from the umbrella window furthes from the protein, after 50 ns of equilibration (the conformation was visually confirmed to be completely unbound).

3.4.2 Results and discussion

REUS simulations were performed, biasing along a distance CV between binding site centre of mass and the lipid headgroup centre of mass (see Fig. 3.6A). The PMF in the absence of umbrella bias was recovered using WHAM, and for wild-type, Kir2.2 shows as a single minimum (labelled as A) approximately 45 kJ/mol deep (see Fig. 3.6B). For PMFs obtained from bound and unbound initial conditions, the profiles overlapped to within 2.5-5 kJ/mol (1-2 k_BT) after 5 μs of REUS simulation. The Kir2.2 binding site for PIP_2 contains 6 positively charged aminoacids; the well-depth obtained is similar to 42 kJ/mol calculated for PIP_2 binding to EGFR juxtamembrane region, which contains 5 positively charged residues and is more disordered [72].

The R182A Kir2.2 mutant has a reduced well-depth of 32 kJ/mol, is rather broad and corresponds to an ablation of the tight binding interaction at $d = 2.8nm$ (see Fig. 3.6C). This binding reduction is in-line with available experimental data: the binding site mutant reduces the number of arginine residues in the binding site from 6 to 5, yet does not completely abrogate binding; indeed it can be crystallized with short-chain PIP_2 [69]. Selected structures of minima A and B show that, while in the mutant the minimum A is still present, the loss of a single coordinating arginine has a dramatic impact on minimum A stability – in the mutant, both minima A and B

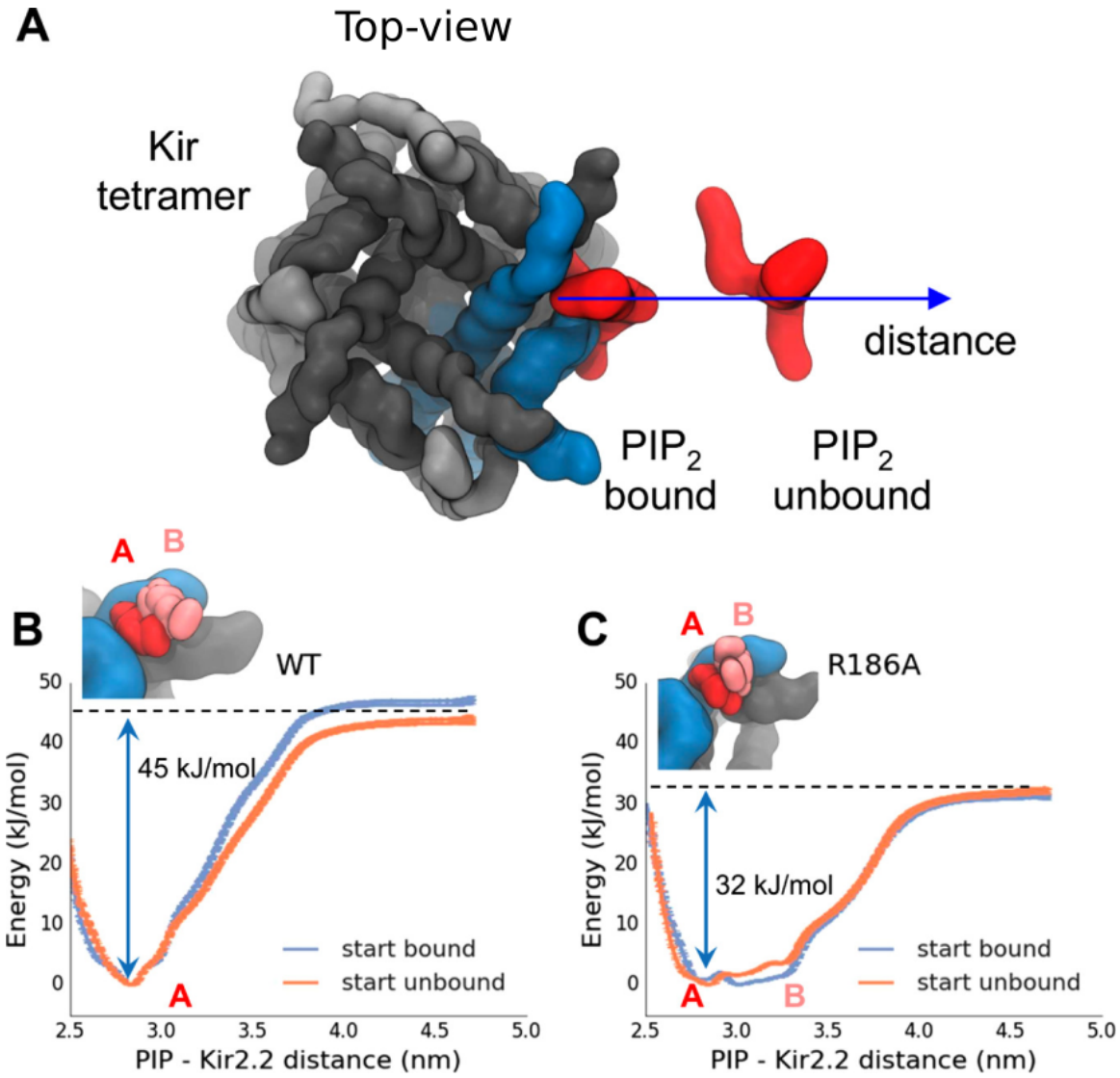


Figure 3.6: Setup and collective variable design of the Kir2.2 PIP_2 binding simulation. (A) Top-view of Kir2.2 complex with one subunit (in blue) highlighted with PIP_2 (in red) either bound or unbound. Blue axis indicates the approximate position of 1D CV along which the bias is applied. (B) PMF obtained for wild-type Kir2.2 with a single minimum at 2.8 nm labelled as A. (C) PMF obtained for mutant R186A Kir2.2 showing 2 minima labelled A and B, corresponding to tightly bound and nearly-bound states, respectively. Presence of state B is undetectable in the WT Kir simulations (left panel). Representative structures of Kir and PIP_2 headgroups for minima A and B (in red and pink, respectively) are shown in insets above (B) and (C). Reproduced with permission from [41], CC-BY 4.0 license.

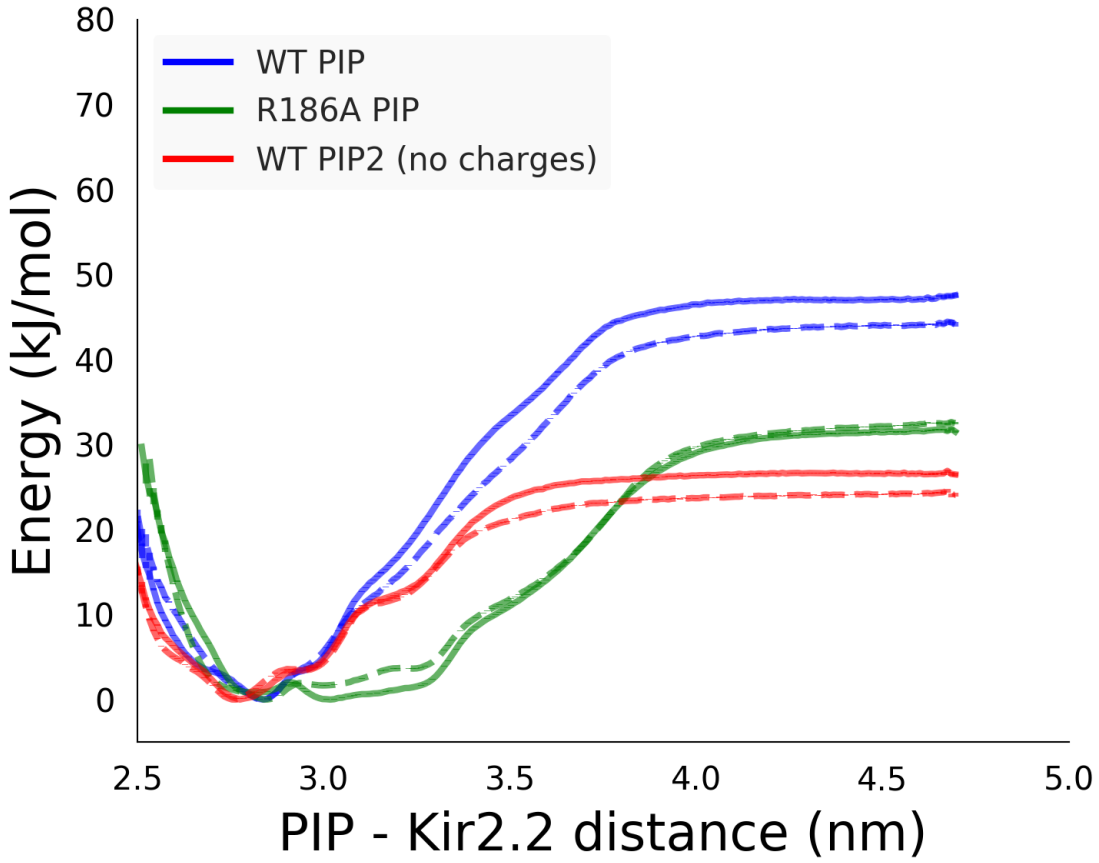


Figure 3.7: Overlap of PMFs obtained for wild-type and mutant Kir2.2 binding to PIP_2 (blue and green curves). Together with a PMF for wt Kir2.2 binding to PIP_2 model with all headgroup charges removed. Solid lines indicate REUS simulations started with all replicas in "bound" configuration, dashed lines indicated an "unbound" configuration. Reproduced with permission from [41], CC-BY 4.0 license.

are approximately equal in energy.

To further understand the charge contribution to the PMF, I set up a simulation with a PIP_2 model with all phosphatidylinositol charges removed: this corresponds to fully-protonated PIP_2 headgroup (see Fig. 3.7). Indeed at physiological conditions, of pH 5-7, the PI headgroup could adopt a number of protonation states [98]. The pK_as of the protonatable hydroxyl groups could be further affected by local chemical environment, for example with PIP_2 bound to a protein. The shape of the PMF obtained for uncharged PIP_2 is similar to that of WT; however, the well-depth is

considerably reduced to about -20 kJ/mol and in range with the mutant Kir2.2 value. The simulation results suggest a 1/3 reduction in binding, as judged from the reduced PMF depth. This is in line with available experimental data for the Kir2.2 and Kir.1 channels [115]: the PIP_2 remains bound to, and can be crystallized with, the mutant channel [69].

3.4.3 Conclusions and future work

Here I show a converged PMF for the association of the polyanionic PIP_2 lipid with Kir2.2 wild-type and mutant channels. I observe a reduced but not completely abrogated binding to the mutant site, in agreement with experiments where partial binding is seen.

The binding site in Kir2.2 is relatively well exposed, making a definition of a 1D binding path straightforward. Future work could focus on the more enclosed binding sites, where finding a simple 1D-path CV into the solvent, out of the binding pocket, may not be possible.

A biologically-interesting angle to explore is the role of lipid specificity in the binding. Kir-type channels are generally selective for specific types of phosphoinositides, most Kir family members preferring $PI(4,5)P_2$ over $PI(3,4,5)P_3$, and $PI(3,4)P_2$; exceptionally Kir6.2 can be activated by a metabolic intermediate, long-chain acyl-CoA, presumably via a nonspecific phosphoinositide binding [167]. Given the simplicity of the charge model used in MARTINI, it would be interesting to see if:

- the experimentally observed specificity can be reproduced,
- what is the accuracy of all-atom force fields (with more sophisticated charge models) for capturing this behaviour.

Another angle that might be interesting is an all-atom protein-lipid binding PMFs. A comparison with coarse-grained simulations could be illuminating, especially given

the high computational cost of running all-atom simulations. To date, to my knowledge, no reliably converged all-atom protein-lipid PMF is available. It is likely that that recent improvement to the enhanced sampling methods (metadynamics or replica exchange umbrella sampling), together with collective variable development, could make this calculation possible.

Chapter 4

Coarse-grained helix-helix dimerization

In this chapter I will consider protein-protein interactions in the bilayer, at a coarse-grained resolution level. As the simplest system, with abundant experimental data, I will study a a transmembrane (TM) helix-helix dimer, the Glycophorin A fragment [122]. I will also investigate a series of TM helix dimers from the Receptor Tyrosine Kinase family, spanning a range of different experimental affinities. The purpose of the work described in this chapter is to understand the association energetics and check if the available force fields are adequate to model simple helix-helix association in membranes.

4.1 Introduction

4.1.1 Glycophorin A

The interaction between two TM helices is the minimal unit allowing the study of membrane protein assembly. Glycophorin A fragment is a prototypical helix-helix transmembrane (TM) dimer that has become the workhorse model system in this

field. Anecdotally, it is abundant in human red blood cells from which it has been historically obtained. With a wealth of experimental data available on both the structural properties and the association energetics, Glycophorin A has been a key system to benchmark simulations against [122, 108, 60, 61, 44, 3, 81, 172, 194].

Structural models of the GpA dimer have been experimentally obtained by X-ray crystallography on samples derived from lipid cubic phase (LCP) [194], as well as by solid-state [185] and solution [122] NMR (see Fig. 4.1). The experimental structures are remarkably consistent, despite the range of methods and data collection conditions: the dimer is symmetric, left-handed, crossing at an angle of about -20 to -40 degrees, packed via a conserved GXXXG motif [107]. Mutants exist that place a bulky-sidechain residue in the dimerization motif, disrupting the interface. These mutants have been studied experimentally, via biochemical methods [81, 60, 61] and *in silico*, via coarse-grained and all-atom simulations[179, 75, 195, 153, 32, 161, 90, 103].

Numerous biophysical and biochemical characterizations of the GpA dimer have been performed, notably including energetic measurements for dimerization free energy. Those measurements were performed in detergent micelles using analytically centrifugation [61] and in POPC bilayers using a FRET-based assay [81, 80, 172]. While both methods point to a stable dimer, the dimerization free energy estimates range from -16 to -51 kJ/mol [81] (using a reference concentration of 1 molecule/nm^2), depending on the experimental conditions.

Simulation studies have used GpA as a model for assessing the quality of force fields, sampling methods and resolution levels. Different levels of resolution were used, starting with the simplest implicit membrane models. Sugita and Im used replica exchange to characterize the GpA dimerization landscape, within an implicit membrane model, correctly identifying the native bound state [110]. Similarly, a CHARMM-derived implicit membrane model 1 (IMM1) was used by Lazaridis to study GpA

dimerization [216]. A step-up in resolution, explicit coarse-grained force fields were used to study membrane helix dimerization. Janosi used a Monte-Carlo simulation with MARTINI force field to study GpA dimerization [90]; unbiased MD simulations were used by Psachoulia revealing which contacts mediate the dimer binding in Glycophorin A [161]. Of particular interest are studies where the dimerization energetics were probed directly. Sengupta calculated a coarse-grained helix-helix dimerization PMF using centre-of-mass distance as a collective variable [179]. More recently, a novel, collective-variable free approach to studying helix-helix binding in the membrane was proposed by Hummer and co-workers using Mga2 TM dimer as test system at all-atom resolution[29, 24].

It is unlikely that study of more complex, multi-span transmembrane proteins is possible if any force field or sampling limitations exist for the "simple" case of Glycophorin A.

4.1.2 Receptor Tyrosine kinases

Another important group of membrane proteins is receptor tyrosine kinases (RTKs) family. RTKs are cell-surface receptors involved in signal transduction through the membrane. RTKs frequently bind peptide ligands such as cytokines and hormones. Mutations to this class of receptors can have a profound effect on cell-signalling, affecting gene expression. RTKs are typically found in developmental pathways and their mutants are found in many cancers. RTKs have a large extracellular region, which is typically involved in ligand binding. The intracellular domain contains the juxtamembrane region and receptor tyrosine kinase domain, which is involved in signal propagation. Connecting the domains is a transmembrane (TM) region, which is thought to part-take in RTKs dimerization and signal transduction. Upon receptor activation, for example by ligand binding to the extracellular domain, a complex set of changes occurs involving the TM and intramembrane domains.

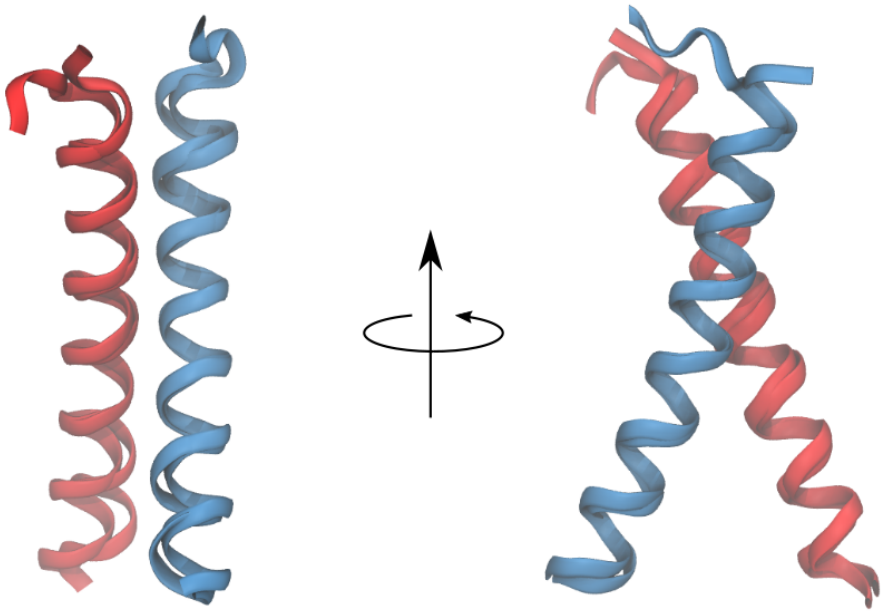


Figure 4.1: Experimental structures of Glycophorin A, obtained in detergent micelle (PDB code 1AFO) and derived from LCP X-ray crystallography (PDB code 5EH4) are essentially indistinguishable. Chain A in blue, chain B in red. Showing only residue 69–97, as used in simulations. The alignment was done on residues 70–97, as the 5EH4 structure does not have residue number 69 resolved.

The RTKs helix TM dimers have been characterized by structural biology methods. NMR structures of related EphA1 and EphA2 have been determined in detergent and found to form a right-handed helix dimer [17, 16]. Numerous human RTKs contain a heptad repeat motif ($LX_3GX_2AX_3VX_2L$, for example in the left-handed EphA2 dimer) mediating the interaction [34]; however, its conservation is not strictly enforced [55]. The apparent loose conservation highlights the possibility of multiple regions mediating the interaction and multiple binding arrangements [109, 16]. It is possible that different interaction modes correspond to inactive and active bound forms of the receptor, as suggested by mutagenesis work on EphA [180].

A series of studies have underscored the role of the TM region in dimerization and activation of various RTKs, including FGFR [203, 171] and EGFR [4, 46]. A mutational study of EphA2 (a member of the RTK family) suggested a possible activation mechanism involving a population switch between two different bound TM configurations. It appears that RTK TM region dimerization is of key importance to activation [105, 127, 109], therefore understating the energetics of TM helix-helix is relevant to understanding the activation mechanism [70].

The interaction strength between TM regions of selected members of this family has been quantified experimentally. In this section, I focus on experimental work by Finger at el. [55] where a qualitative TOXCAT assay is used to estimate dimerization propensities for 50 TM helix homo-dimer constructs. A TOXCAT bioassay provided an affinity ranking, using Glycophorin A and its dimerization-disrupting mutant to calibrate the fully-bound and fully-unbound states, respectively: wild-type GpA is a tightly-associated dimer giving a dimerization propensity of 1, the dimerization-disrupting G83I mutant has a 0 dimerization propensity. In this chapter, I have selected four representative constructs (roughly equally spaced along the TOXCAT-scale) at 1, 0.55, 0.25 and 0.13 dimerization propensity – see Table 4.1 for selected protein names, their PDB codes and TM construct sequences. Then I calculated the

Table 4.1: Selected entries of the TOXCAT assay results by Finger et al. used in this study, together with wild-type and mutant Glycophorin A interaction propensities used to calibrate the scale of the assay. Table taken and modified from [55, Table 1]

Protein	PDB	Sequence	Interaction propensity
GpA wild-type	1AFO	RASLIIFGVMAGVIGTILIN	1 ± 0.08
GpA G83I	N/A	RASLIIFGVMAIVIGTILIN	0.09 ± 0.00
Her2/ErbB2	2JWA	RASSIISAVVGILLVVVLGVVFGILIGIL	1.08 ± 0.18
Her1/ErbB1	2M0B	EGFRRASIATGMVGALLLLLVVALGIGLFMGIL	0.55 ± 0.07
EphA2	2K9Y	RASLAVIGGVAVGVVLLLVLAGVGFFIGIL	0.25 ± 0.05
EphA1	2K1L	RASIVAVIFGLLLGAALLLGILVFGIL	0.13 ± 0.01

association PMFs for these helix-helix pairs. The reason for the selection was to benchmark the MARTINI 2.1 force field and see if it can reproduce the rank order of dimerization propensities from TOXCAT. Devoid of the GXXXG motif found in Glycophorin A (see next page), and exhibiting different helix handedness, the RTKs provide an interesting test case.

4.2 Glycophorin A

4.2.1 Material and methods

The protein structure used was the detergent-micelle model, PDB code 1AFO [122]. The model was further truncated to include only residues between Ser-69 and Arg-97. The protein was inserted into a POPC bilayer using the *g_membed* tool [213]. The final system had approximately 125 POPC and 1400 water molecules.

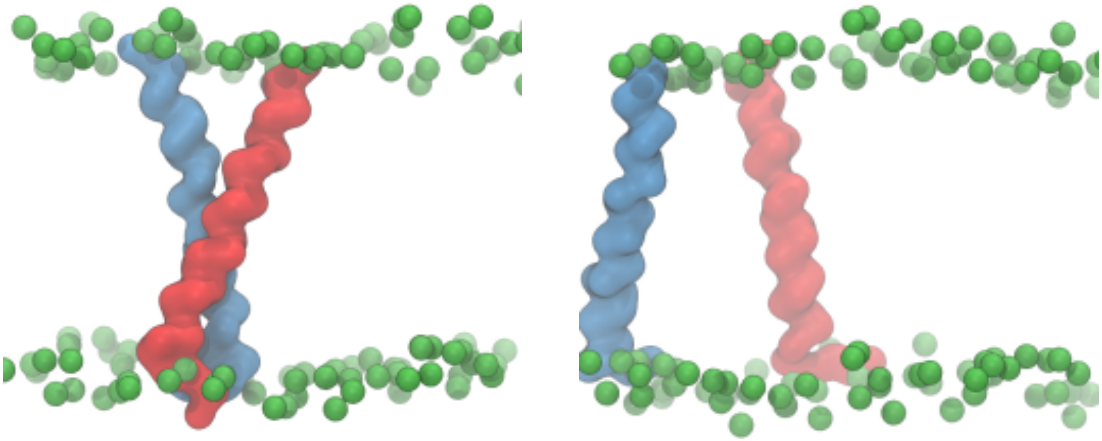
To bias the helix-helix dimerization and estimate the free energy surface for the process, an inter-helical D_{RMS} collective variable was used. Only the helical region of the protein, between residues Glu-72 and Ile-95, with only the non-symmetric protein sidechain beads used to calculate the CV. For details of the D_{RMS} definition and K_d calculation, please refer to the main Methods section. Umbrella sampling simulation was setup with 16 windows, linearly spaced along the D_{RMS} coordinate between 0 and 2.5 nm, with a spring constant of 100 kJ/mol. Exchanges between windows were attempted every 1000 steps.

4.2.2 Results

Here I attempt to compute converged GpA dimerization PMFs, in order to understand the free energy surface and asses the accuracy of MARTINI coarse-grained force field. I have used replica exchange umbrella sampling (REUS) to address some of the challenges sampling free energy surfaces using vanilla, unbiased MD simulations (as discussed in the Methods Chapter).

An appropriate collective variable has to be defined to use with umbrella sampling simulation that describes the helix-helix dimerization reaction. The collective variable has to represent the dissociated/separated state but also discriminate between many possible packing arrangements of the helix-helix interface at short separation.

Here I devised a novel collective variable for biasing protein-protein interactions.



(a) Bound/together (b) Unbound/separate

Figure 4.2: Initial configurations for the coarse-grained Glycophorin A simulations with MARTINI 2.1. The bound structure was initialized with the NMR structure with PDB code 1AFO (showing only residues 69–97 used in simulation). The unbound structure was obtained from the last windows of umbrella sampling simulation, with the helix-helix Glycophorin A dimer full dissociated. Taken with permission from [41].

A set of helix-helix distances was taken from the native structure. Using this vector of native distances as a reference, we then compute the root mean square difference between the reference distances and distances at any simulation step t. Three sets of simulations were started from different initial conditions, to assess convergence. In the 'bound' simulations, all windows were initialized from the native/bound configuration, as observed in the detergent-micelle GpA structure (PDB code 1AFO [122]); in the 'unbound' simulations, all windows were initialized from a fully-dissociated configuration (see Fig. 4.2).

The dimerization PMF estimated from umbrella sampling simulations (see Fig. 4.3) appears to be independent of the initial configuration used. The PMF flattens out at around 2 nm D_{RMS}, indicating that the dimer becomes unbound around that point. The bound well is approximately 35 kJ/mol deeper than the unbound region of the PMF. The free energy surface has four local minima in the bound region, all similar in stability. Importantly, the native minimum (labelled 1 in Fig. 4.3A) appears to be 5 kJ/mol less stable than the near-native minima (labelled 2–4 in Fig. 4.3A). Minima

1 and 2 are native and near-native states, with crossing angles ($-25°$ and $-20°$, respectively) resembling experimental structures ($-23°$ in 1AFO NMR structure and $-20°$ in the 5EH4 crystal structure). Minima 3 and 4 are non-native bound states, with the dimer being approximately parallel, and characterized by crossing angles of $-15°$ and $-5°$, respectively (see Fig 4.3D). Projecting the PMF along the helix-helix COM distance, a single minimum at a non-native separation distance of 0.9 nm is obtained (see Fig. 4.3B).

Previous computational studies have found the native state of GpA to be most stable with MARTINI force field, using umbrella sampling MD [179] and Monte-Carlo simulations [90]. However, these simulations were performed using the helix-helix COM distance (D) as the collective variable. At large helix-helix separations (greater than 1.5 nm), D and D_{RMS} are largely equivalent (see Fig. 4.4). However, at short separations, D is unable to distinguish multiple minima captured by D_{RMS}; in other words, D collapses the state space and is thus a poor collective variable. This is clearly seen in 2D projections of the PMF along D_{RMS} and D (see Fig. 4.4). See Methods 2.2.3 for a discussion of "good" and "bad" collective variables.

One unexpected finding from the simulations is that there is a considerable proportion of non-native dimer present at equilibrium. It is unclear if that effect is real or an artefact of coarse-grained resolution of the model.
Available experimental structures point to a well-defined structure with a negative crossing angle [194], it is possible (however unlikely) that other non-native but bound states exist at equilibrium.

To resolve this issue, I will compare the equilibrium dissociation constant K_d derived from simulations to the values obtained experimentally via the steric trap method. To compare the computed PMF to experiment, the PMF needs to be integrated suitably. Experimentally, the readout captures the fraction of dimer bound. An unbound portion of the PMF is typically chosen as a cutoff value, to defined the

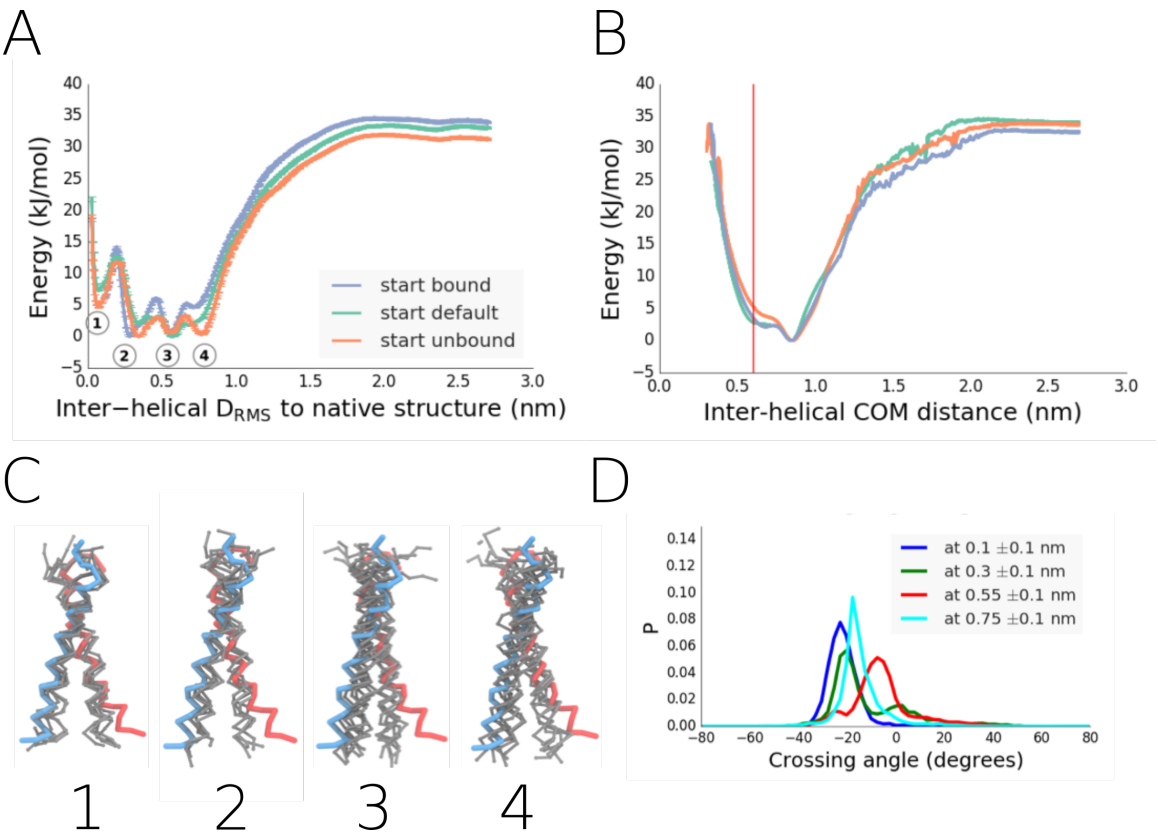


Figure 4.3: Association of Glycophorin A TM fragment helix-helix dimer within a POPC bilayer. (A) Dimerization PMF calculated along inter-helical D_{RMS} as the collective variable (CV)/reaction coordinate (RC) using umbrella sampling (US). Inter-helical D_{RMS} was defined relative to the native dimer structure (PDB code 1AFO).Umbrella sampling simulations were initialized with either: all windows in the native configuration ('bound', blue), all windows in the unbound configuration ('unbound', orange), all windows linearly spread along the CV ('default', cyan). Local minimal are labelled 1–4, with D_{RMS} values of about 0.1, 0.3, 0.55 and 0.75, respectively. (B) PMF obtained by projecting the PMF in (A) along a helix-helix center of mass distance collective variable. Red line at 0.6 nm corresponds to native GpA helix-helix COM distance. (C) Representative structures (shown as wireframe backbone, in grey) from each of the local minima from PMF in (A), labelled 1-4. Experimental 1AFO shown as wire-frame backbone, with helix A in blue and helix B in red. (D) Distributions of helix-helix crossing angle for local minima 1–4 from PMF in (A), crossing angle is defined between vectors connecting C-alpha atoms of residues 78 and 88 in each helix. Taken with permission from [41].

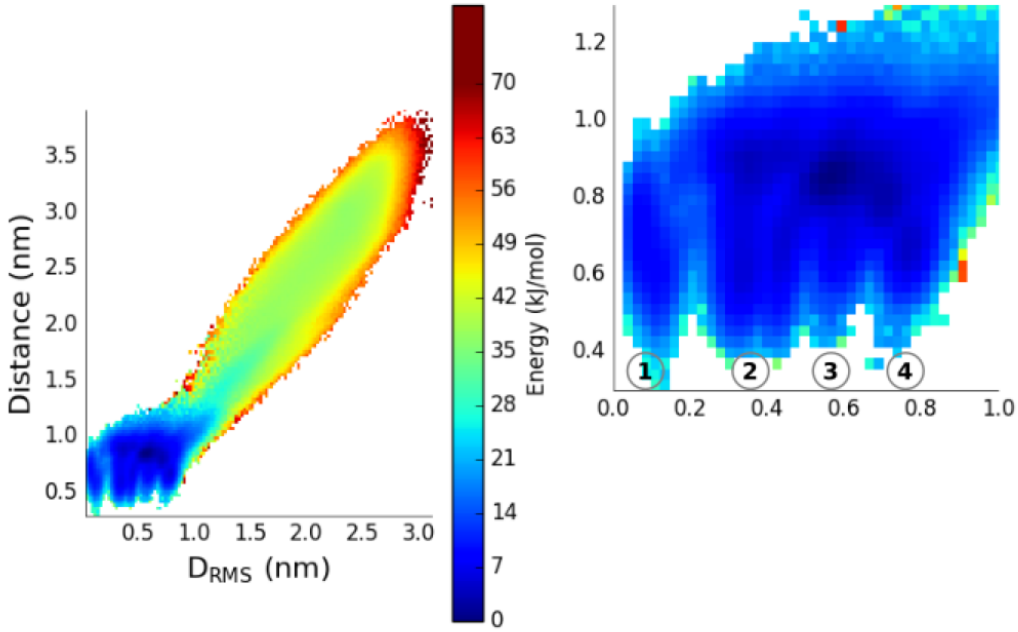


Figure 4.4: Projections of the wild-type Glycophorin A dimerization PMF onto a 2D free energy surface, along the helix-helix centre of mass distance (Y-axis) and inter-helical D_{RMS} to native (X-axis). The inset shows a close-up of the bound region of the PMF, showing that a single helix-helix distance value corresponds to multiple basins along D_{RMS}. Taken with permission from [41].

unbound state, beyond which increasing the dimer separation has no effect on the free energy. Defining the bound state as $< 2\,nm$ D_{RMS} (corresponding to helix-helix D of 2.3 nm), I can integrate the population distribution obtained by inverting the PMF and calculate the K_d for the wild-type Glycophorin A TM construct. We obtain an estimate of 1.1×10^{-6} molecules / nm^2, compared with experimental estimate of 1.3×10^{-9} molecules / nm^2 (the details of this calculation are described in Sub-section 2.4.1). Given the low resolution of the coarse-grained MARTINI model, this may seem like an acceptable difference, however large.

Experimentally the GpA K_d is known to be sensitive to dimerization-disrupting mutants. G83L and G83I mutants introduce a bulky side-chain residues that break-up the tight packing at the GXXXG interface, preventing dimerization in experimental measurements. Leucine and Isoleucine are nearly identical in MARTINI representation, I have performed dimerization PMFs with the G83L starting from 'bound' and 'unbound' initial conditions. The K_d computed for the GpA mutant from these simulations is essentially the same as wild-type, which is apparent from near-identical PMFs (see Fig. 4.5, upper panel) and population distribution plot (see Fig. 4.5, lower panel). The calculated K_d is a strong indication that the binding modes seen in my simulations are inconsistent with experiment. Encouragingly the native minimum (labelled 1) is destabilized by 5 ± 2.5 kJ/mol in the mutant relative to wild-type (see Fig. 4.5A); experimentally this destabilization is much larger, approximately 20 kJ/mol. There could be many reasons for this discrepancy: it could be inherent to the coarse-grained representation, or simply a consequence of force field parameterization that could be improved.

Additionally, the variation in experimental conditions used to measure the K_d make a quantitative comparison difficult: constructs used experimentally vary, and lipid environment can be either an artificial bilayer (pure POPC), a bacterial bilayer or a detergent micelle. I consider the Bowie FRET meastuments most relevant.

Taken together these results suggest that while the dimerized state is more stable than the unbound state, it is predominantly due to non-native binding in simulations.

A necessary but not sufficient criterion for assessing convergence of PMFs is the overlap of collective variable histograms. The histograms obtained from the different initial conditions have a good degree of overlap for both MARTINI 2.1 and 2.2 force field (see Fig. 4.6A).

Furthermore, we calculate the PMF root mean square difference, for the 'bound' and 'unbound' initial conditions, for simulations run with both MARTINI 2.1 and 2.2 models. For wild-type GpA profile obtained with MARTINI 2.1 the PMF RMSD converges to within 2.5 kJ/mol in 3 μs simulation time (see Fig. 4.6B).

To check simulation convergence, I looked at the structural properties of the dimers obtained from simulations. In particular, I calculated crossing-angle distributions for each of the minima identified in the PMFs, starting from different initial conditions. Encouragingly, the crossing angle distributions appear near-overlapping and independent of starting configuration, for both the wild-type and mutant GpA constructs (see Fig. 4.7, left and right column respectively). Having said that the overlap is not perfect, indicating there may be some degeneracy of D_{RMS} relative to crossing-angle Θ. The crossing angle distribution for the native minimum (labelled '1') for the G93L mutant appears to have a significantly lower crossing angle relative to wild-type ($-40°$ vs. $-20°$), likely due to extreme dimer interface distortion caused by the presence of bulky side-chain mutations.

To further investigate the convergence of our Glycophorin A replica exchange simulations, I checked how diffusive the collective variable was over the replica space. If each copy of the system remains close to the umbrella sampling window it started in, it is indicative of poor diffusion (in replica space) and is not desirable. We have used the demux.pl script provided with GROMACS to recover continuous trajectories, removing the effect of exchanges. Time series of the inter-helical D_{RMS} for each replica

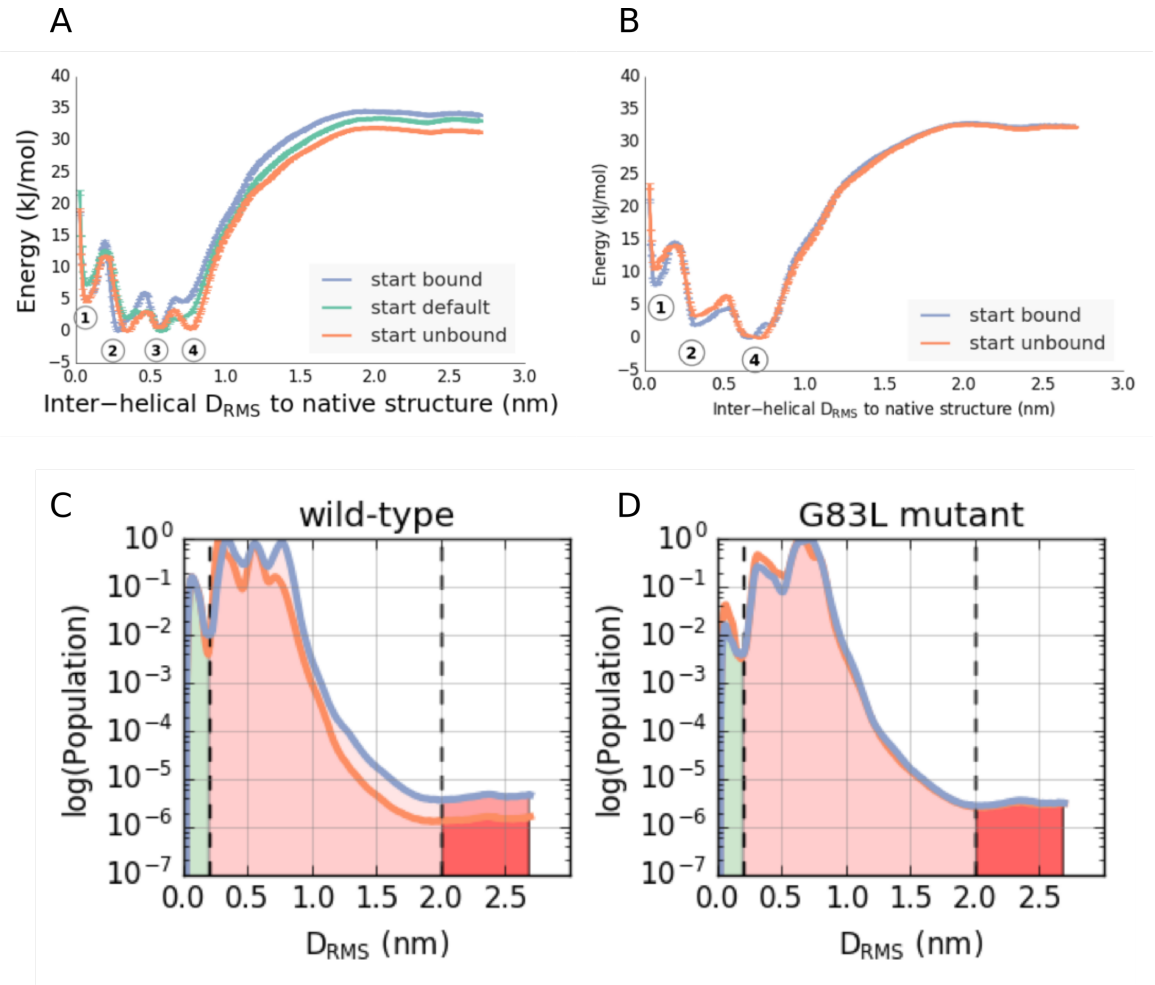


Figure 4.5: Comparison of PMFs for wild-type and dimerization-disrupting mutant (G83L) of Glycophorin A, subplots A and B, respectively. The equivalent minima in the PMFs are labelled 1–4. For the mutant simulations, the umbrella sampling windows were all initialized with either the native configuration ('bound', in blue) or the dissociated configuration ('unbound', in orange). Population distributions obtained by transforming the wild-type and mutant PMFs (C and D, respectively). Defining the bound state as either 0.2 or 2.0 nm D_{RMS}, estimated $\Delta\Delta G$ from G83L becomes 0 and 4 kJ/mol, respectively. Taken with permission from [41].

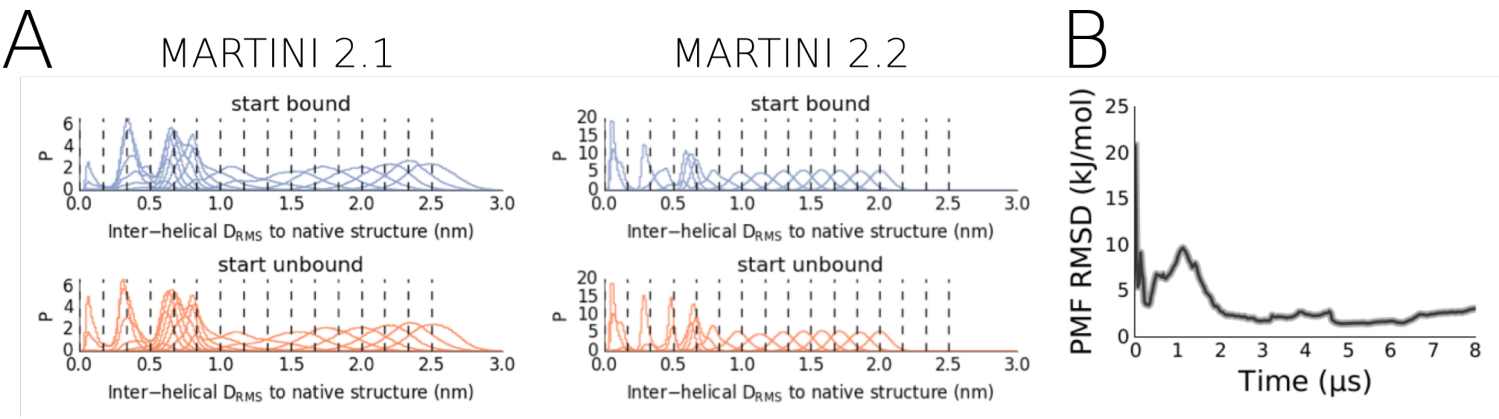


Figure 4.6: Assessing the convergence of the Glycophorin A dimerization PMFs. (A) Histograms of the D_{RMS} distributions for each replica, for MARTINI 2.1 and 2.2 models, for both initial conditions: native, labelled as 'bound' in blue and dissociated, labelled as 'unbound' in orange. (B) The RMSD between the PMFs obtained from the bound and unbound initial conditions, as a function of simulation length. At any time point, only the 2nd half of the simulation was included in the analysis. Taken with permission from [41].

show a number of binding/unbinding transitions and reversible dimerization (see Fig 4.8A and B, for simulations starting from bound and unbound configurations, respectively). The replicas appear diffusive, irrespective of the initial condition. Having said that, individual replicas appear to "get stuck" close to their initial umbrella window (for instance, Replica #3 in Fig. 4.8A), indicating that either inter-helical D_{RMS} is not a perfect collective variable or the dynamics of the system is very slow and even longer simulations would be needed. Future improvements could attempt to quantify the "diffusivity" of such simulations. Here the check was purely visual to confirm that several binding/unbinding transitions occured. More generally to carefully check the quality of the collective variable a committor analysis could be performed, where trajectories are started from the apparent barrier peak – in absence of orthogonal barriers, the trajectories should go to either side of the barrier with a roughly fifty-fifty split. We thus conclude that the simulations are well-converged, with multiple association and dissociation events.

To understand the origin of replicas getting 'stuck' in continuous simulations, we

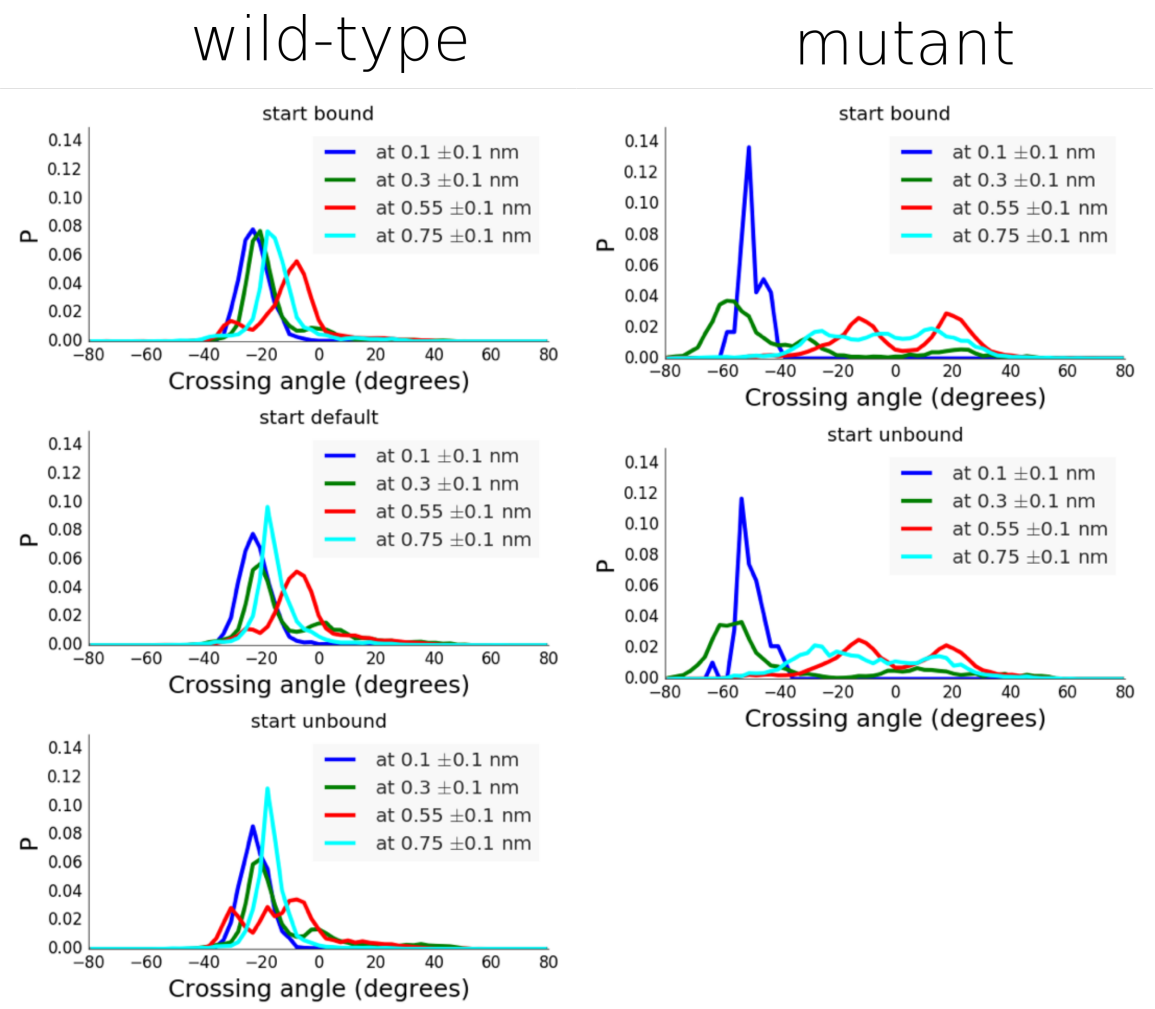
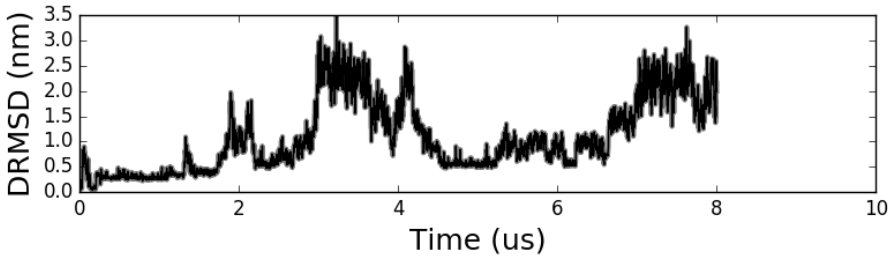
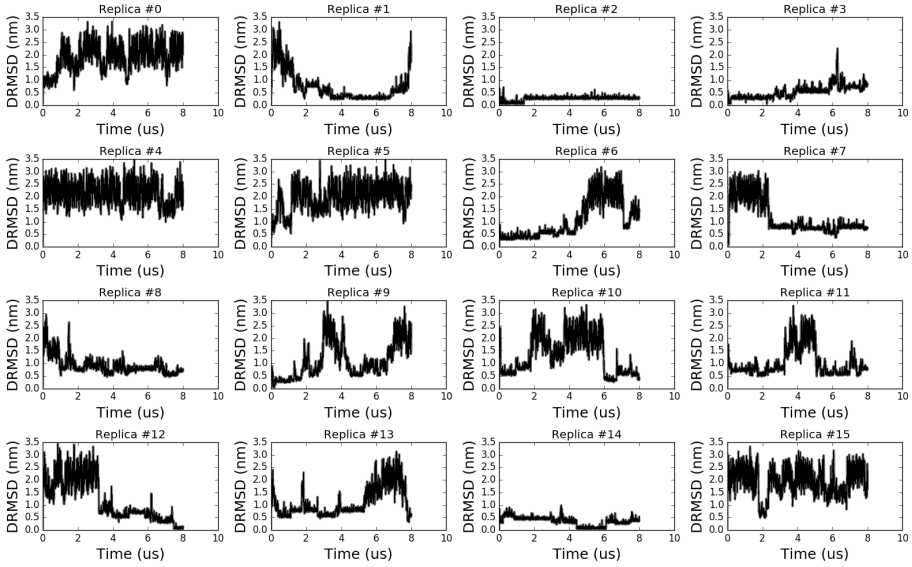


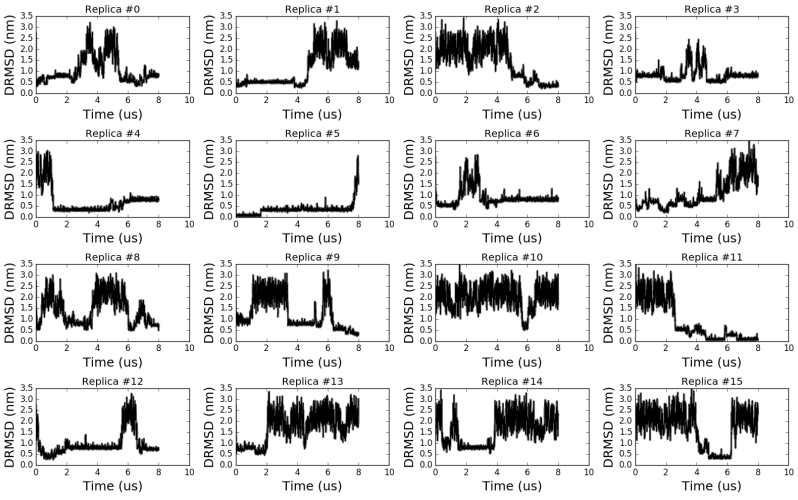
Figure 4.7: Distributions of helix-helix crossing angle for the different minima in Glycophorin A association PMFs, wild-type and mutant constructs on the left- and right-hand panels, respectively. Taken with permission from [41].



(a) Example timeseries taken from replica #9 (started from bound, see (b) below)



(b) Bound



(c) Unbound

Figure 4.8: Continuous trajectories from Glycophorin A (wild-type) simulation with MARTINI 2.1 using inter-helical D_{RMS} as a biasing CV. The replica exchanges were removed using demux script, recovering continuous trajectories through the replica space. (A) Zoom-in on an example of single replica: CV time series from a continuous trajectory (replica 9, start bound) (B-C) All 16 replicas for the bound and unbound initial conditions.

performed 2D PMF projections along different collective variables. For all 3 initial conditions (labelled 'bound', 'unbound' and 'default') a consistent picture emerges (see Fig. 4.9), where the helix-helix center of mass distance is degenerate with respect to the inter-helical D_{RMS}. However, it also seems like the D_{RMS} is degenerate relative to the helix-helix crossing angle Θ. The degeneracy is not entirely unexpected: the distance-based collective variables such as contacts and D_{RMS} are unable to capture handedness. A left-handed and right-handed helix-helix dimer can have similar sets of distances. Collective variables such as RMSD, which are position based, can distinguish handedness but are disadvantageous when driving the dissociation reaction. As a future improvement, one might consider devising a novel, pseudo-1D collective variable that combines D_{RMS} and crossing-angle, such as:

$$X = D_{\mathrm{RMS}} + \lambda \times \Theta \tag{4.1}$$

Where Θ is the helix-helix crossing angle and λ is an appropriate mixing constant that divides the bound and unbound surfaces on a 2D PMF along Θ and D_{RMS}.

4.2.3 Conclusions and future work

Encouragingly the coarse-grained MARTINI model correctly predicts that the global minimum is the bound/dimer state, approximately $-35\,\mathrm{kJ/mol}$ more stable than the dissociated state. However, the most stable bound state is non-native. For GpA TM fragment, the coarse-grained MARTINI model seems unable to distinguish the native and near-native states. Furthermore dimerization disrupting GpA mutant has a minimal effect on the K_d, in contrast to available biophysical data, from either sedimentation equilibration centrifugation or FRET. This points to a force field imbalance, where non-specific protein-protein interactions are too strong, and detailed interactions that stabilize the native state too weak.

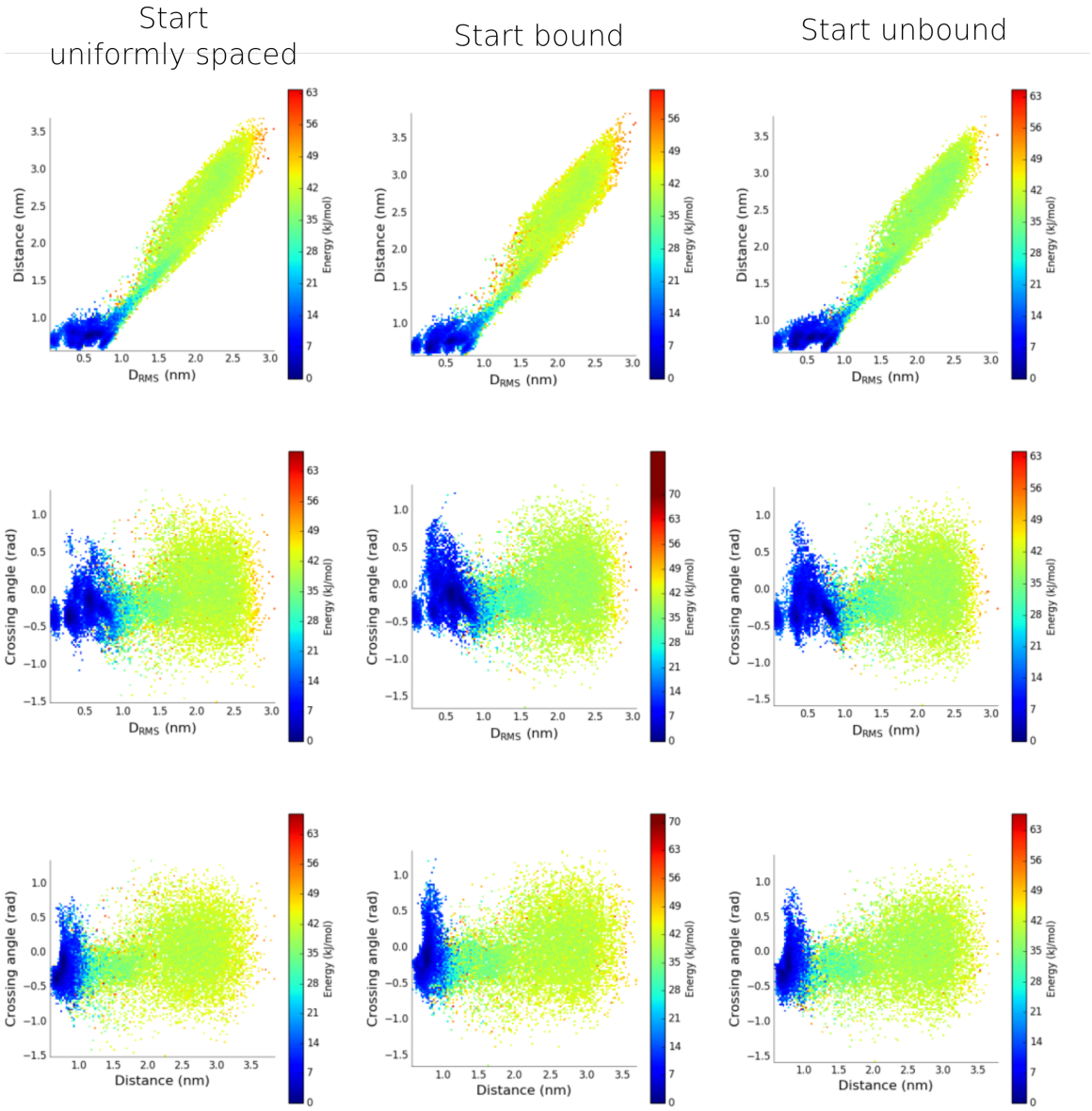


Figure 4.9: Identifying orthogonal collective variables in wild-type Glycophorin A association free energy surface. By projecting the PMFs calculated along: D_{RMS} and helix-helix COM distance (top row), D_{RMS} and helix-helix crossing angle (middle row) and helix-helix crossing angle and helix-helix COM distance (bottom row). Different columns correspond to three initial conditions. Taken with permission from [41].

To further explore if this limitation is inherent to the Glycophorin A, in the next section, I will verify the dimerization energetics of a different class of TM dimers, the receptor tyrosine kinases. To understand if this limitation can be circumvented by using a higher-resolution all-atom model, I will estimate an all-atom Glycophorin A dimerization PMF in the next chapter , where a higher resolution modul eis expector to more accurately match the experiment..

Table 4.2: Summary of methods used in coarse-grained simulations

GROMACS version used	5.1.2 with plumed2.2
Time-step	20 fs
Pressure control	Semi-isotropic in XY and Z, reference pressure of 1 bar, 12 ps coupling constant
Pressure control scheme	Parrinello-Rahman
Temperature control	Stochastic velocity rescaling temperature control (V-rescale), target temperature of 310 K, characteristic time of 1 ps
Lennard-Jones (LJ) interactions	Cut-off at 1.2 nm, switched-off smoothly between 0.9 and 1.2 nm
Coulombic interactions	Particle Mesh Ewald (PME) with real space cutoff of 1.2 nm and grid spacing of 0.12 nm

4.3 Receptor Tyrosine Kinases

4.3.1 Material and methods

The parameters used were taken from MARTINI 2.1 force field. The POPC lipid bilayer used the 5-tail bead model.

PDB structures of the four RTKs presented here (PDB codes: 2JWA, 2K1L, 2K9Y and 2M0B) were taken in their unmodified form. The martinize.py script was used to generate the coarse-grained structure and topology.

Each helix-helix pair was inserted into a POPC bilayer, with the final system contained approximately 128 POPC molecules and 1500 water beads. No ions were used to buffer the system.

The inter-helical D_{RMS} was used to bias along, with the umbrella sampling simulation. Sixteen windows were linearly spaced between 0 and 3.0 nm D_{RMS}, with a spring constant of 100 $\mathrm{kJ/mol/nm^2}$. Backbone beads of each dimer were used to define a set of native helix-helix distances and the D_{RMS} coordinate. Exchanges between neighbouring windows were attempted every 1000 simulation steps.

4.3.2 Results

To estimate the RTK TMs binding energy, the same protocol was used as for the Glycophorin A simulations. Well converged PMFs were obtained for the selected RTK dimers and are shown in Figure 4.10. The native state seen in structural models is not captured in simulations, not even as a local minimum; instead, the global minimum seems to be around 0.5–1.0nm D_{RMS} and corresponds to a range of non-native, bound structures.

The estimated PMFs for RTK TMs association could be used to calculate a dimerization ΔG or K_d, and compared with available experimental measurements. For EphA (PBD code 2K1L) the well-depth of the PMF is ~ 25 kJ/mol, experimentally from FRET measurements the ΔG is ~ 15 kJ/mol [6]. This is a good agreement, given the caveats on experimental and computational limitations in estimating helix-helix dimerization free energies.

The PMF for Her1/ErbB1 seems different than the others, in that it is much smoother, with no apparent native-like minimum but instead with a broad minimum at around 1.1 nm D_{RMS} (see Fig 4.10B). The Her1 construct is the longest used (see Table 4.1) and the entire length is used to define the D_{RMS} collective variable. The disordered tails of the dimer are part of the CV; however, these only introduce broadening as pushing on the helix tails does not drive the reaction. In the future, only the helical fragment of the dimer should be used to define the CV.

Experimentally, both the construct sequence used [202] and the lipid, or lipid-mimicking environment [57, 216] can affect the measured affinity. Case in point is the huge range of values obtained for the model Glycophorin A association energy. Experimentally, dimerization energy of around 30 kJ/mol is obtained in C_8E_5 detergent micelles [59], while a value of about 15 kJ/mol in plasma membrane-derived vesicles [172]. Computationally, the range of values obtained is equally considerable: atomistic simulations in dodecane slab, mimicking the lipid bilayer, yield an estimate of

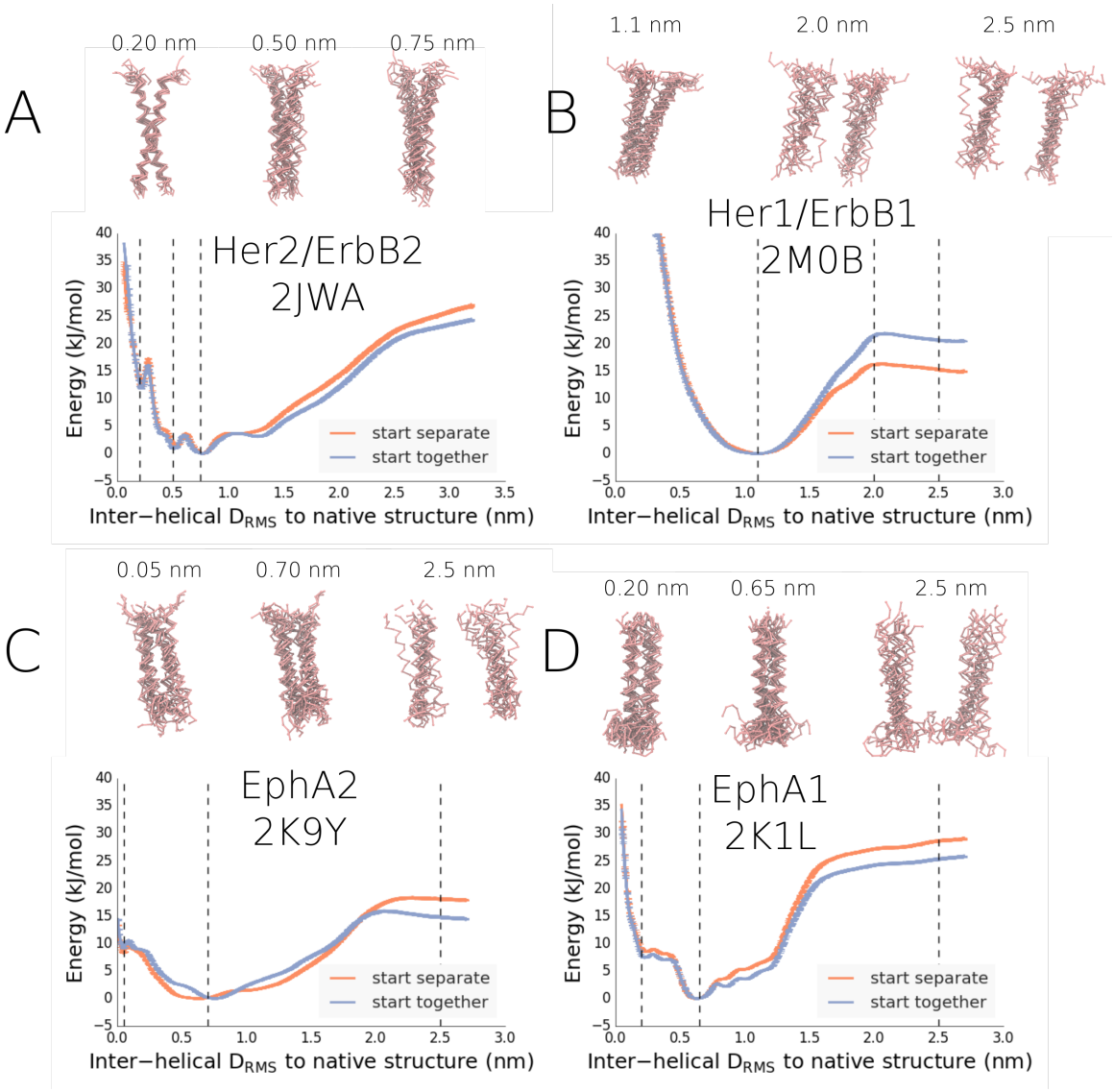


Figure 4.10: Helix-helix dimerization PMFs for 4 selected RTKs, computed along the inter-helix D_{RMS} collective variable. Each of the 4 sub-figures is showing a PMF calculated from "start together" and "start separate" initial conditions. Dashed vertical lines indicate characteristic points along the PMF, with representative structures above the plot.

approx. 45 kJ/mol [75], while coarse-grained simulations in lipid bilayer models give estimates around 30 [90] to 40 kJ/mol [179].

4.3.3 Conclusions and future work

Force field accuracy is important as it allows us to have confidence in the simulation-generated predictions. To understand how well the MARTINI 2.1 force field can model helix-helix interactions in the membrane, a comparison to experimental values is essential, such as the RTK dimerization affinity ranking generated by the TOXCAT assay.

Comparing simulations to experiment requires that a reasonable care is taken to ensure simulation convergence; otherwise, it is not clear if the results obtained are due to insufficient sampling or force field effects. The work presented in this chapter suggests that for a series of helix-helix dimers, with very different relative TOXCAT affinities, the MARTINI 2.1 force field yields similar PMFs: with a non-native bound configurations that are consistently more stable than the native configuration. This set of 4 TM helix dimers could be the first helix-helix dimerization force field benchmark. Future work could include extending this set to additional helix-helix pairs, including dimerization-disrupting mutants when there ase known and have been characterized experimentally. To facilitate comparison to experiment K_d or ΔG values could be calculated, this was not done here because all the dimers show similar behaviour and the TOXCAT assay provides only a semi-quantitative view of their affinity. Another improvement could involve an experimental reference based on a biophysical measurement, rather than a bioassay. While this would be further from a biological situation but where a greater control over experimental conditions is possible. Controlling for the lipid environment, or the temperature would all be of interest for force field calibration.

Chapter 5

All-atom GpA dimerization

Building on the findings from previous chapters, I apply the enhanced sampling methodology to study TM helix dimerization at all-atom resolution. This level of resolution is believed to provide sufficient level of detail to account for all relevant interaction types, avoiding approximations introduced by coarse-grained force fields.

The focus of this chapter will be understanding the energetics of the helix-helix interaction for GpA TM fragment in a POPC bilayer. Surprisingly, I find that despite the higher resolution offered by the all-atom force field, the dimer dissociates in equilibrium conditions, a dramatic difference relative to available experimental data. I speculate about the origin of this discrepancy, discuss why it has not been observed before in comparable simulations, and propose two simple corrections to stabilize the GpA helix dimer in POPC membrane with CHARMM36 force field.

5.1 Glycophorin A

5.1.1 Introduction

The GXXXG motif mediating Glycophorin A dimerization involves tight-packing of Glycine residues, and potential formation of a C_α-hydrogen interaction [139]. Such atomic-detail interactions may be critical to getting the system model behaving correctly in simulation and are inherently hard (if not impossible) to capture via a coarse-grained model. All-atom simulations can capture the details of such interactions while remaining at the limit of feasibility for simulation performance. Capturing the atomistic details of a system in a simulation requires a shorter time step, and it also makes the free energy surface more rugged. In a lipid simulation context, this is particularly important as the viscosity of the membrane is quite high. The pioneering work by Chipot used a dodecane 'slab' to mimic the membrane and speedup the lateral diffusion; using an enhanced sampling simulation the authors proposed a dimerization free energy estimate [75]. To address the timescale problem, Pastor at el. used the ANTON super-computer to perform $1\,\mu s$ unbiased simulations, demonstrating the stability of related TM helix pairs, the ErbB1/B2 and EphA [217]. Umbrella sampling simulations were used by Kuznetsov et al. to characterize the binding PMF of GpA in POPC bilayer with GROMOS 43a2 force field [103]; however, their PMF was monotonically decaying along the CV with no discernible minimum.

One of the principal challenges is obtaining a dimerization PMF at all-atom resolution. A collective variable along which the biasing can be performed (either using metadynamics or umbrella sampling) plays a key role. In the studies cited above, the GpA helix-helix PMFs were all biased along a helix center-of-mass distance. While it is a simple CV that can be readily interpreted, it has been shown that the dimerization landscape has barriers orthogonal to it: examples of CVs along which orthogonal barriers are present include the helix-helix crossing and relative helix twist angles

[110]. In the previous chapter on coarse-grained GpA simulations, I have shown the existence of a deep minimum for GpA in MARTINI 2.1 force field as a function of the helix-helix centre of mass distance. However, that minimum corresponds to a non-native, parallel arrangement of helices. I have shown that inter-helical D_{RMS} is a better collective variable than the centre-of-mass distance in this system [41].

Comparison to experiment remains another challenge. The steric trap experiments cannot exclude the existence of a non-native bound dimer. It is only indirectly inferred to be non-existent or near non-existed from experiments on mutants that show no detectable binding – that is assuming that the mutation only disrupts the native bound state and has no effect on a non-native bound state. A further challenge is for the simulation PMFs to be reasonably converged. In the previous chapter, a working definition of convergence was that the PMFs started from different initial configurations should be identical. Specifically, I define two initial conditions: a native "bound" configuration, as seen in the experimental structures, and an unbound, separated configuration with the helices at an appreciable distance apart generated by a constant-force pulling simulation. In the previous chapter, convergence for a coarse-grained MARTINI force field was obtained after over $5\,\mu s$ of REUS simulation [41].

A popular force field for macromolecular simulations is CHARMM36 [83]. This force field is known to faithfully reproduce various experimental lipid bilayer properties [94, 149] as well as the properties of folded globular proteins [83, 155, 84]. In this chapter, I used CHARMM36 to study the dimerization energetics of Glycphorin A TM fragment. To this aim, I have performed REUS simulations with biasing along D_{RMS} collective variable (see Methods section for the definition).

The estimated free energy surface is characterized by a native minimum separated by a steep 10-20 kJ/mol barrier, which has not been seen in previous work using centre-of-mass distance CV. Structural interpretation of the barrier is the breakage of

a tightly packaged GXXXG motif: representative structures on either side of the barrier differ primarily at this motif. However, the native state is only a local minimum, which is in marked disagreement with a large body experimental data. In simulation, the dimer is *unstable* at equilibrium in phosphatidylcholine (POPC) bilayer and dissociates.

I have tested a number of force field corrections to bring the force field in line with available experimental data. The simplest correction involves modifying the protein-lipid van der Waals (vdW) cross-interaction, by scaling down the vdW potential ϵ parameters. A modest correction of this type, makes the bound state more favourable than the unbound; furthermore, the native bound state is more stable than the competing non-native bound configurations.

5.1.2 Methods

5.1.2.1 Simulation methods

All simulations used GROMACS simulation code version 4.6.7 (http://www.gromacs.org)[160]. GROMACS was patched with the PLUMED 2.2-hrex for enhanced sampling functionality [196]. Parrinello-Rahman barostat [148] was used to maintain a reference pressure of 1 bar. The pressure coupling was done separately for the XY and the Z box dimensions, using a coupling constant of 5 ps. Stochastic velocity rescaling [19] was used to keep the temperature close to a reference value of 310 K, corresponding to the physiological temperature, using a relaxation time of 1 ps. Lennard-Jones interactions between particles were shifted to zero and cutoff between 0.9 and 1.2 nm. Particle-mesh Ewald method [37] was used to calculate long-range electrostatic interactions with a grid spacing of 0.12 nm, and the short-range electrostatic interactions cut-off at 1.2 nm.

NAMD version 2.12 [154] was used to run simulations in the NVT ensemble. Electrostatics and van der Waals settings used were similar to those used in GROMACS. Nonbonded interactions were truncated beyond 1.2 nm; between 1.0 and 1.2 a force-based switching function was applied. Non-bonded interaction pair-list was updated every 10 time-steps, the pair-list was calculated with a cutoff of 1.6 nm. Electrostatic forces and interactions were calculated with the Particle-mesh Ewald method, using a 6th-order spline and grid spacing of 0.1 nm. The simulation was run using Langevin dynamics integrator, using a 2 fs time step and friction coefficient of 1.0 ps^{-1}.

The standard implementation of temperature replica exchange was used in GROMACS [192]. Temperature ladder used is shown in Table 5.1. GROMACS 4.6.7 with PLUMED 2.2-hrex patch was used to run the solute tempering (REST2) simulations [205]. Exchanges between replicas were attempted every 1000 MD steps; acceptance probability was in the range of $0.05 - 0.20$. Interactions involving the protein and

Code	Force field	Method and ensemble	Ladder spacing	No. of lipids
Gromacs 4.6.7	CHARMM36	T-REMD NPT	45 windows, 310-450K	78 POPC
Gromacs 4.6.7	CHARMM36	REST2 NPT	16 windows, 300-400K T_{eff}	112 POPC
Gromacs 4.6.7	AMBER	REST2 NPT	45 windows, 300-450K T_{eff}	112 POPC
NAMD 2.12	CHARMM36	T-REMD NVT	40 windows, 300-450K	78 POPC

Table 5.1: Summary of replica exchange simulations performed. Taken with permission from [40].

palmitoyl oleoyl phosphatidylcholine (POPC) atoms were scaled using the parameter λ, for the REST2 simulatinos.

Protein and lipid force field used the CHARMM36 model [83] with the TIP3-CHARMM water parameters[141, 123].

5.1.2.2 System preparation

PDB entry 1AFO [122] was used to define the initial structure for Glycophorin A TM simulations [122]. The flexible N- and C-term residues were removed, and only residues between Ser-69 and Arg-97 were simulated. A square POPC bilayer was used after the protein was inserted the final system contained 112 POPC lipids, equally distributed between both leaflets, and around 4000 explicit solvent molecules. The dimensions of the system were approximately 6.5 nm in the X and Y dimensions. (see Fig. 5.1) For selected simulations (as specified) a rectangular lipid patch was used, which contained only 78 POPC molecules. The dimensions of the smaller system were 3.2 x 6.4 nm in the X and Y dimensions. Unless otherwise stated, the larger, square system with 112 POPC lipids was used in the simulations.

5.1.2.3 Potential of mean force calculations

D_{RMS} collective variable was defined using the Glycophorin TM fragment encompassing the GXXXG motif, which mediates the binding. Heavy, non-symmetric atoms of residues 78-88 were used to define the distance matrix. Only the all-atom distances

above the lower cutoff of 0.1 nm and below the upper cutoff of 0.6 nm were included in the collective variable calculation. D_{RMS} is thus defined as:

$$D_{\mathrm{RMS}}(X^A, X^B) = \frac{1}{N_{cont}} \sum_{i \neq j \in contacts} [d(x_i^a, x_j^a) - d(x_i^b, x_j^b)]^2 \tag{5.1}$$

where $d(x_i^a, x_j^a)$ is the distance between the coordinates x_i and x_j of atoms i and j in configuration X^A; and summation runs over all atom pairs (i, j) specified in the contact list.

An umbrella sampling replica exchange (REUS) simulation was configured with 40 windows spaced non-uniformly between D_{RMS} of 0 and 2.5 nm. Umbrella center positions and spring constants used are given in the Appendix. Exchanges between neighbouring replicas were permitted to aid sampling. Vector connecting sC_α atoms of residues 78 to 88 were used to define the crossing angle between the helices.

The umbrella sampling simulations were unbiased using the WHAM method [100]. Grossfield lab WHAM code, version 2.0.9 (http://membrane.urmc.rochester.edu/content/wham) was used to unbias the simulations, first half of each simulation was discarded as equilibration.

Estimation of the dissociation constant K_d was done using the equation below:

$$\frac{1}{K_d} = 2\pi \int_0^b exp[-\beta F(r)] dr \tag{5.2}$$

Where K_d is in units of molecule/nm^2, β is $\frac{1}{k_{\mathrm{B}}T}$, F(r) is the dimerization PMF along r, the radial centre-of-mass distance CV, with the entropy correction $k_{\mathrm{B}}T \log(r)$ added. With that correction, there should be some distance b above which $F(r)$ is constant, which defines the bound state. It is assumed that $F(r) = 0$ for large distances, so that the above integral converges, which can be ensured by adding a suitable constant to $F(r)$.

5.1.3 Analysis and visualization

VMD was used for visualization and rendering VMD [86], analysis was performed using MDAnalysis [134].

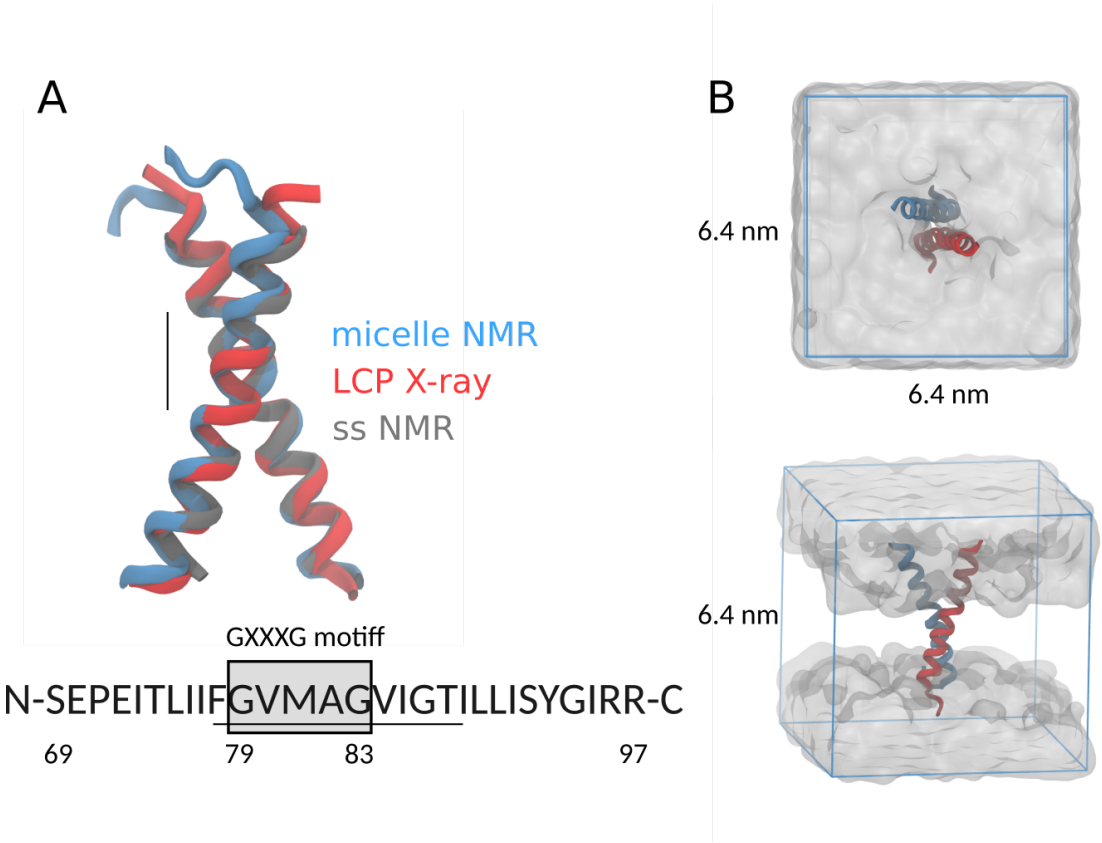


Figure 5.1: (A) Available structural models for GpA TM dimer structure, showing the fragment used in simulation in cartoon representation. Thin black line, near the structure and in the sequence, indicates the region of the protein used to define the inter-helical D_{RMS}. The GXXXG motif is boxed on the sequence plot. (B) The square system box, with approximately 128 POPC lipids. Explicit solvent is rendered as a transparent surface, the lipid bilayer is not shown for clarity, protein shown in cartoon representation. Taken with permission from [40].

5.1.4 Results and discussion

5.1.4.1 Potentials of mean force

To estimate the dimerization free energy landscape for GpA, I performed REUS simulations of the dimer biasing along the inter-helical D_{RMS} coordinate. To check if the result is independent of initial conditions, I have started two sets of simulations: one initialized from the bound, native conformation taken from the experimental structures (termed "start together") and the other set initialized from an unbound structure in all the windows (termed "start separate"). This setup is equivalent to

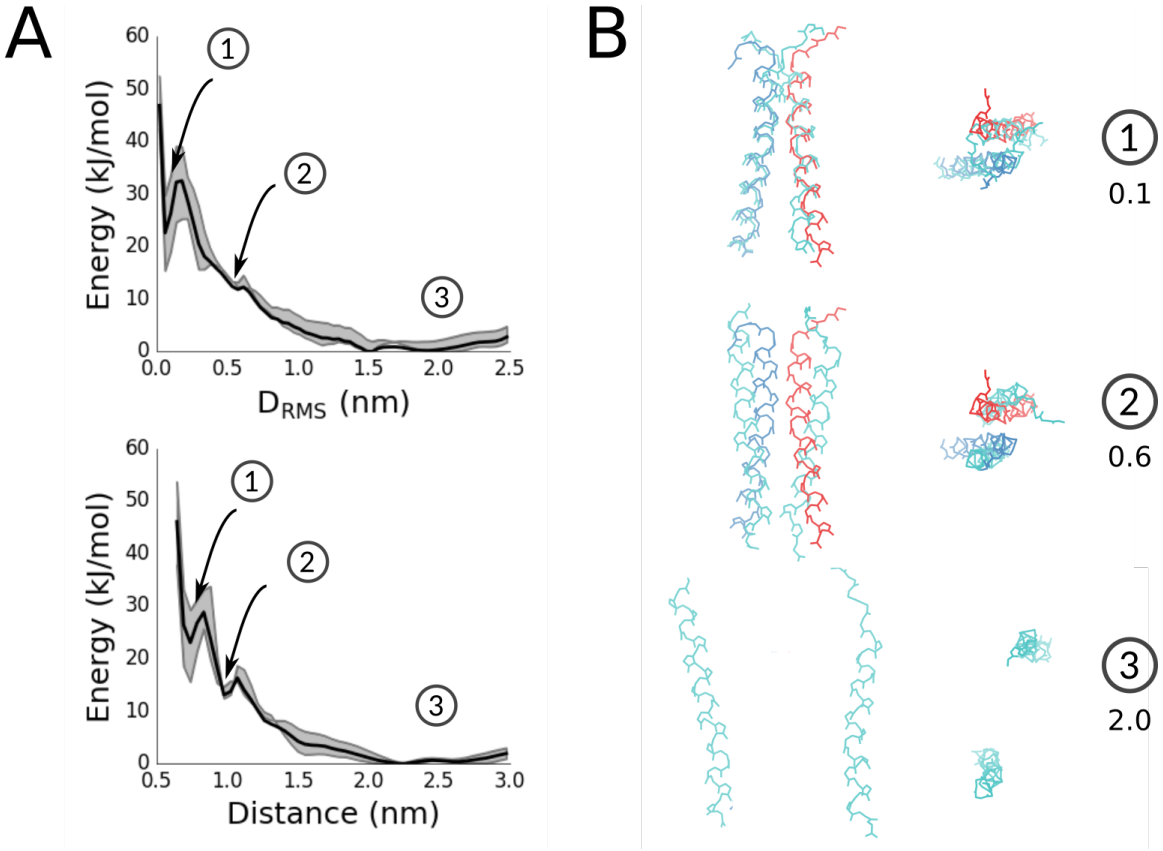


Figure 5.2: (A) Potential of mean force calculated along inter-helical D_{RMS} (upper panel) and projected along D (lower panel). With characteristic minima labelled 1–3. The black line indicates the pooled PMF obtained from "start together" and "start unbound simulations", the thin lines indicate the PMF obtained from just the "start together" and "start separate" simulations. (B) Representative structures for minima 1–3, with their D_{RMS} value, show in cyan backbone representation. Experimental 1AFO structure shown in blue and red, for chains A and B respectively. Side and top views. Taken with permission from [40].

the one described in the earlier Chapter of this thesis describing the work on coarse-grained GpA dimerization [41].

The PMFs were estimated using WHAM by unbiasing the REUS simulations and are shown in Fig. 5.2A and B, projected along D_{RMS} and centre-of-mass distance (D) reaction coordinates, respectively. Despite nearly 0.5 μs of simulation time the PMFs are not completely overlapping. During the simulation the helices stay folded all the time, so the sampling challenge is instead related to the high-viscosity of the membrane.

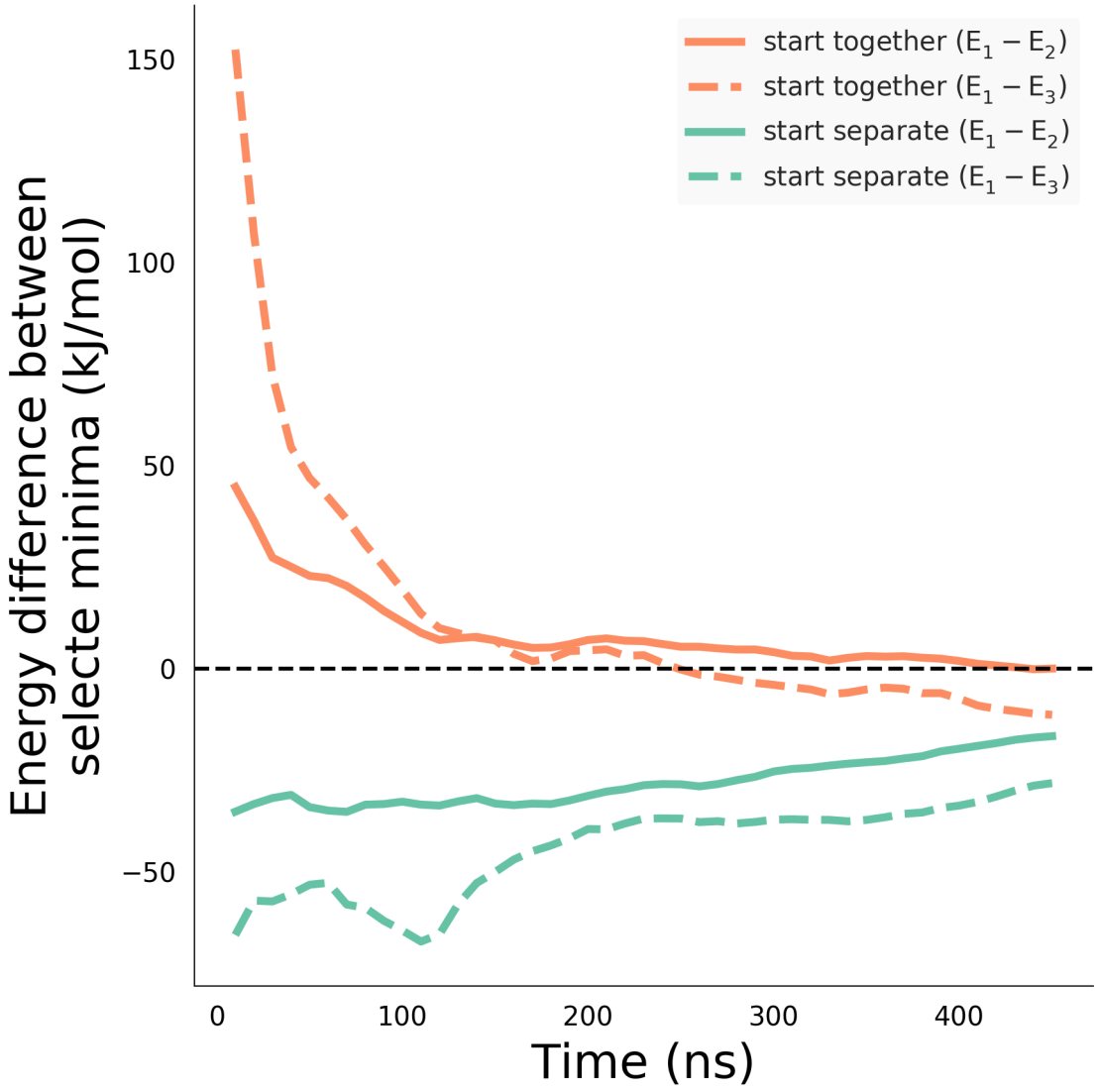


Figure 5.3: Test for convergence of the PMFs showing energy differences between minima as a function of simulation time. Energies of minima "1", "2" and "3" are designated as E_1, E_2 and E_3. Taken with permission from [40].

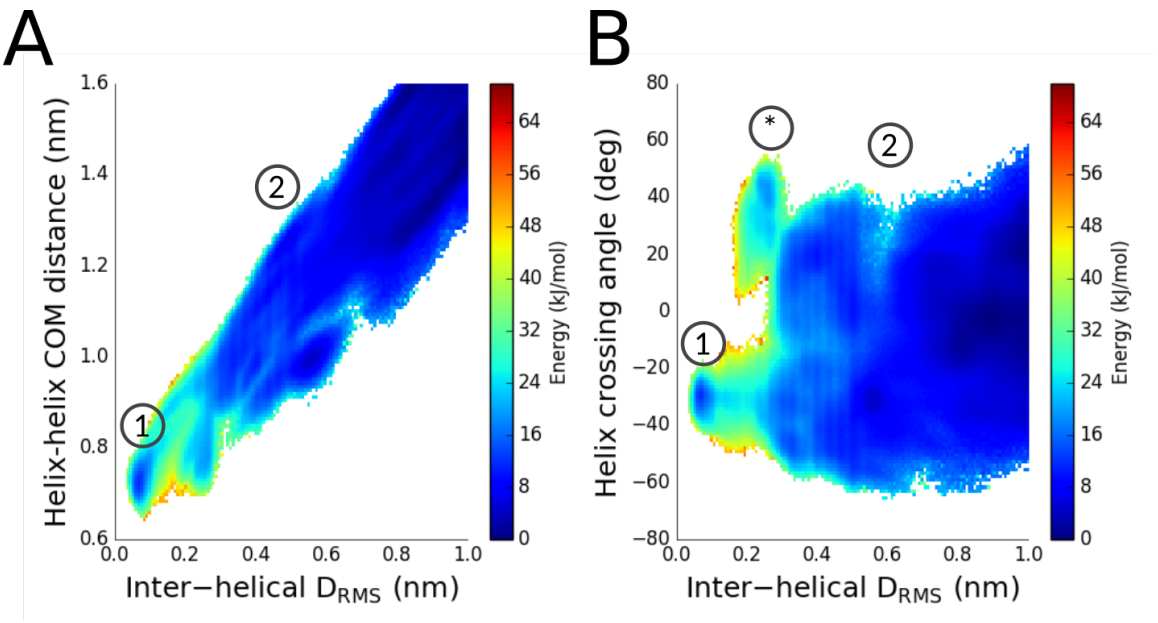


Figure 5.4: 2D PMF for GpA dimerization projected along inter-helical D_{RMS} and D (A) as well as D_{RMS} and helix crossing angle (B). The bound native and non-native states are bound as 1 and 2, respectively. Asterisk in (B) indicates an additional non-native bound state with inverted handedness of the helix dimer. The figure shows data pooled from the "start together" and "start separate" simulations. Taken with permission from [40].

The PMF contains two regions labelled as "1" and "2". With increasing simulation time, the free energy gap between those regions becomes smaller, independent of initial condition (see Fig. 5.3). This suggests that slow diffusion within the bilayer limits convergence. The PMFs started from the two different initial conditions are both pointing to the unbound state being more stable by approximately 20-30 kJ/mol. The two bound states, native and non-native, labelled "1" and "2" are separated by an energy barrier, approximately 10 kJ/mol when projected along D_{RMS} (Fig. 5.2A). Structurally this barrier is related to breaking of a tight interaction at the GXXXG motif.

The existence of this barrier may explain why TM helix dissociation was not observed in long, unbiased MD simulations of a related TM helix dimer [217]. The barrier has not been seen in previous work, in both coarse-grained [179] and all-atom simulations [103], which used D as a reaction coordinate. In my simulations,

a smaller barrier of about 5 kJ/mol is observed when projected onto D, the centre-of-mass distance (Fig. 5.2B), which may help to explain why this barrier was not prominent in previously calculated PMFs. Alternatively, the absence of the barrier is just due to insufficient sampling. Another possibility is that the barrier simply is not present in the other force fields.

D is a worse collective variable than D_{RMS} for studying GpA dimerization: multiple minima along D_{RMS} exist corresponding to a D value of 0.7 nm, as apparent from a 2D PMF projection (Fig. 5.4). At large separation between helices D and D_{RMS} behave similarly. Having said that, D_{RMS} is not a perfect collective variable either. Like any distance-based CV, D_{RMS} is unable to distinguish states with different handedness. Thus, it should come as no surprise that D_{RMS} appears degenerate with respect to helix-helix crossing angle θ: a D_{RMS} value of 0.25 nm corresponds to two basins along θ (Fig. 5.4B).

5.1.4.2 Additional simulations

The apparent instability of GpA dimer in CHARMM36 force field was surprising. It is a state-of-the-art force field used by many researchers, and this result is inconsistent with experimental data available for GpA [199]. I have run additional simulations, without umbrella sampling and calculated with a different simulation code and other enhanced sampling techniques, to confirm that this dissociation can be robustly observed. In the replica exchange simulations I performed, an unbound fraction of GpA appears within 50-300 ns in the 1st replica, corresponding to physiological temperature conditions (see Fig. 5.5). Two flavours of replica exchange simulations were used: standard temperature replica exchange (T-REMD) and solute tempering replica exchange (REST2) (see Fig. 5.5). The dissociation appears to be system-size independent, occurring also in a smaller rectangular lipid patch system. There is also no dependence on the MD code used: T-REMD simulations run with NAMD in the

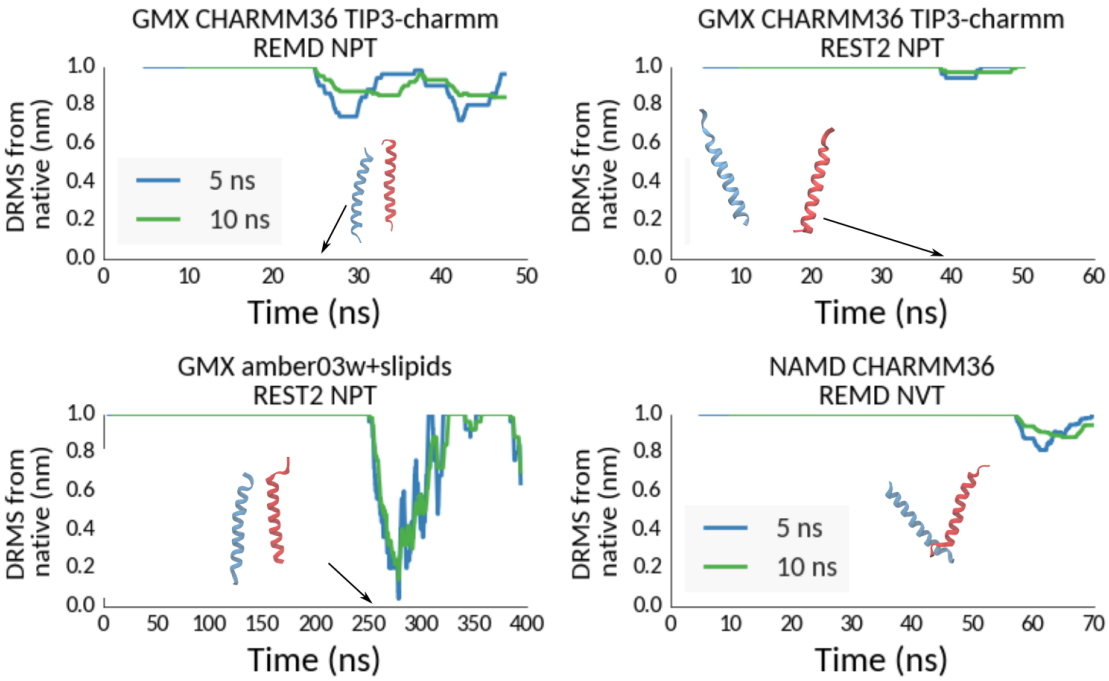


Figure 5.5: Dissociation of GpA TM dimer in simulation at physiological temperature (310K). Fraction-of-dimer timeseries at 5 and 10 ns lag times from replica exchange simulations, where "dimer" is defined as inter-helical $D_{\mathrm{RMS}} < 1.0nm$). Simulations were performed using GROMACS with the CHARMM36 force field at NPT ensemble. Both the T-REMD and REST2 methods show the dimer dissociate in 30-50 ns (upper row). This behaviour is not specific to CHARMM, AMBER Slipids simulation run using GROMACS with the REST2 method shows the same behaviour (lower row, left column). Dissociation also occurs in NAMD simulation code, with CHARMM36 force field. Further details available in Table 5.1. Taken with permission from [40].

NVT ensemble show the dimer dissociating. Amber Slipids is another widely-used parameter set for protein-lipid simulations [89, 88]. During a REST2 simulation setup with Amber ff03w and Slipids, I observe the GpA dimer dissociation in the 1st replica within 250 ns.

To estimate the population of the bound and unbound states in the 1st replica, a much longer replica exchange simulation would be needed. However, just observing the existence of any unbound population in the 1st replica is inconsistent with available experimental data. Available experimental data suggests that no detectable unbound fraction should be present at physiological temperature, independent of lipid type used or experimental method.

5.1.4.3 Checking lipid behaviour in simulation

Lipid simulations can be particularly sensitive to the choice of parameters, in particular, the treatment of long-range dispersion and electrostatic interactions. For example, cut-off electrostatics can result in freezing of DPPC bilayers [150]; instead, using PME electrostatics is strongly recommended. Alan Mark and co-workers recently showed that MD simulation code changes can cause subtle artefacts: using twin-range cut-off for van der Waals and Coulombic interactions with a new version of GROMACS (5.1.2) can push DPPC to form liquid-order like phase above its transition temperature [165]. DPPC is particularly sensitive as its phase transition temperature is 315 K, very close to the temperatures at which simulations tend to be run at. In my simulations I have used exclusively POPC which should be less sensitive to parameter choices, with its transition temperature T_c at 271 K.

In the REST2 simulations, a scaling factor λ is used to adjust the various energy-function terms of the atoms in the "hot" group (lipids in our case).

Scaling factor of $\lambda = 1.0$ corresponds to the unmodified energy function, scaling factor of 0 means that the group of atoms has been "turned off" and is no longer

λ	T_{eff}
1.0	300
0.75	400
0.5	600
0.375	800

Table 5.2: Example values showing the relationship between λ and T_{eff}, $\lambda = T_0/T_i$

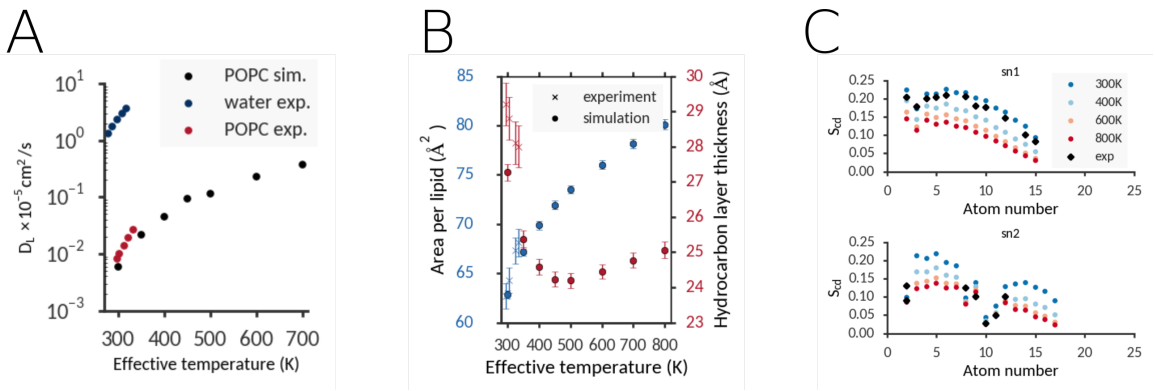


Figure 5.6: Pure POPC bilayer properties as function of scaling factor λ (also referred to as effective temperature). (A) Diffusion coefficient as function of temperature, experimentally for water [200] and POPC [54]. For the POPC simulation, effective temperature was used. (B) Bilayer thickness (red), and area per lipid (blue), experimentally as function of temperature and as function of effective temperature for simulations. Experimental data on water and lipid diffusion, bilayer thickness was taken from [99]. (C) Lipid tail deuterium order parameters from simulation and experiment. Experimental NMR data (in black) was collected at 300 K [177]. Taken with permission from [40].

interacting with other components of the system. Decreasing the value of the scaling factor λ is thus commonly thought to make the affected group of atoms "hotter". Therefore, for better chemical intuition, λ is sometimes re-written as T_{eff}, effective temperature (see Table 5.2 for example values). I performed simulations of pure POPC bilayers in explicit solvent with a range of scaling factors λ to ensure that the REST2 method does not perturb the expected lipid properties. Critical properties of the bilayer were checked: area per lipid, lateral diffusion coefficient and S_2 lipid-tail order parameters. All the parameters appeared to be within experimental error, for the replica run with $\lambda = 1$. With decreasing λ, lipid diffusion becomes faster and disorder of the tails increases; this is also accompanied by an increased area-per-lipid

and thinning of the bilayer (Fig. 5.6A-C)). Visual inspection of the pure POPC bilayer REST2 trajectories showed no bilayer defects (such as holes) over 100 ns timescale as the λ was decreased.

5.1.4.4 Proposed force field fixes

The PMF and additional replica exchange simulations illustrate a deficiency in the CHARMM36 force field. To propose fix, I want to start by a short discussion on force field development. The protein and lipid parts of all-atom force fields are optimized separately to a large extent: the lipid force field is optimized to match experimental observables for lipid/water systems, while the protein force field is optimized towards protein/water properties. Mixed protein-lipid systems are rarely used in parametrization but the behaviour is expected to be reasonable, provided the protein-water and lipid-water parametrization is sound.

Protein-lipid force field can be evaluated directly by comparing to appropriate experimental data [119, 118, 157]. The challenge is related to the sparsity of experimental data for mixed, protein-lipid systems that can be converged in simulation. The experimental observable that is easiest to compare to is side-chain analog or amino-acid partitioning, which is typically measured as partitioning between water and either a lipid or octanol. However, the limitation of this approach is that it may not be fully transferable to larger peptides and proteins; also, it does not test the strength of protein-protein interaction in the membrane [212, 211, 209, 1]. Given the complexities outlined above, I will aim to make a minimum correction that ensures that GpA stability in simulation in line with experimental data.

This correction could be accomplished, for example, by modifying the lipid force field in a way that changes the bilayer properties and indirectly stabilizes the TM dimer. However, the lipid force field currently reproduces many of experimental observables and any modifications to it would require extensive validation of an already

well-optimized parameter set. A similar argument applies to modifying the protein force field. The only remaining possibility is introducing a correction by modifying the protein-lipid interactions. The protein-lipid interaction strength can be tuned using a single parameter κ, used in the combination rule:

$$\epsilon_{ij} = \kappa(\epsilon_i \epsilon_j)^{1/2}, \tag{5.3}$$

where ϵ_{ij} is the Lennard-Jones potenital depth for atoms i and j, κ is a scaling factor, and ϵ_{ij} is the off-diagonal Lennard-Jones contact energy for atoms i and j (i and j are protein and lipid atoms, respectively). Thus, only atom pairs involving one protein and one lipid atom have their standard combination rule adjusted by the factor κ. In contrast to λ, which affects all the nonbonded interactions of a selected group, κ affects only those between protein and lipids.

To test the effect of introducing the protein-lipid correction, I re-weighted the PMFs obtained in the previous section. Those PMFs were obtained using the original, unmodified CHARMM36 force field and will be termed $\kappa = 1.0$ The PMFs were re-weighted using a scaling factor κ of 0.95 and 0.9 scaling down the van der Waals interaction between protein and lipid components (see Fig. 5.7A). Smaller scaling factors, such as $\kappa = 0.8$, were not necessary since I already obtained a stabilization of the bound state in the PMF, and the native minimum.

Further umbrella sampling simulations were then setup, using the last frames from the $\kappa = 1.0$ simulations, with the new force field scaled with $\kappa = 0.9$, since the re-weighted PMFs were quite noisy. I confirmed by re-sampling at $\kappa = 0.9$ the improvement predicted by re-weighting the simulation (Fig. 5.8).

While the main benefit of this correction is its simplicity, I cannot exclude that it has some unexpected side-effects. Other properties may be affected including amino-acid partitioning between water and the bilayer. I do not expect these properties

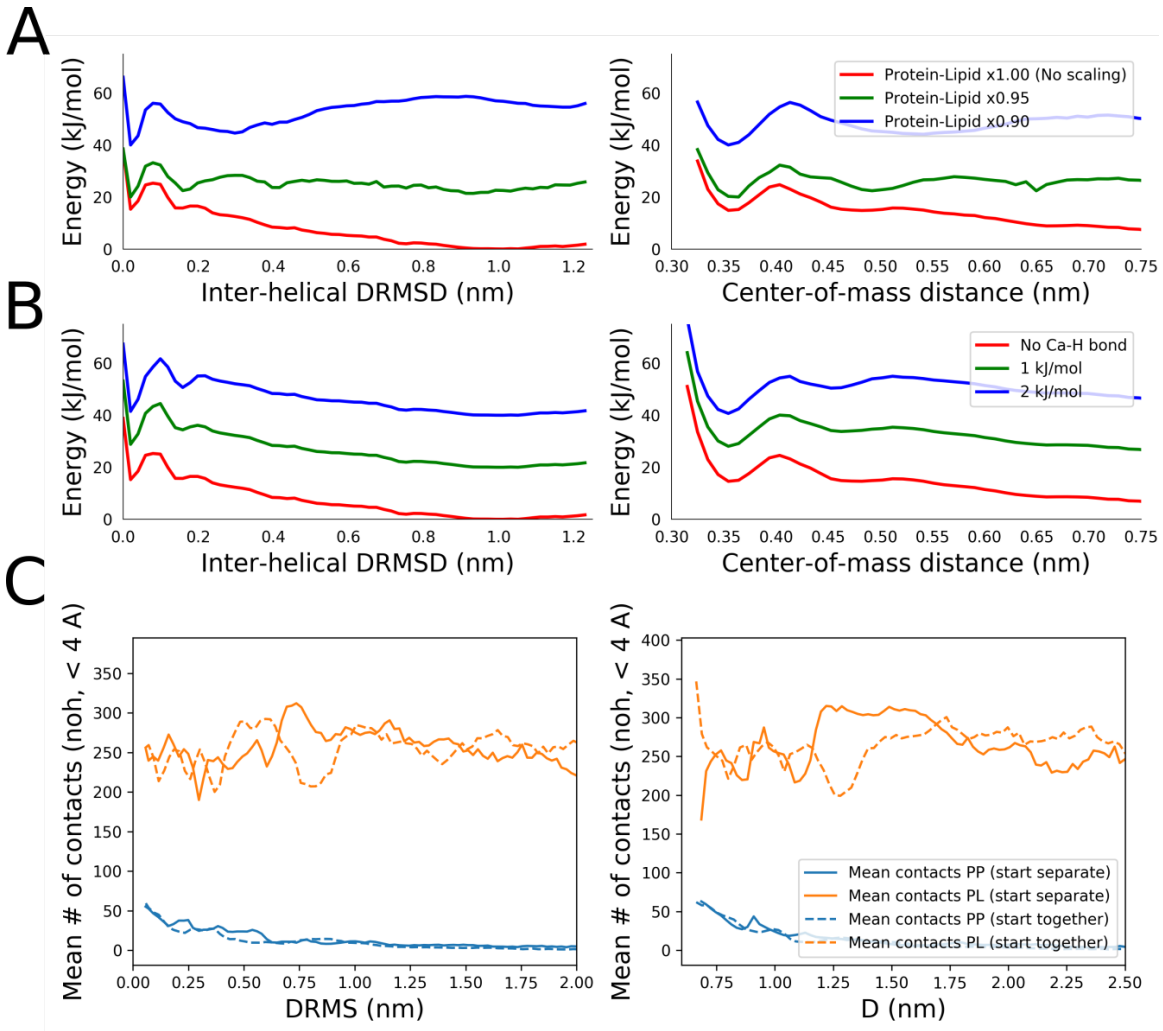


Figure 5.7: Changing the CHARMM36 force field parameters to stabilize the native state of GpA dimer, projected onto the inter-helical D_{RMS} (left column) or helix centre-of-mass distance D (right-column) (A) Scaling the ϵ of the van der Waals interaction between protein and lipid atom-types, by a factor of 1.0, 0.95 and 0.9. Scaling at 0.9 was obtained by running an independent REUS simulation, while scaling at 0.95 was obtained by re-weighting the simulation run with 1.0 scaling (corresponding to the original force field). The re-weightings derived from the 1.0 and 0.9 κ-scaling runs are self-consistent (see Fig. 5.8 below). (B) Correction by introducing a C_α– hydrogen bond-like term: re-weighting simulations using the original CHARMM36 force field with 0, 1 or 2 kJ/mol per formed C_α-hydrogen bond between the monomers (see main text for C_α– hydrogen bond definition). (C) Change in the number of protein-protein contacts (heavy atoms, within 0.5 nm cutoff) as the function of helical separation. Taken with permission from [40].

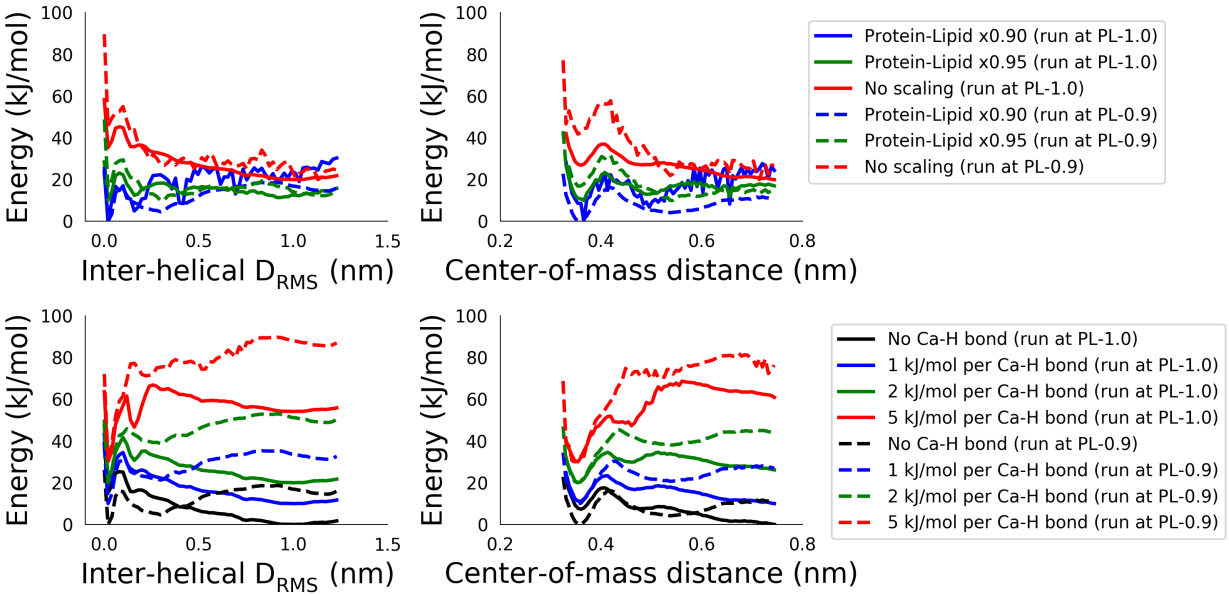


Figure 5.8: Re-weighting the PMF along two collective variables for simulation run at κ 0.9 and 1.0 (with and without the scaling correction). The projection is done along inter-helical D_{RMS} and D. Upper and lower row are fixes involving protein-lipid interaction scaling and introduction of a C_α-hydrogen bond. The re-weightings are consistent: the PMF obtained by obtaining simulation run at $\kappa = 1.0$ and re-weighting onto $\kappa = 0.9$ is largely the same as obtained by running the simulation at $\kappa = 0.9$ (and vice-versa). Taken with permission from [40].

to be strongly affected: the magnitude of the proposed correction is small and the water-bilayer insertion free energies are typically quite high.

Alternatively, it is possible that a more specific interaction mediates the GpA dimerization and is not captured in force field. GXXXG motif mediates GpA binding and it has been suggested that C_α–hydrogen bonds are formed between the monomers, stabilizing the bound complex [139]. This type of bonding may also be found in other TM helix dimers [168, 117]. These C_α–hydrogen bonds form between C_α–H bond and acceptors with a π system, such backbone carbonyl groups [64, 152]. Therfore I tested if introducing this correction stabilizes the native, bound GpA configuration. I re-weighted the PMFs obtained with the original force field: C_α–H hydrogen was considered to be formed when the hydrogen was within 0.32 nm of either a hydroxyl or a carbonyl oxygen in the other monomer. 1, 2 or 5 kJ/mol were given for each interaction formed. I have found that this correction only has a localized effect: it stabilizes only the native minimum, changing only that region of the PMF. (Fig. 5.7B)

Using GpA TM helix dimer alone it is impossible to say which of the two corrections proposed is more appropriate. A larger number of examples would have to be tested to evaluate the appropriate strength of the C_α–H interactions. It has been estimated that the magnitude of this interaction is 7.9-10.6 kJ/mol per bond in proteins [173]. In light of this, the correction seems reasonable in magnitude but this interaction must be, at least partially, already accounted for by existing electrostatic terms in the force field.

5.1.4.5 Comparison to a coarse-grained model

Comparing the PMFs obtained using the all-atom CHARMM36 and the coarse-grained MARTINI force fields reveal interesting differences. The all-atom PMF correctly identifies an native minimum, and finds it to be more stable than the non-native

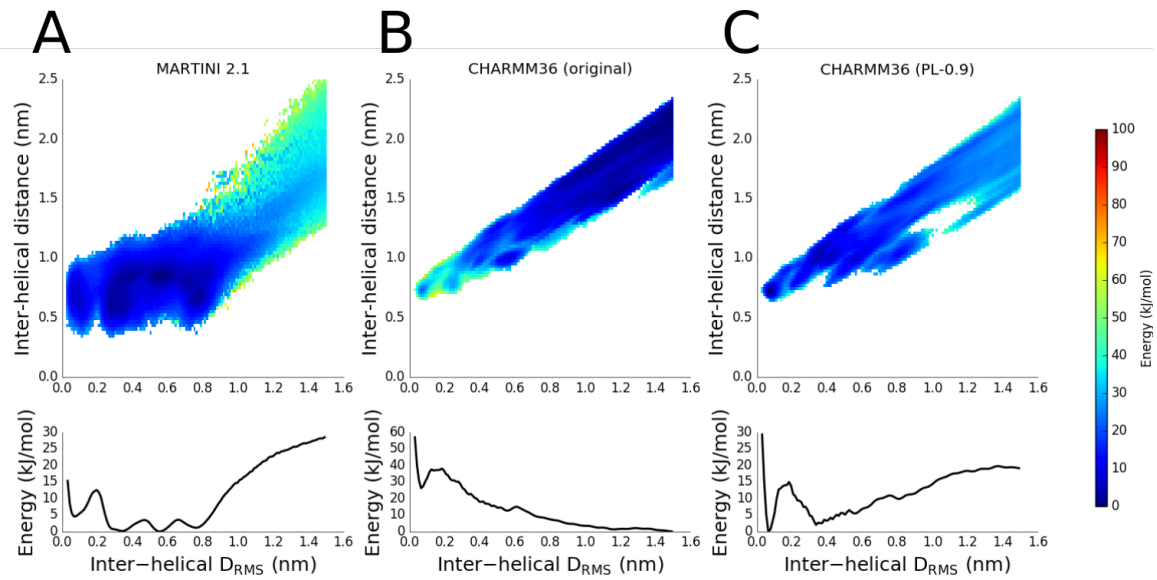


Figure 5.9: Comparing the dimerization energy landscape for the coarse-grained and all-atom force field. 2D PMF projected along inter-helical D_{RMS} and D (upper row) as well as the 1D PMF projected along D_{RMS}, for the coarse-grained MARTINI2.1 (A), all-atom CHARMM36 (B) and the modified $\kappa = 0.9$ CHARMM36 (C). Taken with permission from [40].

bound. However, the all-atom PMF incorrectly predicts the bound state to be overall unfavourable. In contrast, MARTINI correctly predicts that the GpA dimer state is stable, i.e. in agreement with experiment, but is unable to distinguish the native and non-native bound states, which appear to be of similar energy [41] (see Fig 5.9). I have shown that stabilizing the dimer in all-atom simulations can be adjusted by changing the protein-lipid interactions. The all-atom model appears to be able to discriminate the native and non-native bound states naturally.

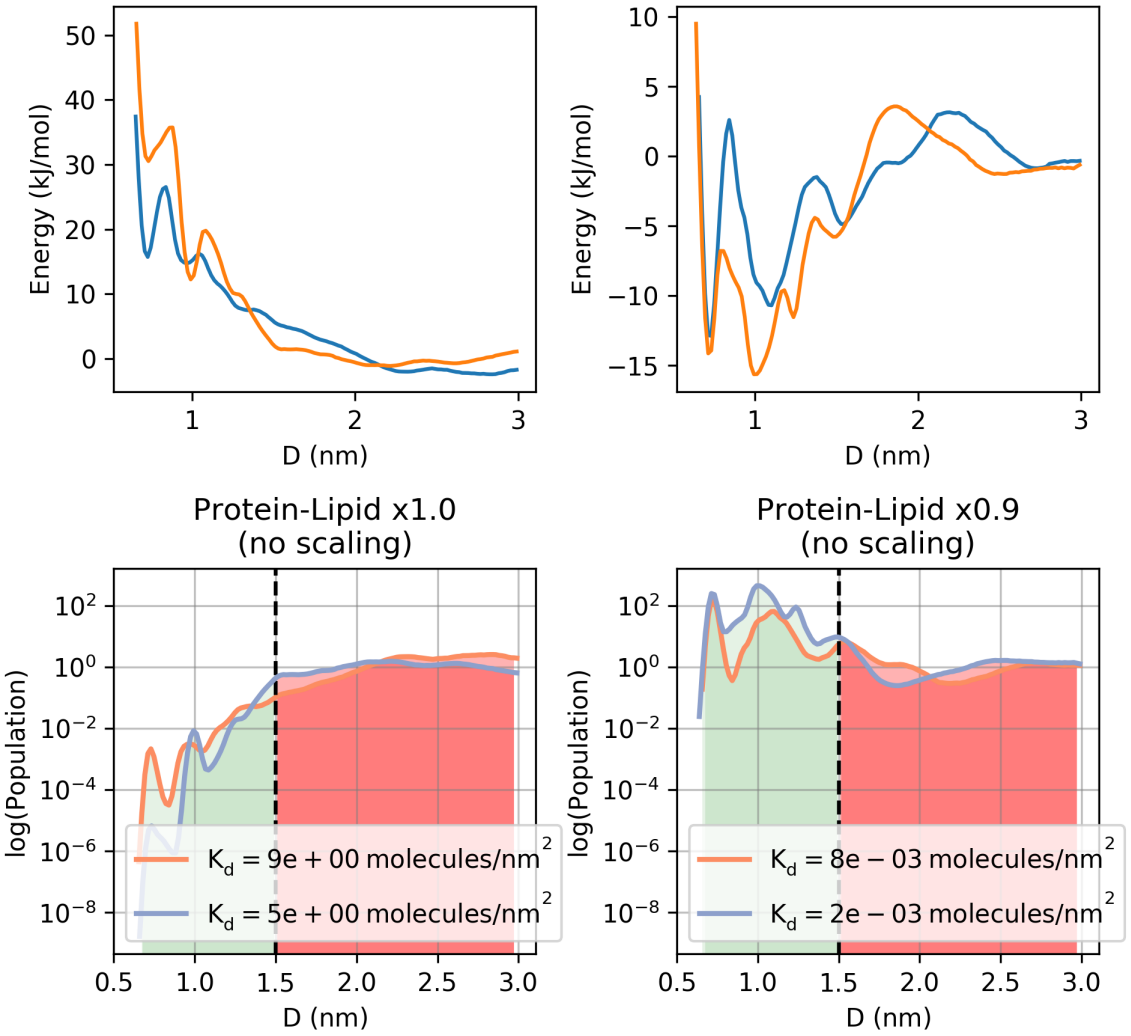


Figure 5.10: Estimating GpA TM dimerization K_d from the PMFs projected along the radial D centre-of-mass distance CV, for the original CHARMM36 force field (left-column) and the modified, $\kappa = 0.9$ CHARMM36 force field (right-column). For the equation used to estimate the K_d from the potential distribution integral see Eq. 5.2. Taken with permission from [40].

5.1.5 Conclusions and future work

Force field parameterization relies in part on correctly capturing the energetics of small chemical building blocks. Properties such as partitioning coefficients and hydration free energies are frequently calculated for small peptides [212, 211, 209, 1]. It is then hoped that agreement obtained for these small species is "transferable" onto larger macromolecules. In this chapter, I have explicitly tested if a well-characterized system such as Glycophorin A TM dimer is correctly described by current all-atom force fields. To study membrane protein folding, it is essential to capture (in computational models) the properties of this and other model systems. I have used state-of-the-art enhanced sampling methods and collective variables to approach this challenging problem. Due to high bilayer viscosity obtaining equilibrium sampling is not trivial.

I find that the CHARMM36 force field captures the helical structure of the GpA dimer. The existence of a native minimum is also correctly represented and not found in coarse-grained models tested (MARTINI 2.1 and 2.2). With the introduction of a force field correction, I was able to stabilize the dimerized structure as a global energy minimum. The proposed corrections include either re-balancing the protein-lipid van der Waals interaction ϵ or introduction of an C_α-hydrogen bond term. Due to the separate derivation of protein and lipid force fields, and likely errors emerging when the models are mixed, I favour the protein-lipid correction.

Experimentally, only the dimerization K_d value can be obtained. The PMF obtained from the simulation can be reduced to a single K_d value and compared to experiment. It is worth noting that an infinite number of PMFs can be reduced to a single K_d value. Thus, the choice of force field correction strategy is up to the researcher: the problem is under-determined in the absence of further constraints, such as K_d values for other proteins, like GpA mutants.

Future work could include a more detailed examination of the protein and lipid components of the CHARMM36 force fields. An open avenue of investigation is cap-

turing the effect of dimerization-disrupting mutations in GpA at all-atom resolution. Also, other helix-helix TM dimers could be tested for the agreement of their dimerization K_d between experiment and simulation.

Chapter 6

Conclusions

6.1 Coarse-grained protein-lipid interaction

Protein-lipid interactions are important in mediating function of proteins embedded in the membrane, as well as, those peripherally associated with the lipid bilayer.

Cardiolipin binds to the mitochondrial Adenine Nucleotide Translocator (ANT) and is required for its function in the powerhouse of the cell, the mitochondrion. Similarly, phosphatidylinositide binding to sites on Kir2.2 (and related channels) modulates the channel activity. Through MD simulations, I have estimated a well-converged binding PMF, for both the ANT-Cardiolipin and the Kir2.2-PIP_2 complexes, using the coarse-grained MARTINI model. The model seems to be in qualitative agreement with experimental data, showing that certain protein mutants have a reduced binding affinity for the associated lipids. The resolution and accuracy of the MARTINI 2.1 model seems adequate to study the questions of protein-lipid binding within the membrane, including the energetic aspect.

6.2 Coarse-grained helix-helix interactions

I obtained well-converged PMFs for the association of TM helices with the coarse-grained MARTINI model. The estimate was in part enabled by using a novel collective variable describing the dimerization reaction, the inter-helical D_{RMS}. For Glycophorin A, I found the that the coarse-grained model captures the dimer state as more stable, in agreement with available experimental data. However, the dominant form is a non-native dimer with the helices arranged in an approximately parallel fashion. Additionally, an experimentally characterized mutant of Glycophorin A, which disrupts dimerization had no effect on the helix binding in the simulation. Both of these observations suggest a force field imbalance: non-specific protein-protein interactions are too strong, while detailed interactions that stabilize the native state are too weak.

This idea is consistent by the work on the Receptor Tyrosine Kinases (RTKs). Here I simulated a series of RTK TM helix dimers, with a range of different affinities measured in a TOXCAT experiment, predicting to see a large difference in computed affinities. The dimerization PMFs from simulation have similar well-depths, with a non-native bound state being the global energy minimum, as in the case of Glycophorin A. Barring the quantitative nature of the TOXCAT assay, what I found suggests that non-specific protein-protein interactions are too strong in the coarse-grained MARTINI force field.

It is unclear if the uncovered limitations are inherent to the coarse-grained model or if an improvement, via re-parametrization is possible. The absence of a native minimum in the PMFs suggests that perhaps the resolution is too coarse to capture the tight interactions mediating helix-helix packing. However, further exploration and improvement of the coarse-grained model lies outside of the scope of this thesis.

6.3 All-atom helix-helix interactions

Including atomistic detail in the dimerization simulations, I expected to see a better agreement with experiment than with coarse-grained model.

Surprisingly, an assessment of AMBER Slipids and a detailed study of CHARMM36 Glycophorin TM dimerization revealed that both force fields predict the dissociated state to be more stable than the dimerized state. This is in stark contrast to available experimental data: while the experimental measurements differ in estimates of dimerization affinity, all point to a stable dimer. Encouragingly, the all-atom model was able to capture the native bound energy basin.

I have applied a range of enhanced sampling methods to characterized the dimerization free energy surface. Another surprising feature was a high-energy barrier separating native and non-native bound states. This feature was not observed previously and may explain why the dimers appeared stable in previously published unbiased MD simulations, even during ultra-long simulations on custom hardware.

I hypothesised that the reason for this discrepancy was a too strong protein-lipid interaction in the CHARMM36 model or absence of a C_α–hydrogen bonding. Including either of these straightforward corrections improved the PMF and made it qualitatively consistent with experiment.

Using just the K_d for a single helix pair, it is impossible to decide which correction is more appropriate. Further study could focus on the Glycophorin A mutants; alternately, different helix-helix dimers could be probed, such as RTKs, as done for the coarse-grained systems. However, this would currently require considerable computing resources and lies beyond the scope of this thesis.

6.4 Membrane simulations using models of different resolution

All-atom models account for all degrees of freedom, have the ability to capture all the relevant physics and thus be most accurate. In the context of this thesis, the GXXXG interaction mediating the Glycophorin A interaction likely requires all-atom representation to capture all the detail. Coarse-grained models are expected to have more difficulty, as some likely-relevant interactions, such as the C_α–hydrogen bonding cannot be modelled explicitly. However, what has emerged in the context of this thesis is that all-atom models have limitations that may be difficult to identify. My simulations demonstrated a force field deficiency, uncovered using enhanced sampling methodology. First of all, this suggests that further such deficiencies likely exist and are difficult to identify for similar reasons. Secondly, improving on those deficiencies remains challenging: where does one start to improve when all the used parametrization model systems show the force field in agreement with experiment? Are further force-field tweaks needed on the protein-water and lipid-water interaction description?

Improving all-atom force fields is challenging due to the long iteration time per refinement, related to the cost of running test simulations. Coarse-grained force fields are much better in this regard. A larger number of test systems can be run faster, increasing the force field development "velocity". Indeed, in this thesis, only the coarse-grained MARTINI model correctly predicts that Glycophorin A is a stable dimer. Perhaps it is possible to capture all the relevant physics using a coarse-grained model, for helix-helix interactions in the membrane? Indeed, much coarser force fields than the MARTINI model have been used to explore the physics of protein oligomerization in membranes via dissipative particle dynamics (DPD) simulations [8]. While the coarse-grained MARTINI model seems promising, it should also be noted that it predicted essentially the same stability for a non-dimerizing Glycophorin A mutant.

References

[1] M. P. Aliste and D. P. Tieleman. "Computer simulation of partitioning of ten pentapeptides Ace-WLXLL at the cyclohexane/water and phospholipid/water interfaces". In: *BMC Biochem.* 6 (Dec. 2005), p. 30.

[2] M. S. Almen et al. "Mapping the human membrane proteome: a majority of the human membrane proteins can be classified according to function and evolutionary origin". In: *BMC Biol.* 7 (Aug. 2009), p. 50.

[3] V. Anbazhagan and D. Schneider. "The membrane environment modulates self-association of the human GpA TM domain-Implications for membrane protein folding and transmembrane signaling". In: *BBA Biomemb.* 1798.10 (2010), pp. 1899–1907.

[4] A. Arkhipov et al. "Architecture and membrane interactions of the EGF receptor". In: *Cell* 152 (2013), pp. 557–569.

[5] C. Arnarez, S. J. Marrink, and X. Periole. "Identification of cardiolipin binding sites on cytochrome c oxidase at the entrance of proton channels". In: *Sci. Rep.* 3 (2013), p. 1263.

[6] E. O. Artemenko et al. "Transmembrane domain of EphA1 receptor forms dimers in membrane-like environment". In: *Biochim. Biophys. Acta* 1778 (2008), pp. 2361–2367.

[7] A. Barducci, G. Bussi, and M. Parrinello. "Well-tempered metadynamics: a smoothly converging and tunable free-energy method". In: *Phys. Rev. Lett.* 100.2 (Jan. 2008), p. 020603.

[8] A. Benjamini and B. Smit. "Lipid mediated packing of transmembrane helices - a dissipative particle dynamics study". In: *Soft Matter* 9.9 (2013), pp. 2673–2683.

[9] N. Ben-Tal et al. "Free Energy of Amide Hydrogen Bond Formation in Vacuum, in Water, and in Liquid Alkane Solution". In: *J. Phys. Chem. B* 101.3 (1997), pp. 450–457.

[10] M. J. Berardi et al. "Mitochondrial uncoupling protein 2 structure determined by NMR molecular fragment searching". In: *Nature* 476.7358 (July 2011), pp. 109–113.

[11] H. J. C. Berendsen et al. "Molecular dynamics with coupling to an external bath". In: *J. Chem. Phys.* 81 (1984), pp. 3684–3690.

[12] J. D. Bernal. "The Bakerian Lecture, 1962. The Structure of Liquids". In: *Proc. R. Soc. Lon. Ser. A Math. Phys. Eng. Sci.* 280.1382 (1964), pp. 299–322.

[13] H. Bertil. *Ion Channels of Excitable Membranes*. Sinauer Associates is an imprint of Oxford University Press, 2001.

[14] R. B. Best and G. Hummer. "Coordinate-dependent diffusion in protein folding". In: *Proc. Natl. Acad. Sci. USA* 107.3 (Jan. 2010), pp. 1088–1093.

[15] Klaus Beyer and Martin Klingenberg. "ADP/ATP carrier protein from beef heart mitochondria has high amounts of tightly bound cardiolipin, as revealed by phosphorus-31 nuclear magnetic resonance". In: *Biochemistry* 24.15 (1985). PMID: 2996583, pp. 3821–3826.

[16] E. V. Bocharov et al. “Left-handed dimer of EphA2 transmembrane domain: helix packing diversity among receptor tyrosine kinases”. In: *Biophys. J.* 98.5 (2010), pp. 881–889.

[17] E. V. Bocharov et al. “Spatial structure and pH-dependent conformational diversity of dimeric transmembrane domain of the receptor tyrosine kinase EphA1”. In: *J. Biol. Chem.* 283.43 (2008), pp. 29385–29395.

[18] G. Bussi, D. Donadio, and M. Parrinello. “Canonical sampling through velocity rescaling”. In: *J. Chem. Phys.* 126.1 (2007), p. 014101.

[19] G. Bussi, D. Donadio, and M. Parrinello. “Canonical sampling through velocity rescaling”. In: *J. Chem. Phys.* 126.1 (Jan. 2007), p. 014101.

[20] P. Cesare et al. “Ion channels gated by heat”. In: *Proc. Natl. Acad. Sci. USA* 96.14 (July 1999), pp. 7658–7663.

[21] M. Chabre, R. Cone, and H. Saibil. “Biophysics: is rhodopsin dimeric in native retinal rods?” In: *Nature* 426.6962 (Nov. 2003), pp. 30–31.

[22] L. Chen et al. “Measuring the Energetics of Membrane Protein Dimerization in Mammalian Membranes”. In: *J. Am. Chem. Soc.* 132.10 (2010). PMID: 20158179, pp. 3628–3635.

[23] W. W. Cheng, D. Enkvetchakul, and C. G. Nichols. “KirBac1.1: it’s an inward rectifying potassium channel”. In: *J. Gen. Physiol.* 133.3 (2009), pp. 295–305.

[24] E. Chiavazzo et al. “Intrinsic map dynamics exploration for uncharted effective free-energy landscapes”. In: *Proc. Natl. Acad. Sci. USA* 114.28 (July 2017), E5494–E5503.

[25] S. M. Claypool et al. “Cardiolipin defines the interactome of the major ADP/ATP carrier protein of the mitochondrial inner membrane”. In: *J. Cell Biol.* 182.5 (2008), pp. 937–950.

[26] R. A. Cone. “Rotational diffusion of rhodopsin in the visual receptor membrane”. In: *Nature New Biol.* 236.63 (Mar. 1972), pp. 39–43.

[27] W. D. Cornell et al. “A second generation force field for the simulation of proteins and nucleic acids”. In: *J. Am. Chem. Soc.* 117 (1995), pp. 5179–5197.

[28] Ü. Coskun et al. “Regulation of human EGF receptor by lipids”. In: *Proc. Natl. Acad. Sci. USA* 108 (2011), pp. 9044–9048.

[29] R. Covino et al. “A Eukaryotic Sensor for Membrane Lipid Saturation”. In: *Mol. Cell* 63.1 (July 2016), pp. 49–59.

[30] C. A. Cukras, I. Jeliazkova, and C. G. Nichols. “Structural and functional determinants of conserved lipid interaction domains of inward rectifying Kir6.2 channels”. In: *J. Gen. Physiol.* 119.6 (2002), pp. 581–591.

[31] C. A. Cukras, I. Jeliazkova, and C. G. Nichols. “The role of NH2-terminal positive charges in the activity of inward rectifier K-ATP channels”. In: *J. Gen. Physiol.* 120.3 (2002), pp. 437–446.

[32] J. M. Cuthbertson, P. J. Bond, and M. S. P. Sansom. “Transmembrane helix-helix interactions: comparative simulations of the glycophorin A dimer”. In: *Biochemistry* 45 (2006), pp. 14298–14310.

[33] F. Cymer, G. von Heijne, and S. H. White. “Mechanisms of integral membrane protein insertion and folding”. In: *J. Mol. Biol.* 427.5 (Mar. 2015), pp. 999–1022.

[34] F. Cymer, A. Veerappan, and D. Schneider. “Transmembrane helix-helix interactions are modulated by the sequence context and by lipid bilayer properties”. In: *Biochim. Biophys. Acta* 1818 (2012), pp. 963–973.

[35] M. Dahlberg and A. Maliniak. “Mechanical Properties of Coarse-Grained Bilayers Formed by Cardiolipin and Zwitterionic Lipids”. In: *J. Chem. Theory Comput.* 6.5 (May 2010), pp. 1638–1649.

[36] R. E. Dalbey and C. Robinson. “Protein translocation into and across the bacterial plasma membrane and the plant thylakoid membrane”. In: *Trends Biochem. Sci.* 24.1 (Jan. 1999), pp. 17–22.

[37] T. Darden, D. York, and L. Pedersen. “Particle mesh Ewald: An N log(N) method for Ewald sums in large systems”. In: *J. Chem. Phys.* 98.12 (1993), pp. 10089–10092.

[38] R. Dawaliby et al. “Allosteric regulation of G protein-coupled receptor activity by phospholipids”. In: *Nature Chem. Biol.* 12 (2015), pp. 35–39.

[39] D. H. De Jong et al. “Determining equilibrium constants for dimerization reactions from molecular dynamics simulations”. In: *J. Comput. Chem.* 32.9 (July 2011), pp. 1919–1928.

[40] J. Domański et al. “Balancing Force Field Protein-Lipid Interactions To Capture Transmembrane Helix-Helix Association”. In: *J. Chem. Theory Comput.* (Feb. 2018).

[41] J. Domański et al. “Convergence and Sampling in Determining Free Energy Landscapes for Membrane Protein Association”. In: *J. Phys. Chem. B* 121.15 (2017). PMID: 27807980, pp. 3364–3375.

[42] C. A. Doupnik, N. Davidson, and H. A. Lester. “The inward rectifier potassium channel family”. In: *Curr. Opin. Neurobiol.* 5 (1995), pp. 268–277.

[43] A. L. Duncan, A. J. Robinson, and J. E. Walker. “Cardiolipin binds selectively but transiently to conserved lysine residues in the rotor of metazoan ATP synthases”. In: *Proc. Natl. Acad. Sci. USA* 113.31 (Aug. 2016), pp. 8687–8692.

[44] M. T. Duong et al. “Changes in apparent free energy of helix-helix dimerization in a biological membrane due to point mutations”. In: *J. Mol. Biol.* 371.2 (2007), pp. 422–434.

[45] A. Z. Ebie and K. G. Fleming. "Dimerization of the erythropoietin receptor transmembrane domain in micelles". In: *J. Mol. Biol.* 366.2 (2007), pp. 517–524.

[46] N. F. Endres et al. "Conformational coupling across the plasma membrane in activation of the EGF receptor". In: *Cell* 152.3 (2013), pp. 543–556.

[47] D. Enkvetchakul et al. "Functional characterization of a prokaryotic Kir channel". In: *J. Biol. Chem.* 279.45 (2004), pp. 47076–47080.

[48] U. Essmann et al. "A smooth particle mesh Ewald method". In: *J. Chem. Phys.* 103.19 (1995), pp. 8577–8593.

[49] P. P. Ewald. "Die Berechnung optischer und elektrostatischer Gitterpotentiale". In: *Ann. Phys. (Berl.)* 369.3 (1921), pp. 253–287.

[50] L. Fagerberg et al. "Prediction of the human membrane proteome". In: *Proteomics* 10.6 (Mar. 2010), pp. 1141–1149.

[51] S. Faham et al. "Side-chain contributions to membrane protein structure and stability". In: *J. Mol. Biol.* 335 (2004), pp. 297–305.

[52] Z. Fan and J. C. Makielski. "Anionic phospholipids activate ATP-sensitive potassium channels". In: *J. Biol. Chem.* 272.9 (1997), pp. 5388–5395.

[53] Z. Fan and J. C. Makielski. "Phosphoinositides decrease ATP sensitivity of the cardiac ATP-sensitive K(+) channel. A molecular probe for the mechanism of ATP-sensitive inhibition". In: *J. Gen. Physiol.* 114.2 (1999), pp. 251–69.

[54] A. Filippov, G. Orädd, and G. Lindblom. "The effect of cholesterol on the lateral diffusion of phospholipids in oriented bilayers". In: *Biophys. J.* 84 (2003), pp. 3079–3086.

[55] C. Finger, C. Escher, and D. Schneider. “The single transmembrane domain of human tyrosine kinases encode self interactions”. In: *Science Signal.* 2 (2009), ra56, 1–8.

[56] L. E. Fisher, D. M. Engelman, and J. N. Sturgis. “Detergents modulate dimerization, but not helicity, of the glycophorin A transmembrane domain”. In: *J. Mol. Biol.* 293 (1999), pp. 639–651.

[57] L. E. Fisher, D. M. Engelman, and J. N. Sturgis. “Effect of detergents on the association of the glycophorin a transmembrane helix”. In: *Biophys. J.* 85 (2003), pp. 3097–3105.

[58] M. Fitzner et al. “Communication: Truncated non-bonded potentials can yield unphysical behavior in molecular dynamics simulations of interfaces”. In: *J. Chem. Phys.* 147.12 (Sept. 2017), p. 121102.

[59] K. G. Fleming. “Standardizing the free energy change of transmembrane helix-helix interactions”. In: *J. Mol. Biol.* 323.3 (2002), pp. 563–571.

[60] K. G. Fleming, A. L. Ackerman, and D. M. Engelman. “The effect of point mutations on the free energy of transmembrane alpha-helix dimerization”. In: *J. Mol. Biol.* 272.2 (1997), pp. 266–275.

[61] K. G. Fleming and D. M. Engelman. “Specificity in transmembrane helix-helix interactions can define a hierarchy of stability for sequence variants”. In: *Proc. Natl. Acad. Sci. USA* 98.25 (2001), pp. 14340–14344.

[62] D. Fraser-Pitt and D. O’Neil. “Cystic fibrosis - a multiorgan protein misfolding disease”. In: *Future Sci. OA* 1.2 (Sept. 2015), FSO57.

[63] D. Frenkel and B. Smit. *Understanding Molecular Simulation: From Algorithms to Applications (Computational Science Series, Vol 1).* Academic Press, 2001.

[64] R. D. Gautam. “C-H...O and other weak hydrogen bonds. From crystal engineering to virtual screening”. In: *Chem. Commun.* (2005), pp. 2995–3001.

[65] K. A. Gromek et al. “The oligomeric states of the purified sigma-1 receptor are stabilized by ligands”. In: *J. Biol. Chem.* 289.29 (July 2014), pp. 20333–20344.

[66] W. F. van Gunsteren et al. *Biomolecular Simulation: The GROMOS96 Manual and User Guide*. Groningen and Zurich: Biomos and Hochschulverlag AG an der ETH Zurich, 1996.

[67] A. Haghighi et al. “Sengers syndrome: six novel AGK mutations in seven new families and review of the phenotypic and mutational spectrum of 29 patients”. In: *Otphanet J. Rare Dis.* 9.1 (Aug. 2014), p. 119.

[68] S. B. Hansen. “Lipid agonism: The PIP2 paradigm of ligand-gated ion channels”. In: *Biochim. Biophys. Acta* 1851.5 (2015), pp. 620–628.

[69] S. B. Hansen, X. Tao, and R. Mackinnon. “Structural basis of PIP2 activation of the classical inward rectifier K+ channel Kir2.2”. In: *Nature* 477 (2011), pp. 495–498.

[70] L. J. He and K. Hristova. “Physical-chemical principles underlying RTK activation, and their implications for human disease”. In: *Biochim. Biophys. Acta* 1818.4 (2012), pp. 995–1005.

[71] G. Hedger and M. S. P. Sansom. “Lipid interaction sites on channels, transporters and receptors: Recent insights from molecular dynamics simulations”. In: *Biochim. Biophys. Acta* 1858.10 (Oct. 2016), pp. 2390–2400.

[72] G. Hedger et al. “Free Energy Landscape of Lipid Interactions with Regulatory Binding Sites on the Transmembrane Domain of the EGF Receptor”. In: *J. Phys. Chem. B* 120.33 (Aug. 2016), pp. 8154–8163.

[73] G. Hedger et al. “Lipid loving ANTs: cardiolipin interactions and the organization of the adenine nucleotide translocase in model mitochondrial membranes”. In: *Biochemistry* 55.45 (Nov. 2016), pp. 6238–6249.

[74] L. Heginbotham, L. Kolmakova-Partensky, and C. Miller. “Functional Reconstitution of a prokaryotic K+ channel.” In: *J. Gen. Physiol.* 111 (1998), pp. 741–749.

[75] J. Hénin, A. Pohorille, and C. Chipot. “Insights into the recognition and association of transmembrane α-helices. The free energy of α-helix dimerization in Glycophorin A”. In: *J. Am. Chem. Soc.* 127 (2005), pp. 8478–8484.

[76] B. Hess et al. “LINCS: a linear constraint solver for molecular simulations”. In: *J. Comput. Chem.* 18.12 (1997), pp. 1463–1472.

[77] T. Hessa et al. “Molecular code for transmembrane-helix recognition by the Sec61 translocon”. In: *Nature* 450.7172 (2007), 1026–U2.

[78] T. Hessa et al. “Recognition of transmembrane helices by the endoplasmic reticulum translocon”. In: *Nature* 433 (2005), pp. 377–381.

[79] R. W. Hockney. “The potential calculation and some applications”. In: *J. Comput. Phys.* 9 (1970), pp. 135–211.

[80] H. Hong, Y. C. Chang, and J. U. Bowie. “Measuring transmembrane helix interaction strengths in lipid bilayers using steric trapping”. In: *Methods Mol. Biol.* 1063 (2013), pp. 37–56.

[81] H. Hong et al. “Method to measure strong protein-protein interactions in lipid bilayers using a steric trap”. In: *Proc. Natl. Acad. Sci. USA* 107.46 (Nov. 2010), pp. 19802–19807.

[82] A. L. Hopkins and C. R. Groom. “The druggable genome”. In: *Nat. Rev. Drug Discov.* 1.9 (Sept. 2002), pp. 727–730.

[83] J. Huang and A. D. MacKerell. "CHARMM36 all-atom additive protein force field: validation based on comparison to NMR data". In: *J. Comput. Chem.* 34.25 (Sept. 2013), pp. 2135–2145.

[84] J. Huang et al. "CHARMM36m: an improved force field for folded and intrinsically disordered proteins". In: *Nat. Methods* 14.1 (Jan. 2017), pp. 71–73.

[85] J. S. Hub, B. L. de Groot, and D. van der Spoel. "gwham A free weighted histogram analysis implementation Including robust error and autocorrelation estimates". In: *J. Chem. Theor. Comput.* 6.12 (2010), pp. 3713–3720.

[86] W. Humphrey, A. Dalke, and K. Schulten. "VMD: visual molecular dynamics". In: *J. Mol. Graph.* 14.1 (Feb. 1996), pp. 33–38.

[87] S. Inagaki et al. "Modulation of the interaction between neurotensin receptor NTS1 and Gq protein by lipid". In: *J. Mol. Biol.* 417.1–2 (2012), pp. 95–111.

[88] J. P. Jambeck and A. P. Lyubartsev. "An Extension and Further Validation of an All-Atomistic Force Field for Biological Membranes". In: *J. Chem. Theory Comput.* 8.8 (Aug. 2012), pp. 2938–2948.

[89] J. P. Jambeck and A. P. Lyubartsev. "Derivation and systematic validation of a refined all-atom force field for phosphatidylcholine lipids". In: *J. Phys. Chem. B* 116.10 (Mar. 2012), pp. 3164–3179.

[90] L. Janosi, A. Prakash, and M. Doxastakis. "Lipid-modulated sequence-specific association of glycophorin A in membranes". In: *Biophys. J.* 99.1 (2010), pp. 284–292.

[91] S. A. John, J. N. Weiss, and B. Ribalet. "ATP sensitivity of ATP-sensitive K+ channels: role of the gamma phosphate group of ATP and the R50 residue of mouse Kir6.2". In: *J. Physiol.* 568.3 (2005), pp. 931–940.

[92] D. H. de Jong et al. "Improved parameters for the Martini coarse-grained protein force field". In: *J. Chem. Theor. Comput.* 9.1 (2013), pp. 687–697.

[93] W. L. Jorgensen, D. S. Maxwell, and J. Tirado-Rives. "Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids". In: *J. Amer. Chem. Soc.* 118 (1996), pp. 11225–11236.

[94] J. B. Klauda et al. "Update of the CHARMM all-atom additive force field for lipids: validation on six lipid types". In: *J. Phys. Chem. B* 114.23 (June 2010), pp. 7830–7843.

[95] M. Klingenberg. "Cardiolipin and mitochondrial carriers". In: *Biochim. Biophys. Acta* 1788.10 (2009), pp. 2048–2058.

[96] M. Klingenberg. "The ADP and ATP transport in mitochondria and its carrier". In: *BBA Biomemb.* 1778.10 (2008), pp. 1978–2021.

[97] J. R. Knowles. "Enzyme-catalyzed phosphoryl transfer reactions". In: *Annu. Rev. Biochem.* 49 (1980), pp. 877–919.

[98] E. E. Kooijman et al. "Ionization properties of phosphatidylinositol polyphosphates in mixed model membranes". In: *Biochemistry* 48 (2009), pp. 9360–9371.

[99] N. Kucerka, M. P. Nieh, and J. Katsaras. "Fluid phase lipid areas and bilayer thicknesses of commonly used phosphatidylcholines as a function of temperature". In: *Biochim. Biophys. Acta* 1808.11 (Nov. 2011), pp. 2761–2771.

[100] S. Kumar et al. "The weighted histogram analysis method for free-energy calculations on biomolecules. I. The method". In: *J. Comput. Chem.* 13.8 (1992), pp. 1011–1021.

[101] E. R. S. Kunji et al. "The transport mechanism of the mitochondrial ADP/ATP carrier". In: *BBA - Mol. Cell Res.* 1863.10 (2016). Channels and transporters in cell metabolism, pp. 2379–2393.

[102] A. Kuo et al. "Crystal structure of the potassium channel KirBac1.1 in the closed state". In: *Science* 300 (2003), pp. 1922–1926.

[103] A. S. Kuznetsov et al. “Adaptable lipid matrix promotes protein-protein association in membranes”. In: *J. Chem. Theor. Comput.* 11.9 (2015), pp. 4415–4426.

[104] A. Laio and M. Parrinello. “Escaping free-energy minima”. In: *Proc. Natl. Acad. Sci. USA* 99.20 (Oct. 2002), pp. 12562–12566.

[105] D. Langosch and I.T. Arkin. “Interaction and conformational dynamics of membrane-spanning protein helices”. In: *Prot. Sci.* 18 (2009), pp. 1343–1358.

[106] A. R. Leach. *Molecular modelling. Principles and applications*. 1st ed. Longman, 1996.

[107] M. A. Lemmon et al. “A dimerization motif for transmembrane alpha-helices”. In: *Nat. Struct. Biol.* 1.3 (Mar. 1994), pp. 157–163.

[108] M. A. Lemmon et al. “Sequence specificity in the dimerisation of transmembrane a-helices.” In: *Biochemistry* 31 (1992), pp. 12719–12725.

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

[110] P. C. Li et al. “Multidimensional umbrella sampling and replica-exchange molecular dynamics simulations for structure prediction of transmembrane helix dimers”. In: *J. Comput. Chem.* 35.4 (Feb. 2014), pp. 300–308.

[111] P. A. Liebman and G. Entine. “Lateral diffusion of visual pigment in photorecptor disk membranes”. In: *Science* 185.4149 (Aug. 1974), pp. 457–459.

[112] K. Lindorff-Larsen et al. “How fast-folding proteins fold”. In: *Science* 334.6055 (Oct. 2011), pp. 517–520.

[113] P. Liu et al. “Replica exchange with solute tempering: a method for sampling biological systems in explicit water”. In: *Proc. Natl. Acad. Sci. USA* 102.39 (Sept. 2005), pp. 13749–13754.

[114] D. E. Logothetis et al. “Phosphoinositide-mediated gating of inwardly rectifying K(+) channels”. In: *Pflugers Arch.* 455.1 (2007), pp. 83–95.

[115] C. M. B. Lopes et al. “Alterations in conserved Kir channel-PIP2 interactions underlie channelopathies”. In: *Neuron* 34.6 (2002), pp. 933–944.

[116] C. A. Lopez et al. “Martini coarse-grained force field: extension to carbohydrates”. In: *J. Chem. Theor. Comput.* 5.12 (2009), pp. 3195–3210.

[117] J. L. Lorieau, J. M. Louis, and A. Bax. “The complete influenza hemagglutinin fusion domain adopts a tight helical hairpin arrangement at the lipid-water interface”. In: *Proc. Natl. Acad. Sci. USA* 107 (2010), pp. 11341–11346.

[118] J. L. MacCallum, W. F. Bennett, and D. P. Tieleman. “Distribution of amino acids in a lipid bilayer from computer simulations”. In: *Biophys. J.* 94.9 (May 2008), pp. 3393–3404.

[119] J. L. MacCallum, W. F. Bennett, and D. P. Tieleman. “Partitioning of amino acid side chains into lipid bilayers: results from computer simulations and comparison to experiment”. In: *J. Gen. Physiol.* 129.5 (May 2007), pp. 371–377.

[120] C. A. Mack. “Fifty Years of Moore’s Law”. In: *IEEE Trans. Semicond. Manuf.* 24.2 (May 2011), pp. 202–207.

[121] K. R. MacKenzie and K. G. Fleming. “Association energetics of membrane spanning alpha-helices”. In: *Curr. Opin. Struct. Biol.* 18.4 (2008), pp. 412–419.

[122] K. R. MacKenzie, J. H. Prestegard, and D. M. Engelman. “A transmembrane helix dimer: structure and implications.” In: *Science* 276 (1997), pp. 131–133.

[123] A. D. MacKerell et al. “All-atom empirical potential for molecular modeling and dynamics studies of proteins”. In: *J. Phys. Chem. B* 102.18 (Apr. 1998), pp. 3586–3616.

[124] S. J. Marrink, A. H. de Vries, and A. E. Mark. “Coarse grained model for semiquantitative lipid simulations”. In: *J. Phys. Chem. B* 108 (2004), pp. 750–760.

[125] S. J. Marrink et al. “The MARTINI force field: coarse grained model for biomolecular simulations”. In: *J. Phys. Chem. B.* 111 (2007), pp. 7812–7824.

[126] Y. Matsunaga et al. “Dimensionality of Collective Variables for Describing Conformational Changes of a Multi-Domain Protein”. In: *J. Phys. Chem. Lett.* 7.8 (Apr. 2016), pp. 1446–1451.

[127] E. E. Matthews, M. Zoonens, and D. M. Engelman. “Dynamic helix interactions in transmembrane signaling”. In: *Cell* 127 (2006), pp. 447–450.

[128] D. C. Mayer et al. “Glycophorin B is the erythrocyte receptor of Plasmodium falciparum erythrocyte-binding ligand, EBL-1”. In: *Proc. Natl. Acad. Sci. U.S.A.* 106.13 (Mar. 2009), pp. 5348–5352.

[129] J. A. Mayr et al. “Lack of the Mitochondrial Protein Acylglycerol Kinase Causes Sengers Syndrome”. In: *Am. J. Hum. Genet.* 90.2 (2012), pp. 314–320.

[130] J.A. McCammon, B.R. Gelin, and M. Karplus. “Dynamics of folded proteins”. In: *Nature* 267 (1977), 585–???

[131] S. McLaughlin et al. “PIP2 and proteins: interactions, organization, and information flow”. In: *Annu. Rev. Biophys. Biomol. Struct.* 31 (2002), pp. 151–175.

[132] G. van Meer, D. R. Voelker, and G. W. Feigenson. “Membrane lipids: where they are and how they behave”. In: *Nat. Rev. Mol. Cell Biol.* 9.2 (Feb. 2008), pp. 112–124.

[133] I. E. Michailidis et al. “Phosphatidylinositol-4,5-bisphosphate regulates epidermal growth factor receptor activation”. In: *Pflugers Arch.* 461 (2011), pp. 387–397.

[134] N. Michaud-Agrawal et al. “MDAnalysis: a toolkit for the analysis of molecular dynamics simulations”. In: *J. Comput. Chem.* 32.10 (July 2011), pp. 2319–2327.

[135] S. Miyazawa and R. L. Jernigan. “Self-consistent estimation of inter-residue protein contact energies based on an equilibrium mixture approximation of residues”. In: *Proteins* 34.1 (Jan. 1999), pp. 49–68.

[136] L. Monticelli et al. “The MARTINI coarse grained force field: extension to proteins”. In: *J. Chem. Theor. Comp.* 4 (2008), pp. 819–834.

[137] C. P. Moon and K. G. Fleming. “Side-chain hydrophobicity scale derived from transmembrane protein folding into lipid bilayers”. In: *Proc. Natl. Acad. Sci. USA* 108.25 (2011), pp. 10174–10177.

[138] G. E. Moore. “Cramming more components onto integrated circuits”. In: *Electronics* 38.8 (Apr. 1965).

[139] B. K. Mueller, S. Subramaniam, and A. Senes. “A frequent, GXXXG-mediated, transmembrane association motif is optimized for the formation of interhelical Cα-H hydrogen bonds”. In: *Proc. Natl. Acad. Sci. U. S. A.* 111 (2014), E888–E895.

[140] F. B. Naughton, A. C. Kalli, and M. S. P. Sansom. “Association of peripheral membrane proteins with membranes: Free energy of binding of GRP1 PH domain with PIP-containing model bilayers”. In: *J. Phys. Chem. Lett.* 7 (2016), pp. 1219–1224.

[141] E. Neria, S. Fischer, and M. Karplus. “Simulation of activation free energies in molecular systems”. In: *J. Chem. Phys.* 105.5 (1996), pp. 1902–1921.

[142] C. G. Nichols and A. N. Lopatin. “Inward rectifier potassium channels.” In: *Annu. Rev. Physiol.* 59 (1997), pp. 171–191.

[143] M. Nishida and R. MacKinnon. “Structural basis of inward rectification: cytoplasmic pore of the G protein-gated inward rectifier GIRK1 at 1.8 Å resolution”. In: *Cell* 111.7 (2002), pp. 957–65.

[144] M. Nishida et al. “Crystal structure of a Kir3.1-prokaryotic Kir channel chimera”. In: *EMBO J.* 26 (2007), pp. 4005–4015.

[145] H. Nury et al. “Structural basis for lipid-mediated interactions between mitochondrial ADP/ATP carrier monomers”. In: *FEBS Lett.* 579 (2005), pp. 6031–6036.

[146] F. Palmieri. “Mitochondrial transporters of the SLC25 family and associated diseases: a review”. In: *J. Inherit. Metab. Dis.* 37.4 (July 2014), pp. 565–575.

[147] M. Parrinello and A. Rahman. “Crystal Structure and Pair Potentials: A Molecular-Dynamics Study”. In: *Phys. Rev. Lett.* 45 (14 Oct. 1980), pp. 1196–1199.

[148] M. Parrinello and A. Rahman. “Polymorphic transitions in single crystals: A new molecular dynamics method”. In: *J. Appl. Phys.* 52.12 (1981), pp. 7182–7190.

[149] R. W. Pastor and A. D. Mackerell. “Development of the CHARMM Force Field for Lipids”. In: *J. Phys. Chem. Lett.* 2.13 (2011), pp. 1526–1532.

[150] M. Patra et al. “Molecular dynamics simulations of lipid bilayers: major artifacts due to truncating electrostatic interactions”. In: *Biophys. J.* 84.6 (June 2003), pp. 3636–3645.

[151] E. Pebay-Peyroula et al. “Structure of mitochondrial ADP/ATP carrier in complex with carboxyatractyloside”. In: *Nature* 426 (2003), pp. 39–44.

[152] F. A. Perras et al. “Observation of CH-pi Interactions between Methyl and Carbonyl Groups in Proteins”. In: *Angew. Chem. Int. Ed. Engl.* (May 2017).

[153] H. I. Petrache et al. “Modulation of glycophorin A transmembrane helix interactions by lipid bilayers: Molecular dynamics calculations”. In: *J. Mol. Biol.* 302 (2000), pp. 727–746.

[154] J. C. Phillips et al. “Scalable molecular dynamics with NAMD”. In: *J. Comput. Chem.* 26.16 (Dec. 2005), pp. 1781–1802.

[155] S. Piana, K. Lindorff-Larsen, and D. E. Shaw. “How robust are protein folding simulations with respect to force field parameterization?” In: *Biophys. J.* 100.9 (May 2011), pp. L47–49.

[156] K. Pluhackova et al. “A Critical Comparison of Biomembrane Force Fields: Structure and Dynamics of Model DMPC, POPC, and POPE Bilayers”. In: *J. Phys. Chem. B* 120.16 (Apr. 2016), pp. 3888–3903.

[157] T. V. Pogorelov et al. “Partitioning of amino acids into a model membrane: capturing the interface”. In: *J. Phys. Chem. B* 118.6 (Feb. 2014), pp. 1481–1492.

[158] J. L. Popot and D. M. Engelman. “Membrane protein folding and oligomerization: the two-stage model”. In: *Biochemistry* 29.17 (May 1990), pp. 4031–4037.

[159] J. L. Popot and D. M. Engelman. “Membranes Do Not Tell Proteins How To Fold”. In: *Biochemistry* 55.1 (Jan. 2016), pp. 5–18.

[160] S. Pronk et al. “GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit”. In: *Bioinformatics* 29.7 (Apr. 2013), pp. 845–854.

[161] E. Psachoulia et al. “Helix-helix interactions in membrane proteins: coarse grained simulations of glycophorin helix dimerization”. In: *Biochemistry* 47 (2008), pp. 10503–105012.

[162] A. Radzicka and R. Wolfenden. “Comparing the polarities of the amino acids: side-chain distribution coefficients between the vapor phase, cyclohexane, 1-octanol, and neutral aqueous solution”. In: *Biochemistry* 27 (1988), pp. 1664–1670.

[163] M. E. Reid. “MNS blood group system: a review”. In: *Immunohematology* 25.3 (2009), pp. 95–101.

[164] F. Reimann and F. M. Ashcroft. “Inwardly rectifying potassium channels”. In: *Curr. Opin. Cell Biol.* 11 (1999), pp. 503–508.

[165] S. Reisser et al. “Real Cost of Speed: The Effect of a Time-Saving Multiple-Time-Stepping Algorithm on the Accuracy of Molecular Dynamics Simulations”. In: *J. Chem. Theory Comput.* 13.6 (June 2017), pp. 2367–2372.

[166] B. Ribalet et al. “ATP-sensitive K+ channels: regulation of bursting by the sulphonylurea receptor, PIP2 and regions of Kir6.2”. In: *J. Physiol.* 571.2 (2006), pp. 303–317.

[167] T. Rohacs et al. “Specificity of activation by phosphoinositides determines lipid regulation of Kir channels”. In: *Proc. Natl. Acad. Sci. USA* 100.2 (Jan. 2003), pp. 745–750.

[168] W. P. Russ and D. M. Engelman. “The GXXXG motif: a framework for transmembrane helix-helix association”. In: *J. Mol. Biol.* 296.3 (Feb. 2000), pp. 911–919.

[169] W. P. Russ and D. M. Engelman. “TOXCAT: a measure of transmembrane helix association in a biological membrane”. In: *Proc. Natl. Acad. Sci. USA* 96 (1999), pp. 863–868.

[170] J. P. Ryckaert, G. Ciccotti, and H. J. C. Berendsen. "Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes". In: *J. Comput. Phys.* 23.3 (1977), pp. 327–341.

[171] S. Sarabipour and K. Hristova. "FGFR3 transmembrane domain interactions persist in the presence of its extracellular domain". In: *Biophys. J.* 105.1 (2013), pp. 165–171.

[172] S. Sarabipour and K. Hristova. "Glycophorin A transmembrane domain dimerization in plasma membrane vesicles derived from CHO, HEK 293T, and A431 cells". In: *Biochim. Biophys. Acta* 1828.8 (Aug. 2013), pp. 1829–1833.

[173] S. Scheiner, T. Kar, and Y. Gu. "Strength of the $C_\alpha H..O$ hydrogen bond of amino acid residues". In: *J. Biol. Chem.* (2001), pp. 9832–9837.

[174] M. Schlame and M. Ren. "The role of cardiolipin in the structural organization of mitochondrial membranes". In: *Biochim. Biophys. Acta* 1788.10 (Oct. 2009), pp. 2080–2083.

[175] M. R. Schmidt et al. "Simulation-based prediction of phosphatidylinositol 4,5-bisphosphate binding to an ion channel". In: *Biochemistry* 52.2 (2013), pp. 279–281.

[176] D. Schulze et al. "Phosphatidylinositol 4,5-bisphosphate (PIP2) modulation of ATP and pH sensitivity in Kir channels. A tale of an active and a silent PIP2 site in the N terminus". In: *J. Biol. Chem.* 278.12 (2003), pp. 10500–5.

[177] J. Seelig and N. Waespe-Sarcevic. "Molecular order in cis and trans unsaturated phospholipid bilayers". In: *Biochemistry* 17.16 (Aug. 1978), pp. 3310–3315.

[178] M. M. Seibert et al. "Reproducible polypeptide folding and structure prediction using molecular dynamics simulations". In: *J. Mol. Biol.* 354.1 (Nov. 2005), pp. 173–183.

[179] D. Sengupta and S. J. Marrink. "Lipid-mediated interactions tune the association of glycophorin A helix and its disruptive mutants in membranes". In: *Phys. Chem. Chem. Phys.* 12.40 (2010), pp. 12987–12996.

[180] G. V. Sharonov et al. "Point mutations in dimerization motifs of transmembrane domain stabilize active or inactive state of the EphA2 receptor tyrosine kinase". In: *J. Biol. Chem.* 289 (2014), pp. 14955–14964.

[181] D. E. Shaw et al. "Anton, a Special-purpose Machine for Molecular Dynamics Simulation". In: *Commun. ACM* 51.7 (July 2008), pp. 91–97.

[182] M. R. Shirts and J. D. Chodera. "Statistically optimal analysis of samples from multiple equilibrium states". In: *J. Chem. Phys.* 129.12 (Sept. 2008), p. 124105.

[183] S. L. Shyng et al. "Structural determinants of PIP(2) regulation of inward rectifier K(ATP) channels". In: *J. Gen. Physiol.* 116.5 (2000), pp. 599–608.

[184] P. Siekevitz. "Powerhouse of the Cell". In: *Scientific American* 197.1 (1957), pp. 131–144.

[185] S. O. Smith et al. "Implications of threonine hydrogen bonding in the glycophorin A transmembrane helix dimer". In: *Biophys. J.* 82.5 (May 2002), pp. 2476–2486.

[186] M. Souaille and B. Roux. "Extension to the weighted histogram analysis method: combining umbrella sampling with free energy calculations". In: *Comput. Phys. Commun.* 135.1 (2001), pp. 40–57.

[187] A. M. Stanley et al. "Energetics of outer membrane phospholipase A (OMPLA) dimerization". In: *J. Mol. Biol.* 358 (2006), pp. 120–131.

[188] P. J. Stansfeld, E. E. Jefferies, and M. S. P. Sansom. "Multiscale simulations reveal conserved patterns of lipid interactions with aquaporins." In: *Structure* 21 (2013), pp. 810–819.

[189] P. J. Stansfeld et al. “MemProtMD: automated insertion of membrane protein structures into explicit lipid membranes”. In: *Structure* 23 (2015), pp. 1350–1361.

[190] P. J. Stansfeld et al. “The PIP2 binding site in Kir channels: definition by multi-scale biomolecular simulations”. In: *Biochemistry* 48 (2009), pp. 10926–10933.

[191] F.H. Stillinger and A. Rahman. “Improved simulation of liquid water by molecular dynamics.” In: *J. Chem. Phys.* 60 (1974), pp. 1545–1557.

[192] Y. Sugita and Y. Okamoto. “Replica-exchange molecular dynamics method for protein folding”. In: *Chem. Phys. Lett.* 314.1–2 (1999), pp. 141–151.

[193] P. Tiwary and M. Parrinello. “A time-independent free energy estimator for metadynamics”. In: *J Phys Chem B* 119.3 (Jan. 2015), pp. 736–742.

[194] R. Trenker, M. E. Call, and M. J. Call. “Crystal Structure of the Glycophorin A Transmembrane Dimer in Lipidic Cubic Phase”. In: *J. Am. Chem. Soc.* 137.50 (Dec. 2015), pp. 15676–15679.

[195] H. R. Treutlein et al. “The glycophorin A transmembrane domain dimer: Sequence specific propensity for a right handed supercoil of helices.” In: *Biochemistry* 31 (1992), pp. 12726–12733.

[196] G. A. Tribello et al. “PLUMED2: New feathers for an old bird”. In: *Comput. Phys. Commun.* 185.2 (2014), pp. 604–613.

[197] M. B. Ulmschneider et al. “Spontaneous transmembrane helix insertion thermodynamically mimics translocon-guided insertion”. In: *Nat. Commun.* 5 (Sept. 2014), p. 4863.

[198] F. I. Valiyaveetil, Y. Zhou, and R. MacKinnon. “Lipids in the structure, folding and function of the KcsA channel”. In: *Biochemistry* 41 (2002), pp. 10771–10777.

[199] K. Vanommeslaeghe and A. D. MacKerell. "CHARMM additive and polarizable force fields for biophysics and computer-aided drug design". In: *Biochim. Biophys. Acta* 1850.5 (May 2015), pp. 861–871.

[200] C. Vega et al. "What ice can teach us about water interactions: a critical comparison of the performance of different water models". In: *Faraday Discuss.* 141 (2009), pp. 251–276.

[201] L. Verlet. "Computer "Experiments" on Classical Fluids. I. Thermodynamical Properties of Lennard-Jones Molecules". In: *Phys. Rev.* 159 (1 July 1967), pp. 98–103.

[202] P. E. Volynsky et al. "Computer simulations and modeling-assisted ToxR screening in deciphering 3D structures of transmembrane alpha-helical dimers: ephrin receptor A1". In: *Phys. Biol.* 7 (2010), p. 16014.

[203] P. E. Volynsky et al. "Role of dimerization efficiency of transmembrane domains in activation of fibroblast growth factor receptor 3". In: *J. Amer. Chem. Soc.* 135.22 (2013), pp. 8105–8108.

[204] C. K. Wan, W. Han, and Y. D. Wu. "Parameterization of PACE Force Field for Membrane Environment and Simulation of Helical Peptides and Helix-Helix Association". In: *J. Chem. Theory Comput.* 8.1 (Jan. 2012), pp. 300–313.

[205] L. Wang, R. A. Friesner, and B. J. Berne. "Replica exchange with solute scaling: a more efficient version of replica exchange with solute tempering (REST2)". In: *J. Phys. Chem. B* 115.30 (Aug. 2011), pp. 9431–9438.

[206] T. A. Wassenaar et al. "Computational lipidomics with Insane: a versatile tool for generating custom membranes for molecular simulations". In: *J. Chem. Theor. Comput.* 11.5 (2015), pp. 2144–2155.

[207] C. L. Wennberg et al. “Direct-Space Corrections Enable Fast and Accurate Lorentz-Berthelot Combination Rule Lennard-Jones Lattice Summation”. In: *J. Chem. Theory. Comput.* 11.12 (Dec. 2015), pp. 5737–5746.

[208] S. H. White. “The progress of membrane protein structure determination”. In: *Protein Sci.* 13.7 (July 2004), pp. 1948–1949.

[209] S. H. White and W. C. Wimley. “Hydrophobic interactions of peptides with membrane interfaces”. In: *Biochim. Biophys. Acta* 1376.3 (Nov. 1998), pp. 339–352.

[210] S. H. White and W. C. Wimley. “Membrane protein folding and stability: physical principles”. In: *Annu. Rev. Biophys. Biomol. Struct.* 28 (1999), pp. 319–365.

[211] C. W. Wimley and S. H. White. “Experimentally determined hydrophobicity scale for proteins at membrane interfaces.” In: *Nature Struct. Biol.* 3.10 (1996), pp. 842–848.

[212] W. C. Wimley and S. H. White. “Membrane partitioning: distinguishing bilayer effects from the hydrophobic effect”. In: *Biochemistry* 32.25 (June 1993), pp. 6307–6312.

[213] M. G. Wolf et al. “g_membed: Efficient insertion of a membrane protein into an equilibrated lipid bilayer with minimal perturbation”. In: *J. Comput. Chem.* 31 (2010), pp. 2169–2174.

[214] H. Yin and A. D. Flynn. “Drugging Membrane Protein Interactions”. In: *Annu. Rev. Biomed. Eng.* 18 (July 2016), pp. 51–76.

[215] H. L. Zhang et al. “Activation of inwardly rectifying K+ channels by distinct PtdIns(4,5)P-2 interactions”. In: *Nat. Cell Biol.* 1.3 (1999), pp. 183–188.

[216] J. Zhang and T. Lazaridis. “Calculating the free energy of association of transmembrane helices”. In: *Biophys. J.* 91.5 (2006), pp. 1710–1723.

[217] L. Zhang et al. “Prediction, refinement, and persistency of transmembrane helix dimers in lipid bilayers using implicit and explicit solvent/lipid representations: microsecond molecular dynamics simulations of ErbB1/B2 and EphA1”. In: *Proteins* 81.3 (Mar. 2013), pp. 365–376.