TAMPEREEN TEKNILLINEN YLIOPISTO
TAMPERE UNIVERSITY OF TECHNOLOGY

EMMI TURPPA

A COMPUTATIONAL STUDY ON THE MEMBRANE SCULPTING MECHANISM OF MISSING-IN-METASTASIS

Master of Science thesis

Examiners: Prof. Ilpo Vattulainen and
Dr. Maria Kalimeri
Examiners and topic approved by the
Faculty Council of the Faculty of
Natural Sciences
on 17th August 2016

ABSTRACT

EMMI TURPPA: A computational study on the membrane sculpting mechanism of missing-in-metastasis
Tampere University of Technology
Master of Science thesis, 69 pages, 5 Appendix pages
February 2017
Master's Degree Programme in Science and Engineering
Major: Advanced Engineering Physics
Examiners: Prof. Ilpo Vattulainen and Dr. Maria Kalimeri
Keywords: missing-in-metastasis, BAR, IMD, I-BAR, protein-lipid interactions, membrane sculpting, curvature generation, curvature sensing, molecular dynamics, protein flexibility, clustering, insertion

Missing-in-metastasis (MIM) is an adaptor protein that connects the actin cytoskeleton to the plasma membrane. Its N-terminal domain, known as MIM IMD, is a member of the Bin-amphiphysin-Rvs (BAR) domain family, known for their ability to generate membrane curvature. More specifically, it is an inverse BAR (I-BAR) domain, generating negative curvature. MIM IMD is implicated in the formation of lamellipodia and filopodia, disassembly of actin stress fibers, and maintenance of adherence junctions. However, the exact membrane sculpting mechanism it employs has remained elusive.

So far MIM has been studied only experimentally. In this thesis, for the first time we employ molecular dynamics (MD) simulations to computationally address the membrane sculpting mechanism of MIM. We employ both atomistic and mesoscopic scale simulations, examining the behaviour of MIM IMD in the presence of lipid bilayers of different properties.

First, inspection of the crystal structure of the domain under study revealed that it cannot generate negative curvature by simply imposing its intrinsic curvature on the membrane. Introducing dynamics, we found that MIM IMD is actually considerably more flexible as compared to other BAR domains. Moreover, we discovered that MIM IMD can acquire a positive conformation, which may enable its suggested ability to sense and couple with positive membrane curvature. However, our study does not support the proposition that it would sense curvature via inserting its N-terminal amphipathic helix to a membrane. Additionally, our study reveals that significant protein-lipid interactions between the domain and lipids are driven by electrostatic interactions, which further induce clustering of phosphatidylinositol 4,5-biphosphate (PI(4,5)P_2). We suggest the PI(4,5)P_2-clustering may have a significant role in the curvature generation mechanism, due to increase of membrane fluidity.

TIIVISTELMÄ

EMMI TURPPA: Laskennallinen tutkimus missing-in-metastasis'in mekanismista muovata solukalvoa
Tampereen teknillinen yliopisto
Diplomityö, 69 sivua, 5 liitesivua
Helmikuu 2017
Teknis-luonnontieteellinen koulutusohjelma
Pääaine: Teknillinen fysiikka
Tarkastajat: Prof. Ilpo Vattulainen ja tutk.toht. Maria Kalimeri
Avainsanat: missing-in-metastasis, BAR, IMD, I-BAR, proteiini-lipidi-vuorovaikutukset, solukalvon kaareuttaminen, solukalvon kaarevuuden aistiminen, molekyylidynamiikka, proteiinin elastisuus, klusterointi

Missing-in-metastasis (MIM) on aktiini-solutukirangan solukalvoon yhdistävä adaptoriproteiini. Sen MIM IMDnä tunnettu aminopään osa kuuluu Bin-amphiphysin-Rvs -domainien (BAR) sukuun, jotka tunnetaan niiden kyvystä kaareuttaa solukalvoja. Erityisesti se kuuluu käänteisten BAR-domainien (I-BAR) ryhmään, jotka tuottavat negatiivista kaarevuutta. MIM IMDn on todettu muodostavan solun levyjalkoja ja filopodeja, hajottavan stressisäikeitä sekä osallistuvan vyöliitosten ylläpitoon. Sen käyttämä solukalvon kaareuttamismekanismi on kuitenkin edelleen hämärän peitossa.

Tässä diplomityössä lähestymme MIM-proteiinin solukalvon kaareuttamismekanismia laskennallisesti molekyylidynamiikka-simulaatioiden kautta. Aiemmin kyseistä proteiinia on tutkittu vain kokeellisesti. Toteutamme sekä atomistisen että mesoskooppisen tason simulaatioita ja tutkimme MIM IMDn käyttäytymistä ominaisuuksiltaan erilaisten lipidi-kaksoiskalvojen yhteydessä.

MIM IMDn kiderakenteen tarkastelu osoittaa, että se ei kykene kaareuttamaan solukalvoa pelkän luontaisen kaarevuutensa avulla. Simulaatiomme sen sijaan paljastavat sen olevan huomattavasti elastisempi muihin BAR-domaineihin verrattuna. Edelleen tutkimuksemme paljastaa, että MIM IMD voi omaksua positiivisesti kaareutuneen olomuodon, mikä mahdollistaa sille aiemmin ehdotetun kyvyn havaita ja sitoa positiivisesti kaareutuneita solukalvoja. Tämä työ ei toisaalta tue toista ehdotusta, jonka mukaan se havaitsisi positiivista kaarevuutta sen aminopään amfipaattisen alfa-kierteisen osan kautta työntämällä tämän kalvoon. Lisäksi osoitamme, että sähköiset vuorovaikutukset ohjaavat merkittäviä proteiini-lipidi-vuorovaikutuksia ja aiheuttavat fosfatidyyli-inositoli 4,5-bifosfaattien (PI(4,5)P_2) klusteroitumista. Ehdotamme, että PI(4,5)P_2-klusteroinnilla on merkittävä rooli solukalvon kaareuttamisessa, sillä se lisää solukalvon nestemäisyyttä.

PREFACE

This Master of Science thesis was written between March and August 2016 as a part of the research conducted in the Biological Physics and Soft Matter group, in the Department of Physics in the Tampere University of Technology (TUT). The purpose of this study was to inspect the membrane sculpting mechanism employed by a protein called missing-in-metastasis (MIM). The computing facilities were provided by the Finnish IT Center for Science (CSC) and Tampere Center for Scientific Computing (TCSC).

I would like to thank my examiner professor Ilpo Vattulainen for all the support and for giving me the opportunity to conduct research on such compelling and essential subjects as a part of his group. I also wish to thank my examiner and supervisor Maria Kalimeri for all the guidance, her contributions in performing part of the simulations, proofreading and teaching me so much about research.

The whole group has been supportive and ready to help with any problems. Especially, I would like to thank my colleagues in the Batcave-office, Miia Niemelä, Sami Rissanen and Heikki Mikkolainen, for the advice, laughs and updates in sports. I would like to express special thanks also to Matti Javanainen for sharing his expertise regarding lipids.

Furthermore, I would like to thank my family for the support and for bringing me up so thirsty for knowledge. I thank my brothers for teaching me that one can do cool stuff with computers, which essentially resulted in my interest in scientific computing. Finally, I would like to thank Atte Antikainen for encouraging and supporting me throughout this project.

Tampere, 18.1.2017

Emmi Turppa

TABLE OF CONTENTS

1. Introduction 1

2. Biological Background 3
  2.1 The Actin Cytoskeleton 3
    2.1.1 Structure 4
    2.1.2 Function in Eukaryotic Cells 5
  2.2 Bin-Amphiphysin-Rvs Domains 6
    2.2.1 General Structure and Function 7
    2.2.2 Inverse Bin-Amphiphysin-Rvs Domains 8
  2.3 Missing-In-Metastasis 9
    2.3.1 Structure 9
    2.3.2 Function in Eukaryotic Cells 11
    2.3.3 Relation to Cancer 13

3. Molecular Dynamics 14
  3.1 The Classical Limit of Quantum Theory 15
  3.2 Force Field 16
    3.2.1 Bonded Interactions 17
    3.2.2 Non-Bonded Interactions 19
    3.2.3 All-Atom versus Coarse-Grained Force Fields 20
  3.3 Energy Minimization 21
  3.4 Molecular Dynamics Pseudo Code 22
  3.5 Temperature and Pressure Coupling 23
  3.6 Periodic Boundary Conditions 26
  3.7 Equations of Motion 26
  3.8 Limitations 27

4. Analysis Methods 29
  4.1 Diffusion Coefficients for Lipids 29
  4.2 Area Per Lipid 30
  4.3 Radial Distribution Function 30
  4.4 Density Mapping 31
  4.5 Surface Tension 31
  4.6 Cluster Analysis 32
  4.7 Principal Component Analysis 33
  4.8 Lipid-binding Residues 34
  4.9 Protein Insertion into the Lipid Bilayer 34
  4.10 Membrane Curvature 35

5. Simulation Models of Missing-In-Metastasis 37
  5.1 Simulation Systems 37
  5.2 Simulation Parameters 41

6. Results and Discussion 42
  6.1 Intrinsic Protein Curvature and Flexibility 43
    6.1.1 Intrinsic Curvature of the Crystal Structure 43
    6.1.2 MIM IMD is Remarkably Flexible 44
  6.2 Membrane Curvature Generation by MIM IMD 46
    6.2.1 MIM IMD Slows Down Lipid Diffusion 47
    6.2.2 Lipid-binding is Driven by Electrostatic Interactions 48
    6.2.3 MIM IMD Induces Clustering of PI(4,5)P_2 50

6.2.4 A Single MIM IMD Cannot Efficiently Induce Local Curvature . 53
6.3 Curvature Sensing Ability of MIM IMD 55
6.3.1 MIM IMD Does Not Insert Its N-Terminal Helix 55
6.3.2 MIM IMD May Couple with Positively Curved Bilayers by Acquiring a Positively Curved Conformation 56
6.3.3 The N-termini of MIM IMD do not Prefer Insertion 58

7. Conclusions 60

Bibliography 62

APPENDIX A. Lipid Diffusion in the Absence of Protein 70

APPENDIX B. Minimum Distances in Protein-Lipid Interactions 71

APPENDIX C. Lipid-binding Interface of MIM IMD 72

APPENDIX D. Density Maps 73

APPENDIX E. Average Distal End Insertion on High Tension Bilayers 74

LIST OF ABBREVIATIONS

2D	Two-dimensional
3D	Three-dimensional
AA	All-atom
ABP	Actin-binding protein
ADP	Adenosine diphosphate
ALA	Alanine
Arp2/3	Actin-related protein-2/3
ARG	Arginine
ASN	Asparagine
ASP	Aspartic acid
ATP	Adenosine triphosphate
BAR	Bin-amphiphysin-Rvs
CSC	The Finnish IT Center of Science
CG	Coarse-grained
CHARMM36	Chemistry at Harvard Molecular Mechanics 36 (a force field)
DNA	Deoxyribonucleic acid
DPH	1,6-diphenyl 1,3,5-hexatriene
EN	Elastic network
F-BAR	A positively curving BAR-domain first identified as a homology of FER and CIP4 proteins
FF	Force field
GLN	Glutamine
GLY	Glycine
GTP	Guanosine triphosphate
I-BAR	A BAR-domain that induces inverse, i.e. negative, curvature, see also IMD
IMD	IRSp53/MIM homology domain, see also I-BAR
IRSp53	Insulin receptor tyrosine kinase substrate p53
LINCS	Linear constraint solver
LJ	Lennard-Jones
LYS	Lysine
MD	Molecular dynamics
MIM	Missing-in-metastasis
MSD	Mean squared displacement
mRNA	Messenger ribonucleic acid

MTSS1	Metastasis suppressor 1, a gene encoding MIM
N-BAR	A positively curving BAR-domain containing an N-terminal amphipathic helix
NPT	Isothermal-isobaric ensemble: constant number of atoms, pressure and temperature
NVT	Canonical ensemble: constant number of atoms, volume and temperature
PBC	Periodic boundary condition
PCA	Principal component analysis
PDB	Protein data bank
$PI(4,5)P_2$	Phosphatidylinositol 4,5-biphosphate
POPC	1-palmitoyl-2-oleoyl-*sn*-glycero-3-phosphocholine
POPE	1-palmitoyl-2-oleoyl-*sn*-glycero-3-phosphoethanolamine
POPS	1-palmitoyl-2-oleoyl-*sn*-glycero-3-phosphoserine
RDF	Radial distribution function
RMSD	Root mean squared deviation
SER	Serine
TCSC	Tampere Center for Scientific Computing
TEM	Transendothelial cell macroaperture
THR	Threonine
TUT	Tampere University of Technology
WH2	Wiscott-Aldrich syndrome protein homology 2

1. INTRODUCTION

The interplay between the plasma membrane and the intracellular actin-protein network, namely the actin cytoskeleton, is vital for the cell structure and motility as well as cell division [1, p. 914]. However, actin requires adaptor proteins to connect it to a membrane and to stabilize the linkage.

Missing-in-metastasis (MIM) is an adaptor protein which links the actin cytoskeleton to the plasma membrane in many eukaryotic cells. It has been suggested as a metastasis suppressor in a multitude of cancers, such as bladder cancer [2]. Interestingly, it has also been suggested as a metastasis driver in some other cases of cancer, for instance in a subset of human melanomas [3]. In order to understand the context-dependent function of MIM, it is necessary to understand the mechanism it employs.

MIM has an N-terminal membrane sculpting domain, which generates negative curvature [4]. This domain is conserved in both MIM and another actin-binding protein called insulin receptor tyrosine kinase substrate p53 (IRSp53). Consequently, it is usually abbreviated as IMD (IRSp53 and MIM homology Domain). *In vitro*, MIM IMD deforms membranes into tubules, on whose inner surface it binds to [5]. *In vivo*, MIM IMD is associated in the formation of protrusions in the cell, disassembly of actin stress fibers, and maintenance of adherence junctions [5–8].

The membrane sculpting ability of MIM IMD is strongly associated with electrostatic interactions between itself and a lipid bilayer [4, 7]. However, the exact membrane sculpting mechanism has remained elusive. Additionally, although somewhat counterintuitively, MIM IMD has been proposed to sense positively curved membranes and localize on them [5, 9]. This sensing mechanism is also poorly understood. In this thesis we study the membrane sculpting mechanism along with the sensing mechanism that controls MIM's localization within the cell.

This thesis probes the interactions between MIM IMD and a lipid bilayer via molecular dynamics (MD) simulations. MD is a numerical method for solving the trajectory of a classical many-body system [10, p. 63]. The basic idea is to iteratively solve Newton's equations of motion for each particle in a system and thus computationally simulate its behaviour in predefined conditions [11, p. 2]. It allows us to inspect otherwise inaccessible nanoscale phenomena.

To date, MD simulations have come a long way, being able to simulate systems with vast amounts of atoms for several milliseconds. Due to the constant increase of computational power, MD performance continues to grow. For this thesis, we performed MD simulations of systems containing hundreds of thousands of atoms. In the near future, in the biological context, we expect MD to be able to handle systems at the more complex cellular level [12].

This thesis is a part of the research conducted in the Biological Physics and Soft Matter group in the Tampere University of Technology (TUT). In the seven chapters of this thesis, we first discuss the relevant biological theory and methodology, then the studied systems and results. The following chapter concentrates on the biological background, starting from the more general concepts and narrowing down to MIM. Subsequently, the third chapter focuses on the theory of MD, while the fourth chapter describes the applied analysis methods. The fifth chapter presents our simulation systems along with the applied simulation parameters. Chapter six displays the obtained results and discusses their significance in the light of preceding experimental results. Finally, chapter seven summarizes the study while discussing future prospects for related research.

2. BIOLOGICAL BACKGROUND

Actin is a protein that plays an important role in eukaryotic cells. Most eukaryotic cells have an actin-rich cortex underneath their plasma membrane along with actin monomers spread throughout the cell interior [13, p.594]. This actin cytoskeleton is a part of the highly flexible supporting framework of the cell. Together with other cytosolic members it regulates many of the cell's vital functions, such as motility [13, p. 590].

The plasma membrane serves as an important interface between the cell and its surroundings. It is formed by lipids that contain a hydrophobic head and one or two hydrophobic tails [13, p. 365; 14, p. 38–39]. Lipids self-assemble into bilayers, such as the plasma membrane, to seal their hydrophobic chains away from water contact [13, p. 367; 14, p. 77]. Lipid bilayers also compartmentalize different organelles within the cytosol [13, p. 364].

The actin cytoskeleton is connected to lipid bilayers intracellularly by an exquisite set of adaptor proteins. One of them is missing-in-metastasis (MIM), whose N-terminal domain is the main target of our study. In this chapter we review the biological context of MIM, discussing the current understanding of its function and introducing the open question that this study addresses.

2.1 The Actin Cytoskeleton

The actin cytoskeleton is an essential cellular structure present in nearly all eukaryotic cells, providing the cell with structure and motility [1, p. 914]. It consists of actin monomers that associate to construct actin filaments, also known as microfilaments, that can be arranged into various flexible structures [1, p. 911].

In this section we delve into the basic constituents of the actin cytoskeleton and its functions in eukaryotic cells. We begin from its basic subunits, the actin monomers, and see how they build up to the elaborate actin cytoskeleton.

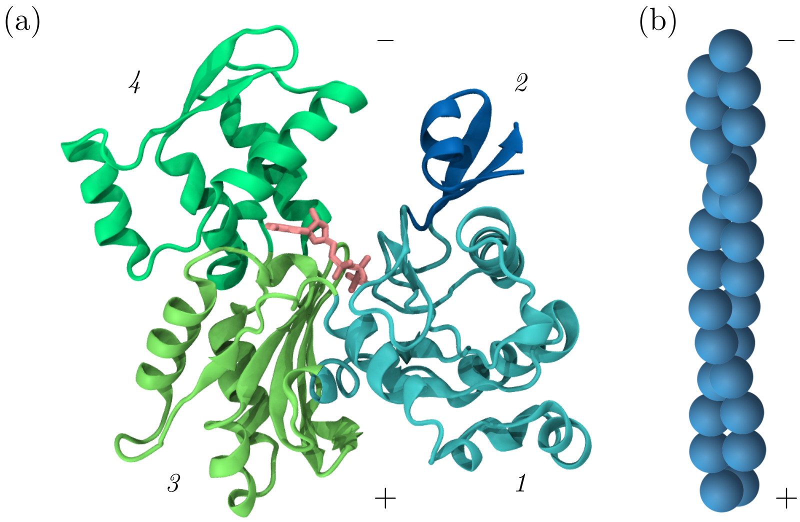

Figure 2.1 The structure of an actin (a) monomer and (b) filament. In (a) the two larger subdomains of G-actin viewed in green and blue are divided into four subdomains in total, represented by different shades and numbered respectively. The ATP bound to the actin monomer is shown in pink in the middle. In (b) each blue sphere represents an actin monomer with bound ADP. Figures adapted from [1, p. 899].

2.1.1 Structure

The most ground-laying constituent of the actin cytoskeleton is the globular actin monomer, also known as G-actin. Consisting of 375 amino acids, it additionally carries a nucleotide, either an adenosine triphosphate (ATP) or adenosine diphosphate (ADP) molecule in a cleft in its center [1, p. 898].

G-actin is structurally polar, as illustrated in figure 2.1 (a) [1, p. 898; 13, p. 591]. The ATP- or ADP-binding cleft opens out toward the minus end, while the other end is respectively known as the plus end [1, p. 898].

The actin filaments, also called F-actin, assemble from multiple actin monomers forming two intertwined polymers, i.e. a double helix [13, p. 592]. When forming filaments the monomers are stacked upon each other, minus end to plus end, so that the filament, too, has structural polarity as illustrated in figure 2.1 (b) [1, p. 898; 13, p. 591]. A filament tends to grow faster from its plus end due to faster dynamics as compared to the minus end, where a monomer undergoes a more hindering conformational change upon polymerization [1, p. 898, 902].

A free G-actin carries an ATP molecule, which may hydrolyse into an ADP molecule upon binding onto an actin filament [13, p. 591–592]. The hydrolysis, however, degrades the binding affinity between two adjacent monomers, making the filaments inherently unstable.

A G-actin bound with an ADP molecule is more likely to dissociate from either end of the filament than an ATP-bound one [1, p. 903]. Hence, the G-actin concentration around a forming actin filament affects whether it is more likely to be assembling or disassembling: if the rate of addition of the ATP-bound monomers is faster than the rate of hydrolysis, the ATPs do not have enough time to hydrolyse and the filament will assemble. This happens above the so called critical concentration [1, p. 901]. On the other hand, with low concentration the binding affinity is deteriorating faster than new monomers are added, making the filament more likely to disassemble. The critical concentration is often higher for the already slow-growing minus end than for the fast-growing plus end, leading the filament to assemble at the plus end while shrinking from the minus end in a process called treadmilling [1, p. 901, 903].

Due to the co-action of the rates of ATP hydrolysis and filament assembly, the actin cytoskeleton is very adaptive and flexible. The filaments can both assemble and disassemble rapidly from either end, which is very useful for several purposes in the cell.

2.1.2 Function in Eukaryotic Cells

The actin cytoskeleton has versatile functions in the eukaryotic cells concerning the shape and motion of the cell and the cell's mechanical properties. It contributes to muscle contraction, signalling, and cell division [1, p. 911; 13, p. 590]. Depending on their function, actin filaments arrange in different conformations as illustrated in figure 2.2 (a).

In cell crawling, i.e., when a cell crawls over a surface, actin filaments form protrusions like lamellipodia and filopodia that stick outward from the cell center [13, p. 594]. These protrusions probe the surroundings, adhere to the surface, and drag the rest of the cell onwards [13, p. 596]. In lamellipodia actin filaments form a sheet-like branched mesh, while filopodia are more like spikes with parallel filaments [1, p. 911]. These protrusions have an important role, e.g., in the head of the signal transmitting axons in neurons [13, p. 594–595]. Actin also has a main role in the microvilli formation in the epithelial cells in the intestine [13, p. 590].

Actin filaments can additionally form contractile bundles, which together with a protein called myosin enable the contraction of muscle cells [13, p. 597]. In non-muscle cells, such contractile bundles are called stress fibers, and they contribute to cell motility. During cell division actin forms a contractile ring that pinches the

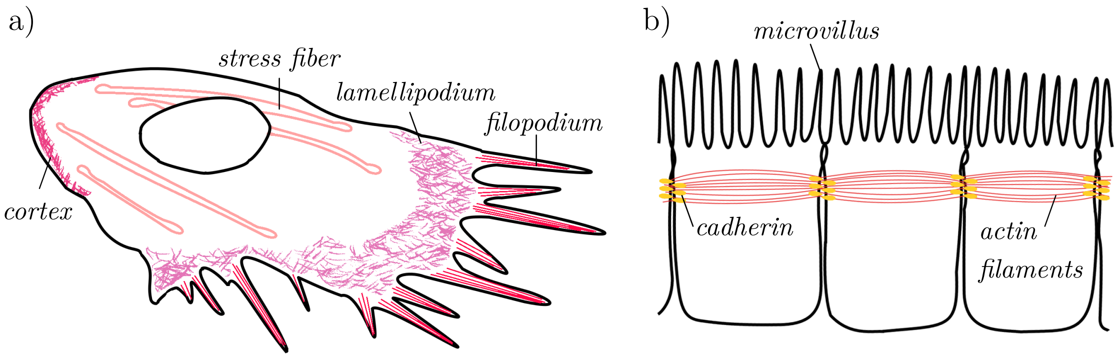

Figure 2.2 (a) Different actin filament conformations, represented in different colours. Figure adapted from [1, p. 911]. (b) Adherence junctions in epithelial cells, which additionally exhibit microvilli. The actin in the microvilli is not shown for simplicity. Figure adapted from [1, p. 1044].

two daughter cells apart in cytokinesis [13, p. 590]. On the cortex of the cell, actin filaments form a gel-like network [1, p. 911].

Moreover, actin is essential in adherens junctions, where adjacent cells are sealed together via actin bundles located at their cytosolic membrane surface. Adherens junctions form when cadherin molecules in the plasma membrane of adjacent cells bind to each other extracellularly, while intracellular linker proteins attach them to actin filaments [13, p. 704]. The filaments often form a continuous belt reaching through all connected epithelial cells as seen in figure 2.2 (b), allowing them to contract, which permits tubular tissue structures with essential roles for instance in embryonic development [13, p. 703–704].

All the functions performed by actin filaments are guided by actin-binding proteins (ABPs). ABPs are able to modify the length, organization, assembly, and dynamics of the filaments, facilitating their function [1, p. 904–905]. They even enable the nucleation of a filament by bringing actin monomers together to form a nucleus, a seed, from which to grow [1, p. 906]. ABPs consist of multiple protein families. The one most relevant for our study is introduced in the following section.

2.2 Bin-Amphiphysin-Rvs Domains

Proteins containing a Bin-Amphiphysin-Rvs (BAR) domain sculpt the plasma membrane, inducing invaginations and protrusions indispensable in processes like cytokinesis, endocytosis, fission of synaptic vesicles, and migration [15–17]. BAR domains are ABPs that, in addition to the plasma membrane, also interact with the actin cytoskeleton, also necessary in the above mentioned cell functions.

BAR domains have been shown to deform membranes into tubules *in vitro*, in the absence of actin [5, 15, 18]. In this section we will get a first grasp onto how the domains function and especially direct our interest towards a subfamily known as inverse BAR domains.

2.2.1 General Structure and Function

BAR domains are dimers possessing a crescent-like shape. In the so called canonical BAR domain both of the two monomers consist of three α-helices [5]. BAR domains have both an actin- and a membrane-binding-interface, the latter typically consisting of positively charged amino acids that are attracted to negatively charged membranes [5, 19, 20].

There have been several studies suggesting that BAR domains can deform membranes via one or more of three different mechanisms: electrostatic interactions between the charged residues of the domain and the lipids, insertion of an amphipathic helix, and scaffolding [5, 15, 20, 21]. In insertion, an amphipathic helix, acting like a wedge, inserts into the bilayer and increases the area of one lipid monolayer. However, the efficiency of this mechanism has been questioned due to the low density of amphiphatic helices in the protein domains [22]. In scaffolding, one or more domains provide a mould for the membrane [23].

Studies also suggest that BAR domains can sense membrane curvature [5, 19, 20, 22, 24]. Curvature sensing essentially signifies an increased affinity towards curved membranes. Amphipathic helices have been suggested to be an important tool for this function [22]. The sensing ability is argued to depend on the amino acid sequence at the polar face of a motif: a polar face abundant with uncharged polar residues implies curvature sensitivity, while a strongly charged lipid-binding interface binds to oppositely charged lipids irrespective of the membrane curvature [22]. Both mechanisms require specific features in the lipid composition. Consequently, the lipid composition also affects the curvature sensing activity, leading the domain to localize onto specific membrane sites in the cell [22]. However, there has also been experimental evidence that BAR domains even without amphipathic helices are able to sense membrane curvature [15].

The diversity of the BAR domain family stems from the plethora of different lengths, intrinsic curvatures and membrane-binding affinities of its members [15]. In general BAR domains can be divided into three subsets: N-BARs that contain an N-terminal amphipathic helix, F-BARs (homology of FER/CIP4 proteins) that contain exceptionally five α-helices per monomer, and finally the inverse BARs (I-

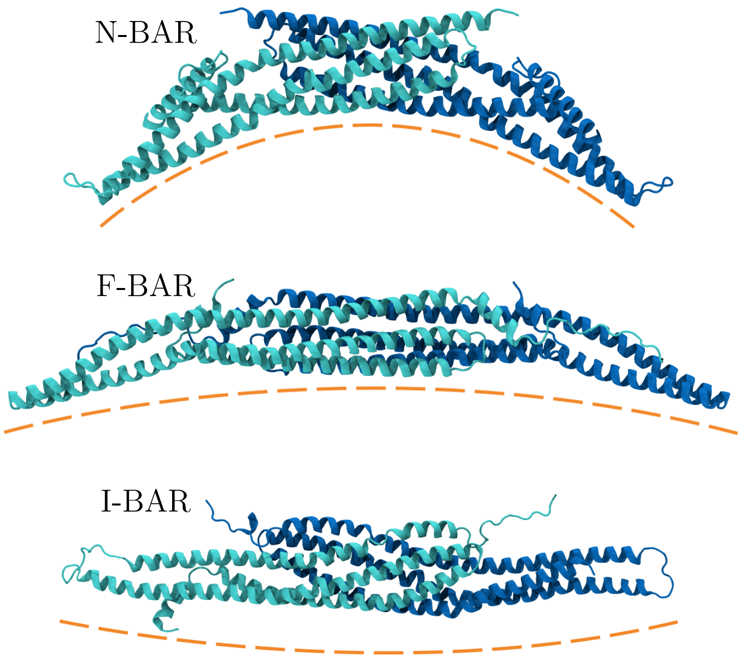

Figure 2.3 A ribbon representation of each type of BAR domain with the two monomers in different colours. The orange dashed line indicates the domains' intrinsic curvature of the lipid-binding interface. Figure adapted from [15].

BAR) that display a zeppelin-like shape [5]. While the membrane-binding-surface of N-BARs and F-BARs show a concave shape, the membrane-binding-interface of I-BARs, in contrast, is convex as illustrated in figure 2.3. They are all indicated to produce the kind of membrane curvature that matches their intrinsic curvature. The curvature induced by N-BARs and F-BARs is known as positive curvature, while I-BARs generate negative curvature.

2.2.2 Inverse Bin-Amphiphysin-Rvs Domains

The I-BAR domain is also known as the IRSp53/MIM homology domain (IMD) due to insulin receptor tyrosine kinase substrate p53 (IRSp53) and missing-in-metastasis (MIM), the proteins it was first identified in [25]. IMDs bind to the inner leaflet of membrane tubules *in vitro* [5]. Consisting of two monomers of three anti-parallel α-helices that are intertwined into a dimer, IMDs obey the canonical BAR domain structure. In contrast to N-BARs and F-BARs, the membrane sculpting mechanism of I-BARs is poorly understood.

All I-BAR domains are structurally similar, yet they can be divided into IRSp53-

like and MIM-like domains [24], which notably differ in their amino acid sequences [4]. Henceforth, we will concentrate on the less studied MIM IMD.

2.3 Missing-In-Metastasis

Missing-in-metastasis is an I-BAR domain protein, encoded by the gene named metastasis suppressor 1 (MTSS1). MIM was originally identified by means of a modified version of the messenger ribonucleic acid (mRNA) differential display technique and reported as a potential metastasis suppressor, as its expression in certain metastatic, i.e. spreading, cancer cell lines was found to be downregulated [2].

In normal non-metastatic human tissues, MIM is expressed widely: for instance in the bladder, spleen, thymus, testis, prostate, uterus, colon, and peripheral blood [2, 26]. Generally speaking, MIM appears to be present only in vertebrates [26]. In particular, experimental studies have been performed with mice, in which MIM is highly expressed in embryonic muscles, heart and postmitotic neurons as well as in adult liver, kidneys, and cerebellar Purkinje cells [8, 26, 27].

It is essential to recognize the biological features and previous findings regarding the protein under study. In this section we converse on reported results of the structure and function of the whole MIM protein, along with its relevance with regard to cancer. Regarding the structure and function, we will especially focus on MIM IMD, since this domain is our main interest in this thesis.

2.3.1 Structure

MIM is a 759 amino acid [6, 27] protein consisting of an N-terminal IMD, a C-terminal Wiscott-Aldrich syndrome protein homology 2 (WH2) domain and an intermediate regulatory part [4, 24]. The IMD and WH2 domain are the basis of the membrane and actin related functionalities of MIM.

The IMD of MIM is roughly 183 Å in length and 30 Å at its widest in diameter [4]. Both monomers of this homodimer consist of 250 amino acids that form three antiparallel α-helices [4]. The two subunits are also antiparallel with respect to each other, interacting extensively through a 2941 Å^2 contact area [4]. However, they leave a fairly large 1396 Å^3 water-filled cavity in the middle of the dimer [4]. Primarily hydrophobic side chains are directed toward this cavity, but also several polar amino acids are oriented towards it so that they form stabilizing non-covalent salt bridges between the two dimers [4].

Interestingly, the two subunits are not exactly identical in the crystallographically resolved structure as seen in figure 2.4. Although their sequences are identical,

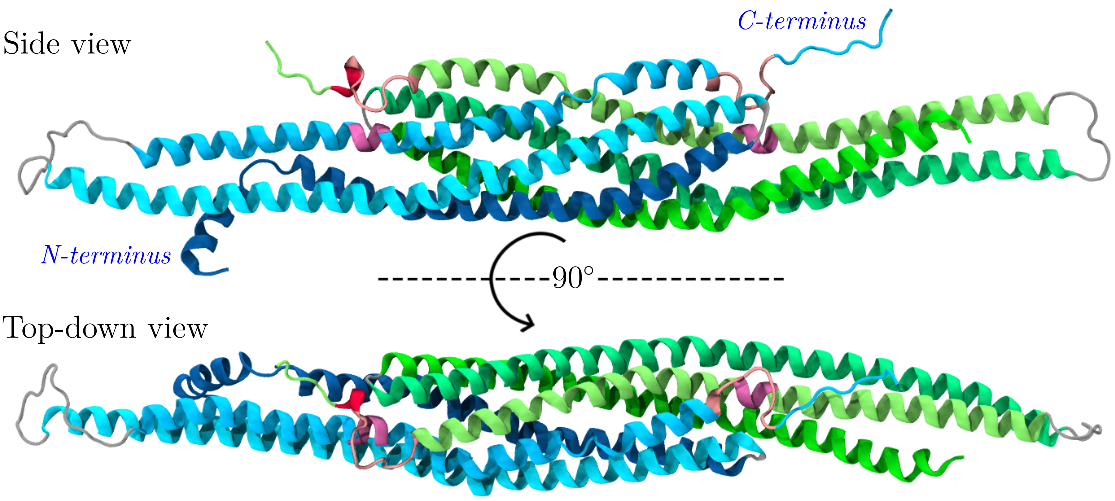

Figure 2.4 The IMD of MIM from two perspectives. The two monomers are coloured in shades of blue and green, different shades representing different helices. The C- and N-termini are marked for the blue-coloured monomer. Marked in red is the helix that is missing from the respective part of the other monomer. In the C-termini, the sequence marked in pink indicates the so called flap loop. Finally, presented in magenta, are the signature sequences. Notice the downwards tilted N-terminal helix in the blue monomer.

they differ in conformation: the N-terminal helix is protruding from the membrane-binding interface only in one monomer. Additionally, while one subunit exhibits at the tip of its C-terminus a short helix, in the other one this helix is unfolded [4]. On the other hand the 9 amino acid sequence preceding the aforementioned helix is found similar in both monomers and it connects the IMD to other domains of MIM [4]. This so called flap loop covers the IMD's signature sequence in the other monomer, burying some charged amino acids [4]. The burial of charged amino acids might imply an important function in this part of the domain, possibly regarding protein-protein interactions [4].

The IMD of MIM additionally contains clusters of positively charged amino acids at the distal ends of the domain [4,7]. These positively charged residues are essential when interacting with negatively charged lipids.

The WH2 domain at the C-terminus of MIM consists of 31 amino acids, being exceptionally large with respect to prototypical WH2 domains [4, 24]. Moreover, its location is also uncommon as compared to other cytoskeletal proteins, as it is not accompanied by other identical motifs nor immediately preceded by proline-rich sequences [4, 24]. The WH2 domain is comprised of an N-terminal amphipathic α-helix, importantly exhibiting a hydrophobic surface, and a C-terminal extended region [4].

2.3.2 Function in Eukaryotic Cells

The two domains residing in the opposite termini of MIM execute virtually all its functions: while the WH2 motif binds actin monomers, the IMD interacts with actin filaments and the plasma membrane [6–8, 24, 27]. *In vitro*, overexpression of MIM has been discovered to promote protrusions like filopodia and lamellipodia as well as disassembly of actin stress fibers [6, 7]. The mechanism by which MIM operates has been examined in diverse ways and there has been multiple suggestions regarding its purpose *in vivo*.

In experiments MIM IMDs deform membranes into tubules of 60 nm diameter [5]. The membrane-deforming ability of MIM IMD is strongly associated with the electrostatic interaction between its positively charged residues and negatively charged lipids [4,5,7,9,24]. *In vitro* experiments have additionally shown that IMDs induce clustering of the negatively charged phosphatidylinositol 4,5-bisphosphate (PI(4,5)P_2) lipids upon membrane-binding [5]. The electrostatic interactions have a crucial role also in this activity [5]. The clustering of a lipid type alters the membrane properties and thus the clustering potentially has an important role in the membrane-deforming ability of MIM IMD [5]. However, the mechanism by which MIM IMD generates membrane curvature remains elusive.

The IMD is likely to employ additional mechanisms in localization to favourable binding-sites besides utilizing its positive residues, because compromised electrostatic interactions do not notably alter the IMDs' co-sedimentation with membranes [5]. Consequently, the IMD has been argued to insert its N-terminal amphipathic helix into the membrane [5,9]. Although the helix has not been found to affect the membrane-binding affinity, the membrane has been shown to be more fluid in the presence of an altered IMD that lacks this helix as compared to when a wild-type MIM is present [5]. This may indicate that the helix is inserted into the bilayer [28]. The potentially inserting helix is argued to sense flat and positive membrane curvature, while the electrostatic interactions sense negative curvature [5, 9, 22]. Thus the helix, along with the IMD's positively charged residues, would enable MIM's specific localization [9]. This helix is not, however, found critical for bending membranes negatively [5]. On the contrary, it has been found to inhibit the sculpting effect [29].

In addition to membranes, the IMD of MIM binds to actin filaments [4, 7]. The F-actin-binding sites of the IMD are expressed in a remarkably large, roughly 4 nm long region along helix-2 [7]. After mutational experiments, the interaction with actin filaments was revealed to happen through non-specific interaction, via a non-specific binding site [4, 7]. Interestingly, the PI(4,5)P_2-binding sites were discovered to overlap with the F-actin-binding sites, indicating that the two binding functions might be competing with each other [7].

Additionally, IMDs have been suggested to bundle actin filaments, but this argument is controversial. In contrast to some previous studies, Mattila et al. [7] showed that the F-actin bundling activity of IMD is weak at physiological salt concentration, while in low salt concentrations the aggregation of IMDs and hence the bundling effect is increased significantly. Consequently, they suggest that the results supporting the assumed F-actin bundling activity is due to experiments performed in non-physiological salt concentrations. All the same, the IMD is able to induce membrane protrusions independently of actin bundling [7].

In addition to actin and lipids, MIM IMD adheres to both inactivated and activated guanosine triphosphate (GTP) binding Rac-protein [24]. An activated Rac promotes actin protrusions during the formation of adherens junctions [1, p. 1043]. In kidney epithelium MIM is argued to contribute to the maintenance of these adherens junctions, to which it has been shown to localize via its IMD in cultured epithelial cells [8]. The binding affinity of MIM IMD to Rac-protein further supports MIM's role in adherens junctions, and MIM has also been observed to induce actin filament accumulation at these junction sites [8]. Actin assembly by MIM was suggested important for stabilizing cadherin adhesions, which are essential in this type of cell-cell contacts [8].

MIM has additionally been connected to the Arp2/3 complex in the endothelium and kidney epithelium [8, 9]. In the firstly mentioned case, MIM was discovered to drive the closure of transendothelial cell macroaperture (TEM) tunnels, involved in dispersing bacteria by rupturing the host endothelium barrier [9]. The IMD was found to promote the accumulation of MIM to the edges of the tunnels, while other domains triggered Arp2/3-dependent actin polymerization [9].

In contrast to the IMD, the WH2 motif is able to bind to actin monomers [4, 27]. Being longer than the usual WH2 domains, MIM WH2 can interact with all four subdomains of actin [4]. Its main function is likely to recruit MIM to specific cytoskeletal networks [4] and to regulate actin filament assembly [27].

Furthermore, co-localization of a seven subunit actin-related protein-2/3 (Arp2/3) complex in MIM-induced protrusions has been observed [6]. The Arp2/3 complex is a known regulator of the actin cytoskeleton, so the co-localization indicates that the two proteins might work together in assembling actin [6]. Moreover, inactivation of Arp2/3 has been shown to reduce the actin assembling by MIM [8]. Thus MIM is suggested to mediate Arp2/3-dependent intercellular actin filament assembly [8].

2.3.3 Relation to Cancer

The life of a cell is normally controlled by extracellular signals [13, p. 642]. Depending on the kind of signal it does or does not receive, the cell can either survive and continue to proliferate or die by apoptosis [13, p. 640–643]. Cancer, on the other hand, evolves from cells rebelling these rules: they survive and proliferate when they should not, and invade into regions intended for other cells [13, p. 718]. Malignant tumours usually arise due to a mutation in the cell's deoxyribonucleic acid (DNA), which induces properties promoting the cell's survival [13, p. 721]. As a result of natural selection among the cells – and possibly new mutations – the cancerous cell proliferates, eventually developing into a tumour [13, p. 721].

The evolution of tumours is regulated by tumour promoters and suppressors. MIM was originally suggested as a metastasis suppressor in bladder cancer [2], meaning that the expression of MIM might inhibit the spread of tumours in bladder cancer. Furthermore, MIM has been connected as a potential suppressor for instance in metastatic prostate cancer, breast cancer, ovarian cancer, acute myeloid leukemia, and hepatocellular carcinoma, the latter being the most common type of malignant liver cancer [2, 30]. Fundamentally, low expression of MIM has been connected to poor prognosis in these diseases, while high expression of MIM has lead to favourable outcomes [30].

Tested in mice, reduced MIM expression led to abnormalities in the kidney tissue, possibly due to weakening of adherens junctions [8]. Weakened contacts enable the migration of the supposedly immobile cells and promote their invasive properties, which might explain the role of MIM as a metastasis suppressor [8].

Interestingly, in a subset of human melanomas and in colorectal cancer MIM has demonstrated the completely opposite role of a metastasis driver [3, 30]. These findings imply that MIM might have a context-dependent function [30].

Woodings et al. [6] argue that MIM might be a downstream regulator of actin assembly in tyrosine kinase signalling, which transmits extracellular signals into the cell and regulates, e.g., its proliferation, differentiation, and survival [13, p. 555], which again could explain the mechanism by which MIM is involved in metastasis. Moreover, MIM has been identified as a sonic hedgehog-responsive gene, that is, having a role in sonic hedgehog signalling, which is critical during development and carcinogenesis, providing yet another opportunity for MIM to affect the evolution of tumours [31].

3. MOLECULAR DYNAMICS

The experimental means of studying the phenomena at the molecular level are diverse yet often indirect. In the scale of one billionth of a meter, which is roughly a ten-thousandth of the thickness of a human hair [32, p. 31], physical phenomena occur in the time scale of one billionth of a second. How can we measure these incredibly small-scaled events our eyes could never keep up with? Molecular dynamics (MD) offers a solution: it is a computational method that permits us to simulate the details of nanoscale events.

MD is a simulation technique used for studying the effect of forces on a classical many-body system [10, p. 63]. MD simulation is like a numerical replica of real experiments: the simulation system under study, or the "sample", is prepared and measured in a controlled environment for a desired amount of time [10, p. 63–64; 33, p. 3]. The preparative phase in a simulation consists of constructing the model system, relaxing it through energy minimization, and permitting it to equilibrate. Then, the measurement can be performed in conditions determined by the simulation parameters.

While in the 1970's and 1980's some early-state MD simulations were performed for a small number of atoms for some hundreds of picoseconds, nowadays simulations can be performed for millions of atoms and last for several milliseconds [34, p. 9; 35]. Within the next decades, the constant improvement of computational performance is expected to aid MD simulations to move into even larger systems and longer time scales, allowing MD to rise from the molecular scale to the more complex cellular level [12, 35].

In this study we use the GROMACS software [11, 36, 37] to perform our simulations. It is a widely used software package especially intended for simulating systems composed of biological molecules [34, p. 11; 37]. In this chapter we will approach the theory behind MD starting from the general concepts and moving down to the details of its implementation.

3.1 The Classical Limit of Quantum Theory

As we zoom into the level of atoms, quantum mechanics becomes relevant, yet MD treats atoms by the rules of classical physics. The reason why this is possible, can be understood by inspecting the expectation values of the relevant operators.

Quantum mechanically the expectation value of an arbitrary operator A is defined as

$$\langle A\rangle = \int \psi^*(x,t)A\psi(x,t)dx, \tag{3.1}$$

where $\psi(x,t)$ is the probability of finding a particle at position x at time t and $\psi^*(x,t)$ is its complex conjugate [38, p. 103]. Thus, by differentiation, we can derive

$$\frac{d\langle A\rangle}{dt} = \left\langle \frac{\partial A}{\partial t} \right\rangle_t + \frac{i}{\hbar}\langle [H,A]\rangle, \tag{3.2}$$

where $\hbar = h/2\pi$ is the reduced Planck constant, $[H,A]$ is the commutation relation between the Hamiltonian $H = \frac{p^2}{2m} + V(x)$, including mass m, and A [38, p. 103]. This reduces down to

$$\frac{d\langle A\rangle}{dt} = \frac{i}{\hbar}\langle [H,A]\rangle, \tag{3.3}$$

when A does not have an explicit time dependence [38, p. 103; 39, p. 180].

Applying equation (3.3) to position x and momentum p, yields

$$\begin{aligned} \frac{d\langle x\rangle}{dt} &= \frac{\langle p\rangle}{m}, & (3.4)\\ \frac{d\langle p\rangle}{dt} &= -\left\langle \frac{dV(x)}{dx} \right\rangle, & (3.5) \end{aligned}$$

respectively, where V is the potential energy function [38, p. 104; 39, p. 180–181]. Now, differentiating equation (3.4) once more and inserting (3.5) in the resulting equation yields

$$m\frac{d^2\langle x\rangle}{dt^2} = \frac{d\langle p\rangle}{dt} = -\left\langle \frac{dV(x)}{dx} \right\rangle, \tag{3.6}$$

which already looks similar to Newton's second law since for a conservative force

$F = -dV(x)/dx$ [38, p. 105]. However, in order to approximate

$$\left\langle \frac{dV(x)}{dx} \right\rangle \approx \frac{dV(\langle x \rangle)}{d\langle x \rangle}, \tag{3.7}$$

it is required that the potential energy varies slowly and $(\Delta x^2) = \langle (x - \langle x \rangle)^2 \rangle$ is small [38, p. 105]. Then the Taylor expansion for the force becomes

$$\begin{aligned} F(x) &= F(\langle x \rangle) + (x - \langle x \rangle)F'(\langle x \rangle) + \frac{(x - \langle x \rangle)^2}{2!}F''(\langle x \rangle) + \ldots, \\ &\approx F(\langle x \rangle) \end{aligned} \tag{3.8}$$

and so equation (3.6) takes the form

$$m\frac{d^2\langle x \rangle}{dt^2} = -\frac{dV(\langle x \rangle)}{d\langle x \rangle}, \tag{3.9}$$

which is Newton's second law [38, p. 105]. This approximation is valid even for some sharply localized subatomic particles [38, p. 105], allowing us therefore to assume a classical behaviour for a molecular many-body system. Nevertheless, there are still some exceptions such as the light hydrogen atoms that cannot be properly described by this approximation [11, p. 3]. We reserve a comment on that in section 3.8.

3.2 Force Field

In order to obtain the trajectory for multiple atoms, we need to know which forces affect their motion and how much. These forces are essentially defined by the potential function. A force field (FF) consists of a set of potential energy terms and their parameters [11, p. 113; 40].

Not all forces whose effects we see in everyday life are relevant in the nanoscale. For instance, gravity has very little effect on atoms [32, p. 15]. Since one atomic unit is of the order of 10^{-27} kg, you can imagine that the gravitational force is also scaled down by that order of magnitude, becoming negligible with respect to other forces. In contrast, electrostatic forces, like those described by Coulomb's law and van der Waals forces pose a considerable effect.

The terms of a potential function can be divided based on whether they describe bonded or non-bonded interactions [11, p. 67]. The total potential energy is the sum of the different terms

$$U_{tot} = U_{bonded} + U_{non-bonded}, \tag{3.10}$$

where both U_{bonded} and $U_{non-bonded}$ also consist of multiple integrated potential energy terms, described in sections 3.2.1 and 3.2.2 respectively. Additionally, force fields in general can be roughly divided into all-atom (AA) and coarse-grained (CG) [11, p. 113-117] as discussed in section 3.2.3. In this study we use the Chemistry at HARvard Molecular Mechanics 36 (CHARMM36) force field [41] for all-atom simulations and the Martini force field [42] for coarse-grained simulations.

3.2.1 Bonded Interactions

Bonded interactions are calculated on the basis of a fixed list of chemically bonded atoms [11, p. 73]. These interactions can be divided into bond stretching, bond angle bending, and dihedral angle twisting, which correspond to 2-body, 3-body, and 4-body interactions, respectively [11, p. 73; 40] as exhibited in figure 3.1.

The covalent bond between two atoms i and j is modelled as a spring which follows Hooke's law. Therefore, bond stretching is described by a harmonic potential of the form

$$U_{bond} = \sum_{bonds} \frac{1}{2} k_{ij,b} (r_{ij} - r_{ij,0})^2, \tag{3.11}$$

where $k_{ij,b}$ is the force constant, r_{ij} is the distance of the two atoms, and $r_{ij,0}$ is their reference distance [11, p. 73].

Bond angle refers to the angle formed by two bonds that connect three atoms i, j, and k together, as illustrated in figure 3.1 (b). The bond angle vibration is also modelled with a harmonic potential, as a function of the bond angle θ_{ijk}

$$U_{angle} = \sum_{angles} \frac{1}{2} k_{ijk,\theta} (\theta_{ijk} - \theta_{ijk,0})^2, \tag{3.12}$$

where $k_{ijk,\theta}$ is the force constant and $\theta_{ijk,0}$ the reference angle [11, p. 76]. The

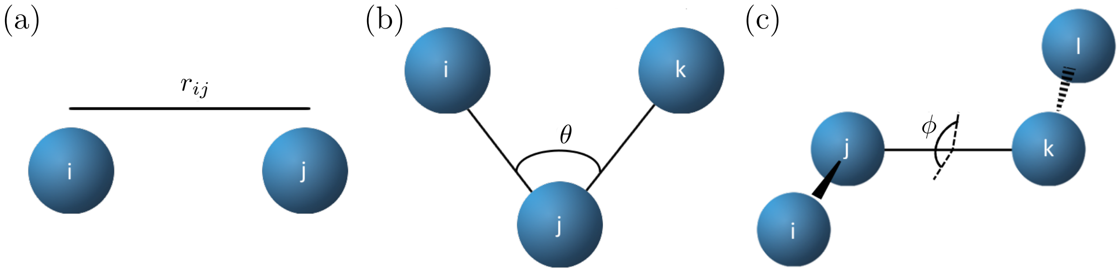

Figure 3.1 Models for (a) 2-body, (b) 3-body and (c) 4-body bonded interactions.

CHARMM36 force field additionally applies a Urey-Bradley potential term, which essentially is a harmonic correction term on the distance between the end-atoms i and k [43]. It enhances the accuracy in processing the vibrational spectra [41]. It is added to the overall potential function as

$$U_{UB} = \sum_{UB} \frac{1}{2} k_{ijk,UB} (r_{ik} - r_{ik,0})^2, \tag{3.13}$$

where $k_{ijk,UB}$ is again the force constant [11, p. 78; 40; 43].

Finally, the dihedral interactions are related to the three-dimensional (3D) arrangement of groups consisting of four atoms bonded to each other sequentially. They can be divided into improper and proper dihedrals.

Improper dihedrals are represented also harmonically by

$$U_{improper} = \sum_{impropers} \frac{1}{2} k_{\omega} (\omega - \omega_0)^2, \tag{3.14}$$

where k_ω is the force constant, ω the torsion angle and ω_0 its reference value [11, p. 79; 40]. They aim to preserve chirality or to prevent deformation of planar structures by bending [40, 41].

In contrast, proper dihedrals account for the normal dihedral interactions [11, p. 80]. Let us denote the dihedral angle between the planes ijk and jkl by ϕ, as marked in figure 3.1 (c). The angle ϕ is zero for *cis* conformation [11, p. 80]. The proper dihedrals are represented by the term

$$U_{dihedral} = \sum_{dihedrals} k_{\phi} (1 + \cos(n\phi - \delta)), \tag{3.15}$$

where k_ϕ is the force constant, n is the periodicity, and δ defines the phase [41]. Furthermore, as a more recent addition, CHARMM36 force field includes also a CMAP-potential term

$$U_{CMAP} = \sum_{residues} u_{CMAP}(\Phi, \Psi), \tag{3.16}$$

which acts as a correction for protein backbone dihedral angles Φ and Ψ [43]. Thus its purpose is to enhance the conformational properties of protein backbones [43].

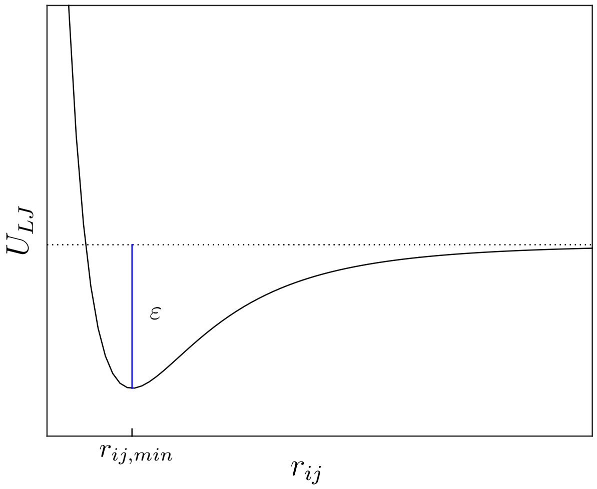

Figure 3.2 Lennard-Jones potential with respect to the distance r_{ij} between two atoms. The minimum potential of the magnitude -ε with respect to the zero level marked with the horizontal dotted line is achieved at $r_{ij,min}$.

3.2.2 Non-Bonded Interactions

The non-bonded interactions are calculated based on a regularly updated list of non-bonded atoms located within a designated distance, i.e., a neighbour list [11, p. 67; 37]. These interactions consist of Lennard-Jones (LJ) or the quite similar Buckingham interactions and Coulomb interactions [11, p. 67; 40]. Either of the first two can be used to provide a combined repulsion and dispersion term, while the latest denotes the Coulomb term [11, p. 67].

The Lennard-Jones potential, chosen in the CHARMM36 force field, describes the van der Waals interactions between a pair of atoms [40, 41]. As illustrated in figure 3.2, the repulsive and attractive forces cancel each other at a distance $r_{ij,min}$, at the bottom of the potential well where the Lennard-Jones force $F_{LJ} = 0$. Thus $r_{ij,min}$ is the distance they crave to settle at. For a system of multiple atoms, and hence multiple atom pairs, the total LJ potential is given by

$$U_{LJ} = \sum_{i<j} \varepsilon_{ij} \left[\left(\frac{r_{ij,min}}{r_{ij}} \right)^{12} - 2 \left(\frac{r_{ij,min}}{r_{ij}} \right)^{6} \right], \tag{3.17}$$

where the LJ potential between each atom pair ij is only calculated once [43, 44, p. 12]. The parameter ε_{ij} describes the depth of the potential well and r_{ij} is the distance between the two atoms i and j.

Coulomb's law describes the interaction between two charged particles. The force between the particles is repulsive if they have charges of the same sign and attractive with charges of opposite signs. The closer the particles, the greater the force. The total potential due to this interaction is described by

$$U_{Coulomb} = \sum_{i<j} k \frac{q_i q_j}{r_{ij}}, \tag{3.18}$$

where q_i and q_j are the charges of the atoms and k is Coulomb's constant [44, p. 12].

The LJ potential is inversely proportional to large powers of distance, thus being non-trivial only for the closest of atom pairs. Therefore, it is only computed for pairs within a certain cut-off distance from each other [43]. In contrast, the Coulomb interaction is non-negligible also at long distances, but the computation is speeded up by a special grid-based method named particle-mesh Ewald, which fundamentally utilizes the Fourier transformation [43]. Also, to avoid needlessly computing the interactions of bonded atoms multiple times through multiple force field terms, the chemically bonded atom pairs are excluded from the non-bonded potential calculation [43].

3.2.3 All-Atom versus Coarse-Grained Force Fields

In all-atom simulations the forces and equations of motion are solved for each atom explicitly – on every single time step. Obviously this results in considerably time-consuming computations when simulating biological molecules, such as proteins that consist of multiple amino acids, already containing several atoms. Therefore, they require days of computing even when using several computers simultaneously. In contrast, in coarse-grained simulations one can bundle groups of atoms into single centers of force or beads, calculate their effective interactions, and thus reduce the computational cost significantly [11, p. 116; 45].

For instance, in the Martini model four water molecules, each naturally consisting of three atoms, form a single water bead [42]. Consequently, if our system was to include N water molecules, i.e., $3N$ oxygen and hydrogen atoms combined, the Martini model would only require $\frac{N}{4}$ water beads. Naturally, both the atomistic and bead-based simulations require their own force fields, since the forces acting on one CG bead cannot be modelled with the forces acting on a single atom.

Nonetheless, albeit often producing results with comparable accuracy to atomistic models, CG models have limited chemical and spatial resolution [42, 45]. Coarse graining also affects the balance between entropy and enthalpy due to the reduced number of degrees of freedom [42, 46]. Essentially, although the Martini force field has the parameters to produce accurate free energies, the enthalpic and entropic contributions to free energies might be incorrect [42, 46, 47, p. 540].

CG simulations smoothen the energy landscape, resulting in enhanced sampling of the energy landscape in a given simulation time, which in turn speeds up the kinetics of the system [42, 46, 47, p. 541]. In the Martini model this hastens the simulation, with a semi-universal factor of about four with respect to all-atom simulations, yet this speed-up depends on the type of molecule and should be treated with care [42].

Selecting between an all-atom and a coarse-grained force field depends on the question we want to answer. For instance, an all-atom force field is required when looking into the trajectory of an amino-acid side-chain, while a coarse-grained model may be preferred when dealing with protein-protein interactions over a long timescale.

3.3 Energy Minimization

At the very beginning of the study, we need to construct the simulation system by adding water molecules and ions around the biomolecules under study. This assembly is not a result of actual interactions and may therefore cause the MD simulation to fail due to tremendous forces caused by atomic clashes [11, p. 2]. Hence an energy minimization is required. In this study the steepest descent method is used for this purpose.

The steepest descent method is a robust algorithm where, after calculating the forces, the new positions $\mathbf{r}$ for each atom are calculated according to

$$\mathbf{r}_{n+1} = \mathbf{r}_n + \frac{\mathbf{F}_n}{max(|\mathbf{F}_n|)} h_n, \tag{3.19}$$

where $max(|\mathbf{F}_n|)$ indicates the largest in magnitude force component and h_n is the maximum displacement the atom is allowed to make during one minimization step [11, p. 51]. The forces are calculated according to the potential function and the relevant parameters, both defined by the force field [11, p. 113].

This method does not take into consideration the previous minimization steps as it goes on, but it always takes steps in the direction of the negative potential gradient, i.e. the force [11, p. 6]. The calculation stops as the absolute value of the maximum force gradient becomes smaller than a specific predefined value or when the user-set number of minimization steps have been performed [11, p. 51]. The minimization is generally considered successful if it has stopped due to the firstly mentioned condition and if the potential energy is of the order of 10^{-5} kJ/mol or smaller.

The next stage is the also preparative equilibration process, i.e., relaxing the system in the given thermodynamic conditions. Finally comes the actual production run. Next we will introduce the routine implemented in both of these stages.

3.4 Molecular Dynamics Pseudo Code

Computationally the basic idea of MD is to apply Newton's equations of motion on a system of N particles

$$m_i \frac{\partial^2 \mathbf{r}_i}{\partial t^2} = \mathbf{F_i}, \tag{3.20}$$

where $i = 1, 2, ..., N$ is the particle index and $\mathbf{r}_i$ is the position in a 3D space [11, p. 2]. The forces follow the relation

$$\mathbf{F_i} = -\frac{\partial V}{\partial \mathbf{r}_i}, \tag{3.21}$$

where V is again the potential energy function [11, p. 2]. The equations are iterated over time to obtain the trajectory of the system.

To write a computer program based on this idea, we need to take into account plenty of things. Equations (3.20) and (3.21) together already require the masses of the atoms, the initial coordinates and velocities, a time step that will produce sufficiently accurate results, and the potential function to calculate the forces acting on each atom. Additionally, there are different algorithms [48] to generate new coordinates and velocities from the previous ones. Furthermore, we need to account for parameters like temperature and pressure and some technical details, such as the cut-off distance of the interactions.

The general flow of a simulation is sketched in figure 3.3. The initialization phase essentially defines the simulation system and simulation parameters. The initial coordinates are obtained from the biological structure of the molecules, but the initial velocities can be generated randomly by the program from the Maxwell-Boltzmann distribution at a specified temperature [11, p. 17], in accordance with the stochastic nature of atomic motion. The topology contains the atom and molecule characteristics as well as the force field [11, p. 17, 119]. In addition, the run parameters define for example the time step, the number of steps, and temperature [49].

After the initialization phase, we loop over the desired number of time steps, first calculating the forces and then solving the equations of motion for each step [11, p. 16]. As a result we get the coordinates and velocities at each time step, in other

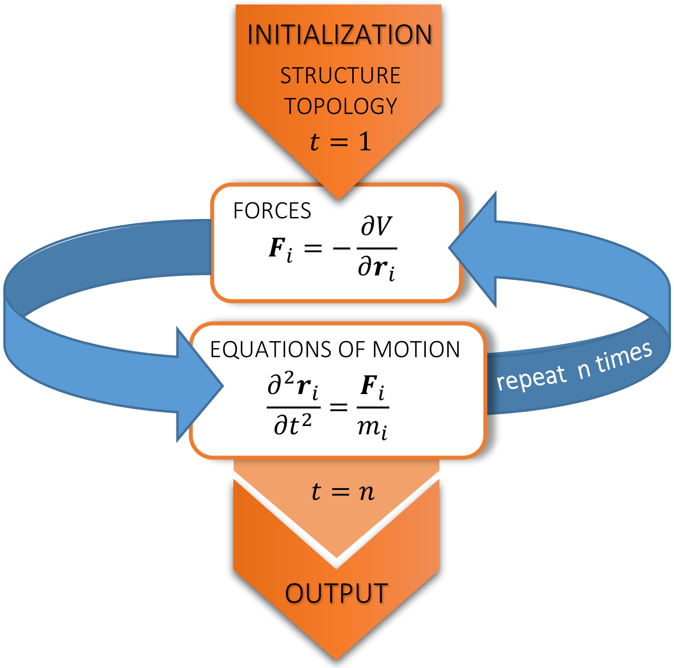
INITIALIZATION
STRUCTURE
TOPOLOGY
$t = 1$
FORCES
$\boldsymbol{F}_i = -\frac{\partial V}{\partial \boldsymbol{r}_i}$
EQUATIONS OF MOTION
$\frac{\partial^2 \boldsymbol{r}_i}{\partial t^2} = \frac{\boldsymbol{F}_i}{m_i}$
repeat n times
$t = n$
OUTPUT

Figure 3.3 General MD simulation flowchart, adapted from [11, p. 16]. The simulation is initialized with the information of the structure and topology. The loop (blue arrows) starts with the calculation of the forces, after which the conformation is updated by numerically solving the equations of motion. The output gives the new coordinates, velocities, and possible additional values.

words the trajectory.

The first essential requirement for an MD simulation is the form of the potential energy function, defined by the FF as discussed in section 3.2. Next, we will converse on the additional features required in the equilibration and actual simulation.

3.5 Temperature and Pressure Coupling

Naturally, the effects of ambient conditions must be considered. When simulating biomolecules one often wants the system to have somewhat the same temperature and pressure as *in vivo* – in order to get relevant results.

When performing a simulation in the canonical ensemble (NVT), the number of particles N, volume V, and temperature T are kept constant. Alternatively if, instead of the volume, one keeps the pressure P constant, one samples in the isothermal-isobaric ensemble (NPT). There are some different approaches to implement either one of these conditions in simulations [11, p. 30–42]. Here, during equilibration, we use the Berendsen [50] or the velocity-rescaling algorithm [51] for temperature coupling and the firstly mentioned in pressure coupling, while during production we use the Nosé-Hoover temperature coupling [52, 53] and Parrinello-Rahman pressure coupling [54,55].

From the statistical mechanics viewpoint, a system can be subjected to a certain constant temperature by coupling it to an external heat bath of the desired temperature T_0 [10, p. 140]. The Berendsen algorithm imitates weak coupling to such a heat bath by correcting the temperature deviation from T_0 according to the equation

$$\frac{dT}{dt} = \frac{T_0 - T}{\tau_T}, \tag{3.22}$$

where τ_T is a time constant with which T decays exponentially [11, p. 31]. However, this algorithm reduces fluctuations of the kinetic energy and thus does not produce a proper canonical ensemble [11, p. 31]. Yet the error is proportional to $1/N$, so the effect for large systems is rather minute [11, p. 31]. It is also highly efficient in relaxing a system to the desired temperature, which makes it an excellent choice for equilibrating the system [11, p. 32].

Temperature coupling via velocity-rescaling is similar to the Berendsen algorithm, but the kinetic energy is rescaled according to

$$dK = (K_0 - K)\frac{dt}{\tau_T} + 2\sqrt{\frac{KK_0}{N_f}}\frac{dW}{\sqrt{\tau_T}}, \tag{3.23}$$

where K is the kinetic energy corresponding to T, K_0 respectively corresponding to T_0, N_f the number of degrees of freedom, and dW a Wiener process [11, p. 32]. This method produces a correct canonical ensemble [11, p. 32].

Similarly, Berendsen pressure coupling follows the equation

$$\frac{d\mathbf{P}}{dt} = \frac{\mathbf{P}_0 - \mathbf{P}}{\tau_p}, \tag{3.24}$$

where τ_p is the respective time constant for pressure as τ is for temperature [11, p. 36]. In practice the pressure is kept constant by scaling the coordinates and simulation box vectors (see section 3.6) at every pressure coupling step n_{PC} with a scaling matrix μ with components

$$\mu_{ij} = \delta_{ij} - \frac{n_{PC}\Delta t}{3\tau_p}\beta_{ij}[P_{ij,0} - P_{ij}(t)], \tag{3.25}$$

where β is the isothermal compressibility of the system [11, p. 36].

After energy minimization and equilibration, sampling in the correct canonical ensemble is essential for the production phase. For temperature this can be achieved by utilizing the Nosé-Hoover temperature coupling [11, p. 33]. In this algorithm an additional friction term is added to the equations of motion [11, p. 33]. The friction force is proportional to each particle's velocity and to a friction parameter ξ, which has its own momentum p_ξ [11, p. 33]. Thus, the extra term inserted in the equations of motion is of the form

$$\frac{\mathbf{F}_{i,NH}}{m_i} = -\frac{p_\xi}{Q}\frac{d\mathbf{r}_i}{dt}, \tag{3.26}$$

where Q is the mass parameter of the heat bath [11, p. 33]. The parameter ξ also has an equation of motion of its own

$$\frac{dp_\xi}{dt} = (T - T_0), \tag{3.27}$$

where T is the temperature of the current calculation step [11, p. 33].

It is also important to note that although the Nosé-Hoover thermostat produces a proper canonical ensemble, it results in an oscillatory relaxation of the system [11, p. 33]. In other words the equilibration with this algorithm would take much longer time, which is why we choose the Berendsen algorithm in the equilibration phase.

Finally, the Parrinello-Rahman barostat – similarly to the Nosé-Hoover thermostat – also adds an extra term to the equations of motion

$$\frac{\mathbf{F}_{i,PR}}{m_i} = -\mathbf{M}\frac{d\mathbf{r}_i}{dt}, \tag{3.28}$$

where

$$M = \mathbf{b}^{-1}\left[\mathbf{b}\frac{d\mathbf{b}'}{dt} + \frac{d\mathbf{b}}{dt}\mathbf{b}'\right]\mathbf{b}'^{-1}, \tag{3.29}$$

with $\mathbf{b}$ representing the simulation box vectors [11, p. 37]. The box vectors are rescaled following their own equation of motion

$$\frac{d\mathbf{b}^2}{dt^2} = V\mathbf{W}^{-1}\mathbf{b}'^{-1}(\mathbf{P} - \mathbf{P}_0), \tag{3.30}$$

where $\mathbf{W}$ is the mass parameter matrix and $\mathbf{P}$ the pressure matrix at the current calculation step [11, p. 37]. Similarly as for the Nosé-Hoover algorithm, the Parrinello-Rahman pressure coupling produces the correct NPT ensemble while the Berendsen barostat is much faster in equilibrating the system [11, p. 37].

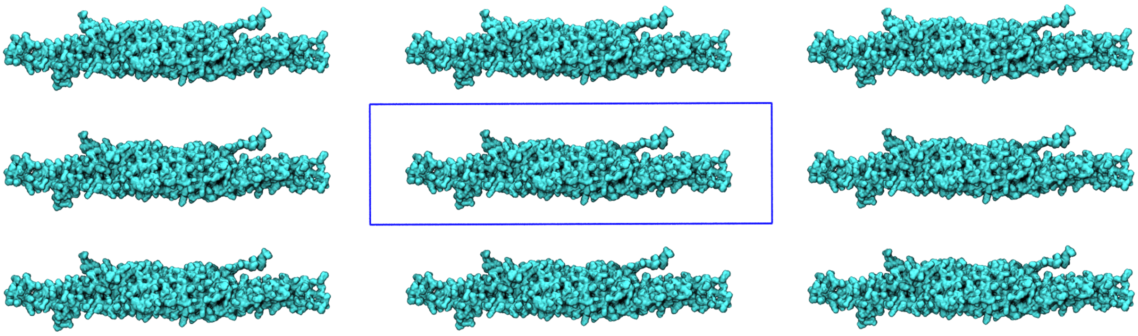

Figure 3.4 A two-dimensional illustration of periodic boundary conditions. The original simulation box is framed in blue and surrounded by eight periodic images.

3.6 Periodic Boundary Conditions

A simulation system consists of a finite number of particles placed inside a designated volume, known as the simulation box. This box is surrounded by endless copies of itself in all directions, making the system pseudo infinite. Periodic boundary conditions (PBCs) determine how the system interacts with its periodic images. If a particle was to cross a boundary, it would pop back in from the opposite boundary immediately. The interactions with the periodic images are limited by the minimum-image convention, which requires that only one image of each atom can be considered for the interactions [10, p. 40; 11, p. 5]. Thus, the cut-off radius of the forces cannot be larger than half the box size [11, p. 5]. Figure 3.4 shows an illustration of periodic boundary conditions.

There are also a number of box geometries to choose from [11, p. 13]. In this study we use the simplest, rectangular box.

3.7 Equations of Motion

The trajectories of the atoms are computed by iterating the equations of motion. The general form of the equations of motion presented in equation (3.20) can be rewritten in two first order equations

$$\frac{d\mathbf{r}_i}{dt} = \mathbf{v}_i, \tag{3.31}$$

$$\frac{d\mathbf{v}_i}{dt} = \frac{\mathbf{F}_i}{m_i}, \tag{3.32}$$

where in the latter we have divided the original equation by m_i [11, p. 16]. These equations can be numerically integrated via different algorithms to produce the dynamics of the system.

The velocity Verlet integrator, which we have chosen for our study, requires one set of initial velocities and positions at time $t = 0$ to compute their new values at $t + \Delta t$ [11, p. 27]. The equations are then iterated until we reach the desired simulation time. The equations for each time step i are

$$\begin{aligned} \mathbf{v}_i(t + \Delta t) &= \mathbf{v}_i(t) + \frac{\Delta t}{2m}[\mathbf{F}_i(t) + \mathbf{F}_i(t + \Delta t)], \quad (3.33) \\ \mathbf{r}_i(t + \Delta t) &= \mathbf{r}_i(t) + \Delta t \mathbf{v}_i(t) + \frac{\Delta t^2}{2m}\mathbf{F}_i(t), \quad (3.34) \end{aligned}$$

as can be derived from the Taylor expansions of equations (3.31) and (3.32) [11, p. 27].

3.8 Limitations

To correctly utilize a method and the results it produces one must always acknowledge its limitations. The limitations mostly arise from the computational implementations and fundamental numerical nature of MD simulations.

Although in section 3.1 we saw that the molecular behaviour even for subatomic particles can be assumed as classical, this estimation can be problematic especially with light-weighted hydrogens and high frequency vibrations of covalent bonds [11, p. 3]. The statistical mechanics of their vibrational motion becomes very different as their frequency exceeds $k_B T/h$ [11, p. 3]. GROMACS confronts this problem, for example, by making corrections to the system's total internal energy [11, p. 3].

Some limitations also arise from the use of force fields. First, they are approximate, due to assuming the forces to be pair-additive [11, p. 4]. Secondly, they use cut-off radii which means that only the interactions with the very nearest neighbours are accounted for [11, p. 4–5]. Finally, they do not consider the electronic motions, so electrons are left on their ground state which can bias the interactions [11, p. 4]. However, this motion is usually irrelevant for the questions approached via MD.

Periodic boundary conditions, too, produce artefacts. For instance, water molecules and ions behave periodically under periodic boundary conditions, which is unnatural [11, p. 5]. The impact of periodic boundary conditions must be carefully considered, yet the error is assumed to be significantly smaller than that of a system with unnatural boundaries with vacuum [11, p. 11].

Finally, let us consider the numerical nature of MD. In order to acquire accurate results with any numerical method, the time step Δt, over which the algorithm evaluates the development of position and velocity, needs to be sufficiently small.

For velocity Verlet the order of Δt should be femtoseconds (10^{-15} s), meaning that a simulation of one microsecond would take roughly 10^8–10^9 simulation steps. A simulation of such many steps will be computationally expensive.

4. ANALYSIS METHODS

In this study we aim to specify the mechanism employed by MIM IMD to deform membranes. To reach this goal, we analyse MD simulations, probing the effects of MIM IMD on lipid behaviour, investigating the effects of protein-lipid interactions on the domain itself and examining its flexibility.

In this chapter we present the analysis methodology employed in this thesis. The GROMACS package already offers a multitude of analysis tools to study MD simulations. In addition to the GROMACS tools specified in each section, we used Matlab [56] for complementary calculations and to plot figures.

4.1 Diffusion Coefficients for Lipids

The lateral diffusion of lipids was quantified via their diffusion coefficient D. For each system the diffusion coefficient for each lipid species was computed separately on both bilayer leaflets using the GROMACS tool `gmx msd`, which computes both the mean squared displacement (MSD) with respect to time and the diffusion coefficient. The mean squared displacement is defined as

$$MSD(t) = \langle (\mathbf{r}(t) - \mathbf{r}(t_0))^2 \rangle, \tag{4.1}$$

where $\mathbf{r}(t)$ is the particle position at time t and $\mathbf{r}(t_0)$ respectively at t_0 [57]. The diffusion coefficient D can be derived from the mean squared displacement via a linear fit

$$D = \lim_{t \to \infty} \frac{1}{2dt} MSD(t) \tag{4.2}$$

where d is the dimensionality, e.g., 2 in two dimensions [57].

In our analysis, the mean squared displacement was calculated over the first 250 ns and the coefficients were fitted starting from the lag time of 25 ns to avoid subdiffusion effects [58]. Upon computing, the lateral center of mass movement of the leaflet was removed. We also confirmed that the diffusion had saturated by plotting the mean square displacement with respect to time intervals.

For simulations of a lipid bilayer without protein, we calculated the mean D for each lipid species by averaging the values of the two leaflets. This allowed us to enhance the statistics. The uncertainties for the mean values were obtained using the formula for the propagation of uncertainty

$$\Delta z = \sqrt{\left(\frac{\partial f}{\partial x}\right)^2 (\Delta x)^2 + \left(\frac{\partial f}{\partial y}\right)^2 (\Delta y)^2}, \tag{4.3}$$

where f is the arithmetic mean $\frac{x+y}{2}$, and Δx and Δy are the uncertainties for two measured values x and y, respectively [59]. Let us denote the uncertainties for the two computed D as ΔD_1 and ΔD_2. Thus, we obtain

$$\Delta D = \frac{1}{2}\sqrt{(\Delta D_1)^2 + (\Delta D_2)^2} \tag{4.4}$$

for the uncertainty of the mean value.

4.2 Area Per Lipid

Area per lipid is a measure of lipid bilayer density. In this study we used it to gain extra insight into the behaviour of lipids. The area per lipid is defined using the box vectors $X(t)$ and $Y(t)$, in x and y dimensions respectively, as

$$A_{lipid} = \frac{X(t) * Y(t)}{N/2}, \tag{4.5}$$

where $N/2$ is half the total number of lipids in the system, i.e., approximately the number of lipids per monolayer.

For each system simulated in NPT conditions, the box vectors were first extracted with `gmx energy`, and then inserted into equation (4.5). For systems simulated in NVT conditions, the box vectors remain constant and consequently the area per lipid also remains the same throughout the simulation.

4.3 Radial Distribution Function

The radial distribution function (RDF) $g(\mathbf{r})$ describes the probability of finding a particle at a distance r from one or several reference particles. It is defined as the ratio between the probability of finding a particle in a given region and the analogous probability if the particles were uniformly distributed [60]. Mathematically it is defined

$$g(\mathbf{r}) = \frac{dN/N}{dV/V}, \tag{4.6}$$

where dN is the number of particles in a given region dV at position $\mathbf{r}$, and N the total number of particles in the total volume V [60]. RDF values higher than unity imply increased probability of finding the particle at distance r, while values below unity indicate decreased probability [60]. We used RDF to examine whether lipids of the same kind aggregate to form clusters.

We computed the RDFs in the plane of the lipid bilayer for each lipid species using the GROMACS tool `gmx rdf`. The distances between lipid molecules were calculated between their centers of mass. The RDFs were averaged starting from 20 ns till the end of the simulations studied.

4.4 Density Mapping

Density map also describes the probability of finding particles at specific locations, but instead of a curve, a map is plotted. We computed the two-dimensional (2D) number density maps for each lipid species to complement the results obtained via the RDFs and to quantify the lipid positions with respect to the protein domain. The basic idea is to compute the number of particles at each position at each time step and finally average the number of observations. As with the RDFs, the density maps were also calculated in the plane of the lipid bilayer, for individual leaflets.

We employed the GROMACS tool `gmx densmap` for this analysis. Before the actual calculation for simulations including a MIM IMD, we centered the protein domain and removed its center of mass movement and rotation by fitting it in the structure it had at the beginning of the trajectory. The resulting densities were divided by the number of molecules of the lipid species in question to have results comparable between species. The density maps were averaged over the trajectories, excluding the first 20 ns.

4.5 Surface Tension

Surface tension affects lipid behaviour within a membrane, and consequently may affect the protein-lipid interactions we study in this thesis. The surface tension can be calculated according to

$$\gamma(t) = \frac{L_z}{n}\left\{P_{zz}(t) - \frac{P_{xx}(t) + P_{yy}(t)}{2}\right\}, \tag{4.7}$$

where L_z is the height of the simulation box, n is the number of bilayer surfaces, and the P_{xx}, P_{yy} and P_{zz} are the pressure values at time t in x, y, and z dimensions respectively [11, p. 38].

The surface tension of the lipid bilayer was obtained by the GROMACS tool `gmx energy`. We averaged the obtained values over the whole trajectories.

4.6 Cluster Analysis

Cluster analysis is a data mining tool, which was employed here to probe the flexibility of MIM IMD in solution. In cluster analysis of a protein domain, its conformation at each time step is compared to conformations at different time steps and similar conformations are classified into the same cluster [61, p. 42]. One measure of similarity is the pair-wise root mean squared deviation (RMSD) of atom positions between two conformations [61, p. 42–43]. The RMSD is defined as

$$RMSD(t) = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(\mathbf{r}_i(t_1) - \mathbf{r}_i(t_2))^2}, \tag{4.8}$$

where N is the number of atoms used for the calculation, $\mathbf{r}_i(t_1)$ is the position of the ith atom at time t_1 and $\mathbf{r}_i(t_2)$ its position at t_2 [61, p. 39]. If the two structures are not within a certain cut-off distance from each other, they are classified into different clusters. Prior to the RMSD, all conformations were aligned on the first one to remove the center of mass movement.

Two different clustering algorithms were used: the leader algorithm and the gromos algorithm. The leader algorithm is sequence-dependent, reading through the trajectory only once. Every cluster has a leading case, *a leader*, which is the first member of the cluster [62, p. 75]. The analysis process comprises only a few fundamental steps [62, p. 75–76]. For a simulation with n time steps they are

1. **$\mathbf{t = 0}$**. Classify the conformation into the first cluster. Assign the conformation as the leading case of the first cluster.
2. **Set $\mathbf{t = t+1}$**. Compare to existing cluster leaders sequentially.
    (a) If the conformation is within the cut-off distance from the leading case

of some cluster, assign the conformation to that cluster.

(b) If the conformation is *not* within the cut-off distance from any leading case, assign the conformation as a leading case to a new cluster.

Repeat this step until all trajectory frames have been processed.

3. $\mathbf{t} = \mathbf{n}$. All frames have been classified into a cluster.

This computation was performed in R [63].

The results obtained by leader algorithm were fitted into the equation

$$N = N_{max}(1 - e^{-\frac{t}{\tau}}), \tag{4.9}$$

where N_{max} is the maximum number of clusters and τ a characteristic time constant. The fitting was performed with the Matlab function `lsqcurvefit`, which outputs an estimate of both N_{max} and τ.

In the gromos algorithm the RMSD is calculated for each conformation with respect to all other conformations [64]. Then the number of conformations within the cut-off distance, i.e. the number of neighbours is calculated for each conformation. The conformation with the most neighbours is chosen as the center of a cluster, and the cluster comprises its neighbours as well [64]. These structures are then eliminated from the pool of conformations and the procedure is applied to the rest of the conformations until the pool is empty [64]. Cluster analysis by the gromos algorithm was performed with the GROMACS tool `gmx cluster`.

4.7 Principal Component Analysis

Principal component analysis (PCA) filters the domain's largest collective modes of fluctuation, *the principal components*, from the trajectory given as input [65]. Principal component analysis was performed to complement the results gained through cluster analysis. With PCA we can inspect protein flexibility also on a lipid bilayer.

Covariance describes the linear relationship between two variables, in this case atom positions. Its sign corresponds to the sign of the slope of a linear equation describing the relationship [66]. We used the GROMACS tool `gmx covar` to find MIM IMD's eigenvectors of the covariance matrix for the trajectory of the domain backbone.

Subsequently, the GROMACS tool `gmx anaeig` was used to analyse the obtained eigenvectors. The `gmx anaeig` calculates the two most extreme projections along the trajectory. This tool was also used to interpolate 13 conformations between the two extremes. We only analysed the two largest eigenvalues, because the remaining eigenvalues were substantially smaller.

4.8 Lipid-binding Residues

The lipid-binding residues were determined by computing the minimum distances between the protein residues and the lipids. For each protein residue we calculated the minimum distance from any lipid of the selected group using the GROMACS tool `gmx mindist`.

We used only heavy atoms for both the protein residues and lipids to omit noise caused by the light hydrogen atoms. To omit random unstable interactions, we required that a protein residue would be in contact with any of the desired lipids for at least 10 ns in total. The cut-off distance for contacts was set to 3.2 Å, which is also the distance limit for hydrogen bonds [67].

4.9 Protein Insertion into the Lipid Bilayer

A protein residue was considered inserted, if at any time t it was found beneath the averaged z-coordinate of the lipid phosphorus atoms that were within a 1 nm radius of the residue in the plane of the membrane, i.e., in the xy-plane. An example is provided in figure 4.1.

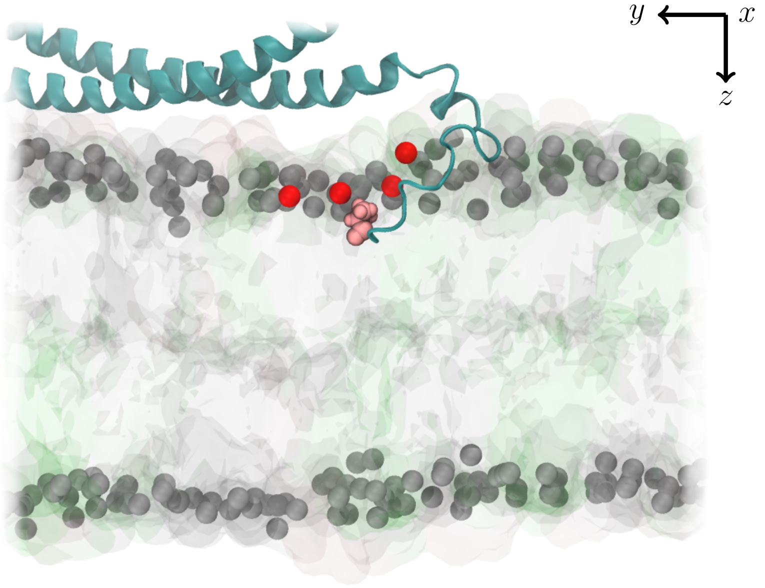

Figure 4.1 An example of an inserted protein residue, presented in pink on top of the cartoon representation of the protein. Phosphorus atoms within the 1 nm radius on the xy-plane are shown in red, while other phosphorus atoms are presented in gray.

We computed the z-dimensional distance between the center of mass of the protein residue and the averaged z-coordinate of the phosphorus atoms with respect to time. Finally, we averaged the distances over all inserting residues to get a pervasive idea of the insertion. The coordinates were extracted from the trajectories with the GROMACS tool `gmx traj`.

4.10 Membrane Curvature

The radius of membrane curvature was estimated at each time step by fitting a circle in the bilayer coordinates, using the Matlab function `circfit` [68]. We used only the phosphorus atoms of the leaflet which is in contact with the protein domain.

In our systems the curvature of a membrane is in the yz-plane, as seen in figure 4.2. Thus we only needed to assess the y and z coordinates of a bilayer. The fit was computed over averaged coordinates. For the averaging, the bilayer was divided into 5 Å wide segments along the y-dimension, and an average coordinate was computed for each of them. Due to the deviation of z-coordinates along the x-dimension, using the coordinates averaged over a high number of phosphorus atoms gives more reliable fits.

The general equation of a circle with center at (y_0,z_0) and radius r is

$$(y-y_0)^2+(z-z_0)^2 = r^2 \tag{4.10}$$

$$y^2+z^2-2y_0y-2z_0z+y_0^2+z_0^2-r^2 = 0. \tag{4.11}$$

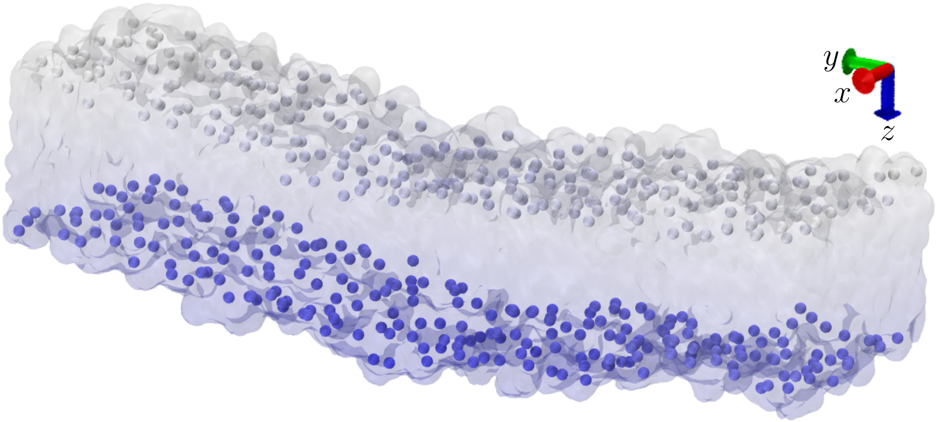

Figure 4.2 A bilayer with lipid phosphorus atoms presented as spheres. The coloring ranging from white to blue indicates the z-coordinate value, i.e., the coordinate along the axis normal to the bilayer.

This can be reformatted into

$$\begin{aligned} y^2 + z^2 + a_1 y + a_2 z + a_3 &= 0 && (4.12) \\ a_1 y + a_2 z + a_3 &= -(y^2 + z^2) && (4.13) \end{aligned}$$

where we have used the denotations

$$a_1 = -2y_0, \qquad a_2 = -2z_0, \qquad a_3 = y_0^2 + z_0^2 - r^2. \tag{4.14}$$

This form allows us to write the equation of a circle as a matrix equation

$$\begin{bmatrix} y_1 & z_1 & 1 \\ y_2 & z_2 & 1 \\ \vdots & \vdots & \vdots \\ y_N & z_N & 1 \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ a_3 \end{bmatrix} = \begin{bmatrix} -(y_1^2 + z_1^2) \\ -(y_2^2 + z_2^2) \\ \vdots \\ -(y_N^2 + z_N^2) \end{bmatrix}$$

for N atoms. Thus, by solving the equation (4.10) with respect to the vector $[a_1, a_2, a_3]^T$, we obtained the center coordinates and the radius of the fitted circle according to

$$y_0 = -\frac{1}{2}a_1, \qquad z_0 = -\frac{1}{2}a_2, \qquad r = \sqrt{y_0^2 + z_0^2 - a_3}. \tag{4.15}$$

These were obtained directly from the equation set (4.14).

A circle can be fitted to any set of coordinates with smaller or larger error. In order to disregard unreliable estimates – such as those for very straight or manifoldly bent bilayers – we set an error tolerance of $0.05r$. Hence, if any of the bilayer atoms were beyond this distance from the fitted circle, the fit was ignored.

5. SIMULATION MODELS OF MISSING-IN-METASTASIS

For a simulation system to properly model the physical phenomena of interest, we need to include the right components and optimize the simulation parameters. Simultaneously, the computational cost should be minimized. In biological simulations the functions of the biomolecules need to be considered in constructing the systems. Additionally, since biological systems comprise a vast number of atoms, the phenomena of interest might occur beyond the accessible timescale. In conjunction with the limitations of MD, this sometimes creates a need for innovative simulation systems.

In this chapter we introduce our simulation systems and the simulation parameters. All of our simulations were performed with the GROMACS software (versions 5.1.1 or 5.0.4). The computing resources for the purposes of this thesis were provided by the Finnish IT Center for Science (CSC) and Tampere Center for Scientific Computing (TCSC).

5.1 Simulation Systems

We are only interested in examining MIM's N-terminal IMD. Additionally, simulating the whole protein would increase the system size to a large extent, making it computationally expensive. Thus, only the IMD is included in our systems. Basically, we study the behaviour of MIM IMD on lipid bilayers of different properties.

In our study we use MIM IMD with the protein data bank (PDB) identifier 2D1L, which has been isolated from the house mouse [4]. Its crystal structure has been determined by x-ray diffraction to a 1.85 Å resolution [4]. The missing residues, four at the C-terminus of one monomer and eight at that of the other, as well as residues ASP155–SER168, were reconstructed using MODELLER which predicts the three-dimensional structure of protein sequences via homology modelling [69].

In our models we use a symmetric lipid bilayer. Although *in vivo* the asymmetry of lipid bilayers is vital for their functionality [1, p. 574], it is not a crucial property when we are studying a protein that interacts primarily with just one bilayer leaflet. More importantly, we want to ensure that the lipid bilayer is free of intrinsic tension.

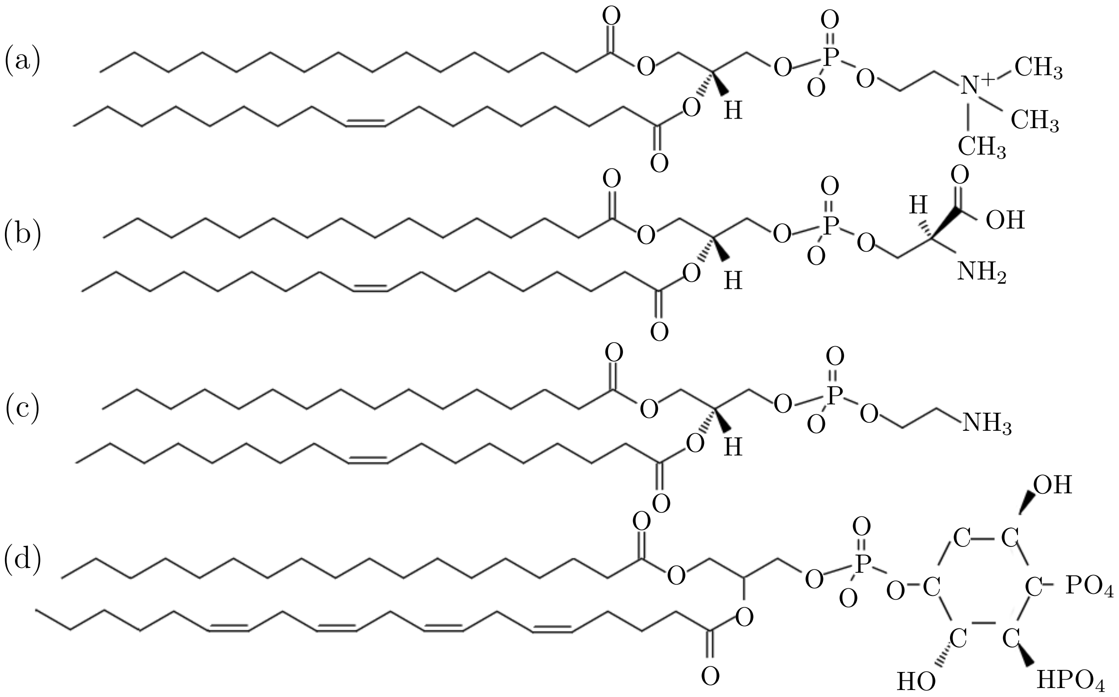

Figure 5.1 Molecular structure of (a) POPC, (b) POPS, (c) POPE, and (d) PI(4,5)P_2. Figure adapted from [71, 73].

Our bilayer mimics roughly the composition of the cytosolic side of the plasma membrane [13, p. 371; 70]. However, to have results comparable with experimental studies [5, 7], we omit cholesterol and sphingomyelin from our model. The bilayer was constructed with CHARMM-GUI for both all-atom and coarse-grained simulations [71, 72]. It is composed of 50 mol% 1-palmitoyl-2-oleoyl-*sn*-glycero-3-phosphocholine (POPC), 20 mol% 1-palmitoyl-2-oleoyl-*sn*-glycero-3-phosphoserine (POPS), 20 mol% 1-palmitoyl-2-oleoyl-*sn*-glycero-3-phosphoethanolamine (POPE) and 10 mol% PI(4,5)P_2 on both monolayers. The molecular lipid structures are illustrated in figure 5.1.

Periodic boundary conditions restrain the curvature of a bilayer, the generation of which is the most fundamental function of MIM IMD. Therefore, two kinds of bilayers were constructed and used, as illustrated in figure 5.2. First, a bilayer pseudo-infinite in two dimensions, affected by the strain imposed on it by PBCs (figure 5.2 (a)). Secondly, a free-edged bilayer, pseudo-infinite only in one dimension, whose curvature is not constrained by PBCs (figure 5.2 (b)). We additionally used tense bilayers.

When a symmetric bilayer is bent positively, the upper, convex, monolayer is stretched. Due to the increased area per lipid, it is easier for a protein to insert

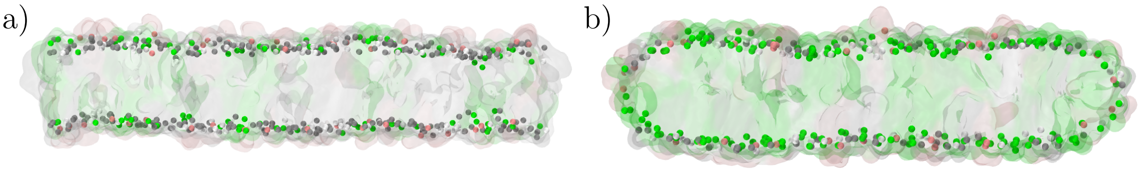

Figure 5.2 In our simulation systems we employed a) a bilayer pseudo-infinite in two dimensions and b) a free-edged bilayer pseudo-infinite in one dimension. The two dimensions in question are the one parallel to its longitudinal axis and the one in the direction perpendicular to the plane of the paper.

between the lipids. Therefore we model positively curved membranes with tense bilayers, where the increased area per lipid corresponds to the less densely packed lipid heads in the upper leaflet of a positively curved membrane. Using these bilayers we aim to probe MIM IMD's ability to sense positive curvature.

The most abundant ion inside a cell is the positive potassium ion (K^+) with a 140 mM concentration, while the most common negative ion is chloride (Cl^-) with a 5–15 mM concentration [13, p. 388]. To enhance the biological relevance of our studies, we include these prevalent ions into our systems. However, in CG simulations we use sodium instead of potassium, since no force field parameters are provided for potassium in the Martini force field.

We employ virtually three degrees of ion concentrations. First, the so called zero ion concentration, where ions are added only to neutralize the total charge of the system. Secondly, we have systems employing additionally the physiological 140 mM concentration of positive ions, with enough negative ions to neutralize the total charge. Finally, to amplify the effects the ions might have on the dynamics, we employ a few systems with excessive ion concentrations.

The simulation systems used in this study are gathered in table 5.1. The simulations can be divided into six groups with different purposes.

First, two all-atom simulations with different ion concentrations were performed for a bilayer in the absence of MIM IMD. These simulations provided us with reference to distinguish the consequences of protein-lipid interactions from lipid-lipid interactions. Secondly, we performed three AA simulations of a MIM IMD on bilayer, with different ion concentrations. Two of these systems had a free-edged bilayer (see figure 5.2 (b)). Subsequently, three coarse-grained simulations were performed with different constrains on the domain: one with an entirely constrained, rigid MIM IMD, another with an elastic network (EN) on the domain, which preserved its principal structure, and the third with no constrains whatsoever. The bilayer was again free-edged in these systems. We performed similar simulations also with

Table 5.1 Simulation systems used in this study. IMD refers to MIM IMD, unless specified explicitly. While one ion type indicates that positive ions have been added just enough to neutralize the total charge of the whole system, two ion types indicate that additional positive ions have been added with negative ones enough to again neutralize the system. The positive ion concentration $c_{pos.ion}$ represents the concentration of the additional positive ions. See figure 5.2 for bilayer types a and b.

System	**Ions**	$c_{pos.ion}$ **(M)**	**FF type**	**Ensemble**	**Bilayer type**	**Simulation time (ns)**
Bilayer	K	–	AA	NPT	a	500
Bilayer	K,Cl	0.21	AA	NPT	a	500
IMD on bilayer	K,Cl	0.14	AA	NPT	a	300
IMD on bilayer	K	–	AA	NPT	b	220
IMD on bilayer	K,Cl	0.21	AA	NPT	b	200
Rigid IMD on bilayer	Na,Cl	0.03	CG	NPT	b	100
IMD (EN) on bilayer	Na,Cl	0.03	CG	NPT	b	500
IMD on bilayer	Na,Cl	0.03	CG	NPT	b	1000
Rigid N-BAR on bilayer	Na,Cl	0.03	CG	NPT	b	150
N-BAR (EN) on bilayer	Na,Cl	0.03	CG	NPT	b	500
N-BAR on bilayer	Na,Cl	0.03	CG	NPT	b	500
IMD on bilayer	Na	–	CG	NVT	a	1000
IMD on tense bilayer	Na	–	CG	NVT	a	1000
IMD on tense bilayer	Na	–	CG	NVT	a	1000
IMD on bilayer[1]	K	0.14	AA	NPT	a	250
IMD in solution	K,Cl	0.006	AA	NPT	–	154
IRSp53 IMD in solution	Cl	–	AA	NPT	–	154
N-BAR in solution	K	–	AA	NPT	–	154

N-BAR domain, for reference.

Three CG simulations with different tensions on the bilayer were performed. The bilayer was prepared by simulating it under a predetermined lateral pressure to produce a surface tension on it. The applied pressure, resulting area per lipid and surface tension for each case are presented in table 5.2. Next a MIM IMD was incorporated into the system, which was then simulated for 1 μs. These systems were simulated exceptionally in NVT, to preserve the obtained membrane tension.

In one of the simulations the putative inserting parts of MIM IMD were initially forced inside the bilayer, to probe whether these sequences prefer to stay inserted and whether it induces any conformational changes in the domain.

Finally, we performed three simulations with a BAR domain in solution, where in addition to MIM IMD we observed the behaviour of the IMD of IRSp53 and N-BAR

---

[1] A simulation with the putative inserting parts forced inside the bilayer.

Table 5.2 The pressure applied laterally on the bilayer prior to simulating with MIM IMD and the corresponding area per lipid and average surface tension.

Applied pressure (bar)	**Area per lipid (nm^2)**	**Average surface tension (bar·nm)**
1	0.64	-53.23
500	0.85	397.18
1000	1.01	536.34

domain. Since the two latter needed some ions to neutralize the overall charge of the system, some ions were added with MIM IMD as well. These simulations primarily provided information about the flexibility of BAR domains with respect to each other, but also information on how the behaviour of MIM IMD changes when in contact with a bilayer.

5.2 Simulation Parameters

All simulations were performed in a 310 K temperature, most in NPT conditions and three in NVT to maintain incresed membrane tension. The force field of choice in all-atom simulations was CHARMM36 and in coarse-grained simulations the Martini force field.

Each system went through energy minimization and equilibration before the production run. In energy minimization we used the steepest descent algorithm.

In the equilibration phase of all-atom simulations, the Berendsen algorithm was used for both pressure and temperature coupling. Subsequently, in the production phase the Parrinello-Rahman pressure coupling and Nosé-Hoover temperature coupling were chosen. The hydrogen-bonds were constrained with the linear constraint solver (LINCS) algorithm [74], and a time step of 2 fs was used.

In equilibrating coarse-grained simulations, the Berendsen algorithm was again used for pressure coupling in NPT systems, while the v-rescale method was used instead for temperature coupling. In the subsequent production phase we used the same coupling methods. The time step in these systems was set to 20 fs.

For neighbour searching in AA simulations we used the Verlet cutoff-scheme, while in CG simulations the group cutoff-scheme was employed. In all simulations a cut-off distance of 1.2 nm was used for both Coulomb and van der Waals interactions.

Additionally, two different pressure coupling types were used in the systems simulated in NPT, depending on the bilayer type. With the free-edged bilayer pressure coupling was isotropic, while with the other type of bilayer it was semi-isotropic, which is the usual choice for membrane simulations.

6. RESULTS AND DISCUSSION

MIM is implicated in the *in vivo* and *in vitro* formation of protrusion like filopodia and lamellipodia [6, 7]. Experiments have revealed that the membrane deforming activity is driven by its N-terminal IMD [5, 24]. However, the exact mechanism of membrane curvature generation has remained elusive.

It has been argued that BAR domains induce membrane curvature via a combination of electrostatic interactions, shallow insertion of amphipathic helices and scaffolding [15]. Electrostatic interactions drive the association of the positively charged membrane-binding interface of a BAR domain with negatively charged bilayers. Thereafter, insertion or scaffolding may take place. Insertion increases the area of a lipid monolayer and, thus, promotes positive membrane curvature [22]. In scaffolding, on the other hand, the protein domain imposes its curvature on the membrane, providing it a mould. While insertion appears an unlikely mechanism for the negatively curving MIM IMD, its convex lipid-binding interface appears to fit the idea of scaffolding well.

Before MIM IMD can deform membranes, it has to localize to the correct binding sites. In addition to the positively charged amino acids attracted to negatively charged lipid headgroups, MIM IMD is suggested to employ additional mechanisms to adhere onto membranes [5]. *In vitro* experiments suggest that it inserts its N-terminal amphipathic helix among the acyl chains of membrane lipids [5]. The principal suggestion is that MIM IMD senses negative curvature through electrostatic interactions, while the amphipathic helix senses positively curved and flat membranes [5, 9]. However, the molecular details of the proposed curvature sensing mechanisms are still poorly understood.

In this chapter we probe the mechanisms of curvature generation and sensing employed by MIM IMD. Specifically, we inspect the flexibility of the domain, the protein-lipid interactions, the possibility of insertion and how these contribute to the curvature generation and sensing abilities of MIM IMD. For this purpose the analysis methods reviewed in chapter 4 were employed to study the MD simulations presented in chapter 5. VMD [75] was used for the reviewed snapshots of our systems.

6.1 Intrinsic Protein Curvature and Flexibility

The flexibility of a protein directly relates to the curvature it can generate, as well as the employed mechanism. BAR domains are presumed to be rather rigid [76]. Consequently, they are argued to induce membrane curvature by imposing their intrinsic curvature on the lipid bilayer. However, the argued ability of MIM IMD to sense positive curvature indicates a flexible structure.

MIM IMD has been discovered to induce tubules *in vitro* of roughly 60 nm in diameter [5], which corresponds approximately to an inner radius of 26 nm, when the average bilayer thickness is estimated to be 4 nm. In the following we examine whether a rigid MIM IMD might induce the curvature observed experimentally, and probe the level of its flexibility.

6.1.1 Intrinsic Curvature of the Crystal Structure

We begin by examining the intrinsic curvature of the crystal structure of MIM IMD. The radius of the intrinsic curvature was found to be roughly 39 nm, as illustrated in figure 6.1. The minimum diameter of a tubule induced by hypothetically rigid MIM IMDs, which bind to the tubule's inner surface, would then be expected to be about 86 nm. Thus, the crystal structure of MIM IMD does not possess a steep enough curvature to induce the tubules observed experimentally. This result suggests that MIM IMD is either rather flexible, or that multiple IMDs assemble in a way that enhances the sculpting effect.

Furthermore, IMDs have been found to induce observable curvature with a lower protein density as compared to N-BAR domains [29], whose crystal structure obviously has a steeper intrinsic curvature (see figure 2.3). This implies that the intrinsic curvature is not a sufficient indicator of membrane sculpting ability. What is more, it may imply that the IMDs do not need to assemble as densely as N-BAR domains to sufficiently accumulate the curving effect through assembly. To get a better idea of MIM IMD's flexibility, we move on to introduce some dynamics.

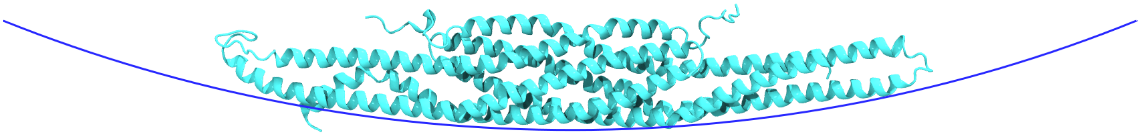

Figure 6.1 The intrinsic curvature of the MIM IMD crystal structure demonstrated by placing it on an arc of a blue circle whose radius of curvature is 39 nm.

6.1.2 MIM IMD is Remarkably Flexible

Cluster analysis was employed to evaluate MIM IMD's degree of flexibility. The analysis was performed on all-atom simulations in the absence of a lipid bilayer. For meaningful results, MIM IMD was compared with two additional BAR domains, the IMD of IRSp53 and an N-BAR domain. For cross validation both leader and gromos algorithms, discussed in section 4.6, were used with several cut-off distances.

In all cases, a clearly larger number of cluster centroids was found for MIM IMD than for the other two domains, as seen in table 6.1. This result implies that MIM IMD is the most flexible of the three, for the explored timescale.

Let us inspect more closely the results obtained with a 2.5 Å cut-off. Figure 6.2 displays the increase in the number of clusters versus time, as obtained via the leader algorithm. The result underlines the flexibility of MIM IMD as compared to the other two domains. Figure 6.3 additionally shows the cluster centroids found for MIM IMD via the gromos algorithm, revealing that its flexibility is mainly pronounced in its distal ends.

In order to complement these results, principal component analysis was performed on all-atom simulations, revealing MIM IMD's largest collective modes of fluctuation. First, PCA was performed on a MIM IMD in solution. The resulting principal components, displayed in figure 6.4 (a), demonstrate that the flexibility of the domain is localized at its distal ends, similarly as observed via cluster analysis. Note that the principal components exhibit greater or equal negative curvature as compared to the crystal structure.

We proceeded to inspect how the presence of a bilayer alters the behaviour of MIM IMD. The first principal component exhibits increased flexibility of the distal ends parallel to the bilayer normal, as seen in figure 6.4 (b). Interestingly, it also clearly displays both positive and negative curvature of the domain, although

Table 6.1 Number of cluster centroids found for the BAR domains of three proteins, via two clustering algorithms and several cut-offs.

		Cut-off				
Algorithm	**Protein**	**1.5**	**2.0**	**2.5**	**3.0**	**3.5**
Leader	MIM		109	35	16	9
	IRSp53		72	21	9	4
	N-BAR		57	18	9	5
Gromos	MIM	291	56	19		
	IRSp53	216	32	8		
	N-BAR	180	30	10		

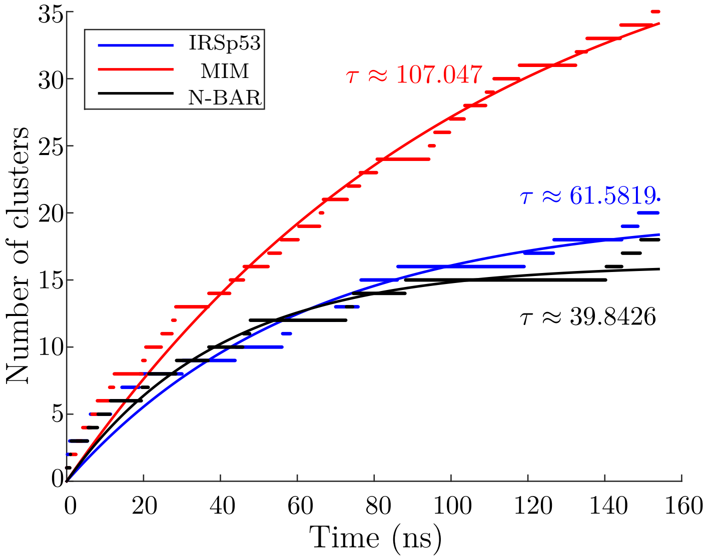

Figure 6.2 Number of clusters with respect to simulation time for three different BAR domains. The results have been fitted to an equation of the form $N = N_{max}(1 - e^{-t/\tau})$, and the respective τ are exhibited beside the curves. Here the clusters have been analysed using the leader algorithm with a 2.5 Å cut-off.

negative is more pronounced. This implies that MIM IMD can couple with positively curved membranes. Additionally in both components, as the domain acquires a more positively curved conformation, one of the putative inserting helices protrudes from the rest of the dimer, unlike in solution. Both components also show flexibility on the plane of the bilayer, which might be relevant to both protein-protein and protein-lipid interactions.

Finally, PCA was also performed on a MIM IMD on a free-edged bilayer, showing similar results as above on a pseudo-infinite one (data not shown). The second principal component here additionally demonstrates that the domain can also acquire an s-shape in the plane of the lipid bilayer. This conformation is also observed in CG simulations, as we will see in subsection 6.2.3.

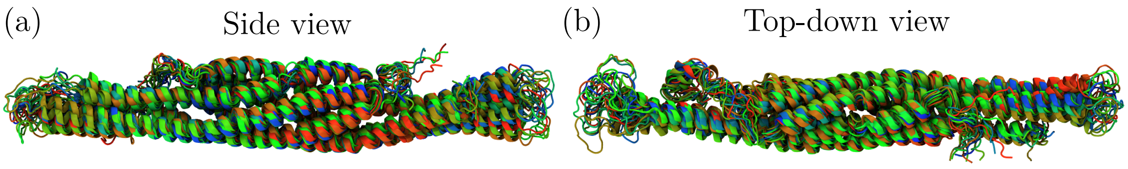

Figure 6.3 Cartoon representation of cluster centroids obtained with the gromos algorithm, using a 2.5 Å cut-off, viewed (a) from the side and (b) top-down. Each centroid is represented in a different colour.

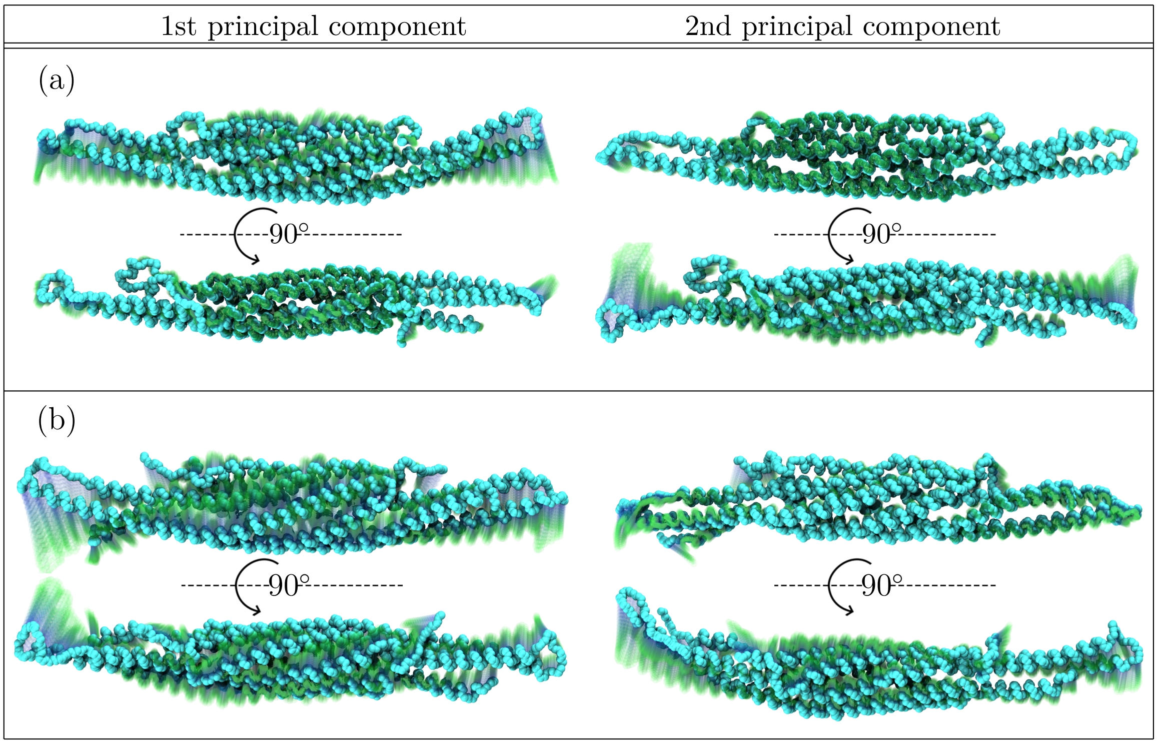

Figure 6.4 The trajectories of the two most common principal components, when MIM IMD was (a) in solution and (b) on a lipid bilayer. For each principal component, one of the extreme conformations is shown in cyan, while the rest of the mode trajectory is shown in faded colours ranging from blue to green. Each principal component is viewed both from the side and from above.

Overall, both cluster analysis and PCA support that MIM IMD is rather flexible, especially in comparison to other BAR domains. The positively curved conformations observed via PCA suggest its ability to couple to positively curved membranes, thus promoting its indicated ability to sense positive curvature. Strikingly, the putative inserting helix protrudes as the domain bends positively, implying that MIM IMD may sense positive curvature via insertion. In the negatively curved principal components, on the other hand, it does not protrude, so the localization onto negatively curved membranes is presumably driven by mere electrostatic interactions.

6.2 Membrane Curvature Generation by MIM IMD

While insertion is evidently unlikely to promote negative curvature, scaffolding and electrostatic interactions remain the likely membrane sculpting mechanisms in the case of MIM IMD. However, the extent to which these two mechanisms contribute to the membrane sculpting has remained an open question.

Membrane properties define how much energy is required to deform it. A change in these properties, like fluidity, might enhance or reduce the sculpting effect. Therefore, we also need to consider any indirect effects MIM IMD might impose on the lipid bilayer via protein-lipid interactions.

Furthermore, due to the lower concentration of MIM IMD required to produce an observable curvature as compared to N-BAR [29], the protein-protein interactions may not be as important in the case of MIM IMD. Thus it is possible that a single MIM IMD first generates local negative curvature, possibly luring other MIM IMDs to that binding site. The accumulation of local curvature would then result in the observable curvature. Hence, in this section we will focus on the membrane sculpting ability of a single MIM IMD.

We start this section by inspecting the effect MIM IMD has on lipids, along with specifying the residues controlling these protein-lipid interactions. Finally we study whether a single MIM IMD can generate negative curvature.

6.2.1 MIM IMD Slows Down Lipid Diffusion

To evaluate the level of protein-lipid interactions, we inspected the effect MIM IMD has on the diffusion of lipids. For a quantitative analysis, we determined the lateral diffusion coefficients D for each lipid species present in our systems.

To gain reference, we first determined the diffusion coefficients in two all-atom simulations containing only a bilayer, without MIM IMD. The two systems had different ion concentrations, so that we could also inspect the effect of ion concentration. The obtained diffusion coefficients are exhibited in appendix A.

The excess concentration of the positive potassium ions slows down the diffusion of each lipid species moderately. PI(4,5)P_2 seems to have the smallest diffusion coefficient in both systems, which is reasonable considering its large size and mass. Additionally, the diffusion coefficients of POPC and POPE obtained with high ion concentration are in good agreement with experimental values, found roughly in the span $(4-5)10^{-8}$ cm^2/s and $(5-5.5)10^{-8}$ cm^2/s respectively [77]. These experimental results imply that the D of POPE should rather be greater than that of POPC and not the other way around as in our systems. This deviation, however, should be attributed to finite size effect due to the reduced number of POPE.

After the control systems, the diffusion coefficients were computed for a bilayer which was in contact with a MIM IMD via its upper leaflet. In this system we had a physiological ion concentration of potassium (0.14 M). The results are displayed in figure 6.5.

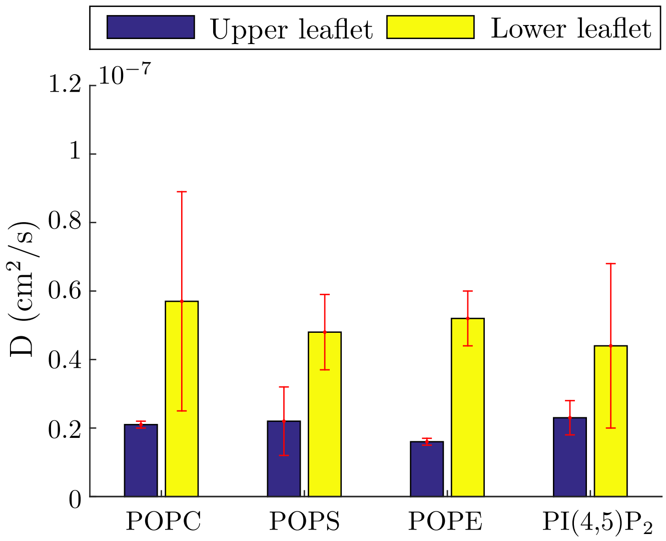

Figure 6.5 Diffusion coefficients in the presence of a MIM IMD for a pseudo-infinite lipid bilayer. The upper leaflet is in contact with the protein domain. The area per lipid is 0.61 nm^2.

Notice how the lower leaflet exhibits diffusion coefficients comparative to those observed in the absence of MIM IMD, since it is not in contact with the domain. In contrast, on the upper leaflet, MIM IMD clearly slows down the lipid diffusion, surpassing the effect of ion concentration. We can thus infer that the protein-lipid interactions are definitely significant.

6.2.2 Lipid-binding is Driven by Electrostatic Interactions

Knowing that the protein-lipid interactions are substantial, we naturally want to know which residues play the main role in them. Furthermore, electrostatic interactions are strongly implicated in the membrane-binding ability of MIM IMD, that binds to negatively charged membranes [4, 19, 24]. Therefore, we are especially interested in the residues binding to the most negatively charged component of our bilayers, PI(4,5)P$_2$. For this purpose an all-atom simulation of MIM IMD on a lipid bilayer was studied.

The resulting PI(4,5)P$_2$-binding residues are featured in figure 6.6 and total contact times in table 6.2. Additionally, the minimum distance each residue achieves with respect to any PI(4,5)P$_2$ is presented in appendix B.

The PI(4,5)P$_2$-binding residues can be divided into five zones, or three zones per monomer (see figure 6.6). However, due to insufficient sampling, the PI(4,5)P$_2$-binding residues are not exactly the same in both monomers. For instance we notice

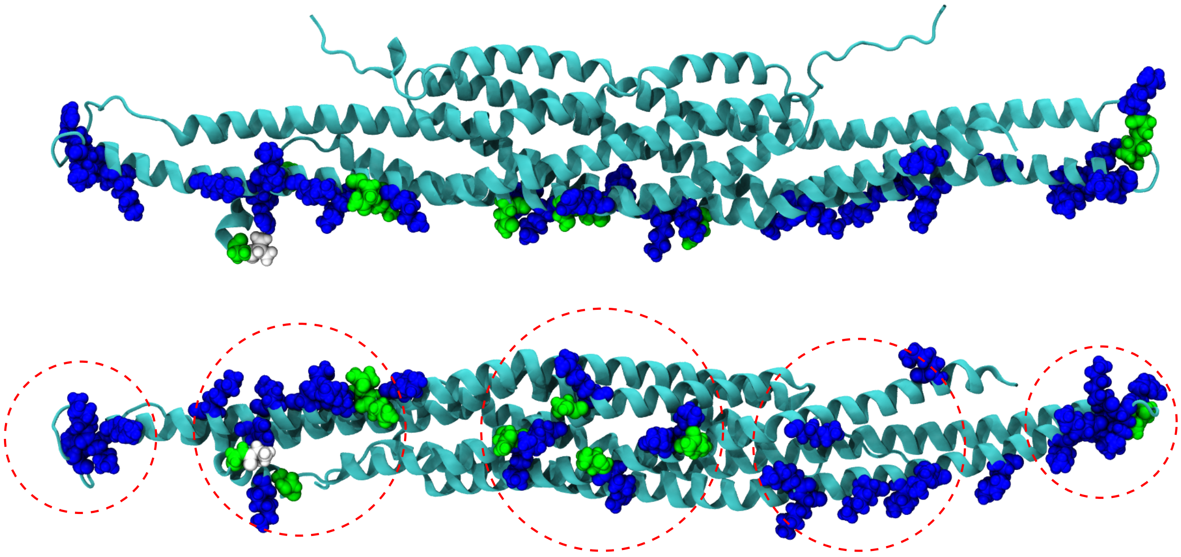

Figure 6.6 PI(4,5)$_2$-binding residues as viewed from the side (top) and from below (bottom). Blue colour represents positively charged amino acids (arginine (ARG), lysine (LYS)), white denotes hydrophobic (alanine (ALA)) and green indicates polar uncharged amino acids (serine (SER), threonine (THR), asparagine (ASN), glutamine (GLN)) and glycine (GLY). As indicated by the red dashed circles, the binding residues can be divided roughly into five zones.

that the N-termini exhibit different PI(4,5)P$_2$-binding residues, probably due to the downwards tilted orientation of the other.

The PI(4,5)P$_2$-binding residues are mostly positively charged, containing also some polar uncharged residues and one hydrophobic residue. As expected, these residues comprise no negatively charged amino acids. A significant 58% fraction of the positively charged PI(4,5)P$_2$-binding residues was bound to PI(4,5)P$_2$ for more than 50% of the simulation. Also the hydrophobic alanine is in contact with PI(4,5)P$_2$ for about 67% of the simulation time.

To uncover the role of PI(4,5)P$_2$-binding residues with respect to other lipid-binding residues, we defined protein residues bound to any lipid with the same requirements. The results reveal the whole lipid-binding interface, featured in appendix C. Additionally the minimum distance each residue achieves with respect to any lipid is presented in appendix B. We note that roughly 86% of all lipid-binding residues are either positively charged or uncharged but polar.

Importantly, we discover that the PI(4,5)P$_2$-binding amino acids comprise approximately 62% of all lipid-binding residues. Hence, the protein-PI(4,5)P$_2$ interactions are in main responsibility of the membrane-binding activity of MIM IMD.

Table 6.2 PI(4,5)P_2-binding residues and the total time they spent in contact with any PI(4,5)P_2 molecule $t_{PI(4,5)P_2}$. The total simulation time was 300 ns.

Residue	**Amino acid**	$t_{PI(4,5)P_2}$ **(ns)**
7	LYS	65.58
24	LYS	100.52
35	ASN	31.5
36	LYS	277.58
39	LYS	220.52
45	ARG	134.58
46	THR	128.04
116	LYS	92.76
117	LYS	196.66
124	LYS	245.14
128	LYS	54.1
131	LYS	29.14
139	LYS	26.18
146	LYS	154.28
149	LYS	212.64
150	LYS	148.92
152	LYS	193.62
153	LYS	199.04
157	GLN	18.44
158	GLY	13.28
159	ARG	234.56
251	ALA	203.02
252	GLY	23.02
260	LYS	241.8
263	SER	145.84
288	ASN	25.12
289	LYS	226.38
292	LYS	112.88
295	SER	56.92
298	ARG	285.84
370	LYS	180.98
373	ASN	13.74
374	GLN	53.66
377	LYS	239.44
381	LYS	167.66
385	LYS	103.42
387	ARG	186.4
392	LYS	137.64
402	LYS	216.18
405	LYS	216.32
406	LYS	12.32
412	ARG	209.14

Consequently, we infer that the protein-lipid interactions of MIM IMD are mainly electrostatic.

6.2.3 MIM IMD Induces Clustering of PI(4,5)P_2

The slowing down of lipid diffusion may indicate other changes in lipid behaviour induced by MIM IMD. *In vitro*, IMDs have been found to cluster PI(4,5)P_2 upon membrane binding [5]. Thus, our next step is to probe whether MIM IMD induces lipid clustering.

We studied whether any of the different lipid species form clusters in the presence of MIM IMD. CG simulations were employed to extend the accessible time scales. To quantify lipid clustering, radial distribution functions and density maps were calculated.

Both the RDF and density maps show clustering of PI(4,5)P_2, while the other lipid species show no such behaviour (see appendix D for density maps for each lipid species). From the lipid density map in figure 6.7 we see that the PI(4,5)P_2

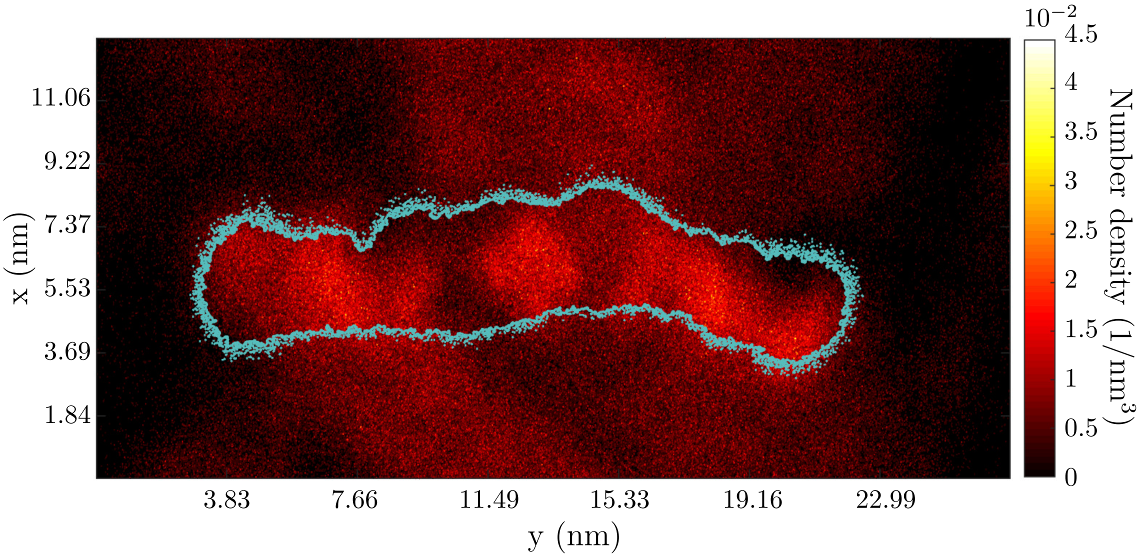

Figure 6.7 PI(4,5)P_2 density normalized with the number of PI(4,5)P_2-molecules. The outline of protein density is drawn in light blue to indicate the position of PI(4,5)P_2-clusters with respect to the protein.

clusters beneath nearly the whole domain. Especially, we notice that the cluster positions are in accordance with the PI(4,5)P_2-binding zones in figure 6.6. Thus the clustering is induced by the electrostatic interactions.

Furthermore, we studied the clustering of lipid species on tense membranes. The details regarding these systems are displayed in table 5.2. The resulting PI(4,5)P_2-density maps of the tense lipid bilayers are presented in figure 6.8. We note that on tense membranes MIM IMD seems to acquire an s-shape on the plane of the bilayer.

In these systems we notice the same behaviour as before: the PI(4,5)P_2-molecules cluster beneath the PI(4,5)P_2-binding residues, while the other lipid species do not cluster. However, the clustering effect is more pronounced on tense bilayers as revealed by RDF, reviewed in figure 6.9.

The amplification of PI(4,5)P_2-clustering with surface tension can be understood by considering this phenomenon in terms of bilayer density. In general, increasing the surface tension on a bilayer creates a larger area per lipid. With a sparser bilayer, the relatively large PI(4,5)P_2-molecules have more space to migrate near the PI(4,5)P_2-binding residues. This feature is reflected also in the increase in PI(4,5)P_2 diffusion coefficients with respect to surface tension, as displayed in figure 6.9 (b).

Yet, we do not observe any difference in the PI(4,5)P_2-clustering effect between the tensest bilayer and the other, less tense bilayer in figure 6.9. From the simulation trajectories we note that the domain is very positively curved on the high tension

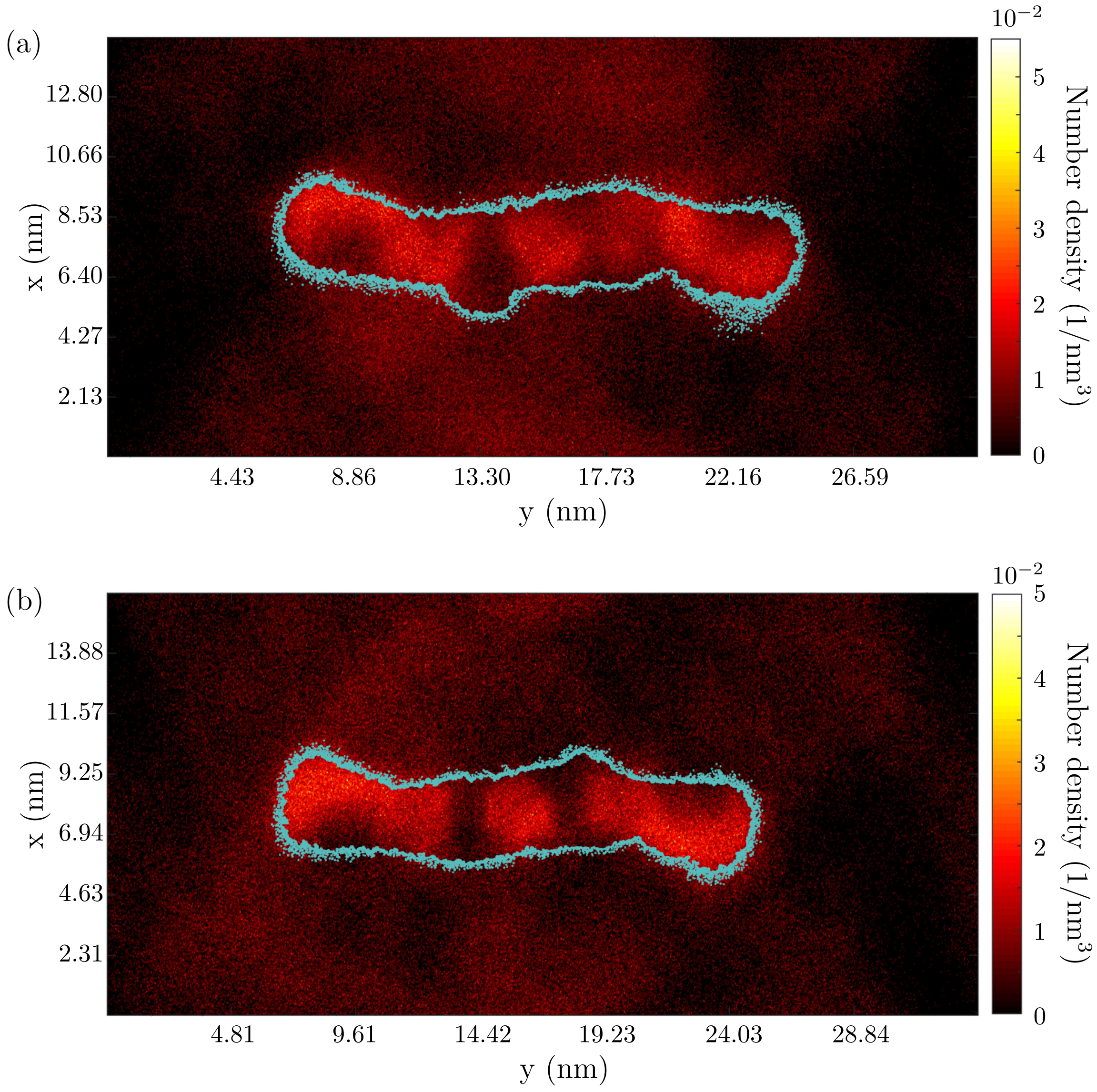

Figure 6.8 PI(4,5)P_2 density maps, normalized with the number of PI(4,5)P_2-molecules, for a system with an area per lipid of (a) 0.85 nm^2 and (b) 1.01 nm^2.

bilayers (see figure 6.12), effectively reducing the contact area. Therefore, speeding up the lipid diffusion while reducing the contact area causes that the clustering effect no more enhances after a certain surface tension.

Overall, the clustering of PI(4,5)P_2 may be vital for the membrane sculpting effect of MIM IMD. PI(4,5)P_2 has a polyunsaturated fatty acid tail, displaying many kinks due to the carbon-carbon double bonds [13, p. 369]. This refrains the lipids from packing together compactly, which in turn increases membrane fluidity [13, 78, p.

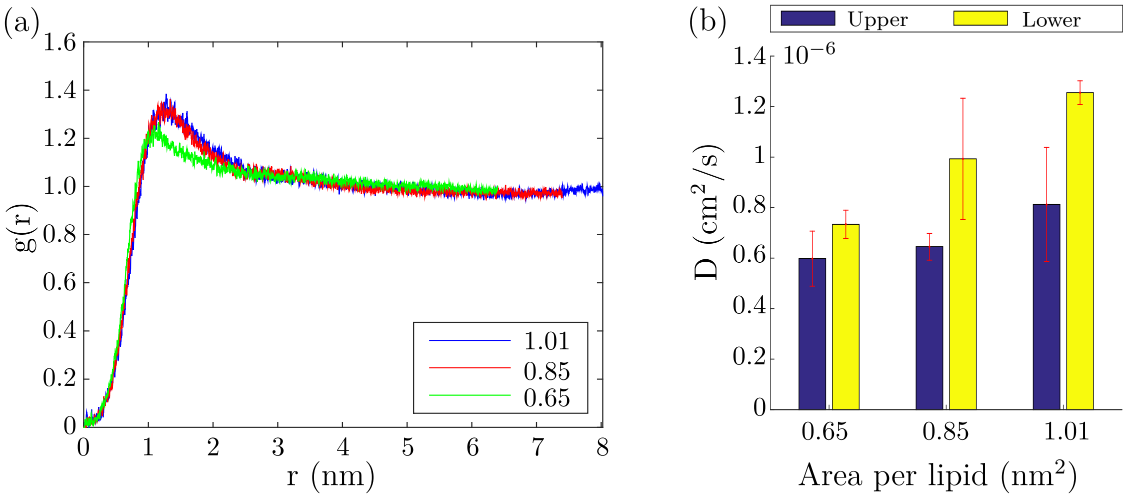

Figure 6.9 (a) RDFs $g(r)$ of the upper lipid monolayer PI(4,5)P$_2$. The curves are averaged over the time span 20–1000 ns for the three systems with different area per lipid, stated in the legends. We used a moving average of 5 to smooth the curves. (b) The diffusion coefficients of PI(4,5)P$_2$ for both leaflets, in each of the three systems. Notice that, due to the speeded up dynamics of CG simulations, these values are not comparable to those obtained from AA simulations. The upper leaflet is in contact with MIM IMD.

370]. Hence, PI(4,5)P$_2$-clustering could increase membrane fluidity locally, reducing the energy required to bend the membrane.

6.2.4 A Single MIM IMD Cannot Efficiently Induce Local Curvature

Finally for MIM IMD's membrane sculpting mechanism, we studied the local curvature induced by a single MIM IMD. For comparison, the local sculpting effect of the IMD of IRSp53 and an N-BAR domain were also examined. In these CG simulations, we employed a free-edged bilayer so that the periodic boundary conditions would not constrain the curvature.

The average radii of curvature to the desired direction, obtained with unconstrained domains, are displayed in figure 6.10. It is easy to see that the IMD of IRSp53 induces a steeper curvature than MIM IMD, which is in accordance to experiments [5]. Additionally, the IMD of IRSp53 is able to induce negative curvature for half of the time, while MIM IMD can do so for only a 26.9 % time fraction.

Comparing to the positively curving N-BAR domain, MIM IMD can induce slightly steeper curvature, but is again defeated in curvature duration. Additionally, we note that the protein-lipid interactions between the N-BAR domain and

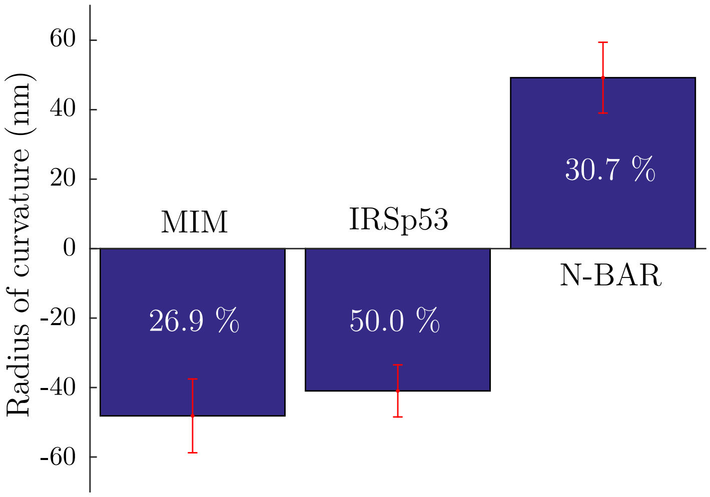

Figure 6.10 Average radius of curvature for three different unconstrained BAR domains: MIM, IRSp53 and N-BAR. The percentage indicates the time fraction when the bilayer was curved negatively (MIM and IRSp53) or positively (N-BAR), and the red bars indicate the standard deviation among those radii.

the bilayer were not sufficient to bend the bilayer, causing the domain to misbehave. However, N-BAR domains usually deform membranes by forming compact lattices of multiple domains [21]. Simulations with a rigid N-BAR domain, or with an elastic network applied on it, display steeper curvature of less than 40 nm radius, while similar simulations of MIM IMD show curvature above that radius (data not shown). Knowing that N-BAR domain is more rigid than MIM IMD (figure 6.10), we can conclude that a single N-BAR domain may be more efficient in inducing local curvature.

As an additional note, a constrained MIM IMD induces steeper curvature than an unconstrained one, and for longer times (data not shown). It is, however, noteworthy that the radius of curvature is above 40 nm in all cases and thus too mild to generate the experimentally observed curvature [5]. As discovered in subsection 6.1.2, MIM IMD is remarkably more flexible than the other two domains, so the fact that a single MIM IMD is unable to induce local curvature efficiently should not be surprising.

To conclude this section, a single MIM IMD cannot efficiently generate negative membrane curvature. On the other hand, BAR domains tend to work in greater groups to deform membranes [15,20,21], so to obtain a more throughout comprehension of MIM IMD's sculpting mechanism, multiple MIM IMDs should be studied, which is beyond the scope of this thesis.

6.3 Curvature Sensing Ability of MIM IMD

In addition to negatively curved membranes, MIM IMD is implicated in the sensing of both positively curved and flat membranes [5, 9]. In general, insertion of amphipathic helices has been suggested as a mechanism in sensing positive membrane curvature [22]. Also MIM IMD has been proposed to insert its N-terminal amphipathic helix into the membrane based on an experiment *in vitro* [5]. The experiment in question was conducted by studying the fluorescence anisotropy of membrane probe 1,6-diphenyl 1,3,5-hexatriene (DPH), which localizes into the hydrophobic core of a membrane [5]. A significant decrease in the DPH anisotropy was observed when the N-terminal helix, consisting of the first 11 residues of the polypeptide chain, was removed from the domain [5]. Consequently, this putative inserting helix is suggested to drive the positive curvature sensing of MIM IMD.

In subsection 6.1.2, PCA already revealed the ability of MIM IMD to acquire positively curved conformations and simultaneously protrude its putative inserting helix. On that account, we proceed to investigate the positive curvature sensing ability of MIM IMD. We begin by probing whether the putative inserting helix of MIM IMD inserts into the lipid bilayer and how deep. Then we proceed to investigate its behaviour on models of positively curved membranes.

6.3.1 MIM IMD Does Not Insert Its N-Terminal Helix

As far as protein insertion is concerned, we focus on MIM IMD's putative inserting helix. For reference, we compute the average insertion depth of the N-terminal helix of N-BAR domain, which has been confirmed to insert the bilayer [79]. More specifically, we use their hydrophobic residues, because they are the most likely ones to insert due to the hydrophobic effect. We inspect CG simulations in NPT conditions and employ free-edged bilayers to readily allow for the increase in monolayer area.

For both domains, figure 6.11 displays the distance with respect to lipid phosphorus heads, averaged over all hydrophobic residues that inserted at least once throughout the trajectory. The results clearly show that MIM IMD does not insert its N-terminal helix. Additionally, we note that the insertion times for the inserting residues of the N-BAR domain were hundreds of nanoseconds, while MIM IMD would insert its residues for at best 25 ns, which emphasizes the difference.

To elaborate, we also analysed all-atom simulations of MIM IMD on a free-edged bilayer, providing atomic resolution. Again, MIM IMD fails to insert its N-terminal helix. The insertion times are diminutive. Nevertheless, we remark the slightly less

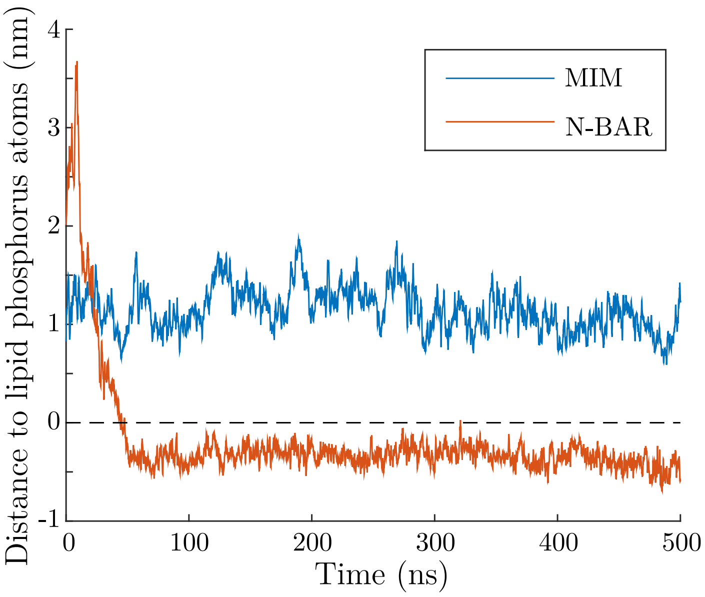

Figure 6.11 Average distance with respect to lipid phosphorus atoms observed for the hydrophobic residues of the putative inserting sequence in MIM and in N-BAR. The average is computed over the residues that inserted at least once. The dashed line represents the bilayer surface. We used a moving average of five points to smooth the curves.

transient insertion of the distal ends observed in these systems.

The simulations observed here do not reproduce the experimentally observed insertion. On the other hand, insertion has been implied in positive membrane curvature association, while in these simulation systems the lipid bilayer does not express any specific curvature. Consequently, we proceed to examine the behaviour of MIM IMD in the presence of a lipid bilayer modelling positive curvature.

6.3.2 MIM IMD May Couple with Positively Curved Bilayers by Acquiring a Positively Curved Conformation

Since insertion is primarily suggested as a mechanism for sensing positive membrane curvature, we naturally want to inspect how MIM IMD behaves on positively curved bilayers. The outer leaflet of a positively curved membrane is considered stretched, and the increased area per lipid makes insertion a reasonable curvature sensing mechanism. Hence, we model positively curved membranes with high tension bilayers, displaying an increased area per lipid. The details of the tense bilayers again remain visible in table 5.2.

Once again we observe no significant insertion of the N-terminal helix. Nevertheless, we do observe superficial insertion of some distal end residues. While with the

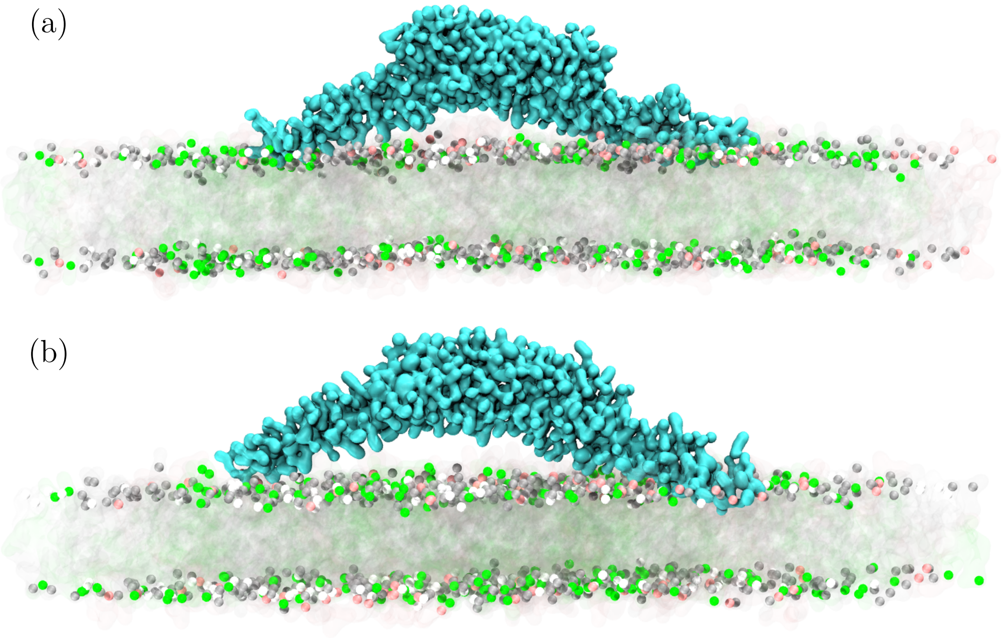

Figure 6.12 The final configuration of the system with an area per lipid of (a) 0.85 nm^2 and (b) 1.01 nm^2. The protein residues that were inserted at least once are shown in dark blue. On the lipid bilayer, the phosphorus atoms are shown in gray (POPC), white (POPE), green (POPS) and pink ($PI(4,5)P_2$).

lower surface tension only one distal end inserted for over 100 ns, with the higher surface tension both distal ends did so. However, they only insert among the lipid phosphorus heads (see appendix E), which is an insignificant depth comparing to what we observed with the N-BAR domain in figure 6.11.

Most importantly, MIM IMD immediately acquires, and maintains throughout the simulations, a positively curved conformation illustrated in figure 6.12. Complemented by our PCA results for a MIM IMD on a bilayer in subsection 6.1.2, this observation further supports that MIM IMD can sense positive membrane curvature.

Nonetheless, notice that here the bilayer is constrained by PBCs and the constant volume of the simulation box, which increase the required energy for membrane sculpting. The positive curvature acquired by MIM IMD suggests that upon binding to positively curved membranes it would first attempt to bring its lipid-binding residues in contact with the membrane by adapting to its shape. Whether it would then attempt to curve it negatively, cannot be said based on these simulations.

Our simulations of MIM IMD on both free-edged and tense lipid bilayers have not exhibited insertion. We proceed to test whether the putative inserting helix has a favourable conformation for insertion in the first place.

6.3.3 The N-termini of MIM IMD do not Prefer Insertion

MIM IMD might require a conformational change upon membrane-binding, to enable the insertion. However, the time-scales required to simulate conformational changes likely exceed the ones accessible with MD. Therefore, we probe whether the putative inserting helices even have a favourable conformation for insertion by performing a simulation with the N-termini forced inside the bilayer. If these sequences remain inserted, it is a hint that they may insert *in vivo*, too. Additionally, this approach may force the domain to show signs of the possible conformational change it experiences. The starting and ending configurations are illustrated in figure 6.13.

The inserted helices immediately start to retrieve back to the surface of the bilayer. As seen in the ending configuration in figure 6.13 (b), the inserted helices

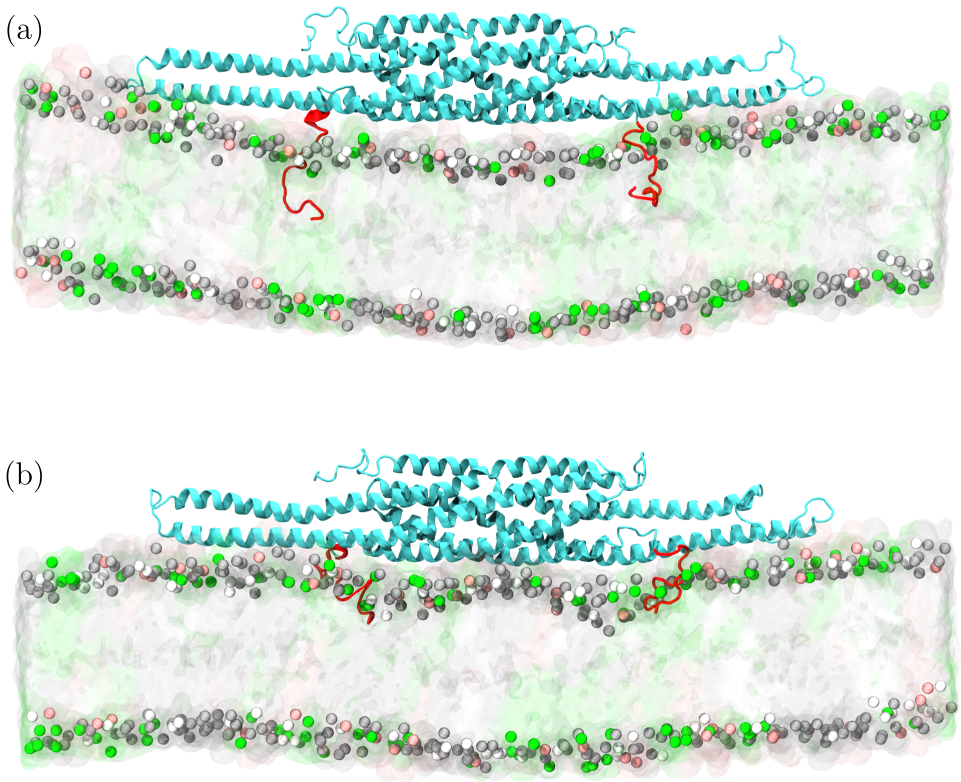

Figure 6.13 The (a) starting and (b) ending configuration of the simulation where the putative inserting helices are forced inside the bilayer. The sequences forced to insert inside the lipid bilayer are presented in red, while the rest of the domain is shown in cyan. The phosphorus atoms of the lipids are shown in gray (POPC), white (POPE), green (POPS), and pink (PI(4,5)P_2) to indicate the surface of the bilayer.

prefer to lie on the surface of the lipid polar heads rather than among the hydrophobic tails.

Regarding the starting conformation displayed in figure 6.13 (a), we note that the N-terminal helices have unfolded already during energy minimization. They only recover their helical conformation after withdrawing to the bilayer surface.

We additionally tried to pull the putative inserting helices inside the bilayer starting from an undisturbed configuration and then applying an external force on them. This, however, was unsuccessful since the domain started to unfold (data not shown).

There are many possible reasons as to why we do not inspect the insertion implied by experiments. On the one hand, coarse-grained simulations lack atomic resolution, so all-atom simulations of MIM IMD on a tense membrane should be studied. On the other hand, it is possible that the wild-type MIM IMD had experienced a conformational change in the DPH-anistropy experiment, undetectable in MD simulations due to restricted timescales, transforming the protein-lipid interactions altogether and hence reducing DPH-anisotropy. We also note that for a domain generating negative curvature it seems biologically unoptimized that it should penetrate the bilayer, since increasing the monolayer area promotes positive curvature. We leave these aspects to be studied in future research.

7. CONCLUSIONS

In this thesis we employed both atomistic and mesoscopic scale MD simulations to study the membrane sculpting mechanism of MIM. The membrane sculpting activity of MIM is driven by its N-terminal IMD [5, 7]. Consequently, we focused on the interactions between MIM IMD and membranes of different properties. A few simulations containing either MIM IMD or a lipid bilayer in solution were also performed for additional information and reference.

Virtually, three kinds of lipid bilayers were employed. First, we had a pseudo-infinite bilayer mimicking the large plasma membrane of a eukaryotic cell. Secondly, we used a free-edged bilayer, whose curvature is not influenced by periodic boundary conditions. Finally, we employed tense pseudo-infinite membranes, indirectly modelling the curvature of the convex upper leaflet of a positively bent membrane.

We began by inspecting the inherent flexibility of MIM IMD. The intrinsic curvature of the crystal structure of MIM IMD was found too mild to induce the membrane tubules observed in experiments [5]. Instead, cluster analysis and PCA revealed that MIM IMD is remarkably flexible as compared to other BAR domains.

Subsequently, we probed MIM IMD's membrane sculpting mechanism, which is the main interest of this thesis. First, by determining lipid diffusion coefficients both in the presence and in the absence of MIM IMD, we found that the protein-lipid interactions slowed down the lipid diffusion significantly. Inspecting lipid-binding residues, the PI(4,5)P_2-binding amino acids were found to play the main role in the protein-lipid interactions. Especially, the strong electrostatic interactions were discovered to induce PI(4,5)P_2-clustering beneath the protein domain. This is in accordance with experimental results [5].

PI(4,5)P_2-clustering could be an indirect but important means in the membrane sculpting mechanism of MIM IMD. Lipids with polyunsaturated tails promote membrane fluidity [13, p. 370]. Thus, the agglomeration of PI(4,5)P_2 could make the membrane locally more bendable. RDFs additionally implied the clustering effect to be enhanced on tense membranes, which suggests that PI(4,5)P_2-clustering could be important in deforming positively curved membranes. Usually, phosphoinositides are considered to promote positive curvature due to their inverted conical shape [1, 80, 81, p. 594], so the role of PI(4,5)P_2 in promoting the opposite curvature, suggested here, may open up new prospects for research.

Comparing the local membrane sculpting effect of MIM IMD, IMD of IRSp53, and N-BAR domain additionally showed that a single MIM IMD cannot effectively bend a membrane. On the other hand, BAR domains are known to work in larger groups upon bending membranes [15, 20, 21]. Hence in the future, it is important to study the sculpting effect of multiple MIM IMDs to gain more information about its curvature generation mechanism. It should be probed whether the MIM IMDs prefer to stay in close contact with each other or not, since it has been experimentally implied that MIM IMD promotes an observable sculpting effect with a clearly lower concentration as compared to N-BAR domains [29]. It will also be of great interest to observe the clustering effect of PI(4,5)P_2 in such studies.

Finally, we examined whether MIM IMD can sense positive curvature, as implied by *in vitro* experiments [5, 9]. Through the aforementioned PCA, we detected that MIM IMD can acquire a positively curved conformation. This was subsequently supported by simulations on tense membranes. This finding supports that MIM IMD can sense positive membrane curvature, since this conformation allows it to couple with the positive curvature.

Our simulations however, do not reveal whether MIM IMD is able to generate negative curvature on positively curved membranes. In the future it may be a challenge to set up a simulation system with a positively curved lipid bilayer whose curvature is not constrained by PBCs. However, it could reveal whether MIM IMD can also deform positively curved membranes, or whether it only senses the curvature.

We did not observe any significant insertion of MIM IMDs N-terminal helix, implied by experiments to insert the bilayer [5], nor any insertion of other parts of the protein matrix. The possibility of insertion should be further studied in all-atom resolution on tense membranes. These simulations could also clarify which interactions cause MIM IMD to acquire a positive conformation on tense membranes. If still no insertion is detected, one should scope other reasons possibly responsible for the experimental result. For instance, possible conformational changes in the tested altered MIM IMD, perhaps combined with the very superficial insertion of the distal ends which we observed on tense membranes.

Additionally, the crystal structure of MIM IMD has an asymmetrical conformation between the two monomers, which may not represent the natural *in vivo*-conformation of MIM IMD. Instead it might be a product of the methods used in defining the crystal structure. This should be considered in future simulations.

BIBLIOGRAPHY

[1] B. Alberts, A. Johnson, J. Lewis, M. D., R. M., R. K., and P. Walter, *Molecular Biology of the Cell.* Garland Science, 6 ed., 2015.

[2] Y. Lee, J. A. Macoska, S. Korenchuk, and K. J. Pienta, "MIM, a potential metastasis suppressor gene in bladder cancer," *Neoplasia*, vol. 4, pp. 291–294, 2002. doi:10.1038/sj.neo.7900231.

[3] K. D. Mertz, G. Pathria, C. Wagner, J. Saarikangas, A. Sboner, J. Romanov, M. Gschaider, F. Lenz, F. Neumann, W. Schreiner, M. Nemethova, A. Glassmann, P. Lappalainen, G. Stingl, J. V. Small, D. Fink, L. Chin, and S. N. Wagner, "MTSS1 is a metastasis driver in a subset of human melanomas," *Nature Communications*, vol. 5:3465, 2014. doi:10.1038/ncomms4465.

[4] S. H. Lee, F. Kerff, D. Chereau, F. Ferron, A. Klug, and R. Dominguez, "Structural basis for the actin-binding function of missing-in-metastasis," *Structure*, vol. 15, pp. 145–155, 2007. doi:10.1016/j.str.2006.12.005.

[5] J. Saarikangas, H. Zhao, A. Pykäläinen, P. Laurinmäki, P. K. Mattila, P. K. J. Kinnunen, S. J. Butcher, and P. Lappalainen, "Molecular mechanisms of membrane deformation by I-BAR domain proteins," *Current Biology*, vol. 19, pp. 95–107, 2009. doi:10.1016/j.cub.2008.12.029.

[6] J. A. Woodings, S. J. Sharp, and L. M. Machesky, "MIM-B, a putative metastasis suppressor protein, binds to actin and to protein tyrosine phosphatase δ," *Biochemical Journal*, vol. 371, pp. 463–471, 2003. doi:10.1042/BJ20021962.

[7] P. K. Mattila, A. Pykäläinen, J. Saarikangas, V. O. Paavilainen, H. Vihinen, E. Jokitalo, and P. Lappalainen, "Missing-in-metastasis and IRSp53 deform PI(4,5)P_2-rich membranes by an inverse BAR domain-like mechanism," *The Journal of Cell Biology*, vol. 176, pp. 953–964, 2007. doi:10.1083/jcb.200609176.

[8] J. Saarikangas, P. K. Mattila, V. M., M. Bovellan, J. Hakanen, J. Calzada-Wack, M. Tost, L. Jennen, B. Rathkolb, W. Hans, M. Horsch, M. E. Hyvönen, N. Perälä, H. Fuchs, V. Gailus-Durner, I. Esposito, E. Wolf, M. Hrabé de Angelis, M. J. Frilander, H. Savilahti, H. Sariola, K. Sainio, S. Lehtonen, J. Taipale, M. Salminen, and P. Lappalainen, "Missing-in-metastasis MIM/MTSS1 promotes actin assembly at intercellular junctions and is required for integrity of kidney epithelia," *Journal of Cell Science*, vol. 124, pp. 1245–1255, 2010. doi:10.1242/jcs.082610.

[9] M. P. Maddugoda, C. Stefani, D. Gonzalez-Rodriguez, J. Saarikangas, S. Torrino, S. Janel, P. Munro, A. Doye, F. Prodon, M. Aurrand-Lions, P. L. Goossens, F. Lafont, P. Bassereau, P. Lappalainen, F. Brochard, and E. Lemichez, “cAMP signaling by anthrax edema toxin induces transendothelial cell tunnels, which are resealed by MIM via arp2/3-driven actin polymerization,” *Cell Host & Microbe*, vol. 10, pp. 464–474, 2011. doi:10.1016/j.chom.2011.09.014.

[10] D. Frenkell and B. Smit, *Understanding Molecular Simulation: From Algorithms to Applications.* Academic Press, 2 ed., 2002.

[11] M. Abraham, D. van der Spoel, E. Lindahl, B. Hess, and the GROMACS development team, *GROMACS User Manual version 5.1.1.* 2015. www.gromacs.org.

[12] M. Karplus and J. A. McCammon, “Molecular dynamics simulations of biomolecules,” *Nature Structural Biology*, vol. 9, pp. 646–652, 2002. doi:10.1038/nsb0902-646.

[13] B. Alberts, D. Bray, K. Hopkin, A. Johnson, J. Lewis, M. Raff, K. Roberts, and P. Walter, *Essential Cell Biology.* Garland Science, 3 ed., 2010.

[14] E. Haug, O. Sand, Ø. V. Sjaastad, and K. C. Toverud, *Ihmisen fysiologia.* Sanoma Pro Oy, 1-5 ed., 2012.

[15] M. Simunovic, G. A. Voth, A. Callan-Jones, and P. Bassereau, “When physics takes over: BAR proteins and membrane curvature,” *Trends in Cell Biology*, vol. 25, pp. 780–792, 2015. doi:10.1016/j.tcb.2015.09.005.

[16] A. Arkhipov, Y. Yin, and K. Schulten, “Four-scale description of membrane sculpting by BAR domains,” *Biophysical Journal*, vol. 95, pp. 2806–2821, 2008. doi:10.1529/biophysj.108.132563.

[17] A. Suarez, T. Ueno, R. Huebner, J. M. McCaffery, and T. Inoue, “Bin/Amphiphysin/Rvs (BAR) family members bend membranes in cells,” *Scientific Reports*, vol. 4:4693, 2014. doi:10.1038/srep04693.

[18] K. Farsad, N. Ringstad, K. Takei, S. R. Floyd, K. Rose, and P. De Camilli, “Generation of high curvature membranes mediated by direct endophilin bilayer interactions,” *The Journal of Cell Biology*, vol. 155, pp. 193–200, 2001. doi:10.1083/jcb.200107075.

[19] B. J. Peter, H. M. Kent, I. G. Mills, Y. Vallis, P. J. G. Butler, P. R. Evans, and H. T. McMahon, “BAR domains as sensors of membrane curvature: the amphiphysin BAR structure,” *Science*, vol. 303, pp. 495–499, 2004. doi:10.1126/science.1092586.

[20] H. Yu and K. Schulten, "Membrane sculpting by F-BAR domains studied by molecular dynamics simulations," *PLoS Computational Biology*, vol. 9:e1002892, 2013. doi:10.1371/journal.pcbi.1002892.

[21] Y. Yin, A. Arkhipov, and K. Schulten, "Simulations of membrane tubulation by lattices of amphiphysin N-BAR domains," *Structure*, vol. 17, pp. 882–892, 2009. doi:10.1016/j.str.2009.03.016.

[22] G. Drin and B. Antonny, "Amphipathic helices and membrane curvature," *FEBS Letters*, vol. 584, pp. 1840–1847, 2009. doi:10.1016/j.febslet.2009.10.022.

[23] F. Campelo, G. Fabrikant, H. T. McMahon, and M. M. Kozlov, "Modeling membrane shaping by proteins: focus on EHD2 and N-BAR domains," *FEBS Letters*, vol. 584, pp. 1830–1839, 2009. doi:10.1016/j.febslet.2009.10.023.

[24] G. Scita, S. Confalonieri, P. Lappalainen, and S. Suetsugu, "IRSp53: crossing the road of membrane and actin dynamics in the formation of membrane protrusions," *Trends in Cell Biology*, vol. 18, pp. 52–60, 2007. doi:10.1016/j.tcb.2007.12.002.

[25] A. Yamagishi, M. Masuda, T. Ohki, H. Onishi, and N. Mochizuki, "A novel actin bundling/filopodium-forming domain conserved in insulin receptor tyrosine kinase substrate p53 and missing in metastasis protein," *The Journal of Biological Chemistry*, vol. 15, pp. 14929–14936, 2004. doi:10.1074/jbc.M309408200.

[26] L. Machesky and S. A. Johnston, "MIM: a multifunctional scaffold protein," *Journal of Molecular Medicine*, vol. 85, pp. 569–576, 2007. doi:10.1007/s00109-007-0207-0.

[27] P. K. Mattila, M. Salminen, T. Yamashiro, and P. Lappalainen, "Mouse MIM, a tissue-specific regulator of cytoskeletal dynamics, interacts with ATP-actin monomers through its C-terminal WH2 domain," *The Journal of Biological Chemistry*, vol. 278, pp. 8452–8459, 2003. doi:10.1074/jcb.M212113200.

[28] H. Zhao and P. Lappalainen, "A simple guide to biochemical approaches for analyzing protein-lipid interactions," *Molecular Biology of the Cell*, vol. 23, pp. 2823–2830, 2012. doi:10.1091/mbc.E11-07-0645.

[29] Z. Chen, Z. Shi, and T. Baumgart, "Regulation of membrane-shape transitions induced by I-BAR domains," *Biophysical Journal*, vol. 109, pp. 298–307, 2015. doi:10.1016/j.bpj.2015.06.010.

[30] M. Schemionek, B. Kharabi Masouleh, Y. Klaile, U. Krug, K. Hebestreit, C. Schubert, M. Dugas, T. Büchner, B. Wörmann, W. Hiddemann, W. E. Berdel, T. H. Brümmendorf, C. Müller-Tidow, and S. Koschmieder, "Identification of the adapter molecule MTSS1 as a potential oncogene-specific tumor suppressor in acute myeloid leukemia," *PLoS ONE*, vol. 10(5):e0125783, 2015. doi:10.1371/journal.pone.0125783.

[31] C. A. Callahan, T. Ofstad, L. Horng, J. K. Wang, H. H. Zhen, P. A. Coulombe, and A. E. Oro, "MIM/BEG4, a Sonic hedgehog-responsive gene that potentiates Gli-dependent transcription," *Genes & Development*, vol. 18, pp. 2724–2729, 2004. doi:10.1101/gad.1221804.

[32] B. Rogers, S. Pennathur, and J. Adams, *Nanotechnology: Understanding Small Systems.* CRC Press, 2 ed., 2011.

[33] D. Rapaport, *Art of Molecular Dynamics Simulation.* Cambridge University Press, 2 ed., 2004.

[34] T. Sclick, *Molecular Modeling and Simulation: An Interdisciplinary Guide.* Springer, 2 ed., 2010.

[35] R. O. Dror, M. Ø. Jensen, D. W. Brohani, and D. E. Shaw, "Exploring atomic resolution physiology on a femtosecond to millisecond timescale using molecular dynamics simulations," *The Journal of General Physiology*, vol. 135, pp. 555–562, 2010. doi:10.1085/jgp.200910373.

[36] M. J. Abraham, M. T., R. Schulz, S. Páll, J. C. Smith, B. Hess, and E. Lindahl, "GROMACS: High performance molecular simulations through multi-level parallelism fom laptops to supercomputers," *SoftwareX*, vol. 1–2, pp. 19–25, 2015. doi:10.1016/j.softx.2015.06.001.

[37] D. Van Der Spoel, B. Hess, G. Groenhof, and A. E. Mark, "GROMACS: fast, flexible, and free," *Journal of Computational Chemistry*, vol. 26, pp. 1701–1718, 2005. doi:10.1002/jcc.20291.

[38] S. Gasiorowicz, *Quantum Physics.* John Wiley & Sons, Inc., 3 ed., 2003.

[39] R. Shankar, *Principles of Quantum Mechanics.* Plenum Press, 2 ed., 1994.

[40] M. A. González, "Force fields and molecular dynamics simulations," *Collection SFN*, vol. 12, pp. 169–200, 2011. doi:10.1051/sfn/201112009.

[41] A. D. MacKerell Jr., B. Brooks, C. L. Brooks III, L. Nilsson, B. Roux, Y. Won, and M. Karplus, “CHARMM: the energy function and its parameterization,” in *Encyclopedia of Computational Chemistry* (P. von Ragué Schleyer *et al.*, eds.), vol. 1, pp. 271–277, John Wiley and Sons, Ltd., 2002.

[42] S. J. Marrink and D. P. Tieleman, “Perspective on the Martini model,” *Chemical Society Reviews*, vol. 42, pp. 6801–6822, 2013. Open Access Article.

[43] “CHARMM Tutorial: The Energy Function.” [WWW]. Available at `http://www.charmmtutorial.org/index.php/The_Energy_Function`. Accessed on March 28th, 2016.

[44] R. Schleif, “A concise guide to charmm and the analysis of protein structure and function.” [WWW], 2013. Available at `http://gene.bio.jhu.edu/book.pdf`. Accessed on August 26th, 2016.

[45] B. A. Merchant and J. D. Madura, “A review of coarse-grained molecular dynamics techniques to access extended spatial and temporal scales in biomolecular simulations,” *Annual Reports in Computational Chemistry*, vol. 7, pp. 67–87, 2011. doi:10.1016/B978-0-444-53835-2.00003-1.

[46] S. J. Marrink, H. J. Risselada, S. Yefimov, D. P. Tieleman, and A. H. de Vried, “The MARTINI force field: coarse grained model for biomolecular simulations,” *The Journal of Physical Chemistry*, vol. 111, pp. 7812–7824, 2007. doi:10.1021/jp071097f.

[47] X. Periole and S. J. Marrink, “The Martini coarse-grained force field,” in *Methods in molecular biology* (L. Monticelli and E. Salonen, eds.), vol. 924, pp. 533–565, Springer, 2013.

[48] N. Giordano and H. Nakanishi, *Computational Physics*. Pearson Education, Inc., 2 ed., 2006.

[49] “GROMACS Online Reference: mdp options.” [WWW]. Version 5.0.7. Available at `http://manual.gromacs.org/online/mdp_opt.html`. Accessed on March 17th, 2016.

[50] H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, A. DiNola, and J. R. Haak, “Molecular dynamics with coupling to an external bath,” *The Journal of Chemical Physics*, vol. 81, pp. 3684–3690, 1984. doi:10.1063/1.448118.

[51] G. Bussi, D. Donadio, and M. Parrinello, “Canonical sampling through velocity rescaling,” *The Journal of Chemical Physics*, vol. 126:014101, 2007. doi:10.1063/1.2408420.

[52] S. Nosé, “A molecular dynamics method for simulations in the canonical ensemble,” *Molecular Physics*, vol. 52, pp. 255–268, 1984. doi:10.1080/00268978400101201.

[53] W. G. Hoover, “Canonical dynamics: equilibrium phase-space distributions,” *Physical Review A*, vol. 31, pp. 1695–1697, 1985. doi:10.1080/PhysRevA.31.1695.

[54] M. Parrinello and A. Rahman, “Polymorphic transitions in single crystals: a new molecular dynamics method,” *Journal of Applied Physics*, vol. 52, pp. 7182–7190, 1998. doi:10.1063/1.328693.

[55] S. Nosé and M. L. Klein, “Constant pressure molecular dynamics for molecular systems,” *Molecular Physics*, vol. 50, pp. 1055–1076, 1983. doi:10.1080/00268978300102851.

[56] MATLAB, *version 8.4 (R2014b).* Natick, Massachusetts: The MathWorks Inc., 2014.

[57] L. Zhigilei. University of Virginia., “Introduction to Atomistic Simulations.” [WWW]. Available at `http://people.virginia.edu/~lz2n/mse627/notes/Diffusion.pdf`. Accessed on August 18th, 2016.

[58] J.-H. Jeon, H. Martinez-Seara, M. Javanainen, and R. Metzler, “Anomalous diffusion of phospholipids and cholesterols in a lipid bilayer and its origins,” *Physical Review Letters*, vol. 109, p. 188103, 2012. doi:10.1103/PhysRevLett.109.188103.

[59] University of Pennysylvania. Department of Physics & Astronomy., “Lab Manual: Averaging, Errors and Uncertainty.” [WWW]. Available at `http://virgo-physics.sas.upenn.edu/uglabs/lab_manual/Error_Analysis.pdf`. Accessed on June 16th, 2016.

[60] K. Younge, B. Johnston, C. Christenson, A. Bohara, J. Jacobson, N. M. Butler, and P. Saulnier, “The use of radial distribution and pair-correlation functions to analyze and describe biological aggregations,” *Limnology and Oceanography: Methods*, vol. 4, 2006. doi:10.4319/lom.2006.4.382.

[61] M. Kalimeri, *Are thermophilic proteins rigid or flexible? An in silico investigation.* PhD thesis, Paris Diderot University, 2014. 122 pages.

[62] J. A. Hartigan, *Clustering Algorithms.* John Wiley & Sons, Inc., 1975.

[63] R Development Core Team (2015), “R: A Language and Environment for Statistical Computing.” R Foundation for Statistical Computing, Vienna, Austria. Available online at `http://www.R-project.org/`.

[64] X. Daura, K. Gademann, B. Jaun, D. Seebach, W. F. van Gunsteren, and A. E. Mark, “Peptide folding: when simulation meets experiment,” *Angewandte Chemie-International Edition in English*, vol. 38, no. 1-2, pp. 236–240, 1999.

[65] A. Kitao and N. Go, “Investigating protein dynamics in collective coordinate space,” *Current Opinion in Structural Biology*, vol. 9, pp. 164–169, 1999. doi:10.1016/S0959-440X(99)80023-2.

[66] E. W. Weisstein. MathWorld – A Wolfram Web Resource., “Covariance.” [WWW]. Available at `http://mathworld.wolfram.com/Covariance.html`. Accessed on August 18th, 2016.

[67] E. Martz, K. Oberholser, W. DeLano, E. Hodis. Protopedia – Life in 3D., “Hydrogen Bonds.” [WWW]. Available at `http://proteopedia.org/wiki/index.php/Hydrogen_bonds`. Accessed on August 18th, 2016.

[68] Izhak Bucher (1991), “Circle fit.” Available online at `https://www.mathworks.com/matlabcentral/fileexchange/5557-circle-fit/content/circfit.m`.

[69] N. Eswar, B. Webb, M. A. Marti-Renom, M. S. Madhusudhan, D. Eramian, M.-y. Shen, U. Pieper, and A. Sali, “Comparative protein structure modeling using MODELLER,” *Current Protocols in Protein Science*, vol. 50, 2007. doi:10.1002/0471140864.ps0209s50.

[70] G. van Meer, D. R. Voelker, and G. W. Feigenson, “Membrane lipids: where they are and how they behave,” *Nature Reviews Molecular Cell Biology*, vol. 9, 2008. doi:10.1038/nrm2330.

[71] “CHARMM-GUI: Membrane Builder.” [WWW]. Available at `http://www.charmm-gui.org/?doc=input/membrane_only`. Accessed on June 2th, 2016.

[72] “CHARMM-GUI: Martini Bilayer Maker.” [WWW]. Available at `http://www.charmm-gui.org/?doc=input/mbilayer`. Accessed on June 7th, 2016.

[73] “LIPID MAPS Structure Database (LMSD).” [WWW]. Available at `http://www.lipidmaps.org/data/structure/LMSDSearch.php`. Accessed on June 2th, 2016.

[74] B. Hess, H. Bekker, H. J. C. Berendsen, and J. G. E. M. Fraaije, “LINCS: a linear constraint solver for molecular simulations,” *Journal of Computational Chemistry*, vol. 18, pp. 1463–1472, 1997. doi:10.1002/(SICI)1096-987X(199709)18:12<1463::AID-JCC4>3.0.CO;2-H.

[75] W. Humphrey, A. Dalke, and K. Schulten, “VMD – Visual Molecular Dynamics,” *Journal of Molecular Graphics*, vol. 14, pp. 33–38, 1996.

[76] V. A. Frolov and J. Zimmerberg, “Flexible scaffolding made of rigid BARs,” *Cell*, vol. 132, pp. 727–729, 2008. doi:10.1016/j.cell.2008.02.025.

[77] K. Pluhackova, S. A. Kirsch, J. Han, L. Sun, Z. Jiang, T. Unruh, and R. A. Böckmann, “A critical comparison of biomembrane force fields: structure and dynamics of model DMPC, POPC and POPE bilayers,” *The Journal of Physical Chemistry*, vol. 120, pp. 3888–3903, 2016. doi:10.1021/acs.jpcb.6b01870.

[78] W. Rawicz, K. C. Olbrich, T. McIntosh, D. Needham, and E. Evans, “Effect of chain length and unsaturation on elasticity of lipid bilayers,” *Biophysical Journal*, vol. 79, pp. 328–339, 2000. doi:10.1016/S0006-3495(00)76295-3.

[79] J. L. Gallop, C. C. Jao, H. M. Kent, P. J. G. Butler, P. R. Evans, R. Langen, and H. T. McMahon, “Mechanism of endophilin N-BAR domain-mediated membrane curvature,” *The EMBO Journal*, vol. 25, pp. 2898–2910, 2006. doi:10.1038/sj.emboj.7601174.

[80] H. T. McMahon and B. E., “Membrane curvature at a glance,” *Journal of Cell Science*, vol. 128, pp. 1065–1070, 2015. doi:10.1242/jcs.114454.

[81] P. A. Janmey and P. K. J. Kinnunen, “Biophysical properties of lipids and dynamic membranes,” *Trends in Cell Biology*, vol. 16, pp. 538–546, 2006. doi:10.1016/j.tcb.2006.08.009.

APPENDIX A. LIPID DIFFUSION IN THE ABSENCE OF PROTEIN

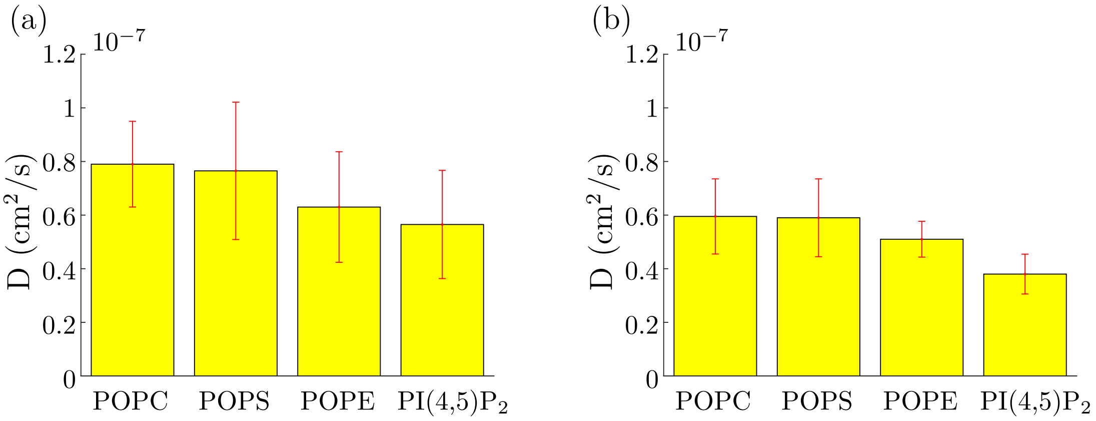

Figure A1 Diffusion coefficients in the absence of MIM IMD (a) with minimum amount of potassium ions to neutralize the solution (mean area per lipid 0.62 nm^2) and (b) with excess potassium and chloride ions (mean area per lipid 0.61 nm^2). The values are averaged over both leaflets.

APPENDIX B. MINIMUM DISTANCES IN PROTEIN-LIPID INTERACTIONS

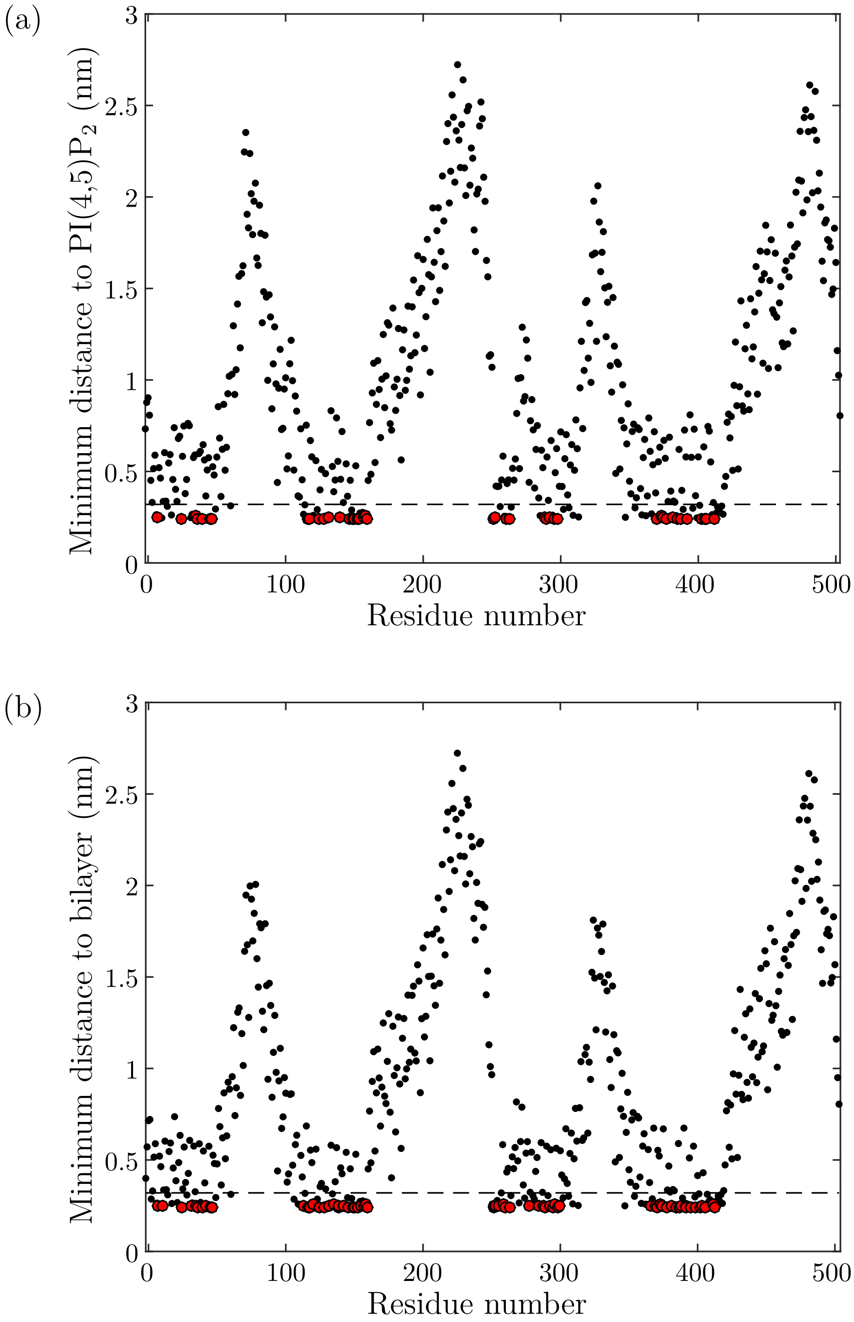

Figure B1 Minimum distance of each residue to (a) any PI(4,5)P$_2$ molecule and (b) any lipid. The values for PI(4,5)P$_2$/lipid-binding residues are marked in red (total contact time at least 10 ns). The dashed line indicates the 3.2 Å cut-off distance.

APPENDIX C. LIPID-BINDING INTERFACE OF MIM IMD

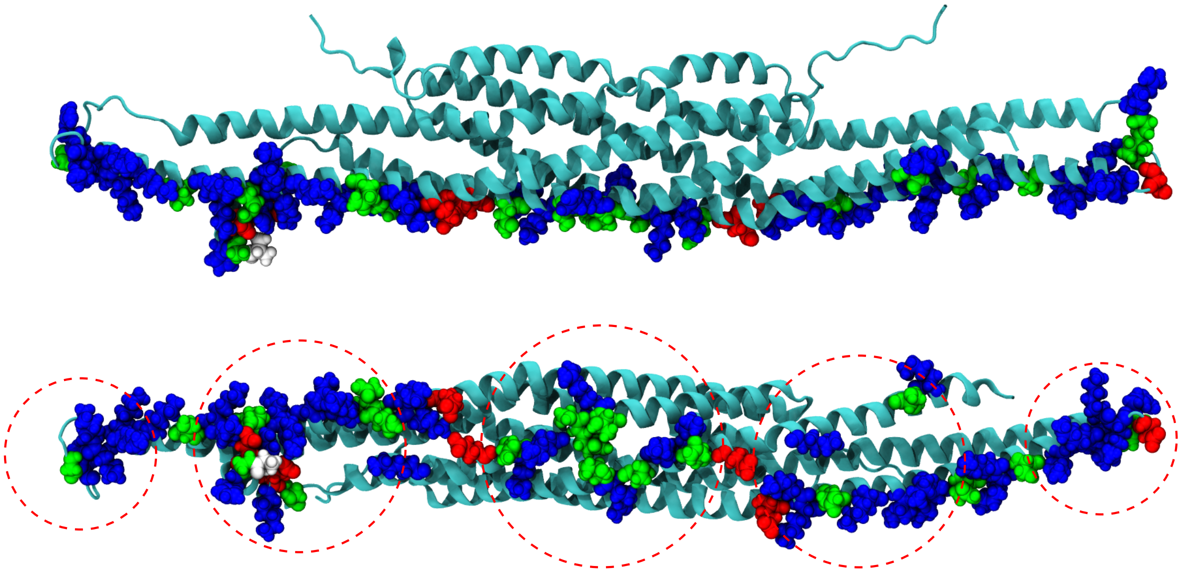

Figure C1 Lipid-binding residues as viewed from the side (top) and from below (bottom). Blue colour represents positively charged amino acids (arginine, lysine), white denotes hydrophobic (alanine), and green indicates polar uncharged amino acids (serine, threonine, asparagine, glutamine) and glycine. The red dashed circles indicate the position zones of PI(4,5)P_2-binding residues for easier comparison.

APPENDIX D. DENSITY MAPS

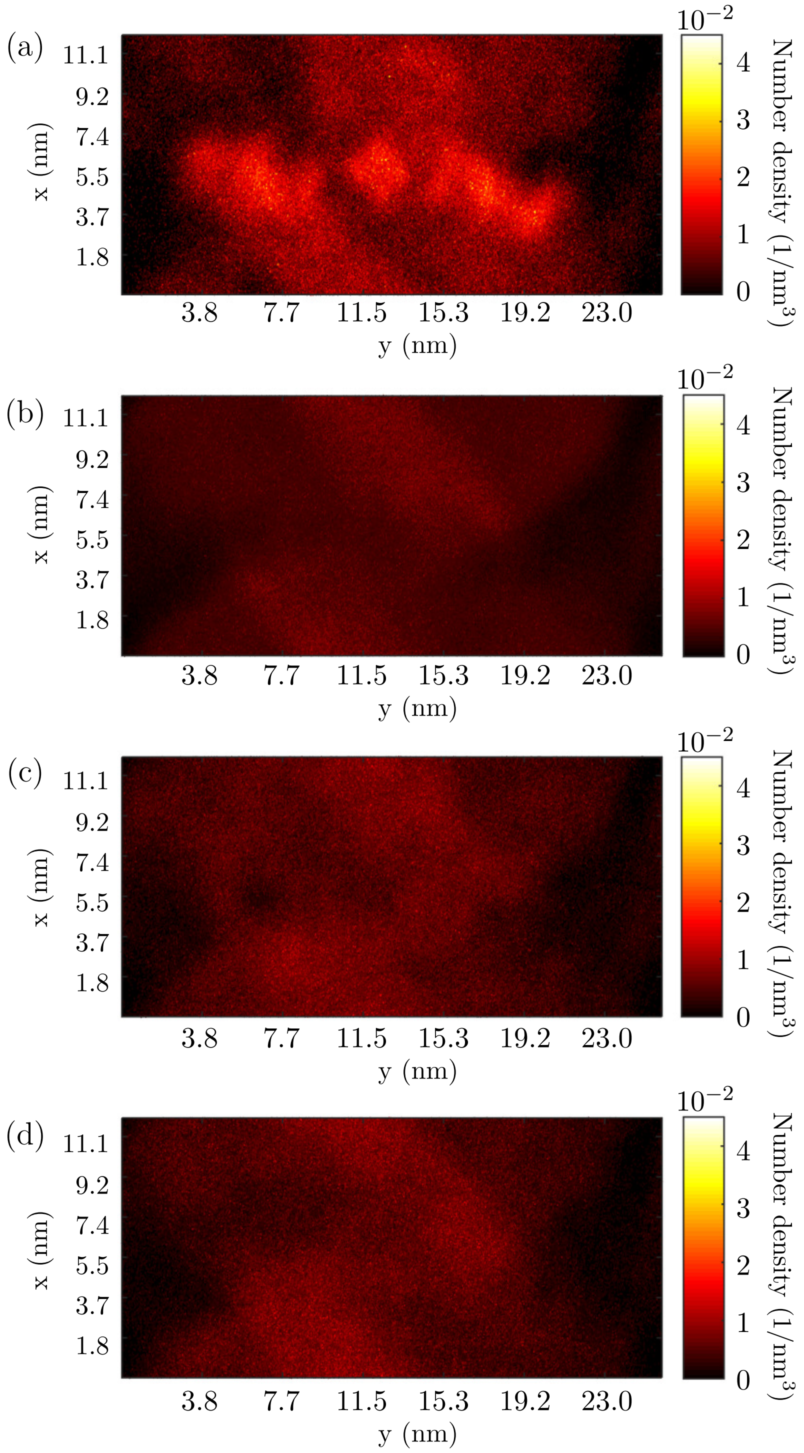

Figure D1 (a) PI(4,5)P$_2$, (b) POPC, (c) POPS, and (d) POPE density map on a tensionless bilayer in contact with MIM IMD, normalized with the number of the lipids in question.

APPENDIX E. AVERAGE DISTAL END INSERTION ON HIGH TENSION BILAYERS

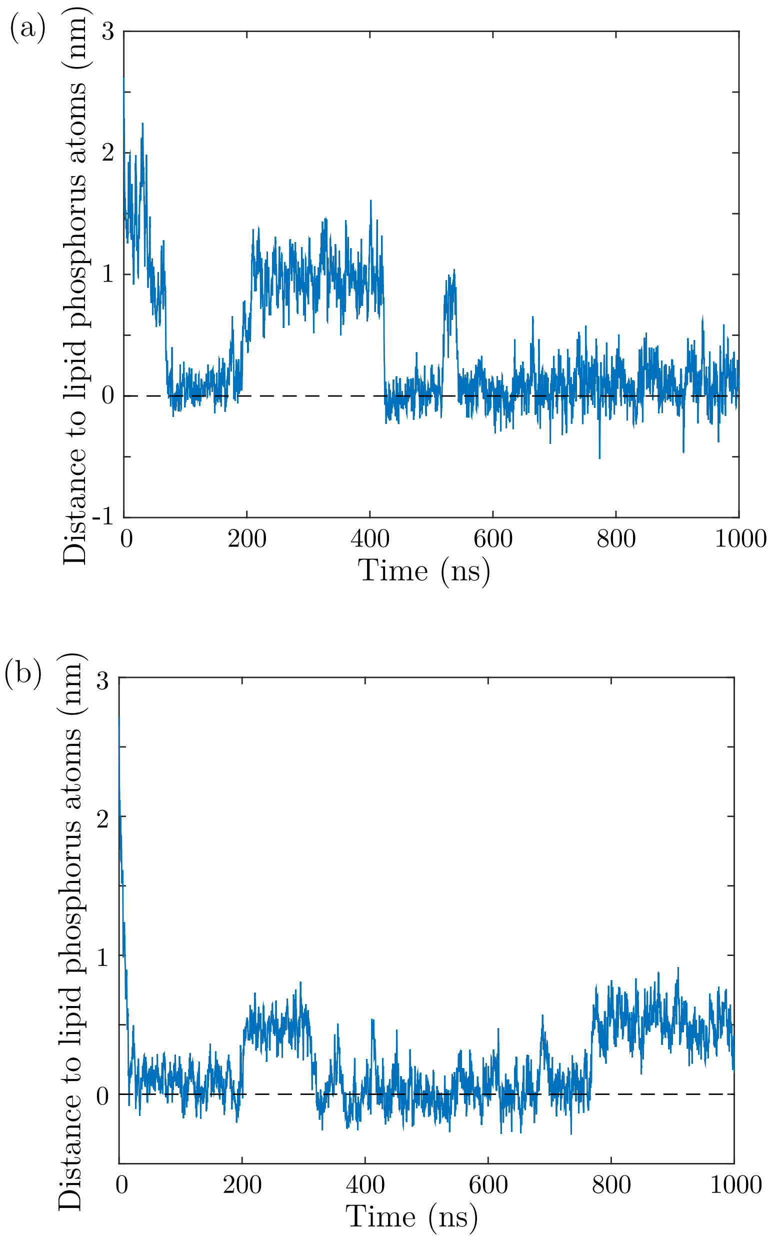

Figure E1 Average insertion depth of inserting hydrophobic residues of the distal ends of MIM IMD, on a tense membrane with an area per lipid of (a) 0.85 nm^2 and (b) 1.01 nm^2. The dashed line indicates the position of lipid phosphorus atoms.