2016

Molecular Simulations of Protein-Induced Membrane Remodeling

Ryan Patrick Bradley
University of Pennsylvania, bradleyrp@gmail.com

Molecular Simulations of Protein-Induced Membrane Remodeling

Abstract
Membranes organize much of the cell and host a great deal of molecular machinery required to integrate signals from the outside, regulate the surrounding matrix, change shape, move, and grow. Understanding how a dense forest of proteins, sugars, and biomarkers modulates the shape of the cell is necessary to produce more detailed, accurate predictions of cell behavior, particularly in the studies of cell signaling processes that lead to oncogenesis. In this dissertation, I will present a series of molecular models which, when combined with continuum models and both in vitro and in vivo experiments, describe the molecular basis for membrane morphology changes. In particular, we investigate the mechanisms by which proteins assemble on a bilayer undergoing thermal fluctuations. This work serves to quantify and explain a series of biophysical experiments in molecular detail, and contributes to the development of multiscale models for predicting cell fate.

Degree Type
Dissertation

Degree Name
Doctor of Philosophy (PhD)

Graduate Group
Chemical and Biomolecular Engineering

First Advisor
Ravi Radhakrishnan

Keywords
intracellular trafficking, membrane proteins, molecular dynamics, protein mechanics

Subject Categories
Biochemistry | Biophysics | Condensed Matter Physics

MOLECULAR SIMULATIONS OF PROTEIN-INDUCED MEMBRANE REMODELING

Ryan Patrick Bradley

A DISSERTATION

in

Chemical and Biomolecular Engineering

Presented to the Faculties of the University of Pennsylvania in Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy

2016

Supervisor of Dissertation:
Ravi Radhakrishnan, Professor of Bioengineering

Graduate Group Chairperson:
Talid Sinno, Professor of Chemical and Biomolecular Engineering

---

Dissertation Committee:
Wei Guo, Professor of Biology
Paul Janmey, Professor of Physiology
Daniel Hammer, Professor of Bioengineering
Tobias Baumgart, Associate Professor of Chemistry
Robert Riggleman, Assistant Professor of Chemical and Biomolecular Engineering

ACKNOWLEDGMENTS

This dissertation would not have been possible without the funding, collaboration, guidance, and support of many of my friends at the University of Pennsylvania and colleagues in the scientific community. I would like to thank the National Science Foundation for support from the Graduate Research Fellowship, along with the Ashton Fellowship, the National Institues of Health training grant in structural biology, and computational resources provided by grants from the Extreme Science and Engineering Discovery Environment. This dissertation also benefited greatly from collaboration with research groups led by Professors Tobias Baumgart, Wei Guo, and Paul Janmey here at the University of Pennsylvania. I am sincerely grateful to my advisor, Ravi, for his patient guidance and encouragement. I am also very lucky to have collaborated with all of the members of the Radhakrishnan lab, in particular Richard Tourdot, Ramakrishnan Natesan, David Slochower, and Joe Jordan. It has been the highest honor to study such fascinating subjects with such wonderful people. I would also like to thank my family members for their support and encouragement over the years, most especially my companion Lisa Musz.

ABSTRACT

MOLECULAR SIMULATIONS OF PROTEIN-INDUCED MEMBRANE REMODELING

Ryan Patrick Bradley

Ravi Radhakrishnan

KEYWORDS: membrane proteins, intracellular trafficking, curvature focusing, coarse-grained molecular dynamics simulations, statistical mechanics, protein mechanics

Membranes organize much of the cell and host a great deal of molecular machinery required to integrate signals from the outside, regulate the surrounding matrix, change shape, move, and grow. Understanding how a dense forest of proteins, sugars, and biomarkers modulates the shape of the cell is necessary to produce more detailed, accurate predictions of cell behavior, particularly in the studies of cell signaling processes that lead to oncogenesis. In this dissertation, I will present a series of molecular models which, when combined with continuum models and both *in vitro* and *in vivo* experiments, describe the molecular basis for membrane morphology changes. In particular, we investigate the mechanisms by which proteins assemble on a bilayer undergoing thermal fluctuations. This work serves to quantify and explain a series of biophysical experiments in molecular detail, and contributes to the development of multiscale models for predicting cell fate.

Contents

List of Tables

List of Figures

ABBREVIATIONS

MD	Molecular Dynamics
CGMD	Coarse-Grained Molecular Dynamics
ENTH	Epsin N-terminal homology domain
PIP_2	Phosphatidylinositol-4,5-biphosphate
CME	Clathrin-Mediated Endocytosis
BAR	Bin/Amphiphysin/Rvs
ECM	Extracellular Matrix
LUV	Large unilamellar vesicle
GUV	Giant unilamellar vesicle
MC	Monte Carlo
DTMC	Dynamically Triangulated Monte Carlo
ENM	Elastic network model
PCA	Principal component analysis
XSEDE	Extreme Science and Engineering Discovery Environment
GROMACS	Groningen Machine for Chemical Simulations
RMSD	Root-mean-squared deviation
MSE	Mean-squared error

NOTATION

k_B	Boltzmann constant
T	Temperature
β	$= 1/(k_B T)$
$\boldsymbol{\kappa}$	Isotropic bending modulus of a fluid membrane
$\bar{\boldsymbol{\kappa}}$	Gaussian bending modulus
$\boldsymbol{\gamma}$	Isotropic tension of membrane including membrane area elasticity
$\langle \dots \rangle$	Ensemble average
$\mathcal{H}_{\text{el}}$	Elastic energy of the surface
$\mathcal{C}_1, \mathcal{C}_2$	Principal curvatures
H	$= (\mathcal{C}_1 + \mathcal{C}_2)/2$ — Mean curvature
K	$= \mathcal{C}_1 \mathcal{C}_2$ — Gaussian curvature
H_0	Spontaneous curvature
$C_0(x, y)$	Spontaneous curvature field
$C_{0,max}$	Peak spontaneous curvature
σ_a	Standard deviation of a Gaussian function**a** direction
τ	Time coupling constant in a molecular dynamics simulation
A	Area of a membrane
$\mathbf{q}$	Wavevector

Chapter 1

Introduction

This dissertation describes a multifaceted investigation of how proteins remodel cell membranes. This question has a lengthy history, stretching back to the first observation of cells by Robert Hooke and Antonie van Leeuwenhoek in 1665 [111]. One of the key subjects of this work is the lipid bilayer, which possesses a remarkable combination of strength and elasticity, serving as a two-dimensional fluid encrusted with a large amount of proteins, sugars, and other biomolecules [91]. The peculiar size and composition of the lipid bilayer itself was not identified until Meyer and Overton discovered that anesthetic potency correlated very strongly with partition coefficients in olive oil, proposing that such compounds work at a very specific lipophilic-hydrophilic site in their "lipoid theory of narcosis" [135] in the early 20th century.

Needless to say, our understanding of lipid bilayers and the proteins that organize their function — and by extension cellular functions — has advanced signficantly since then, even though many basic questions about its structure and function are still unanswered. This dissertation seeks to modestly advance the history of the study of cell membranes using modern tools and minimal models to explain a crucial set of experiments and lend insight into our understanding of the rich set of biophysical processes hosted by cell membranes.

1.1 Sequence of this thesis

This document is organized into seven chapters and one appendix. This chapter outlines the common themes that run throughout the work. Chapter 2 is an extensive review of the coarse-grained simulation methodology used in the remainder of the thesis. The following chapters address several biological systems, namely the formation of lamellaepodia in cells (chapter 3, based on [380]), the generation of curvature required for endocytosis (chapter 4, based on [44]), and phosphoinositide-ion binding which may influence cell signaling cascades (chapter 5). In chapter 6 I will reflect on the process by which molecular measurements are extended via the mesoscale

to comment on experiments and macroscopic observables drawing on both the "invadopodia" and "endocytosis" stories. In the final chapter (7), I will outline the new modes of inquiry that these models have opened up, and suggest further uses for these modeling strategies.

1.2 Fundamental themes

While the dissertation is broadly organized around several questions about how cell membranes work, there are several threads that run throughout the thesis.

Cellular transport. Intracellular trafficking events and cell membrane morphology change are essential processes that determine how cells interact with their environments. A key hypothesis that informs most of the investigation in this thesis is the claim that form precedes function, and that specific (read: non-flat) membrane morphologies are necessary to facilitate the organization and trafficking of biolmolecules, particularly the cargo and signals which ultimately determine cell fate. Several chapters in this dissertation will consider a *cartoon* of these intracellular trafficking processes, one of which is drawn in figure (1.1).

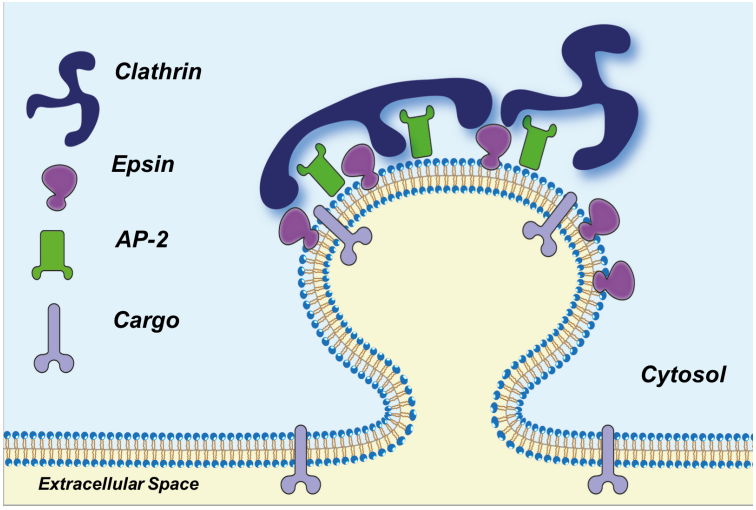

Figure 1.1: A cartoon depiction of clatrin-mediated endocytosis, including only three of the many adaptor proteins.

Much of our understanding of these trafficking process requires that we assign chemical identities and concentrations to the proteins which act in concert to generate these structures.

Statistical mechanics. This thesis depends on the field of *statistical mechanics*, a subset of theoretical physics which seeks to show how the thermodynamic properties

of large systems emerge from uncertainty about its physical conditions. We employ molecular dynamics as a metaphorical "microscope" for investigating the collective action of protein-membrane systems, and the theoretical edifice provided by statistical mechanics allows us to match our nano-scale observations to macroscopic observables.

Biological observations. If chemical physics married to statistical mechanics provides the *general* framework for this work, then biological observations are responsible for identifying the *specific* phenomena which we model. This is to say that even though we perform "theory" in this work, its ultimate purpose is a practical one. We wish to put accurate, sometimes precise numbers to observable phenomena, particularly those found in *in vitro* experiments supplied by expert collaborators in the chemistry and biology departments. The specificity conferred by matching biological experiments to our simulations — which admittedly contain orders of magnitude *less* information, measured at orders of magnitude *higher* precision — ensures that our investigation can be useful to the scientific community at large.

Harmonic springs. One specific mathematical device appears with an uncanny regularity throughout much of the methods used in this dissertation: harmonic springs, or more generally, quadratic equations. The concept of an *harmonic* or Hookean spring (in which energy goes as $\mathrm{k}(\mathrm{x}-\mathrm{x}_0)^2$ for some constant k) is found in several places. Each molecular simulation uses this approximation for covalent bonds. Several coarse-grained protein simulations use this approximation to model the interactions between tertiary structures in a protein. Measurements of protein motion in the spectrin experiment in the final chapter rely on a quasi-harmonic assumption that is largely true in equilibrium simulations. And finally, we precisely measure protein-induced curvature fields in 4 using a correction to the *equipartition theorem*, which depends on the assumption that we can write the energy as in a quadratic form. It's difficult to overstate the utility that we gain from modeling physical phenomena as harmonic springs.

Entropy and fluctuations. The simulation methods used in this dissertation complement experiments because they provide a description of fluctuation systems below the diffraction limit. For that reason, we can use simulations to test theories about what happens *in vitro* with higher precision and more detail. However, at such small length scales, entropy and fluctuations play an important role in determining the overall behavior of the system. In particular, this dissertation will consider how to distinguishing thermal fluctuations from protein-induced curvature. More broadly, entropic effects are a common, and important theme. For example, coarse-grained simulations trade enthalpy for entropy, and lipid bilayers self-assemble thanks to a delicate balance of hydrophilic and hydrophobic forces which are strongly influenced by entropy within the bilayer.

Estimation. Finally, one of the themes that runs through each chapter is the idea that we can reasonably estimate physical quantities from many different experiments, with varying degrees of uncertainty. This sounds like an obvious claim, but as we will see in chapter 6, combining various estimations at different time- and length-scales is essential to building holistic models that can integrate many incomplete but accurate views of the cell. A famous statistician, George E.P. Box once explained that "Since all models are wrong the scientist cannot obtain a "correct" one by excessive elaboration." [42]. We hope to keep this under advisement.

1.3 Novelty and extensibility of this work

The investigations described in this dissertation depend on a massive amount of scientific infrastructure. All of the target systems are observed by combinations of *in vitro* and *in vivo* experiments. The simulations are performed on supercomputing platforms provided by the University of Pennsylvania and the extreme science and engineering discovery environment (XSEDE). These platforms run software developed over decades, designed to efficiently and accurately simulate molecular systems, particularly the `GROMACS` integrator [342], the `MARTINI` coarse-grained force field [212], and the `CHARMM` atomistic force field [50]. These simulations were analyzed and visualized using a suite of open-source software also developed over years, with hundreds of thousands of hours in collective development time.

In this thesis I have sought to apply many of these tools to add molecular detail to our understanding of protein-membrane remodeling processes. The novelty of this research originates in the particular blend of theory and simulation that I have used to address open problems in biology. However, it is my sincere hope that this work become *less* novel in the future, and specifically, that many of the computing tools outlined this thesis be applied to more and more detailed molecular systems so that computational biophysics investigators may improve their biophysical predictions in service of answering questions about cell behavior.

Chapter 2

Coarse-Grained Models for Protein-Cell Membrane Interactions

The following coarse-grained molecular dynamics review is adapted from "Coarse-Grained Models for Protein-Cell Membrane Interactions" [43] by Ryan Bradley and Ravi Radhakrishnan

The physiological properties of biological soft matter are the product of collective interactions which span many time and length scales. Recent computational modeling efforts have helped illuminate experiments which characterize the ways in which proteins modulate membrane physics. Linking these models across time and length scales in a multiscale model explains how atomistic information propagates to larger scales. This paper reviews continuum modeling and coarse-grained molecular dynamics methods which connect atomistic simulations and single-molecule experiments with the observed microscopic or mesoscale properties of soft-matter systems essential to our understanding of cells, particularly those involved in sculpting and remodeling cell membranes.

2.1 Introduction

In more than three decades since the first molecular dynamics simulation of a protein [218], molecular dynamics methods have emerged as an effective tool for simulating biological soft matter thanks to the careful development of force fields and simulation methods. The synthesis of models for soft-matter physics and protein dynamics provides great insight into a wide range of biological processes important to understanding human health.

Molecular dynamics simulations at both atomistic and coarser levels of detail allow

us to probe the properties of complex biomolecular systems with numerical methods. While molecular dynamics simulations collapse many degrees of freedom into relatively few, they are nevertheless capable of reproducing a host of important physical phenomena that result from the collective action of complex particles. In addition to separate applications in soft matter and protein systems, many simulation studies have investigated the crucial interactions between lipid bilayers and the proteins that remodel them. These interactions are crucial to a wide range of cellular processes including membrane remodeling in endocytosis [10], the action of protein-gated ion channels [39, 201, 370] the assembly of membrane proteins [163, 289, 299, 338], mediation of membrane fusion [24, 108, 300], and the activation of membrane-protein-based signaling networks [21, 256]. The interactions of proteins with lipid bilayers are vital to our understanding of these phenomena; this necessesitates the use of models that span several time- and length-scales (delineated in figure 2.1), as well as careful matching to experimental results.

Simulations have become more powerful in recent years thanks to a combination of increased computer power, advanced sampling methods, distributed computing, and specialized hardware [104, 106, 167]. However, it is the parameterization of force fields which are capable of matching experimental data at multiple length scales, from X-ray scattering data to protein crystal structures, that makes these models into useful microscopes for studying cell systems.

Coarse-grained molecular dynamics simulations employ intermediate resolution in order to balance chemical detail with system size. They offer sufficient size to study membrane-remodeling events while retaining the ability to self-assemble. Because they are capable of simulating mesoscopic length-scales, they make contact with a wider variety of experiments, many of which lack the precision to easily inform small atomistic models.

A complete coarse-grained model must include two components: a mapping from atomistic structures to coarse-grained "beads" and a set of potentials which describe the interactions between beads. The former defines the geometry or length-scale of the resulting model, while the latter defines the "force field". The parameterization of the force field is essential to the performance of the model, which is only relevant insofar as it can reproduce experimental observables. Recent improvements to widely-disseminated force fields have strengthened both their transferability — the ability to use a model on a novel biophysical system with straightforward parameterization — as well as tunability, the ability to customize a model to match a desired quantity [92, 149, 330].

In this review, we will describe the characteristic methods for developing CGMD models, namely the "bottom-up" force-matching, and "top-down" free energy-based approaches. We will illustrate the myriad ways in which these models can reproduce protein dynamics, bilayer physics, and experiments which probe protein-membrane interactions. We will survey the applications of coarse-grained models to protein-membrane interactions, and describe the ways in which CGMD simulations make

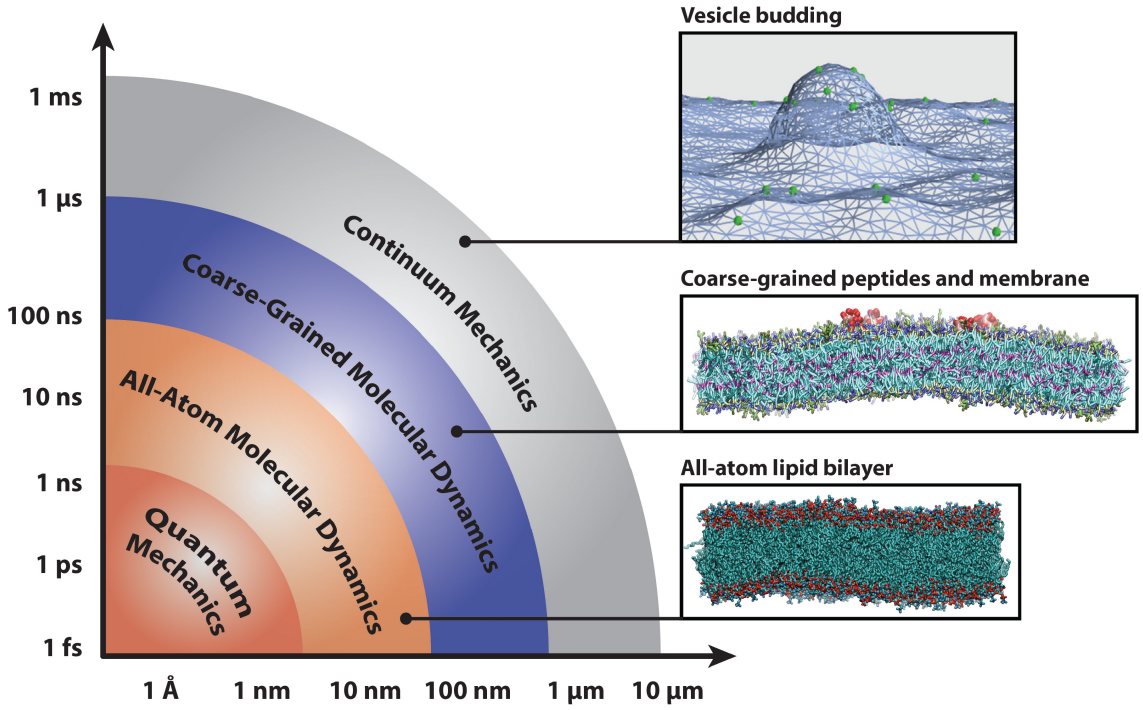

Figure 2.1: Diagram of computational methods for studying biophysical systems across a range of time- and length-scales. Representative snapshots depict an all-atom lipid bilayer, peptides embedded in a coarse-grained bilayer, and proteins remodeling a continuum mechanics membrane model. Bilayers were simulated with the CHARMM36 [166] and MARTINI [212] force fields, and rendered with Visual Molecular Dynamics [142].

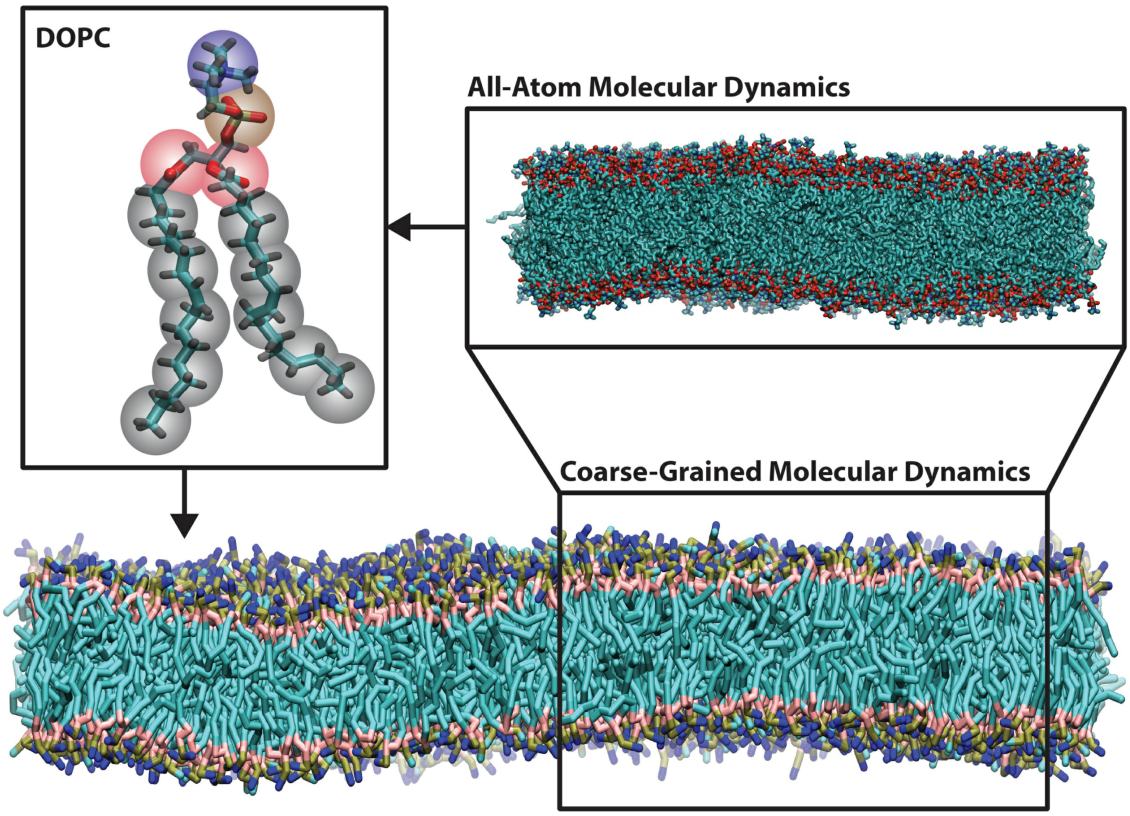

Figure 2.2: Representative snapshots of all-atom (upper right) and Martini coarse-grained (bottom) molecular dynamics simulations of a 4:1 dioleoylphosphatidylcholine with dioleoylphospatidylserine (DOPC/DOPS) bilayer. The upper left shows the coarse-grained mapping of a single DOPC lipid, with beads colored by bead type (gray for hydrocarbons-, pink for glycerol-, brown for phosphate- and blue for choline-type). The all-atom system contains 800 lipids, while the coarse-grained system contains 3,200 lipids (water molecules are not pictured here). Bilayers were simulated with the CHARMM36 [166] and Martini [212] force fields, and rendered with Visual Molecular Dynamics [142].

contact with experiments and simulations at larger and smaller length scales. While the array of coarse-grained tools can be used to probe cell-biology problems *in silico*, there is also much to be gained from studying these models as a whole, by studying the communication of information between different length scales in biological processes.

2.2 Methods for parameterizing coarse-grained force fields

The defining feature of a coarse-grained biophysical model is the length-scale at which chemical components are modeled; such a model necessarily lumps many atomic degrees of freedom into a single coarse-grained bead. As with any classical MD approach, a CGMD model treats molecules classically, integrating Newton's laws of motion according to potentials which define the forces between each bead in the system.

$$m_i \frac{\partial^2 \mathbf{r}_i}{\partial t^2} = \mathbf{F}_i,\ \mathbf{F}_i = -\frac{\partial V}{\partial \mathbf{r}_i},\ i = 1 \ldots N$$

These equations describe the motion of N particles, each with mass m_i, experiencing a force F_i due to a potential energy function V, itself a function of the configuration of all atoms in the system which are close enough to exert a measurable force. Several software packages are capable of integrating these equations, including the popular GROMACS [342], NAMD [251], CHARMM [166], and AMBER [356] packages. Many of the coarse-grained methods utilize one of these integrators to perform simulations.

Molecular dynamics simulations makes contact with observables like temperature and pressure via statistical mechanics. Temperature is defined by kinetic energy of the particles, while macroscopic pressure is defined by the average of the molecular virial [34] as follows.

$$\frac{1}{2} N_{df} k_B T = E_{kin},\ E_{kin} = \frac{1}{2} \sum_i^N m_i \mathbf{v}_i \cdot \mathbf{v}_i$$

$$\mathbf{P} = \frac{2}{\mathscr{V}} \left[E_{kin} - \mathbf{\Xi} \right],\ \Xi = -\frac{1}{2} \sum_{i<j} \mathbf{r}_{ij} \cdot \mathbf{F}_{ij}$$

In this equation, $\mathscr{V}$ is the volume of the system, E_{kin} is the kinetic energy, $\mathbf{r}_{ij}$ is the distance vector between particles i and j, $\mathbf{F}_{ij}$ is the corresponding force, N_{df} is the number of degrees of freedom ($3N - 3$ for N particles, minus any constraints), and Ξ is virial. The choice of these forces and the physical quantities they represent – dispersion forces, electrostatics, and bonded forces – define the model and determine its ability to reproduce observed physical phenomena. In this section we will first summarize the early advances in coarse-graining and then review three representative coarse-grained models built from structure-based, force-based, and energy-based force-fields, respectively. Because coarse-graining requires a simplification of many degrees of freedom, it is impossible to build a model which simulateneously repro-

duces all of the geometric, thermodynamic, and kinetics features of a physical system. To build a coarse-grained model, it is therefore necessary to choose which physical properties are essential to the behavior of the target system. We can classify the most popular models by which property they aim to reproduce, namely the geometry of the system (structure-based), the distribution of forces between particles, or thermodynamic properties (energy-based). The representative models described in sections 2.2.2 through 2.2.4 each take these approaches, though there is significant overlap and these are not the only suitable coarse-grained methods. In fact, much of the power of coarse-graining method lies in its flexibility and the ease with which it can be adapted to new applications.

2.2.1 Early Coarse-grained models and dissipative particle dynamics

The development of coarse-grained models for interfacial systems was made possible by the need to bridge detailed atomistic simulations with continuum methods. The seminal coarse-grained modeling approaches drew from many different methods, including both Monte Carlo and molecular dynamics integration schemes, lattice and off-lattice models, and either hard sphere or Lennard-Jones potentials. While a comprehensive summary of these modeling approaches is beyond the scope of this review, early development of these models and connections to earlier work is summarized in a pair of reviews [230, 294]. In general, early coarse-grained models can be classified by the number of molecules that are mapped onto a single coarse-grained particle.

Early models mapped a single molecule onto one coarse-grained particle in order to simulate spontaneous phase separation. Larson employed a Monte Carlo scheme [176] to simulate oil-water-amphiphile systems on two- and three-dimenaional cubic lattices, while Smit and coworkers used molecular dynamics to simulate these systems using Lennard-Jones particles [310]. In 1998, Goetz and Lipowsky modeled surfactant molecules by Lennard-Jones spheres connected by harmonic bonds in order to simulate the self-assembly of bilayers and micelles and calculate the resulting stress and density profiles [114]. With a molecular representation they calculated the bending rigidity of the bilayer from its fluctuation spectra, and demonstrate that these models are able to reproduce both bending and protrusion modes [115].

To reach larger time- and length-scales, the dissipative particle dynamics (DPD) method uses a much coarser mapping, in which one site may represent many molecules in a small fluid volume [94, 138]. There are three types of forces present in DPD models: a conserved soft repulsion force, pairwise dissipation forces, and pairwise random forces. The balance of dissipation and random forces provides the thermostat for the DPD model, and since this thermostat preserves the momentum of individual particles, these models provide correct hydrodynamic behavior. In addition to using a coarser mapping, DPD simulations use a longer time-step due to the use of soft repulsion forces. It is necessary to match the observed compressibility in a DPD simulation to the target fluid in order to study the phase behavior and interfacial tension of the model fluid [120]. The DPD method has been applied to biological lipid

bilayers [350], membrane fusion processes [118, 300], and bilayers with proteins [348], and its connections to the mesoscale have been reviewed extensively [116, 121, 349].

It is clear that the full spectrum of coarse-grained modeling approaches contains contributions from several different fields. Early coarse-grained simulations were made possible by advances in computer hardware, which made it possible to simulate larger interfacial systems at finer levels of detail. These simulations began to bridge the gap between the atomistic simulation of lipids and protein systems with mesoscale, statistical mechanics models for membranes. In this section we have cited some of the milestones in coarse-grained model development. To review more recent coarse-grained simulation methods, and to limit the scope of this review, we will now turn our attention to three models which represent the structure, force, and energy-matching approaches. This list is by no means comprehensive, and there are many other coarse-grained models for simulating biological, interfacial systems available in the literature.

2.2.2 Structure and energy matching in the CMM-CG model

In the early 2000s, Klein and coworkers developed a coarse-grained model for phospholipid bilayers by matching the structural and thermodynamic properties of water, hydrocarbons and lipid amphiphile to experimental measurements and all-atom simulations. The resulting force field, titled CMM-CG, has been used to investigate a range of polymer systems as well as those containing nonionic liquids and lipids.

The model was originally developed to reproduce structural properties of a dimyristoylphosphatidylcholine (DMPC) bilayer [295]. This requires careful assembly of water, hydrocarbon, and amphiphilic components. Given that water is the largest constituent of many soft-matter systems, and indeed makes most condensed matter systems truly "soft", it is necessary to reproduce both its structure and phase transitions. The CMM-CG model maps three water molecules onto a single bead.

Non-bonded forces are modeled with general Lennard-Jones (LJ) potentials with a potential well depth ($\epsilon_{\alpha\beta}$) and zero-position ($\sigma_{\alpha\beta}$) which is tuned to reproduce the desired structure and thermodynamic properties of the target system. The softer 12-4 potential was used to model dispersion forces in water by matching the melting temperature, density, and vapor pressure observed in bulk and thin-film test simulations.

$$V(r_{ij}) = \frac{3\sqrt{3}}{2}\epsilon_{\alpha\beta}\left\{\left(\frac{\sigma_{\alpha\beta}}{r_{ij}}\right)^9 - \left(\frac{\sigma_{\alpha\beta}}{r_{ij}}\right)^6\right\} \text{ (non-bonded)} \tag{2.1}$$

$$V(r_{ij}) = \frac{27}{4}\epsilon_{\alpha\beta}\left\{\left(\frac{\sigma_{\alpha\beta}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{\alpha\beta}}{r_{ij}}\right)^4\right\} \text{ (water)} \tag{2.2}$$

In the CMM-CG model, the well depth in water-water interactions was chosen to simultaneously provide a melting temperature of 212.1K, a boiling temperature

of 373.15K, and a minimum well-depth at 4.58Å, which is necessary to recover the correct density of water at 303.15K in a three-water bead [295]. To model soft-matter systems, it is necessary to include hydrophobic chemical components; in this case represented by a collection of n-alkanes. Simulations of alkanes in the CHARMM force field [97] provide target structural data for harmonic bond and angle potentials given by equation 2.3. These harmonic bond and angle are standard among many molecular dynamics force fields; a comprehensive summary of many common potential functions can be found in the GROMACS manual [341, 342].

$$V_{bond}(r) = \frac{k_{bond}}{2}\left(r - r_{eq}\right)^2, \; V_{bend}(\theta) = \frac{k_{bend}}{2}\left(\theta - \theta_{eq}\right)^2 \tag{2.3}$$

In equation 2.3, $\mathrm{V_{bond}}$ and $\mathrm{V_{bend}}$ represent the contribution of the bond stretching and bending to the potential energy function, $\mathrm{k_{bond}}$ and $\mathrm{k_{bend}}$ are the corresponding length and angle stiffnesses, and $\mathrm{r_{eq}}$ and θ_{eq} are the equilibrium bond length and angle, respectively. Tuning the equilibrium values and spring constants specifies the structure and fluctuations of individual molecules. Likewise, the inter-molecular structure and thermodynamics of the target fluid depends on non-bonded interactions, which are modeled with a 9-6 LJ potential. Nonbonded parameters were chosen to reproduce phase separation with water, as well as experimental bulk density and vapor pressure measurements.

Finally, to assemble hydrophobic, hydrophillic, and water components into a viable model for amphiphiles, non-bonded parameters must be chosen to reproduce the physics of lipid bilayers. Classic coarse-grained methods propose pair potentials between CG beads according to the Boltzmann inversion method. In this method, a pair correlation function, or radial distribution function (RDF) g(r) defines the probability of finding a particle at distance r from a reference particle such that the conditional probability of finding the particle is $\rho(\mathrm{r}) = \rho\mathrm{g}(\mathrm{r})$, where ρ is the average number density of the fluid. This pair correlation function may be calculated by analyzing an atomistic trajectory mapped onto coarse-grained beads. A potential of mean force (PMF) between CG beads is then estimated by equation 2.4 where $\mathrm{g_{aa}}(\mathrm{r})$ is the RDF measured from atomistic simulation, $\mathrm{k_B}$ is the Boltzmann constant, T is absolute temperature, and α_{n} is a scaling factor (corresponding to the nth iteration of the estimate) designed to include the effect of interactions with the (necessarily) heterogeneous environment.

$$V_n(r) = \alpha_n \left\{-k_B T \ln(g_{aa}(r)\right\} \tag{2.4}$$

The Boltzmann inversion method may be iteratively corrected according to equation 2.5 to correct the tabulated potentials until the pair-correlation functions for the atomistic and coarse-grained systems agree.

$$V_{n+1}(r) = V_n(r) + k_B T \ln \frac{g_n(r)}{g_{aa(r)}} \tag{2.5}$$

In practice, since the pair correlation function computed in an homogeneous environment is not equivalent to the potential of mean force in an inhomogeneous environment, it is necessary to include the effects of correlated contributions from the surrounding environment, as well as contributions from bonded intra-molecular forces. To account for these contributions, a reverse Monte Carlo (MC) method proposed by Lyubartsev and Laaksonen [203] is used to construct an effective pair potential. In the canonical ensemble, we may write the expectation for an observable RDF as:

$$\langle g_n(r_i)\rangle = \frac{\int g_n(r_i)e^{-\beta V}dr}{\int e^{-\beta V}dr}$$

where β is the inverse temperature $(\mathrm{k_B T})^{-1}$. In the following formulation, j indexes atoms or sites, the $\mathrm{r_i}$ refer to a set of inter-site distances which define the RDF, and n indexes the iterated calculations of the RDF given by $\mathrm{g_n(r_i)}$. Taking the partial derivative gives the following fluctuation formula.

$$\frac{\partial \langle g_n(r_i)\rangle}{\partial V(r_j)} = -\beta\left[\langle g_n(r_i)g_n(r_j)\rangle - \langle g_n(r_i)\rangle\langle g_n(r_j)\rangle\right]$$

This relates changes in the coarse-grained RDF to changes in tabulated potentials while considering cross-correlations between subject particles and their environment. Linearizing this equation gives a solution according to equation 2.6, which will yield self-consistent pair potentials in agreement with the atomistic data.

$$\langle g_n(r_i)\rangle - g_{target}(r_i) = \sum_{j=1}^{N}\frac{\partial \langle g_n(r_i)\rangle}{\partial V_n(r_j)}\triangle V_n(r_j) \tag{2.6}$$

It can be shown that iteratively solving for these potentials $\mathrm{V_n}$ with Monte Carlo methods (known as reverse Monte Carlo) can generate effective pair potentials which reproduce soft matter properties for a particular system. However, this method suffers from reduced transferability because it depends on a particular thermodynamic ensemble. That is, the target RDF includes information about temperature, density, and most importantly, composition, which limits its applicability to novel systems. For this reason it is necessary to test the model against thermodynamic data.

The relevant thermodynamic property in amphiphilic systems is the surface tension, which can be calculated from MD simulation by equation 2.7, where $\mathrm{L_z}$ is the box-size normal to the interface and $\mathrm{P_{ij}}$ is the ij component of the pressure tensor. Its condensed phase analog is the interfacial tension, which can be estimated with a combination of experiment and theory [74], however these measurements are subject to large errors.

$$\boldsymbol{\gamma} = \frac{L_z}{2}\left\langle P_{zz} - \frac{P_{xx}+P_{yy}}{2}\right\rangle \tag{2.7}$$

Matching the area-per-lipid and bilayer density (or electron density profile if matching to AAMD simulations or neutron diffraction) confirms that the CGMD has the correct structure. Tuning non-bonded cross-terms between lipid headgroups and hydrophilic tails will influence the observed surface tension or its conjugate variable, area-per-lipid, in simulations under zero tension. Capturing both condensed phase structure and energies in this way is necessary for building an accurate model. Contact with experiments will be discussed further in section 2.2.5.

2.2.3 Force matching with the MS-CG model

Gregory Voth and coworkers have proposed the concept of force-matching to develop a rigorous coarse-grained force field directly from forces measured in all-atom simulations. This is necessary, they argue, because other coarse-grained approaches suffer from reduced transferability compared to all-atom counterparts, namely because the coarse-grained simulation may not contain the correct thermodynamic ensemble. Insofar as the multi-body coarse-grained PMF is derived from structure factors which depend on temperature, pressure, and composition, they cannot be transferred to new systems.

To avoid this problem, they propose a variational method in which a coarse-grained force field is systematically developed from all-atom simulations under the correct thermodynamic ensemble [151]. In the statistical framework developed by Izvekov and Voth [150,151,291], it is possible to develop the exact many-body coarse-grained PMF from a trajectory of atomistic forces with a sufficiently detailed basis function.

In the original force-matching method developed by Ercolessi and Adams [93], a set of parameters defining classical forces of a pre-defined form are optimized by minimizing their squared difference from reference forces provided by *ab initio* simulation. This fitting procedure becomes intractable with the many components found in biochemical systems. To circumvent this optimization problem, Voth *et al.* have designed the force field to be linear in the fitting parameters by constraining their choices of basic functions to those that have zero derivatives between mesh points. This allows one to optimize the force field by finding the least-squares solution to an overdetermined system of linear equations.

They start with a collection of sampled configurations from an atomistic simulation of the target system and calculate the reference forces between atoms of a particular type. After decomposing their target force into a short-ranged part approximated by a cubic spline and a long-ranged Coulomb part they solve the overdetermined set of linear equations given by equation 2.8.

$$\sum_{\beta=1}^{K}\sum_{j=1}^{N_\beta}\left(-f(r_{\alpha il,\beta jl},\{r_{\alpha\beta,\boldsymbol{\kappa}}\},\{f_{\alpha\beta,\boldsymbol{\kappa}}\},\{f''_{\alpha\beta,\boldsymbol{\kappa}}\})-\frac{q_{\alpha\beta}}{r^2_{\alpha il,\beta jl}}\right)\mathbf{n}_{\alpha il,\beta jl}=\mathbf{F}^{\mathbf{ref}}_{\alpha il} \tag{2.8}$$

In equation 2.8, the $\{r_{\alpha\beta,\boldsymbol{\kappa}}\}$ correspond to the spline mesh at points $\boldsymbol{\kappa}$ for pairs of atoms of type α and β, while $\{f_{\alpha\beta,\boldsymbol{\kappa}} f''_{\alpha\beta,\boldsymbol{\kappa}}\}$ are spline parameters that ensure continuous derivatives $f'(r)$ at the mesh points and define the short-ranged part of the force. The subscript αil labels the i-th atom of type α in the l-th sampled atomic configuration. Solving these equations minimizes the Euclidean norm of vectors of residuals, and can be solved on a minimal set of atomistic snapshots using a singular value decomposition (SVD) algorithm [31]. By adding the Coulomb term to the short-ranged potential above, this technique allows for the inclusion of explicit electrostatics. The MS-CG model reproduces site-to-site RDFs from atomistic MD simulations in the as well as the density profile perpendicular to the bilayer normal in DMPC bilayers [151].

The MS-CG method has been extended to access still larger time- and length-scales using an approach called hybrid analytic-systematic (HAS) coarse-graining. In this method, the MS-CG force field provides the in-plane center-of-mass lipid interaction potentials while an analytic Gay-Berne (GB) liquid crystal model describes any inter-monolayer and out-of-plane interactions. The GB liquid crystal model uses ellipsoidal particles which interact with an anisotropic form of the Lennard-Jones 12-6 potential, and has been successfully applied to higher-resolution coarse-grained modeling with explicit water [243]. In the HAS approach, however, the GB interactions replace those with explicit solvent, providing significant computational efficiency. This model successfully self-assembles and reproduces the undulation spectrum, tensionless area per lipid, and area compressibility modulus in agreement with experimental measurements. It has been used to simulate a 200 nm liposome [16] and N-BAR protein remodeling of a liposome [13].

2.2.4 The energy-based approach of the Martini force field

The Martini force field developed by Siewert-Jan Marrink and co-workers eschews systematic structure-matching in pursuit of a maximally transferable force field which is parameterized in a "top-down" manner, designed to encode information about the free energy of the chemical components, thereby increasing the range of thermodynamic ensembles over which the model is valid. To date, it has been used to study a broad range of biological soft-matter systems described in sections 2.2.5, 2.3.2, and 2.4.

The Martini model employs a four-to-one mapping of water and non-hydrogen atoms onto a single a bead, except in ring-like structures, which preserve geometry with a finer mapping. Molecules are built from relatively few bead types which are categorized by polarity (polar, non-polar, apolar, and charged). Each type is further distinguished by hydrogen bonding capabilities (donor, acceptor, both, or none) as well as a score describing the level of polarity. Like the CMM-CG and MS-CG models, Lennard-Jones parameters for non-bonded interactions are tuned for each pair of particles. These potentials are shifted to mimic a distance-dependent screening effect, and increase computational efficiency. Charged groups interact via a Coulomb potential $U_{elec}(r) = \frac{q_i q_j}{4\pi\epsilon_0\epsilon_r r}$ with a low relative dielectric of ($\epsilon_r = 15$) for

explicit screening. This allows the use of full charges while reproducing salt structure factors seen in previous atomistic studies [212] as well as the hydration shell identified by neutron diffraction studies [63]. Non-bonded interactions for all bead types are tuned to semi-quantitatively match basic measurements of density and compressibility [187].

Bonded interactions are specified by potential energy functions which model bonds, angles, dihedrals, and impropers with harmonic functions, with relatively weak force constants to match flexibility of target molecules at the fine-grained resolutions.

$$
\begin{aligned}
V_b &= \frac{1}{2}K_b(d_{ij} - d_b)^2 \text{ (bond)} \\
V_a &= \frac{1}{2}K_a(\cos(\phi_{ijk}) - \cos(\phi_a))^2 \text{ (angle)} \\
V_d &= K_d(1 + \cos(\theta_{ijkl} - \theta_d))^2 \text{ (dihedral)} \\
V_{id} &= K_{id}(\theta_{ijkl} - \theta_{id})^2 \text{ (improper dihedral)}
\end{aligned}
$$

Here V represents the component of the potential energy function arising from bond, angle, dihedral, and improper dihedral contributions, the set $\{\mathrm{K_b}, \mathrm{K_a}, \mathrm{K_d}, \mathrm{K_{id}}\}$ represents the corresponding stiffness constants, and the set $\{\mathrm{d_b}, \phi_\mathrm{a}, \theta_\mathrm{d}, \theta_\mathrm{id}\}$ represents the equilibrium values for these interactions. Dihedral potentials are only implemented for peptide backbones. Alkanes are constructed to reproduce dihedral, bond, and angle parameters given by atomistic simulations in the GROMOS force field [59]. Simulations of small ice cubes surrounded by water show that Martini ice is in equilibrium with liquid water at 290K and melts within 5K of this temperature. Like many CG models, Martini water becomes supercooled as the temperature is lowered, failing to freeze spontaneously until 240K [210].

The defining feature of the Martini force field is the selection of non-bonded parameters which are optimized to reproduce thermodynamic measurements in the condensed phase. Specifically, the Martini model semi-quantitatively reproduces the free energy of hydration, the free energy of vaporization, and the partitioning free energies between water and a collection of organic phases, obtained from the equilibrium densities in both phases: $\Delta\mathrm{G}^{\mathrm{oil/aqueous}} = \mathrm{k_B T} \ln(\rho_{\mathrm{oil}}/\rho_{\mathrm{aqueous}})$. These calculations require long MD simulations of two-phase systems with very dilute concentrations of the target substance. Results agree to within $2\mathrm{k_B T}$ for many of these properties [212].

Systematic tuning to experimental partitioning free energies was used to select particle types for amino acids, represented with up to four beads which correspond to the polar character of the amino acid. Additionally, the pre-determined secondary structure modulates the character of the beads; backbone hydrogen bonds found in helices have reduced polar character [212]. While the building blocks for the Martini model were chosen to match thermodynamic data in general, any application of Martini model can be optimized by comparison to AA simulation, especially when designing bonded interactions to reproduce the protein structure.

In the Martini model, calculation of the partitioning free energy of water in hexadecane agrees with the measurement of $25\,\mathrm{kJ{-}mol^{-1}}$ observed in Fischer titration [282]. More broadly, a combination of experiments and predictive modeling efforts have quantified partitioning free energies for other molecules [84, 352] and contributed to the parameter choices made by Marrink *et al.* [212]. Spin-echo nuclear magnetic resonance experiments provide self-diffusion data for water and alkanes [81, 169].

The hydration free energy can be calculated by comparing the partitioning of target molecules between liquid and vapor water phases, while the vaporization free energy can be calculated from the simulation of liquid-vapor equilibrium. In these simulations concentrations of 0.01 mole fraction provide a reasonable approximation of infinitely dilute solutions. A comparison of thermodynamic properties by Baron *et al.* [25] showed that this CG model tends to overestimate the water-oil repulsion with free energies of vaporization and hydration which are systematically high, but still follow the correct trend.

The surface tension measures the free energy cost of adding area to the interface between solvents. Simulations were compared to drop volume tensiometry measurements with good agreement for water, vapor, and dodecane mixtures [8]. Electron density profiles from X-ray diffraction data on multilamellar arrays of bilayers provide a measure of the thickness. While neutron scattering is weaker, specific deuteration of different lipid components gives a local contrast agent without chemical modifications [232].

It is clear that the coarse-grained models described in this review often share the same target data. In this section, we have summarized the most important experiments which inform the Martini model in order to demonstrate the breadth of the physics which these data capture. While the three CG models reviewed in this paper produce extremely rich physics, they draw on data sets with orders of magnitude greater detail and information. Some of the key differences between these models are summarized in table 2.1. For this reason, there is no one "correct" method for incorporating these data into an accurate coarse-grained model. Indeed, it is impossible for any coarse-grained model to simultaneously match thermodynamic, structural, and kinetic features perfectly. Therefore, it is necessary for model developers to choose specific experimental results which are relevant to the desired application. In the next section we will take a deeper look at how these experiments can be used in a general coarse-grained model.

2.2.5 Reproducing experiments in coarse grained models

The coarse-grained model-development process described in the preceding sections is often iterative. Beginning with first-principles estimates for potential energy functions, these functions are initially parameterized from fundamental structure and thermodynamics measurements. Since the function between these parameters and higher-order experimental observables is unknown by definition, the model is then

Model	Key Methods	Key Target Data
CMM-CG [295, 301]	structure matching, energy matching, Boltzmann inversion, reverse Monte Carlo	density distributions, interfacial tension, area per lipid, bending modulus, area compressibility modulus, lipid order parameters
MS-CG [16, 151]	bottom-up force matching, variational optimization, cubic spline basis functions, hybrid analytic-systematic coarse-graining, screened electrostatics	atomistic site-to-site radial distribution functions, density distributions, bending modulus, area compressibility modulus, lipid diffusion rates
Martini [210, 212]	top-down energy matching, potential of mean force between phases, bilayer stress profile, free energy of lipid desorption or flip-flop, short-range electrostatics	free energy of hydration, free energy of vaporization, partitioning free energies, surface tension, interfacial tension, density distributions, bending modulus, area per lipid

Table 2.1: Summary of key modeling calculations and target data for representative coarse grained models discussed in sections 2.2.2, 2.2.3, and 2.2.4. This list is not exhaustive, however, and these models reproduce a wide range of experimental data.

iteratively tuned to match these experimental target data. In this section we will summarize the range of experiments which can validate a coarse-grained model.

Titration. One of the guiding principles for energy-based coarse-graining is that the model should show the correct partitioning free energies of its constituents, since this property describes the attraction or repulsion between phases. This quantity can be measured by titration methods which determine the density of one species in a fluid of the other, at equilibrium. For example, Karl Fischer titration of water in hydrocarbons [33, 282] is the target data for alkane-water interactions in the Martini model. Salt solution diffraction experiments show two hydration shells around these ions, and this justifies the inclusion of hydration-shell waters in coarse-grained ion models [63]. These methods may be extended to more complex systems, namely the partitioning of amphiphilic solutes onto membranes, using titration calorimetry [272, 338, 339].

Magnetic resonance spin echo. The self-diffusion coefficient is the diffusion rate of a particle when its chemical potential gradient is zero, given by $\mathrm{D}_\mathrm{i}^* = \mathrm{D}_\mathrm{i}(\partial \ln \mathrm{c}_\mathrm{i}/\partial \ln \mathrm{a}_\mathrm{i})$ where c_i is concentration, a_i is activity, and D_i is the diffusion coefficient of the species [148]. Because the diffusion rate reflects the chemical potential, modeling it correctly helps ensure the correct equilibrium density of the fluid. The self-diffusion coefficient for a particle type i may be measured in simulations via Einstein's diffusion equation, which relates the diffusion constant to the mean-squared displacement of the particle over time. It is given by the following equation, where N is the number of particles, $\mathrm{r}_\mathrm{j}(\mathrm{t})$ is the position of particle j and time t and the brackets denote an

ensemble average over all starting times.

$$D = \lim_{t\to\infty} \frac{1}{6Nt} \left\langle \sum_{j=1}^{N} \left(r_j(t) - r_j(0)\right)^2 \right\rangle$$

Self-diffusion coefficients can be measured with magnetic resonance spin echo [33, 81, 169]. In a nuclear magnetic resonance (NMR) spin-echo experiment, the self-diffusion is a proportionality constant which relates the logarithm of echo intensity and strength of the magnetic field gradient. These experiments make direct contact with CGMD simulations, which quantify diffusion rates by the slope of the mean-squared displacement. For example, the Martini polarizable water molecule matches experimental diffusion coefficients [371].

Neutron and X-ray scattering. The behavior of salts in water has a strong effect on the electrostatics of any coarse-grained system. Because neutrons are electrically neutral and interact only with the nucleus of a particular atom, they are subject to very short-ranged interactions and therefore penetrate the sample very efficiently. In scattering experiments, incident particles impart and receive momentum energy from the sample and scatter at a measurable angle. Measuring the number, angle, and energy (the scattering intensity) of the diffracted neutron provides a time- and space-dependent auto-correlation function via Fourier transform. In this way, neutron scattering is capable of measuring the density and time-dependent correlations of a fluid. This resulting dynamical structure factor can be directly compared with the radius of gyration in order to tune non-bonded interactions in simulations. These methods are reviewed by Fischer *et al.* [100]. Neutron scattering can provide a density profile of the bilayer, however only recent experimental methods have measured this density in fully hydrated bilayers [170].

X-rays may be used in place of neutrons in order to measure the electron density profiles of lamellae in vesicles and stacks. Small-angle X-ray scattering (SAXS) is used to quantify molecule size, low-angle X-ray scattering (LAXS) provides electron density profiles, and wide-angle X-ray scattering (WAXS) can measure in-plane features of the bilayer, including micro-domains [225, 232]. Both neutron and X-ray scattering can provide density profiles which describe the thickness of a bilayer. Combined with volume information, these can be used to estimate the area-per-lipid, which is a crucial structural feature of a bilayer simulation. The range of scattering experiments is often difficult to match to biologically relevant systems, and different methods often produce conflicting results. Specific algorithms for fitting these data to CGMD simulations are under development [250].

Significant effort must be devoted to faithfully reproducing the area-per-lipid in bilayer simulations not only to match experimental structure measurements, but because interfacial area and surface tension are conjugate thermodynamic variables. Recent re-parameterization of the CHARMM force field for lipids [166] optimizes the

lipid partial charges and Lennard-Jones parameters using QM calculations and experimental data to optimize the area per lipid in the tensionless ensemble. A survey of integrator parameters in GROMACS for a number of force fields [253] also provides guidance for choosing correct short- and long-range cutoffs and water models necessary to achieve the correct lipid areas and tensions in the correct phase.

Nuclear magnetic resonance. In order to characterize the order in lipid hydrophobic tails, one can measure the angle between a chemical bond and the bilayer normal. This gives the second-rank order parameter $P_2 = \frac{1}{2}(3\cos^2\theta - 1)$. Deuterium magnetic resonance (DMR) experiments use selective deuteration of carbon atoms in the hydrophobic tails. The order parameter is a function of the residual quadropole coupling value. DMR does not require sonication, in contrast with spin-label NMR measurements [287]. Coarse-grained lipids models do not include all of the available order-parameter data because they possess fewer degrees of freedom in their tails; instead, angles measured in CGMD tails may be tested for agreement with atomistic simulations, which must reproduce the order parameter.

2.2.6 Phase transition and tension measurements.

The area per lipid and surface tension are conjugate variables; all other things being equal, fixing one should fix the other, for a particular thermodynamic ensemble. Given that these properties depend strongly on the chemical composition (lipid type, hydration level) of a particular system, reproducing the phase transition temperatures is a useful indicator that the model is robust.

Most phase transition data is provided by measuring density via scattering or order parameters via NMR at various temperatures. To observe the phases directly, epifluorescence microscopy improves upon these measurements by making it possible to resolve microdomains and phase coexistence in monolayers [219]. Cryotransmission electron microscopy (cryo-TEM) and differential scanning calorimetry are also used to characterize the phase behavior of bilayers [303]. Densitometry, and acoustic measurements have been used to study dipalmitoylphosphatidylcholine (DPPC) multilamellar vesicles, extracting the temperature-dependence of phase transitions and plausible theories for the kinetics of these transitions [162]. This topic is described further in section 2.5.

In monolayer systems, the pressure-area isotherm can be measured via Langmuir trough or captive bubble surfactometer. These measurements can be mimicked in coarse-grained simulations in order to validate the model against experiments, quantify finite size effects, and investigate the effects of composition on the isotherm [22] and the dynamics of monolayer buckling [23, 301]. Since bilayers are unstable when lipids include fewer than 9 carbon atoms per tail, the observation of spontaneous pore formation in the Martini model indicates that the balance of hydrophobic repulsion and entropic repulsion qualitatively matches experiments. This balance is also

confirmed via calculation of the lateral stress profile, which is compared to atomistic simulation [212].

More generally, calculations of line tension in bilayers with mixed compositions agree semi-quantitatively with those measured by fitting fluorescence microscopy of mixed composition giant unilamellar vesicles (GUVs) to elastic theory, or micropipette aspiration [28, 332]. Another measure of bilayer stability is the water permeation rate. Permeability coefficients from vesicles can be measured via microturbimetry and differential scanning calorimetry. These rates also help set the time-scale of the coarse-grained simulation [210].

Simulations have the ability to measure local pressure from the first moment of the stress profile [314]. When averaged across the bilayer plane, this gives the stress profile, which may be compared to atomistic simulations. The integral of the stress profile along the bilayer normal z between z_1 and z_2 across the mid-plane z_0 also quantifies the intrinsic curvature C_0 via $\boldsymbol{\kappa} C_0 = \int_{z_1}^{z_2} (z - z_0)\Sigma(z)dz$. In this formulation, the local pressure tensor $\mathbf{P}(r)$ gives the lateral pressure profile by $\Sigma(z) = \frac{1}{2}(\mathbf{P}_{xx} + \mathbf{P}_{yy}) - \mathbf{P}_{zz}$. These calculations have made it possible to tune the balance of forces in Martini model bilayers in qualitative agreement with atomistic simulation [212]. Given an estimate for bending rigidity, the resulting spontaneous curvature values agree with fluorescence spectroscopy of supported lipid bilayers experiments for several lipid types [124]. Calculation of the local stress tensor also makes it possible to study the position-dependent stress profile generated by membrane-protein systems, such as the mechano-sensitive protein channel MscL [242, 370]. Applying a similar method to the study of the stress profile in atomistic simulations makes it possible to quantify the both the chemical and entropic contributions to the tension [190].

The integral of the lateral stress profile itself gives the surface tension ($\sigma = -\int_{z_1}^{z_2} \Sigma(z)dz$) of the bilayer. Because the tension and its conjugate variable area-per-lipid determine the phase of the system, there has been much debate on choosing the correct ensemble for biological simulations of lipid bilayers. In experiments, bilayers adjust their area per lipid to minimize contact between water and hydrophobic lipid tails. For bilayers with no spontaneous curvature, this gives free energy minimum $(\partial G/\partial A)_{A_0} = \boldsymbol{\gamma}$ which must be zero. Some have argued that coupling between area and thickness may introduce another variable, changing the free energy dependence to include non-zero tension. However, in practice, this coupling is very weak, and simulation studies find good agreement with experiments when using zero tension [98, 152, 209, 315].

Measuring elastic properties. Lipid bilayers possess an incredible combination of material properties which make them ideally suited to hosting biophysical processes and compartmentalizing the cell. Several experimental methods are able to measure the elastic properties of a bilayer. Many of these methods use the Helfrich model to interpret their results [131]. In this model, the membrane is treated as an infinitesimally thin elastic sheet with energy terms from bending, Gaussian curvature, and

surface tension. It is given by equation 2.9 where $\boldsymbol{\kappa}$ is the bending modulus, H is the mean curvature, H_0 is the intrinsic curvature (zero for a symmetric bilayer), G is the Gaussian curvature, $\bar{\boldsymbol{\kappa}}_{\mathrm{G}}$ is the Gaussian bending modulus, and $\boldsymbol{\gamma}$ is surface tension.

$$\mathscr{H}_{\mathrm{el}} = \int \left\{ \frac{\boldsymbol{\kappa}}{2} \left(2H - H_0\right)^2 + \bar{\boldsymbol{\kappa}}_G G + \boldsymbol{\gamma} \right\} dA \tag{2.9}$$

The Helfrich approximation forms the basis of many mesoscale models, briefly described in section 2.12. The bending rigidity $\boldsymbol{\kappa}$ describes the energy required to bend the membrane to a unit curvature, and is usually estimated at $\sim 20\mathrm{k_BT}$ for biological lipid bilayers. Because the Gaussian curvature is invariant under deformations, it is relevant only to topology changes in the membrane, i.e. vesicle fusion or phase transitions, or when the Gaussian rigidity varies along the spatial coordinate. The bending modulus (or bending rigidity) can be measured in a number of ways. The Fourier transform of the Helfrich Hamiltonian given by equation 2.10 measures the height-height auto-correlation function (otherwise known as the undulation spectrum). For bilayers with nearly zero surface tension ($\sim$0.1 N/m), it is possible to fit this function in the low-q regime where $\boldsymbol{\kappa}$ may be extracted as a pre-factor.

$$\left\langle |u_{\mathrm{und}}(q)|^2 \right\rangle = \frac{k_B T}{A(\boldsymbol{\kappa} q^4 + \boldsymbol{\gamma} q^2)} \tag{2.10}$$

Here q is the transformed variable defining the Fourier transform, and $\mathrm{u}^2_{\mathrm{und}}(\mathrm{q})$ is the height-height auto-correlation function. These fluctuations can be directly calculated from video phase contrast microscopy which quantifies the bilayer shape changes according to the fluctuation spectra given by equation 2.10 [239]. Shear flow experiments on giant vesicles can be used to measure the bending rigidity by relating the deformation of the vesicle to the flow in a theory which includes thermal membrane undulations [68]. Micropipette aspiration experiments provide a measure for bending rigidity as well as the area compressibility modulus which is given by $\mathrm{K_A} = \mathrm{A}_0^\star(\partial\boldsymbol{\gamma}/\partial\mathrm{A}_0)_{\mathrm{T}}$ where A_0 is the area per lipid and $\mathrm{A}_0^\star$ is the area per lipid at the free energy minimum. This study indicates that bending rigidity increases with the number of carbons [264]. In addition to the bending modes, contributions to the energy arise from peristaltic modes corresponding to fluctuations in the inter-leaflet distances, and protrusion modes corresponding to lipid motion normal to the bilayer plane (and therefore high wave-number modes in equation 2.10). In most cases, these modes are decoupled from the bending modes. Due to the high resolution provided by bilayer simulations, recent studies have characterized these undulations, peristaltic motions, and area compressibilities for comparison with experiment and a better understanding of membrane elasticity [45, 98, 127, 189, 209].

Cholesterol molecules, present in biological membranes, induce changes in bilayer elasticity and a reduction in the area per headgroup. This effect has been investigated using a number of atomistic force fields, and can be reproduced in coarse-grained models [212]. Experiments which use fluorescent quenching methods can be used to

Property	Experimental Method	Simulation Measurement
partition coefficient	titration calorimetry	potential of mean force of a particle pulled between phases
self-diffusion coefficient	magnetic resonance spin echo	mean-squared displacement
electron density profile	X-ray scattering	electron density
area per lipid	neutron scattering	area measurement (bilayer mid-plane)
lipid order parameter	nuclear magnetic resonance (NMR)	lipid tail angles to the bilayer normal
phase transition temperature	cryotransmission electron microscopy (cryo-TEM)	structure factor
pressure-area isotherm	Langmuir trough, captive bubble surfactometer	pressure tensor, area measurement
line tension	fluorescence microscopy of GUVs, micropipette aspiration	pressure tensor
bending rigidity	video phase contrast microscopy, GUV shear flow	height-height fluctuation spectrum

Table 2.2: Summary of corresponding experimental methods and simulation measurements which may be used match key physical properties of soft matter systems.

investigate the phase coexistence in bilayers with cholesterol [67,223], providing useful target data for cholesterol coarse-graining.

2.2.7 Summary

In this section we have reviewed many of the experiments which inform coarse-grained models for soft-matter systems; these methods are briefly summarized in Table 2.2. It is important to note that coarse-grained models can be designed to reproduce many other experimental methods beyond the scope of this review. Soft-matter experiments are rich in information about the structure and dynamics of the components of many biological systems, however it is far from easy to design models which reproduce these quantities. Even when the experimental data are clear, there is no guarantee that a coarse-grained model will be able to capture the nuance and context of these experiments without careful adjustment and attention to the limits of the model. The limits of coarse-grained simulations of soft matter will be discussed in the next section.

2.2.8 Assessing CGMD model performance

In the preceding sections we have described the construction and verification of representative coarse-grained models. Comparison to basic experiments verifies that the model is capable of matching basic physical properties of soft matter systems. In the remainder of this review we will discuss the validation of these models on more complex biomolecular systems. Any coarse-grained model is only useful insofar as it can reproduce the physics of a complex system. In this sense, the model must be carefully designed to match experiments, and more importantly, answer a clear question about a biophysical system. In this section we will note some of the limitations inherent to the coarse-grained modeling approach.

The clearest limitations of a coarse-grained model are the result of discarded degrees of freedom. For example, the earliest iterations of the CMM-CG model lacked explicit electrostatic interactions. The standard (non-polarizable) Martini model includes electrostatics, however they are highly screened, and thus imprecise compared to all-atom methods. In the residue-based coarse-graining approach used to model membrane remodeling by N-BAR (see section 2.5.1), an artificially low dielectric constant was necessary to reproduce the electrostatic interactions between the protein and bilayer [9, 10, 373].

It has been noted that the Martini method is a free energy method and therefore includes significant entropy loss owing to the loss of degrees of freedom in the coarse grained mapping. Enthalpy terms compensate for this, however the entropy/enthalpy balance may be upset, affecting temperature dependence. Secondary structure is also static in this model, however many applications can show the relative movement of secondary structure elements. For example, the tension-driven activation of a mechanosensitive channel can be resolved by CGMD [199, 370]. Both the inability to model protein conformational change and the challenge of reproducing bilayer physics without an accurate entropy-enthalpy balance are both consequences of the loss of detail in a CGMD model, and are therefore common to all of the methods discussed in this review. Similarly, a coarse representation of amino acids often obscures the chemical detail responsible for protein function. We will describe CGMD models for a number of protein-mediated processes in section 2.4.

A major objective of coarse-grained modeling approaches is to design transferable or universal force fields capable of modeling novel systems with minimal modification. All-atom force fields do not require extensive tuning, partly because they already contain explicit degrees of freedom for first principles-based interactions. However, not all degrees of freedom are created equal, and a major challenge of coarse-graining is deciding which degrees of freedom are essential to the system of interest. The force-matching (MS-CG) approach addresses this limitation by computing a coarse-grained force field directly from an atomistic trajectory of the target system. A recent survey of MS-CG peptide models indicates that transferability between systems is strongest in the low-energy regions of the free energy landscape [330]. The contrasting bottom-up and and top-down coarse-graining methods highlight the tension between including

unnecessary detail, and designing universal models. As with any modeling strategy, the choice of a highly-tuned yet non-transferable model versus a general, transferable model will depend on the system of interest.

Coarse-grained molecular dynamics also often fail to reproduce both correct or even self-consistent kinetics. Martini simulations use a timestep of 20-40 fs, which is effectively three- or four-fold longer in real time due to the smoothed interactions between CG beads. This speed-up factor may change with system composition. Since CGMD force fields may independently reproduce either the potential energy or free energy of different parts of the underlying system, they may produce incompatible kinetics within the same system. By collapsing many degrees of freedom into relatively few, this model makes a number of compromises. Lack of atomistic detail prevents strong hydrogen bond networks from forming, and this in turn generates lateral diffusion rates which are much higher than normal. For that reason, the time-scale of a CG simulation must be calibrated to the diffusion in an AAMD simulation to *a posteriori* determine the duration of the simulation. As a result of model approximations, the relative kinetics for subsets of particles in the same CGMD simulations may not be preserved. Loss of detail also prevents accurate conformational sampling of proteins and accurate reproduction of lipid order factors. Many of these disadvantages may be mitigated by comparison to atomistic systems, experiments, and even mesoscale continuum models. And, despite these limitations, coarse-grained models are able to predict and describe an amazing array of biomolecular systems. Before we discuss these applications, we must first explain the ways in which protein detail can be added to the system.

2.3 Modeling Proteins

The challenges in reproducing the physics of soft-matter systems with a coarse-grained force-field are significantly larger when introducing proteins. In addition to introducing a multiplicity of additional chemical interactions with lipids, it is a significant challenge to successfully model the internal structure and dynamics of membrane proteins. All-atom molecular dynamics simulations provide the best means of capturing these motions, and recent years have seem sophisticated methods for matching these simulations to experiments, NMR in particular. Thanks in large part to the study of all-atom MD simulations of proteins, coarse-grained force fields have incorporated parameters for amino acids. The resulting CGMD applications rely on the wide body of atomistic protein simulation. However, comparing the results of simulations at atomistic and coarse resolutions tells us precisely which kinds of chemistry and nanoscale physics manifests itself at much larger length scales, and how this information is propagated. This information can often be used to guide the design of more useful atomistic simulations.

2.3.1 Atomistic simulations of proteins

The earliest simulations of proteins in atomistic detail employed first principles to study proteins *in vacuo* and in solution. Since the first simulation of a protein in 1976 [218], protein simulations have characterized the conformations of proteins, DNA and other biomolecules. A comprehensive review of atomistic protein simulations is beyond the scope of this review, however summaries of significant progress in the field have been compiled in a series of recent reviews [1, 19, 83, 86, 112, 157, 158, 344]. The CHARMM, GROMOS, AMBER, and OPLS-AA protein force fields are each capable of simulating biomolecules with similar, but not exact results. Recent studies have systematically compared these force fields to each other and relevant experimental in order to validate both the molecular dynamics method in general, and the force field tools in specific [30, 123, 192, 252, 253, 311]. The methods for designing coarse-grained force fields described in section 2.2 rely heavily upon these force fields.

Atomistic simulations are more "literal" than CGMD simulations, in that they seek to reproduce a particular thermodynamic ensemble with no ambiguity (beyond that of the reference state) in the definitions of physical quantities such as length, time, force, and energy. To this end, they employ first principles often gleaned from quantum mechanical theory and experiment, namely electrostatic potentials, dipole moments, and dimerization energies. Atomistic protein simulations also seek to reproduce a host of target data, also used in CGMD parameterization, including spectroscopy data, thermodynamic data such as solvation free energy, and heats of vaporization, and X-ray, electron, and microwave diffraction structures. With meticulous matching to experiment and quantum mechanical theory, and consequent application in parameterization of CGMD force fields, atomistic simulations provide a filter through which these experimental data inform CGMD models.

Having matured in recent years, atomistic simulations are now capable of accurately modeling protein folding [55, 78, 119, 167, 283], predicting protein-ligand docking to guide drug design [58, 65, 86, 188, 198, 279, 317, 360], and understanding protein mechanical properties [178, 186]. Notably, simulations have recently been used to study enzyme binding processes [51], cooperative protein folding [185], the molecular motors [164], solvent behavior in the ribosome exit tunnel [200], and protein-DNA binding [177, 197].

Enhanced Sampling Methods

Despite these successes, atomistic molecular dynamics simulations are tempered by the primary disadvantage of atomistic protein simulation: accessing physically relevant time-scales. While much work has been devoted to parallelization algorithms, use of graphical processing units (GPUs), and the development of larger, more specialized, and massively distributed supercomputers [167], these efforts are unlikely to provide access to large biomolecular systems for more than milliseconds of real time.

The most straightforward way to improve the sampling of a molecular dynamics simulation is to simulate multiple copies of the same system using slightly different starting configurations. This provides a more robust sample, albeit at the same cost as the original simulation. To efficiently extend the atomistic methodology to longer time-scales, a collection of enhanced sampling methods have been developed. For example, graph-based geometric methods, probabilistic road maps, and Markov models may be used to better understand protein dynamics and kinematics by discarding uncorrelated, high-frequency atomic motions [112]. Elastic network models and normal mode analysis methods reveal collective motion and allosteric mechanisms in good agreement with NMR and X-ray scattering data [19], and often in conjunction with standard coarse-graining [202, 247, 378]. Methods such as transition path sampling, transition interface sampling, forward flux sampling, and weighted ensembles provide additional access to longer time-scales in atomistic systems [383]. In replica exchange molecular dynamics, multiple weakly-coupled simulations of the same system are exchanged between temperatures to escape kinetic traps [325]. Biasing potentials may be used to generate non-Boltzmann-distributed ensembles from which equilibrium properties may be calculated in steered molecular dynamics [110]. Metadynamics [174], temperature-accelerated molecular dynamics [208], and other free energy perturbation methods [58, 73] and even Monte Carlo methods [155, 268, 376]. Perhaps the most thorough extension of atomistic molecular dynamics is realized by coupling simulation with NMR measurements, which enhance the sampling of the simulation at the longer time-scales possible in experiments [280, 377].

While these methods seek to extend the range of a particular simulation method, it is also possible to optimize existing coarse-grained approaches using a relative entropy-based method [293]. Many of these enhanced sampling methods discussed in this section may also be used in combination to tackle specific biological modeling problems. Common to each is a reduction in the number of degrees of freedom, making it possible to simulate large atomistic systems in great detail. In that sense, coarse-grained molecular dynamics simulations are another enhanced sampling method, in this case ideally suited to extending the size and duration of protein simulations.

Atomistic simulations of membrane proteins

Atomistic simulations of membrane proteins provide a direct link between coarse-grained descriptions of large soft-matter systems and the high levels of detail available from atomistic simulations. Reviews of membrane-protein simulations include descriptions of both fine and coarse resolutions, often describing the ways in which these simulations can be linked [11, 191].

The study of G-protein-coupled receptors (GPCR) provides the prototype for atomistic investigation because the action of these transmembrane proteins is modulated by membrane environment. A recent review of atomistic GPCR simulations [277] emphasizes the importance of developing accurate models for lipids and protein oligomerization in order to produce models which can inform experiments

and future drug design. In a similar application, circular dichroism experiments at high temperature showed that some transmembrane peptides are thermostable, allowing elevated temperature simulations which quantified the pathway by which these proteins partition into membranes [339].

Due to the added difficulty of accessing biologically relevant time-scales for both proteins and bilayers in combined systems, atomistic membrane protein simulations are often combined with other methods. Atomistic simulation augmented with Monte Carlo methods was used to accelerate lipid equilibration to investigate hydrophobic mismatch near helical peptides [372]. Amphipathic polymers which stabilize membrane proteins in solution employed combined all-atom and coarse-grained resolution with a back-mapping scheme to probe particle assembly, in agreement with small-angle neutron scattering [249]. In a study of phospholipase, coarse-grained simulations served as seeds for atomistic simulations which improved conformational sampling of the peptide [359]. Other studies used atomistic simulation to test lipoprotein complexes against SAXS data [299]. For bilayer systems in general, reverse coarse-graining makes it possible to connect CGMD models to low angle X-ray scattering (LAXS) measurements of bilayer geometry by confirming these geometries in atomistic simulations [250]. Likewise, the back-mapping from simulations of the antimicrobial peptide alamethicin from the Martini model to the CHARMM27 all-atom force field helped confirm that this peptide loses its helical character during aggregation, in agreement with NMR measurements [329].

The abundance of atomistic membrane-protein simulation reflects the usefulness of high-resolution simulation of membrane-associated proteins. Coarse-grained simulations of proteins serve to extend these studies to larger systems and biologically relevant time-scales.

2.3.2 Parameterization of coarse-grained proteins

Introducing protein detail to a coarse-grained force field requires an accurate model for both the structure and dynamics of the protein itself, as well as the interactions with surrounding lipids and solvent which remain faithful to experimental observations. In this section we will summarize the development of coarse-grained protein models, the experiments they match, and their integration into popular force-fields.

Structure-based coarse-grained protein modeling

While coarse-grained simulations have difficulty reproducing secondary structural transformations, it is possible to recover accurate conformational sampling by a reverse-transformation from the CGMD level to the atomistic one. Atomistic simulations of back-mapped CGMD structures can recover the conformational properties of the original atomistic system. In this procedure, back-mapped atoms are randomly placed near their corresponding coarse-grained bead. The cener of mass of

these atoms is then restrained to the position of the coarse-grained bead. The system may be relaxed by a simulated annealing procedure to minimize large or unphysical forces, stochastically sample the conformation space, and gradually introduce inter- and intra-molecular potentials that are consistent with the all-atom model. This method has been used to generate atomistic structures of simple peptides and transmembrane proteins from coarse-grained trajectories [36, 276, 351]. The back-mapping procedure also quantifies the information loss from coarse-graining, providing a useful way to validate a CG model against a more robust atomistic force field or extend a CG trajectory to include greater detail.

The earliest coarse-grained proteins were based on the Gō model in which each amino acid is represented by a single bead which attracts or repels the other beads in the model according to interactions in the ground state. These models sought to investigate protein folding mechanisms [184]. Many of these models use non-standard molecular dynamics techniques. For example, discontinuous molecular dynamics was used to study the aggregation of peptides in implicit solvent [90, 236], while Brownian dynamics simulations have been used to study crowding effects in the GroEL-GroES chaperonin system [89]. Elastic network models have found wide application in flexible fitting methods which add detail to low resolution cryo-EM measurements [20]. Coarse-graining with empirical potentials is a common method for protein structure prediction and protein design [60, 130].

Many of these coarse-grained approaches lack the chemical specificity necessary to study protein aggregation and association with lipid bilayers. The coarse-grained models described in section 2.2 have been modified to include this detail in a number of ways.

There are many ways to generate intermolecular interactions for CG proteins. A common data set for generating non-bonded parameters in coarse-grained proteins is surface tension and density data for side-chain analogues. The surface tension characterizes the energetics of amino acid interactions at a vacuum interface; this quantity is a useful proxy for the attractive forces that mediate the interactions with water-lipid interfaces. The amino acid model developed by Klein and coworkers uses surface tension as target data and showed that the solvent accessible surface area (SASA) of the resulting protein models agreed with atomistic simulations [75]. The model was also able to recognize the native protein structures from a set of decoys. In another approach, Han *et al.* fit dihedral potentials for a test set of small molecules, and tuned the force field to match self-solvation free energies and hydration free energies across a representative sample of organic molecules, finding good agreement to atomistic simulation [126]. In contrast, the model by Basdevant *et al.* [26] used a r^{-6} repulsive term with a Gaussian attractive term to represent non-bonded forces between amino acids, parameterized from atomistic simulation.

In an early extension of the Martini model [210] to proteins, Schulten and coworkers used residue-based coarse-graining (RBCG) as an intermediate scale in a multiscale model for membrane bending by BAR domains [9], described further in

section 2.5.1. In this system, as well as applications to lipoprotein particles, the authors selected bead types from the Martini building blocks according to polarity and charge [298, 299]. The authors also made minor modifications to protein bonded parameters to match atomistic simulations of their target systems [299]. Further coarse-graining by the shape-based coarse-graining method (SBCG) extended the model further but discarded electrostatics, modeling lipids with only three beads and describing proteins with an elastic network model derived via iterative Boltzmann inversion [9, 105, 373].

Coarse-grained simulations developed by Voth and co-workers employs a Hybrid Analytical Systematic (HAS) model parameterized according to the MS-CG algorithm described in section 2.2.3. The HAS model is based on the Gay-Berne ellipsoid particle model which allows a single bead to represent a lipid. Lipid-protein interactions are modeled with a single Lennard-Jones term, and electrostatics are modeled with exponential screening according to Debye-Hückel theory, which is used to recover the polarizability lost during coarse-graining. A recent application of this model to N-BAR proteins tuned Lennard-Jones parameters for interactions between the membrane and amphipathic helix to match the atomistic peptide folding free-energy and empirical binding calculations [13, 226, 322]. Other recent modeling efforts have been extended to include DNA and RNA [117].

Martini Proteins

In the Martini force field, amino acids are mapped onto as many as five beads, one of which represents the polypeptide backbone. Residues with rings (His,Phe,Tyr,Trp) use a finer mapping and improper dihedral terms to preserve the topology of these rings. Intra-amino acid bonded potentials – that is, bonds, angles, and dihedrals – have equilibrium values equal to the average of distributions measured from all bonded amino acid pairs found in a representative sample of 2000 proteins from the protein data bank (PDB). These were sorted by helix, coil, and extended secondary structure, as measured by the DSSP ("define secondary structure of proteins") prediction algorithm [156] so that the Martini model includes the effect of secondary structure on the apparent hydrophobicity and polarity of its constituent particles. This secondary structure remains fixed through the simulation, therefore the Martini model cannot sample secondary structure changes. However, it is possible to reconstitute atomistic details from a coarse-grained simulation using a "back-mapping" procedure similar to simulated annealing. This method has been demonstrated on simulations of the WALP transmembrane protein [276].

Both protein-protein and protein-lipid interactions are modulated by the non-bonded parameters, which are assigned via selection of sidechain bead types. This procedure has several parts. First, the bead types must partition between oil and water phases consistent with experiments in which the distribution coefficients of amino-acid analogs were calculated with NMR [257] and dynamic vapor pressure measurements [363]. These experiments are similar to those used to generate the

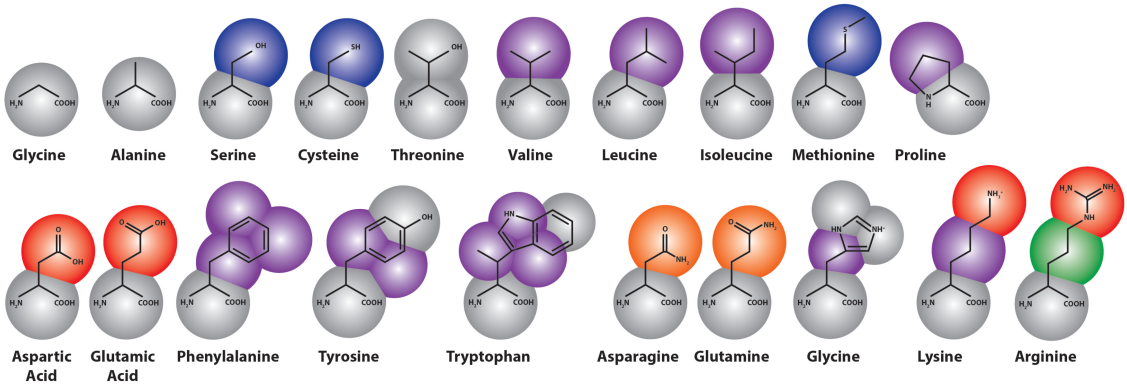

Figure 2.3: Coarse-grained representation of the Martini model extension to amino acids [228], colored by bead type (where purple is apolar, blue and green are intermediate, gray and orange are polar and red represents charged particles).

partitioning free energies of the Martini alkane building blocks. In this case, a free energy perturbation method (FEP) called thermodynamic integration was used to provide a more sophisticated measure of the $\triangle G^{oil/aqueous}$ for each side-chain analog. In thermodynamic integration, a coupling parameter λ weights the addition of a single particle to the Hamiltonian. The derivative of this Hamiltonian can be numerically integrated to obtain the free energy difference for adding that particle. In this case, the partitioning coefficient is calculated from thermodynamic integration of the addition of side chain analogs to water and decane boxes [70]. A modification to the Bennett acceptance ratio method was used to calculate the free energy difference and associated errors [302].

A second validation of the protein force field is given by the potential of mean force (PMF) of the amino acid interactions with a lipid bilayer. The potential of mean force quantifies the free energy landscape according to a fixed coordinate, in this case given by the distance of the amino acid from the center of the bilayer. Umbrella sampling and the weighted histogram analysis method (WHAM) [171] were used to generate the PMFs for comparison to atomistic PMFs calculated by MacCallum *et al.* using the OPLS protein force field [204].

It is generally difficult to compute a PMF analog from experiments, especially for peptides. However, a recent study calculated the PMFs of penta-peptides of the form Ac-WLXLL (where X is any one of the twenty natural amino acids). The free energies of partitioning of the variable residue were calculated using a thermodynamic cycle which included the free energy change of displacement from the membrane via umbrella sampling and the alchemical introduction of the particle via thermodynamic integration [305]. These values were consistent with a measure of hydrophobicity called the Wimley-White scale, which groups amino acids into five categories according to their partition coefficients as measured by a combination of equilibrium dialysis and quantitative reverse-phase HPLC for peptide hydrophobicity at palmi-

toyloleoylphosphatidylcholine (POPC) bilayers [362].

In addition to reproducing the correct association with lipid bilayers, proteins must associate with themselves in a physical way. To that end, association constants given by $K_{ij} = \frac{1}{C} \times \frac{P_{bond}}{P_{free}}$ where C is a concentration correction and the P_{bond} and P_{free} are the probabilities of finding a pair in a given bound or unbound state [70]. These were distinguished by a solvent accessible surface area (SASA) calculation, in which areas below a particular threshold indicate that the residues are contacting. The dimerization free energy was also computed directly from equation 2.11 according to a radial distribution function given by the PMF over the distance between side chains [70].

$$\triangle G^{dim} = -k_B T \ln \frac{4\pi R_{max}^3 \int_0^{r_c} r^2 g(r) dr}{3v^{\phi} \int_{r_c}^{R_{max}} r^2 g(r) dr} \tag{2.11}$$

In this equation, $g(r) = e^{-PMF(r)/k_B T}$, R_{max} is the maximum distance between monomers, r_c is the dimer-monomer cutoff distance, and $v^{\phi}$is the standard volume, 1 mol-L^{-1}. The dimerization free energy agrees with that measured in atomistic simulations in OPLS and GROMOS in test systems of amino acid pairs [69]. The dimerization free energy cannot be measured directly by experiment, however a host of knowledge based potentials have been designed to quantify the $\triangle G^{dim}$ by ranking the co-occurrence between pairs of amino acids in known protein structures [227]. While these results may be influenced by the presence of a hydrophobic environment inside the body of the protein, it nevertheless provides a useful benchmark.

Recent improvements to the Martini model's protein parameters [70] have included refinements to the free energy methods described above, as well as slight changes to the bead types for non-charged polar residues. Additionally, the development of a polarizable Martini force field has made it possible to improve the polar but neutral Asn, Gln, Ser, and Thr residues. The polarizable extension to the Martini model includes a fluctuating dipole resembling the Drude oscillator in which two partial charges are tethered to a polarized bead, and interact via a Coulomb function only. The dipole momentum is adjusted via harmonic angle and distance potentials. The resulting model therefore includes orientational polarizability, which makes it possible to more accurately model electrostatic interactions, particularly in transmembrane pores and antimicrobial peptide applications [371].

In this section, we have described the ways in which the individual Martini building blocks were adapted to include the interactions between systems of proteins and lipids. Careful parameterization of these building blocks ensures that the model is capable of reproducing the complex behavior of many biomolecular systems. Applications to richer biological problems validates the model while providing molecular insight into experiments. For example, an extension of the Martini protein force field to model the aggregation of amyloid-like peptides [290].

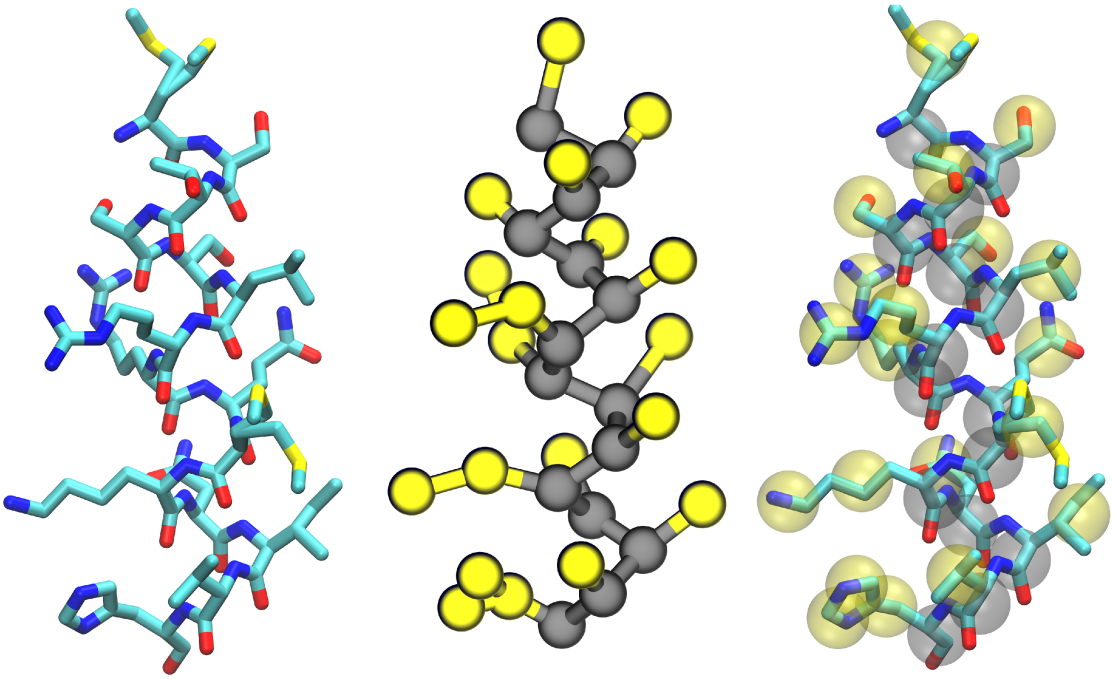

Figure 2.4: An example protein helix in all-atom (left) and Martini coarse-grained representations (center, backbone beads in gray and side-chain beads in yellow) with both images merged (right) to show how the fine-grained structure is mapped onto the coarse-grained beads. This image was rendered with Visual Molecular Dynamics [142].

2.3.3 Improvements to protein models

While most of the efforts to incorporate proteins into coarse-grained simulations of soft matter has rightly focused on parameterizing the interactions of amino acids with water, lipids, and each other, cutting edge development of more advanced force fields has explored the possibility of capturing conformational sampling in coarse-grained models. If CGMD can accurately explain protein-bilayer interactions, peptide self-assembly, and protein binding, then it is reasonable to see whether these methods can also model internal structural changes that guide the biological functions of many proteins. This is an extension of the previous challenge: to accurately capture peptide-peptide interactions and their relationship to the intra-molecular (bonded) forces which makes the complex conformational equilibria of polypeptides possible.

2.4 Membrane-Protein Applications

In sections 2.2 and 2.3 we have shown how coarse-grained molecular dynamics simulations are constructed from chemical components which match molecular experiments. We see that these models are capable of reproducing the fundamental properties of the systems they mimic, including protein structure and dynamics, peptide-bilayer interactions, and the geometry and elasticity of membranes. In this section we will show how these models can reproduce the behavior of far more complicated biophysical systems, yielding insight to experiments and elucidating the molecular mechanism by which proteins interact with cell membranes.

2.5 Simulations of biological membranes

While the parameterization of any soft matter CGMD force field includes validation of membrane fluidity and geometry, these models must also capture the condensed matter properties of biological membranes. Any simulation which seeks to quantify protein-mediated membrane properties must be capable of mimicking the properties of a bare membrane.

The earliest Martini model simulations sought to reproduce the complex phase behavior of bio-mimetic membranes. Simulations of phosphatidylcholines of differing lengths separated into gel and liquid phases in a small temperature range and semi-quantitatively matched experimental phase transitions, which were modulated by their relative concentrations [96]. Formation of non-lamellar phases is essential for modeling the first steps of membrane fusion, and can be induced in Martini bilayers by varying temperature and hydration levels. Simulations of mixed dioleylphosphatidylcholine with dioleylphosphatidylethanolamine (DOPC/DOPE) bilayers mimicked hexagonal, inverted hexagonal, and rhombohedral phases according to X-ray diffraction experiments [263, 367] with precise control of hydration levels [214]. Fur-

ther tests have studied the temperature dependence of the the fluid-gel phase transition for DPPC bilayers, providing thermodynamic parameters, namely estimates for the line tension and entropy difference of the fluid-gel interface [213]. These measurements connect simulations with non-equilibrium experiments using X-ray diffraction with pressure-jump relaxation [286], and temperature scanning calorimetric, densitometric, and acoustic measurements, providing insight into the kinetics of these phase transitions [162]. A study of the effects of lipid compositions identified lipids with varying levels of saturation that lower the line tension at domain interfaces in the bilayer [28, 281, 332].

Having established that coarse-grained bilayers exhibit the phase behavior features of bilayers observed in experiments, researchers began to study peptide-bilayer interactions. Simulations of the influenza HA fusion peptide revealed a bi-continuous cubic phase by stabilizing stalk/pore complexes in agreement with *in vitro* measurements that show the peptide lowers the lamellar-to-inverted hexagonal phase transition temperature [108]. Simulations of antimicrobial peptides show that they adhere to bilayers, assemble into amphipathic nanotubes, and extrude lipids from the bilayer [161]. Antimicrobial peptide aggregates also induce long-range order in phosphatidylglycerol domains, in agreement with atomic force and TIRF experiments [254].

Having demonstrated that the Martini model accurately predicts these phase transitions, these models have been extended to systems which simulate vesicle fusion. Initial studies of vesicle fusion events show a branched pathway for fusion in which stalk-like structures may either form a fusion pore or slowly fuse via a hemi-fused state [159, 211], and subsequent study estimates the free energy barrier to fusion and show that kinetics of the early stages of fusion are determined by the energy of solvent-exposed lipid tails [309]. Simulations of lung surfactant protein show the mechanism by which they facilitate the formation of a lipid bridge in vesicle fusion [23, 24, 85]. Likewise, monolayer simulations in the Martini model [22, 23] and a recent coarse-grained force field by Shinoda *et al.* [301] have been used to explore monolayer buckling. In addition to coarse-grained MD, a mesoscale method called dissipative particle dynamics can be used to study vesicle fusion [300]. Mixtures of double-stranded DNA and lipids called lipoplexes are potential transfection vectors which have a lower toxicity than viral vectors. Martini simulations have matched observations from SAXS and other experiments which observed a lamellar to inverse-hexagonal phase transition [64, 85].

Lipid rafts have been known to provide additional compartmentalization of the cell membrane, serving to organize and direct the action of biomolecular complexes. CGMD simulations have shown that thickness mismatches between phases are communicated to opposing leaflets and assist in guiding rafts together or stabilizing a registered geometry [245]. Martini simulations show broad agreement with NMR measurements [269] which show that cholesterol preference for saturated tails drives phase separation.

Finally, simulations of biological membranes are not limited to models which in-

clude explicit solvent particles. For example, Deserno and co-workers have applied a solvent-free coarse-grained model with three beads per lipid [62] to study membrane remodeling by generic viral capsids or colloids, finding that such particles attract due to their induced membrane curvature [267].

A generic implicit model by Brown and co-workers has been used to study the effects of protein inclusions in lipid bilayers [47, 361]. Likewise, a solvent-free combination of the MS-CG method with short-ranged coarse-grained potentials has been used to simulate liposomes [322]. Recent efforts have tuned implicit coarse-grained bilayer models to reproduce bilayer stress profiles [312]. The range of coarse-grained approaches for simulating lipid bilayers makes it possible for researchers to select the desirable level of detail necessary to study a biological system of interest. In the next section we will review one such example in which protein-protein and protein-membrane interactions at different length scales act in concert to remodel the membrane.

2.5.1 Modeling membrane bending

Membrane remodeling by Bin/Amphiphysin/Rvs (BAR) domains provides an archetypical application of the combined soft-matter and protein coarse-graining methods described above [9, 10, 373]. Highly conserved, ubiquitously expressed BAR domains bend cellular membranes from the cytosol, participating in endocytosis, vesicle fusion, cell-cell fusion, and also apoptosis. *In vivo*, members of the BAR domain, such as N-BAR, form high-curvature tubes with a low radius of ($\sim$50 nm), while *in vitro* they form vesicles from liposomes. N-BAR includes an N-terminal amphipathic helices (helix-0) which may either scaffold the charged lipids which contact it, or induce an area asymmetry (or both) in service of bending the membrane.

All-atom molecular dynamics simulations of N-BAR bending a 7:3 dioleoylphosphatidylcholine with dioleoylphospatidylserine (DOPC/DOPS) membrane showed that the protein stabilized local curvature [9, 10, 373]. These simulations were used to quantify the flexibility and tertiary structure of a single domain. These results were used to construct bonds between coarse-grained beads via Boltzmann inversion in the so-called residue-based coarse-graining (RBCG) method, which uses non-bonded forces adapted from the original Martini lipid force field [9, 105, 299]. A low dielectric constant ($\epsilon = 1$) was necessary to reproduce the strong electrostatic contacts responsible for bending the membrane to adhere to the N-BAR surface.

The RBCG simulations tested staggered and ordered arrangements of six N-BAR domains on a membrane patch, finding that only the former yielded a stable, global bending mode. Further coarse-graining under the shape-based (SBCG) method, in which lipids of $\sim$150 atoms are represented by two beads connected by a harmonic spring, extended these simulations to still larger time scales (5 μs). Non-bonded LJ parameters in this model were tuned to reproduce area-per-lipid and bilayer thickness measurements. The SBCG simulations also provided an estimate of the membrane

bending modulus by measuring the force exerted on the edges of a bilayer tube, confirming that the SBCG N-BAR domains have sufficient energy to bend the membrane. To explore the dynamics of membrane bending, the measured curvature from SBCG provided the intrinsic curvature parameter for a continuum elastic membrane model which included membrane bending, stretching, and viscous drag forces [9]. Testing the sensitivity of these parameters showed that drag forces determine the damping of the remodeled membrane. This study matched the structure seen in cryo-TEM images of the tubules which failed to locate the precise ordering and orientation of the four inserted amphipathic helices but found that helix-0 interactions are degenerate, dynamic, and necessary for stabilizing the lattice [226]. In the absence of dimers of amphipathic helices, BAR domain oligomers show reduced orientational order and cannot form a stable lattice. CGMD simulations of liposomes shows that N-BAR forms a lattice with both higher order and higher density, suggesting that this density is necessary to induce high, stable curvature *in vivo* [66].

In a related system, the protein epsin is hypothesized to sense and induce curvature while recruiting accessory proteins in the early stages of clathrin-mediated endocytosis (CME). The highly-conserved epsin N-terminal homology domain (ENTH) binds phosphatidylinositol 4,5-bisphosphate (PIP_2) by inserting an N-terminal amphipathic helix similar to that found in the BAR domains (helix-0). This helix becomes helical upon membrane binding according to circular dichroism, and ENTH domains were found to tubulate liposomes *in vivo* [101].

According to spin-labeled electron paramagnetic resonance (EPR) spectroscopy measurements, helix-0 becomes structured when binding the PIP_2 headgroup [172]. Further EPR studies and AAMD simulation provides a more detailed description of helix-0 penetration and key distances between ENTH domains, and these parameters were integrated into a CGMD model for ENTH-induced tubulation [173]. This model showed that heterogeneous lattice reduces the anisotropy of the spontaneous curvature and tends to frustrate tubule formation, thus explaining experimental observations that high initial concentration of ENTH domains tends to form vesicles (which have isotropic curvature by definition), while adding ENTH to preformed membrane tubules crystallizes proteins in a more ordered, helical pattern [173]. It is important to note that the protein epsin and the family of BAR domains are only two examples of a diverse set of membrane remodeling proteins. Other proteins have been shown to induce curvature as well, likely via different mechanisms. For example, the protein α-synuclein, the protein implicated in Parkinson's disease, induces negative Gaussian curvature according to coarse-grained molecular dynamics simulations which were then matched to low-angle X-ray scattering data which highlight the thinning effect on the bilayer [48]. The mechanism of protein-induced curvature sensing and generation has been recently reviewed by Baumgart *et al.* [27].

2.5.2 Lipid bilayers support protein assembly and function

While many proteins actively remodel bilayers during biological processes, in contrasting mechanisms the bilayer (or liquid-liquid interface) provides a substrate for protein assembly. The CMM-CG model developed by Klein and coworkers found early application in modeling the interaction of synthetic hydraphiles with bilayers [319] and the assembly of peptide nanotubes at oil-water interfaces [163]. Transmembrane peptides demonstrate the ability to sort lipids by chain length when they are smaller than the bilayer thickness [237,238]. These simulations show that the peptide induces a meniscus which depletes water from the peptide and encourages bilayer fusion, thus explaining experiments which show that these peptides can induce a transition from the lamellar to inverted phase. The protein force field extension by DeVane *et al.* was used to simulate the behavior of hydrophobins, proteins which self-assemble at air-water interfaces [57] in agreement with experimental measurements of adsorption and desorption free energies of comparable molecules. Tilt angles and helix-helix association of transmembrane peptides modeled with the Martini force field agreed with those measured from solid-state NMR [228, 289]. Lateral diffusion rates measured by fluorescence correlation spectroscopy (on confocal laser-scanning microscopy) quantify the diffusion rate of transmembrane proteins in bilayers; this rate is modulated by membrane thickness and composition, with little effect from lipid headgroup [259].

The structure of the HIV-1 virion is another valuable candidate for coarse-grained study because it shows a relatively complex morphology that is generated from the components of a single polypeptide, about which much is known. In a recent multiscale simulation, Ayton and Voth used CGMD to reproduce structural features of the virion as observed by electron microscopy and cryotomography. Noting that only enhanced interactions the C-terminal capsid domain were sufficient to stabilize the hexameric lattice on the immature virion, they performed AAMD simulations to validate the CGMD model, showing that close-contact sites have a PMF well that is twice as deep in the wild type compared to mutants which show reduced viral infectivity in cells and particle defects under transmission EM [18]. Coarse-grained simulations and PMF calculations of lipid-mediated protein interactions have also been used to study the hydrophilic shielding of proteins within a bilayer [71, 72].

In an advanced application, CGMD has been applied to the study of protein-gated ion channels. Simulations of the plug domain in SecY shows that the introduction of a disulfide bond is sufficient to open the channel, explaining unrestricted translocation seen in experiments using a disulfide-immobilized plug domain [201]. Simulations of the mechano-sensitive protein channel MscL characterize the decrease in liposome stress as the channel activates [199]. Simulations of voltage-gated potassium (Kv) channels have characterized the closed structure while matching experimental constraints [336], including pore radius measurements, electrophysiology observations, and histidine scanning. Simulations have also investigated carbon nanotube-lipid interactions [355, 365], confinement of copolymers [128], and pore formation by dendrimers [179–181].

The class of G-protein coupled receptors (GPCRs) represents another relevant protein-membrane mechanism. Martini simulations of a particular GPCR, rhodopsin, show that it self-assembles via a hydrophobic mismatch mechanism by matching simulations to EPR and FRET experiments with bilayers of varying thickness [248]. Free energy profiles calculated from Martini simulations of the glycophorin A (GpA) show that mutations to this transmembrane alpha-helical protein disrupt its association in bilayers, but do not abolish it, suggesting that non-specific aggregates are possible [38,40,256,289]. These results agree with experiments which quantify the self-assembly of GpA using sedimentation equilibrium analytical centrifugation, FRET, and thiol disulfide interchange experiments [206]. Studies of high-density lipoprotein "nanodiscs" self-assembled on bilayers and matched temperature-dependent swelling of the particle observed in SAXS measurements and hydrophobic mismatch at the protein-lipid interface observed via solid state NMR [298, 299, 354].

In addition to the multiscale model for membrane bending by BAR domain proteins described in section 2.5.1, there has been considerable effort to connect atomistic, coarse-grained, and mesoscale models to study other protein-membrane systems. For example, a hybrid molecular mechanics/coarse-grained (MM/CG) model has been used to simultaneously improve the resolution of corase grained systems and extend atomistic ones to larger scales. In this approach, soft boundary potentials divide the atomistic and coarse-grained representations with an overlapping interface region. These simulations include stochastic and frictional forces due to the solvent along with cross-potentials designed to distribute coarse-grained forces across their constituent atoms in the interface region. The hybrid approach retains key microscopic details including hydrogen bond networks and the root mean squared fluctuations (RMSF) of the protein structure. It has found application in the study of enzyme active sites [234] and outer membrane proteases [233]. The hybrid approach has also made it possible to describe the ligand binding site for GPCRs in good agreement with atomistic simulation, suggesting a future role for this method in drug design [182]. Other hybrid approaches merge atomistic models with continuum methods in order to study protein-nucleic acid complexes [168] and membrane-peptide association [145]. Finally, hybrid models which use virtual sites in the interface region have been developed to bridge the Martini model with atomistic force fields [275]. The development of these hybrid modeling strategies makes it possible to customize coarse-grained simulations to include atomistic detail when necessary, broadening the range of possible applications to include specific protein-ligand binding. The hybrid approach further improves the flow of information between coarse-grained and atomistic representations by explicitly merging both models in the same simulation.

2.5.3 Extending to the Mesoscale

In much the same way that atomistic simulations inform coarse-grained models (and vice versa), coarse-grained models also make contact with mesoscale continuum mechanics models. Common to many of the membrane applications is the use of the

Helfrich Hamiltonian [131] given by equation 2.12, in which the membrane energy $\mathscr{H}_{\mathrm{el}}$ is modeled as an infinitesimally thin elastic sheet .

$$\mathscr{H}_{\mathrm{el}} = \int \left\{ \frac{\boldsymbol{\kappa}}{2} \left(H - H_0\right)^2 + \bar{\boldsymbol{\kappa}}_G G + \sigma \right\} dA \tag{2.12}$$

Equation 2.12 is integrated over the surface area A and consists of terms which account for surface tension σ as well as bending and Gaussian curvature, with energies given by their respective bending rigidities $\boldsymbol{\kappa}$ and $\bar{\boldsymbol{\kappa}}_{\mathrm{G}}$. In this formulation, the mean curvature $\mathrm{H} = \mathrm{c}_1 + \mathrm{c}_2$ and Gaussian curvature $\mathrm{K} = \mathrm{c}_1\mathrm{c}_2$ where c_1 and c_2 are the principal radii of curvature. Membrane remodeling enters the Helfrich in two places: via the spontaneous (or intrinsic) curvature term H_0, and also by modulating the bending rigidity of the underlying membrane.

Theoretical study of the Helfrich model has been used to explain vesicle configurations [288] and the effects of undulations on membrane elasticity [102, 132]. Protein-induced deformations have been added to these models, which can then make contact with CGMD simulations. For example, the theory predicts an elastic response to cylindrical protein inclusions which can be matched to ion channel experiments and further resolved with CGMD [46, 47, 327, 361]. The Helfrich model may also be coupled to mesoscopic solvent models in order to include the effects of hydrodynamics on membrane motion [15].

While modeling protein assembly on lipid bilayers is a prime candidate for coarse-grained simulation, continuum methods often augment CGMD simulations. For example, multiscale study of the BAR domain proteins uses coarse-grained simulations as a bridge to continuum methods to understand the time-scales required for membrane bending by N-BAR [9]. A host of coupling algorithms may be used to bridge the gap between atomistic simulation and continuum methods. High resolution simulations provide the chemical detail, while numerical methods make it possible to apply them to realistic models [14, 56].

Continuum mechanics simulations do not necessarily require coupling to atomistic or coarse-grained simulation, however. There is much to be learned from minimal mesoscale models which make contact with experiments. There are chemical and mechanical similarities between many membrane remodeling proteins [27]. Minimal mesoscale models can resolve the partitioning behavior of curvature-inducing proteins and the energetics of bud formation [2, 3, 196]. By modifying the direction, strength, and anisotropy of the spontaneous curvature induced by a particular membrane-remodeling protein, these models can predict the geometry and energetics of the resulting cellular morphologies. The wide range of continuum methods described in this section suggest that mesoscale simulation and coarse-grained simulations may be employed together or separately to characterize membrane-remodeling events.

2.6 Conclusions and Future Directions

This review has surveyed the ways in which coarse-grained molecular dynamics simulations provide a crucial bridge between the chemical detail found in atomistic simulations of membrane proteins, and the biologically relevant time- and length-scales accessible by continuum methods. Coarse-grained molecular dynamics simulations are ideally suited to simulating soft-matter systems relevant to biology because they are efficient enough to represent diverse cellular morphologies, but descriptive enough to distinguish the energetics and geometry of systems with different lipid compositions and, amazingly, differences in protein sequence and structure.

As is evidenced by the competing methods for designing and tuning coarse-grained force fields, there is no single coarse-grained method which can produce the same description as an atomistic one. The choice between structure, thermodynamic, and force-matching coarse-graining strategies depends strongly on the system of interest, computational resources, target experimental data, and most importantly, the question which the model must answer. Though many of these methods are able to reproduce some combination of structural and thermodynamic data, there is no guarantee that a naïve coarse-grained simulation will produce accurate results. Careful matching to theory, simulation, and experiments ensures that a particular model is physically accurate. More importantly, contact between these methods provides perspective on the physics of biological processes. In particular, we see that lipid bilayers mediate a host of cell processes, from the action of mechanically-sensitive ion channels, to morphology-generating membrane-remodeling, to the activation of complex cell-signaling networks by membrane-associated proteins. Future study of protein-membrane systems with coarse-grained methods will depend on synthesizing our understanding of soft matter systems with biology and biochemistry. This field of study has the potential to improve human health by resolving cell biological process at high resolution, and moreover, guiding the design of new treatment strategies.

Chapter 3

The protein Exo70 drives cell morphogenesis

The following chapter is adapted from "Exo70 Generates Membrane Curvature for Morphogenesis and Cell Migration" [380]

Dynamic changes in the shape of the plasma membrane are required for many processes essential to cell function, particularly cell migration and morphogenesis. In later chapters, we will explore endocytosis events in the interest of understanding how cells internalize cargo and regulate cell surface receptors. In this chapter, however, we will consider the topological inverse of this problem, namely: how do cells create protrusions?

3.1 The function of the exocyst

Cell shapes are determined by a variety of mechanisms, commonly facilitated by the self-assembly of proteins which sense, induce, and stabilize particular shapes [107,222, 381]. The prominent Bin/Amphiphysin/Rvs (BAR) family of proteins provide a well-characterized example. They grasp the membrane with concave, positively-charged surfaces and induce tubular extensions from synthetic vesicles *in vitro* and create invaginations towards ths the cytoplasm during e.g. endocytosis events [107, 222]. We typically consider curvature from the point of view of the proteins which are either decoated *outside* of large vesicles or occupy the cytoplasm of the cell and bind to the negatively-charged inner leaflet of the cell membrane.

The BAR domain proteins are complemented by the inverse BAR (I-BAR) proteins which bind membranes with a *convex* surface and have the opposite effect [216, 285, 379]. That is, they induce lumen-directed tubules in vesicles and surface protrusions in cells. We classify membrane bending by sign: "positive curvature" bends lipid bilayers *towards* the protein (e.g. BAR domains), while "negative curva-

ture" pushes *away* from the proteins (e.g. I-BAR) [222, 381]. The exocyst complex consists of proteins Sec3, Sec5, Sec6, Sec8, Sec10, Sec15, Exo70, and Exo84, mediates the tethering of secretory vesicles at the membrane in endocytosis and cell-surface expansion [129, 141, 231]. The exocyst plays a role in epithelia formation, cytokinesis, and neurite branching. Recent investigations show that the exocyst is also involved in cell migration [193, 197, 265, 271, 318, 328, 382]. A particular member of the exocyst, Exo70, is known to directly interact with the Arpc1 subunit of the Arp2/3 complex and kinetically stimulate actin polymerization and branching at the leading edges of migrating cells [193, 197, 382].

The protein, named for its molecular weight, has a long rod-like structure that binds to $PI(4,5)P_2$ through positively charged residues on its surface [80,125,194,229]. Overexpression of Exo70 — but not any other subunit of the exocyst complex — induces filopodia formation in cells independent of its function in exocytosis [382].

Parallel experiments and models

In this chapter I will outline the results of a collaborative effort to join biochemical and cell biological analyses to a multiscale model for membrane deformations by Exo70. This collaboration was initiated by Professor Wei Guo, who identified curvature induction by Exo70 and asked us to model it without indicating the direction of the induced curvature (positive versus negative). In the following sections we will describe the molecular, biochemical, and biological measurements used to characterize this system *in silico*, *in vitro*, and *in vivo*. The mapping between the molecular model and experiment is reserved for chapter 6, where we will discuss strategies for integrating results from multiple scales. We locate the mechanism of Exo70 curvature induction alongside the convex, negative curvature-inducers, and in so doing, detail a new case of protein-membrane remodeling in morphogenesis and directional cell migration.

3.2 Methods

Molecular dynamics simulations were previously used to study the BAR domain-induced positive curvature and resolve at the molecular level the mechanisms by which BAR domains sculpt the lipid bilayers [9, 17, 373]. Recently, coarse-grained molecular dynamics (CGMD) simulations were employed to study the molecular interactions of Epsin N-terminal homology domain oligomerization and association with membranes leading to the stabilization of tubular membrane geometries [173]. Here, we employed similar approaches to investigate the interaction between Exo70 and the membrane.

We test whether the minimal oligomerized state (namely a dimer) shows enhanced curvature induction in comparison to the monomer. If the enhancement is present at the dimer level, this effect is expected to be accentuated further in the case of higher-order oligomers. The crystal structure of mouse Exo70 (amino acids $85 - 653$)

features a 170×35Å rod composed of an α-helical bundles slightly curved in the middle [125, 229], and the N-terminal, noncrystallized region is predicted to be a coiled-coil structure. A coarse-grained Exo70 model was constructed according to the crystal structure in the protein data bank (pdb code `2PFT`) [229]. Structures for the unresolved loops and alpha-helical N-terminus were constructed with `MODELLER` [353]. The resulting structure was relaxed using a short, 10 ns all-atom simulation under the `CHARMM27` force field in the `GROMACS` molecular dynamics simulation package [37]. A representative snapshot was then coarse-grained for use in the MARTINI force field [228]. To maintain secondary structure, elastic bonds were applied to all backbone beads within $0.5 - 0.9$ nm with a force of $500\,\mathrm{kJ}\,/\,\mathrm{mol}\cdot\mathrm{nm}^2$, except for the N-terminal tail, for which we used a much weaker $5\,\mathrm{kJ}\,/\,\mathrm{mol}\cdot\mathrm{nm}^2$ force constant for residues $1 - 20$ and $50\,\mathrm{kJ}\,/\,\mathrm{mol}\cdot\mathrm{nm}^2$ for residues $21 - 90$.

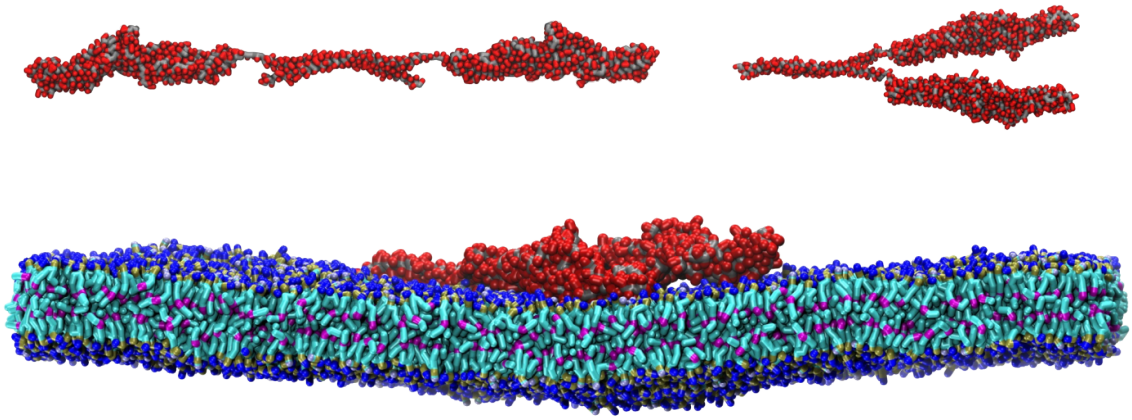

Figure 3.1: Coarse-grained models for the antiparallel (top left) and parallel (top right) Exo70 dimers. These models were each added to the bilayer in order to produce configurations that resemble the side-view of the parallel dimer pictured below. Atomistic simulations (not pictured) were used to tune an elastic network model (ENM) used to retain the protein secondary structure in each monomer.

Wild type Exo70 was modeled in a parallel or anti-parallel dimer conformation while the Exo70(K571A/E572A) mutant was modeled in an anti-parallel conformation and compared to the corresponding wild type control. Exo70($\Delta 1 - 75$) model was effectively a monomer as the dimerization domain is considerably shortened. Each modeled Exo70 variant was attached to an equilibrated bilayer containing 12,800 lipids with a 4:1 DOPC:DOPS ratio. The charge (-1) on DOPS partly serves as a surrogate for the higher charge (-4) on PIP_2 at a lower concentration. The systems were minimized and simulated with a standard MARTINI time step of 40 fs for 100 ns. Due to the smoothed energy landscape in the MARTINI force field, this corresponds to a real time of $3 - 4$ times longer. The Berendsen thermostat and barostat maintained a temperature of 310K and semi-isotropic pressure coupling of 1.0 bar both parallel and perpendicular to the bilayer plane with coupling time constants of

1.0 and 0.2 ps, respectively. Simulations were solvated with 600,000 water beads, corresponding to 4 water molecules each, and counter-ions were added at a physiological concentration of 150 mM. All other parameters were set per the MARTINI force field specifications [228]. All simulations were performed on supercomputing platforms at the Texas Advanced Computing Center (TACC). The coarse-grained MD (CGMD) simulations of the protein-membrane system were run on parallel architectures with $48 - 46$ processors.

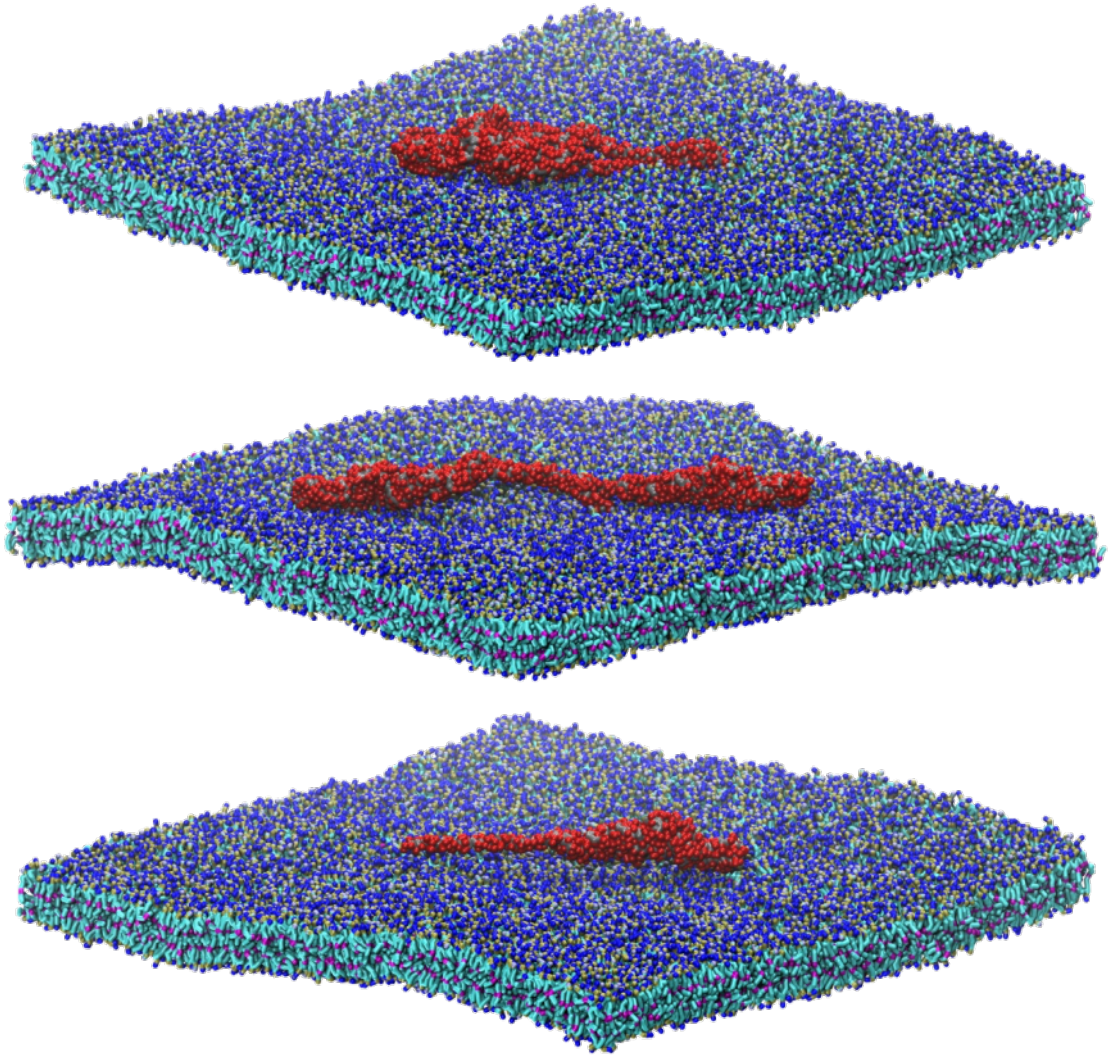

Figure 3.2: Orthographic views of three simulations: the parallel dimer (top), the antiparallel dimer (middle), and the monomer (bottom). Supplemental video `SV3.1` provides a video snapshot of the antiparallel dimer producing a negative curvature deformation.

3.2.1 Characterizing curvature: the surface-fitting algorithm

To characterize the induced curvature from these CGMD simulations, we have modeled the height of the mid-plane of the bilayer as a two-dimensional anisotropic Gaussian function centered and oriented with respect to the location of the protein (see below). This method provides a general description of the induced curvature by separately measuring the position, extent, degree of anisotropy, magnitude of curvature of the associated lipids and fluctuations due to dynamics/membrane undulations.

Since Exo70 dynamically remodels the associated bilayer, it was necessary to analyze the average as well as the fluctuations about the average in bilayer curvature to deconvolute the curvature effect from the undulation modes of the bilayer. Specifically, we must distinguish between negative curvature induction and the random fluctuations in the height of the bilayer. To this end, we first constructed a surface corresponding to the mid-plane of the bilayer and aligned the protein along its principal axis in each frame of the simulation. Then, for each frame, we divided the bilayer mid-plane surface into 5×5 nm patches and considered the patches that were within 10 nm of any atom of the protein. Analysis of a control membrane with no associated protein indicates that this filter effectively removes local undulations, which add noise to our estimate of curvature without changing the overall result. That is, the measured curvature of our control simulation is effectively infinite. The heights of these relevant mid-plane patches were then fit (frame-by-frame, see supplemental video `SV3.2`) to a two-dimensional Gaussian function given by:

$$z(x, y) = A_0 \exp\left\{\frac{(x - x_0)^2}{2\sigma_x^2}\right\} \exp\left\{\frac{(y - y_0)^2}{2\sigma_y^2}\right\} \tag{3.1}$$

Here, x and y are orthogonal axes chosen with respect to the alignment of the protein. The quantities σ_{x} and σ_{y} are the standard deviations in orthogonal directions, each corresponding to the "width" or extent of curvature of the fitted Gaussian parallel and perpendicular to the aligned axis of the protein. The maximum curvature is induced at the position where the height is higher than the average. This position is not necessarily at the center of the protein. The fits to equation 4.12 from the frames are depicted in supplemental video `SV3.2`, clearly showing negative curvature induced by the protein. Note that this fitting procedure is a conservative estimate of the negative curvature because it always predicts a curvature smaller than the actual curvature observed in the simulations. We calculate the strength of the induced curvature according to the following expression for mean curvature,

$$H = \frac{(1 + z_x^2)z_{yy} + (1 + z_y^2)z_{xx} - 2z_x z_y z_{xy}}{(1 + z_x^2 + z_y^2)^{\frac{3}{2}}} \tag{3.2}$$

Here z_{xy} terms correspond to the partial derivatives of the height of the fitted function. Thus, for each frame we have obtained a maximum mean curvature, i.e.

$H(x_0, y_0)$ from equation 4.12, and values and which characterize the parallel and perpendicular extent of curvature. In order to esimate the average deformation induced by the protein on the membrane, we compute the average deformation energy E (in units of $k_B T$) using the expression:

$$\langle E \rangle = \int_0^{\sigma_x} dx \int_0^{\sigma_y} dy \left[\langle H_{max} \rangle \exp\left\{ \frac{(x - x_0)^2}{2\sigma_x^2} \right\} \exp\left\{ \frac{(y - y_0)^2}{2\sigma_y^2} \right\} \right]^2 \tag{3.3}$$

Here $\langle E \rangle$ represents the ensemble average over all the frames of the CGMD simulation.

3.3 Results

The measurements described in 3.2.1 are summarized in table 3.1.

system	$\langle H_{max} \rangle (nm^{-1})$	$\langle \sigma_x \rangle$ (nm)	$\langle \sigma_y \rangle$ (nm)	radius (nm)	$\langle E \rangle$
anti-parallel dimer	-9.4×10^{-3}	11.2	17.4	53	0.0537
parallel dimer	-12.4×10^{-3}	8.6	16.1	40	0.0667
Exo70 ($\Delta 1 - 75$) monomer	-6.2×10^{-3}	9.9	14.6	81	0.0174
Exo70 (K571A/E572A) dimer	-5.9×10^{-3}	11.1	17.2	84	0.0208
control	-4.4×10^{-3}	18.6	38.9	112	N/A

Table 3.1: Maximum mean curvatures, extents, corresponding tubule radii, and deformation energies based on curvature measurements described in 3.2.1.

These values suggest that Exo70 dimers induce negative curvature, and that this curvature is stronger than those observed for the free bilayer, mutant anti-parallel dimer, and monomer simulations. We also find that he curvature induced by Exo70 dimers is *anisotropic* (that is, $\sigma_x < \sigma_y$). Analysis of the protein and lipid conformations in our simulations revealed that the induction of the curvature was mediated by the positively charged residues throughout the surface of the Exo70 dimers that interact with the negatively charged phospholipids, which suggests that curvature induction requires a scaffold of at least two linked Exo70 monomers that act in concert to remodel the associated lipid bilayer from its native planar state.

In chapter 6 we will describe the multiscale modeling procedure required to turn these deformation measurements into clear predictions for morphology changes. For the remainder of this chapter we will discuss the biological consequences of the result.

3.3.1 Recapitulating experiments

Experiments performed by our collaborators and reported in Zhao *et al* [380] jibes neatly with our curvature estimates. These curvatures agree with both *in vitro* and *in vivo* results from this work, briefly summarized below.

Exo70 induces inward-pointing membrane tubules. Our collaborators tested whether Exo70 produces protrusions similar to I-BAR [217, 366, 382] when incubated with large unilamellar vesicles (LUVs) containing 30% PIP_2. Transmission electron microscopy (EM) showed that LUVs incubated with wild-type Exo70 displayed tubular invaginations which projected toward the lumen, with dimensions (average diameter of $68.8 \pm 13.8\,\mathrm{nm}$) similar to those produced by an I-BAR domain. Interestingly an Exo70 mutant (K571A/E572A) which is defective in PIP_2 binding produced no such tubules. This result agrees with our findings that the same mutant produced *weaker* negative curvature than the wild-type.

Tubulation requires oligomers. Gel-filtration chromatography of recombinant wild-type Exo70 identified a range of potential oligomers while deleting the coiled-coil domain ($\Delta 1 - 75$) produced monomers. We mimicked this ologomerization-defective mutant by simulating a single monomer. This simulation also showed reduced negative curvature. Our dimer model represent the minimal oligomer that we could test using these simulations; in the event that Exo70 natively forms higher-order oligomers, we would expect the curvature to be even stronger.

Exo70 induces actin-free membrane protrusions in cells. These *in vitro* measurements are the primary point of contact with our models, however our colleagues also extended these to an *in vivo* system. Using time-lapse fluorescence microscopy, they found that overexpressing GFP-Exo70 in B16F1 cells generated filopodia protrustions, a substantial fraction of which ($\sim 20\%$) lacked F-actin. This result is topologically equivalent to the formation of inward-pointing tubules in LUVs (however in this case the protein is found in the cytoplasm, and the filopodia project *outward*).

Exo70 mediates leading edge protrusion and directional cell migration. Knockdown of endogenous Exo70 in human MDA-MB-231 cell lines hampered the cells' ability to form lamellipodial protrusions measured by Arp3 and F-actin staining. The two Exo70 mutants which failed to deform the membrane *in vitro* (K571A/E572A and $\Delta 1 - 75$) also failed to recover these protrusions, however the wild type reconstituted them. The mutants experienced slower migration, suggesting that the protrusion formation mediated by Exo70 plays a role in directional cell migration.

3.4 Biological consequences

To summarize these results, our collaborators have found that Exo70 generates negative curvature *in vivo*, and that this curvature depends on its ability to form higher-order oligomers and its competency to bind PIP_2. In cells, overexpression leads to increated filopodial protrusions, many of which lack actin at the tip. Similarly, cells with Exo70 inhibition or mutations which weaken its curvature induction produce fewer filapodia and fail to migrate directionally.

Given these findings in combination with the mechanisms suggested by the *in silico* portion of the study, we conjecture that "oligomerization-mediated scaffolding" drives membrane curvature induction. This claim is parsimonious with both our experimental findings and the wide body of literature that suggests that other proteins generate curvature by creating similar scaffolds [41, 107, 222, 226]. With its rod-like structure, propensity to bind PIP_2, and oligomerization state, the role of Exo70 resembles the well-characterized BAR domain proteins.

Despite its similar structure and function, however, Exo70 distinguishes itself from the BAR domain proteins in a number of ways. More specifically, it is implicated in cytoskeletal activity [79, 285]. It directly stimulates Arp2/3-mediated actin polymerization and branching [197,382], the latter of which is widely believed to push against the cell membrane in order to form protrusions. Recent experiments involving I-BAR and F-BAR suggest that membrane curvature facilitates protrusions [122, 216, 217, 366]. Hence we hypothesize that Exo70 may perform a similar function, creating space at the leading edge which accomodates actin polymerization and branching mediated by the Arp2/3 complex.

Given its canny ability to independently remodel membranes, along with its regulation by small GTPases and kinases [146, 265] and interactions with actin branching and polymerizing proteins, it is likely that Exo70 plays an important role in facilitating morphology change in a cellular context. For that reason, further study of the molecular mechanism by which it bends membranes, as well as a full accounting of its binding partners and effectors will be important to characterizing the morphogenesis pathways that it influences.

3.5 Refining the model

Our molecular model for Exo70-membrane interactions gives us the opportunity to ask more refined questions about its role in exocytosis and cell morphogenesis. Since this work was published, we have refined our model by including explicit interactions with a single PIP_2 molecule hypothesized to be necessary to adhere the protein to the bilayer. While the original simulations relied solely on electrostatic effects with mildly-charged lipids like DOPS (-1e), simulations depicted in figure 3.3 include explicit PIP_2 at so-called "infinite-dilution", that is, with a single PIP_2 per monomer.

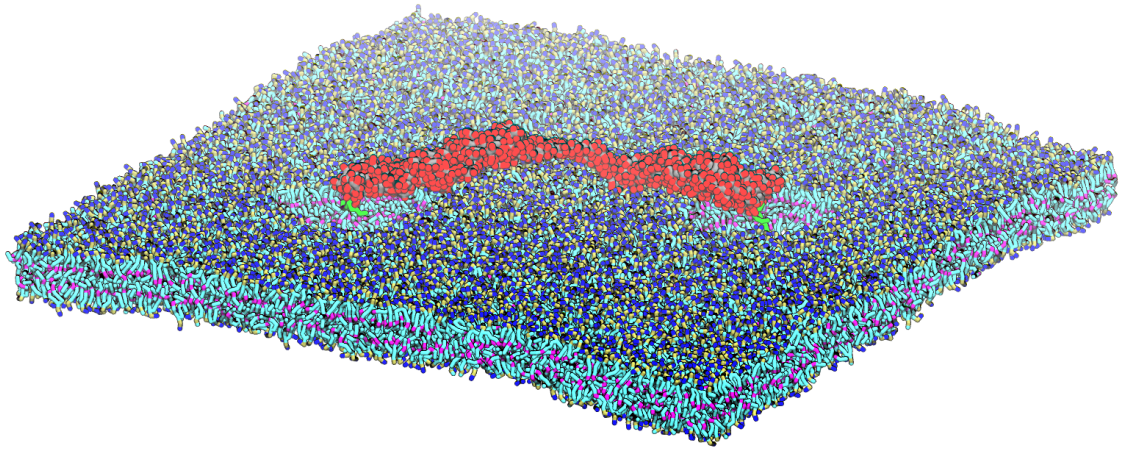

Figure 3.3: Snapshot of a simulation containing the Exo70 anti-parallel dimer interacting with two PIP_2 molecules. We have removed lipids within 6 nm to highlight the associated PIP_2, highlighted in green.

These simulations use a model for PIP_2 generated using instructions provided by [212] along with guidance from structure-matching between our hypothesized model and simulations of PIP_2 reported in chapter 5. Simulations of Exo70 in the absence of PIP_2 indicated that the protein would eventually detach. In these simulations, the protein remains bound to the bilayer.

Proof of concept. In this study, we used a surface-fitting algorithm to estimate the protein-induced deformation fields. This algorithm was also employed to estimate the curvatures induced by simulations of ENTH domains in a related study [333]. We have found that our original estimates agree with more precise values computed in chapter 4 (due to the curvature-undulation coupling method). This work demonstrates the usefulness of a "naïve" measurement of curvature, but more importantly, it also demonstrates that it's possible to link mutations in a single protein to more profound changes in cell morphology. These findings substantiate the goal of this work, namely, adding molecular detail to our understanding of protein-membrane morphology changes. Mutations provide straightforward examples of how we can add molecular detail to our calculations, however the protein identities, concentrations, and association geometries also provide the kinds of specificity necessary to predict *in vivo* experiments. We will explore the logic and purpose of this multiscale modeling strategy further in chapter 6.

Chapter 4

Curvature-undulation coupling as a basis for curvature sensing and generation in bilayer membranes

This chapter is adapted from a paper currently in press [44].

In the following study, we present coarse-grained molecular dynamics simulations of the epsin N-terminal homology (ENTH) domain interacting with a lipid bilayer and demonstrate a rigorous theoretical formalism and analysis method for computing the induced curvature field in varying concentrations of the protein in the dilute limit. Our theory is based on the description of the height-height undulation spectrum in the presence of a curvature field. We formulated an objective function to compare the acquired undulation spectrum from the simulations to that of the theory. We recover the curvature field parameters by minimizing the objective function even in the limit where the protein-induced membrane curvature is of the same order as the amplitude due to thermal undulations. The coupling between curvature and undulations leads to significant predictions: (1) under dilute conditions, the proteins can sense a site of spontaneous curvature at distances much larger than their size; (2) as the density of proteins increase the coupling focuses and stabilizes the curvature field to the site of the proteins; (3) the mapping of the protein localization and the induction of a stable curvature is a cooperative process, which can be described through a Hill function.

4.1 Introduction

Cellular membranes host many processes essential to life, all of which depend on the unique physical and chemical properties of the lipids and proteins that comprise, adhere to, and remodel them. Membrane shapes provide unique microenvironments across organelles, and distinctions between them confer high selectivity for trafficking

and signaling processes along with more potent, local concentrations of important signaling molecules [284]. Understanding how proteins sculpt lipid bilayers is vital to our understanding of how cell membranes modulate cell signaling pathways and consequent cell fate.

In most curvature-driven or curvature associated cellular processes, three length scales in the plane of the membrane, namely those corresponding to the lipids ($\sim 1\,\mathrm{nm}$), small proteins ($\sim 5 - 50\,\mathrm{nm}$), and cytoskeleton ($\sim 1\,\mu\mathrm{m}$), are all instantiated by a rich complex of partners. While *in vitro* experiments tend to isolate only a few components of larger trafficking machinery [278], and *in vivo* experiments can test the role of individual components [136], it is clear that common cellular trafficking events involve the concerted action of many proteins. A typical example is clathrin-mediated endocytosis (CME), which depends on a score of scaffolding proteins, adapters, and signaling enzymes [220]. In keeping with its diverse functionality, it has been argued that physiological curvature generation can occur through many mechanisms [221], each of which may carry unique energy costs [323].

The intrinsic shape of constituent lipids influences larger-scale bilayer shapes [61]. By themselves, lipids are thought to generate curvature *cooperatively*, when organized into domains with distinct boundaries [331]. In some cases protein mediated processes, for example, the action of flipases, is necessary to create lipid composition differences across the opposing leaflets of a bilayer which can generate differences in membrane curvature [76]. In other examples of cooperative lipid "domains" induced curvature, lipid composition can vary across organelles, between membrane structures like tubules and vesicles [345] and even within small nanoscale lipid rafts created by phase demixing in a single membrane region [215]. There is evidence that lipid shape may be coupled to the shape of the bilayer, but precisely quantifying this coupling through experiments or simulations is challenging [61, 139, 274]. Such experiments delineate mechanisms by which lipids sense particular preferred curvatures.

Several mechanisms of protein-mediated curvature induction on bilayers have been proposed. Curvature can be generated by enrichment of transmembrane proteins [246], steric pressure (i.e. "protein crowding") at the bilayer surface [324], and hydrophobic insertion by peripheral proteins [66]. Larger protein lattices can often be directly imaged via electron microscopy or mapped using scattering or correlation spectroscopy experiments. In some cases, tour-de-force simulations have been conducted to match with experiments to identify Bin/Amphiphysin/Rvs (BAR) domain lattices on membrane tubules [374], aggregation of light harvesting proteins on highly curved membranes [140], and bilayer thinning caused by α-synuclein [48].

The actin cytoskeleton influences endocytic pathways by interacting with a host of adapter proteins and curvature-inducers, namely amphiphysin and dynamin [147]. Actin dynamics are tightly linked to the recruitment of these proteins which are thought to generate curvatures necessary to initiate trafficking events [369]. Actin dynamics may also influence membrane curvature by a physical mechanism: by modulating membrane tension, which is conjugate to its excess area [292].

Several experiments have focused on highly concentrated, homogeneous systems which provide unambiguous evidence for membrane shape changes. Many proteins can bend bilayers if they are enriched to a high concentration. For instance, above 20% surface coverage, green fluorescent protein can bend bilayers, despite any evidence that it plays a role in physiological membrane remodeling [324]. While this evidence indicates that protein surface coverage is relevant to bilayer shapes, tightly-regulated membrane shapes may be generated more efficiently *in vivo* at protein surface coverages well below this threshold, by proteins with a more specific membrane-bending functionality [316].

More specifically, curvature-inducing proteins enriched to a moderate concentration may initiate membrane bending events necessary for intracellular transport or cell motility. For example, the protein FBP17 localizes to low-tension membrane invaginations at the leading edge of migrating cells in order to activate actin assembly [337]. *In vitro* experiments show that both membrane tension and bound protein density may work together to initiate the membrane shape changes that lead to trafficking events [297]. Each of these cases supports a tripartite relationship between tension, protein density, and membrane shape change, yet each of these phenomena operate at different length and energy scales on the bilayer. There is likewise evidence that a member of the exocyst (Exo70) drives the formation of outward-pointing membrane protrusions (negative curvature) and promotes actin branching at the sites of lamellapodia formation which may be associated with cell migration and tumor invasion [197, 380].

Regardless of the specific protein composition, both simulations and experiments yield important insights into protein-induced membrane bending. Simulations indicate that the ordering of both BAR and ENTH domains influences the morphology of the bilayer to which they are attached. For example, ordered helical lattices of these proteins generate tubules while disordered proteins generate vesicles (albeit at a higher concentration) [12, 226]. Simulations can accurately match the lattice found in cryoEM experiments in order to predict the optimal angle between protein constituents [9]. On the other hand, free energy calculations show that BAR domain self-associations are inhibited by tension [304].

These studies help to explain the features of *mature* curved morphologies or protein lattices, however it is necessary to characterize the original protein-protein associations and membrane interactions which generate *nascent* morphologies. In order to identify the smallest possible building blocks for curvature generation, one must test whether particular molecular components can bend the bilayer at moderate, physiological concentrations. Oftentimes, these concentrations give surface coverages which are well below the diffraction limit and resist clear imaging. An additional complication comes from thermal noise. The softness of lipid bilayers produces many long-wavelength undulations which may *appear* indistinguishable from an induced curvature-field at a given location at a given instant. Predictive models for membrane curvature require a careful accounting of the properties which influence bilayer tension, rigidity, and excess membrane area, because these influence the nature of the

undulations.

In this study we offer a first step towards a minimal, predictive model for membrane bending by studying the undulations in a planar bilayer with modest concentrations of a commonly studied curvature-inducing protein domain. We focus on a single exemplar of curvature generation at the intermediate scale: the epsin N-terminal homology domain (ENTH), which induces curvature via an inserted amphipathic helix (helix-0 or **H0**) [101]. Amphipathic helices are the ideal candidates for distinguishing curvature sensing and generation since it appears they can do both, they assemble more quickly than larger complexes due to their size, and they presume to remodel the bilayer with a mechanism that may be less sensitive to ordering parameters [82, 153].

Through coarse-grained molecular dynamics simulations, we quantify curvature fields induced by ENTH domains at three concentrations, and distinguish this curvature from background undulations by applying a fluctuation theory based on a continuum model to our simulations. We show that ENTH domains induce curvature cooperatively and *focus* background undulations into coherent curvature fields over distances which are much larger than the proteins themselves.

4.2 Methods

To quantify the dependence of bilayer curvature on protein density, we carried out simulations with 1, 4, or 8 coarse-grained ENTH domains according to the methods employed in previous simulations of ENTH [335] and Exo70 [380]. We adhere these proteins to a bare bilayer containing a total of 12,800 lipids at a 4:1 composition of DOPC and DOPS at a spacing of roughly 15 nm. Each protein is bound to a single $\mathrm{PI}(4,5)\mathrm{P}_2$ molecule parameterized from atomistic simulations which remains bound to **H0** near the known binding pocket [101]. Simulations lacking $\mathrm{PI}(4,5)\mathrm{P}_2$ failed to show measurable curvature and in some cases the ENTH domains detached from the bilayer (results not shown). This result is consistent with experiments that show that ENTH domains require binding [137] and motivates our use of a single $\mathrm{PI}(4,5)\mathrm{P}_2$ which is sufficient to retain ENTH at the bilayer in our model.

Bilayer simulations proceed for 500 ns in MARTINI time ($\sim 2\,\mu$s in real time) with a timestep of 40 fs sampled every 160 ps. The Berendsen thermostat and barostat maintained a temperature of 310 K and semi-isotropic pressure coupling of 1.0 bar both parallel and perpendicular to the bilayer plane with coupling time constants of 1.0 and 0.2 ps, respectively [34]. Simulations were solvated with $\sim 700,000$ water beads which provide more than 20 nm between periodic images in the normal direction. Simulations data are read by the MDAnalysis toolkit [224] and analyzed by in-house analysis codes which use `SciPy` functions for optimization [343]. A typical dynamics run required 48 hours of computing time on a 48−core (3 nodes, 16 processors each), therefore requiring 1,920 CPU hours. The aggregate time for the project required 8,000 CPU hours of computing. Figure 4.1 provides snapshots of our protein-membrane systems.

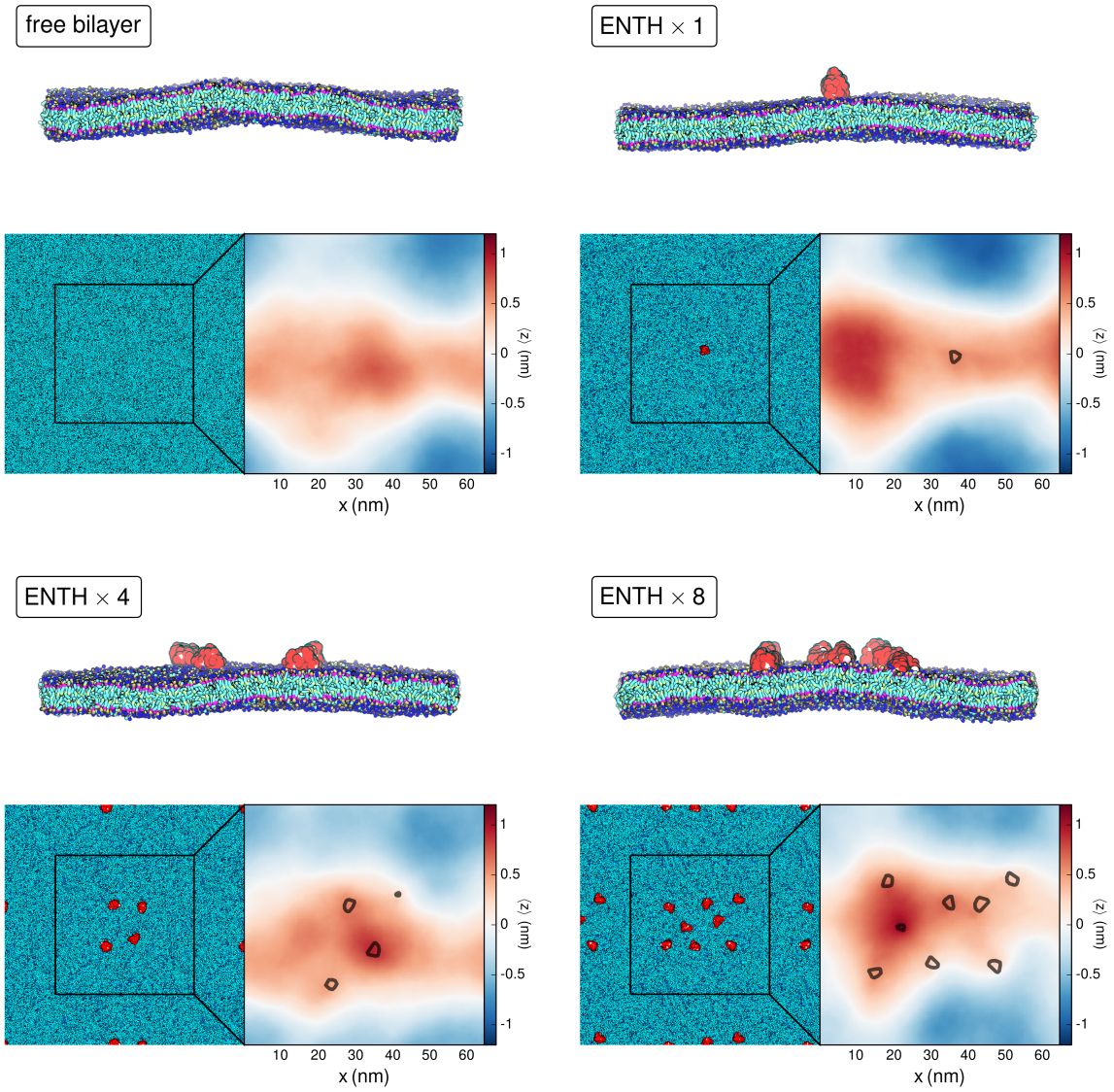

Figure 4.1: Simulation snapshots showing a side profile, top view under periodic boundary conditions, and average height profile (z) of a free bilayer along with either one, four, or eight ENTH domains (red). We omit water and counterions for clarity. The top view shows the size of the periodic simulation box (black square).

These simulations have a spatial extent of roughly $\sim 65 \times 65\,\mathrm{nm}$ which is large enough to accomodate a rich collection of proteins in a biological context, but which is still too small to form even the smallest sub-cellular membrane shapes, e.g. a small vesicle or complete tubule. However the length of our system is large enough for us to capture the effect of membrane undulation modes that are 10 times longer than the size of the proteins. In order to relate the coupling between protein-induced curvature and the membrane undulations, we employ the Canham-Helfrich [131] free energy functional, which has been used to describe the plasma membrane in a variety of biological contexts [261]. We have not reported the results of a system-size scaling analysis in this study, however, we describe this procedure in section 7.3. Through this functional, we describe the bilayer height fluctuations according to its physical properties — namely the bending modulus and surface tension — and the spontaneous curvature field of adhered proteins as well as the temperature. In order to directly apply the theoretical analysis on our molecular dynamics simulations, we interpolate each leaflet of the bilayer at a length scale of 0.5 nm resulting in height profiles of 128×128 for our square bilayers of $65 \times 65\,\mathrm{nm}^2$. The leaflet average gives the heights of the bilayer midplane; the Fourier transform of these heights provides the values of $\mathrm{h}_{0,\mathbf{q}}$ given in equation 4.8.

Inferring the coarse-grained time-scale

To compute the diffusion coefficients of our proteins, we must first establish the correct timescale for our simulations. Due to a smoothed energy landscape, coarse-grained simulations appear to be faster than their clock time. The true timescale of the simulation can be found by matching a common physical parameter to an experiment. To this end, we compute the diffusion rate of the most common lipid in our simulated bilayers, DOPC, and compare it to its experimental value. The diffusion rates in MARTINI time are pictured in figure 4.2.

Experiments which measure DOPC diffusion using NMR estimate a rate of $\sim 11.5\mu\mathrm{ms}^{-1}$ at 303K [99]. We observe a DOPC diffusion rate of $39.1\mu\mathrm{ms}^{-1}$ in our simulations. The ratio of these rates gives a time dilation of 3.61 so that the 500 ns simulation represents $1.8\,\mu$s in real time. This is similar to the fourfold conversion factor recommended by the MARTINI force field authors [212].

4.2.1 Curvature-undulation coupling

We treat the energetics of a biological membrane with the Canham-Helfrich Hamiltonian [54, 131].

$$\mathcal{H}_{\mathrm{el}} = \int \left(\frac{\boldsymbol{\kappa}}{2} \left(2H - H_0\right)^2 + \overline{\boldsymbol{\kappa}} K + \boldsymbol{\gamma} \right) dA. \tag{4.1}$$

We parameterize the membrane in the Monge gauge (that is, the small-slope limit where $|\nabla \mathrm{h}|^2 \ll 1$) where its position is given by positions $\mathbf{r} = \{\mathrm{x}, \mathrm{y}, \mathrm{h}(\mathrm{x}, \mathrm{y})\}$. This leads to the following linearized energy functional for a surface with zero spontaneous

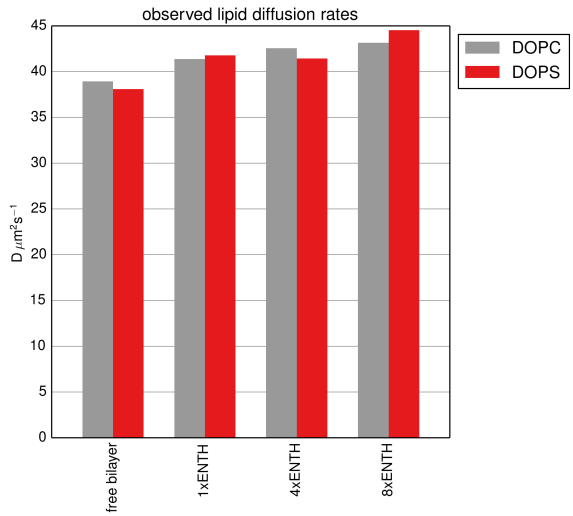

Figure 4.2: Observed lipid diffusion rates for each lipid in coarse-grained time. We compare the average diffusion coefficient for DOPC, the most abundant lipid, to experiments to confirm the expected fourfold speedup.

curvature:

$$\langle \mathcal{H}_{\mathrm{el}} \rangle = \int \frac{\boldsymbol{\kappa}(\mathbf{r})}{2} \left(\nabla^2 h(\mathbf{r}) \right)^2 \mathbf{dr} + \frac{\boldsymbol{\gamma}(\mathbf{r})}{2} \int \left(\nabla h(\mathbf{r}) \right)^2 \mathbf{dr}. \tag{4.2}$$

We define the Fourier transform of the bilayer midplane relative to the wavevector $\mathbf{q} = (\mathrm{q_x}, \mathrm{q_y}) = 2\pi(\mathrm{n_x}/\mathrm{L_x}, \mathrm{n_y}/\mathrm{L_y})$ where $\mathrm{A} = \mathrm{L_x} \times \mathrm{L_y}$ is the projected area and $\mathrm{n_x}, \mathrm{n_y} \in \mathbb{Z}$.

$$h(\mathbf{r}) = \sum_{\mathbf{q}} h_{\mathbf{q}} \exp(i\mathbf{q} \cdot \mathbf{r}). \tag{4.3}$$

Free bilayer

In the case of a free bilayer, the spontaneous curvature $\mathrm{C}_0(\mathbf{r})$ is zero everywhere. In the most general case, the bending rigidity may be inhomogeneous such that $\boldsymbol{\kappa}(\mathbf{q}) = \boldsymbol{\kappa}(\mathrm{q_x}, \mathrm{q_y})$. We will treat tension and bending rigidity as constants for the calculations in this chapter. Under the Fourier transform, this constant becomes a Dirac delta function such that $\boldsymbol{\kappa}(\mathbf{q} + \mathbf{q}') = \boldsymbol{\kappa}\delta_{\mathbf{q}+\mathbf{q}',0}$ which we will denote $\boldsymbol{\kappa}_{\mathbf{q}+\mathbf{q}'}$. We also assume a constant tension denoted $\boldsymbol{\gamma}_{\mathbf{q}+\mathbf{q}'}$. Given that the wavevectors are symmetric ($\mathbf{q} = -\mathbf{q}$), assuming constant bending rigidity and tension allows us to write the energy as a sum over the wavevector ($\mathbf{q}$). The Fourier transform of equation 4.2 gives the energy in the frequency space as follows:

$$
\begin{aligned}
\langle \mathcal{H}_{\mathrm{el}} \rangle &= \int dA \sum_{\mathbf{q},\mathbf{q}'} \langle h_{\mathbf{q}} h_{\mathbf{q}'} \rangle \exp(i(\mathbf{q}+\mathbf{q}')\cdot\mathbf{r}) \left\{ \frac{1}{2}\boldsymbol{\kappa}(\mathbf{q}+\mathbf{q}')\mathbf{q}^2\mathbf{q}'^2 + \frac{1}{2}\boldsymbol{\gamma}(\mathbf{q}+\mathbf{q}')\mathbf{q}\cdot\mathbf{q}' \right\}, \\
&= \frac{A}{2} \sum_{\mathbf{q},\mathbf{q}'} h_{\mathbf{q}} h_{\mathbf{q}'} \delta_{\mathbf{q}+\mathbf{q}',0} \left\{ \boldsymbol{\kappa}_{\mathbf{q}+\mathbf{q}'}\mathbf{q}^2\mathbf{q}'^2 + \boldsymbol{\gamma}_{\mathbf{q}+\mathbf{q}'}\mathbf{q}\cdot\mathbf{q}' \right\}, \\
&= \frac{A}{2} \sum_{\mathbf{q}} \langle |h_{\mathbf{q}}| \rangle^2 \left\{ \boldsymbol{\kappa}\mathbf{q}^4 + \boldsymbol{\gamma}\mathbf{q}^2 \right\}.
\end{aligned}
\tag{4.4}
$$

According to the equipartition theorem, each independent undulation mode posseses an energy of $\frac{1}{2}\mathrm{k_B T}$. Applying equipartition to our energy functional gives the well-known height-height undulation spectrum:

$$
\left\langle |h_{\mathbf{q}}|^2 \right\rangle = \frac{k_B T}{A\left[\boldsymbol{\kappa}\mathbf{q}^4 + \boldsymbol{\gamma}\mathbf{q}^2\right]}. \tag{4.5}
$$

Bilayers with spontaneous curvature

In case the spontaneous curvature is nonzero, we define:

$$
\begin{aligned}
C_0(\mathbf{r}) &= -\nabla^2 h_0(\mathbf{r}), \text{ and} \\
C_{0,\mathbf{q}} &= \sum_{\mathbf{q}} C_0(\mathbf{r}) \exp(-i\mathbf{q}\cdot\mathbf{r}).
\end{aligned}
\tag{4.6}
$$

We transform equation 4.2 as follows.

$$
\begin{aligned}
\langle \mathcal{H}_{\mathrm{el}} \rangle &= \frac{1}{2} \int dA \sum_{\mathbf{q},\mathbf{q}'} \left\{ \boldsymbol{\kappa}(\mathbf{r}) \Big(\nabla^2 h(\mathbf{r}) - C_0(\mathbf{r}) \Big)^2 + \boldsymbol{\gamma}(\mathbf{r}) (\nabla h(\mathbf{r}))^2 \right\}, \\
&= \frac{1}{2} \int dA \sum_{\mathbf{q},\mathbf{q}'} \left\{ \boldsymbol{\kappa}_{\mathbf{q}+\mathbf{q}'} \left[(\nabla^2 h_{\mathbf{q}})^2 - 2C_{0,\mathbf{q}} \nabla^2 h_{\mathbf{q}} + {C_{0,\mathbf{q}}}^2 \right] + \boldsymbol{\gamma}_{\mathbf{q}+\mathbf{q}'} (\nabla h_{\mathbf{q}})^2 \right\}, \\
&= \frac{A}{2} \sum_{\mathbf{q},\mathbf{q}'} \Big\{ \boldsymbol{\kappa}_{\mathbf{q}+\mathbf{q}'} \Big[\mathbf{q}^2\mathbf{q}'^2 \langle h_{\mathbf{q}} h_{\mathbf{q}'} \rangle + \mathbf{q}^2 \langle h_{\mathbf{q}} C_{0,\mathbf{q}'} \rangle \\
&\qquad + \mathbf{q}'^2 \langle h_{\mathbf{q}} C_{0,\mathbf{q}} \rangle + \langle C_{0,\mathbf{q}} C_{0,\mathbf{q}'} \rangle \Big] \\
&\qquad + \mathbf{q}\mathbf{q}' \langle h_{\mathbf{q}} h_{\mathbf{q}'} \rangle \boldsymbol{\gamma}_{\mathbf{q}+\mathbf{q}'} \Big\}.
\end{aligned}
\tag{4.7}
$$

Assuming homogeneous bending rigidity and tension ensures that

$\boldsymbol{\kappa}_{\mathbf{q}+\mathbf{q}'} = \boldsymbol{\kappa}\delta_{\mathbf{q}+\mathbf{q}',0}$. Choosing $\mathbf{q}' = -\mathbf{q}$ allows us to write the energy in terms of a single wavevector, as follows:

$$\langle \mathcal{H}_{\mathrm{el}} \rangle = \frac{A}{2} \sum_{\mathbf{q}} \left\{ \boldsymbol{\kappa} \left[\mathbf{q}^4 \langle h_{\mathbf{q}}^2 \rangle + 2\mathbf{q}^2 \langle h_{\mathbf{q}} C_{0,\mathbf{q}} \rangle + \langle C_{0,\mathbf{q}}^2 \rangle \right] + \boldsymbol{\gamma} \mathbf{q}^{\mathbf{2}} \langle h_{\mathbf{q}}^2 \rangle \right\} \tag{4.8}$$

Note that some treatments include a higher order bending term $\sim \boldsymbol{\kappa} {\mathrm{C}_0}^2 (\nabla \mathrm{h})^2$ in equation 4.1 which couples spontaneous curvature to the height fluctuations. Continuum simulations show that the eigenmodes deviate no more than 1.1 degrees from plane waves, therefore we neglect this term for the remainder of the analysis [4].

Equation 4.8 encodes the expected spectrum of height-height undulations of a fluid bilayer. It has traditionally been used to interpret the fluctuations of a bare (protein-free) bilayer at small and large length scales using continuum methods [261] as well as molecular models [45] in the case where $\mathrm{C}_0(\mathrm{x}, \mathrm{y}) = 0$ and the spectrum is given by equation 4.5; our simulations of the protein-free and tensionless bilayer adhere closely to this behavior (see figure 4.3). Encouraged by this agreement between the molecular simulations and the continuum model for the protein-free system, we hypothesized that curvatures computed from molecular simulations in a bilayer with curvature-inducing proteins will obey the relationship in equation 4.8, and that we can therefore estimate protein-induced deformation fields (C_0) from the bilayer height-height undulation spectrum observed in the simulations.

We have previously estimated curvature-fields induced in protein-bilayer simulations for ENTH domains [333] and the protein Exo70 [380] by directly analyzing the statistics of deformations in the simulation trajectories. These studies also revealed that the induced curvature is of the same scale as natural undulations making it difficult to convolve the two effects. It is for the same reason that, we believe, single molecule experiments on these proteins fail to reveal any direct evidence of their curvature sensing or induction [139].

Discrete undulatory energy levels

At moderate temperature, the system lacks the energy to explore all possible high-frequency undulation modes evenly as the high-frequency vibrations become "frozen-out". To account for this effect, and thereby quantify the accessible frequencies of vibration in our system, we define discrete energy levels $\mathrm{E_n} = \mathrm{n}\hbar\omega$. Our discretized energy can be written with the following Hamiltonian where $\beta = 1/\mathrm{k_B T}$:

$$\langle \mathcal{H}_{\mathrm{osc}} \rangle = \frac{\hbar\omega}{\exp(\beta\hbar\omega) - 1}. \tag{4.9}$$

We treat the accessible frequencies according to the dispersion relation $\omega = \mathrm{c}\mathbf{q}\mathrm{k_B T}$, which assumes that frequencies are proportional to the wavevector via a proportion-

ality constant (c) such that:

$$\langle \beta \mathcal{H}_{\text{osc}} \rangle = \frac{\beta \hbar c \mathbf{q}}{\exp(\beta \hbar c \mathbf{q}) - 1} \tag{4.10}$$

The formalism in equation 4.8 accounts for the ways in which the imposed protein curvature field is convolved with thermal noise to create a distinct undulation spectrum that satisfies our modified expression for thermal equipartition. These relations yeild an intuitive method for determining the "best" $\mathrm{C}_0(\mathrm{x}, \mathrm{y})$ in equation 4.8 that is also consistent with our corrected equipartition theorem. Such a field would ensure that the ratio between $\mathcal{H}_{\text{el}}$ and $\mathcal{H}_{\text{osc}}$ is as close to unity as possible. If we penalize deviations from unity on a logarithmic scale, across all wavevectors, this yields the following objective function:

$$L = \sum_{\mathbf{q} < \mathbf{q}_{\text{cut}}} \left(\log \langle \mathcal{H}_{\text{el}} \rangle - \log \langle \mathcal{H}_{\text{osc}} \rangle \right)^2. \tag{4.11}$$

4.2.2 Analysis

In order to quantify protein-induced bilayer curvatures, we extract two types of data from our coarse-grained simulations. First, lipid centers of mass from both monolayers are interpolated and then averaged to generate a bilayer midplane surface at regularly-spaced intervals in the `XY`-plane. Interpolations are computed using a linear interpolator provided by `LinearNDInterpolator` in `scipy.interpolate`. When modeling the curvature field using a single "dimple", its position is irrelevant under the Fourier transform as long as the field remains smooth at the periodic boundaries. However, we also wish to test trial functions under the hypothesis that *each protein* may induce an individual (presumably smaller) dimple. In this case, the relative positions of these dimples may influence the quality of our fit, so we also save the protein centers of mass. Under the *dynamic protein-fields* hypothesis, we center a dimple on each protein. We find almost no difference between using the instantaneous versus average protein positions — this is most likely because the proteins do not diffuse very far — however the protein-fields hypothesis tends to perform somewhat better than the single field hypotheses at higher concentrations. Both lipid and protein positions are read directly from `GROMACS` trajectories via the `MDAnalysis Python` library [224] while all subsequent analysis is performed using `SciPy` [343] and in-house `Python` codes described in appendix A.

Standard undulation spectra

The analysis presented in the above utilizes the spectra according to equation 4.8 and equation 4.9, which are all provided as a dimensionless ratio between the oscillator and elastic energies. This formalism mangifies the low-amplitude, high-energy wavevectors

in order to compare their energy distributions with equipartition, however the strict undulation spectra depicted in figure 4.3 may be more familiar to the reader. The $\mathbf{q}^4$-scaling of the height-height autocorrelation is a hallmark of undulating lipid bilayers. Figure 4.3 shows the scaling according to equation 4.5. In section 5.3 we will find that small differences in height autocorrelation for intermediate wavevectors will be essential for matching our bilayer fluctuations to equipartition.

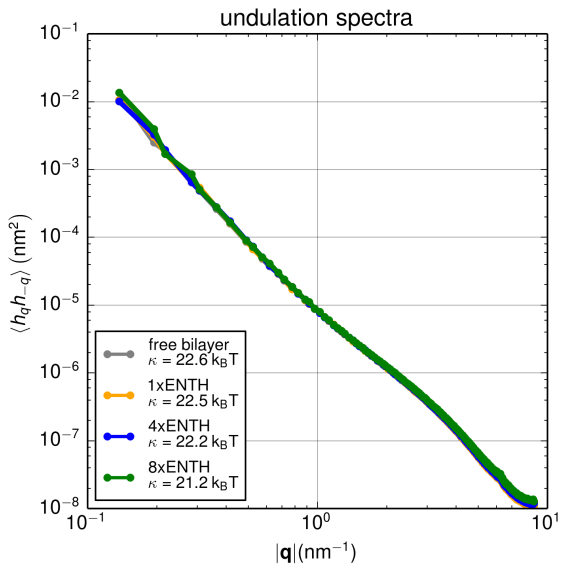

Figure 4.3: Height-height undulation spectra for each simulation according to equation 4.5. Even without a spontaneous curvature field, the undulations follow the standardard $\mathbf{q}^{-4}$ scaling with apparent bending rigidities noted in the legend.

Trial functions

We treat the field as a two-dimensional Gaussian function with the follwing form:

$$C_0(x, y) = C_{0,max} \exp\left(\frac{-(x - x_i)^2}{2{\sigma_x}^2}\right) \exp\left(\frac{-(y - y_i)^2}{2{\sigma_y}^2}\right). \tag{4.12}$$

The midplane surface and protein positions, now discretized to a resolution of 0.5 nm in the `XY`-plane are then Fourier-transformed using `numpy.fft.fft`. We generate trial functions for our hypothetical curvature fields on the same grid as the midplane surface. Our simulations are $\sim 65 \times 65\,\mathrm{nm}^2$ giving a grid with $\sim 130 \times 130$ components. Our trial functions sample the following ranges:

$$
\begin{aligned}
C_{0,max} &\in \{0.0, 0.001, 0.002, 0.005, 0.01, \\
&\quad 0.014, 0.018, 0.02, 0.024, 0.028, 0.032\}, \\
\sigma &\in \{1, 2, 4, 6, 8, 10, 12, 18, 24\} \text{ (single field)}, \\
\sigma &\in \{1, 2, 3, 4, 5, 6, 7, 8, 9, 10\} \text{ (protein fields)}, \\
\sigma &= \sigma_x = \sigma_y \text{ (isotropic)}.
\end{aligned} \tag{4.13}
$$

These parameters define $\mathrm{C_0(x, y)}$ in equation 4.12 where $(\mathrm{x_i}, \mathrm{y_i})$ denotes the instantaneous center of mass for each protein $\mathrm{i} \in \{1, \ldots, \mathrm{N}\}$ for the multiple dynamic protein-fields hypothesis, otherwise it is set to the geometric center of the simulation box in the single-field scenario.

Frequency filter

We select wavevectors above a cutoff of $1\,\mathrm{nm}^{-1}$ for two reasons. First, the data are interpolated to a resolution of 0.5 nm and we expect that high-frequency wavevectors near this resolution will be somewhat unreliable. More importantly, the interlipid spacing on the bilayer lies between $0.5 - 1\,\mathrm{nm}$ and we do not expect the functional form of the bending energy provided in the continuum theory to be valid at this scale. It is possible to analyze the fluctuations at arbitrarily high wavevectors, however there is little reason to expect that such high-frequency vibrations are coupled to the strengths of curvatures that we measure in this study [45]. Therefore, in all subsequent steps, we consider only wavevectors above $1\,\mathrm{nm}^{-1}$. We expect that higher wavevectors are not relevant to the curvature-coupling problem, most notably because the bilayer midplane is not precisely defined at length scales smaller than the inter-lipid spacing.

Optimizations

The trial functions, height values, and areas each represent variables present in equation 4.8. The remaining variables, namely $\boldsymbol{\kappa}$, $\boldsymbol{\gamma}$ and c, are free parameters, with respect to which we minimize the objective function $\mathrm{L} =$ defined in equation 4.11. Minimization is performed using `scipy.optimize.fmin`. We judge the accuracy of the hypothetical curvature field (our trial functions) by the mean squared residual (which we call mean-squared error or MSE). These mean squared errors are visualized in figure 4.7 for both the multiple dynamic protein-fields hypothesis and in figure 4.8 for the single-field hypothesis. We call such landscapes "fitness landscapes".

All analyses were completed using `Python 2.7`. A copy of the source code is available upon request from the authors.

Implementation

To implement this curvature-undulation coupling method, we first extract the mid-plane heights and then propose many different curvature fields as trial functions. There are three parameters needed to describe our fields: the number of proteins per field, the maximum curvature (hereafter curvature *strength*) ($C_{0,\max}\,\mathrm{nm}^{-1}$), and the spatial extent of the isotropic curvature field(s) ($\sigma\,\mathrm{nm}^2$). We employ two protein-field mappings. The first is a many-to-one mapping, one field for all of the proteins in the simulation, called the *single-field* hypothesis. The second is a one-to-one mapping in which each protein is represented by a single-field which tracks the instantaneous position of the protein's center of mass. The result is a collection of individual Gaussian functions given by equation 4.12 centered at the locations of each protein, which allows for anisotropy as well as higher order poles in the overall shape of the field. We call this the *multiple dynamic protein-fields* hypothesis. In both the single-field and multiple dynamic protein-fields scenarios we minimize L with respect to $C_{0,\max}$ and σ in equation 4.12. We rank these fields based on the value of the residual associated with minimizing the objective function written in equation 4.11. Associated codes are available in modules described in the appendix A.

4.3 Results

Before applying the formal method described above, it is useful to inspect the height profiles of protein-laden bilayers because they provide a qualitative picture of how the proteins can influence the bilayer shape. The ensemble-averaged snapshots of bilayer heights depicted in figure 4.1 show an interesting trend: membrane-attached ENTH domains almost always occupy above-average bilayer heights. Bilayer undulations have no preferred lateral position; their constituent waves have uniformly-distributed phase angles. That the proteins occupy the peaks of the underlying deformation is no coincidence. As the number of proteins (concentration) increases (from zero to eight ENTH domains), these proteins appear to *focus* the undulations underneath them. This process creates an average deflection of $\sim$ 1nm in our trajectories, but the average deformation obscures a significantly stronger dynamic focusing effect which is most apparent when viewing the dynamics, see supplementary video `SV4.1`. This video shows stronger deflections, centered on the proteins which wax and wane throughout the trajectory. More importantly, these deformations are mobile, orbiting the proteins over time. Based on the observed dynamics in the trajectories, we hypothesize that increasing protein concentration increasingly focuses these background thermal undulations to the vicinity of the proteins.

We depict the the bilayer height profiles for each simulation in supplemental video `SV4.1` with a snapshot given in figure 4.4. These movies show a single $65 \times 65\,\mathrm{nm}^2$ simulation box outlined in black, flanked by periodic images, which help to emphasize the dynamics of the long-mode undulations which tend to orbit or focus on the proteins (labeled in black). Red areas have above-average heights while blue areas

are below the average. Compared to figure 4.1, the scale bar indicates that the spontaneous deflection is larger than the average deflection.

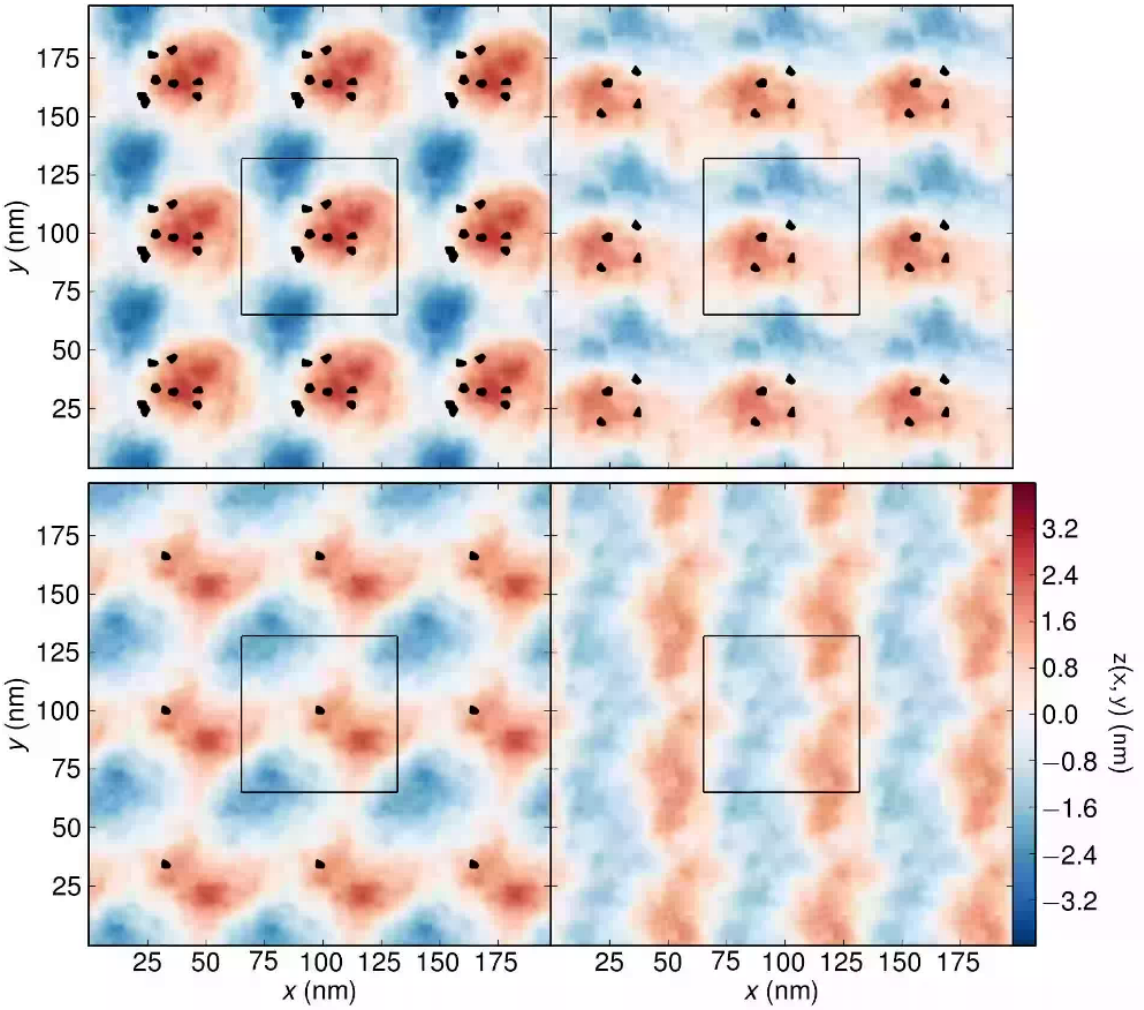

Figure 4.4: Snapshot of supplemental video `SV4.1` which shows the bilayer fluctuations for our simulations. This snapshot depicts the bilayer deflection either towards the protein (up, red) or away from it (down, blue). Average protein positions are indicated with black marks. We include images of the simulation under periodic boundary conditions (PBCs); the center box represents the size of the simulation.

The results of the curvature-undulation coupling analysis in equation 4.8 is presented in figure 4.5, which depicts the energy distribution among undulation modes indexed by the wavevector $\mathbf{q}$. The data presented here corresponds to the case of the optimal field which minimizes the objective function L in equation 4.11. The energy distribution for different modes ($\mathbf{q}$) show good agreement with equipartition, which posits that these modes should have uniformly distributed energies of $k_B T$. The bold lines depict the average over multiple wavevectors with the same magnitude, but different directions on the tangent plane. We find that our predicted deformation fields

are in good agreement with equipartition. We also see from figure 4.5 that proteins soften the bilayer and may produce a slight increase in the apparent tension $\boldsymbol{\gamma}$. Insofar as proteins suppress undulations, this result is consistent with the observation that inhomogeneities in the elastic modulus may lead to a softening of the bilayer [235]. Having found agreement between the best-fit fields and equipartition, we now turn to analyzing the features of these deformations.

The strength, extent and error are shown in figure 4.6. These parameters show several striking conclusions. The results for the single-field hypotheses imply that this scenario, namely describing the deformation field induced by the protein as a single anisotropic Gaussian field, is not a good representation across all the systems. While this model describes the one and four ENTH systems to the same degree of MSE as the other systems, the MSEs for the free bilayer and that for the eight ENTH systems are almost twice the other systems. This conclusion is not surprising, for in visualizing the deformation modes, it is clear that the deformations are both dynamic and possess modes higher than just a monopolar deformation. From a quantitative standpoint, the single-field model is therefore useful only as an internal control in order to assess and compare the MSEs of optimal fields to those that are less than optimal.

Qualitatively however, the single-field model suggests that the influence of the induced curvature field increases with increasing proteins, which is consistent with the visualization of the deformation modes in video `SV4.1`.

The multiple dynamic protein-fields hypothesis provides a more mechanistic picture that is also quantitatively accurate (and robust). First, it is evident that as the protein concentration increases, the curvature strength does so as well. The relationship between the number of proteins and the curvature field strength ($C_{0,\max}$) indicates positive cooperativity in curvature induction, suggesting a collective phenomena in curvature induction, (see red bars *BDFH* in figure 4.6). The maximum curvature approaches a plateau when four or more ENTH domains are present. With the approach of the plateau in the curvature field strength, the spatial extents of the individual fields also stabilize, while the overall spaial extent of the combined field (due to all the proteins) increases. This scenario is evident from the spatial maps of the curvature fields in the panels below. In order to establish the degree of cooperativity, we analyze $C_{0,\max}$ and σ versus the number of proteins in terms of a cooperative Hill reaction $nA_1 \rightleftharpoons A_n$ with an equilibirum association constant K_n [77]. By coupling the stable induction of curvature exclusively with A_n, we can determine the degree of cooperativity n by fitting the deformation strength (D_s) as a function of the number of protein domains (or the dynamic protein-fields), n_p: that is, we compute $D_s(n_p) = \sum_{i=1}^{n_p} \int {C_0}^2 dA \approx n_p \pi (C_{0,\max})^2 \sigma^2/2$, where the index i runs over n_p and $dA = dxdy$; the approximate equality evaluates the integral assuming each dynamic protein-field i is an independent Gaussian function. The plot of D_s versus n_p in figure 4.6, see symbols, is fit to the Hill cooperativity expression (solid-red-line) [77]:

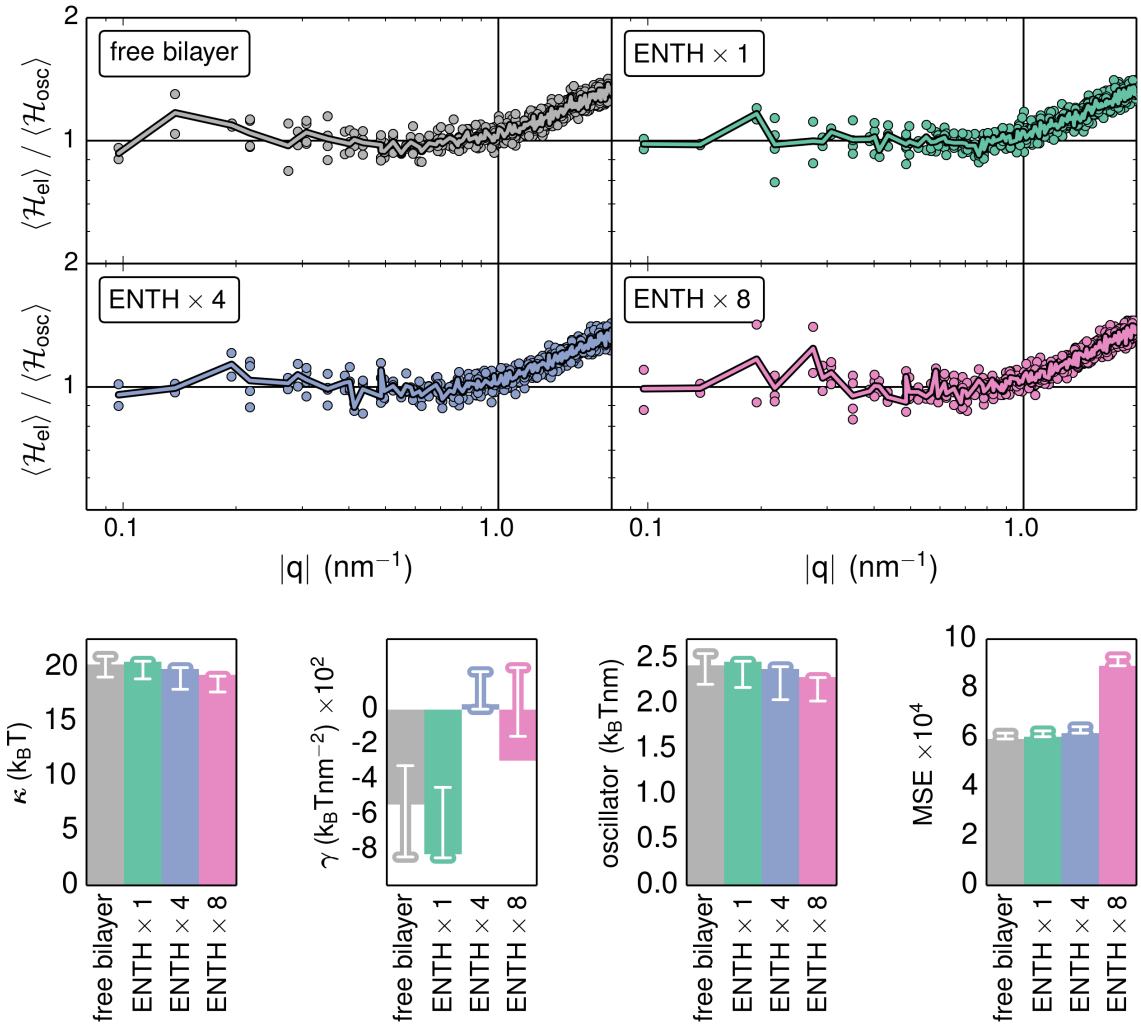

Figure 4.5: Energy distribution according to the wavevector ($\mathbf{q}$) for each simulation (top). Relevant modes lie to the left of $1\,\mathrm{nm}^{-1}$ and show good agreement with equipartition, which posits that these modes should have uniformly distributed energies of k_BT. Errors are somewhat higher for systems containing more proteins. Also depicted in the bottom row are bilayer properties including $\boldsymbol{\kappa}$ and $\boldsymbol{\gamma}$ found in equation 4.8, the harmonic oscillator constant (c) found in equation 4.10, and the resulting mean-squared error (MSE) taken from the residual difference between the spectra and k_BT; the renormalized bending moduli and tension fall within experimental ranges (see section 4.3.2). Error bars for $\boldsymbol{\kappa}$, $\boldsymbol{\gamma}$, and c (as well as the measurements shown in figure 4.6 below) are derived from fitted values observed for hypotheses that lie within four percent of the minimum observed MSE (this four percent range sets the error bars on the MSE).

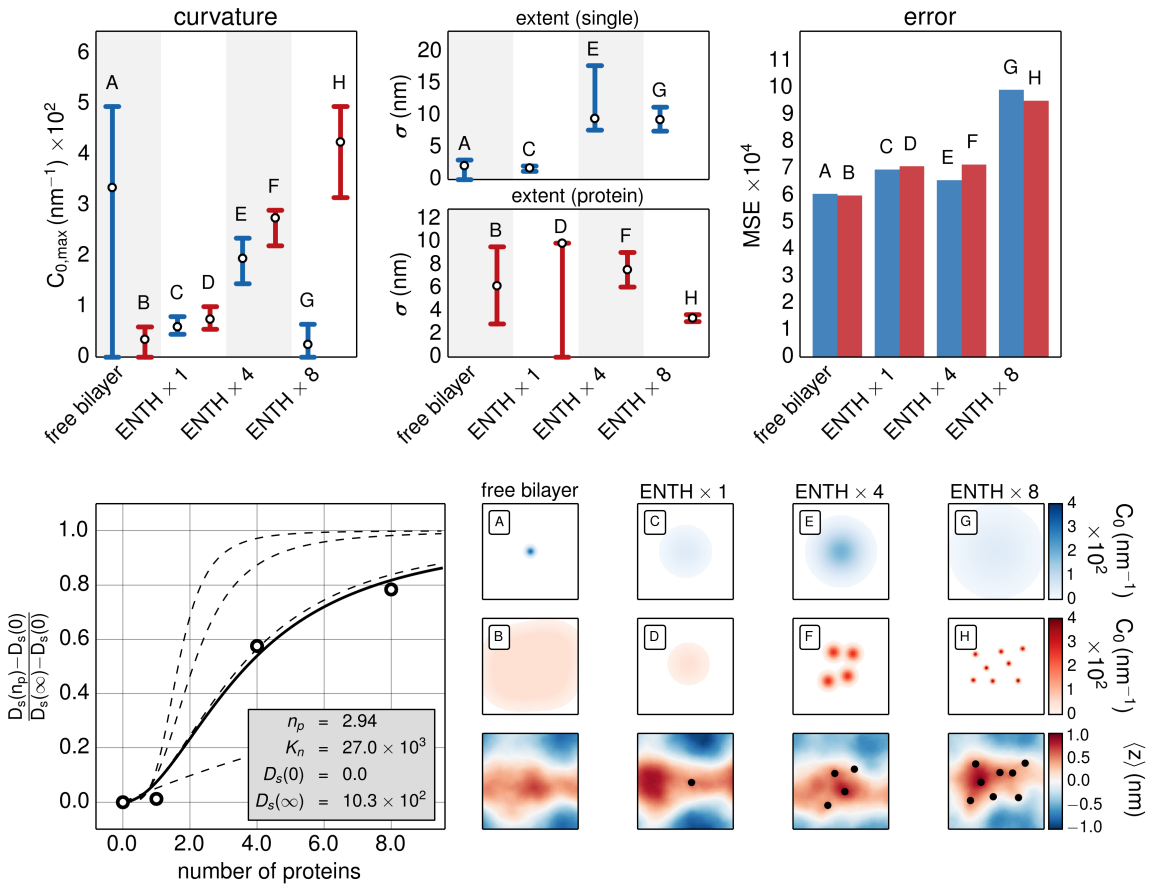

Figure 4.6: Curvature fields for bilayers with ENTH domains inferred from minimization of the objective function in the curvature-undulation coupling analysis. The parameters that minimize the objective function represent the closest fit to the adherence of equipartition. Curvature strength and extent for both single-field and multiple dynamic protein-fields mappings are depicted alongside their errors (top). In the plot titled "bending" (see bottom row), we show how our results for the computed curvature field parameters as a function the number of proteins can be used to analyze the degree of cooperativity between proteins in inducing a stable curvature field, see symbols. The Hill analysis (described in the text) provides a Hill coefficient of ~ 3 and the fit to the Hill equation is depicted as the solid black line. The dashed lines are fits to the Hill equation for $n \in \{2, 3, 4, 5\}$ and are provided to show that $n = 3$ is a distinctly close fit to the computed results. We also show spatial maps of these fields (middle) and the resulting average height profiles (bottom). Error bars are computed by taking the variation in either curvature or extent which corresponds to a four percent increase in the MSE.

$$\frac{D_s(n_p) - D_s(0)}{D_s(\infty) - D_s(0)} = \frac{nK_n[A_1]^n}{(1 + nK_n[A_1]^n)},$$

to yield the value of the "best fit" parameters: n $\sim$ 2.94, which clearly establishes the degree of cooperativity (namely, n = 3) among the ENTH domains in inducing a stable curvature; here, $[A_1]$ is the concentration of A_1, which is simply taken as n_p.

The results that we have enumerated so far — bilayer properties and estimated best-fit deformation fields — are sufficient to identify threshold concentrations, estimate the saturation level of curvature, and demonstrate that curvature induction depends on concentration resembling that of a cooperatibe Hill process. These estimates reduce the size of our survey to a single, consensus hypothesis. Analyzing the MSE (or fitness) landscape of less-precise hypotheses allows us to describe how robust our atom-to-field mapping procedure for the determination of the curvature fields can be. The robustness of our estimates is best assessed by the error landscapes over the two parameters of our trial functions. The landscapes for the multiple dynamic protein-fields trial functions are shown in figure 4.7, which plots the fitness landscape in terms of the MSE for all of the trial fields.

Figure 4.7 demonstrates that with the parameter sweeps we have carried out for the trial functions for the curvature field, many curvature fields may be nearly equally compatible with equipartition. This helps to put our consensus fields in the proper context and also quantifies the robustness/sensitivity of our analysis in identifying the curvature-field strengths from molecular dynamics simulations. Each landscape shows an inverse relationship between curvature and extent, suggesting that higher curvature looks like a lower extent through the lens of the undulation spectra. The finite length scale of our simulations means that we cannot probe wavelengths above 92 nm. This sets the floor on measurable curvature at $\sim 0.005\,\mathrm{nm}^{-1}$, which is the maximum curvature of a sine wave with the maximum wavelength. This floor corresponds to the consensus curvature observed in the error landscapes for free bilayers. We find that this minimum-observable curvature has a large extent on the free bilayer, suggesting that this is the background spontaneous curvature for our system.

4.3.1 Mappings between molecular- and meso-scales

The energy spectra in figure 4.5 along with the increase in errors with higher concentrations suggest that the single-field model performs performs slightly poorer in comparison to the multiple dynamic protein-fields model. We believe that these errors can be reduced by considering a wider variety of trial functions with dipole, quadrupole contributions etc., and by testing spatially inhomogeneous $\boldsymbol{\kappa}$ fields. The trial functions provide an important link between these molecular simulations and the mesoscale data and make it possible to predict e.g. the formation of tubules in an *in vitro* experiment [333]. While both the precision and specificity of our molecular-scale curvature estimates depend strongly on the trial functions, continuum methods use

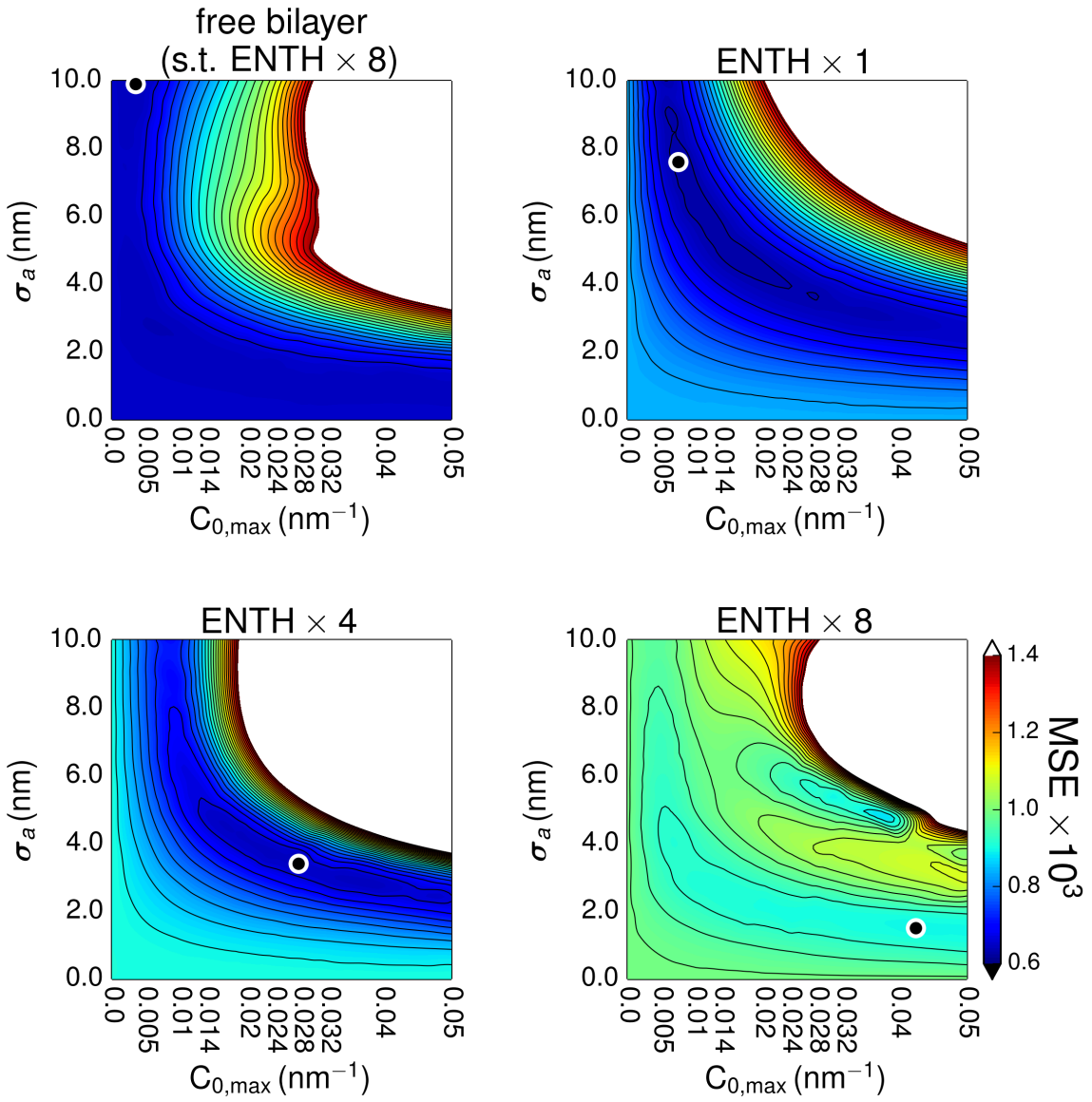

Figure 4.7: Error (fitness) landscapes show the mean-squared errors (MSE) for each trial function with a one-to-one protein-field mapping (in the multiple dynamic protein-fields hypothesis). Each point on the landscape corresponds to a unique spectrum. The black dots depict the best-fitting fields, and correspond to the spectra given in figure 4.5. These landscapes are smooth, somewhat flat, and monotone, which suggests that the strength and extent of curvature are partly fungible. Error bars in figure 4.6 are computed by taking the variation in either curvature or extent which corresponds to a two percent increase in the MSE.

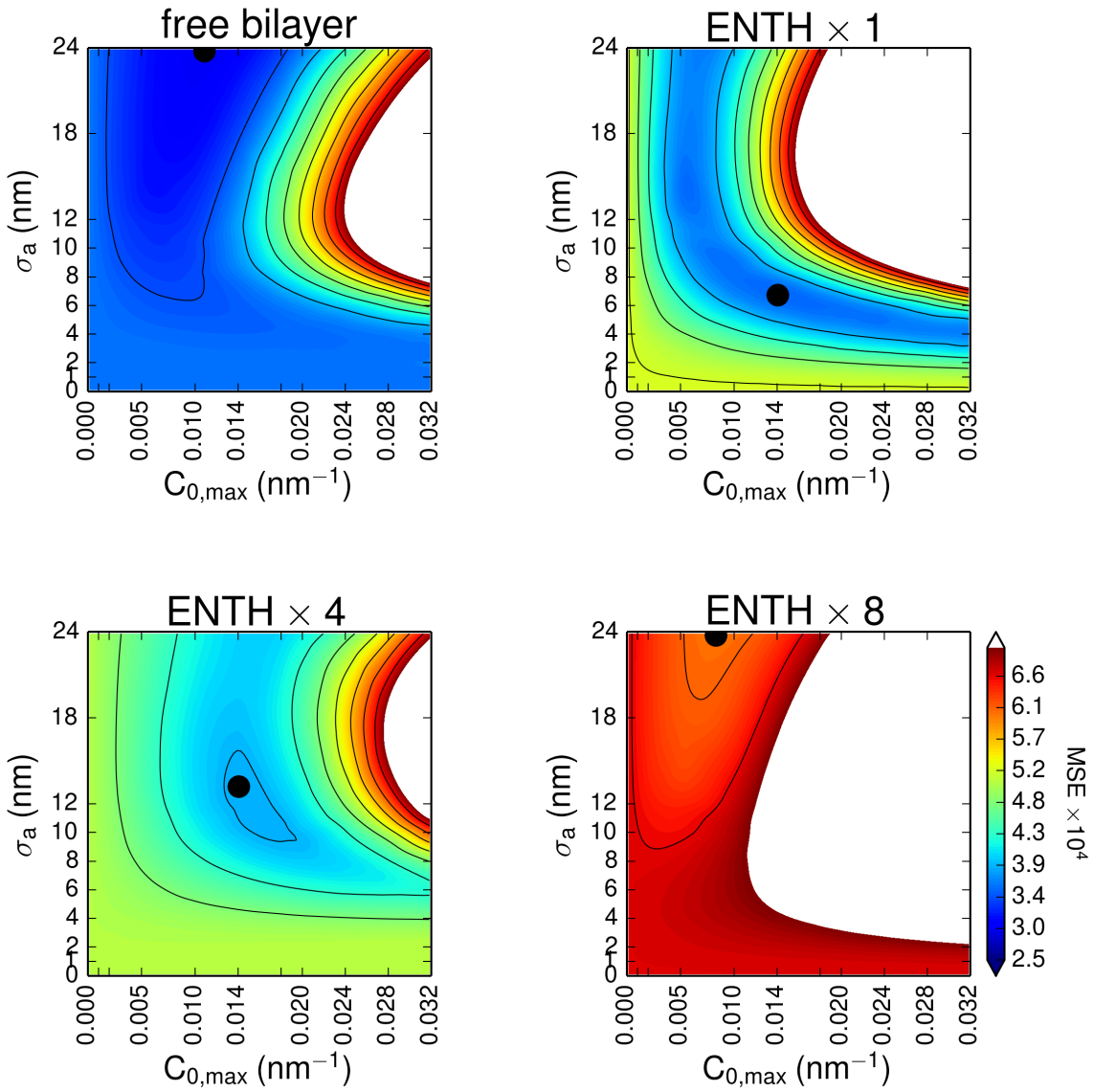

Figure 4.8: Fitness landscapes for the single-field hypotheses. This map describes the agreement between observation and theory for a single trial curvature field per simulation (hence $0 - 8$ proteins per field).

larger length scales and typically model several proteins with the single-field hypothesis. Since it lacks molecular detail, we must make sure that information gleaned from the multiple dynamic protein fields hypothesis is applied to continuum scales. We thereby ensure that single fields used in mesoscale models have the appropriate curvature, strength, extent, and overall bending energy.

4.3.2 Renormalized tension

We consistently observe a bending rigidity of $\boldsymbol{\kappa} \simeq 20\,\mathrm{k_B T}$, in good agreement with experiments described in chapter 2. Despite using a semi-isotropic pressure coupling to create tensionless conditions on average, we nevertheless fit a negative tension as low as $\boldsymbol{\gamma} = -8\,\mathrm{k_B T nm^{-2}}$ in the fluctuation analysis. A small atomistic simulation typically experiences an average tension as low as $10\,\mu\mathrm{Nm^{-1}}$, however due to its size, it will experience fluctuations which are much larger, typically about $100\,\mathrm{mNm^{-1}}$. Given that the large standard deviation, our estimates for tension are less reliable than those for bending rigidity.

4.4 Discussion

Our estimates of protein-induced deformation fields have demonstrated three important components of the molecular-to-mesoscale mapping. The mapping identifies the relationship between concentration and curvature and the underlying degree of cooperativity involved in stable curvature-field induction. While we have identified cooperative curvature induction by ENTH domains, it is also possible that other protein compositions create different transitions from weak to strong curvature, by primarily interacting with membrane undulations. The level of maximum curvature strength provides an upper limit on the strength of curvature that the protein domains can create intrinsically due to their interactions with the bilayer. This ceiling may be used to conclude that a particular protein cannot create more highly-curved morphologies without help from extrinsic factors such as stabilizing scaffolds. This may also help to distinguish the curvature induction at very high densities from protein coats that initiate bending and trafficking events *in vivo*. In particular, we find that ENTH domains have a relatively high ceiling of intrinsic curvature of $\sim 0.03\,\mathrm{nm^{-1}}$ which suggests that they are capable of stabilizing membrane tubules with the minimum diameter determined by the size of the lipids, a result which is consistent with *in vitro* observations [101]. In contrast to previous estimates that ENTH domains lack the energy or cannot pack tightly enough to bend the membrane [323], we estimate that ENTH domains can reach curvature strength saturation at moderate densities with more than enough space to accommodate the remainder of the protein epsin. In order to extend our conclusions to the mesoscale, we must demonstrate that many small protein-induced deformations can sum to one with a larger extent in order to remodel the bilayer at physiologically relevant length scales. This is the third

component of our mapping, and it has been ably demonstrated by several findings at the mesoscale using continuum models, where we have employed spatially diffusing curvature fields in fluid membranes to study the emergence and stabilization of various membrane morphologies quantitatively consistent with in vitro and in vivo observations [261, 333, 334, 380].

Our methodology and predictions also identify an important biophysical mechanism which is relevant to general membrane remodeling processes. That is, even a small $\sim$ nm-sized domain is capable of coupling a deformation field to background thermal undulations, and moreover, this coupling allows small proteins to affect bilayer fluctuations with a much larger wavelength ($\sim$ 10s of nm). This provides a possible mechanism for protein enrichment at the sites of membrane remodeling in which the sensing mechanism occurs through the curvature undulation coupling at distances much larger than the size of the protein. Consistent with our curvature sensing hypothesis, we find that the diffusion coefficient of the individual ENTH domains is a strong function of the background curvature of the membrane and of the presence of other proteins in the neighborhood (see figures 4.9 and 4.10). We note that the study of the diffusion coefficient also rationalizes the stability of the cooperatively induced curvature field due to multiple proteins. Based on the computed values of D, we estimate that in a typical trajectory of $\mathrm{t} =\sim 2\mu\mathrm{s}$ that we harvest, each domain can move a characteristic distance of $\sqrt{4\mathrm{Dt}} =\sim 10\,\mathrm{nm}$, which is long enough for sufficient conformational sampling in the region of the induced-field. Hence, the fact that the induced-field is persistent in this timescale implies its stability within these timescales. Alternatively, we can utilize the Hill analysis presented above to estimate the stability of the induced-field using the relationship $\mathrm{K_n} = \exp(-\mu_\mathrm{n}/\mathrm{k_B T})$ [77], where μ_n is the chemical potential of $\mathrm{A_n}$, which represents the stability of one assembled (or cooperatively-induced) curvature field. Converting from number of proteins to mole fraction, we estimate that $\mu_\mathrm{n} = -35\mathrm{k_B T}$ (or ~ -21kcal/mol at $\mathrm{T} = 310\mathrm{K}$).

Our models make contact with experiments in two ways: (1) they recapitulate known features of protein-protein and protein-lipid association and (2) they predict membrane geometries observed in experiments. We find that ENTH domains remain bound to the bilayer for the duration of the simulation without any customization or modification to the force field typically necessary to tailor simpler coarse-grained models to specific experiments. This result comports with the slow measured off-rate of $1\,\mathrm{s}^{-1}$ [273] and its nanomolar membrane binding affinity [137]. We attribute this persistent binding to favorable interactions between the inserted helix-0 and a single associated $\mathrm{PI}(4,5)\mathrm{P}_2$, which provides a binding enthalpy of $-14\,\mathrm{k_B T}$ according to isothermal titration calorimetry [101]. The $\mathrm{PI}(4,5)\mathrm{P}_2$-protein association is stable throughout our simulation, and the associated ENTH domains have diffusion rates as low as $1\mu\mathrm{m}^2\mathrm{s}^{-1}$. This matches membrane-bound epsin diffusion measurements under total internal reflection fluorescence (TIRF) microscopy [273] which report 2D lateral diffusion rates of $1.5\mu\mathrm{m}^2\mathrm{s}^{-1}$. We also find that ENTH domains generate deformations with a degree of cooperativity $\mathrm{n} = 3$ (see figure 4.6) that suggests that tetramers are required to bend the bilayer. This observation is parsimonious with physiological con-

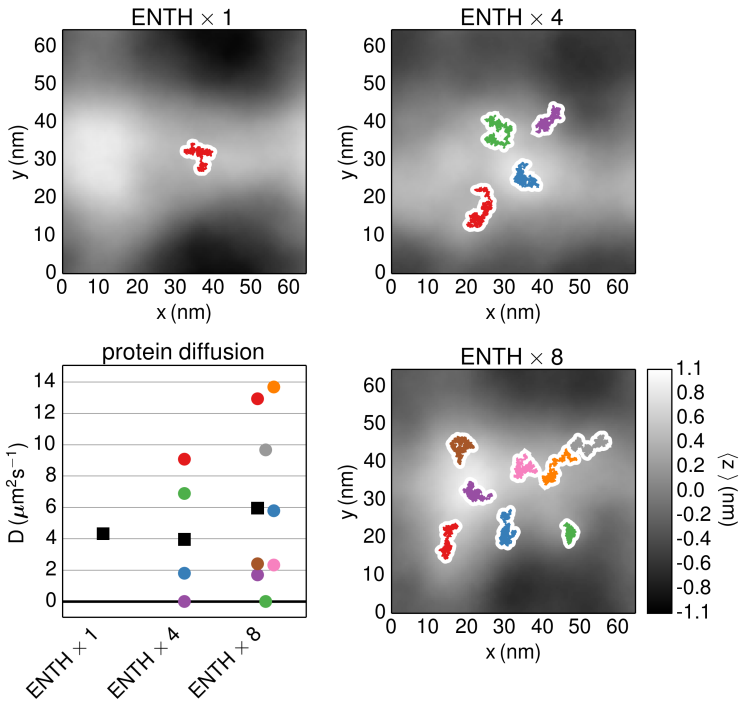

Figure 4.9: Protein diffusion coefficients (colored symbols) and average protein diffusion (black squares) for each protein-membrane system. We compute the absolute diffusion rate by first measuring DOPC diffusion and comparing it to experiments (coarse-grained time is 3.6 faster than real time in these simulations, see section 4.2). We find that proteins diffuse roughly half as fast as the surrounding lipids, but that this diffusion rate is influenced by the protein's position in the average deformation field. Figure 4.10 depicts the mean-squared displacement curves used to compute these diffusion coefficients. We have set the diffusion to zero for proteins which show anomalous diffusion at intermediate timescales.

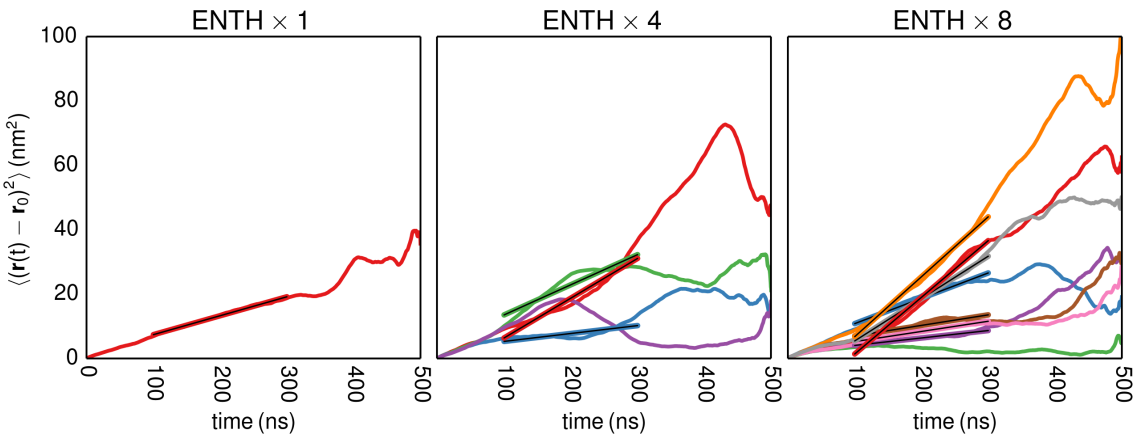

Figure 4.10: Mean-squared displacement curves used to generate the diffusion coefficients summarized in figure 4.9. We extract the two-dimensional diffusion coefficient D according to the relation $\langle(\mathbf{r}(t)-\mathbf{r}_0)^2\rangle \propto 4Dt$. The nearly-linear displacement curves suggests that the proteins undergo Brownian diffusion, however some proteins in the $8\times$ ENTH system move faster than the typical diffusion of the phosphoinositide binding partner. This suggests that some of the protein motion may be active, and is likely due to the underlying chemical potential gradient due to curvature sensing. Linear fits are performed on protein motion which occurs between $100-300$ ns and are depicted with black lines for those proteins with positive displacements. The colors on these plots match those shown in figure 4.9.

straints on epsin concentrations at the sites of clathrin-coated pit formation. Epsins must be enriched above its native intracellular concentration in order to bend the bilayer, however ENTH concentrations far above the threshold coat might exceed the CLAP-binding domains on the overarching protein coat, which are known to be spaced at ~ 18 nm [103], or the availability of $PI(4,5)P_2$, which must also be enriched at the sites of coat formation according to estimates from our previous work [333]; this average separation is equivalent to that found in our four-ENTH simulation, and our analysis shows that this system can generate the necessary deformation. The relatively low accessibility of both $PI(4,5)P_2$ and binding sites on the coat suggest that a higher degree of cooperativity would impose additional free energy costs on coat formation. At lower concentrations exemplified by our single-ENTH domain system, we observe far less curvature, in agreement with single-molecule tracking experiments that indicate that single ENTH domains lack the ability to sense low curvatures [139]. These features — stable binding, slow diffusion, and cooperative membrane bending — demonstrate that our protein senses curvature collectively, a process which is necessary to explain the self-assembly required to enrich local concentrations of ENTH domains to sufficient concentrations to trigger trafficking events in the cell. In addition to matching these features of protein-lipid association, we find that our predicted membrane maximum curvatures and additional bending energies agree with experimental measurements of ENTH-decorated liposomes, which have radii of 10 nm [173]. The predicted maximum mean curvature at saturation is $\sim 0.03\,\mathrm{nm}^{-1}$ which is above the estimated threshold to vesiculate the membrane at higher densities [2, 333].

While this study treats the bending rigidity of the protein-bilayer system as a constant, we expect this curvature coupling effect to become even stronger when proteins are enriched to higher concentrations and hence produce a material with a more heterogeneous bending modulus; an extension of the curvature-undulation coupling theory to include bilayers with spatially heterogeneous material properties such as bending rigidity will be pursued in the future. By distinguishing deformations at the single-molecule level on a large bilayer, we have demonstrated that this method is sensitive to the molecular details of a particular protein-membrane system. For this reason we expect it to provide a useful method for classifying proteins which sense or induce curvature at physiological concentrations and interpreting *in vitro* studies of membrane remodeling as well as *in vivo* studies of how such proteins localize to regions of preferred curvature [337]. In particular, our calculations explain how small proteins can modulate bilayer fluctuations at long distances and how relatively small deformation fields are capable of coupling to the background thermal undulations of the bilayer. In this way, proteins *focus* membrane curvature in service of generating more curvature and attracting protein adapters which sense this curvature, thereby creating a conducive entropy-dominated mechanism for the creation of large assemblies of curvature-inducing protein domains. In concert with other more recognized mechanisms (such as phosphoinositide clustering [357], see chapter 5), we speculate that the curvature-undulation coupling process could be responsible for protein recruitment. Indeed we have shown in previous work that the chemical potential (or propensity) for such proteins to be recruited to sites of curvature depends on the strength of the protein-curvature field; this coupling quantifies the curvature sensing property of the protein through the chemical potential [333, 334], through which intracellular trafficking events may be orchestrated in order to influence cell phenotype. More specifically, the ability of both curvature-inducing proteins and the applied tension to impact undulating modes of large wavelengths implies that the curvature-undulation coupling mechanism may explain why curvature-inducing/sensing proteins can also sense membrane tension, thereby making them conducive to mechanosensing.

Chapter 5

Phosphoinositides

Chapters 2 through 6 are restricted to length scales of tens of nanometers and above. In this chaper, we will increase our spatial and temporal resolution to Angströms and femtoseconds in order to explore the highly-charged phosphoinositides which play crucial roles in cellular signaling, but which cannot be resolved with coarse-grained simulations. The work described in this chaper was performed jointly with Dr. David Slochower and in close collaboration with Professor Paul Janmey.

5.1 Phosphoinositides orchestrate many signals on cell membranes

Biological membranes have a rich composition the spatial distribution, function, and interactions of which have been the subject of much scientific inquiry and debate [346]. Cell membranes are defined by their primary constituent, phospholipids, which act as a substrate for a wide range of biomolecules that organize themselves along the surfaces it creates. However, the building blocks of the membrane material itself is also part of the biochemical *foreground* in part because it hosts phosphoinositides (PPIs), a class of phospholipids with an inositol ring that provides binding sites for up to three additional phosphate groups.

The multiplicity and relative positions of these groups confer an incredible biochemical specificity [183]. One molecule in particular, phosphatidylinositol 4,5-bisphosphate (PIP_2) is unique in that it acts as a beacon for hundreds of cytosolic proteins and regulates a number of processes that require them [109]. Defects in PPI signaling are implicated in a number of melanoma, lung, breast, and ovarian cancers [52, 95, 160]. Since the 1960s, investigators have observed cation-mediated domains in plasma membrane extracts [134]. Recent study has located PIP_2 in such domains, finding that they host a manifold higher concentration compared to background [144, 308, 340, 357].

5.1.1 Characterizing lipid clusters

In the remainder of this chapter we will investigate the hypothesis that cation-lipid binding affects the motion and binding partners of PIP_2, and specifically that strong bonds formed with calcium in particular can influence the structure and dynamics of the bilayer. The overhead view of diffusing lipids shown in figure 5.2 suggests that the PIP_2 molecules move slower than the other lipids. We often turn to abstractions of our simulated bilayers in order to understand the motion of the molecules inside them. The resulting images can provide intuition about the molecular motions that give rise to the altered material properties measured more precisely in subsequent exposition.

5.1.2 Physiological bilayers

For this study, we have constructed twelve simulations. The focus of our analysis will be so-called "asymmetric" bilayers designed to have a lipid composition that mimics the inner leaflet of the plasma membrane, and is distinct from the outer leaflet. We also include simulations which control for the effects of cholesterol, the identity and charge of counterions, the charge state of PIP_2, and the phosphate positions on the inositol ring. By including a diverse set of compositions in this study, we hope to underscore the chemical specificity of processes that depend on PIP_2.

5.2 Atomistic simulations resolve PIP_2-cation binding

We turn to atomistic molecular dynamics simulations in order to overcome the diffraction limit, which makes it impossible to resolve lipid-ion binding in atomic detail. We select the `CHARMM36` force field [97,166] and use `GROMACS 4.6.3` to simulate our model bilayers [342]. We used updated force field parameters for PIP_2 based on quantum calculations [306]. We have investigated twelve distinct compositions. Each bilayer has 400 lipids (including cholesterol in the physiological bilayers) in each leaflet, contains positive and negative counterions to reach a charge neutrality at concentrations from $150 - 225$mM, and solvated with enough water to ensure at least 5 nm separation between periodic images. Our "physiological" bilayers have an asymmetric composition with 75% POPC and 25% cholesterol in the outer leaflet. The inner leaflet was composed of 50% DOPE, 25% cholesterol, 15% DOPS, and 10% PIP_2. Symmetric bilayers contained a 4:1 mixure of DOPC and DOPS along with 10% PIP_2 in both leaflets.

In this simulation protocol, a randomized grid of 400 lipid structures was assembled for each leaflet and composition and arranged with a regular 1 nm spacing. The lipids were fixed with mild position restraints of $500\,\mathrm{kJ}\,/\,\mathrm{mol}\cdot\mathrm{nm}^2$ in the normal direction and then "packed" into a bilayer using a vacuum equilibration procedure in order to ensure that no lipids flipped to the opposite leaflet. A number of simi-

lar bilayer-construction tools are available [313], however our procedure is seamless with the tools described in appendix A. The resulting bilayer configurations were then solvated with water and then one of four different cations (sodium, potassium, magnesium, or calcium) and chloride ions. As per instructions provided by Klauda *et al* [166], we employed `CHARMM` "special" water (`TIPS3P`) which includes Lennard-Jones interactions on water hydrogen atoms in order to prevent artefacts in lipid motion.

Our systems were equilibrated for at least 20 ns followed by production runs lasting 80 ns. These systems were held at a temperature of 310K using the velocity-rescaling thermostat due to Bussi *et al* [53] with a coupling frequency of $0.5\,\mathrm{ps}^{-1}$ and atmospheric pressure via the Parrinello-Rahman barostat with a coupling constant of $2.0\,\mathrm{ps}^{-1}$. The `LINCS` algorithm contrained hydrogen bond distances allowing for a 2 fs timestep. Van der Waals forces were switched off smoothly from $0.8 - 1.2\,\mathrm{nm}$, and electrostatics were computed according to the particle-mesh Ewald summation (PME) with a Fourier spacing of 0.16 nm.

These simulations were performed over a timespan of three years on several supercomputing platforms provided by XSEDE, including `LONESTAR`, `COMET`, `GORDON`, and `TRESTLES`. Simulations contained up to 360,000 atoms including water and counterions and achieved simulation speeds of up to 16 ns / day on 6×16-processor platforms. The simulations ran for an aggregate production time over $1\,\mu$s.

5.2.1 Analysis

A typical simulation snapshot is provided in figure 5.1. All such snapshots, including supplemental videos, have been rendered with Visual Molecular Dynamics (VMD) [142]. An example simulation is depicted in supplemental video `SV5.1`.

In 5.3 we will measure a number of bilayer properties using a set of analysis codes written in `Python` which make use of several libraries, namely `MDAnalysis` [224] for reading and selecting molecules, `NumPy` [343] and `Scipy` [154] for creating Delaunay triangulations and Vornoi tesselations of the lipids. We extract two kinds of data from our simulations. First, lipid-ion binding measurements given in section 5.3.1 are computed by counting the numbers of close ions using the pairwise distance calculator available in `SciPy` modified for use under periodic boundary conditions (PBCs). Second, material properties are computed on a triangulated mesh of the lipid centers of mass (also modified to obey PBCs). This mesh allows us to compute the individual lipid areas, while a regular interpolation of these points provides the undulation spectra and bending rigidities. An example of our mesh is depicted in figure 5.3 for a bilayer with a physiological composition.

Bilayer simulations were constructed using codes outlined in appendix A.

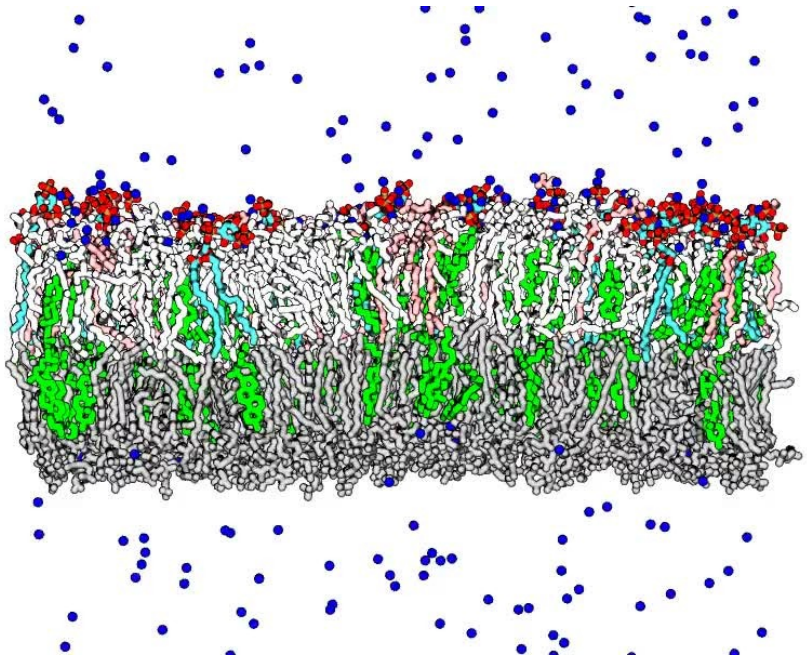

Figure 5.1: Simulation snapshot of a bilayer containing 800 lipids with PIP_2 found in the innner leaflet (top) along with DOPE (white), DOPS (red), and cholesterol (green). The outer leaflet (bottom) contains POPC (gray) and cholesterol. PIP_2 molecules are colored by atom (e.g. carbon is cyan, oxygen is red, phosphorus is tan). Calcium cations are shown in blue while water and chloride ions are hidden.

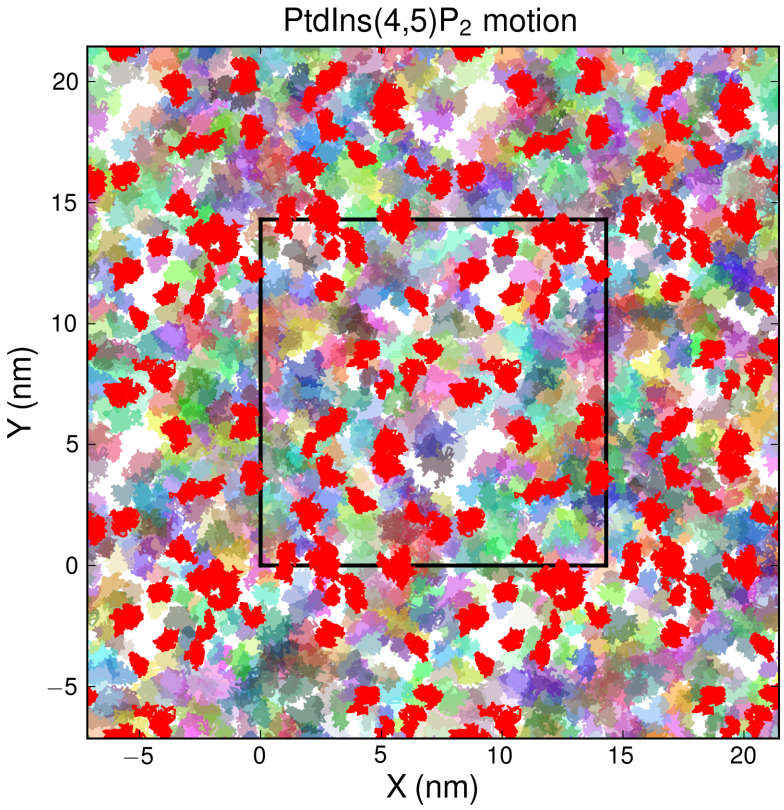

Figure 5.2: A depiction of lipid diffusion in which PIP_2 trajectories (over 50 ns) are highlighted in red, while other lipids have distinct per-molecule colors. The periodic boundary is drawn in black in the center of the box. In general, PIP_2 molecules diffuse slower and form pairs and triplets with each other, typically bridged by divalent cations.

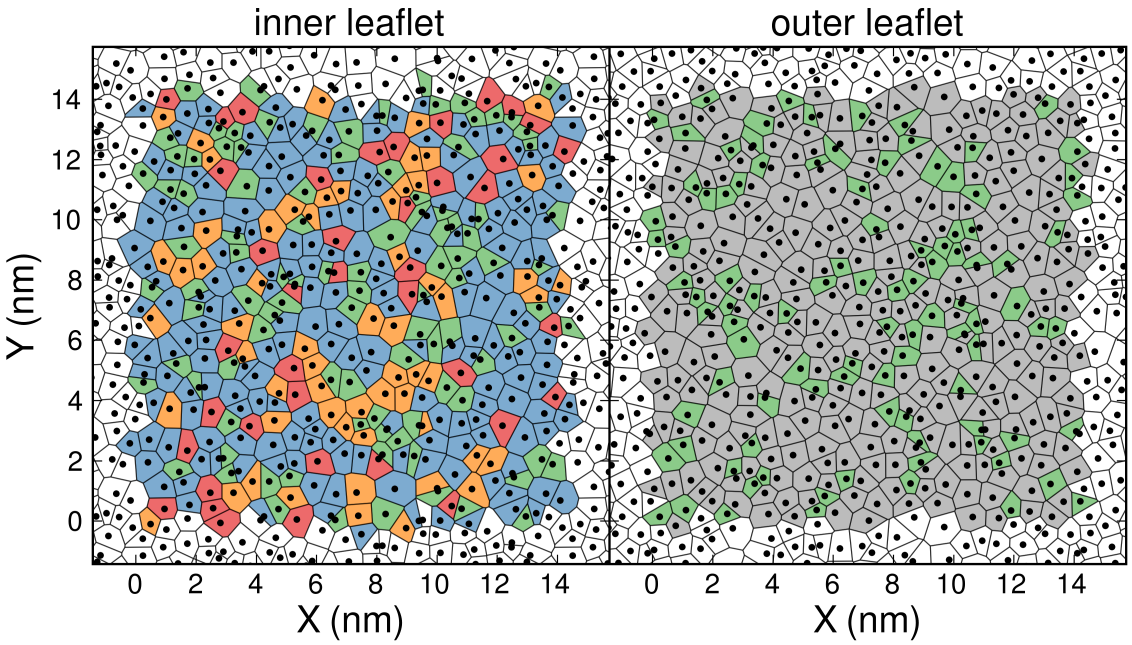

Figure 5.3: A representation of the Voronoi mesh for a physiological bilayer. We show the diester phosphate group with a black dot. These positions are used to generate a Voronoi tiling of the surface using the algorithm provided by `SciPy` with modifications for periodic boundary conditions. The unit cell of the mesh is colored according to lipid identity, where DOPE is blue, DOPS is orange, cholesterol is green, POPC is gray, and PIP_2 molecules are red. These mesh structures were used to generate the measurements described in 5.3.2.

5.3 Results

In this section we will outline two kinds of results: (1) *molecular* measurements such as ion-lipid binding distances and (2) *bilayer* properties computed from the triangulated meshes of the membrane.

5.3.1 Molecular measurements

Counting close ions

In order to understand the effects of divalent cations on a PIP_2-laden bilayer, we must first classify the types of bonds they form with the phosphate groups attached to the inositol ring (hereafter, the "headgroup"). We do this by counting ions found within a cutoff distance spanning a range of $1.8 - 5.0$ Å from any atom in a PIP_2 molecule. An example of this measurement is pictured in figure 5.4.

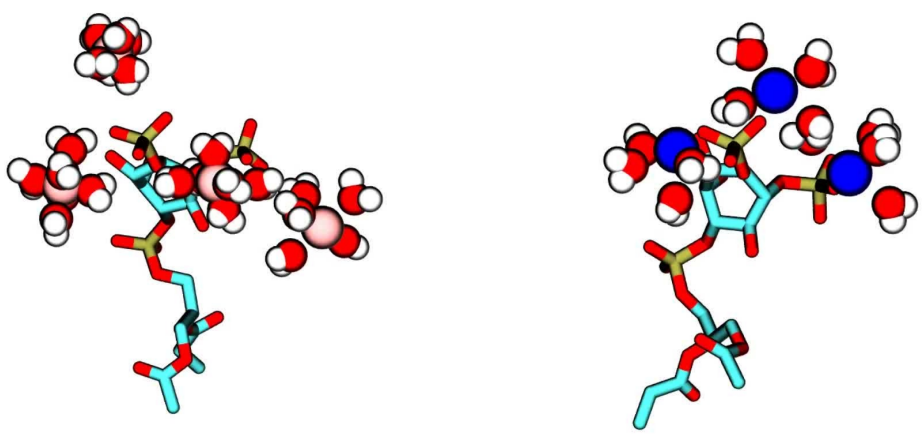

Figure 5.4: Snapshots of separate simulations of PIP_2 with magnesium (left, pink) and calcium (right, blue) along with hydrated water. This image is a snapshot of from supplementary video `SV5.2` which shows that the dehydrated calcium bonds are more persistent. The hydophobic tails are omitted here.

We classify bound ions by their distance from PIP_2 molecules and the number of lipids that can also be found in this distance. Figure 5.5 shows these counts for all twelve simulations, distinguished by the number of neighboring lipids that fall within the cutoff distance.

These counts capture a striking result: namely that calcium ions form a tight, nearly ionic ≤ 2.2Å bond with PIP_2 while other ions generally do not. If we expand our cutoff distance to 4.6Å (see figure 5.6) we find far more ions in the zone. The

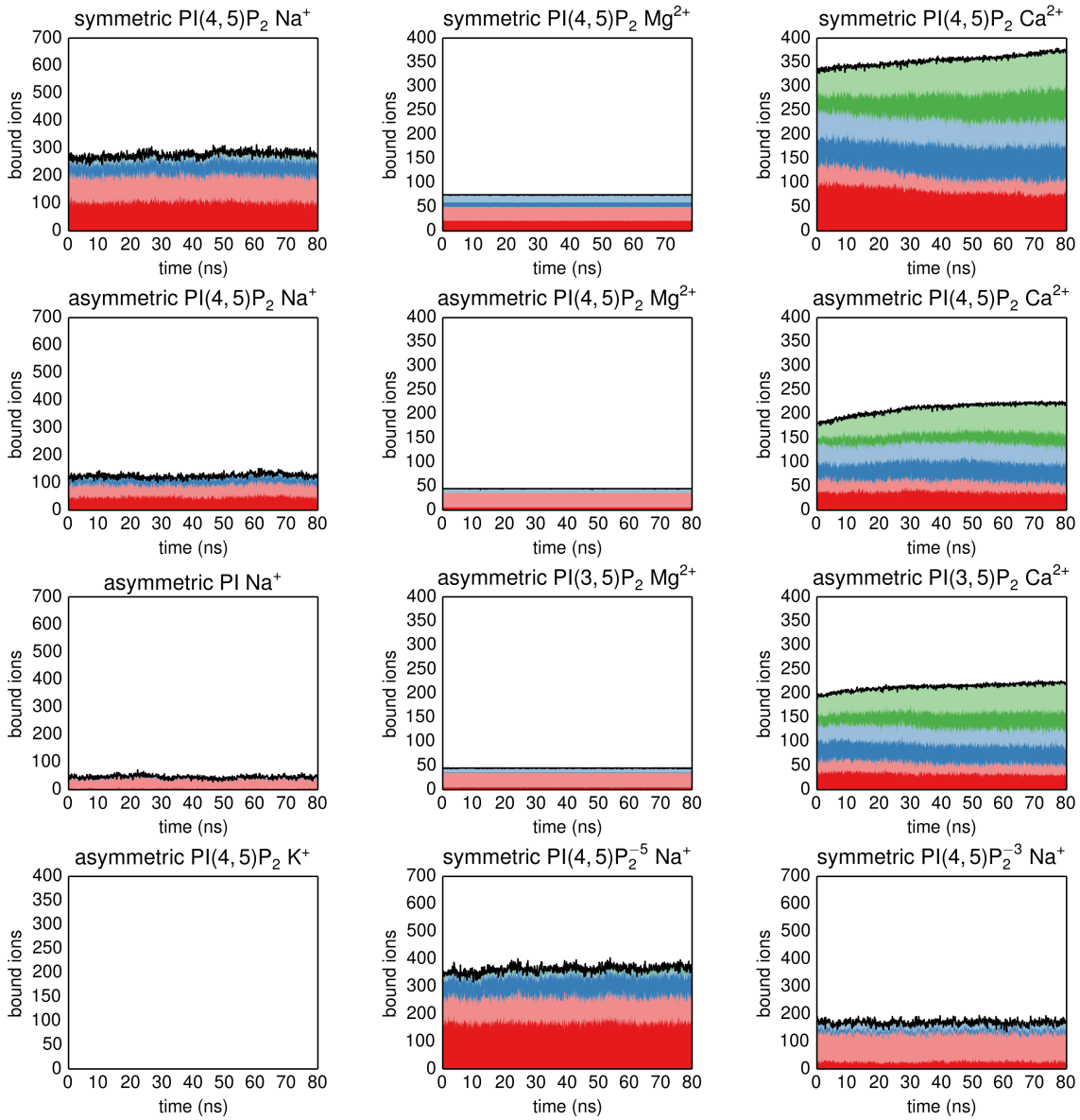

Figure 5.5: Counts of closely-bound ≤ 2.2Å ions distinguished by the number of lipids which are also bound to these ions. Red bands indicate that the ion is bound to only a single lipid; blue bands indicate that the ion is shared with exactly *two* lipids, and green bands count the ions bound to three or more lipids. Within each colored band, the dark portion counts the number of ions that are bound to at least one (or in the case of the red bands, exactly one) PIP_2 molecule. We find that divalent cations form far more close bonds with all of the lipids in the bilayer, and that these bonds typically include PIP_2.

high $-4e$ charge on PIP_2 ensures that most of the ions in the simulation are drawn to the bilayer, also possibly forming an electric double layer [133].

We also observe the differences in calcium and magnesium binding in figure 5.4 (see also supplementary video `SV5.2`), which shows that magnesium ions tend to associate with PIP_2 along with intervening waters while calcium ions bind directly to the phoshate groups. This stronger bond effectively creates a larger effective mass for diffusing lipids connected by ions. In the results below we will call this "lipid bridging". This bridging effect can also be seen in figure 5.6 and figure 5.5, where the green bands, indicating ions bound to at least three lipids, grow slowly over time. This demonstrates that the bilayer is "charging" on the relatively slow, 100 ns timescale of the simulation. Observations of a dehydrated calcium-PIP_2 bond agrees with single-molecule free energy calculations which show that magnesium is more difficult to dehydrate [307].

Changes in headgroup conformations

The identity of the divalent cation also influences the conformation of the inositol ring as measured by its two degrees of freedom: tilt and rotation. The tilt angle determines how far the ring extends into the solvent, while the rotation angle determines which inositol positions (particularly positions two versus four) are exposed to the solvent. These angles are summarized in figure 5.7.

The distributions of both tilt and rotation angles have implications for protein binding because they determine the relative accessibility of the phosphate groups to binding sites on potential membrane-binding proteins. 5.8 shows these distributions for the physiological simulations. It has several distinct features. First, the tilt angle distributions show that the group is usually tilted "forward" (if the reader imagines the 5-phosphate on the left). However, simulations with calcium show more "backward" tilts those containing magnesium or sodium.

In contrast, the rotation angle determines whether the 5-phosphate projects into the solvent or becomes sequestered underneath the headgroup. We also find more negative rotation angles, but in this case they indicate that the 5-phosphate position is rotated *down* towards the bilayer, possibly engaged in ion-lipid bridging. Ion-lipid bridging along with varying headgroup shapes can also be seen as a dynamic process depicted in supplemental videos `SV5.3` and `SV5.4`.

5.3.2 Bilayer measurements

In 5.3.1 we provided evidence that calcium-PIP_2 bonds are qualitatively different than bonds with other ions, even with the same charge (e.g. magnesium). Here we will investigate these implications to a scale which includes many lipids acting in concert. In particular, we hypothesize that these strong bonds make the effective diffusive unit larger in proportion to the number of additional lipids that are associated with the

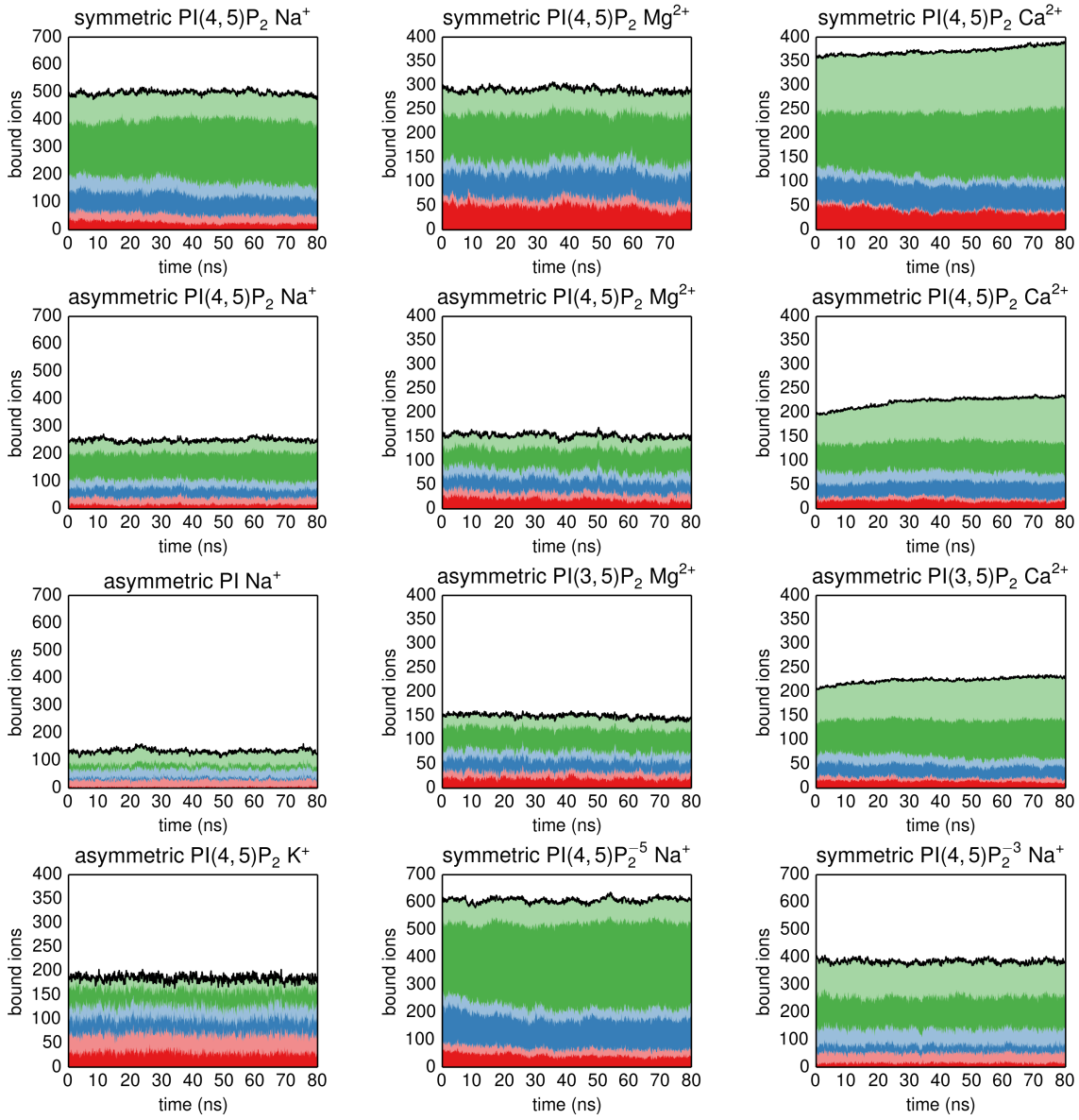

Figure 5.6: This figure uses the same coloring scheme as figure 5.5, with a larger cutoff distance of 4.6Å. At this larger distance we find more similarities in the numbers of bound ions, however PIP_2 still accrues more calcium than magnesium.

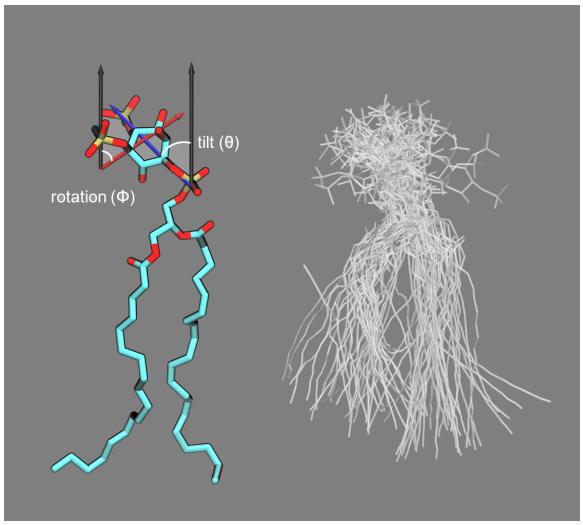

Figure 5.7: A description of the headgroup rotation and tilt angles (left) alongside a depiction of the PIP_2 motion (right), which includes aligned PIP_2 structures to demonstrate the wide range of headgroup motion.

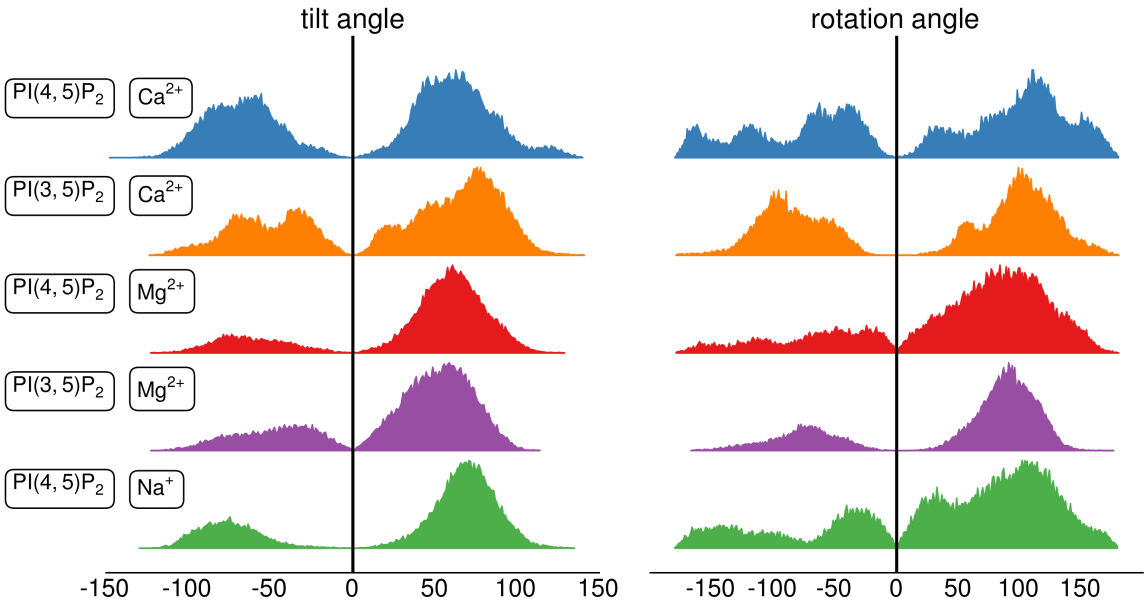

Figure 5.8: Distributions of tilte (θ) and rotation (ϕ) angles for 80 ns simulations of PIP_2 in asymmetric bilayers. See figure 5.7 for a definition of these angles. In this measurement, positive angles indicate that the headgroup is titled *back* or that the 5-phosphate on the inositol ring is pointed *down*.

PIP_2 molecule. We call the latter "ion-lipid bridging".

Diffusion

In good agreement with experiments [357], we find that calcium slows lipid diffusion. Figure 5.9 shows the overall diffusion rates for the physiological bilayers, as well as the individual lipid diffusion rates. We find that calcium slows diffusion, and PIP_2 (except for the PI control) always diffuse slower than the other lipids. This result supports our hypothesis that ion-lipid bridging forms strong bonds which give small clusters of PIP_2 molecules a larger effective mass, hence slowing their lateral motion.

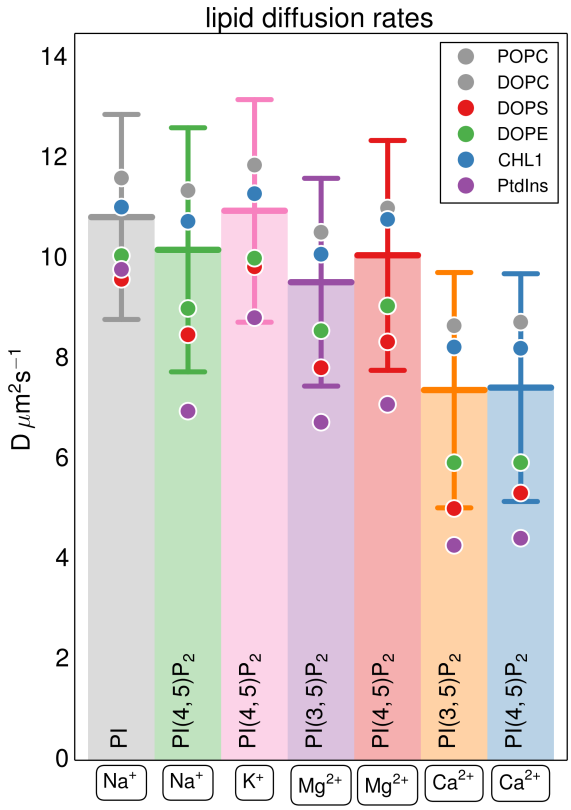

Figure 5.9: Total lipid diffusion rates computed from mean-squared displacement curves over 80 ns simulations. We find that calcium significantly slows diffusion compared to magnesium and control simulations. We also find that PIP_2 (purple dots) diffuses slower than the other species. Error bars are generated from the standard deviation of the diffusion coefficient across all molecules in the system.

Observing changes in lipid areas. The *area per lipid* is an important biophysical quantity because it encodes crucial information about both the density and rigidity of lipid bilayers. It is typically measured by experiments (see chaper 2) and is one of the crucial points of contact between molecular models and experiments. We have

computed the lipid areas in two ways: first in the projected plane, largely to control for the effects of the size of the PIP_2 headgroup, and secondly, in three dimensions, in order to understand the areas in each monolayer/leaflet. The projected lipid areas provide evidence that calcium condenses the bilayer area. We observe a reduction of $\sim 1.5\,\text{Å}^2$ per lipid between simulations containing calcium and magnesium, however this area change is distributed over *all of the lipids* in the bilayer. Since PIP_2 is only a minor (10%) constituent of the inner leaflet, this is a significant change, equivalent to $\sim 15\text{Å}^2$ per PIP_2, greater than the average difference between POPC and POPE, attributable solely to the chemical identity of the cation.

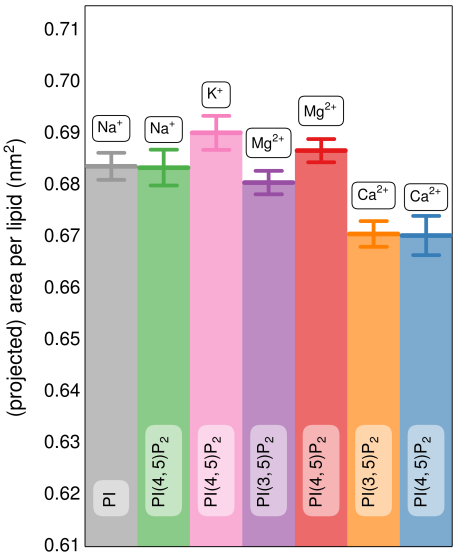

Figure 5.10: Projected (i.e. two-dimensional) areas per lipid for physiological bilayers. We find that calcium reduces the area per lipid. Even though the differences are small, they are distributed over all lipids in the bilayer, and are typically modulated only by cation identity or phosphate position. Error bars are computed from the standard deviation over 10 ns samples of the trajectory.

A novel feature of our simulations is the use of asymmetric lipid compositions. That is, the inner leaflet contains charged lipids like PIP_2 and DOPS while the outer leaflet contains only POPC (and cholesterol, which is found in both leaflets). Using the triangulated meshes described in 5.2.1 we have computed the three-dimensional areas of each leaflet in order to quantify the communication between them. Figure 5.11 shows the difference in monolayer leaflet areas between the inner and outer leaflets.

Having constructed these bilayers with equal numbers of lipids in both leaflets, we typically find that the inner leaflet has a larger area. We nevertheless conclude from these measurements that calcium shrinks the area of bilayers containing PIP_2, and that this condensing effect can be transmitted to the outer leaflet as well. This provides a possible mechanism for cation-induced domain formation on GUVs and supported lipid monolayers [357]. If we estimate that highly charged lipids diffuse slowly enough (see 5.3.2) or that PIP_2 form clusters (see the evidence described in

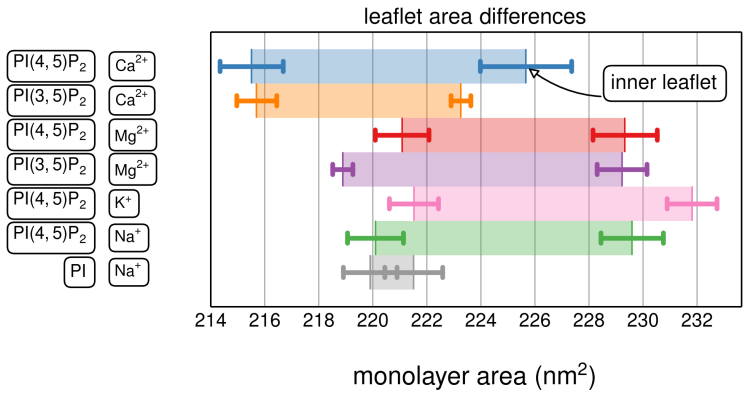

Figure 5.11: Monolayer leaflet areas computed from the Delaunay mesh between lipid centers of mass under periodic boundary conditions. Activated PIP_2 molecules show a larger difference in leaflet areas than the control PtdIns (PI) simulation. This extra area has implications for lipid packing. Error bars are computed from the standard deviation of 10 ns samples of the trajectory.

5.1.1) then we can infer that the sites of PIP_2 enrichment may give rise to an excess area that drives domain formation.

Bending rigidity We compute height-height autocorrelations as per methods described in chapter 4. This provides the undulation spectra from which we can extract the bending rigidities summarized in figure 5.12. We find that $PI(3,5)P_2$ tends to stiffen the bilayer compared to $PI(4,5)P_2$. This calculation has a relatively high uncertainty, hence this calculation and further exporation of the formation of nanodomains would be well-served by mapping the specific PIP_2-cation interactions onto a coarse-grained force field, particularly if the close association and screening effects between PIP_2 and divalent cations could be made explicit in those force fields.

5.4 Discussion

In this chapter we have outlined a number of findings that support experiments which show calcium-induced lipid clustering. The most direct evidence in support of the lipid clustering hypothesis comes from observations of the ion behavior. Not only do calcium ions bind more tightly than magnesium, but these bonds also typically expel water, in good agreement with attenuated total reflection Fourier transform infrared (ATR-FTIR) experiments [357] and single-molecule calcuations [307]. These same experiments show a slowed diffusion that we also observe in our simulation.

Observations of ion charging (see figure 5.5), cations which are shared between lipids (i.e. "bridging"), and dehydration of tightly-bound calcium lend support to the hypothesis that calcium induces domains by connecting PIP_2 molecules together, thereby slowing their diffusion, and changing their interactions with other lipids in

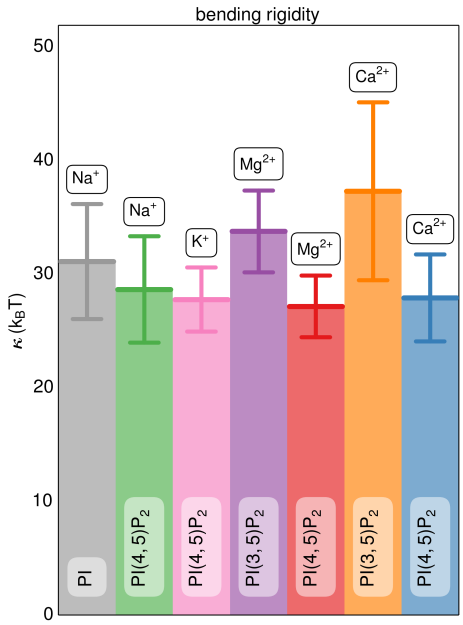

Figure 5.12: Bending rigidity ($\boldsymbol{\kappa}$) computed from the height-height undulation spectra for the physical bilayers. We find that the bilayers containing $PI(4,5)P_2$ are softer than those containing $PI(3,5)P_2$. Error bars are computed from the standard deviation over 10 ns samples of the trajectory.

the bilayer.

Our bilayer measurements support this theory by finding changes in both the two-dimensional lipid area and the three-dimensional leaflet area, both of which have implications for lipid packing. Modeling efforts to understand domain formation suggest that the precise balance between total molecular volume and area may determine how lipids sort into domains or stabilize the interface between membrane domains [49].

5.4.1 Future directions

Observed differences in lipid packing have implications for PIP_2 binding to proteins. Background PIP_2 concentration may be too low to recruit proteins for intracellular trafficking events [183], in which case any local enrichment may trigger these events by recruiting the necessary protein components. Proteins, and amphipathic domains, are known to sense packing defects and differences in lipid area [244].

An accurate model for PIP_2 is essential to many of the models discussed in this dissertation. We expect that the unique features of PIP_2 quantified in this study — in particular the geometry of the headgroup, the lipid area, and the effects of ion-lipid bridging — may determine its ability to separate into domains or select particular protein binding partners. Our understanding of the biological function of PIP_2 would benefit from further studies of protein-PIP_2 interactions, particularly

with actin nucleating and branching proteins [270].

Moreover, a well-characterized atomistic model for PIP_2 could be more easily extended to coarse-grained modeling efforts to explicate the formation of these domains. In the next chapter (6), we will review the multiscale modeling method. Our study of PIP_2 provides a useful, albeit challening, multiscale modeling target due to the coupling between strong electrostatic bonds and much slower lipid and protein reorganization on the bilayer.

Chapter 6

Multiscale coupling

If we think of biological systems as communication systems, then the frequency of this communication varies over *twenty* orders of magnitude in time and *ten* orders of magnitude in space. In this chapter we will draw on the findings described in chapters 3 and 4 in order to characterize the multiscale mapping strategy necessary to apply molecular models to *in vitro* and cell experiments.

6.1 Multiscale coupling strategies

The review of coarse-grained molecular dynamics simulations provided in chapter 2 describes the most limited kind of multiscale coupling: from atomistic simulations to slightly-larger coarse-grained ones. While this coupling strategy has made all of the research in this dissertation tractable, it represents a very *modest* type of coupling. Recall that coarse-grained simulations of lipid bilayers aim to reproduce many sensitive biophysical parameters such as area-per-lipid, the stress distribution across the bilayer, and lipid phase behavior very precisely. For that reason, the most rigorous parameterization methods require an enormous amount of "bandwidth" between simulations, often, as is the case for the multiscale coupling algorithms [240], requiring independent atomistic simulations of target systems. The multiscale efforts described in this thesis represent the opposite strategy in which we pass relatively little information between scales. This strategy is far more common in the biological literature, which often must contend with far greater measurement uncertainty while spanning larger gaps in time and space to make useful models.

Typically, molecular dynamics barely extend beyond hundreds of nanometers and millisecond simulation times for small proteins. Modeling biological systems at larger length scales often requires a combination of several strategies, including solutions to ordinary and partial differential equations, descriptions of cell signaling networks in Boolean or agent-based models, and stochastic models [266]. Larger target systems often require multiple modeling strategies at once, and are typically tuned to study

specific problems in, for instance, tumor metastasis [255, 358]. As a general rule, larger target systems require investigators to compartmentalize many parts of their theory in order to minimize the information that is passed between components in the model. This serves two functions: it ensures that fine-grained measurements are properly sampled before being used in a coarser model, and it usually indicates that the models are computationally and experimentally tractable.

6.2 Contacting experiments

To explore the multiscale modeling method, we will treat the models described earlier in this thesis as "case studies" in model-building that comment on opposite intracellular transport fluxes but jointly rely on a similar modeling framework. Both projects were inspired by questions raised by *in vitro* experiments performed by collaborators working under Professors Wei Guo and Tobias Baumgart. The first study showed that the N-terminal domain of an endocytic adapter protein epsin (ENTH) is sensitive to the curvature of the supported lipid bilayers to which it adheres [139]. The second study was prompted by confocal microscopy images which indicated that the protein Exo70 could generate inward-pointing tubules on large unilamellar vesicles [380]. The biological context of for these systems are provided in chapters 4 (ENTH) and 3 (Exo70). The biological details of both processes provide the target for our multiscale modeling efforts.

That said, it is also important to note that efforts to link experiments and theory across multiple scales often turns the experiments themselves into modeling "targets". Even though our ostensible goal is to describe a problem in biology, we often spend as much effort verifying and validating our theory. In some limit, the question of theory and experiment collapses into a single problem.

Dimensionality reduction. The number of "bits" of information (given by a `0` or `1`) encoded by these experiments is fairly low, especially in comparison to the rich chemistry experiments required to build the molecular models in the first place. In the following exposition, we will find that matching theory and experiments almost always requires a massive reduction in the dimensionality of our data set. This happens both *within* experiments and *between* them. A single microscopy image might include billions of molecules, even if the readout is as simple as answering the question: "do the proteins generate protrusions?" The act of "compressing" these data into a single claim is an essential part of most scientific inquiry, and attending to this process is an important part of multiscale modeling.

Membrane bending at the mesoscale. So far, I have described the multiscale modeling process as a highly general process by which we combine observations at many scales in order to make sense of them. In practice, however, matching experiments

at different scales typically requires a theory that can link them. The degree to which this theory is made explicit often determines the success of the effort, and in this regard, soft-matter biophysics researchers are lucky to have a nearly totalizing theory for how membranes bend: the Helfrich Hamiltonian [131]. Other modeling investigations do not typically have this advantage, and must rely on a Boolean set of (usually verbalized) logical propositions and careful definition of experiments to guide their efforts (i.e. "when protein **A** *activates* protein **B**, then ...”). In this work we seek to mimic the holistic approach used by biologists to describe the cell, while acknowledging the advantages conferred by the relatively elegant continuum theory.

6.3 Mapping between scales

In contrast to the high-bandwith multiscale modeling efforts described in section 6.1, the two examples we consider here have remarkably low bandwidth. That is, we pass only a single function from the molecular to mesoscale models: the concentration-dependent curvature fields. This requires two important assumptions. First, we assume that we have measured the curvature accurately in the molecular models. Second, we assume that our mesoscale model is well-founded and well-sampled so that the conclusions we draw are faithful to the underlying physics. If we imagine that multiscale modeling is a metaphorical signal-processing problem, then we must ensure that each model "trusts" the information coming from the other, and in particular, that the information is stated in the same language.

In the two studies described here, we can be sure that both the continuum and molecular models are speaking the same language because they have been extensively validated. The molecular models obey the Helfrich Hamiltonian down to very small scales (see section 4.2.2) and with physical constants (such as bending rigidity) which agree with experiments (see chapter 2). In previous work from the Radhakrishnan lab, Tourdot *et al* and Ramakrishnan *et al* have shown that continuum simulations and free energy calculations can be effective tools for quantifying the physical forces which constrain membrane shapes, in good agreement with experiments [261,334,335]. These foundations ensure that the individual models are internally consistent, and improve our confidence that combining them can provide useful predictions. In this section we will compare the multiscale modeling methods in both projects, and in the conclusion (section 6.4) we will generalize this process.

6.3.1 Predicting negative curvature in Invadopodia

The molecular model. In chapter 3, we recapitulated a method for estimating the curvature fields generated by dimers of the protein Exo70, a key component of the exocyst. As with many condensed matter, or more generally, entropically-mediated phenomena, our concept of the protein-induced curvature is highly unintuitive. Curvature generated by Exo70 is difficult to visualize in a dynamics video or average

surface structure because of thermal undulations, and the overall bilayer deflection in these simulations amounts to only several nanometers over an extent which is dozens of nanometers long. We employ a surface-fitting procedure described in chapter 3 resulting in fields depicted in figures 6.1 and 6.2.

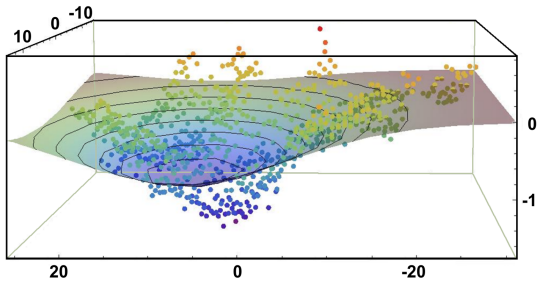

Figure 6.1: A depiction of a single frame of the curvature fitting algorithm in which a two-dimensional Gaussian function is fit to the heights of lipids in the neighborhood (15 nm) of the protein. Downward deflection (blue) corresponds to $z < 0$ and indicates negative curvature. These fits were used to generate average membrane surface shapes shown in figure 6.2.

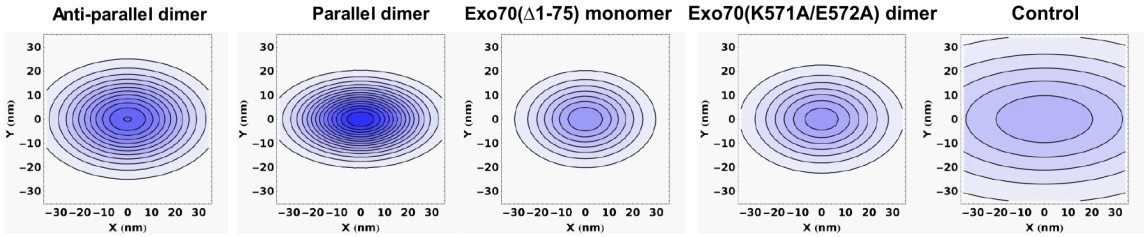

Figure 6.2: A depiction of the two-dimensional Gaussian "dimple" which represents the average curvature and extent of the membrane surfaces measured using the surface fitting method. In this case we find that the dimer systems create stronger, more focused negative curvature fields. The parameters used to draw these fields are summarized in chapter 3.

This measurement lacks the finesse of the method described in 4, but nevertheless provides an explanation for the differences we see in the membrane dynamics. It also crucially recovers an anisotropic curvature field, which is common among rod-shaped proteins that induce curvature [222]. The strengths and extents, summarized earlier in 3.1 are direct inputs to the mesoscale model.

Mesoscale membrane bending on spheres. To test the hypotheses that the curvature fields measured above can generate changes in the shapes of LUVs observed in experiments, we require a mesoscale simulation to bridge the gap delineated in section 6.2. We select a dynamically-triangulated Monte Carlo (DTMC) simulation of the Helfrich Hamiltonian [131] designed to match the size ($\sim$ 500nm) and topology of the spherical vesicles used in the *in vitro* experiments. This procedure was carried out by N. Ramakrishnan with extensive details provided in the supplemental material of

Zhao *et al* [380]. The Hamiltonian is given by:

$$\mathcal{H}_{\mathrm{el}} = \left(\frac{\boldsymbol{\kappa}}{2}\left(2H - H_0\right)^2 + \overline{\boldsymbol{\kappa}}K + \boldsymbol{\gamma}\right) dA. \tag{6.1}$$

Simulation methods based on the Helfrich are phenomenological, and depend on a set of physical parameters. In this framework the membrane is approximated as a two-dimensional fluid, elastic sheet that is large enough to neglect its thickness. We estimate parameters such as bending rigidity ($\boldsymbol{\kappa}$) from experiments. We neglect the Gaussian rigidity $\overline{\boldsymbol{\kappa}}$, which is constant for systems with a fixed topology. Proteins are represented in this model as a nematic field because they have an elongated shape and induce an anisotropic deformation. This produces orientational order on the surface, and is a necessary component of any model that predicts tubule morphologies [260].

The mesoscale simulations produce a binary readout: the presence or absence of inward-pointing tubules on a triangulated sphere. In the mesoscale simulations presented in [380], the authors found that deformation fields corresponding to the wild type dimers created inward-pointing tubules on model LUVs, while the oligomerization-deficient mutant and the PIP_2-binding-deficient mutant failed to create these tubules (see chapter 3).

We define the curvature field according to the anisotropic spontaneous curvature parallel to the protein, given by $\mathrm{H}_0^{\parallel} = \alpha \mathrm{a}_0^{-1}$. In this expression, α represents the strength of the curvature in units of a_0, the smallest length scale in the simulation. For a particular set of physical parameters, the model produces the constraint $\alpha < -0.6$. When this constraint is satisfied, the model produces tubules. Otherwise, the vesicles remain unchanged.

Inferring the protein scaffold. The mesoscale simulations identify a clear threshold for morphology change. To match this threshold to the experiments and molecular simulations, we have to (1) apply the curvature estimate from the simulation and (2) account for the spacing between proteins. According to the mesoscale model, the mean curvature and energy per vertex are given by,

$$H = \frac{H_0^{\parallel}}{2} = \frac{\alpha}{2a_0},\ E_{vertex} = \frac{1}{2}\left(\frac{H_0^{\parallel}}{2}\right)^2 A_v. \tag{6.2}$$

Any measurement of membrane curvature must take place over a particular membrane area. In the mesoscale model we account for the presence of multiple proteins with a factor n such that the mean curvature energy $\langle \mathrm{E} \rangle$ scales linearly with the number of proteins. This gives a vertex energy of $\langle \mathrm{E}_{\mathrm{vertex}} \rangle = \mathrm{n}\langle \mathrm{E} \rangle$ over an average vertex area of $\mathrm{A_v} = \sqrt{3}(1.3\mathrm{a}_0)^2 / 2$. This provides the following expression for n:

$$n = \frac{\sqrt{3}(1.3)^2\alpha^2}{16\langle E \rangle} \tag{6.3}$$

From the molecular model, we estimate the molecular areas of the monomers, parallel dimer, and anti-parallel dimer as $150\,\mathrm{nm}^2$, $200\,\mathrm{nm}^2$, and $240\,\mathrm{nm}^2$, respectively. Based on the vertex areas, the theoretical maximum number of proteins in a particular area is $\mathrm{n}^\star = (\mathrm{A_v} = 294\,\mathrm{nm}^2)\,/\,\texttt{molecular area}$.

This leads to two conclusions. First, the monomer area is too large to pack enough mononomers into the vertex area required to cross the tubulation threshold given by $\alpha < -0.6$. Second, by noting that the anti-parallel dimer system barely satisfies the constraint $\mathrm{n} < \mathrm{n}^\star$, we can rewrite the morphology threshold in terms of its radius of curvature measured in the molecular simulations. This leaves a critical radius of curvature $\mathrm{R}^\star = \langle 2\mathrm{H}^\star_{\mathrm{max}} \rangle^{-1} = 53.4\,\mathrm{nm}$. These results are summarized in figure 6.3.

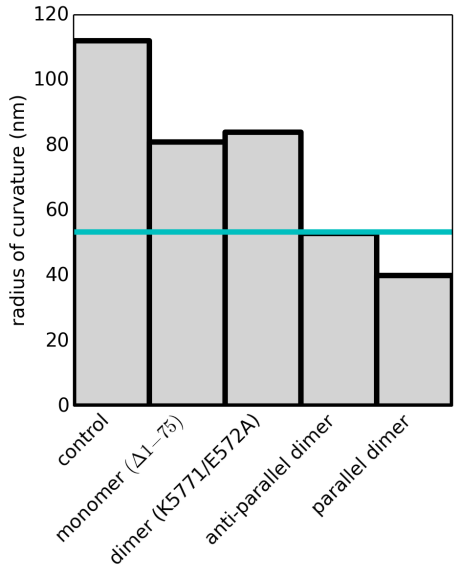

Figure 6.3: Curvature radii computed from membrane shapes under the action of Exo70 monomers, mutants, and dimers. The threshold radius of curvature ($53.4\,\mathrm{nm}^{-1}$) is noted in cyan. Simulations which produce a radius of curvature below this threshold are predicted to generate inward pointing tubules on LUVs.

6.3.2 Predicting phosphoinositide enrichment in endocytic initiation

There is an abundance of evidence that many proteins must assemble in a particular way in order to "trigger" a clathrin-mediated endocytosis (CME) event [88]. In this section, I will summarize our work towards integrating experiments and theory, and in particular, to mapping the free energy landscape of CME. Simulations of the protein Exo70 in the previous section (section 6.3.1) showed that the minimal oligomer (a dimer) could generate sufficient curvature. In the following study we have chosen to compute the concentration-curvature relation with slightly more resolution. We consider simulations containing either 1, 4, or 8 ENTH domains which are smaller and more mobile than Exo70.

Molecular models for ENTH domains. CME is characterized by the presence of a *clathrin coat*, but the clathrin triskelion itself is unable to induce this curvature on its own, and instead relies on several protein adapters [87, 262]. This provides a key hypothesis for this work: accessory proteins such as ENTH domains — known from *in vitro* experiments to sense and induce curvature on the membrane (see chapter 4) — must facilitate the curvature require to generate a mature vesicle *in vivo*.

To shed light on this mechanism we performed coarse-grained molecular dynamics simulations of either one or four ENTH domains on a lipid bilayer, along with a free bilayer which acts as a control. These simulations were published in Tourdot *et al* [333] and were extended significantly in order to generate the subject of chapter 4. These simulations were analyzed with curvature-fitting methods nearly identical to those used to characterize the bilayer bending under the action of Exo70 domains described in chapter 3. The results are summarized in figure 6.4, which illustrates the strength and extent of the fitted curvature fields.

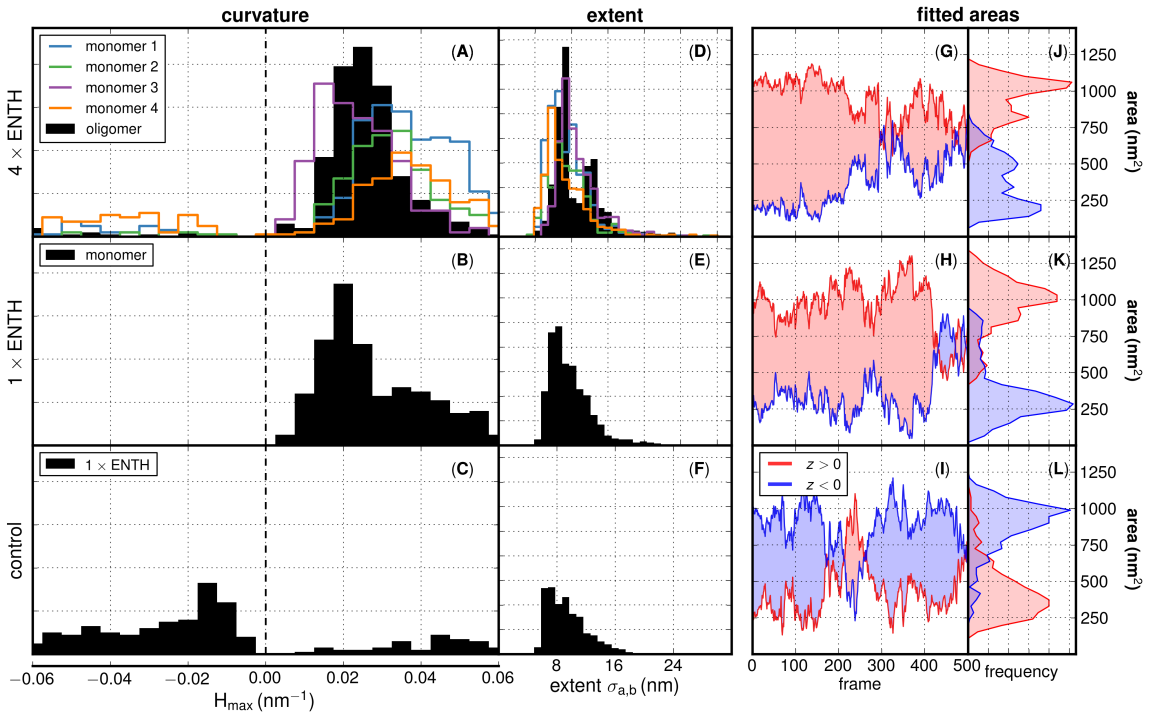

Figure 6.4: Distributions of maximum curvature and extent measured from framewise fits of the membrane surfaces in the neighborhood (15 nm) of ENTH domains. For the 4×ENTH system we also show these distributions for each protein in the simulation. We find stronger positive curvature when proteins are present, and a flatter distribution of curvature on the control (free) bilayer. The right panels show the membrane surface area near the proteins which are either above (red) or below (blue) the average membrane heights, indicating that ENTH domains are typically located on areas with positive deflection (and hence positive curvature).

We summarize these distributions in 6.1, which provides the raw data for our atom-to-field mapping.

system	$\langle H_{max} \rangle$ (nm^{-1})	$\langle \sigma_a + \sigma_b \rangle$ / 2 (nm)
(four) ENTH domains	0.024	10.8
lower left	0.031	9.4
lower right	0.030	9.8
upper left	0.024	11.1
upper right	0.018	9.6
(one) ENTH domain	0.028	9.8
control (no protein)	-0.013	9.8

Table 6.1: A summary of curvature strength and extent measurements according to the curvature fitting algorithm [333].

Identifying the threshold for morphology change. The atom-to-field mapping described above tells us how curvature induction depends on concentration. To predict the formation of vesicles on a planar membrane patch, my colleague Dr. Richard Tourdot computed the free energy landscapes of a membrane based on the Helfrich Hamiltonian [131] and simulated using dynamically-triangulated Monte Carlo methods. A full description of these simulations is provided by [333].

As with the simulations of spherical vesicles, these free energy calculations identified a critical curvature strength required to generate vesicular buds with a constricted neck, specifically $C_0 > 0.7\,a_0^{-1}$. Vesicles without necks form at $C_0 = 0.6\,a_0^{-1}$, which found at the maximum in the free energy change (a barrier height). Interestingly, the free energy landscapes depend somewhat on the size of the coat. Larger coats divide into smaller coats, each of which forms a bud. This gives an optimal coat radius of $r_0 = 4.55a_0$. If we rescale the mesoscale model to the size of a typical clathrin coated vesicle in endothelial cells [165], we set $a_0 = 28.9\,nm$. Therefore, mesoscale simulations mimicking clathrin-coated vesicles must provide enough curvature such that $C_0 > 0.7\,a_0^{-1} = 0.024\,nm^{-1}$. This is considerably less than the curvature observed in molecular simulations, $C_0 = 2H = 0.05\,nm^{-1}$, hence our model predicts that ENTH domains at moderate densities can generate curvature required to induce vesicles.

Closing the loop: predicting endocytosis events. Having constructed a model that matches both experimental measurements of clathrin-coated vesicles *and* free energy calculations of membrane shape change, we can now use this model to make predictions about the conditions necessary to trigger endocytosis.

Using a thermodynamic cycle, we decouple the energy functions for membrane-remodelling from protein-membrane association [3] and find that the free energy of membrane deformation is almost $400\,k_BT$ for a membrane with $\boldsymbol{\kappa} = 20\,k_BT$. This energy barrier must be overcome by attraction between the coat and the membrane. The enthalpy of binding of ENTH domains with its lipid binding partner PIP_2 (in this case represented by its headgroup, $Ins(1,4,5)P_3$) is $-14\,k_BT$ according to isothermal titration calorimetry [101]. However, this experiment does not consider the availability of PIP_2 in a bilayer setting.

Single-molecule kinetics studies show that ENTH dissociation from the bilayer has an off-rate of $k_{off} = 1s^{-1}$ [273]. We estimate the corresponding on-rate with a few assumptions.

1. The translational on-rate is:

$$k_{diff} = 6.023 \times 10^{23} D_{ENTH} \left(R_{ENTH} + R_{PIP_2}\right) = 6 \times 10^7 M^{-1} s^{-1},$$

   based on the radii of ENTH domains and PIP_2 (1 nm and 0.15 nm, respectively), and the finding that domains diffuse at roughly $D_{ENTH} = 100\,\mu m^2 s^{-1}$ [273].

2. ENTH domain binding to PIP_2 requires a coil-helix transition that costs $-3.5\,\text{kcal / mol} = -6k_BT$ [77], contributing a factor of e^{-6} to k_{diff}.

3. ENTH domain binding to PIP_2 is mediated by its embedded helix-0 domain, which contributes to the on-rate in proportion to the ratio of its rotational degrees of freedom in bound versus unbound states (approximately $(R_{PIP_2} / R_{ENTH}) / 4\pi = 0.01$).

These estimates yield $k_{on} = 0.01 k_{diff} e^{-6}$ and $k_d = k_{off} / k_{on} = 6 \times 10^{-4}$M which translates to a free energy of binding of $\Delta G = -7.3\,k_BT$ for ENTH binding to PIP_2 in a bilayer. Therefore, the minimum required number of ENTH domains to clear the energy barrier for vesiculation is $400k_BT / 7.3k_BT = 55$.

This quantity is lower than the number of available clathrin-associated protein-binding (CLAP) domains (115). However, assembling at least 55 epsins at the site of a budding vesicle with a radius of ~ 131 nm would require a local enrichment of PIP_2, which typically exists in the plasma membrane at low concentrations ($\sim 0.05\%$) [183]. This leads us to conclude that PIP_2 must be enriched at the sites of endocytosis. This claim agrees with the emerging consensus that PIP_2 is compartmentalized in particular domains which are essential to its regulatory functions [357].

In summation, our multiscale model reproduces a number of useful features of protein-induced membrane remodeling. The mesoscale model identifies an optimal coat size that agrees with measurements of clathrin-coated vesicles and tests the curvature found in molecular simulations against the free energy costs of creating vesicle buds on an otherwise flat bilayer. The combined multiscale model helps to guide our understanding of exactly which protein adapters are required to trigger endocytosis and make contact with the literature which describes the cell signaling cascades that control cell fate.

While this multiscale model is detailed and predictive, it is almost surely incomplete in many ways. For example, electrostatic interactions between epsin and the highly-charged PIP_2 molecules are only approximations in the coarse-grained model. As we have seen in chapter 5, these interactions depend on the specific chemistry of the PtdIns molecule and its associated ions. Other factors, like membrane lipid heterogeneity, pinning by the cytoskeleton and attachment of additional adapter pro-

teins may influence the model predictions, and help to shed more light on the events that initiate endocytosis.

6.4 Classifying multiscale models

This chapter has outlined two membrane-protein "case studies" in an effort to highlight typical multiscale modeling strategies. Besides using similar biophysics principles, there are several features which are common to both studies.

First, it is important to acknolwedge that these represent relatively *minimal* coupling between scales. Simpler mappings are easier to define and offer fewer opportunities for systematic errors. In our case, soft matter systems are particularly amenable to these kinds of multiscale couplings because their motion can be easily renormalized and because they have many soft degrees of freedom. Other systems, for example problems in hydrodynamics, may require more careful coupling strategies to bridge the gaps between small-and-large or fast-and-slow. In the final chapter (7) we will find that the many stiff degrees of freedom and chemical heterogeneities inherent to protein structures makes them particularly challenging multiscale modeling targets.

Second, because our models communicate with such "low bandwidth", many of their conclusions must be checked for internal consistency. Larger models might test their conclusions for parameter sensitivity; our models tend to use parameters that can be directly validated by experiments and follow statistical mechanics rules that can also be verified. This feature has upsides and downsides. It typically means that we cannot simply search a parameter space for the most parsimonious models, but the upshot is that our theories tend to connect to more experiments.

Lastly, the multiscale coupling described here leaves out large amounts of information in favor of using causal inference to comment on biology. We don't need to know the *exact* concentration of PIP_2 to conclude from our calculations that it must be enriched at the sites of endocytosis. Similarly, we don't need to quantify the *exact* oligomerization state of Exo70 to know that it must self-associate to create protrusions. These kinds of approximations are standard operating procedure in biology experiments.

In the final chapter (7), I will present a series of open questions and modeling opportunities, each of which will require multiscale modeling strategies in order to solve. As with the examples presented here, the models that constitute any multiscale strategy must be highly tuned to **trust** the information that they send and receive. This can only happen when the models are scientifically rigorous.

Chapter 7

Further perspectives on predictive simulations

In the preceding chapters, I hope to have convinced the reader that biophysical simulations provide unique contributions to our study of the molecular basis for cellular mechanics. In addition to the fundamental themes (see section 1.2) present in each chapter, a common, distinguishing feature of this work is that it is tightly woven into a large network of research groups, experiments, and scientific disciplines, each of which plays an important role in ensuring that the consequent research is accurate and comprehensive. In the spirit of these collaborations, I would like to review some of the future possibilities of these methods toward answering more questions in the biological sciences. In this chapter, I will outline onging modeling efforts in two main areas: membrane-protein interactions, and protein dynamics.

7.1 Tuneable nanocarriers for drug delivery

Functionalized nanocarriers (NCs) are a promising strategy for therapeutic and diagnostic applications [296]. In this section, I will briefly summarize contributions to Liu *et al* [195] which incorporate a molecular description of receptor bending and use these estimates in a continuum model for nanocarrier adhesion to endothelial cells. Ongoing study of receptor flexibility is relevant to more detailed models for nanocarrier binding which include the effects of hydrodynamics, summarized briefly below (see section 7.1.4).

7.1.1 Continuum modeling of nanocarrier binding

This study [195] produced free energy calculations of nanocarrier binding using a Metropolis Monte Carlo (MC) model sampled with the weighted histogram analysis method (WHAM). This calculation generates a potential of mean force (PMF)

for nanocarrier binding which depends on a number of biological parameters. These include the density, mobility, and flexure of antigens on the surface, the density of antibodies coated on the surface of the nanocarrier, the size and resistance of the glycocalyx layer, the flow rate over the endothelium, and also the size of the nanocarrier. Careful estimation of these parameters in the manner described in chapter 6 allows for a model which can predict these binding events without fitted parameters. However, some parameters are difficult to measure with experiments and benefit from additional validation. For this reason, we estimate the flexural rigidity of the 20 nm-long ICAM-1 receptor using coarse-grained molecular dynamics.

7.1.2 Molecular measurement of antigen flexure

The protein ICAM-1 is expressed in low levels in vascular endothelial cells and is upregulated during inflammation [326] and, along with other adhesion molecules linked to disease states [6], making it a good candidate for selective drug delivery.

We constructed atomistic models of ICAM-1 according to PDB structures `1IC1` and `1ZXQ` [368]. Since its dimerization state is unknown, we tested both the monomer and a dimer [241]. Coarse-grained molecular dynamics simulations of an ICAM-1 monomer and dimer were prepared using methods similar to those described in chapters 3 and 4. As per recommendations in Montocelli *et al* [228], an elastic network model with a force constant of $500\,\mathrm{kJ}\,/\,\mathrm{mol}\cdot\mathrm{nm}^2$ was applied to all coarse-grained beads within $0.5-0.9\,\mathrm{nm}$ in order to stabilize the higher-order structure of the protein. The simulation trajectories are visualized in figures 7.1 and 7.2, which indicate that the dimer undergoes more constrained motion during our $\sim 400\,\mathrm{ns}$ trajectories (100 ns in MARTINI time).

The MC model for nanocarrier binding treates antigen-antibody interactions according to a Bell model [32], however the bending of the receptor contributes significantly to the free energy of the system, and hence the multivalency of the nanocarrier bonds. To quantify this bending term, we model the receptor as an elastic rod. For small fluctuations, the flexure is given by Hooke's law such that the total bending energy can be written as:

$$U = \frac{1}{2}EI\int_0^L \left(\frac{d\theta}{ds} - \frac{d\theta_0}{ds}\right)^2 ds, \tag{7.1}$$

where s is the contour length, EI is the flexural rigidity, and the rod is parameterized by an angle θ(s) which fluctuates about the mean structure given by θ_0(s). According to the analysis of actin filament flexure described by Gittes *et al* [113], we can decompose these shapes into Fourier modes with corresponding coefficients $\mathrm{a_n}$ and $\mathrm{a_n^0}$ in order to rewrite the bending energy term as:

$$U = \frac{1}{2}EI\sum_{n=1}^{\infty}\left(\frac{n\pi}{L}\right)\left(a_n - a_n^0\right)^2 \tag{7.2}$$

Figure 7.1: Backbone representation of the ICAM-1 monomer from the beginning (red) to the end (blue) of a ~ 400 ns dynamics trajectory.

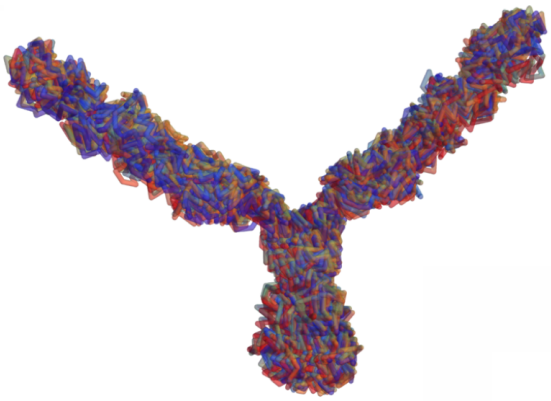

Figure 7.2: Backbone representation of the ICAM-1 dimer over the course of a ~ 400 ns trajectory.

The equipartition theorem tells us that each mode contributes $k_B T / 2$ to the energy, hence we can compute EI from the variance of the Fourier modes expressed in equation 7.2:

$$\mathrm{Var}[a_n] = \langle \left(a_n - a_n^0\right)^2 \rangle = \frac{k_B T}{EI} \left(\frac{L}{n\pi}\right)^2. \tag{7.3}$$

Equation 7.4 relates the flexure to the persistence length (L_p), the distance at which tangent-tangent angles become uncorrelated. Both flexure and persistence length describe the same property: the receptor's resistance to bending, given by:

$$L_p = \frac{EI}{k_B T}, \cos\left(\theta(s) - \theta_0(s)\right) = \exp\left[-\frac{s}{L_p}\right]. \tag{7.4}$$

We parameterize the ICAM-1 filaments by selecting residues at the interstices between its five domains and compute the flexure. According to the first Fourier mode, we compute a flexure of $800\,\mathrm{pNnm}^2$ for the monomer and $12000 - 24000\,\mathrm{pNnm}^2$ for the dimer. These correspond to persistence lengths that range from $200\,\mathrm{nm} - 3\,\mu\mathrm{m}$, meaning that the receptor is softer than actin and microtubules, but stiffer than collagen and P-selectin.

7.1.3 Continuum findings

Having validated the antigen flexure, among many other parameters, the continuum model produced PMFs which matched results from *in vitro* cell culture, endothelial targeting in mice, and atomic force microscopy experiments. The model predicts a threshold of 45% antibody coverage above which nanocarrier binding is stronger than that of a single-antibody. That this threshold was observed in mouse experiments suggests that the model is accurate and may be a useful reference for designing nanocarrier therapies.

7.1.4 Adding hydrodynamics

Work is currently underway to characterize the vibrational spectra of ICAM-1 binding in order to improve our understanding of nanocarrier binding under confinement. In previous work from the Radhakrishnan lab, Yu et al. have shown that a composite generalized Langevin equation (GLE) can be used to describe the motion of a nanocarrier across hydrodynamic regimes, from the bulk to a wall [375]. In order to add molecular specificity to this model, one must characterize the motion of the protein receptor along with the surrounding fluid by computing the velocity autocorrelation function (VACF) and the corresponding memory kernel. This method can yield important insights into the collective conformational motions in proteins [175]; these collective motions contribute one of many forces that appear in the model. This work employs constant number, constant volume, and constant energy (i.e. microcanonical ensemble, or NVE) simulations of ICAM-1 to compute the VACF in order

to refine the nanocarrier model for the particular case of ICAM-mediated binding.

7.2 Disentangling biomolecule conformational changes

A key future-oriented theme for much of the modeling work in this dissertation is the need to understand the organization and dynamics of the cytoskeleton because it plays a key role in cell function. In 2013, I produced models for Dr. Sira Sriswasdi, who was then a student working with David Spicher at the Wistar institute. The goal of this project was to generate a molecular model of minispectrin in order to provide detail to crosslinking experiments. Spectrin is large, flexible molecule that exists primarily as a heterotetramer of various isoforms, and forms a large part of the "membrane skeleton", a two-dimensional network on the inner leaflet of the plasma membrane [205].

Cross-linking experiments coupled with mass spectrometry can resolve the motion of spectrin, which is otherwise too large for crystallography or NMR experiments [320]. Our simulations sought to test the accuracy of homology models for the protein by comparing their motions to that observed by crosslinking. Specifically, we sought to predict whether particular residues would move close enough to crosslink in hour-long duration of a typical experiment.

Molecular simulations Molecular dynamics simulations were used to investigate the flexibility of mini-spectrin subunits in order to provide additional insight into the dynamics mini-spectrin flexibility. The $\alpha 2$ subunit was simulated for 100 ns using the `CHARMM27` all-atom molecular dynamics force field [207] in `GROMACS` (version 4.5.5) [342]. The starting structure, derived from homology models provided by the Spicher lab, was minimized, solvated with water, neutralized with counter-ions, and relaxed under constant volume (NVT) simulation for 100 ps before switching to a constant-pressure (NPT) for a production run lasting 100 ns. All simulation parameters were set as per the standard method [37] with a temperature of 300K and at least 20 Å of water between the protein and the periodic boundary condition.

This simulation procedure was also used to refine the structures of the wild-type mini-spectrin dimer, the `L207P` mutant dimer, and mini-spectrin tetramer. These simulations contained up to 1.8 million atoms. These simulations showed a root mean-squared deviation (RMSD) of up to 11 Å from the homology models.

Principal component analysis To connect these simulations to cross-linking experiments, principal component analysis (PCA) was used to estimate the energy barrier necessary to bring key residues within the 12 Å cross-linking distance in a simulation of a single subunit. In this method, the motion of backbone α-carbon atoms is measured by a covariance matrix of atomic positions. For a system of N particles, diagonalization of the covariance matrix gives a set of eigenvalues and eigenvectors which

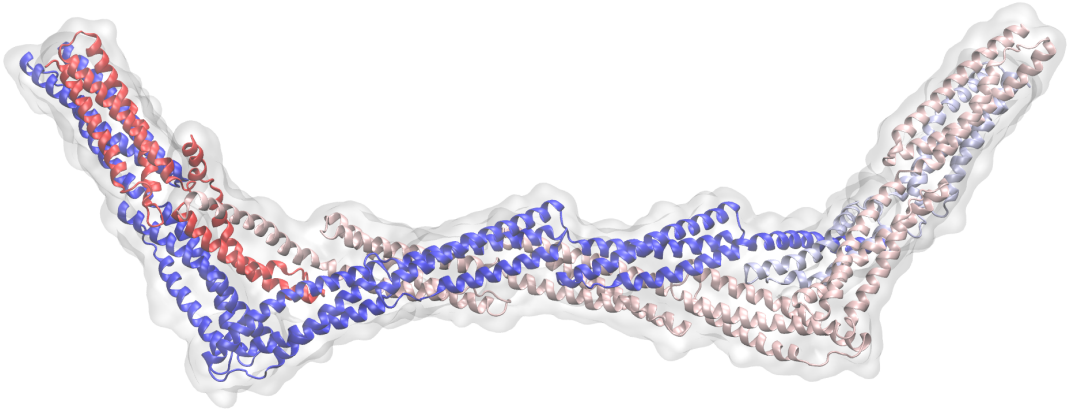

Figure 7.3: Structure of the minispectrin tetramer used in atomistic simulations described in this section. Spectrin is made from similar, heterogeneous repeating units of these filaments.

provide the amplitudes and directions of independent motions in the 3N-dimensional configuration space. The largest eigenvalues correspond to delocalized, low-frequency motions and are often sufficient to describe much of the protein fluctuations. Similar methods have been used to characterize protein structure and function in a variety of systems [7, 19, 35].

Invoking the quasiharmonic approximation to estimate the minimum energy required to bring key residues within the 12 Å cross-linking distance, we project a set of eigenvectors onto the average protein structure, each of which is scaled by a chosen amplitude δ_i requiring an energy equal to $U_i(\delta_i) = \frac{1}{2}k_i\delta_i^2$ where k_i is an effective spring constant for the ith mode. The spring constants are estimated using the equipartition theorem, according to which, each harmonic mode has a stiffness given by $k_i = k_B T / \lambda_i$ where λ_i is the eigenvalue corresponding to the ith mode and k_B is Boltzmann's constant. By projecting a linear combination of these scaled eigenvectors onto the protein's average structure, we calculate the corresponding inter-residue distance.

The total energy to bring the residues to this distance is then the sum of the corresponding $U_i(\delta_i)$. The Broyden-Fletcher-Goldfarb-Shanno (BFGS) minimization algorithm implemented in `NumPy` (version 1.6.2) [343] was used to search for the lowest-energy combination of eigenvectors subject to a harmonic restraint which set a preferred intra-residue gap distance.

Assuming a molecular vibrational frequency of $k_B T / h = 6 \times 10^{12}\,s^{-1}$ we estimate (using transition state theory) that it would be possible to cross an energy barrier of roughly $36\,k_B T$ during an hour-long cross-linking experiment. This suggests that the first link requires a trivial amount of time to crosslink ($< 0.5 k_B T$) while the second

and third links will reach 18.8 Å and 22.4 Å within one hour. The average inter-residue distances for these links were observed to be 15.3 Å, 28.2 Å and 31.3 Å, respectively during the 100 ns simulation.

While estimates of the time-to-crosslink were reasonable when analyzed one-at-a-time, they failed to produce a *consistent* description of the crosslinking. For example, comparing the projection of the simulation trajectories onto the long-mode fluctuations suggested that the structure was not well-equilibrated. Moreover, the PCA calculation showed high uncertainty across segments of the trajectory. This is likely a consequence of the rugged energy landscape that proteins explore.

There is much to be learned by extrapolating molecular simulations to timescales accessible by modern experiments. In the future it may be possible to more closely link atomistic simulations with the results of crosslinking experiment, particularly since zero-length crosslinks are now available [321].

7.3 Predicting endocytosis events

In chapter 4 I have outlined the method by which we may calculate the deformation field by proteins at various concentrations using a theory for curvature-undulation coupling. This method should be extended in several ways in order to improve the predictions and make better contacts with the continuum models.

Anisotropic curvature-inducing proteins First, measuring the curvature fields induced by epsins demonstrates that the method is sensitive. It is able to resolve differences between even one and four ENTH domains which insert helices of $\sim 30\,\mathrm{nm}^2$ at a low concentration (on a bilayer that is over $4000\,\mathrm{nm}^2$). For this reason we expect that it is appropriate for studying curvature induction by protein coats with many compositions that might be observed *in vivo*.

However, the ENTH-domain systems used in that study created curvature fields that were mostly isotropic. Given that many proteins must generate *anisotropic* deformation fields in order to create tubules and lamellipodia, it would be useful to extend this work to elongated curvature fields typical of e.g. BAR domains. We expect that studying these systems would require a richer set of hypothesized curvature fields. Proteins which soften the bilayer may also contribute a negative Gaussian curvature necessary to generate pores, and these curvatures may also be induced by more elaborate multipoles induced by heterogeneous protein coats.

Secondly, it may also be the case that proteins induce curvature more strongly or weakly depending on their arrangement. In a proof-of-principle calculation depicted in figure 7.4, we show the results of an unconstrained search through curvature strengths for a fixed extent of either 2 or 4 nm. The optimization routine is provided by `NumPy` (version 1.7) [343] and `SciPy` (version 0.17) [154] and included in the codes described in the appendix A.

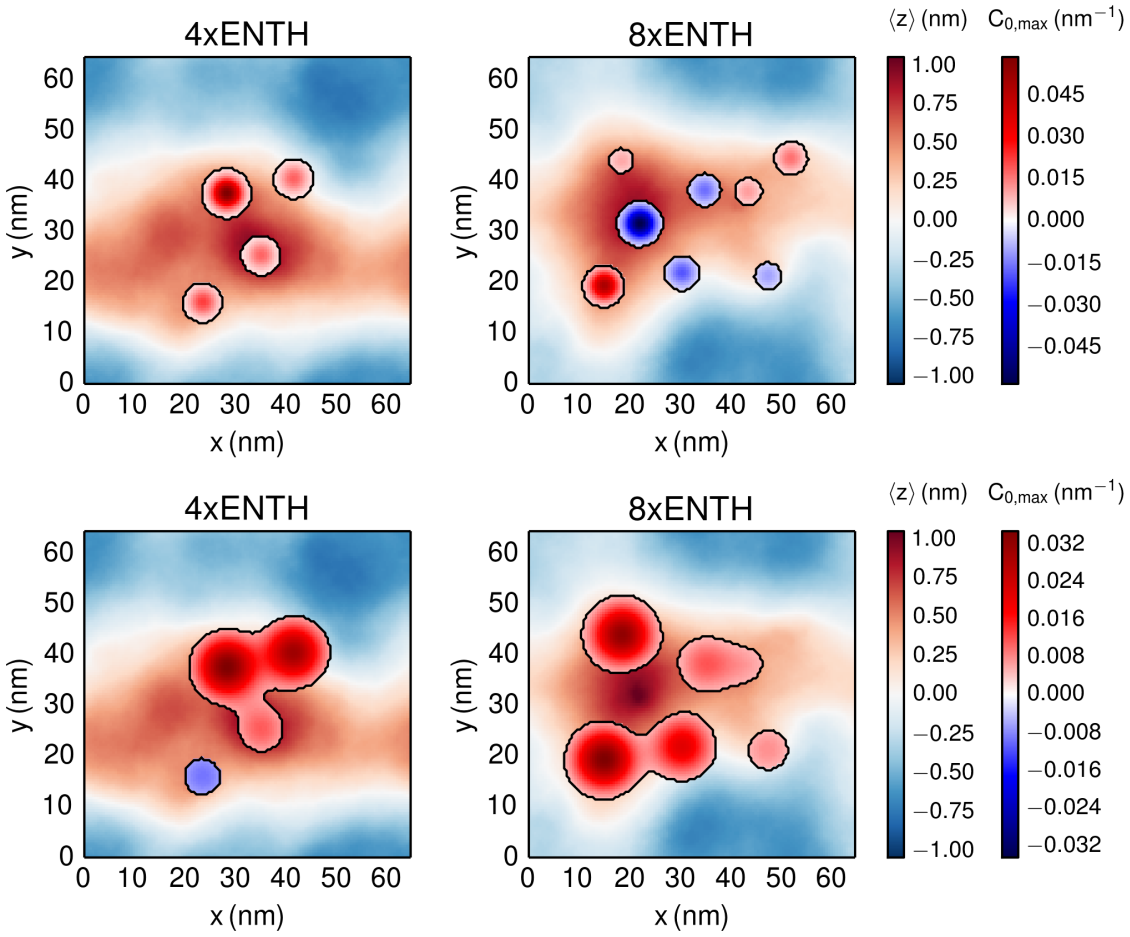

Figure 7.4: Curvature estimates generated by optimizing each dynamic protein field independently, for two different extents. The top panel shows fits for $\sigma = 2\,\mathrm{nm}$ while the bottom panel shows fits for a larger $\sigma = 4\,\mathrm{nm}$ extent. Both curvature and average membrane hights are indicated with colorbars. A similar optimization procedure can be used to add more precision to the results described in chapter 4.

This calculation suggests that the collection of proteins may induce different amounts of cuvature based on their relative positions and distances, as well as their proximity to the locus of the nascent membrane dimple. In chapter 4 we claimed that undulations allow proteins to communicate over distances large enough to explain their self-assembly at the sites of endocytosis. Computing the distance-dependent curvature strengths may make it possible to learn more about how proteins self-assemble *in vivo*. This problem is relevant to many membrane-facilitated recruitment events, which often requires local enrichment of species which are otherwise found in low concentrations.

System-size scaling

Deformation fields measured from molecular simulations can only be used to predict endocytic events with the help of the mesoscale simulations described in chapter 6. In that chapter and other similar studies [335] we employ a one-to-several modeling strategy. It would be useful to match the molecular and mesoscale calculations more explicitly by matching their fluctuations. A proof-of-principle is provided in figure 7.5.

In the above test case, fluctuations from several small, 22×22 mesoscale simulations (codes courtesy of Dr. Richard Tourdot [335]) were matched to the height fluctuations in molecular simulations containing four ENTH domains. To match the size and bending rigidity of the molecular simulations, we rescaled the mesoscale simulations according to a length parameter $\mathrm{a}_0 = 2.26\,\mathrm{nm}^{-1}$ and inferred a curvature field of $0.05\,\mathrm{a}_0 = 0.022\,\mathrm{nm}^{-1}$, which generally agrees with the findings in chapter 4. We expect that this type of fluctuation matching can also answer questions about the finite size of our simulations.

Simulations which use periodic boundary conditions preclude access to waves with a length that is longer than the simulated box. This means that smaller simulations may frustrate the curvature sorting or induction behavior of adhered proteins. Testing the size-dependent fluctuations is tractable by carefully matching fluctuations to larger bilayers simulated the continuum methods employed by Tourdot *et al* [334,335].

Inhomogeneous bending rigidity Much of exposition in chapter 4 relies on the reasonable assumption that bilayers have a relatively homogeneous bending rigidity ($\boldsymbol{\kappa}$). While this is largely true for low and moderate protein concentrations, we hypothesize that the insertion of larger numbers of proteins might stiffen the bilayer. The effects of higher-concentration protein inclusions may be evaluated by proposing an inhomogeneous $\boldsymbol{\kappa}$-field. This problem was explored in [4] for one-dimensional systems and could be applied to our systems. Such a field would cause "mode mixing" whereby the Fourier modes are not exactly the eigenmodes of the system. This means that independent undulation modes may be composed of fluctuations across many wavevectors. While these modes would be non-trivial to compute and interpret, studying the ways

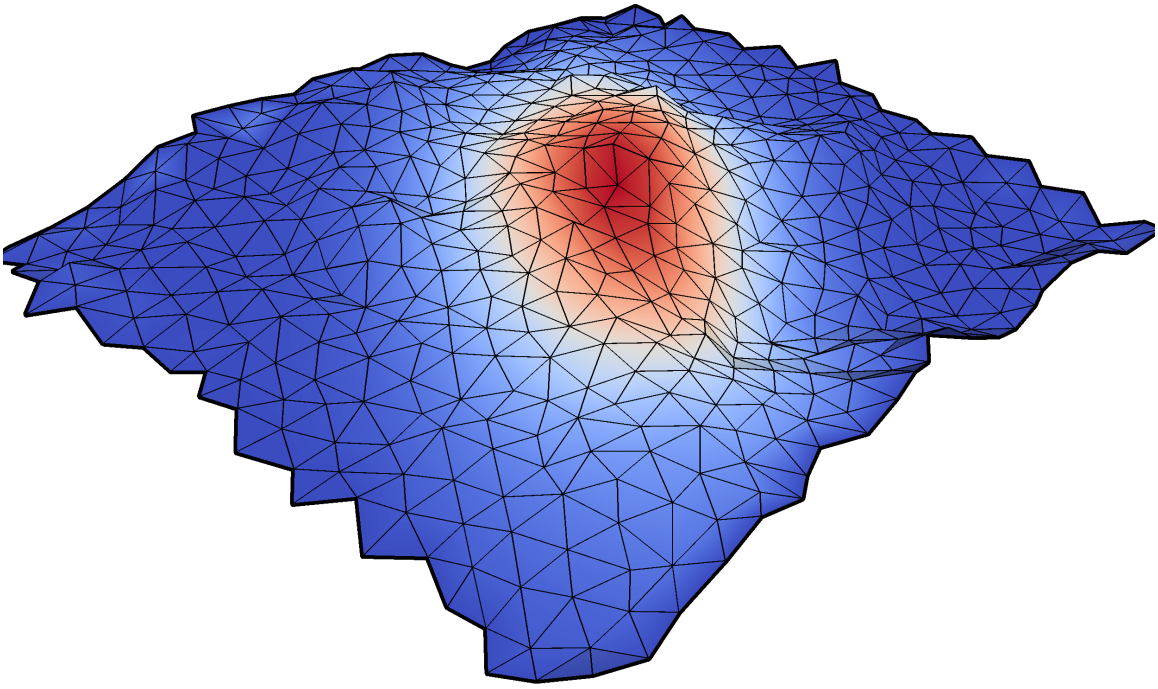

Figure 7.5: A snapshot of the dynamically triangulated Monte Carlo (DTMC) membrane surface designed for comparison to ENTH domain simulations reported in chapter 6. In this proof-of-principle, we search through parameters — specifically the strength and extent of a two-dimensional Gaussian in curvature — in order to identify the mesoscale simulation with the fluctuations that most closely match the 4×ENTH simulations. In this case we estimate a curvature strength of $0.022\,\mathrm{nm}^{-1}$ in general agreement with other estimates.

in which the undulations couple when they are organized by proteins is important to understanding biological membranes which may be laden with a dense coat of biomolecules.

7.4 Conclusions

In this chapter, I have outlined a number of useful modes of inquiry that might extend the methods described in this dissertation to address open challenges in biology and continue the important work of marrying theory to simulation and experiment.

Appendix A

Scalable simulation protocols

Biomolecular modeling efforts require both a robust model and a practical method for generating, organizing, and matching the model to experiments. This chapter will describe the practical matter of managing the data created in the course of making and testing our models. This is not strictly a scientfic question — it's also an engineering challenge.

Collaborative development. The work in this chapter was highly collaborative, and required close cooperation with two colleagues. David Slochower and I jointly developed all of the codes used to analyze lipid bilyers in chapter 5. These codes formed the basis for the membrane portions of the "calculator" codes described below (see section A.5). David and I also cooperatively designed the bilayer construction algorithms used to generate the atomistic bilayer simulations in chapter 5.

The "factory" codes described below (see section A.6) were jointly developed by Joe Jordan and me with the explicit goal of connecting many disparate codes into a single simulation/calculation pipeline. The codes produced in our collaboration were useful training tools for Bachelor's and Master's degree students who collaborated with our lab, and assisted in generating data sets for their research projects. Both David and Joe are contributors to shared codes which I have posted on github at `http://github.com/bradleyrp`.

A.1 Scaling and reproducibility

The scientific inquiry described in the body of this dissertation would be *feasible* — but hardly economical — without writing standardized codes. In addition to conserving time and effort, we have developed codes for simulating and analyzing molecular dynamics simulations with twin purposes:

1. Design simulations protocols that are *scalable* so that they can be produced at

little to no cost beyond the requisite computer time.

2. Create *reproducible* simulations so that other researchers, students, and trainees can use the codes without an extensive computer programming background.

I have helped to develop these codes in order to make this work more accessible to my peers, but there are also two knock-on effects from writing computer programs in this way. First, highly modular and reproducible codes make it easier to perform more elaborate analyses. While the utility of a calculation does not necessarily increase with its complexity, the ability to ask nuanced questions about our modeled systems makes it easier to answer more difficult scientific questions more confidently.

A.2 Open-source codes

The work in this thesis along with the codes described in this appendix depend on a wide-range of open-source codes.

A.2.1 `GROMACS`.

The ***Groningen machine*** for molecular dynamics [342] is one of a handful of widely-adopted molecular dynamics packages which is designed to integrate Newton's equations of motion efficiently on millions of atoms. Despite being designed for biomolecular systems, a large investment in developing the `GROMACS` codes have led to its widespread use in non-biological problems as well. The `GROMACS` website (`http://www.gromacs.org/`) notes that the collected codebase required roughly 500 person-years of labor to produce.

For this reason, it's difficult to overstate the importance of the community effort required to develop such a useful, well-documented, and fully-featured code. The same is true for the small ecosystem of analysis libraries described in the next section. All of the simulation data in this thesis were generated with `GROMACS`, and it serves as a model for all of our efforts in developing scalable codes, not least of all because it is simple and easy to use.

A.2.2 Scientific computing.

The analyses in this dissertation similarly relied on an elaborate library of computing resources, namely the `Python` programming language [347] and the widely-adopted `NumPy` [154] and `SciPy` [343] mathematical libraries. These libraries in particular are capable of running fast analyses, specifically those that require pairwise distance searching, generating triangulated surfaces, and numerical optimization methods. We used the `MDAnalysis` [224] library for parsing simulation trajectories. Lastly, the

calculations described in chapter 4 used `SQLalchemy` [29] to manage the large database of hypotheses.

A.2.3 Visualization.

Biomolecules such as proteins, lipids, and membranes all have intricate geometries. We have used `VMD` [142] in particular to render proteins and membranes, while `MayaVi` [258] and `ParaView` [5] were useful for visualizing triangulated meshes in some cases. Plots, and figures were typically rendered and laid out by using `Mathematica` [364] or `MatPlotLib` [143].

A.3 Cooperative codes

The data generated for this thesis were managed using three codes:

1. A simulator code manages each GROMACS simulation.
2. A calculator code analyzes, interprets, and visualizes the data.
3. A "factory" code creates a pipeline between simulations and calculations

The simulator and calculator codes can be used independently. The simulator only requires a computer running linux, `Python` version `2.7` or higher, and a stock copy of `GROMACS`. The calculator requires scientific libraries described in section A.2.2. The third component, the "factory" serves both the simulator and calculator with a simple web interface, and was created to make it easier for new students to generate larger data sets. It additionally requires the `Django` (the "web framework for perfectionists with deadlines"). The factory has been used to generate dozens of simulations for both graduate students and undergraduate trainees, and has been running continuously for almost a year.

In contrast to many software packages, these codes are meant to be downloaded once per simulation instead of "installed" on a new machine. This has two advantages: extremely rapid deployment on machines with scientific software already installed and parallel development of new simulation procedures. The calculator codes are designed to run on a workstation, while the simulator codes are designed to start simulations on a small workstation and send them to a larger computing platform to perform "production" runs.

A.4 Automatic GROMACS

Automatic procotols for `GROMACS` (which we have named `AUTOMACS`) is a small set of Python [347] codes which which we have uploaded to `github` at

`http://github.com/bradleyrp/automacs`. The codes are packaged with a few protocols from running common simulations. These include:

1. Atomistic proteins in water.
2. Coarse-grained proteins in water using the `MARTINI` [212] force field.
3. Coarse-grained lipid bilayers.
4. Coarse-grained proteins adhered to lipid bilayers.
5. Homology modeling using `MODELLER` [353].

The codes are also designed to be easy to modify and connect to the factory described in A.6. All of the procedures used in other parts of this thesis are available as code modules (we call them "bundles" to distinguish them from typical libraries) which are easy to distribute with separate `git` repositories, and may contain more than one distinct simulation protocol. These codes make it easy to extend the automacs functions to more complicated simulations, and were essential for developing the protein-bilayer simulations described in chapters 3 and 4. An example protocol, designed to produce a coarse-grained lipid bilayer with an arbitrary composition, is reproduced below.

```python
#!/usr/bin/python

settings = """
system name:        CGMD BILAYER
lipid structures:   inputs/cgmd-inputs/
step:               cgmd-bilayer
requires:           cgmd,bilayer
shape:              flat
height:             6
binsize:            1.0
monolayer offset:   1.5
monolayer top:      400
monolayer bottom:   None
composition top:    {"DOPC":0.8,"DOPS":0.2}
composition bottom: None
aspect:             1.0
solvent thickness:  20
lipid ready:        lipid-ready.gro
force field:        martini
cation:             NA+
anion:              CL-
ionic strength:     0.150
sol:                W
ff includes:        ["martini-v2.2","martini-v2.0-lipids",
                     "martini-v2.2-aminoacids","martini-v2.0-ions"]
files:              ["cgmd-inputs/martini-water.gro"]
sources:            ["martini.ff"]
equilibration:      npt-bilayer
"""

from amx import *
init(settings)
try:
    #---development options save your progress
    if not wordspace["under_development"]:
        start(wordspace["step"])
        write_mdp()
        #---create the bilayer
        build_bilayer(name="vacuum-bilayer")
        filecopy(wordspace["step"]+"vacuum-bilayer.gro",wordspace["step"]+"vacuum.gro")
        write_top("vacuum.top")
        minimize("vacuum")
        remove_jump(structure="vacuum-minimized",tpr="em-vacuum-steep",gro="vacuum-nojump")
        #---pack the lipids without flips
        vacuum_pack(structure="vacuum-nojump",name="vacuum-pack",gro="vacuum-packed")
    #---add water
    solvate_bilayer("vacuum-packed")
    write_top("solvate.top")
    minimize("solvate")
    remove_jump(structure="solvate-minimized",tpr="em-solvate-steep",gro="solvate-nojump")
    #---add ions
    counterions("solvate-nojump","solvate",resname="W")
    counterion_renamer("counterions")
    write_top("counterions.top")
    #---final minimization
    minimize("counterions")
    remove_jump(structure="counterions-minimized",
        tpr="em-counterions-steep",gro="counterions-nojump")
    #---center the bilayer in the box
    bilayer_middle(structure="counterions-nojump",gro="system")
    write_mdp()
    #---create groups
    bilayer_sorter(structure="system",ndx="system-groups")
    write_top("system.top")
    #---equilibration
    equilibrate(groups="system-groups")
    #---write steps for continuation
    write_continue_script()
#---if interrupted save the state
except KeyboardInterrupt as e: exception_handler(e,wordspace,all=True)
except Exception as e: exception_handler(e,wordspace,all=True)
```

Automacs also generates fully-documented simulations by using a standard logging

procedure. This means that each simulation comes with a list of GROMACS and terminal commands which can exactly reproduce the simulation from stratch, even without using the original code. Both the simulator and calculator codes use the same machine configuration files, which means that it is easy to deploy systems on supercomputing platforms (e.g. `XSEDE`) as long as you have an allocation on these machines. The benchmarks shown in figure A.1 were generated automatically on `XSEDE` resources using a built-in benchmarking procedure.

Lastly, the automacs code includes upload, download, and token features that make it easy to send new simulations to computation clusters, and later retrive them. Automacs simulations are indexed so that they can be easily incorporated into the calculator, described in the next section.

A.5 Calculations

Making reproducible simulations is useful for extending this work to new biomolecules, while developing robust analysis codes is essential even when studying a single biological system. For that reason, the calculator codes are designed to streamline the requisite bookkeeping and distill a particular calculation down into small, readable components that are independent of the storage operations. The codes can be downloaded from github at `http://github.com/bradleyrp`.

Under our framework, the calculation codes must be written so that they take a standard input, typically a structure and trajectory from a simulation, and return a dictionary of parameters or `NumPy` arrays which are saved in a transportable, binary format using the `Hdf5` library. By way of example, a simple protein root-mean squared deviation (RMSD) calculation is quoted below.

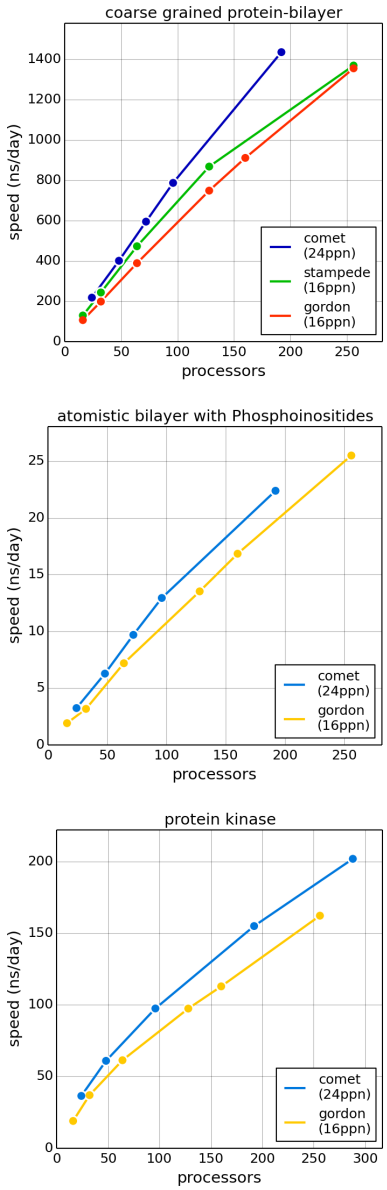

Figure A.1: (Top) Performance benchmarks for a simulation containing eight ENTH domains including 915,691 beads (four heavy atoms each) under the MARTINI coarse-grained force field using `GROMACS` 5. The simulation consists of a square patch roughly 65×65 nm in extent. (Middle) Performance for an atomistic lipid bilayer containing 800 lipids for the Phosphoinositide project. These simulations use the `CHARMM36` [166] force field, include 361,894 atoms, and use `TIP3PS` or `CHARMM` "special water" hence the slightly lower performance compared to the smaller protein kinase simulation below. (Bottom) Performance for a simulation of a protein kinase (BRAF, active form) containing 73,507 atoms using `CHARMM27` [50] in `GROMACS` 5 [342].

```python
#!/usr/bin/python

import time
import numpy as np
import MDAnalysis
from base.tools import status

def protein_rmsd(grofile,trajfile,**kwargs):

    """
    Compute the RMSD of a protein.
    """

    #---unpack
    sn = kwargs["sn"]
    work = kwargs["workspace"]

    #---prepare universe
    slice_name = kwargs["calc"]["slice_name"]
    group = kwargs["calc"]["group"]
    grofile,trajfile = [work.slice(sn)[slice_name][group][i] for i in ["gro","xtc"]]
    uni = MDAnalysis.Universe(work.postdir+grofile,work.postdir+trajfile)
    nframes = len(uni.trajectory)
    protein = uni.select_atoms("protein and name CA")

    #---reference frame
    uni.trajectory[0]
    r0 = protein.coordinates()
    r0 -= np.mean(r0,axis=0)

    #---collect coordinates
    nframes = len(uni.trajectory)
    coords = []
    for fr in range(0,nframes):
        uni.trajectory[fr]
        r1 = protein.coordinates()
        coords.append(r1)

    #---simple RMSD code
    rmsds = []
    for fr in range(nframes):
        status("RMSD",i=fr,looplen=nframes)
        r1 = coords[fr]
        r1 -= mean(r1,axis=0)
        U,s,Vt = np.linalg.svd(np.dot(r0.T,r1))
        signer = np.identity(3)
        signer[2,2] = np.sign(np.linalg.det(np.dot(Vt.T,U)))
        RM = np.dot(dot(U,signer),Vt)
        rmsds.append(np.sqrt(np.mean(np.sum((r0.T-np.dot(RM,r1.T))**2,axis=0))))

    #---save an hdf5 binary
    attrs,result = {},{}
    result["rmsds"] = np.array(rmsds)
    return result,attrs
```

While our framework requires that each calculation is written in a particular format, by using standardized variables, intuitive lookup tables, and a `YAML` dictionary for defining parameter sweeps, the user can focus on writing (and debugging) the calculation instead of searching for files.

Replicating results. The calculations described in this thesis are available in separate modules that can be easily copied between resources or distributed on `github`. This is an essential feature of our framework: calculations should be transparent, easy to read, and more importantly, easy to apply to new (or even old) data. The calculator codes are designed with this goal in mind: *all calculations should be almost effortlessly*

replicable.

A.6 Generating data with the "factory"

The factory joins the simulator and calculator codes into a single web-based interface which the user can deploy privately on a workstation or laptop. It makes it easy for relative novices to generate simulation data, and it also makes it easy to organize the production and analysis of new simulations. The codes can be cloned from github at `http://github.com/bradleyrp`. The user starts with a basic project displayed in figure A.2.

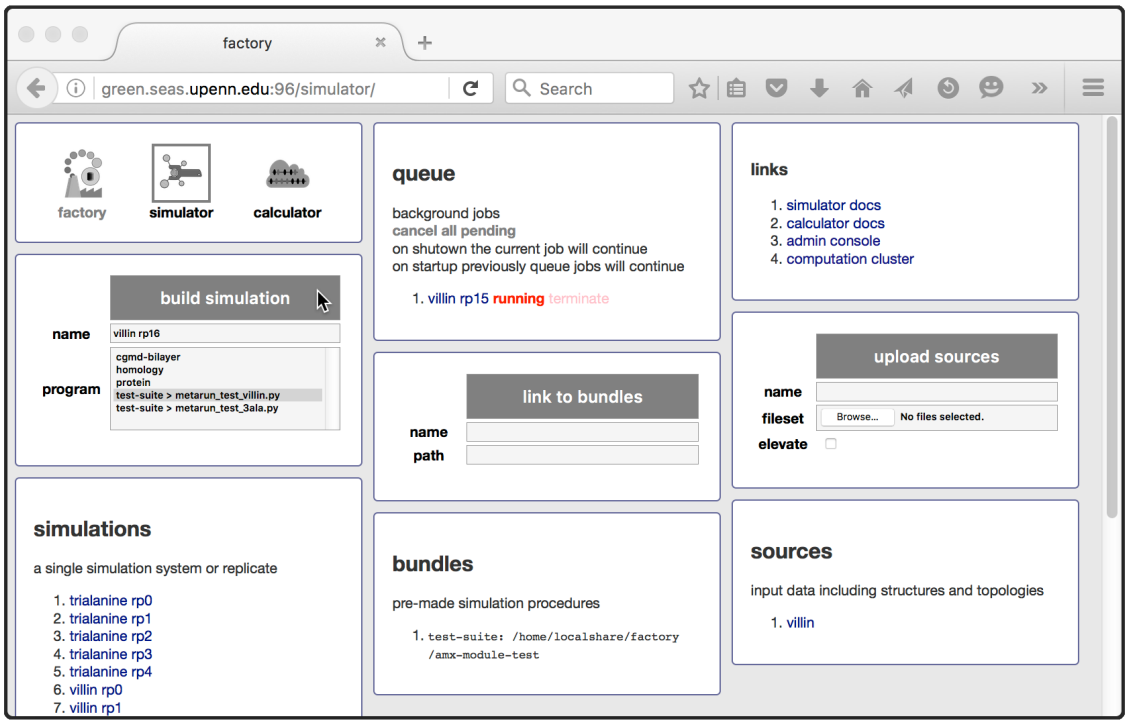

Figure A.2: Start a simulation from a "bundle" — a pre-packaged protocol. In this demonstration, the user is creating a new simulation of the villin headpiece.

The factory detects all of the parameters — even if the user has customized their own simulation protocols — and serves them in a form that can be easily tweaked to e.g. induce a mutation in a protein, make a larger bilayer, or use a different lipid composition. After running the simulation, the codes can produce some basic visualizations automatically. See figure A.3 for an example of a short protein simulation of the villin headpiece.

The factory codes are designed to *automatically* detect the settings for new "bundles", which means that any protocol written in automacs can be used in the factory.

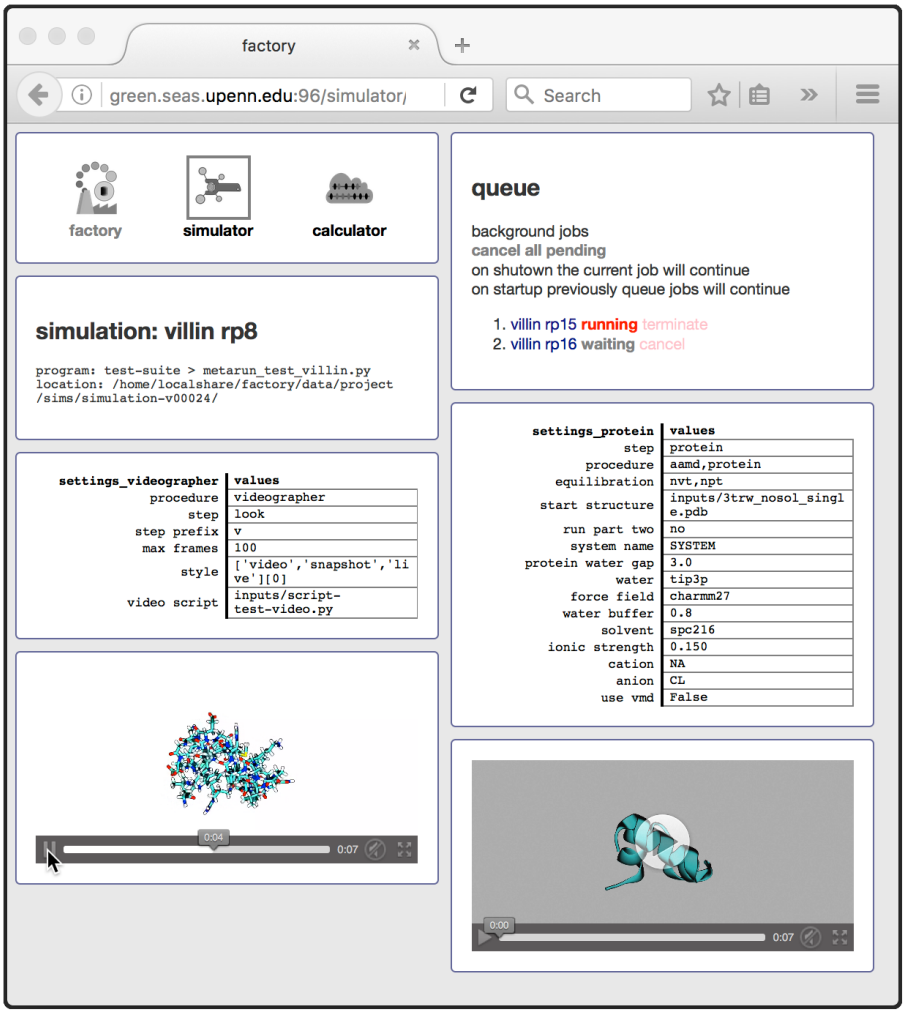

Figure A.3: Some simulation protocols can automatically create images of the data using VMD [142].

Batch calculations. Calculations can be deployed *en masse* using the factory described in the next section. The web interface for the calculations is shown in figure A.4, but it can be deployed entirely from the terminal as well.

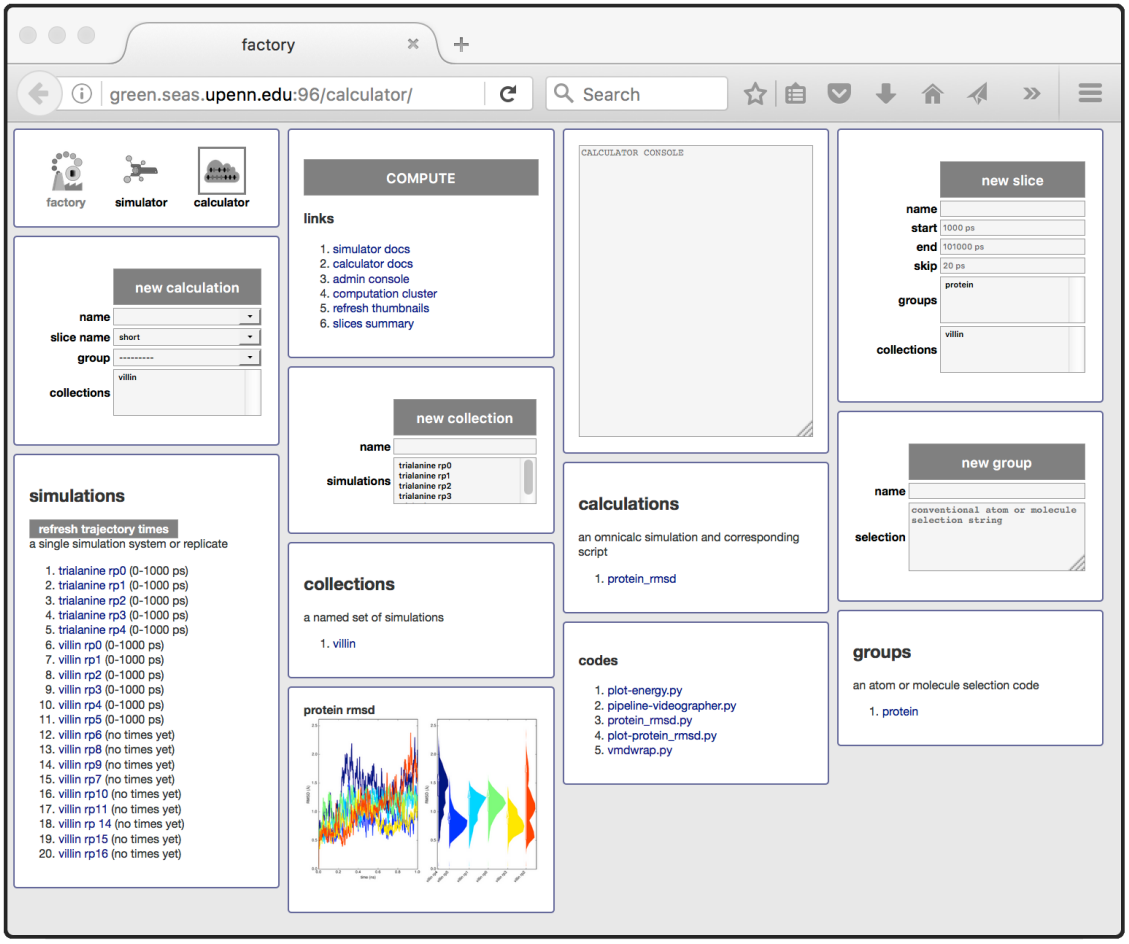

Figure A.4: The calculator web page can be used to apply calculations to many new simulations at once. For example, you can see the results from the protein RMSD calculation in the bottom left.

A.7 Ongoing Development

The three codes described in this chapter are under active development. To date they have been used to train several new students from high school interns to post-docs in order to answer increasingly technical modeling questions. Codes are shared on github and it is our sincere hope that even if these codes are not widely adopted, that they make it easy to extend the work described in this dissertation to more biomolecular systems.

Bibliography

[1] **Adcock, S.**, **J. McCammon** (2006). Molecular dynamics: survey of methods for simulating the activity of proteins. *Chem. Rev.* **106**(5), 1589.

[2] **Agrawal, N. J.**, **J. Nukpezah**, **R. Radhakrishnan** (2010). Minimal Mesoscale Model for Protein-Mediated Vesiculation in Clathrin-Dependent Endocytosis. *PLoS Comput. Biol.* **6**(9), e1000926.

[3] **Agrawal, N. J.**, **R. Radhakrishnan** (2009). Calculation of free energies in fluid membranes subject to heterogeneous curvature fields. *Phys. Rev. E* **80**(1), 011925.

[4] **Agrawal, N. J.**, **R. Radhakrishnan** (2009). Calculation of free energies in fluid membranes subject to heterogeneous curvature fields. *Phys. Rev. E* **80**(1), 011925.

[5] **Ahrens, J.**, **B. Geveci**, **C. Law**, *Visualization Handbook*, chapter ParaView: An End-User Tool for Large Data Visualization (Elsevier, 2005).

[6] **Ala, A.**, **A. Dhillon**, **H. Hodgson** (2002). Role of cell adhesion molecules in leukocyte recruitment in the liver and gut. *Int. J. Exp. Pathol.* **84**(1), 1.

[7] **Amadei, A.**, **A. B. M. Linssen**, **H. J. C. Berendsen** (1993). Essential dynamics of proteins. *Proteins* **17**(4), 412.

[8] **Amaya, J.**, **D. Rana**, **V. Hornof** (2002). Dynamic Interfacial Tension Behavior of Water/Oil Systems Containing In situ-Formed Surfactants. *J. Solution Chem.* **31**, 139.

[9] **Arkhipov, A.**, **Y. Yin**, **K. Schulten** (2008). Four-Scale Description of Membrane Sculpting by BAR Domains. *Biophys. J.* **95**(6), 2806.

[10] **Arkhipov, A.**, **Y. Yin**, **K. Schulten** (2009). Membrane-Bending Mechanism of Amphiphysin N-BAR Domains. *Biophys. J.* **97**(10), 2727.

[11] **Ash, W. L.**, **M. R. Zlomislic**, **E. O. Oloo**, **D. P. Tieleman** (2004). Computer simulations of membrane proteins. *Biochim. Biophys. Acta, Biomembr.* **1666**(1–2), 158.

[12] **Ayton, G. S.**, **P. D. Blood**, **G. A. Voth** (2007). Membrane Remodeling from N-BAR Domain Interactions: Insights from Multi-Scale Simulation. *Biophys. J.* **92**(10), 3595.

[13] **Ayton, G. S.**, **E. Lyman**, **G. A. Voth** (2010). Hierarchical coarse-graining strategy for protein-membrane systems to access mesoscopic scales. *Faraday Discuss.* **144**, 347.

[14] **Ayton, G. S.**, **J. L. McWhirter**, **P. McMurtry**, **G. A. Voth** (2005). Coupling Field Theory with Continuum Mechanics: A Simulation of Domain Formation in Giant Unilamellar Vesicles. *Biophys. J.* **88**(6), 3855.

[15] **Ayton, G. S.**, **J. L. McWhirter**, **G. A. Voth** (2006). A second generation mesoscopic lipid bilayer model: Connections to field-theory descriptions of membranes and nonlocal hydrodynamics. *J. Chem. Phys.* **124**, 064906.

[16] **Ayton, G. S.**, **G. A. Voth** (2009). Hybrid Coarse-Graining Approach for Lipid Bilayers at Large Length and Time Scales. *J. Phys. Chem. B* **113**(13), 4413.

[17] **Ayton, G. S.**, **G. A. Voth** (2009). Systematic multiscale simulation of membrane protein systems. *Curr. Opin. Struct. Biol.* **19**(2), 138.

[18] **Ayton, G. S.**, **G. A. Voth** (2010). Multiscale Computer Simulation of the Immature HIV-1 Virion. *Biophys. J.* **99**(9), 2757.

[19] **Bahar, I.**, **T. R. Lezon**, **L.-W. Yang**, **E. Eyal** (2010). Global Dynamics of Proteins: Bridging Between Structure and Function. *Annu. Rev. Biophys.* **39**(1), 23.

[20] **Bahar, I.**, **A. Rader** (2005). Coarse-grained normal mode analysis in structural biology. *Curr. Opin. Struct. Biol.* **15**(5), 586.

[21] **Balabin, I. A.**, **W. Yang**, **D. N. Beratan** (2009). Coarse-grained modeling of allosteric regulation in protein receptors. *Proc. Natl. Acad. Sci. U. S. A.* **106**(34), 14253.

[22] **Baoukina, S.**, **L. Monticelli**, **S. J. Marrink**, **D. P. Tieleman** (2007). Pressure-Area Isotherm of a Lipid Monolayer from Molecular Dynamics Simulations. *Langmuir* **23**(25), 12617.

[23] **Baoukina, S.**, **L. Monticelli**, **H. J. Risselada**, **S. J. Marrink**, **D. P. Tieleman** (2008). The molecular mechanism of lipid monolayer collapse. *Proc. Natl. Acad. Sci. U.S.A.* **105**(31), 10803.

[24] **Baoukina, S.**, **D. P. Tieleman** (2010). Direct Simulation of Protein-Mediated Vesicle Fusion: Lung Surfactant Protein B. *Biophys. J.* **99**(7), 2134.

[25] **Baron, R.**, **D. Trzesniak**, **A. H. de Vries**, **A. Elsener**, **S. J. Marrink**, **W. F. van Gunsteren** (2007). Comparison of Thermodynamic Properties of Coarse-Grained and Atomic-Level Simulation Models. *ChemPhysChem* **8**(3), 452.

[26] **Basdevant, N.**, **D. Borgis**, **T. Ha-Duong** (2007). A Coarse-Grained Protein-Protein Potential Derived from an All-Atom Force Field. *J. Phys. Chem. B* **111**(31), 9390.

[27] **Baumgart, T.**, **B. R. Capraro**, **C. Zhu**, **S. L. Das** (2011). Thermodynamics and mechanics of membrane curvature generation and sensing by proteins and lipids. *Annu. Rev. Phys. Chem.* **62**, 483.

[28] **Baumgart, T.**, **S. T. Hess**, **W. W. Webb** (2003). Imaging coexisting fluid domains in biomembrane models coupling curvature and line tension. *Nature* **425**(6960), 821.

[29] **Bayer, M.** (2016), `SQLAlchemy`. Accessed at `http://www.sqlalchemy.org/`.

[30] **Beauchamp, K. A.**, **Y.-S. Lin**, **R. Das**, **V. S. Pande** (2012). Are Protein Force Fields Getting Better? A Systematic Benchmark on 524 Diverse NMR Measurements. *J. Chem. Theory Comput.* **8**(4), 1409.

[31] **Beers, K. J.**, *Numerical methods for chemical engineering: applications in Matlab* (Cambridge University Press Cambridge, UK, 2007).

[32] **Bell, G. I.**, **M. Dembo**, **P. Bongrand** (1984). Cell-Adhesion - Competition between Nonspecific Repulsion and Specific Bonding. *Biophys.J.* **45**(6), 1051.

[33] **Ben-Naim, A.**, *Solvation Thermodynamics* (Plenum Press, NY, 1987).

[34] **Berendsen, H. J. C.**, **J. P. M. Postma**, **W. F. van Gunsteren**, **A. DiNola**, **J. R. Haak** (1984). Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* **81**(8), 3684.

[35] **Berendsen, H. J.**, **S. Hayward** (2000). Collective protein dynamics in relation to function. *Curr. Opin. Struct. Biol.* **10**(2), 165.

[36] **Bezkorovaynaya, O.**, **A. Lukyanov**, **K. Kremer**, **C. Peter** (2012). Multiscale simulation of small peptides: Consistent conformational sampling in atomistic and coarse-grained models. *J. Comput. Chem.* **33**(9), 937.

[37] **Bjelkmar, P.**, **P. Larsson**, **M. A. Cuendet**, **B. Hess**, **E. Lindahl** (2010). Implementation of the CHARMM Force Field in GROMACS: Analysis of Protein Stability Effects from Correction Maps, Virtual Interaction Sites, and Water Models. *J. Chem. Theory Comput.* **6**(2), 459.

[38] **Bond, P. J.**, **M. S. P. Sansom** (2006). Insertion and Assembly of Membrane Proteins via Simulation. *J. Am. Chem. Soc.* **128**(8), 2697.

[39] **Bond, P. J.**, **M. S. P. Sansom** (2007). Bilayer deformation by the Kv channel voltage sensor domain revealed by self-assembly simulations. *Proc. Natl. Acad. Sci. U.S.A.* **104**(8), 2631.

[40] **Bond, P. J.**, **C. L. Wee**, **M. S. P. Sansom** (2008). Coarse-Grained Molecular Dynamics Simulations of the Energetics of Helix Insertion into a Lipid Bilayer. *Biochemistry* **47**(43), 11321.

[41] **Boucrot, E.**, **A. Pick**, **G. Camdere**, **N. Liska**, **E. Evergren**, **H. T. McMahon**, **M. M. Kozlov** (2012). Membrane Fission Is Promoted by Insertion of Amphipathic Helices and Is Restricted by Crescent BAR Domains. *Cell* **149**(1), 124.

[42] **Box, G. E. P.** (1976). Science and Statistics. *J. Am. Stat. Assoc.* **71**(356), 791.

[43] **Bradley, R.**, **R. Radhakrishnan** (2013). Coarse-Grained Models for Protein-Cell Membrane Interactions. *Polymers* **5**(3), 890.

[44] **Bradley, R. P.**, **R. Radhakrishnan** (2016). Curvature-undulation coupling as a basis for curvature sensing and generation in bilayer membranes. *Proc. Natl. Acad. Sci. U. S. A.* **In press.**

[45] **Brandt, E. G.**, **A. R. Braun**, **J. N. Sachs**, **J. F. Nagle**, **O. Edholm** (2011). Interpretation of Fluctuation Spectra in Lipid Bilayer Simulations. *Biophys. J.* **100**(9), 2104.

[46] **Brannigan, G.**, **F. L. Brown** (2006). A Consistent Model for Thermal Fluctuations and Protein-Induced Deformations in Lipid Bilayers. *Biophys. J.* **90**(5), 1501.

[47] **Brannigan, G.**, **L.-L. Lin**, **F. Brown** (2006). Implicit solvent simulation models for biomembranes. *Eur. Biophys. J.* **35**(2), 104.

[48] **Braun, A. R.**, **E. Sevcsik**, **P. Chin**, **E. Rhoades**, **S. Tristram-Nagle**, **J. N. Sachs** (2012). α-Synuclein Induces Both Positive Mean Curvature and Negative Gaussian Curvature in Membranes. *J. Am. Chem. Soc.* **134**(5), 2613.

[49] **Brewster, R.**, **S. A. Safran** (2009). Line Active Hybrid Lipids Determine Domain Size in Phase Separation of Saturated and Unsaturated Lipids. *Biophys. J.* **98**(6), L21.

[50] **Brooks, B. R.**, **R. E. Bruccoleri**, **B. D. Olafson**, **D. J. States**, **S. Swaminathan**, **M. Karplus** (1983). CHARMM: A program for macromolecular energy, minimization, and dynamics calculations. *J. Comput. Chem.* **4**(2), 187.

[51] **Buch, I.**, **T. Giorgino**, **G. De Fabritiis** (2011). Complete reconstruction of an enzyme-inhibitor binding process by molecular dynamics simulations. *Proc. Natl. Acad. Sci. U.S.A.* **108**(25), 10184.

[52] **Bunney, T. D.**, **M. Katan** (2010). Phosphoinositide signalling in cancer: beyond PI3K and PTEN. *Nat. Rev. Cancer* **10**(5), 342.

[53] **Bussi, G.**, **D. Donadio**, **M. Parrinello** (2007). Canonical sampling through velocity rescaling. *J. Chem. Phys.* **126**(1), 014101.

[54] **Canham, P.** (1970). The minimum energy of bending as a possible explanation of the biconcave shape of the human red blood cell. *J. Theor. Biol.* **26**(1), 61.

[55] **Cellmer, T.**, **M. Buscaglia**, **E. R. Henry**, **J. Hofrichter**, **W. A. Eaton** (2011). Making connections between ultrafast protein folding kinetics and molecular dynamics simulations. *Proc. Natl. Acad. Sci. U.S.A.* **108**(15), 6103.

[56] **Chang, R.**, **G. S. Ayton**, **G. A. Voth** (2005). Multiscale coupling of mesoscopic-and atomistic-level lipid bilayer simulations. *J. Chem. Phys.* **122**, 244716.

[57] **Cheung, D. L.** (2012). Molecular Simulation of Hydrophobin Adsorption at an Oil-Water Interface. *Langmuir* **28**(23), 8730.

[58] **Chodera, J. D.**, **D. L. Mobley**, **M. R. Shirts**, **R. W. Dixon**, **K. Branson**, **V. S. Pande** (2011). Alchemical free energy methods for drug discovery: progress and challenges. *Curr. Opin. Struct. Biol.* **21**(2), 150.

[59] **Christen, M.**, **P. H. Hünenberger**, **D. Bakowies**, **R. Baron**, **R. Bürgi**, **D. P. Geerke**, **T. N. Heinz**, **M. A. Kastenholz**, **V. Kräutler**, **C. Oostenbrink**, **C. Peter**, **D. Trzesniak**, **W. F. van Gunsteren** (2005). The GROMOS software for biomolecular simulation: GROMOS05. *J. Comput. Chem.* **26**(16), 1719.

[60] **Coluzza, I.** (2011). A Coarse-Grained Approach to Protein Design: Learning from Design to Understand Folding. *PLoS One* **6**(7), e20853.

[61] **Cooke, I. R.**, **M. Deserno** (2006). Coupling between lipid shape and membrane curvature. *Biophys. J.* **91**(2), 487.

[62] **Cooke, I. R.**, **K. Kremer**, **M. Deserno** (2005). Tunable generic model for fluid bilayer membranes. *Phys. Rev. E* **72**(1), 011506.

[63] **Copestake, A. P.**, **G. W. Neilson**, **J. E. Enderby** (1985). The structure of a highly concentrated aqueous solution of lithium chloride. *J. Phys. C: Solid State Phys.* **18**(22), 4211.

[64] **Corsi, J.**, **R. W. Hawtin**, **O. Ces**, **G. S. Attard**, **S. Khalid** (2010). DNA Lipoplexes: Formation of the Inverse Hexagonal Phase Observed by Coarse-Grained Molecular Dynamics Simulation. *Langmuir* **26**(14), 12119.

[65] **Csermely, P.**, **K. Singh Sandhu**, **E. Hazai**, **Z. Hoksza**, **H. J. M. Kiss**, **F. Miozzo**, **D. V. Veres**, **F. Piazza**, **R. Nussinov** (2011). Disordered proteins and network disorder in network descriptions of protein structure, dynamics and function. Hypotheses and a comprehensive review. *Curr. Protein Pept. Sci.* **13**, 19.

[66] **Cui, H.**, **C. Mim**, **F. Vázquez**, **E. Lyman**, **V. Unger**, **G. Voth** (2013). Understanding the Role of Amphipathic Helices in N-BAR Domain Driven Membrane Remodeling. *Biophys. J.* **104**(2), 404.

[67] **De Almeida, R. F.**, **A. Fedorov**, **M. Prieto** (2003). Sphingomyelin/phosphatidylcholine/cholesterol phase diagram: boundaries and composition of lipid rafts. *Biophys. J.* **85**(4), 2406.

[68] **de Haas, K. H.**, **C. Blom**, **D. van den Ende**, **M. H. G. Duits**, **J. Mellema** (1997). Deformation of giant lipid bilayer vesicles in shear flow. *Phys. Rev. E* **56**(6), 7132.

[69] **de Jong, D. H.**, **X. Periole**, **S. J. Marrink** (2012). Dimerization of Amino Acid Side Chains: Lessons from the Comparison of Different Force Fields. *J. Chem. Theory Comput.* **8**(3), 1003.

[70] **de Jong, D. H.**, **G. Singh**, **W. F. D. Bennett**, **C. Arnarez**, **T. A. Wassenaar**, **L. V. Schäfer**, **X. Periole**, **D. P. Tieleman**, **S. J. Marrink** (2012). Improved Parameters for the Martini Coarse-Grained Protein Force Field. *J. Chem. Theory Comput.* **9**(1), 687.

[71] **De Meyer, F. J.-M.**, **M. Venturoli**, **B. Smit** (2008). Molecular simulations of lipid-mediated protein-protein interactions. *Biophys. J.* **95**(4), 1851.

[72] **de Meyer, F. J.-M.**, **J. M. Rodgers**, **T. F. Willems**, **B. Smit** (2010). Molecular Simulation of the Effect of Cholesterol on Lipid-Mediated Protein-Protein Interactions. *Biophys. J.* **99**(11), 3629.

[73] **de Ruiter, A.**, **C. Oostenbrink** (2011). Free energy calculations of protein-ligand interactions. *Curr. Opin. Chem. Biol.* **15**(4), 547.

[74] **Demond, A. H.**, **A. S. Lindner** (1993). Estimation of interfacial tension between organic liquids and water. *Environ. Sci. Technol.* **27**(12), 2318.

[75] **DeVane, R.**, **W. Shinoda**, **P. B. Moore**, **M. L. Klein** (2009). Transferable Coarse Grain Nonbonded Interaction Model for Amino Acids. *J. Chem. Theory Comput.* **5**(8), 2115.

[76] **Devaux, P. F.**, **A. Herrmann**, **N. Ohlwein**, **M. M. Kozlov** (2008). How lipid flippases can modulate membrane structure. *Biochim. Biophys. Acta, Biomembr.* **1778**(7-8), 1591.

[77] **Dill, K. A.**, **S. Bromberg**, *Molecular Driving Forces* (Garland Science, New York, 2011), 2nd edition.

[78] **Dill, K. A.**, **J. L. MacCallum** (2012). The Protein-Folding Problem, 50 Years On. *Science* **338**(6110), 1042.

[79] **Doherty, G. J.**, **H. T. McMahon** (2008). Mediation, Modulation, and Consequences of Membrane-Cytoskeleton Interactions. *Annu. Rev. Biophys.* **37**(1), 65.

[80] **Dong, G.**, **A. H. Hutagalung**, **C. M. Fu**, **P. Novick**, **K. M. Reinisch** (2005). The structures of exocyst subunit Exo70p and the Exo84p C-terminal domains reveal a common motif. *Nat. Struct. Mol. Biol.* **12**(12), 1094.

[81] **Douglass, D. C.**, **D. W. McCall** (1958). Diffusion in Paraffin Hydrocarbons. *J. Phys. Chem.* **62**(9), 1102.

[82] **Drin, G.**, **B. Antonny** (2010). Amphipathic helices and membrane curvature. *FEBS Lett.* **584**(9), 1840.

[83] **Dror, R. O.**, **R. M. Dirks**, **J. P. Grossman**, **H. Xu**, **D. E. Shaw** (2012). Biomolecular Simulation: A Computational Microscope for Molecular Biology. *Annu. Rev. Biophys.* **41**(1), 429.

[84] **Duffy, E. M.**, **W. L. Jorgensen** (2000). Prediction of Properties from Simulations: Free Energies of Solvation in Hexadecane, Octanol, and Water. *J. Am. Chem. Soc.* **122**(12), 2878.

[85] **Duncan, S. L.**, **R. G. Larson** (2010). Folding of lipid monolayers containing lung surfactant proteins $SP-B_{1-25}$ and SP-C studied via coarse-grained molecular dynamics simulations. *Biochim. Biophys. Acta, Biomembr.* **1798**(9), 1632.

[86] **Durrant, J.**, **J. A. McCammon** (2011). Molecular dynamics simulations and drug discovery. *BMC Biology* **9**(1), 71.

[87] **Edeling, M. A.**, **C. Smith**, **D. Owen** (2006). Life of a clathrin coat: insights from clathrin and AP structures. *Nat. Rev. Mol. Cell Biol.* **7**(1), 32.

[88] **Ehrlich, M.**, **W. Boll**, **A. van Oijen**, **R. Hariharan**, **K. Chandran**, **M. L. Nibert**, **T. Kirchhausen** (2004). Endocytosis by random initiation and stabilization of clathrin-coated pits. *Cell* **118**(5), 591.

[89] **Elcock, A. H.** (2003). Atomic-level observation of macromolecular crowding effects: Escape of a protein from the GroEL cage. *Proc. Natl. Acad. Sci. U.S.A.* **100**(5), 2340.

[90] **Emperador, A.**, **O. Carrillo**, **M. Rueda**, **M. Orozco** (2008). Exploring the Suitability of Coarse-Grained Techniques for the Representation of Protein Dynamics. *Biophys. J.* **95**(5), 2127.

[91] **Engelman, D. M.** (2005). Membranes are more mosaic than fluid. *Nature* **438**(7068), 578.

[92] **Engin, O.**, **A. Villa**, **C. Peter**, **M. Sayar** (2011). A Challenge for Peptide Coarse Graining: Transferability of Fragment-Based Models. *Macromol. Theory Simul.* **20**(7), 451.

[93] **Ercolessi, F.**, **J. B. Adams** (1994). Interatomic Potentials from First-Principles Calculations: The Force-Matching Method. *Europhys. Lett.* **26**(8), 583.

[94] **Español, P.**, **P. Warren** (1995). Statistical Mechanics of Dissipative Particle Dynamics. *Europhys. Lett.* **30**(4), 191.

[95] **Faiderbe, S.**, **J. Chagnaud**, **M. Geffard** (1992). Anti-phosphoinositide autoantibodies in sera of cancer patients: isotypic and immunochemical characterization. *Cancer Lett.* **66**(1), 35.

[96] **Faller, R.**, **S.-J. Marrink** (2004). Simulation of Domain Formation in Mixed DLPC-DSPC Bilayers. *Langmuir* **20**(18), 7686.

[97] **Feller, S. E.**, **A. D. MacKerell** (2000). An Improved Empirical Potential Energy Function for Molecular Simulations of Phospholipids. *J. Phys. Chem. B* **104**(31), 7510.

[98] **Feller, S. E.**, **R. W. Pastor** (1999). Constant surface tension simulations of lipid bilayers: the sensitivity of surface areas and compressibilities. *J. Chem. Phys.* **111**, 1281.

[99] **Filippov, A.**, **G. Orädd**, **G. Lindblom** (2003). Influence of Cholesterol and Water Content on Phospholipid Lateral Diffusion in Bilayers. *Langmuir* **19**(16), 6397.

[100] **Fischer, H. E.**, **A. C. Barnes**, **P. S. Salmon** (2006). Neutron and x-ray diffraction studies of liquids and glasses. *Rep. Prog. Phys.* **69**(1), 233.

[101] **Ford, M. G. J.**, **I. G. Mills**, **B. J. Peter**, **Y. Vallis**, **G. J. K. Praefcke**, **P. R. Evans**, **H. T. McMahon** (2002). Curvature of clathrin-coated pits driven by epsin. *Nature* **419**(6905), 361.

[102] **Föster, D.** (1986). On the scale dependence, due to thermal fluctuations, of the elastic properties of membranes. *Phys. Lett. A* **114**(3), 115.

[103] **Fotin, A.**, **Y. Cheng**, **P. Sliz**, **N. Grigorieff**, **S. C. Harrison**, **T. Kirchhausen**, **T. Walz** (2004). Molecular model for a complete clathrin lattice from electron cryomicroscopy. *Nature* **432**(7017), 573.

[104] **Freddolino, P. L.**, **C. B. Harrison**, **Y. Liu**, **K. Schulten** (2010). Challenges in protein-folding simulations. *Nat. Phys.* **6**(10), 751.

[105] **Freddolino, P. L.**, **A. Y. Shih**, **A. Arkhipov**, **Y. Ying**, **Z. Chen**, **K. Schulten**, Application of Residue-Based and Shape-Based Coarse-Graining to Biomolecular Simulations. *Coarse-graining of condensed phase and biomolecular systems* (CRC Press, 2009), p. 299.

[106] **Friedrichs, M. S.**, **P. Eastman**, **V. Vaidyanathan**, **M. Houston**, **S. Legrand**, **A. L. Beberg**, **D. L. Ensign**, **C. M. Bruns**, **V. S. Pande** (2009). Accelerating molecular dynamic simulation on graphics processing units. *J. Comput. Chem.* **30**(6), 864.

[107] **Frost, A.**, **V. M. Unger**, **P. De Camilli** (2009). The BAR Domain Superfamily: Membrane-Molding Macromolecules. *Cell* **137**(2), 191.

[108] **Fuhrmans, M.**, **V. Knecht**, **S. J. Marrink** (2009). A Single Bicontinuous Cubic Phase Induced by Fusion Peptides. *J. Am. Chem. Soc.* **131**(26), 9166.

[109] **Gamper, N.**, **M. S. Shapiro** (2007). Target-specific PIP(2) signalling: how might it work? *J. Physiol.* **582**(Pt 3), 967.

[110] **Genchev, G. Z.**, **M. Källberg**, **G. Gürsoy**, **A. Mittal**, **L. Dubey**, **O. Perisic**, **G. Feng**, **R. Langlois**, **H. Lu** (2009). Mechanical Signaling on the Single Protein Level Studied Using Steered Molecular Dynamics. *Cell Biochem. Biophys.* **55**(3), 141.

[111] **Gest, H.** (2004). The discovery of microorganisms by Robert Hooke and Antoni van Leeuwenhoek, Fellows of The Royal Society. *Notes Rec. R. Soc. Lond.* **58**(2), 187.

[112] **Gipson, B.**, **D. Hsu**, **L. E. Kavraki**, **J.-C. Latombe** (2012). Computational Models of Protein Kinematics and Dynamics: Beyond Simulation. *Annu. Rev. Anal. Chem.* **5**(1), 273.

[113] **Gittes, F.**, **B. Mickey**, **J. Nettleton**, **J. Howard** (1993). Flexural Rigidity of Microtubules and Actin-Filaments Measured from Thermal Fluctuations in Shape. *J. Cell Biol.* **120**(4), 923.

[114] **Goetz, R.**, **R. Lipowsky** (1998). Computer simulations of bilayer membranes: Self-assembly and interfacial tension. *J. Chem. Phys.* **108**, 7397.

[115] **Goetz, R.**, **G. Gompper**, **R. Lipowsky** (1999). Mobility and Elasticity of Self-Assembled Membranes. *Phys. Rev. Lett.* **82**(1), 221.

[116] **Goga, N.**, **A. J. Rzepiela**, **A. H. de Vries**, **S. J. Marrink**, **H. J. C. Berendsen** (2012). Efficient Algorithms for Langevin and DPD Dynamics. *J. Chem. Theory Comput.* **8**(10), 3637.

[117] **Gopal, S. M.**, **S. Mukherjee**, **Y.-M. Cheng**, **M. Feig** (2010). PRIMO/PRIMONA: A coarse-grained model for proteins and nucleic acids that preserves near-atomistic accuracy. *Proteins* **78**(5), 1266.

[118] **Grafmüller, A.**, **J. Shillcock**, **R. Lipowsky** (2007). Pathway of Membrane Fusion with Two Tension-Dependent Energy Barriers. *Phys. Rev. Lett.* **98**(21), 218101.

[119] **Grant, B. J.**, **A. A. Gorfe**, **J. A. McCammon** (2010). Large conformational changes in proteins: signaling and other functions. *Curr. Opin. Struct. Biol.* **20**(2), 142.

[120] **Groot, R. D.**, **P. B. Warren** (1997). Dissipative particle dynamics: Bridging the gap between atomistic and mesoscopic simulation. *J. Chem. Phys.* **107**, 4423.

[121] **Groot, R.**, **K. Rabone** (2001). Mesoscopic Simulation of Cell Membrane Damage, Morphology Change and Rupture by Nonionic Surfactants. *Biophys. J.* **81**(2), 725.

[122] **Guerrier, S.**, **J. Coutinho-Budd**, **T. Sassa**, **A. Gresset**, **N. V. Jordan**, **K. Chen**, **W.-L. Jin**, **A. Frost**, **F. Polleux** (2009). The F-BAR Domain of srGAP2 Induces Membrane Protrusions Required for Neuronal Migration and Morphogenesis. *Cell* **138**(5), 990.

[123] **Guvench, O.**, **J. MacKerell, Alexander D.**, Comparison of Protein Force Fields for Molecular Dynamics Simulations. *Methods Molecular Biology* (edited by **A. Kukol**), volume 443 (Humana Press, 2008), pp. 63–88.

[124] **Hamai, C.**, **T. Yang**, **S. Kataoka**, **P. S. Cremer**, **S. M. Musser** (2006). Effect of Average Phospholipid Curvature on Supported Bilayer Formation on Glass by Vesicle Fusion. *Biophys. J.* **90**(4), 1241.

[125] **Hamburger, Z. A.**, **A. E. Hamburger**, **A. P. West**, **W. I. Weis** (2006). Crystal structure of the S-cerevisiae exocyst component Exo70p. *J. Mol. Biol.* **356**(1), 9.

[126] **Han, W.**, **C.-K. Wan**, **Y.-D. Wu** (2008). Toward a Coarse-Grained Protein Model Coupled with a Coarse-Grained Solvent Model: Solvation Free Energies of Amino Acid Side Chains. *J. Chem. Theory Comput.* **4**(11), 1891.

[127] **Harmandaris, V. A.**, **M. Deserno** (2006). A novel method for measuring the bending rigidity of model lipid membranes by simulating tethers. *J. Chem. Phys.* **125**, 204905.

[128] **Hatakeyama, M.**, **R. Faller** (2007). Coarse-grained simulations of ABA amphiphilic triblock copolymer solutions in thin films. *Phys. Chem. Chem. Phys.* **9**(33), 4662.

[129] **He, B.**, **W. Guo** (2009). The exocyst complex in polarized exocytosis. *Curr.Opin.Cell Biol.* **21**(4), 537.

[130] **Head-Gordon, T.**, **S. Brown** (2003). Minimalist models for protein folding and design. *Curr. Opin. Struct. Biol.* **13**(2), 160.

[131] **Helfrich, W.** (1973). Elastic properties of lipid bilayers: theory and possible experiments. *Z. Naturforsch., C: Biosci.* **28**(11), 693.

[132] **Helfrich, W.** (1985). Effect of thermal undulations on the rigidity of fluid membranes and interfaces. *J. Phys. France* **46**(7), 1263.

[133] **Henderson, D.**, **D. Boda** (2009). Insights from theory and simulation on the electrical double layer. *Phys. Chem. Chem. Phys.* **11**(20), 3822.

[134] **Hendrickson, H.**, **J. Fullington** (1965). Stabilities of metal complexes of phospholipids: Ca(II), Mg(II), and Ni(II) complexes of phosphatidylserine and triphosphoinositide. *Biochemistry* **4**(8), 1599.

[135] **Hintzenstern, U.**, **W. Schwarz**, **M. Goerig**, **H. Petermann** (2002). Development of the "lipoid theory of narcosis" in German-speaking countries in the

19th century: from Bibra/Harless to Meyer/Overton. *Int. Congr. Ser.* **1242**, 609.

[136] **Hirst, J.**, **J. R. Edgar**, **G. H. H. Borner**, **S. Li**, **D. A. Sahlender**, **R. Antrobus**, **M. S. Robinson** (2015). Contributions of epsinR and gadkin to clathrin-mediated intracellular trafficking. *Mol. Biol. Cell* **26**(17), 3085.

[137] **Hom, R. A.**, **M. Vora**, **M. Regner**, **O. M. Subach**, **W. Cho**, **V. V. Verkhusha**, **R. V. Stahelin**, **T. G. Kutateladze** (2007). pH-dependent Binding of the Epsin ENTH Domain and the AP180 ANTH Domain to PI(4,5)P2-containing Bilayers. *J. Mol. Biol.* **373**(2), 412.

[138] **Hoogerbrugge, P. J.**, **J. M. V. A. Koelman** (1992). Simulating Microscopic Hydrodynamic Phenomena with Dissipative Particle Dynamics. *EPL* **19**(3), 155.

[139] **Hsieh, W.-T.**, **C.-J. Hsu**, **B. R. Capraro**, **T. Wu**, **C.-M. Chen**, **S. Yang**, **T. Baumgart** (2012). Curvature Sorting of Peripheral Proteins on Solid-Supported Wavy Membranes. *Langmuir* **28**(35), 12838.

[140] **Hsin, J.**, **J. Gumbart**, **L. G. Trabuco**, **E. Villa**, **P. Qian**, **C. N. Hunter**, **K. Schulten** (2009). Protein-Induced Membrane Curvature Investigated through Molecular Dynamics Flexible Fitting. *Biophys. J.* **97**(1), 321.

[141] **Hsu, S. C.**, **D. TerBush**, **M. Abraham**, **W. Guo** (2004). The exocyst complex in polarized exocytosis. *International Review of Cytology - a Survey of Cell Biology, Vol.233* **233**, 243.

[142] **Humphrey, W.**, **A. Dalke**, **K. Schulten** (1996). VMD: Visual molecular dynamics. *J. Mol. Graphics* **14**(1), 33.

[143] **Hunter, J. D.** (2007). Matplotlib: A 2D graphics environment. *Comput. Sci. Eng.* **9**(3), 90.

[144] **Ilya, L.**, **D. A. Christian**, **Y.-H. Wang**, **J. J. Madara**, **D. E. Discher**, **P. A. Janmey** (2009). Calcium-dependent lateral organization in phosphatidylinositol (4,5) bisphosphate (PIP2)- and cholesterol-containing monolayers. *Biochemistry* **48**(34), 8241.

[145] **Im, W.**, **C. L. Brooks** (2005). Interfacial folding and membrane insertion of designed peptides studied by molecular dynamics simulations. *Proc. Natl. Acad. Sci. U. S. A.* **102**(19), 6771.

[146] **Inoue, M.**, **L. Chang**, **J. Hwang**, **S. H. Chiang**, **A. R. Saltiel** (2003). The exocyst complex is required for targeting of Glut4 to the plasma membrane by insulin. *Nature* **422**(6932), 629.

[147] **Itoh, T.**, **K. S. Erdmann**, **A. Roux**, **B. Habermann**, **H. Werner**, **P. De Camilli** (2005). Dynamin and the Actin Cytoskeleton Cooperatively Regulate Plasma Membrane Invagination by BAR and F-BAR Proteins. *Dev. Cell* **9**(6), 791.

[148] **IUPAC**, *Compendium of Chemical Terminology, 2nd ed.* (Blackwell Scientific Publications, Oxford, 1997), ISBN 978-0865426849.

[149] **Izvekov, S.**, **P. W. Chung**, **B. M. Rice** (2010). The multiscale coarse-graining method: Assessing its accuracy and introducing density dependent coarse-grain potentials. *J. Chem. Phys.* **133**(6), 064109.

[150] **Izvekov, S.**, **A. Violi**, **G. A. Voth** (2005). Systematic Coarse-Graining of Nanoparticle Interactions in Molecular Dynamics Simulation. *J. Phys. Chem. B* **109**(36), 17019.

[151] **Izvekov, S.**, **G. A. Voth** (2005). A Multiscale Coarse-Graining Method for Biomolecular Systems. *J. Phys. Chem. B* **109**(7), 2469.

[152] **Jähnig, F.** (1996). What is the surface tension of a lipid bilayer membrane? *Biophys. J.* **71**(3), 1348.

[153] **Jensen, M. B.**, **V. K. Bhatia**, **C. C. Jao**, **J. E. Rasmussen**, **S. L. Pedersen**, **K. J. Jensen**, **R. Langen**, **D. Stamou** (2011). Membrane Curvature Sensing by Amphipathic Helices: A single liposome studying using α-synuclein and annexin B12. *J. Biol. Chem.* **286**(49), 42603.

[154] **Jones, E.**, **T. Oliphant**, **P. Peterson**, *et al.* (2001), `SciPy`: Open source scientific tools for Python. Http://www.scipy.org/.

[155] **Jónsson, S. Æ.**, **I. Staneva**, **S. Mohanty**, **A. Irbäck** (2012). Monte Carlo Studies of Protein Aggregation. *Phys. Procedia* **34**(0), 49.

[156] **Kabsch, W.**, **C. Sander** (1983). Dictionary of protein secondary structure: Pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* **22**(12), 2577.

[157] **Karplus, M.**, **J. Kuriyan** (2005). Molecular dynamics and protein function. *Proc. Natl. Acad. Sci. U.S.A.* **102**(19), 6679.

[158] **Karplus, M.**, **J. A. McCammon** (2002). Molecular dynamics simulations of biomolecules. *Nat. Struct. Mol. Biol.* **9**(9), 646.

[159] **Kasson, P. M.**, **N. W. Kelley**, **N. Singhal**, **M. Vrljic**, **A. T. Brunger**, **V. S. Pande** (2006). Ensemble molecular dynamics yields submillisecond kinetics and intermediates of membrane fusion. *Proc. Natl. Acad. Sci. U.S.A.* **103**(32), 11916.

[160] **Katso, R.**, **K. Okkenhaug**, **K. Ahmadi**, **S. White**, **J. Timms**, **M. D. Waterfield** (2001). Cellular Function of Phosphoinositide 3-Kinases: Implications for Development, Immunity, Homeostasis, and Cancer. *Annu. Rev. Cell Dev. Biol.* **17**(1), 615.

[161] **Khalfa, A.**, **M. Tarek** (2010). On the Antibacterial Action of Cyclic Peptides: Insights from Coarse-Grained MD Simulations. *J. Phys. Chem. B* **114**(8), 2676.

[162] **Kharakoz, D. P.**, **E. A. Shlyapnikova** (2000). Thermodynamics and Kinetics of the Early Steps of Solid-State Nucleation in the Fluid Lipid Bilayer. *J. Phys. Chem. B* **104**(44), 10368.

[163] **Khurana, E.**, **R. H. DeVane**, **A. Kohlmeyer**, **M. L. Klein** (2008). Probing Peptide Nanotube Self-Assembly at a Liquid-Liquid Interface with Coarse-Grained Molecular Dynamics. *Nano Lett.* **8**(11), 3626.

[164] **Kim, H.**, **J. Hsin**, **Y. Liu**, **P. R. Selvin**, **K. Schulten** (2010). Formation of Salt Bridges Mediates Internal Dimerization of Myosin VI Medial Tail Domain. *Structure* **18**(11), 1443.

[165] **Kirchhausen, T.** (2000). Three ways to make a vesicle. *Nat. Rev. Mol. Cell Biol.* **1**(3), 187.

[166] **Klauda, J. B.**, **R. M. Venable**, **J. A. Freites**, **J. W. O'Connor**, **D. J. Tobias**, **C. Mondragon-Ramirez**, **I. Vorobyov**, **A. D. MacKerell**, **R. W. Pastor** (2010). Update of the CHARMM All-Atom Additive Force Field for Lipids: Validation on Six Lipid Types. *J. Phys. Chem. B* **114**(23), 7830.

[167] **Klepeis, J. L.**, **K. Lindorff-Larsen**, **R. O. Dror**, **D. E. Shaw** (2009). Long-timescale molecular dynamics simulations of protein structure and function. *Curr. Opin. Struct. Biol.* **19**(2), 120.

[168] **Kollman, P. A.**, **I. Massova**, **C. Reyes**, **B. Kuhn**, **S. Huo**, **L. Chong**, **M. Lee**, **T. Lee**, **Y. Duan**, **W. Wang**, **O. Donini**, **P. Cieplak**, **J. Srinivasan**, **D. A. Case**, **T. E. Cheatham** (2000). Calculating Structures and Free Energies of Complex Molecules: Combining Molecular Mechanics and Continuum Models. *Acc. Chem. Res.* **33**(12), 889.

[169] **Krynicki, K.**, **C. D. Green**, **D. W. Sawyer** (1978). Pressure and temperature dependence of self-diffusion in water. *Faraday Discuss. Chem. Soc.* **66**(0), 199.

[170] **Kučerka, N.**, **J. F. Nagle**, **J. N. Sachs**, **S. E. Feller**, **J. Pencer**, **A. Jackson**, **J. Katsaras** (2008). Lipid Bilayer Structure Determined by the Simultaneous Analysis of Neutron and X-Ray Scattering Data. *Biophys. J.* **95**(5), 2356.

[171] **Kumar, S.**, **J. M. Rosenberg**, **D. Bouzida**, **R. H. Swendsen**, **P. A. Kollman** (1992). THE weighted histogram analysis method for free-energy calculations on biomolecules. I. The method. *J. Comput. Chem.* **13**(8), 1011.

[172] **Kweon, D.**, **Y. Shin**, **J. Shin**, **J. Lee**, **J. Lee**, **J. Seo**, **Y. Kim** (2006). Membrane topology of helix 0 of the Epsin N-terminal homology domain. *Mol. Cells* **21**(3), 428.

[173] **Lai, C.-L.**, **C. C. Jao**, **E. Lyman**, **J. L. Gallop**, **B. J. Peter**, **H. T. McMahon**, **R. Langen**, **G. A. Voth** (2012). Membrane Binding and Self-Association of the Epsin N-Terminal Homology Domain. *J. Mol. Biol.* **423**(5), 800.

[174] **Laio, A.**, **F. L. Gervasio** (2008). Metadynamics: a method to simulate rare events and reconstruct the free energy in biophysics, chemistry and material science. *Rep. Prog. Phys.* **71**(12), 126601.

[175] **Lange, O. F.**, **H. Grubmüller** (2006). Collective Langevin dynamics of conformational motions in proteins. *J. Chem. Phys.* **124**(21), 214903.

[176] **Larson, R.** (1988). Monte Carlo lattice simulation of amphiphilic systems in two and three dimensions. *J. Chem. Phys.* **89**, 1642.

[177] **Laughton, C. A.**, **S. A. Harris** (2011). The atomistic simulation of DNA. *WIREs Comput. Mol. Sci.* **1**(4), 590.

[178] **Lee, E.**, **J. Hsin**, **M. Sotomayor**, **G. Comellas**, **K. Schulten** (2009). Discovery Through the Computational Microscope. *Structure* **17**(10), 1295.

[179] **Lee, H.**, **R. G. Larson** (2006). Molecular Dynamics Simulations of PAMAM Dendrimer-Induced Pore Formation in DPPC Bilayers with a Coarse-Grained Model. *J. Phys. Chem. B* **110**(37), 18204.

[180] **Lee, H.**, **R. G. Larson** (2008). Coarse-Grained Molecular Dynamics Studies of the Concentration and Size Dependence of Fifth- and Seventh-Generation PAMAM Dendrimers on Pore Formation in DMPC Bilayer. *J. Phys. Chem. B* **112**(26), 7778.

[181] **Lee, H.**, **R. G. Larson** (2008). Lipid Bilayer Curvature and Pore Formation Induced by Charged Linear Polymers and Dendrimers: The Effect of Molecular Shape. *J. Phys. Chem. B* **112**(39), 12279.

[182] **Leguèbe, M.**, **C. Nguyen**, **L. Capece**, **Z. Hoang**, **A. Giorgetti**, **P. Carloni** (2012). Hybrid Molecular Mechanics/Coarse-Grained Simulations for Structural Prediction of G-Protein Coupled Receptor/Ligand Complexes. *PLoS One* **7**(10), e47332.

[183] **Lemmon, M. A.** (2008). Membrane recognition by phospholipid-binding domains. *Nat. Rev. Mol. Cell Biol.* **9**(2), 99.

[184] **Levy, Y.**, **P. G. Wolynes**, **J. N. Onuchic** (2004). Protein topology determines binding mechanism. *Proc. Natl. Acad. Sci. U.S.A.* **101**(2), 511.

[185] **Lewis, J. I.**, **D. J. Moss**, **T. A. Knotts** (2012). Multiple molecule effects on the cooperativity of protein folding transitions in simulations. *J. Chem. Phys.* **136**(24), 245101.

[186] **Li, H.**, **Y. Cao** (2010). Protein Mechanics: From Single Molecules to Functional Biomaterials. *Acc. Chem. Res.* **43**(10), 1331.

[187] **Lide, D.**, *CRC Handbook of Chemistry and Physics* (CRC press, Boca Raton, FL, 1992).

[188] **Lill, M. A.** (2011). Efficient Incorporation of Protein Flexibility and Dynamics into Molecular Docking Simulations. *Biochemistry* **50**(28), 6157.

[189] **Lindahl, E.**, **O. Edholm** (2000). Mesoscopic undulations and thickness fluctuations in lipid bilayers from molecular dynamics simulations. *Biophys. J.* **79**(1), 426.

[190] **Lindahl, E.**, **O. Edholm** (2000). Spatial and energetic-entropic decomposition of surface tension in lipid bilayers from molecular dynamics simulations. *J. Chem. Phys.* **113**, 3882.

[191] **Lindahl, E.**, **M. S. Sansom** (2008). Membrane proteins: molecular dynamics simulations. *Curr. Opin. Struct. Biol.* **18**(4), 425.

[192] **Lindorff-Larsen, K.**, **P. Maragakis**, **S. Piana**, **M. P. Eastwood**, **R. O. Dror**, **D. E. Shaw** (2012). Systematic Validation of Protein Force Fields against Experimental Data. *PLoS One* **7**(2), e32131.

[193] **Liu, J.**, **Y. Sun**, **D. G. Drubin**, **G. F. Oster** (2009). The Mechanochemistry of Endocytosis. *PLoS Biol.* **7**(9), e1000204.

[194] **Liu, J.**, **X. Zuo**, **P. Yue**, **W. Guo** (2007). Phosphatidylinositol 4,5-bisphosphate mediates the targeting of the exocyst to the plasma membrane for exocytosis in mammalian cells. *Mol. Biol. Cell* **18**(11), 4483.

[195] **Liu, J.**, **R. Bradley**, **D. M. Eckmann**, **P. S. Ayyaswamy**, **R. Radhakrishnan** (2011). Multiscale Modeling of Functionalized Nanocarriers in Targeted Drug Delivery. *Curr. Nanosci.* **7**(5), 727.

[196] **Liu, J.**, **R. Tourdot**, **V. Ramanan**, **N. J. Agrawal**, **R. Radhakrishanan** (2012). Mesoscale simulations of curvature-inducing protein partitioning on lipid bilayer membranes in the presence of mean curvature fields. *Mol. Phys.* **110**(11-12), 1127.

[197] **Liu, L. A.**, **P. Bradley** (2012). Atomistic modeling of protein-DNA interaction specificity: progress and applications. *Curr. Opin. Struct. Biol.* **22**(4), 397.

[198] **Livesay, D. R.** (2010). Protein dynamics: dancing on an ever-changing free energy stage. *Curr. Opin. Pharmacol.* **10**(6), 706.

[199] **Louhivuori, M.**, **H. J. Risselada**, **E. van der Giessen**, **S. J. Marrink** (2010). Release of content through mechano-sensitive gates in pressurized liposomes. *Proc. Natl. Acad. Sci. U.S.A.* **107**(46), 19856.

[200] **Lucent, D.**, **C. D. Snow**, **C. E. Aitken**, **V. S. Pande** (2010). Non-Bulk-Like Solvent Behavior in the Ribosome Exit Tunnel. *PLoS Comput. Biol.* **6**(10), e1000963.

[201] **Lycklama a Nijeholt, J. A.**, **M. Bulacu**, **S. J. Marrink**, **A. J. M. Driessen** (2010). Immobilization of the Plug Domain Inside the SecY Channel Allows Unrestricted Protein Translocation. *J. Biol. Chem.* **285**(31), 23747.

[202] **Lyman, E.**, **J. Pfaendtner**, **G. Voth** (2008). Systematic Multiscale Parameterization of Heterogeneous Elastic Network Models of Proteins. *Biophys. J.* **95**, 4183.

[203] **Lyubartsev, A. P.**, **A. Laaksonen** (1995). Calculation of effective interaction potentials from radial distribution functions: A reverse Monte Carlo approach. *Phys. Rev. E* **52**(4), 3730.

[204] **MacCallum, J. L.**, **W. F. D. Bennett**, **D. P. Tieleman** (2008). Distribution of Amino Acids in a Lipid Bilayer from Computer Simulations. *Biophys. J.* **94**(9), 3393.

[205] **Machnicka, B.**, **R. Grochowalska**, **D. M. Bogusławska**, **A. F. Sikorski**, **M. C. Lecomte** (2011). Spectrin-based skeleton as an actor in cell signaling. *Cell. Mol. Life Sci.* **69**(2), 191.

[206] **MacKenzie, K. R.**, **K. G. Fleming** (2008). Association energetics of membrane spanning α-helices. *Curr. Opin. Struct. Biol.* **18**(4), 412.

[207] **Mackerell, A. D.**, **M. Feig**, **C. L. Brooks** (2004). Extending the treatment of backbone energetics in protein force fields: Limitations of gas-phase quantum mechanics in reproducing protein conformational distributions in molecular dynamics simulations. *J. Comput. Chem.* **25**(11), 1400.

[208] **Maragliano, L.**, **E. Vanden-Eijnden** (2006). A temperature accelerated method for sampling free energy and determining reaction pathways in rare events simulations. *Chem. Phys. Lett.* **426**(1–3), 168.

[209] **Marrink, S. J.**, **A. E. Mark** (2001). Effect of Undulations on Surface Tension in Simulated Bilayers. *J. Phys. Chem. B* **105**(26), 6122.

[210] **Marrink, S. J.**, **A. H. de Vries**, **A. E. Mark** (2004). Coarse Grained Model for Semiquantitative Lipid Simulations. *J. Phys. Chem. B* **108**(2), 750.

[211] **Marrink, S. J.**, **A. E. Mark** (2003). The Mechanism of Vesicle Fusion as Revealed by Molecular Dynamics Simulations. *J. Am. Chem. Soc.* **125**(37), 11144.

[212] **Marrink, S. J.**, **H. J. Risselada**, **S. Yefimov**, **D. P. Tieleman**, **A. H. de Vries** (2007). The MARTINI Force Field: Coarse Grained Model for Biomolecular Simulations. *J. Phys. Chem. B* **111**(27), 7812.

[213] **Marrink, S. J.**, **J. Risselada**, **A. E. Mark** (2005). Simulation of gel phase formation and melting in lipid bilayers using a coarse grained model. *Chem. Phys. Lipids* **135**(2), 223.

[214] **Marrink, S.-J.**, **A. E. Mark** (2004). Molecular View of Hexagonal Phase Formation in Phospholipid Membranes. *Biophys. J.* **87**(6), 3894.

[215] **Marsh, D.** (2009). Cholesterol-induced fluid membrane domains: A compendium of lipid-raft ternary phase diagrams. *Biochim. Biophys. Acta, Biomembr.* **1788**(10), 2114.

[216] **Mattila, P. K.**, **P. Lappalainen** (2008). Filopodia: molecular architecture and cellular functions. *Nat. Rev. Mol. Cell Biol.* **9**(6), 446.

[217] **Mattila, P. K.**, **A. Pykalainen**, **J. Saarikangas**, **V. O. Paavilainen**, **H. Vihinen**, **E. Jokitalo**, **P. Lappalainen** (2007). Missing-in-metastasis and IRSp53 deform PI(4,5)P-2-rich membranes by an inverse BAR domain-like mechanism. *J.Cell Biol.* **176**(7), 953.

[218] **McCammon, J. A.**, **B. R. Gelin**, **M. Karplus** (1977). Dynamics of folded proteins. *Nature* **267**(5612), 585.

[219] **McConnell, H. M.** (1991). Structures and Transitions in Lipid Monolayers at the Air-Water Interface. *Annu. Rev. Phys. Chem.* **42**(1), 171.

[220] **McMahon, H. T.**, **E. Boucrot** (2011). Molecular mechanism and physiological functions of clathrin-mediated endocytosis. *Nat. Rev. Mol. Cell Biol.* **12**(8), 517.

[221] **McMahon, H. T.**, **E. Boucrot** (2015). Membrane curvature at a glance. *J. Cell Sci.* **128**(6), 1065.

[222] **McMahon, H. T.**, **J. L. Gallop** (2005). Membrane curvature and mechanisms of dynamic cell membrane remodelling. *Nature* **438**(7068), 590.

[223] **McMullen, T. P.**, **R. N. Lewis**, **R. N. McElhaney** (2004). Cholesterol-phospholipid interactions, the liquid-ordered phase and lipid rafts in model and biological membranes. *Curr. Opin. Colloid Interface Sci.* **8**(6), 459.

[224] **Michaud-Agrawal, N.**, **E. J. Denning**, **T. B. Woolf**, **O. Beckstein** (2011). MDAnalysis: A Toolkit for the Analysis of Molecular Dynamics Simulations. *J. Comput. Chem.* **32**(10), 2319.

[225] **Mills, T. T.**, **S. Tristram-Nagle**, **F. A. Heberle**, **N. F. Morales**, **J. Zhao**, **J. Wu**, **G. E. Toombes**, **J. F. Nagle**, **G. W. Feigenson** (2008). Liquid-Liquid Domains in Bilayers Detected by Wide Angle X-Ray Scattering. *Biophys. J.* **95**(2), 682.

[226] **Mim, C.**, **H. Cui**, **J. Gawronski-Salerno**, **A. Frost**, **E. Lyman**, **G. Voth**, **V. Unger** (2012). Structural Basis of Membrane Bending by the N-BAR Protein Endophilin. *Cell* **149**(1), 137.

[227] **Miyazawa, S.**, **R. L. Jernigan** (1996). Residue-Residue Potentials with a Favorable Contact Pair Term and an Unfavorable High Packing Density Term, for Simulation and Threading. *J. Mol. Biol.* **256**(3), 623.

[228] **Monticelli, L.**, **S. K. Kandasamy**, **X. Periole**, **R. G. Larson**, **D. P. Tieleman**, **S.-J. Marrink** (2008). The MARTINI Coarse-Grained Force Field: Extension to Proteins. *J. Chem. Theory Comput.* **4**(5), 819.

[229] **Moore, B. A.**, **H. H. Robinson**, **Z. Xu** (2007). The crystal structure of mouse Exo70 reveals unique features of the mammalian exocyst. *J. Mol. Biol.* **371**(2), 410.

[230] **Müller, M.**, **K. Katsov**, **M. Schick** (2003). Coarse-grained models and collective phenomena in membranes: Computer simulation of membrane fusion. *J. Polym. Sci. B Polym. Phys.* **41**(13), 1441.

[231] **Munson, M.**, **P. Novick** (2006). The exocyst defrocked, a framework of rods revealed. *Nat. Struct. Mol. Biol.* **13**(7), 577.

[232] **Nagle, J. F.**, **S. Tristram-Nagle** (2000). Structure of lipid bilayers. *Biochim. Biophys. Acta, Biomembr.* **1469**(3), 159.

[233] **Neri, M.**, **C. Anselmi**, **V. Carnevale**, **A. V. Vargiu**, **P. Carloni** (2006). Molecular dynamics simulations of outer-membrane protease T from E. coli based on a hybrid coarse-grained/atomistic potential. *J. Phys. Condens. Matter* **18**(14), S347.

[234] **Neri, M.**, **C. Anselmi**, **M. Cascella**, **A. Maritan**, **P. Carloni** (2005). Coarse-Grained Model of Proteins Incorporating Atomistic Detail of the Active Site. *Phys. Rev. Lett.* **95**(21), 218102.

[235] **Netz, R. R.**, **P. Pincus** (1995). Inhomogeneous fluid membranes: Segregation, ordering, and effective rigidity. *Phys. Rev. E* **52**(4), 4114.

[236] **Nguyen, H. D.**, **C. K. Hall** (2004). Molecular dynamics simulations of spontaneous fibril formation by random-coil peptides. *Proc. Natl. Acad. Sci. U.S.A.* **101**(46), 16180.

[237] **Nielsen, S. O.**, **B. Ensing**, **V. Ortiz**, **P. B. Moore**, **M. L. Klein** (2005). Lipid Bilayer Perturbations around a Transmembrane Nanotube: A Coarse Grain Molecular Dynamics Study. *Biophys. J.* **88**(6), 3822.

[238] **Nielsen, S. O.**, **C. F. Lopez**, **I. Ivanov**, **P. B. Moore**, **J. C. Shelley**, **M. L. Klein** (2004). Transmembrane Peptide-Induced Lipid Sorting and Mechanism of L_α-to-Inverted Phase Transition Using Coarse-Grain Molecular Dynamics. *Biophys. J.* **87**(4), 2107.

[239] **Niggemann, G.**, **M. Kummrow**, **W. Helfrich** (1995). The Bending Rigidity of Phosphatidylcholine Bilayers: Dependences on Experimental Method, Sample Cell Sealing and Temperature. *J. Phys. II France* **5**(3), 413.

[240] **Noid, W. G.**, **J.-W. Chu**, **G. S. Ayton**, **V. Krishna**, **S. Izvekov**, **G. A. Voth**, **A. Das**, **H. C. Andersen** (2008). The multiscale coarse-graining method. I. A rigorous bridge between atomistic and coarse-grained models. *J. Chem. Phys.* **128**(24), 244114.

[241] **Oh, H.-M.**, **M.-S. Kwon**, **H.-J. Kim**, **B.-H. Jeon**, **H.-R. Kim**, **H.-O. Choi**, **B.-R. Na**, **S.-H. Eom**, **N. W. Song**, **C.-D. Jun** (2011). Intermediate monomer-dimer equilibrium structure of native ICAM-1: Implication for enhanced cell adhesion. *Exp. Cell Res.* **317**(2), 163.

[242] **Ollila, O. H. S.**, **H. J. Risselada**, **M. Louhivuori**, **E. Lindahl**, **I. Vattulainen**, **S. J. Marrink** (2009). 3D Pressure Field in Lipid Membranes and Membrane-Protein Complexes. *Phys. Rev. Lett.* **102**(7), 078101.

[243] **Orsi, M.**, **D. Y. Haubertin**, **W. E. Sanderson**, **J. W. Essex** (2008). A Quantitative Coarse-Grain Model for Lipid Bilayers. *J. Phys. Chem. B* **112**(3), 802.

[244] **Ouberai, M. M.**, **J. Wang**, **M. J. Swann**, **C. Galvagnion**, **T. Guilliams**, **C. M. Dobson**, **M. E. Welland** (2013). α-Synuclein Senses Lipid Packing Defects and Induces Lateral Expansion of Lipids Leading to Membrane Remodelin. *J. Biol. Chem.* **288**(29), 20883.

[245] **Pantano, D. A.**, **P. B. Moore**, **M. L. Klein**, **D. E. Discher** (2011). Raft registration across bilayers in a molecularly detailed model. *Soft Matter* **7**(18), 8182.

[246] **Parton, D.**, **J. Klingelhoefer**, **M. Sansom** (2011). Aggregation of Model Membrane Proteins, Modulated by Hydrophobic Mismatch, Membrane Curvature, and Protein Class. *Biophys. J.* **101**(3), 691.

[247] **Periole, X.**, **M. Cavalli**, **S.-J. Marrink**, **M. A. Ceruso** (2009). Combining an Elastic Network With a Coarse-Grained Molecular Force Field: Structure, Dynamics, and Intermolecular Recognition. *J. Chem. Theory Comput.* **5**(9), 2531.

[248] **Periole, X.**, **T. Huber**, **S.-J. Marrink**, **T. P. Sakmar** (2007). G Protein-Coupled Receptors Self-Assemble in Dynamics Simulations of Model Bilayers. *J. Am. Chem. Soc.* **129**(33), 10126.

[249] **Perlmutter, J. D.**, **W. J. Drasler**, **W. Xie**, **J. Gao**, **J.-L. Popot**, **J. N. Sachs** (2011). All-Atom and Coarse-Grained Molecular Dynamics Simulations of a Membrane Protein Stabilizing Polymer. *Langmuir* **27**(17), 10523.

[250] **Perlmutter, J. D.**, **J. N. Sachs** (2009). Experimental verification of lipid bilayer structure through multi-scale modeling. *Biochim. Biophys. Acta, Biomembr.* **1788**(10), 2284.

[251] **Phillips, J. C.**, **R. Braun**, **W. Wang**, **J. Gumbart**, **E. Tajkhorshid**, **E. Villa**, **C. Chipot**, **R. D. Skeel**, **L. Kalé**, **K. Schulten** (2005). Scalable molecular dynamics with NAMD. *J. Comput. Chem.* **26**(16), 1781.

[252] **Piana, S.**, **K. Lindorff-Larsen**, **D. Shaw** (2011). How Robust Are Protein Folding Simulations with Respect to Force Field Parameterization? *Biophys. J.* **100**(9), L47.

[253] **Piggot, T. J.**, **Á. Piñeiro**, **S. Khalid** (2012). Molecular Dynamics Simulations of Phosphatidylcholine Membranes: A Comparative Force Field Study. *J. Chem. Theory Comput.* **8**(11), 4593.

[254] **Polyansky, A. A.**, **R. Ramaswamy**, **P. E. Volynsky**, **I. F. Sbalzarini**, **S. J. Marrink**, **R. G. Efremov** (2010). Antimicrobial Peptides Induce Growth of Phosphatidylglycerol Domains in a Model Bacterial Membrane. *J. Phys. Chem. Lett.* **1**(20), 3108.

[255] **Powathil, G. G.**, **K. E. Gordon**, **L. A. Hill**, **M. A. Chaplain** (2012). Modelling the effects of cell-cycle heterogeneity on the response of a solid tumour to chemotherapy: Biological insights from a hybrid multiscale cellular automaton model. *J. Theor. Biol.* **308**, 1.

[256] **Psachoulia, E.**, **P. W. Fowler**, **P. J. Bond**, **M. S. P. Sansom** (2008). Helix-Helix Interactions in Membrane Proteins: Coarse-Grained Simulations of Glycophorin A Helix Dimerization. *Biochemistry* **47**(40), 10503.

[257] **Radzicka, A.**, **R. Wolfenden** (1988). Comparing the polarities of the amino acids: side-chain distribution coefficients between the vapor phase, cyclohexane, 1-octanol, and neutral aqueous solution. *Biochemistry* **27**(5), 1664.

[258] **Ramachandran, P.**, **G. Varoquaux** (2011). Mayavi: 3D Visualization of Scientific Data. *Comput. Sci. Eng.* **13**(2), 40.

[259] **Ramadurai, S.**, **A. Holt**, **L. V. Schäfer**, **V. V. Krasnikov**, **D. T. Rijkers**, **S. J. Marrink**, **J. A. Killian**, **B. Poolman** (2010). Influence of Hydrophobic Mismatch and Amino Acid Composition on the Lateral Diffusion of Transmembrane Peptides. *Biophys. J.* **99**(5), 1447.

[260] **Ramakrishnan, N.**, **P. B. Sunil Kumar**, **J. H. Ipsen** (2010). Monte Carlo simulations of fluid vesicles with in-plane orientational ordering. *Phys. Rev. E* **81**(4), 041922.

[261] **Ramakrishnan, N.**, **P. Sunil Kumar**, **R. Radhakrishnan** (2014). Mesoscale computational studies of membrane bilayer remodeling by curvature-inducing proteins. *Phys. Rep.* **543**(1), 1.

[262] **Ramanan, V.**, **N. J. Agrawal**, **J. Liu**, **S. Engles**, **R. Toy**, **R. Radhakrishnan** (2011). Systems biology and physical biology of clathrin-mediated endocytosis. *Integrative Biology* **3**(8), 803.

[263] **Rand, R.**, **N. Fuller** (1994). Structural dimensions and their changes in a reentrant hexagonal-lamellar transition of phospholipids. *Biophys. J.* **66**(6), 2127.

[264] **Rawicz, W.**, **K. Olbrich**, **T. McIntosh**, **D. Needham**, **E. Evans** (2000). Effect of chain length and unsaturation on elasticity of lipid bilayers. *Biophys. J.* **79**(1), 328.

[265] **Ren, J.**, **W. Guo** (2012). ERK1/2 Regulate Exocytosis through Direct Phosphorylation of the Exocyst Component Exo70. *Dev. Cell* **22**(5), 967.

[266] **Resendis-Antonio, O.**, **C. González-Torres**, **G. Jaime-Muñoz**, **C. E. Hernandez-Patiño**, **C. F. Salgado-Muñoz** (2015). Modeling metabolism: A window toward a comprehensive interpretation of networks in cancer. *Semin. Cancer Biol.* **30**, 79.

[267] **Reynwar, B. J.**, **G. Illya**, **V. A. Harmandaris**, **M. M. Muller**, **K. Kremer**, **M. Deserno** (2007). Aggregation and vesiculation of membrane proteins by curvature-mediated interactions. *Nature* **447**(7143), 461.

[268] **Ribeiro, A. A.**, **R. B. de Alencastro** (2013). Mixed Monte Carlo-Molecular Dynamics Simulations of the Prion Protein. *J. Mol. Graphics Modell.* (0), In Press.

[269] **Risselada, H. J.**, **S. J. Marrink** (2008). The molecular face of lipid rafts in model membranes. *Proc. Natl. Acad. Sci. U.S.A.* **105**(45), 17367.

[270] **Rocha-Perugini, V.**, **M. Gordon-Alonso**, **F. SÃąnchez-Madrid** (2014). PIP2: choreographer of actin-adaptor proteins in the HIV-1 dance. *Trends Microbiol.* **22**(7), 379.

[271] **Rosse, C.**, **A. Hatzoglou**, **M. C. Parrini**, **M. A. White**, **P. Chavrier**, **J. Camonis** (2006). RalB mobilizes the exocyst to drive cell migration. *Mol. Biol. Cell* **26**(2), 727.

[272] **Rowe, E. S.**, **F. Zhang**, **T. W. Leung**, **J. S. Parr**, **P. T. Guy** (1998). Thermodynamics of Membrane Partitioning for a Series of n-Alcohols Determined by Titration Calorimetry: Role of Hydrophobic Effects. *Biochemistry* **37**(8), 2430.

[273] **Rozovsky, S.**, **M. B. Forstner**, **H. Sondermann**, **J. T. Groves** (2012). Single Molecule Kinetics of ENTH Binding to Lipid Membranes. *J. Phys. Chem. B* **116**(17), 5122.

[274] **Ryu, Y.-S.**, **I.-H. Lee**, **J.-H. Suh**, **S. C. Park**, **S. Oh**, **L. R. Jordan**, **N. J. Wittenberg**, **S.-H. Oh**, **N. L. Jeon**, **B. Lee**, **A. N. Parikh**, **S.-D. Lee** (2014). Reconstituting ring-rafts in bud-mimicking topography of model membranes. *Nat. Commun.* **5**, .

[275] **Rzepiela, A. J.**, **M. Louhivuori**, **C. Peter**, **S. J. Marrink** (2011). Hybrid simulations: combining atomistic and coarse-grained force fields using virtual sites. *Physical Chemistry Chemical Physics* **13**(22), 10437.

[276] **Rzepiela, A. J.**, **L. V. SchÃďfer**, **N. Goga**, **H. J. Risselada**, **A. H. De Vries**, **S. J. Marrink** (2010). Reconstruction of atomistic details from coarse-grained structures. *J. Comput. Chem.* **31**(6), 1333.

[277] **Sadiq, S.**, **R. Guixa-Gonzalez**, **E. Dainese**, **M. Pastor**, **G. De Fabritiis**, **J. Selent** (2013). Molecular Modeling and Simulation of Membrane Lipid-Mediated Effects on GPCRs. *Curr. Med. Chem.* **20**(1), 22.

[278] **Saleem, M.**, **S. Morlot**, **A. Hohendahl**, **J. Manzi**, **M. Lenz**, **A. Roux** (2015). A balance between membrane elasticity and polymerization energy sets the shape of spherical clathrin coats. *Nat. Commun.* **6**, .

[279] **Salsbury Jr., F. R.** (2010). Molecular dynamics simulations of protein dynamics and their relevance to drug discovery. *Curr. Opin. Pharmacol.* **10**(6), 738.

[280] **Sapienza, P. J.**, **A. L. Lee** (2010). Using NMR to study fast dynamics in proteins: methods and applications. *Curr. Opin. Pharmacol.* **10**(6), 723.

[281] **Schäfer, L. V.**, **S. J. Marrink** (2010). Partitioning of Lipids at Domain Boundaries in Model Membranes. *Biophys. J.* **99**(12), L91.

[282] **Schatzberg, P.** (1963). Solubilities of Water in Several Normal Alkanes from C7 to C161. *J. Phys. Chem.* **67**(4), 776.

[283] **Scheraga, H. A.**, **M. Khalili**, **A. Liwo** (2007). Protein-Folding Dynamics: Overview of Molecular Simulation Techniques. *Annu. Rev. Phys. Chem.* **58**(1), 57.

[284] **Schmick, M.**, **P. Bastiaens** (2014). The Interdependence of Membrane Shape and Cellular Signal Processing. *Cell* **156**(6), 1132.

[285] **Scita, G.**, **S. Confalonieri**, **P. Lappalainen**, **S. Suetsugu** (2008). IRSp53: crossing the road of membrane and actin dynamics in the formation of membrane protrusions. *Trends Cell Biol.* **18**(2), 52.

[286] **Seddon, J. M.**, **A. M. Squires**, **C. E. Conn**, **O. Ces**, **A. J. Heron**, **X. Mulet**, **G. C. Shearman**, **R. H. Templer** (2006). Pressure-jump X-ray studies of liquid crystal transitions in lipids. *Philos. Trans. R. Soc. London, Ser. A* **364**(1847), 2635.

[287] **Seelig, A.**, **J. Seelig** (1974). Dynamic structure of fatty acyl chains in a phospholipid bilayer measured by deuterium magnetic resonance. *Biochemistry* **13**(23), 4839.

[288] **Seifert, U.** (1997). Configurations of fluid membranes and vesicles. *Adv. Phys.* **46**(1), 13.

[289] **Sengupta, D.**, **S. J. Marrink** (2010). Lipid-mediated interactions tune the association of glycophorin A helix and its disruptive mutants in membranes. *Phys. Chem. Chem. Phys.* **12**(40), 12987.

[290] **Seo, M.**, **S. Rauscher**, **R. Pomès**, **D. P. Tieleman** (2012). Improving Internal Peptide Dynamics in the Coarse-Grained MARTINI Model: Toward Large-Scale Simulations of Amyloid- and Elastin-like Peptides. *J. Chem. Theory Comput.* **8**(5), 1774.

[291] **Sergei Izvekov, C. J. B., Michele Parrinello**, **G. A. Voth** (2004). Effective force fields for condensed phase systems from ab initio molecular dynamics simulation: A new method for force-matching. *J. Chem. Phys.* **120**, 10896.

[292] **Sheetz, M. P.** (2001). Cell control by membrane-cytoskeleton adhesion. *Nat. Rev. Mol. Cell Biol.* **2**(5), 392.

[293] **Shell, M. S.** (2008). The relative entropy is fundamental to multiscale and inverse thermodynamic problems. *J. Chem. Phys.* **129**(14), 144108.

[294] **Shelley, J. C.**, **M. Y. Shelley** (2000). Computer simulation of surfactant solutions. *Curr. Opin. Colloid Interface Sci.* **5**(1–2), 101.

[295] **Shelley, J. C.**, **M. Y. Shelley**, **R. C. Reeder**, **S. Bandyopadhyay**, **M. L. Klein** (2001). A Coarse Grain Model for Phospholipid Simulations. *J. Phys. Chem. B* **105**(19), 4464.

[296] **Shi, J.**, **A. R. Votruba**, **O. C. Farokhzad**, **R. Langer** (2010). Nanotechnology in Drug Delivery and Tissue Engineering: From Discovery to Applications. *Nano Lett.* **10**(9), 3223.

[297] **Shi, Z.**, **T. Baumgart** (2015). Membrane tension and peripheral protein density mediate membrane shape transitions. *Nat. Commun.* **6**, .

[298] **Shih, A. Y.**, **A. Arkhipov**, **P. L. Freddolino**, **K. Schulten** (2006). Coarse Grained Protein-Lipid Model with Application to Lipoprotein Particles. *J. Phys. Chem. B* **110**(8), 3674.

[299] **Shih, A. Y.**, **P. L. Freddolino**, **A. Arkhipov**, **K. Schulten** (2007). Assembly of lipoprotein particles revealed by coarse-grained molecular dynamics simulations. *J. Struct. Biol.* **157**(3), 579.

[300] **Shillcock, J.**, Vesicles and Vesicle Fusion: Coarse-Grained Simulations. *Methods in Molecular Biology* (edited by **L. Monticelli**, **E. Salonen**), volume 924 (Humana Press, 2013), pp. 659–697.

[301] **Shinoda, W.**, **R. DeVane**, **M. L. Klein** (2010). Zwitterionic Lipid Assemblies: Molecular Dynamics Studies of Monolayers, Bilayers, and Vesicles Using a New Coarse Grain Force Field. *J. Phys. Chem. B* **114**(20), 6836.

[302] **Shirts, M.**, **J. Chodera** (2008). Statistically optimal analysis of samples from multiple equilibrium states. *J. Chem. Phys.* **129**(12), 124105.

[303] **Siegel, D.**, **R. Epand** (1997). The mechanism of lamellar-to-inverted hexagonal phase transitions in phosphatidylethanolamine: implications for membrane fusion mechanisms. *Biophys. J.* **73**(6), 3089.

[304] **Simunovic, M.**, **A. Srivastava**, **G. A. Voth** (2013). Linear aggregation of proteins on the membrane as a prelude to membrane remodeling. *Proc. Natl. Acad. Sci. U. S. A.* **110**(51), 20396.

[305] **Singh, G.**, **D. P. Tieleman** (2011). Using the Wimley-White Hydrophobicity Scale as a Direct Quantitative Test of Force Fields: The MARTINI Coarse-Grained Model. *J. Chem. Theory Comput.* **7**(7), 2316.

[306] **Slochower, D. R.**, **P. J. Huwe**, **R. Radhakrishnan**, **P. A. Janmey** (2013). Quantum and All-Atom Molecular Dynamics Simulations of Protonation and Divalent Ion Binding to Phosphatidylinositol 4,5-Bisphosphate (PIP(2)). *J. Phys. Chem. B* **117**(28), 8322.

[307] **Slochower, D. R.**, **Y.-H. Wang**, **R. Radhakrishnan**, **P. A. Janmey** (2015). Physical chemistry and membrane properties of two phosphatidylinositol bisphosphate isomers(). *Phys. Chem. Chem. Phys.* **17**(19), 12608.

[308] **Slochower, D. R.**, **Y.-H. Wang**, **R. W. Tourdot**, **R. Radhakrishnan**, **P. A. Janmey** (2014). Counterion-mediated pattern formation in membranes containing anionic lipids. *Adv. Colloid Interface Sci.* **208**, 177.

[309] **Smirnova, Y. G.**, **S.-J. Marrink**, **R. Lipowsky**, **V. Knecht** (2010). Solvent-Exposed Tails as Prestalk Transition States for Membrane Fusion at Low Hydration. *J. Am. Chem. Soc.* **132**(19), 6710.

[310] **Smit, B.**, **P. A. J. Hilbers**, **K. Esselink**, **L. A. M. Rupert**, **N. M. van Os**, **A. G. Schlijper** (1990). Computer simulations of a water/oil interface in the presence of micelles. *Nature* **348**(6302), 624.

[311] **Snow, C. D.**, **H. Nguyen**, **V. S. Pande**, **M. Gruebele** (2002). Absolute comparison of simulated and experimental protein-folding dynamics. *Nature* **420**(6911), 102.

[312] **Sodt, A. J.**, **T. Head-Gordon** (2010). An implicit solvent coarse-grained lipid model with correct stress profile. *J. Chem. Phys.* **132**(20), 205103.

[313] **Sommer, B.** (2013). Membrane Packing Problems: A short Review on computational Membrane Modeling Methods and Tools. *Comput. Struct. Biotechnol. J.* **5**, e201302014.

[314] **Sonne, J.**, **F. Y. Hansen**, **G. H. Peters** (2005). Methodological problems in pressure profile calculations for lipid bilayers. *J. Chem. Phys.* **122**, 124903.

[315] **Sonne, J.**, **M. Ø. Jensen**, **F. Y. Hansen**, **L. Hemmingsen**, **G. H. Peters** (2007). Reparameterization of All-Atom Dipalmitoylphosphatidylcholine Lipid Parameters Enables Simulation of Fluid Bilayers at Zero Tension. *Biophys. J.* **92**(12), 4157.

[316] **Sorre, B.**, **A. Callan-Jones**, **J. Manzi**, **B. Goud**, **J. Prost**, **P. Bassereau**, **A. Roux** (2012). Nature of curvature coupling of amphiphysin with membranes depends on its bound density. *Proc. Natl. Acad. Sci. U. S. A.* **109**(1), 173.

[317] **Sousa, S. F.**, **P. A. Fernandes**, **M. J. Ramos** (2006). Protein-ligand docking: Current status and future challenges. *Proteins* **65**(1), 15.

[318] **Spiczka, K. S.**, **C. Yeaman** (2008). Ral-regulated interaction between Sec5 and paxillin targets Exocyst to focal complexes during cell migration. *J. Cell. Sci.* **121**(17), 2880.

[319] **Srinivas, G.**, **M. L. Klein** (2004). Computational approaches to nanobiotechnology: probing the interaction of synthetic molecules with phospholipid bilayers via a coarse grain model. *Nanotechnology* **15**(9), 1289.

[320] **Sriswasdi, S.**, **S. L. Harper**, **H.-Y. Tang**, **P. G. Gallagher**, **D. W. Speicher** (2014). Probing large conformational rearrangements in wild-type and mutant spectrin using structural mass spectrometry. *Proc. Natl. Acad. Sci. U. S. A.* **111**(5), 1801.

[321] **Sriswasdi, S.**, **S. L. Harper**, **H.-Y. Tang**, **D. W. Speicher** (2014). Enhanced Identification of Zero-Length Chemical Cross-Links Using Label-Free Quantitation and High-Resolution Fragment Ion Spectra. *J. Proteome Res.* **13**(2), 898.

[322] **Srivastava, A.**, **G. A. Voth** (2012). Hybrid Approach for Highly Coarse-Grained Lipid Bilayer Models. *J. Chem. Theory Comput.* **9**(1), 750.

[323] **Stachowiak, J. C.**, **F. M. Brodsky**, **E. A. Miller** (2013). A cost-benefit analysis of the physical mechanisms of membrane curvature. *Nat. Cell Biol.* **15**(9), 1019.

[324] **Stachowiak, J. C.**, **E. M. Schmid**, **C. J. Ryan**, **H. S. Ann**, **D. Y. Sasaki**, **M. B. Sherman**, **P. L. Geissler**, **D. A. Fletcher**, **C. C. Hayden** (2012). Membrane bending by protein-protein crowding. *Nat. Cell Biol.* **14**(9), 944.

[325] **Sugita, Y.**, **Y. Okamoto** (1999). Replica-exchange molecular dynamics method for protein folding. *Chem. Phys. Lett.* **314**(1–2), 141.

[326] **Sumagin, R.**, **I. H. Sarelius** (2007). A role for ICAM-1 in maintenance of leukocyte-endothelial cell rolling interactions in inflamed arterioles. *Am. J. Physiol. Heart Circ. Physiol.* **293**(5), H2786.

[327] **Tang, Y.**, **J. Yoo**, **A. Yethiraj**, **Q. Cui**, **X. Chen** (2008). Mechanosensitive Channels: Insights from Continuum-Based Simulations **52**(1), 1.

[328] **Thapa, N.**, **Y. Sun**, **M. Schramp**, **S. Choi**, **K. Ling**, **R. A. Anderson** (2012). Phosphoinositide Signaling Regulates the Exocyst Complex and Polarized Integrin Trafficking in Directionally Migrating Cells. *Dev. Cell* **22**(1), 116.

[329] **Thøgersen, L.**, **B. Schiøtt**, **T. Vosegaard**, **N. C. Nielsen**, **E. Tajkhorshid** (2008). Peptide Aggregation and Pore Formation in a Lipid Bilayer: A Combined Coarse-Grained and All Atom Molecular Dynamics Study. *Biophys. J.* **95**(9), 4337.

[330] **Thorpe, I. F.**, **D. P. Goldenberg**, **G. A. Voth** (2011). Exploration of Transferability in Multiscale Coarse-Grained Peptide Models. *J. Phys. Chem. B* **115**(41), 11911.

[331] **Tian, A.**, **T. Baumgart** (2009). Sorting of lipids and proteins in membrane curvature gradients. *Biophys. J.* **96**(7), 2676.

[332] **Tian, A.**, **C. Johnson**, **W. Wang**, **T. Baumgart** (2007). Line Tension at Fluid Membrane Domain Boundaries Measured by Micropipette Aspiration. *Phys. Rev. Lett.* **98**(20), 208102.

[333] **Tourdot, R. W.**, **R. P. Bradley**, **N. Ramakrishnan**, **R. Radhakrishnan** (2014). Multiscale computational models in physical systems biology of intracellular trafficking. *IET Syst. Biol.* **8**(5), 198.

[334] **Tourdot, R. W.**, **N. Ramakrishnan**, **T. Baumgart**, **R. Radhakrishnan** (2015). Application of a free-energy-landscape approach to study tension-dependent bilayer tubulation mediated by curvature-inducing proteins. *Phys. Rev. E* **92**(4), 042715.

[335] **Tourdot, R. W.**, **N. Ramakrishnan**, **R. Radhakrishnan** (2014). Defining the free-energy landscape of curvature-inducing proteins on membrane bilayers. *Phys. Rev. E* **90**(2), 022717.

[336] **Treptow, W.**, **S.-J. Marrink**, **M. Tarek** (2008). Gating Motions in Voltage-Gated Potassium Channels Revealed by Coarse-Grained Molecular Dynamics Simulations. *J. Phys. Chem. B* **112**(11), 3277.

[337] **Tsujita, K.**, **T. Takenawa**, **T. Itoh** (2015). Feedback regulation between plasma membrane tension and membrane-bending proteins organizes cell polarity during leading edge formation. *Nat. Cell Biol.* **17**, 749.

[338] **Ulmschneider, J. P.**, **M. Andersson**, **M. B. Ulmschneider** (2011). Determining Peptide Partitioning Properties via Computer Simulation. *J. Membr. Biol.* **239**(1-2), 15.

[339] **Ulmschneider, M. B.**, **J. P. F. Doux**, **J. A. Killian**, **J. C. Smith**, **J. P. Ulmschneider** (2010). Mechanism and Kinetics of Peptide Partitioning into Membranes from All-Atom Simulations of Thermostable Peptides. *J. Am. Chem. Soc.* **132**(10), 3452.

[340] **van den Bogaart, G.**, **K. Meyenberg**, **H. J. Risselada**, **H. Amin**, **K. I. Willig**, **B. E. Hubrich**, **M. Dier**, **S. W. Hell**, **H. Grubmueller**, **U. Diederichsen**, **R. Jahn** (2011). Membrane protein sequestering by ionic protein-lipid interactions. *Nature* **479**(7374), 552.

[341] **van der Spoel, D.**, **E. Lindahl**, **B. Hess**, **A. van Buuren**, **E. Apol**, **P. Meulenhoff**, **D. Tieleman**, **A. Sijbers**, **K. Feenstra**, **R. van Drunen**, **H. Berendsen**, *Gromacs User Manual version 4.0* (2005).

[342] **van der Spoel, D.**, **E. Lindahl**, **B. Hess**, **G. Groenhof**, **A. E. Mark**, **H. J. C. Berendsen** (2005). `GROMACS`: Fast, flexible, and free. *J. Comput. Chem.* **26**(16), 1701.

[343] **van der Walt, S.**, **S. Colbert**, **G. Varoquaux** (2011). The NumPy Array: A Structure for Efficient Numerical Computation. *Comput. Sci. Eng.* **13**(2), 22.

[344] **van Gunsteren, W. F.**, **J. Dolenc**, **A. E. Mark** (2008). Molecular simulation as an aid to experimentalists. *Curr. Opin. Struct. Biol.* **18**(2), 149.

[345] **van Meer, G.**, **A. I. P. M. de Kroon** (2010). Lipid map of the mammalian cell. *J. Cell Sci.* **124**(1), 5.

[346] **van Meer, G.**, **D. R. Voelker**, **G. W. Feigenson** (2008). Membrane lipids: where they are and how they behave. *Nat. Rev. Mol. Cell Biol.* **9**(2), 112.

[347] **Van Rossum, G.**, **F. L. Drake Jr**, *Python reference manual* (Centrum voor Wiskunde en Informatica Amsterdam, 1995).

[348] **Venturoli, M.** (2005). Simulation Studies of Protein-Induced Bilayer Deformations, and Lipid-Induced Protein Tilting, on a Mesoscopic Model for Lipid Bilayers with Embedded Proteins. *Biophys. J.* **88**, 1778.

[349] **Venturoli, M.**, **M. Maddalena Sperotto**, **M. Kranenburg**, **B. Smit** (2006). Mesoscopic models of biological membranes. *Phys. Rep.* **437**(1–2), 1.

[350] **Venturoli, M.**, **B. Smit** (1999). Simulating the self-assembly of model membranes. *PhysChemComm* **2**, 45.

[351] **Villa, A.**, **C. Peter**, **N. F. A. van der Vegt** (2009). Self-assembling dipeptides: conformational sampling in solvent-free coarse-grained simulation. *Physical Chemistry Chemical Physics (Incorporating Faraday Transactions)* **11**, 2077.

[352] **Viswanadhan, V. N.**, **A. K. Ghose**, **U. C. Singh**, **J. J. Wendoloski** (1999). Prediction of Solvation Free Energies of Small Organic Molecules Additive-Constitutive Models Based on Molecular Fingerprints and Atomic Constants. *J. Chem. Inf. Comput. Sci.* **39**(2), 405.

[353] **Šali, A.**, **T. L. Blundell** (1993). Comparative Protein Modelling by Satisfaction of Spatial Restraints. *J. Mol. Biol.* **234**(3), 779.

[354] **Vuorela, T.**, **A. Catte**, **P. S. Niemelä**, **A. Hall**, **M. T. Hyvönen**, **S.-J. Marrink**, **M. Karttunen**, **I. Vattulainen** (2010). Role of Lipids in Spheroidal High Density Lipoproteins. *PLoS Comput. Biol.* **6**(10), e1000964.

[355] **Wallace, E. J.**, **M. S. P. Sansom** (2008). Blocking of Carbon Nanotube Based Nanoinjectors by Lipids: A Simulation Study. *Nano Lett.* **8**(9), 2751.

[356] **Wang, J.**, **R. M. Wolf**, **J. W. Caldwell**, **P. A. Kollman**, **D. A. Case** (2004). Development and testing of a general amber force field. *J. Comput. Chem.* **25**(9), 1157.

[357] **Wang, Y.-H.**, **A. Collins**, **L. Guo**, **K. B. Smith-Dupont**, **F. Gai**, **T. Svitkina**, **P. A. Janmey** (2012). Divalent Cation-Induced Cluster Formation by Polyphosphoinositides in Model Membranes. *J. Am. Chem. Soc.* **134**(7), 3387.

[358] **Wang, Z.**, **J. D. Butner**, **R. Kerketta**, **V. Cristini**, **T. S. Deisboeck** (2015). Simulating cancer growth with multiscale agent-based modeling. *Semin. Cancer Biol.* **30**, 70.

[359] **Wee, C. L.**, **K. Balali-Mood**, **D. Gavaghan**, **M. S. Sansom** (2008). The Interaction of Phospholipase A2 with a Phospholipid Bilayer: Coarse-Grained Molecular Dynamics Simulations. *Biophys. J.* **95**(4), 1649.

[360] **Werner, T.**, **M. B. Morris**, **S. Dastmalchi**, **W. B. Church** (2012). Structural modelling and dynamics of proteins for insights into drug interactions. *Adv. Drug Delivery Rev.* **64**(4), 323.

[361] **West, B.**, **F. L. Brown**, **F. Schmid** (2009). Membrane-Protein Interactions in a Generic Coarse-Grained Model for Lipid Bilayers. *Biophys. J.* **96**(1), 101.

[362] **Wimley, W. C.**, **S. H. White** (1996). Experimentally determined hydrophobicity scale for proteins at membrane interfaces. *Nat. Struct. Mol. Biol.* **3**(10), 842.

[363] **Wolfenden, R.**, **L. Andersson**, **P. M. Cullis**, **C. C. B. Southgate** (1981). Affinities of amino acid side chains for solvent water. *Biochemistry* **20**(4), 849.

[364] **Wolfram Research, Inc.** (2010), Mathematica 8.0.

[365] **Wong-Ekkabut, J.**, **S. Baoukina**, **W. Triampo**, **I.-M. Tang**, **D. P. Tieleman**, **L. Monticelli** (2008). Computer simulation study of fullerene translocation through lipid membranes. *Nat. Nano.* **3**(6), 363.

[366] **Yang, C.**, **M. Hoelzle**, **A. Disanza**, **G. Scita**, **T. Svitkina** (2009). Coordination of Membrane and Actin Cytoskeleton Dynamics during Filopodia Protrusion. *PLoS One* **4**(5), e5678.

[367] **Yang, L.**, **L. Ding**, **H. W. Huang** (2003). New Phases of Phospholipids and Implications to the Membrane Fusion Problem. *Biochemistry* **42**(22), 6631.

[368] **Yang, Y. T.**, **C. D. Jun**, **J. H. Liu**, **R. G. Zhang**, **A. Joachimiak**, **T. A. Springer**, **J. H. Wang** (2004). Structural basis for dimerization of ICAM-1 on the cell surface. *Mol.Cell* **14**(2), 269.

[369] **Yarar, D.**, **C. M. Waterman-Storer**, **S. L. Schmid** (2005). A Dynamic Actin Cytoskeleton Functions at Multiple Stages of Clathrin-mediated Endocytosis. *Mol. Biol. Cell* **16**(2), 964.

[370] **Yefimov, S.**, **E. van der Giessen**, **P. R. Onck**, **S. J. Marrink** (2008). Mechanosensitive Membrane Channels in Action. *Biophys. J.* **94**(8), 2994.

[371] **Yesylevskyy, S.**, **L. Schäfer**, **D. Sengupta**, **S. Marrink** (2010). Polarizable water model for the coarse-grained MARTINI force field. *PLoS Comput. Biol.* **6**(6), e1000810.

[372] **Yin, F.**, **J. T. Kindt** (2010). Atomistic Simulation of Hydrophobic Matching Effects on Lipid Composition near a Helical Peptide Embedded in Mixed-Lipid Bilayers. *J. Phys. Chem. B* **114**(24), 8076.

[373] **Yin, Y.**, **A. Arkhipov**, **K. Schulten** (2009). Simulations of Membrane Tubulation by Lattices of Amphiphysin N-BAR Domains. *Structure* **17**(6), 882.

[374] **Yu, H.**, **K. Schulten** (2013). Membrane Sculpting by F-BAR Domains Studied by Molecular Dynamics Simulations. *PLoS Comput. Biol.* **9**(1), e1002892.

[375] **Yu, H.-Y.**, **D. M. Eckmann**, **P. S. Ayyaswamy**, **R. Radhakrishnan** (2015). Composite generalized Langevin equation for Brownian motion in different hydrodynamic and adhesion regimes. *Phys. Rev. E* **91**(5), 052303.

[376] **Zhang, L.**, **Y. Sun** (2010). Molecular simulation of adsorption and its implications to protein chromatography: A review. *Biochem. Eng. J.* **48**(3), 408.

[377] **Zhang, L.**, **S. Bouguet-Bonnet**, **M. Buck**, Combining NMR and Molecular Dynamics Studies for Insights into the Allostery of Small GTPase-Protein Interactions. *Methods in Molecular Biology* (edited by **A. W. Fenton**), volume 796 (Springer New York, 2012), pp. 235–259–.

[378] **Zhang, Z.**, **J. Pfaendtner**, **A. Grafmüller**, **G. A. Voth** (2009). Defining Coarse-Grained Representations of Large Biomolecules and Biomolecular Complexes from Elastic Network Models. *Biophys. J.* **97**, 2327.

[379] **Zhao, H.**, **A. Pykäläinen**, **P. Lappalainen** (2011). I-BAR domain proteins: linking actin and plasma membrane dynamics. *Curr. Opin. Cell Biol.* **23**(1), 14.

[380] **Zhao, Y.**, **J. Liu**, **C. Yang**, **B. R. Capraro**, **T. Baumgart**, **R. P. Bradley**, **N. Ramakrishnan**, **X. Xu**, **R. Radhakrishnan**, **T. Svitkina**, **W. Guo** (2013). Exo70 Generates Membrane Curvature for Morphogenesis and Cell Migration. *Dev. Cell* **26**(3), 266.

[381] **Zimmerberg, J.**, **M. M. Kozlov** (2006). How proteins produce cellular membrane curvature. *Nat. Rev. Mol. Cell Biol.* **7**(1), 9.

[382] **Zuo, X.**, **J. Zhang**, **Y. Zhang**, **S.-C. Hsu**, **D. Zhou**, **W. Guo** (2006). Exo70 interacts with the Arp2/3 complex and regulates cell migration. *Nat.Cell Biol.* **8**(12), 1383.

[383] **Zwier, M. C.**, **L. T. Chong** (2010). Reaching biological timescales with all-atom molecular dynamics simulations. *Curr. Opin. Pharmacol.* **10**(6), 745.