UNIVERSIDADE DE LISBOA

FACULDADE DE CIÊNCIAS

Molecular modeling and simulation of pH effects in lipid bilayers

Doutoramento em Bioquímica

Bioquímica Teórica

Diogo Ruivo dos Santos Vila Viçosa

Tese orientada por:

Doutor Miguel Machuqueiro e Professora Doutora Maria José Calhorda

Documento especialmente elaborado para a obtenção do grau de doutor

2015

Acknowledgments

My first acknowledgment goes to my supervisors: Dr. Miguel Machuqueiro and Prof. Maria José Calhorda. To Miguel for his friendship, for the patience in all moments I did not agree with him, and for killing 99% of my (bad) ideas. To Prof. Calhorda for involving me in many scientific collaborations, and for all the patience she showed me in explaining many basic concepts of inorganic chemistry that I probably should have learned in my degree.

Dr. António M. Baptista, I would like to thank for all the theoretical support and so many useful discussions of my papers and ideas. Vitor H. Teixeira, who was present during these four years, I thank for being a huge encyclopedia of bugs and errors of all softwares used in our group.

For the long range support, I also thank Paulo J. Costa, Luís C. S. Filipe, João M. Damas, and João Henriques. Rafael Nunes I thank all times he said "Yes, I understood", after a long and boring explanation from my part. Tiago Silva and Andreia Valente, I thank for all the coffee breaks and lunches during the last four years. Pedro B. P. S. Reis, I thank for being the guy who rebuilt the cluster instead of me during the writing of this thesis.

A very special acknowledgment goes to Sofia Carvalho for all the emotional support, for marrying me (she is probably as crazy as I am), and for being always right beside me.

This work would not be possible without the kind welcome of the Inorganic and Theoretical Chemistry group from Centro de Química e Bioquímica and the PhD program in biochemistry of Departamento de Química e Bioquímica. I also acknowledge the financial support from Fundação para a Ciência e Tecnologia through grant SFRH/BD/81017/2011.

I also want to address two more special acknowledgments that I prefer to write in Portuguese:

Aos meus pais, António e Lucinda Vila Viçosa, quero deixar um agradecimento muito especial por toda a educação que me deram e que me permitiu chegar a um ponto da minha vida em que me sinto extremamente realizado e capaz de continuar o meu projecto de vida.

Ao Grupo 23 de Queluz e ao conjunto de dirigentes do mesmo, quero agradecer

todos os momentos felizes que me proporcionaram ao longo destes dezassete anos que me permitiram manter o espírito e a "pica" sempre em alta, sem que nada fosse deixado para trás. Ao grupo e a todos os dirigentes e jovens com quem trabalhei e continuarei a trabalhar, o meu muito obrigado.

Finally, I thank all my friends and family for trying to understand what I really do when I go to the university.

Preface

This PhD was performed in the Inorganic and Theoretical Chemistry group of *Centro de Química e Bioquímica* and I was supervised by Doctor Miguel Machuqueiro and Professor Maria José Calhorda. They are experts in two different fields (molecular modeling and quantum chemistry, respectively) of computational (bio)chemistry which was of crucial importance to my evolution as a scientist. During these four years I was able to learn and implement several theoretical / computational methods used in (bio)chemistry. Molecular mechanics and dynamics, continuum electrostatics and Monte Carlo simulations were the techniques / models used more in this thesis. However, I was also able to learn several aspects of quantum chemistry such as density functional theory, reaction mechanisms, etc.

This thesis

The original idea of this PhD project was to develop a method to simulate the pH effects in biological membranes, in the first year, and then apply it to several biological systems. However, we quickly found out that the development of such methodology would not be so straightforward as initially planned, and we ended up addressing the problem in a stepwise way. Firstly, we needed to solve the issue regarding periodic boundary conditions in Poisson–Boltzmann calculations that were not implemented in any of the available PB-solver softwares (chapter 3). Having solved this problem, we used our approach to calculate the pK_{a} of a dimyristoyl phosphatidylcholine (DMPC) molecule in the bilayer environment:

V. H. Teixeira, D. Vila-Viçosa, A. M. Baptista, and M. Machuqueiro. Protonation of DMPC in a bilayer environment using a linear response approximation. *J. Chem. Theory Comput.*, 10:2176–2184, 2014

Next, we came across another problem regarding the treatment of ionic strength in simulations of significantly charged bilayers (chapter 4). We developed a Poisson–Boltzmann based method to determine the number of explicit ions to be added in those simulations:

D. Vila-Viçosa, V. H. Teixeira, H. A. F. Santos, A. M. Baptista, and M. Machuqueiro. Treatment of ionic strength in biomolecular simulations of charged lipid bilayers. *J. Chem. Theory Comput.*, 10:5483–5492, 2014

We then used the newly developed approaches in the context of the stochastic titration constant-pH molecular dynamics method to simulate the pH effect in lipid bilayers (CpHMD-L) (chapter 5):

D. Vila-Viçosa, V. H. Teixeira, A. M. Baptista, and M. Machuqueiro. Constant-pH MD simulations of an oleic acid bilayer. *J. Chem. Theory Comput.*, 11(5):2367–2376, 2015

Finally, in the beginning of my fourth year, we started to apply the CpHMD-L method to several systems (chapter 6): i) an anionic lipid bilayer; ii) pK_{a} calculations of model peptides in a bilayer environment; and iii) (de)protonation of pHLIP, a peptide that inserts into a lipid bilayer at low pH. These three works are currently being prepared for publication.

Other publications

As mentioned above, during these four years, I was able to learn several methodologies that albeit not essential to this thesis work, allowed me to improve as scientist and to establish several collaborations, both from theoretical and experimental fields.

Quantum mechanics

My first contact with quantum mechanics was before this PhD, through a collaboration with an experimental group, where we tried to understand how certain inorganic molecules were packed in a crystal. Using quantum chemistry methods, we optimized their geometry and compared their final energies:

M. Đaković, D. Vila-Viçosa, M. J. Calhorda, and Z. Popović. Coordination-driven self-assembly of thiocyanate complexes of Co(II), Ni(II) and Cu(II) with picolinamide: a structural and DFT study. *Cryst. Eng. Comm.*, 13(19):5863–5871, 2011

The continuation of this collaboration led to another work where we determined the reaction pathway leading to the experimentally observed products:

M. Đaković, D. Vila-Viçosa, N. A. G. Bandeira, M. J. Calhorda, B. Kozlevčar,

Z. Jagličić, and Z. Popović. Can self-assembly of copper(II) picolinamide building blocks be controlled? *Cryst. Eng. Comm.*, 15(40):8074–8087, 2013

A different collaboration was established with a group from Strasbourg, in which we calculated the reaction profile of the ring opening polymerization of a lactide catalyzed by an inorganic complex with a binuclear center of Zinc(II):

C. Fliedel, D. Vila-Viçosa, M. J. Calhorda, S. Dagorne, and T. Avilés. Dinuclear zinc–n-heterocyclic carbene complexes for either the controlled ring-opening polymerization of lactide or the controlled degradation of polylactide under mild conditions. *Chem. Cat. Chem.*, 6(5):1357–1367, 2014

A final quantum chemistry approach was developed, from a long time collaboration with a group in *Instituto Superior Técnico*, where we calculated the optical properties of boron compounds and rationalized the results with the experimental spectra. The first manuscript was published this year and others are being prepared.

D. Suresh, C. S. B. Gomes, P. S. Lopes, C. A. Figueira, B. Ferreira, P. T. Gomes, R. E. Di Paolo, A. L. Maçanita, M. T. Duarte, A. Charas, J. Morgado, D. Vila-Viçosa, and M. J. Calhorda. Luminescent di-and trinuclear boron complexes based on aromatic iminopyrrolyl spacer ligands: Synthesis, characterization, and application in OLEDs. *Chem.-Eur. J.*, 21:9133–9149, 2015

Molecular modeling

The study of the reversibility of prion protein misfolding was the subject of my master thesis. This work was published during this PhD:

D. Vila-Viçosa, S. R. R. Campos, A. M. Baptista, and M. Machuqueiro. Reversibility of prion misfolding: insights from constant-pH molecular dynamics simulations. *J. Phys. Chem. B*, 116(30):8812–8821, 2012

I was also part of a collaboration with other theoretical group where we studied the pH dependent misfolding of β2-microglobulin, a protein involved in dialysis-related amyloidosis:

S. G. Estácio, H. Krobath, D. Vila-Viçosa, M. Machuqueiro, E. I. Shakhnovich, and P. F. N. Faísca. A simulated intermediate state for folding and aggregation provides

insights into ΔN6 β2-microglobulin amyloidogenic behavior. *PLOS Comp. Biol.*, 10:1–17, 2014

In our group, we also applied the former version of the stochastic titration constant-pH molecular dynamics method to the titration of glutathione:

D. Vila-Viçosa, V. H. Teixeira, H. A. F. Santos, and M. Machuqueiro. Conformational study of GSH and GSSG using constant-pH molecular dynamics simulations. *J. Phys. Chem. B*, 117(25):7507–7517, 2013

And to a synthetic tripodal sugar receptor:

D. Vila-Viçosa, O. Francesconi, and M. Machuqueiro. Why a diaminopyrrolic tripodal receptor binds mannosides in acetonitrile but not in water? *Beilstein J. Org. Chem.*, 10(1):1513–1523, 2014

Abstract

Biological membranes are the first barrier between cell and the extracellular environment, and also their first via of interaction. These biological structures are complex and diverse but share a common feature: they are all supported by a lipid bilayer. This bilayer is often negative and sensitive to pH and ionic strength. Computational / theoretical methods have been used to understand how these two properties influence the bilayer structure. However, the methods available at the beginning of our work had two significant bottlenecks that we attempted to get rid of. First, none of the available Poisson–Boltzmann (PB) solvers were able to deal with both periodicity and pK_{a} calculations. The first project was developed to overcome this technical limitation: we used an available PB solver in a new approach to perform pK_{a} calculations taking into account the system periodicity. Secondly, to simulate highly charged membranes, a proper treatment of the ionic strength is crucial and a full neutralization of the system is probably too rough an approximation. Hence, we developed a PB-based method to determine the number of ions that should be added to the simulations.

With these two problems solved, we developed a new constant-pH molecular dynamics method to deal with charged lipid bilayers (CpHMD-L). This method allows a proper treatment of the periodicity and ionic strength in model membranes. Finally, we applied our methods to three model systems to illustrate the importance of taking into account protonation / conformation coupling in molecular dynamics simulations, in particular when looking at pH dependent phenomena. To the best of our knowledge, this is the only available method that can deal simultaneously with pH considering the protonation / conformation coupling, periodic boundary conditions in protonation free energy calculations, and a careful treatment of ionic strength. This represents a significant improvement in simulations of model biological membranes. It is now possible to move a step forward in the direction of biological membranes since we are now able to simulate a lipid mixture with several different lipids.

Keywords - pH, ionic strength, protonation, membrane, lipid

Resumo

As membranas biológicas são a primeira barreira entre a célula e o ambiente extracelular, sendo também a primeira via de interacção entre as mesmas. Estas estruturas biológicas são complexas mas partilham uma característica comum: são todas suportadas por uma bicamada lipídica. Esta bicamada é quase sempre carregada negativamente e sensível ao pH e à força iónica. A dependência que a estrutura das bicamadas tem destas duas propriedades tem sido muito estudada nos últimos anos. No entanto, o seu detalhe molecular ainda está longe de ser bem compreendido. Os métodos teóricos / computacionais são muito úteis para estudar as bicamadas a este nível. As simulações de dinâmica molecular permitem perceber qual o comportamento dinâmico destas estruturas. Apesar disso, não permitem avaliar qual a influência do pH na estrutura, uma vez que não têm em conta o equilíbrio temporal de protonação / desprotonação. Este equilíbrio pode ser avaliado através de um modelo de Poisson–Boltzmann para calcular as energias livres de protonação de cada resíduo titulável, seguido de uma amostragem de Monte Carlo dos vários estados de protonação desses resíduos. No entanto, esta última abordagem, não permite o estudo do comportamento dinâmico do sistema. A complementaridade dos dois métodos referidos pode ser utilizada para realizar simulações de dinâmica molecular a pH constante. Nestas simulações, a dinâmica do sistema é interrompida, em intervalos de tempo bem definidos, para calcular novos estados de protonação típicos de um determinado valor de pH e de uma conformação. A aplicação de um destes métodos a um sistema lipídico carregado não é trivial e foi, portanto, o principal objectivo deste trabalho de doutoramento: desenvolver uma nova extensão ao método da titulação estocástica [12, 13] para bicamadas lipídicas (CpHMD-L). Para desenvolver esta aplicação, foi necessário ultrapassar dois problemas: a periodicidade nos cálculos de Poisson–Boltzmann e o tratamento da força iónica em membranas carregadas.

No capítulo 3 foi implementado um método que utiliza um programa para resolver a equação de Poisson–Boltzmann, para que os cálculos de energia livre de protonação pudessem ser feitos com periodicidade. A periodicidade já estava implementada no programa que escolhemos, mas foi necessário adaptá-lo para também efectuar cálculos de protonação. Com o método desenvolvido, calculámos

o pK_{a} de uma molécula de DMPC inserida numa membrana de DMPC desprotonado puro utilizando uma aproximação de resposta linear. Este estudo permitiu calcular a que pH é que o DMPC começa a protonar quando se encontra numa membrana pura. De acordo com os nossos resultados, isto só deverá ocorrer a valores de pH muito baixos (2–3) o que significa que, numa simulação de DMPC puro, podemos assumir que este estará desprotonado.

No segundo trabalho deste doutoramento (apresentado no capítulo 4) foi abordado o problema do tratamento da força iónica. Na grande maioria dos trabalhos computacionais realizados com membranas carregadas, estas são simuladas com contra iões suficientes para neutralizar as cargas na caixa de simulação. No entanto, de acordo com vários modelos, isto representa uma concentração local de iões demasiado elevada e pode sobre estabilizar a membrana. Para evitar esta solução, desenvolvemos um método, baseado na equação de Poisson–Boltzmann, no qual esta é utilizada para estimar o número de iões (positivos e negativos) que deve ser utilizado em cada simulação. Isto significa que uma simulação de uma membrana neutra é simulada com menos contra iões do que uma membrana muito carregada. Para testar o nosso método, utilizámos uma mistura de 25% de DMPA em DMPC e a sua transição de fase entre gel (baixas ionizações) e fluido (altas ionizações). Comparando com outros métodos habitualmente utilizados, o nosso foi o único capaz de prever a transição de fase na gama de ionização (ou pH) correcta.

Com estes problemas resolvidos, foi possível desenvolver o novo método CpHMD-L, em que são utilizadas as duas soluções apresentadas nos parágrafos anteriores (capítulo 5). Assim sendo, este método permite a simulação, a pH constante e a uma força iónica definida, de qualquer sistema com periodicidade segundo x e y. Como primeiro teste, utilizámos uma membrana de ácido oleico para a qual existe uma curva de titulação experimental [14]. Deste estudo sabe-se que, à semelhança de outros ácidos gordos de cadeia longa, o ácido oleico está presente na forma de bicamada a valores de pH mais baixos (em que o lípido está presente em ionizações abaixo de 50%) e na forma de micelas para ionizações muito elevadas. Para esta gama de ionizações foi possível reproduzir as curvas de titulação experimentais disponíveis para este lípido.

Depois do método final ter sido desenvolvido, foi possível aplicá-lo a três sistemas muito diferentes entre si. Em primeiro lugar, estudámos a titulação de uma mistura de DMPA e DMPC, em que foi possível corroborar a transição de fase obtida no capítulo 4 por dinâmica molecular clássica e estender a aplicação do método CpHMD-L a lípidos com mais do que dois estados de ionização possíveis (o DMPA pode adoptar com carga 0, -1 ou -2). Através do estudo da variação da estrutura da membrana com o pH, foi possível encontrar indícios de agregação das moléculas de DMPA para ionizações mais baixas. Esta observação está de acordo com resultados experimentais que sugerem que a pH 4.0 ocorre a formação

de microdomínios numa mistura de DMPA e DMPC [15].

Tendo como objectivo aumentar as aplicações a interacções péptido - membrana, estudámos a variação do valor de pK_a de péptidos modelo na interface água - lípido. Apesar de não existirem resultados experimentais para validar directamente os nossos resultados, a variação obtida foi de acordo com o esperado. Em termos gerais, o valor de pK_a de resíduos acídicos aumenta à medida que estes se inserem na membrana, sendo observado o oposto para resíduos positivos (histidina, lisina e N terminal).

O último trabalho apresentado nesta tese diz respeito ao estudo da interacção e inserção em membranas da família de péptidos pHLIP [16]. Na presença de bicamadas lipídicas, estes péptidos permanecem adsorvidos na interface membranar a pH neutro, inserindo na forma de hélice alfa a valores de pH baixos. Devido a este comportamento, é possível direccionar estes péptidos para células cancerígenas devido ao pH acídico existente perto destas. Os dados experimentais obtidos sugerem que a inserção dos péptidos pHLIP pode ser despoletada através da protonação de alguns resíduos chave de aspartato. Assim sendo, utilizámos uma aproximação de resposta linear (também utilizada no capítulo 3) para calcular o pK_a de alguns desses resíduos. Foi então possível observar uma forte correlação entre esses valores de pK_a e os valores de pH ao qual se sabe experimentalmente que os péptidos inserem na membrana (os valores de pK_a obtidos para as quatro variantes estudadas são todos entre 5.5 e 6.5). Para além disso, estudámos uma variante proposta por nós, na qual foi adicionada uma histidina, com o objectivo de tornar o péptido activo apenas numa gama curta de pH. Obtivemos um pK_a abaixo de 4 para essa histidina, o que significa que esta variante do péptido deverá ser activa entre pH $\sim$ 4.0 e $\sim$ 6.0.

Tendo em conta o trabalho desenvolvido nesta tese, foi possível desenvolver um método com o qual se pode estudar qualquer sistema periódico, a um determinado valor de pH e de força iónica. Isto representa um melhoramento significativo nas simulações computacionais de membranas biológicas. É agora possível avançar um pouco mais na direcção da simulação de membranas biológicas através da simulação de misturas de vários lípidos diferentes. Com sistemas maiores, é possível observar a formação de microdomínios, tal como é sugerido na secção 6.2. Os trabalhos desenvolvidos com interacções péptido – membrana representam uma vasta gama de possíveis estudos de proteínas membranares, tais como as proteínas envolvidas na cadeia transportadora de electrões, na qual os protões desempenham o papel essencial na síntese de ATP.

Palavras-chave - pH, força iónica, protonação, membrana, lípido

Contents

List of Figures

List of Tables

List of Abbreviations

Amino acids

Alanine	Ala	A
Arginine	Arg	R
Asparagine	Asn	N
Aspartate	Asp	D
Cysteine	Cys	C
Glutamate	Glu	E
Glutamine	Gln	Q
Glycine	Gly	G
Histidine	His	H
Isoleucine	Ile	I
Leucine	Leu	L
Lysine	Lys	K
Methionine	Met	M
Phenylalanine	Phe	F
Proline	Pro	P
Serine	Ser	S
Threonine	Thr	T
Tryptophan	Trp	W
Tyrosine	Tyr	Y
Valine	Val	V

Lipids

CL	cardiolipin
CTAB	cetyltrimethylammonium bromide
DMPA	dimyristoyl phosphatidic acid
DMPC	dimyristoyl phosphatidylcholine
DOPA	dioleoyl phosphatidic acid
DPPC	dipalmitoyl phosphatidylcholine
PA	phosphatidic acid
PC	phosphatidylcholine
PE	phosphatidylethanolamine
PG	phosphatidylglycerol
PI	phosphatidylinositol
PS	phosphatidylserine
SDS	sodium dodecyl sulfate

Other abbreviations and constants

PB	Poisson–Boltzmann
2D-PBC	two-dimensional periodic boundary conditions
AA	all atom
A_l	area per lipid
CpHMD	constant-pH MD
CpHMD-L	developed extension to CpHMD method to deal with charged lipid bilayers
CTr	C-Terminus
DD	Davis-Dowden method
ϵ	dielectric constant
ϵ_0	permittivity in vacuum
ϵ_{in}	membrane / protein dielectric constant
ϵ_{out}	water dielectric constant
GC	Gouy–Chapman
GRF	generalized reaction field
HF	Hartree–Fock
I	ionic strength

k_B	Boltzmann constant
$\mathbf{L}_\alpha$	liquid crystalline state
$\mathbf{L}_\beta$	crystalline state
l-BFGS	limited memory-Broyden-Fletcher-Goldfarb-Shanno
LRA	linear response approximation
MC	Monte Carlo
MD	molecular dynamics
MM	molecular mechanics
MSD	mean square displacement
N_A	Avogadro's constant
NMR	nuclear magnetic resonance
$n_\mathrm{PB}(\mathrm{A}^+)$	number of cations
$n_\mathrm{PB}(\mathrm{Cl}^-)$	number of anions
NTr	N-terminus
$\mathcal{P}$	critical packing parameter of amphiphilic molecules
PBC	periodic boundary conditions
PEF	potential energy function
pHLIP	pH (low) insertion peptides
$\mathrm{p}K_\mathrm{insertion}$	pK of insertion
$\mathrm{p}K^\mathrm{mod}$	pK_a value of the model compound
PME	particle mesh ewald
RESP	restricted electrostatic potential
RF	reaction field
$\boldsymbol{R}_\mathrm{g}$	radius of gyration
SPC	single point charge
T	temperature
T_C	melting temperatures
TM	transmembrane
τ_prt	duration of the unconstrained MM/MD segment in a CpHMD simulation
τ_rlx	duration of the MM/MD segment to relax the solvent around a frozen solute in a CpHMD simulation
UA	united atom
wt	wild type

Chapter 1

Lipids and their protonation

Contents

> "in saponification it is an acid which combines with the alkali, because the oil or fat contains a hidden acid." Otto Tachenius, 17th century [17].

In the 17th century, Otto Tachenius, a german chemist, already had an educated guess of how saponification reactions occur [17]. Later on, between 1813 and 1818, Michel Chevreul, a french chemist that was a pioneer in many works with lipids, elucidated a little further this reaction and found that *fats are combinations of acids with glycerin* [17] (these fats are nowadays known as glycerolipids - see section 1.1). Soap production was probably the first practical application of lipids. Due to their amphiphilic properties, they are able to solubilize apolar molecules in water (orienting their hydrophobic chain to the apolar molecule and the hydrophilic part to water). Most lipids used in soaps are also tensioactive, meaning that, when solubilized in water, the surface tension is reduced turning the wetting of surfaces easier.

Biological membranes are the first barrier between the cell and the extracellular environment and also the first via of interaction. All membranes present a very complex structure comprising proteic receptors, glycolipids, cholesterol, fatty acids, etc [18–21]. Their composition varies significantly between species, in particular between eukaryotes and prokaryotes, but all membranes have lipids forming a structural fluid with two monolayers forming the so-called lipid bilayer [18–21] (see section 1.2 for more details on lipid supramolecular structure).

1.1 Structure of lipid molecules

Lipid molecules present a wide variety of chemical structures. In 2005, Fahy et al. performed an exhaustive analysis of lipid molecules and grouped them in eight categories [22] (the revised version can be consulted in *http://www.lipidmaps.org/* [23]): fatty acyls, glycerolipids, glycerophospholipids, sphingolipids, sterol lipids, prenol lipids, saccharolipids and polyketides (Table 1.1). Fatty acyls, such as fatty acids (with a carboxylic group), fatty alcohols, esters, etc. include lipids with a single carbon chain (saturated and unsaturated). Glycerolipids derive from glycerol by esterification, glycosylation or other reaction, which replaces one or more alcohol functional groups by substituents. When one of the glycerol alcohol groups is replaced by a phosphate a glycerophospholipid is obtained. These lipids are organized in a separate group since biological membranes are very rich in glycerophospholipids with a large structural variety. All the work developed in this thesis is related to lipids within these three categories. Their biological relevance and previous studies will be discussed in more detail in the next sections.

Table 1.1: Lipid categories according with Ref. 22.

Category	Example (common name)
Fatty acyls	Oleic acid
Glycerolipids	Triacylglycerols
Glycerophospholipids	Dimyristoyl phosphatidylcholine (DMPC)
Sphingolipids	Sphingosine
Sterol lipids	Cholesterol
Prenol lipids	Coenzyme Q10
Saccharolipids	UDP-GlcNAc
Polyketides	Trichostatin

Sphingolipids also include a large variety of compounds, namely ceramide and phosphosphingolipids. They are biologically synthesized by the condensation of an acyl coenzyme A with serine, followed by reduction with NADPH. The resulting sphingamine molecule is formed and then converts into ceramides, phosphosphingolipids, etc. All these molecules share a sphingoid base backbone (in mammals, sphingosine is the most common sphingoid base) which is produced in the first steps of their biosynthesis. Similarly to glycerolipids and glycerophospholipids, the large variability of sphingolipids structure relies on the length of acyl chains, as well as the number and position of the branching units, unsaturations and other functional groups such as hydroxyls [22].

Cholesterol is probably the most famous lipid in the world. From a mediatic point of view there is the "good" cholesterol (in fish, nuts, etc) and the "bad" cholesterol (in burgers, pizzas, etc). From a biological point of view the good cholesterol is stored in high density lipoproteins (HDL) and the bad cholesterol is stored in low density lipoproteins (LDL). When it is not stored in a lipoprotein, cholesterol is an important constituent of biological membranes in mammals. From a biophysical point of view, cholesterol is a molecule that increases the melting temperature of a lipid bilayer. If we think as a theoretical biophysicist, cholesterol decreases the average area per lipid in a simulation box (see chapter 2). For what matters in this section, cholesterol is included in the sterols category. Besides cholesterol, the sterols category also includes other molecules such as testosterone, androgens and bile acids [22].

Prenols are lipid molecules which have an isoprenoid alcohol precursor. They are not structural lipids in biological membranes, having other biological functions in several processes such as in the electron transport chain (ubiquinone). Vitamins E and K are also included in this category [22].

Lipids have many associations with other biological molecules. Among the most important are sugars forming saccharolipids. Many of the membrane lipids are saccharolipids usually with very important roles in cell signaling and regulation processes [22].

The last category of lipids comprises polyketides which are natural compounds with carbonyl and methylene groups derived from several condensations of acetyl coenzyme A [24].

1.1.1 Lipids with charged groups

Among all mentioned categories, in this thesis, we are particularly interested in lipids with charged groups[1], since we studied membranes of glycerophospholipids (with phosphocholines and phosphatidic acids) and of a fatty acid (oleic acid).

Phosphatidylcholines (PC) are the most abundant group of zwitterionic lipids in humans (Figure 1.1). In these molecules, the hydroxyl groups of glycerol have been replaced by two esterified fatty acids (in this work we used myristic, palmitic and oleic acids), and one phosphocholine group, which renders the lipid zwitterionic at neutral pH (the net charge of these lipids is studied in detail in chapter 3). These molecules are able to form spontaneously very stable liposomes and, hence, have been widely used as model system to study lipid bilayers *in vitro* and *in silico* [25–40]. On the other hand, lipids with a negative charge at neutral pH are difficult to handle since they do not form easily a stable bilayer. The strong electrostatic repulsion between the headgroups competes with the hydrophobic interactions between the alkyl chains and the membrane becomes unstable. Phosphatidic acids (PA) are a set of glycerophospholids with a monosubstituted phosphate group in the headgroup (Figure 1.2). This phosphate has two pK_{a} values (see chapter 4 and section 6.2) and its net charge is strongly dependent on solution pH. The study of bilayers of PA is normally performed in PA/PC mixtures in order to decrease the repulsions between PA molecules. Finally, we also studied a single chain mono unsaturated fatty acids with 18 carbon atoms: oleic acid. The protonation of both phosphatidic acids and fatty acids is largely described in the literature and is explored in more detail in sections 1.3.2 and 1.3.4.

1.2 Supramolecular structure of lipids

The goal of this section is to offer some insight on the answer to the question "how do lipid molecules organize when they meet each other?" or, in more general terms "how do amphiphilic molecules organize when they meet each other?" since the

[1] In this group we find the so-called zwitterionic lipids which do not have a significant net charge at neutral pH, but have strong local charges in some of their chemical groups.

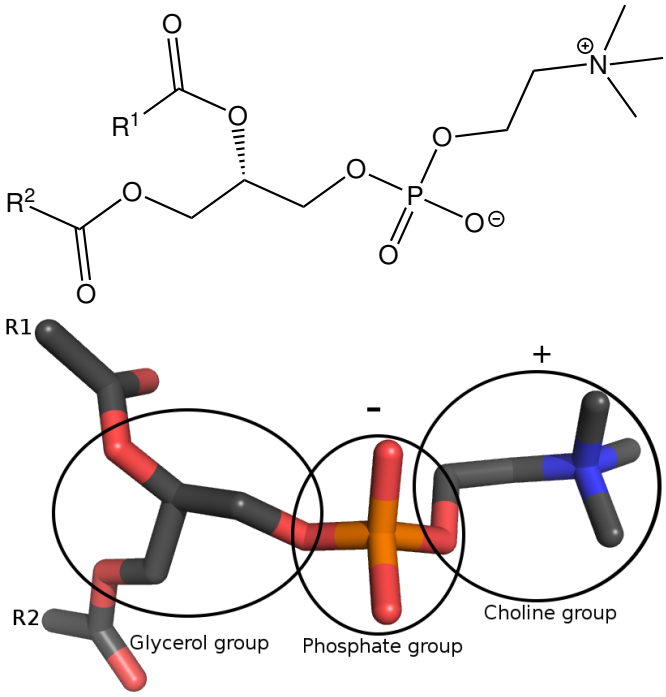


Figure 1.1: Chemical structure of phosphatidylcholines (PCs). PCs have a positively charged choline group, a negatively charged phosphate group and a glycerol moiety. One of the O atoms of glycerol is bonded to the phosphate while the other two are esterified. Hydrogen atoms were omitted for clarity.

thermodynamic principles that govern these interactions are quite similar. Both questions are carefully addressed in Ref. 41 and will be briefly discussed in this section.

Amphiphilic molecules have both polar and apolar groups, so that their interaction is possible in a wide variety of media. Depending on the solvent and on the chemical nature of the amphiphile, these molecules can adopt several supramolecular structures (Figure 1.3). If SDS ($NaC_{12}H_{25}SO_4$) is dissolved in water we shall probably obtain spherical micelles. In a micelle environment the carbon chains are in close contact with each other and the sulfate groups are facing the solvent (water), which makes spherical micelles the most stable supramolecular arrangement, i.e. the one with the lowest energy from a thermodynamic point of view. However, adding a salt (i.e. increasing the ionic strength) will lead to cylindrical micelles (also made of SDS). This happens because the repulsion between the negative headgroups is partially decreased by the presence of counterions and the curvature of the lipid–water interface becomes smaller. In other amphiphilic molecules,

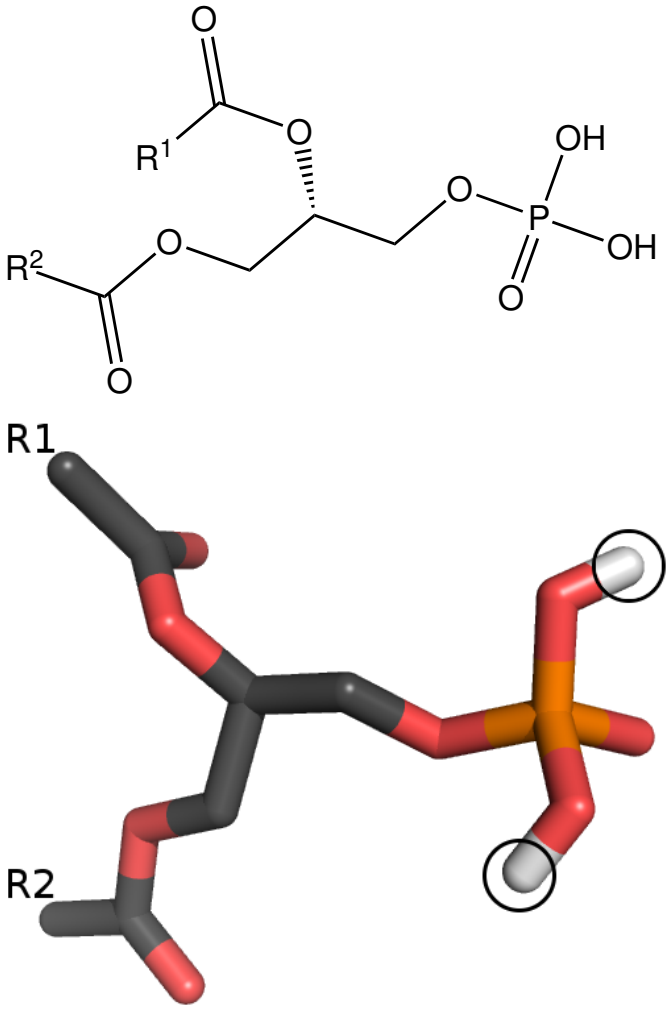


Figure 1.2: The chemical structure of phosphatidic acids is similar to phosphatidylcholines but the choline group is not present. Hence, the two represented protons can either be present or not, depending on the pH. Hydrogen atoms were omitted for clarity.

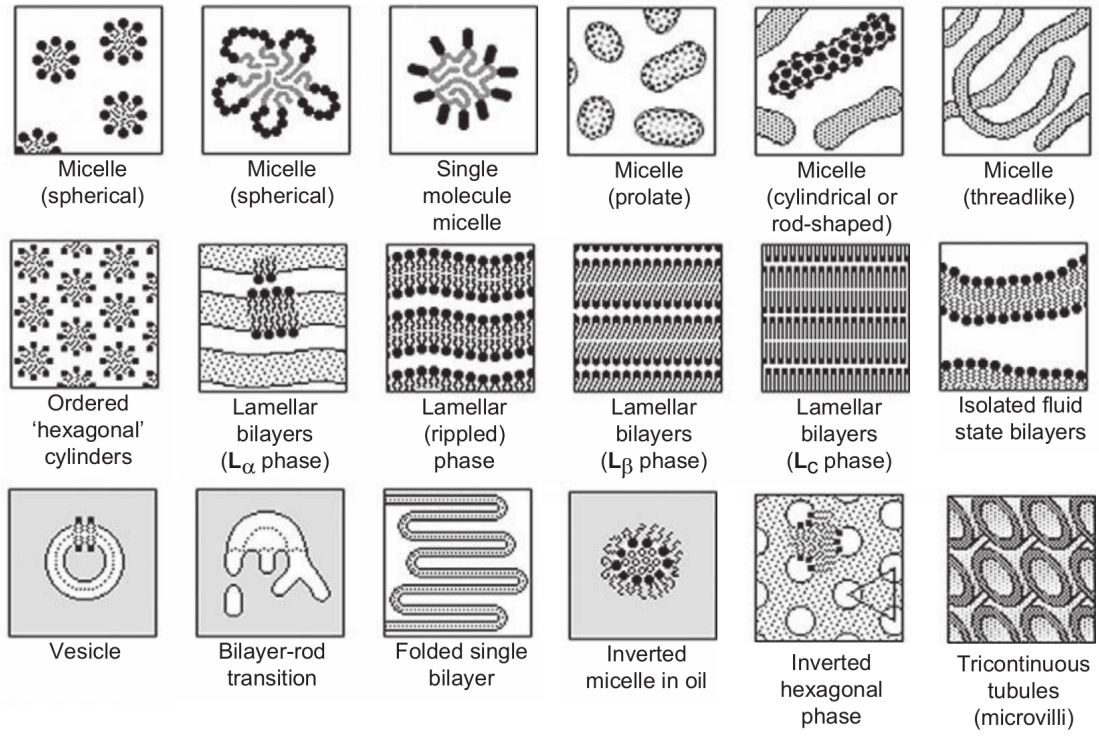


Figure 1.3: Examples of supramolecular structures of amphiphilic molecules. Adapted from Ref. 41.

for example DMPC, the charge repulsion between the zwitterionic headgroup is smaller and the adhesion between the chains is large, which makes bilayers the preferred supramolecular structure. In this case, vesicles surrounded by a bilayer of PC (liposomes) are formed in solution.

There are several factors that affect the way amphiphiles are arranged in aqueous solution. The chemical nature of the amphiphile itself is the most obvious one. A molecule with a small chain will have lower adhesion than a molecule with a longer chain. A charged lipid, such as PA, will have strong repulsions when compared to a zwitterionic lipid such as PC or PE. To understand how this happens, from a geometrical point of view, a critical packing parameter ($\mathcal{P}$) is defined relating the volume of the molecule when it is packed in the preferred supramolecular structure (V_P) with its optimal area (a_0) and its critical chain length (l_c) [41], a_0 being the area of the molecule facing the solvent and l_c the maximum length of the chain. $\mathcal{P}$ is given by:

$$\mathcal{P} = \frac{V_P}{a_o l_c} \tag{1.1}$$

These values are difficult to calculate and will not be determined in this thesis. However, the detailed study of this parameter helps to understand why amphiphilic molecules behave the way they do. In fact, in the simpler cases, there is a direct relation between $\mathcal{P}$ and the adopted supramolecular structure, since there is critical packing shape (the average shape of a single molecule in the supramolecular environment) that can be estimated from $\mathcal{P}$(Table 1.2). In the simplest example:

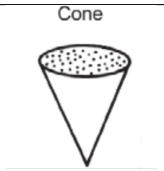
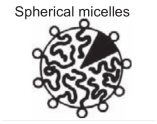
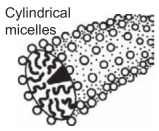
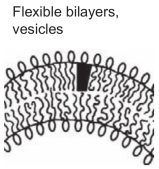
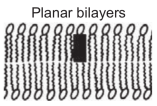
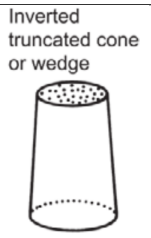
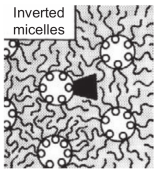
$$V_P = a_o l_c \tag{1.2}$$

$\mathcal{P}$ will be 1 and the critical packing shape will be a cylinder. Hence, a molecule with $\mathcal{P} = 1$ will be organized in planar bilayer, i.e. a large vesicle where the membrane curvature is so small that can be approximated to an infinite surface in a microscopic environment. In all simulations of this thesis, the lipids are assumed to pack in this critical shape, and we have always simulated a small patch of a lipid bilayer in a tetragonal box. This approximation is not very significant since we studied lipids (and mixtures of lipids) which are known to form stable liposomes. One possible exception is oleic acid (chapter 5) but this molecule has only been simulated at pH values where the bilayer is stable. Coming back to SDS, the first example of this chapter, we can compare its critical packing shape with that of DMPC. On the one hand, is has only one, smaller, hydrocarbon chain (12 carbons against the 14 in dimyristoyl chains), which means that l_c decreases. On the other hand, the sulfate groups are negatively charged and will repel each other, so that a_c is large compared to the chain length. These two combined effects lead to a smaller $\mathcal{P}$ value ($< 1/3$), the critical packing shape is a cone and the most stable supramolecular structure is a spherical micelle (Table 1.2). Continuing the analysis of the interesting behavior of SDS, as we mentioned before, if we add a salt, the positive ions will stabilize the interactions between sulfate groups and a_0 will decrease thus leading to an increase in $\mathcal{P}$. The critical packing shape is now a truncated cone and the preferred structure is a cylindrical micelle (Table 1.2). If we make the same reasoning with molecules where the headgroup is small compared to its chains, it can be easily predicted that $\mathcal{P}$ will increase and the adopted structure will be an inverted micelle (this structure is observed in solutions of cardiolipin and phosphatidic acid in the presence of Ca^{2+}, for example [42] - Table 1.2).

In a general way, the factors that influence the packing preferences of amphiphiles are:

- headgroup area
- chain packing
- temperature
- mixtures

Table 1.2: Representative dynamic packing shapes of amphiphilic molecules. Adapted from Ref. 41.

Examples	$\mathcal{P}$	Limit packing shape	Typical supramolecular structure
Single-chained lipids (surfactants) with large headgroup areas: SDS in low salt concentration	$< 1/3$		
Single-chained lipids with small headgroup areas: SDS and CTAB in high salt concentration	$1/3 - 1/2$		
Double-chained lipids with large headgroup areas, fluid chains: PCs , PSs, PGs, PAs, sphingomyelin	$1/2 - 1$		
Double-chained lipids with small headgroup areas, anionic lipids in high salt, saturated frozen chains: PEs, PSs + Ca^{2+}	~ 1		
Double-chained lipids with small headgroup areas, nonionic lipids, poly(cis)unsaturated chains, high T: unsaturated PEs, CLs + Ca^{2+}, PAs + Ca^{2+}, cholesterol	> 1		

SDS, sodium dodecyl sulfate

CTAB, cetyltrimethylammonium bromide

PCs, phosphatidylcholines

PSs, phosphatidylserines

PGs, phosphatidylglycerols

PAs, phosphatidic acids

PEs, phosphatidylethanolamines

CLs, cardiolipins

The first two have already been addressed. However, it is interesting to note that there are several elements that may affect the headgroup area (a_0), not only the aforementioned ionic strength and charge, but also pH (typically by altering the headgroup charge - see section 1.3). Regarding chain packing, we mentioned the chain length, but the presence of unsaturations and branches also alter $\mathcal{P}$ (by decreasing l_c) and, consequently, structural preferences. Temperature can also alter $\mathcal{P}$ in a less intuitive way. In molecules with more hydrophilic headgroups, an increase in the temperature will increase the repulsion between them, thus increasing a_0 and decreasing $\mathcal{P}$ (which in practice will reduce the size of a spherical micelle for example). In spherical micelles of nonionic surfactants, the effect is the opposite; they grow and eventually become cylindrical (probably by lowering a_0). Finally, the presence of more than one lipid in the system complicates the above reasoning. For example, lyso-DMPC (a DMPC with only one hydrocarbon chain) has a $\mathcal{P}$ below 0.5 which means that it forms spontaneously spherical micelles. However, a mixture of lyso-DMPC and DMPC ($0.5 < \mathcal{P} < 1.0$), at low concentrations of lyso-DMPC, the $\mathcal{P}$ of the mixture is also between 0.5 and 1.0 and the bilayer (liposome) will be stable [43]. In fact, in many cases, it is possible to assume that $\mathcal{P}$ of the mixture will take a value between the $\mathcal{P}$ of the two components [43].

This discussion was focused on large structural differences. However, there are more subtle changes of significant biological relevance that can occur without changing the lipid mixture significantly. Lipid bilayers are typically present in two macroscopic states: a more solid state called crystalline (L_β) and a fluid state called liquid crystalline (L_α) [44, 45] (there are more states that can be identified [45] but will not be discussed here). In L_β state the lipid molecules are almost frozen, i.e. it is a very organized state. The L_α is more disorganized and the lipid molecules have more space to move which makes their diffusion higher. From a biological point of view, a membrane in L_β phase is almost impermeable to metabolites which is a good property to maintain gradients. On the other hand, if the cell needs a permeable membrane it is more likely in L_α phase. These two phases can interconvert with a temperature changing: below a certain melting temperature (T_C) a bilayer is in the L_β phase and *vice versa*. T_C is a characteristic of the lipid molecule (or lipid mixture) and, in the case of phospholipids, strongly depends on the chain length and on the headgroup [41, 44]. For example, in PC there is a variation between a T_C of -1 °C with two lauroyl chains and 54 °C with two stearoyl (Table 1.3), as the adhesion is higher for longer chains. Moreover, the introduction of an unsaturation induces disorder which makes easier to change to L_α phase (T_C decreases - Table 1.3).

In the atomistic models used in all works reported in this thesis, all observed structural changes are microscopic, which means that we are not able to sample a real L_α to L_β transition or *vice versa*. Nevertheless, in most cases, large differences

Table 1.3: Melting temperatures (T_C) of some common lipids (in °C). Adapted from Ref. 44

Lipid chains	Melting temperature (T_C)
Dilauroyl	−1
Dimyristoyl	23
Dipalmitoyl	41
Distearoyl	54
Palmitoyl oleyl	−5

induced by pH or ionic strength can be observed. To describe these observations we refer to our microscopic phases as gel (analogous to L_β) and fluid (analogous to L_α). By using this nomenclature, we clearly state the difference between a microscopic description of the two crystalline phases of lipids and their corresponding macroscopic states.

1.3 Charged surfaces and protonation of lipids

Most biological lipid / water interfaces are negatively charged. In fact, positive lipids are rare and only occur in very specific cases [46–48]. The negative charge results from dissociation of a proton from a lipid molecule, leaving behind a negative charge [49]. From an evolutionary point of view it is also advantageous to have negative membranes, in order to attract protons, allowing them to quickly diffuse from one membrane complex to another, in the electron transport chain, for example [14, 50–55]. The membrane charge is influenced by several factors, namely its composition (concentration of negatively charged lipids), ionic strength and pH [49]. The control of this charge is crucial in living systems: if the charge level is too high, the membrane is destabilized and, in the limit, can disrupt or form other supramolecular structures [56]; on the other hand, if the charge is too low, the membrane can became too ordered (condensed) [57] and impermeable to metabolites.

1.3.1 External factors

pH is one of the most relevant properties responsible for membrane charge, since it directly affects the protonation state of lipids. Due to the anionic nature of the lipid / water interfaces, the local concentration of H^+ is often higher than in bulk solution. This fact does not mean that the "local pH" is different, since pH

is a bulk and macroscopic property, which is difficult (or impossible) to define at microscopic scales [58].

pH strongly influences the structure of lipid bilayers and, in some cases (for example long chain fatty acids), it can even induce transitions from membranes to micelles [14, 59, 60]. At low pH values (lower ionizations) fatty acids are easily packed in ordered structures, such as oil droplets or membranes. With pH increase, carboxylic groups became negatively charged, the membrane structure is destabilized and formation of micelles is favorable ($\mathcal{P}$ decreases with pH since the repulsion between headgroups increases). Smaller effects can be observed, for example, in pseudo binary mixtures[2] of DMPA and DMPC, where pH induces a phase transition from fluid (at neutral pH) to a mixed phase, with gel and fluid micro-domains (at pH 4.0) [15]. On the one hand, DMPC is zwitterionic and remains in the fluid phase at both pH 4.0 and 7.0; on the other hand, the protonation state of DMPA increases significantly from pH 7.0 to pH 4.0. It means that it becomes less negative and the repulsion between headgroups is smaller ($\mathcal{P}$ decreases). This leads to a transition to an ordered phase, coupled to a phase separation from DMPC, and the consequent formation of micro-domains.

Ionic strength also has a strong influence in anionic lipid bilayers since cations can replace protons as positive stabilizers of the negative charges. Interestingly, this property may have two opposite effects explored in some detail in chapter 5: first, the partial replacement of protons by cations induces an attenuation of the repulsions between negative charges, which would lead to higher ionizations, and the titration curve deviates to lower pH values (this is the most common effect). However, due to this attenuation, the headgroups come closer, which has the opposite effect.

1.3.2 The role of lipid protonation in biology

Like proteins, sugars, and nucleic acids, lipids also have important roles in several biological processes that are often poorly understood [61]. In mitochondria respiratory chain, lipids play an essential role in maintaining a pH gradient (and, consequently, a membrane potential) across the bilayer, which is responsible for ATP production in complex V (ATPase) [18]. In fact, it is thought that cardiolipin (CL) has the effect of keeping respiratory complexes associated facilitating all respiratory processes [51]. Moreover, in proton diffusion studies between transmembrane channels, is was observed that protons diffuse much faster than the expected for a 3D diffusion [50, 52, 53, 55]. This indicates that protons are kept in the membrane skating between lipids, as had already been suggested several

[2] In this context, a pseudo binary mixture is a mixture of two different lipids but, in this case, one of them have several protonation states.

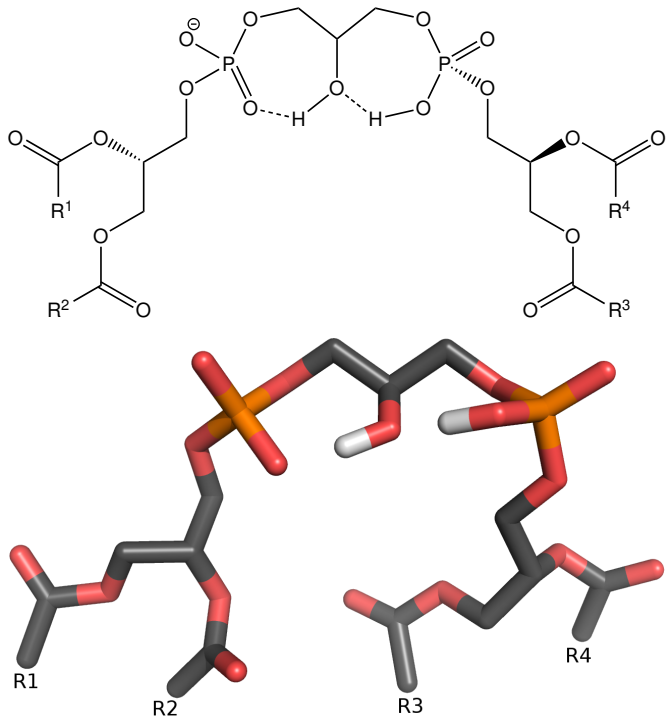


Figure 1.4: Chemical structure of cardiolipins (CLs) in their most probable protonation state at neutral pH. CLs have four acyl chains where two phosphatidyl moieties are linked together through a central glycerol. Hydrogen atoms were omitted for clarity.

years before in the titration of oleic acid [14]. The characteristic negative charge of biological membranes, essential to maintain the flow of "skating protons", is due to a substantial fraction of charged phospholipids, such as PS, PA, PG, phosphatidylinositol (PI), CL.

CLs are an exceptional class of phospholipids, having crucial roles in several biological processes [51, 54, 62–64]. Their structure is unique, since they present four acyl chains, where two phosphatidyl moieties are linked together through a central glycerol group (Figure 1.4) [54, 61, 65–67]. Interestingly, CLs are typically found in the inner membranes of bacteria and mitochondria, with molar fractions from 0.05 to 0.20 [68]. Moreover, CLs are the only lipids conferring a negative charge to these membranes [68], forming lipid domains in bacterial membranes [69]. As previously mentioned in Ref. 51, this indicates a strong relation between these lipids and respiratory complexes. In fact, a deficiency in CLs biosynthesis is associated with Barth's syndrome disease [70]. In this disease, the lipid only has three chains and is not able to form the so-called "acid-anion" structure (Figure 1.4), since the

adhesion between the chains is smaller and the asymmetry destabilizes the desired conformation. Therefore, CLs are not able to retain protons, which leads to a deficient respiratory process.

PAs (Figure 1.2) have a simple structure (a single phosphate moiety in the headgroup) but occupy key positions in many biological processes [71]. They exhibit three possible protonation states (0, -1, and -2), originating a rather complex titration behavior. This lipid is used in living systems as a pH biosensor [72, 73] because, under nutrient depletion conditions, the cytosol pH is lower and PA molecules became less negative (the protonation of phosphomonoester is increased). As a consequence, the positive transcriptional regulator Opi1 is released from the ER membrane to the nucleus, where it inhibits membrane biogenesis. Thereby, pH-modulated PA binding of Opi1 is the mechanistic basis of the coupling between membrane biogenesis and general metabolism [73].

1.3.3 First theoretical models

The negatively charged lipids, present in most biological membranes, originate a negative electrostatic potential, propagated for significant distances away. This electrostatic potential is important in several experimental observations, such as membrane fusion or solute / membrane interactions. The negative potential induces a modified local concentration of negative and positive species (ions for example), which also depend on bulk concentrations of these species and are not trivial to estimate.

One of the simplest methods to study charged membranes is to imagine a flat, infinite, and uniformly charged surface separating a highly polarizable media (water) and a media with a lower dielectric constant (membrane). This crude approximation is one of the few examples where the Poisson–Boltzmann (PB) equation can be solved analytically. This solution is then known as Gouy–Chapman (GC) equation and relates the potential generated by the surface with the distance, its charge density, and ionic strength [74, 75]. This rational can also be extended to model a capacitor, i.e. two surfaces facing each other [41, 49, 76], which can be particularly informative to estimate the electrostatic potential between two membranes during fusion processes. In fact, it was developed a solution where the electrostatic potential between the two surfaces is related with the distance between the two surfaces, ionic strength, solution pH, dissociation constants of titrable groups, and their concentration in the membrane [77]. This simple model gain some detail with the inclusion of explicit ions close to the surface. The dynamic behavior of these ions can be studied with Monte Carlo (MC) simulations to obtain a more accurate concentration profile of the ions (counterions and coions) close to a charged surface [78–80].

With both GC equation and MC sampling the molecular detail of the mem-

brane is not present. The possibility to numerically solve the PB equation allowed the inclusion of molecular detail in the study of electrostatic properties of lipid bilayers [49, 76, 81–83]. Hence, it is now possible to understand how this detail, and the correspondent partial atomic charges, influence the electrostatic potential adjacent to the membrane [81]. As in GC model, the calculated potential can be used to study the concentration profile of ionic species. The PB equation is then used in several applications, such as membrane titration [82] and, more recently, the study of systems with nonuniform dielectric constants [83].

1.3.4 Experimental results

Model systems of lipid bilayers have been used over the past decades [32, 33]. Usually, a lipid bilayer film is suspended in an aqueous solution to form multilamellar vesicles (commonly known as MLVs) which can then be extruded or directly used as biophysical models of lipid bilayers. This approximation is similar in molecular modeling (section 1.3.5), i.e. it is a model system where the bilayer is only formed with lipids, where proteins are (often) absent.

Lipid protonation has been studied in the last decades [84]. In 1966, it was already known that the titration of fatty acids originated unusual and intriguing curves [84]. Later, in a study with an oleic acid suspension, a model was proposed to explain these curves: at low pH the lipid is arranged in oil droplets, at intermediate values in liposomes, and, at higher pH values, where the repulsion between the carboxilates is higher, micelles became the most stable structure [14]. In the same report, a break-through hypothesis was also proposed, stating that anionic lipid headgroups can act as a proton conducting pathway between, for example, respiratory complexes [14]. The titration of fatty acids was the subject of several more recent studies [59, 60, 85–90]. Even though more complex and detailed mechanisms have been proposed, the original interpretation of fatty acid aggregation was maintained [59, 90].

Due to the presence of two phosphate groups and the absence of positive groups, the protonation state of CLs is a matter of debate in the literature [91]. This lipid has three possible protonation states (0, -1, and -2) and their populations depend on pH and other external factors, such as the concentration of CLs [91]. The study of their titration curve was already performed and a high macroscopic $\mathrm{p}K_\mathrm{a}$ value (> 7.0) has been proposed [92]. This means that, at physiological pH, CLs have a net charge of -1, which will strongly influence its conformation in biological membranes, allowing the presence of "acid-anion" conformations (see figure 1.4).

1.3.5 Atomistic MD simulations

MM/MD simulations have been widely used to study the dynamic behavior of model lipid bilayers [93]. However, the study of lipid protonation is a difficult task and is often addressed using models simulated at constant protonation, i.e. ignoring the protonation equilibria in the simulations [66, 83, 94–98]. A possible solution to this rough approximation is the use of the so-called constant-pH MD methods [3, 8, 10–13, 99–131]. These methods have several possible theoretical basis but, in all of them, the conformation and the protonation are coupled in the MD simulation. This means that, in simple terms, pH is an external variable, such as temperature or pressure. A detailed discussion of these methods is beyond the scope of this thesis and focus will be given to constant-pH MD simulations of membrane environments.

Some of the available constant-pH MD methods treat protons as a continuous variable, i.e. the protonation state of a titrable group is a decimal number between 0 and 1 (similarly to λ-dynamics [132]). One of these methods was applied to the titration of a surfactant molecule in a micelle environment [113]. The authors compared the effect of ionic and nonionic micelles in the titration curve. More recently, the same method was applied to the pH-dependent membrane–micelle transition of oleic acid [116, 118], where the authors observed the formation of a pseudo bilayer at low pH values (it is not a model of a bilayer in the most common sense since it does not have periodicity). More recently, the effect of cationic lipids in the titration of fatty acids was also studied using this method [121]. Though it is widely used in the literature, it has not been shown that it samples from the proper ensemble. This means that, when using it, one can not be sure that the sampled conformations (and protonation states) are correct, from a theoretical point of view. A different method, also based in λ-dynamics, was applied to a coarse grained model[3] of oleic acid in micelles and in a membrane environment [117].

1.4 Goals

> "Protein and nucleic acids receive so much attention and hype that a third biological building block, the phospholipid, might well be suffering from an inferiority complex." F. M. Menger, 2005 [19].

The constant-pH MD methods available at the beginning of our work did not allow the treatment of periodicity in protonation free energy calculations. Moreover, the

[3] In a coarse grained model, the atoms are grouped in beads to accelerate the simulations.

treatment of ionic strength in simulations of charged bilayers was often disregarded by simply neutralizing the systems (which is probably far from a sound physical situation). Then, our first goals were:

- To implement PB calculations with a software that allows the correct treatment of periodicity in membrane systems (chapter 3)
- To propose a new theoretical method to treat the ionic strength in simulations of significantly charged lipid bilayers (chapter 4)

With this two advances we were able to proceed to the main goal of this thesis:

- To develop a new extension of the stochastic titration constant-pH MD method [12, 13] that allows for periodicity in pK_{a} calculations and correct treatment of ionic strength (chapter 5)

This method is applicable to several biological systems as illustrated in chapter 6.

With this work, we wanted to provide a method to simulate any periodic system at specific pH and ionic strength values. The possible applications of such method extend from simple anionic membranes to more complex systems, such as peptide / lipid interactions or the simulation of membrane proteins.

Chapter 2

Theory and methods

Contents

> "Theoretical chemistry is a peculiar subject. It is based on an equation that can hardly ever be solved." P. Fowler, 1990 [133].

The author of the above citation was referring to Schrödinger's equation which is not the subject of the present chapter, although it can also be applied to all the basic equations that we shall mention in the following sections: integration of the Newton's equations, the solution of the Poisson–Boltzmann equation, etc. All these equations share a common feature: in most cases, they have to be solved numerically, nowadays using a computer program.

The first step of most theoretical works is to develop a model: do we need to describe electrons? do we need to use explicit water molecules? how large should the membrane be? For example, to understand a chemical reaction you may need to describe all electrons involved in that reaction; to calculate the pK_a of a simple molecule it is probably enough to use an approximate continuum electrostatics model. The goal of this chapter is to provide a simple explanation for the models and simulation techniques used in the remaining chapters of the present thesis.

2.1 Molecular mechanics / molecular dynamics

To characterize the conformational space of small systems such as a lipid patch, a micelle or a protein, we need to choose a model and a sampling method. In all works presented in this thesis we used a molecular mechanics (MM) model to describe the system and molecular dynamics (MD) to sample its conformational space. In this type of models, we use the Born–Oppenheimer approximation, in which it is assumed that the electrons can adapt to the nuclei positions very quickly so we can look only at the nuclei positions. A consequence of this approximation is that the models will lack the explicit effects of polarization. A detailed explanation about MM/MD can be found in references [134] and [135] as we only try to offer a simple and concise explanation about this subject.

2.1.1 Potential energy function

In a MM approach the system is described with a potential energy function (PEF). It describes the atoms and their interactions defined by their nuclei positions while ignoring electron movements. PEFs used in this thesis were the ones associated with the GROMOS [136] force fields. Generally, this function take into account four types of bonded interactions (bond length stretch, bending angles, and torsions of proper and improper dihedral) and two nonbonded interactions (van der

Waals forces and electrostatic interactions):

$$V(\vec{r}^N) = V_b(\vec{r}^N) + V_a(\vec{r}^N) + V_{pd}(\vec{r}^N) + V_{id}(\vec{r}^N) + V_{vdw}(\vec{r}^N) + V_{elect}(\vec{r}^N) \quad (2.1)$$

where $\vec{r}^N$ represents the vectorial positions of all atoms in the system, the first four terms correspond to bonded interactions and the last two correspond to nonbonded interactions.

$V_b(\vec{r}^N)$ is the vector containing all bonds which are described with a harmonic potential:

$$V_b(\vec{r}^N) = \sum_{n=1}^{N_b} \frac{1}{2} K_{b_n} (b_n - b_{0_n})^2 \quad (2.2)$$

where K_{b_n} is the force constant between two atoms, b_n is the distance between them and b_{0_n} is their optimal distance.

$V_a(\vec{r}^N)$ is the angle bend which is described with a harmonic angle potential:

$$V_a(\vec{r}^N) = \sum_{n=1}^{N_\theta} \frac{1}{2} K_{\theta_n} (\theta_n - \theta_{0_n})^2 \quad (2.3)$$

where K_{θ_n} is the force constant associated with angle bend, θ_n is the angle between three atoms and θ_{0_n} is their optimal angle.

$V_{pd}(\vec{r}^N)$ and $V_{id}(\vec{r}^N)$ are the potential energy of the torsion of proper and improper dihedral, respectively. $V_{pd}(\vec{r}^N)$ is described with a periodic function with minima and maxima at regular intervals:

$$V_{pd}(\vec{r}^N) = \sum_{n=1}^{N_\varphi} K_{\varphi_n} [1 + \cos(m_n \varphi_n - \delta_n)] \quad (2.4)$$

where K_{φ_n} is the force constant associated with torsion around the dihedral angle, φ_n is the value of proper dihedral angle, m_n is the multiplicity which stands for the number of minima (and maxima) of energy in one complete rotation and δ_n is the reference maximum and it can be 0 or π.

The last bonded interaction in equation 2.1, $V_{id}(\vec{r}^N)$, regards the improper dihedral which is also described as a harmonic potential:

$$V_{id}(\vec{r}^N) = \sum_{n=1}^{N_\xi} \frac{1}{2} K_{\xi_n} (\xi_n - \xi_{0_n})^2 \quad (2.5)$$

where K_{ξ_n} is the force constant associated with angle torsion, ξ_n is the value of improper dihedral angle between four atoms and ξ_{0_n} is their optimal angle.

The last two terms of the PEF are related with nonbonded interactions: van der Waals forces ($V_{vdw}(\vec{r}^N)$) and electrostatic interactions ($V_{elect}(\vec{r}^N)$).

$V_{vdw}(\vec{r}^N)$ is described as a Lennard-Jones interaction:

$$V_{vdw}(\vec{r}^N) = \sum_{i=1}^{N} \sum_{j>i}^{N} \left(\frac{C_{12_{ij}}}{r_{ij}^{12}} - \frac{C_{6_{ij}}}{r_{ij}^{6}} \right) \tag{2.6}$$

where r_{ij} is the distance between two atoms (i and j) and $C_{12_{ij}}$ and $C_{6_{ij}}$ are interaction parameters. $C_{12_{ij}}$ is related with the repulsion between these two atoms while $C_{6_{ij}}$ is related with their attraction.

$V_{elect}(\vec{r}^N)$ stands for electrostatic interactions which will be further explained (see section 2.1.7 for details). For now, to understand PEF we just assume that this interaction is described by Coulomb's law:

$$V_{elect}(\vec{r}^N) = \sum_{i=1}^{N} \sum_{j>i}^{N} \frac{q_i q_j}{4\pi\epsilon_0 \epsilon_r r_{ij}} \tag{2.7}$$

where r_{ij} is the distance between two atoms (i and j), q_i and q_j are the charges of these atoms, ϵ_0 is the permittivity in vacuum and ϵ_r is the relative dielectric constant.

Both nonbonded interactions are calculated for all pairs in the system separated by more than two covalent bonds. However, there are situations where it is convenient to ignore interactions between atoms that are separated by 3 covalent bonds (1-4 interactions). This is the case of the interaction between two consecutive hydrogen atoms in a benzene ring for example. These 1-4 exceptions are added to a proper list and are also part of the force field (section 2.1.2).

Force

With an energy description of the system (i.e. with $V_b(\vec{r}^N)$ calculated) we now compute the forces that act on each atom:

$$\vec{F_i} = -\frac{\partial f}{\partial \vec{r_i}} V = -\left(\frac{\partial f}{\partial x_i}, \frac{\partial f}{\partial y_i}, \frac{\partial f}{\partial z_i} \right) V = -\nabla_{\vec{r}_i} V \tag{2.8}$$

where $\nabla_{\vec{r}_i}$ is the gradient calculated at the position of atom i, and $\vec{F_i}$ is the resulting force on atom i. With $\vec{F_i}$ calculated, the description of the system is completed and we are now in position to study its conformational space. Methods of energy minimization (section 2.1.4) such as Steepest Descent or l-BFGS and sampling techniques such as MD (section 2.1.3) or Monte Carlo [137] can be used in the system.

2.1.2 Force field

To calculate the PEF of a system, we need optimal bond lengths, force constants for bonds, angle, and dihedral, Lennard–Jones parameters, partial atomic charges, etc. This large set of parameters is called a force field. A force field is a list of these parameters which were calculated according to some rules and should, in principle, reproduce the available experimental results. In fact, from a pragmatic point of view, the force field is as good as its ability to reproduce the experimental data.

All force fields are empirical, which means that the parameters are manually adjusted to reproduce experimental data such as area per lipid [138] or free energy of solvation [136]. In this field, typically, the parameters are first estimated using a quantum method and then some of them are adjusted. There are, however, two groups of molecular dynamics force fields that is worth distinguishing: all atom and united atom (see Figure 2.1). In all-atom force fields, such as AMBER [139–141] and CHARMM [142, 143], all atoms are considered in the PEF. In united-atom force fields such as GROMOS [136, 144–148], the force field used in this thesis, apolar hydrogen atoms are omitted and are treated together with the heavy atoms to which they are bonded. This means that, for example, a CH_3 is treated as a single, slightly larger and heavier carbon atom.

2.1.3 Molecular dynamics

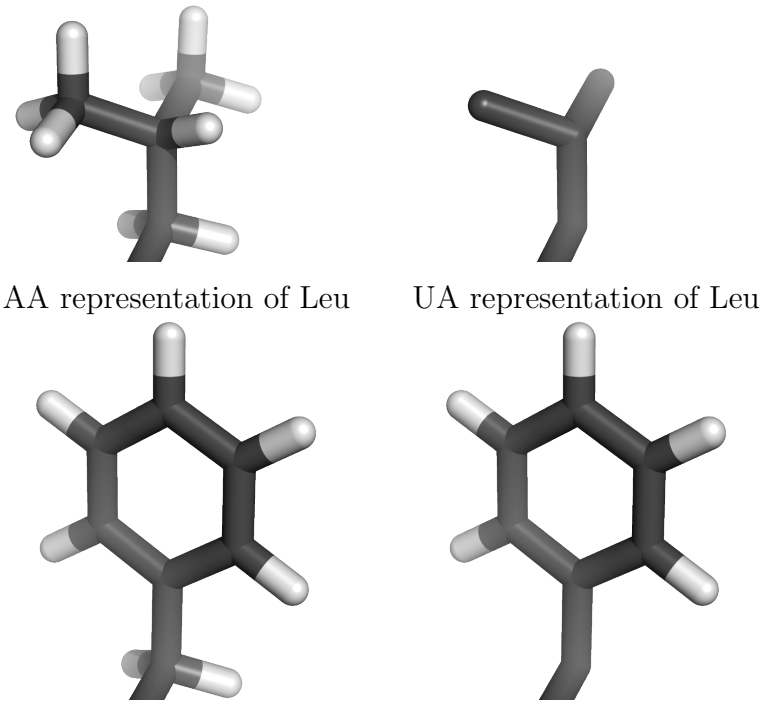
Molecular dynamics (MD) is one of the most used sampling methods in biomolecular simulations. There are several methods that allow a faster sampling of the conformational space such as replica exchange molecular dynamics [149], metadynamics [150, 151], accelerated molecular dynamics [152], etc., but, for all purposes of this thesis, we opted for standard MD. For all these methods, we need a description of the potential energy function as described in the previous sections.

In MD, Newton's equations of motion are used to obtain a trajectory, i.e. a sequence of positions and velocities of each particle of the system along the simulation time (usually in the ns or μs timescale). We start with the acceleration of each particle at instant t by using Newton's second law:

$$\frac{\mathrm{d}\vec{r}_i(t)}{\mathrm{d}t^2} = \frac{\vec{F}_i(t)}{m_i} \tag{2.9}$$

where $\vec{r}_i$ and m_i are the coordinates and mass of atom i at instant t, and $\vec{F}_i$ is the force acting on that atom.

The acceleration vector can then be integrated to obtain velocities and positions at instant $t + \Delta t$. Since this problem does not have an analytic solution we need a numerical method to solve the equations. The numerical method used in this



AA representation of Leu UA representation of Leu

AA representation of Phe UA representation of Phe

Figure 2.1: All atom (AA) and united atom (UA) representations of leucine (Leu) and phenylalanine (Phe) side chains. In Leu, all hydrogen atoms are omitted in UA representation. In Phe, only the two hydrogen atoms of the methylene group are omitted in UA representation.

work was the leap-frog algorithm [153]. We shall not provide too much detail but, in this method, the position at time t and the velocities at time $t + \frac{\Delta t}{2}$ are used to calculate new positions and velocities according to:

$$\vec{r}_i(t + \Delta t) = \vec{r}_i(t) + \vec{v}_i \left(t + \frac{\Delta t}{2} \right) \Delta t \tag{2.10}$$

$$\vec{v}_i \left(t + \frac{\Delta t}{2} \right) = \vec{v}_i \left(t - \frac{\Delta t}{2} \right) + \frac{\vec{F}_i(t) \Delta t}{m_i} \tag{2.11}$$

where Δt is the already mentioned time step. This time step is a relevant parameter of our simulation since a value too small will increase drastically the computational time and a value too large will wrongly describe the dynamic behavior of the system. We chose 2 fs in all our simulations which we consider to be a good compromise (see section 2.1.9 for a discussion of possible problems that could

arise from this choice and how to avoid them). The number of integration steps (a step between t and $t + \Delta t$) is chosen by the user: for example, to simulate a system for 1 ns one would need 500 000 steps.

2.1.4 Energy minimization

We used MD to sample the conformational space of our systems in all works presented in the next chapters. This sampling method is chaotic and the system can easily diverge to nonphysical conformations. To avoid this, the simulations should start from an energy minimum which guarantees the stability of the system in the initiation protocol (see section 2.1.5) since they correspond to high probability states.

The potential energy surface is described with a multidimensional function (PEF) which can be treated as a mathematical function with $3n$ dimensions (n being the number of atoms). This means that to minimize the energy we need a mathematical method to find minima in one dimension of a multidimensional function, which is a point where the gradient of PEF is zero and the determinant of Hessian matrix[1] is positive. In a typical biomolecular simulation, the number of minima is very large and it is virtually impossible to find the absolute minimum. The best solution is probably to use more than one method in sequential steps. We used two energy minimization methods: Steepest Descent and limited memory-Broyden-Fletcher-Goldfarb-Shanno (l-BFGS) [154].

Steepest descent is a robust and easy to implement algorithm in which the energy goes down along the potential energy surface in the direction of the maximum component of the force. This means that this method does not allow to jump between local minima since it will go quickly down to the nearest one (or, more exactly, the one in the direction of the steepest descent). Nevertheless, since it reaches very quickly the minimum (at least using a large convergence criteria) it is a very useful method for an initial step of a more complex minimization protocol or for small corrections in the conformation.

l-BFGS method is an improvement of the original BFGS (lowering the memory cost) which tries to approximate the inverse Hessian matrix but whose detailed explanation is beyond the scope of this text (see detailed information in Ref. 154). Unlike steepest descent, l-BFGS allows to jump between local minima and it is useful for complex systems. The main drawback of this method is the large computational cost.

[1] Hessian matrix is the square matrix of second-order partial derivatives of a function with n variables

2.1.5 Initiation

With the minimization step completed, the system is frozen (all velocities are zero) in an energy minimum (or very close to it). Before starting a production segment of MD, the system should be "heated" carefully to avoid instability and artifacts in the simulation. We start by using a seed to generate random velocities at the chosen temperature. In this first step it is common to use position restraints. This is simply an additional term ($V_{pr}(\vec{r}_i)$) added in equation 2.1 (PEF) which represents an artificial force acting on chosen atoms to limit their movement:

$$V_{pr}(\vec{r}_i) = \frac{1}{2} K_{pr} \left\| \vec{r}_i - \vec{R}_i \right\|^2 \tag{2.12}$$

where $\vec{R}_i$ is the reference position of atom i and K_{pr} is the force constant applied to that atom.

The initiation process is normally performed in several steps starting with large restraints. In the following steps K_{pr} is gradually decreased until the system is completely released. In these steps the coupling constants for temperature and pressure (see section 2.1.8) can be adjusted to avoid particles with a kinetic energy too large and oscillations in box size, respectively. When the system is stable, the initiation step can be stopped and production MD is started.

2.1.6 Periodic boundary conditions

In most cases, the models used in a biomolecular simulation try to approximate the effect of an infinite amount of solvent molecules surrounding the solute (or the membrane in our cases), i.e. they intend to simulate an infinite dilution (or an infinite membrane). However, to avoid large computational times, these models are restricted to a relatively small number of molecules. Hence, to correctly describe the desired effects, most microscopic simulation boxes are simulated with periodic boundary conditions (PBC) (Figure 2.2). In this method, the system is artificially surrounded by copies of itself which means that, when a particle leaves the box from one side it enters the box from the opposite side. This way the use of physical walls in a simulation box is avoided. PBC are particularly relevant in biomolecular models of membranes since they are able to describe an infinite bilayer in x and y directions of a tetragonal box in a infinite dilution z provided that long range electrostatics are treated properly. In this thesis, we give particular relevance to the treatment of PBC when solving the Poisson–Boltzmann equation (see section 2.2 and chapter 3).

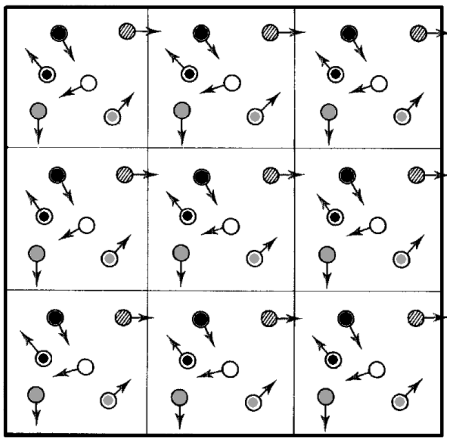


Figure 2.2: Representation of a periodic system in two dimensions simulated with periodic boundary conditions (adapted from Ref. 135).

2.1.7 Nonbonded interactions

We have stated that with PBC we can model an infinite system. However, this creates a problem when dealing with long range interactions since we cannot calculate them all (the number of particles is virtually infinite). The simplest method to solve this problem is by applying a cutoff, i.e. above a certain distance the interactions are truncated. This method works well for Lennard-Jones interactions since they decay to zero at relatively short interatomic distances [41]. However, for electrostatic interactions this is not the case and one would need a very large cutoff value to avoid artifacts. Moreover, even for Lennard-Jones interactions, the computational cost of a simple cutoff is still very large. A twin range method [134] is much more suitable for these cases. In this method, two cutoff values are used and the neighbors of each atom within the first cutoff are calculated in every step and the ones between the two cutoffs are only calculated at regular intervals. The approximation relies on the assumption that this last group of neighbors does not change during that interval. Since the neighboring search is one of the limiting steps in a MD simulation this saves a significant amount of computational time. With this simple trick we solved the problem for Lennard-Jones interaction. However, we still need to take into account the electrostatic contributions of the atoms beyond both cutoffs.

The treatment of long range electrostatics is a matter of broad discussion in the literature which mainly relies on two methods: reaction field (RF) [155] and particle mesh Ewald (PME) [156]. A discussion on these methods can be found in [157] or [134] and references therein. The mathematical treatment of these methods is complex and beyond the scope of this text in which we will restrict to qualitative aspects of both methods.

RF is a method in which the region beyond the cutoff is treated as homogeneous. This homogeneous medium has a dielectric constant that attenuates all electrostatic interactions within the cutoff region. In the generalized reaction field (GRF) method [155] the effect of ionic strength is added, which means that the interactions are also attenuated by it. The approximations made in RF works better in homogeneous systems, hence, in our membrane simulations, there could be some artifacts arising from this method.

PME [156] is a fast algorithm to calculate Ewald summations [158]. Ewald summations were developed to calculate the potential energy of ionic crystals which are periodic and neutral. It considers the effect of periodicity in crystals by considering it to be infinite without assuming an homogeneous region. This method was then applied to MD simulations since MD boxes are also periodic (when using PBC) and inhomogeneous. This means that PME is probably a good approach in a membrane environment since it will mimic an infinite bilayer in x and y directions of a tetragonal box without introducing a homogeneous region beyond a cutoff. However, in z direction it may add an artifact by considering the effect of other artificial membranes in both ways. Moreover, as it happens with crystals, PME boxes must be neutral which represents a problem when the molecules are significantly charged (in chapter 4 we explored this problem in detail).

2.1.8 Temperature / pressure

All MD steps were performed in the NPT ensemble. This means that the number of particles, pressure and temperature were the three thermodynamic properties kept constant in our simulations. The discussion of ensembles relies on the field of statistical thermodynamics and we address the interested reader to more specialized references such as [159].

To keep the thermodynamic temperature and pressure constants, we need to model an exterior bath that can give or extract kinetic energy from the system and that can properly resize the simulation box. We used Berendsen [160] and v-rescale [161] temperature baths and Berendsen [160] pressure bath.

Temperature

In the Berendsen temperature bath [160], the energy of the system is adjusted by scaling the velocities of atoms. This adjustment is proportional to the difference between bath and system temperatures:

$$\frac{\mathrm{d}T(t)}{\mathrm{d}t} = \frac{T_{bath} - T(t)}{\tau} \tag{2.13}$$

where τ is the temperature coupling constant which determines how the bath and system temperatures are coupled. With this method, the system tempera-

ture decays exponentially to the desired temperature. The temperature difference between two steps is given by:

$$\Delta T = \frac{\Delta t}{\tau} \left(T_{bath} - T(t) \right) \tag{2.14}$$

Finally, a scaling factor (λ) applied to the velocities is given by:

$$\lambda = \left[1 + \frac{\Delta t}{\tau} \left(\frac{T_{bath}}{T(t)} - 1 \right) \right]^{\frac{1}{2}} \tag{2.15}$$

The coupling strength depends on the value of τ. A high value of τ means a small temperature coupling and *vice versa*. Usually, a lower value of τ is used in the first steps of simulation to ensure that the temperature converges quickly to the desired value.

Berendsen temperature coupling is simple to implement and its performance is very high. However, with this method, there is no conserved property and it does not sample from a proper ensemble. As an alternative, the v-rescale thermostat [161] was shown to properly sample from the NVT ensemble (constant number of particles, volume and temperature) and it was developed to ensure a correct distribution of kinetic energy. Hence, we decided to use this thermostat in the later chapters of this thesis.

Pressure

In the Berendsen pressure coupling, the system is also coupled to an external bath and the volume of the simulation box (hence, atom positions) is adjusted to maintain the pressure. The velocity of this variation is given by:

$$\frac{\mathrm{d}P(t)}{\mathrm{d}t} = \frac{P_{bath} - P(t)}{\tau_P} \tag{2.16}$$

where τ_P is the coupling constant in analogy to τ present in temperature. The volume of the simulation box is scaled by a factor μ:

$$\mu = 1 - \beta \frac{\Delta t}{3\tau_P} \left(P_{bath} - P(t) \right) \tag{2.17}$$

where β is the isothermal compressibility of the system. In our membrane simulations, analogously to what happens in other works and to proteins, we used the isothermal compressibility of water since we do not have a proper estimation of this value for a membrane water mixture. This approximation is probably rougher in membrane systems than in proteins and, to avoid large oscillations, we use a larger τ_P.

The new atoms positions are given by:

$$\vec{r_i'} = \mu^{\frac{1}{3}} \vec{r_i} \tag{2.18}$$

Once again, a higher value of τ_P, results in a lower coupling of pressure to the desired value and $vice - versa$.

2.1.9 Constraints

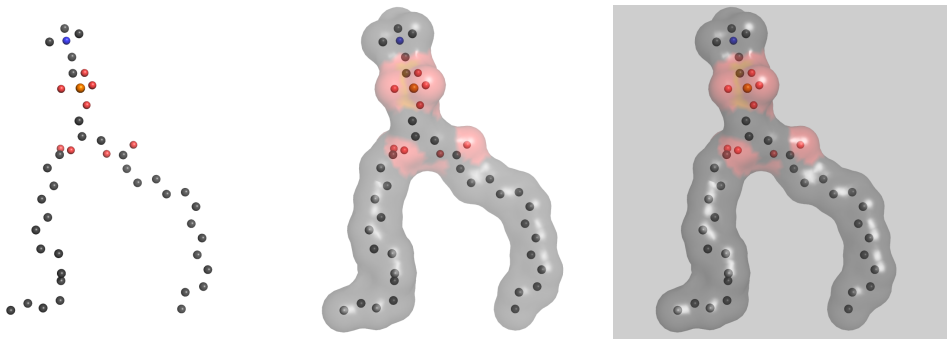
In MD simulations the quality of the results depends on the chosen time step (see section 2.1.3). Using 2 fs (see section 2.1.3) we probably obtain an incorrect description of bond stretching. To solve this issue, maintaining this time step, we used bond constraints. This means that the bond distances are forced to maintain their original values (b_{0_n} in equation 2.2).

To constrain bond distances in the molecules we used the parallel version of LINCS algorithm [162, 163]. In this algorithm bond lengths are allowed to change in the MD step and are restored *a posteriori*. This is performed in two steps where the forces acting in the bond are projected along the original bond and then the bond length is adjusted to b_{0_n} (Figure 1 in Ref. 162).

In some cases, it is also useful to constrain angles as in water molecules. SETTLE algorithm [164] is an analytical solution of another (SHAKE [165]) specifically solved for water molecules. This makes the simulation of water molecules much faster which is particularly relevant since, in most biomolecular simulations, water molecules correspond to more than half of the system.

2.2 Continuum electrostatics and $\mathrm{p}K_\mathrm{a}$ calculations

The electrostatic behavior of a biological system is very useful in the understanding of many biochemical processes [49, 76, 166–169]. In particular, charged and polar groups, which are found ubiquitously in biological macromolecules, have important roles in processes such as enzymatic catalysis, protein stability or biomolecular recognition [170]. Some of these groups have fractional protonation states at some pH values, which means that they are continuously exchanging protons with the solvent as they change their conformation (a possible way to model this is the scope of this thesis and is achieved in chapter 5). In this section we just assume that the bilayer (or any other solute) is frozen and, using the Poisson–Boltzmann (PB) [168, 171–173] equation, we shall discuss how to calculate protonation energies that can then be used to sample protonation states with a Monte Carlo simulation [137] (section 2.3).



(a) DMPC atomic charges (b) DMPC reorganization ϵ_{in} (c) Water reorganization ϵ_{out} and inclusion of ionic strength I

Figure 2.3: Representation of Poisson–Boltzmann model of a DMPC molecule.

2.2.1 Poisson–Boltzmann equation

To calculate the electrostatic potential energy of a bilayer (or any other system — we shall just assume that it is a bilayer for simplicity) with PB equation, some approximations are necessary. In this model, the membrane is described as a continuous region with a single value of dielectric constant ϵ_{in} (Figure 2.3b) and a set of partial atomic charges (Figure 2.3a). The dielectric constant is a measure of how fast can molecules reorganize in response to an electric field. The higher the dielectric constant the more reorganizable is the material. Thus, ϵ_{in} describes an instantaneous reorganization of the dipoles in the membrane. The same approximation is done for the solvent, which is treated implicitly with a single value of dielectric constant ϵ_{out} (Figure 2.3c). Normally, a low value of dielectric constant is used for the membrane and a high value for water (usually 80). The dielectric constant of water is very large due to its ability to quickly adopt a new conformation around a charge. Similarly to what happens in generalized reaction field (section 2.1.7), the ionic strength is treated implicitly as an attenuation of electrostatic interactions between atomic charges (Figure 2.3c).

With our model, we can calculate the electrostatic potential using the linearized form of PB equation:

$$\nabla \cdot [\epsilon(\vec{r})\nabla\Phi(\vec{r})] - \kappa^2(\vec{r})\epsilon(\vec{r})\Phi(\vec{r}) = -4\pi\rho(\vec{r}) \tag{2.19}$$

where $\vec{r}$ is the vector of all atom positions, $\epsilon(\vec{r})$ the dielectric constant, $\Phi(\vec{r})$ the electrostatic potential, $\rho(\vec{r})$ the charge density and $\kappa(\vec{r})$ the so-called reciprocal

$$\begin{array}{ccc} \sum_i^N \mathrm{A}^{\mathrm{sol}}(\vec{0}) + N\mathrm{H}^+ & \xrightarrow{\Delta G^{sol}_{\vec{0}\to\vec{a}}} & \sum_i^N \mathrm{A}^{\mathrm{sol}}(\vec{a}) \\ \Delta G_{\mathrm{sol}\to\mathrm{A}}(\vec{0}) \downarrow & & \downarrow \Delta G_{\mathrm{sol}\to\mathrm{A}}(\vec{a}) \\ \mathrm{A}^{\mathrm{A}}(\vec{0}) + N\mathrm{H}^+ & \xrightarrow{\Delta G^{\mathrm{A}}_{\vec{0}\to\vec{a}}} & \mathrm{A}^{\mathrm{A}}(\vec{a}) \end{array}$$

Figure 2.4: Thermodynamic cycle involving membrane and model compounds.

Debye length ($\kappa(\vec{r})$):

$$\kappa(\vec{r}) = \begin{cases} \left(\dfrac{8\pi e^2 I}{\epsilon_{out} k_{\mathrm{B}} T}\right)^{1/2} & \text{if } \vec{r} \text{ is in an ion accessible region} \\ 0 & \text{otherwise} \end{cases} \tag{2.20}$$

where k_{B} is the Boltzmann constant and T the absolute temperature. $\kappa(\vec{r})$ is zero when the $\vec{r}$ is in an inaccessible region, i.e. the membrane interior. PB equation does not have an analytical solution (unless in very special cases) and must be solved using a numerical method involving, in our case, a cubic lattice and a finite difference procedure [174, 175].

We finally calculate the electrostatic potential energy of our membrane ($\mathcal{V}$) which is defined as the work necessary to bring a certain charge distribution together:

$$\mathcal{V} = \frac{1}{2}\sum_i^N q_i \Phi(\vec{r_i}) \tag{2.21}$$

where the summation runs over all N atoms and q_i is the charge of atom i.

2.2.2 Protonation free energy

Imagine a membrane or a solute A whose protonation state is defined with a vector $\vec{a}$ of 0's and 1's where 0 (1) means that the site is neutral (ionized)[2]. The free energy of converting between some arbitrarily chosen reference state ($\vec{0}$ - usually with all sites neutral) and state $\vec{a}$ with N additional additional protons can be estimated using a thermodynamic cycle (Figure 2.4) and is given by:

$$\Delta G^{\mathrm{A}}_{\vec{0}\to\vec{a}} = \Delta G^{sol}_{\vec{0}\to\vec{a}} + \Delta G_{\mathrm{sol}\to\mathrm{A}}(\vec{a}) - \Delta G_{\mathrm{sol}\to\mathrm{A}}(\vec{0}) \tag{2.22}$$

[2]In the present formalism we ignore the proton tautomerism. To a discussion on this topic see reference [176]

where $\Delta G^{\mathrm{A}}_{\vec{0}\to\vec{a}}$ is the free energy associated with the protonation of N sites in the membrane A environment, $\Delta G_{\mathrm{sol}\to\mathrm{A}}(\vec{0})$ is the free energy associated with the transition from the water to the membrane environment in the $\vec{0}$ state, and analogously for other two terms. $\Delta G^{sol}_{\vec{0}\to\vec{a}}$ is in fact a summation of the free energy of protonation of model compounds. A model compound is a part of the titrable group surrounded by a continuous high dielectric medium (water) and has a characteristic pK_{a} value (pK^{mod}) which can be converted to free energy of protonation using the general equation:

$$\mathrm{p}K = \frac{\Delta G}{2.3 k_{\mathrm{B}} T} \tag{2.23}$$

where, in electrostatic models, the Gibbs free energy of solvation (ΔG) is simply regarded as the electrostatic potential energy (equation 2.21). Finally, from equation 2.22 it can be shown (see Ref. 177 for example) that $\Delta G^{\mathrm{A}}_{\vec{0}\to\vec{a}}$ is given by:

$$\Delta G^{\mathrm{A}}_{\vec{0}\to\vec{a}} = -2.3 k_{\mathrm{B}} T \sum_{i}^{N} a_i \gamma_i \mathrm{p}K_i^{\mathrm{int}} + \sum_{i}^{N}\sum_{j\neq i}^{N} a_i a_j \Delta\mathcal{V}_{ij} \tag{2.24}$$

where γ_i is the charge of site i when it is ionized and $\Delta\mathcal{V}_{ij}$ is the interaction free energy between sites i and j and is given by:

$$\Delta\mathcal{V}_{ij} = \mathcal{V}(1,1) - \mathcal{V}(1,0) - \mathcal{V}(0,1) + \mathcal{V}(0,0) \tag{2.25}$$

pK_i^{int} is the intrinsic pK_{a} of site i which includes the pK^{mod} and the interaction with all other titrating sites in their reference state ($\Delta\mathcal{V}_{int}$):

$$\mathrm{p}K^{\mathrm{int}} = \mathrm{p}K^{\mathrm{mod}} - \frac{1}{2.3 k_{\mathrm{B}} T}\Delta\mathcal{V}_{int} \tag{2.26}$$

$\Delta\mathcal{V}_{int}$ corresponds to $\Delta\Delta\ G_{\mathrm{Solv}} + \Delta\Delta\ G_{\mathrm{Back}}$ in chapter 3, where it is explained how it is calculated. This ends our derivation since we already have an equation for the protonation free energy with terms that can be directly obtained from a PB calculation.

2.3 Monte Carlo

Similarly to previous section, here we kept the assumption that the membrane is frozen. With all electrostatic contributions calculated, we now want a reasonable protonation state for our membrane at a given pH value. The first problem is that the number of titrable sites is often very large which renders the calculation of all $\Delta G^{\mathrm{A}}_{\vec{0}\to\vec{a}}$ virtually impossible. One possible way to circumvent this problem is by

running a Monte Carlo (MC) simulation using Metropolis criterion [137] to sample protonation states [176, 178, 179].

We start by the definition of $\Delta G^{\mathrm{A}}_{\vec{0}\to\vec{a}}(\mathrm{pH})$ which is the pH *dependent* Gibbs free energy of converting state $\vec{0}$ into state $\vec{a}$ [180]:

$$\Delta G^{\mathrm{A}}_{\vec{0}\to\vec{a}}(\mathrm{pH}) = -2.3 k_{\mathrm{B}} T \sum_{i}^{N} a_i \gamma_i (\mathrm{p}K_{\mathrm{int},i} - \mathrm{pH}) + \sum_{i}^{N} \sum_{j\neq i}^{N} a_i a_j \Delta \mathcal{V}_{ij} \tag{2.27}$$

In our notation, the free energy required to change the protonation state from a state $\vec{a_1}$ to another $\vec{a_2}$ can be expressed as a $\Delta\Delta G$:

$$\Delta\Delta G^{\mathrm{A}}_{\vec{a_1}\to\vec{a_2}}(\mathrm{pH}) = \Delta G^{\mathrm{A}}_{\vec{0}\to\vec{a_2}}(\mathrm{pH}) - \Delta G^{\mathrm{A}}_{\vec{0}\to\vec{a_1}}(\mathrm{pH}) \tag{2.28}$$

which will simplify to $\Delta G_{\vec{a_1}\to\vec{a_2}}(\mathrm{pH})$.

The transition from $\vec{a_1}$ to $\vec{a_2}$ is the MC move and, according to the Metropolis criterion the new state is accepted if:

$$\Delta G_{\vec{a_1}\to\vec{a_2}}(\mathrm{pH}) < 0 \tag{2.29}$$

And it is accepted with a probability of

$$e^{-\Delta G_{\vec{a_1}\to\vec{a_2}}(\mathrm{pH})/RT} \qquad \text{if} \qquad \Delta G_{\vec{a_1}\to\vec{a_2}}(\mathrm{pH}) > 0 \tag{2.30}$$

The MC simulation consists of a sequence of MC moves and it was shown that, with this sampling technique, the protonation states are being properly sampled [178, 179]. To obtain a reasonable sampling in a relatively short computational time it is important that the difference between $\vec{a_1}$ and $\vec{a_2}$ is 1 or 0 protons (as it happens with coupled sites).

2.3.1 Coupling between titrable sites

Normally, a MC move corresponds to protonate or deprotonate a single site. However, if two sites are coupled, i.e. if they interact strongly, it can be useful to change the protonation states of both sites in one move. For example, for two carboxylic groups close to each other, at some pH values, in the more stable protonation state one of them will be protonated and the other deprotonated ((0,1) or (1,0)). Moreover, it is expected that the proton can change between the two groups. Since, from one state to the other, both titrable groups formally change their protonation state it is mandatory to consider coupled MC moves to properly model these cases.

2.3.2 Titration curves and probable protonation states

After some equilibration MC moves, we are sampling physically relevant protonation states. In the case of a constant-pH MD simulation (section 2.4) we can choose one of them and continue to the next step. Moreover, it is possible to estimate the titration curve of our molecule by performing MC simulations at several pH values and calculate the average protonation state at each pH. For simple cases, it is also possible to calculate the pK_a values of each protonation state. We are still assuming a rigid structure and the obtained titration curves and pK_a values are inaccurate in many cases.

2.4 Stochastic titration constant-pH MD

In this section we shall briefly introduce the former version of the stochastic titration constant-pH MD method [12, 13]. All other details are included mainly in chapter 5. In this method, the dynamic behavior of the system is simulated by MD, which can properly sample the conformational space of biomolecules with high accuracy and the protonation state of the membrane is sampled using a PB / MC methodology (Figure 2.5). The method works in a cycle with three main steps, with the loop running until the end of the desired simulation time:

- **PB / MC** - the first step is a PB / MC calculation to choose a suitable protonation state for the membrane
- **MM / MD of solvent** - a short MD simulation is then performed with the membrane frozen to allow the water molecules to adapt to the new protonation state
- **MM / MD** - the effective MD simulation is then performed with all the system unfrozen. The last conformation is then used as input to the next PB / MC step

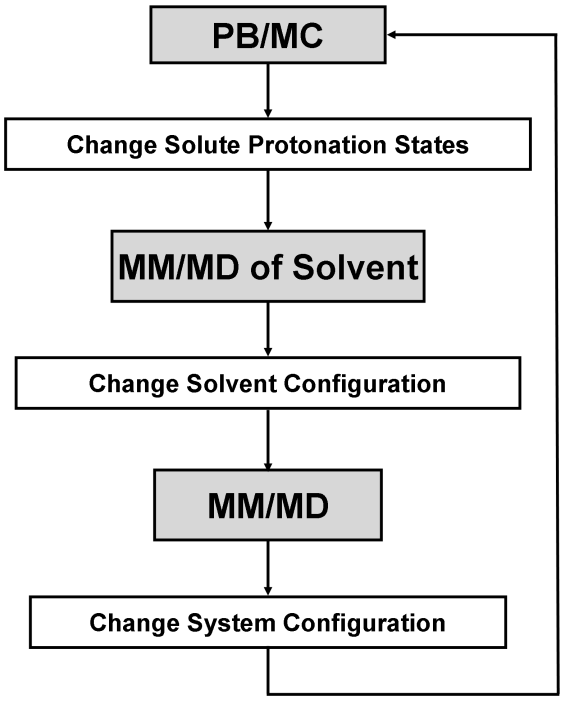
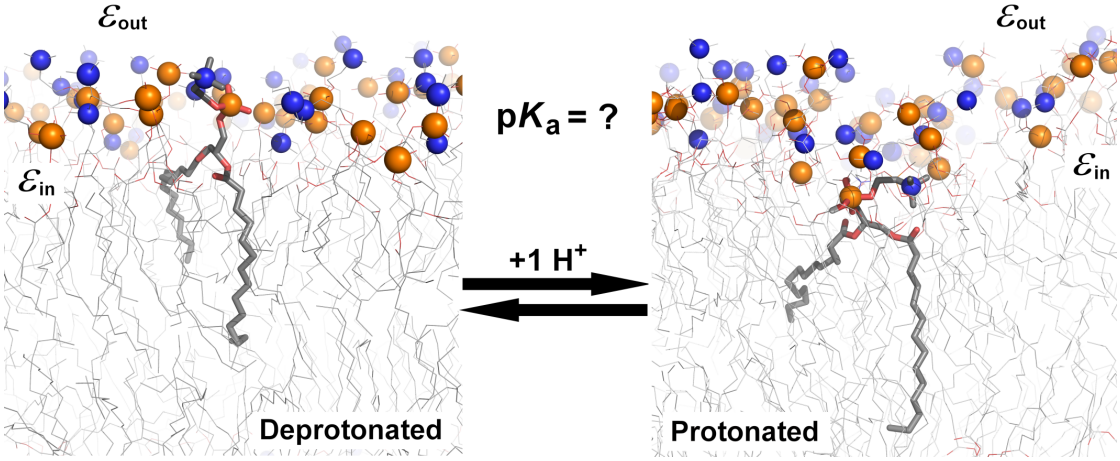


Figure 2.5: Scheme of the stochastic constant-pH MD algorithm (adapted from Ref. 12).

Chapter 3

Protonation of DMPC in a bilayer environment using a linear response approximation

Vitor H. Teixeira, Diogo Vila-Viçosa[1], António M. Baptista, Miguel Machuqueiro



Contents

[1]In this work, I performed some initial simulations and all charge parametrizations. I was also part of all the discussions of the results and of manuscript preparation.

3.1 Context

The first and major goal of this thesis was to develop a constant-pH molecular dynamics method that allows the simulation of a lipid bilayer with periodic boundary conditions (PBC) in x and y directions. However, none of the Poisson–Boltzmann (PB) solvers available at the beginning of this work were able to deal with both PBC and pK_a calculations. Hence, the first project was developed to overcome this technical limitation. We used the DelPhi (an available PB solver with PBC) software in a completely new approach to perform the pK_a calculations which were used in all chapters of this thesis, particularly in chapter 5 where the final constant-pH method is presented. In this chapter, we used a linear response approximation to calculate the pH at which a DMPC molecule in the bilayer environment starts to protonate. This work was published in Journal of Chemical Theory and Computation (2014,10:2176. DOI: 10.1021/ct5000082) and it is available at `http://pubs.acs.org/doi/abs/10.1021/ct5000082`.

3.2 Abstract

pH is a very important property, influencing all important biomolecules such as proteins, nucleic acids, and lipids. The effect of pH on proteins has been the subject of many computational works in recent years. However, the same has not been done for lipids, especially in their most biologically relevant environment: the bilayer. A reason for this is the inherent technical difficulty in dealing with this type of

periodic systems. Here, we tackle this problem by developing a Poisson–Boltzmann based method that takes in consideration the periodic boundary conditions of lipid bilayer patches. We used this approach with a linear response approximation to calculate the pK_a value of a DMPC molecule when diluted in zwitterionic lipids. Our results show that DMPC protonation only becomes relevant at quite low pH values (2–3). However, when it happens, it has a strong impact on lipid conformations, leading to significant heterogeneity in the membrane.

3.3 Introduction

Solution pH is a fundamental physicochemical property that directly influences the structure and function of various types of biological molecules [168, 169, 177, 181–183]. Because of several difficulties inherent to these systems, it was only recently that an increasing number of works started to appear in the literature. Lipid protonation leads to changes in the electrostatic environment, which will affect the lipid bilayer, directly or indirectly interfering in many membrane–mediated biological processes such as membrane fusion, domain formation, drug–liposome interactions, and membrane phase transitions [184–188]. The mitochondrial membrane plays a crucial role in ATP synthesis (respiratory chain) which is dependent on the pH gradient across the membrane together with the membrane potential [189–192].

The interpretation of the pH effect on a lipid bilayer and on its electrostatic environment is a challenging task. Theoretical methods can be very useful to understand this problem at a molecular level. With the constant evolution of computer power and algorithms, larger and more complex model systems have been treated using molecular modeling methods. Simulations with membrane systems are only now being fully explored, because of the extra complexity comparatively to protein systems. The first studies on membrane electrostatics used a simplified theory, which assumes that the charges on phospholipids are uniformly spread over a planar surface and that the solvent is a structureless medium with a uniform dielectric constant (Gouy–Chapman theory) [193]. This simple theory was successfully used to explain the ζ potential of phospholipid vesicles [194], or how ions affect the surface potential above the phospholipid monolayers [195]. It was also shown that, for charged and well organized membranes, the Gouy–Chapman theory is in good agreement with a more detailed system representation with Poisson–Boltzmann (PB) theory [81, 196]. Other works have analyzed the electrostatic interaction of proteins with charged flat surfaces [197] or discretized charged surfaces [198], using the linearized form of the PB equation, to evaluate the best orientation and distance of interaction. Membrane interactions with peptides were also studied using lysine as a model with a mixture of anionic and zwitterionic phospholipids

to access the binding free energy dependence on the number of lysine residues, the salt concentration and percentage of acidic phospholipids [196]. From PB calculations, using several snapshots from molecular dynamics (MD) simulations it was also possible to analyze the potential roughness near the membrane surface [94].

The first attempt to calculate a pK_a shift in a membrane–like environment used a monolayer of docosyl amines treated with two-dimensional periodic boundary conditions (2D-PBC) in an air/water interface [82], with the electrostatic potential obtained from the nonlinear PB equation. More recently, the lipid titration problem has also been addressed in micellar systems [113, 116, 199, 200]. A mix of Monte Carlo (MC) and PB techniques were used trying to determine the pK_a shift of lauric acid in a neutral [199] and ionic micellar environment [199, 200]; in Ref. 199, MC calculations are done in a discrete system that defines a radius for the hydrophobic region and includes the "lipid headgroups" in the solvent region to be given a high dielectric constant. Recently, a constant-pH MD methodology [105] was applied to study the pK_a of lauric acid under different conditions [113, 116] namely in micellar and membranelike supramolecular assemblies. In all these systems the solutes were always completely solvated; therefore, the PBC were not taken into account [113, 116, 199, 200] in the electrostatic calculations.

The objective of this work is to present a PB-based method to estimate the electrostatic potential in an infinite bilayer environment with atomic detail and use it for pK_a calculations. It borrows the theoretical/computational aspects associated with the thermodynamic cycle used in protein pK_a calculations, but requires several additional features inherent to the two-dimensional periodicity of membrane systems. PB calculations were performed on conformations from MD simulations of a patch of phospholipids of finite dimensions. The inclusion of 2D-PBC in the electrostatic potential calculation in the membrane patches allows for an unprecedented realism in PB-based methods used in pK_a calculations of phospholipids. This new method was applied to the known zwitterionic phospholipid 1,2-dimyristoyl-sn-glycero-3-phosphocholine (DMPC). This is a well parametrized system for which there is considerable structural information at the lamellar phase [26–32, 34–40] and solid parameters for most biomolecular force fields [138]. The pK_a of the phosphate group in a DMPC bilayer is not known, but it its usually assumed to be very low, as the deprotonated (charged) form is usually used in MD simulations. We used a linear response approximation (LRA) [201, 202] to calculate the pK_a value of one DMPC diluted in other zwitterionic DMPC molecules, thus providing an estimate of the pH where protonation of this lipid becomes relevant.

3.4 Methods

3.4.1 Theoretical considerations

One standard approach to determine the pK_a of protonatable groups is to calculate electrostatic properties using the PB equation, followed by a calculation of the probability of each protonation state along a pH range [176, 203, 204]. The PB equation can be solved through several methods such as finite differences, boundary element, and finite element. In the finite difference method [175], as implemented in many programs, the potential is iteratively calculated in each grid point until convergence. In this iterative process an initial guess is made for the electrostatic potential at each grid point. The electrostatic potential at the boundaries of the grid does not change, keeping the first initial guess (Dirichlet boundary conditions). The described procedure works well for a localized system, such as a protein in solution (with approximate spherical symmetry), but that represents a problem when dealing with "infinite" systems like biological membranes (with approximate cylindrical symmetry), since the usual methods to estimate the initial potential (Coulombic, dipolar, etc.) are inadequate to take into account the virtually infinite membrane dielectric medium. A simple way to circumvent this problem is to use periodic boundary conditions along the directions parallel to the membrane (2D-PBC). The use of this approach in the calculation of the electrostatic potential allow us to determine pK_a values of phospholipids embedded in bilayers.

Even though DelPhi [205, 206] is able to compute the electrostatic potential using periodicity, no such option exists for the surface calculation. When applied to a periodic MD configuration of the bilayer, this produces an irregular molecular shape near the walls of the finite difference grid, reflecting the way the system happened to be split across the walls of the MD simulation box. When this non-periodic shape is used to assign the dielectric constant over the grid, two types of problems may arise in the bilayer regions near the walls (Figure 3.1). The first problem occurs when a grid point that should be in the low-dielectric region is assigned to the solvent region, because the nearby atomic sphere that would enclose it in the periodic case happens to be located on the opposite side of the grid in the nonperiodic case. The second problem is related to the fact that, in the radius assignment method adopted here (section 3.4.4), some hydrogen atoms are assigned a zero radius while remaining enclosed in the sphere of their non-hydrogen partner (this follows from their Lennard-Jones parameters in the GROMOS 54A7 force field); when such a hydrogen atom is near a wall, the nonperiodic treatment may place it and its partner on opposite sides of the grid, leaving the hydrogen not enclosed in any sphere and assigning its charge to the solvent region. In order to solve these problems, we added a small portion (5% of the box side dimension) of the system atoms in both x and y directions along which 2D-PBC are used (Fig-

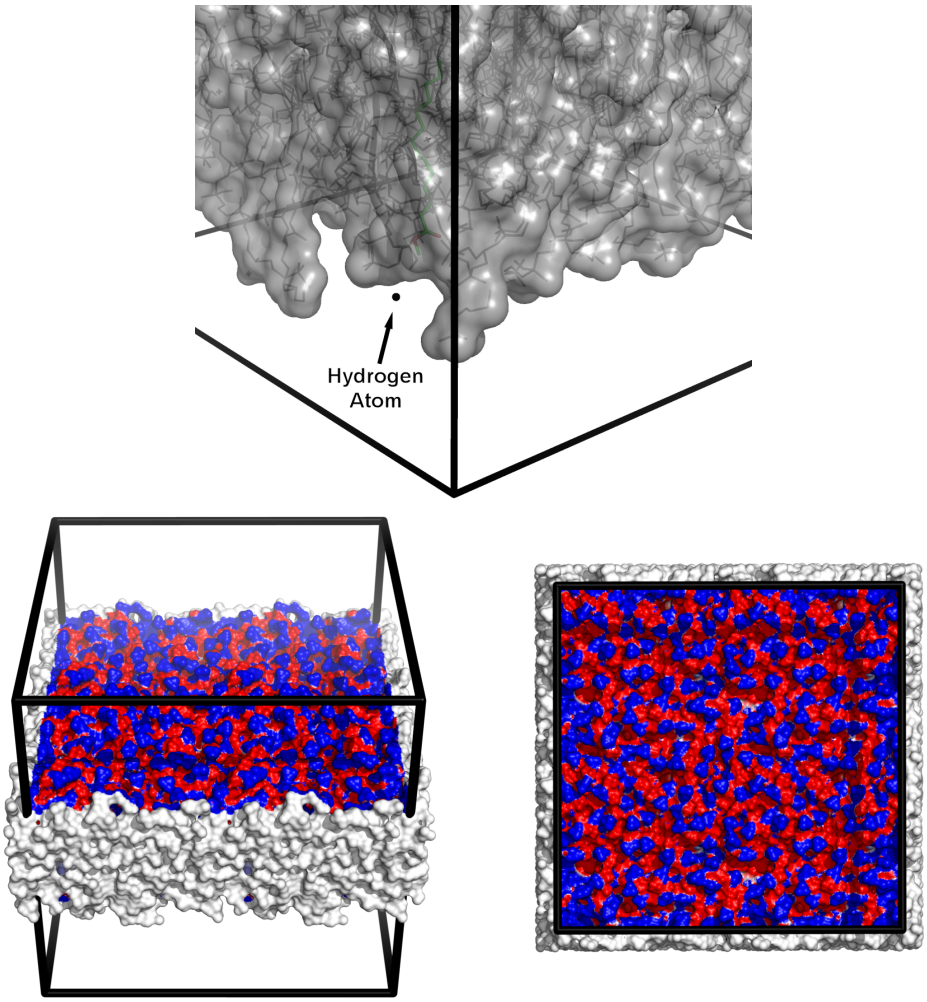


Figure 3.1: The DMPC surface discontinuity problem (top) where a hydrogen atom is left alone with a bond "broken" by PBC. Surface in tilted (bottom left) and top (bottom right) views with a 5% extra layer to ensure the continuity of the molecular surface at the box edges.

ure 3.1), with the added atoms being assigned their usual radius but a zero atomic charge. This procedure ensures the continuity of the bilayer molecular surface at the box walls and avoids placing charges in the solvent.

Another problem is when a titrating group falls on the boundary, split across the box, with atoms appearing on the other side of the box through 2D-PBC. This situation presents difficulties in the application of the focusing procedure, where a second finer and smaller box is used to calculate the electrostatic potential without 2D-PBC. One of the limitations is that the smaller box must be confined inside the larger and coarser one, impairing its application to a residue at the box edge. One simple way of preventing this problem, is to create an equivalent configuration in which the titrating group is centered on the grid. With this x/y centering procedure, the atomic charges will always be in the center of the grid.

Usually proteins can simply be inserted inside a box large enough to enclose it and to properly estimate the electrostatic terms needed to the pK_a calculation. However, in a membrane system, there is 2D-PBC and we should not calculate interactions with groups that are farther than half the x/y grid vector dimension. This problem can be circumvented by calculating pairwise and background interactions only within a given cutoff. This approximation can be considered small because only the titrating group is charged in our pK_a calculation procedure, and therefore the electrostatic potential difference should be small at longer distances. We investigated the influence of cutoff size by comparing the final pK_a value obtained with the largest cutoff possible (30 Å). The calculations were done with the same conformation using different cutoff distances and allowed all 128 DMPC phospholipids to titrate (Figure 3.2). At short distances the pK_a deviations are significant, but as expected, at higher cutoff distances, most relevant sites are included in the calculations and the deviations decrease. Although for this particular membrane bilayer conformation we were able to use a maximum cutoff value of 30 Å, in a NPT MD simulation the system lateral size may become shorter than 60 Å. To prevent these possible problems we choose a safe cutoff value of 25 Å in this work.

Despite the above mentioned procedures, the calculation of pK_a values in phospholipids embedded in a membrane is similar to the calculation in proteins. The methodology used here for the pK_a calculation is similar to the one previously published [203]. To calculate the pK_a value of a phosphate group of a DMPC molecule, we can use equation 3.1

$$\mathrm{p}K_i = \mathrm{p}K_i^{\mathrm{int}} + \sum_{j \neq i} W_{ij} \,, \tag{3.1}$$

where the first term is the intrinsic pK_a and the second term corresponds to the direct interactions between different titrating groups of the system. pK^{int} is the

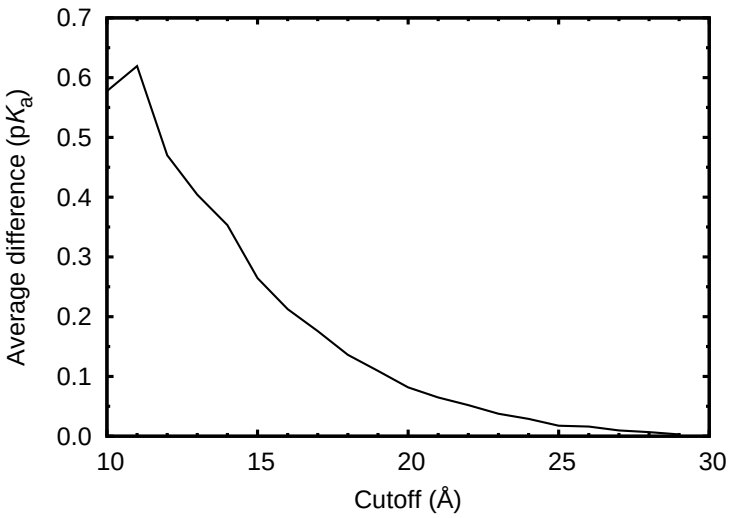


Figure 3.2: Average pK_a values differences at various cutoff radius. The difference was calculated relative to the value obtained with a cutoff value of 30 Å.

pK when all other titrable groups are fixed in a reference state, which, in our case, is the charged state [176] and can be written as:

$$\mathrm{p}K^{\mathrm{int}} = \mathrm{p}K^{\mathrm{mod}} - \gamma \, \frac{(\Delta\Delta G_{\mathrm{Solv}} + \Delta\Delta G_{\mathrm{Back}})}{2.3\, k_{\mathrm{B}} T} \, , \tag{3.2}$$

where pK^{mod} is the pK of the model compound; γ is 1 or -1, depending on the group being cationic or anionic, respectively; $\Delta\Delta G_{\mathrm{Solv}}$ is the solvation free energy difference; $\Delta\Delta G_{\mathrm{Back}}$ is the background contribution difference; k_{B} is the Boltzmann constant; and T is the absolute temperature. The solvation free energy of transferring a molecule between two different dielectric media can be obtained directly from the output of the DelPhi [205, 206] program, using the induced charge method (see Ref. 205 for details). The background contribution is the interaction energy from nontitrating partial charges and can be computed as:

$$\Delta\Delta G_{\mathrm{Back}} = \sum_i q_i [\phi^{\mathrm{C}}_{\mathrm{memb}}(r_i) - \phi^{\mathrm{N}}_{\mathrm{memb}}(r_i)] - \sum_i q_i [\phi^{\mathrm{C}}_{\mathrm{model}}(r_i) - \phi^{\mathrm{N}}_{\mathrm{model}}(r_i)] \, , \tag{3.3}$$

where q_i are the nontitrating partial charges located at r_i, and the superscripts C and N denote charged and neutral states, respectively. The second term in equation 3.3, relative to the model compound, is zero.

The last term in equation 3.1 corresponds to the direct interactions between pairs of titrating groups of the system and is calculated as:

$$W_{ij} = w_{ij}(1,1) - w_{ij}(1,0) - w_{ij}(0,1) + w_{ij}(0,0) \, , \tag{3.4}$$

where $w_{ij}(n_i, n_j)$ is the electrostatic interaction energy between sites i and j. The four terms in equation 3.4 correspond to the possible combinations of the protonation states of a pair of sites and, 1 and 0 stands for protonated or deprotonated, respectively. As an example, the second term in equation 3.4 gives the electrostatic interaction energy when site i is protonated and site j is deprotonated and can be calculated as:

$$w_{ij}(1,0) = \frac{1}{2}\sum_i (Q_i^{\mathrm{C}}\ \phi_i^{j^{\mathrm{N}}}) + \frac{1}{2}\sum_j (Q_j^{\mathrm{N}}\ \phi_j^{i^{\mathrm{C}}})\ , \tag{3.5}$$

where the first sum runs over all atoms of site i and the second over all atoms of site j. Q_x^y is the nth charge of site x in the protonation state y and $\phi_x^{z^y}$ is the electrostatic potential on site x when site z is in protonation state y.

3.4.2 Parametrization of DMPC molecule

There are atomic partial charges for DMPC in the literature [207], but only for the zwitterionic form where the phosphate group is deprotonated. Since our objective is to do pK_{a} calculations of DMPC, we need atomic partial charges of this phospholipid in both protonated and deprotonated forms. We determined a new set of atomic partial charges following a procedure similar to that previously used for the zwitterionic form [207]. DMPC molecule geometry was optimized with Gaussian 03 [208] at the HF/6-31G(d) level, and the resulting electrostatic potential was fitted to atomic coordinates with RESP [209] with P and N fixed to the Mulliken charges. The final atomic charges were slightly adjusted to be consistent with the charge groups of the GROMOS96 54A7 force field (Figure 3.3).

3.4.3 MM/MD

Molecular mechanics/dynamics simulations were performed with GROMACS version 4.0.7 [210, 211]. The GROMOS 54A7 force field [212, 213] was used, which already includes the DMPC corrections [138], together with the atomic partial charges determined in this work. All MD simulations started from an equilibrated lipid bilayer with 128 DMPC molecules and 4038 SPC water molecules [214], totaling 18 770 atoms per system. An orthorhombic simulation box was used, applying periodic boundary conditions in all three directions and the minimum image convention, as usual when using an explicit solvent. The x/y and z box vectors varied around 62 and 69 Å, respectively, ensuring that the membrane does not see its periodic image along the z direction. The simulations were done in the NPT ensemble, using Berendsen heat baths [160] at 310 K with separate couplings for the solute and solvent with a relaxation time of 0.1 ps, while a Berendsen semi-isotropic pressure couple [160] was used to keep the pressure at 1 bar with isothermal compressibility of 4.5×10^{-5} bar^{-1} and a relaxation time of 2.0 ps. All lipid bond

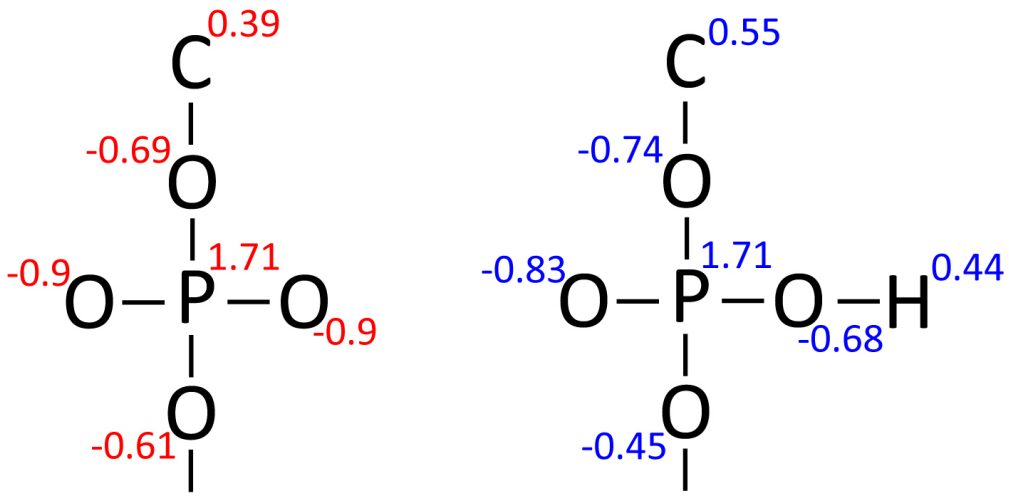


Figure 3.3: Partial atomic charges of the deprotonated (left) and protonated (right) phosphate groups of DMPC.

lengths were constrained using the parallel version of the LINCS algorithm [163], and the SETTLE algorithm [164] was used for water. The equations of motion were integrated using a time step of 2 fs. Nonbonded interactions were treated with a twin-range method, using group-based cutoffs of 8 and 14 Å, updated every 5 steps (10 ps). Long-range electrostatics thus truncated were corrected with a reaction field, using a dielectric constant of 54.0 [215]. No ions were added to the systems.

A three step energy minimization was performed consisting of a steepest-descent calculation followed by another using the l-BFGS algorithm without constraints nor restraints. A final minimization with steepest-descent using constraints in all bonds was performed.

Initialization of the molecular dynamics runs occurred in 2 steps. First, a 50 ps MD simulation was done with all heavy atoms position-restrained with a force constant of 1 000 kJ/(mol nm^2). Initial velocities were taken from a Maxwellian distribution at 310 K. In the second step, a 200 ps MD simulation was done with position restraints in all P atoms using a force constant of 1 000 kJ/(mol nm^2).

Table 3.1 shows the list of production runs performed. System Sys1 presents all 128 lipid molecules in the deprotonated zwitterionic form, while systems Sys2a-e have two DMPC molecules protonated, one in each monolayer. The 10 protonated phospholipids were selected randomly. For analysis, these 10 phospholipids were also chosen in Sys1, even although we could have taken any of the 128 lipids, since they are all equivalent.

Table 3.1: MD simulations performed with DMPC bilayers with the random selected phospholipids in each monolayer. The numbers refer to the topological lipid residue number in the simulations.

System	**State**	**lipid 1**[a]	**lipid 2**[a]
Sys1	Deprot	-	-
Sys2a	Prot	54	65
Sys2b	Prot	4	119
Sys2c	Prot	16	99
Sys2d	Prot	24	75
Sys2e	Prot	50	51

a) lipid 1 and 2 are located in different monolayers

All simulations started from the same pre-equilibrated simulation and were run for 500 ns each. Calculations and analysis were performed in the last 450 ns equilibrated segments of each system (Table 3.1), with conformations saved every 10 ps.

3.4.4 Poisson–Boltzmann / Monte Carlo

The Poisson–Boltzmann (PB) calculations were done with the program DelPhi V5.1 [205, 206], using radii taken from the GROMOS 54A7 force field [212, 213] and partial charges derived in this work (see above). The molecular surface of the membrane was defined by a probe of radius 1.4 Å, the ion exclusion layer was 2.0 Å and the ionic strength was 0.1 M. The dielectric constant of the solvent was 80, and for the membrane, we used values of 2, 4, 6, and 8 (see results and Refs. 202, 216–218 for a discussion of these values). In the coarse grid, we used relaxation parameters of 0.2 and 0.75 to help both the linear and nonlinear iteration convergence processes. Considering the membrane oriented along the xy plane, we explicitly applied PBC in the calculation of the potential in both the x and y directions in the coarse grid (2D-PBC).

A cutoff of 25 Å (see results) was used to calculate the background interactions. The convergence threshold value based on maximum change of potential was set to 0.01. The PB calculations were done in a cubic grid with 61 grid points and a two step focusing [219] where the focus grid was one fourth of the coarser grid size. The conformations were taken from the MD simulations preformed in NPT. Since the dimension of the box changes along the simulation, the space between grid points for the coarser grid must be calculated for every conformation (average values of ∼1.0 and ∼0.25 Å for the coarse and focus grids, respectively). The

Monte Carlo (MC) sampling was performed with the PETIT program [176]. All runs were performed using 10^5 MC cycles, where one cycle consists of sequential state changes over all individual sites.

Conformations for PB/MC calculations were taken from the six MD simulations (Table 3.1) from the equilibrated last 450 ns of each system at intervals of 10 ps. This makes 45 000 frames per system in a total of 270 000 frames. Since the titrating groups (Table 3.1) are to far away from each other, they do not interact and are not included in the pairwise interaction calculations. This way, the groups can be treated independently with increased sampling.

3.4.5 Linear response approximation

The linear response approximation (LRA) has been successfully used to determine pK_a values of amino acid residues in proteins [201, 202]. This method was used here to compute the pK_a value of a phosphate group in a DMPC molecule embedded in a lipid bilayer, and can be calculated using the expression

$$\mathrm{p}K_\mathrm{a} = \tfrac{1}{2}[\langle \mathrm{p}K_\mathrm{a}(c)\rangle_\mathrm{p} + \langle \mathrm{p}K_\mathrm{a}(c)\rangle_\mathrm{d}] \,, \tag{3.6}$$

where p$K_\mathrm{a}(c)$ is the pK_a when the system is fixed in conformation c, and the angled brackets denote averages over the conformations sampled from the MD simulations with the site protonated (p) or deprotonated (d). The pK_a value for each conformation, p$K_\mathrm{a}(c)$, was obtained from MC runs at different pH values using the PB-derived free energy terms (see sections 3.4.1 and 3.4.4).

The reorganization energy [202, 220, 221] is the free energy needed to change the protonated system to configurations typical of the deprotonated state (and vice versa), and can be calculated using the expression

$$\lambda = \tfrac{1}{2}[\langle \mathrm{p}K_\mathrm{a}(c)\rangle_\mathrm{p} - \langle \mathrm{p}K_\mathrm{a}(c)\rangle_\mathrm{d}] \,. \tag{3.7}$$

3.4.6 Analysis

Several tools from the GROMACS software package [210, 211] were used and others were developed in-house. The individual areas per lipid headgroup were calculated using the GridMAT-MD tool [222]. The calculations of correlation-corrected errors for averages were computed using standard methods based on the autocorrelation function of the property measured to determine the number of independent blocks in the simulations [134].

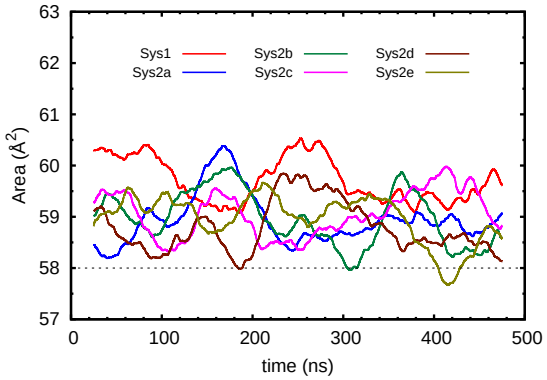


Figure 3.4: A_l valuer of all simulations over time. System Sys1 is pure zwitterionic DMPC, while in systems Sys2a-e there are two protonated DMPC molecules. The dashed line defines the lower limit of the experimental range ($\sim$ 58 Å^2) with the upper limit ($\sim$ 65 Å) being outside of range [26–32, 34–40]. A floating window average (50 ns) was applied for clarity.

3.5 Results and discussion

3.5.1 Effect of protonation on membrane structural features

We evaluated several structural properties of our bilayers and compared them with the available experimental data. The area per lipid (A_l) is probably the most important property to access the system equilibration and to determine the correct liquid state of a lipid bilayer. Figure 3.4 shows the A_l of all simulations over time. This property is difficult to measure experimentally and significantly different values can be obtained from different approaches and methodologies [26–32, 34–40]. Our Sys1 result shows an A_l value in the lower limit, but still within the experimental range (Figure 3.4). This slightly lower A_l value is the result of a small difference between our charge set and the one from Chiu and coworkers [207]. The A_l value of the deprotonated system is slightly higher than the values observed for the protonated ones. The presence of a proton decreases the electrostatic repulsion in the neighbor phosphate groups, stabilizing them through hydrogen bonds. To characterize the environment around the protonated lipid, we envisaged a radial descriptor where we divide the membrane plane in concentric circles (of radii r) centered on a specific phospholipid in a way that each annulus has a width w, and r goes from $\sim$ 3.5 to 20 Å in steps of $\sim$ 0.5 Å (Figure 3.5a). Figure 3.5b illustrates the decrease in the A_l value of the zwitterionic DMPC lipids as they approach the protonated one. The A_l value for this protonated lipid (45.4 Å^2) is remarkably

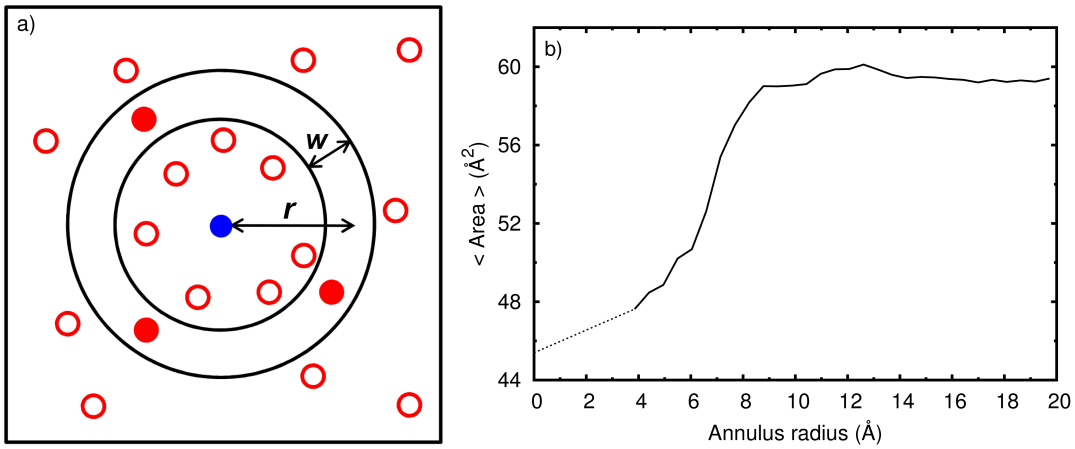


Figure 3.5: Radial A_l description around the lipid of interest: a) graphical representation of the sliding annulus with $r - w/2$ as the smaller radius and $r + w/2$ as the larger radius; b) average A_l value along a sliding annulus of 0.5 Å width. At the zero radius, we plotted the average A_l for the protonated DMPC. The individual areas of each lipid were calculated with the GridMAT-MD tool [222].

low, typical from a gel phase. Thus, protonation of a DMPC molecule induces a local contraction of the membrane.

The diffusion coefficient can be obtained from the Einstein relation by measuring the slope of the mean square deviation (MSD) over time from the linear part of the curve [134]. We obtained the values 6.2×10^{-8} and 7.6×10^{-8} cm^2/s for the deprotonated and protonated, respectively (Figure 3.6). These values are within the experimental range ($(5 - 10) \times 10^{-8}$ cm^2/s) [223]. The small increase in the diffusion coefficient for the protonated DMPC was not expected and indicates some instability factor that is able to overcome the hydrogen bond network formed.

Figure 3.7 shows the order parameter profiles for the protonated and deprotonated forms. Both systems show order parameters a little above the experimental values (Figure 3.7), still all have $|S_{CD}|$ values lower than 0.25, indicating that the aliphatic chains are disordered [38]. The fact that protonation induces a minor disorder to the system while still lowering its A_l prompted us to look at the protonated phosphate groups in detail.

3.5.2 Phosphate distribution along the membrane normal

The phospholipids in membranes are not static, and, although we did not observe flip–flop events, they move considerably. These movements are observed not only

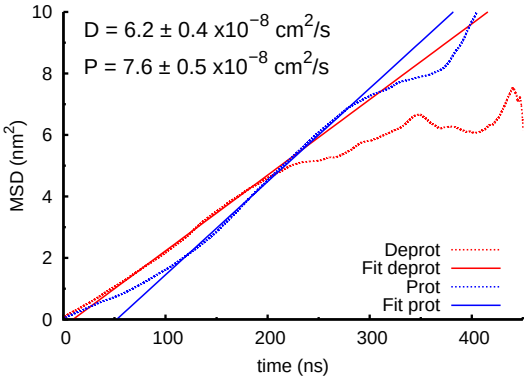


Figure 3.6: MSD, over time, for protonated (blue) and deprotonated (red) DMPC in a membrane. The deprotonated was fitted between 50 and 200 ns, while the protonated one was fitted between 100 and 300 ns.

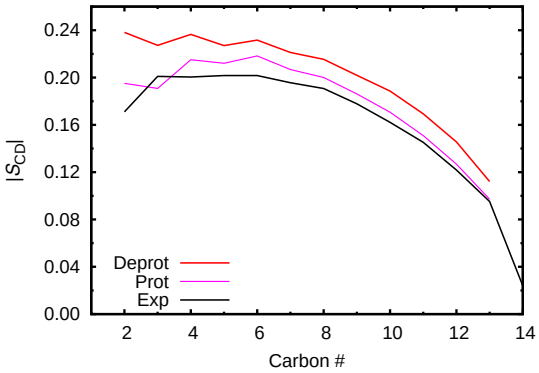


Figure 3.7: Deuterium order parameters averaged over sn-1 and sn-2 chains of DMPC compared with experimental values interpolated to 310 K [31].

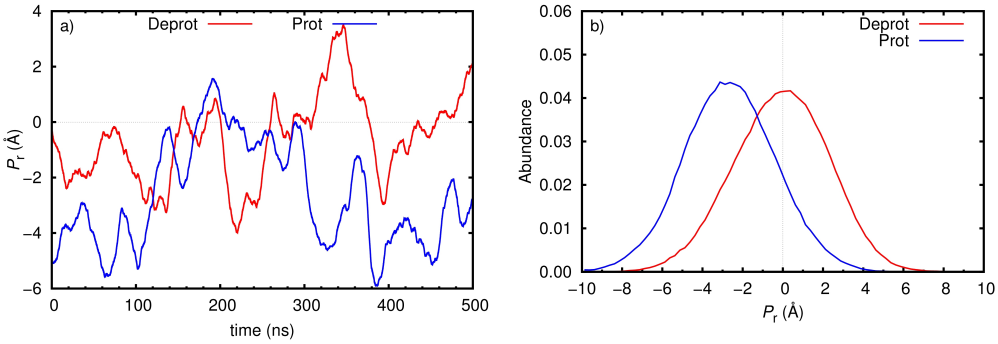


Figure 3.8: P_r distribution relative to the membrane normal: a) floating window average (20 ns) of P_r over time for one specific lipid (lipid 65) in both protonation states; b) the P_r distribution histograms for all lipids studied.

in the membrane plane, but also in the membrane normal. Looking at a specific phospholipid, we can follow the z-position of its P atom, relative to the other 63 in the same monolayer (P_r) (Figure 3.8a). These P_r values oscillate over a large range ($\sim$ 15 Å) for both protonation states (Figure 3.8b). The histograms of P_r distribution illustrate a clear difference for the protonated species, namely, its insertion into the membrane. While the deprotonated phospholipids are distributed around the overall zero position, upon protonation, the distribution moves toward the inside of the membrane $\sim$ 3 Å. The neutralization of the phosphate group in DMPC may be favoring both the desolvation (moving away from the solvent) and the interactions with the ester groups deeper in the membrane. In addition, the choline positive group can now push, with less resistance, the lipid toward the membrane in order to interact directly with the neighboring negatively charged phosphates.

3.5.3 Phosphate protonation induces a local depression

The previous analysis showed that protonation induced a significant insertion of the lipid in the bilayer. In order to characterize the possible membrane deformation, we calculated P_r as we did in the previous section, but using as reference only the phospholipids that fall inside a sliding annulus (red filled circles in Figure 3.5a). The resulting profile (Figure 3.9) shows how the average z-positions of the other phosphate groups in the DMPC bilayer (P_r^*) changes with distance. Similar to what was observed previously (Figure 3.8b), the P_r^* values of the deprotonated system expectedly fluctuate around the position of the lipid of interest. The protonated system added some structural detail regarding the organization ot its neighbors. As we increase the annulus radius, P_r^* increases up to a distance

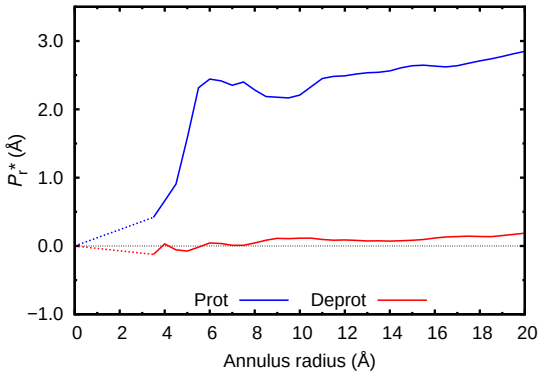


Figure 3.9: Radial depth description around the lipid of interest. P_r^* values are relative to the sliding annulus. The P_r^* values were adjusted so that the zero was assigned to the lipid of interest.

of $\sim 6-7$ Å where it stabilizes (note that the P_r^* values were adjusted so that zero was assigned to the z position of the lipid of interest). At this distance, P_r^* is ~ 2.5 Å and increased to 3.0 Å at higher radius. A similar result has been observed with lauric acid, where its solvent accessible surface decreased upon protonation [113]. The bell-shaped depression in the DMPC bilayer that occurs upon protonation of one lipid (Figure 3.9) is also characterized by internalized deprotonated lipids situated in the first interacting sphere. These lipids seem to be dragged into a deeper region of the bilayer by the protonated DMPC. This is probably only possible because the protonated lipid becomes a hydrogen bond donor in a large excess of hydrogen bond acceptors, and it is able to network with multiple neighbors. The impact of protonation in the interaction of our lipid of interest and its neighbors was already hinted by the average specific area of the protonated DMPC (Figure 3.5b), and it can also be illustrated using a radial distribution function (rdf).

The rdf analysis was done over all phospholipids of interest (Table 3.1) using only their P position. Figure 3.10 shows that the protonated DMPC molecule interacts more strongly with its neighbors than the deprotonated form. The hydrogen bonding and the absence of the negative charge can contribute to a significant decrease in the distance peak observed for the first shell of interacting lipids. This result is in accordance with the low A_l value of protonated DMPC (45.4 Å^2), and it helps to explain the smaller A_l value observed in Sys2, relative to Sys1, although, because of the existence of only one protonated phospholipid per monolayer, this difference is not significant in the final A_l value of Figure 3.4.

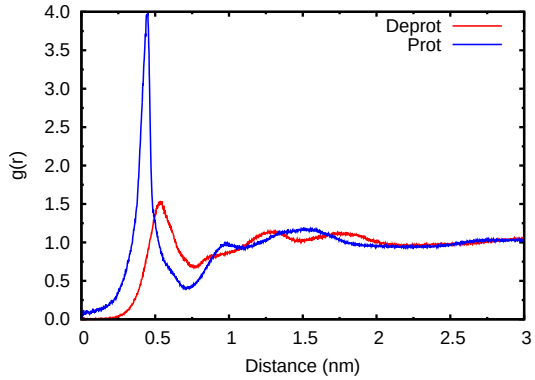


Figure 3.10: Average 2D-radial distribution of deprotonated phospholipids around the lipid of interest. Only the P atoms were used in the calculations.

Table 3.2: pK_a values from LRA at different interior dielectric constants (ϵ_{in}). The reorganization energies (in pK units) are also presented.

ϵ_{in}	pK_a (LRA)	λ (pKunits)
2	5.5 ($\pm$ 1.2)	8.0
4	3.3 ($\pm$ 0.6)	4.3
6	2.6 ($\pm$ 0.4)	3.1
8	2.2 ($\pm$ 0.3)	2.4

3.5.4 pK_a calculation using a linear response approximation

The pK_a of the phosphate group of DMPC embedded in a bilayer is not known, although the first pK_a of phosphate in water is 2.12 [224]. In DMPC, the phosphate group is doubly functionalized being similar to the dimethyl phosphate whose pK_a in water is 1.29 [225], which was used as the pK^{mod} value. Here, we used the LRA method to compute the pK_a of one DMPC molecule in a pure bilayer of deprotonated DMPC. Even though this is not the midpoint pK_a value of a DMPC membrane, it corresponds to the pH at which DMPC starts to become protonated and exhibit heterogeneous behavior. The values computed at different dielectric constants are shown in Table 3.2. The interior dielectric constant (ϵ_{in}) in proteins has been the subject of many discussions [201, 202, 204, 216], but values between 2 and 8 have been adopted within the LRA methodology [202, 216–218]. Although a value of $\epsilon_{in} = 2$ should work if both states have been well sampled *and* if their distributions extensively overlap, this is certainly not the case for the

system under study: indeed, the large local change induced by protonation in the membrane structure (Figure 3.9) points to a poor overlap, corroborated by the significant reorganization energy obtained even at high ϵ_{in} values (Table 3.2). In such cases, the LRA method was found to require a larger ϵ_{in} value (6 or 8) [202, 216]. Therefore, from Table 3.2, we can estimate a pK_{a} value of $\sim$ 2.2–2.6. It is important to note that this is only one estimate of the pH at which the first DMPC gets protonated in an infinite bilayer of deprotonated lipid. Nevertheless, these values support most theoretical MD simulations with pure DMPC, where authors usually ignore phosphate protonation.

We observed a considerable degree of structural deformation upon DMPC protonation, with the protonated lipid sampling different regions along the membrane normal, comparing with its deprotonated counterpart. Figure 3.11 shows how that heterogeneity affects the calculated pK_{a} values. Having the P_{r}^{*} and the pK_{half} values for the phospholipids of interest, we calculated the pK_{a} values using LRA in small portions along the membrane normal. The plot shows the pK_{a} profile along

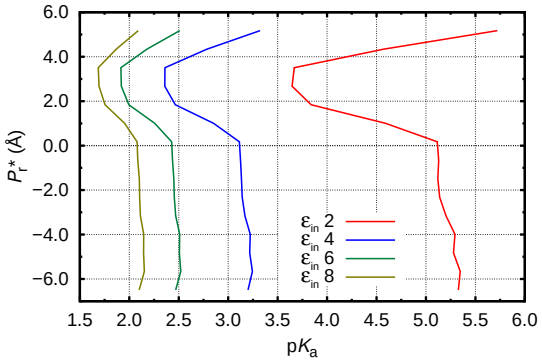


Figure 3.11: pK_{a} values at different P_{r}^{*} positions of the membrane normal. The bin size of the P_{r}^{*} has a fixed dimension of 0.833 Å.

the membrane normal. The lowest pK_{a} was observed at 2-3 Å, which corresponds to an average position of the positive choline groups that can stabilize the negative (deprotonated) phosphate group, lowering its pK_{a} value. Going deeper into the membrane, the pK_{a} values increase as a result of unfavorable electrostatic interactions between the negatively charged phosphate groups. In addition, the protonated form can be stabilized through hydrogen bonds with neighbor phosphate groups and especially with the esters. The high pK_{a} values observed above 3 Å were obtained with a very limited number of protonated conformations, which results in poor sampling. These few conformations were special because they consisted in a protonated phosphate pairing through hydrogen bonding with a

deprotonated neighbor, which resulted in significantly higher pK_a values.

3.6 Conclusions

We present here a new approach to calculate pK_a values in lipid bilayers using Poisson–Boltzmann theory together with periodic boundary conditions. This allows the calculation of the electrostatic properties in different types of periodic systems, such as membranes.

The pK_a calculations were done using LRA with several dielectric constant values for the bilayer. Based on the substantial local rearrangement observed in the MD simulations, a moderately higher ϵ_{in} was selected (6–8). As a result, we estimate that relevant DMPC protonation will only happen at very low pH values (2–3). Therefore, our results support the decision of using zwitterionic DMPC in most theoretical MD simulation studies.

Some important structural properties such as A_l, diffusion coefficient and order parameters were calculated and showed to be within the experimental ranges for liquid-crystalline Lα phase DMPC. The protonation of 1 DMPC molecule (in 64) did not affect the membrane properties significantly, probably because its effect became diluted. However, the protonation led to a decrease in the specific A_l value and to the internalization of the phosphate group, and formation of a local depression. This was favored by the hydrogen bond network that the protonated phosphate is able to do with other phosphates and esters. In addition, the neutral phosphate does not need to be solvated as much as the charged one. One of the consequences of the hydrogen bonding was the dragging of its closest neighbors, which resulted in a bell-shaped local depression.

The approach here presented can easily be generalized to other lipid bilayers. The inclusion of this method within the constant-pH MD framework [3, 8, 10–13, 124–131] will certainly open new research opportunities to characterize, at the molecular level, systems where pH and phase transitions are strongly coupled. Furthermore, if, with pure PC systems, pH seems to play only a minor role, for lipid bilayers containing anionic lipids, this most probably will not be the case.

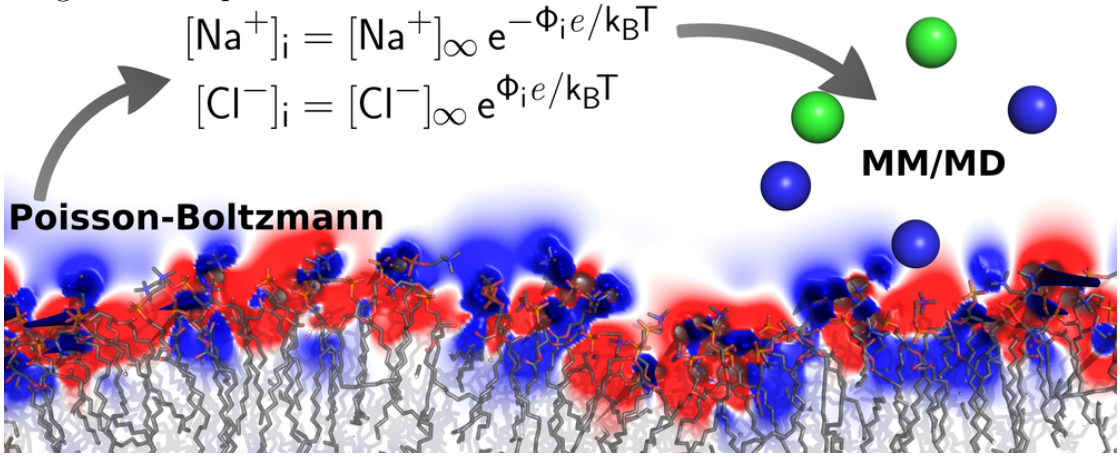
3.7 Acknowledgments

We thank Maria J. Calhorda, João Henriques, Paulo J. Costa, Sara R. R. Campos, and Hugo A. F. Santos for fruitful discussions. We acknowledge financial support from Fundação para a Ciência e a Tecnologia through projects PTDC/QUI-BIQ/113721/2009 and PEst-OE/QUI/UI0612/2013 and grant SFRH/BD/81017/2011.

Chapter 4

Treatment of ionic strength in biomolecular simulations of charged lipid bilayers

Diogo Vila-Viçosa[1], Vitor H. Teixeira, Hugo A. F. Santos, António M. Baptista and Miguel Machuqueiro



Contents

[1]In this work, I performed most of the simulations and analysis. I was also involved in all steps of the manuscript preparation.

4.1 Context

At the end of the first year of this PhD, having already solved the PBC problem, we started the constant-pH MD simulations. However, at this point, we figured out that the treatment of ionic strength could be a problem. We needed to decide which long range electrostatics method should we use and how many ions should we add to our simulations boxes. Having this problem in mind, we developed a Poisson–Boltzmann based method to determine the number of ions that should be added to the simulations. We also compared our approach with other commonly used ones, such as a full system neutralization or no addition of ions. This work was published in Journal of Chemical Theory and Computation (2014,10:5483. DOI: 10.1021/ct500680q) and it is available at `http://pubs.acs.org/doi/abs/10.1021/ct500680q`.

4.2 Abstract

Biological membranes are complex systems that have recently attracted a significant scientific interest. Due to the presence of many different anionic lipids, these membranes are usually negatively charged and sensitive to pH. The protonation states of lipids and the ion distribution close to the bilayer are two of the main challenges in biomolecular simulations of these systems. These two problems have been circumvented by using ionized (deprotonated) anionic lipids and enough counterions to preserve the electroneutrality. In this work, we propose a method based on the Poisson–Boltzmann equation to estimate the counterion and coion concentration close to a lipid bilayer that avoids the need for neutrality at this microscopic scale. The estimated number of ions was tested in molecular dynamics simulations of a 25% DMPA/DMPC lipid bilayer at different ionization levels. Our results show that the system neutralization represents an overestimation of the number of counterions. Consequently, the resulting lipid bilayer becomes too

ordered and practically insensitive to ionization. On the other hand, our proposed approach is able to correctly model the ionization dependent isothermal phase transition of the bilayer observed experimentally. Furthermore, our approach is not too computationally expensive and can easily be used to model diverse charged biomolecular systems in molecular dynamics simulations.

4.3 Introduction

Biological membranes are complex systems that have been broadly studied in recent years [1, 15, 32, 33, 54, 138]. They possess a large variety of lipids, mainly neutral (zwitterionic) and anionic [18]. Most biological membranes are negatively charged and sensitive to pH. Consequently, their protonation state is altered upon small pH changes and can be related with many biological processes such as signal transduction [15, 71–73, 226–228].

Membrane biophysical models, often used to understand the behavior of biological lipid bilayers, are usually simple and mainly composed of pure lipids and, in some cases, of binary or ternary mixtures [19, 21, 93]. Moreover, phosphatidylcholines (PC) are widely used to model membranes since, in general, they are the main component of eukaryotic biological membranes [229] and form very stable lipid bilayers [32, 33]. In particular, 1,2-dimyristoyl-*sn*-glycero-3-phosphocholine (DMPC) is commonly used in theoretical models of lipid bilayers since it is well parametrized in most biomolecular force fields [138], and there is considerable structural information for its L_α lamellar phase [25–40].

In the past decade, molecular dynamics (MD) simulations have been widely used to study lipid bilayers at the molecular level. To increase the realism of the model systems used in such simulations, it is crucial to add anionic lipids such as 1,2-dimyristoyl-*sn*-glycero-3-phosphate (DMPA). The phosphatidic acid (PA) family of phospholipids is particularly sensitive to pH at physiological levels and has an important role in the regulation of several biological processes, [75, 227, 230–232] such as in the Opi1 mediated control of phospholipid synthesis and lipid metabolism [227].

The simulation of charged binary systems (such as DMPA/DMPC mixtures) poses two main challenges: first, the ionization of the bilayer strongly depends on pH and on the DMPA/DMPC ratio [15]; second, the ion distribution close to the membrane has a high impact on the membrane stability and depends on the membrane ionization and on the ionic strength of the solution. These two problems have been addressed by considering lipids to be ionized (deprotonated) and adding the corresponding amount of counterions needed to preserve the electroneutrality, with or without additional ions to describe the bulk ionic strength effect [66, 83, 94–98]. For this purpose, particle mesh Ewald (PME) [156] is commonly used to

describe the long range electrostatic interactions in MD simulations (see Ref. 157 for a recent review on the electrostatic treatment in MD simulations). However, the neutrality of the system might not be appropriate for a highly charged system at this microscopic level. For example, according to Gouy–Chapman (GC) theory and Monte Carlo (MC) simulations, the concentration of counterions and coions close to a charged surface with a charge density $\sigma = -0.0621$ C m^{-2} (approximately 23% of singly charged lipids in a membrane with an average area per lipid of 0.6 nm) only converges to the bulk value of 0.1 M at $\sim$ 3 nm (see Figure 14.8 from Ref. 41). This means that, in these conditions, the electroneutrality is only achieved after this distance from the surface. Hence, to simulate a neutral system one would need to use a very large simulation box (in z direction) which would be too time consuming. To avoid this, it is necessary to model the local effects of the ionic strength (that are strongly related with the electrostatic potential generated by the membrane) without introducing artifacts.

The Poisson–Boltzmann (PB) equation [49, 76, 81–83] (from which GC theory results [74, 75] as a particular case) and MC simulations [78–80] have been widely used to describe the concentration profile of the ions (counterions and coions) close to a charged surface. Both GC and MC simulations of ions often ignore the molecular detail of the charged surface [41]. However, the numerically solved PB equation can be used to describe the potential (and consequently the ion distribution) close to a membrane with atomic detail [41, 49, 83]. This equation provides a mathematical relation between the charge distribution in a solute and the electrostatic potential generated by it in a continuous medium with a specific dielectric constant and salt concentration. This electrostatic potential can then be used to estimate the amount of ions close to the solute, in our case, a lipid bilayer (see section 4.4 for details).

In this work, we propose a method based on the PB equation to estimate the counterion and coion concentration close to a lipid bilayer. The estimated number of ions is then explicitly included in a MD simulation of the system. We performed our simulations using a generalized reaction field (GRF) [155] to treat long range electrostatic interactions and the bulk effect of ionic strength (instead of GRF, one might use PME with a neutralizing uniform background charge density, but such an approach was recently found to introduce artifacts in inhomogeneous systems [233]). We compared the results obtained using our method with simulations without any explicit ions and with neutralized systems simulated with both GRF and PME.

4.4 Theory and methods

4.4.1 Estimation of the number of explicit ions

As stated in the Introduction, we use a PB model to estimate the amount of counterions and coions near a charged membrane. Given the variable size of the MD simulation box, together with the fact that the across-solvent distance between the two monolayers in the periodic box was always only slightly higher than twice the cutoff value used in the GRF treatment of long-range electrostatics (1.4 nm; see section 4.4.2), we decided to compute the amount of ions only up to such distance. This substantially simplifies the computational procedure and corresponds to differences of only tenths of an ion.

According to the Poisson–Boltzmann (PB) theory, the number density of an ion with charge z at position $\mathbf{r}$ is given by [41, 49, 234]:

$$\rho(\mathbf{r}) = \rho_0 \mathrm{e}^{-ze\phi(\mathbf{r})/k_\mathrm{B}T} \tag{4.1}$$

where $\phi(\mathbf{r})$ is the electrostatic potential at $\mathbf{r}$, ρ_0 is the bulk number density, e is the elementary charge, k_B is the Boltzmann constant and T is the absolute temperature. If we discretize the region C encompassed by the cutoff into a set of M small cells, each with volume V and approximately constant electrostatic potential, we obtain a PB-estimate of the total amount of ions of this type in region C:

$$n_\mathrm{PB} = \int_C \rho(\mathbf{r})\mathrm{d}v \approx V N_\mathrm{A} I \sum_{i=1}^{M} \mathrm{e}^{-ze\phi_i/k_\mathrm{B}T} \tag{4.2}$$

where ϕ_i is the potential in cell i, and we used the fact that $\rho_0 = N_\mathrm{A} I$ for a 1:1 salt, where N_A is Avogadro's constant and I is the solution ionic strength. This equation is then used to calculate the number of cations, $n_\mathrm{PB}(\mathrm{A}^+)$, and of anions, $n_\mathrm{PB}(\mathrm{Cl}^-)$, for each of a set of conformations from a MD simulation. The median, which is robust to outliers, of these calculated values is then used as an estimation for the next MD simulation of the same system.

The workflow to calculate the final number of explicit ions ($n_\mathrm{exp}(\mathrm{A}^+)$ and $n_\mathrm{exp}(\mathrm{Cl}^-)$) is represented in Figure 4.1. The first MM/MD simulation is performed by setting n_exp to any number within 0 and enough ions for neutralization of the system (this initial guess does not affect the final converged result). After this first simulation, the n_PB values are calculated as explained above and n_exp are adjusted accordingly. We then determine the number of explicit ions that are closer than 1.4 nm to the membrane in the new set of conformations ($n_\mathrm{eval}(\mathrm{A}^+)$ and $n_\mathrm{eval}(\mathrm{Cl}^-)$). After the calculation of both n_PB and n_eval the two pairs of values are compared and, if necessary, n_exp is adjusted for a new MD simulation. This process continues iteratively until $n_\mathrm{eval} \approx n_\mathrm{PB}$, i.e. the process becomes self-consistent (usually after 2 or 3 iterations).

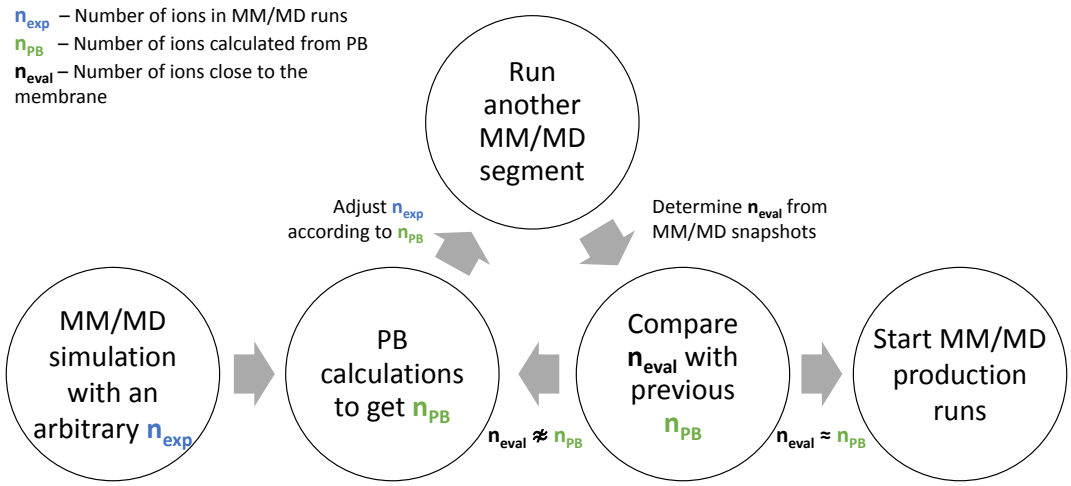


Figure 4.1: Workflow to calculate the final number of explicit ions obtained from the self-consistent procedure.

4.4.2 Methods

Simulations setup

All simulated systems were composed of 128 lipids, 32 DMPA, and 96 DMPC molecules equally distributed by the two monolayers (representing 25% of DMPA and 75% of DMPC). The phosphate group of DMPC was kept deprotonated (zwitterionic lipid) in all simulations. In the case of DMPA, whose headgroup has three possible protonation states (with charge 0, -1 or -2 corresponding to double protonated, single protonated, and deprotonated, respectively), the ionization state was varied from 0 to -64 (Table 4.1). We performed simulations with both GRF and PME treatment for long range electrostatic interactions (Table 4.2). Both Na^+ and K^+ were used as counterions and Cl^- was used as coion in all simulations. In PB^{NaCl} and PB^{KCl}, the number of ions were calculated according to the methodology described in section 4.4.1, while the remaining simulations used either sufficient counterions to neutralize the system (GRF^{Na}, GRF^{K}, PME^{Na} and PME^{K}) or none ($GRF^{no\text{-}ions}$). The number of water molecules were adjusted to maintain a similar size of water region beyond the 1.4 nm cutoff for all systems (around 150 molecules).

Table 4.1: Ionization states distribution in the studied membranes containing 25% DMPA in DMPC.

# DMPA	**Charge**	**Composition**		
		DMPA^{0}	**DMPA^{-1}**	**DMPA^{-2}**
32	0	32	0	0
	-16	16	16	0
	-32	0	32	0
	-48	0	16	16
	-64	0	0	32

Table 4.2: Summary of the simulations performed in this work. The bilayer ionizations studied were 0, −16, −32, −48 and −64. GRF was always used with a fixed 0.1 M ionic strength. All simulations were done in triplicate for 100 ns each.

Name	Electrostatics	counterion	coion	System charge
PB^{NaCl}	GRF	Na^+	Cl^-	Variable[a]
GRF^{Na}	GRF	Na^+	–	0
PME^{Na}	PME	Na^+	–	0
PB^{KCl}	GRF	K^+	Cl^-	Variable[a]
GRF^{K}	GRF	K^+	–	0
PME^{K}	PME	K^+	–	0
$GRF^{no\text{-}ions}$	GRF	–	–	Membrane charge

a) Depends on the number of ions calculated with our self-consistent methodology (see Results and Discussion).

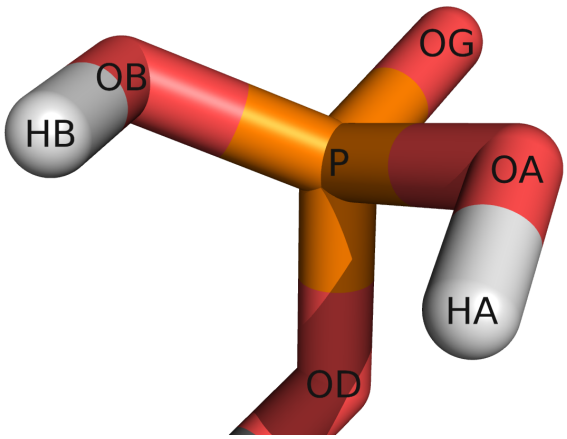


Figure 4.2: Double protonated ($DMPA^0$) form of the phosphate group of DMPA.

Parametrization of atomic charges of DMPC and DMPA

The atomic partial charges for DMPC were taken from the work of Teixeira et al. [1]. For the DMPA molecules the atomic partial charges for neutral, single ionized, and double ionized forms were calculated using a similar procedure: the geometries were optimized with Gaussian 03 [208] at the HF/6-31G(d) level, and the resulting electrostatic potential was fitted to atomic coordinates using RESP [209] with P partial charge fixed to the Mulliken charge. The obtained charges were slightly adjusted to be consistent with the charge groups of the GROMOS96 54A7 force field (Figure 4.2 and Table 4.3).

MM/MD

Molecular mechanics/dynamics (MM/MD) simulations were done with the GROMOS 54A7 force field [212] and a modified version [12, 13] of GROMACS distribution (version 4.0.7 [210, 211]). The Lennard–Jones parameters for K^+ were taken from Ref. 235. The membrane was solvated in a tetragonal box with SPC water molecules [214], with the membrane plane oriented parallel to the x and y axis. All lipid bond-lengths were constrained using the parallel version of the LINCS algorithm [163], and the SETTLE algorithm [164] was used for water. The simulations were done in the NPT ensemble, using the v-rescale thermostat [161] at 310 K with separate couplings for the solute and solvent (including ions, when present) with a relaxation time of 0.1 ps. A semi-isotropic Berendsen [160] pressure couple was used to maintain the pressure constant at 1 bar with a compressibility

Table 4.3: Partial atomic charges of the three protonation states of DMPA.

Atom Name	**DMPA**$^{-2}$	**DMPA**$^{-1}$	**DMPA**^{0}
OA	−0.97	−0.80	−0.63
HA	–	0.43	0.48
P	1.47	1.54	1.17
OB	−0.97	−0.91	−0.63
HB	–	–	0.48
OG	−0.97	−0.91	−0.72
OD	−0.56	−0.35	−0.15

of 4.5×10^{-5} bar^{-1} and a relaxation time of 5.0 ps. An integration step of 2 fs was used in the equations of motion in all MD simulations.

Nonbonded interactions were treated with a twin-range method, using group-based cutoffs of 0.8 and 1.4 nm, updated every 5 steps. Long-range electrostatics were corrected with a GRF [13, 155], using a dielectric constant of 54.0 [215] and an ionic strength of 0.1 M [13]. The particle mesh Ewald (PME) method [156] was also used in a set of simulations. For this method, we used a grid spacing of 0.12 nm, the short–range interactions were computed using a nonbonded pairlist with a single cutoff of 1.0 nm, updated every 10 steps.

The minimization procedure consisted in three sequential steps: starting with up to 10 000 steps using the steepest descent algorithm (unconstrained), followed by 10 000 steps using the l-BFGS integrator (unconstrained), and ended with $\sim$ 300 steps using the steepest descent algorithm (with all bonds constrained).

The same initialization procedure was applied for all systems and consisted of four steps. First, a 50 ps MD simulation with all heavy atoms harmonically restrained with a force constant of 1 000 kJ mol^{-1} nm^{-2}. In the three following steps of 100, 150, and 200 ps MD simulations, position restraints were applied differently for the phosphorus and remaining heavy atoms with force constants of 1 000, 100, and 10 kJ mol^{-1} nm^{-2} and of 100, 10, and 0 kJ mol^{-1} nm^{-2}, respectively. The MD simulations of all bilayer systems were pre-equilibrated for 100 ns, after which three replicates were done for an additional 100 ns.

Ion estimation using a Poisson–Boltzmann model

The Poisson–Boltzmann (PB) calculations were done with the program DelPhi V5.1 [205, 206] using atomic radii taken from GROMOS 54A7 force field [212] and

atomic partial charges as described in the previous sections. The molecular surface of the membrane was defined by a spherical probe of radius 0.14 nm, the ion exclusion layer was 0.2 nm, and the ionic strength was 0.1 M. The dielectric constant of the solvent was 80 and for the membrane we used a value of 2 (similar results were obtained with dielectric constant values between 2 and 8). Considering the membrane oriented along the xy plane, we explicitly applied PBC along the x and y directions. The convergence threshold value based on maximum change of potential was set to 0.01. In the iteration convergence process, we used relaxation parameters for both nonlinear and linear forms of the PB equation with values of 0.15 and 0.20. The PB calculations were performed in a cubic (finite differences) grid with 251 points per side (corresponding to a grid space of ~ 0.05 nm). For these calculations were used 200 conformations from each $\mathrm{PB}^{\mathrm{NaCl}}$ and $\mathrm{PB}^{\mathrm{KCl}}$ simulation.

To estimate the number of ions (n_{PB}), the electrostatic potential (ϕ) of the system is written to a new 3D grid with 50 points per side ($\mathrm{grid}^{\mathrm{ion}}$) and the same size as the finite difference grid. Different number of points per side were tested and consistent results were obtained using between 50 and 80 points. Using equation 4.2, we sum the number of ions n for all $\mathrm{grid}^{\mathrm{ion}}$ points whose shortest distance to any atom of the bilayer is larger than $r_i + 0.2$ nm and shorter than R_c, where r_i is the radii of the atom in the bilayer, 0.2 nm is the ion exclusion layer and $R_c = 1.4$ nm, the GRF cutoff.

Calculation of lipid protonation

The lipid protonation was computed using a PB/MC methodology previously described [1]. PB calculations (with DelPhi V5.1 [205, 206]) were done in a cubic grid with 61 points per side and a two step focusing [219] where the focus grid was one fourth of the coarser grid size. Since the dimension of the box changes along the simulation (the conformations were taken from the MD simulations, performed in NPT), the space between grid points for the coarser grid needs to be calculated for every conformation (average values around 0.1 and 0.025 nm for the coarse and focus grids, respectively). In the coarse grid, we used relaxation parameters of 0.2 and 0.75 for linear and nonlinear iteration convergence process. A cutoff of 2.5 nm was used to calculate the background and pairwise interactions (see reference [1] for more details). The $\mathrm{p}K_{\mathrm{a}}$ values of the model compounds used were 1.29 for DMPC (the $\mathrm{p}K_{\mathrm{a}}$ value of dimethylphosphate) and 1.54 and 6.31 for DMPA (the two $\mathrm{p}K_{\mathrm{a}}$ values of methylphosphate) [225]. Conformations for the protonation calculations were taken from the MD simulations from the equilibrated last 90 ns of each system at intervals of 10 ns. This makes 10 frames per replicate and a total of 30 for each system.

These calculations were performed in several configurations with mixtures of

lipids in different protonation states, meaning that they are necessarily biased toward those inherited protonation states, similarly to what happens in linear response approximation calculations [1, 201, 202]. These limitations are usually circumvented with an increased dielectric constant value, in order to roughly account for the reorganization that would result from ionization changes. Therefore, based on our previous linear response approximation studies using a similar methodology [1, 202], dielectric constants of 8 and 80 were used for the membrane and the solvent, respectively. For a more detailed discussion on this subject, see Refs. 1, 169, 177, 201, 202, 204, 216, 218.

The MC sampling was performed with the PETIT program [176] version 1.6.1 using an absolute temperature of 310 K. All runs were performed using 10^5 MC cycles, where one cycle consists of sequential state changes over all individual sites and pairs of sites with an interaction larger than 2 pH units.

Numbers of ions from Gouy–Chapman theory

For comparison with the PB calculations taking into account the detailed membrane structure, we also computed the numbers of coions and counterions predicted from Gouy–Chapman (GC) theory, in which all properties depend only on the distance from an idealized uniformly charged planar membrane. The number of ions accumulated by an ion species with charge z over an area A up to a distance r along the normal to the membrane is given by:

$$n_{\mathrm{GC}}(r) = \int_{A\times r} \rho(\mathbf{r})\mathrm{d}v = N_{\mathrm{A}} I A \int_0^r \mathrm{e}^{-ze\phi(u)/k_{\mathrm{B}}T}\mathrm{d}u \; . \tag{4.3}$$

For a 1:1 salt, GC theory gives (equation 14.39 from Ref. 41 and equation 4.22 from Ref. 236):

$$\phi(r) = \frac{2k_{\mathrm{B}}T}{e}\log\frac{1+\gamma\mathrm{e}^{-\kappa r}}{1-\gamma\mathrm{e}^{-\kappa r}} \tag{4.4}$$

where:

$$\kappa = \sqrt{\frac{2Ie^2}{\epsilon\epsilon_0 k_{\mathrm{B}}T}} \tag{4.5}$$

is the inverse Debye length and $\gamma = \tanh(e\phi_0/4k_{\mathrm{B}}T)$, with ϕ_0 being the potential at the membrane surface. This potential is related to the membrane charge density σ through the Grahame equation for a 1:1 salt (equation 4.26 from Ref. 236),

$$\sigma = \sqrt{8I\epsilon\epsilon_0 k_{\mathrm{B}}T}\sinh(\phi_0/2) \; , \tag{4.6}$$

meaning that $\gamma = (\sqrt{1+s^2}-1)/s$, with $s = \sigma/\sqrt{8I\epsilon\epsilon_0 k_{\mathrm{B}}T}$. Therefore, $n_{\mathrm{GC}}(r)$ can be easily computed by numerical integration for given values of T, ϵ, I, A and σ.

Table 4.4: Number of ions in PB^{NaCl} and PB^{KCl} simulations.

Counterion	Lipid charge	n_{exp}		n_{eval}		n_{PB}		n_{GC}		System charge	$n_{eval}(A^+) - n_{eval}(Cl^-)^a$
		A^{+b}	Cl^-	A^+	Cl^-	A^+	Cl^-	A^+	Cl^-		
Na^+	0	5	7	5.0	6.3	4.6	6.7	–	–	−2	−1.3
	−16	6	7	6.0	6.3	5.9	7.1	15.4	2.6	−17	−0.3
	−32	8	7	8.0	6.0	7.4	7.0	28.8	2.0	−31	2.0
	−48	10	7	10.0	6.0	9.5	6.9	43.3	1.8	−45	4.0
	−64	12	7	12.0	5.4	12.7	6.7	58.3	1.7	−59	6.6
K^+	0	5	7	5.0	6.4	4.7	7.0	–	–	−2	−1.4
	−16	6	7	6.0	6.2	5.8	7.0	15.4	2.7	−17	−0.2
	−32	8	7	8.0	5.9	7.1	6.8	28.8	1.9	−31	2.1
	−48	10	7	10.0	5.6	9.2	6.6	43.4	1.7	−45	4.4
	−64	12	7	12.0	4.9	12.3	6.7	58.3	1.7	−59	7.1

a) Average charge within the solvent region up to 1.4 nm from the membrane.
b) A^+ stands for either Na^+ or K^+.

Analysis

The last 90 ns of each simulation were used for analysis. Area per lipid (A_l) and deuterium order parameters were obtained using several tools from the GROMACS software package [210, 211]. Membrane thickness and ion distributions were calculated using in-house tools. Membrane thickness was defined as twice the average z distance between each phosphorus atom and the center of mass (also in z) of all phosphorus atoms. All presented errors were computed using standard methods based on the autocorrelation function of the property measured to determine the number of independent blocks in the simulations [134].

4.5 Results and discussion

4.5.1 Ion estimation

PB^{NaCl} and PB^{KCl} simulations were performed with a variable number of ions that were calculated according to the method described in section 4.4.1 (see Table 4.4 and Figure 4.3). In these two groups of simulations the number of counterions (Na^+ or K^+) varied between 5 and 12 and the number of coions (Cl^-) was 7 in all simulations, i.e. the number of coions is almost insensitive to the membrane ionization while the number of counterions always increases with the membrane charge. This is expected since there is usually an accumulation of counterions

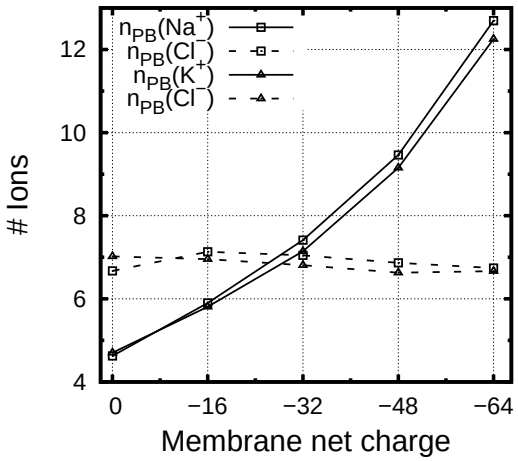


Figure 4.3: Number of counterions ($n_{\mathrm{PB}}(\mathrm{A}^+)$) and coions ($n_{\mathrm{PB}}(\mathrm{Cl}^-)$) in $\mathrm{PB}^{\mathrm{NaCl}}$ and $\mathrm{PB}^{\mathrm{KCl}}$ simulations according to the PB-based methodology.

close to a charged surface (following a Boltzmann distribution - see Theory and Methods). Since the ionization of DMPA varies between 0 and −64, these small number of ions lead to systems with total charges between −2 and −59. Consequently, according to our methodology, a microscopic system with the size of our simulation box ($\sim$ 7–8 nm in z direction) is not globally neutral. In particular, the $\mathrm{PB}^{\mathrm{NaCl}}$ system with 64 negative charges requires 12 Na^+ and 7 Cl^- closer than 1.4 nm, an example particularly far from neutrality. Nevertheless, all membranes were stable during our simulations, without any disruptions. From our results, we can also observe that the number of predicted Na^+ and K^+ is the same in their corresponding simulations. This suggests that these two ions have similar counterion behaviors regardless of their chemical nature and preferred membrane interaction regions. Another nonintuitive result is that the neutral system requires slightly more Cl^- coions than counterions. This observation can be explained by the positively charged layer that is generated in our model by the choline groups of DMPC that can be partially stabilized by the presence of negative ions. However, confirmation of this result would require a robustness analysis (e.g., using different force field parameters), which falls outside the scope of this work. Once again, system neutrality is only achieved farther from this distance from the surface. Table 4.4 also shows the ion estimation using a GC model (n_{GC}). The estimated number of cations is larger than n_{PB}, indicating that the lack of molecular detail is overestimating these values. These results illustrate the need to use PB for these calculations instead of the more simplified and computationally inexpensive GC approach.

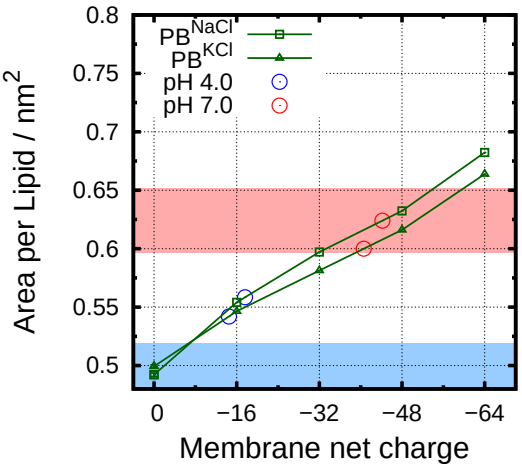


Figure 4.4: Average area per lipid (A_l) of PB^{NaCl} and PB^{KCl} simulations as a function of the membrane net charge. Typical experimental values ranges for gel and liquid crystalline phases of DMPC are marked in blue and red, respectively [25–34]. The ionizations estimated for pH 4 and 7 are also plotted.

4.5.2 Isothermal phase transition

The area per lipid (A_l) values were calculated for PB^{NaCl} and PB^{KCl} simulations (Figure 4.4) by averaging along 270 ns for each system (last 90 ns of each replicate). When the membrane is neutral, the repulsion between the headgroups is minimized and the A_l is very low corresponding to the gel phase of pure DMPC (see blue marked region in Figure 4.4). As we increase the ionization of the bilayer, the A_l also increases, which is not surprising since the electrostatic repulsion between the headgroups is stronger. In the simulations with 32 and 48 negative charges, the A_l is in the range of fluid DMPC [25–34] (see red marked region in Figure 4.4), while above 48 the values exceed this region (0.65 nm^2). With these results, we were able to model an isothermal gel–fluid phase transition in our membrane that is only associated with changes in the ionization state of the DMPA. Additionally, according to the average A_l of our systems, Na^+ and K^+ have a similar impact on our bilayer. Probably, the role of the counterions in our systems is mainly electrostatic and the chemical nature of the ion (Lennard–Jones parameters) plays a secondary role in this property. In fact, the A_l in systems with K^+ is always a bit smaller that the ones with Na^+. The smaller radius of Na^+ allows it to interact with the lipids in the ester region while K^+ stays a little above in the phosphate region minimizing the repulsion between them (see the next section for further structural information).

The pseudobinary phase diagrams for the DMPA/DMPC mixture at several

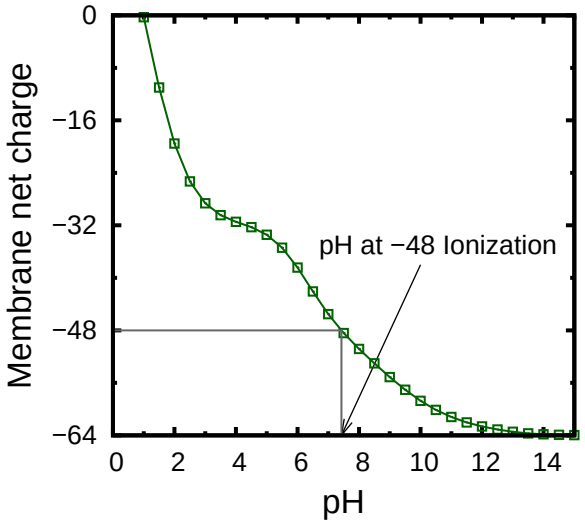


Figure 4.5: Graphical representation of the procedure to determine the pH at which the system has the protonation state used in a PB^{NaCl} simulation at −48 ionization.

pH values has been previously reported [15]. It was shown that, at pH 7 and T = 310 K, the 25% DMPA/DMPC mixture is in a fluid phase. At the same temperature but at pH 4, the mixture is in a gel+fluid phase state and, in some conditions, a macroscopic phase separation is observed. In our simulations, these behaviors occur when the ionization is −48 (fluid phase that is typical of pH 7) and when it is −16 (intermediate phase that may be typical of pH 4). To gain insight into the agreement of our results with the experiment, we performed PB/MC calculations to determine the pH at which the ionization is the same as in the simulation (Figure 4.5). Since the conformations were sampled at the given protonation states, the obtained pH values are extracted from the most reliable regions of the total titration curves. This procedure is repeated for all groups of simulations (Table 4.5). With all the obtained values, we made a plot for each system and then used linear interpolation to determine the value of ionization at pH 4.0 and 7.0 (Figure 4.6). Repeting this procedure to all groups of simulation, we obtained the ionization at pH 4.0 and 7.0 (Table 4.6). These results show that the A_l obtained with our methodology is in excellent agreement with the lipid phases experimentally observed [15].

In order to compare the presented methodology with commonly used approaches, we performed five additional sets of simulations (see Theory and Methods) with both GRF and PME. It was used a set of systems without explicit addition of ions leaving the ionic screening to the long range electrostatic treatment of GRF with

Table 4.5: pH at which the system has the protonation state used in the simulation according to the PB/MC calculations.

Ionization	**GRF$^{\text{no-ions}}$**	**PB$^{\text{NaCl}}$**	**GRF$^{\text{Na}}$**	**PME$^{\text{Na}}$**	**PB$^{\text{KCl}}$**	**GRF$^{\text{K}}$**	**PME$^{\text{K}}$**
0*	2.30	2.55	2.30	2.22	2.52	2.30	2.22
−16	3.97	3.82	4.36	4.15	4.15	4.40	4.61
−32	5.34	5.60	5.95	6.40	5.99	6.89	7.04
−48	7.20	7.43	7.85	8.24	7.86	9.84	10.70
−63#	8.23	8.47	9.51	10.10	10.39	14.77	16.46

* Some DMPC molecules are protonated, which slightly interferes with this ionization.

An ionization of −63 was chosen because protonation never reaches zero, the condition needed to obtain an ionization of −64.

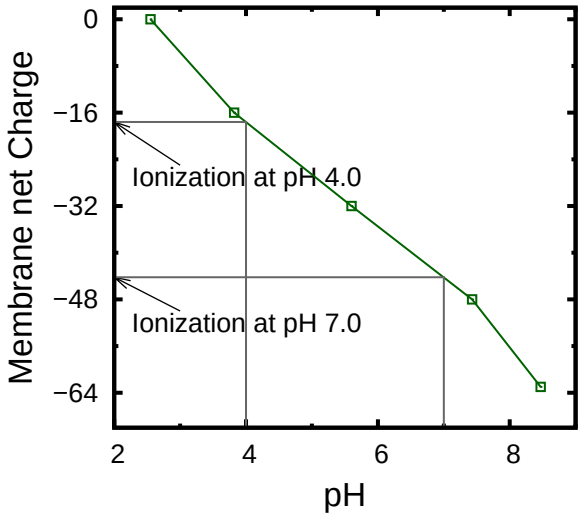


Figure 4.6: Graphical representation of the procedure to determine the value of ionization at pH 4.0 and 7.0 for PB$^{\text{NaCl}}$ simulations.

Table 4.6: Estimated ionization at pH 4.0 and 7.0 from PB$^{\text{NaCl}}$ simulations.

pH	**GRF$^{\text{no-ions}}$**	**PB$^{\text{NaCl}}$**	**GRF$^{\text{Na}}$**	**PME$^{\text{Na}}$**	**PB$^{\text{KCl}}$**	**GRF$^{\text{K}}$**	**PME$^{\text{K}}$**
4	−16.4	−17.6	−13.2	−14.8	−14.5	−13.0	−11.9
7	−46.3	−44.2	−40.8	−37.2	−40.6	−32.6	−31.7

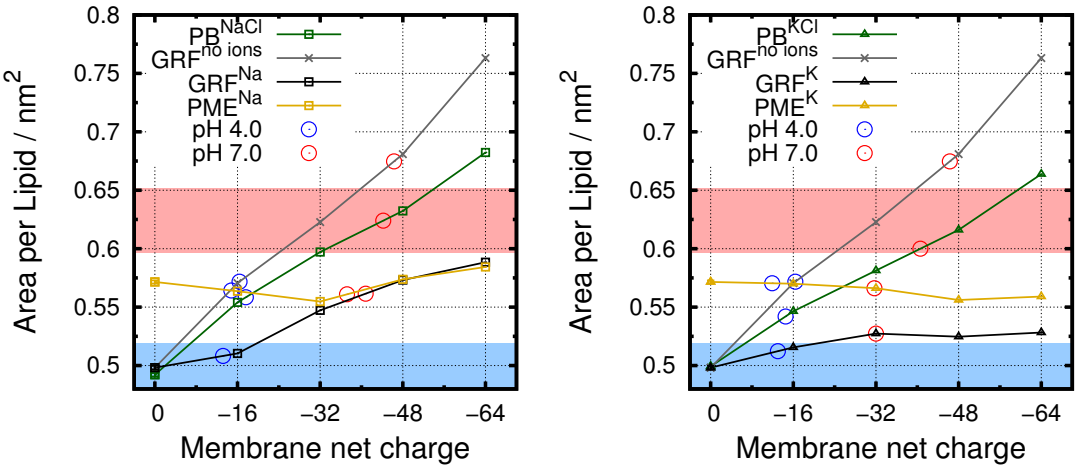


Figure 4.7: Average area per lipid (A_l) for all simulated systems as a function of the membrane net charge using either Na^+ (left panel) or K^+ (right panel) as counterions. Simulations with Na^+ as counterion in the left panel and with K^+ in right. Typical experimental values ranges for gel and liquid crystalline phases of DMPC are marked in blue and red, respectively [25–34]. The ionizations estimated for pH 4 and 7 are also plotted.

I = 0.1 M ($GRF^{no\text{-}ions}$). We also performed simulations with neutral systems and, in these cases, both GRF with I = 0.1 M (GRF^{Na}, GRF^{K}) and PME (PME^{Na}, PME^{K}) were employed. Figure 4.7 shows the A_l per ionization for all these systems using both Na^+ and K^+ counterions. In the $GRF^{no\text{-}ions}$ system the A_l is strongly dependent on the ionization of the bilayer. It varies between a gel-like phase (with $A_l \sim 0.5$ nm^2) and a very fluid phase ($A_l > 0.75$ nm^2) where the membrane is significantly distorted. In these simulations, due to the absence of counterions, the repulsion between the negatively charged headgroups is very high, being only attenuated by an implicit ionic strength screening. At pH 7.0, the obtained A_l (0.67 nm^2) reflects the excessive destabilization present in this set of simulations. This result shows that, when dealing with significantly charged lipid bilayers, the use of counterions is essential. In the neutral systems (GRF^{Na}, GRF^{K}, PME^{Na} and PME^{K}) the A_l curves are not significantly affected by the ionization. This can be explained by the counterions sequestration to the membrane that occur in the systems where they are available. Therefore, the number of stabilizing counterions is probably in excess since, independently of the counterion and the long range electrostatics treatment method used, the membrane is never observed in the fluid phase as expected from experiment [15]. Recently, it has also been observed that a charged membrane moiety rigidifies in an electroneutral system [83].

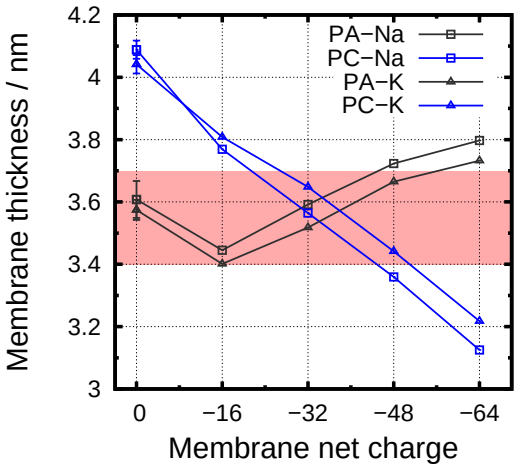


Figure 4.8: Average membrane thickness of PB^{NaCl} and PB^{KCl} simulations as a function of the membrane net charge. Typical experimental values range for liquid crystalline phase of DMPC is shown in red [237].

These results show that the neutrality is inadequate to describe this charged lipid bilayer and its close vicinity as it is usually modeled in MD simulations. On the other hand, with our method the system is usually far from neutrality but is able to correctly model the isothermal phase transition observed in this DMPA/DMPC pseudobinary system [15].

4.5.3 Structural analysis

Membrane thickness is a very helpful structural property, specially when using lipid mixtures since it allows to easily discriminate between the different species. Also, this property can be a good descriptor of the lipid phase transition. Figure 4.8 shows the membrane thickness in the PB^{NaCl} and PB^{KCl} simulations calculated separately for DMPA and DMPC. Interestingly, even though the ionization is centered on the DMPA molecules, we observe a clear phase transition only on DMPC. It seems that protonation of DMPA induces a better packing of the DMPC molecules increasing its thickness from values typical of a fluid phase (3.4–3.7 nm [237]) to values up to $\sim$ 4.1 nm which are probably only observed in more ordered phases. The membrane thickness of DMPA is less sensitive to its ionization and its increase is probably related with the high solvent exposure needed to stabilize its ionized form. As a consequence, DMPC is driven to sample rather lower thickness values ($\sim$ 3.2 nm).

Deuterium order parameter is also a widely used structural property of lipid

bilayers that can easily aid on the discrimination of its constituents. With this property, we follow the phase transition focusing on the packing of the lipid tails. Figure 4.9 shows the order parameters calculated for the PB^{NaCl} and PB^{KCl} simulations. Both DMPA and DMPC molecules show an ionization dependence of the order parameter. However, while DMPC packs more gradually as the membrane ionization decreases, DMPA lipid tails are kept mostly disordered in the ionized systems, only becoming ordered upon full protonation. As observed for the membrane thickness, DMPC is the lipid more sensitive to the phase transition. Once again, at ionization of -48 the S_{CD} values for the plateau region of DMPC (methylene groups 3–6) are within the experimental range for a pure DMPC bilayer (0.17–0.21 [38]).

According to both the membrane thickness and deuterium order parameters, there are no significant differences between using Na^+ or K^+ as counterions in our systems. To better understand the different roles of these counterions, we calculated the counterion distribution around oxygen atoms of phosphates and ester carbonyl groups for both PB^{NaCl} and PB^{KCl} simulations (Figure 4.10). These results clearly show that the counterions are interacting with different regions of the lipids. The smaller Na^+ ions are able to insert into deeper regions of the bilayer interacting preferentially with the ester groups (peak at ~ 0.23 nm). The larger K^+ ions are excluded from these regions and, consequently, remain at a more shallow region interacting preferentially with the phosphate groups (peak at ~ 0.25 nm). With increasing ionization, we observe the gradual release of Na^+ which can be triggered by the increase of membrane fluidity promoting the solvent exposure of the ester regions. On the contrary, the normalized number of K^+ ions interacting with its preferred partners (phosphate groups) does not depend on the ionization, suggesting that this counterion is strongly attached to the membrane. Moreover, there are recent evidence that the current Na^+ ion MM parameters might lead to an overbinding of this ion to PC bilayers [98, 238]. Regardless of these apparent shortcomings in their parametrizations, both ions were able to correctly stabilize the charged lipid bilayers to similar extents, which indicates that the electrostatic counterion effect overcomes their preferred membrane interaction regions.

4.6 Conclusions

Here, we present a method to estimate the number of ions close to a charged lipid bilayer at a given ionic strength. This method is proposed as an alternative to the commonly used approaches based on system neutrality [66, 83, 94–98]. It has been shown that the electrostatic potential generated by a charged surface is propagated over large z direction distances beyond the commonly used MD system sizes [41, 49, 74, 75, 78–82]. Therefore, at this microscopic level the system

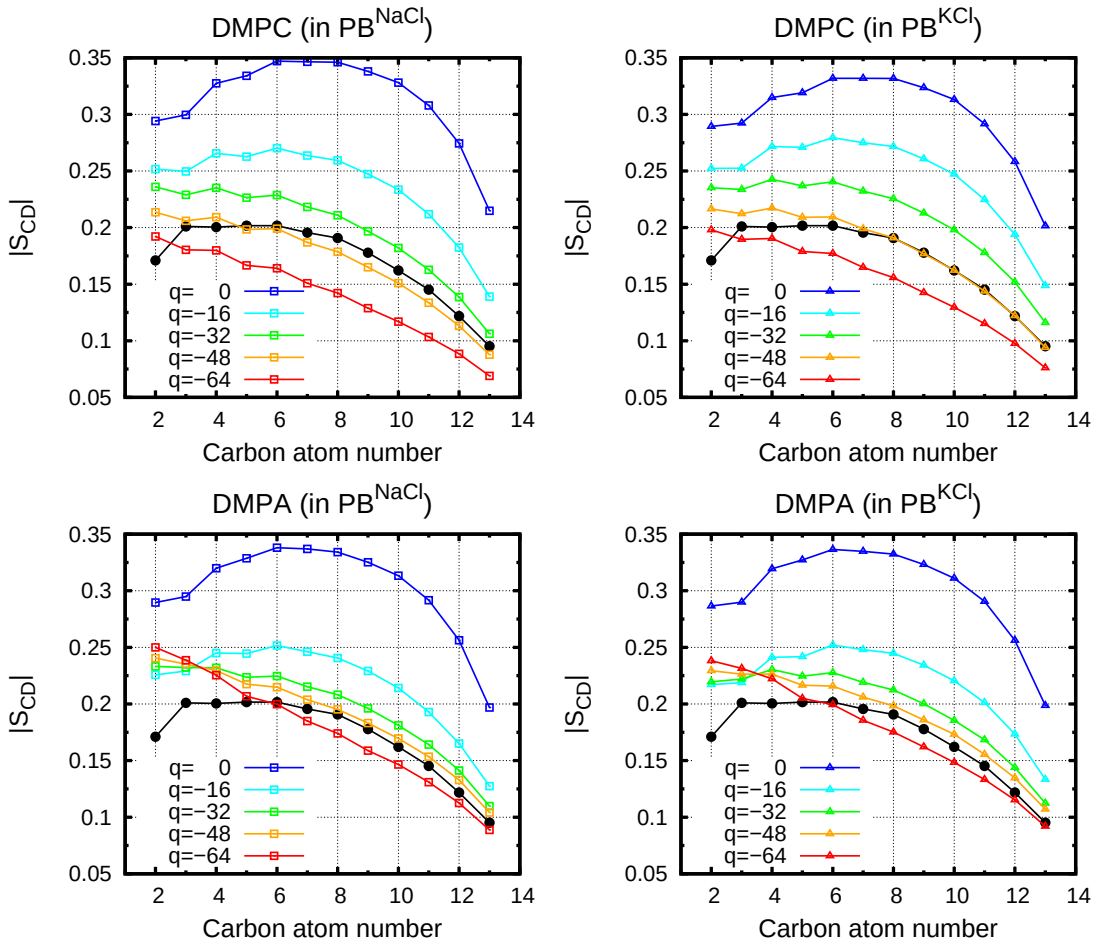


Figure 4.9: Deuterium order parameters for PB^{NaCl} and PB^{KCl} simulations. Experimental curve for DMPC in the liquid crystalline phase is shown in black circles. These values were linearly interpolated to the desired temperature (310 K) from the three available temperature sets (303, 323 and, 338 K) [38].

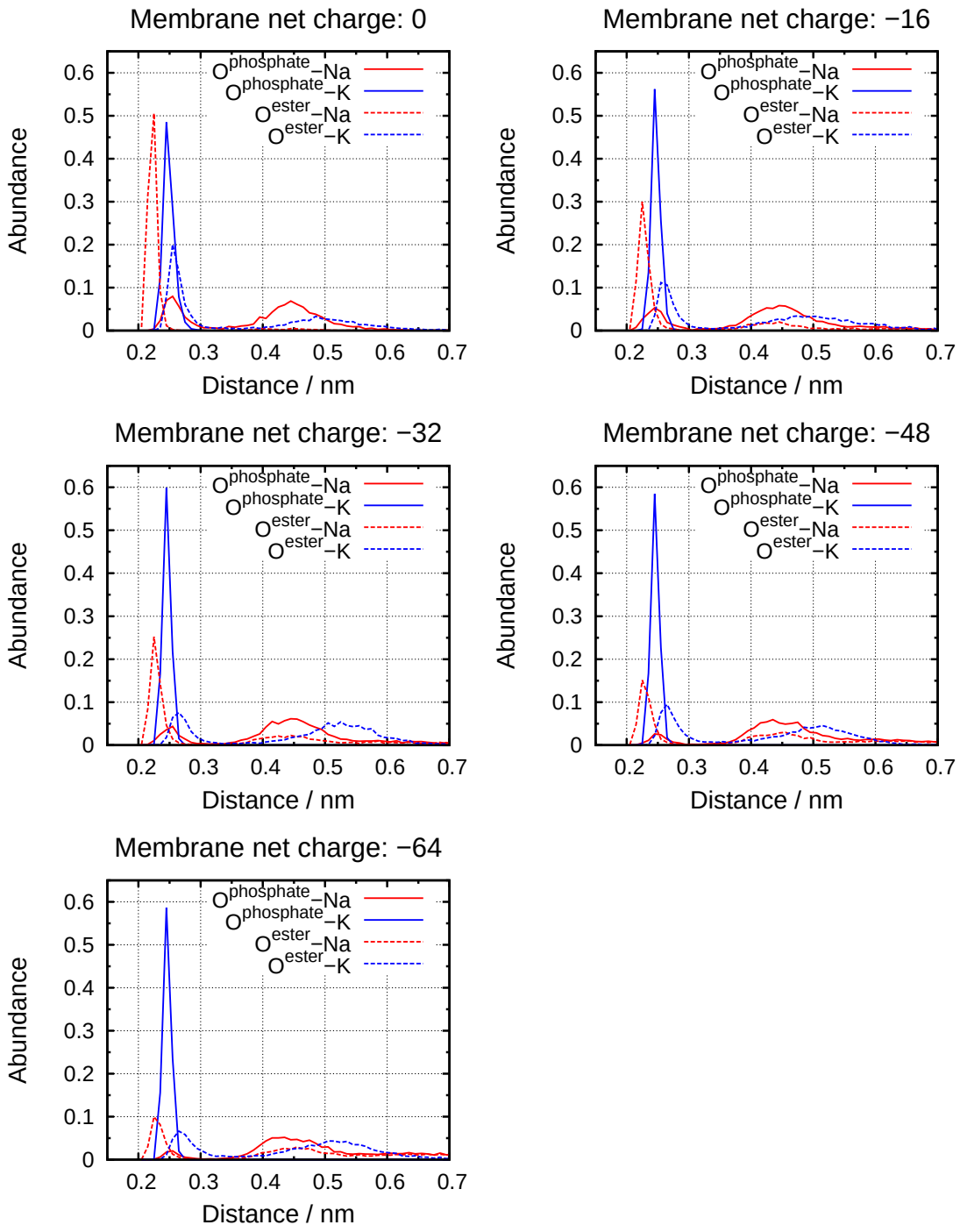


Figure 4.10: Minimum distance histograms for counterions relatively to O atoms of phosphates and esters (for PB^{NaCl} and PB^{KCl} simulations), normalized by the number of ions.

neutrality does not seem appropriate.

Our results show that the system neutralization of an ionized PA/PC bilayer represents an overestimation of the number of counterions. Consequently, the resulting lipid bilayer becomes too ordered and practically insensitive to ionization. A similar observation has been recently made where it was shown that the membrane moiety gets rigidified in an electroneutral system [83]. On the other hand, with our method, we were able to correctly model the ionization dependent isothermal phase transition of a 25% PA/PC bilayer [15]. This transition was observed following several structural properties usually used to describe lipid bilayers, namely area per lipid (A_l), deuterium order parameters, and membrane thickness.

We also compared two treatments of the long range electrostatics: PME and GRF. Using these two approaches in neutral systems, we observed similar and almost ionization insensitive behaviors in terms of A_l. It has been shown that PME in inhomogeneous systems (such as a bilayer-water mixture) with a system net charge leads to significant artifacts [233], rendering this approach inappropriate to model such small and charged systems.

GRF was also tested without any ions and we were able to observe a significant ionization dependent isothermal transition. However, in this group of simulations, the observed A_l values are too large for a typical fluid phase of DMPC and the ionization at which the transition occurs is not in complete agreement with the experimental data [15]. These results indicate that the "reaction field" approach alone does not stabilize enough a bilayer system with a large net charge.

The MD parameters of ions are difficult to obtain and have been the subject of several works in recent years [212, 235]. Despite such efforts, it has been shown that the Na^+ parameters used in this work might lead to an overbinding of this ion to the lipid bilayer [98, 238]. To circumvent these limitations, we performed our simulations with both Na^+ and K^+ as counterions. Interestingly, these ions have different preferential regions of interaction with the lipid molecules: Na^+ prefers to interact at deeper regions near the ester groups, while K^+ stays in a more shallow region close to the phosphates probably due to its larger size. However, with both ions the membranes showed similar behaviors in terms of all structural properties. Therefore, the role as electrostatic counterions is their most relevant effect, overcoming the artifacts that may arise from the MM description.

A PB-based approach to estimate the counterions distribution close to a charged surface can be considered a rough approximation. However, with this method, there is a good description of the heterogeneous charge distribution and the roughness of the membrane surface. Furthermore, our approach is not too computationally expensive and can easily be relevant to model diverse charged biomolecular systems in MD simulations. In the future, this method can also be very impor-

tant to study several pH-dependent events using the so-called constant-pH MD methodologies [3, 8, 10–13, 99–131].

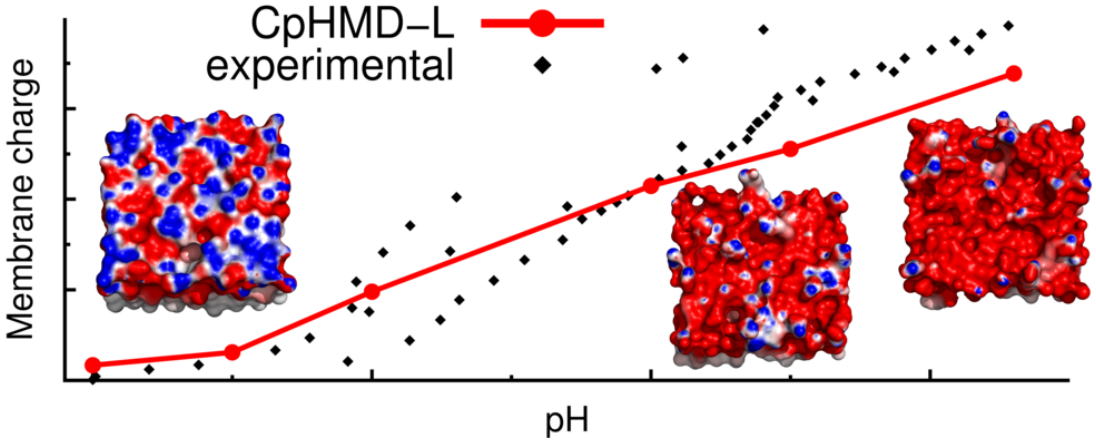
4.7 Acknowledgments

We thank Maria J. Calhorda, Paulo J. Costa, João Henriques, and Sara R. R. Campos for fruitful discussions. We acknowledge financial support from Fundação para a Ciência e a Tecnologia through projects PTDC/QUI-BIQ/113721/2009, PTDC/QEQ-COM/1623/2012 and PEst-OE/QUI/UI0612/2013 and grant SFRH /BD/81017/2011.

Chapter 5

Constant-pH MD simulations of an oleic acid bilayer

Diogo Vila-Viçosa[1], Vitor H. Teixeira, António M. Baptista, Miguel Machuqueiro



Contents

[1]In this work, I performed all the simulations and analysis. I was also involved in all steps of the manuscript preparation.

5.1 Context

This chapter closes the three stage process of developing a new extension to the stochastic titration constant-pH molecular dynamics method (CpHMD-L). This new implementation enables a correct treatment of periodic boundary conditions and ionic strength. In principle, this method is applicable to all systems with periodic boundary conditions in two directions, such as pure membranes, bilayers of lipid mixtures, peptides interacting with membranes, etc. As a first case study we used oleic acid, a simple molecule that spontaneously forms liposomes at some pH values. This work was published in Journal of Chemical Theory and Computation (2015,11:2367. DOI: 10.1021/acs.jctc.5b00095) and it is available at `http://pubs.acs.org/doi/abs/10.1021/acs.jctc.5b00095`.

5.2 Abstract

Oleic acid is a simple molecule with an aliphatic chain and a carboxylic group whose ionization and, consequently, intermolecular interactions are strongly dependent on the solution pH. The titration curve of these molecules was already obtained using different experimental methods, which have shown the lipid bilayer assemblies to be stable between pH 7.0 and 9.0. In this work, we take advantage of our recent implementations of periodic boundary conditions in Poisson–Boltzmann calculations and ionic strength treatment in simulations of charged lipid bilayers,

and we studied the ionization dependent behavior of an oleic acid bilayer using a new extension of the stochastic titration constant-pH MD method. With this new approach, we obtained titration curves that are in good agreement with the experimental data. Also, we were able to estimate the slope of the titration curve from charge fluctuations, which is an important test of thermodynamic consistency for the sampling in a constant-pH MD method. The simulations were performed for ionizations up to 50%, because an experimentally observed macroscopic transition to micelles occurs above this value. As previously seen for a binary mixture of a zwitterionic and an anionic lipid, we were able to reproduce experimental results with simulation boxes usually far from neutrality. This observation further supports the idea that a charged membrane strongly influences the ion distribution in its vicinity and that neutrality is achieved significantly far from the bilayer surface. The good results obtained with this extension of the stochastic titration constant-pH MD method strongly supports its usefulness to sample the coupling between configuration and protonation in these types of biophysical systems. This method stands now as a powerful tool to study more realistic lipid bilayers where pH can influence both the lipids and the solutes interacting with them.

5.3 Introduction

Lipids are one of the most important macromolecules in living systems. They have crucial roles as structural components of biological membranes, in energy storing, and as signaling molecules [18]. Many of the biologically relevant lipids are anionic and their ionization may depend on the solution pH [92]. Hence, pH affects many biological processes involving lipid protonation such as membrane fusion, cholesterol domain formation, drug-liposome interactions and membrane phase transitions [188].

The study of the properties of lipidic structures poses many different challenges when compared to the most studied macromolecules: proteins. Lipid conformations are intrinsically dependent on their macromolecular assemblies, and thus they can not be studied individually. Both experimental and theoretical methods need adjustment to study these molecules in their biologically relevant environment. In particular, molecular modeling techniques require the use of periodic boundary conditions (PBC) to mimic an infinite bilayer [1]. Also, the ionic strength local effects have to be accounted when the system size is as small as a standard simulation box [2]. Nevertheless, it is now possible to model, at atomic level, large and complex lipidic systems using molecular mechanics / dynamics (MM/MD) simulations with explicit solvent [239]. To increase the realism of these models, it is necessary to add anionic lipids whose ionization depends on the solution pH. However, MM/MD simulations deal with pH in a limited way, simply by setting

a fixed protonation state for each titrable group. On the other hand, there are simple models that can estimate the protonation state for a rigid structure, such as those based on the Poisson–Boltzmann (PB) equation [81, 167, 168, 196, 204, 240] and on a Monte Carlo (MC) sampling [176, 178, 179]. The limitations of the previous methods gave rise to the so-called constant-pH MD methods (CpHMD) [3, 8, 10–13, 99–131] that, in some cases, take advantage of the complementarity between MM/MD and PB/MC methods. In these methods, the titrable groups are allowed to change their protonation state along the simulation time. Thus, it is possible to study the protonation / conformation coupling in the system of interest by setting a pH value and running a simulation with pH as an external parameter such as temperature or pressure. These methods are often used with proteins and it is nowadays possible to perform complete titrations of these molecules [8, 10, 13, 101, 104–106, 112, 114, 119, 120, 122, 124–131, 241]. However, the application of such methods to lipid bilayers poses two main challenges: on the one hand, the correct description of the PBC in the PB calculations [1] (in the methods that use them) and, on the other hand, the treatment of the ionic strength effect close to charged lipid bilayers [2]. Recently, a constant-pH MD approach was applied to oleic acid aggregates using a coarse-grained model, [117] treating the protonation of the titrable residues as a continuous variable [105, 107, 108, 111] analogously to λ-dynamics [132]. Using a similar theoretical basis, Shen and coworkers studied the pH-dependent self-assembly of lauric acid at atomistic level [116].

The use of PBC is well established in biomolecular models and can be applied in the electrostatic calculations needed for the stochastic titration constant-pH MD method [1]. However, the treatment of ionic strength is not trivial and different approaches have been proposed. Charge-leveling has been used to preserve the electroneutrality in the context of a constant-pH MD method [115] and applied to several dicarboxylic acids. However, electroneutrality may not be appropriate in some systems. In particular, a charged lipid bilayer induces a local ion concentration that can be far from neutrality and very different from the bulk ionic strength [2, 41, 49, 74, 75, 78–82]. In Ref. 2, we proposed a PB-based method to estimate the number of ions close to a lipid bilayer and tested this estimate in MD simulations of a 25% PA/PC lipid mixture.

In this work, we developed a new extension of the stochastic titration constant-pH MD method [12, 13] to deal with lipid bilayers of oleic acid, taking advantage of the recent implementations of PBC in PB calculations [1] and ionic strength in membrane simulations [2]. Oleic acid is a simple molecule with a carboxylic group and a carbon chain of 18 carbons with one insaturation. The ionization of carboxylic acids and, consequently, their intermolecular interactions are strongly dependent on the solution pH. In particular, oleic acid can be present as either

liposomes or micelles [14, 59, 60, 90, 242]. Moreover, the apparent pK_a value of oleic acid in these aggregated phases is somewhat ill-defined and difficult to estimate [14, 59, 60, 90] and, depending on the used method, can vary between 7.5 at low concentrations inserted in a PC bilayer [85] and 9.85 in pure monolayers [89]. There are also three independent oleic acid titration experiments that report lipid bilayer assemblies between pH 7.0 and 9.0 (above this value a phase transition into micelles is observed) [14, 59, 60]. Here, we performed several constant-pH MD simulations of a bilayer of oleic acid in order to study the pH-dependence of its structure. The total titration curve obtained for the bilayer was compared with the available experimental data measured for this macromolecular assembly.

5.4 Methods

5.4.1 Stochastic titration constant-pH MD

The stochastic titration constant-pH MD method [12, 13] works in a cycle with three main steps, with the loop running until the end of the desired simulation time (Figures 5.1 and 2.5). In the Poisson–Boltzmann / Monte Carlo (PB/MC) step, the protonation states of the solute are sampled using MC calculations [176, 178, 179] with free energies obtained from PB calculations [81, 167, 168, 196, 204, 240]. The final protonation states are then assigned to the solute in the Molecular Mechanics / Molecular Dynamics (MM/MD) steps. The second block consists of a short MM/MD simulation (during a time interval τ_{rlx}) with the solute frozen and in the NVT ensemble with the objective of relaxing the water molecules to the new protonation states. The last step of the cycle is a production MM/MD simulation during a time interval τ_{prt} with all system unfrozen.

In a bilayer environment, the correct description of the periodicity of the system along the membrane plane is crucial. In the MD steps, the periodic boundary conditions (PBC) corrections are readily available in most software packages. However, for PB calculations, the use of 2D-PBC presents a few new challenges [1]. We have addressed the inclusion of 2D-PBC in the PB calculations to study the protonation of a DMPC molecule using a linear response approximation [1]. The same approach was adopted in the current CpHMD implementation for lipidic systems (CpHMD-L).

The treatment of the ionic strength effect on a charged lipid bilayer is not trivial. In particular, it has been suggested by several authors that the use of full neutralization is not the most adequate approach to deal with this problem [41, 49, 74, 75, 78–82]. To simulate a neutral system one would need to use a very large simulation box (in z direction) which would be too time-consuming. Recently, we have developed a method to deal with ionic strength in charged

systems [2]. This method uses a PB model to estimate the amount of counterions and coions near a charged membrane. The amount of ions is computed in the region between the membrane surface and the cutoff value used in the treatment of long-range electrostatics (1.4 nm; see section MM/MD settings). The region beyond the cutoff value is kept relatively small avoiding significant trapping of ions. The PB-estimate of the total amount of ions in this region (n_{PB}) is given by:

$$n_{\mathrm{PB}} = \int_C \rho(\mathbf{r})\mathrm{d}v \approx V N_{\mathrm{A}} I \sum_{i=1}^{M} \mathrm{e}^{-ze\phi_i/k_{\mathrm{B}}T} \tag{5.1}$$

where the region of interest C is discretized into a set of M small cells, each with volume V and approximately constant electrostatic potential, $\rho(\mathbf{r})$ is the number density of an ion at position $\mathbf{r}$ with charge z, ϕ_i is the electrostatic potential in cell i, ρ_0 is the bulk number density, e is the elementary charge, k_{B} is the Boltzmann constant, I is the ionic strength and T is the absolute temperature. The number of cations, $n_{\mathrm{PB}}(\mathrm{Na}^+)$, and of anions, $n_{\mathrm{PB}}(\mathrm{Cl}^-)$, is calculated for each of a set of configurations from a MD simulation. The median, which is robust to outliers, of these calculated values is then used as an estimation for the next MD simulation of the same system.

The final numbers of explicit ions, $n_{\mathrm{exp}}(\mathrm{Na}^+)$ and $n_{\mathrm{exp}}(\mathrm{Cl}^-)$, are determined as follows (Figure 5.1). A first MM/MD simulation is performed by setting n_{exp} to any number between 0 and a sufficient amount to neutralize the system (the initial guess does not affect the final converged result). After this, the first set of n_{PB} values is calculated from equation 5.1 and the number of ions in the simulation (n_{exp}) is adjusted accordingly. We then determine the average numbers of explicit ions that are closer than 1.4 nm to the membrane in the new simulation, $n_{\mathrm{eval}}(\mathrm{Na}^+)$ and $n_{\mathrm{eval}}(\mathrm{Cl}^-)$. After the calculation of both n_{PB} and n_{eval} the two pairs of values are compared and, if necessary, n_{exp} is adjusted for a new MD simulation. This process continues iteratively until $n_{\mathrm{eval}} \approx n_{\mathrm{PB}}$, i.e. the process becomes self-consistent (usually after 2 or 3 iterations).

In this work, we take advantage of these two recent developments [1, 2] regarding the PBC in PB calculations and the explicit ions estimation in charged bilayer systems, and we implement them in an extension to the stochastic titration constant-pH MD method (CpHMD-L).

5.4.2 Charge fluctuations and slope of the titration curve

All protonation states of a system with multiple titration groups are in equilibrium with each other and the solution. Which means that, at a given pH value, the total charge of the system (in our case, a membrane) is not a constant, rather fluctuating around an average value (Figure 5.4a). These fluctuations are related

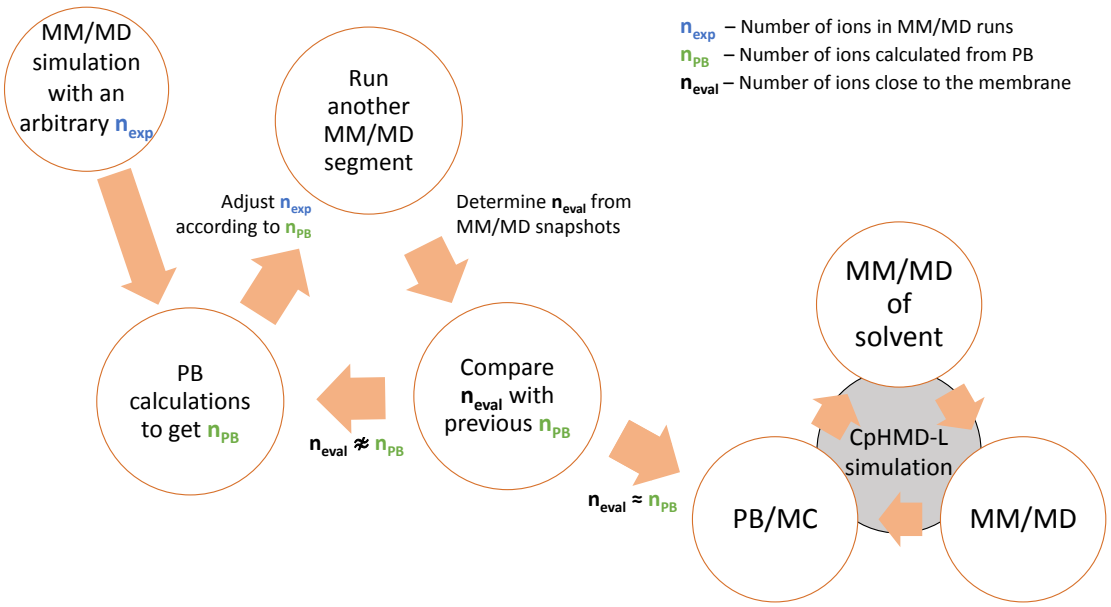


Figure 5.1: Workflow to calculate the final number of explicit ions obtained from the self-consistent procedure (adapted from Ref. 2) in the context of constant-pH MD algorithm.

with the slope of the titration curve through [243]:

$$\frac{\partial \bar{Z}}{\partial \mathrm{pH}} = -\ln(10)\sigma^2 \tag{5.2}$$

where $\bar{Z}$ and σ are, respectively, the average and the standard deviation of the membrane charge at a given pH value. This is a particular case of the general equation holding for the number of bound molecules of a ligand, relating its fluctuation with the derivative of its average with respect to ligand activity [244], which is itself a particular case of the analogous relation for the number of particles in a grand canonical ensemble [245].

The above relation can be used to test if the CpHMD-L method is able to correctly sample the charge fluctuations of a membrane. This was done here by comparing two cubic splines computed from the simulation data using slopes taken either from the MD simulations (through equation 5.2) or directly estimated from the average protonations using the Davis–Dowden method for monotone data [246]. The two splines thus obtained should be very similar if our constant-pH MD method is thermodynamically consistent, as it was found indeed to be the case (Figure 5.4b and the accompanying text).

Table 5.1: Summary of the simulations performed in this work (100 oleic acid molecules were used in all simulations).

I / M	pH	$n_{\mathrm{exp}}(\mathrm{Na}^+)$	$n_{\mathrm{exp}}(\mathrm{Cl}^-)$
0.1	6.00	2	2
	6.50	2	2
	7.00	3	2
	7.25	4	2
0.3	6.00	5	5
	6.50	6	5
	7.00	10	8
	8.00	14	8
0.5	6.00	9	9
	6.50	10	9
	7.00	15	10
	8.00	20	12
	8.50	26	12
	9.30	34	12

5.4.3 Constant-pH MD settings

The CpHMD-L simulations were performed using three ionic strength values (0.1, 0.3, and 0.5 M). The values above 0.1 were chosen in order to increase the membrane stability at higher pH values (see Results and Discussion). It should also be noted that the ionic strength values were not reported in the titration experiments [14, 59, 60]. All simulations were done in triplicate for 100 ns each (Table 5.1). The number of ions in each simulation depends on the ionic strength and on the membrane ionization. All systems were pre-equilibrated using standard MM/MD simulations with ionization from 0% to 50%. For these equilibrations, we used the self-consistent algorithm (Figure 5.1) previously explained. The self-consistency between n_{eval} and n_{PB} values for all CpHMD-L simulations was confirmed *a posteriori* (see section Estimation of the number of ions and system equilibration).

We also tested several τ_{prt} values, 2, 20, and 50 ps, starting from structures equilibrated at 20, 30, 40, and 50% of ionization, from one structure with high area per lipid (A_{l}) and another with low A_{l}. These tests were necessary because the best protonation frequency for lipid bilayers may differ from the one used in

protein systems (usually 2 ps). The best values are always a compromise between computational speed and the convergence of protonation and configuration, which are coupled. The τ_{rlx} value, on the other hand, was the same as for proteins (0.2 ps). Because all these simulations were performed at pH 8.0 and $I = 0.5$ M, in principle, they should converge to the same values (of both A_{l} and ionization—see Results and discussion).

5.4.4 MM/MD settings

MM/MD simulations were done using the GROMOS 54a7 force field [212] and a modified version [12, 13] of GROMACS distribution (version 4.0.7 [210, 211]). The MM parameters for the hydrocarbon chain of oleic acid were taken from the POPC molecule (in Ref. 138) and for the carboxylic group were taken from the glutamate side chain present in GROMOS 54a7 force field [212]. All simulations were performed using Na^+ as a cation, even though we have shown that similar results can be obtained with K^+ [2]. Cl^- was used as anion in all simulations. The membrane was solvated in a tetragonal box with SPC water molecules [214], with the membrane surface oriented parallel to the xy plane. All bond-lengths were constrained using the parallel version of the LINCS algorithm for lipids [163], and the SETTLE algorithm [164] for water. All NPT simulations were done using the v-rescale thermostat [161] at 298 K with separate couplings for the solute and solvent (including ions) with a relaxation time of 0.1 ps. A semi-isotropic Berendsen [160] pressure couple was used to maintain the pressure constant at 1 bar with a compressibility of 4.5×10^{-5} bar^{-1} and a relaxation time of 5.0 ps. In CpHMD-L simulations, the solvent relaxation steps were performed in the NVT ensemble. An integration step of 2 fs was used in the MD equations of motion. Nonbonded interactions were treated with a twin-range method, using group-based cutoffs of 0.8 and 1.4 nm, updated every five steps. Long-range electrostatics were corrected with a generalized reaction field (GRF) [13, 155], using a dielectric constant of 54.0, which has been shown to minimize artifacts due to electrostatic interactions truncation and their approximation by an infinite dielectric continuum [215].

The minimization procedure consisted of three sequential steps: starting with the steepest descent algorithm (unconstrained), followed by 10 000 steps using the l-BFGS integrator (unconstrained), and ended with another step using the steepest descent algorithm (with all bonds constrained).

The same initialization procedure was applied to all systems and consisted of four steps. First, a 50 ps MD simulation was done with all heavy atoms harmonically restrained with a force constant of 1 000 kJ mol^{-1} nm^{-2}. In the three following steps of 100, 150, and 200 ps MD simulations, position restraints were applied differently for the carbon atom of the carboxylic group and remaining

heavy atoms, with force constants of 1 000, 100, and 10 kJ mol^{-1} nm^{-2} and of 100, 10, and 0 kJ mol^{-1} nm^{-2}, respectively.

5.4.5 PB/MC settings for CpHMD-L simulations

The PB calculations were done with the program DelPhi V5.1 [205, 206] using atomic radii and partial charges taken from GROMOS 54a7 force field [212]. The molecular surface of the membrane was defined by a probe of radius 0.14 nm, and the ion exclusion layer was 0.2 nm. In order to avoid discontinuities in the PBC surface (Delphi does not calculate the molecular surface taking PBC into account) we added a small portion (5% of the box side dimension) of the membrane atoms in the x and y directions [1]. PBC were explicitly applied along those two directions in the calculation of the potential. A convergence threshold value based on maximum change of potential of 0.01 k_B T/e was used. All PB calculations in the CpHMD-L workflow were done in a cubic grid with 45 grid points (corresponding to a grid space of $\sim$ 0.08 nm) and a two step focusing [219] where the focus grid was one fourth of the coarser grid size. In the iteration convergence process of finite differences method, we used relaxation parameters for the nonlinear and linear forms of the PB equation with values of respectively 0.75 and 0.20. A cutoff of 1.6 nm was used to calculate the background and pairwise interactions (see Ref. 1 for more details). The pK_a values of the model compounds (carboxylic group) were taken from the pK_a value of acetic acid in solution (4.76) [247]. The dielectric constant of the solvent was 80, and for the membrane, we used a value of 2.

The MC sampling was performed with the PETIT program [176] version 1.6 using an absolute temperature of 298 K. All runs were performed using 10^5 MC cycles, where one cycle consists of sequential state changes over all individual sites and pairs of sites with an interaction larger than 2 pH units.

5.4.6 PB settings for ion estimation

The PB calculations performed for the ion estimation are similar to those described in the previous section. The focusing procedure was not used because we are interested in all grid points closer than 1.4 nm to the membrane. In the iteration convergence process of finite differences method, we used relaxation parameters for the nonlinear and linear forms of the PB equation with values of respectively 0.15 and 0.20. In these calculations, we used 450 configurations from each simulation. To estimate the number of ions (n_{PB}), the electrostatic potential is written to a new 3D grid with 50 points per side, the same size as the finite differences grid [1]. The number of ions were calculated according to equation 5.1.

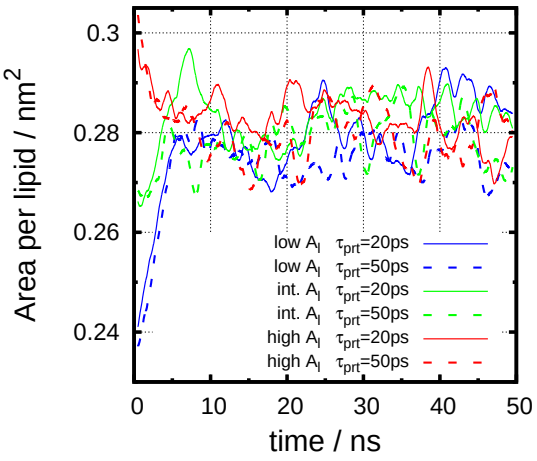


Figure 5.2: Time variation of area per lipid in simulations to calibrate the τ_{prt} parameter. Three groups of simulations are shown, starting from low, intermediate, and high A_{l}.

5.4.7 Analysis

The last 90 ns of each simulation were used for analysis. Several tools from the GROMACS software package [210, 211] were used. Area per lipid (A_{l}) was defined as the average xy area occupied by a lipid molecule. Deuterium order parameters were calculated taking in consideration a bug correction introduced in GROMACS 4.5.4 [248]. All presented errors were computed using standard methods based on the autocorrelation function of the property measured to determine the number of independent blocks in the simulations [134].

5.5 Results and discussion

5.5.1 Calibration of the τ_{prt} parameter

The duration of the production MM/MD step in the constant-pH MD cycle (τ_{prt}) is an important variable in our constant-pH MD simulations, since a value that is too large may originate insufficient sampling of protonation states, and a value that is too small may increase the computational cost and destabilize the solute conformation / configuration (see below). Hence, we performed a set of simulations with three τ_{prt} values (2, 20, and 50 ps—see Methods for details). In these simulations, we checked the convergence of both A_{l} and ionization (Figures 5.2, A.1, and A.2). For $\tau_{\mathrm{prt}} = 2$ ps, the membrane became too unstable and disrupted

in the first nanoseconds of simulation. This destabilization was probably caused by the large oscillations in the membrane charge. Despite the fact that this τ_{prt} value is usually used in simulations of proteins with the stochastic titration constant-pH MD method, [12, 13] it might not be appropriate for a lipid bilayer. The fluctuations of the amino acid side chains are usually in a time scale much faster than lipid diffusion, therefore a larger τ_{prt} value is needed for the lipids to adapt to new protonation states. Not doing so can result in local instabilities and membrane disruption. For both $\tau_{\mathrm{prt}} = 20$ ps and $\tau_{\mathrm{prt}} = 50$ ps, all the simulations were structurally stable during 50 ns and both A_{l} and protonation converged rapidly (in less than 10 ns) to the same values (Figures 5.2, A.1, and A.2). Therefore, both τ_{prt} values were suitable to be used in the present work. We chose the τ_{prt} value of 20 ps mainly due to its more extensive sampling of protonation states and the consequent smaller chance of falling into kinetic traps during the simulations. A smaller τ_{prt} could better capture the protonation / configuration coupling in a lipid bilayer, which can be challenging for lipid molecules other than oleic acid.

5.5.2 Estimation of the number of ions and system equilibration

Ionic strength is a macroscopic property that is very complex to modulate at a microscopic level. As explained in the methods section we used a methodology described in Ref. 2 to deal with the ionic strength effects in our systems. Hence, the number of ions in our simulations is not always the same and it depends both on the ionic strength and on the average degree of ionization. It is possible to confirm the number of ions that should be present in a CpHMD-L simulation comparing the number of explicit ions that we introduced *a priori* (Table 5.1) with an estimation using the PB formalism [2].

The number of ions was calculated by averaging the estimations with the PB-based approach from 450 structures taken from the last 90 ns of each simulation (Figure 5.3). The number of both sodium and chloride ions increases with ionic strength. Hence, at higher ionic strength values, there are more ions available to stabilize the repulsions between the headgroups leading to more stable membranes. Interestingly, despite our membrane being negatively charged, there are always chloride ions. The fact that the number of these ions does not change significantly with ionization, can be explained by two opposite effects. On the one hand, its concentration decreases near the membrane, due to the negative potential generated by the anionic lipids (see the ion concentration variation with the ionization in Figure A.3). On the other hand, the A_{l} (Figure 5.5), hence the volume where the number of ions is estimated, increases, thereby compensating the first effect.

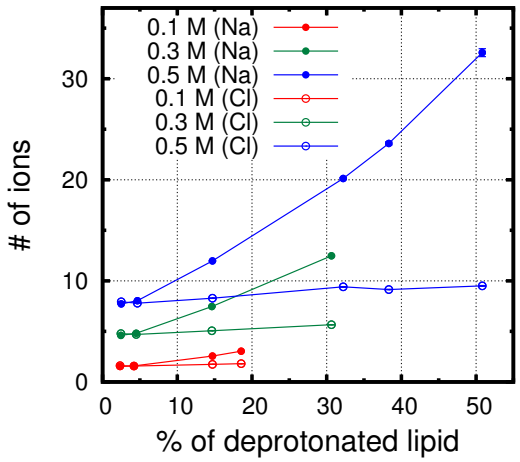


Figure 5.3: Number of counterions and coions in all CpHMD-L simulations at different membrane ionization values. Membrane ionization was obtained from the total titration curves (Figure 5.4).

Sodium ions, which in principle are the more important stabilizers of the negatively charged membranes, are also present in a small number (according to the ionic strength) when the membrane is almost neutral (Figure A.4 for a representation of the PB-estimated electrostatic potential used to estimate the number of ions). This number increases with the bilayer ionization because the counterions are attracted to the more negative membrane.

5.5.3 Membrane titration

The CpHMD-L simulations of an oleic acid bilayer allow us to follow the total membrane charge over time at different pH values (Figure 5.4a). In all simulations, this charge converges very fast (in the subnanosecond time scale) to an average value. In addition, the charge fluctuations around the average are significantly large (more than 15 charge units at high pH values), which emphasizes the need to use the CpHMD-L method with these systems rather than standard MM/MD simulations at constant protonation. As mentioned in the Methods section, the accuracy of these fluctuations can be checked by converting them into slopes of the titration curves and comparing the resulting cubic splines with those obtained with a standard method (Figure 5.4b). The very good agreement between the two cubic splines indicates that our method is correctly sampling not only the average protonation values (see below) but also the charge fluctuations.

The titration curve of oleic acid was computed by averaging the occupancy

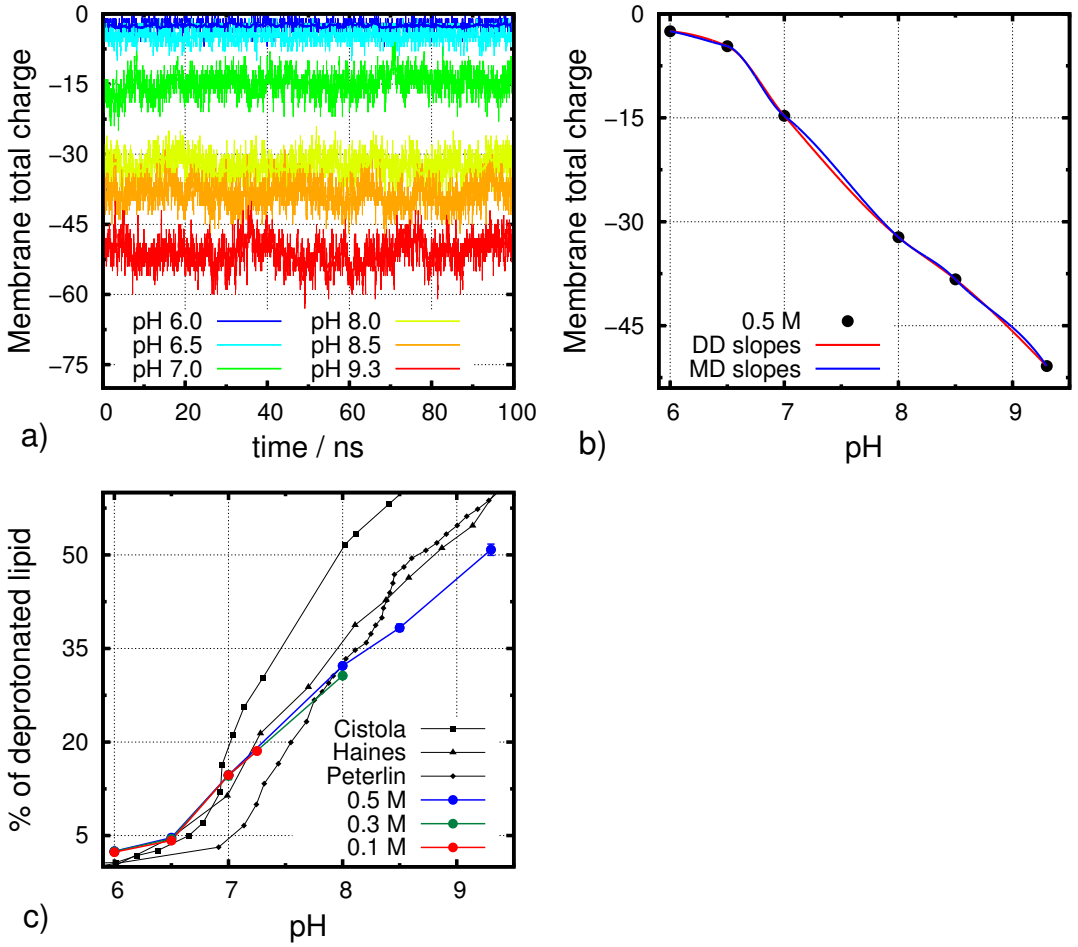


Figure 5.4: Titration of lipid bilayer. Membrane charge vs simulation time for one replicate per pH at $I = 0.5$ M (**a**). Simulated titration curves where the values are connected using cubic splines whose slopes are estimated using either equation 5.2 (MD) or the Davis–Dowden method (DD) (**b**). Simulated and experimental titration curves (**c**). All simulated titration curves were obtained in this work with the CpHMD-L simulations and for the experimental we used the data from Cistola et al. [59], Haines et al. [14] and Peterlin et al. [60].

states of all oleic acid molecules over the last 90 ns of each simulation (Figure 5.4c). Our simulations were done using three ionic strength values (see "Methods") and compared with different experimental results [14, 59, 60]. There are no abrupt changes in the titration curves and there is a large shift between the titration region and the pK^{mod} value (4.76), as expected for such a highly anti-cooperative system.

Haines et al. have measured a potentiometric titration curve at room temperature from pH 12 to pH 6 of a dispersion of 50 mM oleic acid and 10 mM lauric acid [14]. More recently, Peterlin et al. obtained an equilibrium titration curve of a 9.9 mM pure oleic acid suspension at 298 K [60]. Despite the slight discrepancy at lower ionization (Figure 5.4c), there is consistency between these two independent measurements in the bilayer region. We observe good agreement between our titration curves and these experiments for all ionic strength values. In fact, our curves are not significantly shifted upon changes in the ionic strength. This observation may be related with two opposite effects of ionic strength on the protonation of lipids in our model. On the one hand, at a higher ionic strength the interaction between the headgroups are weaker, leading to higher ionization and the consequent deviation of the titration curve to lower pH values. On the other hand, the same effect can approximate the headgroups (lowering the A_{l}) resulting in higher percentage of protonation and deviating the curve to higher pH values.

Cistola et al. have also measured a titration curve of pure oleic acid (80 mM) at 313 K following ^{13}C NMR chemical shifts [59]. This experimental curve is shifted to lower pH values compared to our results and the other experimental curves. These measurements were carried out at higher temperature, and thus it is expected that the average distance between the headgroups and, consequently, the solvent exposure, are higher, favoring the ionized form of these groups.

The experimental titration curves were obtained for a wide range of pH values. However, we were only able to run our simulations at pH values up to 7.25, 8.00, and 9.30 using, respectively, I = 0.1, 0.3, and 0.5 M. Above these values, the membrane became unstable and disrupted due to the strong repulsions between the negatively charged headgroups. At lower ionic strength values, the number of counterions available are not sufficient to stabilize these repulsions even at lower ionizations. The membrane disruption at I = 0.5 M and pH 9.30 is in agreement with the experiments suggesting that above pH 9.0 (8.0 in Ref. 59 due to a higher temperature) a phase transition to micelles occurs. In fact, it is possible that the small deviation observed between our titration curve and the experiments [14, 60] at higher ionizations can be explained by the coexistence of liposomes and micelles in the experimental dispersion. At lower ionization, a phase transition into oil droplets has been experimentally observed [14]. In our simulations, because all systems were built as bilayers, the oil droplet phase representation is charac-

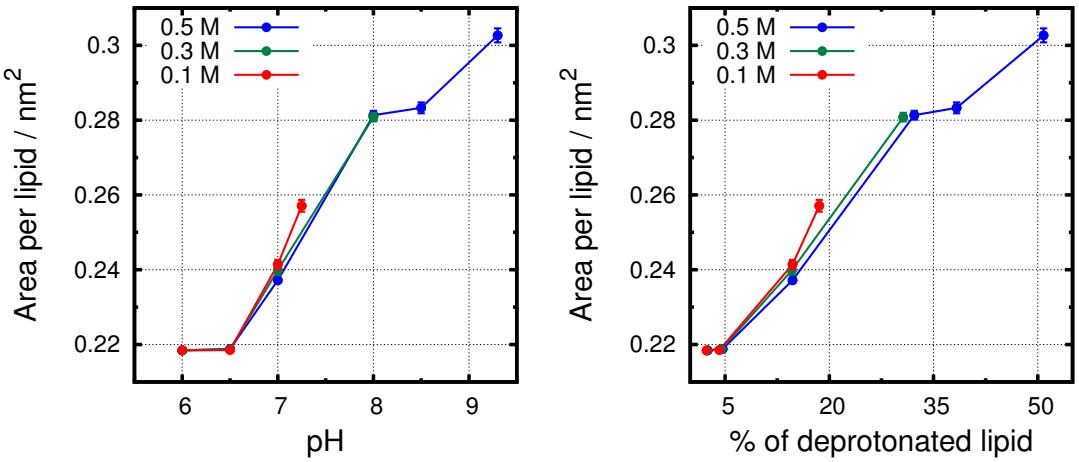


Figure 5.5: Average area per lipid (A_l) at different pH (left panel) and membrane ionization (right panel) values.

terized by tightly packed configurations with slow diffusion rates (Table 5.2 and Figures A.9, A.10, and A.11).

The total system charge varies approximately between −2 (at low pH) and −25 (at high pH) (Figure 5.3), which results in several systems far from neutrality. In our work, we were able to reproduce the experimental titration curve with stable bilayers, further supporting the idea that such small and charged systems do not always need to be neutralized. These findings are in agreement with previous observations about neutrality only being achieved much farther away from a charged membrane [2, 41, 49, 74, 75, 78–82].

5.5.4 Membrane structural parameters

The area per lipid (A_l) is one of the most commonly used properties to describe the behavior of lipidic membranes and directly reflects their phase transitions. From our CpHMD-L simulations, we can follow the x and y dimensions of the simulation box at different pH/ionization values and calculate the A_l of the lipid bilayer (Figure 5.5). This property increases with pH in almost all simulated pH range. This is expected since the ionization of the carboxylic headgroups leads to higher electrostatic repulsions (Figure 5.6). Nevertheless, we observe two plateaus at pH 6.0–6.5 ($\sim$ 3–5% of ionization) and at pH 8.0–8.5 ($\sim$ 30–40%). In the first case, the lipid bilayer is in a very ordered gel phase ($\sim$ 0.22 nm^2) with tightly packed lipid configurations (Figure 5.6, left panel). Only at slightly higher pH values (pH 7.00) it is possible to disrupt these very stable configurations and

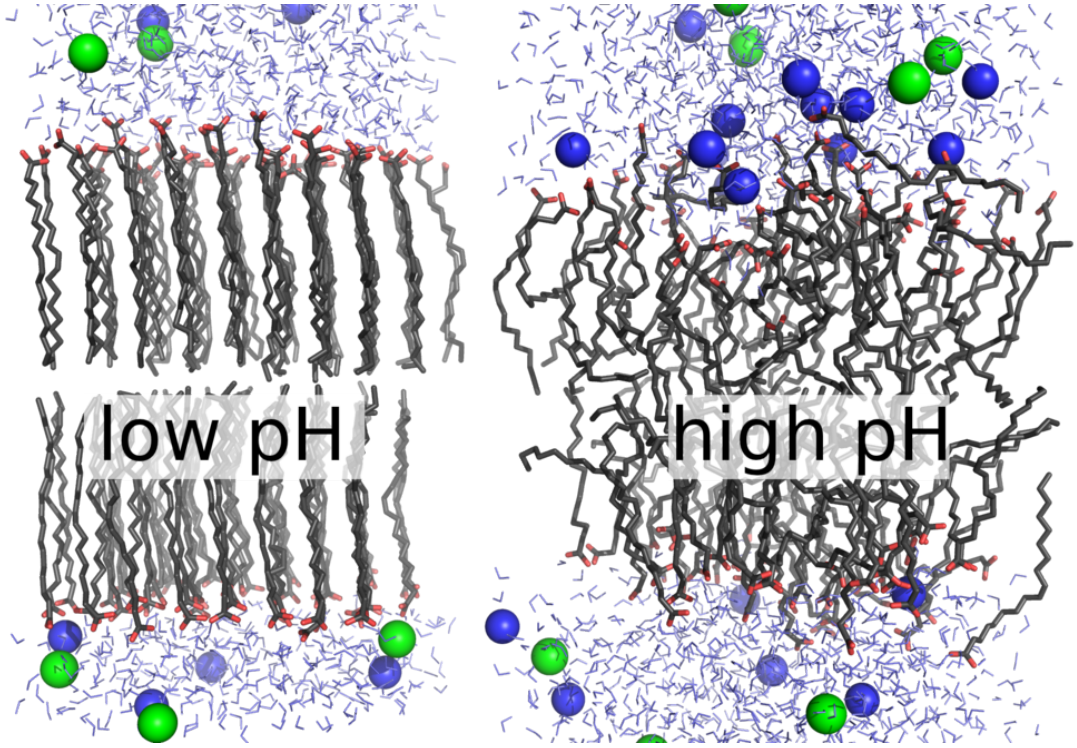


Figure 5.6: Graphical representation of representative configurations of our system at low and high pH values. Oleic acid molecules are shown in gray (C atoms) and red (oxygen atoms) sticks, water molecules in blue, and Cl^- and Na^+ ions in, respectively, green and blue spheres.

increase both ionization and A_l. The second plateau is probably related with the bilayer stability limits, because we observed that the bilayer becomes very unstable at pH 9.3, as pointed above.

As observed for the total titration curves, there is no significant influence of the ionic strength on the A_l. At lower ionic strength, the lipid bilayer gets disrupted at relatively low A_l values due to the lack of enough stabilizing counterions. There is a clear pH-dependent isothermal phase transition between a very ordered gel phase (pH 6.0–6.5) and a highly dynamic fluid phase (pH 9.3). At higher pH values, the lipid bilayer gets disrupted, probably because micelles became the most stable assembly, as shown in many experiments [14, 59, 60]. A detailed study of these micellar assemblies could not be accomplished with our systems construct and is out of the scope of this work.

The isothermal phase transition can also be followed using deuterium order parameters of the oleic acid aliphatic tail (Figure 5.7, A.5 , A.6, A.7, and A.8).

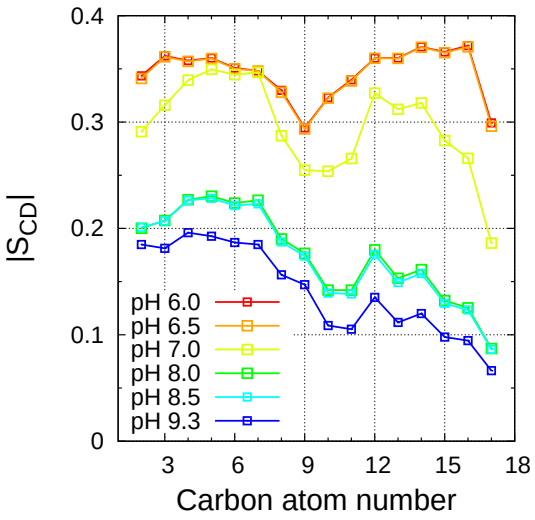


Figure 5.7: Deuterium order parameters for all CpHMD-L simulations at $I = 0.5$ M. Note that the curves for pH 6.0 and 6.5 are almost coincident, as well as those for pH 8.0 and 8.5.

Similarly to what is observed for A_l, the order of the lipidic tails changes drastically with pH/ionization. The S_{CD} values at the very ordered gel phases (0.36 for C5 at pH 6.0–6.5) are significantly higher than the available experimentally measured values for DPPC slightly below the melting temperature (0.23 for C5 at 314 K [249]). On the other hand, at higher pH values, the observed S_{CD} (0.19 for C5 at pH 9.3) are similar to the fluid phase for the oleyl chain of POPC ($\sim$ 0.2 for C5 [250, 251]). Therefore, the order parameters in our simulations also illustrate the pH-dependent isothermal phase transition already indicated by the A_l results.

Lateral diffusion of lipids is a measure of how fast the molecules diffuse along the membrane xy coordinates. In computer simulations, the diffusion coefficient can be obtained from the Einstein relation measuring the slope of the mean square displacement (MSD) over time. The diffusion coefficient values obtained increase with pH (Table 5.2, and Figures A.9, A.10, and A.11), illustrating the isothermal phase transition and an increased fluidity of the membrane. At low pH values, in the more ordered gel phase, the lipids are almost frozen in our time scale, which did not allow us to calculate their diffusion coefficient. This also helps to explain the unusually high order parameters (Figure 5.7), and the presence of an ionization and A_l (Figure 5.5) plateau observed at these pH values. At high pH values, oleic acid can diffuse approximately 1 order of magnitude faster than what has been measured for a POPC bilayer in fluid phase ($0.15 - 0.48 \times 10^{-7}$ cm^2 s^{-1}) [252, 253].

Table 5.2: Diffusion coefficient (10^{-7} cm^2 s^{-1}) obtained for all CpHMD-L simulations.

pH \ I / M	0.1	0.3	0.5
6.00		~ 0	
6.50			
7.00	1.4	0.7	0.3
7.25	4.2	–	–
8.00	–	7.3	6.1
8.50	–	–	6.6
9.30	–	–	7.8

5.5.5 Counterion concentration and distribution near the membrane

The total numbers of ions (n_{PB}) are estimated using a PB-based approach to determine the number of explicit ions to be placed in a MD simulation (n_{exp}). This estimation can also be performed using a Gouy–Chapman (GC) based approach [2] (i.e., an analytical (and, hence, much faster to calculate) solution of the PB equation for an uniformly charged infinite surface). Using these two methods (PB and GC), we determined the cumulative number of counterions as a function of the distance to the membrane (Figure 5.8, left panel) for a CpHMD-L simulation at pH 8.0 and $I = 0.5$ M. The accumulated number of explicit counterions observed in the MD simulation is also shown. The significant agreement between the MD and PB curves was expected because, in our method, their total number of ions are self-consistent. However, the number of ions estimated by the GC approach is significantly higher compared with the PB approach. This discrepancy highlights the limitations of applying a simplified approach such as GC to the irregular molecular surface of a significantly charged oleic acid bilayer with a low dielectric interior (note that the GC model lacks image-charge effects that repel the ions [254]). Apparently, the molecular detail introduced in the PB formalism is important to stabilize the strong interactions between carboxylic groups, hence, lowering the magnitude of the negative electrostatic potential near the membrane.

In our approach, the total number of explicit ions in MD is equal to the PB estimation. Nevertheless, the counterion distribution near the membrane can be compared (Figure 5.8, right panel). The minimum distance histograms of the ex-

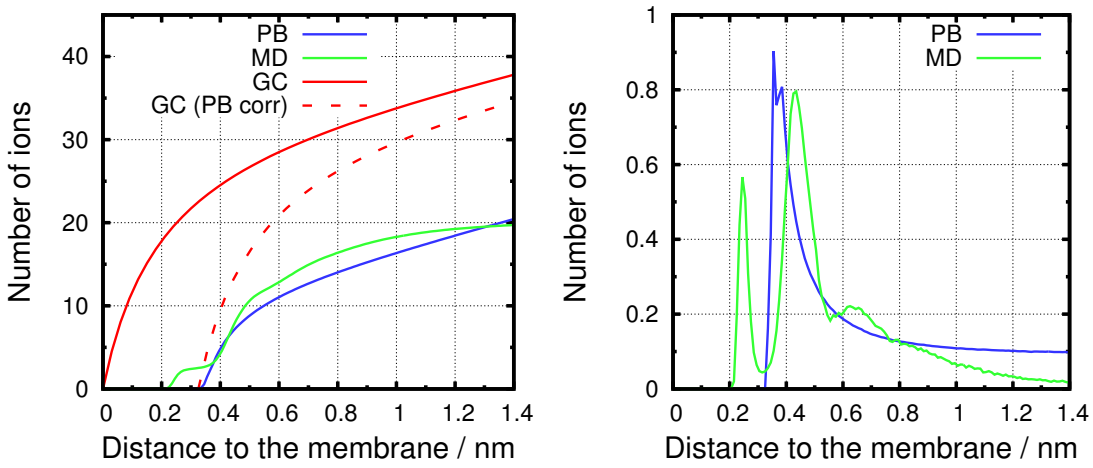


Figure 5.8: Variation of cumulative number (left panel) and minimum distance histograms (right panel) of counterions with distance to the membrane, calculated for a CpHMD-L simulation at pH 8.0 and $I = 0.5$ M with three different methods. The "GC (PB corr)" curve was obtained by assuming that the minimum distance of the ions to the membrane is not zero, as assumed in standard GC theory (e.g., see equation 1 in the Supporting Information of Ref. 2), but rather the closest distance allowed by the PB method used in this work; this corresponds to impose a Stern-like region of ion exclusion.

plicit ions show two preferential positions (around 0.25 and 0.43 nm—Figures A.12, and A.13), corresponding to strong interactions with the lipid headgroups. The difference between these two regions is related with the first water solvation layer of the Na^+ ions (see a radial distribution function and the number of water molecules surrounding Na^+ ions in Figures A.14, and A.15). At 0.43 nm, the ion is interacting with the bilayer and still retains its complete first water shell, while the strong interactions at 0.25 nm implies a partial desolvation of the ion, with the associated free energy cost. This behavior cannot be captured using a mean–field approach such as PB. Therefore, the ion distribution only shows one peak whose position (around 0.36 nm) depends only on the molecular surface and on the width of the ion exclusion layer (0.2 nm). At longer distances (between 1.0 and 1.4 nm), the MD curve diverges from PB showing an unexpected absence of ions in this region. This result is in agreement with recent results suggesting an overbinding of Na^+ ions to lipid bilayers [98]. Nevertheless, PB seems to capture the overall ion distribution profile observed in the MD simulation.

5.6 Conclusions

In the present work, we developed an extension to the stochastic titration constant-pH MD method [12, 13] that is able to deal with PBC in x and y directions of the PB calculations and take into account the explicit ion effects at a given ionic strength value. With the PBC correction, it is possible to avoid electrostatic calculation errors that may arise from surface discontinuities at the box edges and pairwise interactions across 2D-PBC boxes [1]. Because an inadequate treatment of ionic strength effects in charged lipid bilayers may also introduce significant errors, we took advantage of a recently published PB formalism [2] to estimate the number of explicit ions in biomolecular simulations of charged bilayers to avoid such problems. As far as we know, we are presenting the first constant-pH MD simulations of continuous (infinite) lipid bilayers.

The CpHMD-L method was tested in a bilayer of oleic acid, a carboxylic acid with 18 carbons and one insaturation. With this method, we obtained titration curves that are in good agreement with the experimental data. Also, we were able to estimate the slope of the titration curve from charge fluctuations, which is a strong evidence that these are being correctly sampled in our calculations. The simulations were performed for ionizations until ~ 50 % because, above this value, a macroscopic transition to micelles occur. At ionic strength values below 0.5 M, we were not able to obtain a stable bilayer up to ~ 50 % of ionization. This observation is related with the high repulsions between charged headgroups that need to be stabilized by explicit counterions. As previously observed for a binary mixture of a zwitterionic and an anionic lipid, [2] we were able to reproduce experimental results with simulation boxes usually far from neutrality. This observation further supports the idea that a charged membrane strongly influences the ion distribution in its vicinity, and the neutrality is achieved significantly far from the bilayer surface.

The results obtained with this extension of the stochastic titration constant-pH MD method strongly support the importance to sample the coupling between configuration and protonation in these types of biophysical systems [3, 8, 10–13, 99–131]. The CpHMD-L method stands now as a valuable tool to study more realistic lipid bilayers where pH can influence both lipids and solutes that might be interacting in the lipidic environment.

5.7 Acknowledgments

We thank Maria J. Calhorda, Sara R. R. Campos, and Hugo A. F. Santos for fruitful discussions. We acknowledge financial support from Fundação para a Ciência e a Tecnologia through projects PTDC/QUI-BIQ/113721/2009, PTDC/QEQ-COM

/1623/2012, UID/MULTI/00612/2013 and PEst-OE/EQB/LA0004/2011 and grant SFRH/BD/81017/2011.

Chapter 6

Selected extensions to CpHMD-L

Contents

6.1 Context

The last chapter comprises three different works that were only possible with some (or all) technical developments presented in the previous chapters. First,

we present the titration of a DMPA/DMPC bilayer, a natural extension of the CpHMD-L method to deal with phospholipids. With this extension, CpHMD-L is now able to handle phospholipids, such as DMPA, but also other molecules with three different charge states. Secondly, we focused on pK_a calculations of model peptides in the presence of a lipid bilayer. In this project, we investigated how the protonation of typical amino acid residues varies as they are inserted into the bilayer. We started another project regarding simulations of pHLIP peptides. Each peptide of this family inserts in a lipid bilayer as a α-helix at low pH values, while remaining unstructured and adsorbed at neutral pH. In analogy to the work developed in chapter 3, we used a linear response approximation to estimate the pK_a value of some of the peptide key residues in the insertion process. At the end of this chapter, we hope to convince the reader about the usefulness and broad applicability of the developed methodologies.

6.2 Constant-pH MD simulations of DMPA/DMPC lipid bilayers

Hugo A. F. Santos, Diogo Vila-Viçosa[1], Vitor H. Teixeira, António M. Baptista and Miguel Machuqueiro

6.2.1 Introductory remarks

In chapter 4 and section 1.3.2, we have mentioned the biological relevance of DMPA, namely as a pH biosensor [72, 73]. When, under nutrient depletion, cytosol pH decreases, the ionization state of PA is altered reducing the electrostatic interactions with Opi1. This transcription factor is then released to the cytosol and inhibits membrane biogenesis [227, 228, 255]. Hence, it is expected that, in a DMPA/DMPC mixture with molar fractions close to physiological, the titration curve of DMPA has a transition near physiological pH.

Since our method deals with ionic strength [2], we were able to model an ionization dependent isothermal phase transition of a 25% DMPA/DMPC bilayer, in agreement with experimental results [15]. The application of the CpHMD-L method to simulate this membrane poses a new technical challenge, since PA headgroup have two pK^{mod} values (Section 2.2.2), i.e. can have 0, 1 or 2 protons. The goal of this project was to perform a complete pH titration of a 25% DMPA/DMPC bilayer using our CpHMD-L methodology, as it will provide valuable information

[1] In this work, I performed some initial test simulations, all charge parametrizations and some final analysis. I was also part of all discussions and of manuscript preparation (this manuscript has been submitted).

regarding the degree of ionization at physiological pH. This complete titration may also be a validation of the CpHMD-L approach, in the case we also observe the isothermal phase transition. As far as we know, this study constitutes the first CpHMD approach where all phospholipids are allowed to titrate, bringing a new level of realism to MD simulations of such biophysical systems.

6.2.2 Methods

All computational details are analogous to the ones presented in chapter 5. However, the specific settings for this work are described in appendix B.

System setup

The starting structures were obtained from the simulations presented in chapter 4. The lipid bilayer is composed of 128 lipid molecules (32 DMPA and 96 DMPC or 128 DMPC molecules) with different amounts of water molecules and ions. The hydration of the bilayer depends on the average area per lipid (A_l), to assure the existence of a small bulk-like water region not seen by any lipid. This region is kept relatively small, avoiding significant trapping of ions [2, 3]. The 100 % DMPC system was simulated at pH values of 0.0, 1.0 and 2.0 and the 25 % DMPA/DMPC system was simulated at pH values between 0.0 and 9.0 (Table 6.1). In this system, we did an extra simulation at pH 1.5, in order to improve the sampling in this region. Relatively long equilibration periods (20-50 ns) were done in CpHMD-L simulations, followed by 70-90 ns of production runs.

6.2.3 Results and discussion

pH titration

From CpHMD-L simulations, we were able to perform complete pH titration experiments. These titration curves were computed by averaging the occupancy states of all titrable sites over the final equilibrated segments at each pH value. Total membrane charge fluctuates along the simulation (Figure 6.1a). However, these curves fluctuate around an average value without a defined trend, confirming that the simulations are equilibrated. From the average values of these equilibrate curves, we obtained the total titration curves (Figure 6.1b). It is clear that, at pH $>$ 4, PC is completely deprotonated and only PA is sensitive to pH. Between pH 6 and 8, the titration curve appears slightly more abrupt, corresponding to the transition from $DMPA^{-1}$ to $DMPA^{-2}$. Remarkably, our CpHMD-L method is able to correlate this abruptness with large fluctuations in membrane charge (see pH = 7 for example in Figure 6.1a), a phenomenon already observed in an oleic

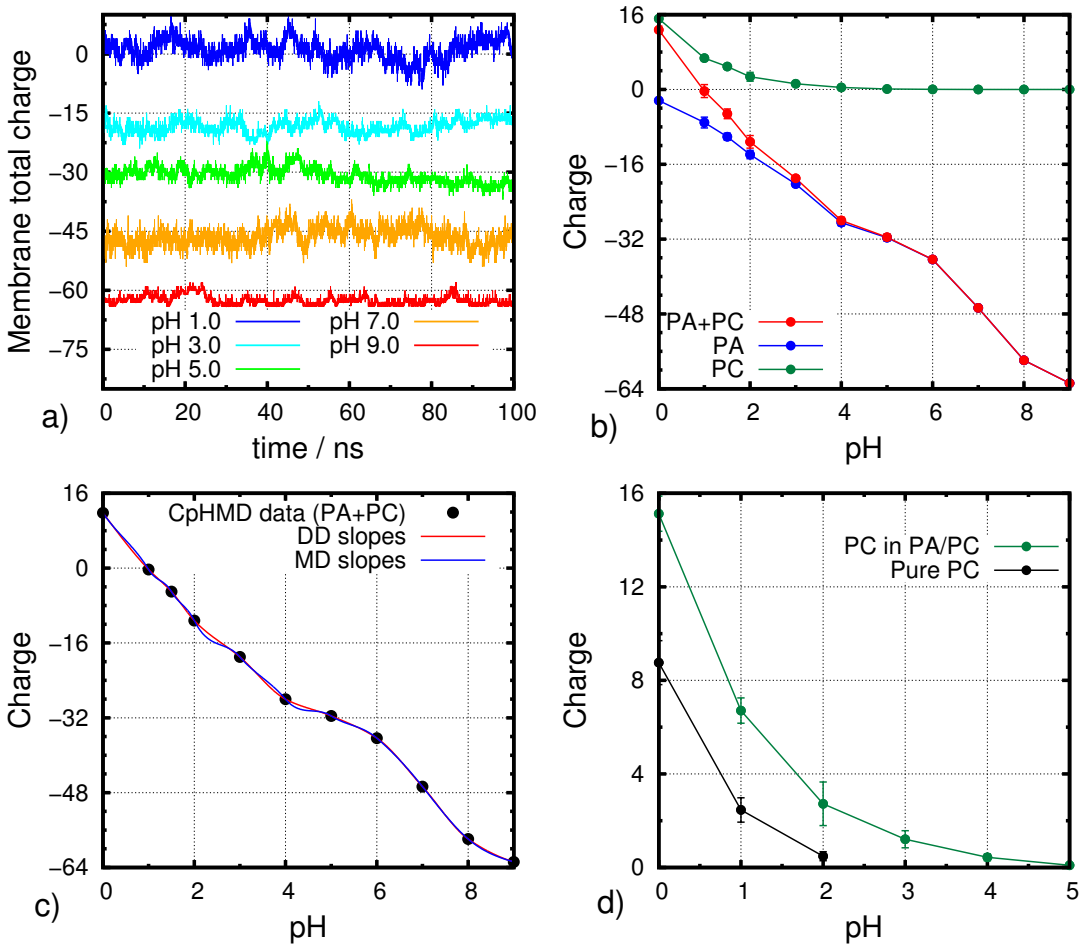


Figure 6.1: pH titration of the 25% DMPA/DMPC lipid bilayer. Membrane charge versus simulation time for one representative replicate per pH (**a**). Simulated titration curves for the whole system and the individual contributions from PA and PC (**b**). Simulated titration curves where the values are connected using cubic splines whose slopes are estimated using either the Davis-Dowden method (DD), or the relation $\partial\bar{Z}/\partial pH = -ln(10)\sigma^2$ (MM) (see chapter 5) (**c**). Comparison between the PC titration curves obtained from pure DMPC and from the 25% DMPA/DMPC mixture (**d**).

Table 6.1: Computed amount of Na^+ and Cl^- ions for the simulated systems at different pH values. "Predicted" refers to the amount of ions estimated by Poisson–Boltzmann that should be visible to the membrane, "Actual" refers to the number of ions added to the simulation, and "Evaluation" is the average number of ions explicitly seen by the membrane in the CpHMD-L simulations.

System	**pH**	**Charge**	**Predicted**		**Actual**		**Evaluation**	
			Na^+	**Cl^-**	**Na^+**	**Cl^-**	**Na^+**	**Cl^-**
PC	0.0	$+8.73 \pm 1.15$	5.27	8.83	5	8	4.88	7.45
	1.0	$+2.53 \pm 0.46$	5.42	8.24	6	8	5.91	7.31
	2.0	$+0.42 \pm 0.10$	5.50	8.11	6	8	5.94	7.22
PA/PC	0.0	$+12.75 \pm 0.82$	4.83	7.84	5	7	4.93	6.75
	1.0	-0.35 ± 1.38	5.05	7.37	5	7	4.96	6.69
	1.5	-5.21 ± 1.04	5.24	7.19	6	7	5.94	6.17
	2.0	-11.21 ± 1.35	5.87	7.10	6	7	6.00	6.50
	3.0	-19.00 ± 0.55	6.22	7.14	7	7	6.99	6.17
	4.0	-28.02 ± 0.69	7.18	6.77	8	7	7.99	6.11
	5.0	-31.60 ± 0.66	7.24	6.85	8	7	8.00	6.16
	6.0	-36.34 ± 0.46	7.80	6.89	8	7	8.00	6.33
	7.0	-46.71 ± 0.57	9.56	6.45	10	6	10.00	4.78
	8.0	-57.91 ± 0.48	11.29	6.39	12	6	12.00	4.45
	9.0	-62.79 ± 0.31	11.79	6.43	12	6	12.00	4.63

acid bilayer [3]. This is an important test of thermodynamic consistency for the sampling of protonation states (Figure 6.1c). At more acidic pH values there is a significant contribution from DMPC to the overall total charge, resulting in positive values at pH < 1.0. In our previous study, we estimated the ionization of the 25% DMPA/DMPC system at pH 4.0 and 7.0 using PB/MC calculations over conformational ensembles extracted from MD simulations (chapter 4 and Ref. 2). At pH 7.0 we estimated the total charge to be ~ -44, which is in excellent agreement with the CpHMD-L value (~ -46). However, at pH 4.0 the PB/MC estimated value (~ -18) is significantly lower than the CpHMD-L value obtained in this work (~ -28), illustrating the dependence of this approach on the conformational sampling quality.

Regarding DMPC, we compared the titration curve of this molecule in a pure

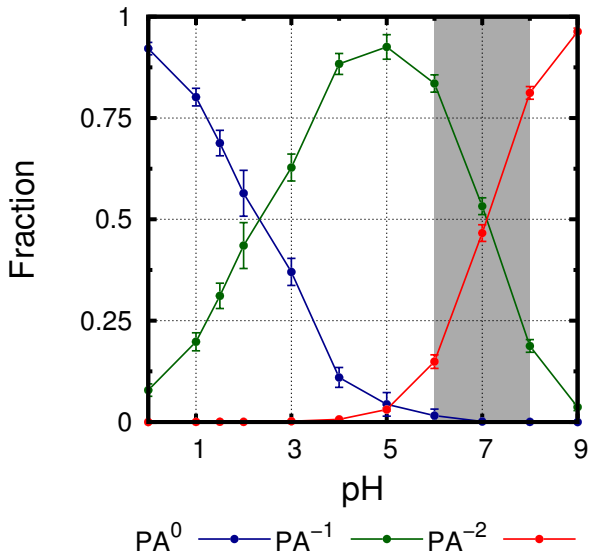


Figure 6.2: pH dependence of the fraction of the three predominant protonation states in DMPA. The physiological pH region was highlighted in gray.

bilayer and in the DMPA/DMPC mixture (Figure 6.1d). The titration curve of pure DMPC indicates that the protonation of the first lipid in an infinite bilayer of zwitterionic DMPC only occurs at pH ∼2.0, which is in agreement with our previous estimation using a linear response approximation method (∼2.2–2.6) (chapter 3 and Ref. 1). The presence of a number of negatively charged DMPA lipids induces a shift in the PC titration curve towards higher pH values, due to the electrostatic repulsion between ionized phosphates, stabilizing the PC^{+1} species with a neutral phosphate group.

Using the DMPA/DMPC CpHMD-L simulations it is possible to know the relative abundances of the three DMPA protonation states ($DMPA^0$, $DMPA^{1-}$, and $DMPA^{2-}$) over the studied pH range (Figure 6.2). From these results, we observe a coexistence region for the three charge states in the pH range $4-6$. Also, there are two main transition regions around pH 2.5 (from $DMPA^0$ to $DMPA^{-1}$) and pH 7.0 (from $DMPA^{-1}$ to $DMPA^{-2}$), consistent with the total titration curve (Figure 6.1c). Interestingly, the transition at physiological pH (Figure 6.2) indicates that any slight changes in the solution pH will have a major impact on the charge densities of membranes containing significant amounts of this lipid. This implicates PA as a key player in the regulation of many cellular events, as suggested by several authors [72, 73, 227, 228].

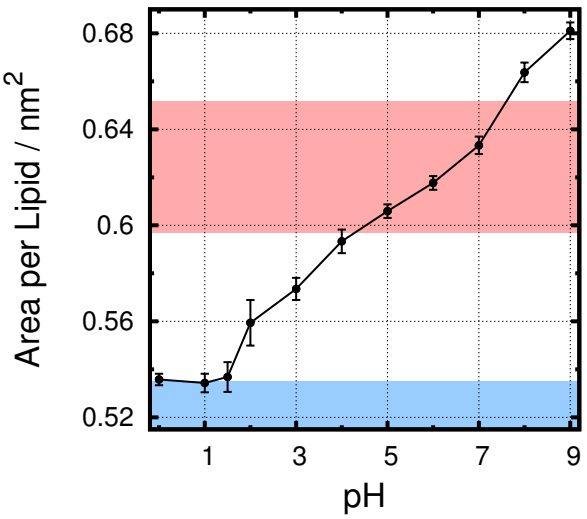


Figure 6.3: A_l variation with pH value for the 25% DMPA/DMPC lipid bilayer. Typical experimental values ranges for gel and liquid crystalline phases of DMPC are marked in blue and red, respectively [25–34].

Structural properties of DMPA/DMPC lipid bilayer

From the previous section we have seen that the protonation behavior of DMPA /DMPC bilayer is being correctly sampled. To further validate our method for lipid bilayer simulations, we determined the area per lipid (A_l) of our membrane as a function of pH (Figure 6.3). The A_l varies from ∼53 Å^2, at low pH values (almost in the typical gel phase region of DMPC), to ∼68 Å^2 at pH 9.0, which is well beyond the fluid phase region for DMPC. These results are in agreement with our previously published data of A_l in several systems with different PA ionizations [2]. Nevertheless, at low pH values, we do not observe A_l values as low as the ones reported for the zero charge PA system (∼49 Å^2) [2]. This is probably because, in CpHMD-L simulations, DMPC was allowed to titrate originating several protonated molecules at low pH values, slightly favoring larger A_l. Garidel and coworkers have reported pseudo-binary phase diagrams for DMPA/DMPC mixture at several pH values [15], and showed that, at pH 7 and $T = 310$ K, the 25% DMPA/DMPC mixture is in a fluid phase. At the same temperature but at pH 4, the mixture is in a mixed gel+fluid phase state and a macroscopic phase separation is observed [15]. Therefore, our results are in agreement with the lipid phases experimentally observed. Beyond A_l, other structural properties such as membrane thickness, deuterium order parameters, and lateral lipid diffusion were determined for this system. Similarly to the observed for A_l, all other properties

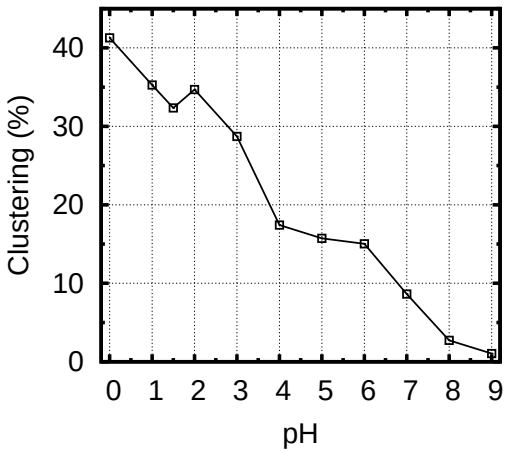


Figure 6.4: DMPA clustering tendency in the 25% DMPA/DMPC mixture measured by calculating the percentage of lipids involved in a cluster. A DMPA lipid is considered to be clustered if it is interacting, within a 5.5 Å cutoff, with at least another PA molecule.

were in accordance with our previous results [2] and / or in agreement with the expected trend.

Several authors have presented experimental data and Monte Carlo simulations suggesting that, in pseudo-binary lipid systems, there is clustering of like molecules and the formation of micro-domains of different compositions [15, 256]. In particular, the PA/PC lipid mixture has been proposed to undergo lipid immiscibility and even phase separation at pH 4.0 [15]. These phenomena can have an important role in the biological function of PA-rich membranes [72, 73, 227, 228]. The size of our 25% DMPA/DMPC lipid systems and the number of PA molecules do not allow the occurrence of clear micro-domains or phase separation. However, in our simulations, we can measure the DMPA aggregation tendency at different pH values by calculating the fraction of lipids involved in PA—PA interactions (Figure 6.4). Two DMPA molecules were considered to be interacting when their distance was below 5.5 Å, based on the distance histogram of DMPA at different pH values (Figure 6.5). With decreasing pH values, PA becomes less ionized and the hydrogen bonding tendency overcomes the electrostatic repulsion, allowing PA clustering. At pH 9.0, there is almost complete ionization and absence of hydrogen bond donors, which leads to disruption of PA aggregates. Figure 6.6 shows a membrane snapshot from the end of one simulation at pH 3.0. Here, we observed a clear DMPA cluster where half of the available PA molecules aggregate (left panel of Figure 6.6). The cluster is stabilized by hydrogen bonding and decreased

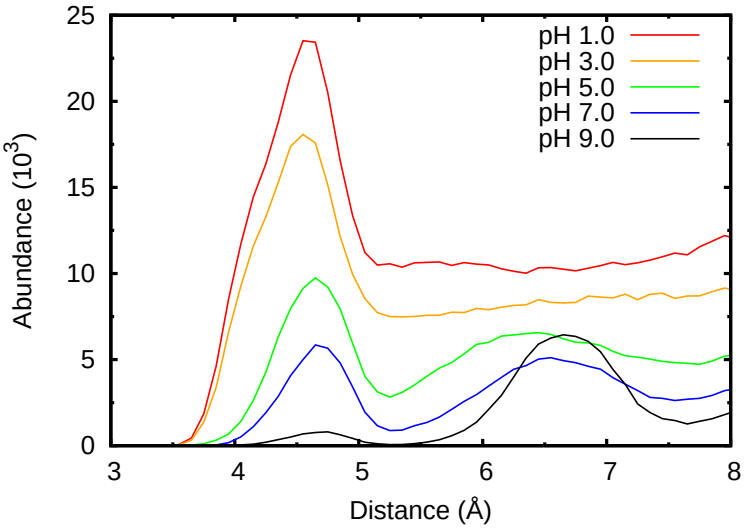


Figure 6.5: Distance histogram of DMPA in the 25% DMPA/DMPC mixture at different pH values. From these curves we chose 5.5 Å as a cutoff to identify PA aggregation.

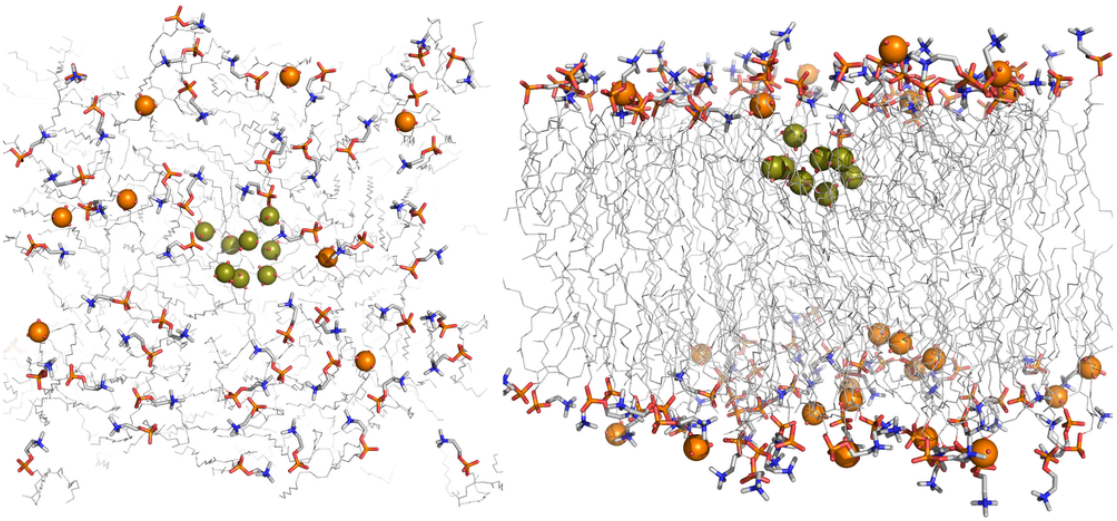


Figure 6.6: Graphical illustration of DMPA aggregation in the 25% DMPA/DMPC mixture at pH 3.0. The top view (left) helps to identify the cluster, and in the side view (right), we can see its preferred location along the membrane normal. The P atoms from DMPA are shown in spheres with different colors, whether they belong to a cluster (olive) or not (orange). DMPC is shown only as sticks.

electrostatic repulsions. The $\sim$ 8 lipid molecules comprising the cluster show an average ionization of -0.24, while the remaining molecules have -0.89. The average of all simulations at pH 3.0 is -0.62. The low ionization of the PA cluster can also help to rationalize its preferred deep inserted location (right panel of Figure 6.6). This internalization is favored by the lower need to solvate the cluster and by hydrogen bonding of DMPA with the ester groups of DMPC, similar to what we have previously observed with a protonated DMPC molecule [1].

6.2.4 Conclusions

This work presents a new extension of the CpHMD-L to study the pH effects on the DMPA/DMPC phospholipid mixture in a membrane model. As far as we know, this is the first CpHMD method used in continuous ("infinite") phospholipid bilayers, with titration of all lipids in the membrane. In our simulations, we observed a transition in the protonation state of DMPA at physiological pH indicating that any slight changes in the solution pH will have a major impact on the charge densities of these membranes, which could implicate PA as a key player in the regulation of many cellular events [72, 73, 227, 228].

Our study showed that the 25% DMPA/DMPC lipid bilayer system undergoes a pH-induced isothermal phase transition, in agreement to our previous work [2]. Also, at pH 4.0 and 7.0, the system is in the correct lipid phase, namely, gel/fluid mixture and completely fluid, respectively [15]. Very recently, a MM/MD study of DOPA in DPPC, where the authors just neutralized their systems with Na^+ ions, resulted in smaller A_l values for the double deprotonated PA^{-2} species compared with PA^{-1}, [257] supporting the idea that a different approach, like the one presented here, is essential to correctly model these systems.

From our simulations, we also investigated the effects of pH in the aggregation tendency of DMPA. Indeed, when lowering the pH values, DMPA became more prone to hydrogen bonding and less prone to electrostatic repulsion, leading to an increase in clustering. Even though our simulation system is not large enough to observe immiscibility or phase separation phenomena, [15, 256] we were able to identify and characterize small PA clusters.

This extension of the CpHMD-L method allows the simulation of lipid bilayers with increased realism, since it is now possible to include anionic lipids which can receive more than one proton in a phospholipid membrane model.

6.3 pK_a values of titrable amino acids at water/membrane interface

Vitor H. Teixeira, Diogo Vila-Viçosa[2], Pedro B. P. S. Reis and Miguel Machuqueiro

6.3.1 Introductory remarks

pH is a crucial physicochemical property that affects most biomolecules. Changes in protonation equilibrium of susceptible sites will modify the electrostatic environment and, consequently, have an effect on the molecular structure, stability and catalysis [168, 169, 177, 182, 183]. The proton avidity of the titrable sites (i.e. their pK_a) is strongly influenced by the level of solvent exposure (buried or exposed) and by the electrostatic interactions with neighbor polar groups. As an example, when a protein unfolds, the pK_a values of several amino acids are significantly shifted [258, 259]. The pK_a values of the typical titrable amino acids can also be influenced by changes in the solvent mixture [131], or by the insertion into a lipid bilayer [16, 218, 260–270]. The protonation behavior of peptides interacting with membranes has also attracted large interest in the context of cell-penetrating peptides [271] and pH-dependent membrane insertion peptides [16, 260–262]. In fact, it was recently showed that a faster membrane permeation is obtained with a positive tryptophan, challenging the electrostatics-based rule, suggesting that the neutral form should be the most stable [272]. Furthermore, to understand the protonation behavior of pH sensitive peptides, such as kyotorphin, close to lipid bilayers it is mandatory to take into consideration their protonation / conformation coupling [218, 266].

Alanine-based pentapeptides are simple, but useful molecules in the field of computational biophysics. These pentapeptides (Ala_2– X –Ala_2, where X is a pH titrating residue) were initially designed in Pace's group [273, 274] to study the effect of the protein environment on the protonation behavior of titrable residues. The experimentally measured pK_a values of these peptides were already used to calibrate the pK_a values of the model compounds needed in PB calculations of CpHMD simulations [10, 129, 131]. Also, the relatively small size of the pentapeptides renders them ideal systems to the study of how lipid membranes can influence pK_a values of pH susceptible amino acid residues.

The main objective of this work is to perform a comprehensive study of how a membrane environment can shift the pK_a values of the common pH-sensitive amino acids (Asp, Glu, His, Lys, Cys, Tyr, and the N- and C-termini). With these systems, we are able to further extend our CpHMD-L in order to study the

[2]In this work, I performed some of the conformational analysis. I was also part of all the discussions of the results and of manuscript preparation (this manuscript is submitted).

protonation behavior of peptides in membrane environments. For this, we used a DMPC membrane and the model Ala-based pentapeptides that have already been well characterized in water by Pace and coworkers [273, 274]. We aim to capture the coupling between conformation / configuration / insertion and protonation at the membrane interface, and focus on the pK_a changes of the peptides along the membrane normal. We were also interested in identifying pH-dependent conformational patterns, lipid region preferences, and kinetic traps. As far as we know, this is the first systematic study aiming to estimate the pK_a values of amino acid residues when a peptide is inserting the membrane.

6.3.2 Methods

Most computational details used in this work are analogous to the ones presented in chapter 5. However, the specific settings for this work are described in appendix C.

System setup

The eight pentapeptides simulated in this work share the same backbone structure (Figure 6.7). The middle Ala residue is replaced by one titrable amino acid with the termini capped with an acetyl or amino group for N- or C-terminus, respectively. This resulted in six peptides, which we will refer as Asp, Cys, Glu, His, Tyr, Lys, and two other were simulated with five Ala residues, titrating (uncapping) either N-terminus (NTr) or C-Terminus (CTr) residues. We simulated the eight pentapeptides in water and in water/membrane interface. To perform water simulations we solvated the peptide with water molecules and simulated this system in a dodecahedral box. For membrane simulations, we built a tetragonal box with 128 DMPC molecules solvated with water molecules and the peptides were placed at approximately 10 Å away from the membrane (Figure 6.7). In all simulations, the protonation state of lipids was kept fixed. We performed several test simulations with CTr peptide, where the lipids were allowed to titrate but this only occurred scarcely at very low pH values. Therefore, we decided to exclude lipid titration form this work, which resulted in a very important increase in computational speed.

6.3.3 Results and discussion

Calibration of model compound pK_a values in water

The new CpHMD-L method uses the DelPhi package [205, 206] to solve the PB equation. However, the model compounds pK_a values previously reported [10, 131] were calibrated using MEAD [275]. Therefore, to take advantage of the new developed methodology to study protein–lipid interactions, we decided to reparametrize

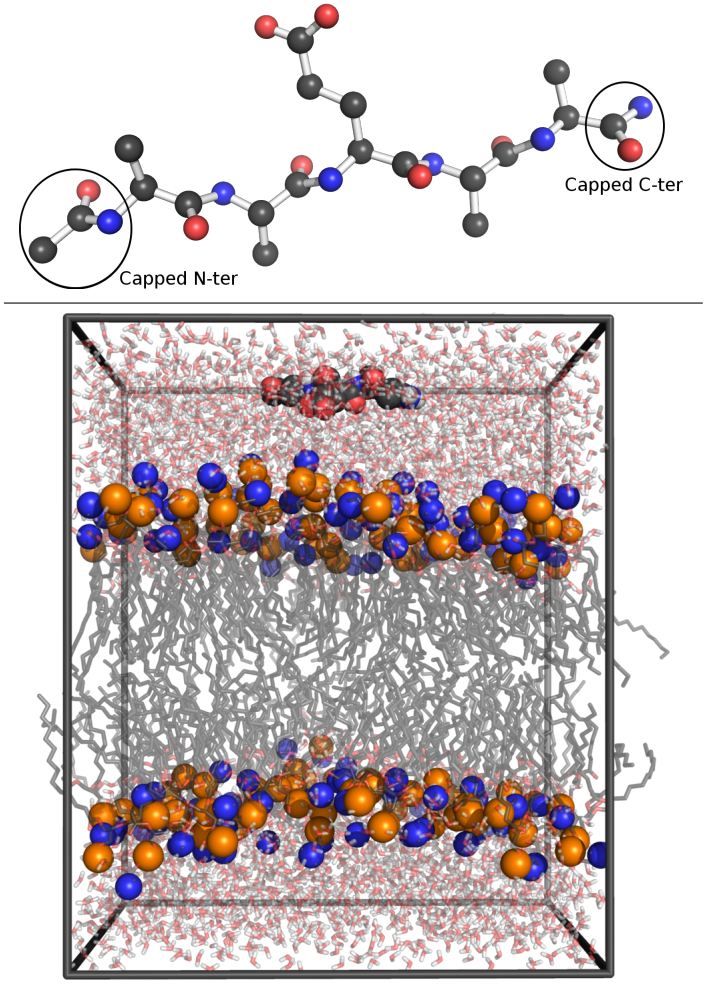


Figure 6.7: Ball and stick representation of the Glu pentapeptide (upper panel) and a starting structure for its simulation in water/membrane interface (bottom panel). Membrane and peptide hydrogen atoms were omitted for clarity.

Table 6.2: Model compound pK_a values in water.

Residue	**Exp**	pK^{mod} **(MEAD)**[a]	**DelPhi**		
			pK_a[b]	**Shift**[c]	pK^{mod}[d]
CTr	3.67	3.01	3.78	−0.11	2.90
Asp	3.94	3.73	3.88	+0.06	3.79
Glu	4.25	4.19	4.24	+0.01	4.20
His	6.54	6.88	6.68	−0.14	6.74
NTr	8.00	7.96	7.97	+0.03	7.99
Cys	8.55	8.58	8.46	+0.09	8.67
Tyr	9.84	9.59	9.84	+0.00	9.59
Lys	10.40	10.48	10.42	−0.02	10.46

[a] Values from Refs. 10, 131.
[b] Obtained from the CpHMD simulations in water.
[c] Calculated as Exp−pK_a.
[d] Calculated as pK^{mod}(Mead) + shift.

the pK_a values of the model compound of the pentapeptides in water studied by Pace and coworkers [273, 274] (Table 6.2). As expected, the differences between the two calibrations are small since the same theoretical model is being used in both approaches. The observed pK_a shifts may be associated with some inherent differences between DelPhi and MEAD. These can be related to numerical errors regarding differences in grid positions, surface algorithms, and convergence criteria. For this reason, in this work, we adopted the model compound pK_a values calibrated using DelPhi.

pH titration of the pentapeptides along membrane normal

In the CpHMD-L simulations the titrable groups are allowed to change their protonation states along the simulation. Therefore, it is possible to evaluate which protonation states are preferred at a given pH value, when the peptide is interacting or even inserting into the membrane. From simulations at three different pH values, and using a Henderson–Hasselbalch fit, we can obtain the pK_a values of each residue. In the membrane simulations, we can separate the system in slices along the membrane normal, where the pK_a values of titrable group can be fitted at different membrane insertion depths (Figure 6.8). In all cases, the pK_a values measured with the peptides away from the membranes (∼10 Å) are lower than their bulk references, which can be explained by the presence of a positive elec-

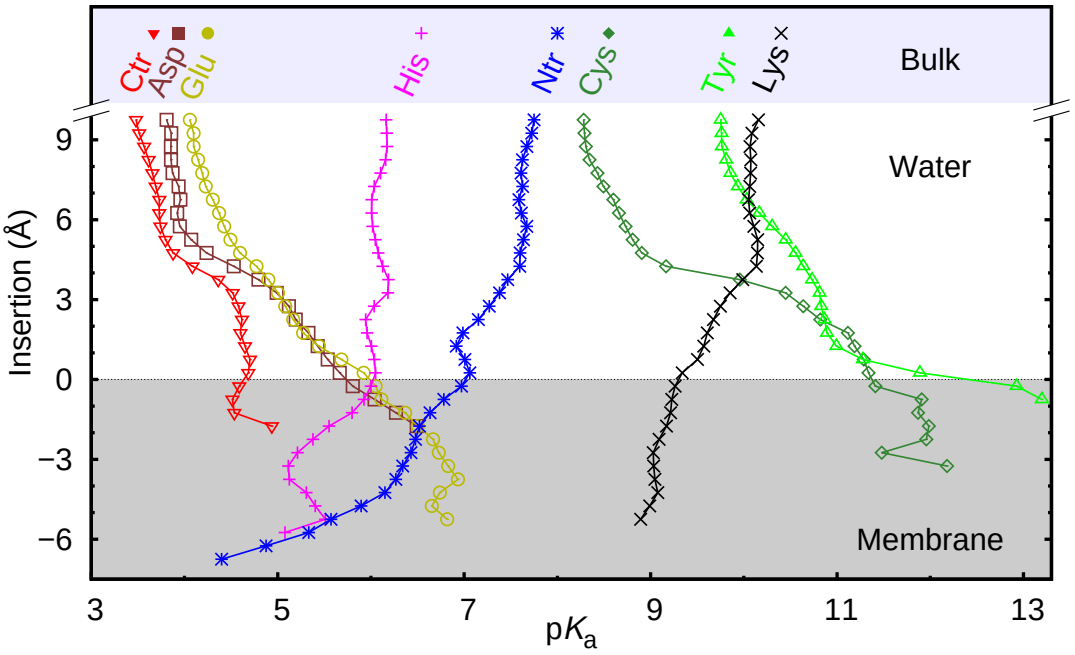


Figure 6.8: Residue pK_a values along the membrane normal. Negative insertion values correspond to deeper membrane insertions, while positive correspond to more shallow locations. The insertion values were measured between the titrable group and the phosphate group from the nearest lipid. The bulk pK_a values of the pentapeptides are shown on top for comparison.

trostatic potential generated by the closest choline groups. This effect induces a lowering of the pK_a values in both anionic and cationic residues.

The membrane insertion leads to an increasing desolvation of the titrable group, resulting in a stabilization of its neutral form. As a consequence, we observe a significant increase in the pK_a values of the anionic groups, while the cationic ones undergo an opposite variation. We were unable to measure the pK_a profile for all residues to the same insertion depth. However, this does not mean the pentapeptides are inserting unequally. This is due to a limitation in our pK_a calculation procedure, where we need enough proton exchange sampling to do the fit. Therefore, in some residues (CTr, Asp, Cys, and Tyr), the peptides are able to insert quite deep into the membrane but, because they do not ionize in those regions, we are not able to estimate their pK_a values. We would probably need to run these simulations at much higher (anionic groups) or lower (cationic groups) pH values and for longer simulation times.

The pK_a shifts of the anionic carboxylic acids range from $\sim$ 1.5 (CTr) to $\sim$ 3.0 pH units (Glu) upon membrane internalization. The large increase in their pK_a values is associated with desolvation, favoring the neutral forms, while the ionized carboxylates are also destabilized in the phosphate and ester group regions. These pK_a shifts have already been observed in pH-triggered fusion proteins, such as the hemagglutinins from Influenza virus [276, 277] and in pH-dependent membrane insertion peptides, like pHLIP [16, 260, 261]. In the case of pHLIP peptide, with two Asp residues in the transmembrane region, the authors report insertion pK values as high as 6.2 [262]. The other two anionic residues, Tyr and Cys, are even more sensitive to membrane insertion and their pK_a values are shifted by $\sim$3.5 pH units. The same arguments given for carboxylic acids are valid here and the slightly larger shifts are probably related to the fact that, in this case, when ionized, the negative charge is concentrated in only one atom. This creates a larger local dipole compared with the charge distribution present in the carboxylate group.

The NTr can be titrated in very deep positions of the membrane ($\sim$ 7 Å below the nearest phosphate group) and, consequently, in this case we measured a significant pK_a shift ($\sim$ 3.5 pH units). This was possible because, as a terminal group, it can reach far deeper in the membrane and its positive form may be stabilized by oxygen atoms from neighbor phosphate and ester groups. In the case of Lys, the protonable group is located at the center of the pentapeptide, where it probably has more conformational restrictions. This reduces the sampling in deeper regions of the membrane and, consequently, makes it difficult to obtain the large shift values observed for NTr. His showed a smaller pK_a shift ($\sim$ 1.0 pH units) probably because, upon ionization, the positive charge is distributed over the whole imidazole group, which becomes easier to stabilize by the neighboring groups. The small increase in the pK_a values around -5 Å insertion can be attributed to the previously mentioned cation stabilization in the region between the phosphate and ester groups.

The pentapeptide structure at the membrane / water interface

In order to investigate the conformational space of the pentapeptide interacting with the membrane, we followed the radius of gyration (R_g) at different insertion regions (Figure 6.9). In most cases, we observed large continuous energy basins outside and in a shallow region of the membrane. Also, in these regions, there is a large overlap between replicates (data not shown). Altogether, these results indicate that most pentapeptides are able to sample their conformational space without many restrictions or kinetic traps, however, some noteworthy cases occurred.

In Glu simulations at pH 4.0, an energy basin with low R_g and insertion slightly

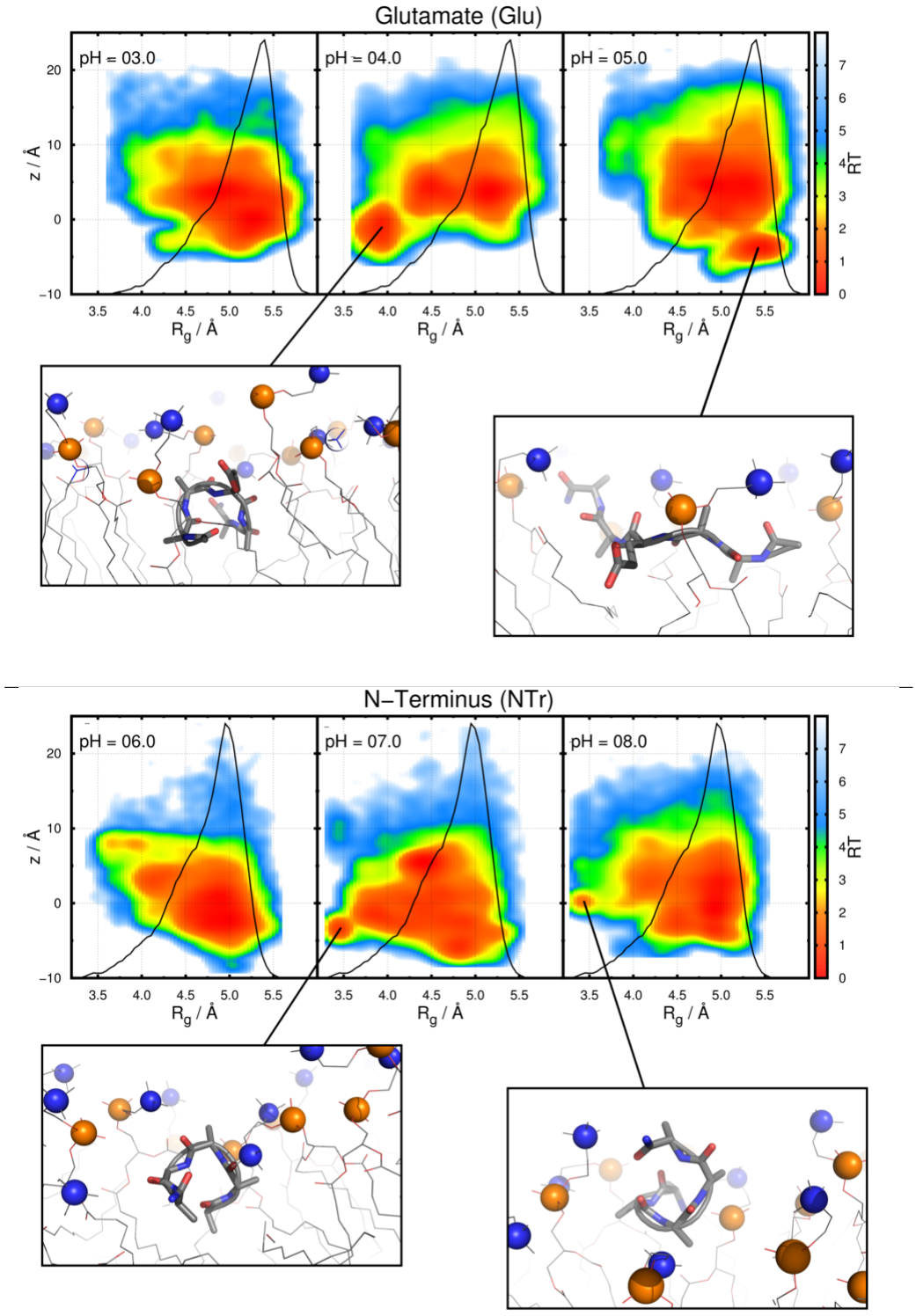


Figure 6.9: Energy landscapes of R_g for Glu (upper panel) and NTr (bottom panel) distributed along the membrane normal (z-axis). The black line refers to the R_g histogram of the pentapeptide in water simulations.

below 0 Å are observed. This basin corresponds to a residual helix structure, which can be stabilized in the membrane environment since it is a low dielectric region. At pH 5.0 there is also a separated basin associated with an extended structure (high R_g), which is partially inserted in the bilayer. In NTr simulations at pH 7.0 and 8.0 it is also observed a residual helix with a very small R_g at slightly inserted regions.

6.3.4 Conclusions

We applied our recently developed CpHMD-L implementation [3] of the CpHMD method [12, 13] to study the protonation of Ala-based pentapeptides incorporating all titrable amino acids at the water/membrane interface. The peptides approaching a DMPC membrane sense the positive charge from the choline groups, which will lower their pK_a values. Membrane insertion leads to desolvation and a clear stabilization of the neutral forms. Consequently, the anionic residues have their pK_a values increased, while the cationic ones are shifted in the opposite direction. Some residues are able to ionize deep in the membrane which allowed us to measure their pK_a values in those regions.

To obtain more complete pK_a profiles and determine more complex conformational behaviors of the studied pentapeptides, we would need more extensive sampling and pH values covered. Nevertheless, the data completely capture their pK_a trends and reveal the molecular bases of the interactions between the titrating groups and the lipid moiety. Furthermore, our results emphasize the need of taking in consideration the protonation equilibria in membrane simulations and that choosing a fixed protonation state can lead to biased and erroneous results.

6.4 pK_a values of key amino acids of pHLIP peptides in water / membrane interface using a linear response approximation

Diogo Vila-Viçosa[3] and Miguel Machuqueiro

6.4.1 Introductory remarks

The pH (low) insertion peptides (pHLIP) is a family of peptides that insert into lipid bilayers at low pH [16, 278–281]. The original version of this peptide was

[3]In this work, I performed all the simulations and analysis. I am also currently preparing a manuscript for submission.

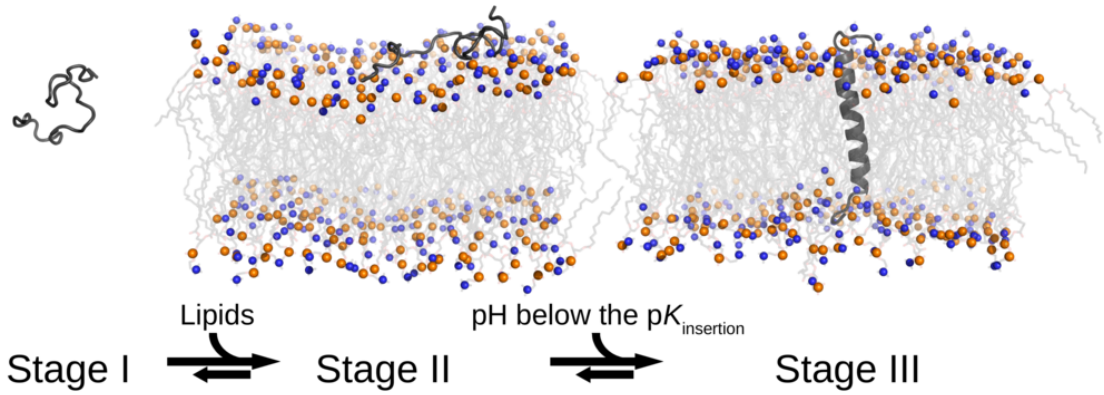


Figure 6.10: Three states of pHLIP interaction with a bilayer.

obtained from a transmembrane helix of bacteriorhodopsin [16]. In pHLIP biophysical studies in the presence of lipid bilayers three states have been identified (Figure 6.10): State I, where the peptide is free in solution and does not adopt a defined structure at any pH value; in the presence of liposomes at neutral pH, the peptide is still unstructured but it stays adsorbed in the bilayer surface (State II); finally, by lowering the pH of this last solution an increase in the Trp fluorescence and a circular dichroism signal characteristic of an α-helix are observed. The higher α-helical content together with the increase of Trp fluorescence strongly indicates that the peptide is now inserted in the bilayer (State III) [279]. Since acidosis is a universal marker for tumor identification, these three states behavior of pHLIP render it a promising molecule to target several drugs and imaging agents to cancer cells [262, 280]. The fine tune of the insertion properties of pHLIP is then essential to target different tissues (with different external pH values).

There is already a large number of pHLIP sequences sharing common features: a transmembrane (TM) and two flanking sequences, with hydrophobic residues and one or more acids (Asp or Glu). Both flanking sequences have several charged residues, which are essential for peptide solubility. In particular, the number of anionic groups in the C-terminus flanking sequence (the terminus that faces the membrane interior [279]) affects the rate of pHLIP insertion into the bilayer [281]. The original TM sequence (`WARYADWLFTTPLLLLDLALLV`) contain many hydrophobic residues (including two Trp responsible for the fluorescence signal) and two Asp residues at positions 14 and 25. These residues play an essential role in the pH-dependent insertion mechanism, since they must be protonated (at least partially) so that the insertion occurs [282]. Following the variation of fluorescence signal with pH, it is possible to measure a pK of insertion (p$K_{\text{insertion}}$), which is probably related with the pK_a of the two Asp acids in the membrane/water interface. In

Table 6.3: Simulated pHLIP sequences.

Variant	Sequence
wt	ACEQNPIYWARYADWLFTTPLLLLDLALLVDADEGT
D14Up	ACEQNPIYWARYDAWLFTTPLLLLDLALLVDADEGT
D14Down	ACEQNPIYWARYAWDLFTTPLLLLDLALLVDADEGT
D14E	ACEQNPIYWARYAEWLFTTPLLLLDLALLVDADEGT
L16H	ACEQNPIYWARYADWHFTTPLLLLDLALLVDADEGT

fact, the p$K_{\text{insertion}}$ of the original pHLIP is 5.96 [260], which seems to be a normal value for an Asp residue at the interface (see section 6.3). Moreover, pHLIPs with single mutations D14E or D25E (Asp− >Glu) show an increase in the p$K_{\text{insertion}}$ of ~ 0.5 pK units [260], which is the difference in the pK_{a} values of Glu and Asp in solution. This suggests that the pK_{a} values of these acids is strongly related with p$K_{\text{insertion}}$. The effect of position change of D14 to 13th and 15th positions which, in principle, changes the solvent exposure, was studied. This results in a decrease in the p$K_{\text{insertion}}$ when the residue is more exposed to the solvent (in the 13th position) and vice versa [261], suggesting a crucial role of D14 pK_{a} in the insertion mechanism.

Despite the large number of tested sequences and experimental studies regarding the mechanism and thermodynamics of pHLIP insertion [281], having the information with molecular detail would be vital for the rational design of new sequences. Here, we used a linear response approximation (see chapter 3 for details) under the assumption that D14 is a key residue for the p$K_{\text{insertion}}$ value. We used four test sequences (see below) and one with a His residue and calculated the variation of the pK_{a} value with z distance to P atoms of the acidic and His residues.

6.4.2 Sequence choice and computational methods

The computational methods are identical to those used in chapter 3. Here we will just point out the differences and provide an explanation for the sequence choice.

The four selected test sequences were mentioned in the introductory remarks (Table 6.3):

- a wild type variant (wt) [262]
- two mutations where the Asp is moved to different positions (D14Up and D14Down) [261]

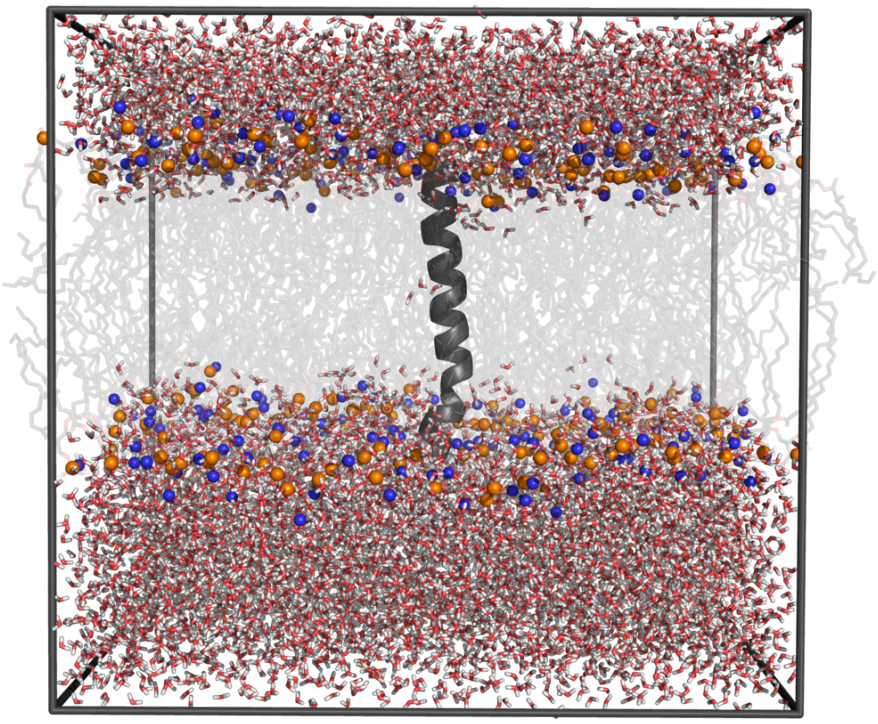


Figure 6.11: Starting structure for a pHLIP/DMPC MD simulation. Lipid tails are omitted for clarity.

- a D14E mutation [260]

In this simulations we worked under two related test hypothesis: the pK_a at water/membrane interface is related with p$K_{insertion}$; the pK_a of D14 is determining for p$K_{insertion}$. With these two hypothesis, we built a 256 lipid membrane with pHLIP peptides already inserted in the bilayer (Figure 6.11). This system was simulated with two protonation states per sequence (with the Asp/Glu acid protonated and deprotonated). Five simulations of 100 ns were performed for each protonation state. From this simulations, the pK_a was calculated using a linear response approximation, as explained in chapter 3. This approach was used instead of CpHMD-L, due to its simplicity and lower computational time. Moreover, since in LRA we force the system to sample high energy protonation states, it is much faster to obtain a proper sample (though the errors in the pK_a values are larger, mainly if this approximation is not valid[4]).

[4]In the linear response approximation it is assumed that the free energy of protonation varies

The Histidine mutation (L16H) is a new proposal from our work. At the moment, all pHLIP variants have one p$K_{\text{insertion}}$ below which the peptide is inserted in the bilayer. This means that it will target all tissues with a pH value below p$K_{\text{insertion}}$, usually tumor cells, but also kidneys. Our goal is to obtain a peptide with two p$K_{\text{insertion}}$ between which the peptide is inserted. This would mean that the peptide will only target tissues with a pH within these values. The rational behind the His mutation is that, if this residue has a pK_{a} below the Asp residue (as we observed for inserted regions in section 6.3), the His will be protonated at lower pH values, which will generate a positive charge, destabilizing the inserted conformation. The difference between the two pK_{a} values should also be large enough to generate a significant population of the peptide, where His and Asp are in their neutral form.

6.4.3 Results and discussion

The MD simulations were performed at constant protonation, which means that the system will be well adapted to the simulated protonation state. In this case, in the ionized states, the residues will be much more exposed to the solvent when compared to the neutral forms. This also means that the conformational space will be very different in the two ensembles, breaking the linear response approximation. To avoid this, we calculated the pK_{a} values as a function of the z distance to P atoms (Figure 6.12). The D14E and D14Down showed approximately the expected behavior (similar to acidic residues in Figure 6.8), i.e. the pK_{a} value increases with the insertion, which means that, as the residue became less exposed to the solvent, it is harder to ionize. On the other hand, in the wt and D14Up variants the pK_{a} decreases with the membrane insertion. This result is nonintuitive and should be further investigated. Nevertheless, this can be explained by an interaction with a positive residue such as Arg 14, whose strength increases with the insertion, lowering the pK_{a} value of the acidic residue.

Building on the hypothesis that the pK_{a} of the acidic residues is related with p$K_{\text{insertion}}$, we determined the pK_{a} value when the residue is in a deeper insertion (Table 6.4). In principle, this pK_{a} value is measured when the solvent exposure is almost zero and the residue stops exchanging protons with water. The obtained pK_{a} values are in strong agreement with experimental p$K_{\text{insertion}}$, which is a significant support to our hypothesis. Nevertheless, these results are preliminary and a detailed conformational analysis should be performed. The consistency of our data should also be checked, i.e. the pK_{a} values should be calculated with more

linearly between the protonated and the deprotonated states. In practice, this means that there should be a conformational overlap and the pK_{a} value calculated in the two states should not be different.

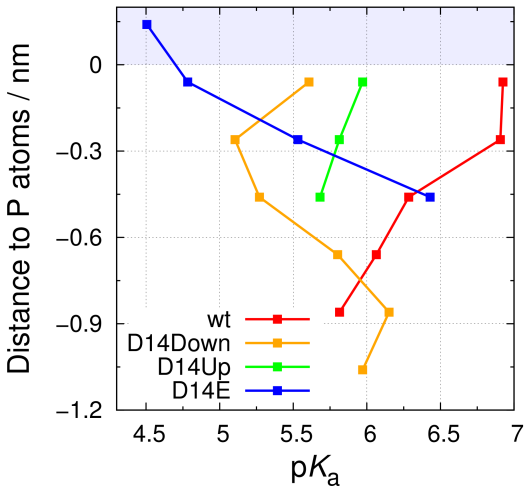


Figure 6.12: Residue pK_a values of pHLIP key residues in known sequences. Negative insertion values correspond to deeper membrane insertions, while positive correspond to more shallow locations. The insertion values were measured between the titrable group and the average position of phosphate groups.

Table 6.4: Comparison between experimental p$K_{\text{insertion}}$ and pK_a values obtained from LRA.

Variant	p$K_{\text{insertion}}$	pK_a (LRA)
wt	5.96	5.81
D14Up	5.65	5.68
D14Down	6.12	5.97
D14E	6.46	6.43

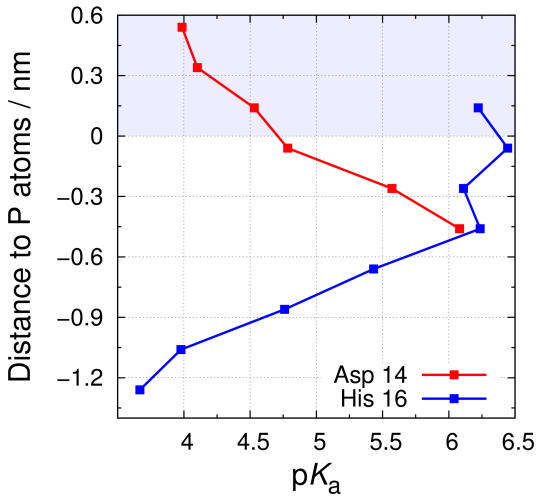


Figure 6.13: Residue pK_a values of pHLIP key residues in L16H variant. Negative insertion values correspond to deeper membrane insertions, while positive correspond to more shallow locations. The insertion values were measured between the titrable group and the average position of phosphate groups.

replicates, since the conformational space of the peptide inside the membrane is sampled very slowly and the results may be affected by the presence of kinetic traps.

Despite the possible sampling problems, we decided to run some initial simulations with the L16H variant and calculate the corresponding pK_a values (Figure 6.13). The obtained profiles are similar to the ones in Figure 6.8. The observed behavior is in agreement with the expected: the pK_a of the His residue decreases with the insertion since the membrane stabilizes the neutral (deprotonated) form, and the pK_a of the Asp residue increases since, in this case, the membrane stabilizes the protonated form. If we now look at the pK_a values in more inserted regions, we see that the Asp residue has a pK_a above 6 and the His has a pK_a below 4. According to the original idea, this would mean that the peptide can have two p$K_{\text{insertion}}$ since below pH 4 the TM region has a positive charge and, above pH 6 became negative, being neutral only between these two values. These results can now trigger new His sequences (or even Lys sequences) in order to fine tune the value of this lower p$K_{\text{insertion}}$.

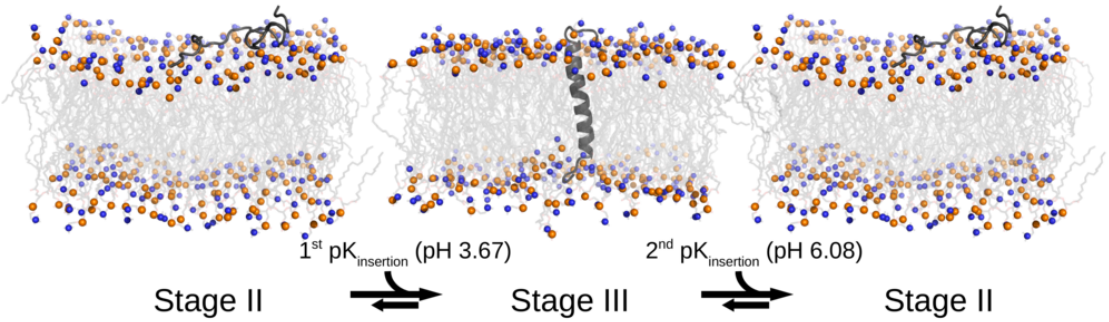


Figure 6.14: Proposal of a new three state model for pHLIP interaction with a bilayer.

6.4.4 Final remarks

In this section, we worked under the hypothesis that the p$K_{\text{insertion}}$ values of pHLIP peptides are related with the pK_a of some key residues. We simulated four known variants and obtained a strong correlation between these two values. Moreover, we proposed a new pHLIP sequence, with the possibility of having two p$K_{\text{insertion}}$ values (Figure 6.14). This would mean that the peptide is only active within these two values, which would increase the therapeutic potential of these molecules. This new variant can now be experimentally tested as a proof of concept. If it works, our computational studies could advance in order to fine tune the pH interval in which the peptide is active. In addition, the CpHMD-L should now be applied to this system since, unlike LRA, this method can capture the coupling between protonation and conformation, increasing the accuracy of the pK_a calculations. Moreover, it will also be important to test whether the protonation of lipids molecules play and important role. Although this work was not performed with the CpHMD-L methodology, it illustrates the importance of considering different protonation states in MD simulations of pH-sensitive systems.

Chapter 7

Conclusions and future perspectives

The main goal of this thesis was to develop a constant-pH molecular dynamics method that allows the simulation of lipid bilayers, i.e., a system with periodic boundary conditions in two directions of a tetragonal box. This single step turns out to be a three stage process which implied the development of: i) a new approach to deal with periodic boundary conditions in Poisson–Boltzmann calculations; ii) a new method to deal with ionic strength in biomolecular simulations; iii) the final development of the CpHMD-L method itself. Hence, we think that our goals were exceeded, since we ended up with three new technical / theoretical advances.

The new software (now using DelPhi, an available Poisson–Boltzmann solver) allows to run protonation free energy calculations with periodic boundary conditions in the finite difference grid used to solve the Poisson–Boltzmann equation. This allowed to estimate for the first time at which pH value a DMPC molecule starts to protonate in a bilayer environment and to perform all other pK_a calculations for this thesis. In the particular case of pHLIP peptides presented at section 6.4, the pK_a calculations enabled us not only to gain insight into the role of some specific amino acids for the insertion process, but also to propose a new sequence that drastically increases the potential applications of this technology.

The treatment of ionic strength in charged membrane simulations is a difficult issue, which is usually solved simply by adding enough counterions to neutralize the system. As shown in chapter 4, this may not be the best approach since it represents a tremendous local ion concentration that far exceeds the value expected from Poisson–Boltzmann or Gouy–Chapman equations. Our method represents a new approach to overcome this problem in simulations of systems with a significant net charge. This has a crucial relevance in charged membranes, such as the ones presented in chapters 4, 5 and section 6.2. Moreover, in full titrations of large proteins (Ref. 125 for example) the accuracy in the calculations of pK_a values could

improve with a careful treatment of the ionic strength, particularly in extreme pH values where the net charge of the protein is higher.

The application of CpHMD-L to these difficult systems illustrates the importance of taking into account the protonation / conformation coupling in molecular dynamics simulations, in particular when looking at pH dependent phenomena. For the best of our knowledge, this is the only available method that can deal simultaneously with pH considering the protonation / conformation coupling, periodic boundary conditions in protonation free energy calculations, and a careful treatment of ionic strength.

Finally, there are a few issues about our method that should be mentioned, namely: the treatment of long range electrostatics, and the high frequency of pK_a calculations needed to assure a fast sampling of protonation states. In the first case, the problem lies on the two most used methods to treat electrostatics. Ewald summations are good for neutral and periodic systems. However, biological systems at a very microscopic scale are rarely neutral or periodic in three directions (for example, membranes can be approximated as periodic systems in two directions but certainly not in the third). On the other hand, reaction field methods assume homogeneity beyond the cutoff value, which is far from the physical reality in a system like a membrane or a large protein. A possible solution to overcome this problem can be the use of much larger cutoff values, at which the electrostatic interactions are almost zero. This will only be achieved with faster algorithms and computers. Interestingly, this is also the solution for our second issue: protonation sampling. This problem will only be solved with longer simulations (orders of magnitudes larger than the ones presented in this work). Nevertheless, in principle, this problem only affects the kinetics of our sampling. This means that the CpHMD-L method can be seen as a sampling method such as Monte Carlo, where the simulation time does not have a strong relation with physical time. Another computational bottleneck of the current implementation derived from the self-consistent method performed to estimate the number of ions. This procedure is time consuming and difficult to automatize. Currently, each step in the self-consistent procedure is manually performed by the user, including the choice (based on the Poisson–Boltzmann calculations) of the number of ions and water molecules in the next MM / MD simulation.

7.1 Future perspectives

The CpHMD-L methodology developed over the last four years allows the simulation of any system with periodic boundary conditions in x and y directions of a tetragonal box and at a defined pH value and ionic strength. This represents a significant improvement in MM / MD models of biological membranes. It is

now possible to move a step forward in the direction of biological membranes since we are able to simulate a lipid mixture with several different lipids. With larger systems it may be possible to understand at an atomic level the pH dependent formation of ordered lipid micro-domains as suggested in section 6.2. The developed methods can also be applied to other systems with titrable groups or a significant net charge, such as the fusion of charged micelles with membranes.

In sections 6.3 and 6.4 we showed two preliminary studies of peptide–membrane interactions. This represents a wide range of possible applications to membrane proteins, like the study of proteins involved in electron transport chains, where the protons play a crucial role being the driving force for ATP synthesis. The molecular detail of cell penetrating peptides (often highly positive) interaction with biological membranes (often negatively charged) can also be addressed. The particular case of pHLIP peptides is already being studied in our group, the next logical step being the study of more variants and, more important, the application of CpHMD-L to these systems. This last step poses a new challenge regarding the protonation sampling since, after insertion, the titrable residues will probably became neutral and the conformation will be kinetically trapped. A possible way to overcome this problem is to use enhanced sampling techniques, such as umbrella sampling that forces the peptide to sample conformations with low probabilities. With this new extension it would be also possible to determine the pH dependent free energy of pHLIP peptide insertion, which should be much more lower at low pH values.

Appendix A

Supporting information for "Constant-pH MD simulations of an oleic acid bilayer" (Chapter 5)

Calibration of the τ_{prt} parameter

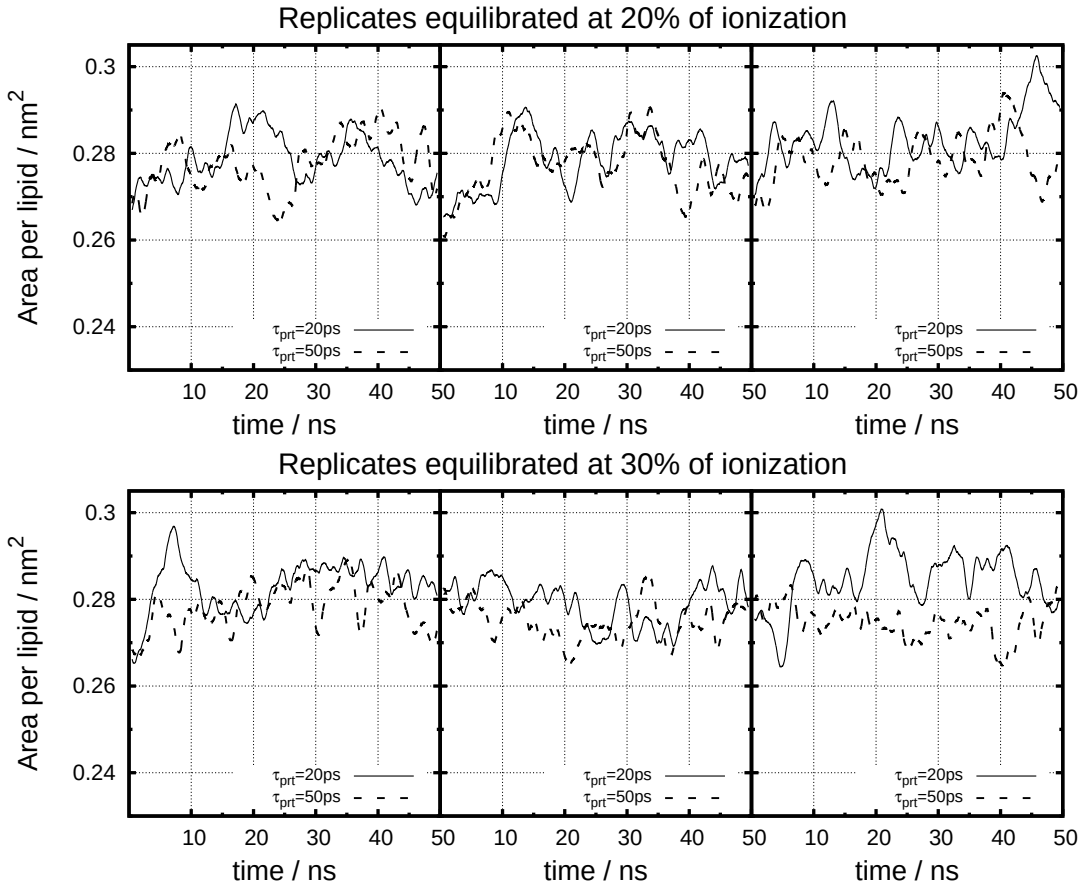
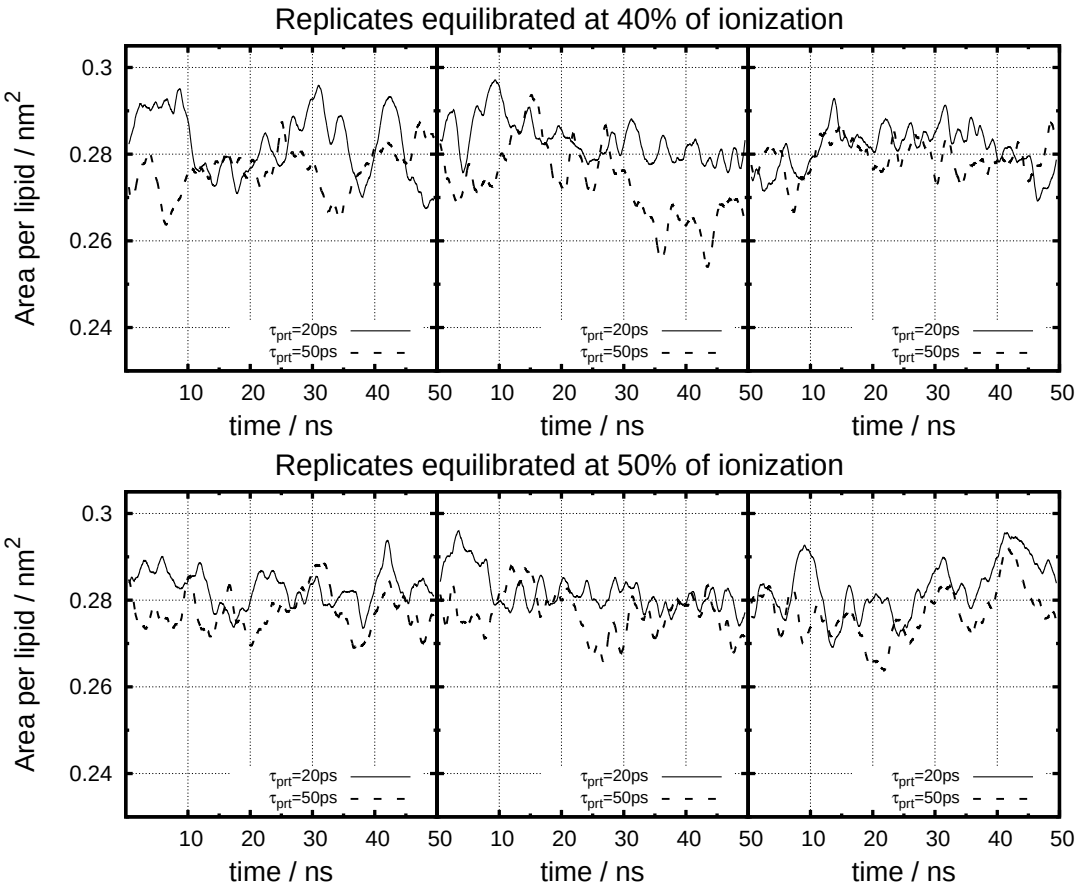


Figure A.1: Time variation of area per lipid in simulations to calibrate the τ_{prt} parameter.


Replicates equilibrated at 40% of ionization
Area per lipid / nm2
0.3
0.28
0.26
0.24
τprt=20ps
τprt=50ps
10
20
30
40
50
time / ns
Replicates equilibrated at 50% of ionization

(cont.)

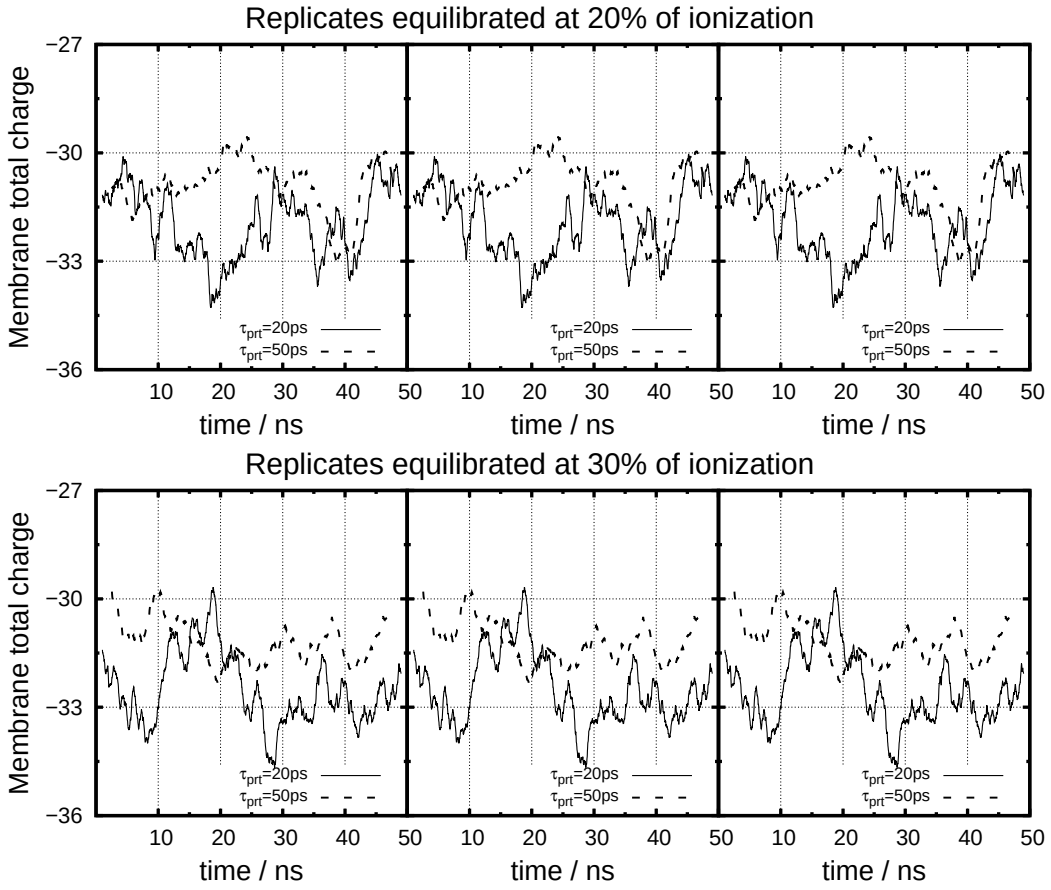
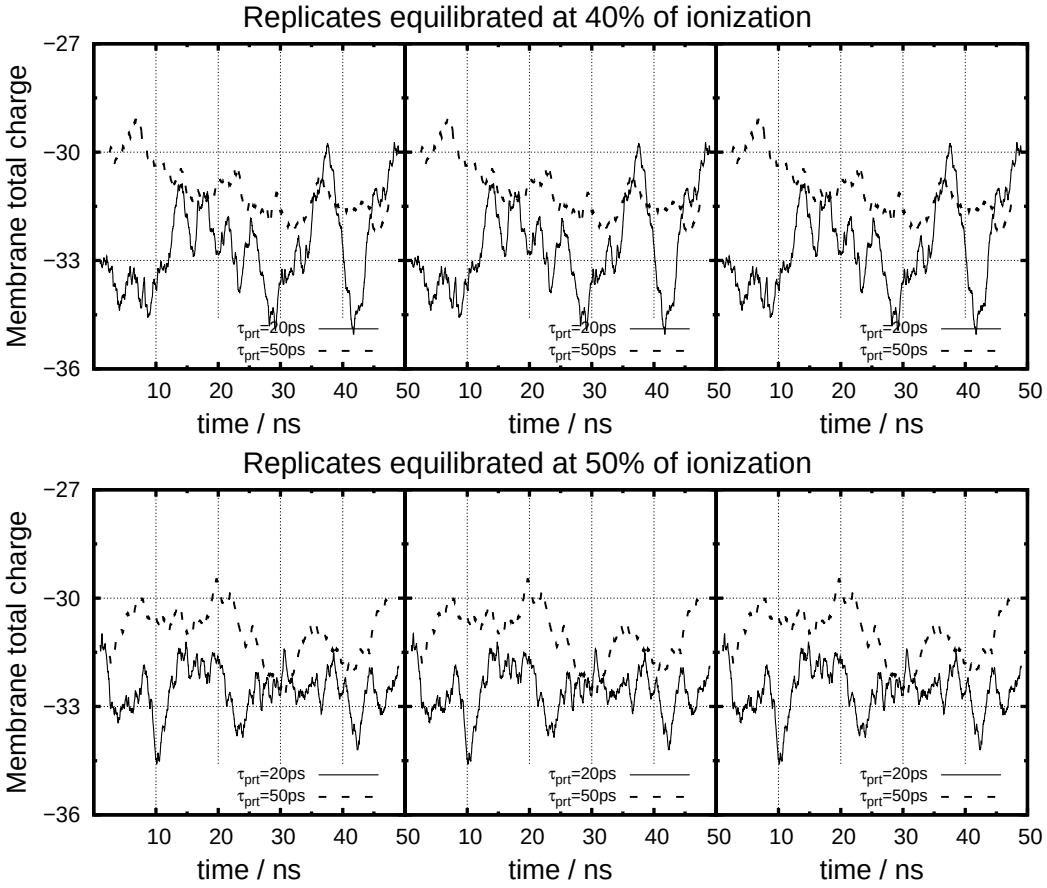


Figure A.2: Time variation of membrane charge in simulations to calibrate the τ_{prt} parameter.


Replicates equilibrated at 40% of ionization
Membrane total charge
−27
−30
−33
−36
10
20
30
40
50
time / ns
τprt=20ps
τprt=50ps
Replicates equilibrated at 50% of ionization
Membrane total charge
−27
−30
−33
−36
10
20
30
40
50
time / ns
τprt=20ps
τprt=50ps

(cont.).

Estimation of the number of ions and system equilibration

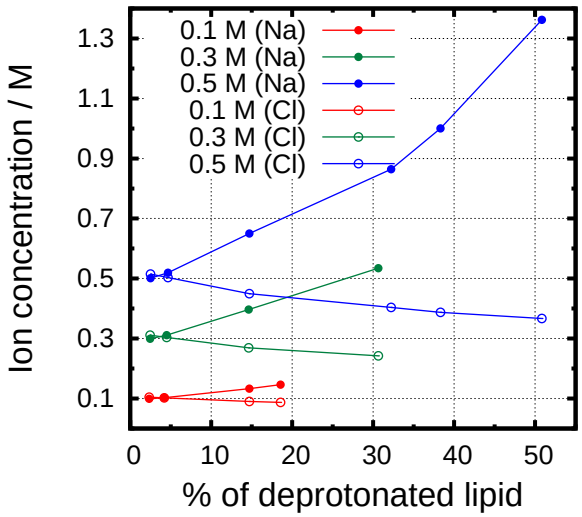


Figure A.3: Concentration of counterions and coions in all CpHMD-L simulations at different membrane ionization values.

PB-estimated electrostatic potential map

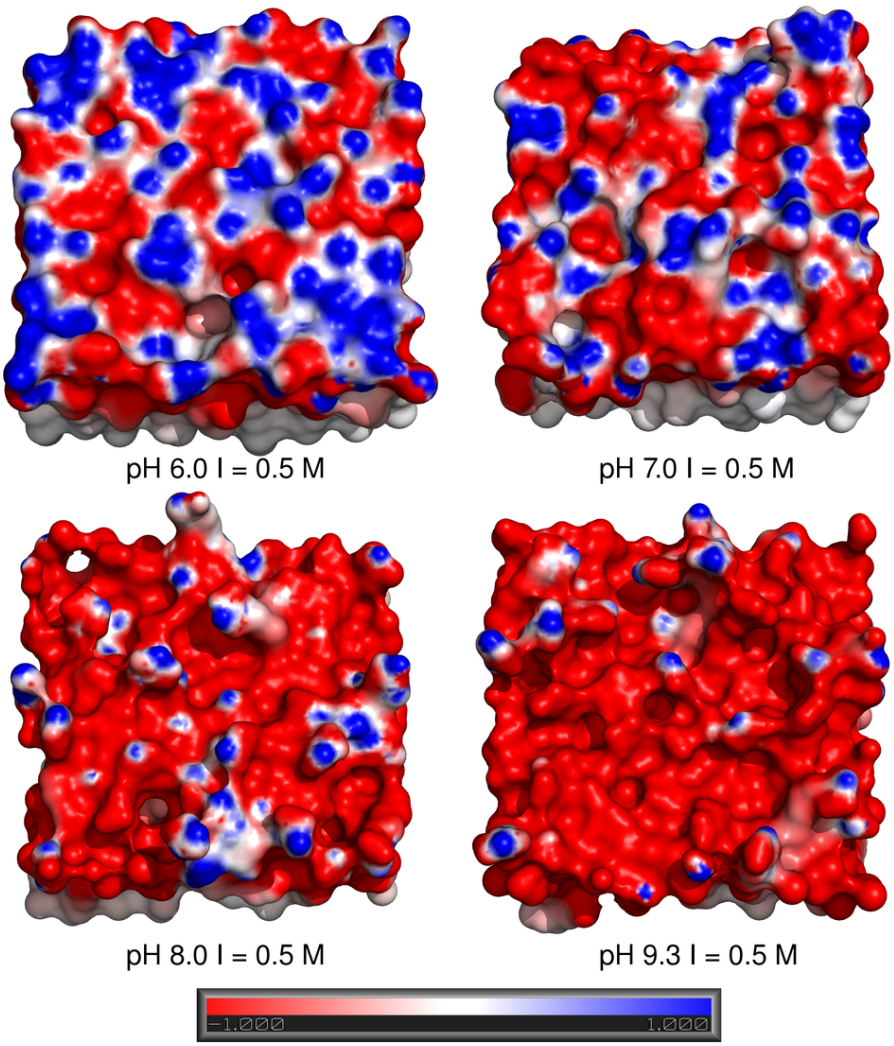


Figure A.4: Representation of the PB-estimated electrostatic potential surface.

Deuterium order parameters of the oleic acid aliphatic tail

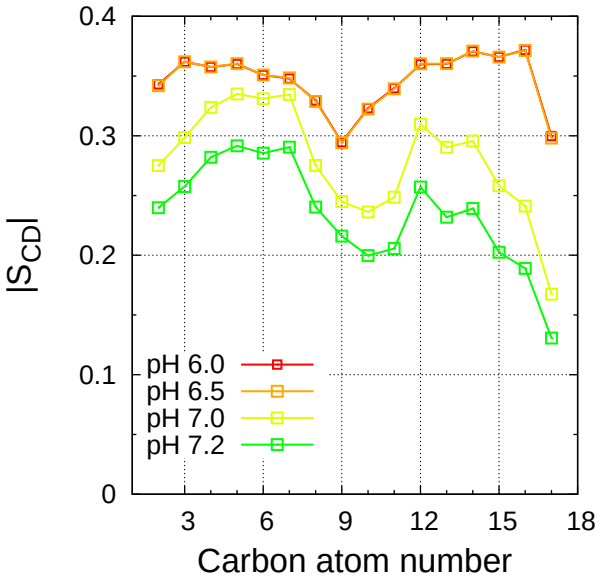


Figure A.5: Deuterium order parameters for all CpHMD-L simulations at I = 0.1 M.

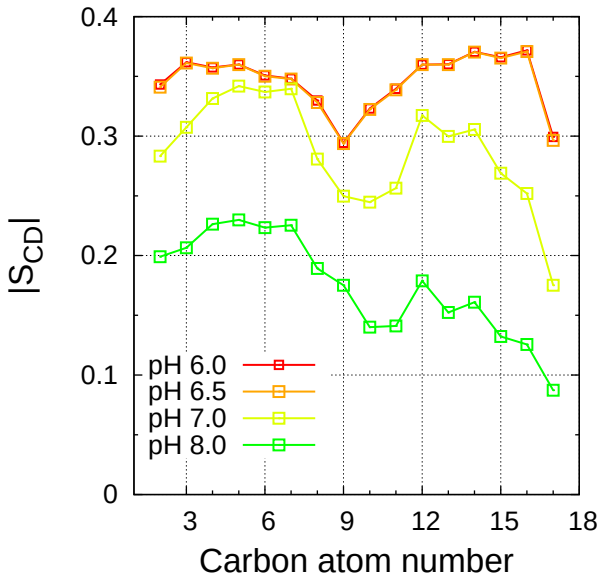


Figure A.6: Deuterium order parameters for all CpHMD-L simulations at I = 0.3 M.

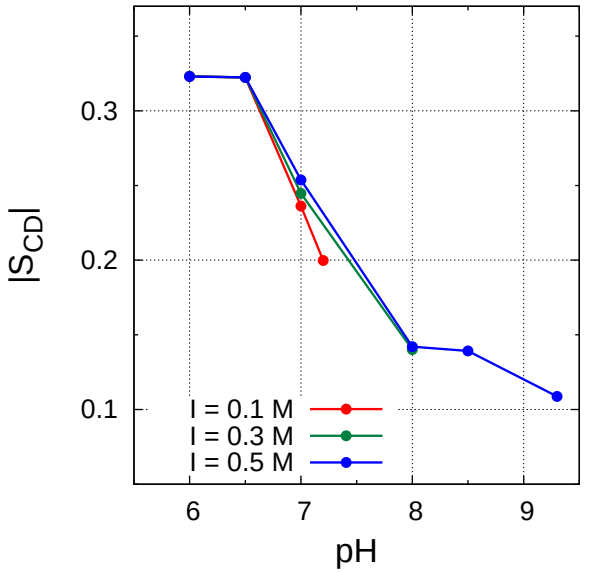


Figure A.7: Deuterium order parameters for the 10^{th} C atom at different pH values.

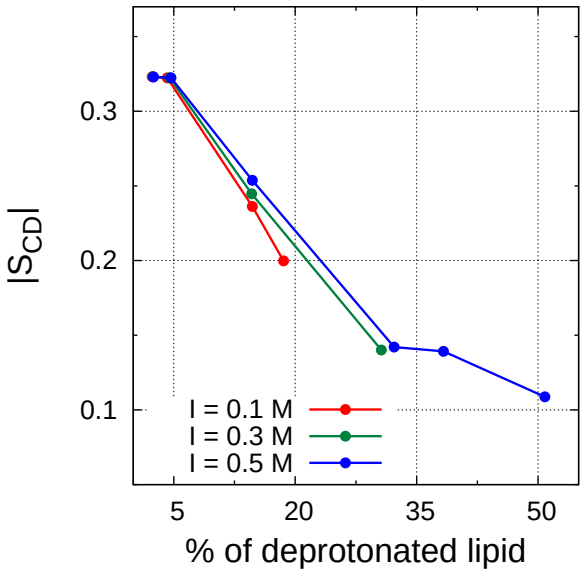


Figure A.8: Deuterium order parameters for the 10^{th} C atom at different membrane ionization values.

Diffusion coefficient values obtained in all CpHMD-L simulations

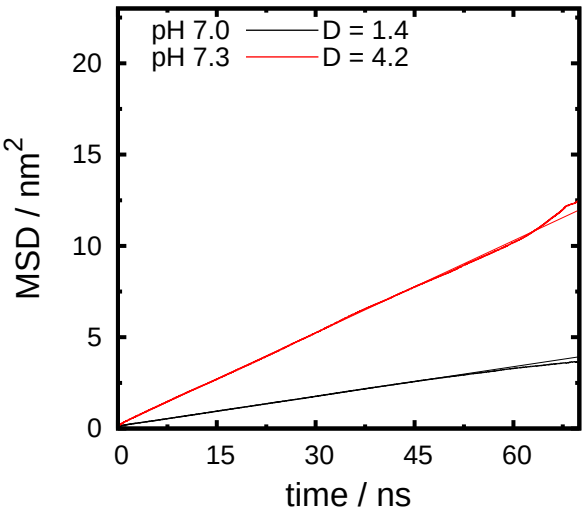


Figure A.9: Mean square displacement (MSD) of all 100 oleic acids over time for all CpHMD-L simulations at $I = 0.1$ M. The diffusion coefficients (10^{-7} cm^2 s^{-1}) were obtained from the slope of the linear portion of the curve (10 - 50 ns).

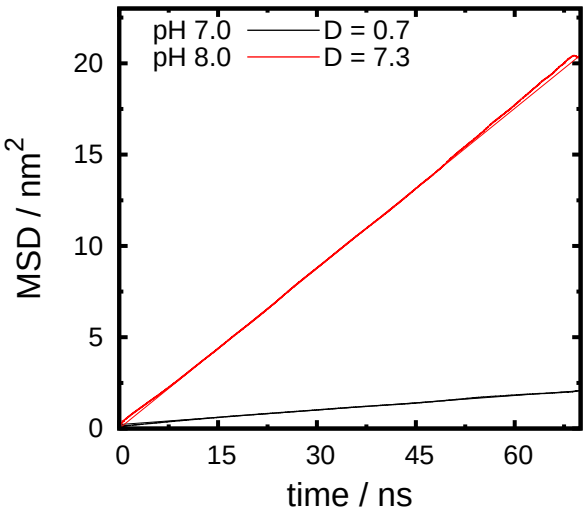


Figure A.10: MSD of all 100 oleic acids over time and diffusion coefficient for all CpHMD-L simulations at $I = 0.3$ M.

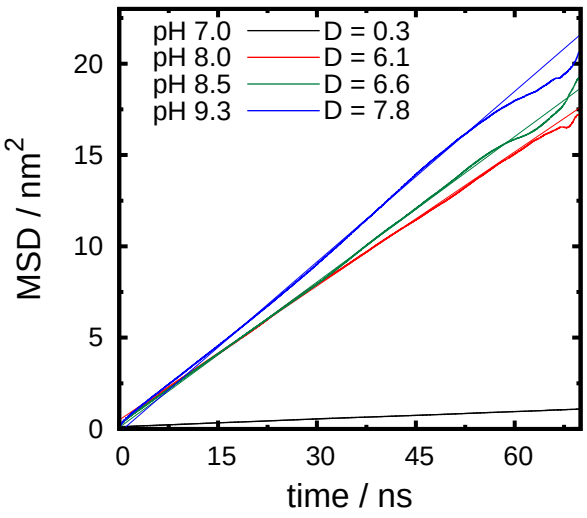


Figure A.11: MSD of all 100 oleic acids over time and diffusion coefficient for all CpHMD-L simulations at $I = 0.5$ M.

Sodium – lipid interaction

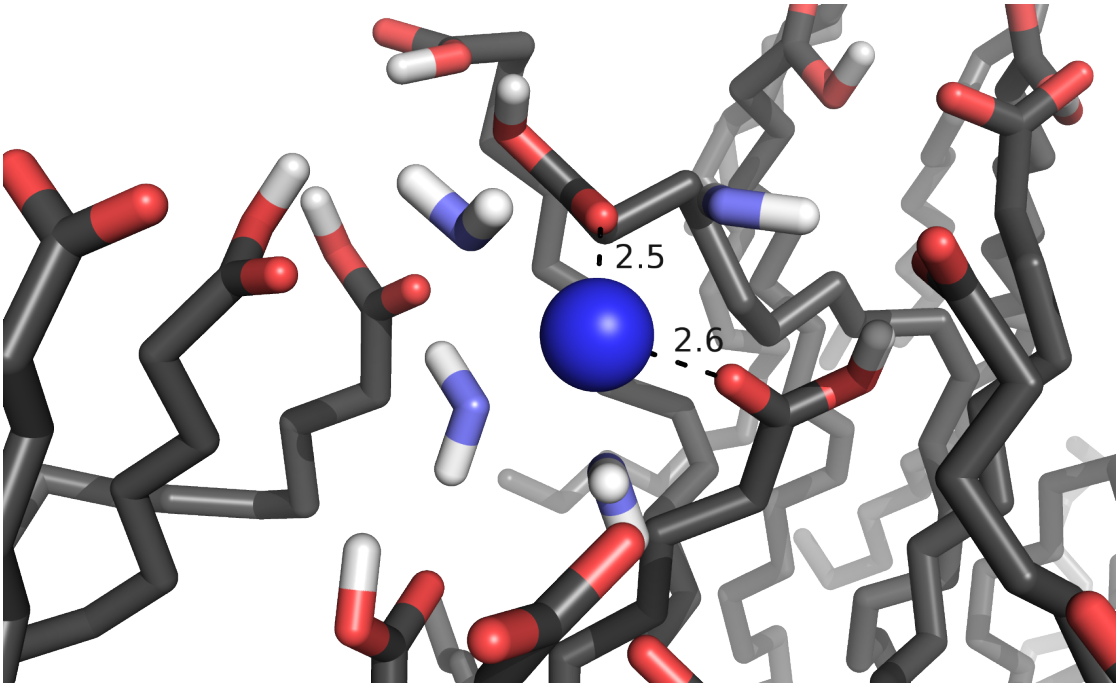


Figure A.12: Representation of a typical Na^+ – lipid interaction at $\sim$ 2.5 Å.

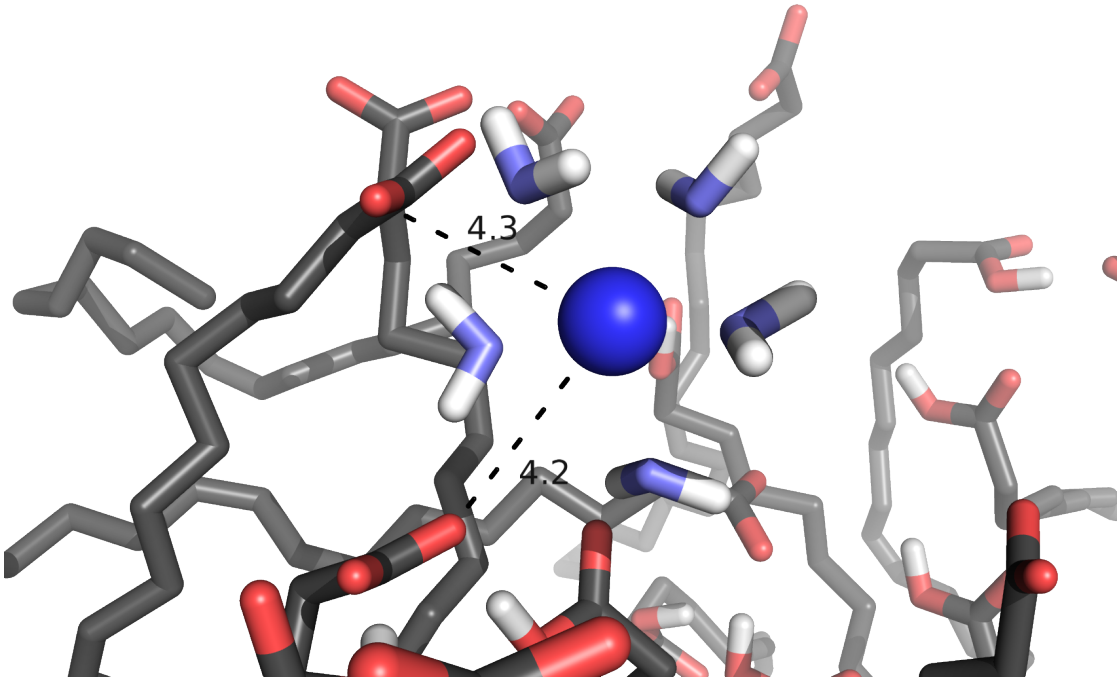


Figure A.13: Representation of a typical Na^+ – lipid interaction at $\sim$ 4.3 Å.

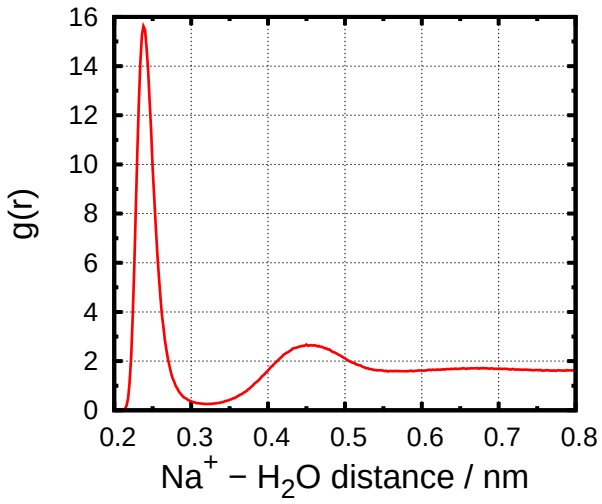


Figure A.14: Radial distribution function ($g(r)$) of water molecules around Na^+ ions.

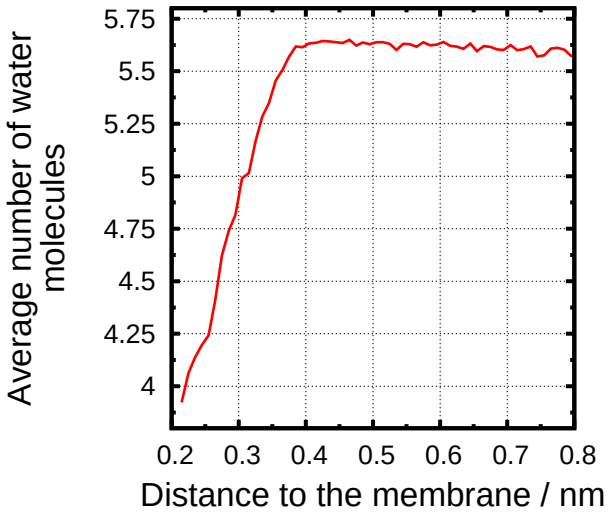


Figure A.15: Average number of water molecules in the first shell of solvation of Na^+ ions as a function of the distance to the membrane.

Appendix B

Methods for "Constant-pH MD simulations of DMPA/DMPC lipid bilayers" (Section 6.2)

CpHMD-L settings

CpHMD-L is an extension of the stochastic titration method, [12, 13] consisting of three successive steps that are cyclically repeated during the simulations: (1) PB/MC to sample the protonation states, (2) MM/MD during a time τ_{rlx} to relax the solvent around a frozen solute, and (3) unrestrained MM/MD during a time τ_{prt}. In our previous work, we showed that $\tau_{\mathrm{rlx}} = 0.2$ ps and $\tau_{\mathrm{prt}} = 20$ ps lead to structurally stable membrane simulations and to extensive sampling of the protonation states, which consequently leads to a smaller chance of falling into kinetic traps during the simulations [3].

Molecular dynamics settings

All MD simulations were performed using the GROMOS 54A7 force field [147] in a modified version of the GROMACS distribution (version 4.0.7) [12, 13] and with the SPC water model [214]. The atomic partial charges of the lipid molecules were the ones previously published [1, 2].

The simulations were done in the NPT ensemble, using the v-rescale thermostat [161] at 310 K with separate couplings for the "solute" (lipids) and solvent (water and ions) with a relaxation time of 0.1 ps. The gel-liquid crystal phase transition temperature of DMPC is $\sim$ 297 K and since DMPC is the major lipid constituent of our membrane systems, an absolute temperature of 310 K assures that the lipid bilayer is mostly in the liquid-crystalline phase state. On the other hand,

the correspondent transition temperature for DMPA is very dependent on the solution pH [15, 283]. A semi-isotropic Berendsen [160] pressure couple was used to maintain the pressure constant at 1 bar with a compressibility of 4.5×10^{-5} bar^{-1} and a relaxation time of 5.0 ps. The bond lengths of all molecules were constraint using using P-LINCS algorithm [163]. An integration step of 2 fs was used for the equations of motion in all MD simulations.

Nonbonded interactions were treated with a twin-range cutoff, with short and long range cutoffs of 8 Å and 14 Å, respectively. The neighbor pair list was updated every 5 steps. Long-range electrostatics interactions were treated using the generalized reaction field method [155], with a relative dielectric constant of 54.0 [215] and ionic strength of 0.1 M, being treated as an external parameter [13].

The energy minimization procedure consisted of 10 000 steps using the steepest descent algorithm (unconstrained) followed by another 10 000 steps using the l-BFGS integrator (unconstrained), and yet another 100 steps using again the steepest descent algorithm (with all bonds constrained). The initiation of all systems consisted of a 50 ps simulation with all heavy atoms harmonically restrained to their respective fixed reference positions with a force constant of 1 000 kJnm^{-2}mol^{-1}. This first initiation step was followed by three other steps of 100 ps, 150 ps, and 200 ps simulations, respectively. Throughout these three last steps, the positions restraints where applied differently to phosphorus atoms and all the other heavy atoms. Hence, the force constants where 1 000, 100, and 10 kJnm^{-2}mol^{-1} for the phosphorus atoms pertaining to the phosphate headgroups of all lipids, whereas all other heavy atoms motion potentials where restrained with force constants of 100, 10, and 0 kJnm^{-2}mol^{-1}.

The MD simulations of all bilayer systems were pre-equilibrated for 100 ns, at which point three replicates were started for 100 ns equilibration runs. The first 20 ns were disregarded and the last 80 ns were used to assess equilibration. These equilibrated systems presented a wide variety of ions and were used to initiate our CpHMD-L simulations. An *a priori* estimation of the ionization at a given pH value was needed to choose the correct system.

PB/MC settings for CpHMD-L simulations

The PB calculations were done with the program DelPhi V5.0 [205, 206] using atomic radii derived from GROMOS 54a7 force field [212] and atomic partial charges from previous works [1, 2]. The molecular surface of the membrane was defined by a probe of radius 0.14 nm, and the ion exclusion layer was 0.2 nm. In order to avoid discontinuities of the molecular surface at the box boundaries, we added a small portion (5% of the box side dimension) of the membrane atoms in the x and y directions [1]. Periodic boundary conditions were explicitly applied along

those two directions in the calculation of the potential. A convergence threshold value based on maximum change of potential of 0.01 $k_B\ T/e$ was used. All PB calculations were done using cubic grids with 61^3 grid points and a two step focusing procedure [219] with successive grid spacing of ~ 0.1 nm and ~ 0.025 nm. In the iteration convergence process of the finite differences method, we used relaxation parameters for the nonlinear and linear forms of the PB equation with values of respectively 0.75 and 0.20. A cutoff of 2.5 nm was used to calculate the background and pairwise interactions (see reference [1] for more details). The pK_a values of the model compounds used were 1.29 for DMPC (the pK_a value of dimethylphosphate) and 1.54 and 6.31 for DMPA (the two pK_a values of methylphosphate) [225]. The dielectric constant of the solvent was 80 and for the membrane we used a value of 2.

The MC sampling was performed with the PETIT program [176] version 1.6.1 using an absolute temperature of 310 K. All runs were performed using 10^5 MC cycles, where one cycle consists of sequential state changes over all individual sites and pairs of sites with an interaction larger than 2 pH units.

PB settings for ion estimation

The PB calculations performed for the ion estimation were similar to the described in previous section. The focusing procedure was not used since we were interested in all grid points closer than 1.4 nm to the membrane. In the iteration convergence process of finite differences method, we used relaxation parameters for the nonlinear and linear forms of the PB equation with values of respectively 0.15 and 0.20. In these calculations, we used 200 configurations from each simulation. To estimate the number of ions, the electrostatic potential was written to a new 3D grid with 50 points per side, the same size as the finite differences grid [1]. The number of cations, Na^+, and of anions, Cl^-(Tables S2 and S3 of Supporting Information), were calculated according to the previously published procedure [2, 3].

Analysis

The equilibrated portions of each simulation were used for analysis. A_l and deuterium order parameters were obtained using several tools from the GROMACS software package [210, 211]. Membrane thickness and DMPA clustering were calculated using in-house tools. Membrane thickness was defined as twice the average z distance between each phosphorus atom and the center of mass (also in z) of all phosphorus atoms. All presented errors were computed using standard methods based on the autocorrelation function of the property measured to determine the

number of independent blocks in the simulations [134].

Appendix C

Methods for "$\mathrm{p}K_\mathrm{a}$ values of titrable amino acids at water/membrane interface" (Section 6.3)

Constant-pH MD settings

In this work, we used both stochastic titration constant-pH MD method (CpHMD) [12, 13, 266] and the recently developed extension (CpHMD-L) that accounts for membrane environments [1, 3]. Our constant-pH MD methodology follows a series of small steps in a repeating cycle [3, 12, 13], starting with a Poisson–Boltzmann / Monte Carlo (PB/MC) calculation to determine the protonation states of the titrable groups for a specific conformation and at a given pH value. Representative protonation states are assigned to the solute and, in the second step, a short (τ_rlx) molecular mechanics / molecular dynamics (MM/MD) simulation in the NVT ensemble is performed with the solute frozen to allow the local water molecules to relax to the new charge distribution. The third step consists of a production run of MM/MD simulation of the unfrozen system (τ_prt). These steps are repeated until the desired simulation time.

Constant-pH MD simulations were performed in eight pentapeptides where all residues were an Ala, with the exception of the middle one which were Asp, Cys, Glu, His, Tyr or Lys. All N- and C-termini were capped with an acetyl or amino groups, respectively. When these termini were to be titrated, they were uncapped and all five residues were Ala. For each pentapeptide, we simulated three pH values and made three (aqueous) or five (membrane) replicates. The simulations were 30 and 100 ns long in water and membrane, respectively. Conformations

were saved every 2 ps. The τ_{prt} was 2 ps for the simulations in water and 20 ps for the simulations in bilayer membranes. A τ_{prt} value of 2 ps has been successfully used with peptides and proteins in aqueous solution [12, 13], while in a membrane environment our method has been shown to work better with larger values, like 10 ps [266] or even 20 ps[3].

MM/MD settings

The GROMACS package version 4.0.7 [210, 211] was used in all MM/MD simulations with the GROMOS 54A7 force field [212, 213] which already includes the corrections for the DMPC molecule [138]. The atomic partial charges were taken from Teixeira et al. [1]. All simulations were done using the v-rescale thermostat [161] at ∼300 K with separate couplings for solute and solvent and a relaxation time of 0.1 ps. The pressure was kept at 1 bar using the Berendsen barostat [160] (isotropic in water and semi-isotropic in the bilayer membrane) with an isothermal compressibility of 4.5×10^{-5} bar^{-1} and a relaxation time of 0.5 ps (water) and 2.0 ps (bilayer membrane). Solutes bond-lengths were constrained using P-LINCS algorithm [163], while for water the SETTLE algorithm was used [164]. Nonbonded interactions were treated with a twin-range method 8/14 Å whose neighbor list was updated every 5 steps (10 ps). Long range electrostatics were treated with a generalized reaction field [155] with a relative dielectric constant of 54 [215] and an ionic strength of 0.1 M [13]. The time equations of motion were integrated with a time step of 2 fs.

In the aqueous systems, the pentapeptides were solvated in a dodecahedral box with periodic boundary conditions and filled with ∼1100 SPC [214] water molecules. The lipid membrane systems started from an equilibrated lipid bilayer of 128 DMPC molecules and was solvated in a tetragonal box with ∼4 700 SPC water molecules, with the membrane surface oriented parallel to the xy plane. Energy minimization was performed in 3 steps that consisted of an initial 10 000 steps of steepest–descent unconstrained calculation, followed by a 2 000 steps l-BFGS algorithm also without constraints and the final step was a ∼50 steps of steepest–descent with all bonds constrained. The aqueous systems were initialized in 3 steps of 100, 150, and 200 ns of MD simulation. Initial velocities were taken from a Maxwell distribution at 300 K. Position restraints were applied to all peptide atoms, all heavy atoms and all CA atoms in the 3 initiation steps, respectively, with a force constant of 1 000 kJ/(mol nm^2). The membrane systems were initialized in four steps of 100, 150, 200, and 250 ns of MD simulation. Position restraints were applied sequentially in all steps to all solute atoms, all heavy atoms, all P atoms (1 000 kJ mol$^-$1 nm$^-$2) and to all P atoms with a lower position restraint force constant (10 kJ mol$^-$1 nm$^-$2).

PB/MC settings

The program DelPhi Version 5.1 [205, 206] was used for the PB calculations using the partial charges referred in the previous section and radii derived [204] from the GROMOS 54A7 force field [212, 213] Lennard–Jones parameters. The molecular surface was defined with a probe of radius 1.4 Å, the ion exclusion layer was 2.0 Å and the ionic strength was 0.1 M. In the calculations of the membrane systems, we added 5% of the box vector dimension in the x and y direction, to ensure continuity in the membrane surface definition (see Ref 1). Periodic boundary conditions were explicitly applied in both x and y directions for the potential calculation in the coarse grid. In water systems, the maximum number of linear iterations was set to 300, while in membrane the coarser grid relaxation parameters were changed to 0.2 and 0.75 for linear and nonlinear PB equations, respectively. A dielectric constant of 2 was used for peptide and membrane and 80 for water. In the lipid membrane systems, we used a cutoff of 25 Å to calculate the background contributions and pairwise interaction [1]. The convergence threshold value based on maximum change of potential was set to 0.01 $k_{\mathrm{B}}\,T/e$. Calculations were done in a cubic grid of 61 grid points and a two step focusing [219] with the focus grid being one fourth of the coarser grid size. These settings result in grid spacings of $\sim$1Å and $\sim$0.25Å for the coarser and focus grids, respectively.

The MC sampling was performed with the PETIT program [176] version 1.6. All runs were performed at an absolute temperature of 300 K using 10^5 MC cycles. Each cycle consists of sequential state changes over all individual sites and pairs of sites with an interaction larger than 2 pK units.

pK_{a} calculation procedure

The pK_{a} values in the membrane systems cannot be calculated similarly to those in water. To capture the average effects of the lipid interactions with the titrating groups, we propose a fractioning of our system in the membrane normal direction. This was achieved by defining a distance in the z-vector to the closest lipid (measured to its phosphor atom) and partitioning it in small slabs (0.5 Å). The pentapeptides reference atoms chosen for distance measurements were the C atoms in the carboxylic groups, N in amines, O in Tyr, S in Cys and Nϵ in His. The closest lipid was chosen instead of the most commonly used phosphor atom average z-positions (from the closest lipid monolayer), in order to avoid erroneous averaging out of the membrane deformation induced by the pentapeptides. Each fraction of our system is associated with a degree of insertion and the pentapeptide conformation/protonation states are separated accordingly. The pK_{a} values for each slab can be calculated using a Henderson–Hasselbalch fit to the weighted

average protonation states from the three pH values simulations. The fact that we are splitting the available conformations can generate limited sampling in some less populated slabs, which we circumvented by simulating five replicates at longer times (100 ns), per pH value. On the other hand, with this approach, we can avoid sampling problems for those cases where the pentapeptide is trapped deep in the membrane. Here, due to the lack of protonation changes (only the neutral form is stable), the resulting conformations are ignored by the pK_a calculation procedure. We also implemented two rules in order to avoid calculating pK_a values in slabs with insufficient sampling, namely, the site needs to be titrating (with protonation different from 0.00 or 1.00) in at least 2 different pH values, and at least a number of 50 conformations per pH value is required in the averaging before the Henderson–Hasselbalch fit.

Bibliography

[1] V. H. Teixeira, D. Vila-Viçosa, A. M. Baptista, and M. Machuqueiro. Protonation of DMPC in a bilayer environment using a linear response approximation. *J. Chem. Theory Comput.*, 10:2176–2184, 2014.

[2] D. Vila-Viçosa, V. H. Teixeira, H. A. F. Santos, A. M. Baptista, and M. Machuqueiro. Treatment of ionic strength in biomolecular simulations of charged lipid bilayers. *J. Chem. Theory Comput.*, 10:5483–5492, 2014.

[3] D. Vila-Viçosa, V. H. Teixeira, A. M. Baptista, and M. Machuqueiro. Constant-pH MD simulations of an oleic acid bilayer. *J. Chem. Theory Comput.*, 11(5):2367–2376, 2015.

[4] M. Đaković, D. Vila-Viçosa, M. J. Calhorda, and Z. Popović. Coordination-driven self-assembly of thiocyanate complexes of Co(II), Ni(II) and Cu(II) with picolinamide: a structural and DFT study. *Cryst. Eng. Comm.*, 13(19):5863–5871, 2011.

[5] M. Đaković, D. Vila-Viçosa, N. A. G. Bandeira, M. J. Calhorda, B. Kozlevčar, Z. Jagličić, and Z. Popović. Can self-assembly of copper(II) picolinamide building blocks be controlled? *Cryst. Eng. Comm.*, 15(40):8074–8087, 2013.

[6] C. Fliedel, D. Vila-Viçosa, M. J. Calhorda, S. Dagorne, and T. Avilés. Dinuclear zinc–n-heterocyclic carbene complexes for either the controlled ring-opening polymerization of lactide or the controlled degradation of polylactide under mild conditions. *Chem. Cat. Chem.*, 6(5):1357–1367, 2014.

[7] D. Suresh, C. S. B. Gomes, P. S. Lopes, C. A. Figueira, B. Ferreira, P. T. Gomes, R. E. Di Paolo, A. L. Maçanita, M. T. Duarte, A. Charas, J. Morgado, D. Vila-Viçosa, and M. J. Calhorda. Luminescent di-and trinuclear boron complexes based on aromatic iminopyrrolyl spacer ligands: Synthesis, characterization, and application in OLEDs. *Chem.-Eur. J.*, 21:9133–9149, 2015.

[8] D. Vila-Viçosa, S. R. R. Campos, A. M. Baptista, and M. Machuqueiro. Reversibility of prion misfolding: insights from constant-pH molecular dynamics simulations. *J. Phys. Chem. B*, 116(30):8812–8821, 2012.

[9] S. G. Estácio, H. Krobath, D. Vila-Viçosa, M. Machuqueiro, E. I. Shakhnovich, and P. F. N. Faísca. A simulated intermediate state for folding and aggregation provides insights into ΔN6 β2-microglobulin amyloidogenic behavior. *PLOS Comp. Biol.*, 10:1–17, 2014.

[10] D. Vila-Viçosa, V. H. Teixeira, H. A. F. Santos, and M. Machuqueiro. Conformational study of GSH and GSSG using constant-pH molecular dynamics simulations. *J. Phys. Chem. B*, 117(25):7507–7517, 2013.

[11] D. Vila-Viçosa, O. Francesconi, and M. Machuqueiro. Why a diaminopyrrolic tripodal receptor binds mannosides in acetonitrile but not in water? *Beilstein J. Org. Chem.*, 10(1):1513–1523, 2014.

[12] A. M. Baptista, V. H. Teixeira, and C. M. Soares. Constant-pH molecular dynamics using stochastic titration. *J. Chem. Phys.*, 117:4184–4200, 2002.

[13] M. Machuqueiro and A. M. Baptista. Constant-pH molecular dynamics with ionic strength effects: Protonation–conformation coupling in decalysine. *J. Phys. Chem. B*, 110:2927–2933, 2006.

[14] T. H. Haines. Anionic lipid headgroups as a proton-conducting pathway along the surface of membranes: a hypothesis. *Proc. Natl. Acad. Sci. USA*, 80:160, 1983.

[15] P. Garidel, C. Johann, and A. Blume. Nonideal mixing and phase separation in phosphatidylcholine-phosphatidic acid mixtures as a function of acyl chain length and pH. *Biophys. J.*, 72:2196–2210, 1997.

[16] J. F. Hunt, P. Rath, K. J. Rothschild, and D. M. Engelman. Spontaneous, pH-dependent membrane insertion of a transbilayer α-helix. *Biochemistry-US*, 36(49):15177–15192, 1997.

[17] P. Lemay. Michel eugene chevreul (1786-1889). *J. Chem. Educ.*, 25(2):62, 1948.

[18] D. L. Nelson and M. M. Cox. *Lehninger Principles of Biochemistry*. New York, New York: WH Freeman and Company, 2005.

[19] F. M. Menger, M. E. Chlebowski, A. L. Galloway, H. Lu, V. A. Seredyuk, J. L. Sorrells, and H. Zhang. A tribute to the phospholipid. *Langmuir*, 21:10336–10341, 2005.

[20] P. V. Escribá, J. M. González-Ros, F. M. Goñi, P. K. J. Kinnunen, L. Vigh, L. Sánchez-Magraner, A. M. Fernández, X. Busquets, I. Horváth, and G. Barceló-Coblijn. Membranes: a meeting point for lipids, proteins and therapies. *J. Cell. Mol. Med.*, 12(3):829–875, 2008.

[21] G. Van Meer, D. R. Voelker, and G. W. Feigenson. Membrane lipids: where they are and how they behave. *Nat. Rev. Mol. Cell. Bio.*, 9:112–124, 2008.

[22] E. Fahy, S. Subramaniam, H. A. Brown, C. K. Glass, A. H. Merrill, R. C. Murphy, C. R. H. Raetz, D. W. Russell, Y. Seyama, W. Shaw, T. Shimizu, F. Spener, G. van Meer, M. S. VanNieuwenhze, S. H. White, J. L. Witztum, and E. A. Dennis. A comprehensive classification system for lipids. *J. Lipid Res.*, 46(5):839–862, 2005.

[23] E. Fahy, S. Subramaniam, R. C. Murphy, M. Nishijima, C. R. H. Raetz, T. Shimizu, F. Spener, G. van Meer, M. J. O. Wakelam, and E. A. Dennis. Update of the lipid maps comprehensive classification system for lipids. *J. Lipid Res.*, 50(Supplement):S9–S14, 2009.

[24] *IUPAC Compendium of chemical terminology.* 2014.

[25] J. F. Nagle and D. A. Wilkinson. Lecithin bilayers. density measurement and molecular interactions. *Biochem. J.*, 23(2):159–175, 1978.

[26] M. J. Janiak, D. M. Small, and G. G. Shipley. Temperature and compositional dependence of the structure of hydrated dimyristoyl lecithin. *J. Biol. Chem.*, 254:6068–6078, 1979.

[27] L. J. Lis, M. McAlister, N. Fuller, R. P. Rand, and V. A. Parsegian. Interactions between neutral phospholipid bilayer membranes. *Biophys. J.*, 37:657, 1982.

[28] B. A. Lewis and D. M. Engelman. Lipid bilayer thickness varies linearly with acyl chain length in fluid phosphatidylcholine vesicles. *J. Mol. Biol.*, 166:211–217, 1983.

[29] R. P. Rand and V. A. Parsegian. Hydration forces between phospholipid bilayers. *Biochem. Biophys. Acta*, 988:351–376, 1989.

[30] B. W. Koenig, H. H. Strey, and K. Gawrisch. Membrane lateral compressibility determined by NMR and X-ray diffraction: effect of acyl chain polyunsaturation. *Biophys. J.*, 73:1954–1966, 1997.

[31] H. I. Petrache, S. Tristram-Nagle, and J. F. Nagle. Fluid phase structure of EPC and DMPC bilayers. *Chem. Phys. Lipids*, 95:83–94, 1998.

[32] J. F. Nagle and S. Tristram-Nagle. Lipid bilayer structure. *Curr. Opin. Struct. Biol.*, 10:474–480, 2000.

[33] J. F. Nagle and S. Tristram-Nagle. Structure of lipid bilayers. *Biochem. Biophys. Acta, Rev. Biomembr.*, 1469(3):159–195, 2000.

[34] N. Kučerka, Y. Liu, N. Chu, H. I. Petrache, S. Tristram-Nagle, and J. F. Nagle. Structure of fully hydrated fluid phase DMPC and DLPC lipid bilayers using X-ray scattering from oriented multilamellar arrays and from unilamellar vesicles. *Biophys. J.*, 88:2626–2637, 2005.

[35] L. R. De Young and K. A. Dill. Solute partitioning into lipid bilayer membranes. *Biochemistry-US*, 27:5281–5289, 1988.

[36] J. M. Smaby, M. M. Momsen, H. L. Brockman, and Rhoderick E. B. Phosphatidylcholine acyl unsaturation modulates the decrease in interfacial elasticity induced by cholesterol. *Biophys. J.*, 73:1492–1505, 1997.

[37] S. C. Costigan, P. J. Booth, and R. H. Templer. Estimations of lipid bilayer geometry in fluid lamellar phases. *Biochem. Biophys. Acta, Biomembr.*, 1468:41–54, 2000.

[38] H. I. Petrache, S. W. Dodd, and M. F. Brown. Area per lipid and acyl length distributions in fluid phosphatidylcholines determined by ^{2}H-NMR spectroscopy. *Biophys. J.*, 79:3172–3192, 2000.

[39] P. Balgavỳ, M. Dubničková, N. Kučerka, M. A. Kiselev, S. P. Yaradaikin, and D. Uhríková. Bilayer thickness and lipid interface area in unilamellar extruded 1, 2-diacylphosphatidylcholine liposomes: a small-angle neutron scattering study. *Biochem. Biophys. Acta, Biomembr.*, 1512:40–52, 2001.

[40] N. Kučerka, M. A. Kiselev, and P. Balgavỳ. Determination of bilayer thickness and lipid surface area in unilamellar dimyristoylphosphatidylcholine vesicles from small-angle neutron scattering curves: a comparison of evaluation methods. *Eur. Biophys. J.*, 33:328–334, 2004.

[41] J. N. Israelachvili. *Intermolecular and surface forces: revised third edition.* Academic press, 2011.

[42] P. R. Cullis and B. de Kruijff. Lipid polymorphism and the functional roles of lipids in biological membranes. *Biochem. Biophys. Acta, Rev. Biomembr.*, 559(4):399–420, 1979.

[43] S. Carnie, J. N. Israelachvili, and B. A. Pailthorpe. Lipid packing and transbilayer asymmetries of mixed lipid vesicles. *Biochem. Biophys. Acta, Biomembr.*, 554(2):340–357, 1979.

[44] D. E. Vance and J. E. Vance. *Biochemistry of lipids, lipoproteins and membranes. Chapter 1.* Elsevier, 1996.

[45] R. E. Goldstein and S. Leibler. Model for lamellar phases of interacting lipid membranes. *Phys. Rev. Lett.*, 61(19):2213, 1988.

[46] A. A. Gurtovenko, M. Patra, M. Karttunen, and I. Vattulainen. Cationic dmpc/dmtap lipid bilayers: molecular dynamics study. *Biophys. J.*, 86(6):3461–3472, 2004.

[47] J. P. F. Doux, B. A. Hall, and J. A. Killian. How lipid headgroups sense the membrane environment: An application of 14N NMR. *Biophys. J.*, 103(6):1245–1253, 2012.

[48] E. M. Jones, M. Dubey, Phillip J. Camp, B. C. Vernon, J. Biernat, E. Mandelkow, J. Majewski, and E. Y. Chi. Interaction of tau protein with model lipid membranes induces tau structural compaction and membrane disruption. *Biochemistry-US*, 51(12):2539–2550, 2012.

[49] D. Andelman. Electrostatic properties of membranes: the Poisson–Boltzmann theory. *Handbook of biological physics*, 1:603–642, 1995.

[50] Y. N. Antonenko and P. Pohl. Coupling of proton source and sink via H^+-migration along the membrane surface as revealed by double patch-clamp experiments. *FEBS Lett.*, 429(2):197–200, 1998.

[51] M. Zhang, E. Mileykovskaya, and W. Dowhan. Gluing the respiratory chain together cardiolipin is required for supercomplex formation in the inner mitochondrial membrane. *J. Biol. Chem.*, 277(46):43553–43556, 2002.

[52] S. Serowy, S. M. Saparov, Y. N. Antonenko, W. Kozlovsky, V. Hagen, and P. Pohl. Structural proton diffusion along lipid bilayers. *Biophys. J.*, 84(2):1031–1037, 2003.

[53] Y. N. Antonenko and P. Pohl. Microinjection in combination with microfluorimetry to study proton diffusion along phospholipid membranes. *Eur. Biophys. J.*, 37(6):865–870, 2008.

[54] T. H. Haines. A new look at cardiolipin. *Biochem. Biophys. Acta, Biomembr.*, 1788:1997–2002, 2009.

[55] L. Ojemyr, T. Sandén, J. Widengren, and P. Brzezinski. Lateral proton transfer between the membrane and a membrane protein. *Biochemistry-US*, 48(10):2173–2179, 2009.

[56] M. P. Goertz, N. Goyal, G. A. Montano, and B. C. Bunker. Lipid bilayer reorganization under extreme pH conditions. *Langmuir*, 27(9):5481–5491, 2011.

[57] A. E. Blaurock and T. J. McIntosh. Structure of the crystalline bilayer in the subgel phase of dipalmitoylphosphatidylglycerol. *Biochemistry-US*, 25(2):299–305, 1986.

[58] D. C. Crans and N. E. Levinger. The conundrum of pH in water nanodroplets: sensing pH in reverse micelle water pools. *Accounts Chem. Res.*, 45(10):1637–1645, 2012.

[59] D. P. Cistola, J. A. Hamilton, D. Jackson, and D. M. Small. Ionization and phase behavior of fatty acids in water: application of the gibbs phase rule. *Biochemistry-US*, 27(6):1881–1888, 1988.

[60] P. Peterlin, V. Arrigler, K. Kogej, S. Svetina, and P. Walde. Growth and shape transformations of giant phospholipid vesicles upon interaction with an aqueous oleic acid suspension. *Chem. Phys. Lipids*, 159(2):67–76, 2009.

[61] T. H. Haines and N. A. Dencher. Cardiolipin: a proton trap for oxidative phosphorylation. *FEBS Lett.*, 528(1):35–39, 2002.

[62] R. N. A. H. Lewis, D. Zweytick, G. Pabst, K. Lohner, and R. N. McElhaney. Calorimetric, X-ray diffraction, and spectroscopic studies of the thermotropic phase behavior and organization of tetramyristoyl cardiolipin membranes. *Biophys. J.*, 92(9):3166–3177, 2007.

[63] D. Acehan, A. Malhotra, Y. Xu, M. Ren, D. L. Stokes, and M. Schlame. Cardiolipin affects the supramolecular organization of ATP synthase in mitochondria. *Biophys. J.*, 100(9):2184–2192, 2011.

[64] R. Arias-Cartin, S. Grimaldi, J. Pommier, P. Lanciano, C. Schaefer, P. Arnoux, G. Giordano, B. Guigliarelli, and A. Magalon. Cardiolipin-based respiratory complex activation in bacteria. *Proc. Natl. Acad. Sci. USA*, 108(19):7781–7786, 2011.

[65] M. Dahlberg and A. Maliniak. Molecular dynamics simulations of cardiolipin bilayers. *J. Phys. Chem. B*, 112(37):11655–11663, 2008.

[66] S. Pöyry, T. Róg, M. Karttunen, and I. Vattulainen. Mitochondrial membranes with mono- and divalent salt: Changes induced by salt ions on structure and dynamics. *J. Phys. Chem. B*, 113:15513–15521, 2009.

[67] T. Róg, H. Martinez-Seara, N. Munck, M. Oresic, M. Karttunen, and I. Vattulainen. Role of cardiolipins in the inner mitochondrial membrane: insight gained through atom-scale simulations. *J. Phys. Chem. B*, 113(11):3413–3422, 2009.

[68] R. Hovius, H. Lambrechts, K. Nicolay, and B. de Kruijff. Improved methods to isolate and subfractionate rat liver mitochondria. lipid composition of the inner and outer membrane. *Biochem. Biophys. Acta, Biomembr.*, 1021(2):217–226, 1990.

[69] K. Matsumoto, J. Kusaka, A. Nishibori, and H. Hara. Lipid domains in bacterial membranes. *Mmol. Microbiol.*, 61(5):1110–1117, 2006.

[70] P. G. Barth, H. R. Scholte, J. A. Berden, J. M. Van der Klei-Van Moorsel, I. E. M. Luyt-Houwen, E. T. V. Veer-Korthof, J. J. Van der Harten, and M. A. Sobotka-Plojhar. An X-linked mitochondrial disease affecting cardiac muscle, skeletal muscle and neutrophil leucocytes. *J. Neurol. Sci.*, 62(1):327–355, 1983.

[71] X. Wang, S. P. Devaiah, W. Zhang, and R. Welti. Signaling functions of phosphatidic acid. *Prog. Lipid Res.*, 45:250–278, 2006.

[72] C. J. R. Loewen, M. L. Gaspar, S. A. Jesch, C. Delon, N. T. Ktistakis, S. A. Henry, and T. P. Levine. Phospholipid metabolism regulated by a transcription factor sensing phosphatidic acid. *Science*, 304:1644–1647, 2004.

[73] B. P. Young, J. J. H. Shin, R. Orij, J. T. Chao, S. C. Li, X. L. Guan, A. Khong, E. Jan, M. R Wenk, W. A. Prinz, G. J. Smits, and C. J. R. Loewen. Phosphatidic acid is a pH biosensor that links membrane biogenesis to metabolism. *Science*, 329:1085–1088, 2010.

[74] F. Lakhdar-Ghazal, J.-L. Tichadou, and J.-F. Tocanne. Effect of pH and monovalent cations on the ionization state of phosphatidylglycerol in monolayers. *Eur. J. Biochem.*, 134:531–537, 1983.

[75] D. H. Mengistu, E. E. Kooijman, and S. May. Ionization properties of mixed lipid membranes: A Gouy–Chapman model of the electrostatic–hydrogen bond switch. *Biochem. Biophys. Acta, Biomembr.*, 1808:1985–1992, 2011.

[76] G. Lamm. The Poisson–Boltzmann equation. *Rev. Comp. Ch.*, 19:147–333, 2003.

[77] B. W. Ninham and V. A. Parsegian. Electrostatic potential between surfaces bearing ionizable groups in ionic equilibrium with physiologic saline solution. *J. Theor. Biol.*, 31(3):405–428, 1971.

[78] G. M. Torrie and J. P. Valleau. A Monte Carlo study of an electrical double layer. *Chem. Phys. Lett.*, 65:343–346, 1979.

[79] B. Jönsson, H. Wennerstroem, and B. Halle. Ion distributions in lamellar liquid crystals. a comparison between results from Monte Carlo simulations and solutions of the Poisson–Boltzmann equation. *J. Phys. Chem.*, 84:2179–2185, 1980.

[80] L. Guldbrand, B. Jönsson, H. Wennerström, and P. Linse. Electrical double layer forces. a Monte Carlo study. *J. Chem. Phys.*, 80:2221–2228, 1984.

[81] R. M. Peitzsch, M. Eisenberg, K. A. Sharp, and S. McLaughlin. Calculations of the electrostatic potential adjacent to model phospholipid bilayers. *Biophys. J.*, 68:729–738, 1995.

[82] J. L. Smart and J. A. McCammon. Surface titration: a continuum electrostatics model. *J. Am. Chem. Soc.*, 118(9):2283–2284, 1996.

[83] R. P. Dias, L. Li, T. A. Soares, and E. Alexov. Modeling the electrostatic potential of asymmetric lipopolysaccharide membranes: The MEMPOT algorithm implemented in DelPhi. *J. Comput. Chem.*, 35:1418–1429, 2014.

[84] H. L. Rosano, K. Breindel, J. H. Schulman, and A. J. Eydt. Mechanism of ionic exchange with carrier molecules through non-aqueous liquid membranes. *J. Colloid Interf. Sci.*, 22(1):58–67, 1966.

[85] D. M. Small, D. J. Cabral, D. P. Cistola, J. S. Parks, and J. A. Hamilton. The ionization behavior of fatty acids and bile acids in micelles and membranes. *Hepatology*, 4(S2):77S–79S, 1984.

[86] M. Meot-Ner, D. E. Elmore, and S. Scheiner. Ionic hydrogen bond effects on the acidities, basicities, solvation, solvent bridging, and self-assembly of carboxylic groups. *J. Am. Chem. Soc.*, 121(33):7625–7635, 1999.

[87] J. R. Kanicky, A. F. Poniatowski, N. R. Mehta, and D. O. Shah. Cooperativity among molecules at interfaces in relation to various technological processes: effect of chain length on the pK_{a} a of fatty acid salt solutions. *Langmuir*, 16(1):172–177, 2000.

[88] C. L. Apel, D. W. Deamer, and M. N. Mautner. Self-assembled vesicles of monocarboxylic acids and alcohols: conditions for stability and for the encapsulation of biopolymers. *Biochem. Biophys. Acta, Biomembr.*, 1559(1):1–9, 2002.

[89] J. R. Kanicky and D. O. Shah. Effect of degree, type, and position of unsaturation on the pk_{a} of long-chain fatty acids. *J. Colloid Interf. Sci.*, 256(1):201–207, 2002.

[90] S. Salentinig, L. Sagalowicz, and O. Glatter. Self-assembled structures and pk_{a} value of oleic acid in systems of biological relevance. *Langmuir*, 26:11670–11679, 2010.

[91] S. Nichols-Smith and T. Kuhl. Electrostatic interactions between model mitochondrial membranes. *Colloid. Surface. B*, 41(2):121–127, 2005.

[92] M. Kates, J.-Y. Syz, D. Gosser, and T. H. Haines. pH-dissociation characteristics of cardiolipin and its 2'-deoxy analogue. *Lipids*, 28(10):877–882, 1993.

[93] D. P. Tieleman, S.-J. Marrink, and H. J. C. Berendsen. A computer perspective of membranes: molecular dynamics studies of lipid bilayer systems. *Biochem. Biophys. Acta, Rev. Biomembr.*, 1331:235–270, 1997.

[94] Z. Li, R. M. Venable, L. A. Rogers, D. Murray, and R. W. Pastor. Molecular dynamics simulations of PIP2 and PIP3 in lipid bilayers: Determination of ring orientation, and the effects of surface roughness on a Poisson–Boltzmann description. *Biophys. J.*, 97:155–163, 2009.

[95] A. Cordomí, J. Prades, J. Frau, O. Vögler, S. S. Funari, J. J. Perez, P. V. Escribá, and F. Barceló. Interactions of fatty acids with phosphatidylethanolamine membranes: X-ray diffraction and molecular dynamics studies. *J. Lipid Res.*, 51:1113–1124, 2010.

[96] K. Lähdesmäki, O. H. Ollila, A. Koivuniemi, P. T. Kovanen, and M. T. Hyvönen. Membrane simulations mimicking acidic pH reveal increased thickness and negative

curvature in a bilayer consisting of lysophosphatidylcholines and free fatty acids. *Biochem. Biophys. Acta, Biomembr.*, 1798:938–946, 2010.

[97] H. Yang, Y. Xu, Z. Gao, Y. Mao, Y. Du, and H. Jiang. Effects of Na^+, K^+, and Ca^{2+} on the structures of anionic lipid bilayers and biological implication. *J. Phys. Chem. B*, 114:16978–16988, 2010.

[98] R. M. Venable, Y. Luo, K. Gawrisch, B. Roux, and R. W. Pastor. Simulations of anionic lipid membranes: Development of interaction-specific ion parameters and validation using NMR data. *J. Phys. Chem. B*, 117:10183–10192, 2013.

[99] J. E. Mertz and B. M. Pettitt. Molecular dynamics at a constant pH. *Int. J. High Perform. C.*, 8(1):47–53, 1994.

[100] A. M. Baptista, P. J. Martel, and S. B. Petersen. Simulation of protein conformational freedom as a function of pH: constant-pH molecular dynamics using implicit titration. *Proteins Struct. Funct. Bioinf.*, 27(4):523–544, 1997.

[101] U. Börjesson and P. H. Hünenberger. Explicit-solvent molecular dynamics simulation at constant pH: methodology and application to small amines. *J. Chem. Phys.*, 114:9706, 2001.

[102] R. Burgi, P. A. Kollman, and W. F. van Gunsteren. Simulating proteins at constant pH: An approach combining molecular dynamics and monte carlo simulation. *Proteins Struct. Funct. Bioinf.*, 47(4):469–480, 2002.

[103] M. Dlugosz and J. M. Antosiewicz. Constant-pH molecular dynamics simulations: a test case of succinic acid. *Chem. Phys.*, 302(1-3):161–170, JUL 12 2004.

[104] M. Dlugosz, J. M. Antosiewicz, and A. D. Robertson. Constant-pH molecular dynamics study of protonation-structure relationship in a heptapeptide derived from ovomucoid third domain. *Phys. Rev. E*, 69(2, Part 1), FEB 2004.

[105] M. S. Lee, F. R. Salsbury, and C. L. Brooks III. Constant-pH molecular dynamics using continuous titration coordinates. *Proteins Struct. Funct. Bioinf.*, 56:738–752, 2004.

[106] J. Mongan, D. A. Case, and J. A. McCammon. Constant pH molecular dynamics in generalized Born implicit solvent. *J. Comput. Chem.*, 25:2038–2048, 2004.

[107] J. Khandogin and C. L. Brooks III. Constant pH molecular dynamics with proton tautomerism. *Biophys. J.*, 89(1):141–157, 2005.

[108] J. Khandogin and C. L. Brooks III. Toward the accurate first-principles prediction of ionization equilibria in proteins. *Biochemistry-US*, 45(31):9363–9373, 2006.

[109] H. A. Stern. Molecular simulation with variable protonation states at constant pH. *J. Chem. Phys.*, 126:164112, 2007.

[110] S. L. Williams, C. A. F. De Oliveira, and J. A. McCammon. Coupling constant pH molecular dynamics with accelerated molecular dynamics. *J. Chem. Theory Comput.*, 6(2):560–568, 2010.

[111] S. Donnini, F. Tegeler, G. Groenhof, and H. Grubmüller. Constant pH molecular dynamics in explicit solvent with λ-dynamics. *J. Chem. Theory Comput.*, 7(6):1962–1978, 2011.

[112] S. G Itoh, A. Damjanović, and B. R. Brooks. pH replica-exchange method based on discrete protonation states. *Proteins Struct. Funct. Bioinf.*, 79(12):3420–3436, 2011.

[113] B. H. Morrow, Y. Wang, J. A. Wallace, P. H. Koenig, and J. K. Shen. Simulating pH titration of a single surfactant in ionic and nonionic surfactant micelles. *J. Phys. Chem. B*, 115(50):14980–14990, 2011.

[114] Y. N. Vorobjev. Potential of mean force of water-proton bath and molecular dynamic simulation of proteins at constant pH. *J. Comput. Chem.*, 33(8):832–842, 2012.

[115] J. A. Wallace and J. K. Shen. Charge-leveling and proper treatment of long-range electrostatics in all-atom molecular dynamics at constant pH. *J. Chem. Phys.*, 137(18):184105, 2012.

[116] B. H. Morrow, P. H. Koenig, and J. K. Shen. Atomistic simulations of pH-dependent self-assembly of micelle and bilayer from fatty acids. *J. Chem. Phys.*, 137:194902–194902, 2012.

[117] W. F. Drew Bennett, A. W. Chen, S. Donnini, G. Groenhof, and D. P. Tieleman. Constant pH simulations with the coarse-grained martini model - application to oleic acid aggregates. *Can. J. Chemistry*, 91(9):839–846, 2013.

[118] B. H. Morrow, P. H. Koenig, and J. K. Shen. Self-assembly and bilayer–micelle transition of fatty acids studied by replica-exchange constant pH molecular dynamics. *Langmuir*, 29(48):14823–14830, 2013.

[119] G. B Goh, B. S Hulbert, H. Zhou, and C. L. Brooks III. Constant pH molecular dynamics of proteins in explicit solvent with proton tautomerism. *Proteins Struct. Funct. Bioinf.*, 82:1319–1331, 2014.

[120] J. Lee, B. T. Miller, A. Damjanovic, and B. R. Brooks. Constant pH molecular dynamics in explicit solvent with enveloping distribution sampling and hamiltonian exchange. *J. Chem. Theory Comput.*, 10(7):2738–2750, 2014.

[121] B. H. Morrow, D. M. Eike, B. P. Murch, P. H. Koenig, and J. K. Shen. Predicting proton titration in cationic micelle and bilayer environments. *J. Chem. Phys.*, 141(8):084714, 2014.

[122] J. M. Swails, D. M. York, and A. E. Roitberg. Constant pH replica exchange molecular dynamics in explicit solvent using discrete protonation states: implementation, testing, and validation. *J. Chem. Theory Comput.*, 10(3):1341–1352, 2014.

[123] Y. Chen and B. Roux. Constant-pH hybrid nonequilibrium molecular dynamics–monte carlo simulation method. *J. Chem. Theory Comput.*, 11(8):3919–3931, 2015.

[124] M. Machuqueiro and A. M. Baptista. The pH-dependent conformational states of kyotorphin: A constant-pH molecular dynamics study. *Biophys. J.*, 92:1836–1845, 2007.

[125] M. Machuqueiro and A. M. Baptista. Acidic range titration of HEWL using a constant-pH molecular dynamics method. *Proteins Struct. Funct. Bioinf.*, 72:289–298, 2008.

[126] S. R. R. Campos and A. M. Baptista. Conformational analysis in a multidimensional energy landscape: Study of an arginylglutamate repeat. *J. Phys. Chem. B*, 113:15989–16001, 2009.

[127] M. Machuqueiro and A. M. Baptista. Molecular dynamics constant-pH and reduction potential: Application to cytochrome c_3. *J. Am. Chem. Soc.*, 131:12586–12594, 2009.

[128] S. R. R. Campos, M. Machuqueiro, and A. M. Baptista. Constant-pH molecular dynamics simulations reveal a β-rich form of the human prion protein at low pH. *J. Phys. Chem. B*, 114:12692–12700, 2010.

[129] M. Machuqueiro and A. M. Baptista. Is the prediction of pK_a values by constant-pH molecular dynamics being hindered by inherited problems? *Proteins Struct. Funct. Bioinf.*, 79:3437–3447, 2011.

[130] J. Henriques, P. J. Costa, M. J. Calhorda, and M. Machuqueiro. Charge parametrization of the DvH-c_3 heme group: Validation using constant-(pH,E) molecular dynamics simulations. *J. Phys. Chem. B*, 117:70–82, 2013.

[131] C. A. Carvalheda, S. R. R. Campos, M. Machuqueiro, and A. M. Baptista. Structural effects of pH and deacylation on surfactant protein c in an organic solvent mixture: a constant-pH MD study. *J. Chem. Inf. Model.*, 53(11):2979–2989, 2013.

[132] X. Kong and C. L. Brooks III. λ-dynamics: A new approach to free energy calculations. *J. Chem. Phys.*, 105(6):2414–2423, 1996.

[133] P. Fowler. Orbital update. *Nature*, 343:222, 1990.

[134] M. P. Allen and D. J. Tildesley. *Computer Simulation of Liquids*. Oxford University Press, USA, 1987.

[135] A. R. Leach. *Molecular modelling: principles and applications*. Addison-Wesley Longman Ltd, 2001.

[136] W. F. van Gunsteren, S. R. Billeter, A. A. Eising, P. H. Hünenberger, P. Krüger, A. E. Mark, W. R. P. Scott, and I. G. Tironi. *Biomolecular Simulation: the GROMOS96 Manual and User Guide*. vdf Hochschulverlag AG an der ETH Zürich, Zürich, 1996.

[137] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller. Equation of state calculations by fast computing machines. *J. Chem. Phys.*, 21(6):1087–1092, 1953.

[138] D. Poger and A. E. Mark. On the validation of molecular dynamics simulations of saturated and cis-monounsaturated phosphatidylcholine lipid bilayers: A comparison with experiment. *J. Chem. Theory Comput.*, 6:325–336, 2010.

[139] S. J. Weiner, P. A. Kollman, D. A. Case, U. C. Singh, C. Ghio, G. Alagona, S. Profeta, and P. Weiner. A new force field for molecular mechanical simulation of nucleic acids and proteins. *J. Am. Chem. Soc.*, 106(3):765–784, 1984.

[140] W. D. Cornell, P. Cieplak, C. I. Bayly, I. R. Gould, K. M. Merz, D. M. Ferguson, D. C. Spellmeyer, T. Fox, J. W. Caldwell, and P. A. Kollman. A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *J. Am. Chem. Soc.*, 117(19):5179–5197, 1995.

[141] V. Hornak, R. Abel, A. Okur, B. Strockbine, A. Roitberg, and C. Simmerling. Comparison of multiple amber force fields and development of improved protein backbone parameters. *Proteins Struct. Funct. Bioinf.*, 65(3):712–725, 2006.

[142] B. R. Brooks, R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan, and M. Karplus. CHARMM: A program for macromolecular energy, minimization, and dynamics calculations. *Journal of Computational Chemistry*, 4(2):187–217, 1983.

[143] A. D. MacKerel Jr, C. L. Brooks III, L. Nilsson, B. Roux, Y. Won, and M. Karplus. CHARMM: The energy function and its parameterization with an overview of the program. *The Encyclopedia of Computational Chemistry*, 1:271–277, 1998.

[144] X. Daura, A. E. Mark, and W. F. Van Gunsteren. Parametrization of aliphatic CHn united atoms of GROMOS96 force field. *J. Comput. Chem.*, 19(5):535–547, 1998.

[145] W. F. van Gunsteren, X. Daura, and A. E. Mark. GROMOS force field. *Encyclopedia of computational chemistry*, 1998.

[146] W. R. P. Scott, P. H. Hünenberger, I. G. Tironi, A. E. Mark, S. R. Billeter, J. Fennen, A. E. Torda, T. Huber, P. Krüger, and W. F. van Gunsteren. The GROMOS biomolecular simulation program package. *J. Phys. Chem. A*, 103(19):3596–3607, 1999.

[147] C. Oostenbrink, A. Villa, A. E. Mark, and W. F. Van Gunsteren. A biomolecular force field based on the free enthalpy of hydration and solvation: The GROMOS force-field parameter sets 53A5 and 53A6. *J. Comput. Chem.*, 25(13):1656–1676, 2004.

[148] C. Oostenbrink, T. A. Soares, N. F. A. Van der Vegt, and W. F. Van Gunsteren. Validation of the 53A6 GROMOS force field. *Eur. Biophys. J.*, 34(4):273–284, 2005.

[149] Y. Sugita and Y. Okamoto. Replica-exchange molecular dynamics method for protein folding. *Chem. Phys. Lett.*, 314(1):141–151, 1999.

[150] G. Bussi, A. Laio, and M. Parrinello. Equilibrium free energies from nonequilibrium metadynamics. *Phys. Rev. Lett.*, 96(9):090601, 2006.

[151] V. Leone, F. Marinelli, P. Carloni, and M. Parrinello. Targeting biomolecular flexibility with metadynamics. *Curr. Opin. Struct. Biol.*, 20(2):148–154, 2010.

[152] D. Hamelberg, J. Mongan, and J. A. McCammon. Accelerated molecular dynamics: a promising and efficient simulation method for biomolecules. *J. Chem. Phys.*, 120(24):11919–11929, 2004.

[153] R. W. Hockney, S. P. Goel, and J. W. Eastwood. Quiet high-resolution computer models of a plasma. *J. Comput. Phys.*, 14(2):148–158, 1974.

[154] R. H. Byrd, P. Lu, J. Nocedal, and C. Zhu. A limited memory algorithm for bound constrained optimization. *SIAM J. Sci. Comput.*, 16(5):1190–1208, 1995.

[155] I. G. Tironi, R. Sperb, P. E. Smith, and W. F. van Gunsteren. A generalized reaction field method for molecular dynamics simulations. *J. Chem. Phys.*, 102:5451–5459, 1995.

[156] T. Darden, D. York, and L. Pedersen. Particle mesh Ewald: An Nlog(N) method for Ewald sums in large systems. *J. Chem. Phys.*, 98:10089–10092, 1993.

[157] G. A. Cisneros, M. Karttunen, P. Ren, and C. Sagui. Classical electrostatics for biomolecular simulations. *Chem. Rev.*, 114:779, 2014.

[158] P. P. Ewald. Die berechnung optischer und elektrostatischer gitterpotentiale. *Ann. Phys.*, 64:253–287, 1921.

[159] T. Hill. *An Introduction to Statistical Thermodynamics*. Dover Publications, 1986.

[160] H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, A. DiNola, and J. R. Haak. Molecular dynamics with coupling to an external bath. *J. Chem. Phys.*, 81:3684–3690, 1984.

[161] G. Bussi, D. Donadio, and M. Parrinello. Canonical sampling through velocity rescaling. *J. Chem. Phys.*, 126:014101, 2007.

[162] B. Hess, H. Bekker, H. J. C. Berendsen, and J. G. E. M. Fraaije. LINCS: a linear constraint solver for molecular simulations. *J. Comput. Chem.*, 18(12):1463–1472, 1997.

[163] B. Hess. P-LINCS: A parallel linear constraint solver for molecular simulation. *J. Chem. Theory Comput.*, 4:116–122, 2008.

[164] S. Miyamoto and P. A. Kollman. SETTLE: An analytical version of the SHAKE and RATTLE algorithm for rigid water models. *J. Comput. Chem.*, 13:952–962, 1992.

[165] J.-P. Ryckaert, G. Ciccotti, and H. J. C. Berendsen. Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *J. Comput. Phys.*, 23(3):327–341, 1977.

[166] A. Warshel and S. T. Russell. Calculations of electrostatic interactions in biological systems and in solutions. *Q. Rev. Biophys.*, 17(03):283–422, 1984.

[167] D. Bashford and M. Karplus. pk_a's of ionizable groups in proteins: atomic detail from a continuum electrostatic model. *Biochemistry-US*, 29(44):10219–10225, 1990.

[168] K. A. Sharp and B. Honig. Electrostatic interaction in macromolecules: theory and applications. *Annu. Rev. Biophys. Bio. Chem.*, 19:301–332, 1990.

[169] A. Warshel. Electrostatic energy and macromolecular function. *Annu. Rev. Biophys. Bio. Chem.*, 20:267–298, 1991.

[170] F. Fogolari, A. Brigo, and H. Molinari. The Poisson–Boltzmann equation for biomolecular electrostatics: a tool for structural biology. *J. Mol. Recognit.*, 15(6):377–392, 2002.

[171] M. K. Gilson. Theory of electrostatic interactions in macromolecules. *Curr. Opin. Struct. Biol.*, 5(2):216–223, 1995.

[172] A. M. Baptista. *Theoretical methods for the simulation of proteins at constant pH.* PhD thesis, Instituto de Tecnologia Química e Biológica, Universidade Nova de Lisboa, Lisboa, 1998.

[173] D. Bashford. Macroscopic electrostatic models for protonation states in proteins. *Front. Biosci*, 9:1082–1099, 2004.

[174] J. Warwicker and H. C. Watson. Calculation of the electric potential in the active site cleft due to α-helix dipoles. *J. Mol. Biol.*, 157(4):671–679, 1982.

[175] I. Klapper, R. Hagstrom, R. Fine, K. Sharp, and B. Honig. Focusing of electric fields in the active site of Cu-Zn superoxide dismutase: Effects of ionic strength and amino-acid modification. *Proteins Struct. Funct. Bioinf.*, 1(1):47–59, 1986.

[176] A. M. Baptista and C. M. Soares. Some theoretical and computational aspects of the inclusion of proton isomerism in the protonation equilibrium of proteins. *J. Phys. Chem. B*, 105:293–309, 2001.

[177] P. J. Martel, A. M. Baptista, and S. B. Petersen. Protein electrostatics. *Biotechnol. Annu. Rev.*, 2:315–372, 1996.

[178] J. Antosiewicz and D. Porschke. The nature of protein dipole moments: experimental and calculated permanent dipole of alpha-chymotrypsin. *Biochemistry-US*, 28(26):10072–10078, 1989.

[179] P. Beroza, D. R. Fredkin, M. Y. Okamura, and G. Feher. Protonation of interacting residues in a protein by a monte carlo method: application to lysozyme and the photosynthetic reaction center of rhodobacter sphaeroides. *Proc. Natl. Acad. Sci. USA*, 88(13):5804–5808, 1991.

[180] P. Beroza and D. A. Case. Including side chain flexibility in continuum electrostatic calculations of protein titration. *J. Phys. Chem.*, 100(51):20156–20163, 1996.

[181] L. Stryer. *Biochemistry.* W. H. Freeman & Company, New York, fourth edition, 1995.

[182] J. B. Matthew. Electrostatic effects in proteins. *Annu. Rev. Biophys. Bio. Chem.*, 14:387–417, 1985.

[183] B. H. Honig, W. L. Hubbell, and R. F. Flewelling. Electrostatic interactions in membranes and proteins. *Annu. Rev. Biophys. Bio. Chem.*, 15(1):163–193, 1986.

[184] L. V. Chernomordik, E. Leikina, V. Frolov, P. Bronk, and J. Zimmerberg. An early stage of membrane fusion mediated by the low pH conformation of influenza hemagglutinin depends upon membrane lipids. *J. Cell Biol.*, 136(1):81–93, 1997.

[185] D. A. Redfern and A. Gericke. pH-dependent domain formation in phosphatidylinositol polyphosphate/phosphatidylcholine mixed vesicles. *J. Lipid Res.*, 46(3):504–515, 2005.

[186] J. M. Carrozzino and M. G. Khaledi. pH effects on drug interactions with lipid bilayers by liposome electrokinetic chromatography. *J. Chromatogr. A*, 1079(1):307–316, 2005.

[187] G. Cevc, A. Watts, and D. Marsh. Titration of the phase transition of phosphatidylserine bilayer membranes. effects of pH, surface electrostatics, ion binding, and head-group hydration. *Biochemistry-US*, 20(17):4955–4965, 1981.

[188] Y. Zhou and R. M. Raphael. Solution pH alters mechanical and electrical properties of phosphatidylcholine membranes: relation between interfacial electrostatics, intramembrane potential, and bending elasticity. *Biophys. J.*, 92(7):2451–2462, 2007.

[189] J. H. Wang. Coupling of proton flux to the hydrolysis and synthesis of ATP. *Annu. Rev. Biophys. Bioeng.*, 12(1):21–34, 1983.

[190] G. Deckers-Hebestreit and K. Altendorf. The F_0F_1-type atp synthases of bacteria: Structure and function of the F_0 complex. *Annu. Rev. Microbiol.*, 50:791–824, 1996.

[191] B. E. Schultz and S. I. Chan. Structures and proton-pumping strategies of mitochondrial respiratory enzymes. *Annu. Rev. Biophys. Biomol. Struct.*, 30:23–65, 2001.

[192] C. von Ballmoos, A. Wiedenmann, and P. Dimroth. Essentials for ATP synthesis by F_0F_1 ATP synthases. *Annu. Rev. Biochem.*, 78:649–672, 2009.

[193] J. O'M. Bockris and A. K. N. Reddy. *Modern Electrochemistry: an introduction to an interdisciplinary area.* Plenum Press, USA, London, 1970.

[194] S. McLaughlin and H. Harary. The hydrophobic adsorption of charged molecules to bilayer membranes: a test of the applicability of the stern equation. *Biochemistry-US*, 15(9):1941–1948, 1976.

[195] R. C. MacDonald and A. D. Bangham. Comparison of double layer potentials in lipid monolayers and lipid bilayer membranes. *J. Membrane Biol.*, 7(1):29–53, 1972.

[196] N. BenTal, B. Honig, R. M. Peitzsch, G. Denisov, and S. McLaughlin. Binding of small basic peptides to membranes containing acidic lipids: Theoretical models and experimental results. *Biophys. J.*, 71(2):561–575, AUG 1996.

[197] B. J. Yoon and A. M. Lenhoff. Computation of the electrostatic interaction energy between a protein and a charged surface. *J. Phys. Chem.*, 96(7):3130–3134, 1992.

[198] D. J. Roush, D. S. Gill, and R. C. Willson. Electrostatic potentials and electrostatic interaction energies of rat cytochrome b5 and a simulated anion-exchange adsorbent surface. *Biophys. J.*, 66(5):1290–1300, 1994.

[199] F. L. B. da Silva, D. Bogren, O. Söderman, T. Åkesson, and B. Jönsson. Titration of fatty acids solubilized in cationic, nonionic, and anionic micelles. theory and experiment. *J. Phys. Chem. B*, 106(13):3515–3522, 2002.

[200] O. Söderman, B. Jönsson, and G. Olofsson. Titration of fatty acids solubilized in cationic and anionic micelles. calorimetry and thermodynamic modeling. *J. Phys. Chem. B*, 110(7):3288–3293, 2006.

[201] F. S. Lee, Z. T. Chu, and A. Warshel. Microscopic and semimicroscopic calculations of electrostatic energies in proteins by the polaris and enzymix programs. *J. Comput. Chem.*, 14:161–185, 1993.

[202] I. Eberini, A. M. Baptista, E. Gianazza, F. Fraternali, and T. Beringhelli. Reorganization in apo-and holo-β-lactoglobulin upon protonation of glu89: Molecular dynamics and pk_{a} calculations. *Proteins Struct. Funct. Bioinf.*, 54:744–758, 2004.

[203] A. M. Baptista, P. J. Martel, and C. M. Soares. Simulation of electron-proton coupling with a Monte Carlo method: Application to cytochrome c_3 using continuum electrostatics. *Biophys. J.*, 76:2978–2998, 1999.

[204] V. H. Teixeira, C. C. Cunha, M. Machuqueiro, A. S. F. Oliveira, B. L. Victor, C. M. Soares, and A. M. Baptista. On the use of different dielectric constants for computing individual and pairwise terms in Poisson-Boltzmann studies of protein ionization equilibrium. *J. Phys. Chem. B*, 109:14691–14706, 2005.

[205] W. Rocchia, S. Sridharan, A. Nicholls, E. Alexov, A. Chiabrera, and B. Honig. Rapid grid-based construction of the molecular surface and the use of induced surface charge to calculate reaction field energies: Applications to the molecular systems and geometric objects. *J. Comput. Chem.*, 23:128–137, 2002.

[206] L. Li, C. Li, S. Sarkar, J. Zhang, S. Witham, Z. Zhang, L. Wang, N. Smith, M. Petukh, and E. Alexov. DelPhi: a comprehensive suite for DelPhi software and associated resources. *BMC Biophys.*, 5(1):9, 2012.

[207] S. W. Chiu, M. Clark, V. Balaji, S. Subramaniam, H. L. Scott, and E. Jakobsson. Incorporation of surface tension into molecular dynamics simulation of an interface: A fluid phase lipid bilayer membrane. *Biophys. J.*, 69(4):1230–1245, 1995.

[208] M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, J. A. Montgomery, Jr., T. Vreven, K. N. Kudin, J. C. Burant, J. M. Millam, S. S. Iyengar, J. Tomasi, V. Barone, B. Mennucci, M. Cossi, G. Scalmani, N. Rega, G. A. Petersson, H. Nakatsuji, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, M. Klene, X. Li, J. E. Knox, H. P. Hratchian, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, P. Y. Ayala, K. Morokuma, G. A. Voth, P. Salvador, J. J. Dannenberg, V. G. Zakrzewski, S. Dapprich, A. D. Daniels, M. C. Strain, O. Farkas, D. K. Malick, A. D. Rabuck, K. Raghavachari, J. B. Foresman, J. V. Ortiz, Q. Cui, A. G. Baboul, S. Clifford, J. Cioslowski, B. B. Stefanov, G. Liu, A. Liashenko, P. Piskorz, I. Komaromi, R. L. Martin, D. J. Fox, T. Keith, M. A. Al-Laham, C. Y. Peng, A. Nanayakkara, M. Challacombe, P. M. W. Gill, B. Johnson, W. Chen, M. W. Wong, C. Gonzalez, and J. A. Pople. Gaussian 03, Revision C.02. Gaussian, Inc., Wallingford, CT, 2004.

[209] C. I. Bayly, P. Cieplak, W. D. Cornell, and P. A. Kollman. A well behaved electrostatic based method using charge restraints for deriving atomic charges: The RESP model. *J. Phys. Chem.*, 97:10269–10280, 1993.

[210] D. van der Spoel, E. Lindahl, B. Hess, G. Groenhof, A. E. Mark, and H. J. C. Berendsen. GROMACS: Fast, flexible, and free. *J. Comput. Chem.*, 26:1701–1718, 2005.

[211] B. Hess, C. Kutzner, D. van der Spoel, and E. Lindahl. GROMACS 4: Algorithms for highly efficient, load-balanced, and scalable molecular simulation. *J. Chem. Theory Comput.*, 4:435–447, 2008.

[212] N. Schmid, A. P. Eichenberger, A. Choutko, S. Riniker, M. Winger, A. E. Mark, and W. F. Van Gunsteren. Definition and testing of the GROMOS force-field versions 54A7 and 54B7. *Eur. Biophys. J.*, 40:843–856, 2011.

[213] W. Huang, Z. Lin, and W. F. van Gunsteren. Validation of the GROMOS 54A7 Force Field with Respect to β-Peptide Folding. *J. Chem. Theory Comput.*, 7(5):1237–1243, 2011.

[214] J. Hermans, H. J. C. Berendsen, W. F. van Gunsteren, and Johan P. M. Postma. A consistent empirical potential for water-protein interactions. *Biopolymers*, 23:1513–1518, 1984.

[215] P. E. Smith and W. F. van Gunsteren. Consistent dielectric properties of the simple point charge and extended point charge water models at 277 and 300 K. *J. Chem. Phys.*, 100:3169–3174, 1994.

[216] C. N. Schutz and A. Warshel. What are the dielectric "constants" of proteins and how to validate electrostatic models? *Proteins: Struct. Funct. Genet.*, 44(4):400–417, 2001.

[217] G. Archontis and T. Simonson. Proton binding to proteins: A free-energy component analysis using a dielectric continuum model. *Biophys. J.*, 88(6):3888–3904, 2005.

[218] M. Machuqueiro, S. R. R. Campos, C. M. Soares, and A. M. Baptista. Membrane-induced conformational changes of kyotorphin revealed by molecular dynamics simulations. *J. Phys. Chem. B*, 114(35):11659–11667, 2010.

[219] M. K. Gilson, K. A. Sharp, and B. Honig. Calculating the eletrostatic potential of molecules in solution: Method and error assessment. *J. Comput. Chem.*, 9:327–335, 1987.

[220] R. A. Marcus and N. Sutin. Electron transfers in chemistry and biology. *Biochem. Biophys. Acta*, 811:265–322, 1985.

[221] A. K. Churg, R. M. Weiss, A. Warshel, and T. Takano. On the action of cytochrome c: correlating geometry changes upon oxidation with activation energies of electron transfer. *J. Phys. Chem.*, 87(10):1683–1694, 1983.

[222] W. J. Allen, J. A. Lemkul, and D. R. Bevan. GridMAT-MD: A grid-based membrane analysis tool for use with molecular dynamics. *J. Comput. Chem.*, 30(12):1952–1958, 2009.

[223] P. F. F. Almeida and W. L. C. Vaz. Lateral diffusion in membranes. *Handbook of biological physics*, 1:305–357, 1995.

[224] J. Bjerrum, G. Schwarzenbach, and L. G. Sillen. *Stability Constants*. Special publication (Chemical Society (Great Britain)), no. 7. Chemical Society, Great Britain, London, 1958.

[225] W. D. Kumler and J. J. Eiler. The acid strength of mono and diesters of phosphoric acid. the n-alkyl esters from methyl to butyl, the esters of biological importance, and the natural guanidine phosphoric acids. *J. Am. Chem. Soc.*, 65:2355–2361, 1943.

[226] H. Träuble and H. Eibl. Electrostatic effects on lipid phase transitions: membrane structure and ionic environment. *Proc. Natl. Acad. Sci. USA*, 71:214–219, 1974.

[227] J. J. H. Shin and C. J. R. Loewen. Putting the pH into phosphatidic acid signaling. *BMC Biol.*, 9:85, 2011.

[228] S. Loew, E. E. Kooijman, and S. May. Increased pH-sensitivity of protein binding to lipid membranes through the electrostatic-hydrogen bond switch. *Chem. Phys. Lipids*, 169:9–18, 2013.

[229] A. I. P. M. de Kroon. Metabolism of phosphatidylcholine and its implications for lipid acyl chain composition in *Saccharomyces cerevisiae*. *Biochem. Biophys. Acta*, 1771(3):343–352, 2007.

[230] H. Eibl and A. Blume. The influence of charge on phosphatidic acid bilayer membranes. *Biochem. Biophys. Acta, Biomembr.*, 553(3):476–488, 1979.

[231] A. Blume and J. Tuchtenhagen. Thermodynamics of ion binding to phosphatidic acid bilayers. titration calorimetry of the heat of dissociation of DMPA. *Biochemistry-US*, 31:4636–4642, 1992.

[232] E. E. Kooijman, D. P. Tieleman, C. Testerink, T. Munnik, Dirk T. S. Rijkers, K. N. J. Burger, and B. de Kruijff. An electrostatic/hydrogen bond switch as the basis for the specific interaction of phosphatidic acid with proteins. *J. Biol. Chem.*, 282:11356–11364, 2007.

[233] J. S. Hub, B. L. Groot, H. Grubmüller, and G. Groenhof. Quantifying artifacts in ewald simulations of inhomogeneous systems with a net charge. *J. Chem. Theory Comput.*, 10:381–390, 2014.

[234] S. McLaughlin. The electrostatic properties of membranes. *Annu. Rev. Biophys. Bio. Chem.*, 18:113–136, 1989.

[235] M. M. Reif and P. H. Hünenberger. Computation of methodology-independent single-ion solvation properties from molecular simulations. IV. optimized Lennard-Jones interaction parameter sets for the alkali and halide ions in water. *J. Chem. Phys.*, 134:144104, 2011.

[236] H. Butt, K. Graf, and M. Kappl. *Physics and chemistry of interfaces*. John Wiley & Sons, 2006.

[237] N. Kučerka, M.-. Nieh, and J. Katsaras. Fluid phase lipid areas and bilayer thicknesses of commonly used phosphatidylcholines as a function of temperature. *Biochem. Biophys. Acta, Biomembr.*, 1808:2761–2771, 2011.

[238] V. Knecht and B. Klasczyk. Specific binding of chloride ions to lipid vesicles and implications at molecular scale. *Biophys. J.*, 104:818, 2013.

[239] C. Hong, P. Tieleman, and Y. Wang. Microsecond molecular dynamics simulations of lipid mixing. *Langmuir*, 30:11993–12001, 2014.

[240] G. M. Ullmann and E.-W. Knapp. Electrostatic models for computing protonation and redox equilibria in proteins. *Eur. Biophys. J.*, 28(7):533–551, 1999.

[241] J. M. Swails and A. E. Roitberg. Enhancing conformation and protonation state sampling of hen egg white lysozyme using pH replica exchange molecular dynamics. *J. Chem. Theory Comput.*, 8(11):4393–4404, 2012.

[242] J. J. Janke, W. F. D. Bennett, and D. P. Tieleman. Oleic acid phase behavior from molecular dynamics simulations. *Langmuir*, 30(35):10661–10667, 2014.

[243] C. Tanford. *Physical chemistry of macromolecules.* John Wiley & Sons, INC., 1961.

[244] J. Wyman. *Binding and linkage: functional chemistry of biological macromolecules.* University Science Books, 1990.

[245] D. Chandler. *Introduction to modern statistical mechanics.* Oxford University Press, USA, 1987.

[246] G. D. Knott. *Interpolating cubic splines.* Birkhäuser Boston, 1999.

[247] J. F. J. Dippy, S. R. C. Hughes, and A. Rozanski. 498. the dissociation constants of some symmetrically disubstituted succinic acids. *J. Chem. Soc.*, pages 2492–2498, 1959.

[248] Chris Neale. Gromacs bug # 1166: g_order is incorrect for unsaturated carbons. `http://redmine.gromacs.org/issues/1166`. (accessed: September 15 2014).

[249] A. Seelig and J. Seelig. Dynamic structure of fatty acyl chains in a phospholipid bilayer measured by deuterium magnetic resonance. *Biochemistry-US*, 13(23):4839–4845, 1974.

[250] J. Seelig and N. Waespe-Sarcevic. Molecular order in cis and trans unsaturated phospholipid bilayers. *Biochemistry-US*, 17(16):3310–3315, 1978.

[251] B. Perly, I. C. P. Smith, and H. C. Jarrell. Effects of the replacement of a double bond by a cyclopropane ring in phosphatidylethanolamines: a deuterium nmr study of phase transitions and molecular organization. *Biochemistry-US*, 24(4):1055–1063, 1985.

[252] L. Guo, J. Y. Har, J. Sankaran, Y. Hong, B. Kannan, and T. Wohland. Molecular diffusion measurement in lipid bilayers over wide concentration ranges: a comparative study. *Chem. Phys. Chem.*, 9(5):721–728, 2008.

[253] R. Macháň and M. Hof. Lipid diffusion in planar membranes investigated by fluorescence correlation spectroscopy. *Biochem. Biophys. Acta, Biomembr.*, 1798(7):1377–1391, 2010.

[254] J. D. Jackson. *Classical electrodynamics.* John Wiley and Sons, INC., 1999.

[255] R. Orij, S. Brul, and G. J. Smits. Intracellular pH is a tightly controlled signal in yeast. *Biochem. Biophys. Acta*, 1810(10):933–944, 2011.

[256] A. K. Hinderliter, J. Huang, and G. W. Feigenson. Detection of phase separation in fluid phosphatidylserine/phosphatidylcholine mixtures. *Biophys. J.*, 67(5):1906, 1994.

[257] U. Kwolek, W. Kulig, P. Wydro, M. Nowakowska, T. Róg, and M. Kepczynski. Effect of phosphatidic acid on biomembrane: Experimental and molecular dynamics simulations study. *J. Phys. Chem. B*, 119(31):10042–10051, 2015.

[258] S. T. Whitten and B. E García-Moreno. pH dependence of stability of staphylococcal nuclease: evidence of substantial electrostatic interactions in the denatured state. *Biochemistry-US*, 39(46):14292–14304, 2000.

[259] C. A. Castaneda, C. A. Fitch, A. Majumdar, V. Khangulov, J. L. Schlessman, and B. E. García-Moreno. Molecular determinants of the pK_a values of Asp and Glu residues in staphylococcal nuclease. *Proteins Struct. Funct. Bioinf.*, 77(3):570–588, 2009.

[260] M. Musial-Siwek, A. Karabadzhak, O. A. Andreev, Y. K. Reshetnyak, and D. M. Engelman. Tuning the insertion properties of pHLIP. *BBA-Biomembranes*, 1798(6):1041–1046, 2010.

[261] J. Fendos, F. N. Barrera, and D. M. Engelman. Aspartate embedding depth affects pHLIP's insertion pK_a. *Biochemistry-US*, 52(27):4595–4604, 2013.

[262] D. Weerakkody, A. Moshnikova, M. S. Thakur, V. Moshnikova, J. Daniels, D. M. Engelman, O. A. Andreev, and Y. K. Reshetnyak. Family of pH (low) insertion peptides for tumor targeting. *Proc. Natl. Acad. Sci. USA*, 110(15):5834–5839, 2013.

[263] A. S. Ladokhin and S. H. White. Interfacial folding and membrane insertion of a designed helical peptide. *Biochemistry-US*, 43(19):5782–5791, 2004.

[264] O. Yuzlenko and T. Lazaridis. Interactions between ionizable amino acid side chains at a lipid bilayer–water interface. *J. Phys. Chem. B*, 115(46):13674–13684, 2011.

[265] N. J. Gleason, V. V. Vostrikov, D. V. Greathouse, and R. E. Koeppe. Buried lysine, but not arginine, titrates and alters transmembrane helix tilt. *Proc. Natl. Acad. Sci. USA*, 110(5):1692–1695, 2013.

[266] P. R. Magalhães, M. Machuqueiro, and A. M. Baptista. Constant-pH molecular dynamics study of kyotorphin in an explicit bilayer. *Biophys. J.*, 108(9):2282–2290, 2015.

[267] C. A. Carvalheda, S. R. R. Campos, and A. M. Baptista. The effect of membrane environment on surfactant protein C stability studied by constant-pH molecular dynamics. *J. Chem. Inf. Model.*, 2015. DOI: 10.1021/acs.jcim.5b00076.

[268] A. Panahi and C. L. Brooks III. Membrane environment modulates the pK_a values of transmembrane helices. *J. Phys. Chem. B*, 119(13):4601–4607, 2015.

[269] A. Kyrychenko, V. Vasquez-Montes, M. B. Ulmschneider, and A. S. Ladokhin. Lipid headgroups modulate membrane insertion of pHLIP peptide. *Biophys. J.*, 108(4):791–794, 2015.

[270] G. Wiedman, W. C. Wimley, and K. Hristova. Testing the limits of rational design by engineering pH sensitivity into membrane-active peptides. *BBA-Biomembranes*, 1848:951–957, 2015.

[271] M.-L. Jobin and I. D. Alves. On the importance of electrostatic interactions between cell penetrating peptides and membranes: A pathway toward tumor cell selectivity? *Biochimie*, 107:154–159, 2014.

[272] A. E. Cardenas, R. Shrestha, L. J. Webb, and R. Elber. Membrane permeation of a peptide: It is better to be positive. *J. Phys. Chem. B*, 2015. DOI: 10.1021/acs.jpcb.5b02122.

[273] R. L. Thurlkill, G. R. Grimsley, J. M. Scholtz, and C. N. Pace. pK values of the ionizable groups of proteins. *Protein Sci.*, 15(5):1214–1218, 2006.

[274] G. R. Grimsley, J. M. Scholtz, and C. N. Pace. A summary of the measured pK values of the ionizable groups in folded proteins. *Protein Sci.*, 18(1):247–251, 2009.

[275] D. Bashford and K. Gerwert. Electrostatic calculations of the pK_a values of ionizable groups in bacteriorhodopsin. *J. Mol. Biol.*, 224:473–486, 1992.

[276] J. D. Lear and W. F. DeGrado. Membrane binding and conformational properties of peptides representing the NH_2 terminus of influenza HA-2. *J. Biol. Chem.*, 262(14):6500–6505, 1987.

[277] M. Rafalski, A. Ortiz, A. Rockwell, L. C. Van Ginkel, J. D. Lear, W. F. DeGrado, and J. Wilschut. Membrane fusion activity of the influenza virus hemagglutinin: interaction of HA2 N-terminal peptides with phospholipid vesicles. *Biochemistry-US*, 30(42):10211–10220, 1991.

[278] Y. K. Reshetnyak, O. A. Andreev, U. Lehnert, and D. M. Engelman. Translocation of molecules into cells by pH-dependent insertion of a transmembrane helix. *Proc. Natl. Acad. Sci. USA*, 103(17):6460–6465, 2006.

[279] Y. K. Reshetnyak, M. Segala, O. A. Andreev, and D. M. Engelman. A monomeric membrane peptide that lives in three worlds: in solution, attached to, and inserted across lipid bilayers. *Biophys. J.*, 93(7):2363–2372, 2007.

[280] L. Yao, J. Daniels, A. Moshnikova, S. Kuznetsov, A. Ahmed, D. M. Engelman, Y. K. Reshetnyak, and O. A. Andreev. pHLIP peptide targets nanogold particles to tumors. *Proc. Natl. Acad. Sci. USA*, 110(2):465–470, 2013.

[281] O. A. Andreev, D. M. Engelman, and Y. K. Reshetnyak. Targeting diseased tissues by pHLIP insertion at low cell surface pH. *Front. Physiol.*, 5, 2014.

[282] Y. K. Reshetnyak, O. A. Andreev, M. Segala, V. S. Markin, and D. M. Engelman. Energetics of peptide (pHLIP) binding to and folding across a lipid bilayer membrane. *Proc. Natl. Acad. Sci. USA*, 105(40):15340–15345, 2008.

[283] Berendsen H. J. C. van der Ploeg P. Molecular dynamics simulation of a bilayer membrane. *J. Chem. Phys.*, 76(6):3271–3275, 1982.