UNIVERSIDADE DE LISBOA
FACULDADE DE CIÊNCIAS
DEPARTAMENTO DE QUÍMICA E BIOQUÍMICA

Improving pK_a calculations of membrane inserting amino acids using replica exchange CpHMD simulations

Pedro de Brito Pires Santos Reis

MESTRADO EM BIOQUÍMICA
Especialização em Bioquímica

Dissertação orientada por:
Doutor Miguel Ângelo dos Santos Machuqueiro
Doutor Diogo Ruivo dos Santos Vila Viçosa

2017

To my grandparents, Minhaia and Zé.

Acknowledgments

Firstly, I would like to thank both my awesome supervisors, Diogo Vila Viçosa and Miguel Machuqueiro. I really could not have asked for better people to guide me through this journey. They are always there to answer my questions and, more importantly, to help me realize which questions were worth asking. They suppress my less fortunate ideas swiftly while granting me the freedom to explore and think for myself. I still have plenty to learn and I would be very lucky if given the opportunity to continue to work with them.

During this past year, it was really easy to get out of bed to go to work. I was surrounded by people who love their job and have a great sense of humor, so working time often blended with breaks. We had fun at the office and worked off the clock. For this, I have to thank not only Diogo and Miguel but also Bruno Victor, Paulo Costa, and Tomás Silva. I will not make personal remarks as I have too much to thank each one of them. I will, however, thank them all for being always willing to help and for making the workplace such an enjoyable spot.

Despite not knowing what my thesis is truly about, my family has obviously also made a very significant contribution to it. I have been blessed to have been born in the midst of such an amazing family. They have always encouraged me to pursue my objectives, to be ambitious and have never restrained me in any way. I hope they all know all their love, unconditional support and teachings do not go by unnoticed and that I genuinely value all they ever did for me.

I also need to thank another person, who by now is a member of my family, my girlfriend Joana Fernandes. She is my better half and helps me to see the world in a different perspective. I thank her for fueling my dreams while making them her own.

Finally, I must thank all my friends in which I also include my supervisors, colleagues, and family. They are pivotal in making non-working hours so fun.

Preface

The work presented in this thesis is the natural continuation of my first project in Inorganic and Theoretical Chemistry group. I have joined this group after immediately liking Miguel Machuqueiro and the subject he taught in Computational Biochemistry class. I then looked up for his research interests and they seemed quite interesting although a bit distant from what I had been learning in my Biochemistry course. Thankfully Miguel accepted me as his undergraduate student and for me it was an instant match both personally and scientifically-wise. I did not know in which area I would like to work after graduation until I met Miguel, so I can say for certain that joining his group was a big event in my life. At the time, Diogo Vila–Viçosa and Vitor Teixeira worked with Miguel and the three began to teach me the molecular modeling tools and theoretical concepts. I have joined a new research project which aimed to understand how the pK_a of titrable amino acids varies at the water/lipid interface. This project yielded my first paper and clearly revealed issues present in the protonation sampling. Upon insertion, the titrating amino acid protonation state did not change for long consecutive periods of time. Constant-pH MD with replica exchange (pHRE) was then developed by Diogo, Miguel and Chris Oostenbrink, to fix this protonation sampling problem. I have accompanied this development by helping to debug and documenting the code. In the present work, we are trying to improve the obtained results by applying this newly developed method. Interacting titrable amino acids seemed like a bigger challenge worth exploring after confirming that we are indeed speeding up our calculations.

After this first project ended, and now as a graduate student, Miguel assigned me the investigation of whether globular highly charged molecules were affected by ionic strength treatment. In a previous work, they had shown that lipid bilayers could only reproduce experimental data if simulated with a proper ionic strength treatment. This project proved to be very time-consuming due to the used system size and number of simulations needed, ending up taking some time from my master's work. In this work I got to work at ITQB with the mastermind behind the CpHMD method António Baptista and a senior group member of the Molecular Simulation Laboratory Sara Campos.

In between running and analyzing simulations from these projects, I have made some contributions to other ventures. My colleague Tomás Silva was investigating the possibility of reducing the long-range cut off to improve performance, and I ran some simulations for him. I also made some preliminary tests for a future investigation to develop a quick method to detect pan-assay interference compounds. During the past year, I have grown an interest in web-development partially due to Miguel's assignment of updating the information on our group website. I got a bit carried away and ended up making a brand new custom website which turned out to be a very didactic endeavor. Miguel also asked me to start

running some long umbrella sampling simulations for a future project of a drug inserting into a buried catalytic center.

I would say I had a productive year for a master student, however to be fair I must state that I owe much of it to the constant support of Miguel and Diogo.

Abstract

pH is one of the most important solution parameters since it influences the properties of solute molecules with labile hydrogen atoms. It plays a major role in most biochemical processes by, among other, inducing protein conformational changes and affecting protein–lipid interactions. Constant-pH molecular dynamics (CpHMD) methods have been used to model such systems since they are able to correctly capture the conformational/protonation coupling. To investigate the pK_a values of titrable amino acids at the water/membrane interface, a previous CpHMD study [1] have shown that, upon insertion, the titrable sites are prone to adopt a neutral state. In that work, we encountered some difficulties to sample ionized conformations in inserted regions with CpHMD, since most residues retained their neutral state upon insertion/desolvation in the timescale of our simulations.

When facing kinetic traps in molecular dynamics simulations, enhanced sampling techniques are a widely used solution. Since our sampling problems are related with a favored protonation state, we implemented a pH-based replica exchange (pHRE)[2]. In this method, a unique pH value is assigned to each simulation replica and attempts to exchange the simulated pH value are periodically performed between replica pairs. The acceptance criterion is dictated by the difference between the exchanging pairs of pH values and protonation states of titrable sites.

Here, we have used the pHRE methodology, a newly developed method to calculate insertion, and more rigorous criteria to define the acceptable protonation sampling, to provide a more accurate description of the membrane influence on the pK_a profiles of titrable amino acids. A more thorough characterization of protein–membrane interactions, membrane deformation and solvation effects is obtained by using a cutoff based insertion method. Since in pHRE, due to replica mixing, all pH values sample similar insertion regions, a larger amount of inserted conformations in the ionized state are obtained. To further improve pHRE sample capability, a high frequency of exchange attempts should be selected. Our efficient pHRE results outperformed previous CpHMD ones, granting more sampling in less simulation time. In the future, pHRE will eventually replace CpHMD as our go-to method to study pH dependent phenomena.

Keywords: replica exchange, pentapeptides, titrable amino acids, constant-pH molecular dynamics, pK_a profiles, membrane interface

Resumo

O pH é um dos parâmetros fisiológicos mais importantes, influenciando as propriedades dos solutos, alterando a distribuição da ocupação de protões lábeis. Caso estes se encontrem em grupos químicos relevantes para a estabilidade conformacional, uma mudança no estado de protonação poderá conduzir a uma transição conformacional significativa. Os principais tipos de biomoléculas contêm grupos tituláveis em zonas estruturalmente importantes, logo uma mudança no seu estado de protonação, e consequentemente na carga, afeta a função das biomoléculas. O facto do pH estar intimamente implicado na maioria dos processos bioquímicos ilustra bem o carácter relevante do pH para a vida, estando também envolvido em doenças como o cancro ou o Alzheimer.

Métodos de dinâmica molecular a pH constante (CpHMD) costumam ser utilizados para modelar sistemas onde a captura correcta do acoplamento das conformações com mudanças no estado de protonação é necessária. Vários métodos de CpHMD foram desenvolvidos e implementados nos mais diversos campos de forças. O CpHMD usado neste trabalho denomina-se titulação estocástica devido aos estados de protonação discretos serem amostrados e aceites pelo critério de Metropolis. Este método de CpHMD foi usado para estudar os valores de pK_a de aminoácidos tituláveis na interface de uma membrana lipídica de fosfatidilcolina. A mudança no ambiente electrostático em redor do resíduo dita a variação do seu pK_a. Ao inserir na membrana apolar, os aminoácidos tendem a estabilizar a sua forma neutra, de tal modo que, a nossa metodologia sentiu dificuldades a amostrar suficientes estados ionizados para ser possível calcular o pK_a em zonas mais inseridas com a confiança desejada.

Geralmente quando se estuda um sistema onde figuram barreiras cinéticas difíceis de transpor usando apenas dinâmica molecular, os métodos de amostragem aumentada são uma solução viável. De entre as técnicas mais comuns, o *replica exchange* (RE) parece ser a mais adequada ao nosso sistema. Métodos como *metadynamics* ou *umbrella sampling*, poderiam auxiliar no aumento da amostragem de estados inseridos contudo provavelmente representam desafios superiores em termos de implementação no contexto do CpHMD. Assim sendo, desenvolvemos o nosso método de RE baseadas em pH (pHRE), complementando o nosso método de CpHMD, tal como já tinha sido desenvolvido por outros grupos na literatura. Nesta metodologia, cada réplica é simulada a um pH único e trocas entre pares de pHs são testadas periodicamente. O critério de aceitação da troca baseia-se na diferença entre os valores de pH dos pares de simulações, bem como na diferença entre estados de protonação dos grupos tituláveis.

Neste trabalho, aplicámos a nova metodologia de pHRE aos pentapéptidos anteriormente estudados, numa tentativa de obter valores de pK_a dos resíduos tituláveis em zonas inseridas que não tenham sido possíveis de determinar face à falta de amostragem. O objectivo foi concluído com sucesso uma vez

que o método de pHRE permite que todos os valores de pH amostrem conformações com inserções semelhantes. Os valores de pH com preferência para o estado carregado, mais moroso de amostrar em configurações inseridas, e que mais dificilmente atingiriam estas inserções em simulações de CpHMD, conseguem desta forma aumentar a qualidade da amostragem nestas zonas. A melhoria da amostragem está portanto diretamente relacionada com a mistura das réplicas que, por sua vez, poderá ser maximizada diminuindo o espaçamento entre os valores de pH escolhidos ou aumentando a periodicidade das tentativas de troca entre replicados. Contudo, aumentar a frequência de trocas poderá ter um efeito secundário nefasto, tendo em conta que caso uma troca entre replicados afete a probabilidade da troca seguinte, pode favorecer ou desfavorecer um estado de protonação particular. Os melhores resultados deste trabalho foram obtidos simulando 4 valores de pH e com um τ_{RE} de 20 ps. Porém, dado que o sistema apenas tinha um resíduo a titular, todas as simulações de pHRE apresentaram uma boa eficiência a nível de mistura de réplicas, apesar do τRE de 100 ps divergir em algumas zonas de inserção das restantes, o que sugere que ainda não terá convergido. De notar, que em dois dos resíduos não foi possível acelerar a amostragem, ficando estas excepções a dever-se ao leque de valores de pH simulados ter sido demasiado limitado.

Um novo método para definir o nível de inserção foi também introduzido neste trabalho. Este foi comparado com os métodos alternativos mais usuais que utilizam todos os lípidos da membrana ou apenas o mais próximo. Usando todos os lípidos, a inserção medida será tão sobrestimada quanto a deformação local causada na membrana durante esse processo. Calculando a inserção apenas com o lípido mais próximo, é possível contornar a deformação, mas um novo problema surge quando o desvio que é provocado nesta referência única não representa corretamente a irregularidade introduzida na membrana. O novo método aqui proposto, faz uso de um *cutoff* até ao qual todos os lípidos são contabilizados. Desta forma, é possível obter valores de inserção correlacionados com o estado de solvatação e minimizar as imprecisões associadas com a troca do lípido mais próximo. Apesar da melhor descrição das interações na zona da interface, este modo de calcular a inserção tem pouca influencia nos perfis de pK_{a}.

O pHRE consegue claramente melhorar a amostragem do CpHMD, bem como os perfis de pK_{a} associados. No futuro, esta metodologia de amostragem aumentada será adoptada, mesmo para estudar processos bioquímicos dependentes do pH cujas limitações de amostragem não sejam tão evidentes quanto as existentes na interface de uma membrana.

Palavras-chave: troca de replicas, pentapéptidos, aminoácidos tituláveis, dinâmica molecular a pH constante, perfis de pK_{a}, interface membranar

Contents

List of Figures

List of Tables

List of Abbreviations

BFGS	Broyden–Fletcher–Goldfarb–Shannol
CE	continuum electrostatics
CpHMD	constant-pH molecular dynamics
DMPC	dimyristoylphosphatidylcholine
GRF	generalized reaction field
LPBE	linearized Poisson–Boltzmann equation
LRA	linear response approximation
PA	phosphatidic acid
PB	Poisson–Boltzmann
PBC	periodic boundary conditions
PC	phosphatidylcholine
PE	phosphatidylethanolamine
pHLIP	pH Low Insertion Peptide
pHRE	replica exchange constant-pH molecular dynamics
PME	particle mesh Ewald
MC	Monte Carlo
MD	molecular dynamics
MM	molecular mechanics
SPC	simple point charge
RE	replica exchange
RF	reaction field
vdW	van der Waals

CHAPTER 1

Introduction

The concentration of hydrogen ions is one of the most important parameters in any solution. It affects the properties of solute molecules with labile hydrogens and since 1909, we call pH to the negative logarithm of $[H^+]$ [3]. In biological systems, the intracelular pH value is tightly regulated because most biomolecules structure and function are heavily influenced by it [4, 5]. Its regulation is necessary to all biochemical processes which depend on protein folding [6, 7, 8], protein-substrate binding [9, 10], protein-protein interactions [11, 12], lipid bilayers properties [13, 14, 15] or lipid-protein interactions [16, 17]. The importance of pH is also well illustrated by its role in many severe human diseases such as cancer [18], Alzheimer's [19], inflammatory bowel disease [20], or celiac disease [21].

The pH effect on protein structures arises from the (de)protonation of the titrable residues. The protonation equilibrium of those sites is influenced by the surrounding electrostatics and as the site charge changes so do the local interactions. In proteins, there are usually multiple titrable residues which influence each other preferred protonation distributions. This complex coupling between interacting residues can induce, and be induced by, large conformational transitions which are difficult to predict[22, 23].

Despite being generally much smaller than proteins and not having as many titrable sites, the study of lipids presents a different challenge. Phosphatidylcholine (PC) and phosphatidylethanolamine (PE), the most abundant lipids in most celullar and subcellular membranes, are zwitterionic and only titrate at pH values far from the physiological pH range [24, 25]. However, biological membranes in order to maintain homeostasis have a dynamically changing number of lipids and proteins, including anionic lipids which can be pH-sensitive [26]. Even a small fraction of phosphatidic acid (PA) in a PA/PC mixture leads to a significant membrane ionization variation within the physiological range, capable of altering mechanical properties such as area per lipid and membrane thickness [27, 28].

Since biological membranes have a very complex asymmetric composition, it is difficult to experimentally follow events of protonation/deprotonation of molecules interacting and inserting in membranes. Optical techniques like fluorescence are the most used to study both model and cell membranes although these only work if the titrating site incites a conformational change which disturbs a fluorescent probe [29]. With the pHLIP peptide, Engelman's group was able to measure an insertion pK value, however, they could not assign it to a specific pK_a value of an individual group due to experimental limitations[30, 31].

To overcome such limitations, over the years several methods have been developed to compute pK_a values. Many of the current methods to study protein titration behavior are based on numerically solving the Poisson-Boltzmann equation [32] and thus treating the bulk aqueous and salt environment as a continuum dielectric. This seemingly severe approximation actually generates quite good solvation free energies and pK_a values for small, rigid molecules [33, 34]. These explicitly incorporate the effect of both desolvation and site-site interactions and are usually obtained from sampling protonation states by Monte Carlo (MC). There are other approaches to determine protonation free energies that also take advantage of single structures and simplified electrostatic models namely the generalized Born [35, 36, 37] or Langevin dipoles [38, 39]. However, as a single conformation is used, rough structural reorganization and protonation-conformation coupling are accounted for in the used dielectric [40, 41]. Some averaging over protein configurations can be done by applying any of these methods to snapshots from an explicit solvent MD simulation, even when the conformational ensemble is biased by the chosen protonation states [42, 43]. This is the primordial ancestor of the so-called constant-pH molecular dynamics (CpHMD), a group of methods which explicitly capture the conformation/protonation coupling.

The first CpHMD method sampled from the grand canonical ensemble and was developed by Mertz and Pettitt in 1994 [44]. They were able to simulate the acid-base equilibrium between acetic acid and water. A few years later, Baptista *et al.* [45] introduced a method based on a potential of mean force using continuous protonation states. This method explored the complementarity between MD and CE calculations: MD samples conformations at fixed protonation, while CE determines the protonation state at fixed conformation. Even though the original version used coarse CE and MD electrostatics, these could easily be replaced by more rigorous methods. Later, Borjesson and Hunenberger proposed the acidostat method [46]. This method was entirely based on MM/MD and also used fractional protonation states. However, the simulated titration curves did not have the expected Henderson–Hasselbalch shape and the theoretical soundness of the method was questioned [47]. A more recent CpHMD method using continuous titration was advanced by Brooks and co-workers [48]. Despite its convergence problems, this λ dynamics [49] inspired method is able to predict pK_a values in good agreement with experimental values.

Meanwhile, CpHMD methods with discrete protonation state models were developed. In 2002 Baptista *et al.* presented the stochastic titration method [50] that inherited the CE and MD complementarity from implicit titration [45]. After a protonation state change yielded from a PB/MC calculation, a brief solvent equilibration is performed with the solute fully restrained. Similar rationale was used by other groups that improved the performance with simpler electrostatic models [51, 52]. Another even less computationally expensive method is the linear response approximation (LRA). This CpHMD alternative defines the pK_a value of a site as the average between the pK_a values of the protonated and deprotonated states obtained from regular MD sampling [53, 54]. Despite having been successfully used to determine pK_a values in proteins [53, 41, 55], this method needs the use of a higher dielectric constant in order to overcome the sampling limitations denoted in the lack of overlap between the two confomational ensembles.

In a previous work, we have applied the stochastic CpHMD method to alanine-based pentapeptides originally designed by Pace's group [56, 57] (Ala_2–X–Ala_2, where X is a pH titrating residue) in an effort to understand the influence of a lipid membrane on the pK_a values of titrable amino acids [1]. Our results showed that assigning a fixed protonation state to membrane interacting titrable amino acids is a biased and erroneous approach. It also became clear that the protonation sampling of less solvent exposed peptidic conformations was poor as protonation states in these tended to be kept constant in consecutive integration steps. To overcome this limitation enhanced sampling techniques are needed.

Generalized ensemble methods [58, 59] such as simulated tempering, [60] and temperature replica exchange molecular dynamics [61] are usually used to avoid kinetic trapping in constant temperature MD simulations. Khandogin *et al.* [62] successfully improved the pK_a convergence of their CpHMD method by incorporating a temperature replica-exchange scheme. However, as our system features a lipid bilayer, increasing the temperature would change system properties like area per lipid and order parameter. Therefore, a pH based RE as the one presented by Roiberg's group [63] was implemented in our CpHMD method. In this methodology, each replica is run at its own pH value and pH exchanges between pairs of replicas are periodically attempted. The criteria of acceptance of this RE scheme is based on the difference between simulated pH values and the protonation states of the titrable sites (see Eq.2.47).

In this work, we apply our newly developed pHRE methodology to the membrane interacting pentapeptides to study titrable amino acids insertion dependent pK_a profiles. To improve the robustness of said profile, we introduce a novel insertion method to account for both the solvation effect an the membrane deformation, and a strict pK_a calculation criteria to better define well sampled regions. Compared to CpHMD, pHRE sampling of inserted conformations in the ionized state is enhanced due to all pH values being able to sample similar regions. Consequently, more extensive pK_a profiles were attained with pHRE.

CHAPTER 2

Methods

2.1 Molecular Mechanics

Molecular modelers wish to unravel the mysteries that molecules present them. To do so, they rely on empirical force fields, relatively simple models that try to capture molecular properties and interactions. Hamiltonians comprise all energy contributions to the system and can be divided into kinetic and potential energy parts. The kinetic energy part is discretized as the sum of the kinetic energies of all particles of the system. In molecular mechanics, the potential energy is described by a functional form that can have as much as thousands of parameters in an effort to reproduce the true Hamiltonian [64]. Force fields do not try to incorporate electronic phenomena and are used to investigate a timescale which quantum chemists can only dream of. Since they disregard individual electronic motion, they can not predict certain properties, while others, like polarizability, are often deliberately ignored to speed up the simulations. Force fields are empirical which means they intend to reproduce a given collection of experimental data and it is unreasonable to ask them to model every property correctly. That is why several flavors were developed, each one with its own parameterization philosophy and applicability [65, 66, 67, 68].

This work was carried out with GROMOS 54A7 force field [69] using GROMACS 4.0.7 software package [68]. GROMOS is a performance and biomolecular focused force field which is parameterized to reproduce thermodynamic properties of small polar molecules and solvation energies of amino acids [70]. GROMOS is an united-atom force field meaning that all non-polar hydrogens are collapsed into the nearest atom which becomes slightly bigger and reflects their charges. For example, an aliphatic methyl group becomes a single CH_3 atom type in a united-atom representation. Contrarily, in an all-atom force field like CHARMM or AMBER, there would be 4 particles of two different atom types. Atom types are a key concept in most force fields as they are the molecular mechanics building blocks allowing for bigger molecules to be designed with pieces from smaller ones. The GROMOS force field has a simple functional form facilitating the extension to other moieties,

$$V(\mathbf{r}) = V^{bond}(\mathbf{r}) + V^{angle}(\mathbf{r}) + V^{proper}(\mathbf{r}) + V^{improper}(\mathbf{r}) + V^{vdW}(\mathbf{r}) + V^{elec}(\mathbf{r}) \tag{2.1}$$

The potential energy is dependent on the atom coordinates (r) and can be thought as having two separate contributions: the bonded and the nonbonded interactions. The bonded interactions are the sum of covalent bonds, bond angle, proper and improper dihedral angle terms. The nonbonded interactions are

the sum of van der Waals (vdW) and electrostatic (elec) interactions. To this physical interactions one can add special terms, for example, to restrain certain properties during a simulation such as the distance of two atoms.

The covalent bond stretching interactions are given by the following equation where b_n is a bond from all existing N_b bonds and $\mathbf{r}_i$, $\mathbf{r}_j$ are the two atomic positions of each bond,

$$V^{bond}(\mathbf{r}; K_b, b_0) = \sum_{n=1}^{N_b} \frac{1}{4} K_{b_n} \left[b_n^2 - b_{0_n}^2 \right]^2 \tag{2.2}$$

$$b_n = r_{ji} = \sqrt[2]{(\mathbf{r}_i - \mathbf{r}_j)^2} \tag{2.3}$$

The bond force constant K_b and the ideal bond length b_0 are intrinsic parameters of each GROMOS bond type, experimentally determined from spectroscopy and X-ray diffraction, respectively [71, 72].

The potential energy of the angle bending interactions is also characterized by two parameters, a force constant K_θ and the ideal bond angle θ_0,

$$V^{angle}(\mathbf{r}; K_\theta, \theta_0) = \sum_{n=1}^{N_\theta} \frac{1}{2} K_{\theta_n} \left[\cos(\theta_n) - \cos(\theta_{0_n}) \right]^2 \tag{2.4}$$

Each θ angle value of the N_θ system angles contributes to the total angle interactions potential energy. These have also been ajusted to experimental data [71, 72]. Both potential energies of covalent bonds and angle interactions are modified versions of the Hooke's law simple harmonic oscillator potential, chosen to improve the calculations performance.

Torsional proper dihedral-angle interactions potentials can be expressed as a cosine series expansion. In GROMOS a one term default is implemented, even though several terms can be added to obtain correct torsional angles energy profiles,

$$V^{proper}(\mathbf{r}; K_\phi, \delta, m) = \sum_{n=1}^{N_\phi} K_{\phi_n} \left[1 + \cos(\delta_n) \cos(m_n \phi_n) \right] \tag{2.5}$$

This periodic function exhibits m minima and energy barriers of height K_ϕ. ϕ is the actual torsion angle value, δ is known as phase shift or phase factor and determines where the minima will be.

Improper torsions are needed to model molecules which have out-of-plane bending motions or to maintain planar structures (e.g. benzene). These type of interactions can also be described using a cosine series expansion, however in GROMOS a harmonic potential is used,

$$V^{improper}(\mathbf{r}; K_\xi, \xi_0) = \sum_{n=1}^{N_\xi} \frac{1}{2} K_{\xi_n} \left[\xi_n - \xi_{0_n} \right]^2 \tag{2.6}$$

In this force field there are only three types of improper torsions defined, applicable to planar groups, tetrahedral centers and heme iron. If one desires to model a compound that has a different kind of improper torsion, one has to define its force constant K_ξ and its ideal improper dihedral angle ξ_0 which are easily obtained with high level quantum calculations.

As aforementioned, the nonbonded interactions are categorized into van der Waals and electrostatic interactions. In this force field, covalently bound atoms and their adjacent atoms do not contribute to the non-bonded part of the Hamiltonian, the reason being that they were already accounted for in the bonded interactions. Third neighbours atoms, covalently bound, that are part of or bound to aromatic rings are also excluded (1–4 exclusions), so is it easier for the improper dihedral interactions to keep those in one plane.

Van der Waals interactions take into account two opposite forces between two interacting atoms. Pauli exclusion principle explains the repulsive contribution. The attractive contribution is due to instantaneous dipoles arising from fluctuations in the electron clouds and bears the name of the man who first explained their origin in 1930, Fritz London [73]. These phenomena need to be modeled by a simple expressions in molecular mechanics. Lennard–Jones 12– 6 is a popular and lightweight van der Waals potential function,

$$V^{vdW}(\mathbf{r};C12,C6) = \sum_{pairsi,j} \left(\frac{C12_{ij}}{r_{ij}^{12}} - \frac{C6_{ij}}{r_{ij}^{6}} \right) \tag{2.7}$$

There are only two adjustable parameters in this equation: $C12_{ij}$ and $C6_{ij}$ which describe, respectively, the faster decaying repulsive forces and the longer ranged attractive interactions between two atoms i and j. Based on the interacting atom types there are three C12 parameter values: one for neutral atoms and two for the usually stronger polar and ionic interactions. GROMACS also features the more accurate Buckingham's potential function but it leads to slower simulations and they were not used in GROMOS parameterization.

Electrostatics are probably the most powerful intermolecular forces. In GROMOS, the charge distribution is treated as partial atomic charges, intended to reproduce the moiety electrostatic potential profile. There are numerous methods to calculate partial atomic charges and since they are not an experimental observable quantity nor can they be directly calculated from a wavefunction, indirect comparisons with a variety of either experimental or quantum mechanics results are acceptable. The charges used in GROMOS were iteratively optimized to reproduce densities and heats of vaporization of a range of pure liquids, and free enthalpies of solvation of small compounds in cyclohexane and water.

The electrostatic interaction between two charge particles can be described by Couloumb's law. However, the calculation of all possible interactions between pairs of charges is too computationally expensive. In GROMOS, beyond a certain cutoff distance r_{rf}, a long-range electrostatic contribution from a homogeneous dielectric reaction field is used. This has an implication on the electrostatic potential function, where besides the cutoff included directly interacting charges there is a attenuation from interactions with an induced field of a continuous dielectric medium outside said r_{rf} cutoff distance,

$$V^{elec}(\mathbf{r};q) = \sum_{pairsi,j} \frac{q_i q_j}{4\pi\varepsilon_0\varepsilon_r} \left[\frac{1}{r_{ij} + k_{rf} r_{ij}^2 - c_{rf}} \right] \tag{2.8}$$

$$k_{rf} = \frac{1}{r_{rj}^3} \frac{\varepsilon_{rf} - \varepsilon_r}{2\varepsilon_{rf} + \varepsilon_r} \tag{2.9}$$

$$c_{rf} = \frac{1}{r_{rf}} \frac{3\varepsilon_{rf}}{2\varepsilon_{rf} + \varepsilon_r} \tag{2.10}$$

where ε_0 is the dielectric constant of vacuum, ε_r and ε_{rf} are relative dielectric constants of the atoms medium and outside the r_{rf} cutoff. The last term ensures in c_{rf} that the electrostatic potential is zero for atoms in the cutoff boundary, thereby reducing the cutoff energy discontinuity problem. Another limitation of the RF methods is their assumption of a spherical mean-field beyond a cutoff, which in inhomogeneous systems, such as lipid bilayers, migh be problematic.

In particle mesh Ewald (PME) two rapidly convergent series replace the slow conditionally convergent long range interactions potential [74]. This is achieved by adding a charge distribution (e.g. Gaussian) to the original point charge which will subsequently be deducted. The added charge distribution is of equal magnitude and opposite sign, weakening the coulombic interactions so that their potential is calculated by a direct summation in the real space. In the reciprocal space, a fast Fourier transform

algorithm subtracts a charge distribution symmetric to the added one as well as the contribution of its interaction with the original point charge [74, 75, 76, 77]. Nonetheless, there are also some PME inherent problems. PME introduces an artificial periodicity in the system and requires a full system neutralization which may be attained by adding an uniform background charge density that induces some artifacts in inhomogeneous systems [78, 79].

Given that both approaches have their limitations and neither of these alternatives is theoretically unquestionable when it comes to modeling lipid bilayers, one should be pragmatic in choosing which one to use. GROMOS has been parameterized with a RF method and it has been previously reported that RF derived methods have been able to reproduce experimental lipidic phases of pure PC and PA/PC membranes [80, 28, 81, 82].

In this work, we used a RF derived method, GRF, that adds the ionic strength I in an implicit manner. [83] Given K different charge species in the system,

$$k^2 = \frac{F^2}{\varepsilon_0 \varepsilon_{rf} RT} \sum_{i=1}^{K} c_i z_i^2 \tag{2.11}$$

$$k_{rf} = \frac{1}{r_{rj}^3} \frac{(\varepsilon_{rf} - \varepsilon_r)(1 + kr_{rf}) + \frac{1}{2}\varepsilon_{rf}(kr_{rf})^2}{(2\varepsilon_{rf} + \varepsilon_r)(1 + kr_{rf}) + \varepsilon_{rf}(kr_{rf})^2} \tag{2.12}$$

$$c_{rf} = \frac{1}{r_{rf}} \frac{3\varepsilon_{rf}(1 + kr_{rf} + \frac{1}{2}\varepsilon_{rf}(kr_{rf})^2)}{(2\varepsilon_{rf} + \varepsilon_r)(1 + kr_{rf}) + \varepsilon_{rf}(kr_{rf})^2} \tag{2.13}$$

where R is the ideal gas constant, T is the absolute temperature, F is the Faraday's constant and c_i and z_i are the molar concentration and charge number of ionic species, respectively. This way, GRF includes the effect of bulk ionic strength which is of the utmost importance when simulating charged lipid membranes [28, 76, 80]. Furthermore, we have used a modified version of GROMACS in which the ion concentration I is an adjustable external parameter such as temperature or pressure [84].

We have now described in detail how to compute the total energy of a system with a force field. Yet, there is an important detail missing, how to model the most important and abundant molecule: water. The water model used in GROMOS parameterization was SPC [85]. SPC is a fairly simple three point charge water model, therefore, being more computationally efficient than most alternatives. It does not display any dummy atoms to improve its electrostatic profile unlike the more accurate OPC or TIP5P [86, 87].

Before simulations can be performed, initial configurations of the system are needed. This initial choice is important since it can dictate the success or failure of the simulation. High-energy interactions may cause instabilities that can escalate into the simulation collapse, frequently portraited as an atom bursting from the simulation box or a unphysical angle or dihedral torsion. These interactions can often be mitigated by performing an energy minimization procedure.

There are several minimization algorithms to choose from. Relative speed and robustness of the methods are two of the most important factors when deciding which ones to use. Virtually all energy minimization procedures intended for molecular mechanics make use of derivatives. The first derivative indicates the direction of the closest local minimum while from the second derivative one can infer inflection points.

Steepest descent is a first-order minimization method since it uses only the first derivative [88]. This iterative method is conceptually intuitive: a new point is calculated, being accepted if its energy is smaller than the current one. A new point could be calculated by a step of arbitrary size along the

direction force, yet GROMACS uses the previously calculated potential energy V and corresponding forces F to determine the following step,

$$r_{n+1} = r_n + \frac{F_n}{\max|F_n|} h_n \tag{2.14}$$

where r is a is a vector of all 3N coordinates and $\max|F_n|$ is the largest absolute value of the force components. h is the maximum displacement, updated every step according to,

$$h_{n+1} = \begin{cases} 1.2h_n, & \text{if } V_{n+1} < V_n. \\ 0.2h_n, & \text{otherwise.} \end{cases} \tag{2.15}$$

The algorithm can run for a certain user specified number of steps or it can stop if the convergence criteria is met.

Limited-memory Broyden–Fletcher–Goldfarb–Shannol (l-BFGS) is a quasi-Newton optimization method [89]. This family of minimizers uses approximations to the inverse Hessian matrix of second derivatives that in molecular mechanics system sizes becomes time-consuming to calculate. These methods converge faster than first-order optimizers when near a minimum, in contrast when far, the minimization could move to higher energy saddle points. To overcome this limitation, a more robust optimizer like steepest descents is usually employed before these more complex and unstable methods.

Energy minimization identifies a minimum energy individual configuration. However, to accurately predict thermodynamic properties, the configurations distribution must be sampled correctly. Thus, the obtained information from this configuration is limited since most biologically relevant systems sample multiple minima. This is where computer simulation methods such as Monte Carlo and molecular dynamics come into play, generating representative systems configurations from a Markov chain [90, 91].

2.2 Molecular Dynamics

Molecular dynamics (MD) is a sampling method from which macroscopic thermodynamic properties can be obtained from simulation time averages. Atoms trajectories are defined by Newton's equations of motion numerically solved using the forces determined by molecular mechanics [92]. Assuming potentials are pairwise additive, the mutual potential energy U_i of a single particle i with all other particles of the system is,

$$U_i = \sum_{i \neq j}^{N} U_{ij} \tag{2.16}$$

from which we can obtain the force action on i by differentiating U_i with respect to the coordinates of particle i,

$$F_i = -\nabla_{r_i} U_i \tag{2.17}$$

Once F_i is found, solving Newton's equation of motion would give us the particle new momentum p, acceleration a, velocity v and position r. Note that the trajectory of all particles in the system are governed by the Hamiltonian, hence,

$$F_i = -\frac{\partial \mathscr{H}}{\partial r_i} = \frac{\partial p_i}{\partial t} = m_i a_i = m_i \frac{\partial v_i}{\partial t} = m_i \frac{\partial^2 r_i}{\partial t^2} \tag{2.18}$$

There are several methods to numerically solve differential equations based on finite differences. They solve the equations step by step in time. Since the force calculation is more computationally expensive

than updating variables to advance a step forward in time, suitable methods to optimize performance should avoid the force evaluation per update step. Open methods satisfy this condition as they predict the next vector position exclusively with quantities already found in previous steps. These methods are less accurate than Gear algorithms which perform a force evaluation in every step, yet the magnitude of this error is hindered by the molecular mechanics approximations incorporated in the potential energy calculation. Leapfrog and Verlet are simpler and more efficient algorithms [93]. The leapfrog scheme is subjectively more intuitive than Verlet's albeit being equivalent,

$$v_{t+\frac{\Delta t}{2}} = v_{t-\frac{\Delta t}{2}} + \frac{F_t}{m}\Delta t \tag{2.19}$$

$$r_{t+\Delta t} = r_t + v_{t+\frac{\Delta t}{2}}\Delta t \tag{2.20}$$

As shown in the above formulation, positions r are determined using the previously determined velocities v. Another important feature of this algorithm is that it can be easily modified to accommodate different equations of motion and to include coupling of external temperature and pressure bath parameters.

Temperature T and pressure P are macroscopic properties that can be accurately calculated,

$$T = \frac{1}{3Nk_B}\sum_{i=1}^{N}\frac{|p_i|^2}{m_i} \tag{2.21}$$

$$P = \frac{1}{V}\left[Nk_BT - \frac{1}{3}\sum_{i=1}^{N}\sum_{j=i+1}^{N} r_{ij}F_{ij}\right] \tag{2.22}$$

As the total momentum p of all particles and interaction forces F are already being computed, there is little additional cost required. Integration errors and truncated long range forces, among others, make these properties drift as the simulation evolves. Multiple methods have been developed to amend this effect. Berendsen has developed an efficient weak coupling methodologies for both temperature and pressure [94] that relate to time according to,

$$\frac{\partial T}{\partial t} = \frac{T_0 - T}{\tau} \tag{2.23}$$

$$\frac{\partial P}{\partial t} = \frac{P_0 - P}{\tau} \tag{2.24}$$

where T_0 and P_0 are the reference bath temperature and pressure, respectively.

The Berendsen thermostat affects the motion of the particles by applying a scaling factor λ every step,

$$\lambda = \sqrt{1 + \frac{\Delta t}{\tau_T}\left(\frac{T_0}{T} - 1\right)} \tag{2.25}$$

τ_T determines how tight is the coupling and dependes on τ, the total heat capacity of the sytem C_V and the total number of degrees of freedom N_{df},

$$\tau_T = \frac{\tau N_{df} k_B}{2C_V} \tag{2.26}$$

This method modifies the equations of motions, adjusting the system temperature too abruptly. As a consequence, the generated configurations are not being sampled from the proper ensemble (NPT). An additional term improves the Berendsen thermostat to ensure a correct kinetic energy distribution. This improved method, named velocity rescaling thermostat, produces a correct ensemble [95].

As expected, Berendsen's pressure coupling method is analogous to the aforementioned [94], making use of the following scaling matrix μ,

$$\mu = \sqrt{1 + \frac{\Delta t}{\tau_P}\beta\left(P - P_0\right)} \tag{2.27}$$

where β is the isothermal compressibility of the system.

It is of paramount importance to be able to control temperature and pressure. Even though ensembles are convertible, usually when simulating biological systems conditions, the configurations should be sampled from the isothermal–isobaric ensemble (NPT). In NPT conditions, all energies are allowed while the number of particles in the system, the average pressure and temperature are constant.

Up until now, we have reviewed how to get proper trajectory ensembles from force field Hamiltonians. In order to get reliable estimates of our systems thermodynamic properties, we also need a sufficient amount of configurations from the most representative states. Therefore, the better the calculations performance, the better the sampling. Most molecular dynamics setups employ a myriad of tricks to boost acceleration, some already mentioned.

Periodic boundary conditions minimize edge effects of a finite system while enabling simulations to have fewer particles and still experience bulk fluid forces. In these systems there are no boundaries, a simulation box is surrounded by images of itself, as a particle leaves the limit of a box it enters the same box on the opposite side. When using PBC, it is necessary to guarantee that the cell size is sufficiently large to avoid artificial periodicity artifacts and prevent molecules to interact with their own images.

Though cubic cells are the most simple and intuitive, in order to get some speed up, in simulations of globular systems usually more elaborate geometries are used such as rhombic dodecahedron or truncated octahedron. These are more similar to a sphere than to a cube, granting an enhanced performance as fewer solvent molecules are needed to fill the box. In particular, the rhombic dodecahedron is the smallest and most regular space-filling unit cell available in GROMACS. It saves about 30% of CPU-time when compared to a cubic box of similar size.

Another common speed up technique is the use of cutoff schemes. As previously stated, GROMOS force fields have been parameterized with a cutoff to deal with electrostatic interactions. Interactions larger than the cutoff are treated as a continuous dielectric medium. With the reaction field method Lennard–Jones potential is simply truncated as it decays rapidly with distance (r^6) and after a few Å of distance its contribution becomes insignificant. Truncating potentials is a trade off between improving the simulation's performance and generating discontinuity instabilities. Not having to evaluate most of the system interactions is essencial, since the number of long range interacting particles is exceptionally bigger than the shorter-ranged. These long range interactions are already being treated in a distance independent manner, with no clear benefit in determining the exact distance between that kind of interacting particles. Moreover, these separated particles will not be directly interacting with a considerable amount of consecutive integration steps. From the above premises arises the twin-range cutoff scheme [96]. A short R_s and a long cutoff R_l lengths are defined. At a user defined constant number of steps, interacting particle pairs within R_s are saved in a so-called neighbor list. Particles with an interaction distance between R_s and R_l have their forces computed and stored until further neighbor list update. The positions and interaction forces of the neighbor list atoms are estimated at every simulation step. Beyond the R_l cutoff a long range electrostatic contribution from a homogeneous dielectric is used, as explained in the previous section when discussing GRF.

Neighbor searching performance can be further improved by adopting charge groups. Charge groups are described as a small set of nearby atoms with a preferably neutral net charge. The distance from a

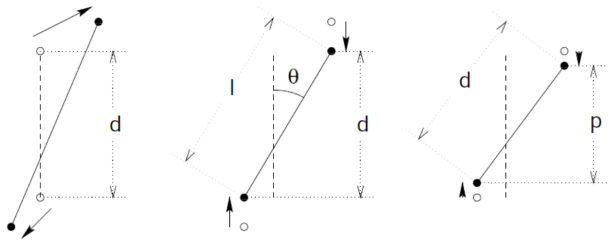

Figure 2.1: LINCS algorithm scheme representation for bond length constraint (adapted from [100])

charge group is defined by its geometrical center. When updating the neighbor list either all or no charge group atoms are added. This prevents an irrealistic charge creation arising from the cutoff scheme usage.

Finally, there is one last obvious way to expedite simulations. Modern computers have multiple CPU cores, it is only natural to try maximizing their workload. Domain decomposition is a parallelization algorithm allowing simulations to be run on distinct cores concurrently [97, 98]. Taking advantage of the fact that particle interactions are confined to its adjacent region, the simulation box can be partitioned into separate cells (domain decomposition cells). Each cell is then assigned to a single processor which will integrate the equations of motion for the particles in it confined. The partitioning is made during the neighbor searching, hence as neighbor searching is based on charge groups, these will also be the domain decomposition units. Coordinates from adjacent cells often have to be communicated so forces dependent on particles from foreign cells can be determined. Communication of particle positions is also required to perform molecular constraints. To constrain bonds and angles is to force them to adopt a specified value. This is a common practice to reduce system instability and gain additional performance. Since some parts of a molecule can reside in different processors, a parallel constraint algorithm is required. GROMACS uses the P-LINCS algorithm, the parallel version of the LINCS algorithm [99]. LINCS non-iteratively corrects bond lengths after an unconstrained update. It is a two-stepped algorithm (see Figure 2.1). First it projects the new bonds on the old bonds and then a correction is applied to the bond rotation [100].

2.3 Monte Carlo

Markov chain Monte Carlo simulations (MCMC), in constrast to MD, yield system properties with no temporal dependence. In spite of the total energy of the system in MC having no kinetic contribution, both techniques can sample from the same NVT ensemble. New configurations can be explored displacing system particles by a small and random 3D distance. All new configurations less energetic than their predecessor are accepted while energy increments can be accepted according to its Boltzmann factor, in the so-called Metropolis criterion [101],

$$P_{acceptance} = \begin{cases} 1 & \text{if } \Delta V < 0 \\ e^{-\Delta V / k_B T} & \text{otherwise.} \end{cases} \tag{2.28}$$

ΔV, the change in potential energy, can be determined with a force field or by alternative models methods. As a matter of fact, MC methods are used by a multitude of fields from traffic simulations to cryptography. In constant-pH molecular dynamics, this methodology is used to sample protonation states (see Protonation Equilibria section).

2.4 Continuum Electrostatics

Explicit water molecules should be used in order to account for correct solvation effects, however, most of the simulation time is spent with these molecules. In some cases, it is advantageous to use a simplified description of the solvent effect. Continuum electrostatics methods provide a welcomed performance improvement while still retaining a somewhat realistic description of an aqueous solvent environment. Two of the most popular CE methods are generalized Born [35] and Poisson–Boltzmann (PB) [102]. The former is faster and allows for the solvation energy to be determined without needing to calculate the electrostatic potential, while the latter is a more accurate solution and the one used in this work. In PB derived methods, the solvent is modeled as a continuum dielectric and the solute is regarded as a rigid body of a usually uniform dielectric constant. Despite the severity of the approximation, this model is widely used to estimate pK_a values, redox potentials[103], and it often yields good equilibrium solvation energies [104, 105].

In a medium of uniform dielectric constant ε, the variation in the electrostatic potential ϕ is related to the charge density ρ by the Poisson equation,

$$\bigtriangledown^2\phi(r) = -\frac{\rho(r)}{\varepsilon_0\varepsilon} \tag{2.29}$$

Poisson equations based on experimental data to reproduce the relation between polarization and electric field. This second-order differential equation reduces to Coulomb's law when applied to set of point charges in a constant dielectric. However, if the dielectric magnitude varies with position, Coulomb's law does not hold true. The Poisson equation does not account for the distribution of ions in solution with electric potential. This ion distribution is a result of their natural thermal motion and electrostatic repulsions, described by a Boltzmann distribution. The Poisson equation can be modified to incorporate these effects, resulting in the Poisson-Boltzmann equation,

$$\bigtriangledown\cdot\varepsilon(r)\bigtriangledown\phi(r) - k'sinh[\phi(r)] = -4\pi\rho(r) \tag{2.30}$$

where the factor $4\pi\varepsilon_0$ has been discarted as in reduced electrostatic units. Debye factor κ is related to k' by the following form,

$$\kappa^2 = \frac{\kappa'^2}{\varepsilon} = \sqrt{\frac{8\pi e^2 I}{\varepsilon k_B T}} \tag{2.31}$$

If the ionic strenght is zero, then the κ' becomes null and the non-linear Poisson–Boltzmann equation reverts to its original Poisson equation. An alternative form of this non-linear equation shown in 2.30 can be obtained by expanding the hyperbolic sine function,

$$\bigtriangledown\cdot\varepsilon(r)\bigtriangledown\phi(r) - \kappa'\phi(r)\Big[1+\frac{\phi(r)^2}{6}+\frac{\phi(r)^4}{120}+\ldots\Big] = -4\pi\rho(r) \tag{2.32}$$

By taking only the first term of the Taylor series expansion, one can obtain the linearized Poisson–Boltzmann equation (LPBE),

$$\bigtriangledown\cdot\varepsilon(r)\bigtriangledown\phi(r) - \kappa'\phi(r) = -4\pi\rho(r) \tag{2.33}$$

The PB equations are typically solved by finite-difference or boundary-element numerical methods [106, 107]. DelPhi employs a finite difference method to solve this equation [108]. This method superimposes a cubic lattice onto the system and each grid point is assigned with values for electrostatic potential, charge density, dielectric constant and ionic strength. Atomic charges are usually taken directly from

the force field and are mapped on the adjacent grid points so that the charge fraction of each grid point reflects their distance to the atomic center. In a cube of side h, the potential ϕ at a grid point j is given by,

$$\phi_j = \frac{\sum \varepsilon_i \phi_i + 4\pi \frac{q_j}{h}}{\sum \varepsilon_i + \kappa_j'^2 f(\phi_j)} \tag{2.34}$$

where q_j is the charge associated with the grid point j. $f(\phi_j)$ takes the value of 1 for the LPBE, while in the non-linear form it is equivalent to the Taylor series of equation 2.32. The featured summations are relative to the potentials ϕ and dielectric contants ε of the neighbour grid points. This feature implies that these potentials need to be obtained by iteratively solving this equation for all the grid points until they converge.

To assign the necessary dielectric constant values, nodes are divided into solvent and solute points. Typically, in biological systems the solvent is treated with a high ε while to the solute is assigned a low dielectric constant. Water has a dielectric of 80 while macromolecules are usually considered to have a ε close to 2. This dielectric should account for the molecules conformation flexibility and polarizability.

Once the electrostatic potential is found, the electrostatic energy is calculated from,

$$G_{elect} = \frac{1}{2} \sum_j q_j \phi_j \tag{2.35}$$

This energy can then be used to study the protonation equilibria of molecules by calculating the protonation free energy ΔG_{prot} as shown in the following section.

2.5 Protonation Equilibria

The protonation free energy is a bit more tricky to calculate than most free energies. Unlike the solvation free energy which can be determined simply by subtracting the G_{elect} in vacuum to the one in solution, in order to determine the ΔG_{prot} one needs to apply a thermodynamic cycle (Figure 2.2) as originally introduced by Bashford and Karplus [109]. A protein site protonation free energy can be written as,

$$\begin{aligned} \Delta G_P^o(A \rightarrow AH) &= \Delta G_{sol}^o(A \rightarrow AH) + \Delta G_{sol \rightarrow P}^o(AH) - \Delta G_{sol \rightarrow P}^o(A) \\ &= \Delta G_{sol}^o(A \rightarrow AH) + \Delta\Delta G_{sol \rightarrow P}^o(A \rightarrow AH) \end{aligned} \tag{2.36}$$

$$\tag{2.37}$$

And as pK_a can be described by,

$$\mathrm{p}K_\mathrm{a} = \frac{\Delta G}{2.3 k_B T} \tag{2.38}$$

$$\begin{array}{ccc} \mathrm{A^{sol}} + \mathrm{H^+} & \xrightarrow{\Delta G^\circ_{\mathrm{sol}}(\mathrm{A}\rightarrow\mathrm{AH})} & \mathrm{AH^{sol}} \\ \Big\downarrow \Delta G^\circ_{\mathrm{sol}\rightarrow\mathrm{P}}(\mathrm{A}) & & \Big\downarrow \Delta G^\circ_{\mathrm{sol}\rightarrow\mathrm{P}}(\mathrm{AH}) \\ \mathrm{A^P} + \mathrm{H^+} & \xrightarrow{\Delta G^\circ_{\mathrm{P}}(\mathrm{A}\rightarrow\mathrm{AH})} & \mathrm{AH^P} \end{array}$$

Figure 2.2: Thermodynamic cycle of protein and model compounds.

one can obtain the pK_a value of a titrable site in a protein from [109],

$$\begin{aligned} \mathrm{p}K_\mathrm{a}(P) &= \mathrm{p}K_\mathrm{a}(sol) + \frac{1}{2.3k_BT}\Delta\Delta G^o_{sol\to P}(AH \to A) \\ &= \mathrm{p}K_\mathrm{int} + \frac{1}{2.3k_BT}\Delta G^o_{interact}(P) \end{aligned} \tag{2.39}$$

(2.40)

where pK_a(sol) refers to the model compound pK_a value in solution (pK_mod). As a protein usually has more than a titrating site, it is beneficial to separate the pH-independent contributions from the interacting titrating sites. pK_int is simply the pK_a(P) when all other titration sites are in a reference state, incorporating the pK_mod and free energy contributions from solvation and interactions with non-titrating sites. The $\Delta G^o_{interact}(P)$ then reflects the energy contribution from the interactions between the titrating sites in the protein, depending on their ionization state. This free energy can be obtained by summation of all ionized interacting sites pairs ij contribution calculated from the electrostatic potential ΔW_{ij} as explained in the previous section.

To calculate the free energy difference between a state **s** and a reference state one needs to take into account all those different sites,

$$\Delta G^o(\mathbf{s}) = -2.3k_BT\sum_i^{N_s} s_i\gamma_i \mathrm{p}K_{\mathrm{int,i}} + \sum_i^{N_s}\sum_{j\neq i}^{N_s} s_i s_j \Delta W_{ij} \tag{2.41}$$

where γ_i is the formal charge of a site i from all existing sites N_s. s_i represents the site charge state. It equals 0 if the site i is neutral and 1 if it is charged.

This ΔG^o energy can then be used to determine the probability p of a charge state (**s**) at a certain pH value,

$$p(\mathbf{s}) = \frac{\exp\Big(-\beta\Delta G^o(\mathbf{s}) - 2.3\sum_i s_i\gamma_i \mathrm{pH}\Big)}{\sum_{s'}\exp\Big(-\beta\Delta G^o(\mathbf{s'}) - 2.3\sum_i s'_i\gamma_i \mathrm{pH}\Big)} \tag{2.42}$$

where β is $\frac{1}{k_BT}$. As the number of sites increases so does the computational cost to determine $p(\mathbf{s})$ and for a regular protein it is not feasible to explicitly determine all possible charge states. However, the probability distribution of **s** can be estimated from MC sampling, and the new trial protonation states are accepted according to Equation 2.28 where,

$$\Delta V = \exp\Big(-2.3\sum_i^{N_s} s_i\gamma_i[\mathrm{p}K_{\mathrm{int,i}} - \mathrm{pH}] + \sum_i^{N_s}\sum_{j\neq i}^{N_s} s_i s_j \Delta W_{ij}\Big) \tag{2.43}$$

2.6 Constant pH Molecular Dynamics

The most common approach attemping the treatment of pH in MM/MD simulations is to use fixed protonation states. These states try to be representative of the pH of interest, yet when this pH is close to the pK_a value of a group, this approximation breaks since different protonation states are almost equally probable. As mentioned in a previous section, PB/MC calculations enable us to study protonation events in rigid conformations. Constant-pH MD methods make use of the complementary between MM/MD conformational sampling and PB/MC protonation sampling to be able to capture the protonation/conformation coupling present in most biological systems.

In this work, we use the stochastic titration method developed by Baptista *et al.* [50]. The algorithm of this method can be divided into three steps:

- **PB/MC** The assignment of the protonation states of the titrable sites are made according to those yielded by the last performed MC step.

- **MM/MD of Solvent** A short MD simulation is run with the solute constrained so the water molecules can adapt to the new charge configuration. This step corrects the energy instabilities introduced by changes in protonation states.

- **MM/MD** A segment of production MD simulation is performed. The last system conformation is selected for the following PB/MC step.

2.7 Replica Exchange Constant pH Molecular Dynamics

Replica Exchange is an enhanced sampling technique designed to overcome kinetic traps in MD simulations. It is commonly used with temperature although in this work we use it to exchange between pH values. The method can be very useful to overcome sampling limitations when titrating residues in confined locations, like when inserted in a lipid bilayer. Replica exchange is conceptually simple: two independent replicas of the same system are running at different conditions, in this case at different pH values. To aid sampling, swaps between the configurations of the two systems are attempted at regular τ_{RE} intervals (in this implementation). This exchange process needs to comply with the detailed balance condition on the transition probability $k_{x\to x'}$,

$$w(x)k_{x\to x'} = w(x')k_{x'\to x} \tag{2.44}$$

where x and x' are the conformations from two distinct replicas i, j and $w(x)$ is the weight factor for the state x. $w(x) \propto p(x)$ and since the replicas are non-interacting, $w(x)$ in the generalized ensemble is given by the product of Boltzmann factors for each replica,

$$w(x) = \exp(-\beta\mathscr{H}_i)\exp(-\beta\mathscr{H}_j) \tag{2.45}$$

As in the grand canonical ensemble of protons the Hamiltonian $\mathscr{H}$ is defined by,

$$\mathscr{H} = K + V + k_B T \mathrm{pH} \ln(10) N^P \tag{2.46}$$

N^P is the number of protonated titrable residues. By applying some basic algebra to Equations 2.44, 2.45 and 2.46 we get the exchange rate Δ,

$$\Delta = \frac{k_{x\to x'}}{k_{x'\to x}} = \exp\Big(\ln(10)(\mathrm{pH}_i - \mathrm{pH}_j)(N_x^P - N_{x'}^P)\Big) \tag{2.47}$$

And yet again, we can use the Metropolis criterion to accept the swap trials with probability $\min(1,\Delta)$.

The generalized ensemble for replica exchange consists of multiple replicas of the original system in the canonical ensemble at different pH values so that there is exactly one replica at each pH value. Replica exchange can be used with any number of replicates, so long as configuration exchanges are performed in a pairwise fashion to ensure overall correct sampling. The number of titrating sites also affects this exchange probability since this method will be applied to CpHMD and N^P will vary along the simulation.

2.8 Simulation Settings

The recently developed CpHMD-L method [24, 110] was used to simulate eight alanine based pentapeptides in which the middle amino acid was changed to titrable (Asp, Cys, His, Lys, Glu, Tyr). The C-terminus and N-terminus were also allowed to titrate in sets of pentalanine peptides simulations. In simulations with other titrable sites, both N- and C-termini were capped with an acetyl or amino groups, respectively. For each pentapeptide, we simulated three and four pH values around their experimental pK_a in water. These systems were simulated for 80 ns production, after 20 ns of equilibration (see Equilibration section in Appendix). In this work, all simulations were performed with GROMOS 54A7 force field [69, 111, 24] and a modified version of GROMACS package version 4.0.7 to include ionic strenght as an external parameter [112, 68, 84]. All MM/MD, PB/MC and CpHMD settings were adapted from Teixeira *et al.* [1] Simulations were performed with replica exchange rate τ_{RE} set to 20 ps and 100 ps while the CpHMD τ_{prt} was 20 ps.

MM/MD settings

The starting conformations for all pentapeptides simulations were taken from the already published CpHMD simulations [1]. The system was built from a previously equilibrated lipid bilayer of 128 DMPC molecules, solvated in a tetragonal box with ∼4 700 SPC water molecules. A 3-step energy minimization was employed. First, 10 000 steps of steepest-descent unconstrained minimization. Followed by 2 000 steps of the low-memory Broyden-Fletcher-Goldfarb-Shanno algorithm. Finishing with ∼ 50 steps of steepest-descent with all bonds constrained. The initialization followed, with a 4-step procedure where position restraints were applied and gradually reduced. MD simulations of 100, 150, 200, and 250 ps with constraints on 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), respectively.

An integration step of 2 fs was used to solve equations of motion. A 8/14 Å twin-range cutoff scheme was implemented with the neighbor list being updated every 5 steps. Long range electrostatics were treated with a generalized reaction field [83] with a relative dielectric constant of 54 [113] and an ionic strength of 0.1 M [84].

The temperature was kept constant at ∼ 300 K with v-rescale thermostat [95]. Separate couplings were applied to solute and solvent and a relaxation time of 0.1 ps was used. The Berendsen barostat [94] set the pressure at 1 bar with an isotropic isothermal compressibility of 4.5×10^{-5} bar^{-1} and a relaxation time of 2.0 ps. All waters were constrained using SETTLE algorithm [114]. All other bonds were constrained with the P-LINCS algorithm [99].

PB/MC settings

DelPhi Version 5.1 [108, 115] was used for the PB calculations using the partial charges taken from ref 103 and radii derived from the Lennard-Jones parameters of GROMOS 54A7 force field [69, 111]. The molecular surface was defined with a probe of radius 1.4 Å and the ion exclusion layer was 2.0 Å. Ionic strength was set to 0.1 M. To ensure continuity in the membrane surface, 5% of the box vector dimension was added in the x and y direction [24]. A dielectric constant of 2 was used for peptide and membrane and 80 for water. Periodic boundary conditions were explicitly applied in both x and y directions for the potential calculation in a coarser grid with relaxation parameters of 0.2 and 0.75 for the linear and nonlinear PB equations, respectively. A cutoff of 25 Å was used to calculate the background contributions and pairwise interaction [24]. The convergence threshold value was set to 0.01 $k_B\ T/e$.

Calculations were done on a cubic grid of 61 grid points and a two step focusing [116]. The coarser grid spacing was $\sim$ 1Å, four times the size of the focus grid of $\sim$ 0.25Å.

The MC sampling was performed using 10^5 MC cycles at 300 K with the PETIT program version 1.6 [117].

2.9 Analysis

2.9.1 pK_a Calculations

Calculating pK_a values of titrable groups from a CpHMD simulation is a fairly simple exercise since it is only necessary to fit the sampled protonation states from the different pH values to a Hill equation,

$$pH = \mathrm{p}K_a + n \log\left(\frac{[A^-]}{[AH]}\right) \tag{2.48}$$

where n is the Hill coefficent. When only one group is titrating, $n = 1$ and we obtain the well-know Henderson–Hasselbalch equation. However, in systems such as the ones used in this work, in which there is a molecule with titrating groups inserting into a lipid membrane, we need a new approach to capture these interactions. In this new method, the pK_a values are measured along the membrane normal where the pentapeptides were fractioned into 0.5 Å slabs according to their insertion [1]. The different criteria to measure the insertion will be discussed in detail in the next section. The pK_a values for each slab are then calculated using a Henderson–Hasselbalch fit, since the pentapeptides only have one titrable group.

The fractioning of conformations by insertion slabs can lead to insufficient sampling in some less populated regions. In our previous work, we calculated the insertion and the resulting slabs with a less stringent criteria which resulted in data with limited sampling, that we have now corrected. In each slab, a replica with 50 ionized conformations is considered to have sufficient sampling. A pH value is included in the pK_a calculations, if at least two replicas have sufficent sampling (not to be confused with RE replicas where each replica has a different pH). Previously, we have included pH values when the sum of ionized conformations over all replicas was above 50. This new criterion ensures that the sampling comes from more than one replicas. To have good estimations of inserted pK_a values one needs slabs with titration information from at least two pH values. The error bars of the calculated pK_a values were determined using a leave-one-out resampling method applied to the fitting protocol.

2.9.2 Insertion

There are multiple ways to atomically define an insertion into a lipid bilayer. For the inserting molecule, one can choose to use its geometric center, a single atom or the geometric center of a reference group. In this work, we chose to follow the insertion of a single atom of interest within the titrating residues since our insertions are only used to study the properties of these residues. We used the C atoms in the carboxylic groups, N in amines, O in Tyr, S in Cys and Nε in His.

Regarding the lipid bilayer reference for the distance measurements, the most common approach uses the phosphorus atoms average z-positions from the closest lipid monolayer. This procedure fails when membranes exhibit significant local deformation. As pentapeptides insert in the membrane they interact and thus displace several lipid molecules (further discussed in Section 3.2). To overcome the artifacts introduced by these local deformations and better correlate insertion with the solvation of the group, an alternative method was used by Teixeira *et al.* [1] in which only the closest phosphorus atom was

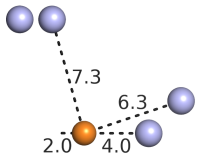

Figure 2.3: Graphical representation of the different insertion criteria where membrane references are depicted in grey and the inserting molecule reference in orange. In this scenario all methods calculate distinct insertion values.

used as the reference. The closest phosphorus atom can be determined using either a 2D (ignoring the z dimension) or a 3D distance, prompting different results as illustrated in Figure 2.3. A 2D xy projection of a non-interacting phosphorus may be closer to the molecule reference than a 3D directly interacting lipid atom.

In the context of this thesis, a new method to measure insertion values of molecules in a lipid bilayer was developed by our group. We propose that a cutoff around our reference atom is the best way to determine its insertion as it enables more phosphorus atoms to be included in the calculation, therefore improving the robustness of the measurement while also being able to avoid the deformation problem. In this approach, when no phosphorus atoms are within the chosen cutoff, we select the nearest phospor to perform the calculation. The insertion is then calculated against the average z-positions of the phosphorus included in the cutoff (or the nearest outside the cutoff). Again, a 2D or 3D cutoff could be used and would result in contrasting insertion values (see Figure 2.3). In all tested systems, a cutoff of 6 Å seemed to give the best tradeoff between a good description of the local deformation and an average number of P atoms within the cutoff with enough robustness of the insertion measurment. For this purpose, I have implemented all these insertion methods in an object-oriented Python 2.7.12 script which was further optimized with Cython 0.26.1 [118]. The debug endeavor was performed with my colleague Tomás Silva. This script also features the thickness mode thoroughly detailed in the following section.

2.9.3 Thickness

Thickness is yet another lipid bilayer property which is vaguely defined. It is possible to find in the literature several experimental thickness reports using different definitions, depending on what is studied and the experimental technique.

Hydrophobic thickness 2 D_C is widely used in cases where transmembrane proteins cause the surrounding lipid bilayer to adjust its hydrocarbon thickness to match the length of the hydrophobic surface of the protein. This thickness is defined by the bilayer's hydrocarbon acyl chains and is assumed to be free of water [119].

D_{HH} corresponds to the distance between the lipid headgroups. It is defined by the distance between the peaks in the electron density profile (EDP) resolved by X-ray scattering [120]. D_{HH} is often taken as the phosphate-phosphate distance D_{PP} as they are similar and both obtained from EDP [121]. D_{HH} is more robust than D_{PP} because it does not depend strongly on the lipid components and the model used to fit the EDP data. However, D_{HH} is influenced by the electron density of inserting/inserted molecules, and thus a not reliable thickness measurment in these cases [122]. The method we used in this work is closely related to D_{PP} as explained below.

The steric thickness $D_{B'}$ is used for interactions with macromolecules or other bilayer and can be calculated from $2D_C + 2D_{H'}$. The assumed headgroup thicknesses $D_{H'}$ are guided by neutron diffraction

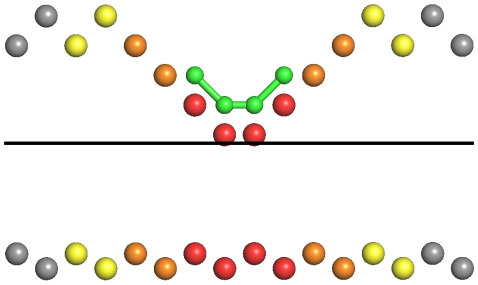

Figure 2.4: Graphical representation of our thickness calculation method. Firstly, the center of the membrane (black line) is determined by averaging the average z-position of both lipid monolayers, excluding all membrane atoms within a defined 2D cutoff (red and orange spheres) from the inserting molecule (green sticks). Then, the membrane atoms are grouped in subsets related to their 2D distance to the peptide, where depicted by different colors. The thickness is the difference between the previously calculated center of the membrane and the average z-position of the membrane atoms of each slab.

but are regarded as arbitrary [123].

The overall bilayer thickness or Luzzati thickness D_B is a simple volume-based measure that can be obtained using neutron scattering [124]. All these different thicknesses are related and a more profound discussion on the topic can be found on references 123 and 125. The presence of an external molecule (like a peptide or a protein) creates a new level of complexity in the thickness calculation. In this case, we no longer have a thickness value, but a thickness profile composed by the lipids radially distributed around the molecule.

Our newly developed method for thickness has the following outline (Figure 2.4): Firstly, the center of the membrane is determined by averaging the average z-position of both lipid monolayers. For this step, only phosphorus atoms beyond a defined 2D cutoff (usually 15 Å) from the inserting molecule are used. This way, heavily deformed lipid regions which could shift the membrane center are disregarded. In the second part of the algorithm, the phosphorus atoms are grouped in slabs according to their 2D distance to a group of interest, a single atom or the whole protein. The thickness is then given by the difference between the center of the membrane and the average z-position of the phosphorus subsets in each slab. The output is a thickness profile for each monolayer, which depends on the distance between phosphorus atoms and the inserting molecule. It may seem unnecessary to calculate the center of the membrane, since one could simply do the second part of the algorithm and defined the thickness as the difference between both monolayers in each phosphorus subset. However, our method was designed not only for inserting small molecules but also for transmembrane peptides (like pHLIP [126, 127]) or proteins. For these systems it is useful to distinguish between the thickness of each monolayer, since they can be significantly different.

CHAPTER 3

Results and Discussion

3.1 Replica Exchange Efficiency

To study the influence of lipid membranes on the pK_a values of titrable amino acids we applied the CpHMD method to the alanine-based pentapepide [1]. We have shown that upon membrane insertion and consequent desolvation, the titrable sites display a clear preference for their neutral forms (Figure B.1). This results in a decreased pK_a value for cationic residues and a shift in the opposite direction for anionic ones. We also observed a general lowering of the pK_a values as the peptides approached the membrane due to influence by choline groups positive charges. In that work, our methodology struggled to sample inserted conformations at the ionized state. To overcome the kinetic traps in protonation sampling, our group developed a pH-based replica exchange (pHRE) [2].

We tested three pentapeptides to investigate the sampling capabilities of this methodology. We chose His and Glu peptides as two typical cationic and anionic models, and C-ter because it showed the worst insertion sampling [1]. The following results and discussion will compare three pHRE simulations of three 100 ns replicas (pH values) with five CpHMD replicas at the same pH values and length.

In the pHRE method, each CpHMD replica is run at a unique pH value and attempts to exchange the simulation pH are performed between pairs of replicas. These exchanges allow replicas that have

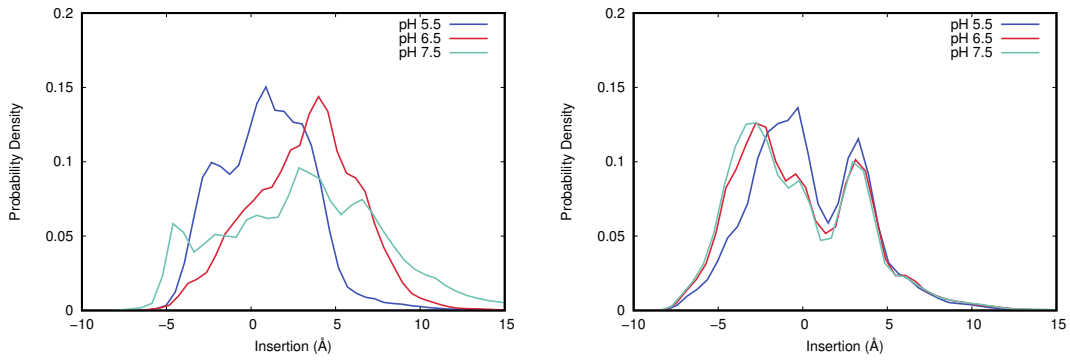

Figure 3.1: Insertion distribution of the His pentapeptide in CpHMD (left) and pHRE (right) with a 100 ps τ_{RE}.

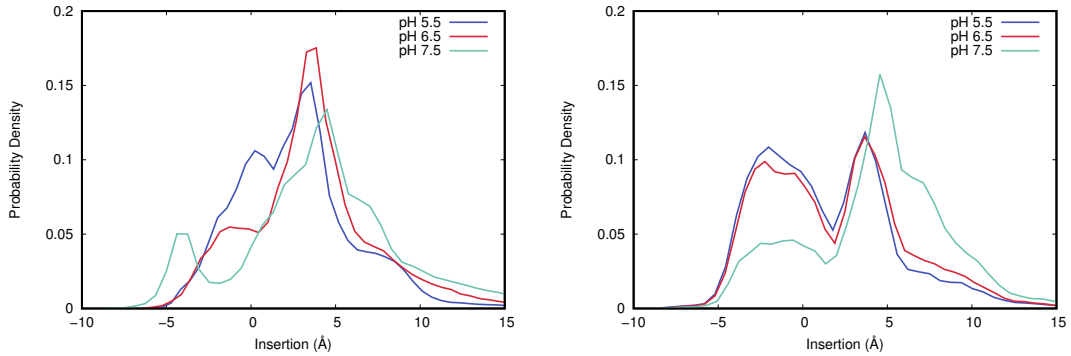

Figure 3.2: Insertion distribution of the Glu pentapeptide in CpHMD (left) and pHRE (right) with a 100 ps τ_{RE}.

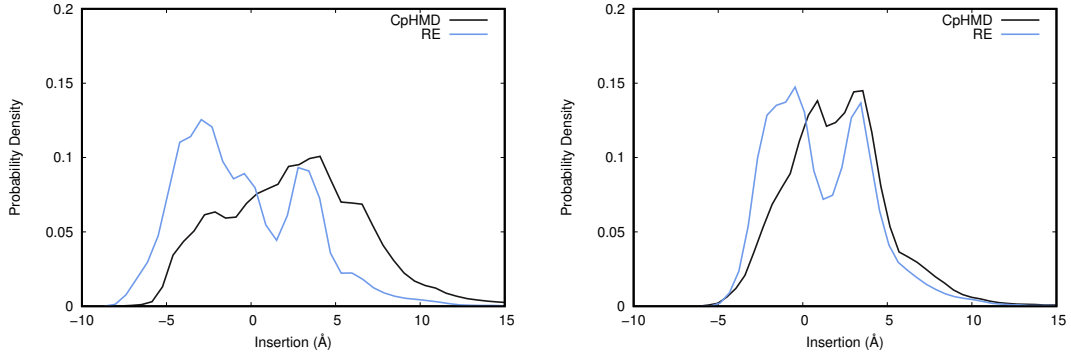

Figure 3.3: Insertion distribution of the His pentapeptide in the deprotonated (left) and protonated (right) states in both CpHMD and pHRE (100 ps τ_{RE}).

reached local energetic minimums to sample from a different energy landscape, hence, all replicas sample a similar portion of the ensemble. In our particular case, all pH values are able to sample conformations with almost equivalent insertions (Figure 3.1, 3.2 and B.2).

In CpHMD, the sampled insertions are not distributed proportionally to the simulation pH value, with the polar environment outside the membrane being preferred, to more inserted conformations. In order to calculate pK_a profiles, we need both protonated and deprotonated states to be well sampled in all regions. In the previous work, a sizable amount of pentapeptides simulations are discarded when calculating pK_a values of inserted configurations since there was an insufficient sampling of the inserted and ionized state. As shown in Figures 3.3, 3.4 and B.3, pHRE simulations were able to even slightly improve the sampling of these less favored configurations in almost half the computational time. Since all simulations in pHRE are (in case there are enough exchanges between pH values) sampling almost equivalent insertions regions, a faster protonation and conformation convergence is expected. This has already been reported by different groups upon extending their CpHMD with a RE scheme [63, 128].

The pHRE developed by Shen's group improved CpHMD protonation sampling which suffered from severe convergence problems stated by the authors [128]. In our case, the faster protonation convergence was not observed due to an increase in the protonation exchange rates (Table 3.1) but probably due to the replica mixing of pHRE. A direct comparison between CpHMD and pHRE simulations has to take into account that the protonation exchange rate varies with insertion. Since the local environment influences the likelihood of protonation states, one can only compare simulations that sampled the same

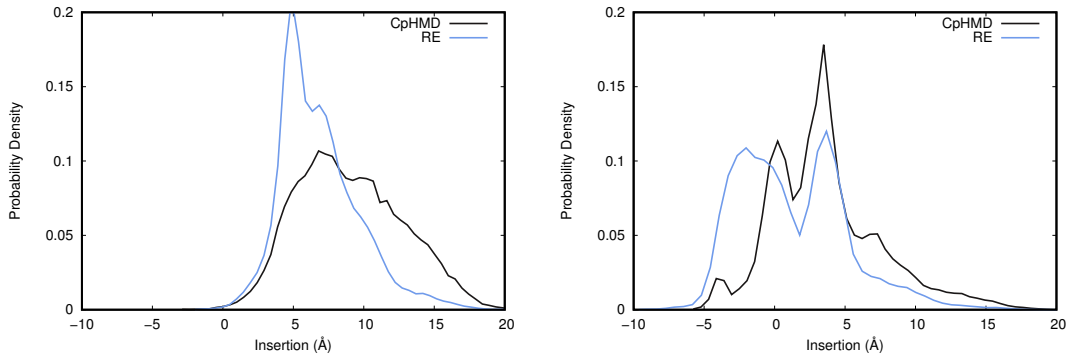

Figure 3.4: Insertion distribution of the Glu pentapeptide in the deprotonated (left) and protonated (right) states in both CpHMD and pHRE (100 ps τ_{RE}).

Table 3.1: Protonation exchange probability of selected replicas from CpHMD and pHRE (100 ps τ_{RE}) which sampled similar insertion regions. These replicas were selected because they displayed very similar insertion distributions. Data is not shown for pH values with no replicas sampling similar insertion regions. The His pentapeptide distributions can be found in Figure B.4. These probabilities are computed with the fraction of accepted protonation exchange attempts performed only in the CpHMD PB/MC step whose periodicity is given by the τ_{prot} (20 ps for both methods simulations), thus excluding protonation states arising from replica exchange.

System	pH	Probability (%)	
		CpHMD	pHRE
HIS	5.50	4.1	3.3
	6.50	23.3	23.9
	7.50	5.0	6.0
GLU	3.00	32.9	28.7
	4.00	-	-
	5.00	5.7	5.6
CTR	3.00	21.3	20.2
	4.00	-	-
	5.00	-	-

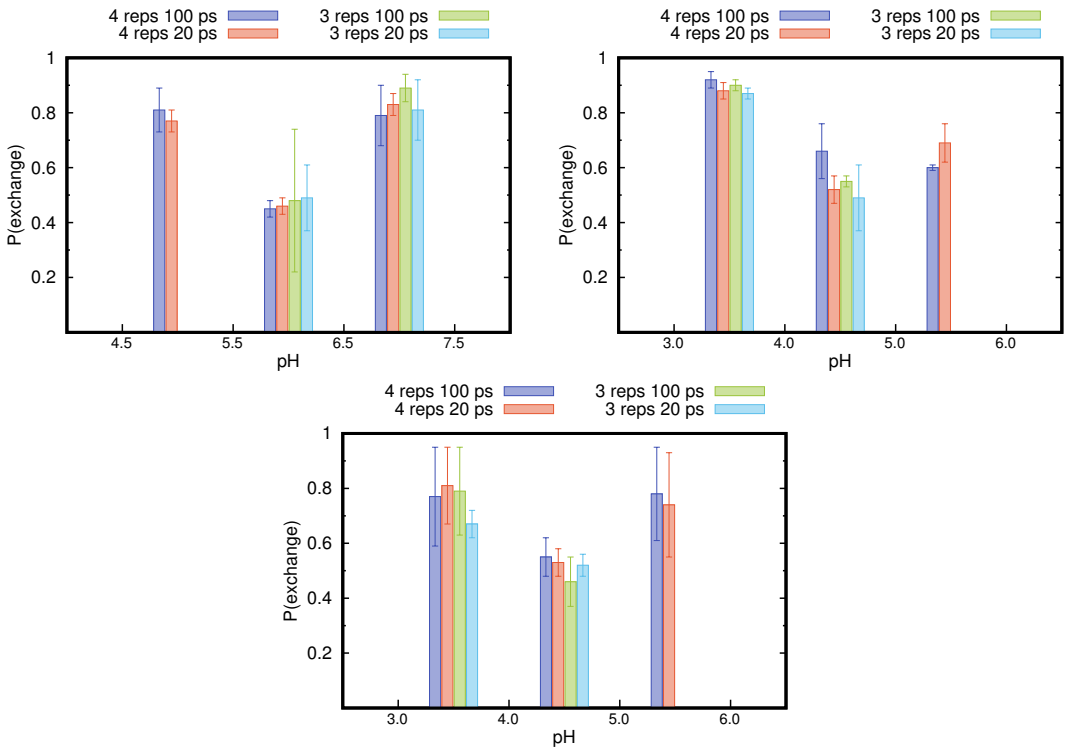

Figure 3.5: Exchange probabilities between pH values in pHRE for the His (left), Glu (right) and C-ter (bottom) pentapeptides.

insertion regions (Figure B.4). Despite speeding up our CpHMD sampling, pHRE is not able to sample more inserted configurations with the current settings. However, we can enhance the pHRE sampling by increasing the frequency of exchanges attempts and the number of pH values.

The pHRE simulations are an improvement over the CpHMD ones due to the mixing of the replicas. The quality of the mixing is then dictated by the exchange rate of replica pairs. In the worst case scenario, the exchange rate is null and both pHRE and CpHMD simulations are equivalent. The exchange criterion (Equation 2.47) is dependent on the differences between the pH values and protonation states of the replica pairs. Since the simulated pentapeptides only have one titrable site, the protonation state difference can at most be one, with a protonated residue in a replica and a deprotonated one in the other. Given our simulations with one pH unit of spacing between replicas, the exchange criteria can only take two values: 0.1 and 1. Therefore, the relatively high exchange rates observed are expected (Figure 3.5). It was also foreseeable that the pairs of replicas further away from the site's pK_a would have higher exchange probabilities, since these are more likely to be in the same protonation state. We tested a His pentapeptide pHRE simulation with a larger spacing (1.5 pH units) between pH values. The exchange probability has then slightly dropped to 0.59 ± 0.10 for the exchange between the 6.0 and 7.5 pH values and to 0.45 ± 0.10 for the pH 4.5 exchange with pH 6.0. Note that changing the number of simulated pH values does not affect the exchange probabilities, since the protonation state distribution of a pH value should not have a great effect on the other simulations. However, a minor deviation could occur considering that the added pH value could shift the insertion sampling which would reshape the observed protonations.

A roundtrip is defined as two complete trips across all pH values from the lower to the higher and vice versa. The number of roundtrips in a nanosecond is yet another good indicator of the quality of

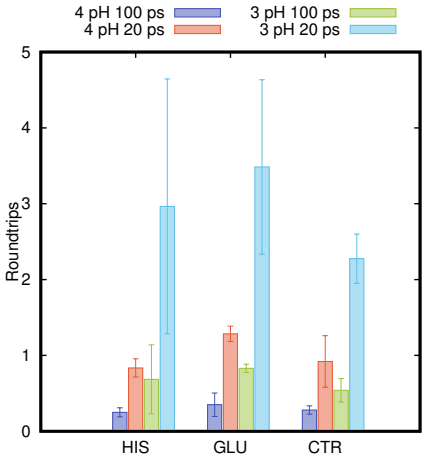

Figure 3.6: Number of roundtrips per nanosecond in all pHRE simulations.

the pHRE mixing. As anticipated by the high exchange rates between pH values, our simulations also exhibit a large amount of roundtrips (Figure 3.6). Contrarily to the exchange probability parameter, the number of pH values used has a clear impact on roundtrips. They decrease as the number of pH values increases. However, this does not mean that adding pH values has a negative influence on the pHRE sampling capability. His, Glu and C-ter feature the same roundtrips, and also exchange probabilities, since the simulated range of pH values were chosen to be approximately equidistant, starting from the residues water pK_a value. Since the sampled insertions were also very similar, the protonation states observed are inevitably identical.

The pHRE methodology introduces a new parameter τ_{RE} that controls the periodicidy of the exchange attempts. Although the exchange probabilities are not influenced by this parameter (Figure 3.5), increasing the number of attempts will also boost the rates of exchange between replicas which in turn, potentiates a higher amount of roundtrips (Figure 3.6). The His pentapeptide with a larger spacing in between three pH values was simulated at τ_{RE} 20 ps. Compared with the 1 pH unit spacing simulation, this one displays fewer roundtrips (0.86 ± 0.09) as a result of the lower exchange probabilities. Even though, the 20 ps τ_{RE} value yielded simulations with more roundtrips than the 100 ps ones, all τ_{RE} values tested present sufficient roundtrips to be considered as having good RE efficiency.

Potential problems related to excessively fast exchange attempts have been noted in not yet published preliminary validity tests of the pHRE method. If the probability of an exchange is affected by the previous swap, simulations may become biased. In this scenario, both an increase and a decrease of some exchange probabilities could take place, depending on whether protonation state convergence or divergence between replicas. Our pentapeptide system does not seem to suffer from this problem, since the protonation exchange rate remains constant after a replica swap (Figure 3.7). In case there was still a protonation bias after 20 ps derived from the replica exchange, a clear trend in the protonation exchange rates would arise. A bias to the preferred protonation state of the new pH value would result in a decreased protonation exchange for the first steps. As the simulation progressed and conformations became less correlated, the protonation exchange would slowly increase until a plateau was reached. There is a clear advantage in choosing the smallest possible τ_{RE} value, since shorter periods between exchange attempts improve the replica mixing. We have shown that a 20 ps τ_{RE} does not exhibit sampling bias problems, and thus we adopt this value for the future, including the rest of the pentapeptides (see Section 3.3) and the tetrapeptides simulations (see Section 5).

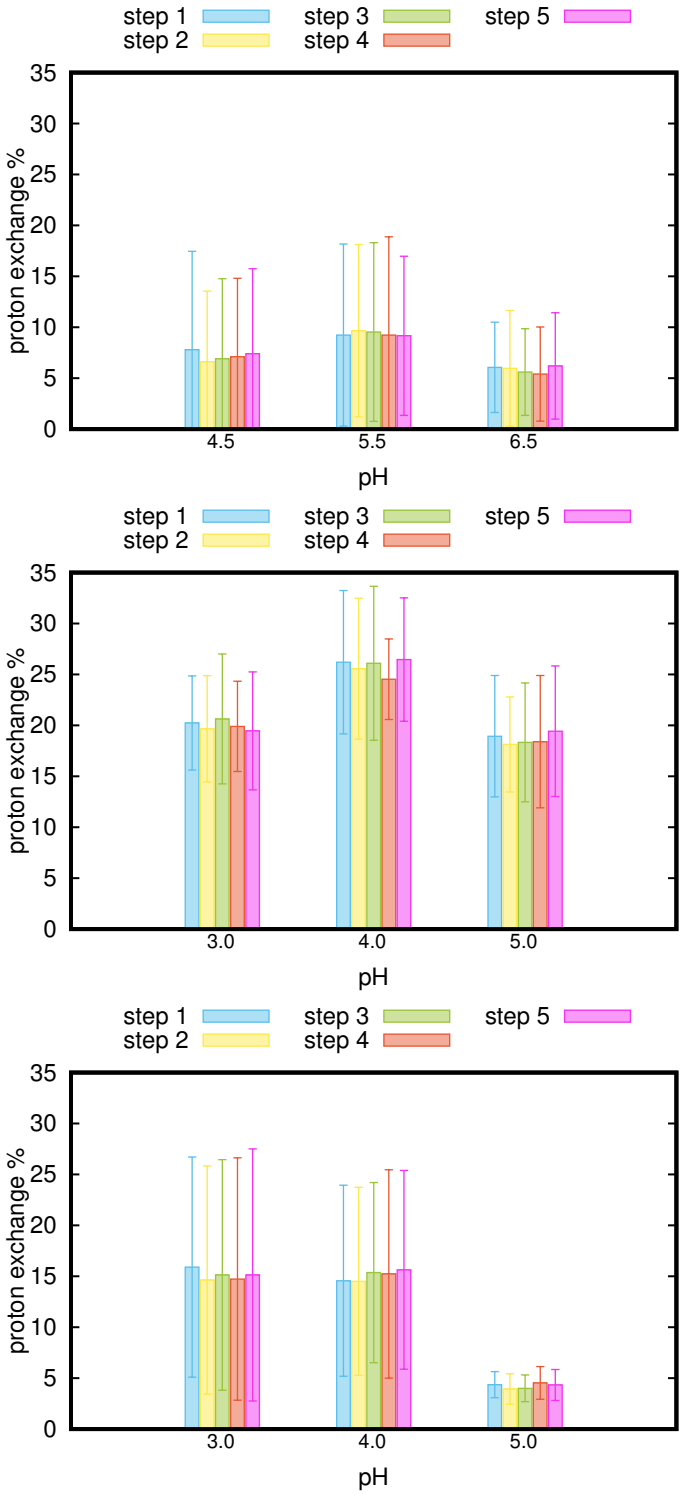

Figure 3.7: Protonation exchange probability of pHRE with τ_{RE} 100 ps in the following steps after a replica exchange for His (top), Glu (middle) and C-ter (bottom) pentapeptides. Step 1 is the protonation exchange attempt immediately following the replica exchange try and step 5 is the last CpHMD PB/MC step before another replica exchange swap is tested.

3.2 Insertion Criteria Influence on the Sampled Space

Membranes are complex systems which contain several lipid molecules. In this work, we use a PC membrane to study its influence on the pK_a values of inserting titrable aminoacids. Thus, a correct insertion measurement is needed in order to obtain the desired pK_a profile. There are several methods to calculate the insertion of a molecule in a membrane. During the insertion process, the membrane does not remain unperturbed, therefore depending on the used method, different insertion values for the same conformation can be obtained. This difference in insertion values, that will be propagated to the pK_a profile calculations can be as large as the membrane deformation.

The membrane thickness profile around the inserting molecule is a good measure to evaluate the degree of deformation of the lipid bilayer. Depending on the residues, each pentapeptide establishes unique interactions with the membrane thus motivating a particular deformation signature (Figure 3.8). However, the deformation magnitude is identical, since all these pentapeptides have a similar size. The membrane thickness profile is, like pK_a, affected by the insertion of the sampled conformations (Figure 3.9, C.1 and C.2). Away from the membrane the peptide has little influence on it, as it approaches the interface more interactions are initiated which results in a bigger membrane deformation.

An insertion value is a measurement related to the difference between the positions of the inserting molecule and its closest monolayer. The inserting molecule position can be described by a single atom, a geometric center or a center of mass. However, the monolayer description is more complex, since each membrane leaftlet is composed by several PC molecules (64 in our model) and a PC molecule features different moieties. Since the PC headgroup is still fairly solvated, we have chosen to adopt only the phosphorus atoms as the reference which also reduces the computation time. The difference in the insertion criteria then lies in including all lipid molecules, a group of them or only one. The most intuitive approach is to use all lipids as we are often tempted to think of monolayers as planes. However, as we have shown, membranes are actually quite deformed by inserting molecules. With this method it is possible to wrongly deem a conformation as inserted if it has displaced its surrounding lipid molecules (Figure 3.10). An alternative method, that could correctly estimate the insertion values of conformations, is to adopt as reference only the closest PC. In Teixeira *et al.*, we have used this approach to be able to capture the membrane deformation in our pK_a profile calculations. As expected, using these two methods in the original data from that study, we can observe that the average approach returns more

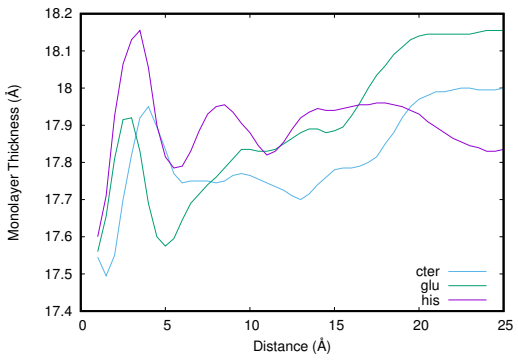

Figure 3.8: Closest monolayer thickness profile of the His, Glu and C-ter peptides. These calculations were calculated using the method described in Section 2.9.3 and a moving average with a 2 Å window size and a step of 0.5 Å.

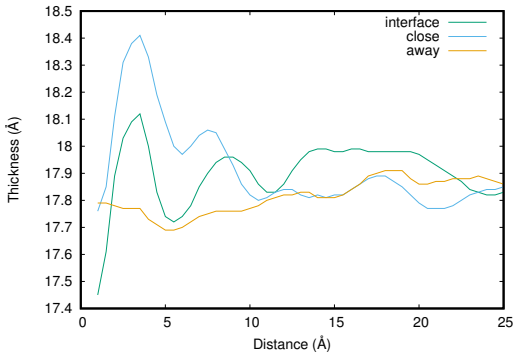

Figure 3.9: Closest interacting monolayer thickness profile of the His peptides in three distinct insertion regions: away from the membrane (at a 10 Å distance), closer (from 5 Å to 10 Å) and in the interface (from -5 Å to 5 Å insertion values). These calculations were calculated using the method described in Section 2.9.3 and a moving average with a 2 Å window size and a step of 0.5 Å.

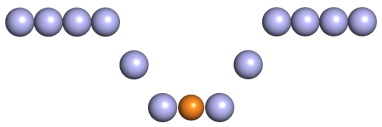

Figure 3.10: Example of a conformation (orange) that would wrongly be considered as inserted if all membrane phosphorus atoms (grey) were used to perform the insertion calculation.

inserted values for the same conformations (Figure 3.11). The resulting pK_a profile becomes artificially more insensitive to the membrane influence (Figure C.3).

While implementing a search algorithm for the closest atom, one has to decide whether it will be based on a two or three dimensional distance. An insertion into a lipid bilayer can be thought as a transition between a polar and an apolar medium. The position of the lipid that interacts the most with our titrable group, best describes its environment (Figure 3.12). In some configurations, choosing the closest 2D lipid is an overestimation of the insertion while the 3D one does not fully capture the aforementioned desolvation effect. Furthermore, both measurements experience a problem derived from representing the membrane with a single molecule. To better illustrate this issue, let us consider a simple example (Figure 3.13). If our reference group is interacting with more than one membrane lipids, and they are approximately equidistant, extremely small position updates are enough to change the attribution of the membrane reference lipid. In conformations such as the one depicted in this example, depending on this attribution, positive or negative insertion values may be reported even though the same solvation state is present. This can be solved by using more than one reference lipid, thus increasing the sampling and robustness of our insertion method. We have adopted a cutoff of around 6 Å since it seemed to minimize the effect, by increasing the average number of atoms used as reference, while still been able to account for membrane deformation (data not shown). If within the defined cutoff no lipids are found, the closest

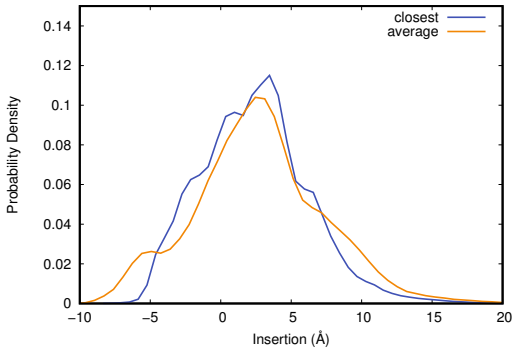

Figure 3.11: Insertion distribution of the His pentapeptide in the CpHMD simulations calculated using all lipids and only the closest methods.

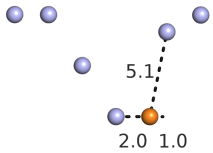

Figure 3.12: Example of a conformation that would wrongly be considered as inserted if a 2D criteria was used to select the closest membrane atom (grey) instead of a 3D one. In this case, the solvation state of the inserting molecule (orange) is closer to a not inserted configuration than a fully inserted one.

3D lipid is used as the reference.

The impact of using a cutoff is considerably shy when compared with the change of an average membrane z-position to a closest lipid-based insertion calculation (Figure 3.11). As depicted in Figure 3.14, only conformations in the interface are influenced (from -5 Å to 5 Å). In this region, we obtain a better description of the desolvation effect due to insertion. This way, we gain more structure in the insertion distribution data. In His pentapeptide, two insertion regions seem to be favored, one at around 2 Å and another at -2 Å, corresponding to an interaction between the titrable group and two phosphate groups (Figure 3.15). The more solvated interaction is preferred as it would be expected since the polar environment is more capable of stabilizing histidine's positive charge. Since anionic residues have less sampling on the interface region and their interaction with phosphate groups is weaker, their insertion distribution is almost not affected by the insertion definition difference (Figure C.4). A case where all presented methods fail except our cutoff approach is shown in Figure 3.16. Using all lipids results in an exaggerated insertion value (-7.7 Å) while choosing only the closest lipid describes the conformation as not inserted (2.7 Å). Our methodology returns a more approximated inserted value (-1.4 Å) considering the solvation state of the titrable site.

Despite the better description of the interactions, the new insertion method effect on the pK_a profile is very small (Figure 3.17). This new insertion definition has a more substantial repercussion in more complex systems such as pHLIP, a lengthier crossmembrane inserting peptide. In the next sections, all insertion calculations will be performed using the distance cutoff averaging method, unless stated otherwise.

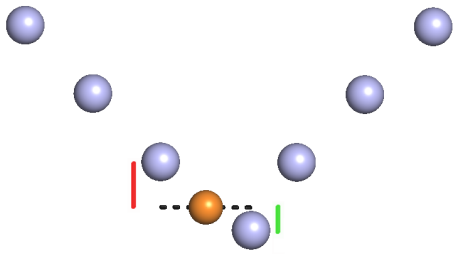

Figure 3.13: Example of a conformation where small movements from the insertion molecule (orange) would abruptly make it interchange between the not inserted and inserted states despite their similar solvation environment.

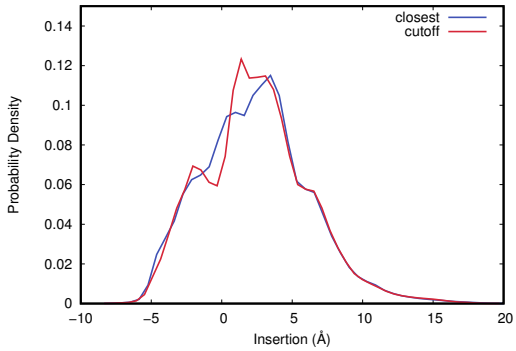

Figure 3.14: Insertion distribution of the His pentapeptide in the CpHMD simulations calculated using only the closest and the cutoff based methods.

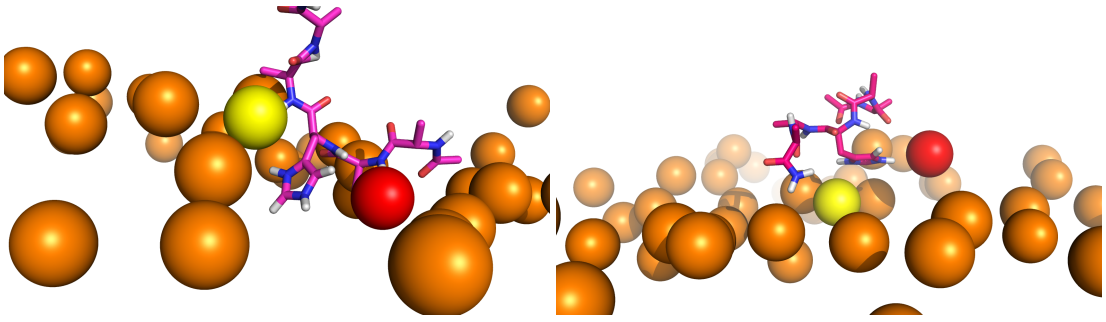

Figure 3.15: His peptide interacting with two phosphate groups (in red the closest one and in yellow the second closest). In the inserted state (left), if the closest atom method was used, an insertion value of 0.24 Å, instead of the more representative -2 Å obtained with the cutoff based method. The clearly not inserted conformation (right) that from the cutoff method is considered to be at 1.3 Å from the membrane, would be inserted -0.71 Å using the closest method.

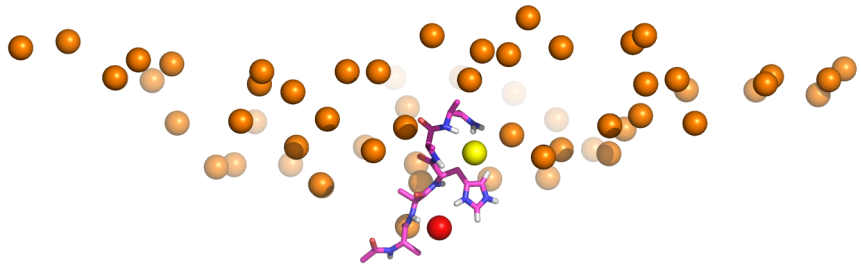

Figure 3.16: His peptide conformation that can take very different insertion states depending on the method used. An exaggerated insertion value (-7.7 Å) is obtained if all the membrane lipids (orange) are used. It can also be categorized as not inserted state (2.7 Å) if only the closest lipid is used. The cutoff based method determined insertion (-1.4 Å) seems like the most representative.

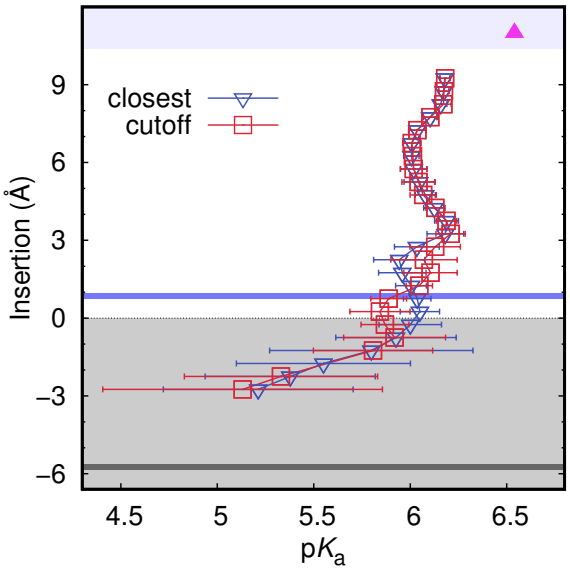

Figure 3.17: pK_a profile using the closest phosphate group and a 6 Å cutoff. The simulations were performed in 3 pH values using CpHMD and pHRE (100 ps τ_{RE}) methods. The water pK_a values of the His residue is shown on top, in a filled pink triangle bullet. The average position of cholines (blue) and the initial carbon of the acylchain (grey) are also shown.

3.3 pHRE as an improvement over CpHMD

In the previous chapters, we showed that pHRE is an improvement over CpHMD sampling and that using a cutoff based method to define insertion leads to a better characterization of the desolvation at the membrane interface. In spite of these changes and the introduction of a more strict pK_a criteria, the conclusions of the original CpHMD study [1] still hold. Profiles obtained with both CpHMD and pHRE are very similar despite the insertion method employed (Figures 3.17, 3.18 and D.1). The new criteria to consider insertion slabs with sufficient sampling was the change that most affected the pK_a profiles. By increasing the robustness of the profile, we have lost data in more inserted regions. Our goal is to explore pHRE's enhanced sampling in order to capture these more inserted pK_a values.

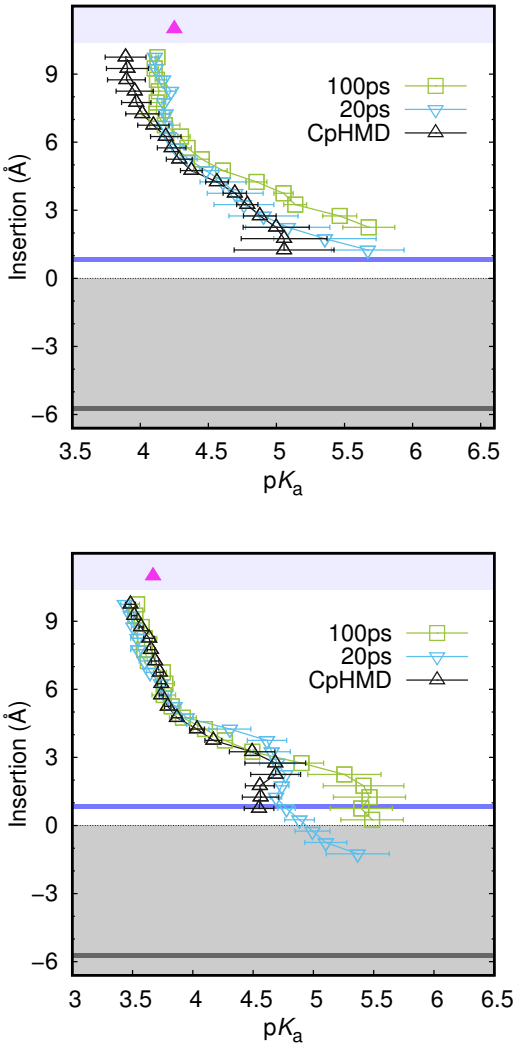

Figure 3.18: pK_a insertion profiles of Glu (top) and C-ter (bottom). The simulations were performed in 3 pH values using CpHMD and pHRE (τ_{RE} = 20 ps and 100 ps) methods. The water pK_a values of these residues are shown on top, in a filled triangle bullet. The average position of cholines (blue) and the initial carbon of the acylchain (grey) are also shown.

As discussed in section 3.1, the 20 ps τ_{RE} promotes a better replica mixing than the 100 ps one. Although similar insertion regions are considered to have enough sampling, a closer pK_a profile to the

CpHMD is obtained with this τ_{RE} value (Figures 3.18 and D.2) which suggests the 100 ps τ_{RE} is not yet converged in the more inserted values or that more replicates are needed. Note that, a proper comparison between CpHMD and pHRE methodologies should include identical total simulation time and number of replicates (see section 5), which is not the case.

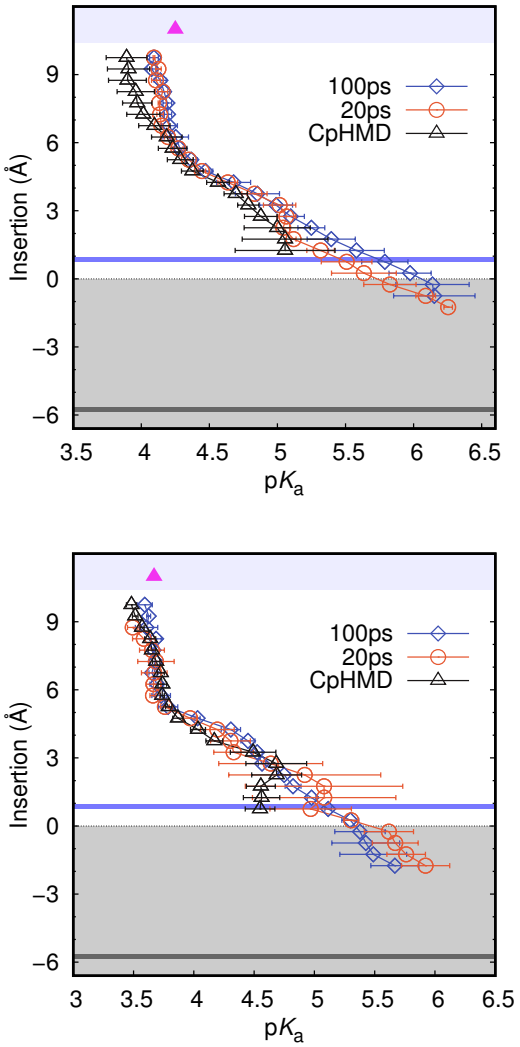

Figure 3.19: pK_a insertion profiles of Glu (top) and C-ter (bottom). The simulations were performed in 4 pH values using CpHMD and pHRE (τ_{RE} = 20 ps and 100 ps) methods. The water pK_a values of these residues are shown on top, in a filled triangle bullet. The average position of cholines (blue) and the initial carbon of the acylchain (grey) are also shown.

Upon membrane insertion, the amino acids tend to favor their neutral state. Despite having sampled many conformations up to 5 Å of insertion (Figures 3.1, 3.2 and B.2), the ionized state shortage crippled the pK_a profiles. To further enhance the sampling of the ionized state, an extra pH value was added to the pHRE simulations. Consequently, a considerable gain was attained in the pK_a profile inserted region (Figures 3.19 and D.2). However, to fully understand how much of the progress was due to increasing simulation time, a comparison with pHRE simulations with 3 pH values and equal total simulation time should be made (see section 5). Again, the 20 ps τ_{RE} is able to better reproduce the CpHMD results and sample slightly more inserted ionized conformations. The added pH value favors this ionized state which

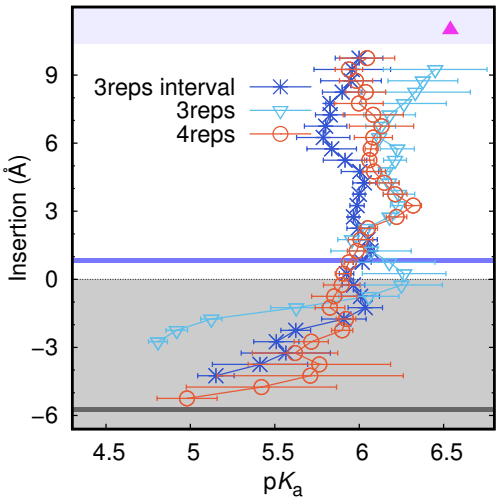

Figure 3.20: pK_a insertion profiles of His peptide. The simulations were performed in 3 pH values with a 1.5 pH units spacing instead of the regular 1 pH in both CpHMD and pHRE (τ_{RE} = 20 ps and 100 ps) methods. The water pK_a values of these residues are shown on top, in a filled triangle bullet. The average position of cholines (blue) and the initial carbon of the acylchain (grey) are also shown.

in turn prefers the water polar medium. Since the RE scheme allows a mixing of the pH values in the different replicas, the probability of the extra pH to sample inserted regions increases. In the CpHMD method, it is likely that the extra pH value would not improve the sampling of the inserted conformations in the ionized state as much. Nonetheless, this reasoning should be tested as it is not supported by enough evidence (see section 5).

Increasing the spacing between pH values was yet another attempt to improve the pK_a profile (Figure 3.20). The high pH values which favor the neutral form were kept while the other two were shifted in opposite direction. This approach did improve the inserted region, however, at the expense of worst sampling of solvated conformations, since two-thirds of the simulation are now more membrane prone and the pH values exchange between replicas have reduced (see section 3.1).

The pK_a profiles of the remaining titrable amino acids were obtained using pHRE simulations with 4 pH values and a 20 ps τ_{RE}. Like the three first tested pentapeptides, all other amino acids were better sampled at inserted configurations (Figures 3.21). The only exception was tyrosine which did not display any improvement on the insertion depth, as shown on the pK_a profile. However, the new Tyr profile presents a modest deviation from the original CpHMD simulations which could also imply a better sampling. Both Tyr and Cys are not able to sample ionized inserted conformations since the range of pH values used (from pH 8 to 11) is short, considering the pK_a shift experienced by these two residues.

pHRE was developed to circumvent possible kinetic traps found in CpHMD simulations. The N-ter pentapeptide, when neutral, showed a persistent locked configuration 3.22 for long periods of time [1]. In these conformations, the terminal amine strongly interacted with the mainchain amine of the second residue, in a hydrogen bond like manner ($\sim$ 0.2 – 0.3 nm distance). These interactions are pH and insertion dependent (Figures 3.23). The pH effect is clear since only a neutral N-terminus is able to present the negative lone pair and accept the interaction from the positive proton in the mainchain of the second residue. The effect of membrane insertion is related with the dielectric media, and the hydrogen bond competition with water molecules. Indeed, in solution, a water hydrogen will be entropically more

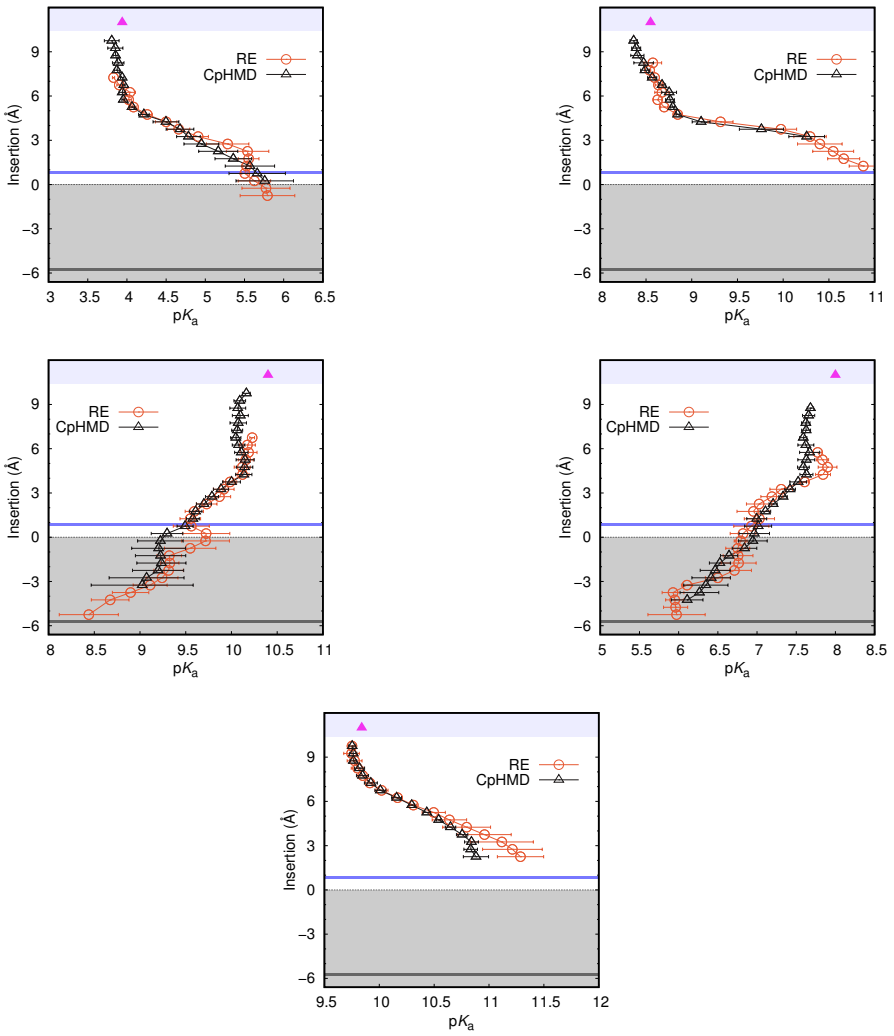

Figure 3.21: pK_a insertion profiles of Asp (top left), Cys (top right), Lys (middle left), N-ter (middle right) and Tyr (bottom). The simulations were performed in 4 pH values using CpHMD and pHRE (τ_{RE} = 20 ps) methods. The water pK_a values of these residues are shown on top, in a filled triangle bullet. The average position of cholines (blue) and the initial carbon of the acylchain (grey) are also shown.

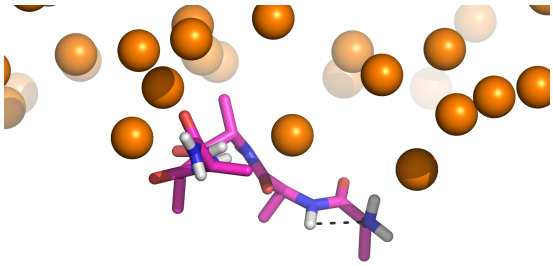

Figure 3.22: Graphical representation of the N-ter peptide (sticks) in a locked conformation where the N-terminus neutral amine is closely interacting with the adjacent residue main chain N–H group. the interaction is highlighted with a black dashed line. Lipids phosphorus atoms are shown as orange spheres.

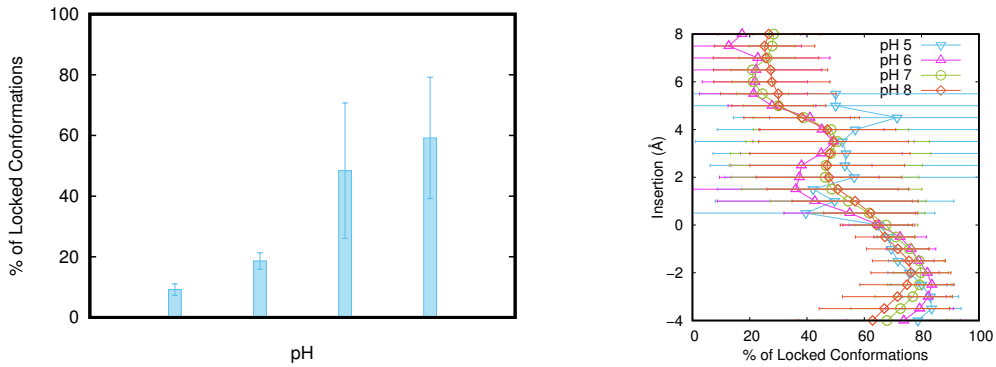

Figure 3.23: Percentage of the locked conformations in N-ter pentapeptide over pH (left). Percentage of neutral locked conformations varying with membrane insertion (right). Only insertions until 4 Å deep are shown because after this value the uncertainty (the error bars) in the calculation becomes too large.

favored to occupy the lone pair slot of the neutral amine drastically decreasing the abundance of the locked conformations, which can be seen as a weak interaction. However, by inserting in the membrane and due to desolvation, this weak interaction becomes almost the best option available, hence leading to the observed locked conformations. The fact that these locked conformations are in such abundance, suggests that they may not be a kinetic trap but rather the global minimum in inserted regions.

CHAPTER 4

Concluding Remarks

We have used a replica exchange constant-pH molecular dynamics methodology to study the pK_a profile of all titrable amino acids at the water/membrane interface. In this method, multiple CpHMD replicas are simulated at different pH values and, at a given time interval τ_{RE}, pairs of replicas try to exchange their pH value. Consequently, all pH values sample similar insertion regions, maximizing the abundance of inserted conformations in the ionized state. Since this type of conformation/ionization was the bottleneck in the calculation of the pK_a profiles, we were able to extend them for most of the amino acids. Nevertheless, we were not able to obtain inserted pK_a values for two residues probably due to a too short simulated pH range.

The best results were obtained using 4 pH values and a 20 ps τ_{RE}. To enhance pHRE sampling ability, one should choose the lowest τ_{RE} value, ensuring no bias in the exchange probability between replicas. The replica mixing can be improved by increasing the frequency of exchange attempts or by reducing the spacing between the pH values of the simulations. Given our simulated pH values spacing and our single titrable site systems, the expected high exchange rates and roundtrips were confirmed. Although all tested τ_{RE} values displayed a good RE efficiency, the pK_a profiles obtained with the 100 ps τ_{RE} simulations suggested these were further away from convergence than the 20 ps ones and probably more replicates are needed.

This work previewed a new method to calculate insertion values of membrane permeating molecules. Unlike the most immediate solution of using the average z-positions of all lipids, this approach is able to capture both membrane deformations and solvation effects. We use the average z-position of all lipids within a cutoff around the inserting molecule. In the past, we have used only the closest lipid, however, using the current more accurate and robust description of the interactions at membrane interface is achieved. Despite having little effect on the pentapeptides pK_a profile, this method has a meaningful impact in studying interactions occurring at the membrane interface.

pHRE was clearly able to improve CpHMD sampling and derived pK_a profiles. In the future, this enhanced sampling technique will be our preferred approach to study pH dependent processes in biomolecules even when the sampling limitations might not be as challenging as in the case of the water/membrane interface.

CHAPTER 5

Ongoing Work

pHRE is currently the cutting-edge methodology of our group and its development, debugging and testing matured throughout the past year. To ensure this thesis presented results of pHRE simulations only a subset of all performed simulations was analyzed. pHRE simulations are slower than the CpHMD for practical reasons since since in our implementation all replicas must be run in the same computer node and the parallelization is limited by the number of available CPU cores in each node.

The comparisons between CpHMD and pHRE simulations presented in this thesis are not completely comparable since CpHMD number of replicates and total simulation time far exceeds the pHRE ones. This maybe responsible for the somewhat shy results obtained in pHRE. However, it does not invalidate the conclusions taken. Furthermore, after the remaining pHRE simulations are analyzed, even better results with at least smaller error bars are expected since we are increasing sampling.

An important control to help the comparison between CpHMD and pHRE is the inclusion of a fourth pH value in the CpHMD simulations. Only after this comparison it is possible to unequivocably attest the superior sampling of pHRE.

The AXXA tetrapeptides (with X= Glu, His) were idealized to study the effect of two titrable amino acids right next to each other. In the case of Glu/Glu or His/His, we should have a simpler effect of electrostatic repulsion added to desolvation effect, leading to larger pK_a shifts for these residues upon membrane insertion. However, for His/Glu, we aim to investigate if these residues insert in the membrane in their neutral forms or if they are able to mutually stabilize their ionizable forms, through a salt bridge. The preliminary results (Figure 5.1) suggest that upon membrane insertion both residues prefer to adopt a neutral state. In case there was a significant contribution of the salt bridge configuration, a severe shift in the pK_a profiles of His and Glu would be noted. Also, these simulations have twice as much sampling (3 replicates of 6 replicas) than the previously presented pHRE pentapeptides simulations. Yet, the pK_a profiles of both Glu and His in tetrapeptides are higly comparable to the ones obtained in pentapeptides. This strongly indicates that our pHRE simulations are rapidly convergent. The presented pK_a values were calculated with the first pK_1 and second pK_2 macroscopic p$K_a$values [129],

$$\langle n \rangle = \frac{10^{-\mathrm{p}K_1 - \mathrm{p}H} + 2 \times 10^{-2\mathrm{p}H}}{10^{-\mathrm{p}K_1 - pK_2} + 10^{-\mathrm{p}K_1 - \mathrm{p}H} + 10^{-2\mathrm{p}H}} \tag{5.1}$$

where $\langle n \rangle$ is the average protonation state of the peptide at a given pH.

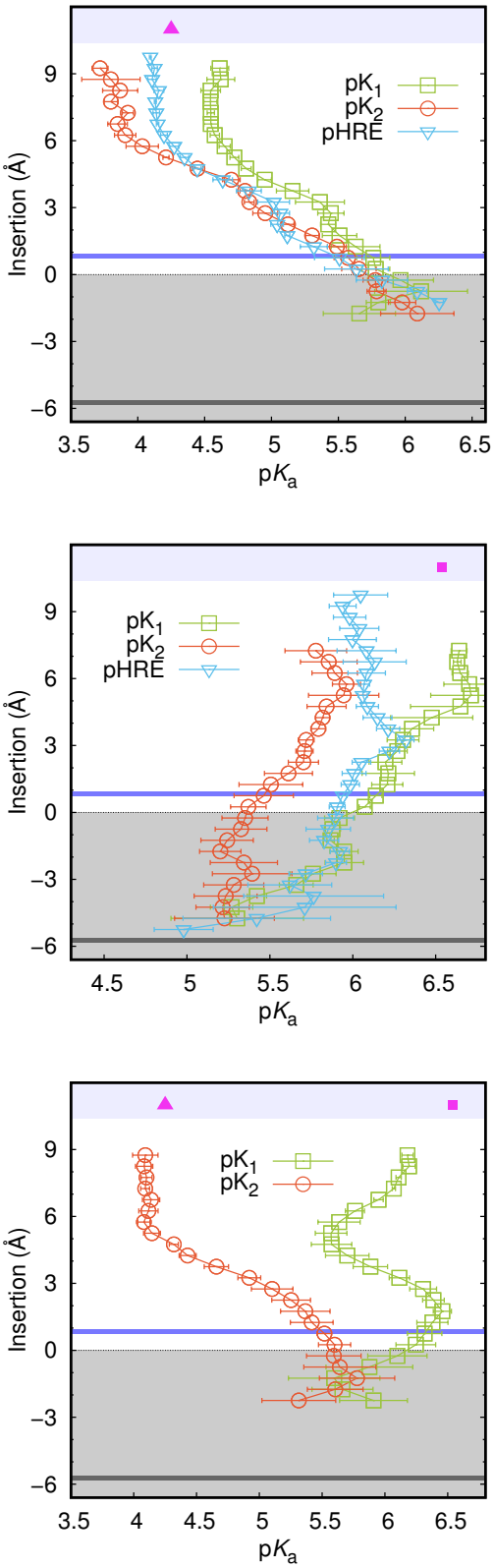

Figure 5.1: pK_a insertion profiles of His-His (top), Glu-Glu and His-Glu amino acids in AXXA tetrapeptides (with X= Glu, His). pHRE (with 4 pH values and a 20ps τ_{RE}) results shown for comparision. The water pK_a values of these residues are shown on top in pink. The average position of cholines (blue) and the initial carbon of the acylchain (grey) are also shown.

Bibliography

[1] V. H. Teixeira, D. Vila Viçosa, P. B. P. S. Reis, and M. Machuqueiro, "PKa Values of Titrable Amino Acids at the Water/Membrane Interface," *Journal of Chemical Theory and Computation*, vol. 12, pp. 930–934, mar 2016.

[2] D. Vila-Viçosa, A. M. Baptista, C. Oostenbrink, and M. Machuqueiro, "A pH replica exchange scheme in the stochastic titration constant-pH MD method," *In preparation*.

[3] S. Sörenson, "Über die messung und die bedeutung der wasserstoffionenkoncentration bei enzymatischen prozessen," *Biochemische Zeitschrift*, no. 21, pp. 131–200, 1909.

[4] M. Stoeckelhuber, A. Noegel, C. Eckerskorn, J. Kohler, D. Rieger, and M. Schleicher, "Structure/function studies on the pH-dependent actin-binding protein hisactophilin in dictyostelium mutants," *J Cell Sci*, no. 109, pp. 1825–1835, 1996.

[5] R. A. Dimitrov and R. R. Crichton, "Self-consistent field approach to protein structure and stability. I. pH dependence of electrostatic contribution," *Proteins*, no. 27, pp. 576–596, 1997.

[6] K. A. Dill, "Dominant forces in protein folding," *Biochemistry*, vol. 29, pp. 7133–7155, aug 1990.

[7] D. E. Anderson, W. J. Becktel, and F. W. Dahlquist, "pH-Induced denaturation of proteins: a single salt bridge contributes 3-5 kcal/mol to the free energy of folding of T4 lysozyme," *Biochemistry*, vol. 29, pp. 2403–2408, mar 1990.

[8] A. Yang and B. Honig, "On the pH Dependence of Protein Stability," *Journal of Molecular Biology*, vol. 231, pp. 459–474, may 1993.

[9] M. Dixon, "The effect of pH on the affinities of enzymes for substrates and inhibitors.," *The Biochemical journal*, vol. 55, pp. 161–70, aug 1953.

[10] C. Hunte, E. Screpanti, M. Venturi, A. Rimon, E. Padan, and H. Michel, "Structure of a Na+/H+ antiporter and insights into mechanism of action and regulation by pH," *Nature*, vol. 435, pp. 1197–1202, jun 2005.

[11] J. Deng, R. T. Toledo, and D. A. Lillard, "Effect of temperature and ph on protein-protein interaction in actomyosin solutions," *Journal of Food Science*, vol. 41, pp. 273–277, mar 1976.

[12] E. Sahin, A. O. Grillo, M. D. Perkins, and C. J. Roberts, "Comparative Effects of pH and Ionic Strength on Protein–Protein Interactions, Unfolding, and Aggregation for IgG1 Antibodies," *Journal of Pharmaceutical Sciences*, vol. 99, pp. 4830–4848, dec 2010.

[13] H. Träuble, M. Teubner, P. Woolley, and H. Eibl, "Electrostatic interactions at charged lipid membranes," *Biophysical Chemistry*, vol. 4, pp. 319–342, jul 1976.

[14] A. Watts, K. Harlos, W. Maschke, and D. Marsh, "Control of the structure and fluidity of phosphatidylglycerol bilayers by pH titration," *Biochimica et Biophysica Acta (BBA) - Biomembranes*, vol. 510, pp. 63–74, jun 1978.

[15] 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*, vol. 20, no. 17, pp. 4955–4965, 1981.

[16] T. C. Vogt and B. Bechinger, "The interactions of histidine-containing amphipathic helical peptide antibiotics with lipid bilayers. The effects of charges and pH.," *The Journal of biological chemistry*, vol. 274, pp. 29115–21, oct 1999.

[17] A. C. Chakrabarti and D. W. Deamer, "Permeability of lipid bilayers to amino acids and phosphate," *Biochimica et Biophysica Acta (BBA) - Biomembranes*, vol. 1111, pp. 171–177, nov 1992.

[18] J. Wike-Hooley, J. Haveman, and H. Reinhold, "The relevance of tumour pH to the treatment of malignant disease," *Radiotherapy and Oncology*, vol. 2, pp. 343–366, dec 1984.

[19] C. D. Syme, R. C. Nadal, S. E. J. Rigby, and J. H. Viles, "Copper binding to the amyloid-beta (Abeta) peptide associated with Alzheimer's disease: folding, coordination geometry, pH dependence, stoichiometry, and affinity of Abeta-(1-28): insights from a range of complementary spectroscopic techniques.," *The Journal of biological chemistry*, vol. 279, pp. 18169–77, apr 2004.

[20] S. G. Nugent, D. Kumar, D. S. Rampton, and D. F. Evans, "Intestinal luminal pH in inflammatory bowel disease: possible determinants and implications for therapy with aminosalicylates and other drugs," *Gut*, vol. 48, no. 4, pp. 571–577, 2001.

[21] B. Fleckenstein, Ø. Molberg, S.-W. Qiao, D. G. Schmid, F. von der Mülbe, K. Elgstøen, G. Jung, and L. M. Sollid, "Gliadin T cell epitope selection by tissue transglutaminase in celiac disease. Role of enzyme specificity and pH influence on the transamidation versus deamidation process.," *The Journal of biological chemistry*, vol. 277, pp. 34109–16, sep 2002.

[22] Y. Goto, L. J. Calciano, and A. L. Finkt, "Acid-induced folding of proteins," *Biochemistry*, vol. 87, pp. 573–577, 1990.

[23] R. Zahn, "The Octapeptide Repeats in Mammalian Prion Protein Constitute a pH-dependent Folding and Aggregation Site," *Journal of Molecular Biology*, vol. 334, pp. 477–488, nov 2003.

[24] 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.*, vol. 10, pp. 2176–2184, 2014.

[25] F. C. Tsui, D. M. Ojcius, and W. L. Hubbell, "The intrinsic pK_a values for phosphatidylserine and phosphatidylethanolamine in phosphatidylcholine host bilayers," *Biophys. J.*, vol. 49, no. 2, pp. 459–468, 1986.

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

[27] 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," *Biochem. J.*, vol. 72, pp. 2196–2210, 1997.

[28] H. A. Santos, D. Vila-Viçosa, V. H. Teixeira, A. M. Baptista, and M. Machuqueiro, "Constant-pH MD simulations of DMPA/DMPC lipid bilayers," *J. Chem. Theory Comput.*, vol. 11, no. 12, pp. 5973–5979, 2015.

[29] A. Klymchenko and R. Kreder, "Fluorescent Probes for Lipid Rafts: From Model Membranes to Living Cells," *Chemistry & Biology*, vol. 21, pp. 97–113, jan 2014.

[30] J. F. Hunt, P. Rath, K. J. Rothschild, and D. M. Engelman, "Spontaneous, pH-dependent membrane insertion of a transbilayer α-helix," *Biochemistry*, vol. 36, no. 49, pp. 15177–15192, 1997.

[31] J. Fendos, F. N. Barrera, and D. M. Engelman, "Aspartate embedding depth affects pHLIP's insertion pK_a," *Biochemistry*, vol. 52, no. 27, pp. 4595–4604, 2013.

[32] N. A. Baker, *Poisson-Boltzmann Methods for Biomolecular Electrostatics*, vol. 383. Academic Press, jan 2004.

[33] M. J. Potter, M. K. Gilson, and J. A. McCammon, "Small molecule pka prediction with continuum electrostatics calculations," *J. Am. Chem. Soc.*, vol. 116, no. 22, pp. 10298–10299, 1994.

[34] L. Wang, L. Li, and E. Alexov, "pKa predictions for proteins, rnas, and dnas with the gaussian dielectric function using delphi pKa," *Proteins: Structure, Function, and Bioinformatics*, vol. 83, pp. 2186–2197, dec 2015.

[35] D. Bashford and D. A. Case, "Generalized born models of macromolecular solvation effects," *Annu. Rev. Phys. Chem*, vol. 51, pp. 129–52, 2000.

[36] M. Feig and C. L. Brooks, *Recent advances in the development and application of implicit solvent models in biomolecule simulations*, vol. 14. Elsevier Current Trends, apr 2004.

[37] W. C. Still, A. Tempczyk, R. C. Hawley, and T. Hendrickson, "Semianalytical treatment of solvation for molecular mechanics and dynamics," *Journal of the American Chemical Society*, vol. 112, pp. 6127–6129, aug 1990.

[38] 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.*, vol. 14, no. 2, pp. 161–185, 1993.

[39] Y. Y. Sham, Z. T. Chu, and A. Warshel, "Consistent calculations of p$k_a$pka's of ionizable residues in proteins: Semi-microscopic and microscopic approaches," *The Journal of Physical Chemistry B*, vol. 101, no. 22, pp. 4458–4472, 1997.

[40] A. Warshel, "Electrostatic energy and macromolecular function," *Annu. Rev. Biophys. Biophys. Chem.*, vol. 20, pp. 267–298, 1991.

[41] I. Eberini, A. 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," vol. 54, no. 4, pp. 744–758, 2004.

[42] A. Koumanov, A. Karshikoff, E. P. Friis, and T. V. Borchert, "Conformational averaging in pK calculations: Improvement and limitations in prediction of ionization properties of proteins," *Journal of Physical Chemistry B*, vol. 105, no. 38, pp. 9339–9344, 2001.

[43] B. Kuhn, P. A. Kollman, and M. Stahl, "Prediction of pK a shifts in proteins using a combination of molecular mechanical and continuum solvent calculations," *Journal of Computational Chemistry*, vol. 25, pp. 1865–1872, nov 2004.

[44] J. E. Mertz and B. M. Pettitt, "Molecular dynamics at a constant ph," *Int. J. High Perform. C.*, vol. 8, no. 1, pp. 47–53, 1994.

[45] 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.*, vol. 27, no. 4, pp. 523–544, 1997.

[46] U. Börjesson and P. Hünenberger, "Explicit-solvent molecular dynamics simulation at constant ph: methodology and application to small amines," *J. Chem. Phys.*, vol. 114, pp. 9706–9719, 2001.

[47] A. M. Baptista, "Comment on "explicit-solvent molecular dynamics simulation at constant pH: Methodology and application to small amines" [J. Chem. Phys. 114, 9706 (2001)]," may 2002.

[48] M. Lee, F. Salsbury Jr, and C. Brooks III, "Constant-ph molecular dynamics using continuous titration coordinates," *Proteins Struct. Funct. Bioinf.*, vol. 56, pp. 738–752, 2004.

[49] X. Kong and C. L. Brooks III, "λ-dynamics: A new approach to free energy calculations," *J. Chem. Phys.*, vol. 105, no. 6, pp. 2414–2423, 1996.

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

[51] M. Dlugosz and J. Antosiewicz, "Constant-ph molecular dynamics simulations: a test case of succinic acid," *Chem. Phys.*, vol. 302, pp. 161–170, 2004.

[52] J. Mongan, D. Case, and J. McCammon, "Constant ph molecular dynamics in generalized born implicit solvent," *J. Comput. Chem.*, vol. 25, pp. 2038–2048, 2004.

[53] F. S. Lee, Z.-T. Chu, M. B. Bolger, and A. Warshel, "Calculations of antibody-antigen interactions: microscopic and semi-microscopic evaluation of the free energies of binding of phosphorylcholine analogs to mcpc603," *Protein Engineering, Design and Selection*, vol. 5, no. 3, pp. 215–228, 1992.

[54] G. Hummer and A. Szabo, "Calculation of free-energy differences from computer simulations of initial and final states," *Journal of Chemical Physics*, vol. 105, p. 2004, aug 1996.

[55] M. Machuqueiro, S. R. Campos, C. M. Soares, and A. M. Baptista, "Membrane-induced conformational changes of kyotorphin revealed by molecular dynamics simulations," *J. Phys. Chem. B*, vol. 114, no. 35, pp. 11659–11667, 2010.

[56] R. L. Thurlkill, G. R. Grimsley, J. M. Scholtz, and C. N. Pace, "pK values of the ionizable groups of proteins," *Protein Sci.*, vol. 15, no. 5, pp. 1214–1218, 2006.

[57] 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.*, vol. 18, no. 1, pp. 247–251, 2009.

[58] A. Mitsutake, Y. Sugita, and Y. Okamoto, "Generalized-ensemble algorithms for molecular simulations of biopolymers," *Peptide Science*, vol. 60, no. 2, pp. 96–123, 2001.

[59] H. Li, M. Fajer, and W. Yang, "Simulated scaling method for localized enhanced sampling and simultaneous "alchemical" free energy simulations: A general method for molecular mechanical, quantum mechanical, and quantum mechanical/molecular mechanical simulations," *The Journal of Chemical Physics*, vol. 126, no. 2, p. 024106, 2007.

[60] E. Marinari and G. Parisi, "Simulated tempering: A new monte carlo scheme," *EPL (Europhysics Letters)*, vol. 19, no. 6, p. 451, 1992.

[61] Y. Sugita and Y. Okamoto, "Replica-exchange molecular dynamics method for protein folding," *Chem. Phys. Lett.*, vol. 314, no. 1, pp. 141–151, 1999.

[62] J. Khandogin and C. L. Brooks, "Toward the accurate first-principles prediction of ionization equilibria in proteins," *Biochemistry*, vol. 45, no. 31, pp. 9363–9373, 2006.

[63] 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.*, vol. 10, no. 3, pp. 1341–1352, 2014.

[64] E. Harder, W. Damm, J. Maple, C. Wu, M. Reboul, J. Y. Xiang, L. Wang, D. Lupyan, M. K. Dahlgren, J. L. Knight, J. W. Kaus, D. S. Cerutti, G. Krilov, W. L. Jorgensen, R. Abel, and R. A. Friesner, "OPLS3: A Force Field Providing Broad Coverage of Drug-like Small Molecules and Proteins," *Journal of Chemical Theory and Computation*, vol. 12, pp. 281–296, jan 2016.

[65] D. Case, R. Betz, D. Cerutti, T. C. III, T. Darden, R. Duke, T. Giese, H. Gohlke, A. Goetz, N. Homeyer, S. Izadi, P. Janowski, J. Kaus, A. Kovalenko, T. Lee, S. LeGrand, P. Li, C. Lin, T. Luchko, R. Luo, B. Madej, D. Mermelstein, K. Merz, G. Monard, H. Nguyen, H. Nguyen, I. Omelyan, A. Onufriev, D. Roe, A. Roitberg, C. Sagui, C. Simmerling, W. Botello-Smith, J. Swails, R. Walker, J. Wang, R. Wolf, X. Wu, L. Xiao, and P. Kollman, *AMBER 2016, University of California, San Francisco.* 2016.

[66] B. R. Brooks, C. L. Brooks, A. D. Mackerell, L. Nilsson, R. J. Petrella, B. Roux, Y. Won, G. Archontis, C. Bartels, S. Boresch, A. Caflisch, L. Caves, Q. Cui, A. R. Dinner, M. Feig, S. Fischer, J. Gao, M. Hodoscek, W. Im, K. Kuczera, T. Lazaridis, J. Ma, V. Ovchinnikov, E. Paci, R. W. Pastor, C. B. Post, J. Z. Pu, M. Schaefer, B. Tidor, R. M. Venable, H. L. Woodcock, X. Wu, W. Yang, D. M. York, M. Karplus, D. York, and M. Karplus, "CHARMM: the biomolecular simulation program.," *Journal of computational chemistry*, vol. 30, pp. 1545–614, jul 2009.

[67] H. Jelger, D. Peter, de Vries, S. J. Marrink, H. Jelger Risselada, S. Yefimov, D. Peter Tieleman, and A. H. de Vries, "The MARTINI force field The MARTINI Force Field: Coarse Grained Model for Biomolecular Simulations," *Journal of Physical Chemistry B*, vol. 111, no. 27, pp. 7812–7824, 2007.

[68] 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.*, vol. 4, no. 3, pp. 435–447, 2008.

[69] N. Schmid, A. Eichenberger, A. Choutko, S. Riniker, M. Winger, A. Mark, and W. Van Gunsteren, "Definition and testing of the GROMOS force-field versions 54A7 and 54B7," *Eur. Biophys. J.*, vol. 40, no. 7, pp. 843–856, 2011.

[70] 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.*, vol. 25, no. 13, pp. 1656–1676, 2004.

[71] W. Van Gunsteren and M. Karplus, "Protein dynamics in solution and in a crystalline environment: a molecular dynamics study," *Biochemistry*, vol. 21, no. 10, pp. 2259–2274, 1982.

[72] G. V. Gurskaya, *Molecular structure of amino acids; determination by X-ray diffraction analysis*. 1968.

[73] R. Eisenschitz and F. London, "Über das verhältnis der van der waalsschen kräfte zu den homöopolaren bindungskräften," *Zeitschrift für Physik*, vol. 60, pp. 491–527, jul 1930.

[74] T. Darden, D. York, and L. Pedersen, "Particle mesh ewald: An n log (n) method for ewald sums in large systems," *J. Chem. Phys.*, vol. 98, pp. 10089–10092, 1993.

[75] A. Y. Toukmaji and J. A. Board, "Ewald summation techniques in perspective: a survey," *Computer Physics Communications*, vol. 95, pp. 73–92, 1996.

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

[77] P. Gibbon and G. Sutmann, "Long-Range Interactions in Many-Particle Simulation," vol. 10, pp. 3–0, 2002.

[78] 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.*, vol. 10, pp. 381–390, 2014.

[79] P. Hünenberger and J. McCammon, "Effect of artificial periodicity in simulations of biomolecules under Ewald boundary conditions: a continuum electrostatics study," *Biophysical Chemistry*, vol. 78, no. 1, pp. 69–88, 1999.

[80] D. Vila-Viçosa, V. H. Teixeira, H. A. F. Santos, A. M. Baptista, and M. Machuqueiro, "On the treatment of ionic strength in biomolecular simulations of charged lipid bilayers," *J. Chem. Theory Comput.*, vol. 10, pp. 5483–5492, 2014.

[81] 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.*, vol. 6, pp. 325–336, JAN 2010.

[82] D. Poger, W. F. Van Gunsteren, and A. E. Mark, "A new force field for simulating phosphatidylcholine bilayers," *J. Comput. Chem.*, vol. 31, no. 6, pp. 1117–1125, 2010.

[83] 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.*, vol. 102, pp. 5451–5459, 1995.

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

[85] H. J. Berendsen, J. P. Postma, W. F. van Gunsteren, and J. Hermans, "Interaction models for water in relation to protein hydration," in *Intermolecular forces*, pp. 331–342, Springer, 1981.

[86] S. Izadi, R. Anandakrishnan, and A. V. Onufriev, "Building Water Models: A Different Approach," *The Journal of Physical Chemistry Letters*, vol. 5, pp. 3863–3871, nov 2014.

[87] M. W. Mahoney and W. L. Jorgensen, "A five-site model for liquid water and the reproduction of the density anomaly by rigid, nonpolarizable potential functions," *The Journal of Chemical Physic*, p. 8910, may 2000.

[88] H. B. Curry, "The method of steepest descent for non-linear minimization problems," *Quarterly of Applied Mathematics*, vol. 2, no. 3, pp. 258–261, 1944.

[89] C. Zhu, R. H. Byrd, P. Lu, and J. Nocedal, "Algorithm 778: L-bfgs-b: Fortran subroutines for large-scale bound-constrained optimization," *ACM Transactions on Mathematical Software (TOMS)*, vol. 23, no. 4, pp. 550–560, 1997.

[90] S. Nosé, "A molecular dynamics method for simulations in the canonical ensemble," *Molecular Physics*, vol. 52, no. 2, pp. 255–268, 1984.

[91] W. K. Hastings, "Monte carlo sampling methods using markov chains and their applications," *Biometrika*, vol. 57, no. 1, pp. 97–109, 1970.

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

[93] H. Berendsen and W. Van Gunsteren, "Practical algorithms for dynamic simulations," *Molecular-dynamics simulation of statistical-mechanical systems*, pp. 43–65, 1986.

[94] 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.*, vol. 81, pp. 3684–3690, 1984.

[95] G. Bussi, D. Donadio, and M. Parrinello, "Canonical sampling through velocity rescaling," *J. Chem. Phys.*, vol. 126, p. 014101, 2007.

[96] W. F. van Gunsteren and H. J. C. Berendsen, "Computer simulation of molecular dynamics: Methodology, applications, and perspectives in chemistry," *Angewandte Chemie International Edition in English*, vol. 29, no. 9, pp. 992–1023, 1990.

[97] S. Liem, D. Brown, and J. Clarke, "Molecular dynamics simulations on distributed memory machines," *Computer Physics Communications*, vol. 67, no. 2, pp. 261 – 267, 1991.

[98] K. J. Bowers, R. O. Dror, and D. E. Shaw, "Zonal methods for the parallel execution of range-limited n-body simulations," *Journal of Computational Physics*, vol. 221, no. 1, pp. 303 – 329, 2007.

[99] B. Hess, "P-LINCS: A Parallel Linear Constraint Solver for Molecular Simulation," *J. Chem. Theory Comput.*, vol. 4, no. 1, pp. 116–122, 2008.

[100] 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.*, vol. 18, pp. 1463–1472, 1997.

[101] 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.*, vol. 21, no. 6, pp. 1087–1092, 1953.

[102] D. Andelman, "Electrostatic properties of membranes: the poisson-boltzmann theory," *Handbook of biological physics*, vol. 1, pp. 603–642, 1995.

[103] 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*, vol. 109, pp. 14691–14706, 2005.

[104] B. Honig, K. Sharp, and A. S. Yang, "Macroscopic models of aqueous solutions: biological and chemical applications," *The Journal of Physical Chemistry*, vol. 97, pp. 1101–1109, feb 1993.

[105] C. J. Cramer and D. G. Truhlar, "Implicit solvation models: equilibria, structure, spectra, and dynamics," *Chemical Reviews*, vol. 99, no. 8, pp. 2161–2200, 1999.

[106] R. J. Zauhar and A. Varnek, "A fast and space-efficient boundary element method for computing electrostatic and hydration effects in large molecules," *Journal of Computational Chemistry*, vol. 17, no. 7, pp. 864–877, 1996.

[107] A. Nicholls and B. Honig, "A rapid finite difference algorithm, utilizing successive over-relaxation to solve the poisson-boltzmann equation," *J. Comput. Chem.*, vol. 12, pp. 435–445, Mar. 1991.

[108] 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.*, vol. 23, no. 1, pp. 128–137, 2002.

[109] D. Bashford and M. Karplus, "pk_a's of ionizable groups in proteins: atomic detail from a continuum electrostatic model," *Biochemistry*, vol. 29, no. 44, pp. 10219–10225, 1990.

[110] 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.*, vol. 11, no. 5, pp. 2367–2376, 2015.

[111] 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.*, vol. 7, no. 5, pp. 1237–1243, 2011.

[112] 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.*, vol. 26, pp. 1701–1718, 2005.

[113] P. Smith and W. van Gunsteren, "Consistent dielectric properties of the simple point charge and extended point charge water models at 277 and 300 K," *J. Chem. Phys.*, vol. 100, pp. 3169–3174, 1994.

[114] S. Miyamoto and P. Kollman, "SETTLE: An analytical version of the SHAKE and RATTLE algorithm for rigid water models," *J. Comput. Chem.*, vol. 13, no. 8, pp. 952–962, 1992.

[115] 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.*, vol. 5, no. 1, p. 9, 2012.

[116] M. Gilson, K. Sharp, and B. Honig, "Calculating the eletrostatic potential of molecules in solution: Method and error assessment," *J. Comput. Chem.*, vol. 9, pp. 327–335, 1987.

[117] 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*, vol. 105, pp. 293–309, 2001.

[118] S. Behnel, R. Bradshaw, C. Citro, L. Dalcin, D. Seljebotn, and K. Smith, "Cython: The best of both worlds," *Computing in Science Engineering*, vol. 13, no. 2, pp. 31 –39, 2011.

[119] L. E. Cybulski and D. de Mendoza, "Bilayer hydrophobic thickness and integral membrane protein function," *Current Protein Peptide Science*, vol. 12, no. 8, pp. 760–766, 2011.

[120] N. Kučerka, M.-P. Nieh, and J. Katsaras, "Fluid phase lipid areas and bilayer thicknesses of commonly used phosphatidylcholines as a function of temperature," *Biochem. Biophys. Acta, Biomembr.*, vol. 1808, pp. 2761–2771, 2011.

[121] N. Garnier, S. Crouzy, and M. Genest, "Molecular Dynamics Simulations of the Transmembrane Domain of the Oncogenic ErbB2 Receptor Dimer in a DMPC Bilayer," *J.Biomol.Struct.Dyn.*, vol. 21, pp. 179–200, oct 2003.

[122] K. Akabori, "Structure Determination of HIV-1 Tat/fluid Phase Membranes and DMPC Ripple Phase Using X-Ray Scattering," *Dissertations*, sep 2014.

[123] J. Nagle and S. Tristram-Nagle, "Structure of lipid bilayers," *Biochem. Biophys. Acta, Biomembr.*, vol. 1469, no. 3, pp. 159–195, 2000.

[124] V. Luzzati and F. Husson, "The structure of the liquid-crystalline phasis of lipid-water systems.," *The Journal of cell biology*, vol. 12, pp. 207–19, feb 1962.

[125] N. Kucerka, J. F. Nagle, J. N. Sachs, S. E. Feller, J. Pencer, A. Jackson, and J. Katsaras, "Lipid bilayer structure determined by the simultaneous analysis of neutron and X-ray scattering data.," *Biophysical journal*, vol. 95, pp. 2356–67, sep 2008.

[126] 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*, vol. 110, no. 15, pp. 5834–5839, 2013.

[127] M. Musial-Siwek, A. Karabadzhak, O. A. Andreev, Y. K. Reshetnyak, and D. M. Engelman, "Tuning the insertion properties of pHLIP," *BBA-Biomembranes*, vol. 1798, no. 6, pp. 1041–1046, 2010.

[128] J. A. Wallace and J. K. Shen, "Continuous constant pH molecular dynamics in explicit solvent with pH-based replica exchange," *Journal of Chemical Theory and Computation*, vol. 7, no. 8, pp. 2617–2629, 2011.

[129] D. Skoog and D. West, “Fundamental of analytical chemistry: Holt-saunders international,” 1982.

APPENDIX A

Equilibration

Our system was built from a previously equilibrated lipid bilayer. Since the pentapeptides started away from the membrane ($\sim$ 10 Å) and they have a limited available conformational space, well equilibrated systems were expected and found.

We have followed properties of both the lipid and the peptide to ensure their equilibration. Here, we show only the analysis performed on the His pentapeptide system since all other behaved similarly. The lipid membrane thickness and area per lipid values fluctuated within an adequate range, as did the radius of gyration of the pentapeptide. Henceforth, the system is apparently well equilibrated from the beginning. However, we have considered the first 20 ns of simulation as equilibration to safeguard any slower properties not checked.

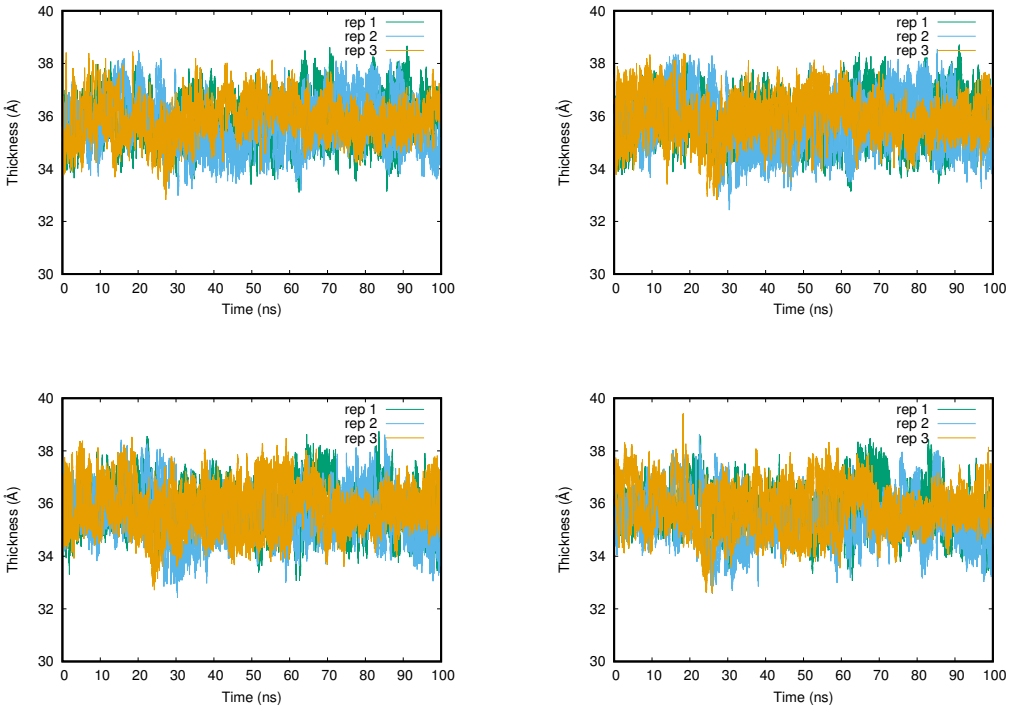

Figure A.1: Time dependent POPC membrane thickness in the presence of the His peptide in the pHRE simulations with 4 pH values (one per plot) and a 100 ps τ_{RE}.

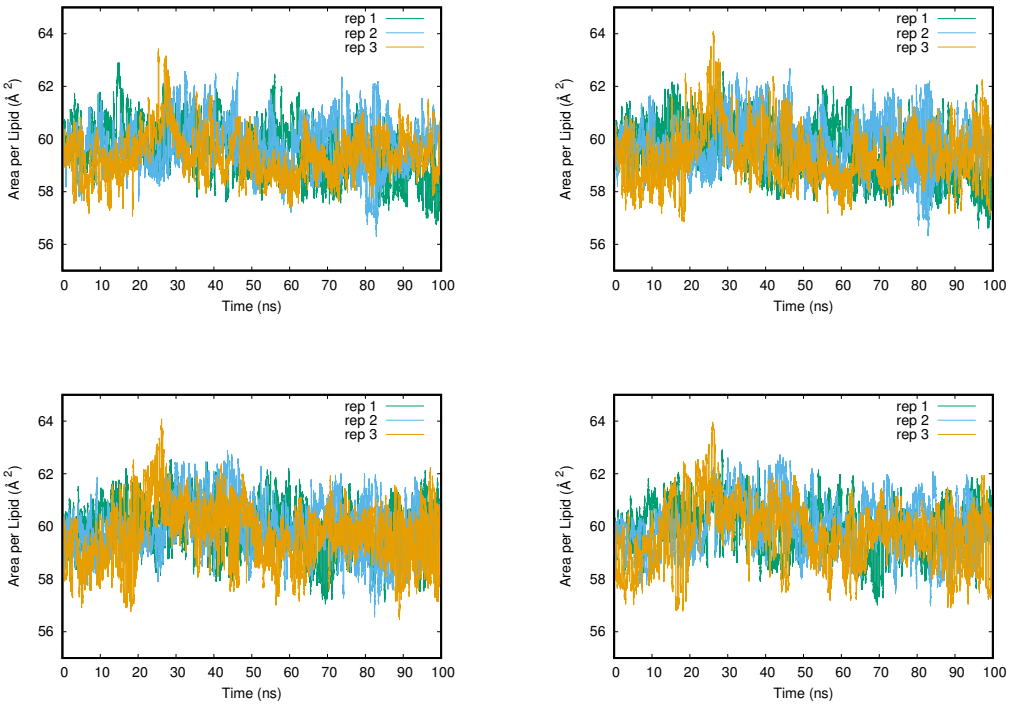

Figure A.2: Time dependent POPC membrane area per lipid in the presence of the His peptide in the pHRE simulations with 4 pH values (one per plot) and a 100 ps τ_{RE}.

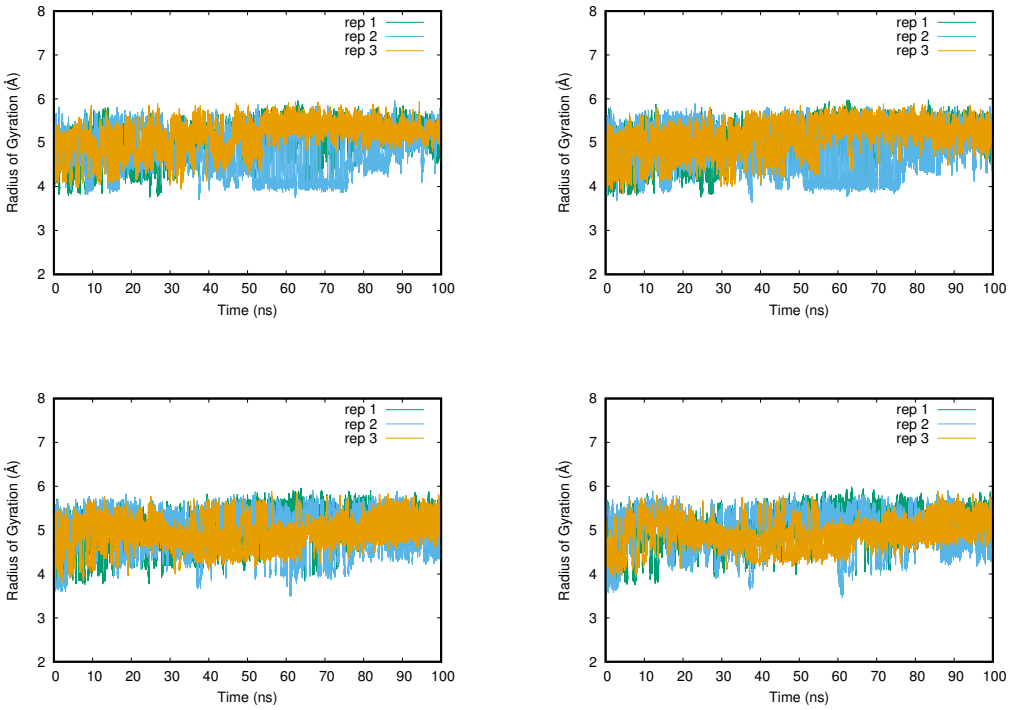

Figure A.3: Time dependent radius of gyration of the His peptide in the pHRE simulations with 4 pH values (one per plot) and a 100 ps τ_{RE}.

APPENDIX B

Replica Exchange Efficiency

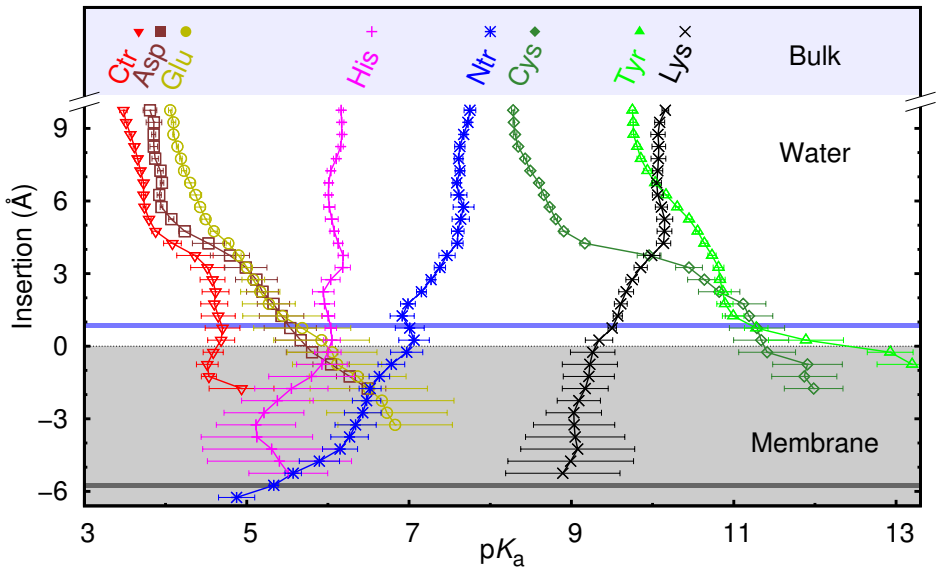

Figure B.1: CpHMD pK_a insertion profiles of all titrable amino acids using the closest phosphate found in ref 1. The water pK_a values of these residues are shown on top. The average position of cholines (blue) and the initial carbon of the acylchain (grey) are also shown.

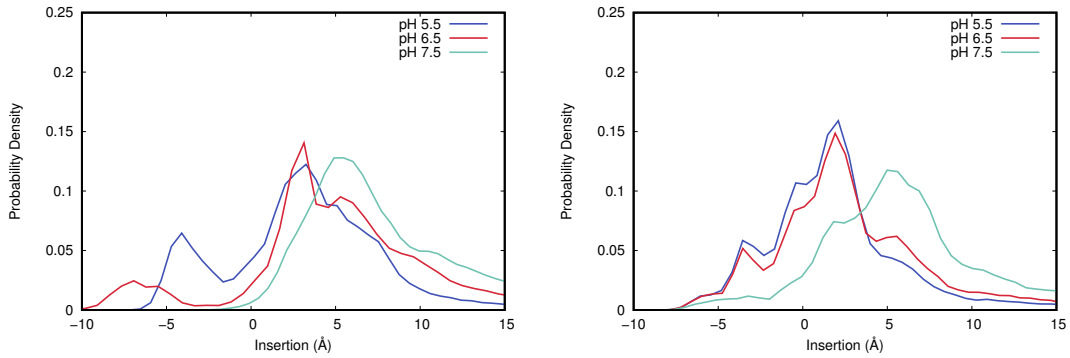

Figure B.2: Insertion distribution of the C-ter pentapeptide in CpHMD (left) and pHRE (right)

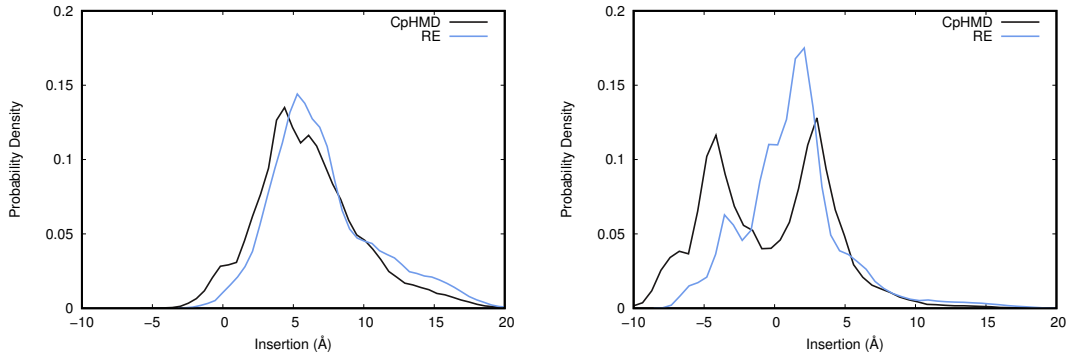

Figure B.3: Insertion distribution of the C-ter pentapeptide in CpHMD and pHRE of the deprotonated (left) and protonated (right) state.

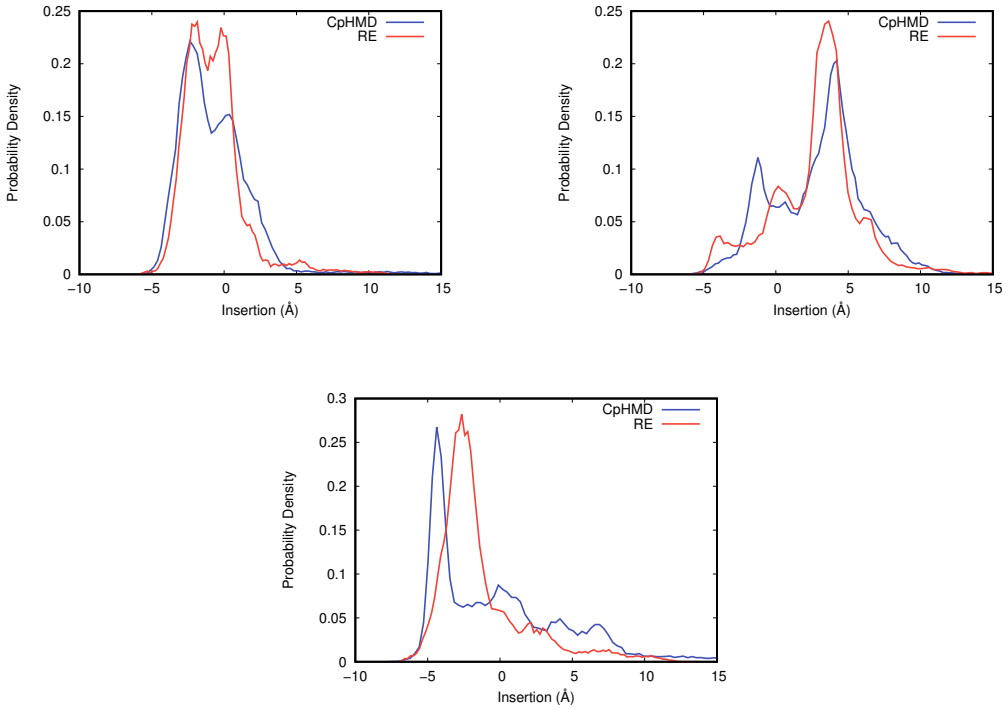

Figure B.4: Insertion distribution of the His pentapeptide CpHMD and pHRE replicas with similar insertion sampling at 5.50 (left), 06.50 (right) and 7.50 (bottom) pH values.

APPENDIX C

Insertion Criteria Influence on the Sampled Space

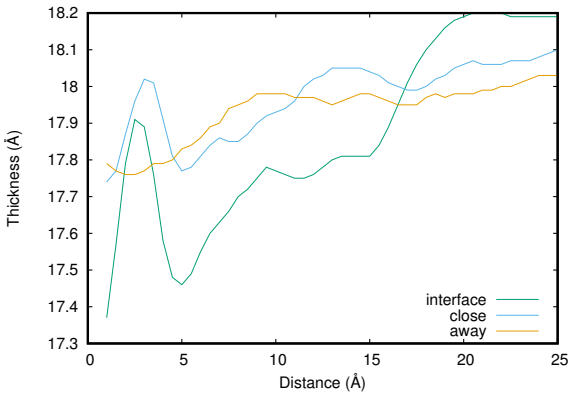

Figure C.1: Closest interacting monolayer thickness profile of the Glu peptides in three distinct insertion regions: away from the membrane (at a 10 Å distance), closer (from 5 Å to 10 Å) and in the interface (from -5 Å to 5 Å insertion values). These calculations were calculated using the method described in Section 2.9.3 and a moving average with a 2 Å window size and a step of 0.5 Å.

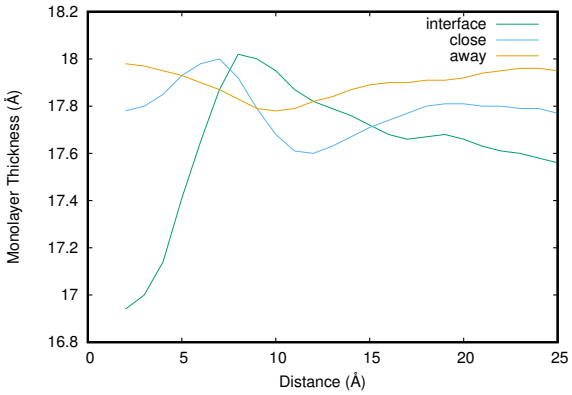

Figure C.2: Closest interacting monolayer thickness profile of the C-ter peptides in three distinct insertion regions: away from the membrane (at a 10 Å distance), closer (from 5 Å to 10 Å) and in the interface (from -5 Å to 5 Å insertion values). These calculations were calculated using the method described in Section 2.9.3 and a moving average with a 2 Å window size and a step of 0.5 Å.

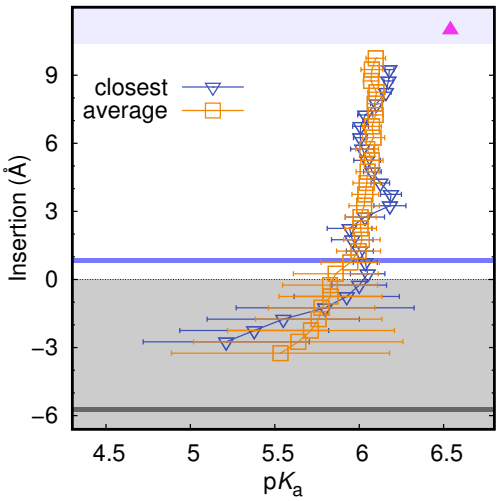

Figure C.3: pK_a insertion profiles of His peptide using the closest phosphate (closest) and all the lipids (average). The simulations were performed in 3 pH values using CpHMD method. The water pK a values of these residues are shown on top, in a filled triangle bullet. The average position of cholines (blue) and the initial carbon of the acylchain (grey) are also shown.

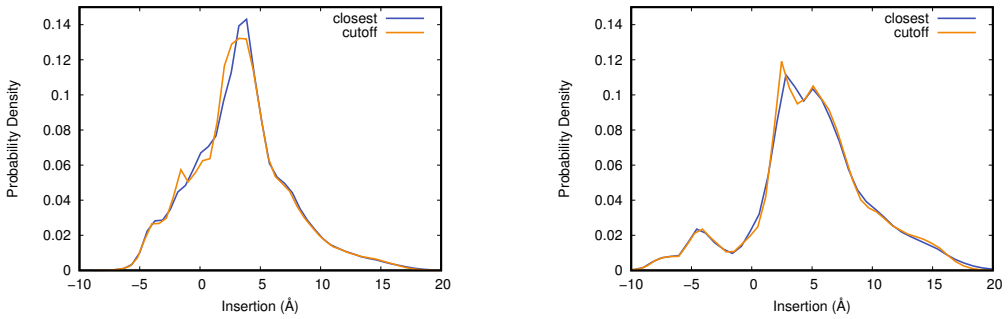

Figure C.4: Insertion distribution of the Glu (left) and C-ter (right) pentapeptide in CpHMD.

APPENDIX D

pHRE as an improvement over CpHMD

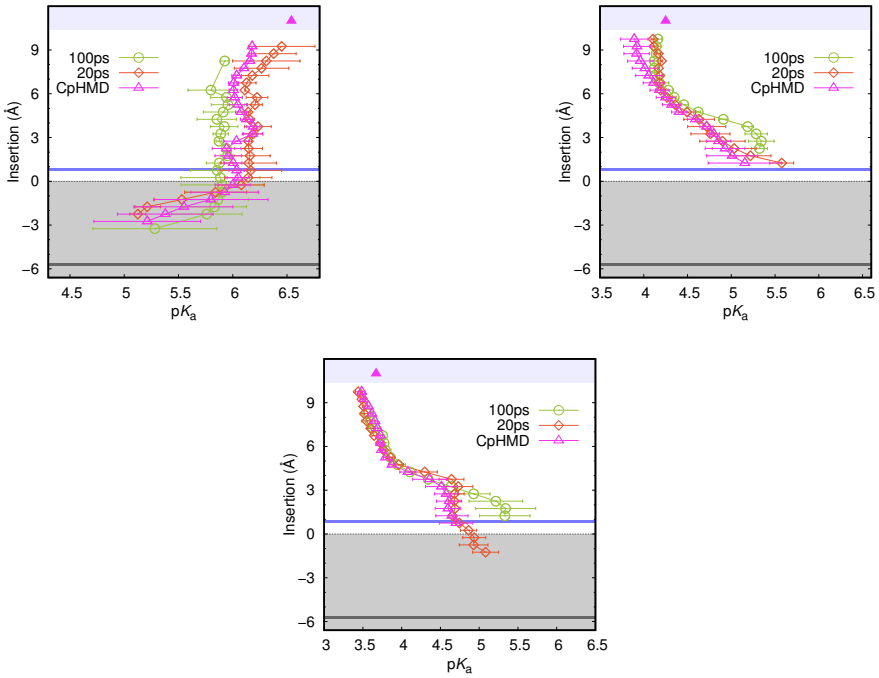

Figure D.1: pK_a insertion profiles of His (left), Glu (right) and C-ter (bottom) peptide using the closest phosphate (closest) and all the lipids (average). The simulations were performed in 3 pH values using CpHMD method. The water pK a values of these residues are shown on top, in a filled triangle bullet. The average position of cholines (blue) and the initial carbon of the acylchain (grey) are also shown.

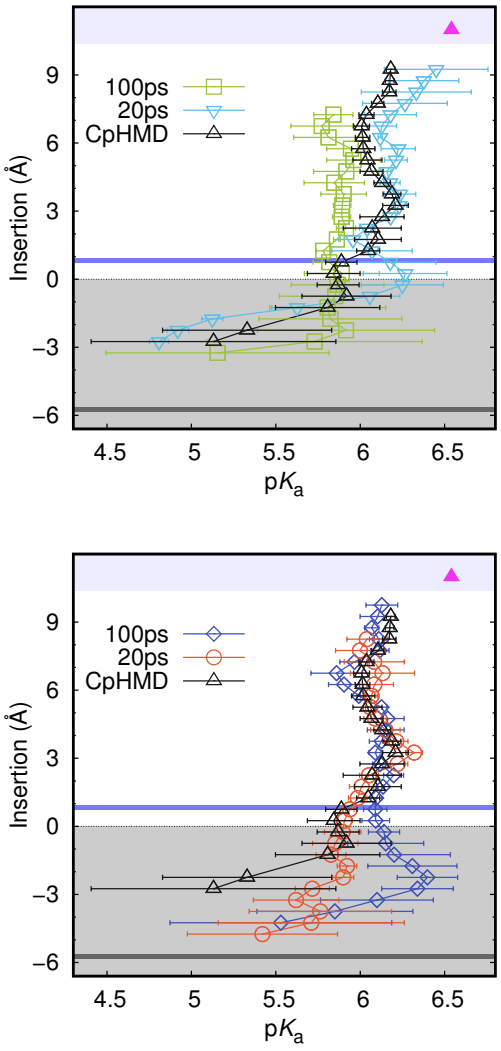

Figure D.2: pK_a insertion profiles of His peptide. The simulations were performed in 3 (top) and 4 (bottom) pH values using CpHMD method. The water pK_a values of these residues are shown on top, in a filled triangle bullet. The average position of cholines (blue) and the initial carbon of the acylchain (grey) are also shown.