UNIVERSIDADE DE LISBOA
FACULDADE DE CIÊNCIAS
DEPARTAMENTO DE QUÍMICA E BIOQUÍMICA

INTERACTION BETWEEN THE NEUROPEPTIDE KYOTORPHIN AND A LIPID BILAYER: A CONSTANT-PH MOLECULAR DYNAMICS STUDY

Pedro Rafael da Silva Álvaro Magalhães

MESTRADO EM BIOQUÍMICA

2009

UNIVERSIDADE DE LISBOA
FACULDADE DE CIÊNCIAS
DEPARTAMENTO DE QUÍMICA E BIOQUÍMICA

INTERACTION BETWEEN THE NEUROPEPTIDE KYOTORPHIN AND A LIPID BILAYER: A CONSTANT-PH MOLECULAR DYNAMICS STUDY

Pedro Rafael da Silva Álvaro Magalhães

Tese orientada por Doutor António M. Baptista e Prof. Doutor António Ferreira

MESTRADO EM BIOQUÍMICA

2009

Now these points of data make a beautiful line.
And we're out of beta. We're releasing on time.

-GLaDOS, Portal

Abstract

Kyotorphin is an endogenous dipeptide (L-tyrosine-L-arginine) with analgesic properties. It shares some characteristics with opioid peptides, but unlike them, does not bind to opiate receptors, possessing instead a specific receptor, which is yet to be identified. It has been hypothesized that the binding process of kyotorphin to its receptor is influenced by the biological membrane, according to the so-called 'membrane catalysis' model. Another important factor regulating the binding process is the solution pH, since it has been shown that kyotorphin has a strong conformational dependence on pH.

We intend to study the effect of pH on the conformational space of kyotorphin in the presence of a biological membrane model. To accomplish this, we employed molecular simulation methods which allow the specification of pH: the so-called constant-pH molecular dynamics methods. One major improvement of the current implementation of the constant-pH molecular dynamics method was the addition of a new module which allows the explicit simulation of membranes. Kyotorphin was simulated in water in the presence and in the absence of a membrane for 100 ns, and within the pH range 2-12.

Here, we present the first constant-pH molecular dynamics simulation of a peptide with the explicit inclusion of a lipid membrane, showing that kyotorphin favors the lipid phase, as has been shown experimentally, and that membrane penetration occurs through both amino acid residues. We also found that kyotorphin's conformational space remains heavily dependent on pH, with few differences being observed between the water and membrane systems. Additionally we have identified some particular cases hitherto unobserved. We have also fitted our peptide to part of the structure of morphine, witnessing an almost perfect overlap, meaning that it should be able to fulfill some of the structural constraints imposed by an opioid-like receptor.

Keywords: kyotorphin, membrane, pH, molecular modeling, constant-pH molecular dynamics

Resumo

A quiotorfina é um dipéptido endógeno com propriedades analgésicas. Foi encontrada pela primeira vez num encéfalo bovino, e desde então foi também isolada a partir de encéfalos de várias espécies, incluindo humanos. O seu efeito analgésico é 4.2 vezes superior ao da Metionina-encefalina, um pentapéptido opióide endógeno. Embora apresente uma actividade opióide, mostrou-se que a quiotorfina não liga aos receptores opióides, o que levou à hipótese de um mecanismo de acção indirecto, através da ligação a um receptor específico. Como ainda não foi identificado o receptor específico, os detalhes do mecanismo de acção da quiotorfina são desconhecidos, tendo sido no entanto sugeridas duas hipóteses: a activação do receptor pela quiotorfina leva à libertação de Met-encefalina, que leva à activação de um receptor opióide-δ; ou a quiotorfina é degradada rapidamente e a L-arginina formada é utilizada pelo enzima NO sintase que leva à formação de NO, que por sua vez induz a libertação de Met-encefalina.

A quiotorfina foi classificada como neuropéptido devido a vários motivos: (i) pode ser formada a partir da biossíntese dos aminoácidos tirosina e arginina ou pelo processamento enzimático de proteínas precursoras; (ii) é degradada rapidamente nos terminais nervosos; (iii) encontra-se distribuída em diversas áreas do cérebro, com níveis mais elevados em zonas envolvidas na inibição da dor; (iv) a sua sequência (L-tirosina-L-arginina) encontra-se presente em outros péptidos analgésicos; (v) é libertada em resposta a estímulos despolarizantes; e (vi) liga a receptores acoplados a mecanismos de mobilização de Ca^{2+}.

Como se trata de um pequeno péptido, a quiotorfina não possui uma conformação prevalecente em solução, apresentando um conjunto de estruturas interconvertíveis. O equilíbrio das populações conformacionais está intensamente dependente do pH da solução, favorecendo conformações em *trans* (mais estendidas) a valores baixos de pH, e conformações *cis* (mais compactas) a valores altos. Como a quiotorfina possui quatro grupos tituláveis, a grande dependência do seu estado de protonação com o pH é algo esperado. A influência do pH na conformação da quiotorfina é portanto de grande interesse para os estudos de interacção com o seu receptor, já que pode ser um factor mediador desta interacção.

Para além do pH, existem outros factores que regulam a interacção da quiotorfina com o seu receptor. Foi hipotetizado, em 1986 por Sargent e Schwyzer, que a membrana lipídica pode ser um factor de extrema importância no processo de ligação de pequenos péptidos aos seus receptores, segundo um modelo que apelidaram de "catálise membranar". Segundo esta teoria, um pequeno péptido com uma grande liberdade conformacional em solução, ao colidir com a membrana, e

influenciado pela difusão e interacções electrostáticas, desenvolveria uma estrutura mais ordenada. A membrana serviria então como um mediador da interacção entre o péptido e o seu receptor, favorecendo uma conformação mais apta à formação do complexo receptor-ligando.

Pretendemos, neste trabalho, estudar a interacção da quiotorfina com uma membrana lipídica num intervalo amplo de pH por métodos computacionais. Embora a abordagem mais comum para a resolução de um problema biológico seja por via experimental, sistemas como o que nos propomos estudar podem beneficiar de uma abordagem teórica/computacional, pelo que optámos por utilizar métodos de simulação molecular. O estudo de péptidos em sistemas membranares requer a consideração simultânea da sua dinâmica conformacional e da protonação/desprotonação dos grupos tituláveis, cujo acoplamento deverá mediar a sua função biológica. Apesar de já ser possível realizar simulações de mecânica/dinâmica molecular de membranas a nível atómico, o tratamento dos grupos tituláveis é bastante limitado nesse tipo de estudos. Com efeito, a conformação e protonação são tratadas de uma forma desacoplada e complementar: os métodos de mecânica/dinâmica molecular tratam adequadamente a dinâmica estrutural de um sistema com estados de protonação fixos, enquanto que os métodos de electrostática simplificados são apropriados para tratar mudanças de protonação num sistema de conformação rígida. Em resposta a este problema, surgiram os métodos de dinâmica molecular a pH constante, que permitem alterações de protonação durante uma simulação de dinâmica molecular. Em vez de definir os estados de protonação fixos que cada centro titulável possuiria a um dado valor de pH, estes métodos permitem que o pH seja definido como um parâmetro externo, como num procedimento experimental típico.

O método de dinâmica molecular a pH constante utilizado neste trabalho é uma variante do método de titulação estocástica previamente utilizado em estudos de proteínas e péptidos em solução. A primeira fase deste trabalho consistiu na adaptação do método para lidar com sistemas membranares, algo que não havia sido implementado ainda. A aplicação desta nova implementação do método de titulação estocástica à interacção da quiotorfina com bicamadas lipídicas é um estudo pioneiro nesta área.

Este trabalho tem então três objectivos principais: (i) a adaptação do método de titulação estocástica para permitir a inclusão explícita de membranas; (ii) a simulação da quiotorfina na presença de uma membrana lipídica a diferentes valores de pH utilizando esta nova implementação do método; e (iii) a comparação dos resultados obtidos com os obtidos previamente por estudos experimentais e computacionais.

Os resultados obtidos mostram que a quiotorfina possui um perfil de titulação com pequenas

diferenças quando em solvente aquoso ou em membrana, e que mostra concordância com o perfil obtido por métodos computacionais semelhantes. As curvas individuais dos quatro grupos tituláveis da quiotorfina exibem diferenças devido à presença da membrana. O grupo terminal amina da quiotorfina, o mais importante devido à sua titulação a valores fisiológicos de pH, possui um valor de pK_a mais baixo na presença da membrana, o que vem contrapor os resultados de outros estudos computacionais. No entanto, convém referir que o nosso método permite fazer simulações fisiologicamente mais realistas, pelo que discutimos as diferenças obtidas entre ambos os métodos.

Relativamente aos aspectos conformacionais da quiotorfina, os resultados que obtivemos são concordantes com os obtidos por outros autores. Existe uma dependência conformacional da quiotorfina com o pH, o que era esperado devido ao número de centros tituláveis do péptido. A tendência para a quiotorfina adoptar estruturas mais compactas ou mais estiradas demonstra uma dependência com o pH de acordo com o que foi previamente observado. Mostramos ainda que existe uma limitação conformacional da quiotorfina quando na presença da membrana, resultando na perda, embora ligeira, de rotações internas, o que está de acordo com o modelo de "catálise membranar".

A inserção da quiotorfina na membrana é também influenciada pelo pH, que afecta inclusivamente a penetração relativa dos seus diferentes grupos. À medida que o pH se torna mais alcalino, existe uma tendência de afastamento do péptido em relação à membrana, o que está de acordo com resultados obtidos experimentalmente, que mostravam que a constante de partição do péptido a valores de pH mais ácidos era mais elevada do que a valores mais alcalinos. Ao observar a inserção do péptido na membrana, deparámo-nos com alguns casos particulares previamente inobservados, nomeadamente afastamentos esporádicos a valores de pH elevados. Discutimos esses casos individualmente, concluindo que estão associados a protonações/desprotonações facultadas pelo método de dinâmica molecular a pH constante utilizado, e que a sua ocorrência é mais uma vez consistente com as partições observadas experimentalmente.

Finalmente, fizemos um *fit* das melhores estruturas da quiotorfina sobre parte da estrutura da morfina, um analgésico opióide cujo receptor está bem descrito, e observámos uma sobreposição quase perfeita, o que demonstra que a quiotorfina é capaz de reproduzir algumas das restrições conformacionais impostas pela presença de um receptor do tipo opióide.

Palavras-chave: quiotorfina, membrana, pH, modelação molecular, dinâmica molecular a pH constante.

Acknowledgments

Ah... the acknowledgments... Surely everyone's favorite part of any thesis or project, reader and writer alike. Although enjoyable, it is also the most difficult section to write and definitely the most unfinished*. Having said that, I would like to attribute this to both a lack of sampling and human error.

First, I would like to thank my supervisor Dr. António Baptista for all the time and patience he devoted to revising this thesis, for everything he has taught (and will teach, hopefully) me, and also for his great disposition and sense of humor.

Next on the list, is Dr. Miguel Machuqueiro, whom I would deeply like to thank for moving to the Faculty of Sciences in Lisbon, and leaving his terribly noisy computer alone and silent. And I guess I should also be thanking him for all of his helpful discussions, teachings, insightful comments, and generally amazing supportive attitude, but I think the computer thing is much more relevant.

I would also like to thank everyone in the Molecular Simulation and Protein Modeling groups at ITQB-UNL for the wonderful work environment and for the fruitful and sometimes nonsensical (but always scientific) discussions. So, I would like to thank, in no particular order: Diana and Luís for their support during my initial refugee period in their room, and Cláudio for allowing me to work in said room and letting me bother everyone with my questions; João for all the tech support help, and also for never reporting my whereabouts to any mental institutions and/or to the proper authorities; Sara for the time spent teaching me how to use the PCA programs and also for being able to laugh at some of my more inappropriate jokes; Sofia for all the cakes, past, present and future, and for her constantly contagious cheerful attitude; Bruno for threatening me with political repercussions if I didn't acknowledge him (on that topic, vote for Bruno Victor for Mayor – thesis public service announcement paid in full by the Bruno Victor for Mayor Commission) as well as for all the help he provided in my most "I think I just broke something, please help" moments; Rita for the weekly visits, the cakes and for being so easy to spook (hah!); and lastly, both my fellow masters colleagues, Carla and Luís Carlos for sharing the burden of being the new guys. Also, I would like to thank everyone for not hitting me that time I accidentally tried to plug my headphones in the reset button.

I would like to thank Professor António Ferreira for the opportunity to work in a simulation group, and also for his many computer related classes, without which I doubt I would have so easily found something I truly enjoy doing. On that note, I would also like to thank all the teachers at the Faculty of Sciences (too many to name), whose classes, although mostly computer unrelated, were at times no less enticing.

I would like to thank my countless, countless friends (data not shown). All five of them are (probably) cheering for me, although that might be due to the fact they can't bear to listen to me talking about this anymore.

Lastly, I would like to thank my parents for funding. And also for their support and firm belief that someday their son might devote his time to something that does not involve a computer.

Thank you all.

* - all acknowledgments were calculated using standard methods; for all of them, the error was computed as the correlation-corrected standard deviation of the average, taking as correlation time the value at which the autocorrelation goes below 0.1

Index

Table index

Figure index

1. Introduction

1.1. Kyotorphin

1.1.1. Discovery and characterization

Kyotorphin (KTP) is an endogenous dipeptide (L-Tyr-L-Arg) (Figure 1.1) with analgesic properties. It was first isolated in 1979, from a bovine brain, by Hiroshi Takagi and his group at the University of Kyoto in Japan. The peptide was named 'kyotorphin' because it was discovered in Kyoto and exhibited endorphin-like properties [**Takagi1979a**].

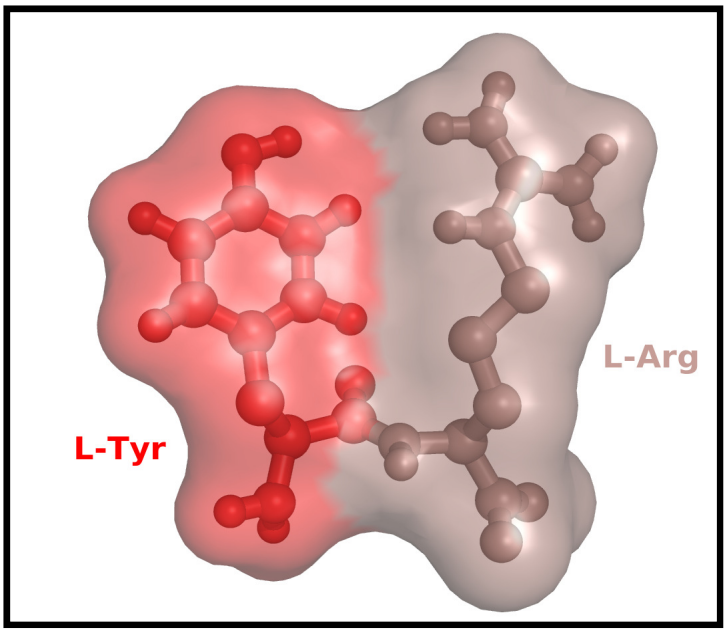

Figure 1.1 – Kyotorphin (KTP) depicted with differently colored amino acid residues.

When it was first isolated, its analgesic effect was measured. An intracisternal administration revealed it was 4.2 times more potent than methionine-enkephalin (Met-enk) [**Takagi1979a**], an endogenous opioid pentapeptide that acts as a neurotransmitter regulating pain transmission in the central nervous system [**Lewis1983**, **Takagi1979b**]. It was shown that, despite its opioid activity, KTP did not bind to the specific opiate receptors (μ-, δ-, κ-), which lead to the assumption that its effect was mediated by enkephalins. It was proposed that KTP acted as a Met-enk releasing agent [**Takagi1979b**], which was

later confirmed when a specific radioimmunoassay showed that Met-enk was released by KTP from the slices of isolated guinea pig striatum and spinal cord [**Shiomi1981a**]. In addition to its higher analgesic potential, KTP exhibited a relatively long lasting effect [**Takagi1979a**], which was attributed to a stabilizing effect on the released Met-enk by a weak inhibition of the enkephalin degrading enzymes [**Takagi1979b**]

In 1980 and 1981, Takagi and his group found that KTP was also present in the brain of rats, mice, guinea pigs, rabbits and humans [**Shiomi1981a**, **Shiomi1981b**, **Ueda1980**], with the highest levels being detected in the lower brain stem, midbrain, hypothalamus, pons plus medulla oblongata and the dorsal part of the spinal cord, all of which are areas known to play a physiological role in the pain inhibitory systems [**Shiomi1981a**].

There are two pathways for the synthesis of KTP; it can be formed either by: (i) biosynthesis from the amino acids tyrosine and arginine, in the nerve terminals, by an ATP-dependent synthetase [**Ueda1986a**, **Ueda1987a**] or (ii) by processing precursor proteins [**Ueda1987a**], via membrane-bound, leupeptin-sensitive 'KTP converting enzymes' [**Ueda1985a**].

Its degradation occurs rapidly through membrane-bound aminopeptidases [**Matsubayashi1984**, **Takagi1979a**, **Ueda1985a**, **Ueda1985b**], which, together with the fact that it was found in areas of the brain related with pain inhibition, and it could be synthesized in the nerve terminals, led to its classification as a neuropeptide. Further studies have since solidified the inclusion of KTP in the neuropeptide family, when the L-Tyr-L-Arg sequence was found to be a common motif in other analgesic peptides [**Amano1984**, **Fukui1983**, **Prasad1995**, **Ueda1987b**], and when it was found that KTP was released in response to depolarizing stimuli and binds to receptors coupled to Ca^{2+} mobilization mechanisms [**Ueda1986b**] [**Ueda1987c**].

1.1.2. Mechanism of action

The analgesic effect induced by kyotorphin can be eliminated through a pretreatment with naloxone, an antagonist that binds to opiate receptors [**Takagi1979a**]. However, in spite of its opioid activity, and as previously mentioned, KTP does not bind to the opiate receptors [**Rackham1982**, **Takagi1979b**], which suggests an indirect mechanism of action via binding to a specific receptor, KTPr [**Ueda1989**, **Ochi2000**]. The presence of this receptor has been confirmed [**Shiomi1981a**, **Takagi1979b**] but it

remains to be identified [**Ueda1989**].

Two possible mechanisms were proposed for the analgesic effect of kyotorphin in the brain:

1. KTP binds to its specific receptor and induces the release of Met-enk [**Rackham1982**, **Shiomi1981a**], which in turn, binds to the μ-receptor and activates the α subunit of the G-protein (G_{α}^{i}) and phospholipase C [**Shiomi1981a**, **Ueda1989**];

2. KTP is rapidly degraded to L-Arg, a substrate for the neuronal nitric oxide synthase (nNOS). The enzyme then produces nitric oxide (NO), a Met-enk release activator [**Arima1997**]. Met-enk then acts as in mechanism 1.

Since the details of the interaction between KTP and its receptor are currently unknown, one must take into consideration opioid peptides or other compounds similar to KTP whose relationship with the receptors is better understood. An example of such a molecule is morphine. The tyrosine residue in KTP, which in other opioid peptides is believed to be involved in the recognition by the receptor [**Lapalu1997**, **Lapalu1998**], can be mapped on morphine, fully matching its aromatic ring [**Machuqueiro2007**]. Another similarity between the two lies in the N-terminus of KTP, which when protonated can form a salt bridge with an anionic group in the receptor, which typically happens in the morphine/receptor interaction [**Patrick2001**]. Also, when the N-terminus group of KTP is protonated, its net charge is positive, which is an important characteristic present in the cationic peptides of dynorphin A and nociceptin/orphanine FQ [**Machuqueiro2007**].

There is a possibility that the L-Tyr residue and the N-terminus group of KTP might play an important role in receptor recognition. Both are protonatable groups (together with two other groups in KTP, the C-terminus and the L-Arg side chain) and as such, are influenced by the pH of the surrounding environment. This raises the question of the influence of pH in the receptor binding process, which we will explore in detail in our studies.

1.1.3. Role of pH

pH is an important factor in biochemical processes. It is responsible for changes in the properties of molecules in biological systems, affecting their structure and function. Specifically, a change in pH will

cause titratable groups to lose or gain protons, thus altering the charge of the titratable group and the net charge of the biomolecule [**Nielsen2009**].

KTP, like most small peptides, does not possess a single prevalent conformation in solution, but rather an ensemble of interconverting structures [**Dyson1991**]. The equilibrium of conformer populations is heavily dependent on pH, meaning that certain pH value changes can alter the distribution and favor specific conformations [**Aburi2002**]. The influence of pH on the conformation of KTP has been studied by different methods, both experimental [**Lopes2006a**] and computational [**Lopes2006a**, **Machuqueiro2007**, **Machuqueiro2010**]. All three studies have shown that KTP displays a preferential conformation in aqueous solution depending on pH: for low values, the *trans* conformation is favored, whereas for high values, the *cis* conformer prevails (Figure 1.2) [**Machuqueiro2007**].

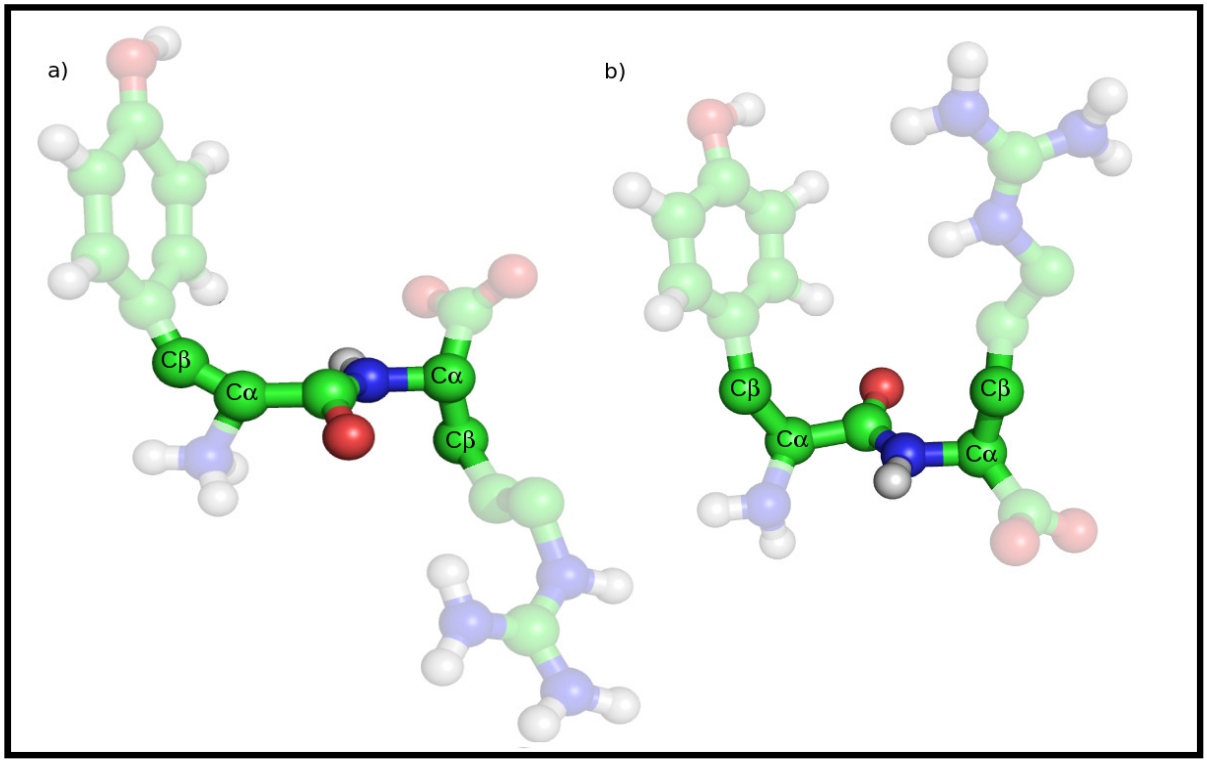

Figure 1.2 – Examples of representative conformations of KTP: the *trans* conformer (a), with the $C\beta_{Tyr}$-$C\alpha_{Tyr}$-$C\alpha_{Arg}$-$C\beta_{Arg}$ dihedral angle closer to 180º and the *cis* conformer (b), with that dihedral angle closer to 0º.

1.1.4. Interaction with biological membranes

In addition to pH, there is another factor of great importance for the interaction between KTP and its receptor: the lipid membrane. Lipid membranes are of vital importance to many biological systems. They interact with different chemical species, from small organic molecules and peptides to proteins and macromolecular complexes.

In 1986, Sargent and Schwyzer [**Sargent1986**] first proposed the so-called "membrane catalysis model", according to which the membrane helps the peptide to attain the necessary conformation for the docking process by raising its local concentration, thus increasing the chance of a successful binding to the receptor. This model, depicted in (Figure 1.3), has the following steps: (1) the amphiphilic peptide **P**, which has little to no order in solution, collides with the membrane and its surface activity is enhanced by diffusion and electrostatic interactions, resulting in the **P'** state; (2) the peptide then suffers internal rotations and insertions between lipids, reducing its translational, conformational and rotational entropy while causing the displacement of bound water (entropy gain). The membrane-bound state **P''** is thus formed; (3) the formation of the (**P-R**) complex between the peptide and its receptor in its inactive conformation takes place; (4) lastly, the concerted rearrangement of the peptide and receptor conformations results in the active peptide-receptor complex (**P-R**)', which in turn will activate the next step in the pathway.

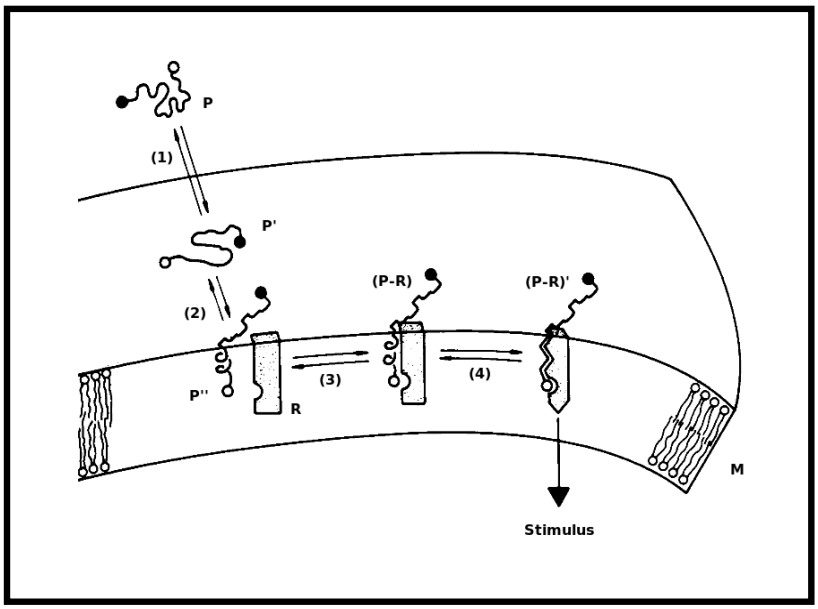

Figure 1.3 – Schematic representation of the membrane catalysis model. Adapted from [**Sargent1986**].

Although KTP exhibits a powerful analgesic effect when administered intracisternally, a systemic introduction shows little activity and only in high concentrations [**Chen1998**]. This is because analgesic peptides like kyotorphin cannot cross the blood brain barrier. Different approaches have been tried over the years to overcome this problem, such as: using an analogue, D-KTP, which is less susceptible to enzymatic degradation [**Matsubayashi1984**]; co-administration of substances that inhibit the degradation of KTP [**Ueda1985b**]; derivatization with lipophilic compounds that facilitate crossing the blood brain barrier [**Lopes2006b**]; or using non-ionic surfactants to increase membrane permeability [**Sakane1989**]. Even though some of the approaches have had some degree of success, KTP has not become more viable as an analgesic drug to justify being marketed.

Even though we know that KTP prefers the membrane environment over the aqueous media (with a high partition coefficient determined experimentally) [**Lopes2006a**], and that the phenolic ring of KTP is exposed for receptor interaction with a well defined orientation relative to the bilayer plane [**Lopes2006a**], at a molecular level very little is known between the interaction between KTP and the lipid membrane. Although the most common approach to any given biological system an experimental one, there might be some advantage in using other methodologies to study the atomic details of this system. In a recent work, soon to be submitted, Machuqueiro *et al.* used computational techniques to study this problem [**Machuqueiro2010**] and observed that KTP is able to penetrate the membrane through both amino acid residues, and witnessed that the peptide loses some internal rotations when interacting with the membrane, which is in accordance with the membrane catalysis hypothesis mentioned. In **Machuqueiro2010**, approximate methods were used and it is our understanding that this system might benefit from a more thorough and detailed computational analysis using more advanced methods that allow the explicit inclusion of pH effects. This approach will also allow the calculation of more accurate pK_a values than the ones determined both by experimental [**Lopes2006a**] and computational methods [**Machuqueiro2007**, **Machuqueiro2010**].

We present, in the next section, a brief historical introduction to the molecular simulation theory and methods, with special emphasis on the methodology used in our work.

1.2. Molecular simulation: theory and methods

The most common approach to any given biological problem is usually an experimental one. Experimental methods are generally the fastest, cheapest and most straightforward option when studying biological systems. However, there are some exceptions to this: systems that do not benefit from such an approach, be it due to their inherent complexity, methodological limitations, time scales involved, or even cost of equipment. Molecular simulation methods are a useful and powerful tool to study molecular systems, particularly how they behave and interact with each other. The system is represented by a model, which is a simplified description through mathematical expressions, and this allows both measurable and unmeasurable properties to be computed, with the former being comparable to experimentally derived ones. This comparison either validates or invalidates the model used [**van Gunsteren1990**]. This model along with the force field used, specifically how accurate they are, govern how reliable a molecular simulation can be. Since this is an area in constant expansion, and with room for improvements, the predictions it is capable of certainly are not accurate enough to abandon the measurement of properties. Any measurement, unless very difficult, should always be favored instead of a prediction by simulation, however, when faced with quantities that are inaccessible by experimental measurements, molecular simulation is a viable alternative, and it should not be seen as a means of replacing experimental determinations, but rather as a way of complementing them.

What follows is an abbreviated historical recount of this scientific area, followed by a brief description of the theory involved as well as an introduction to the different methods we use in our studies.

1.2.1. Historical overview

Predating electronic computers by several years, one of the first models of liquids [**Morrel1936**], involved physical manipulation and analysis of the packing of a large number of gelatin balls, representing the molecules. Later applications of the technique have been described, including its application on the study of assemblies of metal ball bearings kept in motion by mechanical vibration [**Pierański1978**]. However, the use of large numbers of physical objects to represent molecules can be very time-consuming, and there are obvious limitations to the types of interactions between them, and the effects of gravity can never be eliminated [**Allen1987**]. This led to the replacement of the physical

model by a mathematical one, which was made more viable with the invention of the computer.

The first computer simulation of a liquid occurred in 1953 at the Los Alamos National Laboratories in the United States [**Metropolis1953**]. In 1957, Alder and Wainwright simulated a system of hard spheres [**Alder1957**, **Alder1959**], observing their dynamic properties. Later, in 1964, Rahman simulated liquid argon [**Rahman1964**, **van Gunsteren1990**], in the first successful attempt at solving the equations of motion for a set of Lennard-Jones particles [**Allen1987**], and in 1971, Rahman was one of the first to simulate liquid water [**Rahman1971**]. This marked the transition from atomic to molecular liquids and shortly after, the first simulations of rigid molecules [**Barojas1973**], flexible alkanes [**Ryckaert1975**], and a small protein (trypsin inhibitor) [**McCammon1977**] emerged.

As models were developed and as the computational power increased, larger and longer simulations were performed: from small proteins to large protein-DNA complexes in solution, and from picosecond to nano, micro and millisecond timescales. Molecular simulation is an area in constant expansion; over the last 60 years computational performance, measured by the number of FLOPS (Floating Point Operations per second), has been steadily growing [**van Gunsteren1990**] and is now approaching the teraflop barrier (10^{12} FLOPS) [**Giupponi2008**]. As the computational power of the machines increases, by contrast, their size and cost have been decreasing, which has allowed a broader application as more research groups have access to increasingly powerful calculation tools. This means that more complex molecular systems may be simulated over longer periods of time, or that it will be possible to handle more detailed molecular models in the future [**van Gunsteren1990**].

1.2.2. Statistical thermodynamics

When studying complex biomolecular systems, the experimental measurements are performed on macroscopic samples that contain extremely large numbers of atoms or molecules. These measurements depend on the positions and momenta of all particles in the system which, due to the interactions between them, will cause the instantaneous value of the property to fluctuate, resulting in a time average and a molecular average being measured. Computationally, it is also possible to calculate average values of properties by simulating the dynamic behavior of the system. The force acting on each atom can be calculated, and from that, the acceleration can be determined (via Newton's second law). The integration of the equations of motion yields a trajectory that describes how the positions, velocities and accelerations of the particles vary with time and from which the average values of

properties can be determined [**Leach2001**]. Ideally, we would like to simulate and study a solution, i.e. a large system with many interacting particles. This is, however, prohibitive to the calculation of a trajectory. This problem can be solved by replacing the molecular average by looking at a single molecule (or a small set of molecules) whose shorter trajectories when taken together (ensemble average) should be equivalent to the solution case, in accordance with the ergodic hypothesis. Macromolecular systems, represented in this way, i.e. with manageable numbers of atoms or molecules, and using both ensemble average and a time average, can be studied by computer simulation [**Leach2001**]. The simulation generates representative configurations of these small replications in such a way that accurate values of structural and thermodynamic properties can be obtained. Statistical thermodynamics is the mathematical tool that bridges the gap between the properties of individual molecules and the macroscopic thermodynamic properties of bulk matter [**Ben-Naim1992**]. This is accomplished by the application of probability rules to the systems, i.e. mathematical expressions are used to determine the probability of finding the system in a given state, which provides an interpretation of thermodynamic quantities such as temperature, pressure, energy or entropy at a molecular level.

1.2.3. Molecular mechanics

Molecular mechanics (MM) is the study of the motions of atoms as described by Newton's equations of motions, i.e. classical mechanics. In comparison to the more detailed and complex quantum mechanical methods, MM does not take into consideration the electronic motions and the system is described as a set of classical nuclei (Born-Oppenheimer approximation). Without this simplification it would be impossible to write the energy of the nuclear coordinates [**Leach2001**].

The description of complex chemical systems is made possible by the formulation of a realistic atomic model, which aims to understand and predict macroscopic properties based on detailed knowledge on an atomic scale [**van der Spoel2005**]. For different types of atomic model there are specific force fields. A force field is built up from two distinct components: (i) the set of equations (potential functions) used to generate the potential energies, and their derivatives, the forces, and (ii) the parameters used in this set of equations.

$$V(\boldsymbol{r}_1, \boldsymbol{r}_2, \ldots, \boldsymbol{r}_N) = \sum_{\text{bonds}} \frac{1}{2} K_b [b - b_0]^2 + \sum_{\text{angles}} \frac{1}{2} \theta_b [\theta - \theta_0]^2 + \sum_{\substack{\text{improper} \\ \text{dihedrals}}} \frac{1}{2} K_\xi [\xi - \xi_0]^2 + \sum_{\text{dihedrals}} K_\varphi [1 + \cos(n\varphi - \delta)] + \sum_{\text{pairs}} \left[\frac{C_{12}(i,j)}{r_{ij}^{12}} - \frac{C_6(i,j)}{r_{ij}^6} \right] + \frac{q_i q_j}{4\pi\varepsilon_0 \varepsilon_r r_{ij}} \quad (1)$$

A typical MM energy function $V(\boldsymbol{r}_1,\boldsymbol{r}_2,\ldots,\boldsymbol{r}_N)$ is given in Eq.1, which is a function of the positions $\boldsymbol{r}$ of N particles (usually atoms) has several contributions (depicted in Figure 1.4). The first term represents the covalent bond stretching interaction along bond b (Figure 1.4a). It is a harmonic potential, in which the minimum energy bond length b_0 and the force constant K_b vary with the particular type of bond. The second term describes the bond angle bending (three-body) interaction (Figure 1.4b) in a similar form. Two forms are used for the (four-body) dihedral angle interactions (Figure 1.4c): a harmonic term for dihedral angles ξ that are not allowed to make transitions (i.e. dihedral angles within aromatic rings are kept planar, and molecules that could flip over to their mirror images are prevented from doing so) and a sinusoidal term for the other dihedral angles φ, which may take 360 degree turns. The last term is a sum over all pairs of atoms and represents the effective non-bonded interaction, composed of the van der Waals (Figure1.4d) and the Coulomb (Figure1.4e) interaction between atom i and j with charges q_i and q_j at a distance r_{ij} [**van Gunsteren1990**].

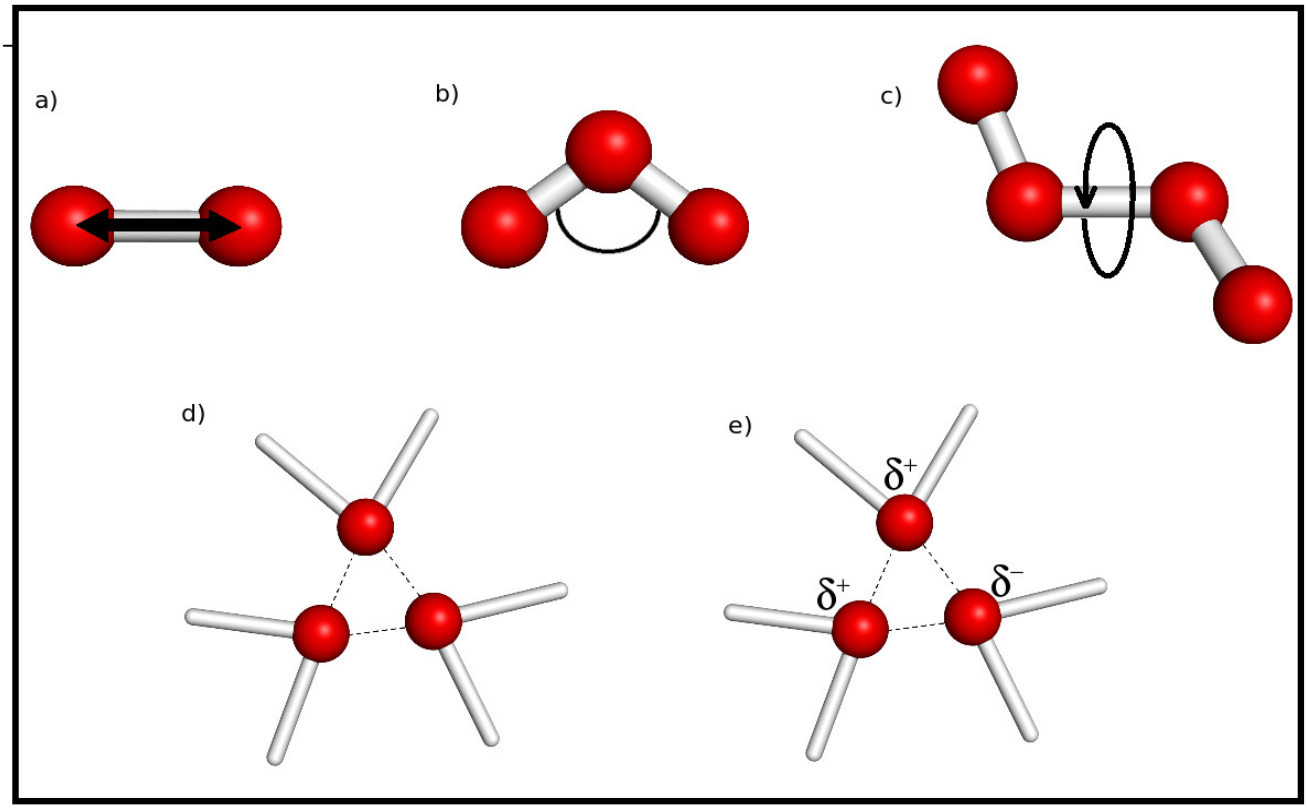

Figure 1.4 – Representation of the five key contributions to a molecular mechanics force field: (a) bond stretching, (b) angle bending, (c) dihedral angle interactions, (d) non-bonded van der Waals forces, and (e) non-bonded electrostatic interactions.

In addition to its functional form, a force field is only truly defined when the parameters, i.e. the various constants in (Eq.1) are specified. These parameters can be determined in one of two ways. The first is to fit them to results of *ab initio* quantum calculations on small molecular clusters. The alternative is to fit the force field parameters to experimental data (crystal structure, nuclear magnetic resonance, etc.) [**van Gunsteren1990**]. Often a combination of both approaches is used.

1.2.4. Molecular dynamics

Once the model and force field have been chosen, a method to search the configuration space has to be selected [**van Gunsteren1990**]. One such method is called molecular dynamics (MD). In MD simulations, Newton's laws of motion are applied to the atoms of a molecular system, generating a representative ensemble of configurations, while also resulting in a dynamic description of the system.

The use of Newton's equations implies that classical mechanics are used to describe the motion of atoms.

$$\boldsymbol{F}_i = m_i a_i \quad \leftrightarrow \quad \boldsymbol{F}_i = m_i \frac{d^2 \boldsymbol{r}_i(t)}{dt^2} \tag{2}$$

$\boldsymbol{F}_i$ denotes the force acting on atom i, and t denotes the time. MD simulation requires the calculation of the gradient of the potential energy $V(\boldsymbol{r})$ which therefore must be a differentiable function of the atomic coordinates $\boldsymbol{r}_i$ (Eq.3).

$$\boldsymbol{F}_i = -\frac{\partial V(r_i, \ldots, r_N)}{\partial r_i} \tag{3}$$

Equations 2 and 3 are numerically solved simultaneously in small time steps for a system of N interacting atoms, and the system is followed for some time, taking care that the temperature and pressure remain at the required values, and the coordinates are written to an output file at regular intervals. The coordinates as a function of time represent a trajectory of the system. After the initial changes, the system will usually reach an equilibrium sate. By averaging over an equilibrium trajectory many macroscopic properties can be extracted from the output file (such as the average potential energy of a system, the radial distribution function in a liquid or the binding constant of an inhibitor to an enzyme). The determination of non-equilibrium properties (e.g. the viscosity of a liquid, diffusion processes in membranes, the dynamics of phase changes, reaction kinetics, or the dynamics of defects in crystals) can also be done.

When a starting configuration is very far from equilibrium, the forces may be excessively large and the MD simulation may fail [**van der Spoel2005**]. A solution to this problem is to slightly change the system coordinates in order to bring it to an energy minimum. The potential energy function of a macromolecular system is a very complex landscape in a large number of dimensions [**Frauenfelder1998**]. It can have a deepest point, a global minimum, and a very large number of local minima, where all derivatives of V with respect to $\boldsymbol{r}$ are zero, and all second derivatives are non-negative. In between the local minima, there are saddle points, i.e. the mountain passes through which the system can migrate from one local minimum to another. Knowledge of all local minima, including the global one, and of all saddle points would enable a description of the relevant structures and

conformations and their free energies, as well as the dynamics of structural transitions. Unfortunately, the dimensionality of the configurational space and the number of local minima is so high that it is impossible to sample the space at a sufficient number of points to obtain a complete survey [**van der Spoel2005**]. Although there is no method that guarantees the determination of the global minimum in any practical amount of time, given a starting configuration it is possible to find the nearest local minimum, through the use of energy minimization methods, such as the steepest descent method, the low-memory Broyden Fletcher-Goldfarb-Shanno (l-BFGS) algorithm [**Press1992**] or the conjugate gradient method. While the first gets close to the local minimum very quickly, the latter gets closer to the minimum, but performs worse when starting in a position far it [**Leach2001**].

1.2.5. Continuum electrostatics

Charged and polar groups are found ubiquitously in biological macromolecules, and electrostatic interactions are of great relevance for many biochemical processes, such as protein structural stability, enzymatic catalysis and biomolecular recognition [**Fogolari2002**]. The study of several biomolecular systems depends on the understanding of these electrostatic forces and their effects. For this purpose, various electrostatics-oriented theoretical models have been developed [**Bashford1990**, **Feig2004**, **Warshel1979**, **Warshel1986**]. Through the use of reasonable simplifications, these models are able to give an "average" of the system description while still being able to capture electrostatic effects. Of widespread use are the continuum electrostatic (CE) models (Figure 1.5) which include the following simplifications:

1. Each time a group (e.g. part of a protein) loses/receives a charged ligand, the solvent and the rest of the system try to adapt to this new situation, in an attempt to compensate any unfavorable new interactions. The use of a dielectric constant ε attenuates the electrostatic interactions, mimicking the reorganization of dipoles. The more reorganizable a material is at the molecular level, the higher its ε value, i.e. a low value (usually between 2-4) is used for ε of the solute atoms, whereas a high value of 80 is normally used for the surrounding solvent [**Fogolari2002**]. Of course since proteins and other macromolecules are heterogeneous, the use of a single ε value is an approximation.

2. Another source of electrostatic reorganization is the redistribution of counterions (i.e. the ions in solution: Na^+, Cl^-, etc.). Instead of using the detailed positions of counterions, an overall measure of their concentrations can be used, i.e. the ionic strength *I*, as done in the Debye-Hückel theory of electrolytes [**Ben-Naim1992**].

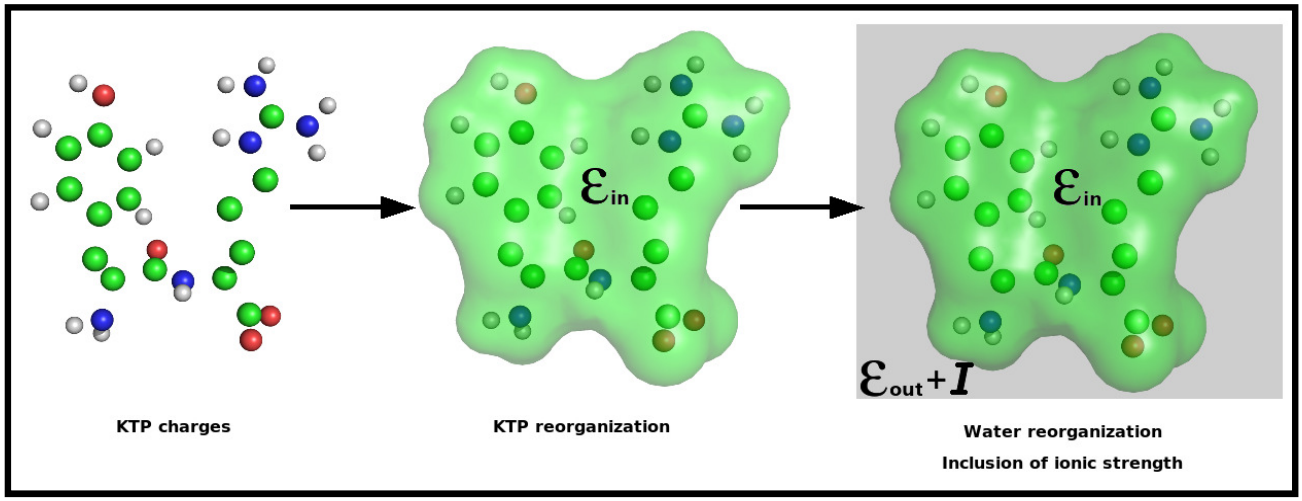

Figure 1.5 – Simplified representation of a CE model of KTP in water with dissolved ions. The model only considers KTP's charges (all bonds are neglected). KTP is mapped as a region with $\varepsilon=\varepsilon_{in}$, and the water reorganization and ionic strength effects are captured by $\varepsilon=\varepsilon_{out}$ and *I*, respectively. In addition, a frontier region surrounds the peptide wherein $\varepsilon=\varepsilon_{out}$ and $I=0$ are considered.

These models treat macromolecules like irregularly-shaped low dielectric objects containing embedded atomic charges, immersed in a high-dielectric (solvent) medium. The boundary between the interior and the exterior (low and high dielectric media, respectively) is a surface defined by the atomic coordinates and radii of the macromolecule [**Bashford2004**].

The method we use to calculate the electrostatic energies is based on the Poisson-Boltzmann (PB) approach. In this method, all solute atoms are explicitly considered as particles with a low dielectric constant and with point partial charges at atomic positions [**Fogolari2002**].

The Poisson equation (Eq.4) [**Ben-Naim1992**] describes a system with different dielectric regions:

$$-\nabla \cdot [\varepsilon(\boldsymbol{r})\nabla\phi(\boldsymbol{r})] = \rho_{\text{tot}}(\boldsymbol{r}) \quad (4)$$

where $\varepsilon(\boldsymbol{r})$ is the dielectric constant at point $\boldsymbol{r}$, $\phi(\boldsymbol{r})$ is the electrostatic potential at point $\boldsymbol{r}$, and $\rho_{\text{tot}}(\boldsymbol{r})$ is the total charge density at point $\boldsymbol{r}$. Its solution gives the electrostatic potential at all points $\boldsymbol{r}$, as a function of $\varepsilon(\boldsymbol{r})$ and $\rho_{\text{tot}}(\boldsymbol{r})$. When including the distribution of counterions in the Poisson equation, we get the Poisson-Boltzmann equation (Eq.5) [**Ben-Naim1992**]:

$$-\nabla \cdot [\varepsilon(\boldsymbol{r})\nabla\phi(\boldsymbol{r})] + \bar{\kappa}^2(\boldsymbol{r})\phi(\boldsymbol{r}) = \rho(\boldsymbol{r}) \quad (5)$$

where $\bar{\kappa}^2(r) = 8\pi e^2 I(r)/RT$, $I(\boldsymbol{r})$ is the ionic strength and $\rho(\boldsymbol{r})$ is the macromolecule charge density at $\boldsymbol{r}$.

The PB equation is usually solved numerically by mapping the system (ρ + ε + I) in a 3D grid, i.e. using a finite difference method, which we will explain in detail in the Methods section.

Once the electrostatic potential is obtained at all points, all electrostatic properties can be obtained from ϕ and ρ, including the electrostatic energy:

$$\mathcal{U} = \sum_i q_i \, \phi(\boldsymbol{r}_i) \quad (6)$$

The electrostatic energy $\mathcal{U}$ only captures the classical electrostatic part of the free energy change of a process. The other contributions, e.g. quantum effects, can be canceled using a thermodynamic cycle (Figure 1.6).

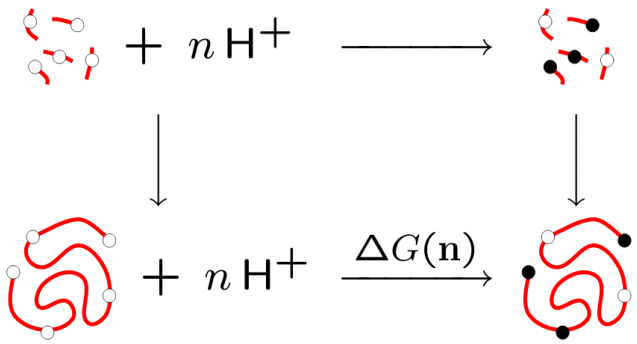

Figure 1.6 – Thermodynamic cycle of the protonation free energies taken from CE calculations.

From the thermodynamic cycle we can write the free energy of protonation as a sum of individual and pairwise terms (Eq.7).

$$\Delta G(\mathbf{n}) = -2.3RT \sum_i n_i \, \mathrm{p}K_i^{\mathrm{int}} + \sum_i \sum_{j<i} (n_i n_j + n_i z_j^o + n_j z_j^o) W_{ij} \qquad (7)$$

where $\mathrm{p}K_i^{\mathrm{int}}$ is the intrinsic $\mathrm{p}K_a$ of site *i* reflecting burying, W_{ij} is the interaction between sites *i* and *j*, n_i is the occupancy of site in the protonation state ***n***, and z_i^o is the charge of site *i* when deprotonated. Both $\mathrm{p}K_i^{\mathrm{int}}$ and W_{ij} can be computed using CE calculations, referring to the cycle in Figure 1.6.

The protonation free energies obtained are then used for sampling protonation states, which allow the calculation of $\mathrm{p}K_a$ values. The "protonation state" is a specification of which sites are protonated and which are deprotonated: if a macromolecule has *N* sites, each with two possible protonation states (empty/occupied), there are 2^N possible protonation states [**Bashford2004**]. In these cases where a system contains many titrating sites, a straightforward scan of the complete configuration space is impossible, in which case a collection of protonation states can be generated by random sampling [**van Gunsteren1990**]. One of the ways to accomplish this is by using Monte Carlo (MC) methods, [**Antosiewicz1989**, **Baptista1999**, **Beroza1991**], so-called because of the role that random numbers play in the method [**Allen1987**]. The MC simulation produces a representative ensemble of protonation states and consists of the following: (i) an initial protonation state is selected and then sites are flipped at random between protonated and deprotonated states; (ii) the flips are then accepted if $\Delta G(\mathbf{n}) \leq 0$ or with a probability when $e^{-\Delta G(\mathbf{n})/RT}$ when $\Delta G(\mathbf{n}) > 0$; (iii) if the criteria are not met, the new configuration is rejected and the previous one is counted again and used as a starting point for the next random displacement [**van Gunsteren1990**].

1.2.6. Constant-pH molecular dynamics

The computational treatment of pH and how it affects biomolecular structure and function has always been particularly challenging. Until recently, computational studies dealt with pH in two simplified

approaches: (i) performing MM/MD simulations of the solute with the protonatable groups set to the states they would normally have in solution at the pH of interest, or (ii) applying simplified electrostatics-oriented theoretical models, such as PB, to a rigid structure and then sampling the protonation states, using MC methods. Both approaches present some problems: the first one fails to consider desolvation and site-site interaction effects caused by the environment, does not allow a change in the protonation state during the MD simulation (also, it is not always clear which state a protonatable group would have at a given pH). Although special MM/MD techniques (free energy methods) exist to address ionization processes, they are computationally very demanding and prohibitive to study several ionizing groups. The second approach does not take into account the structural reorganization and protonation-conformation coupling events. However, and despite their flaws, both approaches share a certain complementarity regarding the treatment of protonation/ deprotonation events: MM/MD methods can properly treat structural dynamics but require the use of a fixed protonation state, while PB/MC-based approaches can efficiently treat multiple protonation equilibrium but require the use of a rigid protein structure.

Recently, an alternative methodology was developed in an attempt to solve this problem. The so-called constant-pH MD methods allow the treatment of pH as an external parameter, like the temperature or pressure, as done in a typical experimental setup. This was first attempted in 1994 by Mertz and Pettitt [**Mertz1994**] when studying the reaction of proton exchange between acetic acid and water. Later, in 1997, a different approach, exploring the complementarity between MM/MD and simplified models was followed in the implicit titration method for constant-pH MD [**Baptista1997**]. This approach used fractional protonation states periodically updated from PB/MC calculations performed along the MD simulation, and efficiently combined both MM/MD and PB/MC and allowed them to do what they do best, namely, conformational sampling at a fixed protonation and protonation sampling at a fixed conformation, respectively. Following a similar rationale of complementarity, in 2002, the stochastic titration method for constant-pH MD [**Baptista2002b**] was proposed. In this method, discrete (non-fractional) protonation states are similarly obtained from PB/MC calculations. The coupling between the MM/MD and PB/MC algorithms ensures a proper Markov sampling from the correct thermodynamic ensemble. Similar stochastic constant-pH MD methods were proposed by Antosiewicz and co-workers [**Walczak2002**] [**Dlugosz2004a**, **Dlugosz2004b**], and by Mongan et al. [**Mongan2004**], with differences lying in the overall use of solvation models (uniform-dielectric Langevin dynamics [**Walczak2002**], analytical continuum solvent potential [**Dlugosz2004a**, **Dlugosz2004b**], and Generalized Born (GB) [**Mongan2004**]) in the MD simulations and the

replacement of the PB model by GB in one of the models [**Mongan2004**]. A different route, not based on simplified models, was followed by Bürgi et al. [**Bürgi2002**], who proposed a method where short segments of MM/MD thermodynamic integration are used to compute the free energies of trial Monte Carlo changes of the protonation states. This constant-pH MD method has a high computational cost, and the sampling efficiency of the short free energy calculations is questionable. The acidostat method proposed by Börjesson and Hünenberger [**Börjesson2001**, **Börjesson2004**] is also entirely based on MM/MD. The method uses fractional protonation states weakly coupled to a proton bath, but its theoretical basis is problematic [**Baptista2002a**]. Another constant-pH MD method using continuous titration coordinates was recently proposed by Brooks and co-workers [**Lee2004**, **Khandogin2005**]. This method uses a GB implicit solvent model and is based on heuristic analogies with lambda-dynamics [**Kong1996**] and linear response approximation (LRA) methods.

In 2006, a new implementation of the stochastic titration method for constant-pH MD was presented [**Machuqueiro2006**]. This version allowed the inclusion of ionic strength effects in the simulations and featured a new reduced titration step, wherein an exclusion list of the sites that are titrating too far away from the pH of interest is created, thus reducing the overall computational cost. In 2008, the same method featured the inclusion of proton tautomerism in the study of a protein [**Machuqueiro2008**].

Of all the different constant-pH MD methods, currently only three are in active use and development: (i) the stochastic titration method of the Molecular Simulation group at ITQB-UNL [**Baptista2002b**], (ii) the method of Charles Brooks' group at the University of Michigan [**Lee2004**], and lastly (iii) the method of Andrew McCammon and David Case's groups at the University of California [**Mongan2004**]. Of the three, only the stochastic titration method allows the explicit inclusion of the solvent, which makes it the perfect candidate for the explicit modeling of lipid membranes.

2. Methods

2.1. Molecular mechanics/molecular dynamics

2.1.1. Molecular model and force field

All simulations were performed with the GROMOS 53A6 force field [**Oostenbrink2004**, **van Gunsteren1990**] for the GROMACS 3.2.1 distribution [**Berendsen1995**, **Lindahl2001**]. The general MD algorithm consists of the following (Figure 2.1):

1. In the first step, the initial conditions of the system are inputed, namely, the potential interaction function V as a function of atom positions, the positions $\boldsymbol{r}$ of all atoms in the system, and the velocities $\boldsymbol{v}$ of all atoms in the system.

2. Next, the forces are computed. The force on any atom is computed by calculating the forces due to bonded interactions (bond stretching, angle bending and proper and improper dihedral interaction terms) plus the forces between non-bonded atom pairs (van der Waals forces and electrostatic interactions).

3. The movement of the atoms is simulated by numerically solving Newton's equations of motion, and thus the configuration is updated.

4. Steps 2 and 3 are repeated until an output step is required. When this occurs, the positions, velocities, energies, temperature, pressure, etc. are written.

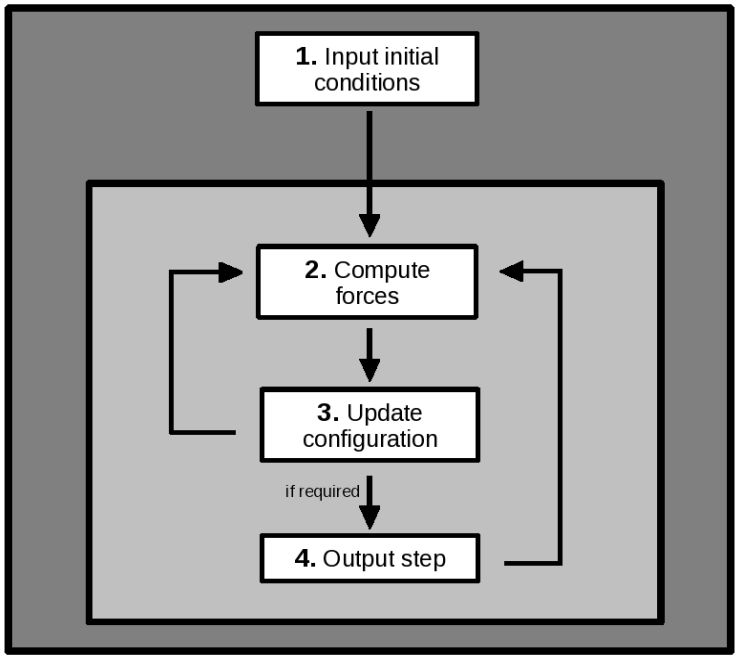

Figure 2.1 – Schematic representation of the MD algorithm. Steps 2, 3 and 4 are repeated for the required number of simulation steps. Adapted from [**van der Spoel2005**].

2.1.2. Boundary conditions

Periodic boundary conditions (PBC) enable a simulation to be performed using a relatively small number of particles, in such a way that the particles experience forces as if they were in bulk fluid [**Leach2001**]. The treatment of boundaries is then necessary in order to minimize edge effects, i.e. in PBC the small molecular system is surrounded by copies of itself without boundaries, to mimic a dilute solution (Fig 2.2). The atoms of the system that is to be simulated are put into a periodic space-filling shaped box, which is treated as if surrounded by 26 ($=3^3-1^3$) identical translated images of itself. The next layer of neighbor images of the central computational box contains $5^3-3^3-1^3=98$ boxes and so on. When a lattice sum of the atomic interactions is to be performed, only the nearest neighbors are taken into account. The filled atom in the central computational box will only interact with atoms or images of atoms that lie within the dashed line. The anisotropy of the interaction due to the cubic shape of the nearest image box can be avoided by the application of a spherical cut-off, like in the twin range method of the previous section. The periodic boundary conditions affect not only the computation of

the forces, but also the positions of the atoms. It is common practice to keep the atoms together, i.e. in the central computational box: when one atom leaves the central box on one side, it enters it with identical velocity at the opposite side at the translated image position [**van Gunsteren1990**].

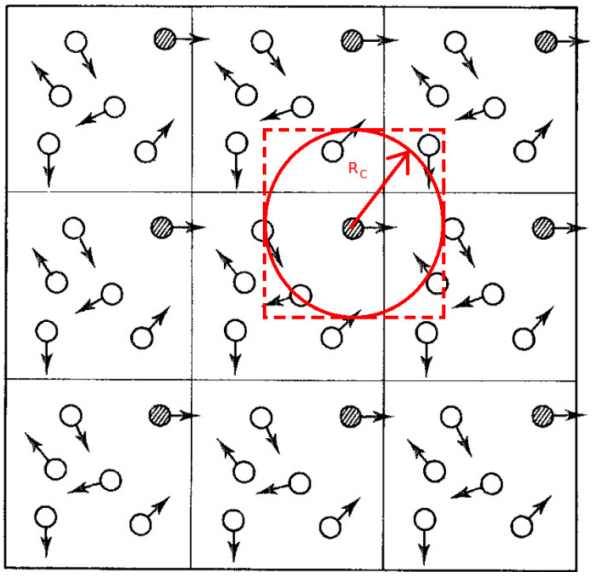

Figure 2.2 – Periodic boundary conditions in two dimensions. The central box is surrounded by eight identical images of itself. Also represented is the cut-off radius RC. The colored atom does not interact with other atoms outside the dashed line. Adapted from [**Leach2001**].

Application of periodic boundary conditions means that in fact a crystal is simulated. For a liquid or solution the periodicity is an artifact of the computation, so the effects should be minimized. An atom should not simultaneously interact with another atom and a periodic image of that atom. Consequently, the cut-off radius R_C must not exceed half of the shortest box vector. However, when simulating a macromolecule this condition alone does not suffice, because a single solvent molecule should not be able to "see" both sides of the macromolecule, which is solved by the spherical cut-offs imposed in the twin range method. Possible distorting effects of the periodic boundary conditions may be traced by simulation of a system in differently shaped boxes (Figure 2.3). The use of truncated octahedral or rhombic dodecahedral boxes instead of cubic or rectangular ones has the advantage of reducing the number of solvent molecules needed to fill the remaining (after insertion of the solute) empty space in the box [**van Gunsteren1990**].

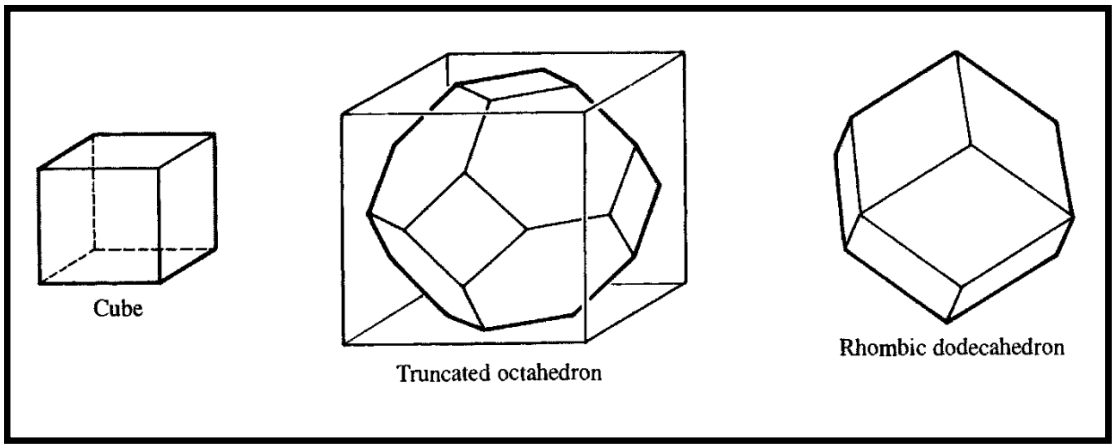

Figure 2.3 – Different periodic cells used in computer simulations. Adapted from [**Leach2001**].

We used a rhombic dodecahedral box for the aqueous system and a rectangular box for the membrane systems. While in the aqueous system, a dodecahedral box was more appropriate to minimize the number of waters present, in the membrane system, a rectangular box was used to the structural constraints imposed by the membrane geometry.

2.1.3. Treatment of long range non-bonded interactions

The summation of the last term in the potential energy function (Eq.1) that covers both types of non-bonded interactions, runs over all atom pairs in the molecular system. Its computation time is proportional to N^2, the square of the number of atoms in the system. Since the computation time of the other parts of the calculation is proportional to N, computational efficiency can be much improved by a reduction of this summation. The simplest procedure is to apply a cut-off criterion for the non-bonded interactions and use a list of neighbor atoms lying within the cut-off, which is only updated every so many simulation steps [**van Gunsteren1990**]. There are several methods that can be used to accomplish this, however we will only present the ones we use in our studies.

To treat the long range van der Waals interactions we used the twin range method [**Berendsen1985**] (Figure 2.4). This method uses two cut-off radii, R_C^1 and R_C^2, where R_C^2 corresponds to the R_C of Figure 2.2. The atoms lying within a distance R_C^1 from atom i are stored in a neighbor list

of atom *i*. The interaction of the atoms *j* for which $R_C^1 < r_{ij} < R_C^2$ with atom *i* are stored in the form of a so-called long range force on atom *i*. At each MD step the non-bonded interaction consists of two contributions: (i) the short range part which is calculated from the neighbor list using the actual atom positions, and (ii) the long range part which is kept fixed during N_C^1 time steps. Neighbor lists and the long range force are then simultaneously updated every N_C^1 time steps. The short range cut-off (or R_C^1) was of 8 Å and the long range cut-off (or R_C^2) was of 14 Å. The neighbor lists were updated every 5 MD steps.

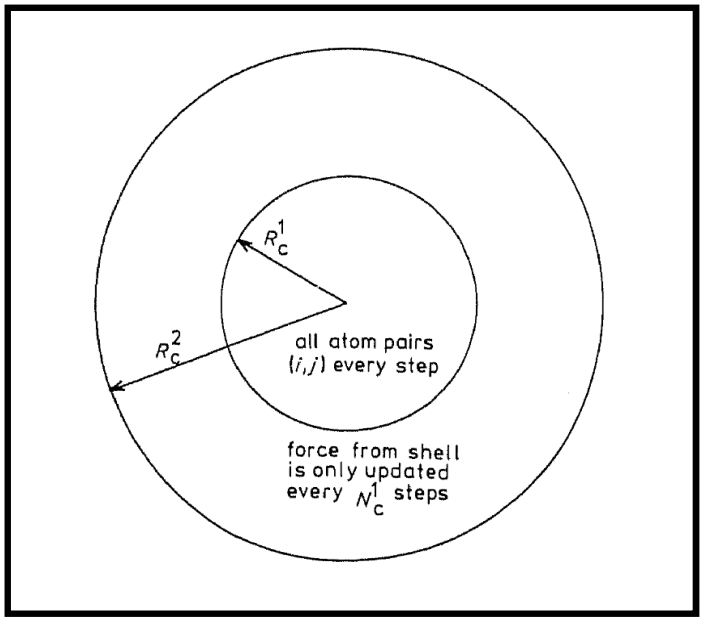

Figure 2.4 – Twin range method. All atom pairs in the inner circle with R_C^1 are calculated every step, while on the outer circle (R_C^2) the force is only updated ever N_C^1 steps. Adapted from [**van Gunsteren1990**].

The long range electrostatic interactions were treated using the Generalized Reaction Field method [**Tironi1995**]. When parts of the system are homogeneous, like the bulk solvent surrounding the solute, the number of atoms or degrees of freedom can be reduced considerably by modeling the homogeneous part as a continuous medium, e.g. a continuous dielectric [**van Gunsteren1990**]. In this type of approximation the system is divided into two parts: (i) an inner region where the atomic charges q_i are explicitly treated (dielectric constant ε_{in}), and (ii) an outer region, beyond the distance R_C^2 (Figure 2.4) which is treated like a continuous dielectric medium with dielectric constant ε_{out} and ionic strength *I*. The values for the dielectric constants used were of 2 and ~80 for the solute and solvent, respectively. The ionic strength was set to 0.1 M.

2.1.4. Temperature and pressure control

When simulating biological systems it is also necessary to control external parameters like the temperature and the pressure. There are several methods for the treatment of such parameters. For the temperature we used the Berendsen temperature coupling [**Berendsen1984**] method, wherein the system is coupled to an external heat bath at the desired temperature. The bath acts as a source of thermal energy, supplying or removing heat from the system as appropriate. The velocities are scaled at each step, such that the rate of change of the instantaneous temperature is proportional to the difference in temperature between the bath and the system:

$$\frac{dT(t)}{dt} = \frac{1}{\tau}(T_{\text{bath}} - T(t)) \tag{8}$$

τ is a coupling parameter whose magnitude determines how tightly the bath and the system are coupled together. This method gives an exponential decay of the system towards the desired temperature. The change in temperature between successive time steps is:

$$\Delta T = \frac{\Delta t}{\tau}(T_{\text{bath}} - T(t)) \tag{9}$$

and the scaling factor for the velocities is thus:

$$\lambda^2 = 1 + \frac{\Delta t}{\tau}\left(\frac{T_{bath}}{T(t)} - 1\right) \tag{10}$$

τ controls the degree of coupling: when large, the coupling is weak, and when τ is small, the coupling is strong. In our simulations, the temperature of KTP, lipids and water molecules was separately coupled to a temperature bath at 310 K, and a relaxation time (τ) of 0.1 ps was used.

The methods used for pressure control are somewhat analogous to the ones used for the temperature. Thus, the system can be coupled to a "pressure bath", much like previously explained for the temperature. The Berendsen pressure coupling method was used, with a rate of change of pressure given by:

$$\frac{dP(t)}{dt} = \frac{1}{\tau_P}(P_{bath} - P(t)) \quad (11)$$

Again, τ_P is the coupling constant and P_{bath} is the pressure of the 'bath'. $P(t)$ is the actual pressure at time t. The volume of the simulation box is scaled by a factor λ, which is equivalent to scaling the atomic coordinates by a factor $\lambda^{1/3}$. Thus:

$$\lambda = 1 - \kappa\frac{\Delta t}{\tau}(P - P_{bath}) \quad (12)$$

where κ is a constant. The new positions are given by:

$$\boldsymbol{r}_i' = \lambda^{\frac{1}{3}}\boldsymbol{r}_i \quad (13)$$

This expression can be applied isotropically (i.e. the scaling factor is equal for all three directions), anisotropically (with the scaling factor calculated independently for the three axes), semi-isotropically (with a scaling factor applied to the *xy* axes and a different one applied to the *z* axis). We used a particular case of the semi-isotropic pressure coupling that applies a surface tension, wherein a normal pressure coupling is used for the *z* axis and the surface tension is directly coupled to the *xy* axes of the box. The *z* pressure component was set to 1 bar, and a surface tension of 25 dynes/cm. A τ_P of 5 ps and a compressibility of 4.5×10^{-5} bar^{-1} were used. The choice of the "correct" surface tension for a given membrane system is not trivial. Experimental values of 23-26 dynes/cm have been estimated by different groups [**Macdonald1987**, **Portlock1992**, **Seelig1987**] and van Gunsteren and co-workers have shown that 25 dynes/cm are required to obtain the correct value for the area per lipid using the GROMOS 53A6 force field [**Chandrasekhar2005**]. The value of 25 dynes/cm was also used in a recent computational study using a membrane model similar to ours [**Machuqueiro2010**].

2.1.5. Constraints and restraints

Constraints are requirements that the system is forced to satisfy, i.e. bonds and angles are forced to adopt specific values throughout a simulation. By constraining individual internal coordinates or combinations of specified coordinates, computing time in MD simulations can be greatly lowered. Restraints, on the other hand, are additional terms included in the force field that 'encourage' bonds and angles to adopt a certain value, by imposing penalties for deviations from that reference value [**Leach2001**].

We use the LINCS algorithm to constrain the bond lengths [**Hess1997**]. This method reset bonds to their correct lengths after an unconstrained update, in a two-step non-iterative process (Figure 2.5). In the first step, the projections of the new bonds on the old bonds are set to zero, and in the second step, a correction is applied for the lengthening of the bonds due to rotation.

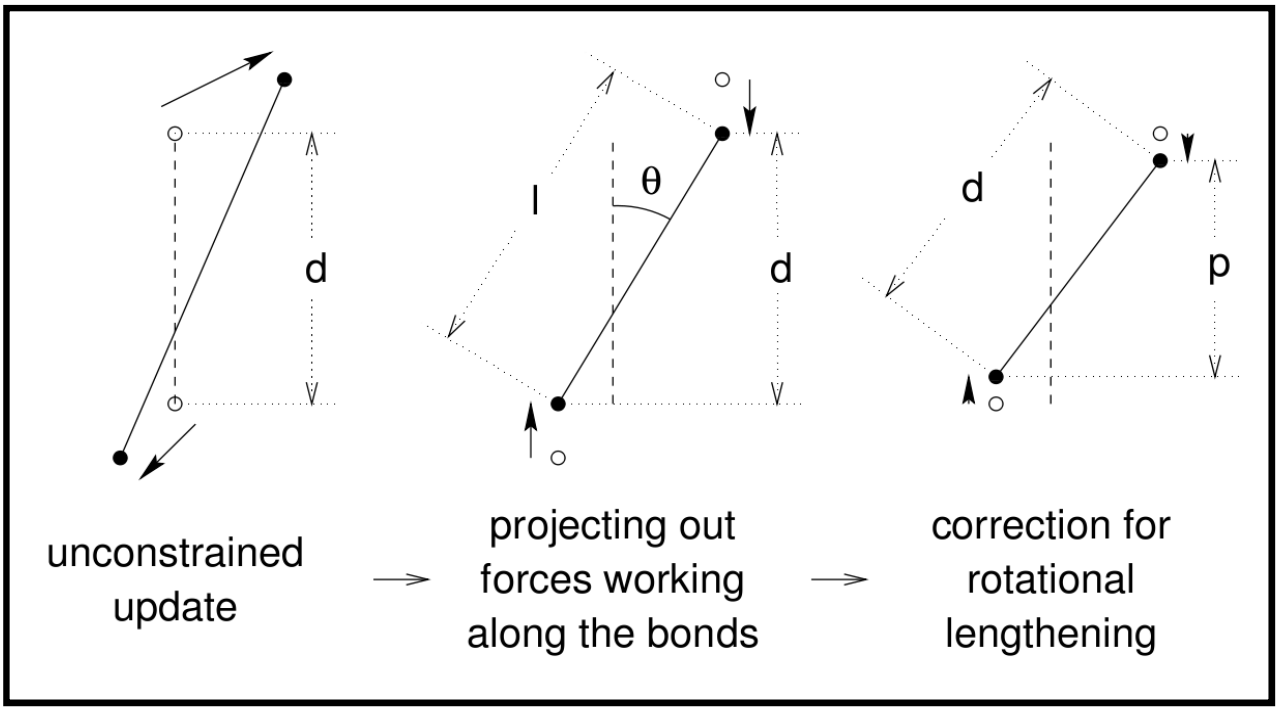

Figure 2.5 – The three position updates needed for one time step. The dashed line is the old bond of length d, and the solid lines are the new bonds ($l = d\cos(\theta)$ and $p = (2d^2\text{-}l^2)^{1/2}$). Taken from [**van der Spoel2005**].

2.2. Poisson-Boltzmann/Monte Carlo

All electrostatic calculations were done using version 2.2.0 of the MEAD (which stands for **M**acroscopic **E**lectrostatics with **A**tomic **D**etail) program [**Bashford1992**].

The Poisson equation (Eq.4) and the Poisson-Boltzmann equation in its linear form are examples of elliptic partial differential equations with complex numerical solutions [**Bashford2004**] whose discussion is beyond the scope of this work. We shall instead focus on one particular type of method to solve the PB equation, the finite-difference method. It is a relatively simple method, of which we shall give a basic description. The rationale of this method is to approximate derivates as differences between function values sampled at points a finite difference apart. In order to accomplish this, the macromolecule is mapped on a three-dimensional grid, allowing the atomic charges to be distributed over grid points. The functions defined on a three-dimensional space are replaced by values defined on a three-dimensional cubic lattice.

We used a particular technique to do this, called focusing [**Gilson1987**]. Since we are only interested in a small region within our peptide (i.e. the titratable site), we map the peptide with a large and coarse grid, and then apply smaller and finer grids, centered on the site of interest, to do the calculations. We used a three-step focusing procedure, first with a 61 x 61 x 61 grid spaced 2 Å then with a 81 x 81 x 81 grid spaced 1 Å (Figure 2.6a) and finally a 65 x 65 x 65 grid spaced 0.25 Å (Figure 2.6b).

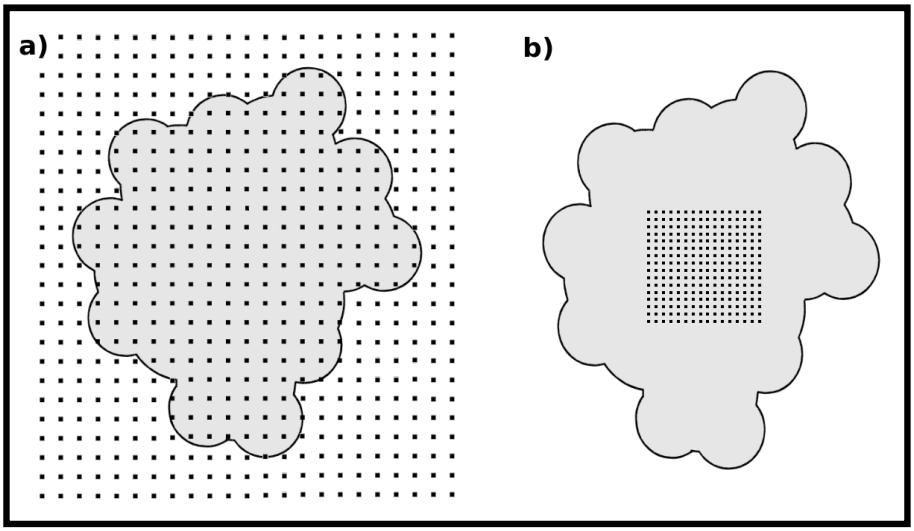

Figure 2.6 – Two step focusing procedure. A larger, coarser grid is first applied, spanning the whole protein/macromolecule (a), followed by a smaller, finer grid on a specific location of interest (b).

The calculations of the protonation states and pK_a values in macromolecules are done using an approach based on model compounds of known pK_a value (Figure 2.7) and electrostatic models to calculate shifts from the model compound values [**Bashford2004**]. The model compound pK_a values we used were adjusted to experimental data available for alanine pentapeptides [**Grimsley2009**, **Thurlkill2006**].

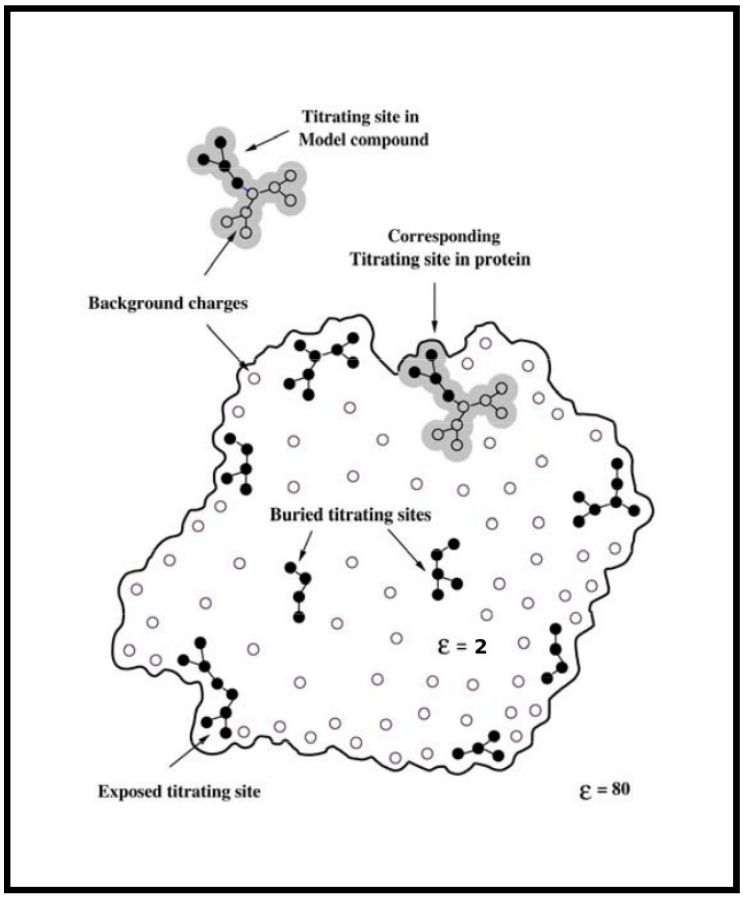

Figure 2.7 – MEAD model for pK_a calculations of proteins. The charges of the titrating site in protein or model compound (black circles in grey region) are changed from protonated to deprotonated values. Also represented are the non-titrating polar groups in both the protein and model compound (open circles) and other ionizable groups in the protein (black circles in white areas). Taken from [**Bashford2004**].

When there is more than one titratable site, as in our case, we must consider the relative energetics of the different protonation states and then the statistics of ensembles over all possible states. This leads to

the predictions of the degree of protonation of each site as a function of pH [**Bashford2004**]. In this type of multi-site titration problem, the calculation of protonation states is usually done by MC methods. The MC runs were performed using 10^5 MC cycles, one cycle consisting of sequential state changes over all individual sites and also all pairs of sites with at least one interaction term above 2 pK_a units [**Baptista1999**]. For this, we used the version 1.5 of the program Petit (**P**roton and **E**lectron **TIT**ration) [**Baptista2001**], developed by Dr. António Baptista.

2.3. Constant-pH molecular dynamics

The constant-pH MD method implemented is the stochastic titration method [**Baptista2002b**]. The method includes the reduced titration step [**Machuqueiro2006**] and the inclusion of tautomers [**Machuqueiro2008**] of recent implementations, as well as further modifications introduced in the course of this work, which we will explain in detail in the results section. The algorithm used, is depicted in Figure 2.8, and consists of different sequential blocks: (a) the first block is a PB/MC calculation, with the protonation states resulting from the last MC step being assigned to the protein; (b) the next block is the solvent-relaxation dynamics, wherein a short MM/MD simulation of the system with a frozen protein allows the solvent to adapt to the new protonation states (the duration of this block is determined by τ_{rlx}); (c) the last block is a full MM/MD simulation of the unconstrained system (the duration of this block is determined by τ_{prt}). Steps (d) and (e) represent the reduced titration step: every n^{th} cycle (in this case, 10^{th}) a full PB/MC calculation of the system is run to assign a fixed state to all the sites whose mean occupancies fall outside a predefined threshold.

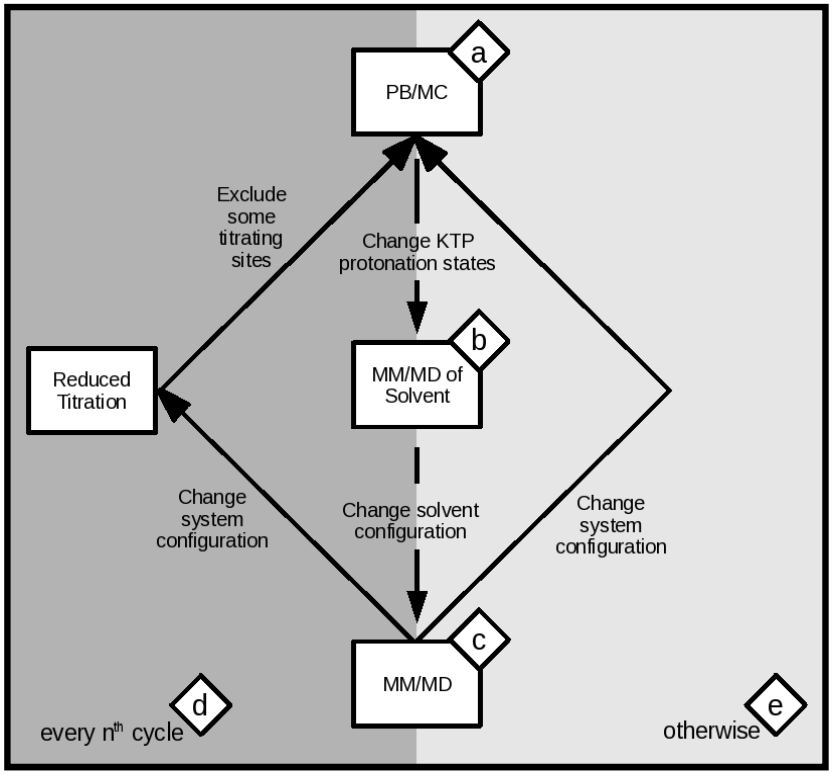

Figure 2.8 – Schematic representation of the constant-pH MD algorithm used.

Each MM/MD block consisted of 5000 effective simulation steps, each with a 2 fs time step, i.e. a total of 10 ps were simulated after each protonation state change. The solvent-relaxation step (Figure 2.8b) was done for 0 ps, i.e. it was effectively bypassed for reasons we will explain in detail in the results section.

2.4. Setup and equilibration

2.4.1. Aqueous system

KTP was built in an extended conformation and placed in the center of a rhombic dodecahedric box. The box was filled with 840 water molecules. The system was first minimized with ~40 steps of steepest descent, followed by ~10^4 steps using the l-BFGS algorithm.

The initiation was done in a 50 ps MD simulation by harmonically restraining all the atoms, followed by another 50 ps simulation with only the Cα atoms restrained. This was done for two different startup systems, one with the N-terminus protonated (charged) and one with the N-terminus deprotonated (neutral). In both systems, the C-terminus was deprotonated and both side-chains were protonated.

Eleven runs of 100 ns were performed (one for each pH value in the 2-12 interval), six starting with the charged system (protonated N-terminus), to simulate the pH range 2-7, and five with the neutral system (unprotonated N-terminus) to simulate the pH range 8-12. The first 5 ns of all simulations considered to correspond to the equilibration of the system, thus being discarded in the subsequent analyses.

2.4.2. Membrane system

The system we used was previously equilibrated for the work of Machuqueiro *et al.* [**Machuqueiro2010**]. The zwitterionic lipid 1,2-dimyristoyl-*sn*-glycero-3-phosphatidylcholine (DMPC) was chosen to build the lipid bilayer (Figure 2.9a). The membrane was composed of two layers of DMPC, with 64 molecules each, according to the scheme in (Figure 2.9), placed in a rectangular box and fully hydrated with 4214 water molecules. The system was minimized first with ~10^3 steps of steepest descent, followed by 10^4 steps using the l-BFGS algorithm. The initiation was achieved by restraining all lipid atoms in a 50 ps MD simulation, followed by another 50 ps simulation with all lipid atoms restrained with a weaker force constant, and finally with 100 ps of unrestrained system. The membrane was simulated for 100 ns and the first 10 ns were discarded. The system we used was already pre-equilibrated for the work of Machuqueiro et. al [**Machuqueiro2010**].

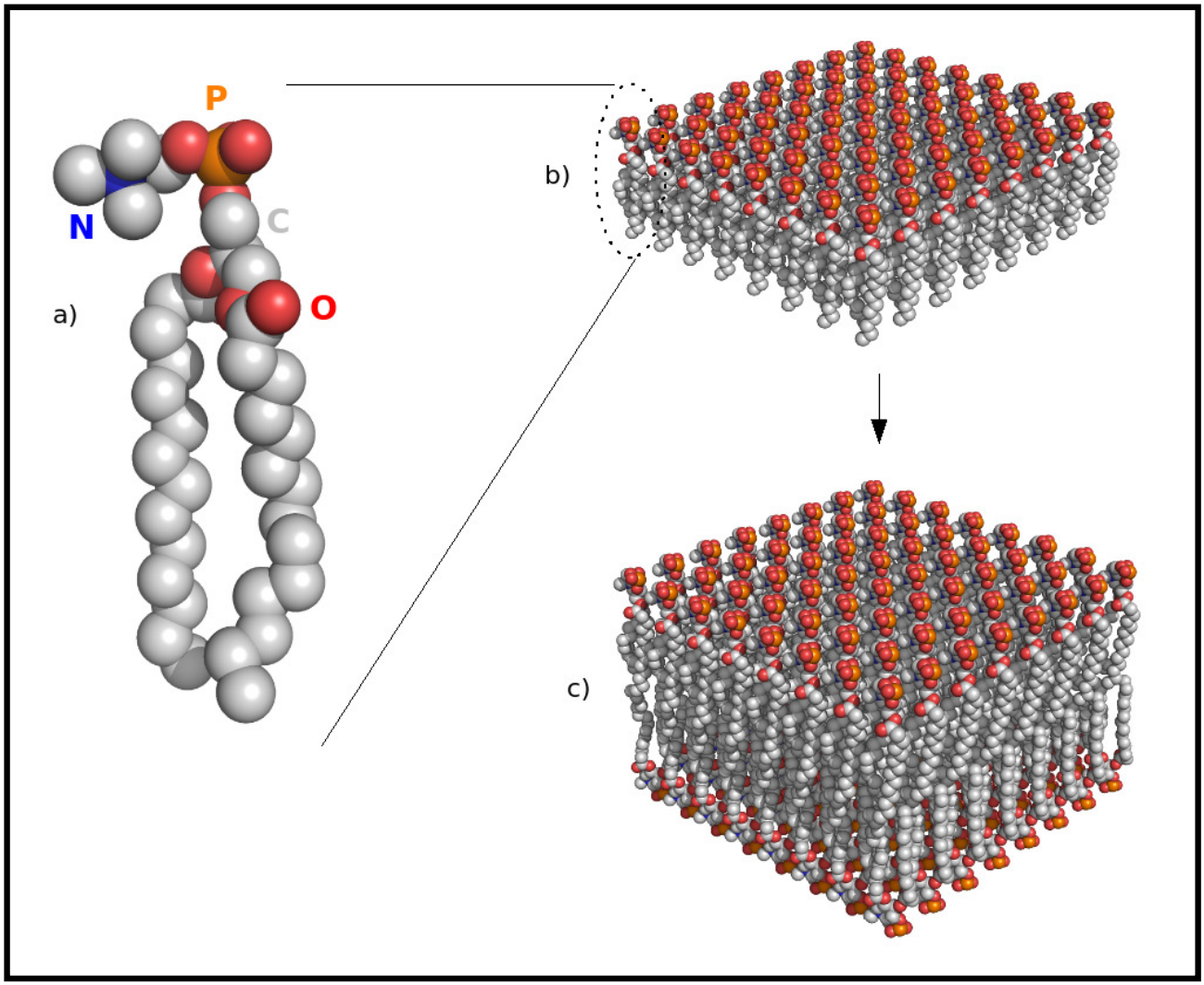

Figure 2.9 – Schematic representation of the membrane building process. Pictured: (a) DMPC lipid molecule, (b) 64 DMPC lipid monolayer, and (c) 128 DMPC lipid bilayer.

2.4.3. KTP-bilayer model system

The dipeptide was built in an extended conformation and placed inside a rectangular box away from the pre-equilibrated membrane mentioned in the previous section. The box was filled with 5640 water molecules. The system was first minimized with ~2500 steps of steepest descent [**Press1992**] followed by ~850 steps using the l-BFGS algorithm [**Press1992**]. This was done for both the charged and neutral

systems.

The initiation was achieved by harmonically restraining all peptide and lipid atoms in a 50 ps MD simulation, followed by another 50 ps with only the lipid atoms restrained, and finally with 50 ps of unrestrained system. This was done for three different startup systems. We performed 33 simulations: three replicates of each pH value in the 2-12 interval. Each simulation was 100 ns long, and the first 10 ns in all cases were discarded in order to get the system equilibrated.

In the membrane calculations, the bilayer was centered in the *z* coordinate of the simulation box while the peptide was centered in the *xy* plane in order to keep all membrane-peptide interactions as homogeneous and realistic as possible.

2.5. General analysis

The dihedral, distance, area and volume analyses were performed using the GROMACS package [**Oostenbrink2004**, **Scott1999**] and in-house tools. pK_a calculations were done using **`interpoly`**, an in-house tool that makes polynomial interpolations. Structural representations were done using PyMOL [**DeLano2002**], and graphics using Gnuplot [**Williams2004**]. Errors were computed using standard methods [**Allen1987**].

The calculations of the non-mass weighed root mean square deviation (RMSD) was done using a modified GROMACS tool (g_rms). The RMSD is the most common measure of the fit between two structures, and is the root mean square distance between pairs of atoms:

$$\mathrm{RMSD} = \sqrt{\frac{\sum_{i=1}^{N_{\mathrm{atoms}}} d_i^2}{N_{\mathrm{atoms}}}} \qquad (14)$$

where N_{atoms} is the number of atoms over which the RMSD is measured and d_i is the distance between the coordinates of atom i in the two structures when they are overlaid [**Leach2001**].

3. Results and discussion

3.1. Membrane equilibration

To ensure that our DMPC bilayer behaved in accordance with a regular fluid phase membrane, we monitored its equilibration by following area and volume per lipid over the length of the simulation.

3.1.1. Area per lipid

The area per lipid (A_L) is evaluated from the *xy* dimensions of the lipid membrane. Figure 3.1 shows that the mean A_L (red line) of 64.2 ± 0.2 Å² is within the experimental range for a DMPC membrane (59.7 – 65.2 Å²) [**Nagle2000**], and slightly higher than the value obtained for the membrane without the peptide present (green line) [**Machuqueiro2010**], as expected, due to the presence of KTP. This difference between the mean A_L value we obtained and that of **Machuqueiro2010** is due to the area that the peptide occupies; since it is known that KTP favors the lipid phase [**Lopes2006a**], it spends most of the simulation time in contact with the membrane, in different conformations, orientations and in different degrees of insertion. Due to this added complexity we could not calculate the mean area of KTP and subtract this value from the total A_L measured.

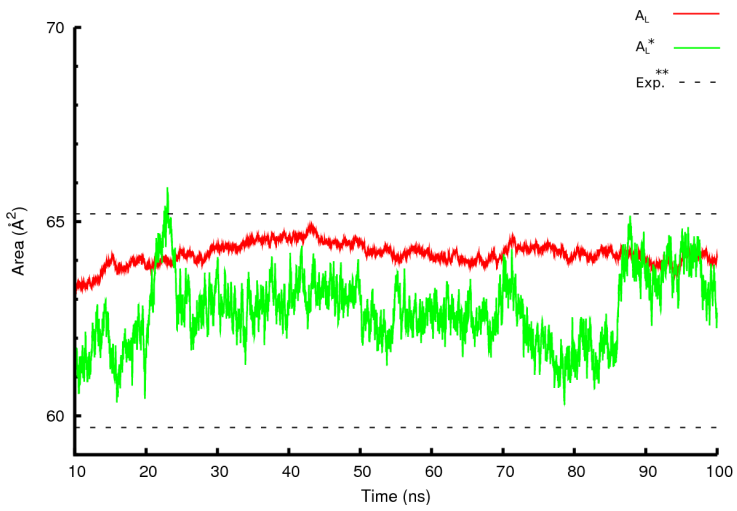

Figure 3.1 – Mean area per lipid of all pH values and all replications. A_L* value is from [**Machuqueiro2010**] and Exp.** value from [**Nagle2000**].

Of all the constant-pH MD settings, the ones that most influenced the A_L of our membrane system were the τ_{rlx} and the τ_{prt}. The values previously used with this methodology [**Baptista2002b**, **Machuqueiro2006**, **Machuqueiro2007**] proved not to work in our system. This was most likely due to a perturbation of the membrane induced by the relaxation step, wherein the water molecules would penetrate the bilayer when the solute was 'frozen' and lead to a collapse, followed by an abrupt increase in the A_L value. We experimented with different values for these two parameters and found that, of all the different scenarios we put forward, the system with $\tau_{rlx} = 0$ ps and $\tau_{prt} = 10$ ps was the one that best allowed an A_L value within the experimental range, i.e. the one that best assured the stability of our membrane. When $\tau_{rlx} = 0$ ps, the solvent-relaxation step is effectively bypassed, which raises some problems in the theoretical framework of our methodology, since the changes in the protonation states resulting from the PB/MC calculations require a 'relaxed solvent' prior to the full system dynamics. While the removal of the solvent-relaxation step might not be theoretically sound, the fact is that for our system this step was a destabilizing factor, and some problems with its use have been previously noted [**Baptista2002b**], since in practical terms, it can be skipped, and when included, it can cause some issues, such as in our case.

An increase in τ_{prt} from 2 (value normally used [**Machuqueiro2006**, **Machuqueiro2007**]) to 10 ps was also necessary to stabilize the membrane, since one of the hypotheses we tried (with $\tau_{rlx} = 0$ ps and $\tau_{prt} = 2$) still led to abnormally high values of A_L. We hypothesized that an increased full system dynamics block after each protonation step was required to allow an adaptation of the membrane to the change in the environment in the absence of a relaxation step. Additionally, this increase in τ_{prt} represented a drop in the protonation sampling, which was somewhat compensated by the number of replicates and the length of the simulation.

3.1.2. Volume per lipid

The volume per lipid (V_L) is evaluated from the *xyz* dimensions of our box, and we can see from Figure 3.2 that it is slightly over the experimental range (1094 – 1101 Å³) [**Nagle2000**]. The V_L of the DMPC membrane alone [**Machuqueiro2010**], however, is well within the experimental range. The mean value (Figure 3.2, red line) was corrected by removing the average volume of KTP which, unlike the area, is much easier to calculate. This average volume of the peptide was calculated using a GROMACS tool

(g_sas) which calculates the total solvent accessible surface area. The difference in V_L between the membrane system with and without (Figure 3.2 blue line) KTP is probably due to a perturbation of the membrane by the peptide. However, even though this perturbation occurs, its effect is very small, since the measured value for the V_L is 1104.14 ± 0.12 Å³, only slightly over the experimental range.

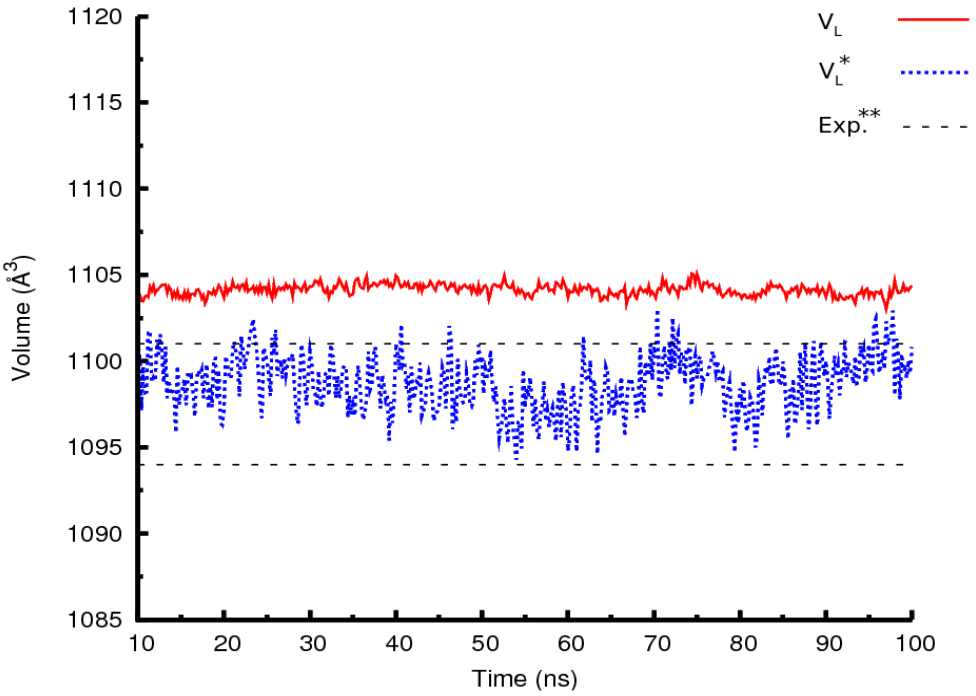

Figure 3.2 – Mean volume per lipid of all pH values and all replications. V_L (red line) also excludes the average volume of KTP. Also depicted, the volume per lipid over time of a DMPC bilayer without KTP (V_L*) (from [**Machuqueiro2010**]). Exp.** value is from [**Nagle2000**].

Together with the area reported in the previous section, this makes us confident that our membrane system is properly equilibrated and fit to be used in further studies.

3.2. Titration curves and pK_a determination

The titration profiles of KTP for both water and membrane systems are depicted in Figure 3.3. KTP has a +2 charge in low pH values, with all sites protonated: N-terminus (+), Tyrosine-OH (0), Arginine-NE(+) and C-terminus (0). When the C-terminus fully deprotonates, around pH ~6, it becomes negative, bringing the total charge of KTP to +1. At pH ~9, the net global charge of KTP becomes zero, which corresponds to the deprotonation of the N-terminus, and is also the isoelectric point (pI) of KTP. Around pH 11, KTP becomes negative as the tyrosine residue completely loses its proton. The titration of arginine is not very relevant in the pH interval we examined.

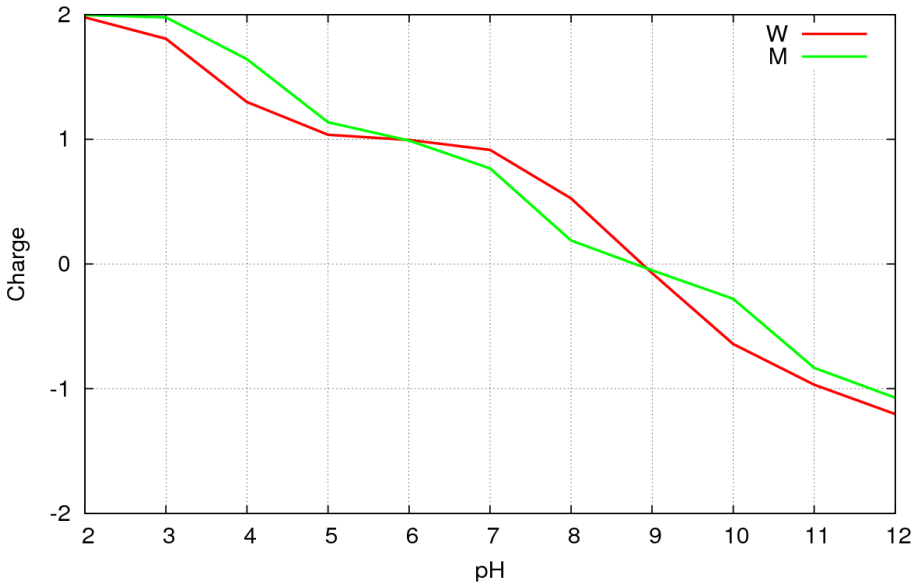

Figure 3.3 – Titration curves of KTP for the water (W) and membrane (M) systems.

The titration curve for the water system closely resembles the one obtained in **Machuqueiro2007**, with a few differences however, namely regarding the titration of the N-terminus, which is slightly shifted to more alkaline pH values in our system. This is most likely due to the force field and readjusted pK_a values for the model compounds [**Grimsley2009**, **Thurlkill2006**] being different in both studies: in **Machuqueiro2007** the molecular regions corresponding to the model compounds were directly assigned the "representative" pK_a values estimated by Nozaki and Tanford [**Nozaki1967**] from a set of analog organic molecules, whereas in this study the model compounds were redefined in terms of their

constituting atoms and assigned pK_a values that are able to reproduce experimental values obtained from potentiometric titrations of single-site alanine pentapeptides [**Grimsley2009**, **Thurlkill2006**], which means that each model compound thus defined reflects directly a single accurate experimental measurement and does not require any approximate reasoning based on chemical similarity.

Next, we looked at the titration curves for the individual titratable sites (Figure 3.4). These were obtained by using a third-order polynomial interpolation through the use of **`interpoly`**. When observing these titration curves (Figure 3.4), the differences between the water and membrane systems are more apparent. Since the N-terminus site is the main responsible for KTP's charge at physiological pH [**Lopes2006a**, **Machuqueiro2007**] its titration behavior will be analyzed first and in greater detail. From Figure 3.4, we can see that the N-terminus in the membrane system is titrating at a lower pH value (7.38) than in the water system (7.96) (Table 3.1). The fact that the pK_a value we determined for the N-terminus in the membrane system is lower than the one for the water system has to be attributed to a stabilization of its neutral (unprotonated) form probably by hydrogen bonding to the phospholipid head groups, or to a destabilization of its positively charged (protonated) form by the choline groups. We then compared our values with those of **Machuqueiro2010**, also determined for the N-terminus group in both water and membrane systems: 7.88 and 8.25, respectively. While these values are relatively higher than in our case, and with the membrane possessing a higher pK_a value than the water system, it is worth mentioning that these were obtained with a different method, which also combines MM/MD and PB/MC but in a different way. The LRA method adopted in that case is a simpler and more approximate method (the MD calculations are executed with fixed pure protonation states instead of simulating pH) than constant-pH MD, which is probably why the values obtained from it are somewhat different. Next we compared our values in water with previously obtained computational and experimental values: 7.14 [**Machuqueiro2007**] and 6.2 [**Lopes2006a**], respectively. The difference between our values and those of **Machuqueiro2007** can be explained by the use of a different force field (GROMOS 53A6 in our case and GROMOS 43A1 [**Scott1999**] in that work) and by the use of more accurate and recent values for the pK_a of the model compounds [**Grimsley2009**, **Thurlkill2006**]. Regarding the difference between the values we obtained for the pK_a of the N-terminus and the one measured in [**Lopes2006a**], it is our understanding that 6.2 is a rather low value for the pK_a of such an exposed (solvated) primary amine group, which can most likely be attributed to the fact that it was obtained by an indirect measurement of a positively charged KTP from an anionic resin.

Regarding the pK_a values obtained for the other sites, we determined values of 9.82 and 10.34 for the tyrosine group and of 3.70 and 4.04 (Table 3.1) for the C-terminus group in water and

membrane systems, respectively. In both cases, the membrane value seems a bit higher, which is probably explained by a higher tendency of the groups to remain neutral in the less polar membrane environment, which increases the pK_a values measured. When comparing these values to those obtained by **Machuqueiro2007** (9.94 and 4,04, for the tyrosine and C-terminus groups, respectively), the differences are very small and most likely due to the different force field and readjusted pK_a values for the model compounds. The difference is higher in the case of the C-terminus because the difference between the new pK_a values [**Grimsley2009**, **Thurlkill2006**] and the old ones [**Nozaki1967**] is also higher.

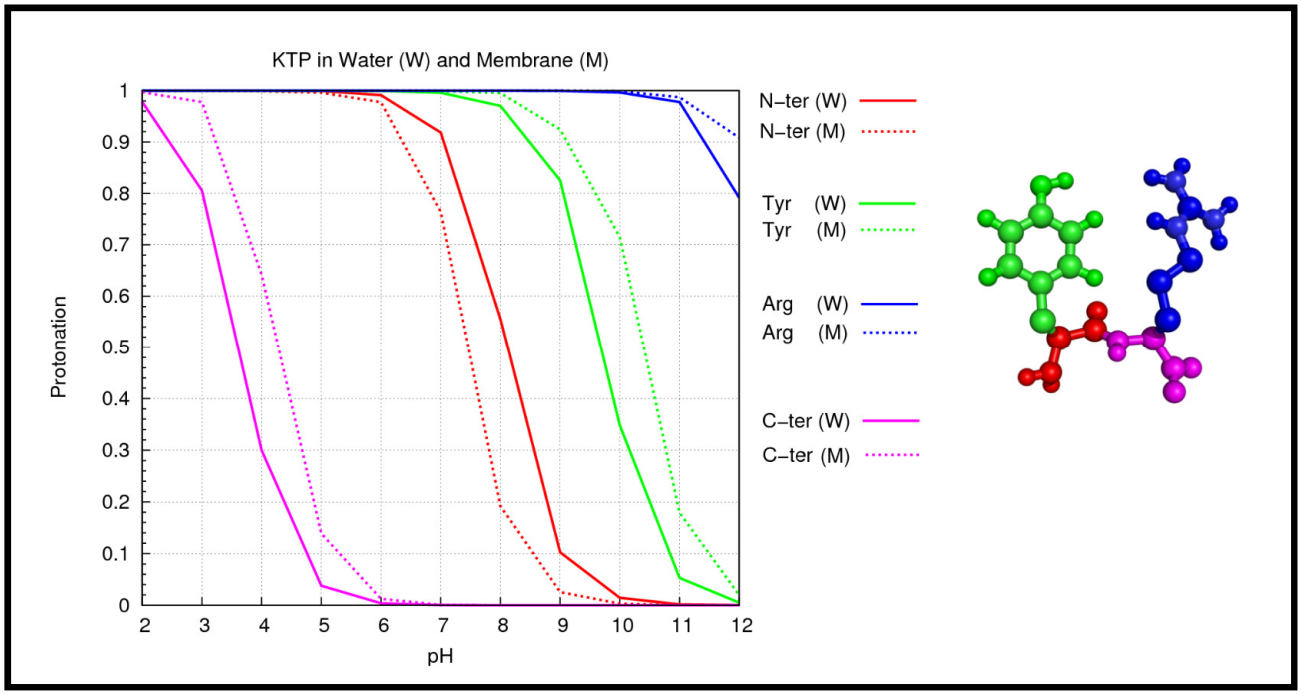

Figure 3.4 – Titration curves of KTP in water (W) and in the presence of a lipid membrane (M). The four titratable sites are distinguished by different colors in the graphic and in the KTP molecule to the right.

Table 3.1 – pK_a values of the four titratable sites for the water and membrane systems.

	Water		**Membrane**	
	pK_a	**Error**	**pK_a**	**Error**
N-ter	7.96	± 0.18	7.38	± 0.39
Tyr	9.82	± 0.24	10.34	± 0.33
Arg	>12	–	>12	–
C-ter	3.70	± 0.25	4.07	± 0.27

3.3. Conformational analysis

We know that KTP has a strong conformational dependence on pH in an aqueous media [**Machuqueiro2007**]. When looking at the $C\beta_{Tyr}$-$C\alpha_{Tyr}$-$C\alpha_{Arg}$-$C\beta_{Arg}$ dihedral angle of KTP in both water (Figure 3.5a) and membrane systems (Figure 3.5b), we can see that they are very similar, and that they exhibit the same trend witnessed in **Machuqueiro2007**, namely, a predominance of the trans conformer at lower pH values ($C\beta_{Tyr}$-$C\alpha_{Tyr}$-$C\alpha_{Arg}$-$C\beta_{Arg}$ dihedral angle < -70º and >100º) and a mixture of *trans* and *cis* conformers as the pH increases, leading to a clear majority of *cis*.

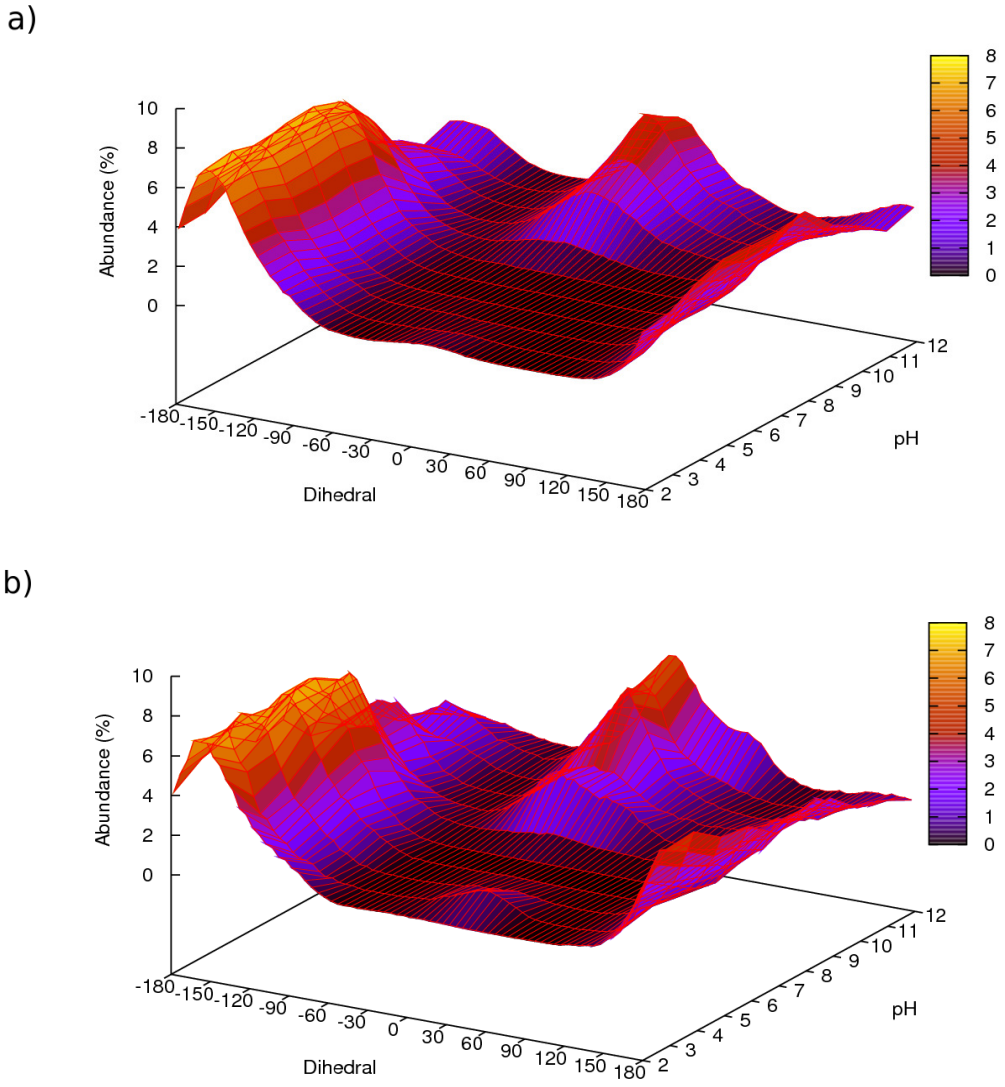

Figure 3.5 – Abundance of $C\beta_{Tyr}$-$C\alpha_{Tyr}$-$C\alpha_{Arg}$-$C\beta_{Arg}$ dihedral of KTP vs. pH in (a) water and (b) membrane systems.

It is worth mentioning that the 3D plot of the water system (Figure 3.5a) appears smoother than its membrane equivalent because of the larger sampling involved (for each pH in the water system, there are 95000 points whereas for the membrane system, there are only about 27000, which means there is a 3:1 difference). The plot for the membrane system has an unusual amount of *cis* conformers at pH = 3, which is attributed to one of the replicates where KTP transitioned from an expected *trans* conformation to a *cis* conformation where it would remain for the rest of the simulation (over 30 ns). Since this phenomenon only occurred in one replicate and only at one pH value, and since we determined it was not caused by a change in protonation (data not shown), we think when the change from *trans* to *cis* occurred (also witnessed at low pH values, but only transiently) its insertion in the membrane caused the system to become 'trapped' in an unusual conformation. This is consistent with the idea that the conformational transitions would become more difficult in the stiffer membrane environment. More sampling is required to determine if this occurrence is a rare event, or if it is a consequence of our system setup. We would also like to observe this particular simulation for a longer period of time, something that was impossible in our time-frame, since the constant-pH MD simulations of systems as large as ours are extremely time consuming. In any case, we will return to this particular case in the next section.

Next we looked at the distance between the side chains of both amino acid residues of KTP, in order to determine if an interaction was occurring. For this we measured the distance between the OH group from the tyrosine phenol ring and the closest NH group from the arginine side chain. We did this for all pH values in our range (2-12) but focused on two of them in particular: pH 6 and 9. These values were chosen because they allowed us to compare our results to those of **Machuqueiro2010**. In **Machuqueiro2010** they used a different method methodology to study the protonation of KTP, namely, the linear response approximation (LRA) [**Eberini2004**], which also combines MM/MD and PB/MC to allow the estimation of pK_a values, but does so in a different way. This means that LRA is a method that makes further approximations and as such does not yield as accurate pK_a values as the constant-pH MD approach. Since LRA does not allow the specification of pH, the study was done with two pure protonation states. These do not exist in a realistic scenario, since at physiological pH values what exists is a mixture of protonation states, and not a fixed representative state. The states used in **Machuqueiro2010** were named KTP^+ and KTP^0, and the only difference between them was the protonation state of the N-terminus group, which was protonated in the first, and deprotonated in the latter. The other titratable groups of KTP were defined as: neutral (protonated) tyrosine group,

positively charged (protonated) arginine group and negatively charged (deprotonated) C-terminus group. According to the titration curves we obtained, it is our understanding that pH 6 and 9 are the ones that best represent these two pure states.

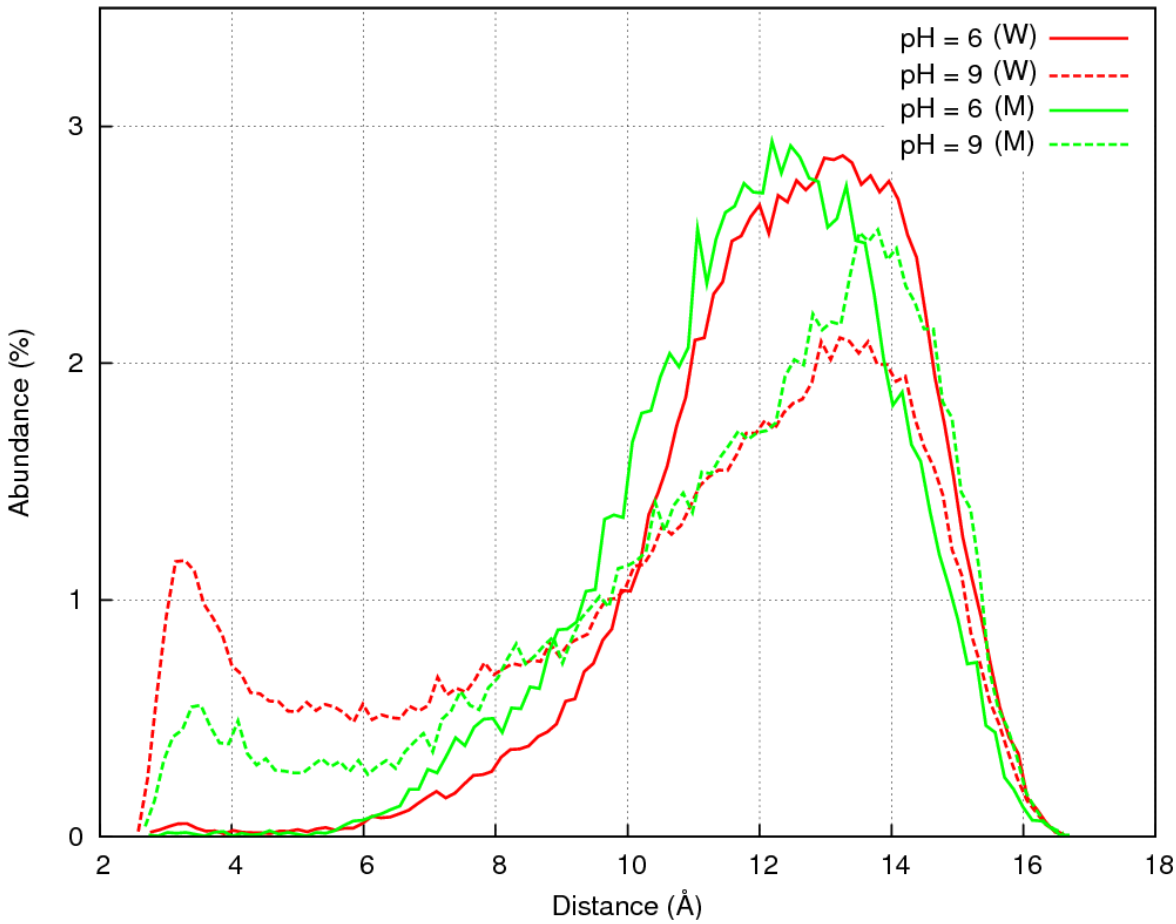

Figure 3.6 – Distance between the side chains of KTP for pH 6 and 9. Depicted, the distribution of distances in water (W) and membrane (M) system.

From Figure 3.6 we can see that at pH = 6, the side chains do not show a relevant interaction in both water and membrane systems, whereas at pH = 9, a small population of interacting side chains begins to appear. When looking at this plot, we decided on a cut-off of 6 Å to define if an interaction was occurring. Bearing in mind that this is only an indirect comparison to **Machuqueiro2010** since at pH 6

and 9 we still have a mixture of protonation states, we decided to take the comparison further and decided to isolate from all of our simulations only the ones that matched the pure protonation states used in **Machuqueiro2010** and obtained Figure 3.7.

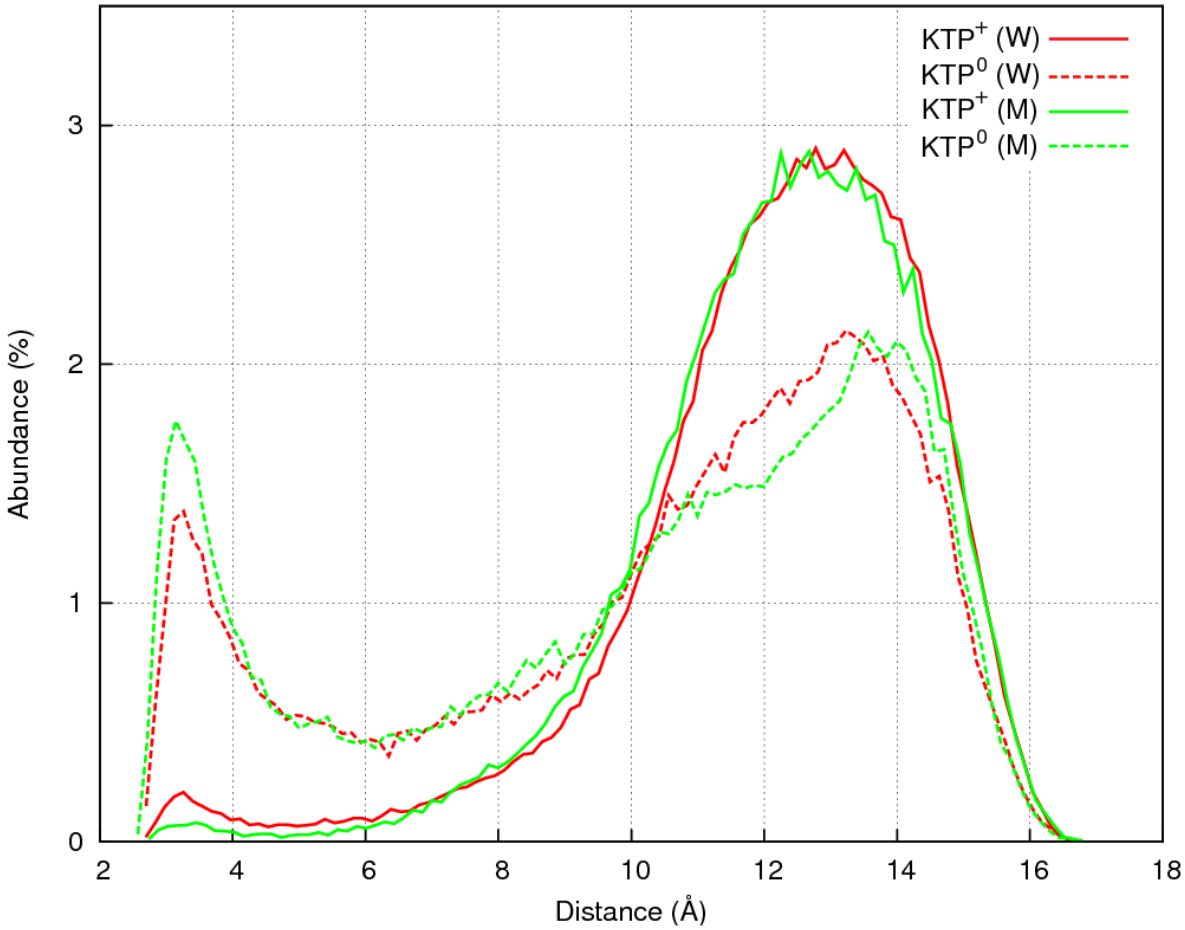

Figure 3.7 – Distance between the side chains of KTP for the KTP^+ and KTP^0 states. Depicted, the distribution of distances in water (W) and membrane (M) system.

In Figure 3.7 we can see the same trend as in the previous plot, but now the difference between the water and membrane systems seems lessened. With further assurance that 6 Å for our cut-off, we decided to use it in the next step. We plotted the amount of *cis* conformers as a function of the pH value, for both water and membrane systems and using the 6 Å cut-off to determine the threshold value for an interaction between both side chains, we determined the amount of *cis* conformers with a packed

structure (Figure 3.8).

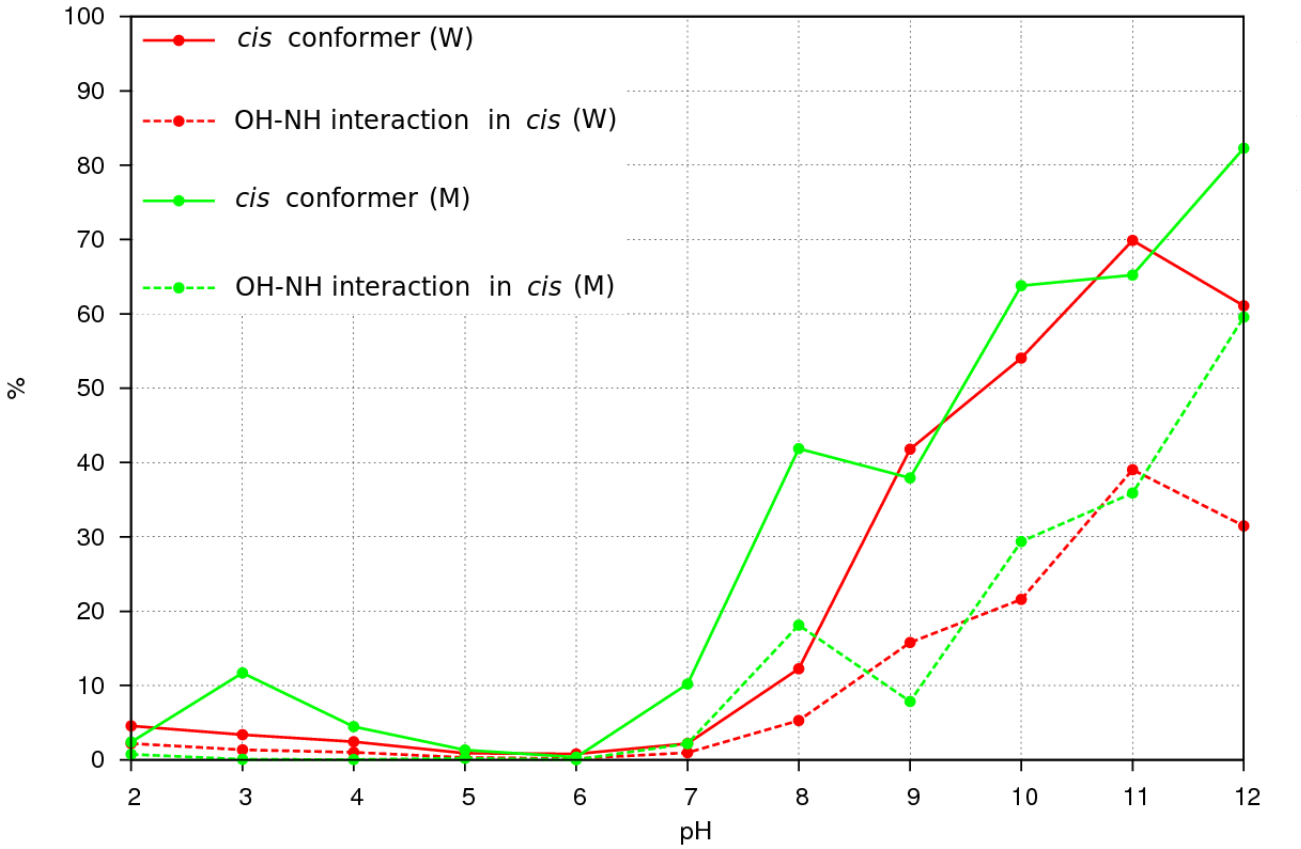

Figure 3.8 – Percentage of packed and extended *cis* conformers vs pH. Depicted are the percentage of *cis* conformers in water (W) and membrane (M) (full lines) and the percentage of compact *cis* conformers (dotted lines).

We can see from Figure 3.8 that there is a clear increase in the *cis* population as the pH becomes more alkaline (~60% in the water system and ~80% in the membrane system). The percentage of *cis* actually decreases from pH 11 to 12 in the water system, something that was not observed in the membrane system or in **Machuqueiro2007**. This can also be observed in Figure 3.5a, where a drop in the *cis* population coincides with an increase of the *trans* conformer in the transition from pH 11 to 12. Also, in Figure 3.8 (dotted lines) the amount of compact *cis* conformers follows a similar trend, and more than half of the *cis* population, in both systems, is in a compact conformation, with both side chains interacting, which means that a considerable amount of *cis* conformers are in an extended

conformation, as observed in **Machuqueiro2007**. We can also see from Figure 3.8 that the unusual amount of *cis* conformations at pH = 3 is mostly in an extended form.

We then looked at the χ_1 torsion angle of tyrosine (composed by the N-C_α-C_β-C_γ atoms). As we can see from Figure 3.9, there are three main minima at -60º, 60º and 180º. In **Machuqueiro2010** a disappearance of the population at 60º was observed for the membrane system for the KTP^+ state. The authors noted that because morphine has an equivalent of the angle of ~-90º, the reduction of the population at 60º will favor the potential to bind an opioid-like receptor. In order to compare our results with these, we again decided to look at the pH values 6 and 9.

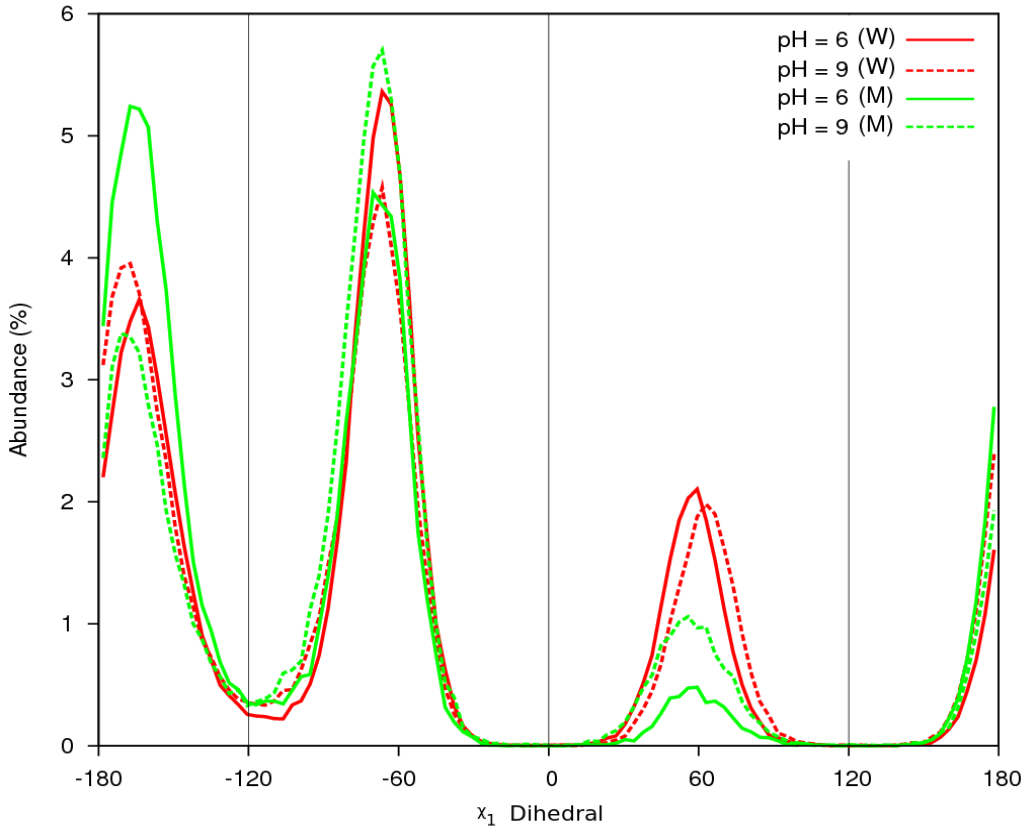

Figure 3.9 – χ_1 dihedral of KTP for water (W) and membrane (M) systems at pH 6 and 9.

Again, in Figure 3.9 we see a good agreement to the results obtained by **Machuqueiro2010**. The χ_1 dihedral population at 60º in the membrane system is lower than in the water system, and this difference is clearly larger at the lower pH value. In order to better interpret the differences between the

populations, we built Table 3.2 (which also exhibits the relative populations of packed and extended *cis* and *trans* conformers).

Table 3.2 – Population percentages of the different KTP conformation groups. Depicted are the values for the water and membrane systems at pH = 6 and pH = 9.

		Water			Membrane		
		cis	***trans***	**Total**	***cis***	***trans***	**Total**
pH=6	χ_1=-60º	.33	45.84	46.18	.24	41.32	41.57
	χ_1=180º	.40	36.26	36.66	.07	54.11	54.19
	χ_1=60º	.08	17.06	17.15	.04	4.19	4.23
	Total	.82	99.16	**100.00**	.35	99.63	**100.00**
	packed	.18	.52	.70	.02	.41	.43
	extended	.64	98.64	99.29	.33	99.22	99.56
	Total	.82	99.16	**100.00**	.35	99.63	**100.00**
pH=9	χ_1=-60º	11.34	30.72	42.07	17.68	36.45	54.14
	χ_1=180º	20.39	20.73	41.13	17.45	17.74	35.19
	χ_1=60º	10.02	6.76	16.79	2.78	7.87	10.65
	Total	41.76	58.22	**100.00**	7.92	2.06	**100.00**
	packed	15.95	.44	16.39	7.85	.32	8.17
	extended	25.81	57.78	83.60	30.07	61.74	91.82
	Total	41.76	58.22	**100.00**	37.92	62.06	**100.00**

From Table 3.2, we can easily observe the difference in the 60º population between the water and membrane systems at both pH values, and indeed confirm that the difference at pH = 6 is much more significant. However, the difference observed by **Machuqueiro2010** was even more significant (from 20.8% to 2.8%). As we have mentioned in the case of the distance between the side chains, this is a

rough comparison between both studies, which is why we need to isolate the proper pure protonation states (Figure 3.10).

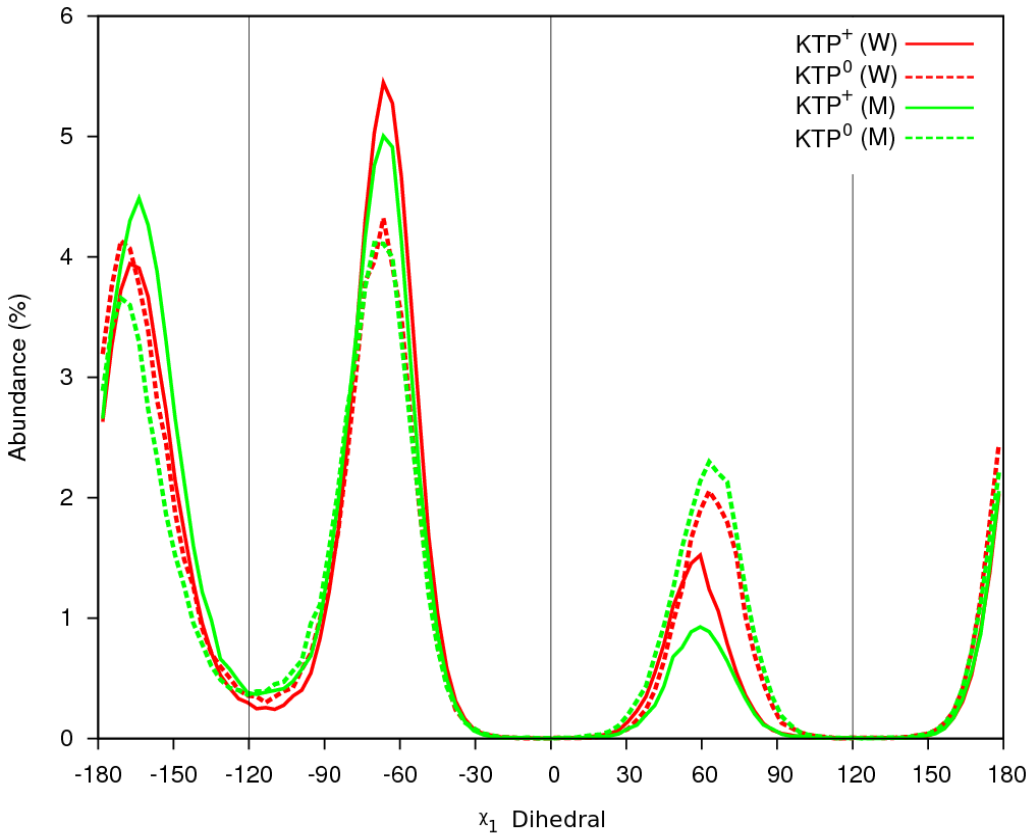

Figure 3.10 – χ_1 torsion angle for the pure states KTP^+ and KTP^0. Depicted are both water and membrane systems.

From Figure 3.10 (and also from the corresponding Table 3.3) we can establish a more direct comparison between our results and those of **Machuqueiro2010**. We observe that the population at χ_1=60º is also lower in the membrane when compared to the water system (for the KTP^+ system). While this is in agreement with the observations made by **Machuqueiro2010**, the difference is very slight, which does not allow any conclusions to be drawn. However, it is worth mentioning that while these results do not show a trend as marked as in that particular study, it should be noted that our method is the one that best mimics a physiological scenario, and the method used by Machuqueiro *et al.* [**Machuqueiro2010**] might be overestimating what happens to the relative populations of the χ_1 dihedral.

Table 3.3 – Population percentages of the different KTP conformation groups. Depicted are the values for the water and membrane systems for the KTP^+ and KTP^0 states.

		Water			Membrane		
		cis	***trans***	**Total**	***cis***	***trans***	**Total**
KTP^+	χ_1=-60º	.86	46.14	47.01	1.33	44.15	45.49
	χ_1=180º	2.54	38.33	40.88	1.58	44.85	46.43
	χ_1=60º	.14	11.95	12.10	.35	7.71	8.07
	Total	3.55	96.43	**100.00**	3.26	96.72	**100.00**
	packed	1.57	.81	2.39	.77	.24	1.02
	extended	1.98	95.62	97.60	2.49	96.48	98.97
	Total	3.55	96.43	**100.00**	3.26	96.72	**100.00**
KTP^0	χ_1=-60º	8.26	31.59	39.85	12.87	28.56	41.43
	χ_1=180º	19.52	23.12	42.65	22.51	14.62	37.13
	χ_1=60º	9.36	8.12	17.49	13.49	7.93	21.42
	Total	37.14	62.84	**100.00**	48.86	51.11	**100.00**
	packed	16.67	.64	17.31	19.73	.22	19.96
	extended	20.47	62.20	82.68	29.13	50.89	80.03
	Total	37.14	62.84	**100.00**	48.86	51.11	**100.00**

3.4. KTP insertion in the membrane

We know from previous studies that KTP has a high partition coefficient for membranes [**Lopes2006a**, **Machuqueiro2010**]. We also know that the insertion occurs via both amino acid residue side-chains, i.e. the phenol group of the tyrosine residue, which is expected due to the nature of the group [**Machuqueiro2010**], and the positively charged arginine side-chain, which seems a bit less intuitive, but was also observed in **Machuqueiro2010**.

In Figure 3.11 we can observe the effect of pH on the relative position of the four groups of KTP and the lipid membrane. The zero is defined as the average z-axis position of the phosphorous atoms of the monolayer that is closest to the peptide. When looking at Figure 3.11**,** the first thing we notice is that, on average, the arginine group is more inserted in the membrane when compared to the tyrosine group. This was not observed in **Machuqueiro2010**, and is most likely due to the constant-pH MD methodology which allows the titration of the sites during the simulation.

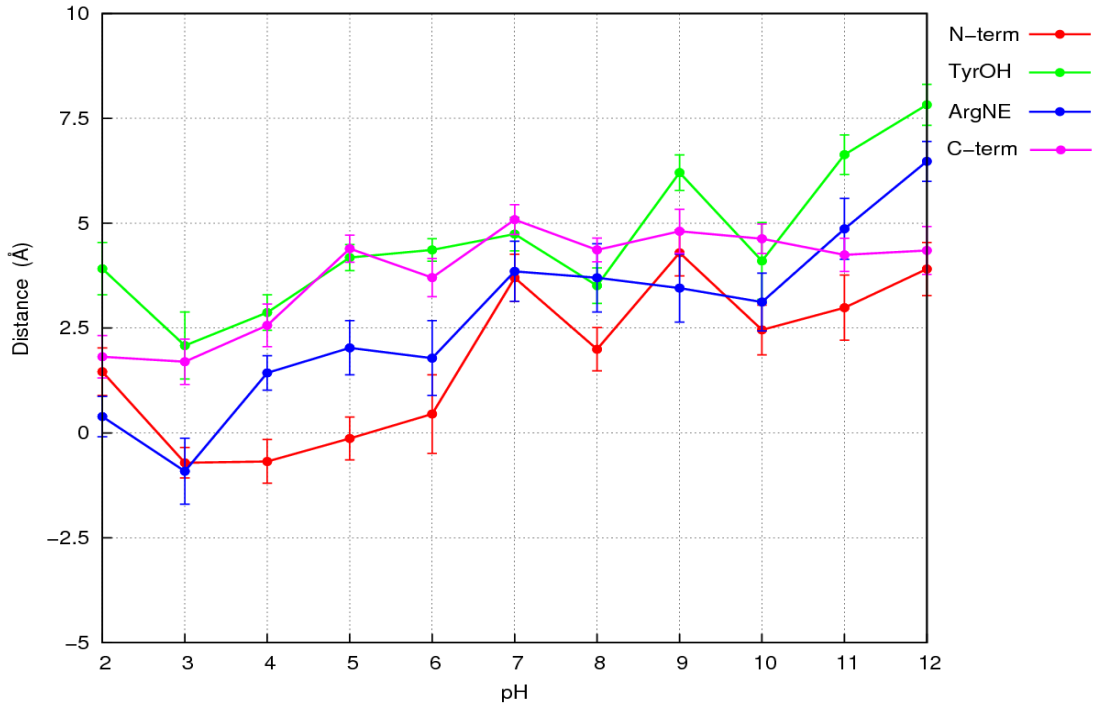

Figure 3.11 – Mean distance between the four groups of KTP and the lipid bilayer at different pH values. The error bars depicted are the error of the mean.

Let us examine each group individually:

1. The N-terminus group is protonated at lower pH values, and as such, possesses a positive charge. The insertion more likely occurs because of a 'dragging' effect of the tyrosine group. When buried in the membrane, the positively charged N-terminus will interact favorably with the negatively charged phosphate groups, and its distancing from the membrane will be made more difficult by the positively charged choline group layer. As the pH increases, the N-terminus loses its proton and at pH ~7 it distances itself from the membrane abruptly, due to the neutral N-terminus conformation being more relevant (pK_a = 7.38) and not as stable next to the phosphate headgroups (and also more able to cross the positively charged choline groups). In its neutral form, the N-terminus can still make some hydrogen bonds with the phosphate groups but is no longer as electrostatically constrained by the cholines and it can distance itself from the membrane.

2. The tyrosine group is in a neutral (protonated) form in most of our pH range. When the pH value becomes ~8, we can observe it moving away from the membrane. This is due to the fact that it begins titrating, and as it loses its proton and bears a negative charge, a repulsion effect occurs between it and the phosphate groups of the membrane, even though the positive charge of the cholines might attenuate this effect, which would explain why there is still some degree of insertion even at high pH values.

3. The arginine group is deeply inserted at lower pH values. As we've mentioned, this is a bit counter-intuitive due to its positive character. We speculate that this deep insertion at acidic pH values is due to a 'drag' effect from the C-terminus group. As the pH increases and reaches alkaline values (pH > 10) the arginine group is again dragged, this time by the tyrosine group, since they are likely in a compact *cis* conformation (see section 3.3). This is effect is much smaller at lower pH values since the amount of compact *cis* is also lower, and if the tyrosine group is leaving the membrane while KTP is an extended, albeit *cis*, conformation, it will not drag the arginine too far from the membrane before it is reprotonated and again reinserts in the membrane.

4. The C-terminus, spends most of its time in a negatively charged conformation. This occurs in the majority of pH values we studied, and as such, its average position (from pH 5 to 12) is by the positive choline layer, without crossing through to the phosphate groups. At pH < 4, it becomes protonated and hence, neutral, and is able to insert more deeply in the membrane.

We can see that all groups are able to penetrate the membrane, even if on average that might not be clear from Figure 3.11. When considering the global titration profile of KTP (Figure 3.3), when looking at Figure 3.11 we can see that it is easier for KTP to internalize when it has a positive charge (below the pI of ~9) rather than when it is in a negative form, which is somewhat counter-intuitive when looking at the lipid membrane as hydrophobic. This is also consistent with the observations of Lopes *et al.* [**Lopes2006a**], which have shown that in a membrane system similar to DMPC, KTP has a higher partition coefficient (inferred from the position of the tyrosine) at lower pH values when it has a positive charge, and a lower partition coefficient at higher pH values, when it loses the positive charge.

We present next a set of plots and snapshots depicting some of the more particular cases we have witnessed. First we have a snapshot taken from a replicate at pH 4, where we can see the N-terminus and tyrosine groups inserted in the membrane while the arginine and C-terminus groups remain at the surface (Figure 3.12b). From Figure 3.12a we can see that this configuration lasted for about 10 ns in this snapshot, which supports our idea that the transitions in a membrane system are much slower than in water (we can also see in the same image that from 60 to 100 ns, the average distance of the four groups remained approximately unchanged).

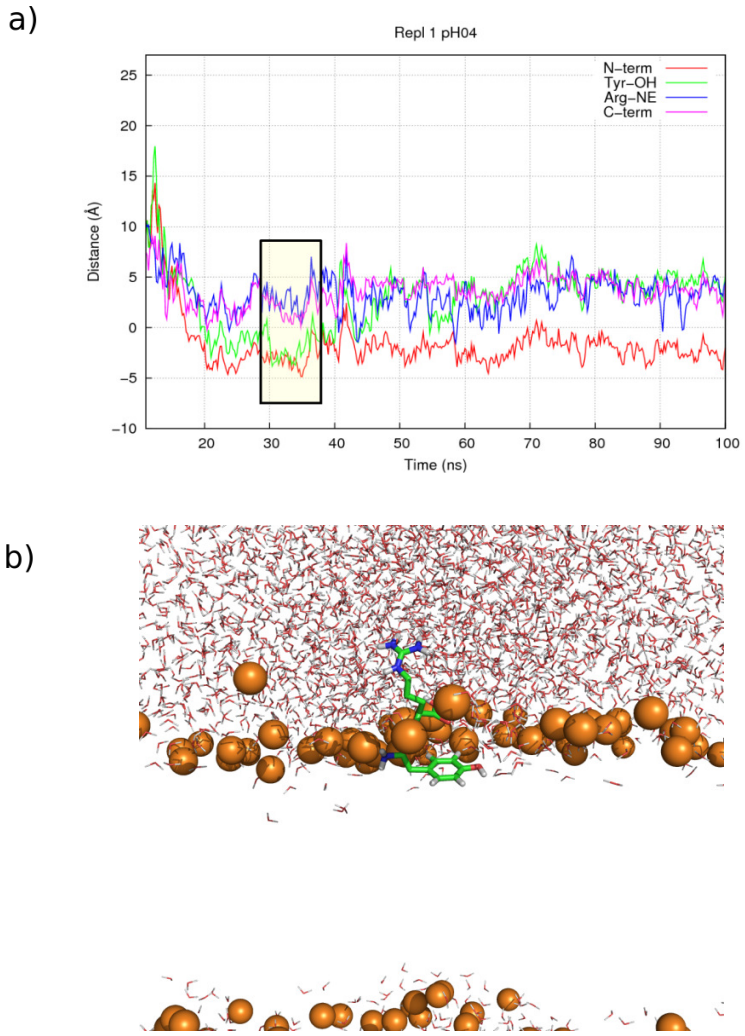

Figure 3.12 – Distance between the four groups of KTP and the lipid membrane (replicate 1 and pH=4). A snapshot of the section highlighted in (a) is presented (b). In the snapshot KTP is shown as stick and colored by element (carbon atoms in green, oxygen atoms in red, nitrogen atoms in blue, and hydrogen atoms in grey), the lipid acyl chains were hidden to help visualize the insertion of KTP and are depicted only as the phosphorous atoms as orange spheres

Next we'll look at another particular occurrence, namely, the deep insertion of the arginine side chain in the membrane, which, again, lasted for approximately 10 ns Figure 3.13a). A snapshot of this insertion is also presented (Figure 3.13b). In order for the arginine group to insert this deep into the membrane, due to a drag effect, the negatively charged C-terminus must also insert in the membrane,

albeit to a lesser extent. In order for this to happen, the also negatively charged phosphate groups must clear a path. This most likely happens through some sort of electrostatic shielding effect by the water molecules of the choline groups, that follow the insertion of the C-terminus into the membrane as the arginine group drags it behind.

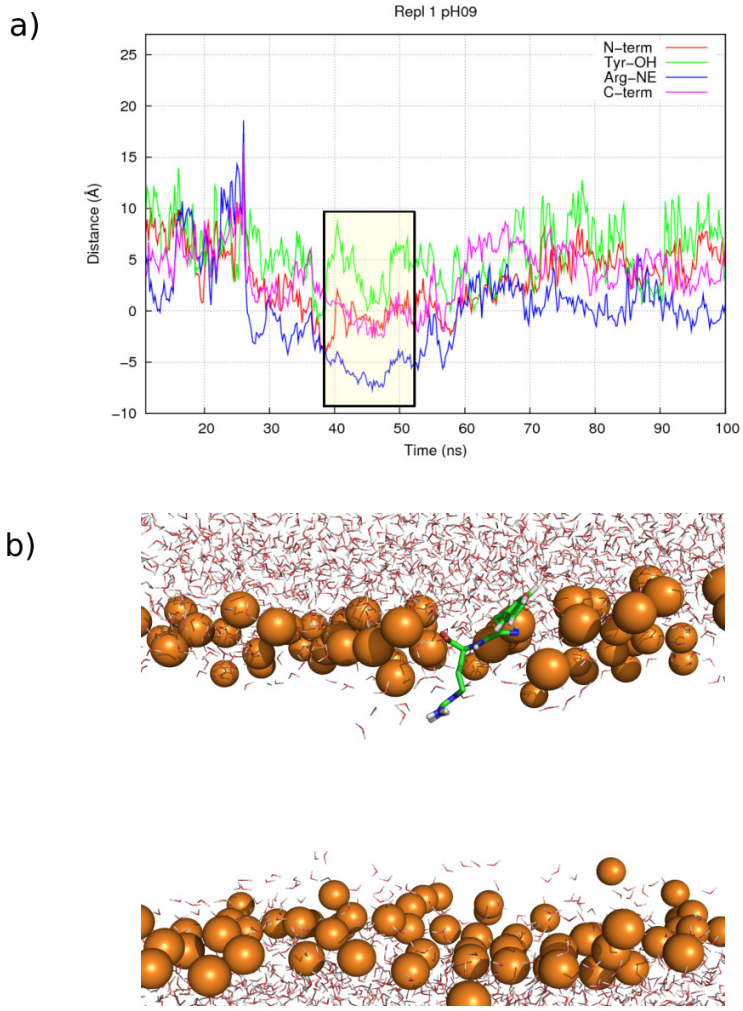

Figure 3.13 – Distance between the four groups of KTP and the lipid membrane (replicate 1 and pH=9). See legend of Figure 3.12 for further details.

In Figure 3.14 we can see an occurrence in one of the replicates, at pH = 10, where both side chains are inserted in the membrane. This level of insertion of both side chains does not occur very often and lasts only for a few nanoseconds, such as in the snapshot presented. In this case, even though this is a *cis* conformer, the peptide is in an extended form. This *cis* conformation induced some destabilization of the neighboring lipid molecules but not at the same level as a packed *cis* conformation would [**Machuqueiro2010**].

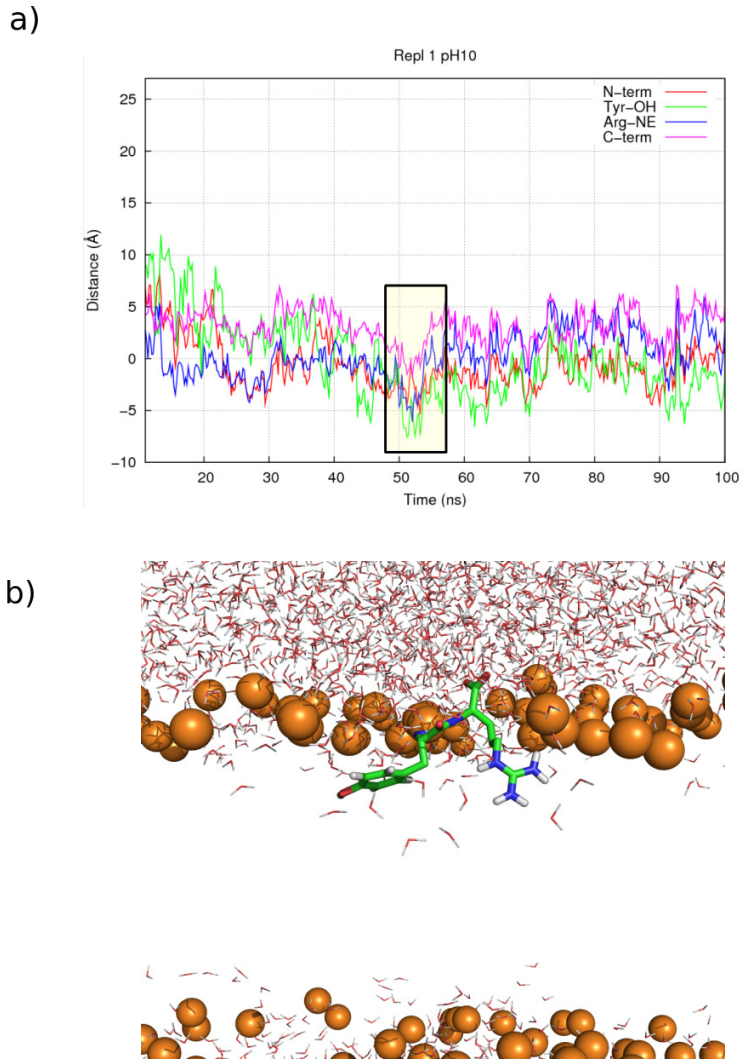

Figure 3.14 – Distance between the four groups of KTP and the lipid membrane (replicate 1 and pH=10). See legend of Figure 3.12 for further details.

Next we present an event that was previously unobserved. In Figure 3.15 we can see the distance between the peptide and the lipid bilayer raise abruptly. We observed this phenomenon at several high pH values (9-12) in some of the replicates, and in a few cases, more than once during a simulation (Figure 3.15a at t ≈ 23 ns and t ≈ 58 ns). The reason why this was not observed by Machuqueiro *et al.* [**Machuqueiro2010**] was because of the protonation states used in that study, where the tyrosine residue was always protonated. It is our understanding that at higher pH values, when the tyrosine group is titrating, the negative charge of the phenolate group would be repelled by the negative charges of the phosphates. The peptide will prefer to be solvated in solution, where the environment will help stabilize it. There is then a reapproximation to the membrane, but rarely a deep insertion. This event is also witnessed at a few lower pH values (pH = 5 and 7) but much less frequently. This led us to believe that since this distancing from the membrane was observed in our constant-pH MD studies and not in the MD analysis of **Machuqueiro2010**, the titration might be introducing some instability into the system, which favors this type of event. Secondly, since this distancing occurs at pH = 5 and 7, but much more frequently at pH > 9, the titration of the tyrosine residue might be making this rare event much more common. At pH = 5 and 7 the C-terminus and N-terminus groups are titrating, respectively, and this creates an instability, which might drive the peptide away from the membrane. The reason why we believe that the tyrosine residue titrating is making this much more common is because its titration is much more critical for repulsion purposes, which raises the level of instability when compared to the titration of the other groups.

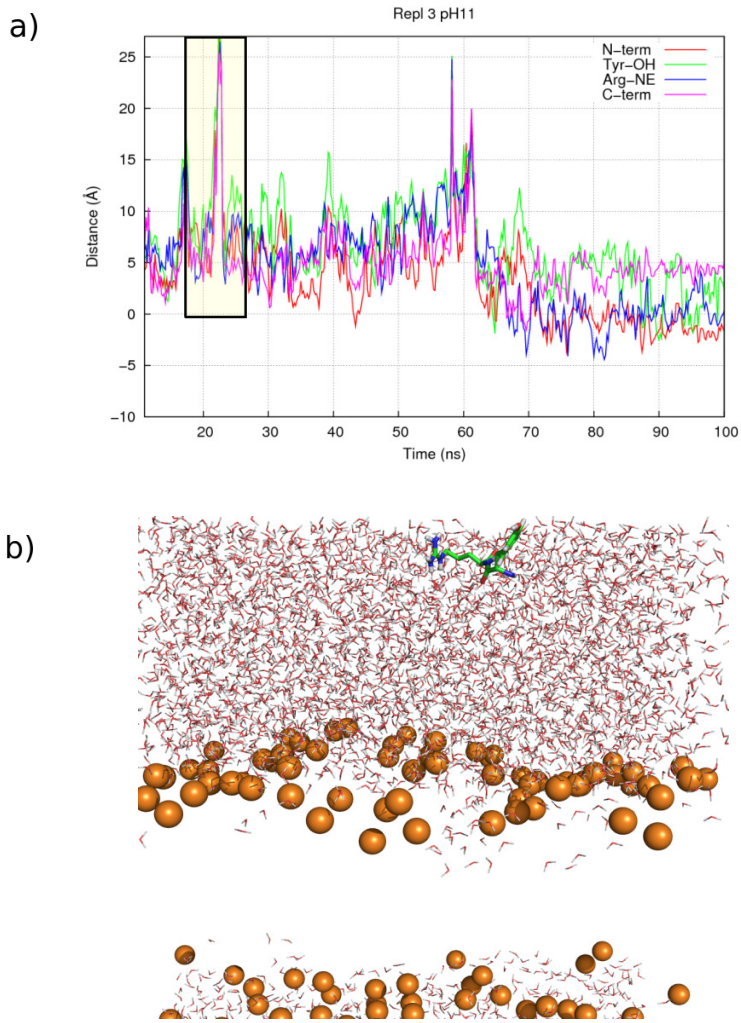

Figure 3.15 – Distance between the four groups of KTP and the lipid membrane (replicate 3 and pH=11). See legend of Figure 3.12 for further details.

Lastly, we present the case mentioned in a previous section, where KTP adopted a *cis* conformation at pH = 3 in one of the replicates. As we can see from Figure 3.16, the peptide is inserted in the membrane and in an extended *cis* conformation. As we have mentioned, this event only occurred in one of the replicates and only at this pH value. We believe that the peptide penetrated the membrane (t ≈ 70 ns) and after what should have been a transient, sporadic shift to a *cis* conformer, it became 'trapped' with that conformation for the remainder of the simulation (~ 30 ns). In order to clarify this, Figure

3.17 shows the same distance between one of the sites of KTP in Figure 3.16a (in this case, the N-terminus) but with the different conformers in distinguishing colors. Without further studies (namely prolonging the simulation and observing KTP's behavior) and more sampling, we cannot be sure if this is an artifact of the method or a rare event, and if so, what triggers it.

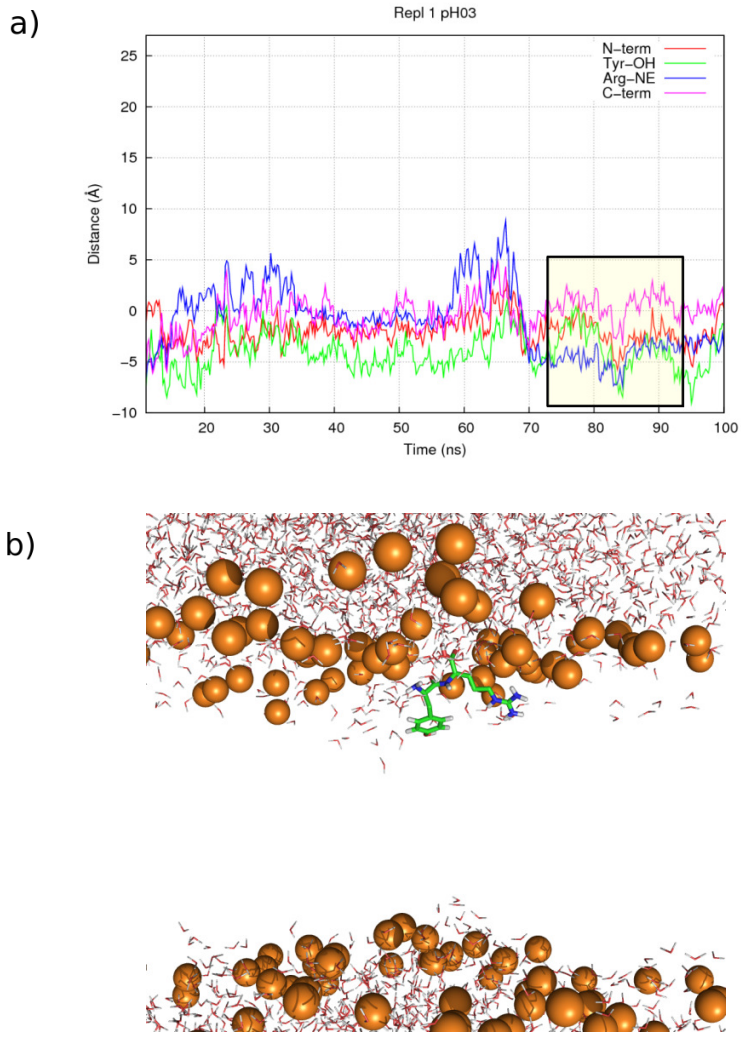

Figure 3.16 – Distance between the four groups of KTP and the lipid membrane (replicate 1 and pH=3). See legend of Figure 3.12 for further details.

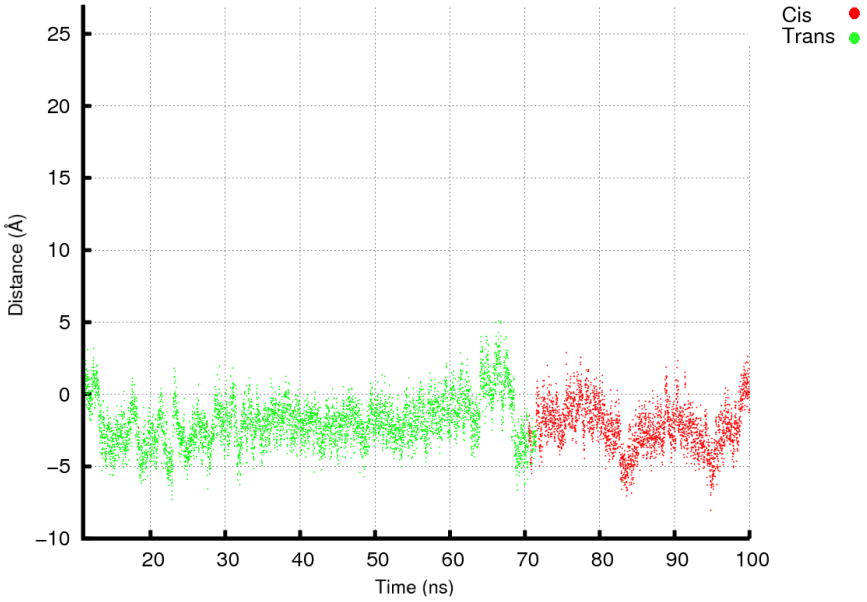

Figure 3.17 – Distance between the N-terminus site of KTP and the lipid membrane for the case in Figure 3.16. Depicted are the conformations of the peptide, in green the *trans* conformer, and in red, the *cis*.

3.5. Morphine-fit

As mentioned in the introduction, since part of KTP has a strong resemblance to morphine, we decided to determine the non-mass-weighed RMSD between all conformations of KTP in both water and membrane systems at all pH values (and replicates, in the case of the membrane system) and the structure of morphine (PDB ID 1Q0Y). In this calculation, the atoms N, Cα, Cβ, Cγ, Cδ1, Cδ2, Cε1, Cε2, Cζ and OH were considered. Due to the possible rotation of the phenol ring, two RMSD calculations were done for each conformation ($C_\delta1/C_\varepsilon1$ exchanged with $C_\delta2/C_\varepsilon2$), and the lowest value was chosen. We applied several cut-offs for the RMSD, and below lowest one (0.2 Å), which contains the best structures, we obtained 56 structures (0,018% of all structures), for the membrane system. These are represented in Figure 3.18, and as we can observe, they overlap perfectly with the morphine ring. This shows that KTP can fulfill the structural constraints present in morphine (see section 1.1.2), something that was already observed in [**Machuqueiro2007**] for the water system. For the water system, we obtained 195 structures below the same cut-off, which again correspond to 0.018% of all structures

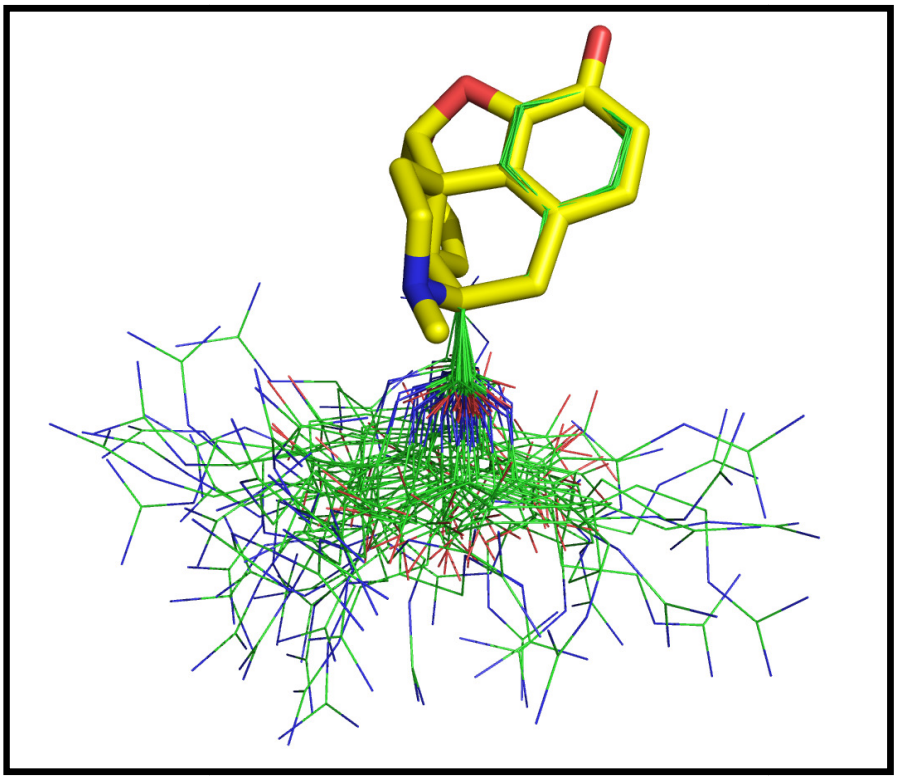

Figure 3.18 – Fit of the 56 best conformations of KTP to morphine.

Next, we plotted the percentage of total structures according to their RMSD values, for both systems (Figure 3.19), and both curves turned out to be very similar, with the water system shifted slightly to the right. We were looking for a more pronounced shift to higher values in the water system, in order to accept that the membrane would induce a more receptor-ready conformation, but since both curves almost overlap, we can make no such claims, and both water and membrane systems allow KTP to adopt conformations that fulfill the constraints present in morphine.

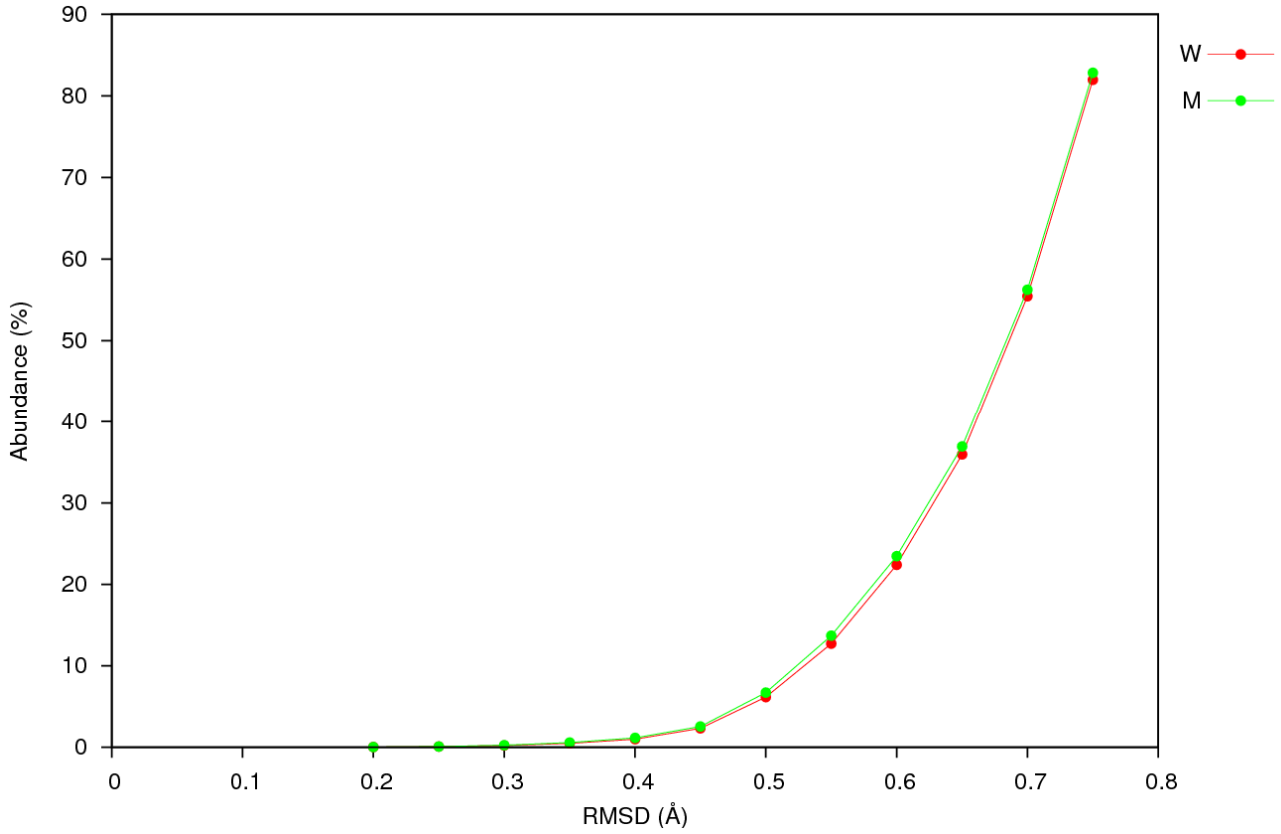

Figure 3.19 – Percentage of structures with different cut-offs of RMSD. Water (W) and membrane (M) systems depicted.

4. Conclusions

At the start of this work, we knew from experimental results that KTP favored a lipid membrane environment [**Lopes2006a**]. We also knew, from previous simulations studies, that it showed a conformational dependence on pH in an aqueous media [**Machuqueiro2007**], and that its interaction with a membrane system somehow limited its conformational freedom [**Machuqueiro2010**]. What we have observed has come to show an agreement with previously made observations as well as shed some light on a few new ones. We determined the pK_a values for all four titratable sites of KTP, and were able to observe the differences due to the presence of the membrane. We determined that the membrane further stabilizes or destabilizes some of the KTP forms and shifts the pK_a values of the individual groups to higher (tyrosine and C-terminus groups) or lower values (N-terminus), with particular emphasis on the latter, since the N-terminus is the main factor responsible for the protonation state of KTP at physiological pH values.

Regarding the conformational aspects of KTP, we determined that dependence of the $C\beta_{Tyr}$-$C\alpha_{Tyr}$-$C\alpha_{Arg}$-$C\beta_{Arg}$ dihedral on pH is witnessed in both water and membrane systems, with few differences being observed between the two. This dependence is in good agreement with the results obtained in **Machuqueiro2007** (for the water system) We have also observed that as the pH value increases, the *cis* conformer is favored, and as such, the interaction between the tyrosine and arginine side chains is evident, while at the same time, a large population of extended *cis* conformers (with the side chains not interacting) being observed in both water and membrane systems. These extended *cis* conformers exhibit their side chains facing the solvent, most likely stabilized by solvation (also witnessed in **Machuqueiro2007**). We also looked at the χ_1 dihedral angle of tyrosine, and observed a small drop in the population at 60º, which is in agreement with recent observations by Machuqueiro *et al.* [**Machuqueiro2010**] which are associated with the loss of some internal rotations of the peptide when interacting with the membrane. This is due to the structural similarity between morphine and some KTP conformations which is in accordance with the lipid catalysis model.

The degree of insertion of KTP in the lipid membrane was also measured, and as expected we observed the phenol group of tyrosine penetrating the membrane, but at a lower depth when compared to the arginine residue, something that we did not expect, and have since speculated about the cause. The methodology we used also allowed us to witness something previously unseen in these studies, which is the distancing of the peptide at high pH values (also at low pH values, but rarely) which we have suggested to be connected to the titration introduced by the constant-pH MD method causing an

instability which leads the peptide away from the membrane.

Lastly, we have fitted the tyrosine group of KTP onto part of the structure of morphine, an opioid compound with a known receptor interaction. What we observed was an almost perfect overlap between some of our structures and morphine, which shows that KTP is able to fulfill some of the structural constraints imposed by receptor binding, something which was also observed by Machuqueiro *et al.* [**Machuqueiro2007**], but for the water system.

In conclusion, our studies have proved to be consistent with previous studies and have shown that the constant-pH MD method used is capable of simulating membrane systems. This gives us a powerful tool that explicitly includes the effect of pH in membrane simulation studies.

References

[**Aburi2002**]

Aburi, M. and Smith, P.E. (2002) A conformational analysis of leucine enkephalin as a function of pH. *Biopolymers* 64:177-188.

[**Alder1957**]

Alder, B.J. and Wainwright, T.E. (1957) Phase transition for a hard sphere system. *J. Chem. Phys.* 27:1208-1209.

[**Alder1959**]

Alder, B.J. and Wainwright, T.E. (1959) Studies in molecular dynamics. I. General method. *J. Chem. Phys.* 31:459-466.

[**Allen1987**]

Allen, M.P. and Tildesley, D.J. (1987) Computer Simulation of Liquids. Oxford University Press, New York.

[**Amano1984**]

Amano, H., Morimoto, Y., Kaneko, S., Takagi, H. (1984) Opioid activity of enkephalin analogues containing the kyotorphin-related structure in the N-terminus. *Neuropharmacology* 23, 4:395-400.

[**Antosiewicz1989**]

Antosiewicz, J. and Porschke, D. (1989) The nature of protein dipole-moments experimental and calculated permanent dipole of α-chymo-trypsin. *Biochemistry* 28:10072–10078.

[**Arima1997**]

Arima, T., Kitamura, Y., Nishiya, T., Taniguchi, T., Takagi, H., Nomura, Y. (1997) Effects of kyotorphin (L-Tyrosyl-L-Arginine) on [3H]NG-Nitro-L-Arginine binding to neuronal nitric oxide synthase in rat brain. *Neurochem. Int.* 30, 6:605-611.

[**Baptista1997**]

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

[**Baptista1999**]

Baptista, A.M., Martel, P.J. and Soares, C.M. (1999) Simulation of electron-proton coupling with a Monte Carlo method: application to cytochrome c_3 using continuum electrostatics. *Biophys. J.* 76:2978-2998.

[**Baptista2001**]

Baptista, A.M., Soares, C.M. (2001) Some theoretical and computational aspects of the inclusion of proton tautomerism in the protonation equilibrium of proteins. *J. Phys. Chem.* 105:293-309.

[**Baptista2002a**]

Baptista, A.M. (2002) Comment on “Explicit-solvent molecular dynamics simulation at constant pH: methodology and application to small amines” [J. Chem. Phys. 114, 9706 (2001)]. *J. Chem. Phys.* 116, 17:7766-7768.

[**Baptista2002b**]

Baptista, A.M., Teixeira, V.H. and Soares, C.M. (2002) Constant-pH molecular dynamics using stochastic titration. *J. Chem. Phys.* 117, 9:4184-4200.

[**Barojas1973**]

Barojas, J., Levesque, D. and Quentrec, B. (1973) Simulation of diatomic homonuclear liquids. *Phys. Rev.* A7:1092-1105.

[**Bashford1990**]

Bashford, D. and Karplus, M. (1990) pK_a's of ionizable groups in proteins: atomic detail from a continuum electrostatic model. *Biochemistry* 29:10219-10225.

[**Bashford1992**]

Bashford, D. and Gerwert, K. (1992) Electrostatic calculations of the pK_a values of ionizable groups in bacteriorhodopsin. *J. Mol. Biol.* 224:473-486.

[**Bashford2004**]

Bashford, D. (2004) Macroscopic electrostatic models for protonation states in proteins. *Front. Biosci.* 9:1082-1099.

[**Ben-Naim1992**]

Ben-Naim, A. (1992) *Statistical thermodynamics for chemists and biochemists*. Plenum Press, New York.

[**Berendsen1984**]

Berendsen, H.J.C., Postma, J.P.M., van Gunsteren, W.F., DiNola, A. and Haak, J.R. (1984) Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* 81, 8:3684-3690.

[**Berendsen1985**]

Berendsen, H.J.C. in Hermans, J. (Ed.) (1985) *Molecular dynamics and protein structure*, Polycrystal Book Service, Western Springs, IL, USA, 18.

[**Berendsen1995**]

Berendsen, H.J.C., van der Spoel, D., van Drunen, R. (1995) GROMACS: a message-passing parallel molecular dynamics implementation. *Comput. Phys. Commun.* 91:43-56.

[**Beroza1991**]

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

[**Börjesson2001**]

Börjesson, U. and Hünenberger, P.H. (2001) Explicit-solvent molecular dynamics simulation at constant pH: methodology and application to small amines. *J. Chem. Phys.* 114, 22:9706-9719.

[**Börjesson2004**]

Börjesson, U. and Hünenberger, P.H. (2004) pH-dependent stability of a decalysine α-helix studied by explicit-solvent molecular dynamics simulations at constant pH. *J. Phys. Chem. B* 108:13551-13559.

[**Burgi2002**]

Bürgi, R., Kollman, P.A., and van Gunsteren, W.F. (2002) Simulating proteins at constant pH: an approach combining molecular dynamics and Monte Carlo simulation. *Proteins: Struct. Funct. Genet.* 47:469-480.

[**Chandrasekhar2005**]

Chandrasekhar, I., Bakowies, D., Glattli, A., Hunenberger, P., Pereira, C. and van Gunsteren W.F. (2005) Molecular dynamics simulation of lipid bilayers with GROMOS96: application of surface tension. *Molecular Simulation* 31:543-548.

[**Chen1998**]

Chen, P., Bodor, N., Wu, W. and Prokai, L. (1998) Strategies to target kyotorphin analogues to the brain. *J. Med. Chem.* 41:3773-3781.

[**DeLano2002**]

DeLano, W.L. (2002) The PyMOL molecular graphics system, www.pymol.org

[**Dlugosz2004a**]

Dlugosz, M., Antosiewicz, J.M., Robertson, A.D. (2004) Constant-pH molecular dynamics study of protonation-structure relationship in a heptapeptide derived from ovomucoid third domain. *Phys. Rev. E.* 69:021915

[**Dlugosz2004b**]

Dlugosz, M., Antosiewicz, J.M. and Robertson, A.D. (2004) Constant-pH molecular dynamics simulations: a test case of succinic acid. *Chem. Phys.* 302:161-170.

[**Dyson1991**]

Dyson, H.J. and Wright, P.E. (1991) Defining solution conformations of small linear peptides. *Annu. Rev. Biophys. Biophys. Chem.* 20:519-538.

[**Eberini2004**]

Eberini, I., Baptista, A.M., Gianazza, E., Fraternali, F. and Beringhelli, T. (2004) Reorganization in apo- and holo-β-lactoglobulin upon protonation of Glu89: molecular dynamics and pK_a calculations. *Proteins: Struct. Funct. Bioinf.* 54:744-758.

[**Feig2004**]

Feig, M. and Brooks, C.L., III (2004) Recent advances in the development and application of implicit solvent models in biomolecule simulations. *Curr. Opin. Struct. Biol.* 14:217-224.

[**Fogolari2002**]

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

[**Frauenfelder1998**]

Frauenfelder, H. and Leeson, D.T. (1998) The energy landscape in non-biological and biological molecules. *Nature Struct. Biol.* 5, 9:757-759.

[**Fukui1983**]

Fukui, K., Shiomi, H., Takagi, H., Hayashi, K., Kiso, Y., Kitagawa, K. (1983) Isolation from bovine brain of a novel analgesic pentapeptide, neo-kyotorphin, containing the Tyr-Arg (kyotorphin) unit. *Neuropharmacology* 22, 2:191-196.

[**Gilson1987**]

Gilson, M.K., Sharp, K.A. and Honig, B. (1987) Calculating the electrostatic potential of molecules in solution: method and error assessment. *J. Comput. Chem*. 9:327-335.

[**Giupponi2008**]

Giupponi, G., Harvey, M.J. and De Fabritiis, G. (2008) The impact of accelerator processors for high-

throughput molecular modeling and simulation. *Drug Discov. Today* 13, 23/24:1052-1058.

[**Grimsley2009**]
Grimsley, G.R., Scholtz, J.M., Pace, C.N. (2009) A summary of the measured p*K* values of the ionizable groups in folded proteins. *Protein Sci.* 18, 1:247-251.

[**Hess1997**]
Hess, B., Bekker, H., Berendsen, H.J.C., Fraaije, J.G.E.M. (1997) LINCS: a linear constraint solver for molecular simulations. *J. Comput. Chem.* 18:1463-1472.

[**Khandogin2005**] Khandogin, J. and Brooks, C.L., III (2005) Constant pH molecular dynamics with proton tautomerism. *Biophys. J.* 89:141-157.

[**Kong1996**]
Kong, X. and Brooks, C.L., III (1996) λ-dynamics: a new approach to free energy calculations. *J. Chem. Phys.* 105, 6:2414-2423.

[**Lapalu1997**]
Lapalu, S., Moisand, C., Mazarguil, H., Cambois, G., Mollereau, C. and Meunier, J.C. (1997) Comparison of the structure-activity relationships of nociceptin and dynorphin A using chimeric peptides. *FEBS Lett.* 417:333-336.

[**Lapalu1998**]
Lapalu, S., Moisand, C., Butour, J.L., Mollereau, C., and Meunier, J.C. (1998) Different domains of the ORL1 and κ-opioid receptors are involved in recognition of nociceptin and dynorphin A. *FEBS Lett.* 427:296-300.

[**Leach2001**]
Leach, A.R. (2001) *Molecular modelling: principles and applications*. Pearson Education Limited, London.

[**Lee2004**]

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

[**Lewis1983**]

Lewis, R.V., Stern, A.S. (1983) Byosynthesis of the enkephalins and enkephalin-containing polypeptides. *Annu. Rev. Pharmacol. Toxicol.* 23:353-372.

[**Lindahl2001**]

Lindahl, E., Hess, B., van der Spoel, D. (2001) GROMACS 3.0: a package for molecular simulation and trajectory analysis. *J. Mol. Model.* 7:306-317.

[**Lopes2006a**]

Lopes, S.C.D.N., Soares, C.M., Baptista, A.M., Goormaghtigh, E., Cabral, B.J.C. and Castanho, M.A.R.B. (2006) Conformational and orientational guidance of the analgesic dipeptide kyotorphin induced by lipidic membranes: putative correlation toward receptor docking. *J. Phys. Chem. B* 110:3385-3394.

[**Lopes2006b**]

Lopes, S.C.D.N., Fedorov, A. and Castanho, M.A.R.B. (2006) Chiral recognition of D-kyotorphin by lipidic membranes: relevance toward improved analgesic efficiency. *ChemMedChem.* 1:723-728.

[**Macdonald1987**]

Macdonald, R.C., Simon, S.A. (1987) Lipid monolayer states and their relationships to bilayers. *Proc. Natl. Acad. Sci. USA* 84:4089-4093.

[**Machuqueiro2006**]

Machuqueiro, M. and Baptista, A.M. (2006) Constant-pH molecular dynamics with ionic strength effects: Protonation-conformation coupling in decalyisine. *J. Phys. Chem. B* 110:2927-2933.

[**Machuqueiro2007**]

Machuqueiro, M. and Baptista, A.M. (2007) The pH-dependent conformational states of kyotorphin: a

constant-pH molecular dynamics study. *Biophys. J.* 92:1836-1845.

[**Machuqueiro2008**]

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

[**Machuqueiro2010**]

Machuqueiro, M., Campos, S.R.R., Soares, C.M. and Baptista, A.M. (2009) The membrane-induced conformational states of kyotorphin: a molecular dynamics study. (to be submitted)

[**Matsubayashi1984**]

Matsubayashi, K., Kojima, C., Kawajiri, S., Ono, K., Takegoshi, T., Ueda, H. and Takagi, H. (1984) Hydrolytic deactivation of kyotorphin by the rodent brain homogenates and sera. *J. Pharmacobiodyn.* 7:479-484.

[**McCammon1977**]

McCammon, J.A., Gelin, B.R. and Karplus, M. (1977) Dynamics of folded proteins. *Nature* 267:585-590.

[**Mertz1994**]

Mertz, J.E. and Pettitt, B.M. (1994) Molecular dynamics at a constant-pH. *Int. J. High Perform. Comput. Appl.* 8, 1:47-53.

[**Metropolis1953**]

Metropolis, N., Rosenbluth, A.W., Rosenbluth, M.N., Teller, A.H. And Teller, E. (1953) Equation of state calculations by fast computing machines. *J. Chem. Phys.* 21:1087-1092.

[**Mongan2004**]

Mongan, J., Case, D.A., McCammon, J.A. (2004) Constant pH molecular dynamics in Generalized Born implicit solvent. *J. Comput. Chem.* 25:2038-2048.

[**Morrel1936**]

Morrel, W.E. and Hildebrand, J.H. (1936) The distribution of molecules in a model liquid. *J. Chem. Phys.* 4:224-227.

[**Nagle2000**]

Nagle, J.F. and Tristram-Nagle, S. (2000) Lipid bilayer structure. *Curr. Opin. Struct. Biol.* 10:474-480.

[**Nielsen2009**]

Nielsen, J.E. (2009) Analyzing enzymatic pH activity profiles and protein titration curves using structure-based pK_a calculations and titration curve fitting. *Methods Enzymol.* 54:233-58.

[**Nozaki1967**]

Nozaki, Y. and Tanford, C. (1967) Examination of titration behavior. *Methods Enzimol.* 11:715:734.

[**Ochi2000**]

Ochi, T., Motoyama, Y. and Goto, T. (2000) The spinal antinociceptive effect of FR1401423 is mediated through kyotorphin receptors. *Life Sciences* 66, 23:2239-2245.

[**Oostenbrink2004**]

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

[**Patrick2001**]

Patrick, G.L. (2001) *An introduction to medicinal chemistry*, 2nd ed. Oxford University Press. New York.

[**Pierański1978**]

Pierański, P., Malecki, J., Kuczynski, W. and Wojciechowski, K. (1978) A hard disc system, an experimental model. *Phil. Mag.* A37:107-115.

[**Portlock1992**]

Portlock, S.H., Lee, Y., Tomich, J.M. and Tamm, L.K. (1992) Insertion and folding of the amino-terminal amphiphilic signal sequences of the mannitol and glucitol permeases of Escherichia coli. *J. Biol. Chem.* 267:11017-11022.

[**Prasad1995**]

Prasad, C. (1995) Bioactive cyclic dipeptides, *Peptides* 16, 1:151-164.

[**Press1992**]

Press, W. H., Teukolsky, S. A., Vetterling, W. T., Flannery, B. P. (1992) *Numerical recipes in C, the art of scientific computing*. 2nd ed, Cambridge University Press, New York.

[**Rackham1982**]

Rackham, A., Wood, P.L. and Hudgin, R.L. (1982) Kyotorphin (tyrosine-arginine): Further evidence for indirect opiate receptor activation. *Life Sciences* 30:1337-1342.

[**Rahman1964**]

Rahman, A. (1964) Correlations in the motion of atoms in liquid argon. *Phys. Rev.* 136A:405-411.

[**Rahman1971**]

Rahman, A. and Stillinger, F.H. (1971) Molecular dynamics study of liquid water. *J. Chem. Phys.* 55:3336-3359.

[**Ryckaert1975**]

Ryckaert, J.P. and Bellemans, A. (1975) Molecular dynamics of liquid n-butane near its boiling point. *Chem. Phys. Lett.* 30:123-125.

[**Sakane1989**]

Sakane, T., Tanaka, C., Yamamoto, A., Hashida, M., Sezaki, H., Ueda, H. and Takagi, H. (1989) The effect of polysorbate 80 on brain uptake and analgesic effect of D-kyotorphin. *Int. J. Pharm.* 57:77-83.

[**Sargent1986**]

Sargent, D.F. and Schwyzer, R. (1986) Membrane lipid phase as catalyst for peptide-receptor interactions. *Proc. Natl. Acad. Sci. USA* 83:5774-5778.

[**Scott1999**]

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

[**Seelig1987**]

Seelig, A. (1987) Local-anesthetics and pressure – a comparison of dibucaine binding to lipid monolayers and bilayers. *Biochim. Biophys. Acta* 899:196-204.

[**Shiomi1981a**]

Shiomi, H., Kuraishi, Y., Ueda, H., Harada, Y., Amano, H. and Takagi, H. (1981) Mechanism of kyotorphin-induced release of Met-enkephalin from pig striatum and spinal cord. *Brain Res.* 221:161-169.

[**Shiomi1981b**]

Shiomi, H., Ueda, H. and Takagi, H. (1981) Isolation and identification of an analgesic opioid dipeptide kyotorphin (Tyr-Arg) from bovine brain. *Neuropharmacology* 20, 7:633-638.

[**Takagi1979a**]

Takagi, H., Shiomi, H., Ueda, H. and Amano, H. (1979) Morphine-like analgesia by a new dipeptide, L-Tyrosyl-L-Arginine (kyotorphin) and its analogue. *Eu. J. Pharm.* 55:109-111.

[**Takagi1979b**]

Takagi, H., Shiomi, H., Ueda, H. and Amano, H. (1979) A novel analgesic dipeptide from bovine brain is a possible Met-enkephalin releaser. *Nature* 282, 5737:410-412.

[**Thurlkill2006**]

Thurlkill, R.L., Grimsley, G.R., Scholtz, J.M. and Pace, C.N. (2006) pK values of the ionizable groups

of proteins. *Protein Sci.* 15:1214-1218.

[**Tironi1995**]

Tironi, I.G., Sperb, R., Smith, P.E., van Gunsteren, W.F. (1995) A generalized reaction field method for molecular dynamics simulations. *J. Chem. Phys*. 102, 13:5451-5459.

[**Ueda1980**]

Ueda, H., Shiomi, H. and Takagi, H. (1980) Regional distribution of a novel analgesic dipeptide kyotorphin (Tyr-Arg) in the rat brain and spinal cord. *Brain Res.* 198, 2:460-464.

[**Ueda1985a**]

Ueda, H., Yoshihara, Y., Nakamura, A., Shiomi, H., Satoh, M. and Takagi, H. (1985) How is kyotorphin (Tyr-Arg) generated in the brain? *Neuropeptides* 5:525-528.

[**Ueda1985b**]

Ueda, H., Ming, G., Hazato, T., Katayama, T. and Takagi, H. (1985) Degradation of kyotorphin by a purified membrane-bound-aminopeptidase from monkey brain: Potentiation of kyotorphin-induced analgesia by a highly effective inhibitor, bestatin. *Life Sciences* 36:1865-1871.

[**Ueda1986a**]

Ueda, H., Matsumoto, S., Yoshihara, Y., Fukushima, N. and Takagi, H. (1986) Uptake and release of kyotorphin in rat brain synaptosomes. *Life Sciences* 38:2405-2411.

[**Ueda1986b**] Ueda, H., Yoshihara, Y. and Takagi, H. (1986) A putative Met-enkephalin releaser, kyotorphin enhances intracellular Ca^{2+} in the synaptosomes. *Biochem. Biophys. Res. Commun*. 137, 2:897-902.

[**Ueda1987a**]

Ueda, H., Yoshihara, Y., Fukushima, N., Shiomi, H., Nakamura, A. and Takagi, H. (1987) Kyotorphin (tyrosine-arginine) synthetase in rat brain synaptosomes. *J. Biol. Chem*. 262, 17:8164-8173.

[**Ueda1987b**]

Ueda, H., Ming, G., M., Satoh, M. and Takagi, H. (1987) Non-opioid analgesia of the neuropeptide, neo-kyotorphin and possible mediation by inhibition of GABA release in the mouse brain. *Peptides* 8:905-909.

[**Ueda1987c**]

Ueda, H., Tamura, S., Satoh, M. and Takagi, H. (1987) Excess release of substance P from the spinal cord of mice during morphine withdrawal and involvement of the enhancement of presynaptic Ca^{2+} entry. *Brain Res*. 425, 1:101-105.

[**Ueda1989**]

Ueda, H., Yoshihara, Y., Misawa, H., Fukushima, N., Katada, T., Ui, M., Takagi, H., Satoh, M. (1989) The kyotorphin (tyrosine-arginine) receptor and selective reconstruction with purified G_i, measured with GTPase and phospholipase C assays. *J. Biol. Chem*. 264, 7:3732-3741.

[**van der Spoel2005**]

van der Spoel, D., Lindahl, E., Hess, B., van Buuren, A.R., Apol, E., Meulenhoff, P.J., Tieleman, D.P., Sijbers, A.L.T.M., Feenstra, K.A., van Drunen, R. and Berendsen, H.J.C. (2005) *Gromacs user manual* version 3.3, www.gromacs.org

[**van Gunsteren1990**]

van Gunsteren, W.F. and Berendsen, H.J.C. (1990) Computer simulation of molecular dynamics: Methodology, applications and perspectives in chemistry. *Angew Chem Int Ed Engl*. 29:992-1023.

[**Walczak2002**]

Walczak, A.M., and Antosiewicz, J.M. (2002) Langevin dynamics of proteins at constant-pH. *Phys. Rev. E*. 66:051911.

[**Warshel1979**]

Warshel, A. (1979) Calculations of chemical processes in solutions. *J. Phys. Chem*. 83, 12:1640-1652.

[**Warshel1986**]

Warshel, A., Sussman, F. and King, G. (1986) Free energy of charges in solvated proteins: microscopic calculations using a reversible charging process. *Biochemistry*. 25:8368-8372.

[**Williams2004**]

Williams, T., Kelley, C., et al. (2004) Gnuplot version 4.2, www.gnuplot.info.