UNIVERSIDADE DE LISBOA
FACULDADE DE CIÊNCIAS

In silico tumor-targeting technologies for the evasion of acidity-induced multidrug resistance

"Documento Definitivo"

Doutoramento em Bioquímica
Bioquímica Teórica

Tomás Fernandes da Silva

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

Documento especialmente elaborado para a obtenção do grau de doutor

2022

UNIVERSIDADE DE LISBOA
FACULDADE DE CIÊNCIAS

In silico tumor-targeting technologies for the evasion of acidity-induced multidrug resistance

Doutoramento em Bioquímica
Bioquímica Teórica

Tomás Fernandes da Silva

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

Júri:
Presidente:

- Doutor Manuel Eduardo Ribeiro Minas da Piedade, Professor Catedrático,Presidente do Departamento de Química e Bioquímica da Faculdade de Ciências da Universidade de Lisboa

Vogais:

- Doutor John Blakely Mertz, Principal Scientist do Modulus Discovery (Estados Unidos)
- Doutor António Manuel Simões Carrão Albuquerque Baptista, Investigador Principal do Instituto de Tecnologia Química e Biológica António Xavier da Universidade Nova de Lisboa
- Doutor Manuel Nuno de Sousa Pereira Simões de Melo, Investigador Auxiliar do Instituto de Tecnologia Química e Biológica António Xavier da Universidade Nova de Lisboa
- Doutor Rodrigo Freire Martins de Almeida, Professor Auxiliar da Faculdade de Ciências da Universidade de Lisboa
- Doutor Miguel Ângelo dos Santos Machuqueiro, Investigador Auxiliar da Faculdade de Ciências da Universidade de Lisboa, Orientador

Documento especialmente elaborado para a obtenção do grau de doutor

Esta tese foi apoiada financeiramente pela FCT através da bolsa SFRH/BD/140886/2018

2022

Acknowledgments

I would like to begin to express my deepest gratitude for being under the guidance of my supervisors: Miguel Machuqueiro and Diogo Vila Viçosa. In their own way, both of them have significantly shaped my values and perspectives on the academic field (and more) and all that it entails. First, I would like to thank Miguel for all the support and opportunities to scientifically and also personally grow with every project and discussion. In these last 6 years, he has been an outstanding mentor and role model of patience, dedication and perseverance in an often inhospitable academic medium. I would also like to extend my thanks to Diogo, as he often times reminded me the (lack of) importance of estimating error bars in (not really) fundamentally wrong insertion methods. On a more serious note, Diogo consistently embodied the versatility, knowledge and curiosity that know no bounds, despite all the trials and tribulations one might find along the way.

I am also thankful to Bruno Victor, Paulo Costa and Pedro Magalhães for all the helpful scientific contributions and daily discussions on the most varied topics ranging from political discourse, football or idiosyncratic stories. They turned this treaded path more fun and interesting at all moments. A special thanks must go to Pedro Reis, for all the help in designing and building all the tools developed and projects discussed as without him, at the very least, our work would be significantly harder. Still, more importantly, for being a friend that endured the same challenges ever since our bachelor's. Last but not least, I would like to extend my thanks to all the people in our group for all the pleasant occasions be it in the workspace or in a conference trip.

On a personal note, I need to sincerely thank my parents, Maria José and José Manuel Silva, and my siblings, Beatriz, Leonor and Dinis. Without their loving and constant support and education, I wouldn't be the person who I am today, nor would I have been able to follow my dreams. I would also like to thank all my friends, for the late night gaming, dinners and jokes, as they were fond memories that helped throughout these last few years. Finally, I cannot begin to express my amusement of leaving my dear girlfriend, Beatriz Pimenta, to last. Even without all her reminders to be included, it would be impossible for me to stand where I am today, realized as a person, without her by my side with her unwavering emotional support.

Finally, I would also to extend my thanks to both the Chemical Biological Science group of the Biosystems and Integrative Sciences Institute and the Department of Chemistry and Biochemistry of the Faculdade Ciências da Universidade de Lisboa. I also acknowledge the financial support of Fundação para a Ciência e Tecnologia through the grant SFRH/BD/140886/2018.

Preface

During my PhD in the Molecular Modelling & Simulation group, I had the privilege to be under the supervision of Doctor Miguel Machuqueiro and Doctor Diogo Vila Viçosa. Under their guidance, I had the opportunity to grow as a scientist and develop the necessary tools to study complex biochemical systems at the molecular level. Their combined expertise on molecular dynamics, quantum mechanics, continuum electrostatics, enhanced sampling techniques, protein-membrane systems, constant-pH MD methods and statistics were fundamental in building this thesis and, more importantly, my current understanding on this thrilling field.

Concerning this thesis, the original project solely focused on the concept of pHLIP peptides interacting in the water-membrane interface, as a continuation of our collaborative work with Dr. Oleg Andreev, Dr. Yana Reshetnyak (Rhode Island, USA) and Dr. Donald Engelman (Yale, USA). However, it quickly evolved into a more concerted and interdisciplinary body of work that encompasses the tumor microenvironment, chemotherapeutics and the challenges of overcoming tumoral multidrug resistance through conjugated therapies. As every proper project, we first established solid groundwork on the limits and advantages of a CpHMD-L protocol coupled with a pHRE scheme to study membrane-interacting biomolecules (Chapter 3). Our challenge was to obtain faster configurational sampling with improved accuracy of our residue pK_a predictions of a key aspartate modulator:

Silva, T. F.D., Vila-Viçosa, D. and Machuqueiro,M. (2021) "Improved Protocol to Tackle the pH Effects on Membrane-Inserting Peptides." J. Chem. Theory Comput., 17 (7), 3830-3840. (`https://doi.org/10.1021/acs.jctc.1c00020`)

The complexity of this endeavor required multiple testing, development of an analytical toolkit for transmembrane systems and several months of work yet, in the end, we obtained a robust, reliable and standard protocol for all our future studies. Concomitantly, we began a collaboration with Professor Yehuda Assaraf (Technion, Israel) on the lysosomal entrapment, a simple yet effective mechanism of tumor resistance against chemotherapeutics, namely Lewis base drugs (Chapter 6). In this work, I had the opportunity to develop a solid background on structural optimization using quantum mechanics calculations and charge set curation necessary for molecular parametrization within the CpHMD-L framework.

Stark, M., Silva, T.F.D., Levin, G., Machuqueiro, M., Assaraf, Y.G. (2020) "The Lysosomotropic Activity of Hydrophobic Weak Base Drugs is Mediated via Their Intercalation into the Lysosomal Membrane", Cells, 9, 1082 (`https://doi:10.3390/cells9051082`)

As a follow-up, I had the pleasure of co-supervising (unofficially) a MSc student, Pedro Suzano, in the application of a novel US-CpHMD method to decipher the molecular minutiae of the membrane crossing of these drugs and their pH-dependent membrane permeabilities (Addendum of Chapter 6). Not only this work allowed me to add another novel and state-of-the-art enhanced sampling methodology to my know-how, but it gave me the learning experience of how to guide another enthusiastic student in a new and challenging field, successfully completing a remarkable project. This work is being prepared for publication:

Suzano, P. M.,Silva, T.F.D., Machuqueiro,M. (2022) "A Computational Method to Calculate pH-Dependent Membrane Permeabilities for Anti-Tumour Drugs", In preparation

Finally, the final steps of my thesis focused on applying all the previous know-how to discern and detail the intramolecular interactions that comprise the pHLIP peptide and an outstanding clinical peptide variant (Var3). These chapters (Chapters 4-5) highlighted the pivotal roles of the key titrating residues and their surrounding vicinity on the peptide *in vivo* performance, as well as the importance of mimicking cellular environments to improve the descriptive and predictive ability of our *in silico* models. To the best of our knowledge, we performed the first study of an explicit CpHMD method with a pH gradient to model a transmembranar system. Moreover, we delved in the supporting role of arginine residues in the definition of the local vicinity of the key aspartate and the overall pHLIP thermodynamic stability. These two works are currently submitted and to be submitted for publications with a special focus on the rational design of these pHLIP peptides:

Silva, T. F.D., Vila-Viçosa, D. and Machuqueiro, M. (2022) "Increase the Realism of *In Silico* pHLIP Peptide Models with a Novel pH Gradient CpHMD Method", J. Chem. Theory Comput., Submitted

Silva, T. F.D., and Machuqueiro, M. "The Supporting Role of a Cationic Residue in the Tumor-Targeting Performance of pHLIP Peptides", to be Submitted

Besides the work performed on my thesis, I had the opportunity to work, in a collaborative study with Prof. Chris Oostenbrink group (Vienna, AT), on the effects of using single atomistic long-range cutoffs with the GROMOS54A7 force field on newer versions of the GROMACS package. This is fundamental work on the validity of a force field parametrization in the context of deprecated features of GROMACS, a popular MD simulation package:

Silva, T. F. D., Vila-Viçosa, D., Reis, P. B. P. S., Victor, B. L., Diem, M., Oostenbrink, C., and Machuqueiro, M. (2018) "The Impact of Using Single Atomistic Long-Range Cutoff Schemes with the GROMOS 54A7 Force Field", J. Chem. Theory Comput., 14, 5823-5833 (`https://doi.org/10.1021/acs.jctc.8b00758`)

Additionally, our groups acquired expertise on molecular dynamics studies of membrane proteins prompted us in publishing a book chapter on reliable pK_a calculations in membrane environments using PypKa, a modern and fast online pK_a calculation tool (`www.pypka.org`):

Oliveira, N. F. B., Reis, P. B. P. S., Silva, T. F. D., Machuqueiro, M. (2021) "pK_a Calculations in Membrane Proteins from Molecular Dynamics Simulations". In: Moreira I.S., Machuqueiro M., Mourão J. (eds) Computational Design of Membrane Proteins. Methods in Molecular Biology, vol 2315, pp 185-195. Humana, New York, NY. (`https://doi.org/10.1007/978-1-0716-1468-6_11`)

I was also a part of a group of experts in distinct *in silico* approaches, such as homology modelling, molecular docking, virtual screening and molecular dynamics, that in collaboratively produced a comprehensive tutorial on a *in silico* characterization pipeline for SARS-COV-2:

Rosário-Ferreira, N., Baptista, S. J., Barreto, C. A. V., Rodrigues, F. E. P., Silva, T. F. D., Ferreira, S. G. F., Vitorino, J. N. M., Melo, R., Victor, B. L., Machuqueiro, M., Moreira, I. S. (2021) "*In Silico* End-to-End Protein-Ligand Interaction Characterization Pipeline: the Case of SARS-CoV-2", ACS Synth. Biol., 10, 11, 3209-3235 (`https://doi.org/10.1021/acssynbio.1c00368`)

More recently, we also participated in a cohort experimental and theoretical study of the molecular details of a bio-engineered DyP-type peroxidase and similar mutants on their increased activity at alkaline pH:

Borges, P. T., Silva, D., Silva, T. F. D., Brissos, V., Cañellas, M., Lucas, M. F., Masgrau, L., Melo, E. P., Machuqueiro, M., Frazão, C., Martins, L. O. (2022) "Unveiling Molecular Details behind Improved

Activity at Neutral to Alkaline pH of an Engineered DyP-type Peroxidase", Comput. Struct. Biotechnol. J., 20, 3899-3910 (`https://doi.org/10.1016/j.csbj.2022.07.032`)

Abstract

The physiology of tumors is tied to MDR mechanisms that hamper chemotherapeutic effects, particularly passive membrane crossing compounds, like hydrophobic Lewis base drugs. Although the lysosomal entrapment phenomena remains to be fully understood, this pH-dependent MDR mechanism induces drug sequestration in the acidic lysosomal lumen. Overcoming the MDR requires multi-pronged therapies, which often overlook an ubiquitous tumor trait: the extracellular acidity of the tumor microenvironment (TME). To address this, pHLIP peptides have emerged as an acidity-selective technology for tumor-targeting drug delivery. We focused on refining our protocols with enhanced sampling techniques and tumor-like features to improve the predictive abilities of the CpHMD-L methodology and augment the realism of these biomolecular models, thus bridging the gap to *in vivo* and cellular conditions. The optimized protocol coupled the CpHMD-L method with a pHRE scheme, providing a robust baseline. Then, we applied the protocol to study the diverging therapeutic efficiency of the *wt* and an over-performing Var3 peptide. A novel implementation of a pH gradient CpHMD-L method successfully reproduced experimental performances, thus elucidating pivotal residues electrostatic networks that dictate peptides thermodynamic stability in TME conditions. A multi-peptide study highlighted the remarkable effects of permuting arginines in modulating the local vicinity of key aspartates. These findings heavily correlate with their tumor-targeting performance, supporting more rational and *in silico*-based approaches to peptide design. Finally, the pH-dependent mechanism of lysosomal entrapment was modelled, hinting at the important role of acidity in Lewis base drugs membrane intercalation. Additional pH-dependent permeability calculations, using a novel US-CpHMD method, identified the TME acidity as an additional MDR defense mechanism that impairs clinical efficiency. It also revealed an intrinsic flaw of these compounds, since they preferably target healthy cells. These findings have important implications in rational drug design, especially of conjugated therapies with pHLIP-like drug delivery systems to overcome these challenges.

Keywords: CpHMD, enhanced sampling, pHLIP, Lewis Base drugs, TME

Resumo

A nível mundial, o cancro é responsável por mais de 10 milhões de mortes por ano. Apesar da existência de várias tipologias, todos os tumores sólidos são caracterizados pela resistência, inata ou adquirida, a diversos fármacos (MDR). Esta propriedade tumoral consiste em múltiplos mecanismos de defesa frequentemente associados a vias metabólicas alteradas; degradação estrutural da matriz extracelular; alterações de propriedades bioquímicas, como a acidez extracelular do microambiente tumoral; e sobre-expressão proteica e de organelos, nomeadamente as proteínas bombas de protões e os lisossomas, respectivamente. Neste último caso, uma elevada concentração de lisossomas exacerba o sequestro lisossomal, onde fármacos anti-tumorais, retidos no lúmen acídico, são degradados ou removidos da célula-alvo. Ainda que este mecanismo molecular não esteja completamente elucidado, este fenómeno dependente do pH é um dos responsáveis pela diminuição da eficiência quimioterapêutica, em especial nos fármacos hidrófobos com bases de Lewis. Avanços tecnólogicos recentes procuram evadir a MDR através de terapias multi-conjugadas focadas em características comuns dos tumores. Ainda assim, a acidez tumoral constitui uma propriedade pouco explorada a este nível. Neste âmbito, o péptido pHLIP emergiu como uma tecnologia específica para a acidez celular, funcionando como um biomarcador tumoral ou um sistema de transporte farmacológico a sítios activos. O péptido *wt*, constítuido por 36 resíduos, adopta uma conformação α-hélice quebrada, a pH's acídicos (< 6.0), inserido ao longo da bicamada lípidica. A sua estabilidade termodinâmica depende fundamentalmente de eventos de (des)protonação de um resíduo-chave titulável (Asp-14) que ocupa preferencialmente a região da interface membrana-água. Estudos anteriores de dinâmica molecular a pH constante (CpHMD) caracterizaram as propriedades moleculares e termodinâmicas deste sistema, realçando a capacidade predictiva do método em reproduzir valores experimentais relativos à potencialidade terapêutica destes péptidos.

Desde o início, esta tese focou-se em melhorar a capacidade descriptiva e predictiva do método de CpHMD-L para estudar sistemas transmembranares complexos. O desenvolvimento e a optimização do nosso protocolo, com técnicas de amostragem aumentada e a introdução de propriedades tipicamente celulares, melhoraram a amostragem configuracional destes sistemas, aproximando os nossos modelos biomoleculares dos modelos celulares e *in vivo*. Numa primeira abordagem, desenvolvemos um protocolo robusto e consistente no estudo de péptidos transmembranares, que acopla um esquema de troca de réplicas com o nosso método de CpHMD-L denominado "pH replica exchange" (pHRE). Este trabalho foi crucial em estabelecer uma plataforma confiável para o desenvolvimento dos restantes projectos reflectidos nesta tese, ao explorar os limites da nossa metodologia. O desenvolvimento deste protocolo focou-se num estudo sistemático do péptido *wt*-pHLIP inserido em bicamadas lípidicas com números de lípidos distintos. Ao avaliar sistemas com tamanho reduzido, identificaram-se problemas sistemáticos de limite de caixa, induzindo deformações membranares que não só afectam significativamente as estimativas de pK_a, tal como aumentam a heterogeneidade conformacional entre replicados. Este efeito deriva da dificuldade em obter amostragem equilibrqada das interacções do Asp-14 com grupos fosfato dos lípidos. Não obstante, o protocolo estabelecido oferece previsões excelentemente corroboradas pelos dados experimentais, o que reforçou a confiança no nosso protocolo para múltiplas aplicações. Uma das primeiras aplicações deste protocolo foi na avaliação da capacidade terapêutica de uma variante (Var3) do pHLIP-*wt*. Esta variante do péptido pHLIP apresenta uma performance clínica dicotómica, visto a Var3 demonstrar experimentalmente performances aquém do péptido *wt* e outras variantes, contudo apresentando melhorias significativas em ambiente *in vivo*. Este resultado contrasta com uma pior performance do péptido *wt*. Neste trabalho, procedemos à implementação inovadora de um gradiente de pH na nossa metodologia de CpHMD-L acoplada a um esquema de pHRE, onde foi possível fixar o pH interno (pH 7.2) e permitir a variação do pH externo numa gama de valores entre 4.0 e 7.0. Através desta metodologia, reproduziu-se com sucesso não só as propriedades experimentais em lipossomas, tal como se quantificou a variação de performance dos péptidos *wt* e Var3, em ambiente tumoral, através de medições de pK_a ao longo da normal da membrana. Adicionalmente, descreveu-se detalhadamente a rede electrostática envolvente dos resíduos-chave (Asp-13 - Var3 ; Asp-14 - *wt*) que governam o equílibrio termodinâmico dos péptidos no ambiente acídico do TME. Este trabalho salientou a importância de incorporar propriedades celulares nos nossos modelos membranares, a fim de obter previsões mais precisas e mimetizar mais realisticamente o microambiente pHLIP-membrana. Na sequência da caracterização destes sistemas, observou-se o efeito de uma arginina próxima dos resíduos-chave que aparentava modular a variação do $\mathrm{p}K_\mathrm{a}$ nas regiões mais profundas da membrana. Deste modo, colocou-se a hipótese do resíduo de arginina ter um efeito fundamental na definição da electrostática local do resíduo aspártico chave. Um estudo multi-peptídico subsequente permitiu elucidar o papel vital da permutação da arginina na modulação da primeira esfera de coordenação do resíduo-chave. Neste trabalho, caracterizámos a forma como o microambiente electrostático do resíduo-chave respondia a diferentes posições da arginina (Arg-10, Arg-14, Arg-15 e Arg-17). Através de simulações de pHRE de diferentes sequências do péptido *wt*, estudámos os perfis de protonação e as interacções inter-(águas e lípidos) e intra-moleculares (Arg-X) de cada aspártico-chave de todas as variantes. Os nossos resultados revelaram o impacto crucial da arginina sobre o comportamento do resíduo aspártico e, por consequência, em promover ou diminuir a estabilidade do péptido de forma directa, via estabilização da forma ionizada do resíduo aniónico através de ligações salinas persistentes, ou de forma indirecta, ao modular a acessibilidade de grupos moleculares vizinhos. Os dados deste estudo correlacionam com os dados experimentais dos nossos colaboradores, ilustrando o papel simbiótico de metodologias experimentais e *in silico* no desenho racional de péptidos. Seguindo uma matriz de estudo sistemática, utilizámos o nosso protocolo pHRE para modelar o mecanismo de sequestro lisossomal dependente do pH em diversos fármacos com bases de Lewis diferenciados por $\mathrm{p}K_\mathrm{a}$s e perfis de polaridade distintos. A caracterização destes fármacos realçou a preponderância da acidez na intercalação destes compostos na membrana celular e na sua eficiência clínica. Por sua vez, a combinação de metodologias *in silico* e *in situ* foi fundamental nesta caracterização e na identificação do mecanismo de desprotonação transiente necessário para a internalização destes fármacos na membrana. Por fim, a implementação de uma metodologia de estado da arte de amostragem aumentada "umbrella sampling-CpHMD" (US-CpHMD) possibilitou a caracterização dos perfis de energia de dois compostos-modelo (Nintedanib e Sunitinib) e das suas permeabilidades dependentes do pH do ambiente celular (lisossomal, tumoral ou normal). Este estudo permitiu identificar a acidez do TME como um mecanismo adicional de defesa contra fármacos anti-tumorais, reduzindo significativamente a permeabilização do Sunitinib na membrana, por oposição ao Nintedanib cuja permeabilidade foi menos afectada. A diferença comportamental entre os compostos reside nas suas polaridade íntrinsicas, sendo o Nintedanib mais hidrófobo, e no $\mathrm{p}K_\mathrm{a}$ do grupo titulável que. sendo mais baixo. facilita a desprotonação do Nintedanib e a consequente perda da carga positiva, tornando o composto mais suave para translocação membranar passiva. Acima de tudo, esta metodologia inovadora permitiu identificar uma falha crítica no desenho racional destes fármacos com bases de Lewis. Não só a sua permeabilização tumoral é bastante reduzida, como a internalização por parte de células saudáveis é consideralvemente superior. Esta inversão do perfil terapêutico pode ser a causa subjacente dos efeitos secundários resultantes da administração de doses elevadas associadas ao tratamento contra o cancro. Estes resultados têm implicações extraordinárias no desenho racional de fármacos anti-tumorais. Como solução, estamos presentemente a testar uma modificação química da base de Lewis por um ácido carboxílico com o intuito de inverter os perfis de permeabilização dependentes do pH, melhorando assim as propriedades clínicas destas terapias.

Keywords: CpHMD, técnicas amostragem aumentada, pHLIP, Fármacos com bases de Lewis, microambiente tumoral

Contents

List of Figures

List of Tables

List of Abbreviations

Å Angstrom
ABC ATP-binding Cassette
AC Acid Ceramidase
AMP Antimicrobial Peptide
ASM Acid Sphingomyelinase
ATB Automated Topology Builder
ATP Adenosine Triphosphate
BCRP Breast Cancer Resistance Protein
BMP Bis(monoacylglycero)phosphate
CE Continuum Electrostatics
CHQ Chloroquine
Clomp Clomipramine
CNSD Central Nervous System Acting Drugs
COG Center of Geometry
CpHMD Constant pH Molecular Dynamics
CV Collective Variable
DDR1 Dimeric Receptor Tyrosine Kinase
DDS Drug Delivery System
DMPC 1–2-dimyristoyl–*sn*–glycero–3–phosphocholine
DNA Deoxyribonucleic Acid
DNR Daunorubicin
DPPC 1–2-dipalmitoyl–*sn*–glycero–3–phosphocholine
ECM Extracellular Matrix
Ethop Ethopropazine
EV Extracellular Vesicles
FF Force Field
FRAP Fluorescence Recovery after Photobleaching
GPA Glycophorin A
GRF Generalized Reaction Field
HEWL Hen Egg-white Lysozyme
HSD Homogeneous Solubility-Diffusion
ILV Intra-lysosomal Vesicle
ISDM Inhomogeneous Solubility Diffusion Model
ITC Isothermal Titration Calorimetry
l–BFGS Limited Memory Broyden–Fletcher–Goldfarb–Shanno
LD Lysosomotropic Drug
LLM Lysosomal Limiting Membrane

LMP Lysosomal Membrane Permeabilization
LSD Lysosomal Storage Disease
LTG Lysotracker Green
LTR Lysotracker Red
LUV Large Unilamellar Vesicles
MC Monte Carlo
MD Molecular Dynamics
MDR Multidrug resistance
MM Molecular Mechanics
MRP1 Multidrug Resistance–Associated Protein 1
NTD Nintedanib
P–gp P–glycoprotein
PB Poisson–Boltzmann
PBC Periodic Boundary Conditions
PC Phosphocholine
PDB Protein Data Bank
PEF Potential Energy Function
pHLIP pH Low Insertion Peptide
pHRE pH Replica-Exchange
PI Phosphatidylinositol
Pimo Pimozide
PLA Phospholipase A
PLC Phospholipase C
PME Particle Mesh Ewald
PMF Potential Mean Force
PNA Peptide Nucleic Acid
POPC 1–palmitoyl–2–oleoyl–*sn*–glycero–3–phosphocholine
PS Phosphatidylserine
QM Quantum Mechanics
RE Replica-Exchange
REMD Replica-Exchange Molecular Dynamics
RF Reaction Field
SIR Siramesine
SUN Sunitinib
TKI Tyrosine Kinase Inhibitor
TM Transmembrane
US Umbrella Sampling
US–CpHMD Umbrella Sampling CpHMD
VCR Vincristine
WHAM Weighted Histogram Analysis Method

Chapter 1

Introduction

In this chapter, we will summarily describe the challenges of a specific class of chemotherapeutics. There is vast literature on different and novel clinical therapies, yet chemotherapy remains as one of the more widely used. Still, critical limitations are found in these therapies and we will address two main approaches: chemically modifying a molecular compound or using specialized and targeted drug delivery systems. In the latter, we will focus on a particular set of transmembrane peptide that has been emerging in the last 10 years as a promising and novel therapeutic tool for both diagnosis and treatment. The pivotal water-membrane interface behaviour remains to be fully understood and every step taken to understand its intrinsic mechanism will achieve more fine and efficient peptide rational design.

1.1 Hallmarks of tumor physiology

1.1.1 Cancer as a global health issue

The advent of medical and biological sciences led to novel therapeutics and diagnostic procedures, and an increase in the overall life expectancy. However, even with these technological developments, humanity still faces a diverse array of health issues and challenges, such as neurodegenerative diseases, viral pandemics and different types of cancer. More specifically, cancer is one of the leading death causes worldwide, with approximately 10 million deaths per year [1]. As per the definition of the North American National Health Institute (NIH), cancer is defined as a "...disease in which some of the body's cells grow uncontrollably and spread to other parts of the body...". Tumor cells possess several traits, such as proliferative signaling, cell death resistance, immune response evasion, tumor-promoting inflammation [2, 3]; however, the provided short definition encapsulates the three defining traits of tumorigenesis: non-specificity, metastasis and aggressiveness (Figure 1.1).

Typically, most cases of cancer manifest themselves as solid tumors that can grow rapidly in size and in any tissue of the body, resulting in several types and variants of the disease. As a result, there is a large ecosystem of tumor cell environments that require distinct clinical approaches, thus hindering the design of broad scale medical treatments. Nevertheless, there are clearly defined structural and physiological differences to normal cells that can be specifically targeted in a clinical context.

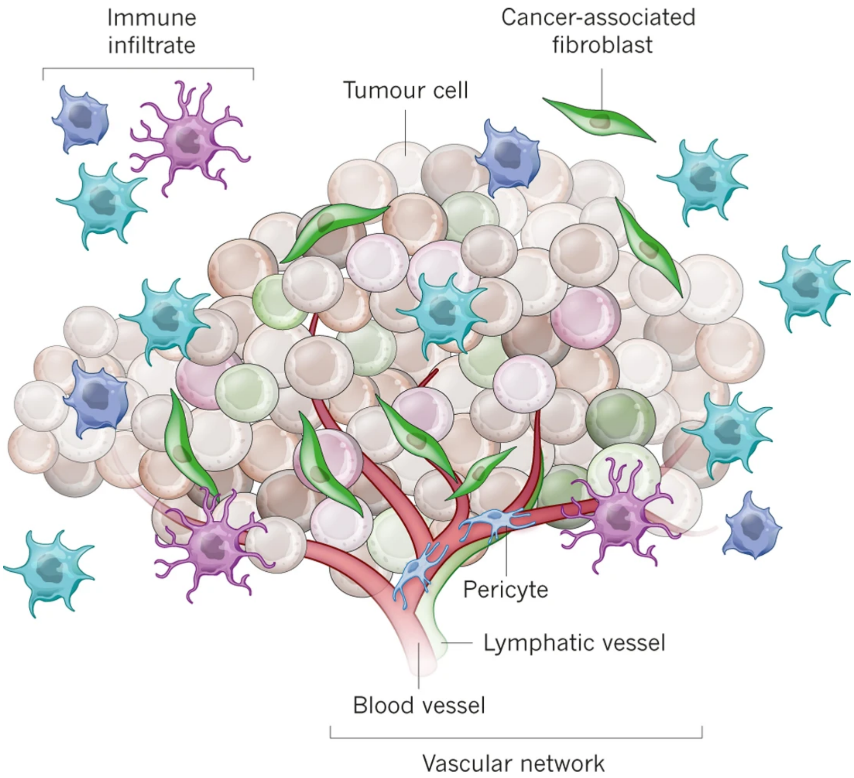


Figure 1.1: Graphical representation of a tumor cell with augmented vascularization of blood and lymphatic vessels. The contrived vascular network hampers a healthy oxygen and nutrient distribution, generating local hypoxia regions and acidity. More aggressive phenotypes exhibit higher acidity and reduced O_2 concentration. Adapted from reference [4].

1.1.2 The Tumor Microenvironment

The tumor microenvironment (TME) has been a well-known factor in cancer development for several decades. The works of Virchow (1863), Paget (1889) and Warburg (1924) spearheaded several TME characterization hypotheses, through patient examinations that highlighted specific biomarkers around tumor lesions, such as inflammation, leukocyte infiltration, hypoxia and enhanced metabolic pathways. In short, the TME is the interstitial region between the solid tumor cell and the surrounding tissue, as shown in Figure 1.2, and it comprises of cellular (fibroblasts, adipose cells, immune-inflammatory cells, organelles and vesicles) and non-cellular components (protein markers, cell factors, the extracellular matrix, external and internal tumor pH, metabolites, etc.) [6, 7]. Furthermore, the TME contains a highly vascularized, yet contrived, network of lymphatic and blood vessels to supply with oxygen, nutrients, while also removing cellular waste [8]. This disorganized tumor vasculature induces a characteristic rapid growth rate, local invasion and metastatic dissemination to other parts of the body [9]. The erratic blood network and hypoxic regions around the tumor cells piqued the interest of Otto Warburg. This led to the discovery of an important biochemical phenomenon called the Warburg effect [10–13]. In tumor cells, the rapid proliferation requires large amounts of energy, hence the glucose uptake dramatically increasing, in the aerobic glycolysis metabolic pathway, concomitant with an increase of fatty acid oxidation to generate ATP [14, 15]. Consequently, an higher proton concentration manifests as a byproduct of the metabolic process. As several biological processes occur within a limited pH range, the tumor cell over-expresses proton pump exporters to preserve the physiological pH in the tumor cytosol. The higher proton efflux to the TME creates an external acidic environment, which coupled with the tumor hypoxia and contrived vascularization, prevents extracellular pH regulation, resulting in a prominent pH gradient between the TME and the tumor cytosol. The pH of the TME ranges from 6.2 to 6.8, [16] intimately correlated with tumor aggressiveness, as more acidic tumors are usually found in more aggressive phenotypes [17]. The TME acidity activates the extracellular matrix proteolytic enzymes that decay the surrounding tissue promoting local invasion [18]. On the other hand, the tumor cytosol is slightly more alkaline compared to a normal cell, as the excess of proton pumpers causes a shift from the physiological pH, normally preserved by homeostasis processes [19, 20]. The occurrence of these phenomena highlights the typical cellular conditions of an important tumor hallmark - the tumor microenvironment acidosis and the role of pH [21, 22].

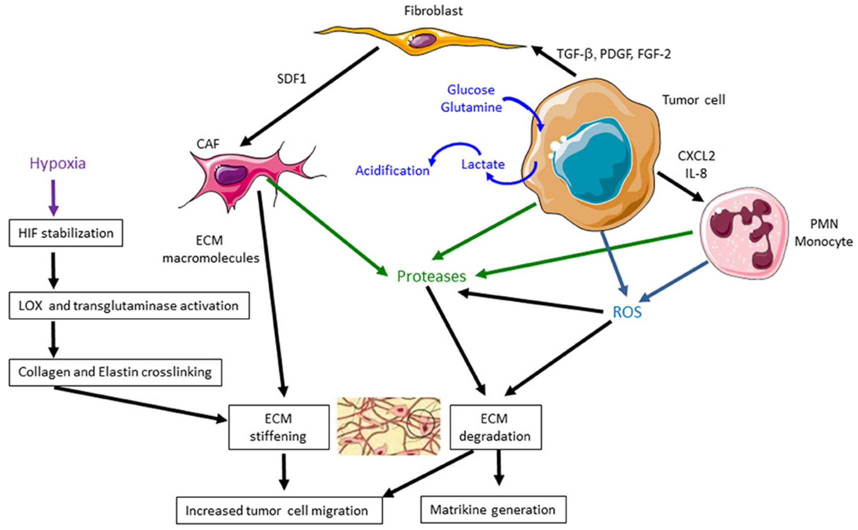


Figure 1.2: Graphical representation of the tumor microenvironment (TME) in a cellular environment. The deregulation of metabolic pathways, augmented vascularization, imbalance of growth factors and extracellular acidity causes the degradation of the extracellular matrix (ECM). The loss of structured ECM provides a fertile foundation for tumor proliferation. Adapted from reference [5].

1.1.3 The biological role of pH

The pH is a fundamental physico-chemical property that acts as a regulator of both biological and chemical processes. It is a measurement of the negative logarithm of the activity of protons (H^+) in a given medium. Most biological reactions occur in a water solvent, which can be acidic ($pH \leq 7$), neutral ($pH \sim 7$) or basic ($pH \geq 7$):

$$\mathrm{pH} = -\log_{10}(\mathrm{a}_{\mathrm{H}^+}) \tag{1.1}$$

Biologically, the medium pH acts as a proton reservoir to acidic and basic group moieties of several biomolecules, thus modulating their activity and respective biological processes. In biological systems, the more noteworthy pH ranges are well preserved in homeostasis processes [23] to maintain the delicate electrostatic balances integral to diverse cell functions. Metabolic pathways, signal transduction, membrane potential, enzymatic catalysis and other molecular level biochemical processes are either regulated by a protonation equilibrium or require protonation transitions of aminoacid residues or chemical group moieties to occur [24]. These protonation transitions, dependent on the functional group $\mathrm{p}K_\mathrm{a}$ and the pH of the medium, modify the electrostatic network, thus affecting the biological function, structure, kinetics and thermodynamic equilibrium. Indeed, it is through this intimate relationship between pH and $\mathrm{p}K_\mathrm{a}$ that biological phenomena sprout.

By its chemical definition, the $\mathrm{p}K_\mathrm{a}$ reflects the strength of an acidic or basic group, hence their tendency to donate or accept a proton. The strength of an acid depends on its acid-dissociation constant:

$$K_a = \left(\frac{[A^-]}{[HA]} \right) \tag{1.2}$$

which then, the $\mathrm{p}K_\mathrm{a}$ can be expressed as the negative logarithm of that same property:

$$\mathrm{p}K_\mathrm{a} = -\log_{10}(K_a) \tag{1.3}$$

where HA and A^- are the protonated acid and its deprotonated conjugated base, respectively. Simply put, the $\mathrm{p}K_\mathrm{a}$ is a measurement of the proton binding affinity of a given group, hence lower $\mathrm{p}K_\mathrm{a}$ values are equivalent to stronger acids and higher values reflect stronger bases. With the pH and $\mathrm{p}K_\mathrm{a}$ definitions, the relationship between both can be expressed through the Henderson-Hasselbalch equation:

$$\mathrm{pH} = \mathrm{p}K_\mathrm{a} - \log_{10}\left(\frac{[HA]}{[A^-]} \right) \tag{1.4}$$

which states that the $\mathrm{p}K_\mathrm{a}$ of an acid is equal to the solution pH when the concentrations of the protonated and deprotonated states of the acid are equal. As previously stated, the relationship between pH and $\mathrm{p}K_\mathrm{a}$ is paramount for biological functions, particularly in the context of proteins and peptides.

Proteins and peptides often are the centerpiece of most cellular processes, due to their plurality of functions and multitude of possible structures and sizes. These biomolecules can operate as catalytic enzymes, promoters of signal transduction and molecular cargo transporters through membrane bilayers, while also being responsible for the cytoskeleton structure of all cells. Although, the aminoacid residues are the building blocks of proteins and peptides, a subgroup of them possess side-chains with ionizable groups, prompting them to play a crucial part in the aforementioned functions. Each of these aminoacids have an individual $\mathrm{p}K_\mathrm{a}$ value (Figure 1.3), hence adopting unique protonation behaviour, at a given solvent pH value, thus operating in different roles. Since the

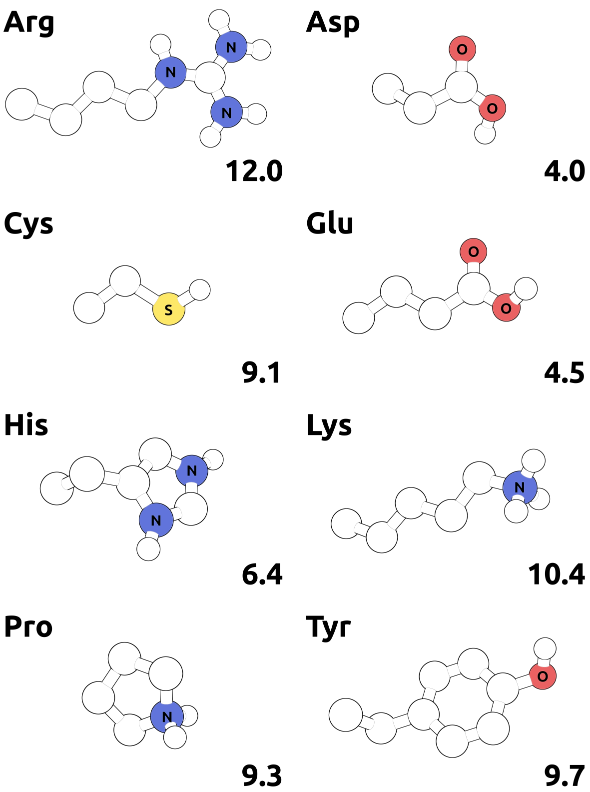


Figure 1.3: Graphical representation of the titrating aminoacid side-chains with the respective pK_a values. The electronegative atoms are highlighted in color (oxygen–red,nitrogen–blue,sulfur–yellow) with their respective atomic letter. Adapted from reference [25]

intra- and extracellular environments require a delicate homeostatic balance, most of the aminoacids are either weak acids or weak bases. Therefore, these titrating residues are often flexible in establishing several types of molecular interactions in a diverse ecosystem of microenvironments. The residue surroundings are composed of their own protein/peptide structure and other neighbouring biomolecules, which can form protein-protein interactions, protein-DNA, protein-RNA, protein-lipid, protein-solvent or protein-ligand. Each interacting group contributes in a certain way to modify the electrostatic environment sensed by the chemical group and, consequently, modulating the group pK_a [26]. The most common molecular interactions that create deviations in the pK_a are charge-charge interactions, charge-dipole and the desolvation effect. The first type occurs between ionized groups/molecules, either with similar or opposing charges, leading to repelling or attractive forces, respectively. When at least two charged groups are closely interacting, their pK_a's will significantly deviate from the respective solution pK_a. The charged-dipole interactions occur between a fully-charged group and a polar one. These type of interactions are commonly found in hydrogen bonds and cation/anion–π interactions and, although not as strongly as charge-charge, they still induce relevant pK_a deviations when present. Finally, the desolvation effect takes place when the ionized state of a titrable residue, such as an aspartic acid, moves from an aqueous polar environment to a more apolar region, like the membrane bilayer core or an apolar protein pocket. As it progresses deeper into the apolar environment, the residue sheds the charge-stabilizing water molecules, thus energetically favoring their neutral form, which is reflected on a pK_a shift [27]. Naturally, a balance between these three effects typically describes the deviations in pK_a values of a given group. And, in turn, it is the combination of a diverse range of molecular interactions and cellular conditions that modulate the biological function.

1.2 Challenges of pH-dependent multidrug resistance

1.2.1 Lewis bases as chemotherapeutics

The diverse and complex nature of the TME hinders the development of more broad therapies for cancer treatment. Yet, all cancer phenotypes share a common physiological trait, the TME acidosis, sometimes with different degrees of acidity. Therefore, high impact clinical therapies have been developed using the ubiquitous TME acidosis as a guide. There is a diverse array of antitumor therapies, including chemotherapy, radiation therapy, surgery and others. Nevertheless, the scope of this thesis is on pH-dependent technologies in tumor environments.

Chemotherapy is one of the most popular anti-tumor therapies worldwide, specially for metastatic tumors, as there are beneficial effects from the systemic drug administration. According to their mechanism of action, different drug classes comprise of alkylating agents to target DNA (cisplatin), molecules that inhibit DNA replication or DNA unravel (inhibitors of topoisomerases I or II) or inhibitors of key enzymes in signaling pathways (tyrosine kinase II) [28, 29]. To obtain optimal therapeutic effect, the drug compound should have high tumor site concentrations. So, many drug design studies perform pharmacokinetics and pharmacodynamics screenings to assess which drugs exhibit higher bioavaliability. Successful screening studies have identified molecules with high hydrophobicity and high pK_a values as some of the best candidates. These molecules named Lewis base drugs are one of the more widely used type of chemotherapeutics.

The term Lewis base derives from the concept of proton-donor bonds, coined by Gilbert Newton Lewis, that identifies Lewis bases as reactive species that donate free electron-pairs and Lewis acids as electron-pair acceptors [30]. These hydrophobic weak bases contains donor atoms such as nitrogen, oxygen, phosphor and sulfur, and typically, Lewis base drugs contain one or more Lewis base groups with pK_a's ranging from 7 to 9. These characteristics entail that the titrable Lewis group acts as a proton donor, which is particularly relevant in solution, as the high pK_a of these hydrophobic molecules facilitates the group protonation (Figure 1.4). The protonated group (positive charge) increases the molecule's polarity which, in turn, improves the molecule's solubility, bioavailability and pharmacokinetics.

Lewis base drugs have been used in the treatment of several type of tumors for several years. Among many, Sunitinib and Nintedanib have been identified as promising drugs to target the tyrosine kinase domains in several protein complexes, specifically for gastrointestinal stromal tumors, renal cell carcinoma [31–35] or non-small cell lung cancer, respectively [29, 36]. Their cell-internalization mechanism hinges on the titration of the Lewis base to promote membrane passive diffusion [22]. The lipid bilayer is a very apolar medium that acts as a filtering barrier for the internal and external cellular environments. Molecule crossing either occurs by active transport, that requires complex protein transporters often counteracting a concentration gradient, or by passive diffusion, either through a protein channel or through passive membrane bilayer crossing. The later process is often limited to small apolar molecules, such as sunitinib and nintedanib (Figure 1.5), as long as they remain in their neutral (deprotonated) state. Although, the physiological pH favours their protonated state, these weak bases transiently deprotonate when interacting with the membrane lipid headgroups (see Chapter 6 for more details). These electrostatic interactions decrease the drugs' pK_a values, prompting the weak bases to lose their charge, and changing to the energetically favourable neutral state that facilitates membrane insertion. In sum, the intrinsic hydrophobicity and neutrality confers a low energy barrier to cross the apolar membrane core, hence improves the compound membrane permeability. In this process, upon reaching the inner membrane monolayer, the exposure to the solvent favours their charged state, since lipid headgroups interactions shift the pK_a to stabilize the protonated state, smoothing the drug release into the tumor cytosol [37].

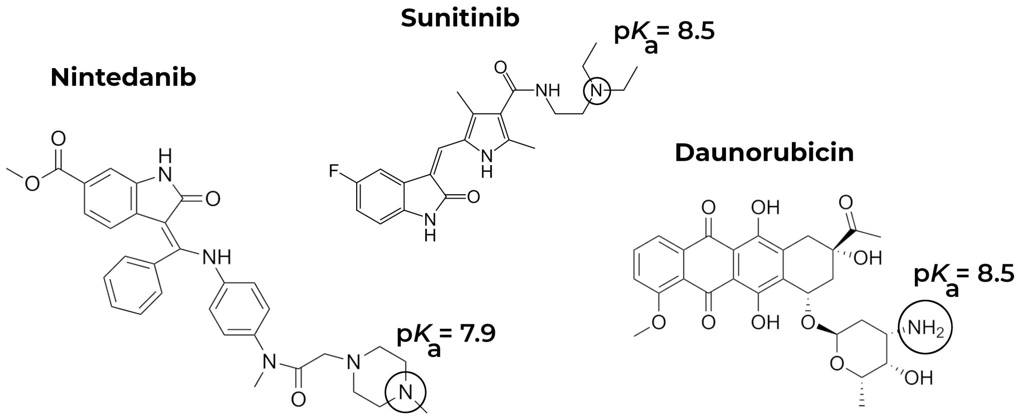


Figure 1.4: Chemical structure schemes of popular Lewis base drugs. The Lewis bases moieties and their respective solution pK_as are highlighted.

1.2.2 Multidrug resistance in cancer treatment

The development of highly efficient tumor treatments prioritizes targeting solely tumor cells without killing healthy normal cells. However, although a molecule may be an excellent inhibitor of a key enzyme, it often lacks tumor specificity, hindering its chemotherapeutic efficacy, thus killing tumor and healthy cells alike. Furthermore, there are patients not responding to treatment due to (1) poor molecule absorption; (2) increased drug excretion; or (3) inability of drug delivery. These issues usually lead to higher drug doses that cause severe side-effects. These occurrences are tied to both acquired and innate resistance to chemotherapeutics, which, when generalized, leads to another hurdle of cancer treatment named multidrug resistance (MDR) [22, 38, 39]. Multidrug resistance is the acquired or innate ability of cancer cells to resist to a large range of unrelated anticancer drugs, through a complex entanglement of diverse structural and molecular mechanisms of action [38]. These mechanisms can range from chemical modifications to drug targets, metabolic pathway changes or structural/physiological modifications to the tumor microenvironment (TME) [22, 40] (Figure 1.6).

Although the gene expression of each cancer genome differs, conferring distinct types resistance to each tumor, there are universal mechanisms of action. One of the more well-known MDR mechanisms is the over-expression of a class of ATP-dependent transporters known as the ATP-binding cassette (ABC) family. The over expression of ABC transporter proteins, like the P-glycoprotein (P-gp) [41], dramatically increases the efflux of drug molecules from the tumor cytosol to the TME. The reduced concentration, at the site of interest, hinders the inhibitor effect, hence impairing the therapeutic efficacy [38, 42, 43]. An obvious solution to this issue would be developing and administrating P-gp inhibitors to impair the drug excretion and allow the therapeutical drug to target the key site. However, promising P-gp inhibitors are lacking in pharmacokinetic properties as they are highly toxic, even in small doses, or non-specific, thus targeting other important enzymes, like the cytochrome P450 [44].

As previously mentioned, multidrug resistance is a complex network of diverse mechanisms working in tandem. Recent studies have highlighted other obstacles in chemotherapy, particularly relevant in drug clearance and poor

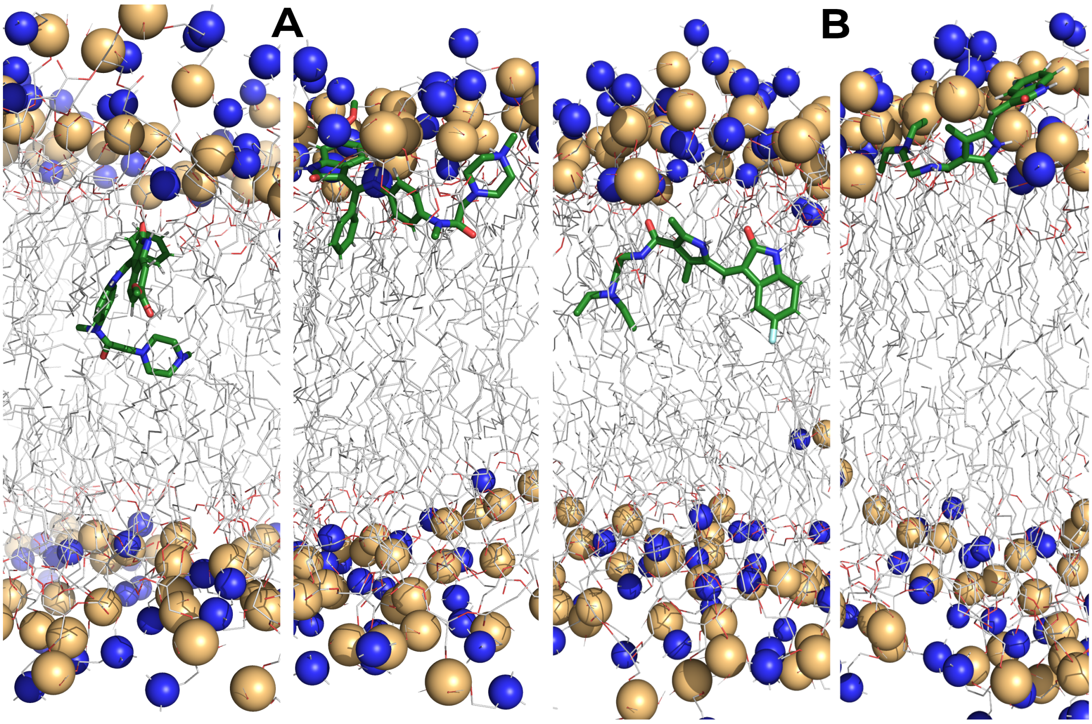


Figure 1.5: Graphical representation of the inserted (left) and water-membrane interface (right) of Nintedanib (A) and Sunitinib (B). The passive membrane diffusion requires the deprotonation of the Lewis base. Adapted from ref. [37].

drug import, which are the extracellular vesicle-based drug sequestration and the acidity of the TME [22, 37, 39, 45, 46]. The extracellular vesicles (EVs) are nanosized particles that disseminate to the interstitial region of tumoral tissue from the tumor cytosol. They are formed by lipid membrane bilayers rich in drug efflux pumps, such as P-gp, MRP1 and BCRP, with reversed orientation [47, 48]. Consequently, these EVs are capable of sequestrating drug molecules into their cargo, impeding them from reaching their target within the tumor cell. Afterwards, the EVs are released to the TME, through exocitosis, where their content is expelled, reducing the intracellular concentration of the drug and hindering the clinical effect. In addition, coupled to these tumor defense mechanisms, there is the over-expression of lysosomes in the tumor cytosol [22].

The lysosome is a very acidic organelle, with an internal pH ranging from 4.5 to 5.0, which generates an even steeper pH gradient against the cytosol relative to the TME gradient. The alkaline pH promotes the internalization of drug molecules, specifically the weak Lewis bases drugs, as they more easily deprotonate then passively diffuse to the lysosome(Figure 1.7). Afterwards, the weak bases are exposed to the acidic lysosomal environment, therefore gaining a positive charge and becoming entrapped within the acidic lumen of the lysosomal compartment [49, 50]. Finally, the lysosome is excreted from the cell, releasing its cargo in the extracellular environment, reducing the drug concentration at the target site.

1.2.3 Targeting the MDR

The challenge of multidrug resistance is crucial in defining successful anticancer strategies. To overcome MDR, there are several promising avenues, though this thesis will only focus on two strategies that tackle the impairment of drug delivery to the site of action. As previously mentioned, the delivery impairment can result from different factors, either from the pharmacokinetics (non-specificity) or drug extrusion to the extracellular environment. One possible solution is the chemical modification of functional group moieties of existing drugs. By identifying the functional groups that hinder targeted cellular intake, either from active or passive diffusion, it

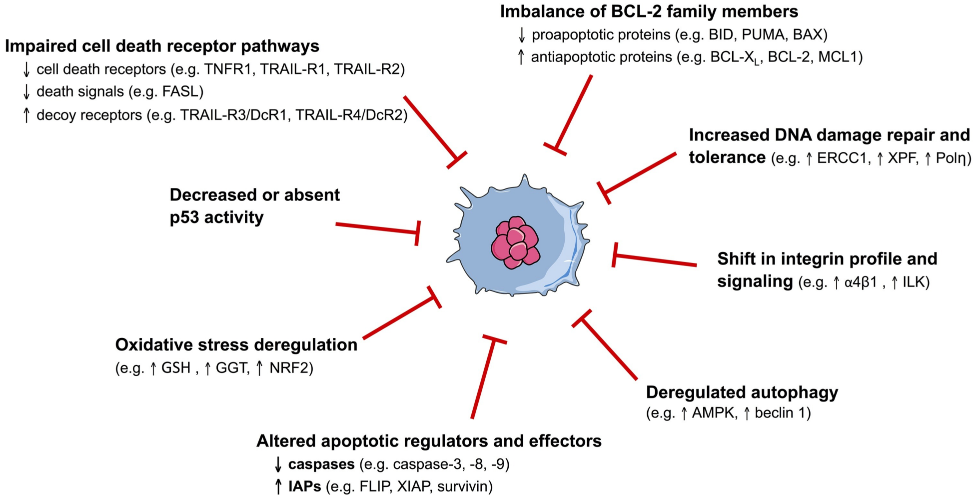


Figure 1.6: Multiple defense mechanisms against cell death induced by chemotherapeutics. Both innate and acquired multi drug resistance (MDR) results from the deregulation of several signaling pathways and environmental conditions. Adapted from reference [22].

is possible to recover clinical efficacy without increasing the dosage (see Chapter 6). Another possible route is the design of target-specific drug delivery systems. Drug delivery systems (DDS) are biomolecular constructs that selectively carry a therapeutic drug to its specific target. Currently, several systems have been developed, and among them there are liposomes, polymers, dendrimers and nanoparticles of distinct materials (ex: gold and hydrogel), with this pivotal targeting ability revitalizing previously discarded drug candidates. These technologies are compatible with molecules that were deemed unspecific and highly toxic, poorly soluble or phenomenal ligands with impaired pharmacokinetics, thus unable to cross vital organ barriers, as the blood-brain barrier. DDS can be classified according to their generation: the 1st generation consisted of nanomaterials constructed as basic liposomes with functionalized surfaces to enhance water solubility; the 2nd generation focused mainly on optimizing both the nanoparticle size and its surface chemistry to facilitate tumor permeation and retention; lastly, the 3rd generation makes use of cellular conditions, such as hypoxia [51], acidity [52] and extracellular matrix proteinases activity, as selective triggers that promote a given change within the DDS structure, prompting the drug release into the tumor [53]. In this last category, some of the more promising technologies can be found, with sustained or targeted-release systems possessing high selectivity which drastically reduces dosage side-effects, enhances therapeutic efficacy and opens more possibilities of property fine-tuning according to the type of cancer [54]. Indeed, not only these DDS can be used for treatment, but also for diagnostics, as these nanoparticles can be incorporated with biomarkers for tumor detection to detect early onset of the disease [55, 56].

Remarkably, the activation modulated drug delivery systems are outstanding in incorporating different intra- and extracellular conditions as unique triggers for targeted drug release. As previously mentioned, the pH range of the tumor microenvironment (TME) varies according to the tumor phenotype, the organ and, more importantly, from healthy tissues [57]. In the last decades, a peptide-based type of activation modulated drug delivery system has been emerging, particularly focusing on using transmembrane proteins and cell penetrating peptides as peptide-drug vectors [52, 58]. Their particular cell-binding/inserting features and potential for efficiency optimization opens a whole new path for pH-dependent DDS.

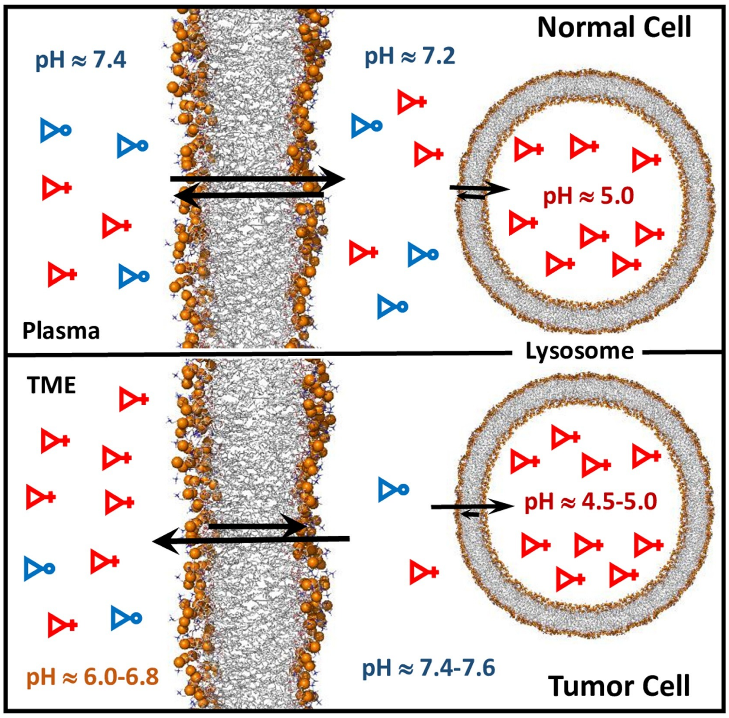


Figure 1.7: Representation of the pH conditions and drug concentrations found in a normal (upper) and tumor (lower) microenvironment. The TME pH gradients between the lysosome and the cytosol heavily favors the lysosomal entrapment, hampering the therapeutical efficacy of the drug. Adapted from reference [22]

.

1.3 Transmembrane peptides as therapeutic systems

1.3.1 Transmembrane Proteins models

Real biological membranes are extraneously complex, as each organism, organ and even tissue is structured with differentiated cells with unique environments. The membrane bilayer of a cell can be composed of a vast array of lipid compositions that intrinsically affects the protein behaviour. Each membrane may incorporate lipids of varying backbones, lipid headgroups, aliphatic chain lengths and their number and type of unsaturations. Biomembranes may also exhibit distinct bilayer asymmetries, with different combinations of phospholipids, even with charged headgroups, sphingolipids or sterols, such as cholesterol in mammals. Meanwhile, their amphiphatic nature promotes the integration of TM proteins that alter their membrane properties, such as fluidity, area per lipid, and bilayer thickness.

So, in early stages of biological studies, it is common to use simpler phospholipid membrane bilayer models. Pure lipid bilayers, such as PC liposomes, are comprised of a highly hydrophobic region composed of lipid acyl-chains capped by two polar headgroup regions. Due to the various types of phospholipids, these models hydrophobic core thickness can range from $\sim$ 20 Å to $\sim$ 35 Å, for short and longer lipids, respectively. The PC-derived lipids (Figure 1.8), like POPC, DMPC or DPPC, are the most common as they are the major component of all animal cell membranes. These PC bilayers exhibit a full membrane bilayer thickness, at 310 K, of $\sim$ 39.1, $\sim$ 36.7 Å and gel phase, respectively [59–61]. The simple lipid models can form different types of structures, like liposomes, rafts, miceles, which provides an important framework for transmembrane protein studies.

Membrane proteins are an integral component of lipid membrane bilayers, either at the cell membrane or in intracellular organelles, representing both 30 % of the known protein genome [62], as well as most of therapeutic drug targets [63]. They can be classified as either transmembrane (TM) or peripheral proteins, with the TM proteins, specifically, being characterized by one or multiple membrane-spanning polypeptide chains. Interestingly, most TM proteins are rich in α-helical or β-sheets, since these secondary structures allow folding patterns that expose their hydrophobic residues to the apolar acyl chains, while minimizing interactions with the polar main-chains [64]. The hydrophobic pairing between the TM domains and the apolar membrane core is energetically favourable and stabilizes the membrane-embedded proteins [65, 66]. In sum, it is their adopted conformations, membrane configurations, and cell distribution that heavily dictates the final biological function. These cellular roles usually involve mediating signal transduction, cell-cell interactions, molecular transport (import and export), enzyme activation and even lipid flip-flopping [67–69].

The variability of TM proteins is immense, yet certain conserved sequence motifs have been identified, whose roles focus on promoting internal packing between peptide chains and the membrane or to further drive their TM domain internalization. These motifs are composed of multiple types of residues found flanking the TM hydrophobic region and they can act as chain-anchors, promote protein-protein or hydrogen-bond interactions or provide geometric specificity to expose other specific folding motifs to different media. Some of these motifs are: the GxxxG motif, found in proteins like glycophorin A (GPA) (Figure 1.9) [70–72]; leucine-rich motifs [73]; Polar-xx-Polar motifs that include Ser, Thr, Glu, Gln, Asp and Asn [74–76]; Gly and Leu zippers, that can be found in the myelin protein zero and the dimeric receptor tyrosine kinase (DDR1), respectively [77]; and Aromatic-xx-Aromatic motifs, typically composed of Trp residues. The presence of these motifs is fundamental for molecular recognition processes that require dynamic and large conformational changes, acting as protein switches for signal transduction with lateral, vertical and rotational motions [78]. Also, these TM domains motifs may be targeted to induce these conformational changes and inhibit certain cascading signal events. Furthermore, at the epicenter of several diseases related to signaling and molecular recognition events, we can find point mutations to polar or aromatic residues that drastically modify the stability of the TM domain, disturbing the

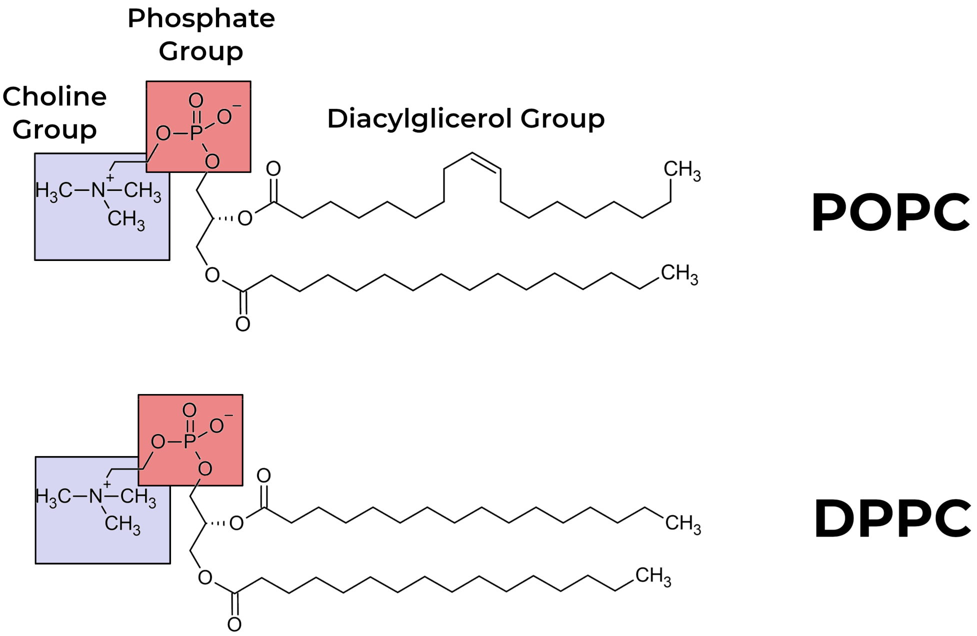


Figure 1.8: Chemical structure scheme representation of the zwitterionic phospholipids: 1-Palmitoyl-2-oleoyl-*sn*-glycero-3-phosphorylcholine (POPC) and 1,2-Dipalmitoyl-*sn*-glycero-3-phosphorylcholine (DPPC). The positive and negative lipid headgroups are highlighted in blue and red, respectively.

proper peptide-membrane configuration, specially in α-helices [79].

1.3.2 Structural design of α-helical transmembrane peptides

Despite their biological relevance and numerous studies, the complex hydrophobic nature of transmembrane proteins and domains still constitutes a challenge for standard crystallographic structural characterizations, assessment of potential therapeutic targets or identification of pathological mutations in key transmembrane chains. Similarly to biological membranes, the use of simplified models, like α-helix transmembrane peptides, paved a novel platform for more detailed mechanistic and structural studies of *wt*-TM systems, TM-based molecular probes or targeted therapeutics. The constitution of an α-helical transmembrane peptide is defined in a two-fold way: the aminoacid sequence and the membrane bilayer composition. The aminoacid sequence composes the motifs and secondary and tertiary structure that shapes the protein/peptide 3D structure and the spatial disposition of key aminoacid residues in the membrane media [80]. The lipid bilayer composition modulates the hydrophobic and polar environments that condition the possible protein-membrane configurations, the rotation of TM segments, orientation of residue side-chains, intra-molecular and inter-molecular interactions with the lipid groups or other neighbouring ligands.

Regarding the peptide sequence, polar and aromatic residue side-chains fulfill the distinct role of interacting with other TM motifs and lipid moieties. Indeed, even a single polar residue flanking one end of the hydrophobic sequence is able to stabilize an α-helical chain in the membrane bilayer, by anchoring itself in the water-membrane interface and balancing their apolar nature exposure to the hydrophobic lipid core [82]. Meanwhile, the hydrophobic stretches are energetically favorable in the lipid membrane core, as previously mentioned, typically spanning in $\sim$ 20 aminoacid residues to match the core hydrophobic length (each α-helix aminoacid length is $\sim$ 1.5 Å), maintaining a straight or slightly tilted α-helical structure. However, when there is a mismatch between the hydrophobic medium and peptide structure, either positive (protein hydrophobic length > bilayer thickness) or negative (protein hydrophobic length < bilayer thickness), the peptide will rearrange itself conformationally to minimize the energy from a thermodynamic point of view. On one hand, if it is a positive

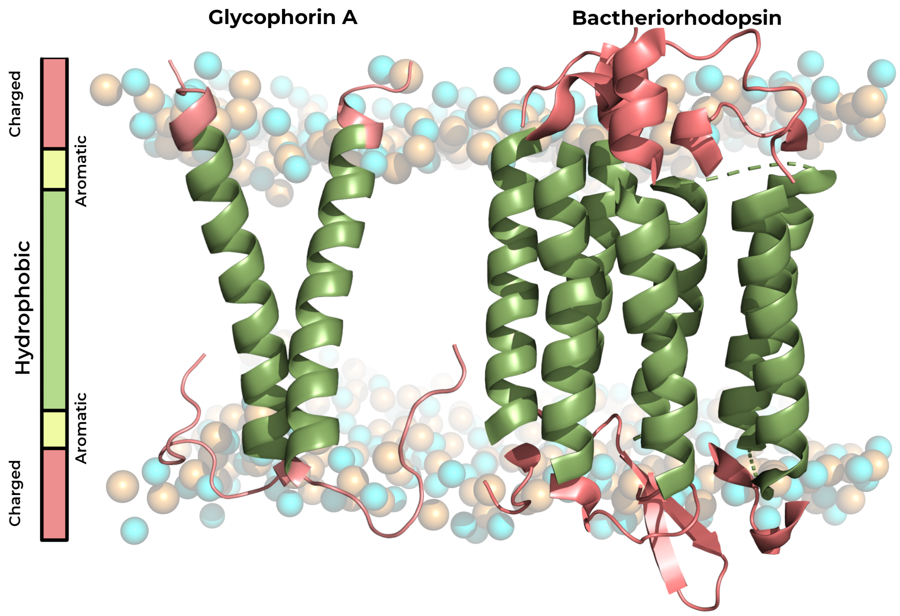


Figure 1.9: Cartoon representation of the dimeric glycophorin A transmembrane protein and the bacteriorhodopsin. The protein transmembrane domains are highlighted in green, while the hydrophilic regions are colored in red. The left rectangle denotes a simple model template for the type of residues composing an α-helix chain of a transmembrane protein/peptide.

mismatch, the peptide may either tilt, to reduce solvent and lipid headgroup exposure, or increase the bilayer thickness by inducing order in the acyl chains, as seen with Trp interactions [83]. On the other hand, if it is a negative mismatch, the helical structure may become more tightly bound (3_{10} helix) or slightly unfold in the C- and N- termini regions to expose itself to the polar medium [84]. Though, if the mismatch is too large, the peptide will exit from the membrane bilayer. Apart from conformational adjustments, the polar pH titrating aminoacid residues (Asp, Glu, Lys and His, when in physiological pH) are more adaptable than the aromatic residues, as the former may even change their protonation state to become more amenable to the medium's polarity (Figure 1.10).

Biophysical studies have also shown that, for transmembrane structure motifs, the transition from a polar medium to the apolar bilayer core requires peptide dehydration, which incurs a great energy penalty. Among all secondary structures, the α-helix minimizes the most the energy penalty associated with peptide internalization [85]. Although it is difficult to experimentally determine the cost of peptide-membrane partitioning, the partitioning of a peptide bond is approximately 1.2 kcal·mol^{-1}, while the H-bond formation associated with the secondary structure reduces the penalty by $\sim$ 0.5 kcal·mol^{-1} per peptide bond [79]. Further studies estimated the tremendous amount of energy ($\sim$ 80 kcal·mol^{-1}) required to unfold a typical 20 amino acid membrane-embedded TM domain. These calculations emphasize the importance of H-bonding stabilization in the secondary structure and the residue side-chains electrostatic interactions with the enveloping lipid headgroups. Precisely, the penalty cost reduction of the α-helix combined with the pivotal roles of polar residues and hydrophobic stretches provides a reliable template to design a standard α-helical transmembrane peptide.

In sum, a standard template, as exemplified in Figure 1.9, for these amphiphilic transmembrane α-helical peptide combines: (1) a hydrophobic core comprised of poly-Leu or Leu-Ala residue stretches, as both possess high hydrophobicity and propensity to form α-helices [86]; (2) the presence of aromatic hydrophobic residues, usually Trp [66, 87], that due to their bulky side-chain, act as anchors and fluorescence probes in the ester region; (3) one or multiple polar residues that may be either titrating or be fully (un)charged in physiological conditions

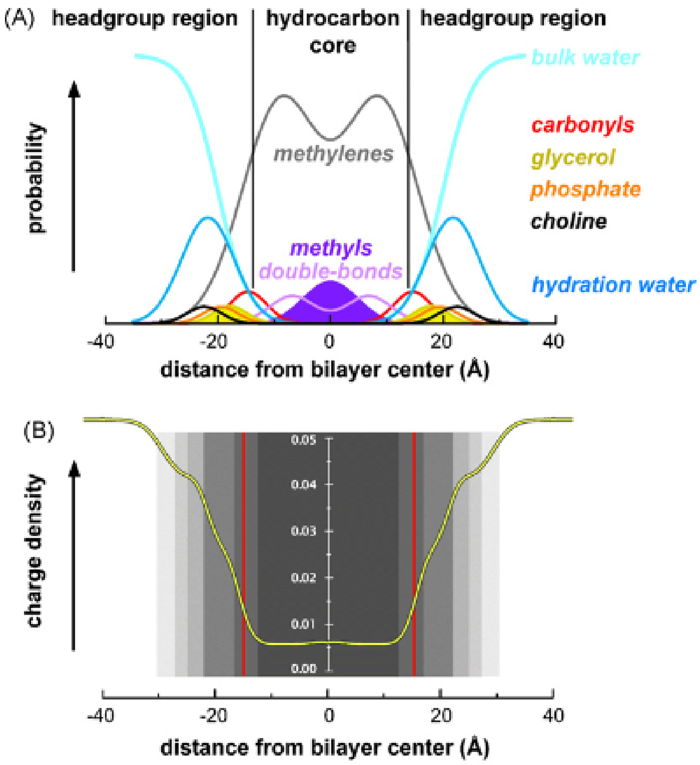


Figure 1.10: Membrane bilayer profiles of the distribution probabilities of chemical moieties at different 1,2-Dioleoyl-*sn*-glycero-3-phosphocholine (DOPC) membrane regions (A); and the charge distribution from the bulk solvent to the hydrophobic membrane core (B). Adapted from ref. [81].

depending on the desired peptide function, thermodynamic stability and insertion/exit kinetics.

1.3.3 Application of peptides as role models

Some peptide design strategies rely only on adopting a single chain of an interesting transmembrane protein as a template and then assess their potential as a therapeutic target. Meanwhile, other approaches induce point mutations in the *wt*-sequences to manipulate several peptide properties, like the secondary structure, to induce hydrophobic mismatching, to promote distinct membrane behaviours, facilitate cargo attachment or improve membrane-tethering. Finally, with the growing knowledge on the structural configuration of TM peptides, peptide rational design may develop unique and fine-tunable strategies that optimize the targeting efficacy, cell entering and retention or cargo transport for particular biomolecular targets. In this subsection, we will tackle a few successful cases of TM peptides for both fundamental and applied research.

The WALP/KALP was one of the first amphiphilic peptides to be synthesized with the purpose of studying protein-lipid interactions [65, 88]. The overall peptide composition (Figure 1.11) adheres to the referenced template as it is comprised of: (1) an hydrophobic n Leu-Ala $(LA)_n$ stretch, with n LA pairs depending on the desired length (Å) of the central stretch, with a $(LA)_8L$ matching the typical ∼25 Å hydrocarbon thickness; (2) either aromatic (Trp) or polar (Lys) residues flank these regions, naming each variant either WALP or KALP, respectively; and (3) the capped N- (acetyl group) and C- (amide group) termini regions. Generally, these peptides exhibit a full α-helix spanning the entire membrane bilayer, though several studies were conducted to assess how distinct flanking residues and central stretch lengths affect the hydrophobic mismatch. First, the central stretch possess a hydrophobicity threshold for transmembrane association, where certain moderate hydrophobic residues (Ser, Gly) are insufficient to comprise a motif for membrane internalization, whereas the poly-Ala, poly-Leu, and poly-Ala-Leu residues suffice [89, 90]. Overall, there is loss of transmembrane orientation if there are significant deviations from these templates. Moreover, the inclusion of titrating residue, like Lys, His, Asp, Glu, within the central stretch confers an amphiphatic nature to the peptides that improves membrane surface adhesion and induces a pH-dependent insertion and folding.

Regarding the flanking residues, the inclusion of either Trp, Lys or both confers significant changes in the membrane-inserting properties of these peptides. In the case of KALP peptides, they incorporate mostly Lys residues, which at the C-terminus region promotes spontaneous transmembrane insertion [91], yet only in basic conditions. Since the Lys solution pK_a is $\sim$ 10.5, a deprotonation event at basic pH values is required [27], otherwise a large energy cost is associated to translocating a positive charge through the membrane. And, although the removal of these Lys flanking residues does not impair peptide internalization, KALP peptides also exhibit a substantial loss of solubility that hampers bioavailability [92]. So, the presence of positively-charged residues in both flanking regions is very helpful in stabilizing the inserted state of shorter peptides, as the side-chain of cationic residues, such as Lys or Arg, will move toward the polar environment in a movement named snorkeling [93–98]. The snorkeling movement helps peptide stabilization in a two-pronged way: the positively charged amine side-chain preferably interacts with the anionic phosphate groups and water molecules in the water-membrane interface [99], therefore anchoring the peptide; (2) simultaneously, these electrostatic interactions reduce the associated energy cost of membrane internalization, thus thermodynamically favouring the inserted state. Concomitant with the snorkeling movement, the peptide may even tilt, dragged by the exerted force of the residue side-chain, depending on the cationic residue sequence position and spatial disposition of the transmembrane segment (see Chapter 5). Finally, the membrane position of these titrating polar residues is of utmost importance, as their surrounding electrostatic microenvironment modulates their pK_a, which affects their overall thermodynamic stability and (de)insertion kinetics, as seen in other peptides (see Section 1.4 and Chapters 4–5) [100–104].

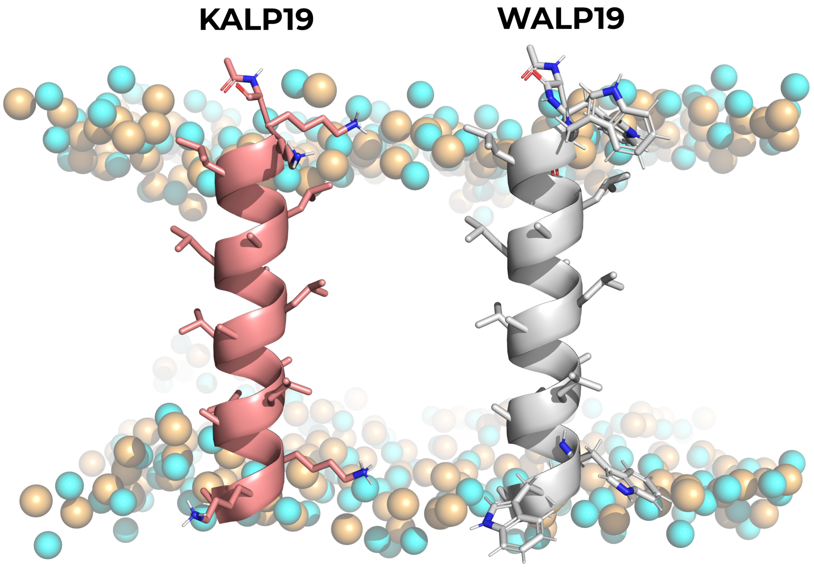


Figure 1.11: α-helical representations of the KALP19 and WALP19 peptides inserted in a POPC membrane bilayer. The Lys and Trp residues are populating the water-membrane interface, while the Leu-Ala motifs populate the membrane core. These aforementioned residues are depicted as colored sticks.

Concerning the WALP peptides, they exhibit similar internalization characteristics, though they possess far more flexibility in stabilizing both shorter (WALP$_{16}$ - GWW(LA)$_5$WWA) and longer peptides (WALP$_{23}$ -

GWW(LA)$_8$LWWA). The aromatic and hydrophobic indole side-chain consistently anchors in the ester region, near the carbonyl groups, resisting displacement both to the solvent, due to their apolar nature, and to the membrane interior, as their bulky side-chain is not energetically favoured for internalization. Therefore, the inclusion of Trp residues in both ends of the hydrophobic stretch confers an adaptable response in these TM peptides, as matching the water-membrane polarity supersedes the hydrophobic mismatching, further imposing a configurational rearrangement.

While WALP/KALP studies were particularly important in assessing the flanking residues role, the side-chains reorientation and lipid interactions and folding pathways, other peptides, like the GALA peptide, clarified the mechanisms of protein aggregation and channel formation for drug and gene delivery. Despite the deceiving name, the GALA peptide is an amphiphilic α-helical peptide constituted of EALA repeats, an N-termini Trp and a central His residue - **W**EAALAEALAEALAE**H**LAEALAEALEALAA. Contrary to the KALP peptides, there are anionic polar residues (Glu) integrating the peptide sequence, conferring an acidic pH preference for both membrane internalization and peptide folding in aqueous solution [105–107]. A sequence of $\sim$30 aminoacids conforms to the established template as it provides a stable transmembrane α-helix without a significant hydrophobic mismatch. The repeat of leucine-alanine residues increases the hydrophobicity, while the Glu repeat imparts a pH-dependent component to the hydrophobicity and a charged side-chain to energetically stabilize the solvent exposure. This solvent exposure stability is fundamental when the peptide is in aqueous solution and, more specifically, when multiple monomers aggregate to form a pore complex that allows free movement of water molecules through the open channel. At acidic pH's ($\sim$5.0), the solution Glu pK_a ($\sim$4.5) allows neutralization that promotes peptide folding and membrane insertion, as the neutral state is more apolar. However, at acidic pH, the GALA peptide is prone to aggregation at high peptide concentrations, even when membrane-inserted [108], due to the cooperativity between protein helices to form channels. At neutral pH, even though it is expected that the desolvation effect and the electrostatic neighbours promote a pK_a shift that favours the protonated state [27], the GALA peptide becomes destabilized, then released from the lipid bilayer. To address this issue, a KALA peptide was designed with the intent to bind and deliver DNA molecules. Regarding GALA, the Glu enriched α-helix impedes the negatively charged DNA backbone from binding and crossing the pore. The KALA repeat not only allows the gene delivery, as the positive charged amine groups bind with the DNA backbone, it also prevents nonspecific aggregation observed in high concentrations of GALA, while also preserving their membrane-perturbing properties [109]. However, the functional pH range of this peptide is limited to more alkaline environments (pH $>$7.5), due to the Lys high pK_a, which hinders its applicability to tumor treatment.

Overall, the major selling point of GALA is its ability to extensively perturb lipid membranes and drive membrane fusion, membrane fragmentation, and more importantly, induce bilayer permeabilization through pore formation [110]. Leakage experiments performed in PC large unilamellar vesicles (LUV) at acidic pH, showed that a 2500/1 lipid/peptide molar ratio promotes the formation of quasi-irreversible pores through the aggregation of, at least, 10 monomers in the membrane bilayer (Figure 1.12) [111]. The permeabilization-inducing ability has several applications from drug or gene delivery to tumor cells, as their acidic pH sensibility targets the TME acidosis.

Both these types of peptides showed remarkable versatility and customizability that provided valuable insight into the positive and negative mismatch conformational responses, the role of flanking residues in the water-membrane interface, side-chain reorientation, pore formation and protein aggregation. In sum, the use of these toy model peptides was paramount in driving forward fundamental research of peptide design of antimicrobial peptides [112, 113], synthetic sequences [65, 105], peptidic venoms [114], fusion peptides [115, 116] and cell-penetrating peptides [117–119], with a wide variety of practical and clinical applications, especially in the case of the pHLIP peptide.

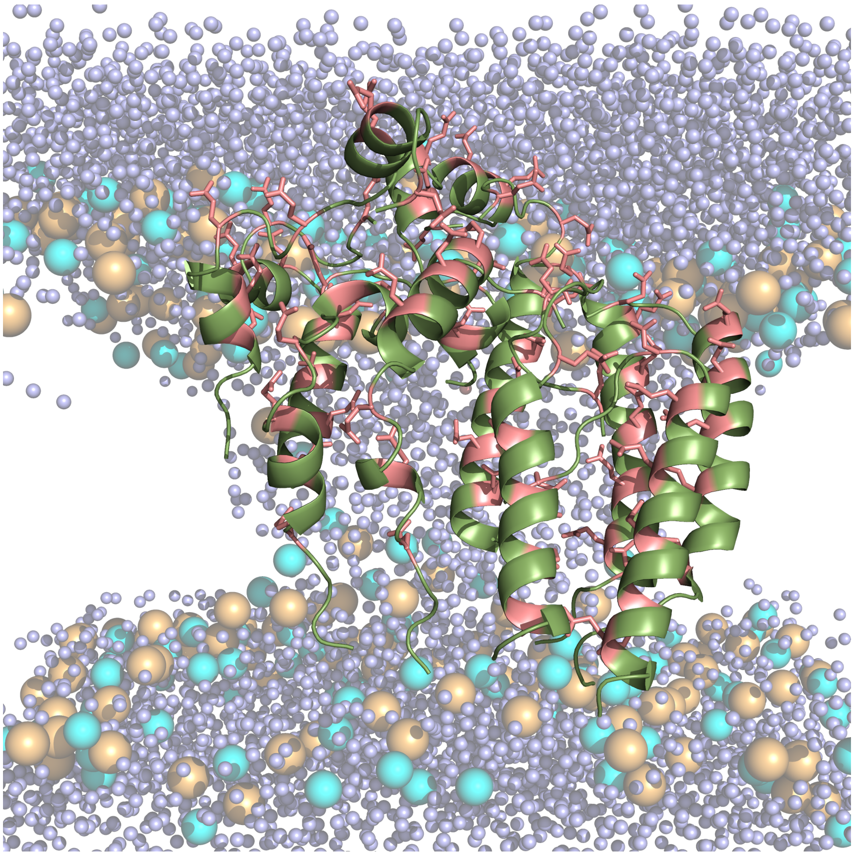


Figure 1.12: Graphical representation of a membrane pore induced by a 10-mer GALA complex at pH 7.0. The glutamate aminoacid residues are highlighted as pink sticks, while the water molecules crossing the membrane pore are shown as light blue spheres. The membrane bilayer is formed by 256 POPC lipids represented by the phosphate (light brown) and choline (cyan) groups.

1.4 pHLIP as a model pH-dependent system

1.4.1 Structural details and potential applications

The pH-low insertion peptides, commonly known as pHLIP, constitute a family of transmembrane α-helical peptides, whose membrane internalization heavily depends on the medium acidity [52, 117, 120]. The 36 aminoacid *wt*-pHLIP was derived from the transmembrane C-helix of the seven-polypeptide chain bacteriorhodopsin (Figure 1.9), the only one that spontaneously equilibrated with the membrane in a pH-dependent manner [121]. For several decades, this peptide family has been thoroughly studied for a wide range of applications. Each termini can be modified to transport molecular cargo with different roles: if attached to the N-terminus, it works as a molecular probe and biomarker, by attaching a fluorophore or using the intrinsic Trp residues, for thermodynamic studies, *in vivo* imaging purposes (Figure 1.13) and tumor cell pH measurements [16, 52, 122–125] ; if attached to a Cys residue in the C-terminus, it behaves as a drug delivery system, particularly in targeting the solid tumor microenvironment. As for drug delivery, the peptide-ligand conjugate can transport therapeutic drugs (amanitin [126], doxorubicin [127]), gene inhibiting peptide nucleic acids (PNA's) [128–130], nanoparticles [131], inhibitors of key cancer-related enzymes [132] and even with other antimicrobial peptides [133].

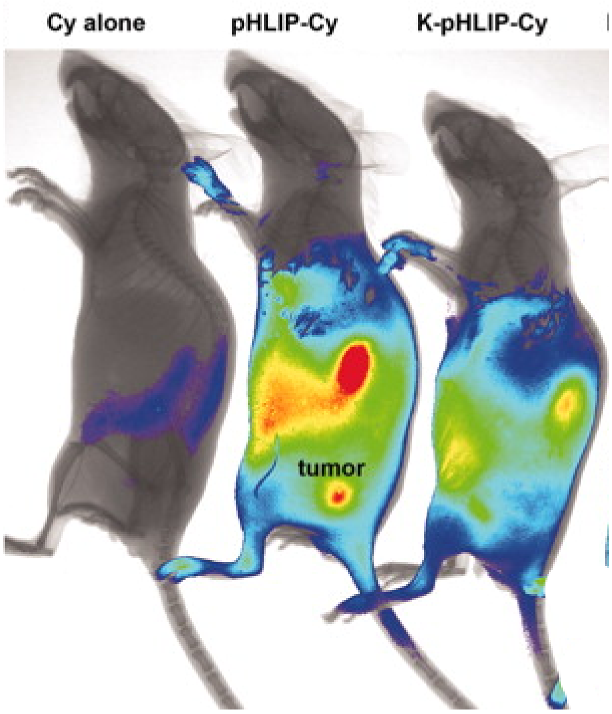


Figure 1.13: Imaging application of two pHLIP peptide variants *wt* and K-pHLIP, as a negative control, on mice presenting tumors on their right flank. Both the tumor and the kidney are highlighted in red as a consequence of their acidic environment. The K-pHLIP variant loses the acidic specificity as the mutated Lys-14 adopts a positive charge that impairs membrane insertion and state III stability. Adapted from reference [52].

Fluorescence spectroscopy and isothermal titration calorimetry (ITC) studies found that pHLIP equilibrates between three major states, depending on environmental conditions: (I) unfolded monomeric state in solution; (II) at high pH values (>7–8), it remains mostly unstructured and adsorbed to the lipid bilayer; (III) at acidic pH (<7), the peptide spontaneously inserts and folds into a kinked α-helix, spanning the whole membrane bilayer (Figure 1.14) [52, 103, 120]. Similarly to other transmembrane peptides, *wt*-pHLIP follows the same template rules as other sequences - ACEQNPIY**W**ARYA**DW**LFTTPLLLL**D**LALLV**D**A**DE**GTG - possessing an hydrophobic central stretch with two flanking regions. However, the similarities are mainly superficial, as pHLIP possesses unique features that modify both thermodynamic and kinetic behaviour. The N-terminus flanking region is rich in aromatic and polar residues, possessing two Trp residues (W-9 and W-15) acting as fluorescence probes, a cationic residue (Arg-11) acting as an anchor, a titrating anionic residue (Asp-14), which is a key the pH sensor, and the Cys-2 for ligand binding purposes. Meanwhile, the C-terminus flanking region comprises of several polar residues, such as the titrating anionic residues (Asp-31, Asp-33, Glu-34 and C-terminus) and an uncharged Thr-35. Interestingly, the hydrophobic stretch is rather shorter (∼16 aa) than other peptides (∼20 aa), only possessing a poly-Leu motif with an interjecting Asp-25 residue, revealing a negative hydrophobic mismatch that is mitigated by the kink induced by Pro-20 [134, 135]. The proline breaks the α-helix turn and allows the hydrophobic segment to be fully embedded in the acyl chains (see Figure 1.14).

Each of these residues plays a fundamental role in the distinct nature of pHLIP compared with other amphipathic peptides. For example, the WALP peptides insertion process is rather pH insensitive, accompanied by spontaneous α-helical folding with membrane adsorption, as previously reviewed. Meanwhile, the pHLIP insertion is mainly pH-driven, requiring the protonation of the carboxylic groups (Asp-14, Asp-25, Asp-31, Asp-33, Glu-34 and the C-terminus) to increase the peptide hydrophobicity and mitigate the energy cost of internalization [101, 120, 136, 137]. Additionally, the relatively lower number of hydrophobic residues also denotes a faster blood clearance and lower affinity for cellular membranes than other TM peptides [120]. This is particularly important, as the insertion kinetics vary distinctively depending on the physical properties of the lipid bilayer [138] and the interactions with the lipid headgroups [139]. Nevertheless, the overall pHLIP-membrane partitioning does not significantly deform the membrane bilayer, only slightly perturbing [103, 104, 140], nor it exhibits pore formation like the GALA peptide. Though the pHLIP insertion is generally uni-directional, with the N-terminus in the outer membrane monolayer and the C-terminus in the inner monolayer, the full folding pathway has yet to be fully discerned. However, there are several models indicating the formation of multiple intermediary steps in the insertion and exit process, depending on the type of peptide sequence, cargo and membrane bilayer [136, 137, 140, 141]

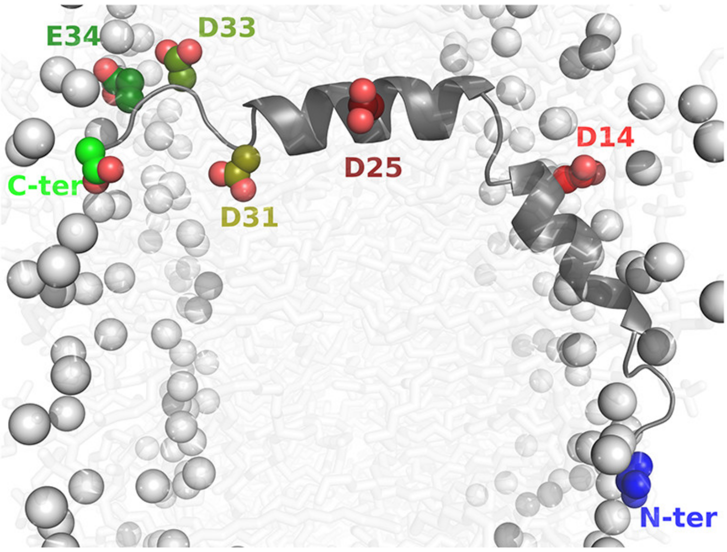


Figure 1.14: Graphical representation of *wt*-pHLIP peptide in a stable state III depicting the characteristic α-helical kink. The peptide is inserted in a POPC membrane bilayer represented by the phosphate headgroups as grey spheres and the acyl chains as transparent sticks. All titrating residues side chains are depicted as colored spheres in their preferred membrane region. The *wt* peptide sequence is: ACE-QNPIYWARYADWLFTTPLLLLDLALLVDADEGTG . Adapted from reference [103]

.

1.4.2 The main modulators of pH-dependence

In the grand scope of the pHLIP tumor targeting ability, the pH-dependent insertion is the defining trait of these peptides, owing it to the several titrating anionic residues in their sequence. Experimental studies have found that the *wt*-pHLIP pK of insertion ($\sim$6.0 to 6.2) [52] correlates well with the TME pH range (6.2 to 6.8), hence its remarkable tumor specificity. To ascertain the molecular impact of these carboxylic acids, fluorescence thermodynamics and computational studies were pivotal in studying the state transitions at different pH values, the stability of the membrane inserted state [140, 142] and the electrostatic interactions governing the pH dependency [102–104, 143, 144]. For instance, the thermodynamic stability can be correlated with the pK of peptide insertion (pK^{ins}), the pH value at which 50% of the peptide population is either in state II or state III, thus in equilibrium between membrane-adsorbed and membrane-inserted.

Using constant-pH molecular dynamics (CpHMD) studies, it was found that the experimentally measured pK^{ins} was in agreement with the individual Asp-14 pK_a behaviour(Figure 1.15-A). Therefore, Asp-14 was found to be the key insertion trigger, as the protonation states heavily dictated the peptide thermodynamic stability [103]. Due to the kinked α-helix of state III, Asp-14 preferably populates beneath the membrane surface, hence operating between two worlds: the apolar membrane core and the polar water-membrane interface. This singular electrostatic vicinity envelops the Asp-14 with diverse interacting molecular groups that modulate its behaviour along its membrane insertion. The vertical peptide movement, associated with internalization and exit processes, conditions the presence of these interacting electrostatic species, like lipid phosphate and choline groups, solvent accessibility and even nearby residues, therefore shifting the pK_a and the preferable protonation states at the relevant pH range [103, 104]. As such, the protonation/deprotonation events promote/hinder the state III membrane stability or the state III–II transitions. Finally, the central Asp-25 appears to mostly help in driving thermodynamically the transition processes between state II and state III. When in state III, the deprotonation of Asp-25 is pivotal to energetically destabilize the inserted state, due to the negative charge in the hydrophobic membrane core. On the other hand, the transition from state II to state III requires the protonation of Asp-25, slowing down the insertion kinetics, as an extra protonation step is required to reduce the polarity of the membrane-embedding peptide structure [102, 103, 143]. In addition, several studies have highlighted the slow kinetics issue (ms–s timescale) [101, 123] of the *wt*-pHLIP, due to the polar-rich C-terminus region in the inner water-membrane interface. The presence of four carboxylic acids requires full protonation/neutralization, in a topologically ordered sequence, to increase the hydrophobicity of the flanking region and favour the membrane crossing event (Figure 1.15-B). Otherwise, there is an insurmountable energy penalty associated with translocating a negative charge across the apolar hydrophobic core [103, 123, 141].

1.4.3 Designing pHLIP variants

Despite the wide range of applications of *wt*-pHLIP, there is a known issue concerning the lack of specificity in targeting acidic tissues. Although the pK^{ins} is within the optimum pH range of the TME, several healthy tissues possess a natural acidity, like the kidneys and stomach, which can be targeted by pHLIP (Figure 1.13). A possible solution focuses on restricting the pH range which promotes a pHLIP inserted state in equilibrium, hence increasing its specificity to a particular acidic environment. The L16H peptide variant (Figure 1.16) was a proof-of-concept sequence that only mutated a Leu with a cationic residue (His) in the N-terminus flanking region [103]. In these studies, the His-16 populated deeper membrane regions than Asp-14 and showed a significant pK_a shift (3.3) that created an insertion pH range, where the two protonation switches (Asp-14 and His-16) were neutral, thus favoured the inserted state thermodynamic stability. At lower pH values, the histidine also protonates, thus gaining a destabilizing positive charge and propelling the peptide to an exit pathway. Although, the induced insertion pH range was not optimal for practical applications, it worked as a template for further peptide variants.

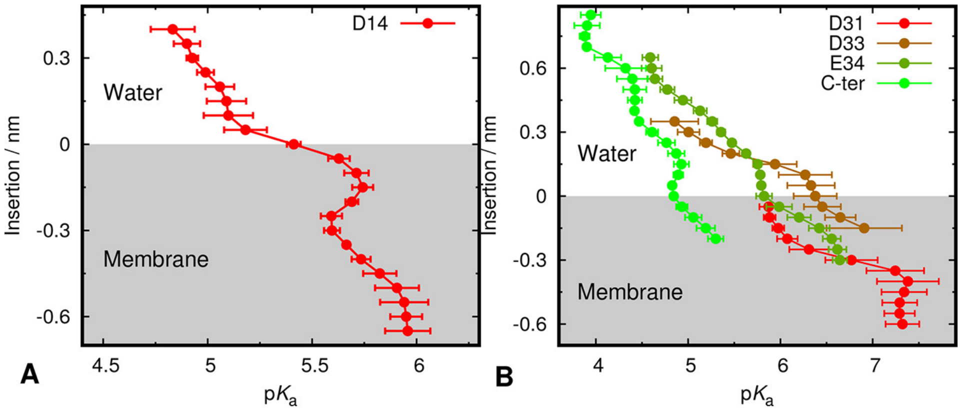


Figure 1.15: pK_a profiles of the key Asp-14 (A) and the C-terminus region carboxylic residues (B) of *wt*-pHLIP along the membrane normal obtained from CpHMD simulations. The final estimated value (5.9±0.1) in the Asp-14 pK_a profile correlates with the experimental pK^{ins} (6.0). In B, the large magnitude of the pK_a shifts is related to their topological order and the dense negative potential generated by them. Each residue requires a transient protonation of an adjacent residue to decrease their own pK_a shift and energetically favour the neutral state, as they reach similar insertion levels, promoting residue insertion. If not, the local negative potential is unfavorable for protonation, the residues are stabilized by the solvent and the peptide remains anchored in the inner membrane monolayer. Adapted from reference [103].

Besides the thermodynamic stability, the therapeutic potential of pHLIP is also intimately tied to the kinetics of insertion, particularly when transporting polar and bulky cargo across the membrane bilayer. If the *wt* peptide membrane translocation (ms–s timescale) is slower than the blood clearance, it hinders the therapeutic

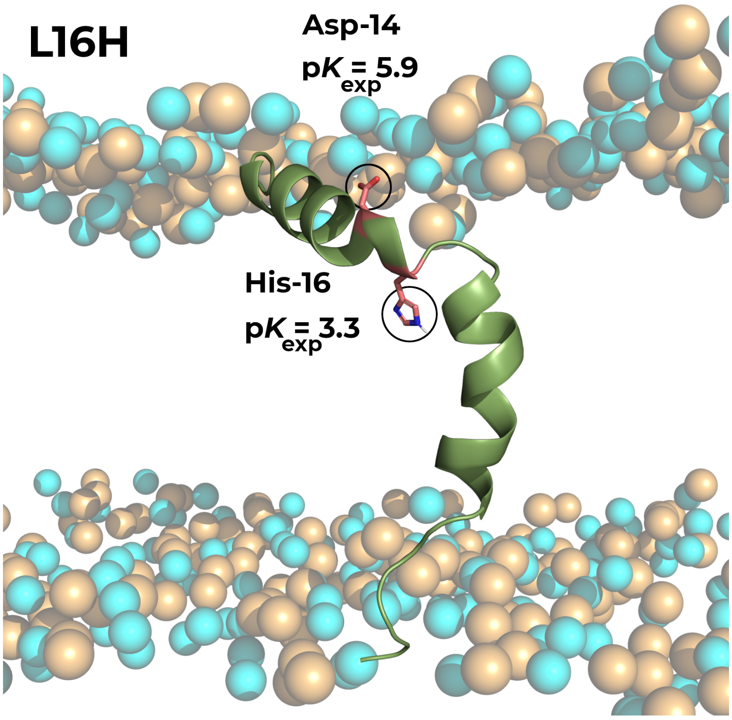


Figure 1.16: Graphical representation of the L16H pHLIP variant. The cationic mutation (His-16) adds an extra protonation trigger (pK^{ins} 3.3) that, in acidic medium, promotes peptide exit processes. The existence of an anionic and cationic protonation triggers creates an insertion pH range, where both residues are neutral and the state III is stable.

applicability. Furthermore, the associated energetic cost of translocation can also be too high to be thermodynamically favorable, as evidenced in several studies [103, 141]. To address these issues, several peptide variants were designed in an attempt to improve the kinetics, stability and translocation ability of pHLIP [123]. From this work, a promising peptide variant (Var3) (Figure 1.17) showed improved insertion kinetics and better therapeutical efficiency in tumor cell lines studies, when compared to the other peptide designs. The Var3 peptide (27 aa) exhibits a shorter α-helix and fewer C-terminus titrating acidic residues, which facilitates the required transient protonation, thus accelerating the insertion kinetics [123]. Interestingly, the lower pK^{ins} (5.0), which was measured in POPC liposomes, would suggest an impaired tumor targeting ability, since only a small fraction of the peptide would internalize at the TME pH. Still, the Var3 outperformed the *wt* in *in vivo* conditions, suggesting a higher *in vivo* pK^{ins} than in the liposomes [125], proving to be far more efficient in practical applications [145, 146]. In sum, the rational design of peptide variants can be advantageous not only for clinical applications, but also for fundamental research. Most of the known molecular details derived from modifying *wt*-pHLIP peptide into a wide array of variants. The peptide sequence fine-tuning mostly included point-mutations in key residues [102], permutation of residue positions [143] or residue inclusion/deletion [103, 123]. By modifying pivotal residues, such as Asp-14, Arg-11 and Asp-25, new fine-tuned variants successfully provided valuable insights into pHLIP structure, function and performance. At the epicenter of these studies, *in silico* methodologies provide a unique comprehension on the molecular minutiae that govern the behaviour of this promising peptide family.

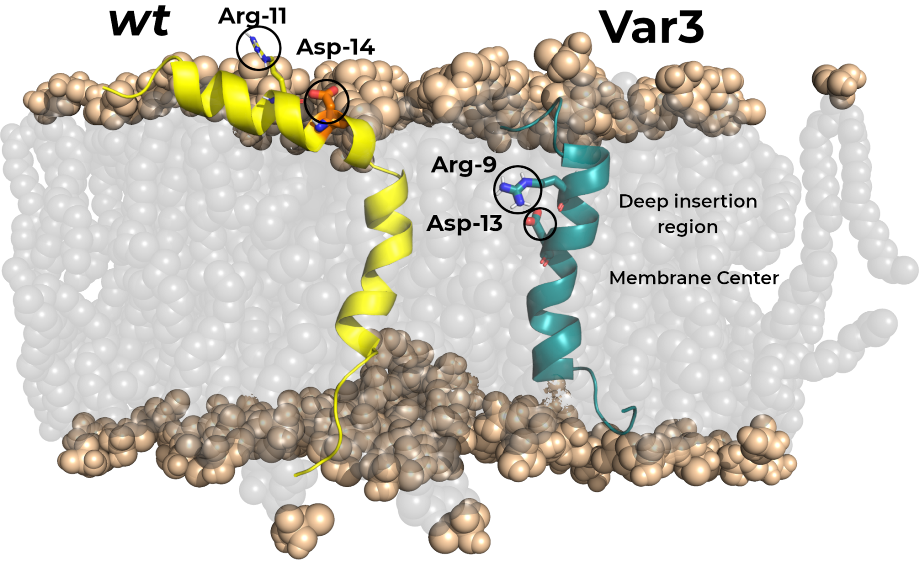


Figure 1.17: Graphical representation of the *wt* and Var3 (ACDDQNPWRAYLDLLFPTDTLLLDLLW) peptide structures, at state III, in a POPC membrane bilayer. The key aspartate and arginine residues are highlighted as sticks and depicted in their preferred membrane region.

1.5 *In silico* methodologies

Experimental studies provide the bulk of knowledge regarding clinical applications, kinetic and thermodynamics processes of transmembrane peptides. However, these often lack in discriminating power regarding the molecular details that explain macro-molecular events. To understand these molecule intricacies, it is often required the use of *in silico* methodologies, imperatively so when it concerns to pH-dependent properties. Constant-pH MD (CpHMD) methodologies define the pH as simulation thermodynamic variable, like the temperature or pressure, and they are paramount to properly model the correct protonation and conformational equilibrium of pH-dependent systems, particularly in peptides and proteins interacting with membranes.

Currently, there are several CpHMD methodologies available that treat the protonation states of chemical groups as either continuous [147–150] or discrete [151–155]; the solvent as implicit or explicit; and are compatible with different force fields. In this thesis, all CpHMD simulations were performed with the stochastic constant-pH molecular dynamics (CpHMD-L), so this is the only method that will be addressed (see Section 2.3). The CpHMD-L method [27, 156–158] is a discrete explicit-solvent methodology that merges two simulation approaches: (1) Poisson-Boltzmann/Monte-Carlo (PB/MC) calculations that estimate the protonation states, for all chosen titrating residues in a single conformation, at a specific pH value; (2) and molecular mechanics/molecular dynamics (MM/MD) simulations which sample system conformations using the previously estimated protonation states. This methodological implementation is based on the stochastic titration method[151, 159] and has proven successful in characterizing numerous systems from globular proteins [156, 160], dendrimers [161], peptides interacting with membranes [27, 103, 104, 162], membrane protein channels [163] and, even titrating membrane bilayers with mixed lipid compositions [158, 164, 165]. Indeed, the method is also capable of quantitatively estimate pK_a values of membrane-inserting moieties (Figure 1.18), hence measuring the pK_a shifts depending on the surrounding medium along the membrane normal, as shown in several works [27, 37, 103, 104, 166]. The work of Teixeira et al. [27] was a seminal study of titrating molecules interacting with membrane bilayers. The pK_a is a macroscopic property, whose meaning breaks in the size scale of MD simulations. Usually, CpHMD simulations obtain information regarding the proton binding affinities of residues

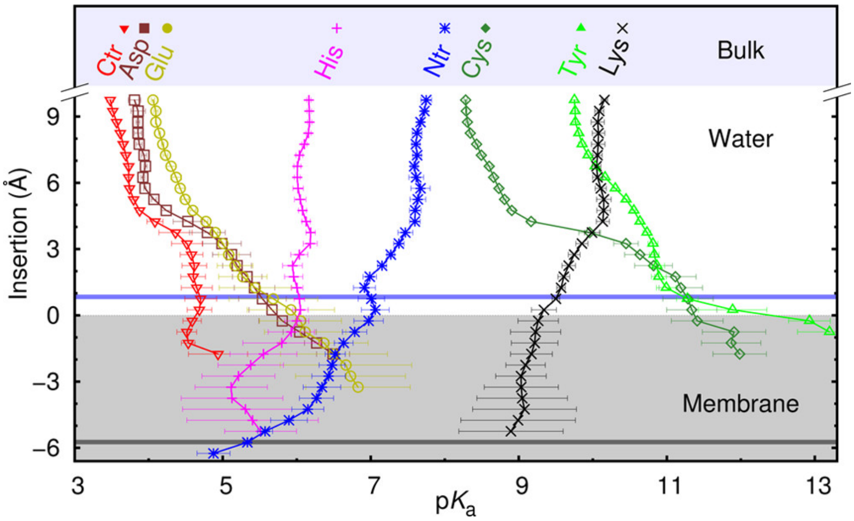


Figure 1.18: pK_a profiles for several aminoacid residues in the water-membrane interface. These calculations were performed from Ala-based pentapeptides simulation (Ala$_2$-X-Ala$_2$) in DMPC membrane bilayers. The anionic residues exhibit a pK_a shift to higher values, while the cationic residues exhibit an inverted pK_a shift along membrane insertion. Both behaviours promote charge neutralization,as a result of the desolvation effect, that energetically is more favorable for membrane insertion. Adapted from reference [27].

that typically correlate well with the macroscopic pK_a values. However, certain environments, such as the water-membrane interface, prove to be difficult to properly sample the protonation/conformation equilibrium, due to lack of sampling of rare events in the ns-μs timescale. Furthermore, when performing single pK_a calculations of a residue in membrane environments, the estimation takes into account all conformations regardless of the electrostatic environment, ranging from solvent exposed residues to deeply membrane-embedded. Consequently, all the information pertaining the residue insertion distribution and protonation equilibrium becomes convoluted into a single value and loses physical meaning. By discerning the protonation equilibrium along the membrane insertion of a residue, we describe a profile of the shifting pK_a of a residue along the membrane internalization. These pK_a profiles provide valuable insight into the thermodynamic and protonation equilibrium of complex peptide-membrane and drug-membrane configurations. Regarding the pHLIP peptide, this model predictive power has elucidated the impact of the key titrating residues positions along the membrane normal upon the α-helix stability and side-chain interactions, and it is even capable of predicting the macroscopic experimental pK^{ins} [103, 104]. As referenced in Section 1.4, transient (de)protonation events are crucial for the exit/insertion processes and state III thermodynamic equilibrium in the membrane [103]. Besides transmembrane peptides, the same methodology describes the behaviour of titrable compound drugs (Lewis bases) in distinct media [37], and even predict the membrane permeabilities at different pH cell conditions (see Chapter 6). In spite of these advantages, the stochastic CpHMD methodology limitations concern on the slow conformational convergence, despite the quick protonation convergence, thus hindering the sampling of rare system events [103], and the difficulty to cross large energy barriers between conformational states, like in drug membrane crossing. These limitations hamper the simulation accuracy, speed and detail that the study of these systems require. To overcome these hurdles, enhanced sampling methodologies were coupled to the stochastic CpHMD methodology. More specifically, our group developed two state-of-the-art implementations of the replica-exchange (RE) [167] and umbrella sampling (US) [168] techniques: the pH-replica exchange (pHRE) [166] and the Umbrella Sampling/constant-pH MD (US-CpHMD) [163] techniques. The pHRE technique improves the protonation/conformation sampling of high energy states, such as ionized residues deep in the membrane, by performing simultaneous n CpHMD simulations (pH replicas), each at different pH values, then exchanging the pH values with adjacent replicas (see section 2.4) [166]. This method presented remarkable results by improving the simulation speed and pK_a accuracy (Figure 1.19) relative to standard CpHMD simulations [104, 166]. On the other hand, the US-CpHMD method is a much slower method compared to the pHRE, although it is capable of overcoming large energy barriers that hamper the standard CpHMD and pHRE methods. The US-CpHMD methodology increases the conformational/protonation sampling by generating n system slices through a pulling procedure, for example along the membrane normal of a transmembrane system. Then each slice consists of a different insertion position of the compound drug, forcibly restrained within the slice range, and the method

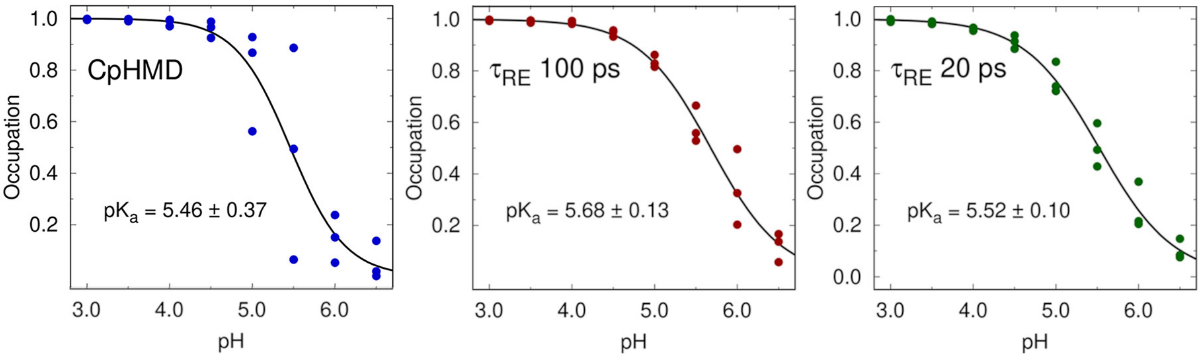


Figure 1.19: Titration curves generated from CpHMD and pHRE methodologies for the Glu-35 in HEWL. These results highlight the increase in accuracy of the titration curves provided by pHRE using two values of τ_{RE}. Adapted from reference [166].

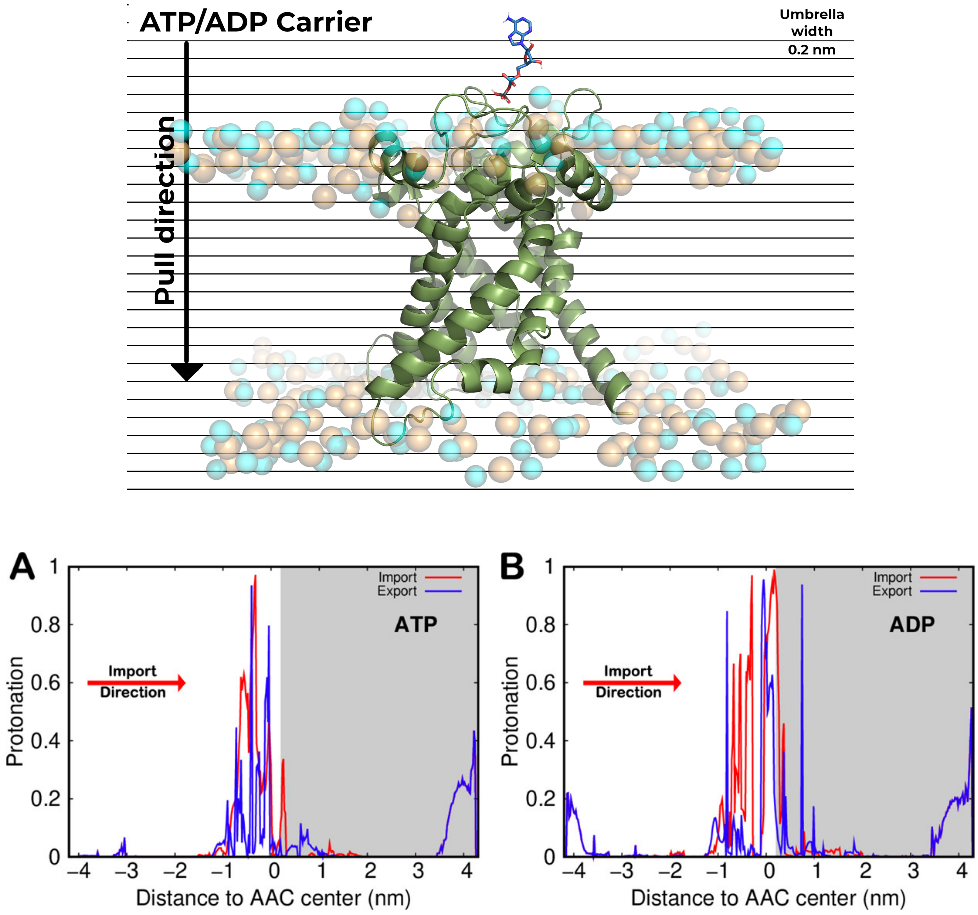


Figure 1.20: In the upper plot, a graphical depiction of the ADP pulling procedure, through the ADP/ATP carrier, required by the US-CpHMD method to generate conformations for each umbrella. Then, each CpHMD simulation is performed on the generated conformation restrained to the umbrella width. The lower plots highlight the protonation profile of the ADP molecule generated along the channel. This type of analysis is only possible through US-CpHMD. Adapted from data obtained in reference [163].

performs individual CpHMD simulations per slice. Although the method incurs a great computational cost due to the large number of simulations/slices, our model is capable of fully describing the process of compound insertion (Figure 1.20), even for unfavorable protonation states, and also calculating membrane permeabilities for drug compounds. For more details, the working principles and applications of these methodologies are described in Section 2.4 and Chapters 3 and 6, respectively.

1.6 Objectives and Goals

This thesis was born from the challenges of understanding the molecular interactions that govern the macroscopic behaviour of biomolecules, in particular, the microscopic minutiae of promising anti-tumor technologies. Constant-pH MD methodologies possess the unique ability of mimicking and predicting *in vivo* pH-dependent phenomena with a remarkable accuracy. Therefore, the use of these methodologies can only be advantageous in propelling and guiding the rational design of novel acidity-targeted anti-tumor technologies. With this in mind, this thesis major purpose can be broken down into three major goals:

(1) assess the implementation of enhanced sampling techniques through the development of a protocol for complex transmembranar systems simulations;

(2) understand the defining electrostatic network and molecular interactions that govern the aforementioned pH-dependent systems: pHLIP peptides and Lewis base drugs;

(3) highlight the potential chemical modifications to key group moieties, in both the chemotherapeutics and drug delivery systems, to improve their therapeutic efficacy and provide crucial information for technological rational design.

By successfully achieving these goals, we hope to provide a reliable framework for enhanced-sampling CpHMD simulations, the required tools for accurate technical analysis to interpret complex transmembrane systems, and to usher new strategies to elude the MDR defense mechanisms, through the rational design of acidity-targeted anti-tumor technologies.

Chapter 2

Methods

In this chapter, a brief review of the theoretical models behind all the following work will be presented. The purpose of this chapter is to contextualize the reader on the advantages, limitations and reasoning of several *in silico* approaches that try to simplify complex biological phenomena into mathematical formulations. In the scope of this thesis, we will focus on molecular mechanics and molecular dynamics, enhanced sampling techniques and briefly on molecular compound parametrization.

2.1 Molecular Mechanics and Molecular Dynamics

In its essence, biomolecular simulations are simplified representations of molecular events that encapsulate the more relevant properties of the system of interest. Depending on the desired properties, the system may be either described in high detail using quantum level calculations, that consider electron and nuclear motions, or in more broad approximations where not even individual atoms are considered, clustering in beads, such as coarse-grain models [169]. All system simulations presented in this thesis used a molecular mechanics (MM) model to describe the system with atomistic detail and a molecular dynamics (MD) approach to sample the system conformational space. In the molecular mechanics model, all atoms are described as points masses in the conformational space only using their nuclei positions. This is possible as the Born-Oppenheimer approximation assumes that the electron movement can be dissociated from the nuclear motion, since the larger relative mass of the nucleus fixates its position, while the fast electronic movement adapts very quickly to the nuclei position. Although this approximation precludes polarization effects, the trade-off in calculation speed enables the sampling of molecular properties in larger time-scales by adding molecular motion through molecular dynamics models, thus also predicting macroscopic effects [170].

2.1.1 Potential energy function

Molecular mechanics methods characterize complex systems using a potential energy function (PEF). The PEF is comprised of a set of mathematical equations that describe distinct energy terms, providing a general physical description capable of predicting a particular state of the system of interest. The general PEF equation used by the GROMOS force field family is as follows:

$$\begin{aligned} V(\vec{\mathrm{r}}^N) = & \sum_{\mathrm{n=1}}^{\mathrm{N_b}} \frac{1}{2} K_{b_n} (b_n - b_{0_n})^2 \\ & + \sum_{\mathrm{n=1}}^{\mathrm{N_\theta}} \frac{1}{2} K_{\theta_n} (\theta_n - \theta_{0_n})^2 \\ & + \sum_{\mathrm{n=1}}^{\mathrm{N_\varphi}} K_{\varphi_n} [1 + cos(m_n \varphi_n - \delta_n)] \\ & + \sum_{\mathrm{n=1}}^{\mathrm{N_\xi}} \frac{1}{2} K_{\xi_n} (\xi_n - \xi_{0_n})^2 \\ & + \sum_{\mathrm{i=1}}^{\mathrm{N}} \sum_{\mathrm{j>1}}^{\mathrm{N}} \left(\frac{\mathrm{C12}_{ij}}{r_{ij}^{12}} - \frac{\mathrm{C6}_{ij}}{r_{ij}^{6}} \right) \\ & + \sum_{\mathrm{i=1}}^{\mathrm{N}} \sum_{\mathrm{j>1}}^{\mathrm{N}} \frac{q_i q_j}{4\pi\varepsilon_0 \varepsilon_r r_{ij}} \end{aligned} \tag{2.1}$$

The PEF is represented in equation 2.1 with each term depending on the vector position of all atoms in the system ($\vec{\mathrm{r}}^N$). It can be split in four terms that describe the bond related contributions and the last two terms represent the non-bonded contributions. The first term (equation 2.2) represents the stretching of all possible atom-pair bonds described with an harmonic potential:

$$\sum_{\mathrm{n=1}}^{\mathrm{N_b}} \frac{1}{2} K_{b_n} (b_n - b_{0_n})^2 \tag{2.2}$$

where the summation of all bond contributions depends on a force constant for each atom pair (K_{b_n}), their bond length (b_n) at a particular instant and their ideal bond length (b_{0_n}). The following term is the angle bend (equation 2.3) and it is similarly described with an harmonic potential, though as a function of the angle between three atoms:

$$\sum_{\mathrm{n=1}}^{\mathrm{N}_\theta} \frac{1}{2} K_{\theta_n} (\theta_n - \theta_{0_n})^2 \tag{2.3}$$

and, as previously, the K_{θ_0} is the force constant of the angle bend, θ_n and θ_{0_n} is the optimal angle. The following two terms represent the potential energy of the torsion of proper (equation 2.4) and improper (equation 2.5) dihedrals between four atoms. The proper dihedral term is expressed as a periodic cosine function described by the following equation:

$$\sum_{\mathrm{n=1}}^{\mathrm{n}_\varphi} K_{\varphi_n} [1 + cos(m_n \varphi_n - \delta_n)] \tag{2.4}$$

with the K_{φ_n} as the force constant of the dihedral angle torsion, φ_n as the proper dihedral angle and m_n as the multiplicity, the number of possible energy maximum or minimum in a 360° torsion rotation, and δ_n is the reference maximum that adopts either 0 or π. Meanwhile, the improper dihedral torsion is defined as follows:

$$\sum_{\mathrm{n=1}}^{\mathrm{N}_\xi} \frac{1}{2} K_{\xi_n} (\xi_n - \xi_{0_n})^2 \tag{2.5}$$

where K_{ξ_n} is the force constant of the torsion, ξ_n is the improper dihedral angle defined by the four atoms and ξ_{0_n} is the optimal torsion angle. Finally, the last PEF terms determine the nonbonded interactions comprised of van der Waals (equation 2.6) and Coulombic forces (equation 2.7). The van der Waals forces can be described as Lennard-Jones interactions as the following equation:

$$\sum_{\mathrm{i=1}}^{\mathrm{N}} \sum_{\mathrm{j>1}}^{\mathrm{N}} \left(\frac{\mathrm{C}_{12_{ij}}}{r_{ij}^{12}} - \frac{\mathrm{C}_{6_{ij}}}{r_{ij}^{6}} \right) \tag{2.6}$$

where r_{ij} comprises the distance between an atom pair ij, $C_{12_{ij}}$ is the repulsion force parameter of the atom pair and $C_{6_{ij}}$ is the attraction force parameter. At last, the electrostatics interactions can be described using Coulomb's law:

$$\sum_{\mathrm{i=1}}^{\mathrm{N}} \sum_{\mathrm{j>1}}^{\mathrm{N}} \frac{q_i q_j}{4\pi\varepsilon_0 \varepsilon_r r_{ij}} \tag{2.7}$$

where r_{ij} is the distance between the atom pair ij, q_i and q_j are the corresponding partial charges of each interacting atom, ε_0 and ε_r are the permittivity in vacuum and the relative dielectric constant of the medium, respectively.

2.1.2 Force Field and Compound Parametrization

In molecular simulations, the PEF characterizes the system of interest and each component is defined by the optimal bond, angle, dihedral, Lennard-Jones parameters and partial atomic charges that depend on the interacting atoms. The set of parameters and rules that solve the PEF is named the force field (FF). And these parameters are often obtained through empirical experimentation to reproduce experimental results, such as the free energy of hydration [171]. There are several types of FF's, though there are two classes worth mentioning: the all-atom and united atom force fields. The main distinction between both is the explicit treatment of apolar hydrogen atoms in the all-atoms force fields, like AMBER and CHARMM, while united-atom FF, such as GROMOS, conflate them with the heavy atoms to which they are bonded. In the case of a methyl group (-CH_3), the hydrogen atoms are omitted and the group is treated as a larger (increased Lennard-Jones parameters) and higher mass carbon atom. In this thesis, the used force field was the GROMOS 54A7 version [172], to which the previously described PEF is related.

As such, each distinct molecule is described as a unique combination of these rules. Additionally, all force fields provide a wide library of already parametrized molecules and chemical groups. For example, aminoacid residues and certain lipid parameters, such as 1-palmitoyl-2-oleoyl-*sn*-glycero-3- phosphocholine (POPC) and 1,2-dimyristoyl-*sn*-glycero-3-phosphocholine (DMPC), are already characterized within the GROMOS FF, however other fringe molecules require a parametrization procedure before they are included in a simulation. In this thesis, a two-pronged parametrization procedure was applied for the following set of Lewis base drugs (Figure 2.1): pimozide, ethopropazine, clomipramine, daunorubicin, nintedanib, sunitinib, and vincristine. First, initial topologies were generated for each drug compound using the automated topology builder (ATB) tool [173]. The ATB server is compatible with a variety of force fields and it generates specific descriptions for diverse types of molecules. The ATB tool protocol uses a multi-step approach based on quantum mechanical (QM) calculations combined with a knowledge-base approach using similar compounds to draw comparison and further optimize the generated parameters. Initially, the ATB protocol generated a base topology for each compound, based on the provided coordinate file (.pdb file). Each topology contains all the atom types, optimized geometries and partial charges. The following step consisted of an optimization protocol at a higher level of theory. Typically, ATB only allows the B3LYP/6-31G* level of theory for molecules with less than 40 atoms, since it requires high computational load from the web-server. So, due to the wide variety of sizes in our compound set and to keep the protocol consistent, a semi-empirical QM calculation in ATB was performed for all molecules regardless of molecule length. This first part of our parametrization procedure provided an initial topology of an optimized

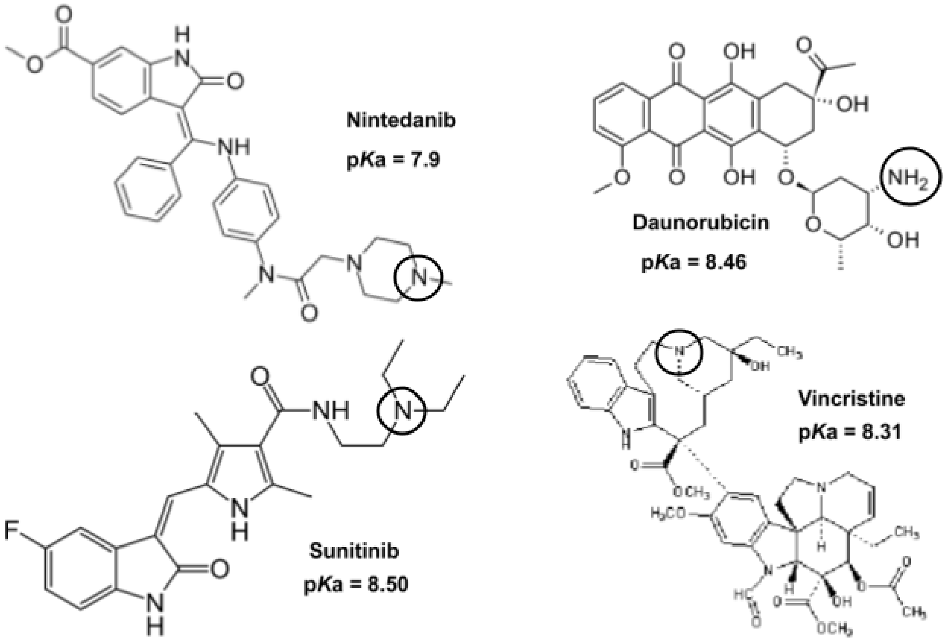


Figure 2.1: Molecular structure schemes of some Lewis Base drugs studied in Chapter 6. The remaining structures can be found in the Appendix for Chapter 6.

geometry with all force field parameters, Lennard-Jones and partial charges, according to the GROMOS rules, which was then further modified in the nonbonded pairs section to exclude all the 1-4 interactions from being considered. Still, the semi-empirical level of theory is insufficient for a proper estimation of the partial charges assigned to each atom. The second part of the parametrization procedure consisted of QM calculations using Gaussian 09 [174]. This procedure calculated the electrostatic potential upon each optimized structured in a two step process: first, with an optimization step using the B3LYP/6-31G* basis set [175] to generate an optimized geometry; then, an additional QM calculation on the optimized structure, using the B3LYP/6-31G functional set, to obtain the atomic partial charges. The estimated charges followed the Merz-Kollmann-Singh scheme [176] and they were manually curated to constrain the formal charges of small charge groups, as per GROMOS rules. To perform constant-pH MD (CpHMD) simulations (see Section 2.3), all titrating molecules require that protonated and deprotonated states of each molecule are included in the force field. Therefore, the partial charges were estimated for the neutral and charged states of all Lewis bases drugs.

2.1.3 Molecular Dynamics

If we wish to characterize a system using molecular mechanics/molecular dynamics (MM/MD) simulations, ideally, we want to accurately sample the most of the conformational space for every molecule. Molecular dynamics simulations achieve this by applying the Newton's laws of motion to every system particle, over a predetermined length of simulation time (ns or μs timescale), thus generating a trajectory. Starting from the MM step, we obtain the PEF that describes the energy of the system as the sum of all the composing terms using the FF of preference. Then, an additional step is required before we add motion to our simulations, as we need to determine the forces applied on each atom to then describe each particle vector of motion (equation 2.8).

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

where $\nabla_{\vec{r_i}}$ is the gradient vector of atom i at a given spatial position with a specific direction, V represents the potential energy function and $\vec{F_i}$ is the applied force on atom i. By computing all the particle forces, we can obtain a possible system state that was sampled from its conformational space.

After obtaining the forces applied on each atom, we are able to solve the Newton's second law of motion that will predict each particle sequence of positions and velocities, at each instant t, along the simulation time (equations 2.9) [177]:

$$\frac{\mathrm{d}^2\vec{\mathrm{r}}_\mathrm{i}(t)}{\mathrm{dt}^2} = \frac{\vec{F_i}(t)}{m_i} \tag{2.9}$$

where $\vec{\mathrm{r}}_\mathrm{i}$ and m_i are the coordinates and the mass of an atom i, respectively, at an instant t, while $\vec{F_i}$ is the force acting on atom i derived from the PEF. Afterward performing the calculations on all the atoms movements by using an integrator, a numerical method that generates solutions to the equations of motion(2.9), we generate new system configurations with updated positions, velocities and energies. In our GROMOS simulations, the more popular integrator is the leap-frog algorithm [178]. The leap-frog algorithm is one of the more simple, efficient and accurate methods to integrate the equations of motion. The algorithm takes the atom position $\vec{r}$ at a time t, and the atom velocity $\vec{v}$ at time $t - \Delta t/2$, then they are updated using the estimated forces $F(t)$ defined by the positions at time t, as shown in the following equations 2.10 and 2.11:

$$\vec{\mathrm{r}}_\mathrm{i}\,(t+\Delta t) = \vec{\mathrm{r}}_\mathrm{i}(t) + \vec{\mathrm{v}}_\mathrm{i}\left(t + \frac{\Delta t}{2}\right)\Delta t \tag{2.10}$$

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

With the Δt as the desired time step, then what the algorithm entails is that the $\vec{r}$ and $\vec{v}$ atom updates are out of phase by $\Delta t/2$. The choice of time step is a particularly important decision and one should ideally use the largest time step possible, since small values are not efficient to generate long simulation times. However, a value too large will incur on integrator errors, as the accuracy of the particle positions will decrease, leading to particle collisions and system instability. In our simulations, we typically use a 2 fs time step as it provides a good balance between integrator efficiency and accuracy.

Periodic Boundary Conditions

Molecular simulations try to predict macroscopic properties of biological systems using a representative small number of molecules, since perfect biological replicas are impossible to achieve due to several constraints, one of which is the system size. The larger the system is, the more computational load is required to compute all the necessary interactions, particularly with the solvent molecules. Most biomolecular simulations try to mimic the effect of an infinite dilution and/or an infinite membrane bilayer in membrane systems. Ideally, a near infinite number of water molecules or an infinitely-sized membrane patch would achieve this purpose, however there would be a prohibitive computational cost associated. Therefore, we impose simulation boxes around the system molecules and apply periodic boundary conditions (PBC). By taking this approach, we are treating the system as if surrounded by copies of the itself, as exemplified in Figure 2.2, so that when a given particle exits the simulation box from one side, an equivalent particle enters the same simulation box from the opposite side, while still carrying the same vector of movement. The advantage of this method is that precludes large solvated

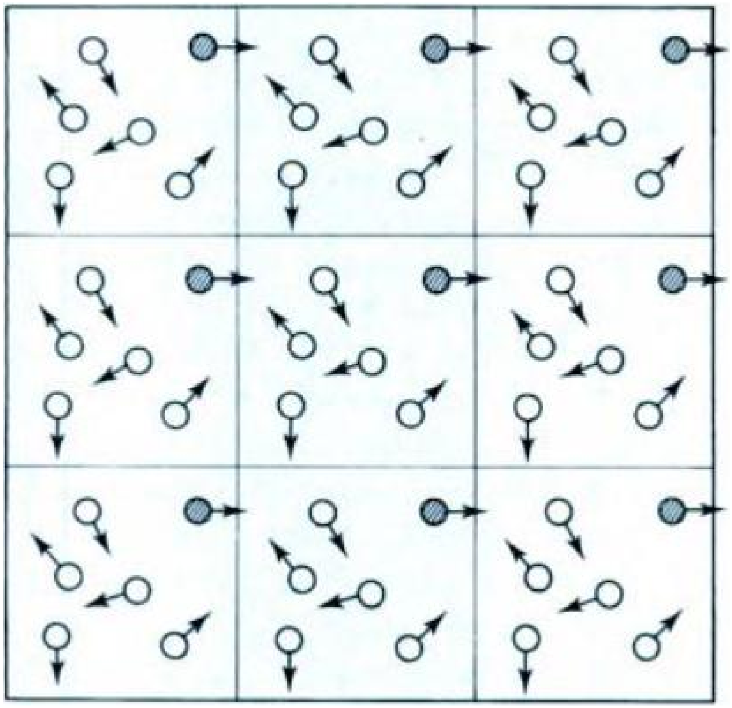


Figure 2.2: Periodic Boundary Conditions set for a two dimensional plane. Each cell represents a copy of the original simulation box. All particles are represented as circles, with the empty circles movement occurring within that particular cell, while full circles transpose the boundary between adjacent cells. Adapted from ref.[169].

simulation boxes, there is a constant number of particles in the box, and the particles experience a similar effect of a bulk solvent. Furthermore, there are several box shapes to be chosen (cubic, hexagonal prism, truncated octahedron, rombic dodecahedron and elongated dodecahedron) that affect the volume to be filled with solvent molecules, hence also affecting the simulation time [169]. Depending on the type of system (globular protein, membrane system, etc..), there are optimum choices to minimize the number of solvent molecules and interactions to be simulated and prioritize the computational resources to the system of interest. Although too small simulation boxes create edge effects, where the molecule distance to the box edge is too short to prevent it from interacting with its image, specially non-bonded interactions. If molecules begin to interact with their image counterparts, the infinite dilution effect is broken. Having successfully chosen an adequate box size and shape, the PBC can be applied in a 3D array or a 2D plane, which is particularly helpful regarding biomolecular simulations of membrane models since, by applying PBC in an xy plane and the infinite dilution on the z axis, we are able to describe a similar environment of a large membrane bilayer.

Treatment of non-bonded interactions

The PBC are relevant in preserving simulation integrity on the infinite effects and to reduce computational load by minimizing the number of possible solvent interactions. Moreover, the non-bonded interactions are the crux of the matter. When applying a PBC, the number of long range interactions, in a near infinite system, would also be virtually infinite, which creates a computational issue when integrating the non-bonded interactions (van der Waals and Coulombic) of all possible atom pairs in the system (see last two terms of PEF in eq. 2.1). The simplest method to circumvent this issue is to apply a distance cutoff around each particle and truncate the interaction calculations beyond that chosen distance. This particular method works well for the Lennard-Jones interactions as, at short inter-atomic distances, they decay to zero faster ($\frac{1}{r^6}$) than Coulombic interactions($\frac{1}{r}$) [170] (see denominators of eq. 2.6and 2.7). Nevertheless, there is still a considerable amount of interactions to be computed within a single cutoff and, historically, stable multi-time-step approaches, like the twin-range method, were used to speed up the simulation by reducing the number of interaction calculations at longer inter-atomic distances. The twin-range method partitions the interactions into two groups defined by two cutoff values: a smaller radius (ex: 0.8 nm) is applied, where all the interactions are updated at a higher frequency (ex: 2 integrator steps), and a larger radius (ex: 1.4 nm) with a smaller update frequency(ex: 5 integrator steps). Yet, recent hardware and software developments led to the deprecation of this scheme [179], particularly in GROMOS simulations, since some associated artifacts have been identified in recent versions of GROMACS and the current computational power also precludes such tricks without significant loss in simulation accuracy and time [180, 181].

Meanwhile, the treatment of electrostatic interactions has been a source of discussion for years, mainly focused on two methodological camps: the reaction field (RF) [182] and the particle mesh Ewald (PME) [183]. The theoretical discussion is long, complex and markedly outside of the scope of this thesis, so we will focus on the used methodology. The RF method central axiom is that, beyond an established cutoff value, homogeneous parts of the system, like the bulk solvent, are akin to a continuous medium, thus they can be modeled as a dielectric medium with a specific dielectric constant [184]. Thus, within the cutoff, electrostatic interactions are calculated explicitly in a medium with dielectric constant (ε_1), while outside electrostatic interactions are encapsulated in an external dielectric constant (ε_2) that has an attenuating effect upon all electrostatic interactions within the cutoff region. Additionally, the generalized reaction (GRF) method takes into account the effects of ionic strength I in the characterization of the continuous medium, that further modulates the interactions attenuating effect.

2.1.4 Energy Minimization

Generally, in MD simulations of biological systems, we want to sample the conformations in a thermodynamic equilibrium, which typically relate to the energy minimum or its surroundings, to predict macroscopic properties.

However, often, the starting configuration of a system is not equilibrated and exhibits atomic/molecular clashes that lead to non-physical descriptions of our system and, consequently, erroneous property predictions. To avoid this, it is necessary to begin our simulations to equilibration and near an energy minimum. We achieve that by travelling through the potential energy surface of the system until we find a stable energy minimum. The potential energy surface is a complex, multidimensional energy surface that depends on the coordinates of all the system particles, i.e. a system configuration, hence it can be treated as a mathematical function with $3n$ dimensions (n is the number of atoms). In biomolecular systems, the number of possible configurations is astronomical, with multiple meta-stable states (energy minima), therefore it is near impossible to determine the absolute minimum. Still, we are able to identify many stable states using energy minimization procedures, which are numerical methods that iteratively change the molecular coordinates to generate new conformations with progressively lower energies until a minimum is reached [169]. In this work, all minimization protocols used one or both of the following methods: the steepest descent and the limited memory Broyden-Fletcher-Goldfarb-Shanno (l-BFGS) method. The steepest descent is one of the more widely used methods of multi-variable function minimization and the algorithm is very robust and easy to implement [177]. As previously defined, in a space of $3n$ dimensions, we can define a vector $\vec{r}$ for all coordinates. The first step calculates the forces $\vec{F}$ and the potential energy, then new positions are obtained with the following equation:

$$\mathrm{r}\vec{_{n+1}} = \vec{\mathrm{r}_n} + \frac{\vec{F_n}}{\max(|\vec{F_n}|)} h_n \tag{2.12}$$

where h_n is the maximum displacement and $\vec{F_n}$ is the applied force, as obtained in equation 2.8 using the potential energy of the system (V). Meanwhile, $\max(|\vec{F_n}|)$ refers to the largest of the absolute values of the force components. In the next step, the forces and energy are once again computed to generate new positions that are accepted if $V_n + 1 < V_n$ and the $h_{n+1} = 1.2h_n$, otherwise the positions are rejected and $h_{n+1} = 0.2h_n$. The algorithm will attempt to generate new positions until one of two conditions is satisfied: either it reaches the user-defined number of steps for the cycle; or once the minimum energy threshold is reached and the energy minimum has converged to a smaller value. The previous description is the implementation done in the GROMACS package that is used in this thesis [177].

Concerning the l-BFGS methodology, the implementation and description is more complex, using the inverse Hessian matrix [185] to move the system to currently generated energy minimum and then, through approximations to the matrix that, at every step, continuously corrects the stored information to search for equivalent local energy minima. This methodology is more computationally intensive than the steepest descent, due to storing in the computer memory the total number of particles multiplied by the correction steps. Nevertheless, this method is very useful in jumping between alternative energy minima in complex systems to optimize the energy minimization process.

2.1.5 Initialization

After a successful minimization step, the system is in an stable state at an energy minimum, yet it remains frozen in space (all velocities are zero). To begin a production segment of MD at the desired conditions, the system needs to be carefully brought to an equilibrium state at the chosen temperature and pressure conditions. Usually, we perform a multi-step initialization procedure that aims at equilibrating each of these properties. First, initial random velocities are assigned to every system particle using a randomly generated seed at the chosen temperature. To prevent simulation issues such as particle clashes, energy instability and artifacts, the first step employs position restraints upon the solute particle movement. The position restraint is an additional term to the PEF (equation 2.1) that limits the possible movement of atom i from a reference position, through an artificial force applied to that atom. In the initialization process, the first step applies a large restraining force that, in subsequent steps, becomes progressively weaker until the system becomes unrestrained.

In this work, the MD simulations were performed in the NPT ensemble (isothermal-isobaric ensemble), which means that the number of particles, temperature and pressure were kept constant throughout our simulations. To sample from this ensemble, during the initialization steps, it is necessary to equilibrate the system temperature and pressure by coupling them to a thermostat and barostat algorithm, respectively.

Temperature coupling

The thermostat algorithm operates as an external heat bath that provides or extracts kinetic energy to the system. There are several thermostats available, like the Nosé-Hoover [186, 187], Berendsen [188], and particularly the v-rescale [189] thermostat that was used throughout this work. The last two are velocity-rescaling methods which ensure that the average kinetic energy of the system is similar to the expected energy at the constant user-defined temperature of the bath. Therefore, there is a difference between the bath and system temperatures that needs to be iteratively adjusted to maintain the average temperature constant:

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

where τ is the temperature coupling constant the defines how tight the bath and system temperatures are coupled. The previous equation shows that the system temperature will progressively converge to the bath temperature at every time step. To assess how much it needs to rescale from one step to the next, the temperature difference between two steps needs to be defined as:

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

Then, to convert these temperature adjustments to practical velocity rescaling, a scaling factor (λ) is required:

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

Overall, the scaling factor will be higher or lower depending on the chosen τ. Smaller τ values are equivalent to tighter coupling and more responsive temperature adjustments, yet it may create more kinetic instabilities. Higher τ values correspond to a looser coupling, which converges more slowly, yet smoothly, to the desired temperature. In the end, we obtain a system that is sampling from the NVT ensemble (constant number of particles, volume and temperature), though not yet from the desired NPT ensemble.

Pressure coupling

The pressure coupling philosophy is similar to the temperature algorithm, since the system is also coupled to an external bath (barostat) and the pressure is kept constant by adjusting the simulation box volume depending on a pressure coupling constant. The frequency of volume adjustments is determined by:

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

where τ_P is the pressure coupling constant, similar to the temperature τ, P_{bath} is the user-defined pressure and $P(t)$ is the system pressure at a time t. Likewise, the simulation box is scaled to adjust the system pressure by a scaling factor (μ):

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

where β is the isothermal compressibility of the system. Similarly to the temperature, smaller values of τ_P produce a stronger pressure coupling and more frequent adjustments. In membrane systems, we need to allow realistic oscillations in xy plane area, therefore, we use a semi-isotropic coupling, where the xy plane and z axis are independently being coupled to the barostat, and larger τ_P values to allow for membrane undulations.

2.2 Continuum electrostatics

In large complex biological systems, the computational cost is drastically increased due to the sheer number of molecules and interactions, even more so if distinct electrostatic descriptions of multiple system conformations are necessary. This is the case of CpHMD simulations, where there are multiple charge and electrostatic variations as a result of the (de)protonation events of titrating residues, at a given pH, along the simulation (see Section 2.3). To address these cases, the concept of continuum electrostatics can be extended to the solute and the solvent using the Poisson–Boltzmann (PB) equation to calculate the protonation energies that are then used to sample the protonation states using a Monte Carlo method (see Section 2.2.3).

2.2.1 Poisson–Boltzmann method

Generally, the implementation of a PB method is based on the assumption that the solute is fixed in a rigid structure with low polarizability within a high polarizability medium (solvent). The polarizability is a reflection of the relative dielectric constant of the medium, where higher dielectric values represent higher polarizability. The dielectric constant of water (ε_{out}) is usually $\sim$80 (high polarizability), therefore it can quickly reorganize in response to an electric field. In practical terms, this entails that water molecules quickly adopt new conformations around point charges. On the other hand, the solute possesses a low dielectric constant (ε_{in}), due to the more rigid nature of a protein or membrane structure inducing a slow conformational reorganization to a given charge. In the same vein as the GRF method, the ionic strength is treated implicitly to account for the attenuation of electrostatic interactions between point charges. Within our model, the electrostatic potential can be calculated using the linearized form of the PB equation:

$$\nabla \cdot [\varepsilon(\vec{\mathrm{r}})\nabla\phi(\vec{\mathrm{r}})] - \kappa^2(\vec{\mathrm{r}})\varepsilon(\vec{\mathrm{r}})\phi(\vec{\mathrm{r}}) = -4\pi\rho(\vec{\mathrm{r}}) \tag{2.18}$$

where ∇ is the gradient vector, $\varepsilon(\vec{r})$ is the dielectric constant at all atom positions, $\nabla\phi(\vec{r})$ is the gradient of electrostatic potential measured in the solute and solvent. $\rho(\vec{r})$ is the charge density distribution of our model. The distribution depicts the fixed point charges that represent the atoms spatial position of charged aminoacid residues in the protein, dipoles from the protein backbone or side-chains of uncharged aminoacid residues. Finally, $\kappa(\vec{r})$ is the Debye length that provides the ionic strength contribution given by:

$$\kappa(\vec{\mathrm{r}}) = \begin{cases} \left(\frac{8\pi e^2 I}{\varepsilon_{out} k_B T}\right)^{1/2}, & \text{if } \vec{r} \text{ is in an ion accessible region} \\ 0, & \text{otherwise} \end{cases} \tag{2.19}$$

where k_B is the Boltzmann constant, T is the absolute temperature and I is the ionic strength of the solvent. This constant adds a non-zero contribution to the screening effect on the charge interactions if the atoms are within an ion accessible region, like in the solvent, otherwise the contribution is 0, such as in the protein or membrane interior. Generally, the PB equation needs to be solved using a numerical method, like the finite difference procedure used in DelPhi [190, 191]. This method replaces the linear equation with a system of linear equations derived from a discrete mapping of the system using three dimensional grids. An initial coarser grid is placed upon the system, with each solute and solvent atom centered on a grid cell and their properties are mapped, like the charge, electrostatic potential, dielectric constant. This provides a general description of the system electrostatic potential. Afterward, a finer grid is iteratively focused on all sites of interest (titrating residue, catalytic site) to obtain a more refined electrostatic potential calculation until convergence is met. In the end, this method provides the necessary electrostatic potential energy of our system (equation 2.20) and, for specific sites of interest, we obtain the measured charge distribution of N point charges (q_i), at positions ($\vec{r}_i$) from an hypothetical zero charge distribution:

$$\nu = \frac{1}{2}\sum_{i}^{N} q_i \phi(\vec{r}_i) \tag{2.20}$$

2.2.2 Protonation free energy

Once the electrostatic potential energy is obtained, one of the possible applications is the calculation of the protonation free energy. The protonation free energy is the required energy to transition from a reference protonation state to a different state. If we consider a single titrable residue (R) in a given medium M environment (protein or membrane), we can assume that the reference state is the neutral state (0) and it changes to a protonated state (p). Then, the protonation free energy associated with the protonation change of R is given by:

$$\Delta \mathrm{G}^{M}_{R_0 \rightarrow R_p} = \Delta \mathrm{G}^{sol}_{R_0 \rightarrow R_p} + \Delta \mathrm{G}_{sol \rightarrow M}(R_p) - \Delta \mathrm{G}_{sol \rightarrow M}(R_0) \tag{2.21}$$

where $\mathrm{G}_{sol \rightarrow M}(R_0)$ and $\mathrm{G}_{sol \rightarrow M}(R_p)$ are the free energy differences from changing the reference and R_p states from a solvent environment to that particular medium, respectively. The $\Delta \mathrm{G}^{sol}_{R_0 \rightarrow R_p}$ term is the free energy of protonation calculated for a model compound. A model compound is defined by a restrict group of atoms of the titrable moiety that has a unique pK_a value (pK^{mod}) experimentally measured in water [192]. We can derive the free energy of protonation in the solvent ($\Delta \mathrm{G}^{sol}_{R_0 \rightarrow R_p}$) using the relation of residue pK_a with the free energy of protonation:

$$\mathrm{p}K = \frac{\Delta \mathrm{G}}{2.3 k_B T} \tag{2.22}$$

where ΔG can be expressed with the electrostatic potential energy (equation 2.20) as:

$$\Delta \mathrm{G} = \frac{1}{2}\sum q_i \phi_i^r \tag{2.23}$$

However, a titrating residue is often surrounded by other titrable sites and background interactions that affect the pK_a. Hence, these interactions need to be taken into account to determine the pK_a value of a particular site R in a medium M:

$$\mathrm{p}K_\mathrm{a}(R^M) = \mathrm{p}K^{int}(R) - \frac{1}{2.3 k_B T}(\Delta\Delta \mathrm{G}_{solv} + \Delta\Delta \mathrm{G}_{background}) \tag{2.24}$$

where the p$K^{int}(R^M)$ is the intrinsic pK_a of R defined by the pK^{mod} and all the other titrating site interactions in their reference state. Yet, this equation would be pH-independent if we only considered all the other titrating sites in their reference state. The inclusion of a $\Delta\Delta \mathrm{G}_{solv}$ and $\Delta\Delta \mathrm{G}_{background}$ provides the difference in solvation free energies and the difference in background energy contributions, respectively. Therefore, the final protonation free energy for a single titrating residue can be described as:

$$\Delta \mathrm{G}^{\mathrm{M}}_{0 \rightarrow \mathrm{p}}(R) = -2.3 k_B T \mathrm{p}K^{int} + (\Delta\Delta \mathrm{G}_{solv} + \Delta\Delta \mathrm{G}_{background}) \tag{2.25}$$

In the case of a protein with multiple titrating sites, we can consider a given system state with N protonation states given by $\vec{a} = (a_1, a_2, ..., a_N)$, where a_i =0 or 1 indicates that site i is either neutral or ionized, respectively. Then the protonation free energy required to move from the reference state ($\vec{0}$), where all sites are neutral, to state $\vec{a}$ in a given medium M (protein or membrane) is given by:

$$\Delta\mathrm{G}^{\mathrm{M}}_{\tilde{0}\rightarrow\tilde{\mathrm{a}}} = -2.3k_BT\sum_{i}^{N}a_i\gamma_i\mathrm{p}K_i^{int} + \sum_{i}^{N}\sum_{j\neq i}^{N}a_ia_j\Delta W_{ij} \tag{2.26}$$

where ΔW_{ij} is the interaction free energy between two ionized sites i and j, γ_i is the charge of a site i when ionized. Furthermore, these equation allows the inclusion of proton tautomerism as referenced in refs. [193, 194]

2.2.3 Sampling of protonation states

After obtaining the protonation free energies using equation 2.26, we need to ascertain what protonation state to assign to the titrating groups of our molecule of interest. For instance, one could calculate the probability of protonation for all n titrating sites at a given conformation, however that would be computationally costly as, for large systems, the large number of titrable residues would render impossible these calculations for all possible 2^n protonation state combinations, even without considering protonation tautomerism. For that purpose, we perform Monte Carlo (MC) simulations to sample for the possible protonation states. A MC simulation has no concept of a trajectory, thus differing from a MD simulation where time connects all system configurations. In a MC simulation, random configurations are generated, from an initial protonation state, through a stochastic process. Then, using a Metropolis criteria [195], the protonation states are either accepted or rejected according to a given probability [193].

First, we need to define the pH-dependent Gibbs free energy of transitioning from state $\vec{0}$ to state $\vec{a}$:

$$\Delta\mathrm{G}^{M}_{\vec{0}\rightarrow\vec{a}}(\mathrm{pH}) = -2.3k_BT\sum_{i}^{N}a_i\gamma_i(\mathrm{p}K_i^{int} - \mathrm{pH}) + \sum_{i}^{N}\sum_{i\neq j}^{N}a_ia_j\Delta W_{ij} \tag{2.27}$$

then, we need to describe free energy associated with changing the protonation states from a configuration $\vec{a_1}$ to a configuration $\vec{a_2}$:

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

The MC move is the transition from one configuration to the other, and the Metropolis criterion states that configuration $\vec{p_2}$ is accepted if:

$$\Delta\mathrm{G}_{\vec{a_1}\rightarrow\vec{a_2}}(\mathrm{pH}) < 0 \tag{2.29}$$

while, if the transition is not energetically favorable - $\Delta\mathrm{G}_{\vec{a_1}\rightarrow\vec{a_2}}(\mathrm{pH}) > 0$ - then it is accepted with a given probability of:

$$\mathrm{p}(\vec{a_1}\rightarrow\vec{a_2}) = e^{-\Delta\mathrm{G}_{\vec{a_1}\rightarrow\vec{a_2}}(\mathrm{pH})/RT} \tag{2.30}$$

Nevertheless, a MC move usually is related to a single site protonating or deprotonating. Yet, this approach is insufficient to sample strongly interacting sites, due to their heavily coupled electrostatic influence on each other. A normal MC simulation only changes one protonation state at the time, which would then create energy barriers between these states and poorly sampling the protonation states. For this purpose, the concept of concomitant protonation state changes between tightly coupled sites was implemented [196]. So, a MC simulation will perform a user-defined *n* number of cycles to generate an equilibrated sampling of protonation states for each titrating group. In the case of the CpHMD method (see section 2.3), the last sampled protonation state is the one assigned to each group. With the protonation states successfully sampled, it is possible to calculate pH-dependent properties, like individual or global titration curves of the biomolecule, single pK_a values of residues of interest, pK_a profiles along a given property (Chapter 3) or even pH-dependent permeabilities (Chapter 6).

2.3 Stochastic Constant-pH Molecular Dynamics Method

In the field of biomolecular simulations, the popularity of computational methods that integrate pH effects and predict pK_a values have been increasing in the last few years with the emergence of novel and user-friendly computational implementations [197, 198]. These implementations encompass different types of pK_a predictors, like the rigid-based methods that perform PB/MC calculations on single system configurations (H++[199], DelPhiPKA[200], MCCE[201], or PypKa [197]). Usually, these methods are very fast and easy to use, yet their accuracy can be found lacking if the system is too complex with a diverse conformational space or if the known system structures are not representative of the desired biomolecule. This rigidity precludes a natural conformational reorganization that would follow a change in protonation. Although changing (increasing) the dielectric constant attenuates the lack of this effect, more often than not buried and strongly interacting residues pK_a's are poorly predicted. Other methods aim to be more robust and accurate by coupling both protonation and conformational sampling to obtain a realistic description of a biological system. These methods are named constant-pH molecular dynamics (CpHMD) methods and they can be discerned in discrete (either protonated or deprotonated) [151–155] and continuous (fraction protonation) methods [147–150]. As mentioned in section 1.5, this work used an extension of the stochastic CpHMD method [151, 156, 159] that includes membrane systems (CpHMD-L) [27, 157, 158].

This CpHMD-L methodology merges the ability to sample discrete protonation states dependent on external pH, using the previously described PB/MC approach, with the robust conformational sampling derived from MM/MD simulations. By working in tandem, the method incorporates the pH as another user-defined thermodynamic parameter like the temperature or pressure. The CpHMD method functions in stop-and-go cycles of three major steps, at a user-defined pH value, until the end of the defined simulation time (Figure 2.3): a PB/MC step; a short solvent relaxation step; and a system MM/MD step. Considering the first iteration of the cycle, an initialized and equilibrated structure, with the titrating residues identified, is submitted for a PB/MC step. As previously explained in section 2.2, these calculations determine the titrating residues protonation states using the provided rigid structure. The assigned protonation states result from the last MC step. Then, the same conformation with new protonation states is the starting structure for MM/MD. An initial short MM/MD simulation aims at "relaxing" the solvent around the solute, which means the interacting water molecules are allowed to reorganize themselves according to the new solute protonation states. During this step, the solute coordinates are fixed to prevent movement and energy instabilities, therefore facilitating the solvent adaptation. Finally, a third step consists of an unrestrained MM/MD simulation of the whole system. New conformations are sampled from the conformational space of the system and the final generated conformation will be used to begin the new CpHMD cycle. The CpHMD simulation ends when it reaches the user-defined number of cycles necessary to complete the simulation time. When studying a biomolecule to predict certain pH-dependent properties, multiple simulations at different pH values are necessary to explore the most of the conformational and protonation spaces.

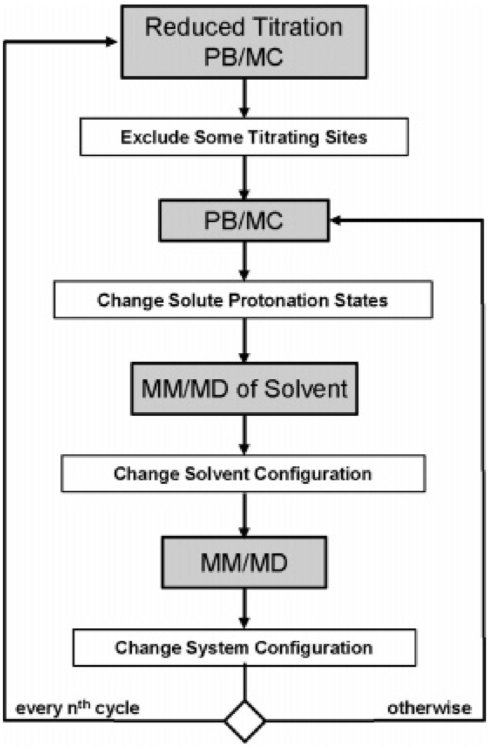


Figure 2.3: Workflow diagram of the standard stochastic CpHMD methodology and the reduced titration scheme. This scheme allows the exclusion of titrating sites, in particularly large systems, whose protonation states are mostly conserved, thus reducing computational cost. The protonation states are recalculated only every nth cycles. Adapted from reference [156].

2.4 Enhanced sampling techniques

The CpHMD methodology provides highly accurate and realistic descriptions of pH effects in molecular systems, due to robust protonation and conformational sampling. However, in large and/or complex systems with a rich and diverse conformational/protonation space, like transmembranar protein systems, these simulations still suffer from oversampling of meta-stable states as often occurs in other MM/MD-based methodologies. The initial structure setup might be stuck in non-relevant energy minima with high energy barriers kinetically trapping the system. This issue is often mitigated by performing multiple simulation replicates from distinct system conformations or extending the simulation time until the system escapes those energy minima, yet computational power and time remains a limiting step to obtain a full description and, more importantly, system convergence. Additionally, when we take into account that pH-dependent systems often have large conformational transitions tied to rare (de)protonation events in the ns-μs timescale, like the exposure of buried titrating residues coupled with major conformational movements, these sampling limitations also extend to the protonation sampling efficiency, causing these hurdles to be even more overwhelming.

When facing these sampling challenges, there are several possible solutions with some being more brute-force approaches, like using more highly optimized and powerful hardware [202], while others are more elegant, as the implementation of enhanced sampling techniques. A multitude of enhanced sampling techniques, like the replica-exchange [167], umbrella sampling [168], metadynamics [203] and others have been devised to improve the free energy landscape sampling, particularly for MD simulations [204, 205]. Yet the scope of this thesis is on CpHMD methods, so our focus will be in briefly reviewing particular techniques that have already been coupled with the stochastic CpHMD-L, thus improving the protonation convergence in tandem with the conformational sampling.

2.4.1 The pH-replica exchange method

The first enhanced sampling method developed for the stochastic titration CpHMD-L method was the pH-replica exchange (pHRE) scheme [166], a derivation of the pH-replica exchange scheme presented by Itoh et. al. [206]. The pHRE methodology operates in the same essence as the temperature REMD [167], where higher energy states, from higher temperature replicas, are exchanged with lower states to obtain more efficient sampling of the free energy landscape. The foundation of the pHRE method is the CpHMD-L methodology, that was expanded in the previous section 2.3, as multiple simultaneous CpHMD-L simulations are run at different pH values, which are aptly named pH replicas. If we consider a given system X of which we have two replicas (i and j) at different protonation states (l and m, respectively), then the probabilities of exchanging one state with the other and vice-verse, with respect of the detailed balance condition, are described as:

$$P(X_i^l)P(X_j^m)w(X_i^l,X_j^m \rightarrow X_i^m,X_j^l) = P(X_i^m)P(X_j^l)w(X_i^m,X_j^l \rightarrow X_i^l,X_j^m) \tag{2.31}$$

where $P(X_i^l)$ denotes the equilibrium probability of X state at a pH$_i$ with protonation state l, with a particular coordinate (q_i) and momentum (p_i) vectors. $w(X_i^l,X_j^m \rightarrow X_i^m,X_j^l)$ is the free energy associated with changing from state X_i^m to the X_j^l state, in other words, the transition probability between states. Further derivation of this equation requires a proper definition of the equilibrium probability of a given state [207]:

$$P(X_i^l) = \frac{1}{\Xi} exp\Big(-\beta(K(p_i,N_i^P)+V(q_i,N_i^P)+k_BT\text{pH}_\mathrm{l} ln10N_i^P)\Big) \tag{2.32}$$

where Ξ is the grand canonical partition function, k_B is the Boltzmann constant, T is the temperature, K is the kinetic energy (which in the CpHMD is not dependent on the protonation state, since the titrable protons are always assigned either with or without charge), V is the potential energy, N_i^P is the number of protonated titrable residues in state X_i. Now, if we derive equation 2.31 with the previous definition for each state, we obtain:

$$\frac{w(X_i^l, X_j^m \rightarrow X_i^m, X_j^l)}{w(X_i^m, X_j^l \rightarrow X_i^l, X_j^m)} = exp(-ln10(\mathrm{pH_m} - \mathrm{pH_l})(N_i^p - N_j^p)) \tag{2.33}$$

then the exchange probability between two replicas is given by:

$$w(X_i^l, X_j^m \rightarrow X_i^m, X_j^l) = min\Big\{1, exp(-ln10(\mathrm{pH_m} - \mathrm{pH_l})(N_i^p - N_j^p))\Big\} \tag{2.34}$$

which defines the Metropolis criterion for exchange acceptance [195]. The criterion only depends on the interval between the exchanging pH replicas and the variation in the number of protonated groups. In our approach, all replica exchange attempts were done during the MM/MD production run, at frequent τ_{RE} intervals that are out of phase with the CpHMD cycle (τ_{prot}) as to keep the replica-exchange and CpHMD Markov chains uncoupled. Additionally, the exchange attempts occur between adjacent pH replicas to increase the probabilities of acceptance, ideally above an acceptance ratio of 20% [206]. Large discrepancies of these parameters between replicas will exponentially decrease the efficiency of the method.

Besides the previously detailed exchange scheme, there are two other pHRE exchange schemes: (1) between both the pH and the protonation states of the replicas; (2) or only the protonation states of the replicas. A discussion in detail can be found in ref. [166], yet one should note that both of these schemes present significant disadvantages, where the acceptance probability of scheme 1 depends on the potential energy difference between two distinct conformations, hence how well two distinct conformations can adopt different protonation states, which typically renders high energy variations, thus lower probabilities; and scheme 2 compiles all the previously referenced dependencies (pH, number of protons and potential energy differences) drastically hindering the method efficiency.

The present pHRE method was initially tested and optimized in a ethylenediamine (EDA) and lysozyme (HEWL) system exhibiting remarkable results in improved efficiency, high accuracy and decreased simulation and computational time compared with the standard stochastic CpHMD-L method (see Figure 1.19 in section 1.5). Still, if provided long enough simulation time, both methodologies sample the same conformational and protonation space, generating similarly accurate results [166]. Although this is a testament to the reliability of the pHRE in reproducing CpHMD-L results, the question remains if the pHRE scheme alone is enough to overcome some of the sampling issues that are present in complex membrane systems. An application of this method to such systems and the answer to this question can be found in chapter 3.

2.4.2 The umbrella sampling constant-pH MD method

Indeed, the pHRE method achieves faster convergence of the protonation equilibria which, in turn, allows longer simulation times as a possible approach to sample pH-dependent rare events. However, in other cases, the method is still limited by the conformational sampling obtained by conventional MD, as the timescale of such events are still out of reach, due to associated large energy barriers. This thesis will focus on a particular process that requires sampling of such events: the membrane crossing of Lewis base drugs. A more intricate description of the issue is presented in chapter 6, though the main issue at hand concerns to the very stable thermodynamic equilibrium reached by these drugs at the water-membrane interface. Sampling of deeper and higher energy conformations is hindered by large energy barriers, thus hampering a proper description of the pH-dependent membrane crossing process, even with the use of pHRE [37]. A novel US-CpHMD protocol that combines an

Umbrella Sampling scheme (US) with our implementation of the stochastic CpHMD method [163] was employed to obtain a more complete description of these drugs membrane crossing process. In theory, the US scheme will overcome the conformational sampling barriers of the system, allowing the CpHMD-L method to sample the protonation states of these new configurations. The protocol consists of two steps: a first step that uses a steered MD protocol to generate conformations across the membrane; and a second step using the US-CpHMD to describe the system.

Umbrella sampling theory

To obtain a more comprehensive understanding of how the US-CpHMD generates conformational sampling, we need to take a look at the theory behind the US and steered MD methods. Both these methodologies are types of enhanced sampling techniques that apply a variable biasing/restraining force to an initial system state. However, the steered MD steers it along a given collective variable (CV) until it reaches another state, while the US method restrains the system to a particular CV value. A CV is a one or multi-dimensional function of coordinates that adopts distinct values for each system meta-stable state. For example, a CV can defined by the degrees of freedom of a residue side-chain [208], the length of a specific molecule bond, local membrane defects, or the insertion level of a molecule entering a protein channel [163] or a membrane bilayer [209–211]. Usually, the choice of a CV is the defining step for both steered and US methods. Yet, more often than not, it is done without prior knowledge, basing solely on chemical or biological intuition to choose an adequate CV to describe the process. However, an unoptimized choice will not capture all the degrees of freedom of the process, undermining the simulation efficiency and, ultimately, the accuracy of the predictions [211, 212].

The theoretical foundation behind both techniques relies on a CV-dependent bias, an additional energy term to the system potential energy function (PEF), to enforce efficient sampling along a chosen CV (X). The sampling is performed across multiple energetically distinct regions of the free energy space named windows and this bias is applied to each window. Therefore, the potential energy of the biased system (V^b) can be described as:

$$V^b(\vec{r}) = V^u(\vec{r}) + w_i(X) \tag{2.35}$$

where $\vec{r}$ describes the position of all atoms, V^u is the unbiased potential energy of the system and w_i is the harmonic bias applied to a particular window i. The w_i is an harmonic bias potential that pulls the molecule along X or to X_i for steered MD and US, respectively. This can be given by:

$$w_i(X) = \frac{K}{2}(X - X_i)^2 \tag{2.36}$$

With K as the strength of the applied bias, where higher values increase the associated penalty cost of diverging movements from X. One should note that an energy penalty does not preclude the group from moving away from the CV. If the free energy of the diverging movement exceeds the penalty cost, then the system will stray from the reference. This is particularly useful to identify very energetically favourable interactions. Although the biased simulation describes the whole process along the CV, a realistic description is meaningless without the unbiased free energies. The unbiased probability distribution of the system (P^u) along X is given by:

$$P_i^u(X) = \frac{\int exp(-\beta V(\vec{r}))\delta(X'(\vec{r}) - X)d^N\vec{r}}{\int exp(-\beta V(\vec{r}))d^N_{\vec{r}}} \tag{2.37}$$

This equation gives the probability of finding a given state of the system with a particular CV ($X'(\vec{r})$) dependent on the defined range of CV values ($\delta(X'(\vec{r}) - X)$) without applying a bias, where $d^N\vec{r}$ refers to the N degrees of freedom of the system. The biased distribution (P_i^b) can be given by:

$$P_i^b(X) = \frac{\int exp(-\beta[V(\vec{r}) + w_i(X'(\vec{r}))])\delta(X'(\vec{r}) - X)d^N\vec{r}}{\int exp(-\beta[V(\vec{r}) + w_i(X'(\vec{r}))])d^N\vec{r}} \tag{2.38}$$

with the bias w_i dependent only on X, thus allowing further derivation as shown in ref [168, 213] until the unbiased free energy (M) for a window i, named the potential of mean force (PMF), is described as:

$$M(X) = -\frac{1}{\beta} ln P_i^b(X) - w_i(X) + F_i \tag{2.39}$$

where F_i is the necessary unbiasing force, at window i, to obtain the unbiased energy. By integrating the PMFs of multiple windows, using a weighted histogram analysis method (WHAM), the individual values of F_i are estimated, thus generating the global PMF profile, as shown in Fig. 2.4. An usual quality measurement for these simulations is the overlap of each window biased distribution with the adjacent windows distributions along the CV, otherwise the ergodicity of the simulation is not respected.

Details of Steered MD

If we consider our membrane bilayer system with a Lewis base drug placed in the bulk solvent, then the characterization of the internalization process uses the z axis as the CV, while defining the pull vector using the geometric center (COG) of two group of atoms, i.e. the titrable amine (or another group moiety), and the membrane center, considering all the methyl groups of the lipid acyl chains. Then, the molecule is pulled along the direction of the vector with a constant force (kJ/mol·nm^2) through the solvent and lipid bilayer with a given rate of nm·ns^{-1}. Through this procedure, multiple generated conformations, at different insertion levels of the molecule, will be the starting conformations for the second step of the protocol. However, one should note that this procedure forcibly pulls a group through the membrane bilayer, therefore even small pull rates can drag water molecules and interacting lipid headgroups, thus inducing large membrane perturbations creating non-physical representations of the system. In these cases, 50 ns MD simulations, with reasonably chosen protonation states, should be performed to allow the membrane bilayer to heal, hence possible membrane perturbations found in the production runs will be the result of relevant system effects. In more extreme cases, it might be even necessary to soften the partial charges of the Lewis base compound before the steered MD to reduce the lipid electrostatic

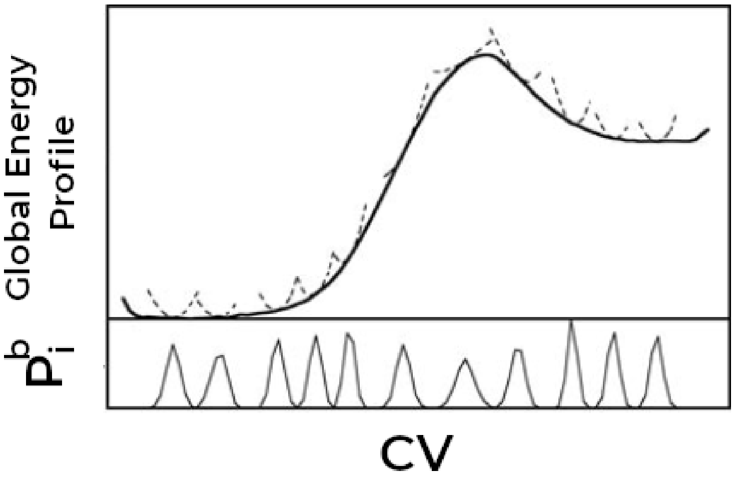


Figure 2.4: Global free energy profile (upper plot) represented in a thick curve line derived from the multiple distributions generated from each window i (bottom plot). Only a third of the distributions are represented for more clarity, hence the lack of window overlapping. Adapted from ref [213].

interactions and smooth the insertion process.

Details of US-CpHMD simulations

The next step is to perform the US-CpHMD simulations (150 ns) to sample both the protonation states and molecule configurations, in distinct membrane media, for physiologically relevant pH values: 4.5 (lysosomal internal pH), 6.0 (TME pH) and 7.5 (physiological pH). Similarly to the pHRE method, the US-CpHMD performs *n* concomitant simulations depending on how fine or broad the desired description of the system. For this system, we generated 20 US windows, each with a 0.2 nm width and a specific starting conformation obtained from the previous steered MD simulation. To reduce the computational cost, we considered only half the simulation box windows - membrane center at 0.0 nm, water-membrane interface at 1.9 nm and bulk solvent at 3.8 nm (Figure 2.5). Though, it is only possible if the following two conditions are met: (1) the insertion and exit processes are mirrored; (2) the molecule has freedom of movement in the membrane center to freely rotate and adopt the mirrored conformation. Indeed, these conditions were met following initial simulations. The final US-CpHMD simulations allow a full characterization of the insertion/exit process of the system at distinct cell pH conditions, identify the associated energy barriers using PMF profiles and estimate each compound permeability for the different environments, as described in section 2.5. The re-weighting procedure applied to estimate other properties along the internalization process was performed using in-house software following the equation 2.39.

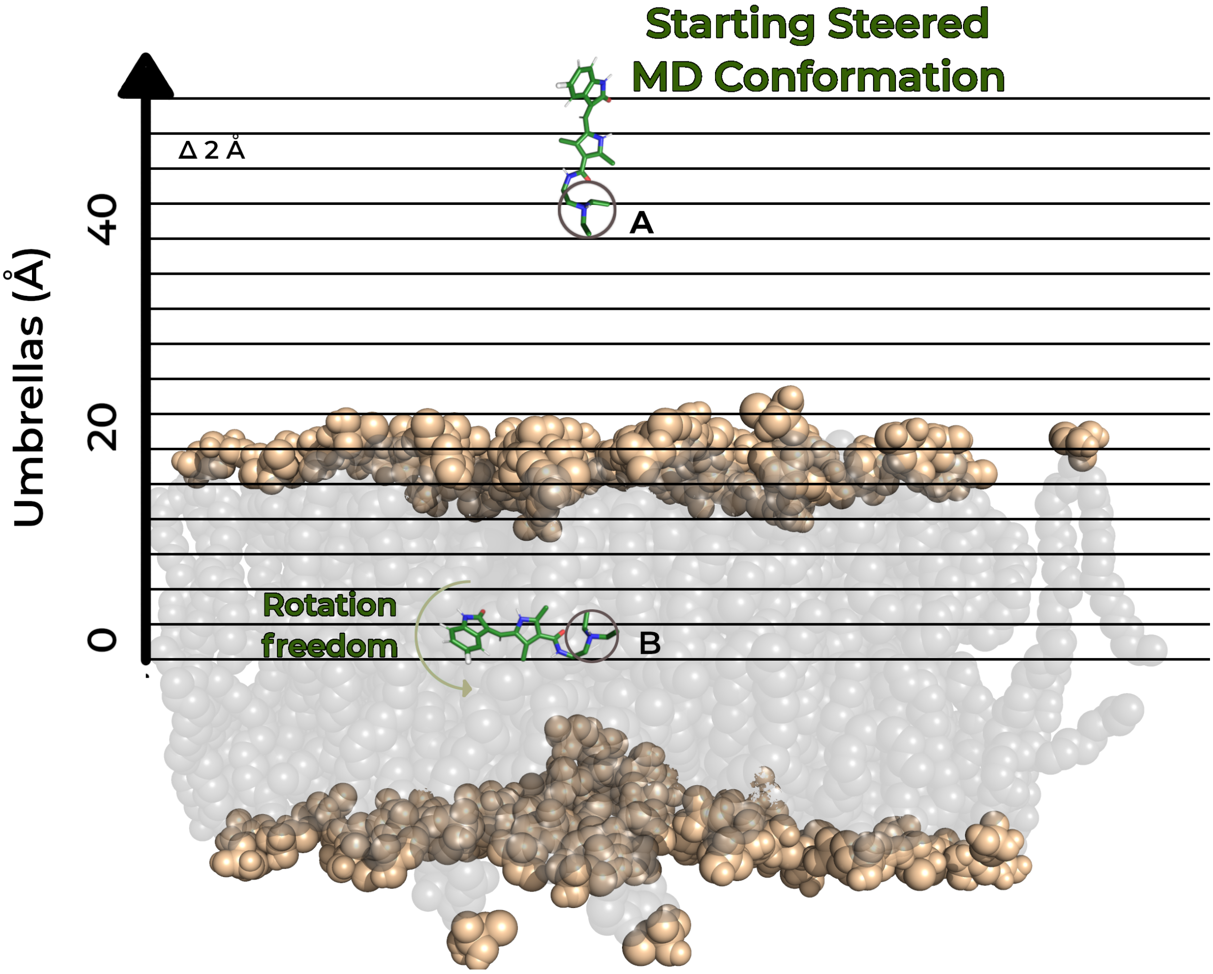


Figure 2.5: Graphical representation of a typical US-CpHMD setup for a Lewis base drug (Sunitinib). The system has been sliced in 40 umbrellas of 2 Å width ranging from bulk solvent (umb 40 Å) until the membrane center (umb 0 Å). The starting conformations for each umbrella were obtained from a steered MD procedure starting in point A until reaching point B.

2.5 Analytical approaches to transmembranar systems

To perform realistic simulations of transmembrane peptides and small membrane-interacting molecules in lipid bilayer systems require state-of-the-art techniques that have been delved upon in the last sections. Due to the complex nature of these systems and these methodologies, specialized analytical approaches are necessary to obtain the desired properties to tackle the proposed goals of this thesis. In this section, we will address a few fundamental analytical tools focused on either membrane-dependent or membrane-specific properties. Other analytical approaches and detailed derivations from the following analyses can be found in each individual chapter Methods section.

2.5.1 Insertion of a functional group

In membrane bilayer systems, the insertion property is merely a representation of the membrane medium sensed by a given group of interest relative to the membrane normal. Typically, the water-membrane interface is represented as 0 nm of insertion and deeper membrane media (ester region, acyl chains) are reflected by negative insertion values. The insertion of a molecule or a group (i.e. titrating residue) is often calculated by using reference atom(s) against the average Z coordinate of all phosphate atoms of the lipid headgroups of the closest monolayer. This particular bilayer thickness calculation method became popular as it directly correlates with experimental data of bilayer thickness (D_{HH}) measured with electron density profiles [214]. If we consider the insertion of a given residue along the membrane normal, electrostatic interactions will occur with the surrounding charged groups, as mentioned in section 1.3, perturbing the local environment of the membrane. The appearance of water-induced membrane deformations will significantly deviate the average Z coordinate value from the reference of a well behaved lipid bilayer. Consequently, affected lipid headgroups cause a poor definition of the membrane surface when using the standard procedure. Now, consider that the group of interest (e.g. an aminoacid residue) is populating the water-membrane interface in a perturbed lipid monolayer. Indeed, if the water-membrane surface sensed by the group is defined as a 0 nm insertion value, then we should only consider the local region defined by the perturbed lipid headgroups, with the biomolecule at the epicenter. Otherwise, the use of bulk lipids in the surface definition would overestimate deeper insertion values, while the group of interest is populating in the lipid headgroup region. Such inaccurate representations would deeply affect future calculations, particularly pK_a profiles.

For this purpose, we devised a new insertion method that takes into account local membrane deformations in the definition of the membrane surface [104]. In these calculations, the complex solute-membrane configuration can be simplified into: the phosphate groups (one phosphorous and four oxygen atoms) and the reference atom(s) of the group of interest. This approximation simplifies the system into two xy planes, each composed of $n/2$ clusters of five atoms (n is the total number of lipids), and our reference group. Moreover, the reference water-membrane surface is the closest monolayer to the group and, additionally, only a small subset of phosphate groups within an annulus region of 6 Å radius, centered on the reference group, are necessary. To obtain a robust definition, at least 10 phosphate group atoms must be within the radius, otherwise the closest atoms to the reference group are used to fulfil this requirement. Then, the membrane surface is defined by the average Z coordinates of the atoms within the annulus region. The group insertion is obtained with the Z difference relative to the membrane surface. These calculations are performed for every conformation along the simulation trajectory.

2.5.2 Determining pK_a profiles

pK_a estimations result from integrating the complete equilibrated protonation data of a given group, from simulations at distinct pH values, into a titration curve. In most cases, this procedure is enough to produce an accurate and representative pK_a value. However, in transmembrane biological systems, a single pK_a value rarely possesses a physical meaning, since it conflates the protonation data of unique media with distinct electrostatic environments. Meanwhile, representative pK_as in each media can describe the electrostatic behaviour experienced by the titrating group during the insertion process. These calculations are called pK_a profiles. There are a few necessary steps to calculate a pK_a profile: (1) generate both insertion and protonation time-series for the trajectories; (2) merge the data according to the correct time stamp; (3) apply a slicing procedure, where multiple insertion bins are generated and then the protonation data is assigned according to their matching insertion value; (4) perform pK_a estimations for every insertion bin using the Henderson-Hasselbach or Hill equation, if there is cooperativity with other titrating residues.

These pK_a profiles allow a fair comparison between distinct membrane-inserting groups, like different peptide variants (see Chapters 4 and 5) or distinct Lewis base drugs (see Chapter 6), even if their properties were estimated at different pH ranges or membrane environments (see Chapter 3). These calculations were useful in evaluating therapeutic efficiency of pHLIP peptides, characterize the membrane affinity at unique pH conditions or assess the sampling quality of systems with different parameters. Moreover, the same slicing procedure allows other properties to be evaluated along the membrane normal, such as the number of interacting water molecules or lipid headgroups, the average protonation, or average distances between interacting groups.

2.5.3 Membrane Half-thickness (deformation) calculations

Similarly to the other properties, the membrane thickness calculations are deeply perturbed by peptide/drug-induced deformations and, as such, average thickness values are not representative of the membrane environment. To overcome this issue, a specific method was devised that depicts the membrane monolayer (half-)thickness profile relative the lipids distance from the peptide. These calculations operate in equivalent conditions as the insertion calculations, where the biomolecule is fully considered as a reference and the phosphate groups define the lipid bilayer.

For each trajectory snapshot, the calculation requires several steps: (1) by centering the membrane on the solute, in a *xy* plane, we first define the bulk region of unperturbed lipids of both monolayers (usually the lipids beyond a 15 Å cutoff distance) and then, using their *Z* coordinates, we define the membrane center; (2) then, we define an annulus region using two starting radii (0 and X Å); (3) using only the phosphate groups within that annulus region, we calculate the average membrane surface value; (4) for each monolayer, the half-thickness value is calculated as the difference between the membrane surface and the membrane center; (5) with the estimated values of this region, the procedure is repeated with increasing radii values, operating as a sliding annulus region until the simulation box limit, thus obtaining two monolayer thickness profiles. These profiles provide a quantitative description of the membrane deformations caused by the inserting biomolecule. One should note that for the equilibrated unperturbed lipid region, the half-thickness values should converge to half the experimental values for that particular type of lipid bilayer [59, 214]. All the previous methods were calculated using the MembIT tool [104] developed in our group (`https://github.com/mms-fcul/MembIT`) and the input parameters can be found in each Results section of the following chapters.

2.5.4 Membrane permeability estimations

Earlier membrane models assumed membrane bilayer homogeneity, although the membrane cell exhibits inhomogeneous features depending on its type. The inhomogeneous solubility diffusion model (ISDM) was devised to study the permeation and diffusion of small molecules taking into account the membrane inhomogeneity. The present model enables comparisons between permeation coefficients and experimental data [168]. The ISDM model uses the Fick's law of diffusion (equation 2.40) with the premise of the Overton's rule, which states that the molecule permeation to a cell hinges on its own lipid solubility.

$$J_i = P_i \nabla c_i \tag{2.40}$$

where J_i is the molecule flux through the membrane (mol·m^{-2}·s^{-1}), P_i is the membrane permeability constant and ∇c_i is the concentration gradient between the solvent and the membrane. Several works expand into more detail the derivations and assumptions necessary [168, 215–217], though for the purpose of a brief review, we follow through to the homogeneous solubility-diffusion model (HSD), since it was an initial approach to estimate membrane permeability:

$$P = \frac{KD}{h} \tag{2.41}$$

where D is the diffusion constant in the membrane, h is the membrane thickness and K is solubility ratio of both phases described as:

$$K = e^{\frac{-\Delta G_{w \longrightarrow m}}{RT}} \tag{2.42}$$

that depends on the free energy of transition from the water to the membrane medium, where R is the gas constant and T is the temperature. Note that this model does not require any intricate membrane detail besides the thickness, hence assuming homogeneity. To introduce that level of detail within permeability estimations, we need to consider that either distinct membrane lipids or conformations may induce particular free energy profiles and diffusion constants. The Smoluchowski equation integrates these parameters depending on one dimension:

$$\frac{\partial p}{\partial t} = \frac{\partial}{\partial z}\left[D_\perp(z) e^{-\frac{1}{k_B T}F(z)} \frac{\partial}{\partial z}(e^{\frac{1}{k_B T}F(z)} P)\right] \tag{2.43}$$

where k_B is the Boltzmann constant, $D_\perp(z)$ and $F(z)$ are the diffusion and free energy profiles along the membrane normal, respectively. This equation can express the permeability across the lipid bilayer thickness as follows:

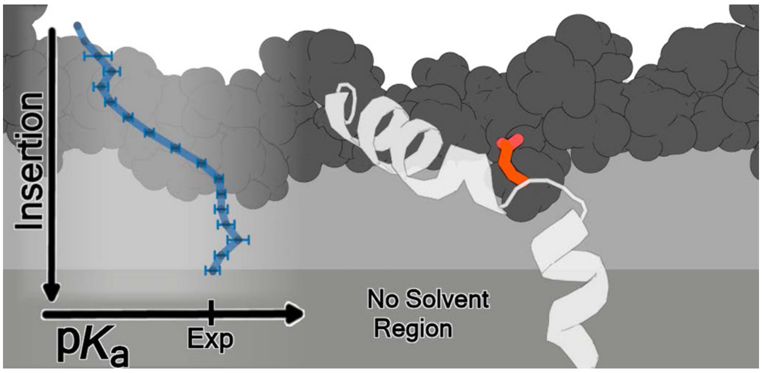
$$\frac{1}{P} = e^{-\frac{1}{k_B T}F_{ref}} \int_{-h/2}^{+h/2} \frac{1}{e^{-\frac{1}{k_B T}F(z)} D_\perp(z)} \partial z \tag{2.44}$$

with the F_{ref} expressing the constant free energy profile in the water medium, which presumes there are no energy barriers to overcome, z is the coordinate along the membrane normal with the origin in the membrane center ($z = 0$). The ISDM approach was used in chapter 6 to determine the pH-dependent permeabilities of two model Lewis base drugs.

Chapter 3

A pH replica exchange protocol for transmembrane peptide simulations

Tomás F. D. Silva[1], Diogo Vila-Viçosa, Miguel Machuqueiro



Contents

---

[1] In this work, I performed all the pHRE simulations, all the analyses, had a significant contribution to the results discussion, and wrote the initial manuscript.

3.1 Context

The first major objective of this thesis was to tackle the conformational and protonation sampling limitations of the CpHMD methodology in pHLIP peptide systems. The first work [103] clarified the role of a key protonation trigger (Asp-14) in the insertion/exit process and the peptide thermodynamic equilibrium in state III. However, sampling limitations hindered further characterization of the system due to the complex nature of peptide-membrane configurations. Using a new state-of-the-art method, the pH-replica exchange (pHRE), we devised a protocol to assess the sampling quality relative to the data obtained with the CpHMD method, at different system conditions. The pHRE method provided a clear enhancement of conformational and protonation sampling, speeding up the simulations compared with standard CpHMD protocols [166]. Still, the simulation time and computational cost were dependent on the system complexity and total size. To optimize the protocol for future pHRE simulations of these peptides (chapters 4 and 5) and other biomolecules (chapter 6), we tested the effects of differently-sized membrane patches on the peptide thermodynamic equilibrium, the accuracy of our pK_a prediction model and overall configurational integrity. The work developed in this chapter was the cornerstone of all the following chapters, without which this thesis could not have been successful. Additionally, our findings were published in the Journal of Chemical Theory and Computation (https://doi.org/10.1021/acs.jctc.1c00020).

3.2 Abstract

Many important biological pathways rely on membrane-interacting peptides or proteins, which can alter the biophysical properties of the cell membrane from simply adsorbing to its surface to undergo a full insertion process. To study these phenomena with atomistic detail, model peptides have been used to refine current computational methodologies. Improvements have been made on force field parameters, enhanced sampling techniques to obtain faster sampling, and the addition of chemical-physical properties, such as pH, whose influence dramatically increases at the water/membrane interface. pHLIP is a peptide that inserts across a membrane bilayer depending on pH due to the presence of a key residue (Asp14), whose acidity-induced protonation triggers the whole process. The complex nature of these peptide/membrane interactions resulted in sampling limitations of the protonation and configurational space albeit using state-of-the-art methods like the constant-pH molecular dynamics. To address this issue and circumvent those limitations, new simulations were performed with our newly-developed pH-replica exchange (pHRE) method using *wt*-pHLIP in different POPC membrane sizes. This technique provided enhanced sampling and allowed the calculation of more complete pK_a profiles for Asp14. The conformational heterogeneity derived from strong electrostatic interactions between Asp14 and the lipid phosphate groups was identified as the source of most pK_a variability. In spite of these persistent and harder-to-equilibrate phosphate interactions, the pK_a profiles at deeper regions (6.0–6.2) still predicted the experimental pK of insertion (6.0), since the electrostatic perturbation decays as the residue inserts further into the membrane. We also observed that reducing the system size leads to membrane deformations where it increasingly loses the ability to accommodate the pHLIP-induced perturbations. This indicates that large membrane patches, like 256 or even 352 lipids, are needed to obtain stable and more realistic pHLIP/membrane systems. These results strengthen our method pK_a predictive and analytical capabilities to study the intricate play of electrostatic effects of the peptide-membrane interface, granting confidence for future applications in similar systems.

3.3 Introduction

Several biological processes and therapeutical technologies resort to membrane-interacting peptides that perturb the membrane, altering its properties such as fluidity or thickness. These molecules can be derived from nature, such as antimicrobial peptides (AMPs)[218] and *wt*-pHLIP[120], or rationally designed such as GALA[219] and WALP/KALP peptides[84]. Different molecules lead to a myriad of interactions from pore formation inducing cell lysis, [219] membrane fluidization, and other phenomena such as (partial or complete) membrane insertion, [84, 120] adsorption to the water-membrane interface or membrane crossing to regulate signaling pathways. [220] These biological phenomena are conducted by several molecular interactions relying on the membrane constitution, apolar environment, pH-dependent triggers for conformational changes, which are always better studied at the atomistic level. However, despite recent improvements on computational resources, methodologies and technological leaps, the molecular interplay in several peptide-membrane systems remains mostly elusive. Thus, well-studied membrane-interacting peptides, such as pHLIP, are a valuable case study to further deepen our understanding of both the dynamics involved in peptide-membrane systems and optimize simulation protocols and analytical methods.

The pH (low) insertion peptides (pHLIP) belong to a family of transmembrane peptides whose insertion and function is dependent on pH[117, 121, 128, 139, 140, 144, 221–223]. The *wt* sequence of this peptide was obtained from the transmembrane C-helix of bacteriorhodopsin [121], while possessing two acidic residues (Asp14 and Asp25) in the transmembrane region and another four anionic residues (Asp31, Asp33, Glu34 and C-ter) in the C-terminus domain. pHLIP biophysical studies have identified and characterized three states, one water soluble state and two in the presence of lipid bilayers[224]. At high pH value, pHLIP accumulates in an adsorbed state at the water/membrane interface, while at slightly acidic pH ($<$6.0), the peptide fully inserts into the lipid bilayer. Due to these thermodynamic and kinetic properties, pHLIP variants have been designed and studied as probes, coupled with imaging techniques, and/or drug-delivery systems, by attaching therapeutic agents to the C-terminus region, to target tumor cells and other inflammatory diseases. [117, 220–222]

In our previous work, we used the stochastic titration constant-pH molecular dynamics (CpHMD) methodology, [151, 156, 159, 225] extended to deal with lipid patches,[27, 158, 164, 165] to explore some of the mechanisms underlying the insertion process and the pH-dependent stability of the pHLIP membrane-inserted state. We showed that, although the insertion process requires the full protonation of Asp14 and Asp25, in perfect agreement with the isothermal calorimetry data [120]), Asp14 is the key residue whose protonation defines the stability of the membrane-inserted state [103]. Due to Asp25 being deeply inserted in the acyl region, it is unable to serve as an effective pH sensor and trigger membrane insertion/exit events. Furthermore, we observed that the acidic residues at the C-terminus domain protonate, at least transiently, following their topological order to properly transverse the lipid bilayer. When charged, these groups are anchored at the water/membrane interface and have a prohibitively high membrane-crossing energetic barrier.

The protonation behavior of pH sensitive molecules interacting with membranes has been attracting a large interest,[27, 37, 103, 226] in particular, due to the development of more advanced CpHMD methodologies able to tackle these systems. Nevertheless, there are fundamental limitations that need to be addressed before our computational data can be directly compared with macroscopic properties. An important issue is the limited sampling obtained at the water/membrane interface.[27] Even though most methodologies can circumvent the kinetic limitation of proton availability, in the presence of a lipid bilayer, the conformational space is slowly sampled and, consequently, this has a tolling effect on the protonation sampling. There are a few strategies to mitigate this, which include the use of long simulation runs with multiple replicates,[27] harmonic biasing methods,[226] or pH replica-exchange schemes.[37] In this regard, our recent development of the pH replica-exchange method (pHRE) improved our ability to sample the protonation and conformational space of several systems,[166] including the protonation of titrable compound drugs at the water/membrane interface. [37]

Another fundamental limitation is related with the fact that pK_a is a macroscopic property and its concept or definition breaks in the microscopic world, like MD-size systems. In our simulations, we calculate proton binding affinities for the titrating groups, which, under normal circumstances, relate well with the macroscopic pK_a values. However, at the water/membrane interface, the macroscopic pK_a values become more difficult to calculate (sampling issues) and with limited usefulness. When attempting to capture the pK_a value at the moment of insertion, single pK_a calculations reflect a value fitted from protonation values obtained from several conformations ranging from deeply inserted to solvent exposed residues, each with distinct local electrostatic environments. Thus, in this scenario, a single pK_a value holds the convoluted information of the protonation equilibrium and the membrane insertion distribution, which most experimental techniques cannot untangle. This problem can be circumvented by using pK_a profiles along the membrane normal for titrating groups inserting a lipid bilayer. [27, 37, 103] This approach allowed us to successfully compare an experimental pK of insertion with the *in silico* pK_a value calculated at the deepest insertion available, even under significant sampling limitations. [103] However, if we solve the protonation/insertion sampling issues, a new fundamental question will arise regarding the level of membrane insertion at which the calculated proton binding affinity correlates better with the insertion pK. At deeper membrane insertions, a proton binding affinity might not have a direct correspondence to a pK value if the titrating group has no real access to water and does not sense the pH in bulk solution.

In this work, we aim at improving our computational protocol which tackles the protonation, membrane-insertion, and conformational spaces of pHLIP peptide. We used our newly developed pHRE method to improve the protonation/conformation sampling in deep membrane regions and assessed it by comparing pHLIP systems built independently with differently-sized lipid patches. We also revised our method to calculate membrane insertion values to make it more robust and compared it with a solvation exposure metric based on the number of explicit water/phosphate contacts within the first solvation shell.

3.4 Methods

3.4.1 System setup and pHRE simulations

Five systems were prepared using the *wt* sequence of pHLIP [103, 123] and a membrane bilayer of 2-oleoyl-1-palmitoyl-*sn*-glycero-3-phosphocholine (POPC) with different sizes/number of total lipid molecules: 128, 160, 192, 256, and 352. All systems were built starting from the previously equilibrated 256 lipids system [103]. We applied a resizing protocol that iteratively removes lipids/water molecules, followed by MD equilibration (20 ns, applying position restraints on the peptide, 1000 kJ/mol nm^2), until the desired size was obtained (192, 160, and 128). The larger lipid system (352) was built starting from a 4x128=512 lipid system, which was then submitted to the same protocol. All systems were equilibrated using MD simulations, followed by 100 ns of CpHMD simulations at pH 6.0. The peptide is folded as a kinked α-helical structure, between the 15th and 18th residues, inserted in the membrane bilayer (Figure 3.1).

All systems were simulated using the pH replica-exchange (pHRE) method.[166] pHRE is an extension of the CpHMD-L methodology [27, 151, 157–159, 164, 165], where an additional step is introduced. The CpHMD-L method can be broken down in three major steps: 1) a Poisson Boltzmann/Monte Carlo (PB/MC) procedure where we obtain new protonation states for titrable groups using PB-derived free energy terms; 2) a small solvent relaxation step where we allow the solvent to adapt to the new protonation states; 3) and a molecular mechanics/molecular dynamics (MM/MD) step where we sample for conformations. In a pHRE simulation, we perform n simultaneous CpHMD-L simulations called replicas, each with an assigned pH value from a given set

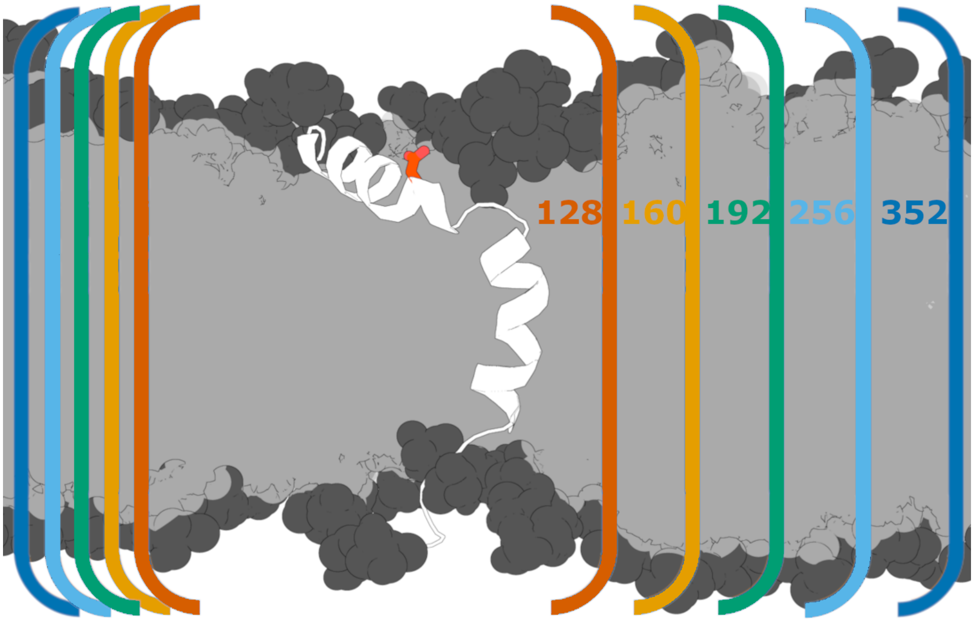


Figure 3.1: Illustrative representation of the 5 *wt*-pHLIP systems build with different sizes, corresponding to different number of lipids. The peptide started from the previously observed kinked α-helical conformation with key Asp14 located at the water/membrane interface.[103]

of n pH values. During the MM/MD step, the simulation is interrupted to attempt an exchange between replicas of adjacent pH values, at a fixed frequency (τ_{RE}). This exchange allows conformations and protonation states at a given pH value (replica) to be obtained from different simulations, thus increasing diversity in our sampling of low and high energy states. An exchange occurs given a probability dependent on the exchanging pH values and the number of protonated titrable groups[166, 206]:

$$P_{acc} = min\left\{1, exp[-(pH_m - pH_l)(N(x_i) - N(x_j))ln10]\right\} \tag{3.1}$$

where pH_m and pH_l are the exchanging pH values, x_i and x_j are the number of protonated groups in each replica. As such, we performed five (128, 160, and 192) and ten (256 and 352) replicates of 100 ns each for the previously referenced systems. Ten replicates were used for the larger systems in order to improve the sampling of pHLIP at more inserted locations. Each replicate is composed of four pH replicas starting from a pH value in the range of 4.0 to 7.0 with a step of 1.0. Each CpHMD cycle was 20 ps (τ_{prt}) and each solvent relaxation step was 0.2 ps (τ_{rlx}). Every exchange attempt was performed every 20 ps. The titrable groups are the (N- and C-) termini and all the acidic residues (Asp14,25,31,33 and Glu34). The starting conformations in each replica were obtained from the final segment of equilibration procedure, with the peptide inserted in the membrane partially exposing Asp14, while Asp25 is deeply inserted in a neutral state (Figure 3.1). [103] For equilibrium properties, the initial 50 ns of each replicate were discarded to allow for the peptide equilibration along the membrane normal (Figure A.1 of Supporting Information).

3.4.2 Molecular dynamics settings

All molecular dynamics simulations were performed with the GROMOS 54A7 force field [172], a modified version [159, 194] of GROMACS 4.0.7 [227] and a Python-based wrapper to implement the pHRE method[166]. Additionally, the treatment of non-bonded interactions used a single cutoff scheme with all pairs below the 14 Å cutoff updated every step.[180] For all long-range interactions, the van der Waals were truncated and the electrostatic interactions were treated by a dielectric constant of 54 [228] and an ionic strength of 0.1 M in the generalized reaction field (GRF) method [159, 182], respectively. Protein and lipid bonds lengths were constrained by P-LINCS algorithm [229], while water molecules were treated as simple point charge (SPC) [230] with the SETTLE algorithm[231]. The integrator time step used was of 2 fs. An NPT ensemble was used with the temperature of the solute and solvent separately coupled to a v-rescale [189] temperature bath at 310 K and a relaxation time of 0.1 ps. The pressure was kept constant at 1 bar using a Parrinello-Rahman barostat [232] with a relaxation time of 5 ps and compressibility of 4.5×10^{-5} bar^{-1}.

3.4.3 Poisson–Boltzmann/Monte-Carlo calculations

Calculations for Poisson–Boltzmann were performed with the program DelPhi V5.1 [191], which used both partial charges and Lennard-Jones parameters, to obtain atom radii at 2 RT,[233] from the GROMOS 54A7 force field. The molecular surface of the membrane-peptide system was characterized using a probe with a 1.4 Å radius, an ion exclusion layer of 2.0 Å and ionic strength set to 0.1 M. The applied dielectric for both solute and solvent was 2 and 80, respectively. In this process, a two step focusing procedure employed a grid with spacing, defined by 91 grid points of approximately 1 and 0.25 Å either in the large or small grids, respectively. The relaxation parameters for the coarse grid were 0.20 and 0.75 for linear and nonlinear iterations, while periodic boundary conditions were applied in the x and y directions. Background interactions calculations were limited to 25 Å and the convergence threshold for the electrostatic potential was 0.01 kT/e[157].

The protonation states were sampled using the PETIT program by applying a Monte Carlo (MC) scheme that uses the free energy terms derived from previous PB calculations [193]. Proton tautomerism was accounted for all titrable groups. $\times 10^5$ MC cycles were performed for each conformation, each cycle corresponds to a trial change of each individual site and pairs of sites with an interaction larger than 2 pK units.

3.4.4 pK_a profile calculation along the membrane normal

To build the pK_a profiles along the membrane normal, multiple pK_a calculations are required at different degrees of insertion. For this purpose, our insertion method separates conformations, sampled from different pH values, into slices according to their relative membrane insertion. This insertion was previously calculated using the average position of the lipids phosphorous atoms within a 6 Å radius from the group of interest [103]. This method improved the issues of local membrane deformations, but often struggled due to a lack of reference atoms within the chosen radius, leading to inconsistencies in the membrane insertion values or heterogeneous conformations in the same slice. Building on top of our previous method, we now use the average Z position of any atom belonging to the phosphate groups (includes the phosphorus and the 4 oxygen atoms) within the cutoff, thus increasing the number of reference atoms and improving our method's robustness. Additionally, we established a requirement of, at least, 10 data points to safeguard the lack of robustness in the previous method when no occurrences were found within the chosen radius and only the closest atom was used. Now, when a minimum of 10 reference atoms (from the phosphate group) are not available within the cutoff radius, then the closest 10 atoms to the reference residue, in a two-dimensional (*x/y*) plane, will be used. This new and more robust method still captures the local membrane deformations induced by the peptide.

The insertion calculations were performed on all conformations, which, along with the respective protonation state, were separated to an insertion bin. In each bin three criteria must be met to assess sampling quality: (1) each slice requires a minimum of 10 conformations of each protonated state; (2) these conformations need to come from, at least, two different pH values and, at least, three replicates must contribute with sampling to each slice; (3) the average protonation, at each pH, must not exceed the value of the previous lower pH (within a 0.05 tolerance), maintaining monotonicity. If these criteria are met, the data is fitted to the Hill equation, thus obtaining a pK_a value for that slice. The slicing procedure previously described was also applied to different combinations of data. We assigned residue protonation to bins of residue insertion or proton donors (water molecules and/or phosphates). We were also interested in assigning proton donors to residue insertion to assess the distribution of phosphates groups and/or water molecules at each insertion level. All these analyses were performed with in-house software (http://mms.rd.ciencias.ulisboa.pt/#software) .

3.4.5 Analyses and error calculations

Membrane thickness values are often derived from the distance between the average Z coordinate of the phosphorous atoms of both monolayers. However, this approach is limited when dealing with a peptide-membrane complex, since an inserted peptide will induce local membrane deformations, which would get diluted, while still affecting the global thickness. To prevent this issue and quantify the local membrane deformations, we used a method that calculates the half thickness, for each monolayer, within two radii from the peptide - an annulus region [103]. Scanning/sliding this annulus radii in the *xy* plane, we can calculate the varying monolayer thickness values by averaging the z positions (to the membrane center) of all phosphorous atoms contained in this region. With this method, we can capture both local deformations and membrane "bulk" (unaffected) regions (defined by lipids at longer distances, > 15 Å). These calculations are applied to all equilibrated conformation snapshots and, at longer radii, the thickness of both monolayers should converge to the same "bulk" value, i.e. half the thickness value for a pure POPC membrane. The experimental POPC half thickness value range was calculated by interpolating from experimental thickness measurements in the fluid range, at different temperatures [59].

All pK_a error values were obtained using a Bayesian bootstrap approach, which avoids several fitting issues by running 1000 bootstraps from our average protonation samples and, for each bootstrap, random weights were assigned to each sample. This procedure also requires the same selection criteria to obtain final pK_a and error values. All other error values were obtained using a simple standard error of the mean.

3.5 Results and discussion

MD simulations of membrane interacting peptides (i.e. pHLIP) face a difficult sampling challenge due to their complex configurational space, especially when pH is included. The use of constant pH MD methods [27, 103, 151, 159] coupled with replica exchange methodologies [166] has shown increased sampling in similar simulation times for several systems,[37] prompting more detailed analysis of complex titrating biological molecules. Previous work done on pHLIP has shown the predictive ability of our model on obtaining the pK_a profile of Asp14 and correlate it to the experimental pK of insertion, despite a lack of protonation/conformation sampling of ionized D14 at deep membrane regions.[103] The application of the pH replica-exchange (pHRE) method [166] seeks to improve the configurational/protonation sampling of the peptide-membrane interactions by applying it to the *wt*-pHLIP in significantly different lipid constructs (Figure 3.1).

In peptide/membrane simulations, the choice of membrane type and size is important since it impacts the simulation speed and, more importantly, the peptide environment, which will define its configuration, conformation, and protonation spaces. A simple insertion analysis of the key Asp14 at different pH values can provide important information on the peptide/membrane equilibration of the different lipid patches (Figure 3.2). Although, the residue average protonation will differ between pH values, its membrane insertion seems to reside consistently within the -2 to -4 Å range across systems, except the 128 lipid membrane system. In this smaller membrane patch, the residue is more inserted at all pH values, suggesting that the peptide-membrane equilibration is significantly perturbed, thus sampling different configurations. The observed effect on the 128 lipid patch is

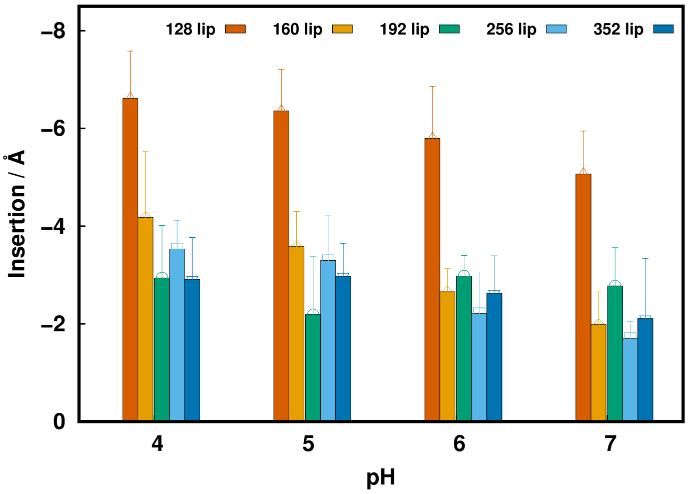


Figure 3.2: Average insertion values of Asp14 for all lipid (lip) size membrane systems at different pH values. The negative sign in the insertion values is a convention for values below the phosphate line (used as the reference). Only the final 80 n s segments of each simulation were used for the average insertion calculations.

most likely caused by box limit effects, due to its short size, causing significant membrane deformations and an altered electrostatic environment for Asp14. These perturbations can also be quantified by calculating the membrane thickness radially around the peptide (Figure 3.3). The monolayer thickness (or half-thickness) observed for the smaller system shows that the lipids further away from the peptide (∼30 Å) are very perturbed and not resembling the typical fluid POPC lipids. The results from membrane insertion and monolayer thickness show that the 128 lipid system is not a good physical model for state III *wt*-pHLIP in POPC.

In Figure 3.3, all systems show significant POPC monolayer thickness deviations at both local (<5 Å) and bulk (30–40 Å) environments. A common feature in all systems is the lack of unperturbed lipids at larger distances, with the magnitude of these deformations inversely-depending on the size of the lipid patch. Usually, larger membrane patches can accommodate most membrane fluctuations, however, in the presence of pHLIP, even with larger membrane patches (352 and 256), there is no complete convergence to the experimental membrane thickness. Also, pure POPC simulations, with these system sizes, result in thickness values inside the experimental range (data not shown).

In the outer monolayer (Figure 3.3A), the peptide induced a relatively small local membrane deformation around Asp14, even though there was a noticeably different impact between membrane patches creating two regimes: lipid patches 352, 192 and 128 exhibit slightly larger deformations (∼2.5 Å) than the remaining systems (256 and 160 lipids), which showed softer distortions (∼0.5 Å). It seems that in the first membrane patch group, Asp14 is establishing stronger (more durable) interactions with lipid headgroups than in the remaining systems, which leads to a more pronounced phosphate-dragging effect, visible in the monolayer thickness profiles. These differences cannot be fully attributed to box size effects, since they do not correlate with the lipid patch size, suggesting that this variability may be due to equilibration issues.

In the inner monolayer (Figure 3.3B), the increased number of anionic residues at the C-terminus leads to larger deformations since they ionize often and require stabilization by a well-solvated environment.[103] Interestingly, in this monolayer, there seems to be a higher correlation between membrane patch size and local perturbations, hinting at the role of bulk lipids (which are truncated at shorter system sizes) in stabilizing the membrane deformations. The inability to accommodate these large bulk deformations by the truncated systems causes system variability dependent of membrane size, which adds to the local variability dependent on the electrostatic

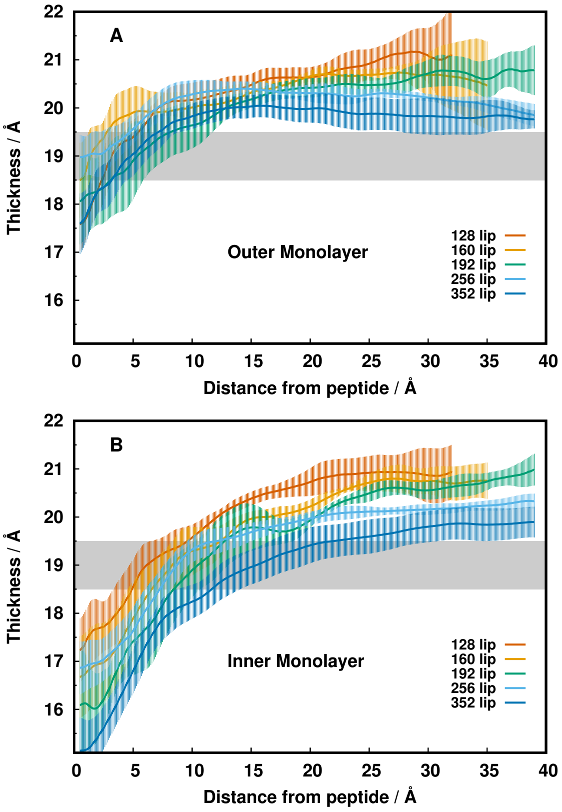


Figure 3.3: Thickness profiles for all systems at pH 6.0 for the outer (A) and inner (B) monolayers. The distance from the peptide is measured, in the X/Y plane, every 0.1 Å. The grey region delimits the monolayer thickness for experimental POPC (19.0 ± 0.5 Å)[59]. A value of 19.5 Å is typical for pure POPC in similar simulation parameters.

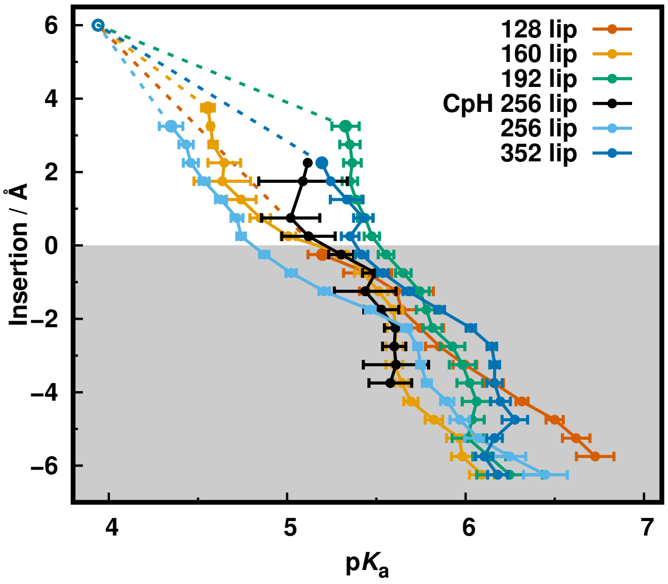


Figure 3.4: pK_a profiles for Asp14 along the membrane insertion for the different lipid size systems. The data obtained using CpHMD simulations[103] is also shown for comparison. The pK_a value (3.94) for an aspartic acid in solution[234] is shown in an arbitrary region away from the membrane, representing bulk water. The membrane is highlighted using a grey-shaded region.

interactions between peptide and lipids. Both these effects can create different meta-stable states of the peptide-membrane configuration corresponding to different equilibrium minima (both conformational and protonation) from which our sampling is obtained. Notwithstanding, this myriad of different peptide/membrane configuration do not seem to have a significant impact on the overall peptide secondary structure (Figure A.2 of Supporting Information).

An effective approach to circumvent the lack of sampling along the membrane normal is to calculate our properties of interest along this vector.[27, 103] Using a system slicing procedure, we can calculate pK_a profiles for Asp14 (Figure 3.4), which we have shown to correlate with the experimental pK of insertion.[103] pK_a profiles of Asp14 were calculated with our new insertion method (see Methods), including the data used in our previous work with CpHMD simulations of *wt*-pHLIP in a 256 POPC membrane bilayer.[103] An evident discrepancy occurs between CpHMD (black) and pHRE with 256 lipids (light blue). The pHRE results show increased sampling, covering a larger insertion range, and a higher degree of confidence, with significantly smaller estimated pK_a error values. The most important improvement of the pHRE methodology is the enhanced ability to sample the ionized form of Asp14 at deeper membrane regions, which was pivotal to perform pK_a calculations. Focusing only on the pHRE profiles, there is an overarching similarity to the expected behaviour of an anionic residue inserting into a membrane [27]. The exception is the 128 lipids profile, which is remarkably distinct from the remaining systems at the deepest sampled insertions. This is expected taking in consideration the significant membrane deformations observed (Figure 3.3), which should have a detrimental effect on the electrostatic environment of the peptide.

The remaining pK_a profiles show significant variability with the exception of the most inserted regions (-5 to -6 Å), where our estimations converge to a pK_a value within the $6.0-6.2$ range (at 5.75 Å insertion), which is a very good estimation of the experimental measurements (pK^{ins} value in the 5.8–6.0 range [123, 223]). The diverging profiles along the membrane normal between all systems are most likely due to the heterogeneity between these systems. Here, we purposely generated different systems, by systematically changing their total number of lipids and quantify the resulting configurational heterogeneity. Although our systems are apparently converged, they are sampling conformations/protonations from slightly different electrostatic environments around Asp14. By slicing the Asp14 insertion values into bins, we grasp the specific environment of each system and quantify its impact directly on the pK_a of Asp14.

The more pronounced variability in the pK_a profiles occurs at the -2 to -4 Å insertion region, where there is poor overlap between the profiles (Figure 3.3). In this particular membrane region, Asp14 has stronger electrostatic interactions with the lipid phosphate groups. The interaction of anionic residues with negative groups favour their neutral (protonated) form, as to lessen the repulsion between negative charges and, concomitantly, shifting the pK_a towards higher values. The different profiles observed in these systems, suggest that the number and/or strength of Asp14–phosphate interactions is different as a consequence of the mentioned configurational heterogeneity. However, such effects begin to soften as the residues insert at deeper regions, leading to convergence of the pK_a profiles. These results indicate that, at these insertion values (-5 to -6 Å), the phosphate interactions are mitigated, with a concomitant decrease in its variability, leading to the desolvation effect as the main factor influencing the pK_a values. Interestingly, even pHLIP structural properties like the local helicity of Asp14 (defined by counting the helical residues in the 12-16 range) can undergo a slicing procedure. The helicity profiles show a significant conformational homogeneity in these inserted regions (Figure A.3 of Supporting Information), which indicates that the pK_a variations are mostly due to differences in peptide/lipids configurations.

To further explore the observed variability in the pK_a profiles, we calculated the average number of contacts between Asp14 and key hydrogen bond donors/acceptors, along the membrane insertion (Figure 3.5). These calculations were performed, for systems with 352, 256, 192 and 160 lipids, using the Asp14 first interaction shells obtained from water RDF analysis (0.42 nm; Figure A.4 of Supporting Information) and phosphate free oxygen probability density analysis (0.52 nm; Figure A.5 of Supporting Information). The average number of interacting partners (both water and phosphate groups) in all systems are steadily decreasing along the membrane insertion, excluding a few low sampling regions, and in agreement with the expected desolvation effect (Figure 3.5A). The profile convergence in the insertion region between 0 and -5 Å, indicates a common interaction shell size of Asp14 along the membrane normal between systems. However, despite the number of partners being similar, there is a varying number of phosphate interactions in this region inversely proportional to the water molecules (Figure 3.5B-C). In particular, we observe a predominance of phosphate interactions in the 352 and, to a lesser extent, in the 192 systems, which will inevitably change the electrostatic environment around Asp14. The phosphate groups create strong interactions with the protonated carboxylic acid and destabilize its ionized form, leading to an overall increase in the pK_a values, which also help explain the observed variability in the pK_a profiles (Figure 3.4).

The average number of phosphate groups in the 160 and 256 lipid patches suggest that these interactions are more transient and less stabilizing than in the 352 and 192 patches (Figure 3.5C). We have observed that this separation between lipid patch sizes (352 and 192 vs. 256 and 160) was also present in the membrane thickness data, where the first group showed pronounced membrane deformation in contrast with the second (Figure 3.3A). We observe larger deformations in the systems where more phosphate groups are within the interaction shell of Asp14 (352 and 192 lipids), probably due to phosphate dragging upon insertion, which can be captured in the monolayer thickness profiles. Hence, the amount of interacting phosphate groups impact both the electrostatic and the structural vicinity of Asp14 creating sampling heterogeneity in our simulations. Nevertheless, at deeper insertions (-5 to -6 Å), the interactions with phosphate groups are significantly decreased (Figure 3.5C), which seems to dilute their perturbing effects and allow for all systems to converge, at least, in terms of the electrostatic environment and the obtained pK_a profiles (Figure 3.4). The fact that, in this region (~ -5.5 Å), we observe very similar pK_a values (6.0–6.2), which are in great agreement with the experimental pK of insertion (6.0[123]), indicates that a simple water desolvation effect is required to mimic the experimental conditions. Moreover, it

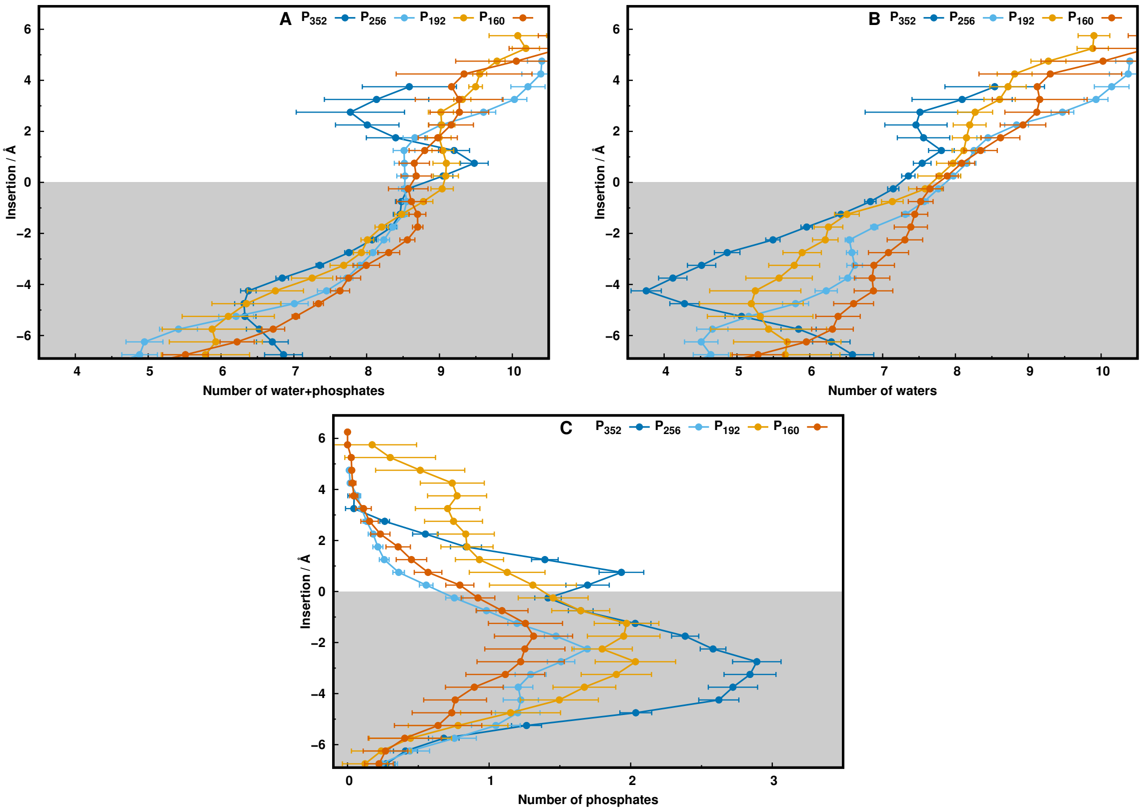


Figure 3.5: Average number of interacting contacts between Asp14 and water/phosphate groups along the residue membrane insertion. The interactions were separated between water+phosphates (A), just water molecules (B), or just phosphate groups (C). The interaction cutoff distances used were 0.42 nm and 0.52 nm for water and phosphate, respectively (see Figures S3 and S4 of Supporting Information for more details). The membrane is highlighted using a grey-shaded region.

seems that the magnitude of phosphate interactions, being more or less realistic, can be simply disregarded when attempting to estimate the experimental pK of insertion.

The pK_a profiles are a useful analytical tool to evaluate the proton-binding affinities and the electrostatic environment along the membrane insertion pathway. Nevertheless, when equilibrium and convergence are attained with a reasonable conformational/protonation sampling, we can extend our analysis to protonation profiles at a given pH, coupled with their insertion distributions (Figure 3.6). In all systems, Asp14 resides mainly in the region immediately below the phosphate groups (0 to-4 Å), with just a few exceptions mostly related with membrane deformation and lack of convergence. In particular, for the 128 lipid membrane patch, the residue was significantly more inserted (-6Å) at all pH values, while also favouring the neutral protonated state compared with the other systems (Figure 3.6A). This evidence corroborates the previous data on the system being significantly perturbed. In the remaining systems, we observe similar protonation profiles along the membrane normal at all pH values, following the desolvation effect pattern, which favors the neutral form (Figure 3.6B-E). As expected, the membrane insertion distributions are pH-dependent with the more acidic pH values favoring deeper regions. In this regard, the 192 lipid system is an exception (Figure 3.6C), as the pH 7 distribution is generally more inserted than the other pH values. This is probably a consequence of the pronounced local membrane deformations which allow for stable conformations of ionized Asp14 in more inserted locations. To some extend, this effect is also present in the 352 lipid system, resulting in an attenuation of the pH dependence (Figure 3.6E). At pH 6, the residue insertion distributions are similar between all systems (excluding the 128 lipid system – Figure A.6 of Supporting Information), while the protonation profiles show heterogeneity in the -2 to -4 Å region (Figure 3.6F), similar to what was observed in the pK_a profiles (Figure 3.4). Nevertheless, these deviations, due to non-equilibrated phosphate interactions, are small and tend to be mitigated in the pK_a calculations when data from different pH values are used.

It is now clear that the observed protonation and configurational variability along the membrane normal sprouts from the Asp14 interactions with phosphate groups that are too strong (Figure A.7 and Table A.1 of Supporting Information) to allow adequate equilibrations within our simulation timescale. These strong interactions become more evident when we evaluate how persistent and prevalent they are throughout the simulation time (Table A.1 of Supporting Information). Our data indicates that the two systems (256 and 352 lipids) are sampling different configurational spaces, where the larger system is favoring the hydrogen bonding between the carboxylic and phosphate groups, which will inevitably affect the electrostatic environment surrounding the Asp14 carboxylic group. It is difficult to assess which of the configurational spaces is more realistic, since there is no experimental data to guide us. Furthermore, these strong interactions may also be exacerbated in our model since they strain the limits of the non-polarizable force field used. Although these interactions are difficult to circumvent, we can still quantify them, by separating our sampling between conformations of Asp14 with only water molecules and conformations which include at least one phosphate group in its first solvation shell. The number of interactions within the interaction shell of the carboxylic acid can be used as a similar measurement to membrane insertion and allowed for the calculation of new pK_a profiles (Figure 3.7). By strongly decreasing the phosphate interaction contribution, excluding these direct contact interactions (Figure 3.7A), all profiles converge to pK_a values ~ 6, as previously noted. On the other hand, when at least one phosphate interaction is enforced, there is an overall pK_a shift towards higher values (Figure 3.7B). In a given sense, the observed pK_a profiles (Figure 3.4) are calculated using data from both sets of conformations, where each system variability derives from the different weights in the populations sampled. Interestingly, the lower limit of interactions necessary for the pK_a calculations is around 2 to 3, indicating that there is a required minimum of stabilizing water/phosphate groups to allow the carboxylic acid ionization. At the limit of desolvation, we find a better match to the experimental pK when the interactions with phosphates are reduced, which also suggest that, at this level of membrane insertion, the realistic phosphate influence should be minimal.

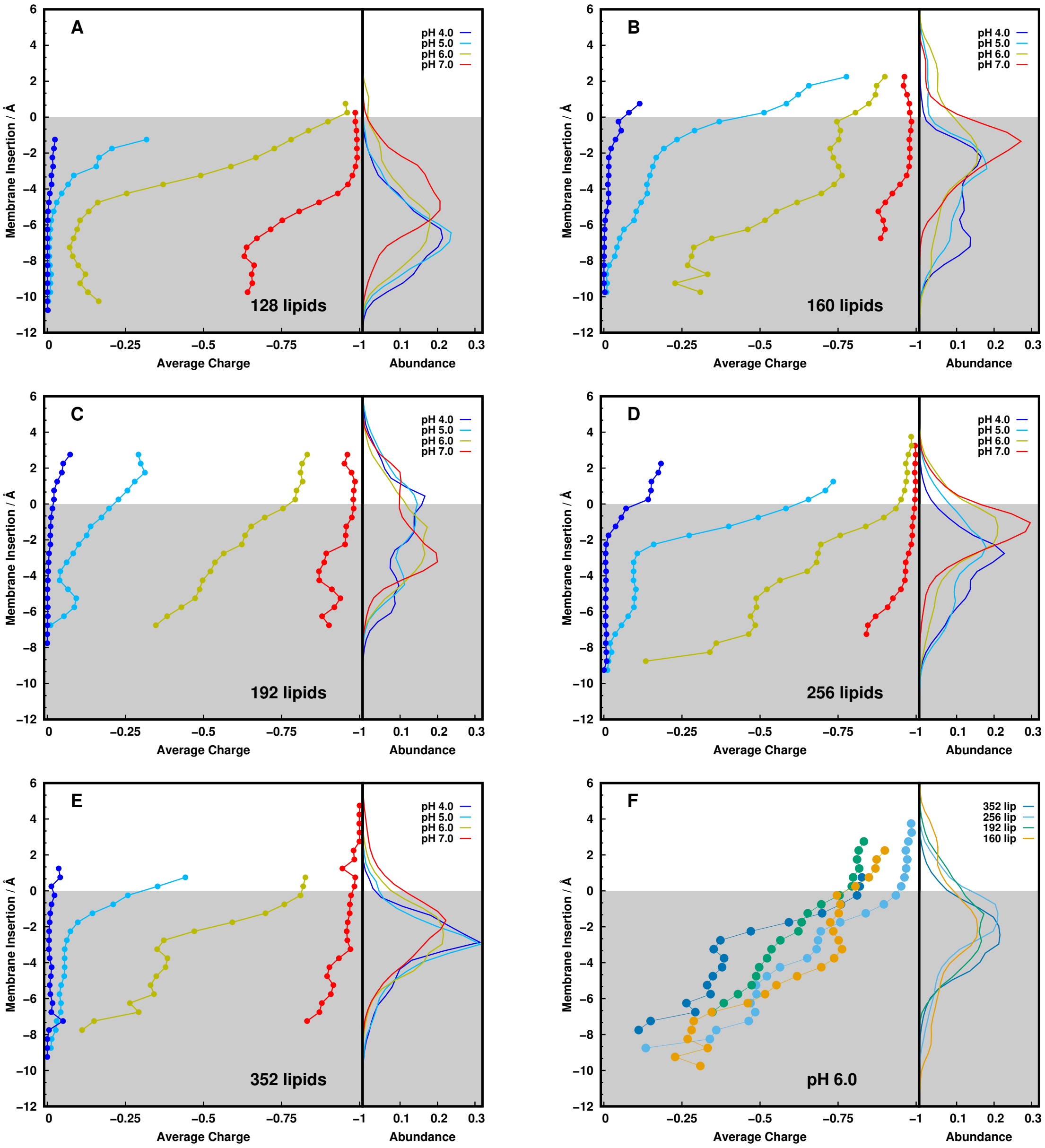


Figure 3.6: Protonation profiles and membrane insertion data of Asp14 at all simulated pH values. The data is presented for all lipidic systems (A–E) and a comparison between systems is also shown at pH 6.0 (F). The membrane is highlighted using a grey-shaded region.

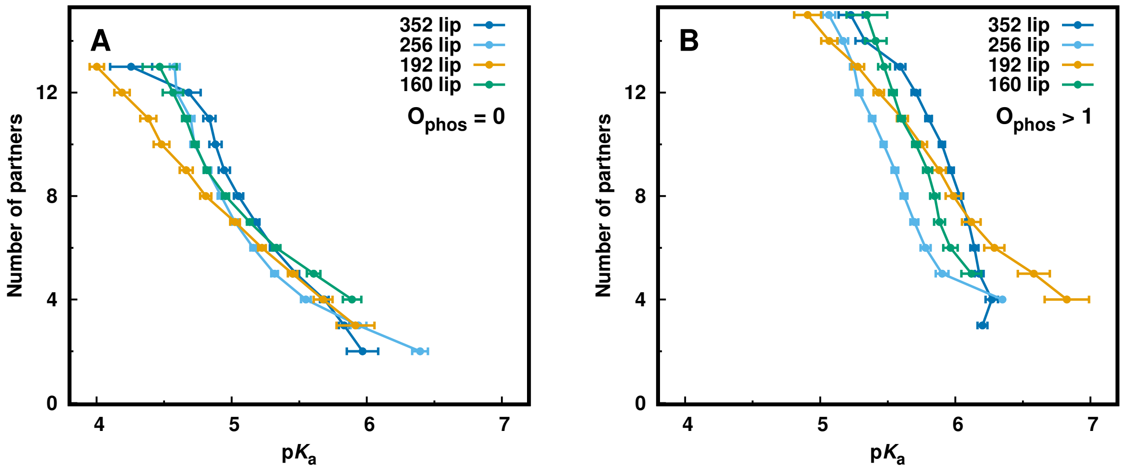


Figure 3.7: pK_a profiles of Asp14 relative to the number of established partners within the interaction shell. The partners are either water molecules and/or phosphate groups. In subplot A, the pK_a calculations were done only on conformations without any phosphate group in the interaction shell of Asp14, while in subplot B, all used conformations have at least 1 phosphate group and water molecules.

3.6 Conclusions

In silico studies of pH-sensitive peptides in membrane environments face several sampling and methodological issues, especially when attempting to directly compare the data with macroscopic properties such as the pKs of membrane insertion. To tackle these challenges, our implementation of a pH-replica exchange (pHRE) methodology improved the overall sampling quality, particularly for disfavored ionized states in deep insertion regions. The distinct variability in pK_a profiles observed previously, using CpHMD simulations of different pHLIP systems, could now be addressed systematically using the pHRE approach. This improved sampling highlights the variability of peptide-membrane interactions, particularly in regions below the water/membrane interface, where interactions with phosphate groups can be exacerbated, thus affecting protonation and pK_a estimations. By creating new system constructs with varying membrane sizes, we observed increased variability in the amount of these interactions, not depending clearly on the system size. This heterogeneity was also observed in several system replicates, propagating the initial building bias and affecting the sampled conformational space. Nevertheless, the existence of such variability is a testament to the conformational complexity of pH-sensitive molecules in membrane environments. Therefore, it is crucial to capture the underlying electrostatics at play that determine peptide thermodynamics, kinetics and protonation of key residues. By slicing our systems according to the Asp14 residue insertion, it was possible to estimate pK_a and average protonation values along the membrane normal, and even discriminate between phosphate and non-phosphate interacting conformations. We observed that, in some systems (352 and 192 lipids) the persistent phosphate interactions are increased, which stabilized the neutral state of Asp14, thus increasing the pK_a values, compared to other systems (256 and 160 lipids). In spite of the non-equilibrated interactions observed mostly in the -2 to -4 Å insertion regions, the pK_a profiles at deeper regions still predict the experimental pK of insertion, since the electrostatic perturbation decays as the residue inserts further. Therefore, the electrostatic variability at those regions do not significantly impact the pK_a estimations at the last moment where residues still sense the bulk water pH.

Many system properties are not directly influenced by these harder-to-equilibrate interactions and readily converge, which allowed us to evaluate box-size effects. Reducing system size affects the membrane overall structure where the peptide induces local and long-range perturbations. The lack of bulk lipids in the more truncated systems hinders the membrane ability to accommodate the pHLIP-induced perturbations, while also enhancing the observed variability. These results support the need for large membrane patches (256 or even 352 lipids) in the setup of these transmembrane peptides, even at the cost of using more computational resources. For future simulations, the pH-replica exchange methodology showed to be a promising enhanced sampling technique, which guarantees a faster sampling of the protonation equilibria of pH-sensitive peptides while interacting/inserting into membranes. The improvements in our predictive ability using pHLIP can leverage the development and expand the pathway to clinically relevant peptide mutants in more complex and realistic membrane models.

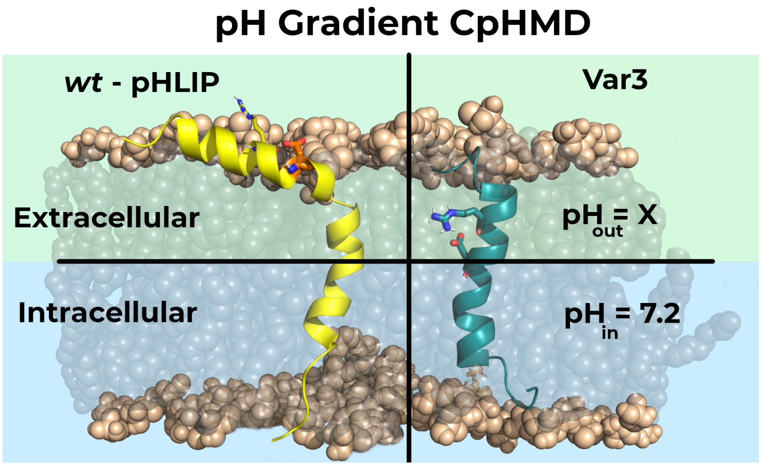
3.7 Acknowledgments

We thank Yana K. Reshetnyak, Oleg A. Andreev, and Don Engelman for fruitful discussions. Special thanks to Pedro B. P. S. Reis for helping with the development of the pK_a profiling tool. We acknowledge financial support from Fundação para a Ciência e a Tecnologia through grants SFRH/BD/140886/2018 and CEECIND/02300/2017 and projects PTDC/BIA-BFS/28419/2017, UIDB/04046/2020, and UIDP/04046/2020. This work also used the EGI infrastructure with the dedicated support of IN2P3-IRES, and NCG-INGRID-PT.

Chapter 4

A comparative study of pH gradient effects on pHLIP peptide variants

Tomás F. D. Silva[1], Diogo Vila-Viçosa, Miguel Machuqueiro



Contents

---

[1] In this work, I performed all the pHRE simulations, all the analyses, had a significant contribution to the results discussion, and wrote the initial manuscript.

4.1 Context

The present chapter successfully undertook one of the proposed goals in applying our newly-developed pHRE protocol to study the molecular intricacies dictating the overall behaviour of transmembrane peptides. The method was applied to the well-understood *wt* peptide and one of its most promising variants (Var3), which experimentally under-performed or excelled depending on the type of media (liposome or *in vivo* tumor cells, respectively). Simpler liposomal models preclude crucial cell properties, like the pH gradient, for the sake of homogeneity. To introduce the pH gradient concept, a previous PB/MC pH gradient implementation by Baptista et. al.[235] was adapted to our CpHMD-L method coupled with the pHRE scheme. In principle, the presence of a pH gradient should significantly alter the peptide-membrane configurations relative to a simple liposomal medium. These changes might impact the microenvironment of key insertion modulators, affecting the peptide thermodynamic stability in the optimum pH range, hence explaining the varying performances. By the end of this chapter, our findings should highlight the robustness and accuracy of our pK_a predictive model in reproducing experimental data, while further rationalizing the molecular minutiae behind the therapeutic performance of pHLIP peptide variants. The work developed throughout this chapter was pivotal in clarifying the roles of particular residues that are usually mutated in peptide design protocols, and establishing the pH gradient implementation as another fundamental step in bridging *in silico* data and *in vivo* cell experiments. This work has been submitted to the Journal of Chemical Theory and Computation.

4.2 Abstract

The pH-low insertion peptides (pHLIP) are pH-dependent membrane inserting peptides, whose function is dependent on the cell microenvironment acidity. Several peptide variants have been designed to improve upon the *wt*-sequence, particularly the state transition kinetics and the selectivity for tumor pH. The variant 3 (Var3) peptide is a 27 residue long peptide, with a key titrating residue (Asp-13), that despite showing a modest performance in liposomes (pK^ins $\sim$5.0), it excelled in tumor cell experiments. To help rationalize these results, we focused on the pH gradient in the cell membrane which is one of the crucial properties that are not present in liposomes. We extended our CpHMD-L method and its pH replica-exchange (pHRE) implementation to include a pH gradient and mimic the pHLIP-membrane microenvironment in a cell where the internal pH is fixed (pH 7.2) and the external pH is allowed to change. We showed that, by properly modeling the pH-gradient, we can correctly predict the experimentally observed loss and gain of performance in tumor cells experiments by the *wt* and Var3 sequences, respectively. In sum, the pH gradient implementation allowed for more accurate and realistic pK_a estimations and was a pivotal step in bridging the *in silico* data and the *in vivo* cell experiments.

4.3 Introduction

The pH-low insertion peptides (pHLIP) constitute a family of long transmembrane peptides ($\sim$27–36 amino acid residues), whose function and folding depend on the extra-cellular acidity to adopt a membrane-inserted state[52, 101, 121, 123, 136, 222]. This transmembrane configuration is characterized by an α-helical fold, usually adopted at pH values below $\sim$6.0[101, 103], and a kink in the water-membrane region [103]. At more basic conditions, pHLIP peptides typically unfold to a random coil conformation, either adsorbed to the membrane surface (pH 7.0 to 8.0) or in solution (pH > 8.0)[52]. Although the folding process has been challenging to describe, as shown in recent work regarding the possible insertion transition states [236], the characteristic pH dependency relates to key titrating residues identified as regulators of this insertion process, such as Asp-14 in the *wt* peptide, and their local interactions with the surrounding environment. Indeed, during the design process of several variant sequences, an aspartate residue located in a key location was found to be essential [123]. In equilibrium, this residue is able to access the water-membrane interface region, at least transiently, and act as a pH probe, sensing when pH is high enough to trigger deprotonation and membrane exiting.

pHLIP *wt* peptide has shown promising targeting features for tumors, inflammation tissues and ischemia [221, 222] or as a model peptide to study the molecular interactions regulating transmembrane peptide thermodynamic stability[103, 104, 139, 237]. However, widespread therapeutical use of *wt* pHLIP has not been adopted owing to a lack of tumor specificity and slow kinetics of insertion in *in vivo* studies [136]. To overcome these limitations, a systematic study with sixteen pHLIP peptide variants was designed in the Andreev Lab with a thorough assessment of their liposome membrane partitioning, peptide stability, and membrane insertion pK values (pK^{ins})[123]. In subsequent *in vivo* studies, the tumor targeting ability and organ distribution of some of these variants were also measured [238]. Among all tested peptides, Variant 3 (Var3) was a distinct case that showed poorer performance than the *wt* peptide in liposome models, however, it excelled in coupling faster kinetics with improved tumor/kidney target ratio in *in vivo* studies[123, 124, 238]. The Var3 peptide is a 27 amino-acid long peptide with a key aspartate (Asp-13), a shorter α-helix, and fewer C-terminus titrating acidic residues. On one hand, this reduced number of anionic residues in the C-terminus region improves the kinetics rate of membrane insertions by facilitating their transient protonation. On the other hand, the smaller pK^{ins} (5.0), measured in liposomes[123], would suggest that the tumor-targeting ability is impaired as only a small fraction of the peptide would accumulate in tumor cells (pH $\sim$ 6.0–6.8). Nevertheless, fluorescence imaging data has shown improved therapeutic index relative to the *wt*[125], reduced liver accumulation, faster organ clearance[124], the wider time window for imaging, and the best tumor/organ ratio of all variants[123]. This data strongly indicates a higher pK^{ins} *in vivo* than the one observed for liposomes.

The pHLIP therapeutic performance depends on a good match between the pK^{ins} value and the tumor microenvironment (TME) acidity[22, 40], yet the aforementioned results highlight a fundamental flaw in extrapolating liposome pK^{ins} to cells and *in vivo* experiments. Indeed, promising peptide sequences may have been discarded on the account of poor liposomal performance or the pK^{ins} falling outside the optimum pH region. Nevertheless, studies with liposomes are quite standard and provide essential data to validate *in silico* results, which, as expected, use very simple membrane models[103, 104, 140].

In silico studies of pHLIP peptides have successfully provided novel insights, with molecular-level detail, into pHLIP function and structure, prompting the rational design of fine-tuned variants. Several studies focused on the structural stability [140, 142], (de-)insertion kinetics, possible meta-stable states and the electrostatic interactions modulating the pH dependency[102–104, 143]. Most of them were performed using 2-oleoyl-1-palmitoyl-*sn*-glycero-3-phosphocholine (POPC) membrane bilayers to mimic liposomal conditions, such as ionic strength, pH, and peptide state)[140, 142, 239]. Particularly, constant-pH molecular dynamics (CpHMD) simulations are advantageous as the residues are titrating at specific pH values, allowing the conformational and protonation sampling to be coupled [27, 37, 103, 104, 151, 158, 159, 163–165, 180]. These methods promote the study of complex peptide-membrane configurations and the impact of the key titrating residues positioning along the membrane normal on helical (un-)folding and side-chain interactions. The electrostatic environment around the key aspartate residues changes with the peptide movement, favoring either insertion/exiting processes through protonation/deprotonation events, respectively. Additionally, the transient protonation of C-terminus anionic residues seems to play a major role in the associated kinetics of transition between the inserted and adsorbed states [101, 103]. Our previous work focused on describing the electrostatic network dictating the pHLIP-membrane thermodynamic equilibrium in state III of *wt* and its L16H peptide variant [103]. We also improved the peptide-membrane configuration sampling by coupling a replica-exchange scheme to the CpHMD methodology (pHRE method) [104] and validated it by comparing multiple pK_{a} profile calculations of Asp-14, in differently sized membrane patches, with the experimental pK^{ins} 6.0.

The computational models we developed are state-of-the-art in the study of pH-sensitive peptides interacting with lipid bilayers. Nevertheless, there are still several avenues that can, and should, be improved to increase the realism of the process. Indeed, our model should mimic more closely crucial cell features that are usually modified in tumors. As an example, the transmembrane pH gradient, that exists between the intracellular and extracellular compartments [235, 240], has an increased role in tumors, due to the TME formation[22]. Furthermore, this gradient is absent in liposomes, affecting pHLIP-membrane equilibrium, $\mathrm{p}K_\mathrm{a}$ shifts in key residues, and, ultimately, the accuracy of our predictive model when comparing liposomes and cell experiments. The development of a transmembrane pH gradient method within pHRE is, to the best of our knowledge, the first effort to perform pH gradient simulations of transmembrane peptides. This newly-developed protocol seeks to mimic *in situ* conditions to improve our pHLIP-membrane model accuracy and robustness while helping bridge the gap between *in silico* models, liposomes, and living cells.

We successfully implemented the pH gradient within our pHRE methodology and significantly improved our model of the local residue interactions that define peptide performance in cellular environments. We also evaluated how this more realistic description of the process may impact our model's accuracy and predictive ability. With this goal, we calculated $\mathrm{p}K_\mathrm{a}$ profiles, in gradient and non-gradient conditions, and performed a detailed quantification of the interactions between the key aspartate and the neighboring electrostatic partners along the membrane normal. This systematic analysis of *wt* and Var3 peptide variants under a pH gradient setup proved pivotal in improving our knowledge of the local changes in the interaction networks that propelled their distinct *in vivo* performances.

4.4 Methods

4.4.1 System setup and pH-gradient implementation

For this work, two systems were simulated using a pre-equilibrated membrane-pHLIP structure composed of the *wt* (ACEQNPIYWA**R**YA**D**WLFTTPLLLLDLALLVDADEGT)[52] and the Var3 (ACDDQNPW**R**AYL**D**LLFPTDTLLLDLLW) sequences[123] across a 256 (*wt*) or a 160 (Var3) membrane bilayer of 2-oleoyl-1-plamitoyl-*sn*-glycero-3-phosphocholine (POPC) molecules, respectively. For the *wt* system, the peptide was already placed in a kinked alpha-helical fold, creating two segments above the 15th and below the 18th residues[103]. For the Var3 sequence, the peptide was allowed to equilibrate through a two-fold optimization protocol: first, a molecular dynamics (MD) simulation using position restraints (1000 $\mathrm{kJ/mol\ nm^2}$) on the peptide to allow the lipids to accommodate the peptide conformation; second, an unrestrained CpHMD simulation, at pH 6.0, to equilibrate both the conformation and the protonation states of the titrating residues.

pHRE is an enhanced sampling method extended from the CpHMD-L methodology[27, 151, 158, 159, 165], which can be described in three modules: a Poisson Boltzmann/Monte Carlo (PB/MC) step from which new protonation states are generated for titrable groups using PB-derived free energy terms; a solvent relaxation step where solvent molecules (SPC water model) are allowed to adapt to the new protonation states; and a final molecular mechanics/molecular dynamics (MM/MD) production step where new conformations are sampled. The Baptista's PB/MC implementation of the pH gradient[235] was integrated into our pHRE methodology. In this setup, the membrane center defines the border between the two regions with different pH values. Prior to the simulation, an outer pH ($\mathrm{pH_{out}}$) is assigned to every residue whose starting Z position is above the membrane center, while residues below the membrane center are assigned to an inner pH ($\mathrm{pH_{in}}$).

To perform non-gradient pHRE simulations, n simultaneous CpHMD-L simulations, known as pH replicas, are running, each with an assigned pH value. During the MM/MD procedure, the replicas are stopped to allow a pH exchange attempt between adjacent values, with a fixed frequency (τ_{RE}). When using the pHRE gradient setup, only the external ($\mathrm{pH_{out}}$) is chosen from the set of n pH values, since $\mathrm{pH_{in}}$ is fixed at the physiological value (7.2).

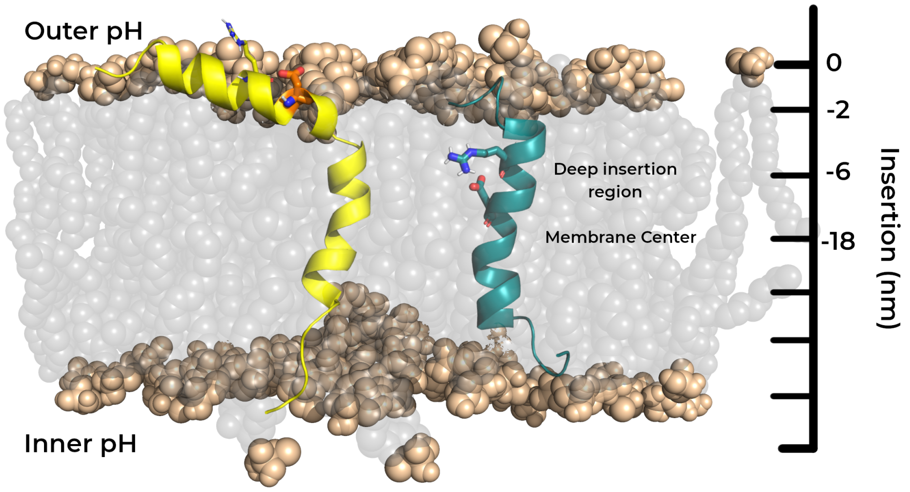


Figure 4.1: Graphical representation of *wt* (left) and Var3 (right) peptides inserted in a POPC bilayer. Although the two peptides are illustrated together, in the same lipid bilayer, they are usually modelled in separate membrane patches. The *wt* and Var3 peptides are represented in yellow and teal, respectively, while the phosphate groups (both the P and O atoms) of the membrane are represented as light brown spheres. The key pH sensor residues, Asp-14 and Asp-13, and the two important arginine residues (Arg-11 and Arg-9) are also highlighted as sticks.

Therefore, only $\mathrm{pH_{out}}$ is allowed to exchange between replicas. If a pH replica-exchange move is accepted, according to the probability given by Equation 1, conformation and protonation are swapped between different pH values, increasing the variability of our sampling at low and high energy states for every replica.

$$P_{acc} = min\left\{1, exp[-(\mathrm{pH}_m - \mathrm{pH}_l)(N(x_i) - N(x_j))ln10]\right\} \tag{4.1}$$

$\mathrm{pH_m}$ and $\mathrm{pH_l}$ are the exchanging pH values, $N(x_i)$ and $N(x_j)$ are the number of protonated groups for the x_i and x_j states. Ten replicates of 100 ns each were simulated for both peptide systems in the gradient and non-gradient setups, running a total of 4 system combinations: *wt* gradient and *wt* non-gradient ; var3 gradient and var3 non-gradient. The initial 50 ns of each replicate were discarded to ensure a good equilibration of the structural properties. Each replicate consisted of four pH replicas, each value given from the range of 4.0 to 7.0 with a step of 1.0. Each CpHMD cycle was 20 ps (τ_{prot}), with a relaxation step of 0.2 ps (τ_{rlx}) and the pH exchanges were attempted at every 20 ps (τ_{RE}) out of phase from τ_{prot}[166].

The titrating residues were the (N- and C-) termini and: *wt* – Asp-14, 25, 31, 33, and Glu34; Var3 – Cys2, Asp3, 4, 13, 19, 24, and 27. The peptide starting conformations, for each replicate, were obtained from the final segment of the previously mentioned equilibration protocol. The *wt* peptide showed the typical conformation with a partial loss of helical content around the residues 15-18, as previously observed[103]. However, Var3 showed a significantly different peptide conformation, probably due to its different size and amino acid sequence (Figure 4.1).

4.4.2 Molecular Dynamics Settings

Every system was simulated using a modified version [156, 194] of the GROMACS 4.0.7 package[227], the GROMOS 54A7 force field[172] and a Python-based wrapper to apply the pH replica-exchange method[104, 166]. The non-bonded interactions treatment was done with a single-cutoff scheme updating the forces, at every step, for all pairs below a 14 Å cutoff[180]. Regarding long-range interactions, the van der Waals interactions were truncated at 14 Å, while the electrostatic interactions were treated with a generalized reaction field (GRF) method, using a dielectric constant of 54 [228] and ionic strength of 0.1 M. Lipid and peptide bond lengths were constrained using the P-LINCS algorithm [229] and water molecules are treated as simple point charges (SPC) [230], using the SETTLE algorithm [231]. In MD simulations, the integrator time step used was 2 fs and conformations were generated from an NPT ensemble. The v-rescale temperature bath [189], at 310 K, was coupled separately to the solute (peptide and membrane) and solvent with a relaxation time of 0.1 ps. The system pressure was kept constant at 1 bar with a Parrinello-Rahman barostat [232], with a relaxation time of 5 ps and compressibility of 4.5×10^{-5} bar^{-1}.

4.4.3 Poisson-Boltzmann/Monte-Carlo Simulations

Poisson-Boltzmann calculations were performed with the Delphi V5.1 program [191] using both partial charges and Lennard-Jones parameters of the GROMOS 54A7 force field to derive the atom radii at 2 RT[233]. The peptide-membrane molecular surface was described using a 1.4 Å radius probe, an ion-exclusion layer of 2.0 Å, and ionic strength of 0.1 M. The dielectric was set as 2 and 80, for solute and solvent, respectively. The conducted two-step focusing procedure employed two 91 point grids, where the coarse grid had a $\sim$1 Å spacing between grid points, and the smaller grid had a spacing of $\sim$0.25 Å. In the coarse grid, the relaxation parameters were 0.20 and 0.75 for linear and nonlinear iterations, respectively, while periodic boundary conditions were applied in the *xy* plane. Background interaction calculations were truncated at 25 Å and the electrostatic potential convergence threshold was 0.01 kT/e.

Monte-Carlo (MC) calculations sample the protonation states of all titrating residues using a modified version of the PETIT program that implements the pH gradient setup, as explained in the cited protocol [235]. In this new version, the program requires a pre-assignment of each titrating site to one of two proton baths, each with a distinct pH value, representing the inner monolayer (pH_{in}) or the outer (pH_{out}) monolayer [235, 240]. The site assignment required a previous insertion analysis, for each titrating residue using the unrestrained CpHMD equilibration data, to determine which monolayer was being populated on average to choose the correct proton bath. The analysis was done for both peptides, where Asp25 (*wt*) and Asp-19 (Var3) assignment needed to be more cautious since their close position to the membrane center allows crossing between both monolayers. Due to previous data on the role of the membrane center aspartate [103] and their average positions, both residues were assigned to the outer monolayer, where Asp25 and Asp-19 may impact the membrane insertion of Asp-14 and Asp-13, respectively. Regarding the non-pH gradient simulations, all titrating residues are exposed to a single proton bath at a given pH value in the previously established pH range. The PB protonation energy terms and the assigned pH of the residue (pH gradient) or the solution pH (non-pH gradient) are used in a Metropolis scheme to calculate the probability of protonation change. Proton tautomerism was taken into account for all titrable groups. For each conformation, 10^5 MC cycles were performed, where each cycle corresponds to a trial change of each individual site and pairs of sites with an interaction larger than 2 p*K* units.

4.4.4 Analyses and error calculations

In this work, Asp-13 (Var3) and Asp-14 (*wt*) protonation is the major membrane insertion trigger. These events are modulated by the phosphate interaction shell and other relevant properties, affected by the peptide-membrane configuration, that require specialized calculations.

The equilibration of membrane-related properties, such as bilayer thickness and residue membrane insertion, in peptide-membrane complexes, is very convoluted due to the local deformation induced by the peptide. The standard bilayer thickness calculations use the arithmetic difference of the average Z positions of both monolayer lipids, hence affected lipids dilute the bilayer thickness values and prevent a fair assessment of the membrane health. To overcome this issue, our equilibration analyses focused on two membrane regions: the local deformation and the 'bulk' unaffected lipids - lipid distance to the peptide > 15 Å. In our approach, we quantify the local deformation and discriminate these two regions by calculating the half thickness values, for each monolayer, of an annulus region - defined by two radii centered on the peptide. An annulus scan on the *xy* plane describes the membrane monolayer outline which, at longer distances, should converge to the experimental POPC half thickness range. All equilibrated conformation snapshots are considered in the calculations and the experimental POPC half thickness range was obtained by interpolating from experimental bilayer thickness measurements in the fluid range at different temperatures [214]. The membrane local deformation also presented in this work was calculated as the difference between the local half thickness and the half thickness of the bulk region (beyond the 15 Å cutoff).

To obtain the necessary residue membrane insertion data, analyses are performed by defining the closest membrane monolayer surface as the average Z position of the lipid phosphate group (P and O) atoms within a 6 Å radius from the group of interest, and then calculating the relative position of the residue to the reference. Additionally, to properly account for membrane deformations, the chosen radius has to be small enough to exclude atoms outside of the perturbation affecting the average estimation. Note that an excessively small radius has the downside of lacking enough atoms to properly characterize the membrane interface. To overcome this issue, our current protocol enforces a minimum of 10 atoms (phosphorous and/or oxygen), within the cutoff radius, to define the average Z coordinate for the membrane surface. Otherwise, the 10 closest atoms to the group of interest, in a two-dimensional (*x*/*y*) plane, will be used regardless of the cutoff distance.

The calculated insertion time-series is then coupled to the corresponding residue protonation states to obtain pK_a and protonation profiles, a representation of the residue pK_a or average protonation values along the membrane normal. These profiles are achieved by assigning the residue protonation values, of each conformation, to an insertion bin according to their respective insertion value. By discriminating the protonation data along insertion windows, it is possible to estimate the residue proton binding affinity (pK_a) or average protonation (at each pH value) at those particular membrane regions. For reproducibility and robustness purposes, all calculations must fulfill a few criteria: (1) each insertion slice (at each pH value and replicate) must have a minimum of 10 conformations of each protonated state; (2) at least three replicates need to contribute for the conformational sampling at each pH value and at least two pH values are required for the fit; (3) to ensure monotonicity, the average protonation should not fluctuate (by 0.05) above the average of the previous lower pH. If the conditions are met, the average protonations at each pH are calculated, then they are fitted to the Hill equation to obtain the pK_a values. For protonation profiles, only the first two criteria are required. The profiling procedure is very useful to ascertain the influence of the changing electrostatic environment on protonation by measuring other residue properties along the membrane depth. Therefore, the procedure was also applied to distances between the residues of interest and neighboring electrostatic partners, such as phosphate groups, water molecules, and other residues. The GROMACS tool package was used to calculate these distances. All analyses were done using in-house software (`http://mms.rd.ciencias.ulisboa.pt/#software`) and the GROMACS package.

To circumvent fitting issues, all pK_a error values were estimated with a Bayesian bootstrap approach that performed 1000 bootstraps from our average protonation samples and, in each bootstrap, each sample was assigned a random weight. This approach must fulfill the previously applied criteria to obtain the final pK_a and error values. A simple standard error of the mean method was used to estimate all the other properties' error values.

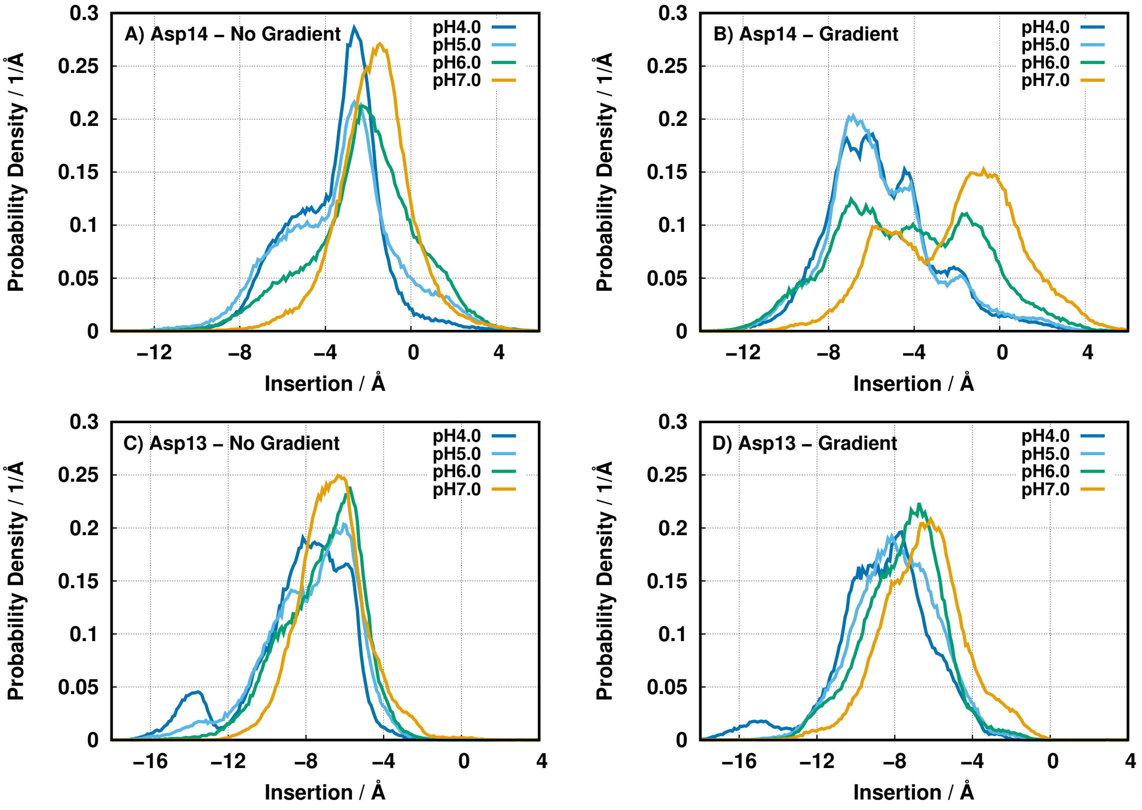


Figure 4.2: Probability density function of residues Asp-14 (A and B) and Asp-13 (C and D) populating a given insertion region in a non-gradient (A and C) and gradient setup (B and D) at all pH values studied. Negative insertion values correspond to the membrane interior.

4.5 Results and discussion

4.5.1 Peptide/Membrane Structural Analyses and Equilibration

Liposome studies have proven to be a reliable and affordable technique to evaluate the features of new transmembrane (TM) peptides sequences in the early development stage. However, most liposomes lack the environmental and complex traits that characterize the cell microenvironment, hindering the transferability of the TM peptides' performance to *in vivo* conditions[134]. Similarly, *in silico* models also suffer from the same constrictions. Most CpHMD methodologies mimic those liposomes' proton availability by applying an equal pH value to the inner and outer monolayer environments. However, with the implementation of a pH gradient in pHRE, it is possible to decrease the gap between most liposome- and cell-like conditions. The *wt*-pHLIP peptide has already been successfully studied in a gradient-free environment [103, 104], where pHRE simulations reproduced the data from experiments performed in liposomes. In a pH gradient setup, the sampling of the peptide-membrane configuration space should be distinct from the previous simulations ensembles, as the sampled protonation space is strongly coupled to the conformation sampling, reflecting on the peptide and key residue properties.

The *wt* Asp-14 residue membrane insertion positions clearly highlight the pH gradient method by propelling Asp-14 to sample more abundantly the deeper membrane regions (-10 to -4 Å) relative to the no-gradient setup (Figure 4.2A–B). This difference is noticeable even at pH 7.0, where the pH gradient effect is minimal and the two configurational ensembles should tend to converge. This indicates that the pH replica-exchange protocol is effectively promoting the mixing between conformational ensembles and allowing a slightly better sampling of

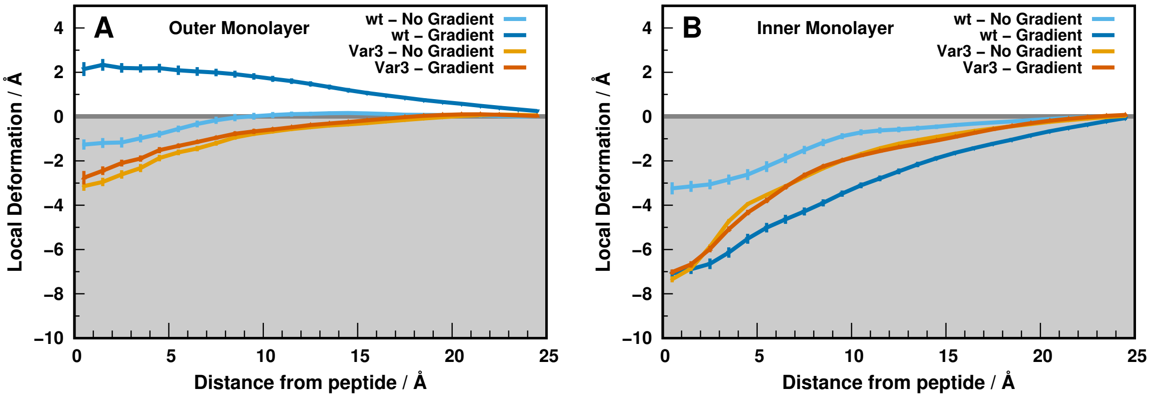


Figure 4.3: Outer (A) and inner (B) local monolayer deformations induced by the *wt* (blue/cyan) and Var3 (orange/brown) peptides, in both setup conditions at pH 6.0. These deformations values are calculated from the half thickness values calculated along the *xy* plane distance to the peptides (Figure B.1 of Supporting Information). The maximum distance shown (~ 25 Å) corresponds to the unperturbed 'bulk' lipids. In the replica-exchange scheme, the configurational sampling is shared across all pH replicas, hence the local deformation profiles at pH 6.0 do not differ significantly from pH 4.0, 5.0, and 7.0 (see Figures B.2–B.4 of Supporting Information). The grey-shaded region corresponds to the membrane interior. The error bars were obtained from all replicates using the standard error of the mean (SEM) and calculated/represented only every 1 Å for clarity.

the deeper regions at this high pH value. The Var3 Asp-13 residue, which is structurally and functionally equivalent to Asp-14 in *wt*, possesses a neighbouring arginine residue (Arg-9), that is directly above its location in the α-helix (see Figure 4.1). The shorter peptide length (27 residues) imparts a different peptide structural disposition in the membrane since the shorter helix does not grant much leeway to span the peptide across the bilayer and promotes a higher level of membrane insertion (Figure 4.2C–D). For this peptide, the presence of a pH gradient in the membrane seems to have only a small effect on the Asp-13 position in the membrane.

The distinct membrane behaviors of *wt* Asp-14 may be the result of rearrangements in the peptide-induced local monolayer deformations, structural changes of the peptide itself, or a combination of the two phenomena. Looking at the local membrane deformation (Figure 4.3), it is evident that both the inner and outer peptide-membrane configurations are significantly altered by the pH gradient at pH 6.0. In the gradient setup, the higher pH_{in} (7.2) induces significant deprotonation of the anionic residues (Asp31, Asp33, Glu34, and C-ter), prompting a larger deformation due to charge repulsion with the phosphate groups coupled with increased water solvation. This membrane invagination of the inner monolayer seems to trigger a small loss of helical content in the peptide segment (Figure B.5 of Supporting Information) and also affects the peptide-membrane equilibration in the outer monolayer. We observe a small protrusion in the outer monolayer (~2 Å) enveloping the peptide, suggesting an increase in the distance spanned between the two endpoints of the transmembrane segment, which was confirmed by a small decrease in the peptide helical content (Figure B.5 of Supporting Information). In the Var3 peptide, the gradient setup does not change the already large local membrane perturbation, compared to the non-gradient setup (Figure 4.3). Furthermore, this shorter variant requires more structural unfolding of the α-helix (mainly on the C-terminus) to cross the lipid bilayer (Figure B.5 of Supporting Information). There are small differences between setups, even at pH 7.0 where the gradient effect should have dissipated. This suggests that the Var3 C-terminus region helicity is quite sensitive to the equilibration procedure. However, these different configurational ensembles do not seem to have any major impact on the remaining structural properties around Asp-13 (Figures 4.2C–D and B.1–B.6 of Supporting Information).

Overall, the *wt* peptide in state III seems more thermodynamically unstable in a gradient than in the non-gradient

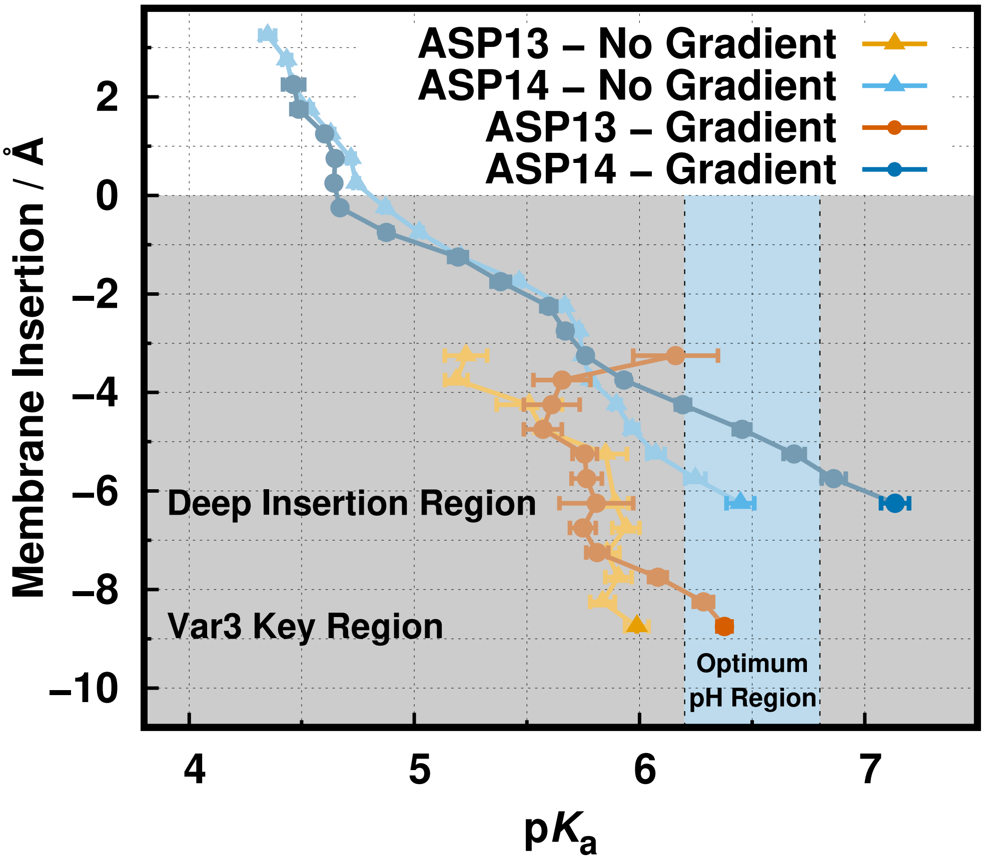


Figure 4.4: pK_a profiles of Asp-14 (*wt*) and Asp-13 (Var3) along the membrane normal for both simulation setups: gradient and no gradient. The white and grey-shaded regions correspond to the water phase and membrane interior, respectively. The light blue vertical stripe identifies the pH region ideal for TME selection. The Asp-14 no gradient data was adapted from reference[104].

setup. These high energy configurations seem to be predominant in the Var3 peptide, independently of the gradient setup used. The membrane deformations observed may alter the electrostatic network around the Asp residues, possibly changing their insertion pK_a values and respective peptide performance.

4.5.2 Asp-14 and Asp-13 Membrane Insertion pK_a Profiles

The therapeutic performance of pHLIP peptides strongly depends on their ability to specifically penetrate the tumor membrane at acidic pH environments (6.2 to 6.8). Therefore, we need to calculate the pK_a profiles of the key aspartate residues (Asp-14 and Asp-13 for *wt* and Var3, respectively) and compare them with the available experimental data, being either the peptide *in vivo* performance or the lysosome pK^{ins} values. The Asp-14 pK_a profiles are remarkably similar in both gradient setups (Figure 4.4), notably following the same expected trend where the pK_a shifts towards higher values, induced by desolvation effects [27]. However, in the key deep membrane region (-5 to -6 Å), both profiles diverge to rather distinct pK_a values, 6.4±0.1 and 7.1±0.1 for non-gradient and gradient, respectively. This indicates that different peptide-membrane configurations are being sampled at those residue insertions, confirming our initial assessment based on structural analysis. Moreover, these pools of configurations being sampled at the deeper insertion regions seem to be very homogeneous and/or well-mixed by the pHRE protocol in terms of both the conformations and the protonation states, which may explain the relatively small error bars in the pK_a profiles. Interestingly, the gradient setup insertion pK_a value falls outside the optimum pH region (light blue region), being in qualitative agreement with the experimental loss of performance of the *wt* peptide in cell experiments[123]. Indeed, this performance can be correlated to the Asp-14 pK_a shift and the overall thermodynamic stability. The significant pK_a shift ($+\sim$0.7 pK units) in the gradient setup indicates an increase in interactions with negatively charged groups, such as oxygen atoms from phosphates, fewer interactions with positively charged residues, such as the nearby Arg-11, or even choline groups, or a larger desolvation effect at similar membrane depths.

The gradient and non-gradient Asp-13 (Var3) profiles follow the same pK_a increase, like the *wt* peptide (Figure 4.4). There is a clear shift in the region sampled by this peptide (Figure B.6 of Supporting Information), which results in an incomplete profile at more shallow (well-solvated) regions and a pK^{ins} region that is deeper (-8 to -9 Å) than in the *wt* peptide (-5 to -6 Å). The calculated pK_a values (6.0 ± 0.1 and 6.4 ± 0.1 for non-gradient and gradient, respectively) at the deepest insertion region are in qualitative agreement with the experimental data since the gradient value falls within the TME optimum pH region[145, 146]. Nevertheless, the pK_a value estimated for liposome-like conditions is overestimated relative to the experimental pK^{ins} (5.0). The pK_a value of an unperturbed Asp residue at the water/membrane interface is usually $>$6[27, 103]. Therefore, we argue that to obtain such a lower value (5.0) it is required an interaction by a positive residue, like Arg-9, which our model does not seem to fully capture. A such contribution would need to be selective for the non-gradient setup, the one that seems to fall short of the experimental data.

The pK_a values of the key Asp residues change in response to desolvation (membrane insertion) and to the neighboring electrostatic interactions. For these peptides, the most relevant players that establish strong interactions are the lipids phosphate and choline groups, the water molecules, and the nearby arginine residue (Figure 4.5). The observed *wt* Asp-14 pK_a shift in the gradient setup ($+\sim$0.7 pK units) originates from a distinguishable electrostatic balance between this residue and the neighboring electrostatic groups. Since the number of interacting phosphate groups is the same in the two setups in the key membrane region (-5 to -6 Å) and Arg-9 remains relatively far away (Figure 4.5A and E), this effect seems to result from a slightly more pronounced desolvation effect coupled with a loss of interacting cholines relative to the non-gradient setup (Figure 4.5C and G). The fewer choline groups within the first interaction shell may be the determinant factor, resulting in a less positive Asp-14 electrostatic vicinity in a pH gradient setup, while embedded in an apolar membrane environment, which typically favors the protonated state of carboxylic acids, hence higher proton binding affinities.

Concerning the Var3 behavior in both setups, the major distinguishing factor stems from more abundant phosphate contributions in the gradient setup at the pK^{ins} region (-8 to -9 Å) (Figure 4.5B). This effect can be somewhat counteracted by the slightly closer Arg-11 (Figure 4.5F), but the final pK_a shift in the gradient setup ($+\sim$0.4 pK units), suggests otherwise. The remaining electrostatic contributions (Figure 4.5D and H) are indistinguishable between setups. As mentioned above, the main disagreement between our simulations and the experimental data available is in the non-gradient setup, where the experimental pK^{ins} (5.0) is being overestimated by 1 pK unit in the calculations. To reproduce the experimental data, the non-gradient simulations would require a larger contribution from a cationic partner (most likely Arg-11), which is not being correctly captured by our force field and/or our CpHMD-L simulations. If properly sampled, these interactions should reflect a stronger positive influence on Asp-13 that stabilizes its ionized form, shifting down the pK_a value, bringing it closer to the experimental pK^{ins} (5.0).

4.6 Conclusions

The pK_a values and protonation states reflect the electrostatic environment that a given titrating residue is sensing. The membrane (de)insertion process is triggered by key pH-dependent residues in the *wt* and Var3 peptides

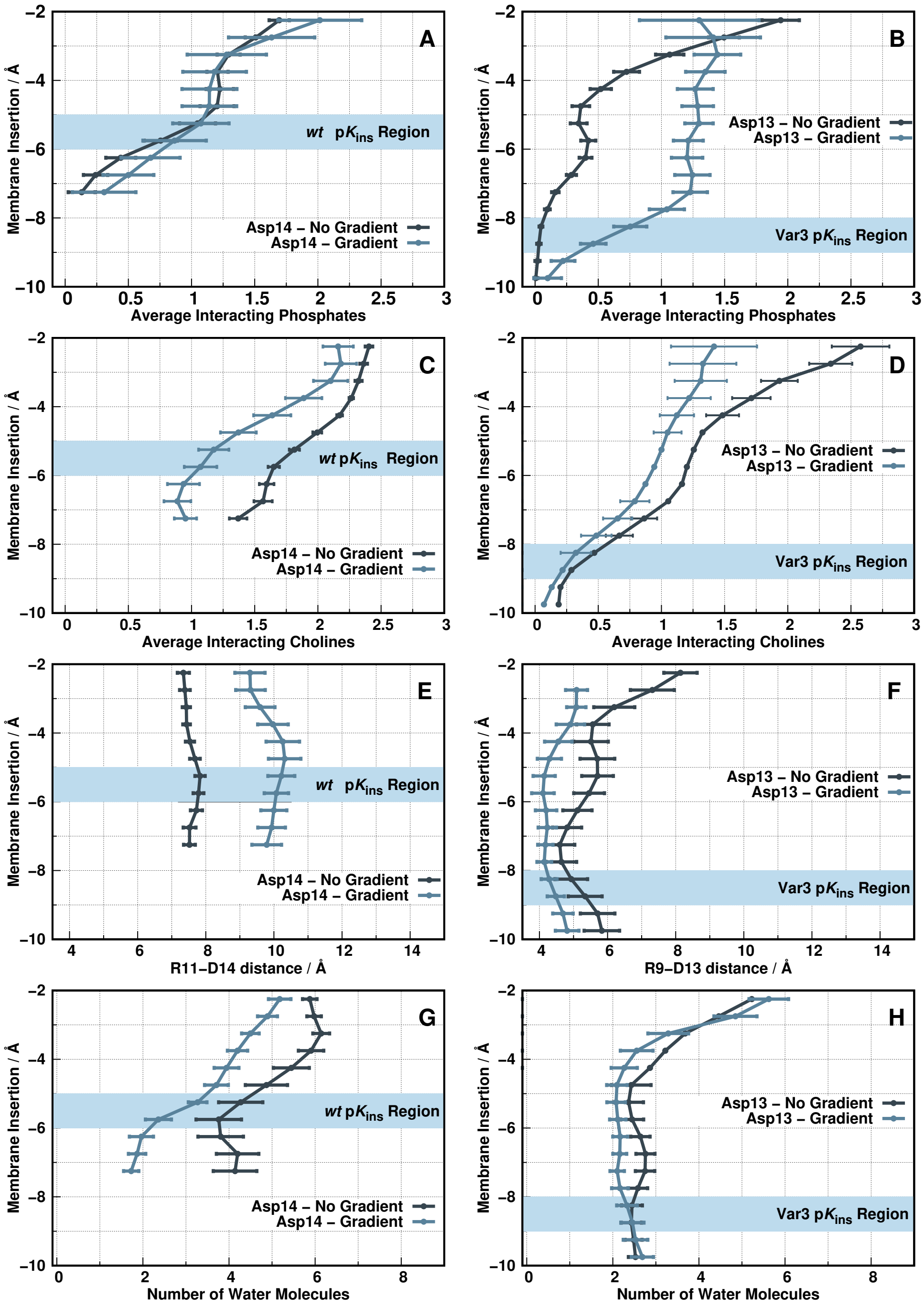


Figure 4.5: Electrostatic interactions of *wt* Asp-14 (A,C,E,G) and Var3 Asp-13 (B,D,F,H) with the surrounding molecular partners at pH 6.0 for both simulation setups. The interacting partners include the phosphate groups (A,B), choline groups (C,D), Arg-11/9 (E,F), and water molecules/desolvation (G,H). The phosphate/choline plots show the average number of interacting groups along the membrane normal. The arginine plots show the average interaction distance between Asp-14–Arg-11 and Asp-13–Arg-9 for *wt* and Var3, respectively. The desolvation plots show the average number of interacting water molecules with Asp-14 and Asp-13, respectively. The light blue horizontal stripe identifies the ideal insertion region for each peptide pK^{ins} estimation.

(Asp-14 and Asp-13, respectively), which interact with several local electrostatic partners. These include the lipids phosphate and choline groups, water molecules, and other neighboring residues, such as nearby arginines. The proton binding affinities of the aspartate residues are the product of a fine trade-off between all these interactions and an incomplete model of these contributions may result in wrong pK_a estimations and poor experimental correlation.

In this work, we extend the framework of our CpHMD-L methodology to include a membrane pH gradient setup. This novel protocol increases the level of realism when modeling cell membranes, particularly in the TME conditions where the exterior pH is significantly acidified, while the interior one is kept relatively stable (∼7.2–7.4). We showed that this pH-gradient setup impacts the configurational/protonation space of both peptides studied and helped us to rationalize the observable loss of performance of the *wt* sequence in tumor cell experiments. The Var3 peptide results also confirmed that the pH gradient setup is required to promote membrane insertion in tumor cells, which is in excellent agreement with the experiments. In sum, the pH gradient implementation was a pivotal step in bridging the *in silico* data to the *in vivo* experiments and identifying important electrostatic partners in pHLIP peptides. One of such partners, the ”anchor” arginine residues, may have an important role as a direct modulator of the aspartate electrostatic vicinity and, possibly, the overall peptide thermodynamic stability. Although it is just a plausible hypothesis, it may warrant a future systematic study focused on the role of the arginine position in these peptides and how it modulates the surrounding peptide environment.

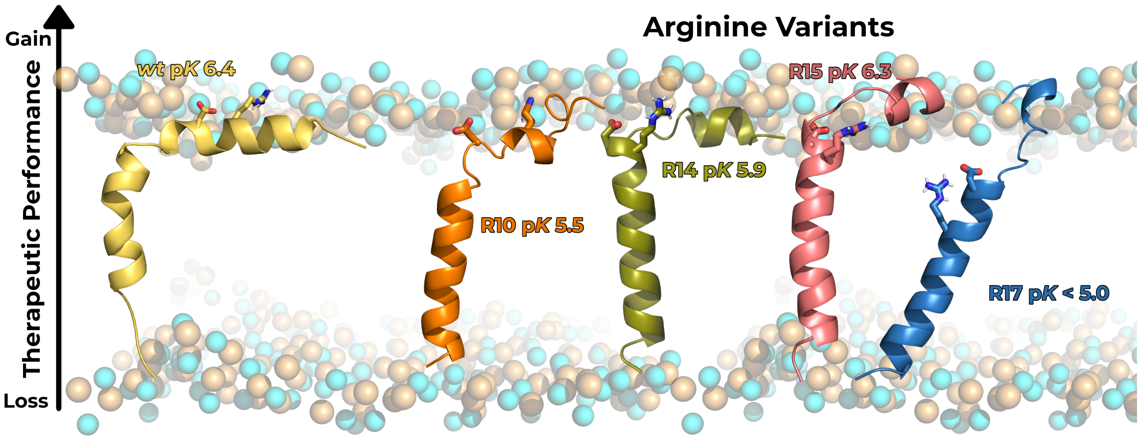
4.7 Acknowledgements

We thank Yana K. Reshetnyak, Oleg A. Andreev, and Don Engelman for fruitful discussions. Special thanks to Pedro B. P. S. Reis for helping with the development of the pK_a profiling and MembIt tools. We acknowledge financial support from Fundação para a Ciência e a Tecnologia through grants SFRH/BD/140886/2018 and CEECIND/02300/2017 and projects UIDB/04046/2020, and UIDP/04046/2020. This work benefited from services and resources provided by the EGI-ACE project (receiving funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement No. 101017567), with the dedicated support from the CESGA and IN2P3-IRES resource providers.

Chapter 5

The Role of Arginine residues in Modulating the Tumor-Targeting Performance of pHLIP Peptides

Tomás F. D. Silva[1], Miguel Machuqueiro



Contents

---

[1] In this work, I performed all the pHRE simulations, all the analyses, had a significant contribution to the results discussion, and wrote the initial manuscript.

5.1 Context

This following chapter fully delves into the notion of pH-dependent peptide design. In the previous chapter, we found that the behaviour of two unique peptides strongly depended in the intramolecular interactions of the titrating key aspartates. More specifically, we identified the crucial role of the arginine residues in defining the electrostatic microenvironment that modulated the therapeutic performances. In collaboration with the Andreev Lab (Rhode Island University, USA), a comparative study was done on the varying effects of permuting an arginine in different sequence positions upon peptide thermodynamic stability at distinct pH values. Our findings are the culmination of extensive work in developing multiple analytical tools and state-of-the-art protocols to expand the usefulness of *in silico* methods to drive a rational approach to pH-dependent peptide design. The final draft of this work is to be complemented with updated experimental findings from the Andreev Lab and submitted to an high-impact international journal.

5.2 Abstract

Most processes at the water-membrane interface often involve protonation events in proteins or peptides that trigger important biological functions and events. This is the working principle behind the pHLIP peptide technology. A key titrating aspartate (Asp-14 in *wt*) is required to protonate to induce the insertion process, increase its thermodynamic stability when membrane-embedded, and trigger the peptide's overall clinical functionality. At the core of pHLIP properties, the aspartate pK_a and protonation are a consequence of the residue side-chain sensing the changing surrounding environment. In this work, we characterized how the microenvironment of the key aspartate residue (Asp-13 in these pHLIP variants) can be modulated by a simple point mutation of a cationic residue (ArgX) at distinct sequence positions (Arg-10, Arg-14, Arg-15, and Arg-17). We used pHRE simulations to study the molecular details and the protonation behavior of each variant in the membrane-inserted equilibrium (state III). We found that the arginine residue has a major effect when it is topologically available for a direct salt bridge formation with the Asp-13. In several of the remaining variants, the arginine counteracts the lack of direct electrostatic impact with an indirect modification the local microenvironment. It promotes or hinders other electrostatic players to coexist in the interaction shell. These results strengthens the important role of cationic residues in modulating the therapeutic performance of pHLIP peptides.

5.3 Introduction

The peptides and membrane interactions are vital for several biological processes, such as molecular transport, signalling pathways and cell membrane integrity[67–69]. And, even though these processes are at the core of a wide array of research areas, the molecular interactions between protein and lipids are still hard to fully characterize in complex systems. Simple peptide models are widely popular as they are customizable, they mimic defining traits of membrane proteins and there are limitless combinations of peptides with membrane models to study different types of biological systems. The advent of these peptide models, such as GALA[105] and WALP[86], propelled the study of transmembrane peptide design, from which were identified several residue sequence patterns that determine membrane folding and insertion, for example: hydrophobic residues stretches (alanine and leucine) and outward hydrophilic residues (lysines, arginines, aspartates, glutamates)[241]. Several computational studies focused on modifying the peptide length, charged residues and hydrophobic stretches to provide molecular insight into biophysical phenomena, ranging from the peptide structural disposition in the membrane[104, 242], the formation of membrane pores[219, 243], determine possible folding pathways[100, 244] and identify key residues for peptide function[103, 245, 246].

One of the more clinically relevant peptide models is the pH-low insertion peptide (pHLIP). The pHLIP family is characterized by long ($<\sim$40 amino acids) membrane inserting peptides, whose distinguishing trait from other transmembrane model peptides is their acidity-dependent insertion[52, 123, 132, 222, 224]. pHLIP peptides possess a kinked α-helical fold, more commonly occurring below pH 6.0, otherwise the peptide adopts a random coil conformation either adsorbed to the membrane surface (pH 7.0 to 8.0) or in solution (pH > 8.0)[52]. The pH dependency results from the titrating residues that populate the water-membrane interface. By fluctuating between the phosphate region and the deeper ester region, the key residue (Asp14) undergoes (de)protonation events, which either promotes insertion or membrane exiting[103]. The proton binding affinity is a measurement of the necessary energy to protonate a given residue and there are various factors that affect this property: peptide movement, (un-)folding, intermolecular (membrane lipids) and intramolecular (side-chains) interactions[104, 246]. Although the Asp14 dictates the stability of the inserted state, other titrating anionic residues play a role in the stability, but mostly in the kinetics of transition between states[103]. In our previous work, we characterized and identified the electrostatics interactions dictating the pK_a of *wt* pHLIP[103, 104] and a more clinically relevant peptide (Var3)[123, 247] in a simple liposomal model and in a cell-like pH gradient setup[248]. We highlighted the necessity of including *in vivo*-like conditions to accurately assess the therapeutic potential of transmembrane peptide models and we also described the intramolecular interactions that dictate their thermodynamic membrane stability. One of these fundamental interactions relates to the key aspartate and a neighbouring arginine residue. Each peptide revealed distinct interaction properties, thus distinct aspartate electrostatic microenvironment, impacting the residue proton binding affinity. This arginine interaction has also been reported in other studies, highlighting its role with both the aspartate and lipid headgroups[246].

For many years, the role of cationic residues has been discussed in the context of studies about cell-penetrating peptides. They have identified a critical impact in inducing/hindering peptide insertion through morphological membrane alterations[97, 249], while also affecting the peptide structure and position within the membrane[96], including the possibility of promoting pore formations and membrane permeability[243, 249]. These effects hinge on the cationic residues behaviour in the membrane, specially for arginines and lysines, as they remain positively charged as they move along the membrane normal[95, 98]. When an α-helical peptide inserts in a membrane, the cationic residue is dragged from a energetically favorable solvent environment to an apolar lipid medium. Depending on the lipid bilayer region placement, the residue may interact with anionic (e.g. phosphate) groups, effectively working as a peptide anchor, or it can snorkel to minimize the energy cost associated with membrane embedding, pulling the peptide with it, as seen in KALP peptides[94].

Several studies have strengthen the significance of cationic residues in transmembrane peptide models and their ability to modulate the peptide-membrane equilibrium. Furthermore, the presence of more charged groups near key peptide residues directly impacts the protonation and pK_a of such residues in pHLIP. Therefore, we aim at characterizing and assessing the impact of cationic residues in the vicinity of the key trigger residues in the context of transmembrane peptide design, while also discussing the impact on the peptide-membrane equilibria, folding stability, key residues pK_a, and other relevant properties.

5.4 Methods

5.4.1 System setup and pHRE simulations

Four pHLIP variant systems were prepared, composed of 32 amino acid residues and derived from the *wt* sequence (Table 5.1). The pHLIP-membrane setups were built using as template the previous simulations of the *wt* system[104] (Figure 5.1). In each setup, the peptide variant was generated as a full α-helical structure inserted in a 256 2-oleoyl-1-plamitoyl-*sn*-glycero-3-phosphocholine (POPC) membrane bilayer. The initial

Table 5.1: pHLIP peptides sequences of the *wt* and arginine variants. The arginine and the titrating residues (including the termini) are highlighted in bold.

Variant	Sequence
wt	**A**C**E**QNPIYWA**R**YA**D**WLFTTPLLLL**D**LALLV**D**A**DE**G**T**
R10	**AD**NNPFIYA**R**YA**D**LTTFPLLLL**D**LALLV**D**W**DD**
R14	**AD**NNPFIYATYA**DR**LTFPLLLL**D**LALLV**D**W**DD**
R15	**AD**NNPFIYATYA**D**L**R**TFPLLLL**D**LALLV**D**W**DD**
R17	**AD**NNPFIYATYA**D**LTF**R**PLLLL**D**LALLV**D**W**DD**

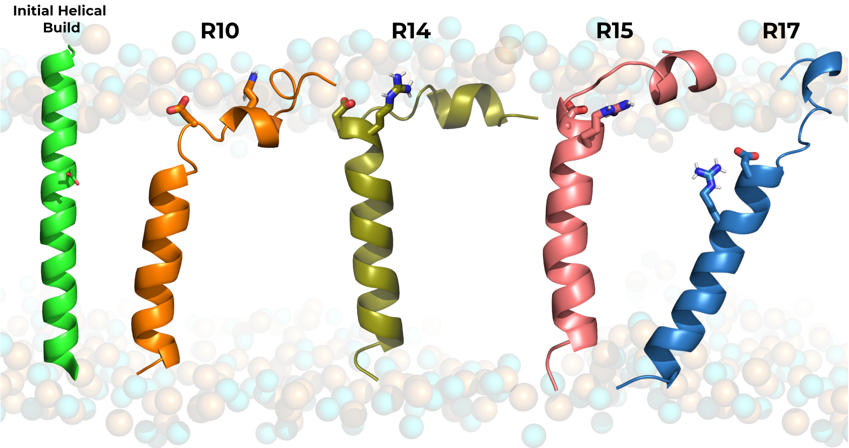


Figure 5.1: Graphical representation of the peptide conformations in the POPC membrane bilayer obtained from the CpHMD equilibration procedure. Each peptide variant is shown in cartoon with their respective color (R10 - orange; R14 - olive green; R15 - pink; R17 - blue). The unbiased full α-helical initial conformation is depicted in light green. The key Asp and Arg residues are shown in sticks.

structures were built with the key aspartate (Asp13) placed in the water-membrane interface. Although *wt* pHLIP variants, these initial structures aimed at a more unbiased approach to the system setup and equilibration protocol, since their equilibrium conformations for the inserted state may differ. After the setup, all systems were submitted to both minimization and initialization procedures, followed by a two-step equilibration protocol: the first step consisted of molecular dynamics (MD) simulations (100 ns), with the titrating residues protonation states chosen as neutral (if membrane-inserted) or charged (if solvent exposed) to promote a smoother accommodation of the surrounding lipids. Additionally, distances restraints (1000 kJ$\cdot$mol$^{-1}\cdot$nm^{-2}) were applied to preserve the integrity of the α-helix hydrogen bonds of every nth-nth+4 residues, starting in the 17th until the 28th residues of the C-terminus region, which corresponds to the region located in the membrane core. The first step of equilibration procedure using these restraints, resulted in most peptides N-terminus segment converging to the kinked α-helical conformation (Figure 5.1), similar to what has been shown for the *wt* peptide[103]. The second step of the protocol consisted of an 100 ns unrestrained constant-pH molecular dynamics (CpHMD) simulation, at pH 6.0, to enable residue titration and equilibration of both the conformation and protonation states of the titrating residues.

All systems were simulated using the pH-replica exchange (pHRE) method[104, 166]. The pHRE is an extension to the CpHMD-L methodology[27, 103, 151, 158, 159, 165, 166] that employs a replica-exchange enhanced sampling technique[167]. This scheme consists of a four-step cycle of *n* simultaneous CpHMD simulations (pH replicas), assigned to a pH value within a given pH range, that occurs as such: a Poisson Boltzmann/Monte Carlo

(PB/MC) calculation; following a molecular mechanics/molecular dynamics (MM/MD) solvent relaxation step; then a final (MM/MD) simulation, with a pH exchange step within the previous step framework. The MC calculations assign the new protonation states, using the PB-derived free energies from the system conformation of the previous cycle. The relaxation step allows the solvent molecules to accommodate to the new charged states, avoiding nonphysical spikes in the system potential energy. The final MM/MD step samples new system conformations, using the calculated protonation states. During the MM/MD step, the simulation is stopped and a pH exchange attempt occurs, at a fixed frequency of 20 ps (τ_{RE}), lagging 10 ps from τ_{prot}, between adjacent pH replicas. If the replica-exchange is accepted, according to the probability given by equation 1, the conformations and protonation states are swapped between the replicas' pH values, thus increasing the sampling variability at both low and high energy states in every replica.

$$P_{acc} = min\left\{1, exp[-(\mathrm{pH}_m - \mathrm{pH}_l)(N(x_i) - N(x_j))ln10]\right\} \tag{5.1}$$

$\mathrm{pH_m}$ and $\mathrm{pH_l}$ are the exchanging pH values, x_i and x_j are the number of protonated groups. For all systems, five replicates of 100 ns were performed, each replicate consisting of four pH replicas. The assigned pH values were in the 5.00 to 7.25 pH range, with a 0.75 pH step. The chosen pH range differs from previous works[103, 104], as according to the previous equation, a smaller pH gap between replicas improves the probability and pH exchange. In these simulations, the average exchange efficiency was of 40 % across all systems. Each replica CpHMD cycle consisted of 20 ps (tau_{prot}) steps, whereas the relaxation step was of 0.2 ps (tau_{rlx}). All systems were titrating the N- and C- termini and the acidic residues highlighted in Table 5.1. In all systems, each replicate starting conformation was obtained from the final segments of the CpHMD equilibration protocol.

5.4.2 MM/MD and CpHMD Settings

All CpHMD and pHRE simulations used a modified version[156, 194] of the GROMACS 5.1.5 package[250], the GROMOS 54A7 force field[172], while a Python-based wrapper was used to apply the pH replica-exchange method[104, 166]. Meanwhile, the restrained MD equilibration simulations were performed using the GROMACS 2020.1 package with the GROMOS 54A7 force field[172].

A single-cutoff scheme was applied for the treatment of non-bonded interactions. The forces were updated every step, for all pairs under a 14 Å cutoff[180]. Pertaining the long-range interactions, the van der Waals forces were truncated at 14 Å, while a generalized reaction field (GRF) method, with a dielectric constant of 54[228] and ionic strength of 0.1 M, was used to treat the Coulombic interactions. Both peptide and lipid bond lengths were constrained using the P-LINCS algorithm[229] and the water molecule model used was the simple point charge (SPC)[230], whose bonds were constrained with the SETTLE algorithm[231]. The integrator time step, for all MD simulations, was 2 fs and the conformations were sampled from an NPT ensemble. The used temperature bath scheme was the v-rescale[189], at 310 K with a relaxation time of 0.1 ps, coupled to the solute (peptide and membrane) and solvent separately. The system pressure was kept constant with a Parrinello-Rahman barostat[232], at 1 bar, with a relaxation time of 5 ps and a compressibility of 4.5×10^{-5} bar^{-1}.

5.4.3 Poisson-Boltzmann/Monte-Carlo Simulations

The Delphi V5.1 program[191] was used to perform Poisson-Boltzmann calculation. The atom radii were obtained from the Lennard-Jones parameters of the GROMOS 54A7 force field, using a 2 RT energy cutoff[233]. The atomic partial charges were used directly from the same force field. The peptide-membrane molecular surface was defined by the following parameters: a 1.4 Å radius probe, an ion-exclusion layer of 2.0 Å and a ionic strength of 0.1 M. The dielectric constants used were 2 and 80, for the solute and solvent, respectively. To calculate the electrostatic potential, a two-step focusing procedure was conducted with two 91 grid points. The coarse grid had a $\sim$1 Å spacing between the grid points, while the smaller grid had $\sim$0.25 Å. The defined relaxation parameters were 0.20 and 0.75, for linear and non-linear interaction, respectively. Periodic boundary conditions, for lipid bilayer systems, were applied in the x and y directions. Background interaction calculations were truncated at 25 Å and the electrostatic potential convergence threshold was 0.01 kT/e[157].

The PETIT program performed the MC calculations of the residues protonation states, using the free energy terms obtained from the PB calculations[193]. Proton tautomerism was accounted for all titrable groups. For each conformation, 10^5 MC cycles were performed and each cycle corresponds to a trial change of each individual site and pairs of sites with an interaction larger than 2 pK units.

5.4.4 Structural characterization of the arginine variants

A proper configurational and local description of these peptide/membrane systems requires both structural and electrostatic analytical tools. Therefore, all systems were evaluated regarding their electrostatic properties, such as average protonation, pK_a of insertion (pK_a^ins) [103, 104] and the complete pK_a profiles of each peptide key Asp13. The most common structural characterization consists of secondary structure analysis, membrane bilayer thickness, the Asp13 membrane insertion and its intramolecular distances to the neighbouring groups. This set of analyses clarifies both the configurational and local changes between each variant peptide conformational arrangement.

In this work, the membrane insertion of the Asp13 residue was used as a guideline to discriminate other properties' behaviours along the membrane normal. The membrane insertion of a given residue is defined by the relative difference between the average Z coordinates of the membrane surface reference and the residue of interest. The membrane surface reference is defined by a minimum of 10 atoms of the neighbouring lipid phosphate group, within a 6 Å radius in the xy plane, from the residue of interest. This data can be paired to the insertion values by their time stamps followed by a slicing procedure, using 0.5 Å insertion bins, where the pertinent data is assigned to the corresponding insertion level, hence obtaining a given property insertion profile.

We used membrane thickness calculations to quantify the local membrane deformations [103, 104]. The method performs half-thickness calculations, for each monolayer, within an annulus region. This region was 0.5 Å wide as it was defined within two radii centered on the peptide. Using this annulus, a scanning procedure is performed on the xy plane of the membrane monolayer, as both radii are simultaneously increased in a 0.5 Å step. With this approach, we describe both local deformations and membrane "bulk" (unaffected) regions (lipids usually at distances $>$15 Å). The presented membrane local deformations were calculated as the difference between the local half-thicknesses and the bulk region half-thickness (beyond the 15 Å cutoff). These calculations were done on all equilibrated conformation snapshots and, at the bulk regions, the thickness of both monolayers should converge to the same value, i.e. half the thickness value for a pure POPC membrane. The experimental POPC half thickness value range was calculated by interpolating from experimental thickness measurements in the fluid range, at different temperatures[214].

To characterize the key aspartate residue microenvironment at distinct membrane media, we need to assess and identify the neighbouring groups within the Asp13 first interaction shell. For each peptide variant, we calculated: the number of interacting lipid phosphate and choline groups; the ArgX-Asp13 side chain interactions; and the number of hydrogen bonds established with water molecules. All these and other system properties were calculated as time series and as a property insertion profile. The first interaction shell cutoff value (0.52 nm) was defined from the RDF distributions for water, phosphate, and choline groups, obtained from our previous work[104].

5.4.5 pK_a profile calculations and electrostatic contributions

The pK_a profiles are an important tool to assess and interpret the local electrostatics and how it affects the proton binding affinity of a pH-sensing residue. To that effect, the pK_a calculations must fulfil the following criteria: (1) each insertion bin must possess a minimum of 10 data points of each protonated state, at each pH value and for each replicate; (2) in the pK_a fit procedure, all conformational sampling data must originate from, at least, three replicates and each replicate requires data from, at least, two replicas to avoid sampling bias; (3) in a titration curve, the average protonation, at a given pH value, should not be higher (by 0.05) than the average of the previous lower pH, thus ensuring monotonicity. By fulfilling these conditions, the average protonations of each pH replica are calculated and then fitted to the Hill equation to derive the pK_a values.

A semi-quantitative analysis was also performed to ascertain how each electrostatic partner contributed to the pK_a of the key aspartate. This analysis required the following steps: 1) slicing procedure of each referenced neighbouring group data to the insertion profiles, for all pH values; 2) then, all pH-dependent data was used to perform a linear interpolation with the *interpolate* tool of the *scipy* module[251]. Then, we estimated each property value for the corresponding pK_a at all insertion levels; 3) the *Random Forest Regressor* algorithm of the *scikit-learn* module[252] was applied. A data array (90x5) was constructed based on the properties data of all peptide variants, including the *wt* data from a previous work[104]. The data consisted of all values present in each electrostatic partner profile (4 independent features) and their corresponding pK_a values (dependent value). The estimator generated several classifying decision trees predictions from several sub-samples of the data set, obtained from a bootstrap resampling method with replacement. Then, it calculated the average of all generated outputs to improve the prediction and it determined the relative importance ranking of each feature for the model. The hyper-parameters used were 2500 trees ($n_{estimator}$) with a max_depth of 20.

5.4.6 Analyses and error calculations

All analyses pertaining secondary structure, distance measurements, number of interactions between groups of interest and property time series were performed using the GROMACS tool package. Further analysis was performed using in house software (`http://mms.rd.ciencias.ulisboa.pt/#software`) and the specified python modules.

All pK_a error values were calculated using a Bayesian bootstrap approach. These estimations prevent fitting issues by executing 1000 bootstraps from our average protonation samples. In each bootstrap, random weights were assigned to each sample. This procedure also requires the same selection criteria (mentioned above) to obtain final pK_a and error values. For all other properties, the error bars reflect the property standard error of the mean.

5.5 Results and discussion

5.5.1 Structural characterization of the pHLIP variants

The therapeutic function of pHLIP can be modulated using small mutations to the peptide sequence, specially to the residues in the transmembrane (TM) region (10^{th} to 30^{th} residue)[101, 102, 123, 137, 141]. These residues dictate the thermodynamic stability of the inserted state, hence any mutation can disturb the electrostatic balance of key titrating residues (Asp-13), the lipophilicity of this region and, ultimately, the thermodynamic equilibrium of the state III peptide-membrane configurations. Cationic residues can modify the electrostatic nature of the TM region and the insertion/exit pathways[246], therefore, we designed four *wt* peptide variants, each with a distinct arginine position relative to the key Asp-13 (Table 5.1). Interestingly, all peptides slowly converged to a similar structure, not very different from the typical *wt* α-helical conformations displaying the characteristic kink near the water-membrane interface (Figures 5.1 and Figure C.1 of Supporting Information). The most important peptide and membrane properties equilibrated relatively fast, with convergence obtained after the initial 30 ns, which were discarded (Figures C.2–C.13 of Supporting Information).

The peptides structural characterization highlights unique effects of each arginine permutation on their structural stability (Figure 5.2A) and the local Asp-13 vicinity, in particular their specific interactions with the key Arg residues (Figure 5.2B). The peptide variants distinct folding patterns suggest that arginine mutations placed lower in the sequence (R14, R15 and R17) progressively induce larger hydrophobic mismatches, as they increasingly expose the C-terminus hydrophobic flanking regions to the water-membrane interface (Figure 5.2C-D), leading to more thermodynamically unstable states. More pronounced peptide tilting (Figure 5.2E) and helical unfolding (Figure 5.2A and Figure C.1 of Supporting Information) promote internalization of the hydrophobic stretch (Pro-18 to Leu-24) to mitigate these mismatch effects. This is further evidenced by the progressively deeper positions (negative values) of the central Leu-21 (Figure 5.2C), with the exception of the R17 variant where the significant peptide tilting ($\approx$30°) counteracts the TM region vertical movement (positive values). Overall, the energy penalty associated with the internalization of the N- and C-termini charged polar residues outweighs the partial helical unfolding and structural tilting, favoring these conformational rearrangements.

Interestingly, the R10 folding pattern contrasts the other peptides, as placing the positive guanidinium group higher in the sequence creates a TM hydrophobic mismatch in the opposite direction. The Arg-10 position inverts the observed helical loss of the hydrophobic TM stretches (18^{th}-21^{th} and 22^{th}-24^{th}) (Figure C.1 of Supporting Information) due to a closer proximity of the TM stretch to the polar environment of the outer water-membrane interface (Figure 5.2D). This proximity triggers helical loss of the hydrophobic stretch (18^{th}-21^{th}) to stabilize near the acyl chains ($\approx$ -1.4 at pH 5.75 in Figure 5.2C). The resulting pH-dependent helical loss near the termini regions is in agreement with the higher solvent exposure of the titrating polar residues, hence stabilizing the peptide structure by matching the flanking regions polarity (Figures 5.2A and C.1 of Supporting Information). Succinctly, these arginine residues functionally work as positive anchors that, depending on their sequence position, either propel(pull) the peptide to(from) the hydrophobic membrane core and inner water-membrane interface region. These TM region hydrophobic mismatches depend on the position permutation and strongly affect the stability of the peptide-membrane configuration, the peptide tilting and the degree of α-helix folding in the flanking regions (Figures 5.2A,E and C.1 of Supporting Information).

Major and minor (local) peptide movements are intertwined to define transmembrane pHLIP configurations and the local electrostatic vicinity of the key Asp-13. The structural disposition of the peptides imparts distinct Asp-13 membrane behaviours, populating either shallow membrane regions (R10) or below the ester region (R14, R15 and R17) (Figure 5.2D and Figure C.15 of Supporting Information). The internalization of a polar charged residue deeply perturbs the membrane bilayer, as water molecules and lipid headgroups typically form a stabilizing polar shell. Deeper Asp-13 residues should induce larger deformations, yet our results show that the deeper R14 and R17 variants induce smaller membrane perturbations, while the R15 and the more shallow R10 cause pronounced perturbations (Figures 5.2D and 5.3). The decoupling between major peptide structure and lipid bilayer deformations warrant a look at the local Asp-13 environment.

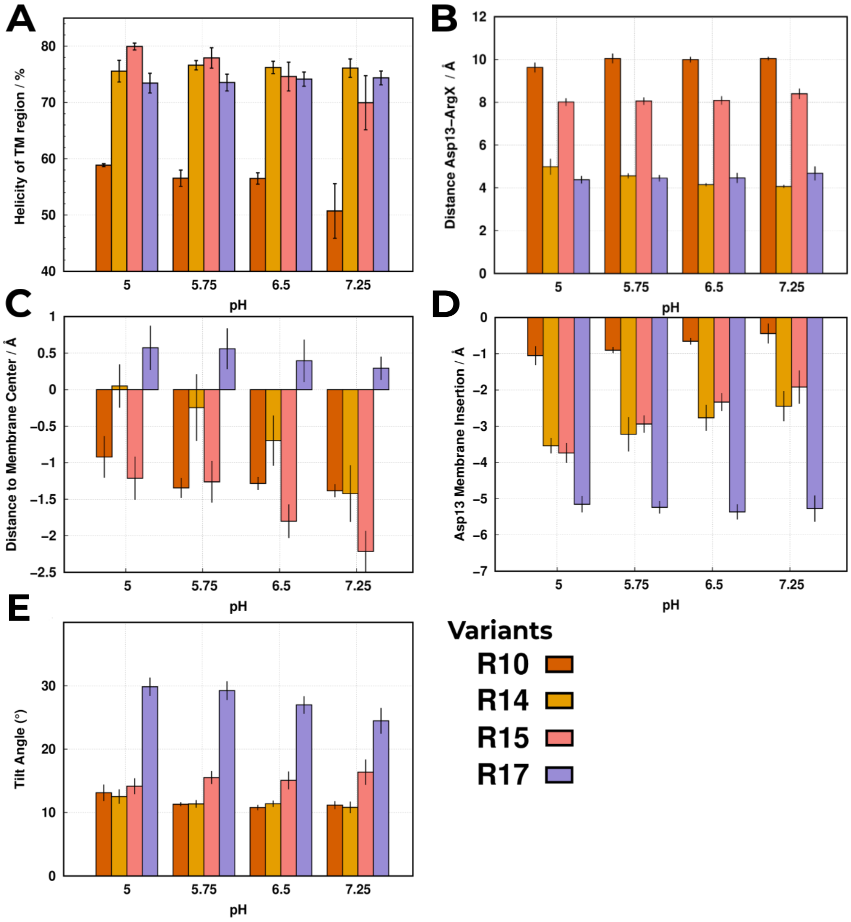


Figure 5.2: Structural analyses performed for all variant peptides using the equilibrated segments of the pHRE simulations at all pH values. Each analysis shows the following calculation: (A) the total average helicity (α-helix) percentage of the TM region (10^{th} to 30^{th}); (B) Average Asp13–ArgX distance at each pH value; (C) Average distance of the most central residue (Leu-21) to the membrane center, where positive/negative values are closer to the outer/inner monolayer; (D) the average membrane insertion of Asp13, where the negative values indicate the membrane depth; and (E) the average tilt angle of the TM segment relative to the membrane normal (see Figure C.14 of Supporting Information for a schematic representation). All shown error bars reflect the SEM.

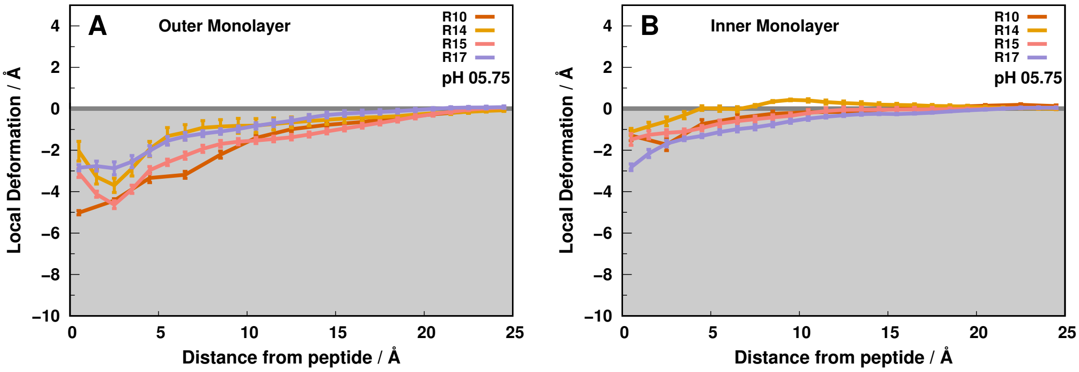


Figure 5.3: Outer (A) and inner (B) local monolayer deformation values in the *xy* plane along peptide distance. Each colored trend represents the membrane deformations induced by one of the peptide variants, at pH 5.75, until the 'bulk' lipids (15 to 25 Å). The error bars represent a standard error of the mean at every 1.0 Å.

Although all peptides exhibit unfolding events, the Asp-13 region remains remarkably conserved throughout the simulations (Figure C.1 of Supporting Information), therefore preserving the lack (R10,R15) or presence (R14,R17) of tight aspartate-arginine interactions (Figures 5.2B). The prevalent formation of salt-bridge interactions between the guanidinium and the carboxylate groups occur as the residues sequence positions (1 and 4 residues apart) place the residue side-chains side-by-side or in the top/bottom positions of the α-helix (Figure C.14 of Supporting Information). Then, these salt-bridges facilitate membrane embedding, despite their polar nature, as the charge neutralization decreases the need for a stabilizing solvation shell. Still, these peptides induce small deformations (∼-2 to -3 Å - R14 and R17), as transient increases in measured Asp13-ArgX distances hint at the arginine side-chain breaking the salt-bridge and snorkeling away to the water-membrane interface (Figures C.2-C.5 of Supporting Information). This snorkeling movement tilts the TM segment in ≈15○ and ≈30° for R14 and R17, respectively (Figure 5.2E), to minimize solvent exposure of the hydrophobic flanks, as previously noted, and inducing small membrane invaginations in the inner membrane monolayer (Figure 5.3B). The pronounced outer membrane monolayer perturbations induced by R10 and R15 (Figure 5.3A), result from the membrane internalization of a well-solvated aspartate. The residues sequence positions prevent a spatial arrangement of the α-helix that favours tight ArgX interactions, hindering the aspartate stabilization through a salt-bridge (Figure 5.2B). Consequently, choline headgroups and water molecules become the stronger interaction partners, promoting the deformation of the local lipid monolayer.

Overall, the destabilization of the water-membrane interfaces seems to mostly depend on the ability of the peptide to stabilize its Asp-13 negative charge. When the Arg side chain is available to interact with Asp-13 (R14 and R17), the peptides insert deeper in the membrane and minimize the water-induced deformations. Otherwise, structural constraints hamper the salt-bridge neutralization, inducing more pronounced deformations and less stable peptide-membrane configurations. Nevertheless, the arginine position is pivotal in stabilizing the peptide/membrane configuration as deeper positions experience more snorkeling events that pull the hydrophobic TM segment upward to the apolar membrane core. Altogether, the structural characterization of these peptides pinpoints a very important role of the Arg position in modulating the Asp-13 electrostatic environment.

5.5.2 The proton binding affinity and electrostatic shell of Asp13

These peptides thermodynamic stability strongly depends on the (de)protonation of Asp-13 to promote/hinder the insertion and exit processes. The proton binding affinity of an amino-acid residue, in a membrane bilayer environment, is defined by the strength of the surrounding electrostatic interactions and the level of access to the

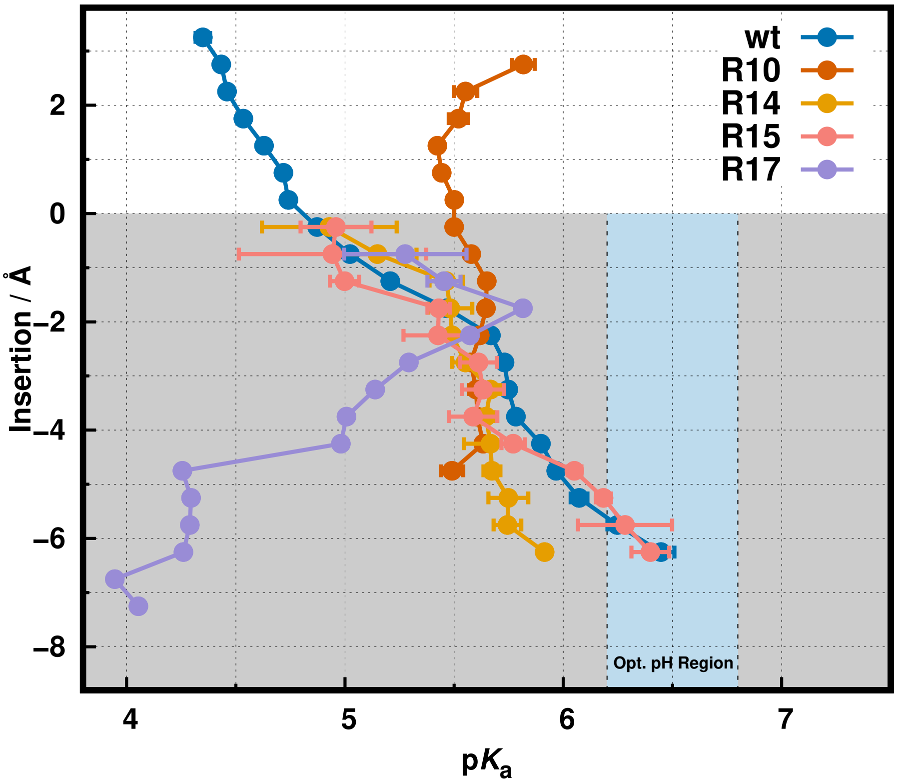


Figure 5.4: pK_a profiles of *wt*-pHLIP and its Arg variants. Each pK_a trend shows the shift along the membrane normal. The white and grey-shaded regions correspond to the water phase and membrane interior, respectively. The light blue vertical stripe identifies the pH region ideal for TME selection. The *wt*-pHLIP data was adapted from reference[104]. The R17 profile is merely indicative of the pK_a shift trend as the error values (not shown) were higher than one pK unit.

solvent[27, 103, 104]. Despite the complexity of depicting the different possible states of the diverse peptide-membrane configurational ecosystem, the insertion property of a residue is an indirect measurement of the peptide-membrane equilibrium, as each distinct insertion level represents a given medium (solvent, water-membrane interface, membrane core). Therefore, it is possible to accurately predict the pK_a behaviour, for a given residue, along the membrane normal (Figure 5.4).

In a first general assessment, we observe that all peptides exhibit distinct pK_a behaviours according to their own unique microenvironment. As expected, most of them exhibit a similar trend to the *wt*-peptide[103, 104], where the pK_a shifts towards higher values (Figure 5.4) induced by desolvation effects[27], with the exceptions of R10 and R17. Notably, in the deep membrane region (-5 to -6 Å), all variant peptides diverge to distinct pK_a values. This phenomena corroborates our structural assessment that each peptide interacts and adopts distinct peptide-membrane configurations, in similar media, according to their respective arginine position. As to detail the distinct microenvironments enveloping each Asp-13, we estimated the individual profiles for relevant electrostatic interactions (Figure 5.5), as discussed in the Methods section.

Although the pK_a profiles diverge in the deep membrane regions, both the structural and electrostatic analysis (Figure 5.2B and 5.5A) hint at two behavioral modalities regarding the presence or lack of tight arginine interactions. The lack of tight arginine interactions indicate a certain structural similarity of the R10 and R15 profiles with the *wt*. Indeed, our predicted R15 profile (pK_a^{ins} = 6.3±0.1) exhibits a remarkably similar behaviour to the model *wt* profile (pK_a^{ins} = 6.4±0.1) at deep membrane regions(-5 to -6 Å), despite its small deviation from

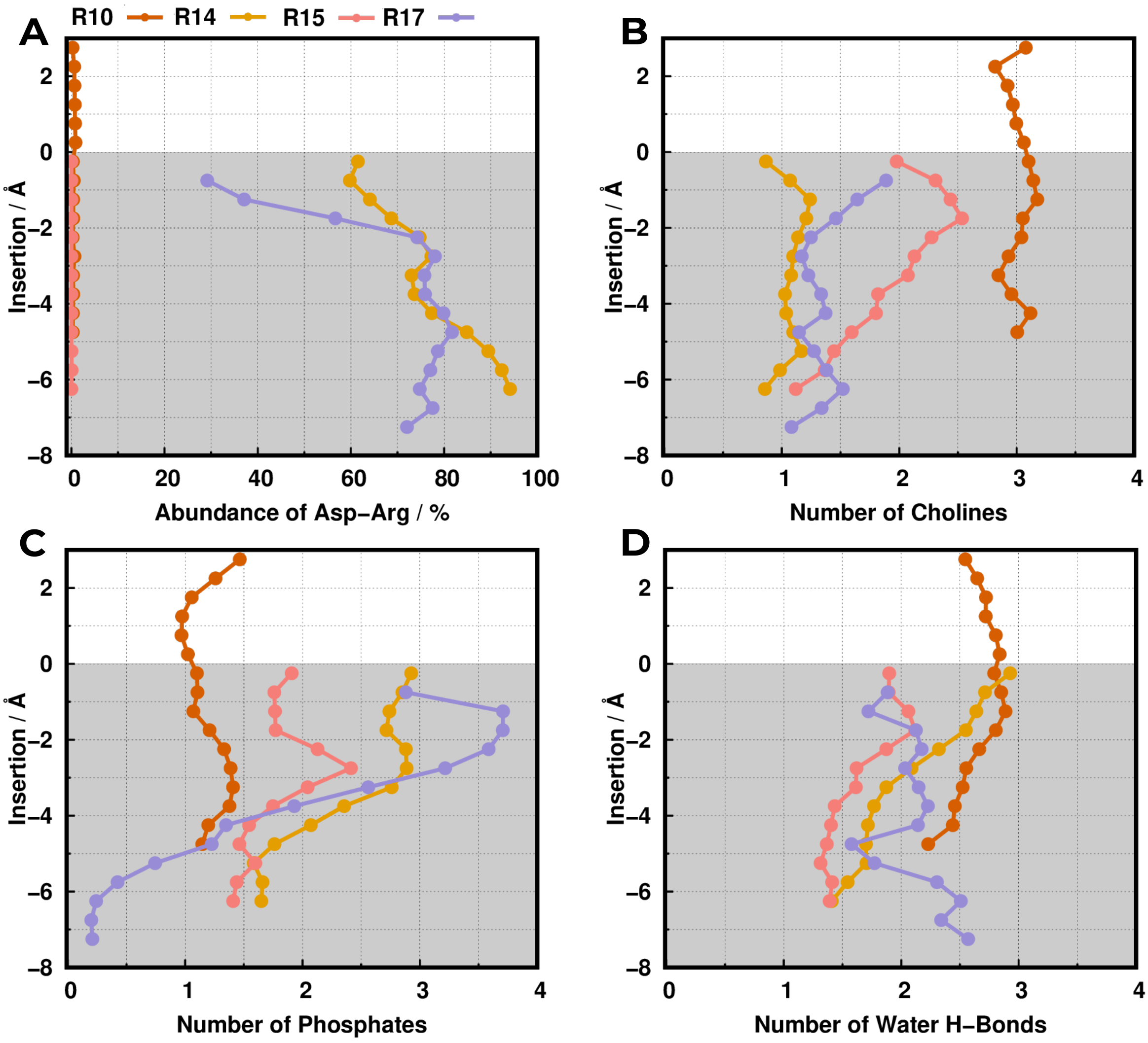


Figure 5.5: Property profiles relative to each neighbouring electrostatic partner. The abundance of the ArgX-Asp13 interaction (A), the number of contacts with the phosphate groups (B), cholines (C) and water H-bonds (D) were all calculated using a distance cutoff of 0.52 nm. The white and grey-shaded regions correspond to the water phase and membrane interior, respectively.

experimental insertion pK (pK^{ins} = 5.6 - 5.9). The omission of tight arginine interactions (Figure 5.2B and 5.5A) and a constant balance of choline and phosphate groups within the interaction shell (Figure 5.5B,C) further hints at analogous electrostatic environments with the *wt* peptide[104, 248]. The previously discussed structural characteristics (Figure 5.2A) attenuates the impact of distinct sequence positions, thus sampling equivalent peptide-membrane configurations in equilibrium.

On a different note, the R10 peptide possesses a unique insertion-insensitive profile, with a smaller pK_a shift (pK_a^{ins} = 5.5 $\pm$0.1), as a result of an homogeneous environment along the membrane normal. Interestingly, this behaviour stems from the notable structural arrangements that impart large membrane deformations (Figure 5.3) and more solvent accessibility (Figure 5.5D). Even if this effect could be decreased by the absence of the arginine residue (Figure 5.5A), the gain of influence by the interacting cholines (Figure 5.5B) results in a more positive solvation shell for Asp-13, that typically favours the ionized state, thus lower proton binding affinities. However, our predictions are in disagreement with the experimental data (pK^{ins} = 5.8 - 6.1), particularly regarding the peptide structural stability that should exhibit a $\approx$ 1.5x higher helical content than R10. In our models, it is still particularly challenging to fully reproduce experimental data, since limitations in the force-field or in our CpHMD-L method are hindering the full description of the complex configurational spaces of these systems.

Regarding the R14 and R17 peptides, the structural analysis highlighted tighter Asp13-ArgX interactions, hence we expect some divergence in the pK_a profiles relative to the previous variants. The R14 shared some structural similarities with R15 peptide, noted by the low Δ%helicity$_{R15-R14}$ (<5% – Figure 5.2A) and analogous membrane monolayer perturbations (Figure 5.3). These similarities permeated to their pK_a profiles, sampling similar peptide-membrane configurations, only diverging in the deeper membrane regions (-5 to -6 Å). The distinction in the proton binding affinities stems from a complete rearrangement of the interaction shell, rooted in the abundant presence of a salt-bridge interaction along the residue internalization (Figure 5.5A). The positively charged group shapes the lipid headgroups spatial disposition, attracting stabilizing water molecules (Figure 5.2D) and the more negatively charged phosphate groups to the shell boundary (Figure 5.5C), to the detriment of the positive choline groups(Figure 5.5B). These concerted electrostatic effects depict a more stable peptide configuration at environmental conditions that favour the ionized aspartate, hinted by the slightly smaller pK_a shift (pK_a^{ins} = 5.9 $\pm$0.1).

While the R14 peptide depicts a fair balance between diverse electrostatic effects, the R17 peptide is an evident outlier concerning peptide behaviour, with the pK_a^{ins} ($<$ 5.0) prominently shifting in the opposite direction (Figure 5.4). The trend suggests an overwhelming positive environment that overcomes the expected desolvation effect in the apolar membrane regions. Indeed, the interaction shell is characterized by progressively more frequent (Figure 5.5A) and tight (Figure 5.2B) arginine interactions, which energetically favor the stabilization of the anionic state of Asp-13. This phenomenon results from the thermodynamically stable peptide structure (Figure 5.2A) promoting side-chain interactions that preclude a large solvation shell and causing smaller membrane perturbations (Figure 5.3). Consequently, these arginine interactions far outweigh other electrostatic contributions, as noted by the pronounced decay of phosphate interactions and lack of competing choline and water interactions (Figure 5.5A,C), specially at deep membrane regions. In contrast, the R14 peptide exhibits a more homogeneous local environment along the membrane normal, even when strongly interacting with the arginine. The lower proton binding affinity ($<$ 5.0) stems from the neutralizing salt-bridge interactions suggesting that the peptide only remains in a thermodynamically stable state III at very acidic conditions. Furthermore, in larger time-scales, the Arg-17 snorkeling events may break the salt-bridge interaction with the ionized aspartate, creating further peptide instability that promote the exit process.

5.5.3 Which electrostatic interactions drive the Asp-13 pK_a shift

A residue pK_a derives from the delicate trade-off between the electrostatic contributions of several interacting partners within the solvation shell. Accordingly, different permutations of these effects, due to changes in the peptide microenvironment, results in distinct pK_a shifts. Nonetheless, the impact of these partner permutations is difficult to estimate, as certain neighbouring interactions may have more prominent effects in the residue proton binding affinity than others. Therefore, we used a Random Forest Regressor method (see more details in the Methods section) to quantify the contributions of each electrostatic feature in the overall Asp-13 pK_a values of these pHLIP variants (Table 5.2) .

Features	Phosphate	Arginine	Choline	H-bonds	R^2
Coefficient	0.57	0.07	0.17	0.19	0.85

Table 5.2: Electrostatic feature importance ranking obtained for our pHLIP variant peptide models. The calculations were done using a Random Forest Regressor module, as explained in the Methods section. The R^2 is the determination coefficient that evaluates the model predictive ability.

In our previous works[103, 104, 248], we determined that the phosphate groups along with the desolvation effect were the major factors for the anionic residues pK_a shifts. These observations are in agreement with the semi-quantitative feature estimation, as the model gives a larger weight to these features (0.57 and 0.19 for phosphate groups and water hydrogen bonds, respectively). Surprisingly, the arginine contribution (0.07) seems strikingly low for the model, even though our structural data highlights how the arginine sequence position heavily shapes the electrostatic microenvironment and overall peptide stability. Although unexpected, this model only assumes direct charge contributions, hence their indirect impact in modulating the restrictive Asp13 interaction shell is not taken into account. As a result, some features may be exacerbated, such as the phosphate groups. Note that our model tries to estimate the contribution of each property within the residue interaction shell (0.52 nm), whose volume can only be occupied by a limited number of charged particles. When a phosphate group is tightly interacting with the aspartate, it is simultaneously shielding the aspartate from the nearby cholines as exemplified in the R14 contributions profiles (Figure 5.5). The spatial composition of these groups are intricately correlated to each other, leading to the information of a property change being already encoded in the others, exacerbating the estimation. Nevertheless, this analysis was still key in weighing the importance of the group moieties that modulate the Asp-13 pK_a, being in a qualitative agreement with our previous discussions on the key role of the phosphate groups. Overall, these results show that a thorough structural and electrostatic analysis is pivotal to obtain a detailed picture of the molecular intricacies at play.

5.6 Conclusions

The peptide-membrane configuration and the interactions between key residues modulate the delicate balance between structural and electrostatic effects. The ionic interactions between the key Asp-13 and Arg-X residues define the favourable thermodynamic states, while these same configurations reorganize the local electrostatic environment sensed by the residue pair. This equilibrium is fundamental to the acidity-dependent therapeutic ability of the pHLIP peptide. In this work, we performed a multi-pronged structural characterization of pHLIP peptides with distinct arginine residue positions (R10, R14, R15 and R17) and their impact upon the proton binding affinity of the key Asp-13. The pH replica-exchange (pHRE) simulations revealed both unique structural and electrostatic features in each arginine permutation. The first major result was the inverse correlation between the aspartate solvent exposure and the interaction with cationic groups, particularly with the arginine side-chain.

We also observed that deeper arginine positions typically pull the aspartate away from the water-membrane interface through salt-bridge charge neutralization, although dependent on helical folding and the residues' side-chains topological proximity. In our model, the main features initially seemed to depend linearly with the arginine position, peptide folding and membrane stability. However, we showed that a more complex and intricate electrostatic interaction network does indeed modulate the proton binding affinity across different media. In a therapeutic potential analysis of these peptide variants, only the R15 peptide ($\mathrm{p}K_\mathrm{a}^\mathrm{ins}$ = 6.4±0.1) seems to behave similarly to the *wt* peptide as the peptide-membrane configurations are thermodynamically stable within the optimum pH region (6.2–6.8). All the remaining peptides $\mathrm{p}K_\mathrm{a}$ shifts hint at either a complete loss of peptide stability at acidic pH values (R17 – $\mathrm{p}K_\mathrm{a}^\mathrm{ins} <5.0$) or poorer targeting ability (R10 – $\mathrm{p}K_\mathrm{a}^\mathrm{ins}$ = 5.5±0.1 and R14 – $\mathrm{p}K_\mathrm{a}^\mathrm{ins}$ = 5.9±0.1) in this acidic range. Nevertheless, the other peptides showed unique behaviours, ranging from inverted $\mathrm{p}K_\mathrm{a}$ shifts (R17), insertion-insensitivity (R10) to minute changes in deep membrane regions (R14), that helped us interpret the role of cationic residues in transmembrane domains.

In sum, the position of the arginine group is fundamental in defining the first interaction shell of the titrating Asp-13. Most of the proton binding affinity contributions result from the phosphate groups configurational reorganization within the shell region, which is also complemented by the interactions with other electrostatic players. The arginine, when available topologically, seems to also play a more explicit role stabilizing the anionic state of the aspartate. Overall, cationic residues can be an important feature for peptide-membrane equilibria in transmembrane peptides and, while the aspartate is the key residue that determines the therapeutic performance of each pHLIP variant, the arginine position can play a decisive supporting role in fine-tuning these clinically relevant peptides.

5.7 Acknowledgements

We thank Diogo Vila Viçosa, Yana K. Reshetnyak, Oleg A. Andreev, and Don Engelman for fruitful discussions. Special thanks to Pedro B. P. S. Reis for helping with the development of the $\mathrm{p}K_\mathrm{a}$ profiling tool. We acknowledge financial support from Fundação para a Ciência e a Tecnologia through grants SFRH/BD/140886/2018 and CEECIND/02300/2017, and projects UIDB/04046/2020 and UIDP/04046/2020. This work also used the EGI infrastructure with the dedicated support of IN2P3-IRES, and NCG-INGRID-PT.

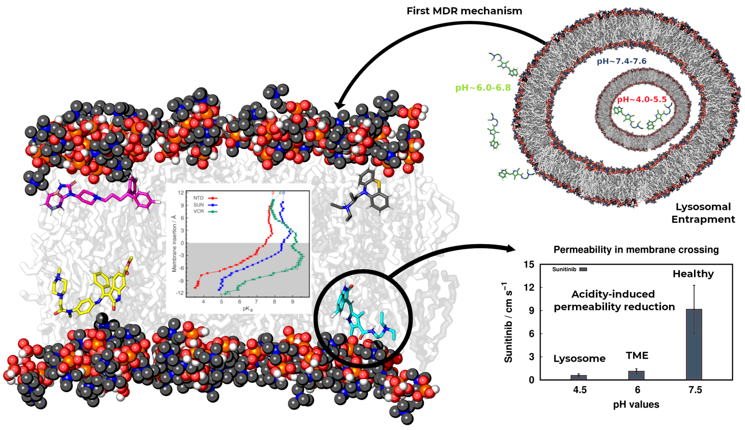
5.7.1 Supporting Information

Average helicity content per residue of peptide variants. Time series data of all peptide variants at all pH values of: arginine interactions with Asp-13; the Leu-21 distance to the membrane center; and the peptide tilt angle relative to the membrane normal. Graphical representation of all peptide variants in a POPC membrane bilayer, highlighting the peptide tilt angle feature. Probability density function of Asp-13 insertion for all variants at all pH values.

Chapter 6

Molecular insights into hydrophobic weak base antitumor drugs

Michal Stark , Tomás F. D. Silva[1], Guy Levin , Miguel Machuqueiro and Yehuda G. Assaraf



Contents

---

[1] In this work, I performed most compound parametrizations, all pHRE simulations and related analyses. I also helped in the system setup and analyses concerning the addendum work, while partially contributing to the results discussion and manuscript preparation.

6.1 Context

This last chapter encompasses two distinct bodies of work centered on Lewis base drugs. The first work focuses on a cooperative paper, published in Cells (https://doi.org/10.3390/cells9051082), that sheds light on the molecular and cellular details of lysosomal entrapment. Although the framework of this thesis is purely theoretical and computational, this work highlights the symbiotic nature of *in silico* and experimental methodologies in deciphering a pivotal defense mechanism of multidrug resistance, particularly the characterization of the transient deprotonation phenomenon across multiple drug compounds. Although Lewis base drugs compose a plethora of medical applications, our follow-up work focused mostly on the antitumoral compounds (Sunitinib and Nintedanib). Aside from the antitumor properties, these two drugs contrasted nicely with each other regarding their polarity, membrane avidity and protonation profiles, as shown in the previous data, thus presenting themselves as good models to characterize the membrane internalization process.

Such a remarkable challenge prompted us to apply our novel US-CpHMD methodology as presented in the MSc thesis work of Pedro Suzano (http://hdl.handle.net/10451/53778). To the best of our knowledge, this is the first known US-CpHMD application to describe a passive membrane crossing process across three unique pH-media (lysosome, TME and healthy cell) and estimate their pH-dependent permeabilities for both compounds. Our findings explain the acidity effect on both compounds permeabilization and how it heavily impairs the tumor internalization of Sunitinib. Furthermore, our results may explain known side-effects of chemotherapeutics concomitant with necessary higher dosage administration and the acquired drug resistance in these treatments. In the end of this chapter, it should be clear that US-CpHMD is a robust methodology to characterize pH-dependent permeabilities and membrane crossing processes, and that Sunitinib and other similar Lewis base drugs may have a critical design flaw that significantly undermines their chemotherapeutic potential. Additionally, a critical step of our future research should be centered on chemically modifying these titrable groups, possibly with acidic groups similarly to the studied TM peptides, to invert their protonation behaviour in these media and improve the design of pH-specific antitumor drugs.

6.2 Abstract

Lipophilic weak base therapeutic agents, termed lysosomotropic drugs (LDs), undergo marked sequestration and concentration within lysosomes, hence altering lysosomal functions. This lysosomal drug entrapment has been described as a luminal drug compartmentalization. Consistent with our recent finding that LDs inflict a pH-dependent membrane fluidization, we herein demonstrate that LDs undergo intercalation and concentration within lysosomal membranes. The latter was revealed experimentally and computationally using: a) confocal microscopy of fluorescent compounds and drugs within lysosomal membranes, and b) molecular dynamics modeling of the pH-dependent membrane insertion and accumulation of an assortment of LDs including anticancer drugs. Based on the multiple functions of the lysosome as a central nutrient sensory hub and a degradation center, we discuss the molecular mechanisms underlying the altered morphology and impaired lysosomal functions as a consequence of lysosomal intercalation of LDs. Our findings bear important implications for drug design, drug induced lysosomal damage, diseases and pertaining therapeutics.

6.3 Introduction

Lysosomes are terminal degradation centers for various proteins and organelles in eukaryotic cells; however, they have emerged in recent years as central cellular sensory hubs regulating differentiation, cell division and apoptosis in response to multiple cues [253, 254]. Lysosomes have been shown to promote tumorigenesis and cancer progression [255, 256] and have been implicated in various anticancer drug resistance mechanisms [39]. Impaired lysosomal function can lead to various non-malignant lysosomal storage diseases (LSDs) [257]. Although LSDs are hereditary, lysosomal dysfunction can also arise following drug treatment, which may alter the morphology, size and pH of lysosomes [258–260]. The majority of lysosomes have an acidic luminal pH between 4.7 and 4.9 [261, 262], hence entrapping a vast array of prognostic and therapeutic agents, all of which are hydrophobic weakly basic compounds [39] termed lysosomotropic drugs (LDs) [263]. Previous studies have found that LDs display a common denominator: beyond marked sequestration in lysosomes, LDs inflict lysosomal membrane permeabilization (LMP) [264–268]. In this respect, we as well as others have shown that multiple compounds that induce membrane permeabilization are typically *bona fide* membrane fluidizing agents which enhance the passive diffusion rates of lipid-soluble compounds and drugs [269–272]. Specifically, sunitinib (SUN) and siramesine (SIR), which we have shown to enhance membrane fluidity [50], are both known to induce LMP [273, 274].

Recent studies have shown the marked sequestration of anticancer LDs in lysosomes, including: Nintedanib (NTD), a tyrosine kinase inhibitor (TKI) used against non-small cell lung cancer [36]; SUN, an antiangiogenic TKI used as the first line treatment of renal cell carcinoma [34, 35]; Daunorubicin (DNR), a DNA intercalator used in the treatment of acute myeloid leukemia [275]; and the MEK inhibitors trametinib and refametinib, used in the treatment of pancreatic ductal adenocarcinoma [276]. However, although lysosomal compartmentalization of structurally and mechanistically distinct LDs is well documented, it is referred to as luminal drug sequestration [263]. Herein, we aimed at demonstrating that LDs accumulate in lysosomes via their intercalation into lysosomal membranes. Along this vein, original studies have demonstrated the concentration of ionizable hydrophobic drugs in cell membranes, liposomes and lipids [276–278]. The concentration of LDs in both the lysosomal limiting membrane (LLM) and intra-lysosomal vesicles (ILVs) was revealed by confocal laser microscopy images of fluorescent dyes and chemotherapeutic drugs as well as constant-pH molecular dynamic (CpHMD) modeling of the pH-dependent membrane insertion and accumulation of an assortment of therapeutic LDs. These findings have important implications for drug-induced lysosomal alterations, lysosomal diseases, novel drug design and pertaining targeted therapeutics.

6.4 Materials and Methods

6.4.1 Chemicals

The central nervous system acting drugs (CNSDs) ethopropazine (Ethop, profenamine), clomipramine (Clomp, anafranil) and pimozide (Pimo), as well as DNR, chloroquine (CHQ) and the DNA dye Hoechst 33342 were from Sigma Aldrich (St. Louis, MO, USA). NTD was from Enzo life sciences (New York, NY, USA). The viable fluorescent lysosomal probes LysoTracker Red DND-99 (LTR) and LysoTracker Green DND-26 (LTG) were from Life Technologies (Grand Isle, NY, USA). Vacuolin-1 was from Santa Cruz Biotechnology (Dallas, TX, USA).

6.4.2 Tissue culture

Human osteosarcoma U2OS cells (American Tissue Culture Collection, Manassas, VA) were maintained in RPMI-1640 medium supplemented with 10% fetal bovine serum (Gibco, Life Technologies), 2 mM glutamine, 100 units/ml penicillin G and 100 μg/ml streptomycin sulfate (Biological Industries, Beit-Haemek, Israel) in a humidified atmosphere of 5% CO2 at 37°C. U2OS cells stably expressing a GFP-tagged lysosomal associated membrane protein 1 (LAMP1-mGFP, a gift from Esteban Dell'Angelica, Addgene plasmid # 34831) [279] were grown in the presence of 650 μg/ml G418 (Calbiochem, EMD Chemicals, San Diego, CA, USA). The

Clomp-resistant subline Clomp^R was selected and maintained under 10μM Clomp.

6.4.3 Growth inhibition assay

U2OS cells were seeded in 96-well plates at 1000 cells/well in triplicates (in 0.1 mL growth medium/well). The following day, CNSDs were added at a 1/3 log increasing concentrations (0-100 μM), and cell survival was determined after 72 h, using a colorimetric cell proliferation kit according to the instructions of the manufacturer (XTT, Biological Industries). Percent cell survival was calculated relative to drug-free controls.

6.4.4 Live imaging

Confocal live imaging of lysosomal size and number

Day 1: U2OS and Clomp^R cells were seeded in Eppendorf's black Cell Imaging 24-well plates (Hamburg, Germany), with Clomp^R cells immediately supplemented with 10 μM Clomp. Day 2: CNSDs were added to U2OS cells at 10 μM (Clomp and Ethop) or 3 μM (Pimo), with 0.1% DMSO (the drug solvent) used as the drug-free control. Day 3: CHQ was added to U2OS cells at 20 μM. Day 5: Cells were washed with fresh medium and incubated for 45 min in the presence of 300 nM LTR and 1 μg/ml Hoechst 33342 in the dark. Cells were then washed with fresh growth medium and visualized using a confocal Zeiss LSM 710 inverted microscope (×63 magnification, Oberkochen, Germany) during incubation at 37°C in an atmosphere of 5% CO_2.

InCell Lysosomal Quantification Analysis

Cells were seeded in duplicates and treated as described above. Fluorescence was then visualized and recorded by an InCell Analyzer 2000 fluorescence microscope (version 3.7.3, GE Healthcare Bio-Sciences, Pittsburgh, PA, USA) during incubation at 37°C in an atmosphere of 5% CO2. Sixteen random fields were captured from each well (i.e. 32 fields per treatment). Several parameters, including the number of lysosomes per cell, their fluorescence intensity and lysosome area were determined using the InCell Investigator software. Results were obtained as mean numbers per field. Total cellular lysosomal volume was calculated as follows: lysosome radii (r) were mathematically derived from the mean lysosomal area (A), $A = \pi r^2$. Then, lysosomal volume (V) was calculated, $V = 4/3\pi r^3$. Finally, the volume values were multiplied by the number of lysosomes per cell. The experiment was performed three independent times. Figure D.3 depicts the results of a representative experiment, while Figure 1 g-h depicts the mean values of all three experiments ± SD.

Confocal visualization of fluorescent compounds in lysosomal membranes

For enlargement of lysosomes, we utilized either vacuolin-1 or Clomp. Parental U2OS and Clomp^R cells were seeded in Eppendorf's black Cell Imaging 24-well plates. The following day, prior to confocal imaging, the growth medium was replaced and cells were treated with 1 μM vacuolin-1 for 2 h and supplemented with either 300 nM LTR, 300 nM LTG, 10μM DNR or 10μM NTD for the last 45 min. Cells were washed with fresh growth medium and visualized using a confocal Zeiss LSM 710 inverted microscope (×63 magnification) during incubation at 37°C in an atmosphere of 5% CO_2.

LAMP1-mGFP, LTG and NTD were excited at 488 nm, LTR and DNR at 514 nm and Hoechst 33342 at 405 nm. Confocal microscopy images were processed using the ZEN (2.3 black edition) software. Fluorescence measurement along the diameter of individual lysosomes was performed using the ImageJ software (version 1.52i, Wayne Rasband National Institute of Health, USA).

6.4.5 CpHMD simulations

System setup and Simulation Settings

We adapted the lipid bilayer composed of 128 molecules of 1,2-dimyristoyl-*sn*-glycero-3--phosphocholine (DMPC) lipids from a previous study [27] and built a new system for each drug, i.e., Pimo, Ethop, Clomp, DNR, NTD, SUN and vincristine (VCR). The molecules were placed in bulk water at approximately 15 Å away from the membrane surface, to allow for unbiased membrane insertion in every replicate. LDs are apolar compounds that, when placed in a vicinity of a membrane, tend to interact avidly with lipids and insert into the lipid bilayer in an attempt to minimize the contact between their hydrophobic parts and water. For simplicity and due to membrane symmetry, we can distinguish three preferred positions for the LDs: (1) being away from the membrane, at the water phase, (2) residing at the water/membrane interface, stabilized by some electrostatic interactions with the lipid head groups, and (3) inserted deeply inside the membrane, hence minimizing the contact with water. In our simulations, the first position is rather undesirable (most molecules are significantly apolar) and, after some equilibration time, the compounds are already exchanging between positions 2 and 3. VCR and DNR are more polar and, hence, require significantly longer times for the simulations to equilibrate and allow for membrane insertion events. To obtain meaningful data for insertion statistics, we extended the normal initial equilibration time for the compounds (i.e, 100 ns), twice for VCR and three times for DNR.

Molecular dynamics (MD) simulations were carried out using the GROMACS 4.0.7 package [227]. The GROMOS 54A7 force field [172] was used together with the SPC water model [230]. The initial topologies for all drugs were built with the Automated Topology Builder (ATB) [173] and manually curated. These topologies were modified in the pairs section in order to exclude all 1-4 interactions in the aromatic moieties, analogously to the GROMOS rules for conjugated rings. The charge sets were derived from Merz-Singh-Kollman analysis using the electrostatic potential calculated in structures optimized with Gaussian 09 [174] using the B3LYP functional [175] and 6-31G* basis set. The compounds were prepared in both neutral and protonated (cationic) states.

Poisson-Boltzmann/Monte Carlo Calculations

The Poisson-Boltzmann calculations were performed using Delphi V5.1 [191], using the radii calculated from the GROMOS 54A7 Lennard-Jones parameters [233]. The molecular surface of the system (DMPC + drug) was defined using a probe with a radius of 1.4 Å, while the ion exclusion layer was of 2.0 Å and the ionic strength was set to 0.1 M. The dielectric constants used were 2 and 80 for the solute and solvent, respectively. Each PB calculation requires a two-step focusing procedure where, initially, we centered the titrating group in a cubic grid with 61 points with approximately 1 Å spacing (coarse grid), followed by a focusing step, reducing the spacing between grid points to 0.25 Å (focus grid). The relaxation factors used in the linear and nonlinear iteration processes for the coarse grid were 0.20 and 0.75, respectively. Periodic boundary conditions were set for the x and y directions. Background interactions were calculated up to a 25 Å cut-off, with a convergence threshold of 0.01 for the electrostatic potential.

Monte Carlo calculations were performed with the PETIT program (version 1.6) [193] using the free-energy terms obtained from the PB calculations and to sample the protonation states of each compound.

pH Replica Exchange Settings

The pH replica exchange (pHRE) method [166] is based on the stochastic titration constant-pH MD method (CpHMD) [151], which enables the treatment of pH effects as an external parameter in MD simulation, while allowing frequent exchanges attempts of pH values between two adjacent replicas (CpHMD simulations). Five replicates of 200 ns simulations were performed for all systems using pHRE. The simulations were extended to 300 and 500 ns for VCR and DNR, respectively, in order to improve their sampling. Each replicate is composed of four replicas, which were assigned an initial pH value from a molecule specific pH range: Clomp and Ethop (7.0 to 10.0), Pimo (6.3 to 9.3), SUN and NTD (5.5 to 8.5), VCR (5.7 to 8.7) and DNR (7.6 to 10.6), with a 1.0 pH unit step. The frequency of exchange attempts was every 20 ps.

Membrane Insertion Procedure

To perform pK_a calculations along the membrane normal for each LD, we separated our conformations according to the relative insertion of their titrable amino groups against the average Z position of the membrane phosphate group atoms that are within a 6 Å radius from the compound in a two-dimensional plane (x/y axis). If within this radius. a minimum of 10 phosphate group atoms (phosphorous and oxygen) is not obtained, then the absolute closest 10 atoms are used. This method accounts for possible local membrane deformations, since it uses the neighboring lipids to obtain these relative positions. After obtaining the insertion values for the titrating group in all conformations (please note that the insertion is not for the whole LD, but only for the titrating group), we separated each, and the respective protonation state, to an insertion bin. In each bin, we applied two criteria to assess sample quality before performing a fit to the Henderson-Hasselbalch (HH) equation. In the current study, each bin required a set of, at least, two different pH values from two different replicates to obtain their average protonation values. To obtain successful HH fits and good pK_a estimations, the data also need to display monotonicity (protonation is required to decrease with pH increase). These criteria allow for robust calculations of the pK_a values and of its standard error of the mean. All analyses were performed using in-house software and scripting (`http://mms.rd.ciencias.ulisboa.pt/#software`).

6.4.6 Statistical analyses

For quantification of changes in lysosomal parameters following CNSDs treatment, a 2-tailed 2-sample unequal variance student t-test was calculated per experiment, and a 2-tailed paired t-test was used for the average of the three experimental repeats. p values ≤ 0.05 were considered statistically significant and are mentioned in the figure legends.

The pK_a error values shown were obtained using a bootstrap approach [280]. However, to avoid fitting issues, a Bayesian bootstrap[281] was used where we ran 1000 bootstraps from our average protonation samples and, in each bootstrap, random weights for each sample were generated. For consistency, we also applied the two previous criteria to this procedure before calculating the final pK_a and error values. The protonation error values were calculated using a simple standard error of the mean.

6.5 Results

6.5.1 Live imaging

We have recently proposed that hydrophobic weakly basic LDs intercalate into the LLM via their lipophilic polyaromatic ring structure (illustrated in Fig. D.1), with their basic residues protruding into the acidic lysosomal lumen, where they undergo protonation [50]. Along this vein, using fluorescence recovery after photobleaching (FRAP), we recently demonstrated that markedly sequestered LDs inflict lysosomal membrane fluidization [50]. Towards deciphering the molecular mechanism underlying lysosomal membrane fluidization, we herein aimed at demonstrating the physical intercalation of weakly basic hydrophobic anticancer drugs and other therapeutic agents into the LLM. In order to enhance our ability to observe lysosomal membrane confinement of LDs, we first expanded the lysosomal compartment using pharmacological agents without triggering cell death; for this purpose, we chose therapeutic LDs that are widely used as a chronic treatment while displaying low toxicity. In this respect, most CNSDs are hydrophobic weak bases since they were strategically designed to cross the blood-brain-barrier [282] and induce a long-lasting effect [283]. The CNSDs that we used here were as follows: Pimo, an anti-psychotic drug of the diphenylbutylpiperidine class [284], which was shown to induce the formation of autophagosomes and autolysosomes [285, 286]. The tricyclic anti-depressant Clomp [287], which was identified as a lysosomotropic agent via a High Content Screening Assay for Identifying Lysosomotropic Compounds [288]; and the anti-Parkinsonian agent Ethop of the phenothiazine family [289], which, together with Pimo and Clomp, was identified as a functional inhibitor of acid sphingomyelinase, termed FIASMA [290, 291], hence suggested to interact with the lysosomal membrane.

A 72 h cytotoxicity assay was performed on human osteosarcoma U2OS cells to assess the desired non-cytotoxic drug concentration for each CNSD, following which the IC_{10} (drug concentration resulting in a 10% growth inhibition) values were determined for Pimo, Ethop and Clomp (i.e. 3μM, 10μM and 10μM, respectively). Using fluorescence microscopy imaging, we evaluated the lysosomotropic effect of the CNSDs on the number and size of lysosomes in U2OS cells treated for 72 h with the abovementioned concentrations of LDs. Cells were then viably stained with the established lysosomal marker LTR and analyzed using live confocal microscopy (Figure 6.1, a-f). CHQ was used as a *bona fide* positive control for its well-established lysosomotropic activity that induces a marked expansion of the lysosomal compartment (Figure 6.1b) [292, 293]. All three CNSDs induced a substantial increase in both lysosome size and lysosome number per cell, as reflected by the viably LTR-stained organelles (Figure 6.1, compare the red LTR fluorescence in b-f to a). This was also corroborated with U2OS cells overexpressing a LAMP1-mGFP, yielding the same LLM staining of the enlarged lysosomes (Figure D.2). One can notice the fluorescent ring-like structures that are captured by the laser scanning through the mid-section of lysosomes containing the green fluorescent LAMP1; the same fluorescent circular structures were observed with the viable LTR staining (Figure 6.1 a-f, insets, *white arrows*), revealing the confinement of this lysosomal probe to the LLM, rather than to the lumen. These findings are in accord with our recent hypothesis that LDs accumulate in the LLM, and could explain the LD-dependent membrane fluidization effect we recently uncovered [50]. Since these fluorescent circular organelle structures are difficult to visualize in the relatively small lysosomes under physiological conditions, we performed a quantitative analysis to evaluate the lysosomotropic effect of the CNSDs and determine which compound is the most effective in increasing lysosome size. We repeated the 72 h treatment with the CNSDs and undertook LTR staining and imaging, using an InCell fluorescence microscope and analyzer. The software can quantify the number of lysosomes per cell, their fluorescence intensity and size (i.e. area in μm^2). Based on these quantitative parameters, we calculated the total cellular lysosomal volume, as detailed under Materials and Methods (Figure 6.1, g-h and Figure D.3). While all three CNSDs induced remarkable effects resulting in a median increase of 10-20-fold in the total lysosomal volume per cell, Clomp exhibited the strongest and most stable effect on lysosomal size (i.e. 3.2 ± 0.013-fold over the drug-free DMSO control, with a median area of 3.03 μm^2, Figure D.3). We therefore established a Clomp-resistant subline named $Clomp^R$ by single step continuous selection with 10 μM Clomp. This subline displayed a stable expansion of the lysosomal compartment as long as cells were grown under drug selective conditions, although to a lesser extent than cells treated with Clomp for 72 h (Figure 6.1, f-h).

In order to demonstrate LLM localization of *bona fide* fluorescent dyes and naturally fluorescent chemotherapeutic drugs, we utilized, in addition to Clomp, vacuolin-1. The latter is a small molecule which was shown to expand lysosomes by inhibiting fusion between autophagosomes and lysosomes [294] and by inhibiting Ca^{2+}-dependent lysosomal exocytosis [295]. $Clomp^R$ cells were loaded with LTR and visualized by confocal microscopy using focus stacking to record lysosomal images at different focal planes (i.e. Z-stacks, Figure 6.2). As the scan proceeded through the lysosome-containing intracellular zone, the appearance of lysosomes shifted from a fluorescent ball-shape to a ring-like structure and back again (Figure 6.2, a-f). The same visuals were captured in U2OS cells treated with vacuolin-1 (Figure 6.2, g-n). However, vacuolin-1 appeared to produce larger lysosomes with thinner membranes (Figure 6.2, compare a-f to g-n). To ascertain that this finding is not a peculiar trait of the viable LTR dye, and to confirm the relevance of this finding to the clinic, we stained cells with expanded lysosomes with the fluorescent agent LTG (Figure 6.3, a-b) and the naturally fluorescent FDA-approved chemotherapeutic LDs DNR (Figure 6.3, c-e), a DNA intercalating agent [296], and NTD (Figure 6.3, f-h), an oral tyrosine kinase inhibitor [297, 298]. Both anticancer drugs displayed a marked incorporation into the LLM as evident from the fluorescent ring-like structures. Expectedly, fluorescence microscopy analysis of individual lysosomes showed that the periphery of lysosomes i.e., the membrane, was brightly fluorescent by DNR. By contrast, a marked decrease in fluorescence within the lumen of lysosomes was noticeable, down to background levels, albeit the luminal fluorescence was sometimes above the background levels, especially when lysosomes were small. This fluorescence distinction between the brightly fluorescent LLM and the less fluorescent lysosomal lumen was much more prominent, the bigger the lysosomes were (Figure 6.4). However, unlike the symmetric round rings observed following LTR, LTG and DNR staining, NTD-loaded lysosomes displayed highly distorted non-circular morphologies with sharp edges and an unequal fluorescence distribution in the LLM (Figure 6.3 compare a-e to f-h). An additional interesting observation is the expansion of ILVs by vacuolin-1 (Figure 6.3c) and the apparent accumulation of DNR therein, also notable with LTR in Figure 6.2g-n (yellow arrow), and further examples in Figure D.4.

6.5.2 CpHMD simulations of LDs insertion into a phospholipid bilayer

LDs are very hydrophobic weak bases, poorly soluble in water, with the ability to deprotonate and passively cross biological membranes. The pH values encountered in the different cellular compartments should have a major impact on the diffusion rates of these compounds across membranes. To address this point, we used CpHMD simulations coupled with a replica exchange scheme [166] of several LDs interacting with a membrane model at different pH values. The pH range used in each simulation was chosen to capture the pK_a shifts induced by the membrane environment according to the pK_a values of the LDs in water. The LDs were set up on the water phase to avoid any conformational bias towards the membrane. In our timescale, it was expected that these hydrophobic compounds would equilibrate at the water/membrane interface, which indeed occurred relatively fast with most LDs. However, convergence could only be obtained with DNR after extending the duration of the simulations to 500 ns, which allowed membrane insertion events. DNR is more polar than the other LDs used in this study, and the energy barrier for its membrane insertion is higher, creating a kinetic trap, due to our initial unbiased set up protocol. In our simulations, no exchange of LDs between monolayers were observed, which is not surprising, since this is probably the rate-limiting step for the diffusion-based membrane crossing, occurring at a slower timescale. Consequently, the LDs/membrane monolayer interactions model the drug accumulation in each type of membrane, dictated by the pH value at the water phase. We observed that deprotonation is a prerequisite for all compounds to allow for their membrane insertion, as evident from the decrease in the average charge of the LDs at deeper insertion depths (Figure 6.5a-f and Figure D.5a-b, left subplots) and the shifts towards lower values in the pK_a profiles along the membrane insertion (Figure 6.6). A simple way of tracking the preferred positions of the LDs in the membrane is to analyze their insertion distributions, which we used to calculate the protonation and pK_a profiles (Figure 6.5 a-f and Figure D.5 a-b, right subplots). It should be noted that these insertion values correspond only to the relative position of the titrable amino group and not the entire molecule, which, in this case, seems appropriate since this is the only group that can “sense” pH. The membrane insertion histograms obtained reveal that pH and, consequently, the protonation state, have a major effect on the preferred membrane depth locations of the LDs. In their protonated forms, the CNSDs preferably accumulate slightly below the phosphate region of the phospholipids and, upon deprotonation, they undergo deeper insertion into the lipid

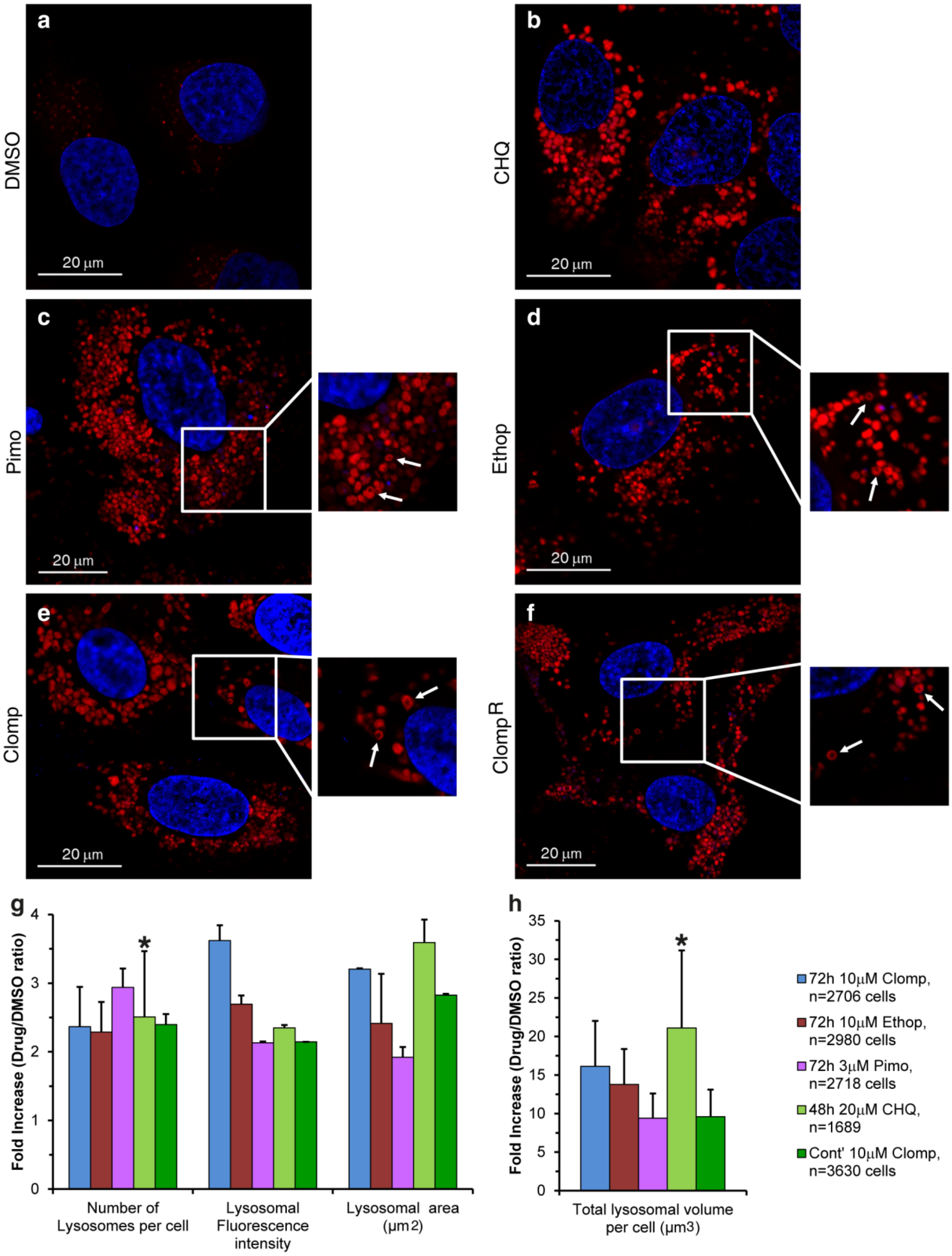


Figure 6.1: The effect of CNSDs on the number and size of lysosomes. U2OS cells were seeded in black glass bottom plates and treated for 72 h with either 10 μM Clomp, 10 μM Ethop, 30 μM Pimo or 0.1% DMSO as the drug free control. In addition, Clomp^R cells continuously grown in 10 μM Clomp were used, as well as U2OS cells treated for 48 h with 20 μM CHQ as a positive lysosomotropic control. Nuclei were stained with the viable DNA dye Hoechst 33342 (blue fluorescence) and lysosomes with the lysosomal probe LTR (red fluorescence). **(a-f)**: Representative images captured using a confocal Zeiss LSM 710 microscope (×63 magnification). Insets: **(a)** show the hardly visible drug free control lysosomes (upper inset) and over-enhanced lysosomes for better visibility (bottom inset), while **(b-f)** show examples of ring-like structures indicated by white arrows. All LTR images are representative of data collected from at least three independent experiments. **(g-h)** Cells were captured using an InCell Analyzer 2000 fluorescence microscope, and lysosomes were analyzed using the InCell Investigator software. Histograms depict the average median values obtained from three independent experiments ±S.D. All p values < 0.045 except for the ones indicated by an asterisk. See also Figure D.3.

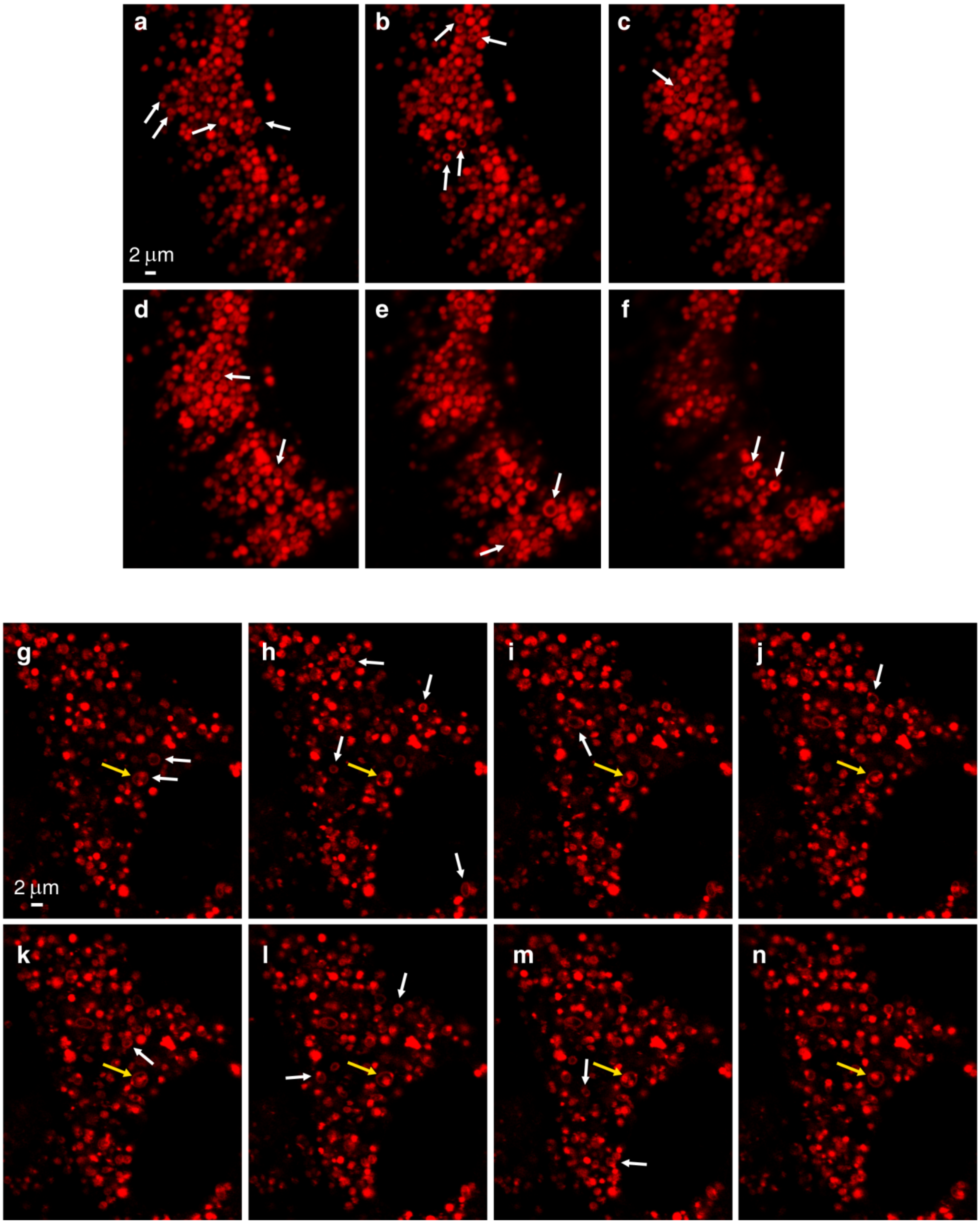


Figure 6.2: Z-Stack analysis of LTR-loaded lysosomes. Clomp^R cells (**a-f**) or vacuolin-1-treated U2OS cells (**g-n**) were labeled with LTR and scanned with a confocal Zeiss LSM 710 microscope (×63 magnification) using focus stacking with 0.2 μm slices. White arrows indicate ring-like structures as they first appear. The yellow arrow points to a lysosome harboring internal vesicles.

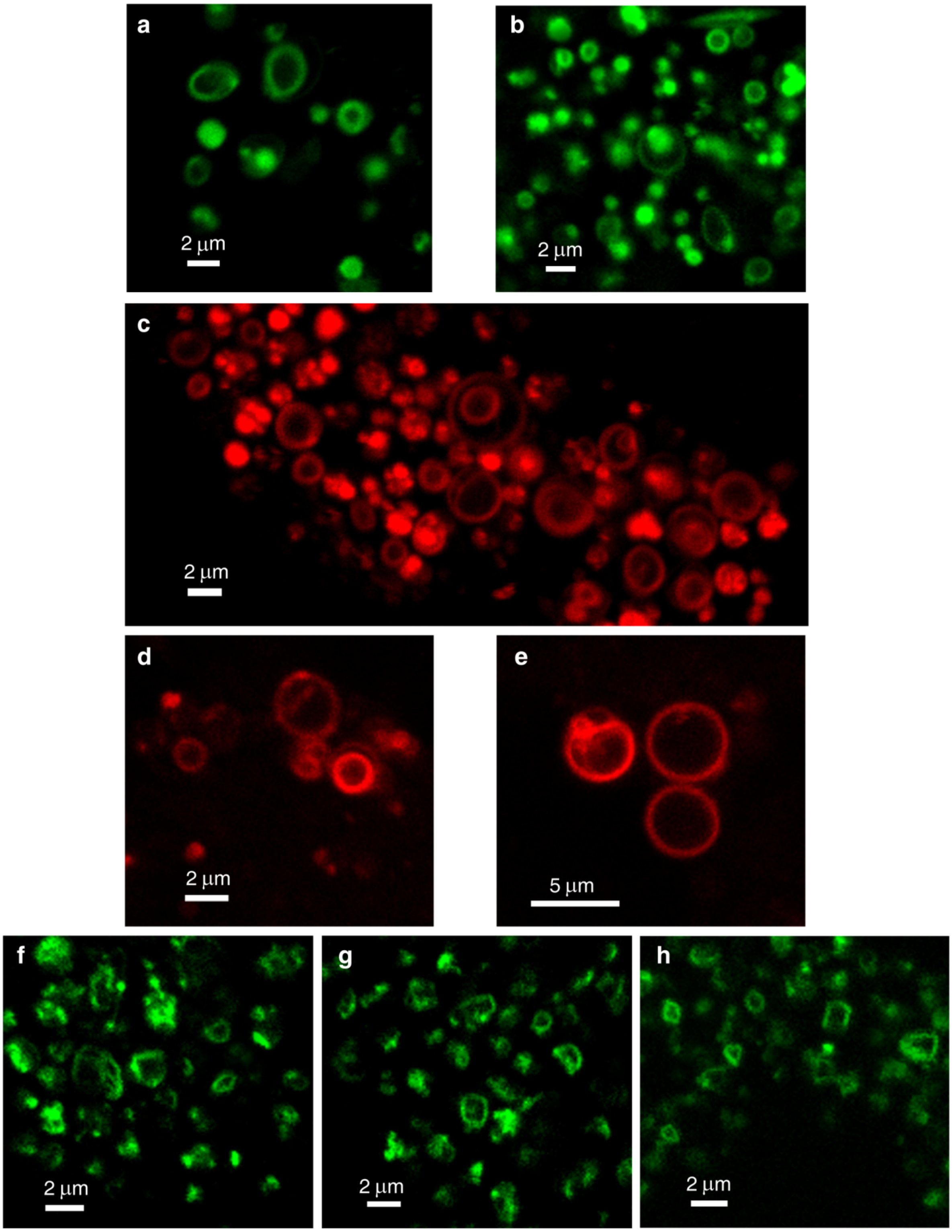


Figure 6.3: Lysosomal membrane staining with naturally fluorescent lysosomotropic compounds. Vacuolin-1-treated U2OS cells were loaded for 45 min with LTG (**a-b**), DNR (**c-e**) or NTD (**f-h**) and captured with a confocal Zeiss LSM 710 microscope (×63 magnification). All images are representative of data collected from at least three independent experiments. See also Figure D.4.

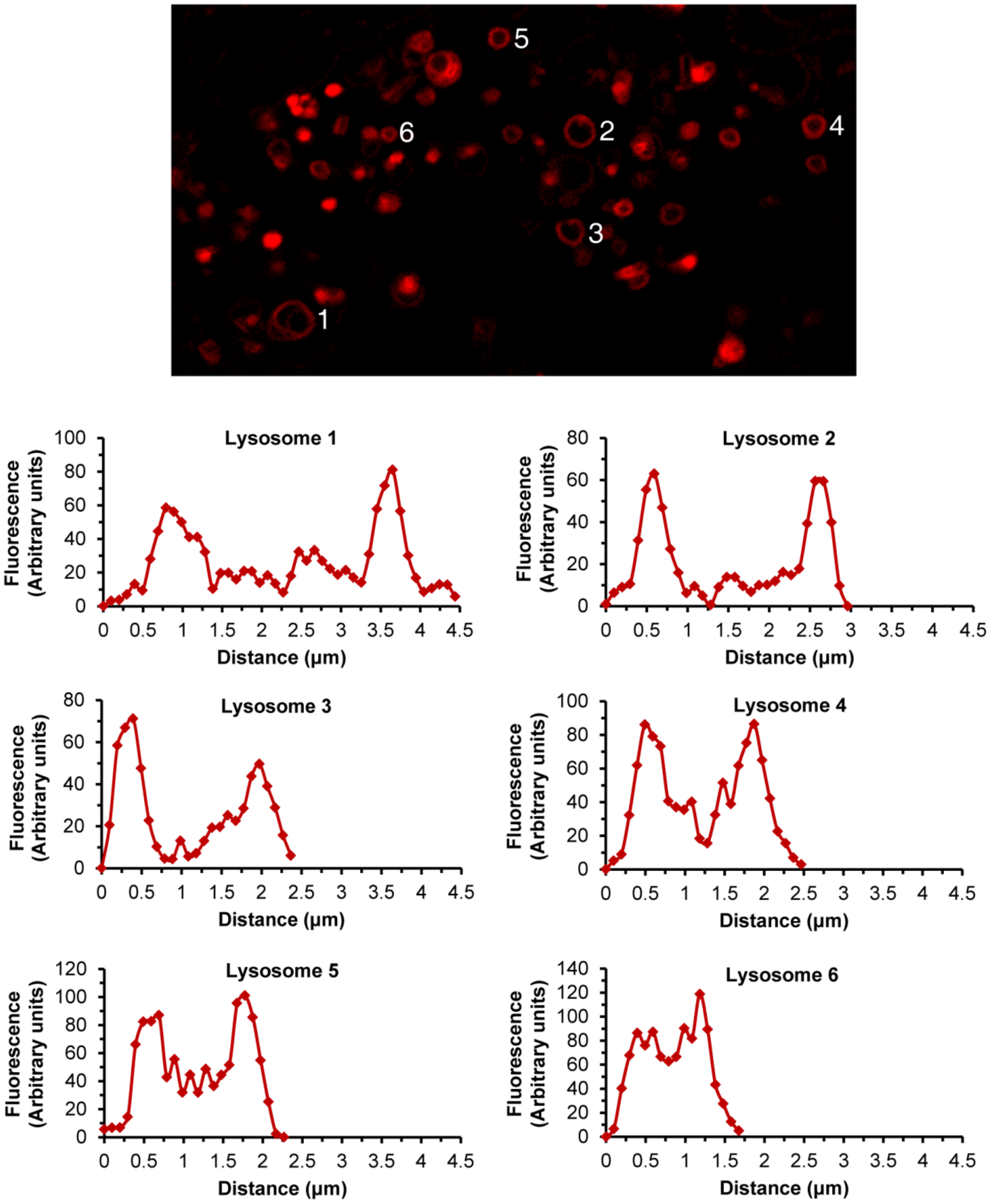


Figure 6.4: Lysosomal DNR fluorescence analysis. Vacuolin-1-treated U2OS cells were loaded for 45 min with DNR and captured with a confocal Zeiss LSM 710 microscope (×63 magnification). Individual lysosomes were analyzed for fluorescence intensity along their diameter using the ImageJ software. Each graph corresponds to a lysosome numbered in the photo.

bilayer,∼10 Å below the phosphate region (Figure 6.5a-c). Clomp and Ethop, at pH 7, and Pimo, at pH 6, became entrapped at the water/membrane interface (∼3 Å below the phosphate atoms). The antitumor drugs SUN, VCR and DNR displayed similar insertion distributions and appear to be fully entrapped at the interface at pH 6.5, 6.7 and 9.6, respectively (Figure D.5a-b and Figure 6.5e), indicating that below these pH values, insertion is highly unfavorable. NTD appears to deprotonate easily and inserts deeply into the membrane, and only at pH 5.5 (the lowest value we used for this LD) the protonated species is found to partially localize at the water/membrane interface (Figure 6.5d). We calculated the proton binding affinities along the membrane normal, which we termed pK_a values for simplicity, from the protonation data (Figure 6.6). These profiles illustrate how the pK_a values of these LDs markedly change upon interaction with the lipid bilayer, achieving shifts of 3-4 units below the water pK_a values. We selected several representative conformational snapshots of LDs interacting with the membrane to illustrate preferred membrane locations (Figure D.5c-d and Figure D.6). The selection process of representative LD conformational snapshots was based on the histogram peaks displayed in Figure 6.5 and Figure D.5a-b which, consequently, correspond to the different energy minima. For each LD, the left panel shows more inserted conformations typical of higher pH values, where the Lewis base is deprotonated, and the right panel shows the water/interface positions, more common at the lower pH values, where most molecules are protonated and unable to penetrate the membrane.

6.6 Discussion

It is well documented that LDs readily traverse biomembranes and accumulate within lysosomes, hence altering lysosome size, morphology, and function [46, 299, 300]. Specifically, LDs were reported to alter lysosomal enzyme levels and impaired their function - however, the exact intra-lysosomal localization of LDs was not reported. To this end, we here demonstrate that LDs intercalate into lysosomal membranes and highly concentrate therein. We thus discuss how this membrane confinement of LDs underlies their lysosomotropic mechanisms of action.

The pK_a values of the hydrophobic weakly basic LDs studied herein are near or well above the physiological ∼ pH 7.4 (Figure D.1, [298, 301–308], resulting in their partial, to almost total, protonation in the aqueous phase (i.e. bloodstream and cytosol). Protonation increases as they approach the membrane (i.e. lipid bilayer or lipid droplet) due to stabilization of their charged form, as reflected by the upsurge in their values and average charge (Figure 6.5) [27, 226]. Although, the compounds have high log p values (Figure D.1), [298, 305, 308–311] and poor water solubility at this pH, a problem that is overcome in the bloodstream by their binding to the highly abundant serum albumin [312], the charged forms are unable to passively cross biological membranes. According to the pH-partition hypothesis, an ionizable molecule can only diffuse across lipid bilayer membranes in its neutral form [313]; hence, deprotonation is required during membrane insertion. Indeed, a drastic decline in the pK_a profiles of the LDs was observed, accompanied by the complete neutralization of their charge. Deprotonation is supported by the desolvation process occurring during membrane insertion as the contact with the aqueous phase diminishes, a process that stabilizes their neutral forms [27]. This effect has been shown for titratable amino acids during their insertion into a lipid bilayer, using a recently developed computational methodology, where the authors observed either a pK_a increase or decrease for the anionic or cationic groups, respectively [27].

Abundance data from the insertion analyses reveal that our tested LDs were predominantly confined to the membrane, and this was recently shown by molecular dynamics (MD) simulations and the COSMOmic approach for various amphiphilic molecules [314]. The authors' calculations localized most compounds just below the phospholipid head groups, in agreement with our simulations herein. The LDs CHQ, imatinib and nilotinib were shown to accumulate in lysosomes at concentrations 1000-fold higher than their extracellular drug concentration within 2 h [314]. Such high concentrations of positively charged drugs at the water-interface of the bilayer can

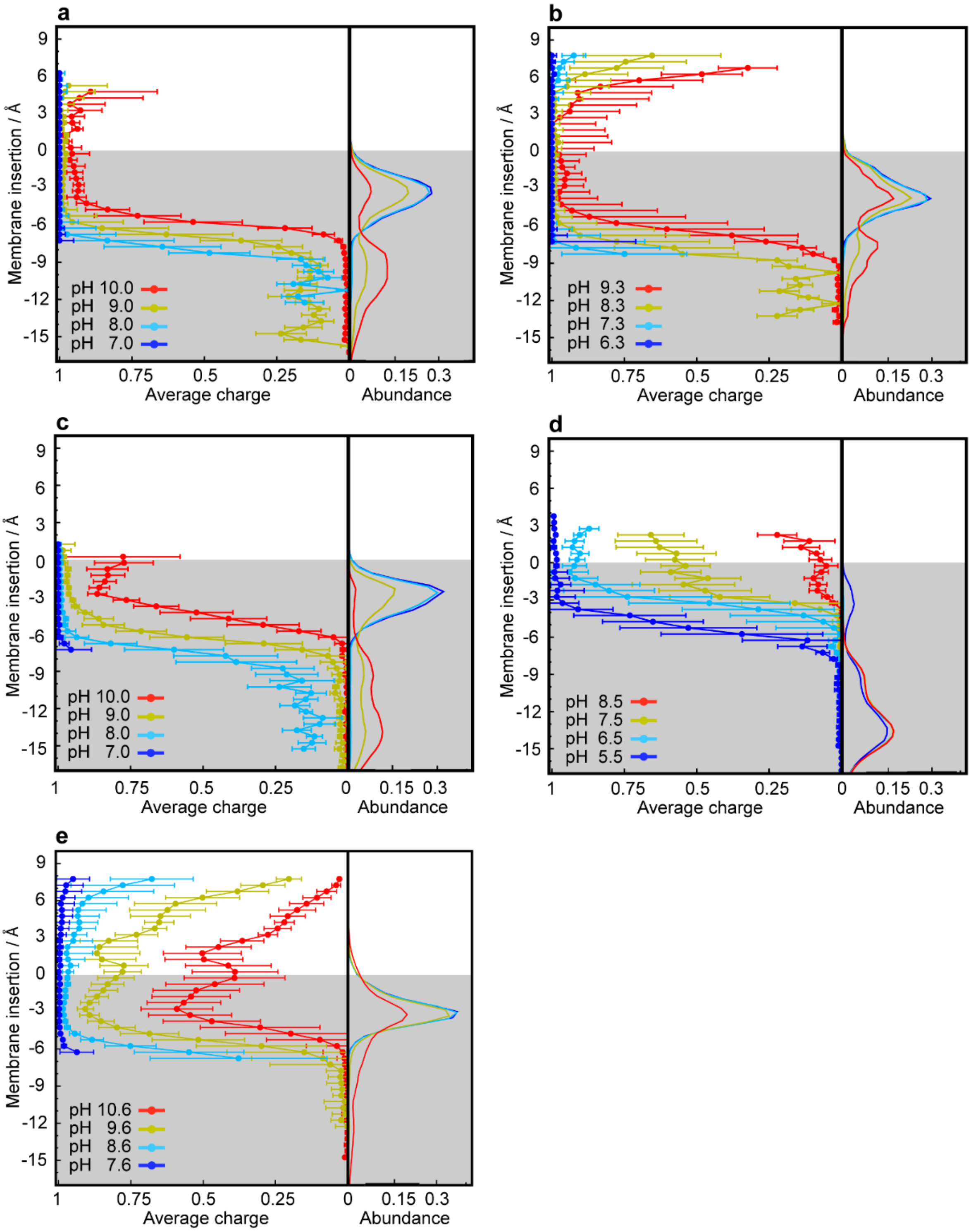


Figure 6.5: CpHMD simulations of membrane insertion of lysosomotropic compounds. Average protonation (left subplots) and abundance histograms (right subplots) are depicted along the membrane insertion axis for Ethop (**a**), Pimo (**b**), Clomp (**c**), NTD (**d**) and DNR (**e**). The grey-shaded area represents the membrane internal region below the phosphate groups, which are used as insertion references (see Materials and Methods).

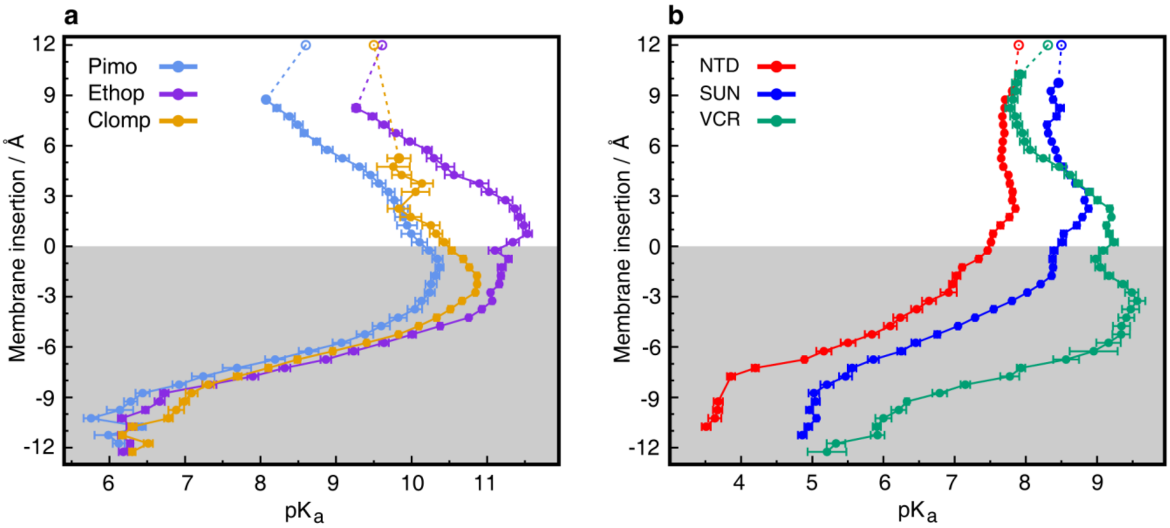


Figure 6.6: Predicted shifts in pK_a values during membrane insertion. pK_a profiles along the membrane normal for the CNSDs (**a**) and anticancer drugs (**b**). The unfilled points represent the experimental solution pK_a values (Figure D.1). The grey-shaded area represents the membrane internal region below the phosphate atoms, which are used as insertion references (see Materials and Methods).

alter the delicate electrostatics balance in the lipid headgroups, inducing repulsion between the choline groups of phospholipids and increase the distance between neighbor lipids (area per lipid), which results in a marked increase in membrane fluidity [315]. The latter is in complete concordance with reports of increased membrane fluidity induced by SUN and SIR [50] and DNR [316], as well as with the observed LMP induced by LDs [317]. In both cases, the fluidization effect was pH-dependent, a finding that correlates with the current results, where the abundance of LDs at the water/membrane interface was more prominent at acidic pH, where all compounds have the highest charge due to immediate protonation. The latter requires that LDs reside, at least transiently, within the lipid bilayer, with their basic residues protruding into the acidic lysosomal lumen - hence, being in proximity to the aqueous phase is a prerequisite for protonation. Upon protonation, the weakly basic compounds are entrapped since the energetic barrier for their diffusion across the bilayer center is too high. For example, the LD propranolol, at its protonated form, was calculated to face an energetic barrier of $\sim$12 kcal·mol^{-1} when encountering the hydrocarbon core of the bilayer [226]. Indeed, neutron diffraction experiments positioned propranolol within a DMPC lipid bilayer, with its naphthalene moiety partitioned into the hydrophobic core and its charged amine side chain in the phospholipid head group region [317]. The protonated charged form is further stabilized by forming a surface ion pair with the zwitterionic lipid headgroups [279, 318], thus anchoring the molecule in place. Furthermore, pH-dependent ionic interactions between the weak base local anesthetic tetracaine and the head groups of phosphatidylcholine were demonstrated by conformational changes in the lipids' head groups. This was inferred from changes in the NMR spectra of ^{2}H in the trimethylammonium group and ^{31}P in the phosphate group of phosphatidylcholine [319]. Indeed, CpHMD simulations at the lowest pH reveal that, except for NTD, our tested LDs do not penetrate the bilayer deep enough to cross over to the other side. In concordance with these experimental and computational results, we here show by confocal microscopy imaging the intra-membrane fluorescence of the lysosomotropic compounds LTR, LTG, DNR and NTD.

Fluorescent ring-like structures captured by confocal microscopy, following lysosomal expansion by Clomp or vacuolin-1, were easily detectable for all naturally fluorescent compounds. It should be noted that we did not use the naturally green fluorescent SUN for this fluorescence microscopy since it induces lysosomal photodestruction via rapid formation of reactive oxygen species and lysosomal membrane rupture [35]. Beyond the need to use a laser scanning microscope, in order to capture the luminal region of the lysosomes, the limiting factors in visualizing fluorescent lysosomal "rings" are lysosomal diameter, lysosomal membrane thickness and fluorescence intensity. The three parameters influence the actual ability to detect the reduction in fluorescence in the lumen area and distinguish between the lumen and the membrane region. In this respect, vacuolin-1 produced larger lysosomes with thinner membranes and lower fluorescence intensities than Clomp; hence it is an optimal agent for visualization of fluorescent lysosomal rings. The confined localization detected with the fluorescent molecules could only be achieved if they were to partition predominantly into the membrane compartment. As drug molecules accumulate at the water/membrane interface they can detach completely from the bilayer and move into the lysosomal lumen. This can be interpreted from Figure 6.3c and Figure D.4, presenting intra-lysosomal fluorescent membranes (i.e. a ring within a ring) following vacuolin-1 and LD staining. These structures are composed of portions of the plasma membrane, destined for degradation, that reach the lysosome in the form of small ILVs [320, 321]. For DNR to reach and accumulate within these vesicles it must first move through the luminal aqueous phase, hence being fully detached from the LLM. Apart from LLM, ILVs present an additional accumulation domain for LDs that can contribute to the sequestration of- and consequently resistance to LDs.

As mentioned above, although LDs cannot evade the lysosomal acidic trap by simple diffusion, they can be actively translocated from lysosomes into the cytoplasm by the Niemann-Pick C1 (NPC1) protein [322, 323], a LLM transporter exporting cholesterol into intracellular destinations. The widely used lysosomal probes LTR and neutral red [324] were shown to undergo displacement from fibroblast lysosomes almost completely within 12 h, whereas there was little change in their lysosomal concentration in NPC1 $^{-/-}$ cells within the tested 24 h period [322]. Except for the viable staining of lysosomes for their imaging, LTR has been used in several published assays as a tool to quantify the degree of lysosomotropism of a given compound [288, 317, 325, 326]. These assays are based on quantification of the LTR displacement from lysosomes by exposure of the cells to increasing concentrations of the tested compounds. However, the exact mechanism of LTR displacement was not revealed. Benoit Lemieux *et al.,* and Johannes Kornhuber *et al.,* did explore the possible contribution of lysosomal alkalization by LDs to the lysosomal displacement of LTR but concluded that this is a pH-independent process [317, 326]. The competitive nature of LTR displacement by other LDs indicates the involvement of a binding step in addition to the membrane insertion, as was previously suggested [279, 318, 319]. LTR reaches a steady-state equilibrium of ionic binding-dissociation to the phospholipid head group, which is disrupted at high competitor concentrations (i.e. >100 molar excess over LTR). With lower rebinding possibilities, it is readily available for clearance by NPC1. Taken together, these results suggest that hydrophobic weakly basic compounds can induce LTR displacement from lysosomes by enhancing NPC1 efflux activity. Indeed, several lysosomotropic amines were reported to stimulate the function of NPC1 in the clearance of lysosomal contents [322]. Enhanced membrane fluidization induced by highly membrane-accumulated LDs [50] might contribute to the increased activity of NPC1, as was shown for the G protein-coupled receptor serotonin 2B [327] and for epidermal growth factor receptor [328]. Another example of an LD as a substrate of NPC1 is the chemotherapeutic drug leelamine, a lipophilic diterpene amine phytochemical with a pK_a of 9.9, that was shown by molecular docking analysis to bind NPC1 at the cholesterol-binding pocket [329] and inhibit its cholesterol export activity from lysosomes [330].

While LTR, LTG and DNR appear to exhibit an even lateral distribution within the round LLM, NTD induced distorted lysosomal membrane structures presenting uneven fluorescence distribution. Although the exposure time to NTD was quite short (i.e. 45 min), NTD accumulated in the lysosomal membrane at a high concentration to induce an impairment of the overall structure of the bilayer. It is possible that NTD accumulated in the LLM at concentrations that exceeded its solubility and hence precipitated. This hypothesis was proposed to account for the observation that certain hydrophobic amines accumulated in lysosomes at concentrations that were significantly higher than computationally predicted [331, 332]. If, indeed, NTD precipitated in the lysosomal membrane, it could account for the uneven NTD fluorescence captured by confocal microscopy imaging. MD simulations suggest that asymmetric drug incorporation into a lipid bilayer (i.e. confinement to only one leaflet), can induce local changes in membrane curvature [333]. All the currently tested compounds are suggested to accumulate in the inner leaflet of the LLM; however, with the exception of NTD, they do not induce changes in lysosomal membrane morphology. Insertion analysis revealed that NTD, in its neutral form, was deeply inserted in the lipid tail region (around 12-15 Å from the membrane surface). Moreover, at the lowest pH studied (5.5) and contrary to all other LDs, a significant amount of NTD was observed in its neutral state. The presence of both protonation forms of NTD at this low pH indicates that this LD can simultaneously inflict opposing effects at different membrane regions. On the one hand, NTD behaves like cholesterol that mostly accumulates at the tail regions of biological membranes, where it induces ordered domains that locally rigidify the bilayer structure [334, 335], whereas at the water/membrane interface it presumably induces membrane fluidization like SUN, SIR [50] and other LDs [336, 337]. This dual effect is detrimental to the bilayer structure and to integral and peripheral proteins therein.

In addition to the membrane localization of LDs predicted by various simulations, the accumulation of an assortment of amphiphilic weakly basic compounds in the inner leaflet of the LLM was suggested based upon their pharmacological effects. Specifically, inhibition of intra-lysosomal enzymes including acid sphingomyelinase (ASM) [290, 291, 338], acid ceramidase (AC) [339], phospholipases A (PLA) and C (PLC) [340, 341]. These lipases are transiently bound to the inner leaflet of the LLM via ionic interactions. The latter induce conformational changes in these enzymes and the consequent exposure of a hydrophobic surface, which allows the binding of the lipid substrate [342–344]. The ionic interactions were shown to depend on the presence of negatively charged phospholipids and acidic pH, which were also required for the inhibitory effect of LDs on the above enzymatic activities [291, 341, 343, 345]. The current view is based on the competitive binding of protonated LDs to the negatively charged phospholipids, which leads to neutralization of the charge of the inner LLM surface and consequent detachment of the above lysosomal lipases. Consequently, these lipases shift to their closed inactive forms, leading to a marked decrease in their enzymatic activities. Another contributing factor for the loss of these enzyme's activities is the decrease in the protein levels of ASM, AC and PLA, presumably due to their subjection to intra-lysosomal proteolysis [339, 344, 346]. Impaired lipid catabolism leads to lysosomal lipid accumulation and consequent lipidosis, a hallmark of LSD [260]. A 2-day treatment with gentamicin, an antibiotic LD which causes renal impairment in ∼30% of treated patients [347], induced renal cortical lysosomal phospholipidosis in adult rats, which was associated with decreased PLC activity [348]. This effect of LDs on membrane-bound enzymes is opposite to their impact on the lysosomal transmembrane protein NPC1 [349]. NPC1 is protected from the digestive environment of the lysosomal lumen by membrane phospholipids, whereas, ASM, AC, PLA and membrane phospholipids are protected from proteolytic digestion by the lysosomal glycocalyx [320, 350]. The glycocalyx is a thick coat of glycoproteins proposed to have an average thickness of ∼8 nm, as determined by electron microscopy [351]. The carbohydrate luminal domains of lysosomal integral membrane proteins together with various LLM-associated proteins form a dense, highly *N*-glycosylated barrier, which is negatively charged and lines the inner surface of the LLM [350, 352, 353]. The apparent charge neutralization of the lysosomal luminal membrane leaflet surface by LDs disrupts major protein-lipid interactions and induces the collapse of the glycocalyx, following which, ASM, AC and PLA become subjected to proteolysis by luminal enzymes.

Several amendments should be implemented in the above sequence of events. Negatively charged lipids, specifically phosphatidylserine (PS), phosphatidylinositol (PI) and Bis (monoacylglycero)phosphate (BMP,lyso-bis-phosphatidic acid)- comprise only a small fraction of the total lysosomal lipidome (i.e. <10 %) [354, 355]. BMP is a unique negatively charged phospholipid found only in late endosomes and lysosomes [356]. It was found to be enriched 32-fold in the lysosome fraction over the homogenate of rat liver [357]; however, immuno-electron microscopy revealed that 99% of lysosomal BMP is localized to the ILV [358]. PS is mostly found in the inner leaflet of the plasma membrane [355], where it does not disturb the stability of the negatively charged glycocalyx on the outer cell surface [359]. It is reasonable to assume that PS does not reside in the inner leaflet of the LLM since it could disrupt the heavily negatively charged lysosomal glycocalyx. It is more likely that PS is part of ILV membranes originating from the plasma membrane. Moreover, the incorporation of the negatively charged phospholipid phosphatidic acid into isolated lysosomes osmotically destabilized them [360], implying that the LLM should harbor only minute amounts or no negative lipids at all. The requirement for negatively charged phospholipids for the membrane binding and activity of ASM, AC and PLA [291, 343, 344], together with the ILV localization of such lipids implies that lipid catabolism occurs mainly or entirely on the surface of ILVs [320, 321, 361]. Since our confocal images show that LDs also accumulate in the membranes of ILV, it is easy to understand how LDs inhibit the activity of lysosomal lipases. The negative charge on the surface of ILVs is neutralized by the high concentration of positively charged LDs incorporated into the outer leaflet of the ILV bilayer, hence presumably disrupting the binding of lipases. Additionally, LDs could ionically bind negative lipids, hence masking lipid substrates from their specific lipases.

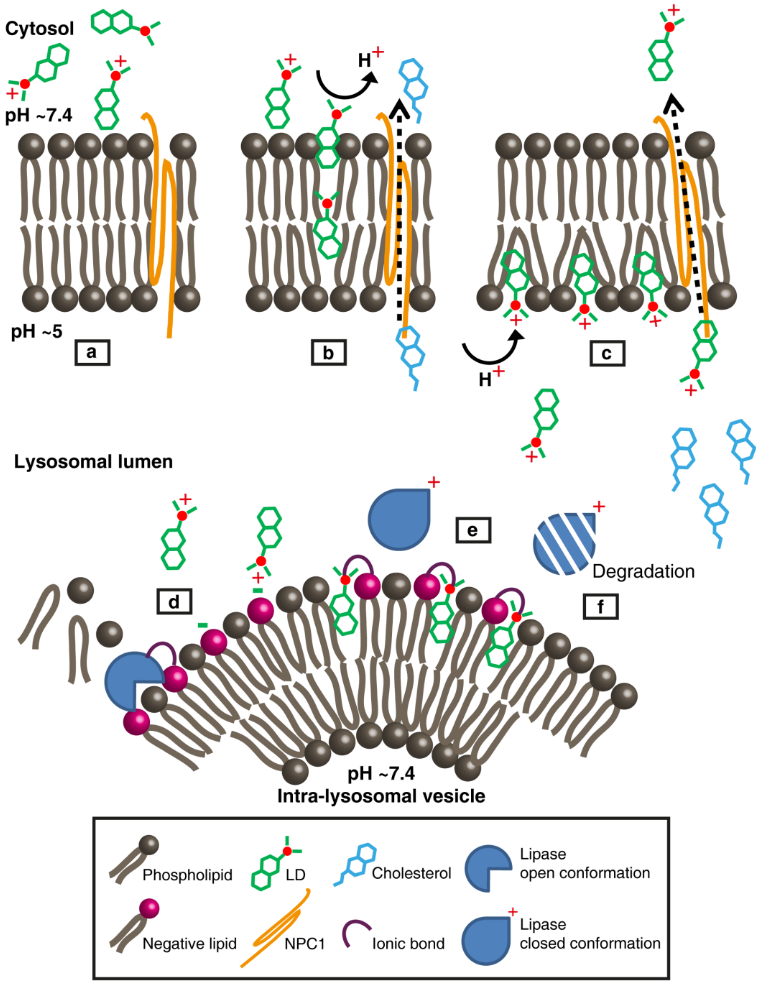


Figure 6.7: Model representation of lysosomotropic drug accumulation within the lysosomal membranes and its lysosomotropic effects. **(a)** An LD reaches the lysosomal outer leaflet surface in its protonated form. The LLM is illustrated with neutral phospholipids (e.g. phosphatidylcholine and sphingomyelin) and the transmembrane protein NPC1. For simplicity reasons, the lysosomal glycocalyx is not shown. **(b)** The hydrophobic nature of the LD promotes its insertion into the bilayer, during which it undergoes complete deprotonation. Under physiological conditions, NPC1 exports cholesterol to the cytosol. **(c)** LD molecules diffuse through the LLM until their amine residues encounter the acidic lysosomal lumen, where they undergo protonation. As more LD molecules accumulate, they induce enhanced bilayer fluidity, thus enhancing NPC1 activity and competing with cholesterol for binding to NPC1. While most LD molecules concentrate within the inner leaflet of the LLM above the phospholipid head groups, some detach and move through the aqueous lumen. **(d)** The LD reaches the intra-lysosomal vesicle (ILV) surface, where various lipases (e.g. acid sphingomyelinase (ASM), acid ceramidase (AC), and phospholipase A (PLA)) are electrostatically bound to the membrane's negatively charged lipids, while degrading their lipid substrates. **(e)** The LD inserts into the ILV's outer leaflet and electrostatically interacts with negatively charged lipids, thereby abolishing the binding of lipases. **(f)** Lipases, in their closed-inactive form, are more rapidly degraded by luminal proteases. As a result of decreased lipase activity, phospholipids accumulate within the lumen of lysosomes, leading to lipidosis. The latter, along with the accumulation of cholesterol, is the hallmark of drug induced- and hereditary LSD.

Figure 6.7 depicts a proposed model summarizing the information gained by MD simulations, intra-lysosomal enzymatic pharmacology and confocal imaging of LDs within lysosomes. Collectively, these combined results converge to the conclusion that LDs intercalate and accumulate in lysosomal membranes hence inflicting their deleterious impact on the function of lysosome enzymes. In conclusion, lipophilic weak base therapeutics, known as LDs, undergo marked sequestration and concentration within lysosomal membranes, inducing alterations in membrane fluidity and integrity. The lysosomal membrane concentration of LDs provides the first molecular basis for the disruption of the lysosomal central signaling functions and cue sensing. Our findings bear important implications for lysosomal diseases and lysosomal damage as well as the future design of targeted therapeutic agents that evade lysosomal membrane entrapment.

Supplementary Materials: The following are available in the Appendix, Figure D.1: Structure and physicochemical properties of various compounds employed in the current study, Figure D.2: Increased size and number of LAMP1-mGFP labeled lysosomes following treatment with CNSDs, Figure D.3: Representative quantification of the increase in lysosomal parameters following CNSDs treatment, Figure D.4: ILVs contain marked levels of the anticancer drugs DNR and NTD, Figure D.5: Selected conformations of LDs inserted into a DMPC membrane model.

6.7 Acknowledgments

M.S and Y.G.A. conceived and supervised the study; M.M. conceived and supervised the CpHMD simulations; M.S.performed the microscope imaging and analyses; T.F.D.S conducted the CpHMD calculations and analyses as well as wrote the computational methods section; G.L. performed the *in vitro* experiments; M.S., Y.G.A and M.M. wrote the paper.

This research received no external funding. We extend our gratitude to Drs. Nitsan Dahan and Yael Lupu-Haber for their assistance with the quantitative analysis using the InCell 2000 analyzer microscope and InCell investigator software.

We acknowledge the financial support from FCT (SFRH/BD/140886/2018, CEECIND/02300/2017 and UID/MULTI/04046/2019). Computational analyses are based upon work from COST Action STRATAGEM (CA17104), supported by COST (European Cooperation in Science and Technology).

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

6.8 Addendum

Following the detailed description of the LDs membrane partitioning and the required pH-dependent transient deprotonation of several Lewis bases, our work naturally evolved to fully characterize the membrane crossing process. Although useful, the pHRE scheme was unable to overcome large energy barriers associated with the diffusion-based membrane crossing, as evidenced by the previous results. Nevertheless, the implementation of an US-CpHMD methodology [163] enables the sampling of these high energy configurational states, by forcefully driving the system along the membrane normal, coupled with chemical moiety titration. For these systems, we choose the Nintedanib (NTD) and Sunitinib (SUN) molecules as model compounds for hydrophobic Lewis Base drugs. As explained in the Method section 2.4, the system was setup following a steered-MD protocol to generate the starting conformations for the US-CpHMD simulations. The simulations were performed at pH 4.5, 6.0, and 7.5, as these values reflect the pH environment of the lysosomal lumen, the TME and the physiological pH of an healthy cell, respectively.

These Lewis Base drugs are hydrophobic weak bases that possess the ability to (de)protonate according to the pH media of the cellular compartment. Consequently, their diffusion rate of membrane crossing should be significantly impacted by their pH-dependent behaviour. Therefore, we performed US-CpHMD simulations of a more hydrophobic and a more hydrophilic Lewis Base drugs to model the selectivity and permeability of these drugs in distinct cellular compartments.

The full internalization process for both SUN (B) and NTD (A) is described by the potential mean force (PMF) profiles (Figure 6.8). Assuming that the free energy in bulk water solvent is 0, we observe significant differences in both molecules. The slightly more hydrophobic NTD consistently exhibits a higher membrane avidity than the SUN, across all pH values. This should be expected, since not only the overall molecule is more apolar, but the lower shifted NTD pK_a(Figure 6.6) facilitates the necessary deprotonation events. Additionally, NTD exhibits a clear pH-independent preference to become stable in the membrane ester region, as denoted by the free energy minimum in the $\sim$ 0.8 to 1.4nm range (Figure 6.8B), further corroborated by previous experimental and CpHMD simulation findings that highlight the NTD ability to accumulate in the membrane and inducing lysosomal expansion (Figure 6.3). Comparatively, SUN exhibits an overall higher energy profile, hinting at a less favorable membrane partitioning. Furthermore, as evidenced by previous experimental and computational results, SUN accumulates closer to the water-membrane interface than NTD (1.4 to 1.6 nm), more easily transitioning to more solvent exposed regions. These results are consequence of the higher pK_a values (Figure 6.6) that facilitates protonation events, impairing the charged molecule insertion to deeper membrane regions. Overall, both the molecule intrinsic polarity and the titrating group pK_a modulate the membrane avidity, the preferable accumulation region and possibly the diffusion-rate, as the energy barriers to cross the membrane center are fairly exacerbated in more acidic environments (pH 4.5 and 6.0) and for the slightly more polar molecule (SUN). To confirm our reasoning, we used the Inhomogeneous Solubility Diffusion Model (ISDM) procedure to estimate the pH-dependent permeabilities for both compounds from their PMF profiles (Figure 6.9).

There are two distinct trends that correlate very well with previous observations, with the NTD exhibiting overall higher permeability values relative to SUN across all pH values (28–32 $cm \cdot s^{-1}$). For NTD, the permeability is only slightly hampered at the most acidic environments (Figure 6.9 pH 4.5), such as the lysosomal lumen, suggesting that the mostly pH-insensitive NTD freely traverses the membrane bilayer in these cellular compartments and, more importantly, capable to evade, at least partially, the lysosomal entrapment. On the other hand, SUN permeability strongly depends on the compartment pH to deprotonate, further increasing their hydrophobicity and become more membrane amenable. These findings clearly depict a multidrug resistance mechanism for tumors based on the TME acidity. The membrane crossing process for drugs with similar protonation and polarity to SUN will be significantly hampered as the acidic environment will prevent the

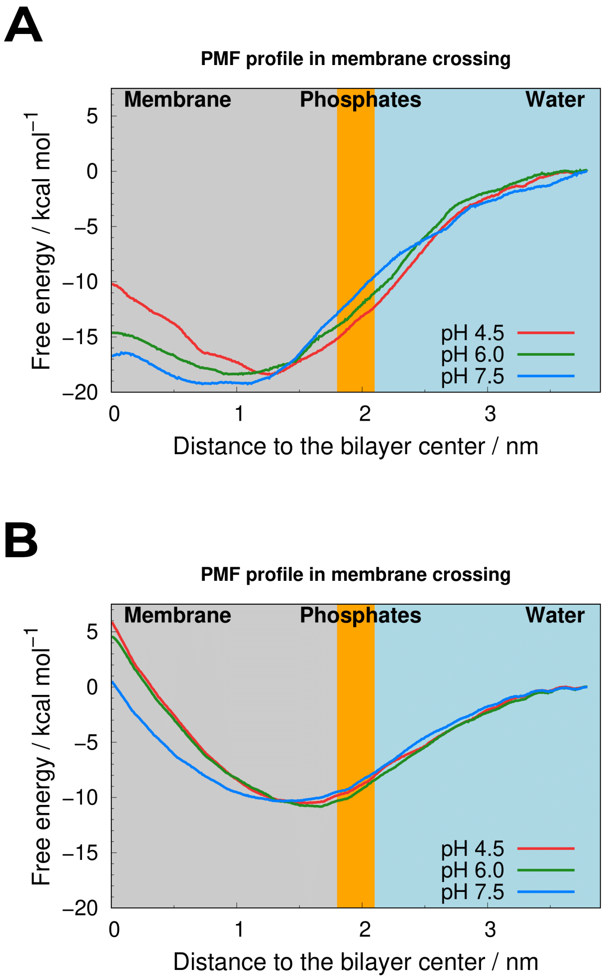


Figure 6.8: Potential of mean force (PMF) profiles for Nintedanib (A) and Sunitinib (B) at different pH values. The free energy calculations were performed using WHAM, with each x-axis value corresponding to an umbrella value. The light blue shaded region (4.0 nm) depicts the solvent. The orange shaded region (2.0 nm) represents the water-membrane interface and the experimental range of fluid DMPC. The grey shaded region represents the hydrocarbon membrane center. This figure was adapted from reference [362].

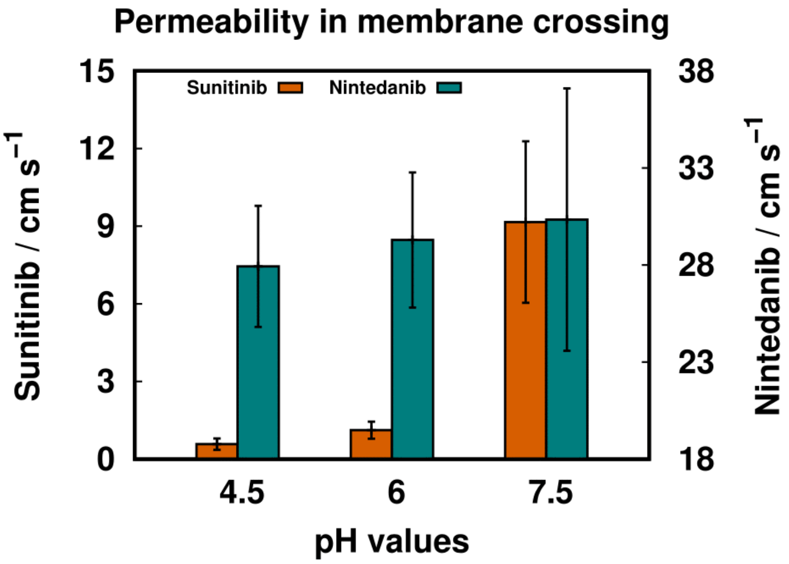


Figure 6.9: Membrane permeability coefficients estimated using the ISDM procedure for SUN (orange) and NTD (green). The error bars were calculated using the jackknife leave-one-out resampling method. This figure was adapted from reference [362].

necessary transient deprotonation, significantly decreasing the permeability. Furthermore, even if the molecule is able to cross into the cytosol, the more basic pH environment promotes the deprotonation and lysosomal capture. At that point, the molecule will once again be exposed to a steep acidic environment, becoming entrapped in the lysosome until it is expelled from the tumor. These two mechanisms are fundamental for tumor cells to prevent these drugs from accumulating at the cytosol in the necessary therapeutic doses. Our hypothesis further expands to the possible side effects of these chemotherapeutics. As higher dosages are administered to counteract these defense mechanisms, the SUN concentration overload in the extracellular environment will preferably target healthy cells (Figure 6.9 pH 7.5) relative to the tumor cells.

In sum, we have characterized the membrane crossing process and the permeabilities of two model compounds of Lewis base drugs (NTD and SUN). Our findings highlighted the presence of two defense mechanisms of multidrug resistance that heavily depend on acidic environment of two distinct media: the TME and the lysosomal lumen. The first defense mechanism prevents the tumor internalization of molecular compounds similar to SUN, as their polarity and proton binding affinities heavily favour the water-membrane interface; the second line of defense entraps the few molecules capable of traversing the membrane inside the acidic lysosomal lumen and expels them via exocytosis. Future research might find these type of chemotherapeutics as perfect candidates to be chemically modified in rational drug design protocols. Our proposal is focused on modifying the amine titrable group into a chemical moiety that becomes neutral at acidic pH's, like carboxylic groups, therefore inverting the protonation and polarity profiles. The advent of enhanced sampling methodologies and protocols, like the US-CpHMD method, will be pivotal in guiding the design of pH-specific anti-tumor drugs and, hopefully, overcoming a cardinal flaw in current chemotherapeutics.

Chapter 7

Conclusions and future perspectives

7.1 Concluding remarks

The main purpose of this thesis was to delve into the underlying molecular mechanisms that govern relevant pH-dependent anti-tumor technologies, which intimately interact in the water-membrane interface. Namely, we studied a class of chemotherapeutics susceptible to pH-dependent defense mechanisms of tumor cell multidrug resistance (MDR) and an acidity-selective transmembrane peptide that aims at overcoming them through conjugated therapies. The first step focused on establishing a reliable methodological framework for our future endeavors. We successfully achieved that with the extension of our CpHMD-L methodology to be compatible with enhanced sampling schemes (US-CpHMD and pHRE) and the integration of cell-like features (pH gradient). Furthermore, we developed multiple protocols and specialized analytical tools to consistently obtain accurate and robust predictions of experimental data. Initially, we devised a multi-step pHRE protocol to deal with a complex transmembrane peptide (*wt*-pHLIP) that was further extended to deal with other membrane crossing biomolecules. A systematic testing with varied complex pHLIP-membrane system sizes allowed the optimization of the system build setup and the pHRE simulation parameters. These optimization steps improved sampling efficiency and convergence speed without sacrificing robustness and accuracy of our configurational sampling. Smaller membrane patches (128 lipids patch) significantly reduced computational load and simulation time, though at the cost of convergence of important properties, such as membrane thickness, peptide thermodynamic stability and pK_a prediction accuracy. Larger system sizes (256 and 352 lipids patch) offered more robust and accurate results, while exhibiting significant computational burdens relative to the standard CpHMD-L method. One must note that the lipid patch size can be further fine-tuned according to the peptide length. Additionally, even if the pHRE scheme significantly sped up the sampling of complex transmembrane systems, the ability to sample rare events remained similarly limited as the CpHMD-L method. Nevertheless, the optimized pHRE protocol and the developed analytical tools became the go-to method for complex membrane systems and the cornerstone of this thesis.

The success of our protocol predicated our consequent studies on the intricate molecular networks that dictate the acidic tumor targeting abilities of the pHLIP peptides. Our motivation was based on the several known experimental roadblocks in improving the *wt* sequence slow kinetics and lack of pH-specificity to tumor cells. Despite the trials and errors, the design of the Var3 peptide was as successful as it was intriguing, since the typical liposomal and cell studies showed incongruous findings. Yet, the remarkable therapeutic improvement over the *wt* peptide was definite. The extension of our CpHMD-L method with the pHRE scheme to incorporate a membrane pH gradient illustrated the limitations of extrapolating the pK^ins of simple liposomal models to cell conditions, and also the necessity of augmenting the realism of our *in silico models* to improve its predictive capabilities. To the best of our knowledge, this is the first successful application of an explicit CpHMD method with a pH gradient on a transmembranar peptide system. Our results highlighted that the pH gradient not only is required for the membrane insertion of Var3 to successfully target the TME, as well as it significantly hinders the *wt* stability in acidic environments. These findings were in excellent agreement with the experimental data. Furthermore, our pK_a predictions highlighted the distinct local electrostatic vicinities enveloping the key aspartate residues (Asp-14 and Asp-13), hinting at a possible pivotal role of the anchor arginine residues (Arg-11 and Arg-9, respectively). The application of our protocol in rationalizing the underlying molecular effects that discerned both peptides provided invaluable data for future optimization of the Var3 sequence. It also propelled us in testing our hypothesis on the arginine role, by simulating multiple *wt* variants with distinct arginine permutations. Indeed, the arginine assumes a fundamental role in defining the electrostatic vicinity of the key aspartate, either by indirectly modulating the proximity of lipid headgroups and the solvent accessibility or by forming tight salt-bridge interactions with the aspartate residue. Conversely, these effects modify the key aspartate proton binding affinity that deeply impacts the overall peptide thermodynamic stability, hence affecting the clinical performance in targeting the TME acidic conditions. The versatility of the pHLIP technology in operating as a drug delivery system and a tumoral biomarker strongly depends on the thermodynamic stability of the membrane-inserted state at the desired pH range. Hence, the development and production of new pHLIP sequences is not a trivial matter and the residue placement, specifically charged and polar residues, has lasting impacts in the viability of newly designed peptides. In sum, the *in silico* methodologies presented in this thesis provide a reliable framework for developing and testing new peptide sequences with accurate predictions of biological phenomena prior to experimental trials.

In the same vein, these methodologies were applied to the study of passive membrane crossing chemotherapeutics, namely Lewis base drugs. These commercialized compounds have been widely used in cancer treatment, though particularly susceptible to a particular mechanism of MDR: lysosomal entrapment. In a collaborative work, the application of CpHMD-L simulations with a pHRE scheme described the transient deprotonation phenomenon required for these drugs membrane partitioning. While their intrinsic hydrophobicity induces a certain membrane avidity, their high pK_a confers a charged nature to the titrable group that prevents membrane internalization. Our data have shown that compounds with higher pK_as tend to accumulate near the water-membrane interface and be effectively sequestrated inside acidic compartments, like the lysosomal lumen. On the other hand, molecules with smaller pK_as and more hydrophobic profiles appear to be more membrane amenable, capable of intercalating in deeper membrane regions. To provide a better assessment of the mechanism of passive membrane crossing, we applied a novel US-CpHMD methodology to study two model anti-tumor compounds: sunitinib (SUN) and nintedanib (NTD). The US-CpHMD method ability to overcome large energy barriers enable the estimation of full PMF profiles along the membrane bilayer. These profiles confirmed our initial findings that molecules with similar polarity and protonation profiles to NTD are energetically more favorable to traverse the apolar hydrocarbon region, while SUN-like molecules need to overcome larger energy barriers, particularly at more acidic environments, to reach the membrane center. Indeed, permeability calculations corroborated how acidic environments significantly impair SUN ability to cross the membrane bilayer. Interestingly, if environmental acidity impairs their membrane crossing ability, then the pH range of the TME effectively acts as a first defense mechanism of MDR that hampers chemotherapeutics internalization. Moreover, the extracellular environment of healthy tissue is more favorable than the tumor cells, which coupled to higher drug dosage administration, aggravates the known side-effects of chemotherapeutics. Summarily, our work has characterized the molecular intricacies of Lewis base drugs facing two distinct pH-dependent MDR mechanisms: the TME acidity and the lysosomal entrapment. This was only possible by further applying a novel method US-CpHMD which provided PMF profiles that discerned the energy barriers associated with different pH media and the permeability estimations that neatly correlated with the experimental data and the characterization process. More importantly, we have exposed a cardinal flaw in the drug design of these commercialized drug compounds as they preferably targets healthy tissue in detriment of solid tumor cells.

7.2 Future perspectives

As a final note, this thesis has shown, throughout every chapter, the reliability and accuracy of *in silico* models in predicting experimental data and explaining biological phenomena, proving themselves as pivotal complementary strategies for rational peptide and drug design. More specifically, we have highlighted the integral role of pH-dependency in tumor features and how it is often overlooked in the development of novel drugs and technologies.

The tumor MDR has multiple defense mechanisms ranging from altered metabolic pathways to environmental effects, like the aforementioned pH-dependent mechanisms. Therefore, our proposal to address these challenges focuses on two main avenues: chemically modifying the Lewis bases and improving the realism of our pHLIP-membrane model. Regarding the chemotherapeutics, we will modify the compounds' Lewis base for an anionic group that should invert the permeability profile, therefore increasing the membrane avidity in more acidic environments. At the moment, simulations are being performed on a modified sunitinib, where the titrable tertiary amine was exchanged for a carboxylic group as a proof of concept. Concerning the pHLIP peptides, despite the significant improvements of Var3, other properties can be further fine-tuned: improve the insertion and exit kinetics; determine the energy barriers associated with transposing a wide array of cytotoxic agents, like Lewis base drugs; how to modify the sequence to facilitate pharmacological delivery; and improving the peptide selectivity by permuting the L16H variant with other cationic residues to regulate the peptide insertion in a clinically relevant pH range. Meanwhile, we also aim at improving the realism of our models. The pH gradient implementation was a pivotal, yet only one step in bridging the *in silico* and cell worlds. Additional methodological work needs to be addressed to incorporate cholesterol or other distinct lipid mixtures in our model. Previous work has been done on titrating anionic PA/PC lipid bilayers[165], though that protocol needs to be fully incorporated into our current methodological approach. Furthermore, other biologically relevant lipids will require extensive parametrization in the current or other more compatible force fields (CHARMM/AMBER) to avoid configurational sampling pitfalls.

The applications of pHRE and US-CpHMD methodologies showed a clear improvement on the study of complex systems relative to the standard CpHMD-L method. Both methodologies provided either faster protonation/conformational sampling or the ability to sample higher energy configurations. The devised protocols and analytical tools will open our horizons for more complex and larger pH-dependent biological systems, like ATP/ADP transporters, the OmpG transmembrane protein, or the Aquaporin-1 protein channels. Additionally, the CpHMD-L method will be further expanded and optimized by coupling enhanced sampling schemes together, like the replica-exchange umbrella sampling (REUS), or implementing other techniques, such as metadynamics. Particularly, the implementation of the REUS scheme will combine the best features of the two methods, which should improve and speed up the convergence of quality sampling of higher energy configurations. These additional techniques will establish the CpHMD-L method as a package for a diverse array of biomolecular simulations. In the end, these methodologies will continue to evolve as they push the boundaries of our knowledge on the role of pH in biological systems.

Appendix A

Supporting Information for Chapter 3

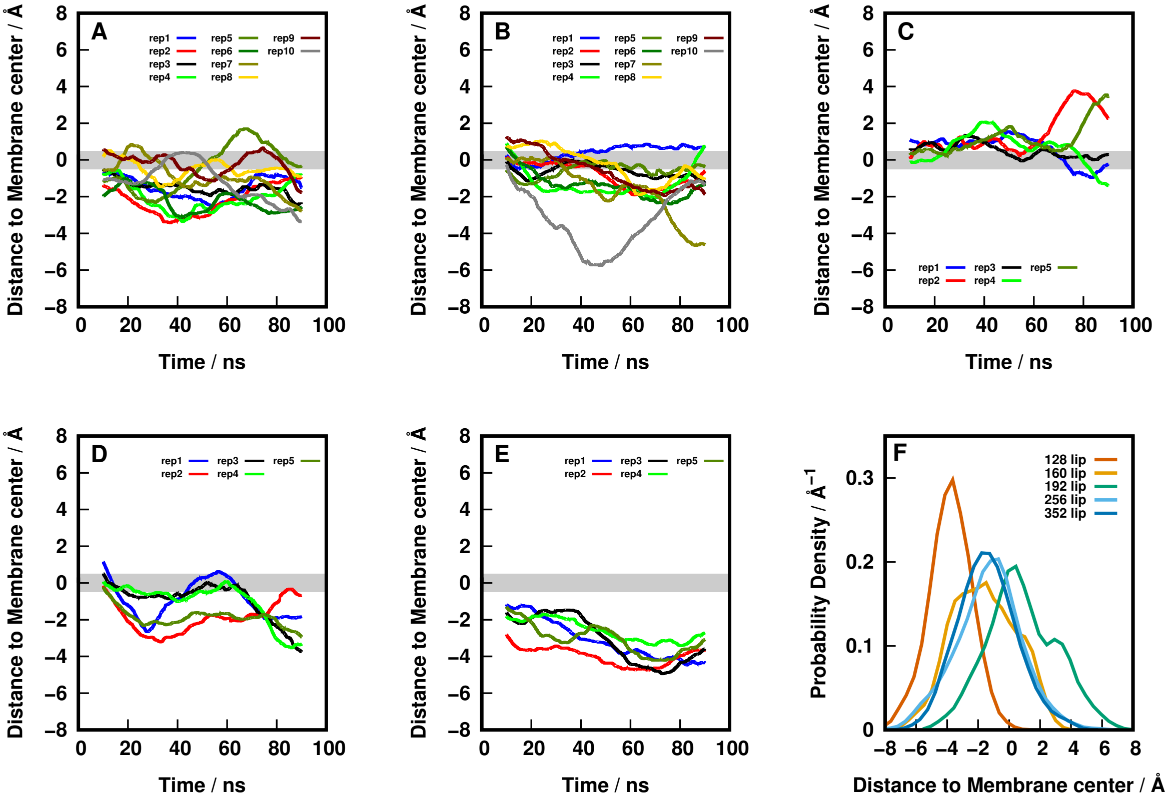


Figure A.1: Time series plot of Asp25 residue distance relative to the membrane center for all simulated systems at pH 6.0 (352 lipids, A; 256 lipids, B; 192 lipids, C; 160 lipids, D; 128 lipids, E). This residue is known to sample the membrane center in unperturbed membrane systems.[103] At 0 Å, the grey region represents the membrane center, while positive distance values indicate the residue is at the N-terminus membrane monolayer, while negative distance values refer to the C-terminus monolayer. Probability density distribution calculated with the previous data for every lipid membrane system (F). The time series show the overall peptide vertical movement along the simulations and the distributions depict the average peptide vertical position using Asp25 as reference.

System	H-bond Abund.	H-bond persistence (ps)
256	21.4 ± 6.6	135.6 ± 42.9
352	34.1 ± 7.7	205.0 ± 60.2

Table A.1: The abundance and persistence of the hydrogen bonds between the Asp14 carboxylic oxygen atoms and the free oxygen atoms of the phosphate groups in the two larger lipidic systems. The H-bond abundance was calculated over the complete equilibrated replicas in all replicates, using a cutoff of 0.3 nm to define a stable hydrogen bond. The H-bond persistence is the average length of all simulation segments where the hydrogen bond occurs.

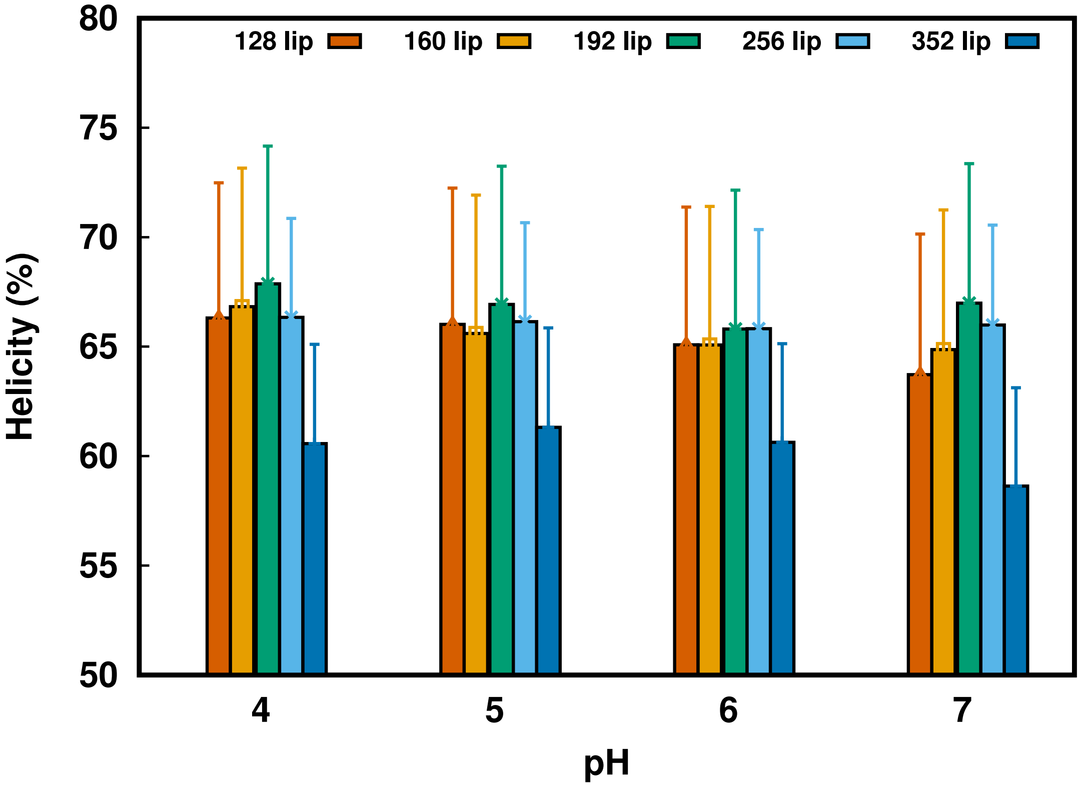


Figure A.2: Average helicity percentage of *wt*-pHLIP in state III obtained through DSSP analysis, at each pH value, for every lipid membrane system. The represented helicity values are a measurement of peptide folding equilibrium in the sampled conformational space. The errors bar shown were obtained by calculating the standard error of the mean.

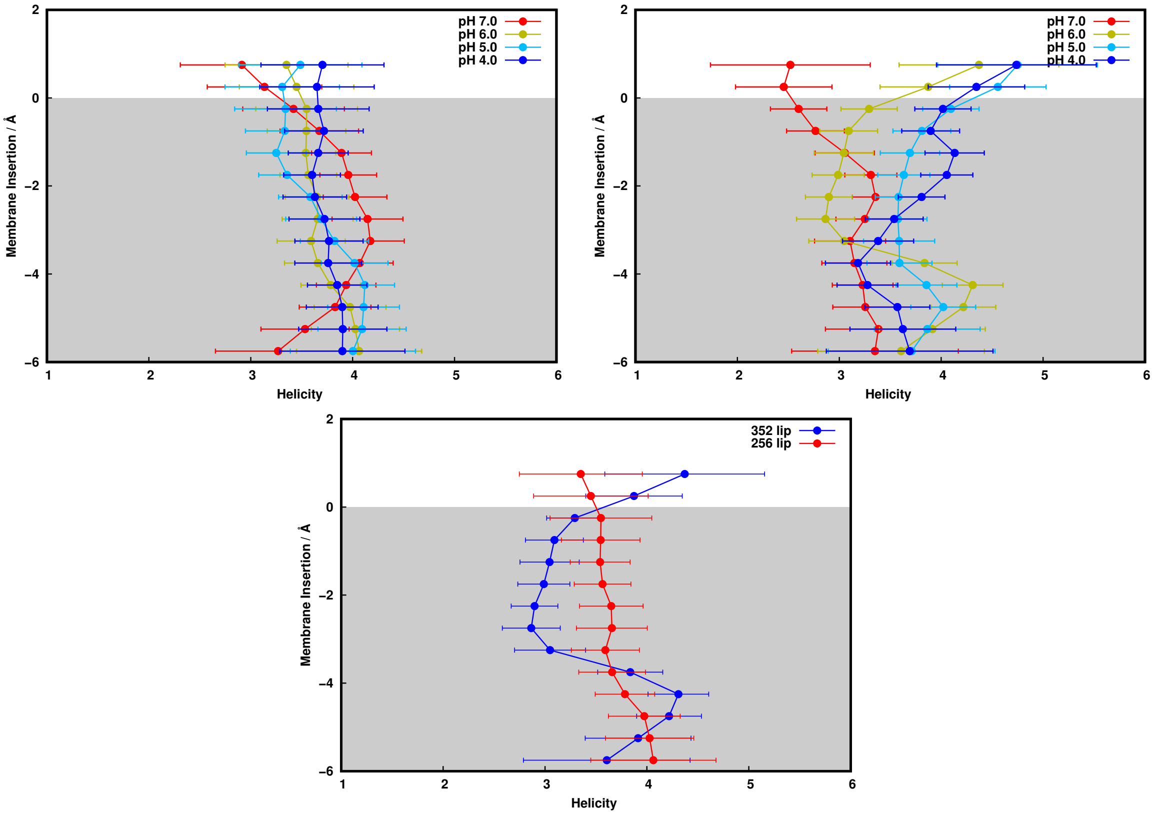


Figure A.3: Helicity profiles for the Asp14 region of pHLIP (residues 12-16) at different pH values for the 256 (top left) and 352 (top right) lipid systems. A direct comparison between the two systems at pH 6.0 is also shown (bottom plot). The helicity is quantified by counting the number of residues in helical secondary structure classifications from the DSSP criteria.[363]

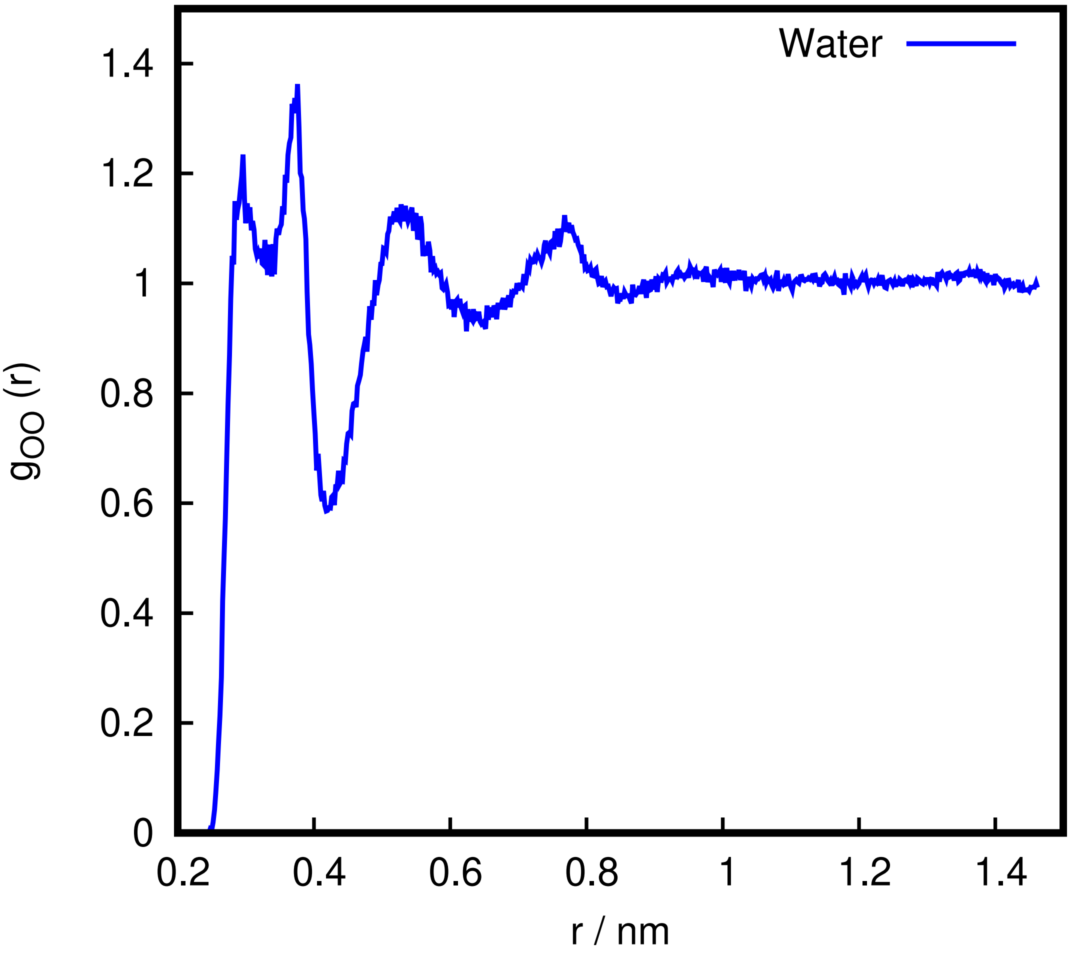


Figure A.4: Water radial distribution function (RDF) from an aspartate pentapeptide in a 128 DMPC membrane bilayer. The data was extracted from previous published simulations.[27] Both water oxygens were used for the calculations. The first, second, and third solvation shells are at 0.42, 0.65 and 0.85 nm respectively. The two peaks inside the first solvation shell appear due to established hydrogen bonds with either oxygen of the carboxylic group.

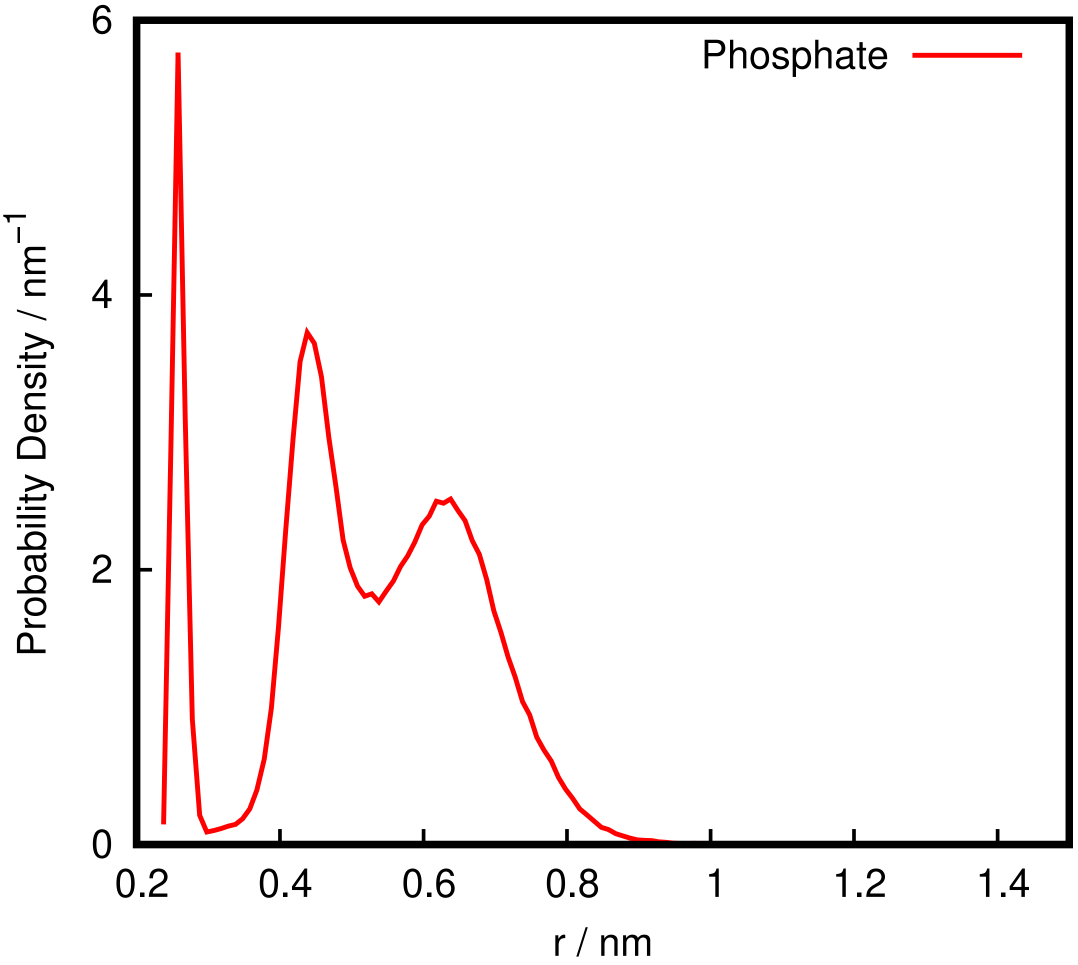


Figure A.5: Minimum phosphate distance distributions for pHLIP Asp14 residue in the 256 POPC membrane bilayer. Both free oxygens of the phosphate group were used for the calculations. The interaction shell at 0.52 nm encapsulates all local interactions with the phosphate groups that compete with water molecules and include both direct and water-mediated interactions.

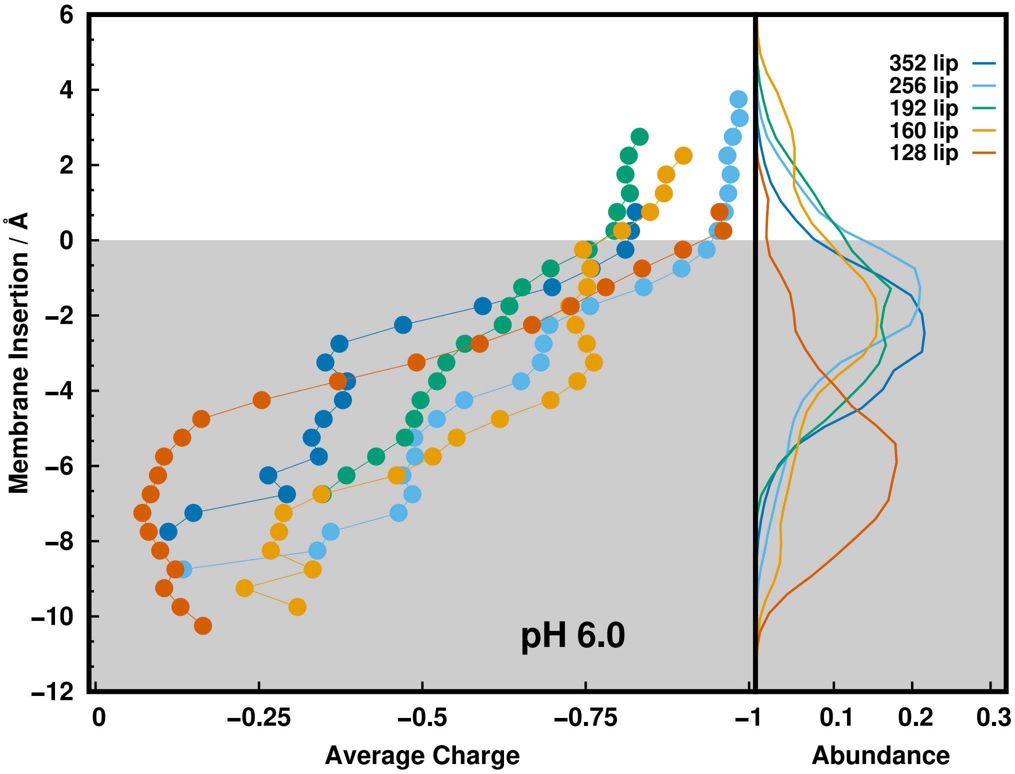


Figure A.6: Protonation profiles and membrane insertion distributions of Asp14 at pH 6.0 for all simulated systems, including the one with 128 lipids. The membrane is highlighted using a grey-shaded region. The negative insertion values indicate residue insertion relative to the water/membrane interface defined by the phosphate groups (0 Å). The positive insertion values indicate that the residue is solvated by bulk water.

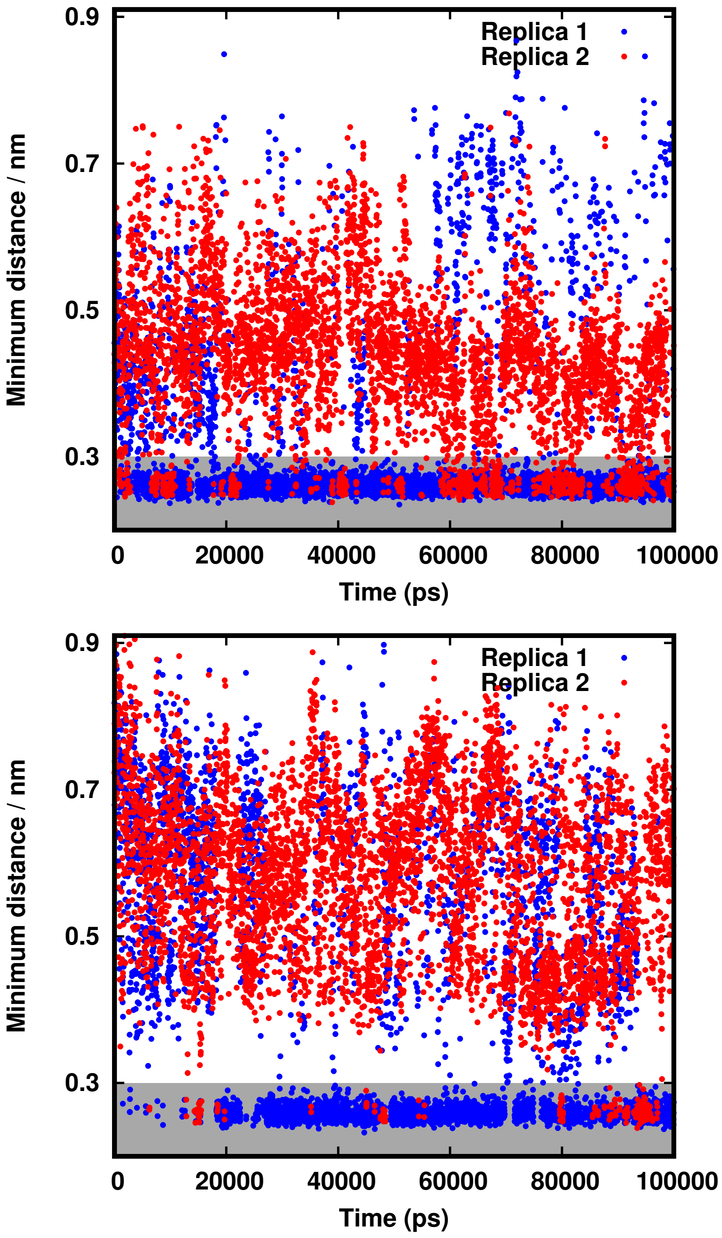


Figure A.7: Time series plot of the minimum distance between the Asp14 carboxylic oxygen atoms and the free oxygen atoms of the phosphate groups. The data was obtained from two replicas of the 352 (upper plot) and 256 (bottom plot) lipid size system, illustrating two very distinct behaviours in our simulations. In replica 1 (blue dots), we observe a "lock" state behaviour, with Asp14 persistently interacting with a phosphate group most of the simulation time. Meanwhile, in replica 2 (red dots), the exhibited behaviour reveals an interchange between "on/off" states, where Asp14 is able to establish and break hydrogen bonds throughout the simulation. The grey shaded region highlights the range of distances (below 0.3 nm) defining a stable hydrogen bond between Asp14 and a phosphate group.

Appendix B

Supporting Information for Chapter 4

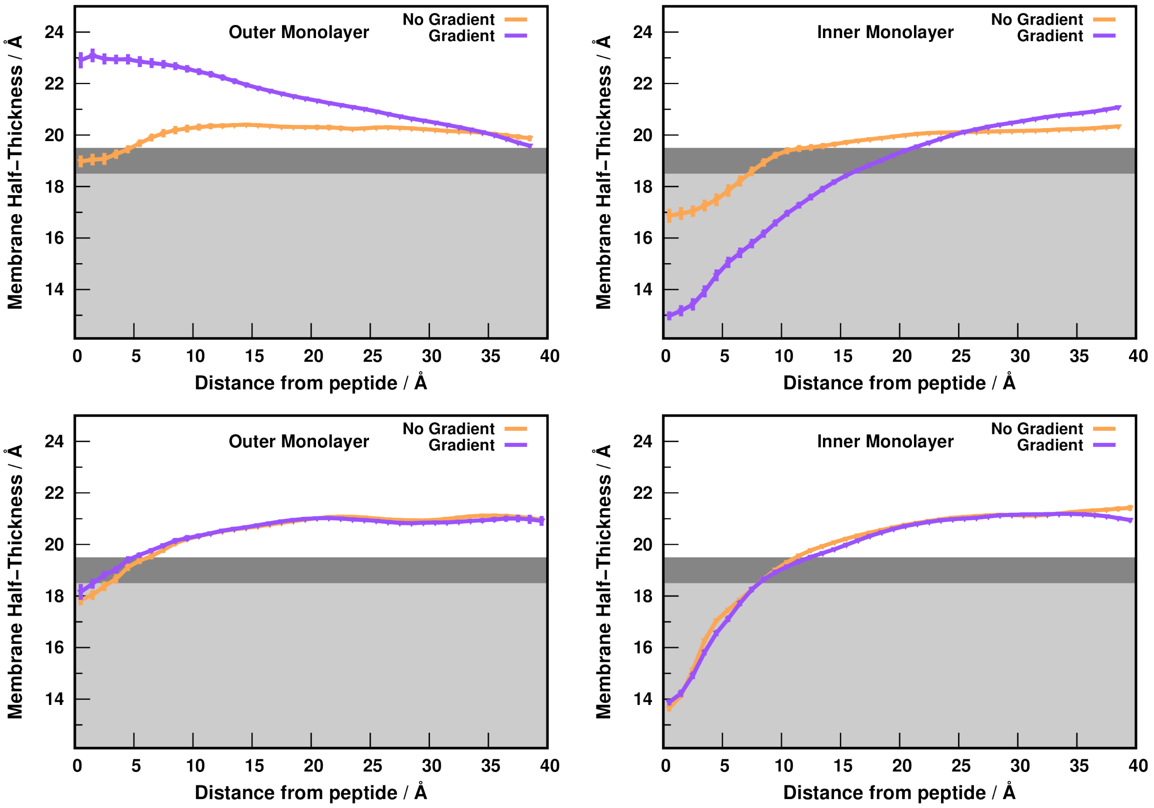


Figure B.1: Outer (left) and inner (right) monolayer half-thickness values in the *xy* plane along the *wt* (top) and Var3 (bottom) peptides distance. The representations show the membrane deformations induced by the peptides, in both setup conditions at pH 6.0, up to the 'bulk' lipids (15–40 Å away). The error bars were calculated from the SEM using all replicates every 0.5 Å and they were plotted every 1 Å for visual clarity. The grey shaded region is the membrane interior, while the dark grey region corresponds to the experimental range of fluid POPC (~19 Å for each monolayer)[214].

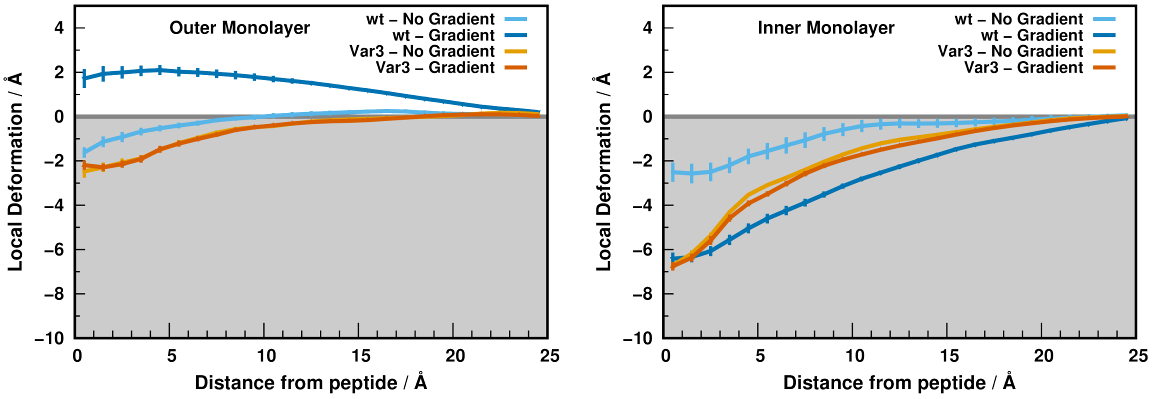


Figure B.2: Outer (left) and inner (right) local monolayer deformations induced by the *wt* (blue/cyan) and Var3 (orange/brown) peptides, in both setup conditions at pH 4.0. These deformations values are calculated from the half thickness values calculated along the *xy* plane distance to the peptides. The maximum distance shown (∼ 25 Å) corresponds to the unperturbed 'bulk' lipids. The error bars were calculated every 1 Å, using all replicates, with a SEM method and they were plotted every 1 Å for visual clarity. The grey-shaded region corresponds to the membrane interior.

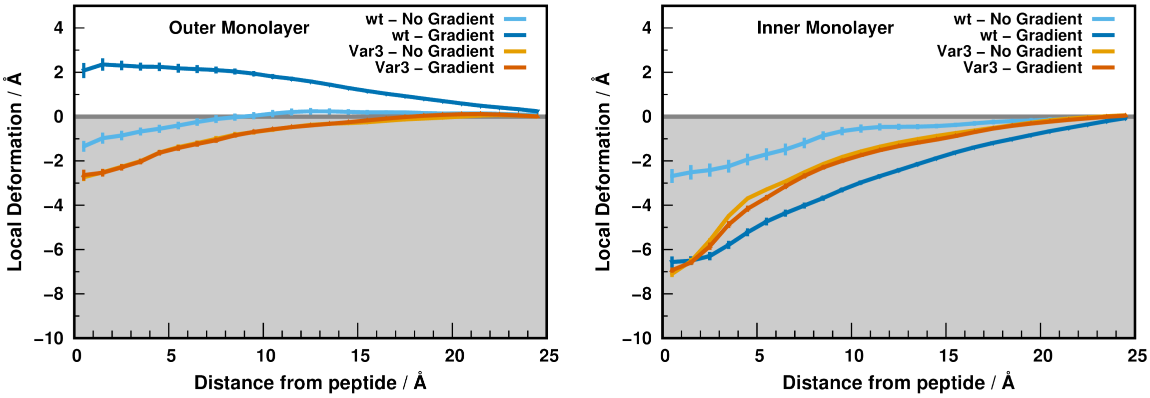


Figure B.3: Outer (left) and inner (right) local monolayer deformations induced by the *wt* (blue/cyan) and Var3 (orange/brown) peptides, in both setup conditions at pH 5.0. These deformations values are calculated from the half thickness values calculated along the *xy* plane distance to the peptides. The maximum distance shown (∼ 25 Å) corresponds to the unperturbed 'bulk' lipids. The error bars were calculated every 1 Å, using all replicates, with a SEM method and they were plotted every 1 Å for visual clarity. The grey-shaded region corresponds to the membrane interior

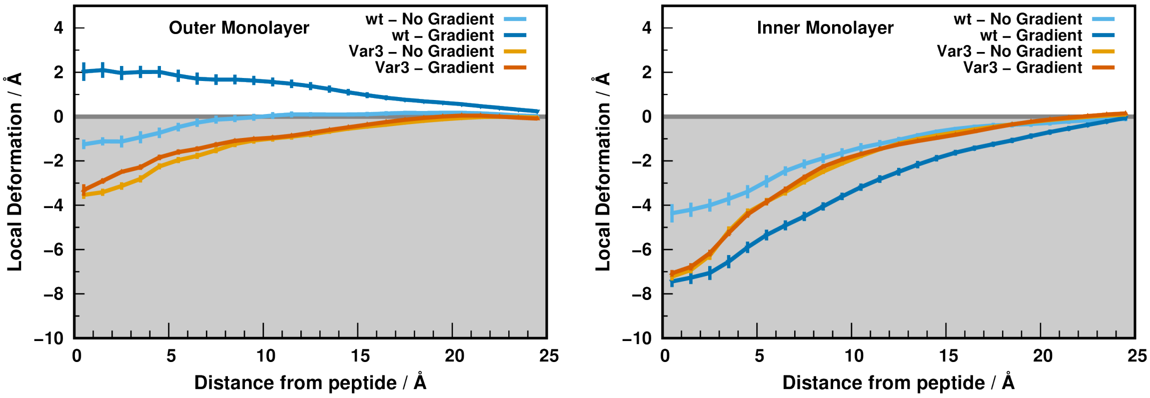


Figure B.4: Outer (left) and inner (right) local monolayer deformations induced by the *wt* (blue/cyan) and Var3 (orange/brown) peptides, in both setup conditions at pH 7.0. These deformations values are calculated from the half thickness values calculated along the *xy* plane distance to the peptides. The maximum distance shown (∼ 25 Å) corresponds to the unperturbed 'bulk' lipids. The error bars were calculated every 1 Å, using all replicates, with a SEM method and they were plotted every 1 Å for visual clarity. The grey-shaded region corresponds to the membrane interior

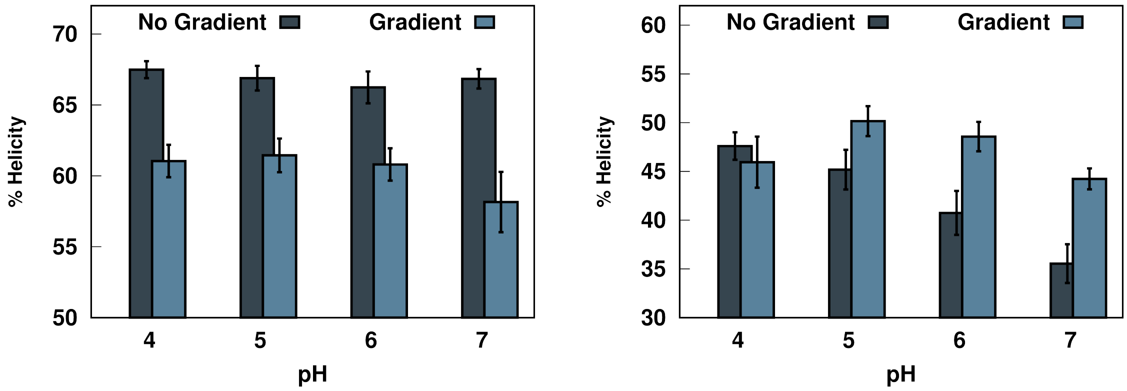


Figure B.5: Percentage of α-helix content of *wt* (left) and Var3 (right) peptides in both gradient and non-gradient pH setups, at all pH values. Note that, in the gradient setup, pH refers to the pH_{out}, since pH_{in} is fixed at 7.2. The average percentage of helicity was calculated using the equilibrated segment of all replicates. The error values were calculated with a standard error of the mean (SEM).

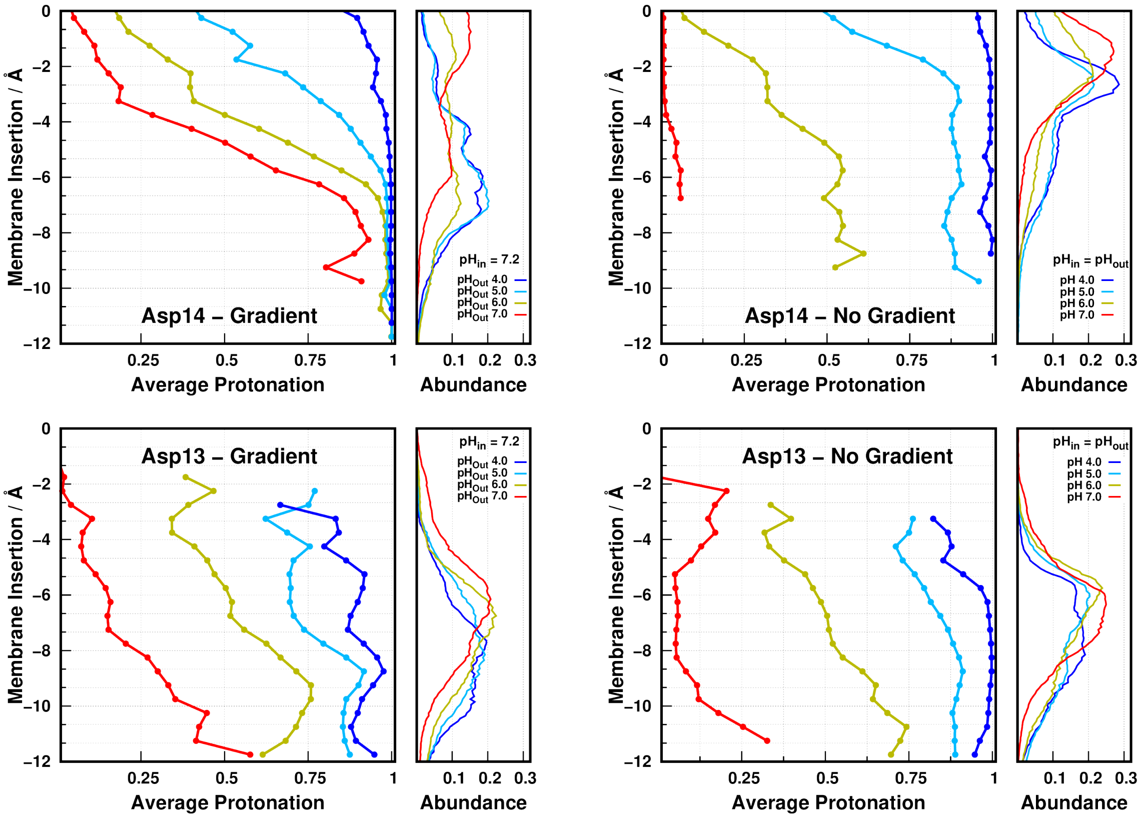


Figure B.6: Protonation profiles of Asp14 (*wt*; top) and Asp13 (Var3; bottom) in a Gradient (left) and no Gradient (right) setup. The profiles indicate the average protonation of the Asp residue, at each pH value, for each insertion slice (left sub-plots) and the histogram distribution of residue insertion in the membrane for each pH value (right sub-plots). The protonation errors bars were omitted for clarity and the average error size was ∼0.1.

Appendix C

Supporting Information for Chapter 5

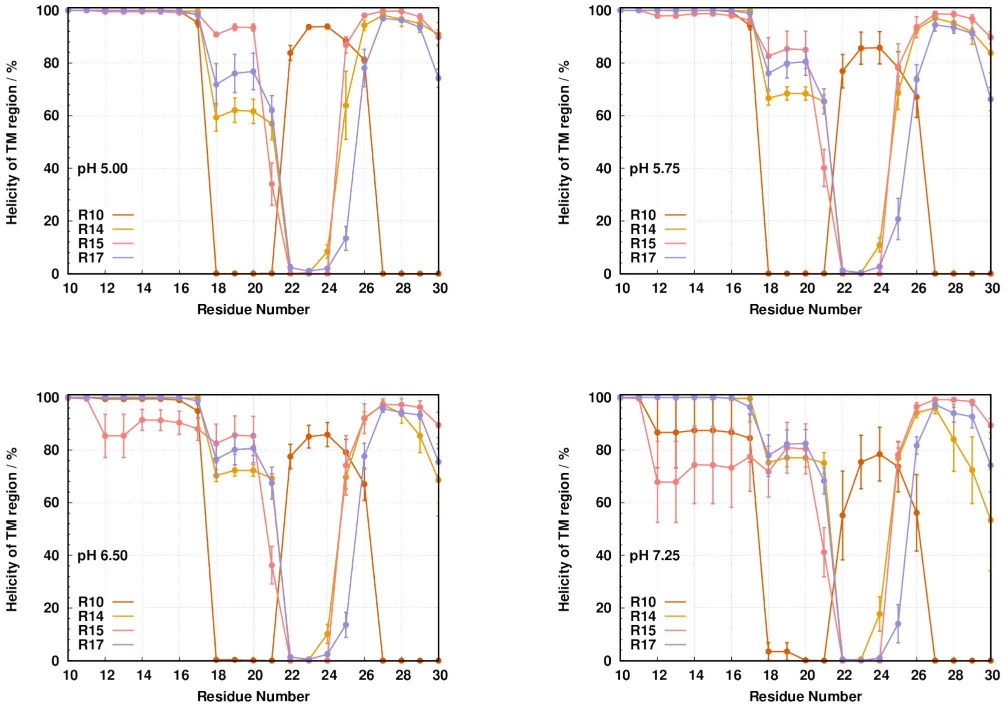


Figure C.1: Average helicity (α-helix) percentage per residue number for all peptide variants, at all pH values. We focused our analysis on the transmembrane region (10^{th}–30^{th} residue). The error bars were obtained from the SEM.

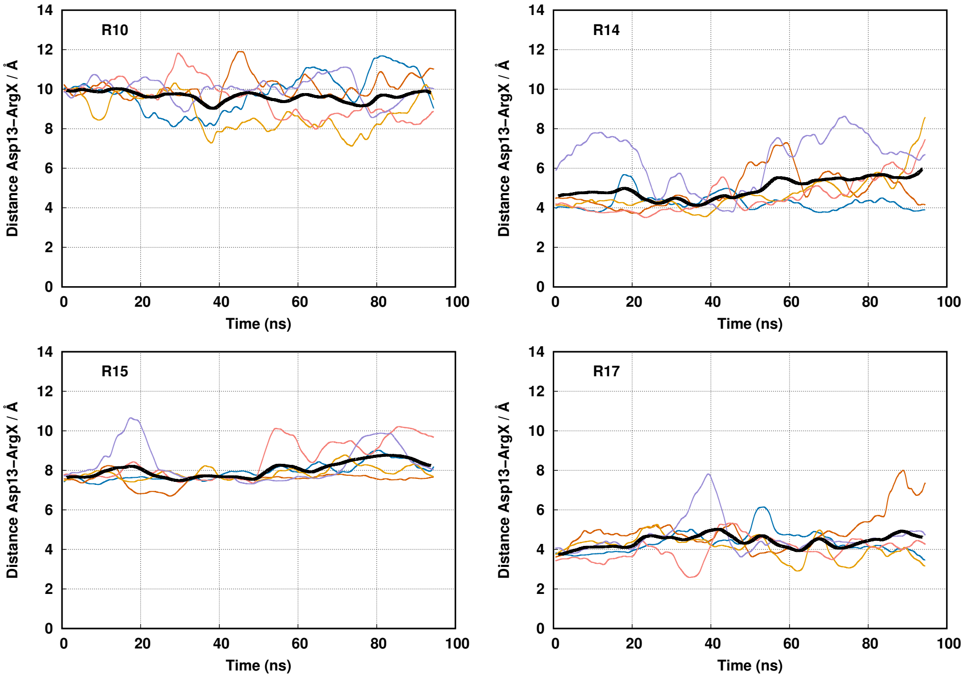


Figure C.2: Time series data of the interaction distance between Asp13 and the key Arg of all peptide variants at **pH 5.00**. In each plot, the colored trends represent a different replicate, while the black trend computes the average of all replicates. All data shown was obtained from a sliding window average (bin size of 1000) to reduce the data noise.

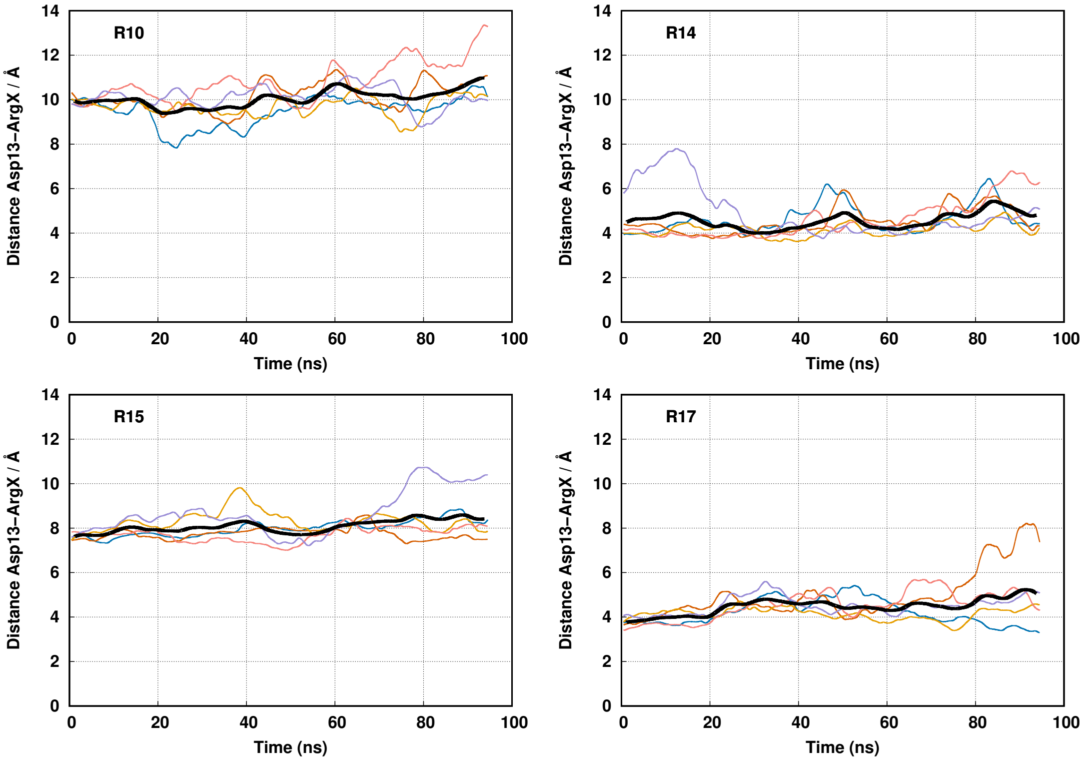


Figure C.3: Time series data of the interaction distance between Asp13 and the key Arg of all peptide variants at **pH 5.75**. In each plot, the colored trends represent a different replicate, while the black trend computes the average of all replicates. All data shown was obtained from a sliding window average (bin size of 1000) to reduce the data noise.

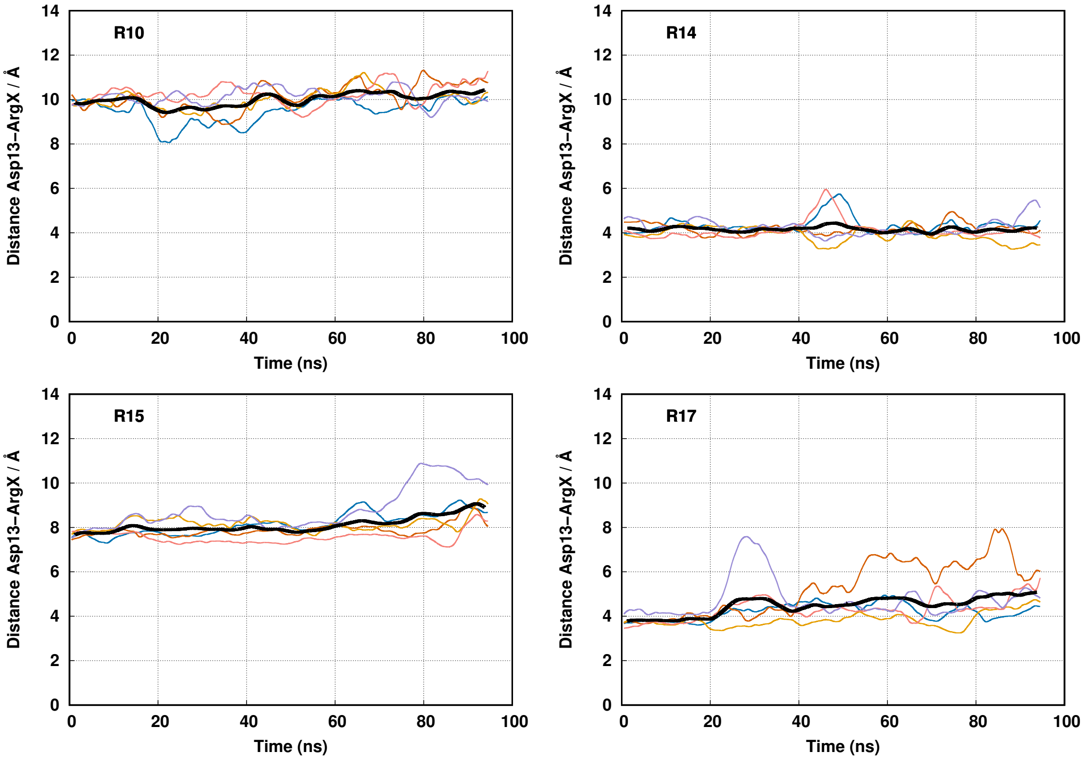


Figure C.4: Time series data of the interaction distance between Asp13 and the key Arg of all peptide variants at **pH 6.50**. In each plot, the colored trends represent a different replicate, while the black trend computes the average of all replicates. All data shown was obtained from a sliding window average (bin size of 1000) to reduce the data noise.

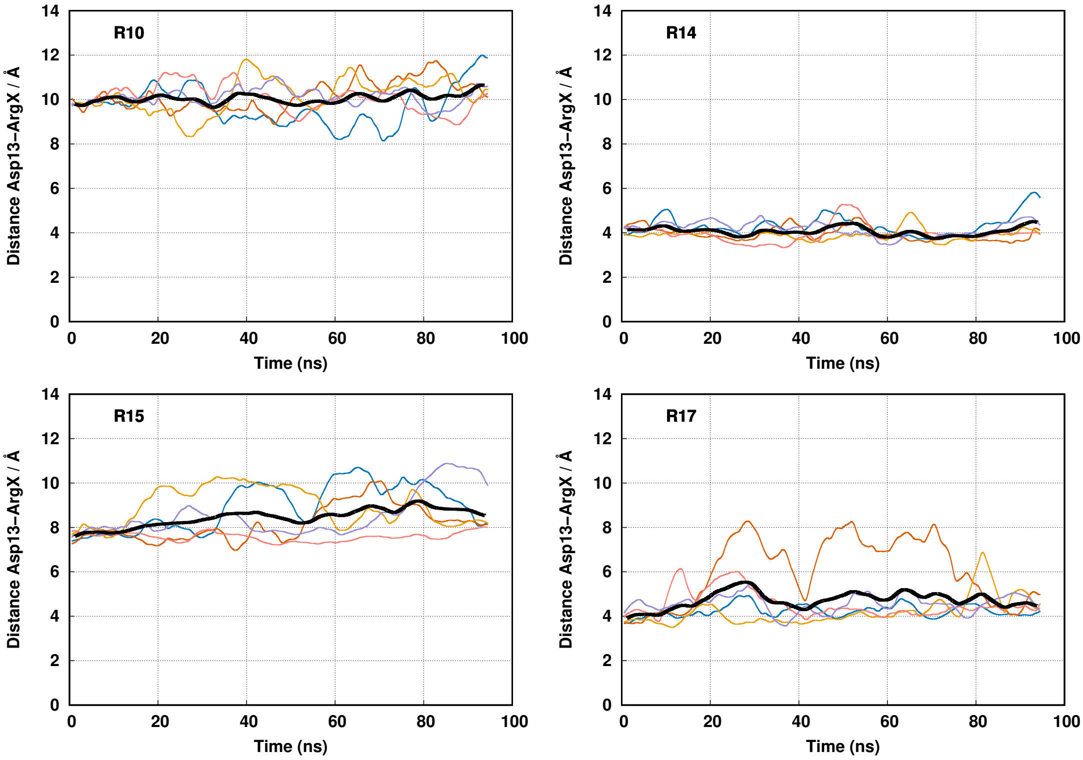


Figure C.5: Time series data of the interaction distance between Asp13 and the key Arg of all peptide variants at **pH 7.25**. In each plot, the colored trends represent a different replicate, while the black trend computes the average of all replicates. All data shown was obtained from a sliding window average (bin size of 1000) to reduce the data noise.

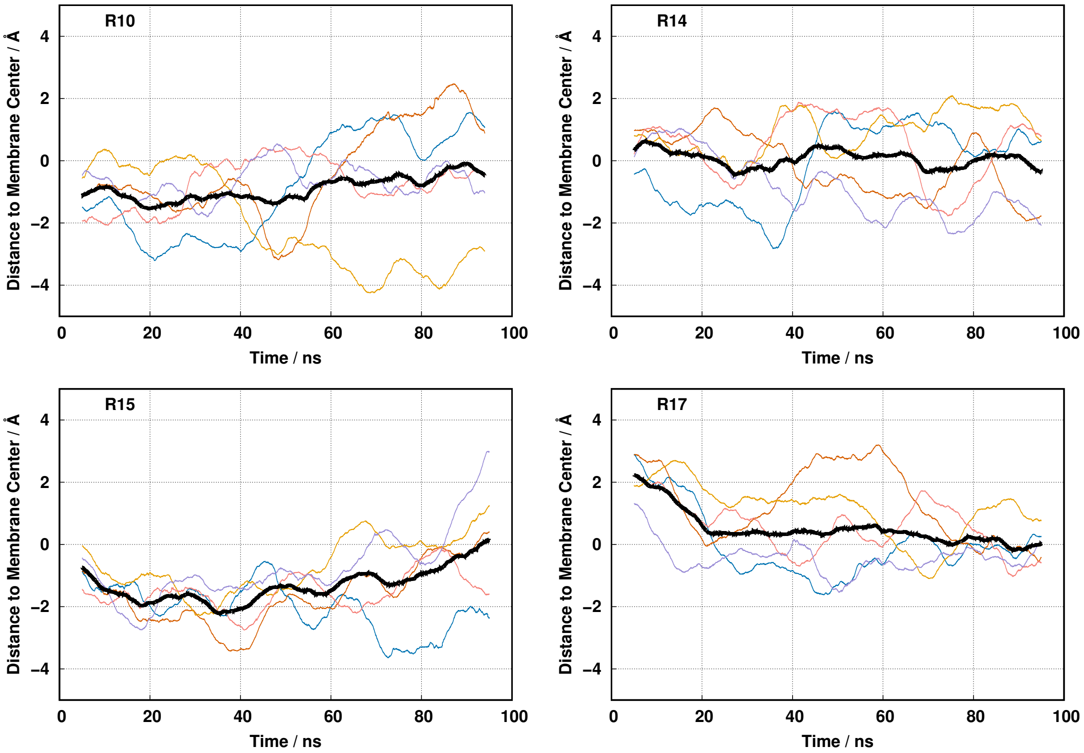


Figure C.6: Time series data of the Leu21 distance to the membrane center, for all peptide variants, at **pH 5.00**. In each plot, the colored trends represent a different replicate, while the black trend computes the average of all replicates. All data shown was obtained from a sliding window average (bin size of 1000) to reduce the data noise.

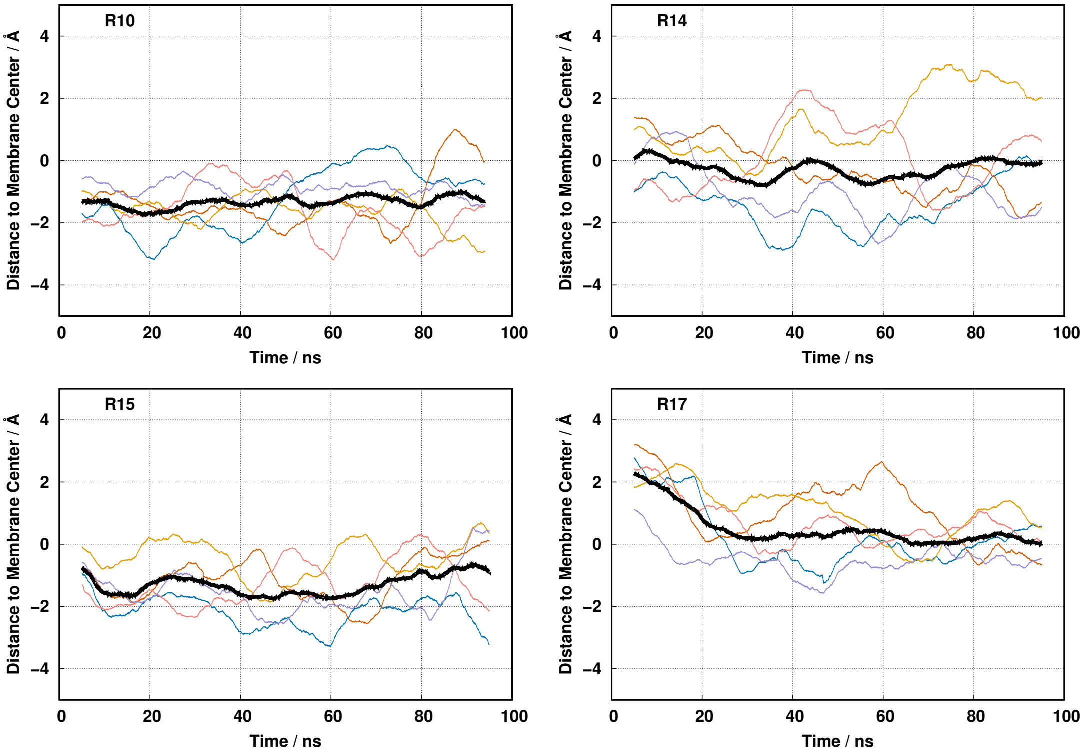


Figure C.7: Time series data of the Leu21 distance to the membrane center, for all peptide variants, at **pH 5.75**. In each plot, the colored trends represent a different replicate, while the black trend computes the average of all replicates. All data shown was obtained from a sliding window average (bin size of 1000) to reduce the data noise.

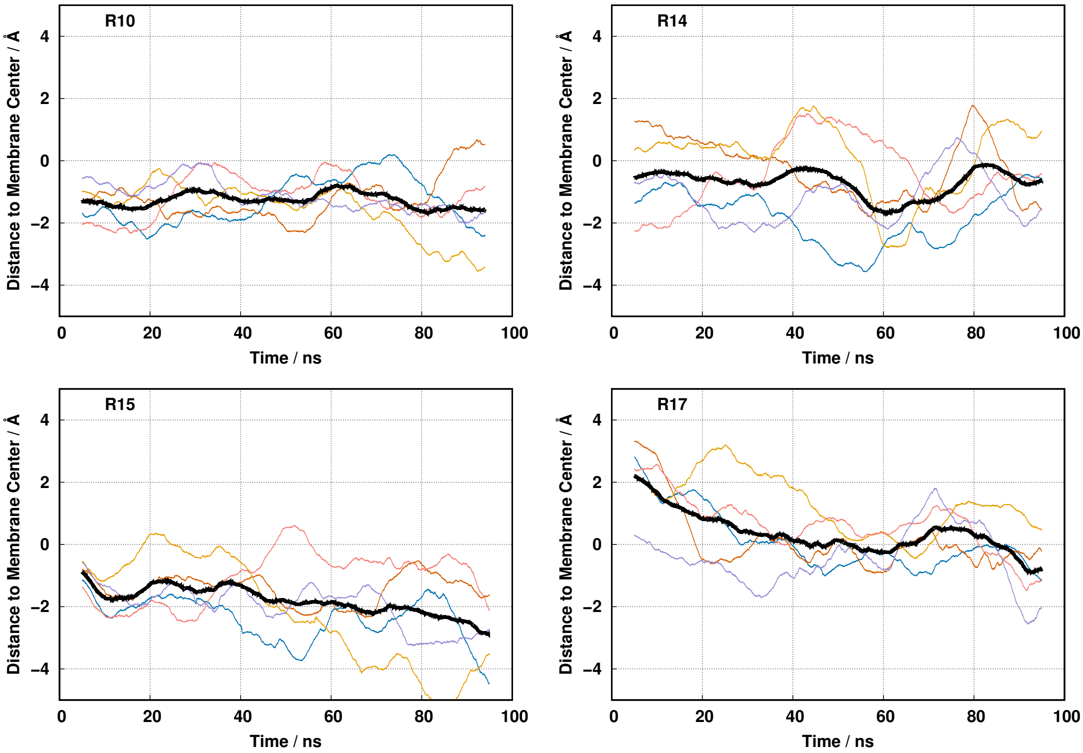


Figure C.8: Time series data of the Leu21 distance from the membrane center for all peptide variants, at **pH 6.50**. In each plot, the colored trends represent a different replicate, while the black trend computes the average of all replicates. All data shown was obtained from a smoothing procedure over the original data, using a moving window average (bin size of 1000) to reduce the data noise.

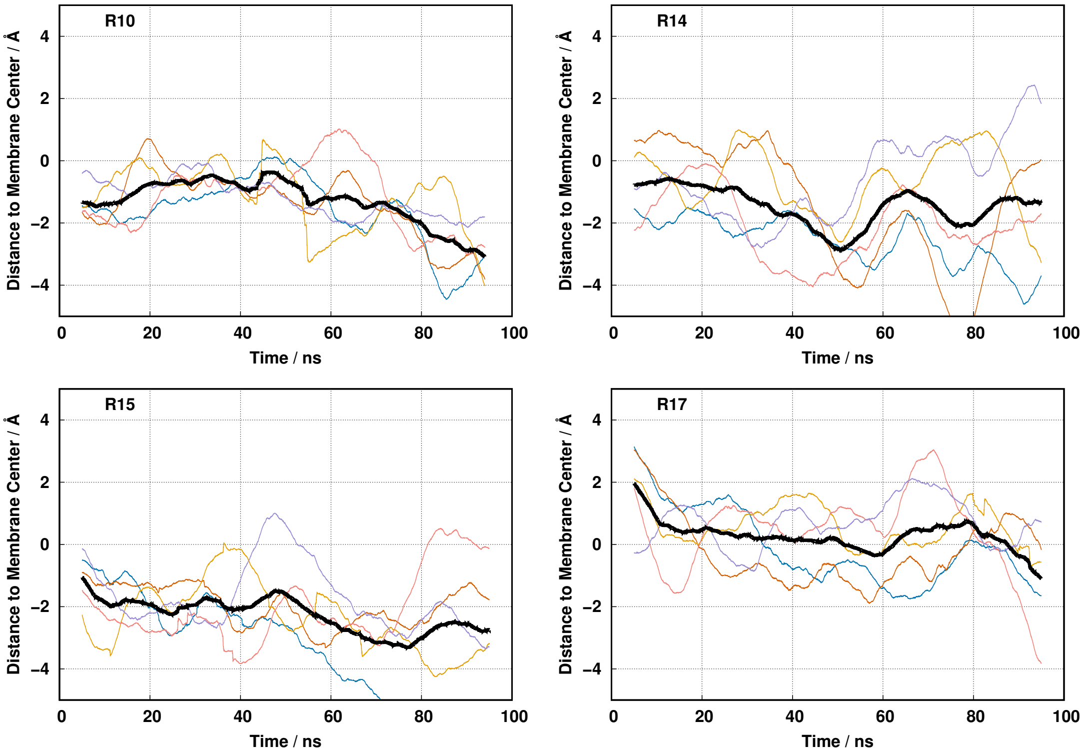


Figure C.9: Time series data of the Leu21 distance from the membrane center for all peptide variants, at **pH 7.25**. In each plot, the colored trends represent a different replicate, while the black trend computes the average of all replicates. All data shown was obtained from a smoothing procedure over the original data, using a moving window average (bin size of 1000) to reduce the data noise.

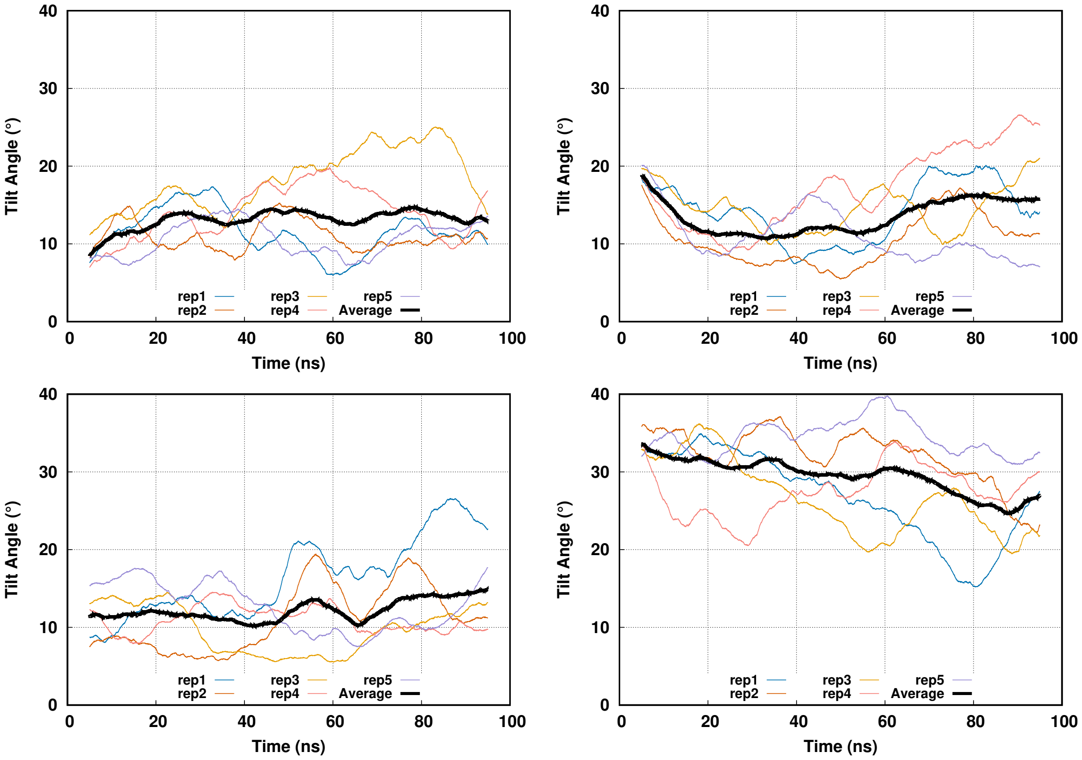


Figure C.10: Time series data of the peptide tilt angle (degrees °), between the H2 (22^{th} to 30^{th}) helical segment and the membrane normal, of each corresponding peptide variant at a representative pH value (**pH 5.00**). The helical segment was identified as the regions with higher helical content after the central Leu21 until the C-terminus. In each plot, the colored trends represent a different replicate, while the black trend computes the average of all replicates. All data shown was obtained from a smoothing procedure over the original data, using a moving window average (bin size of 1000) to reduce the data noise.

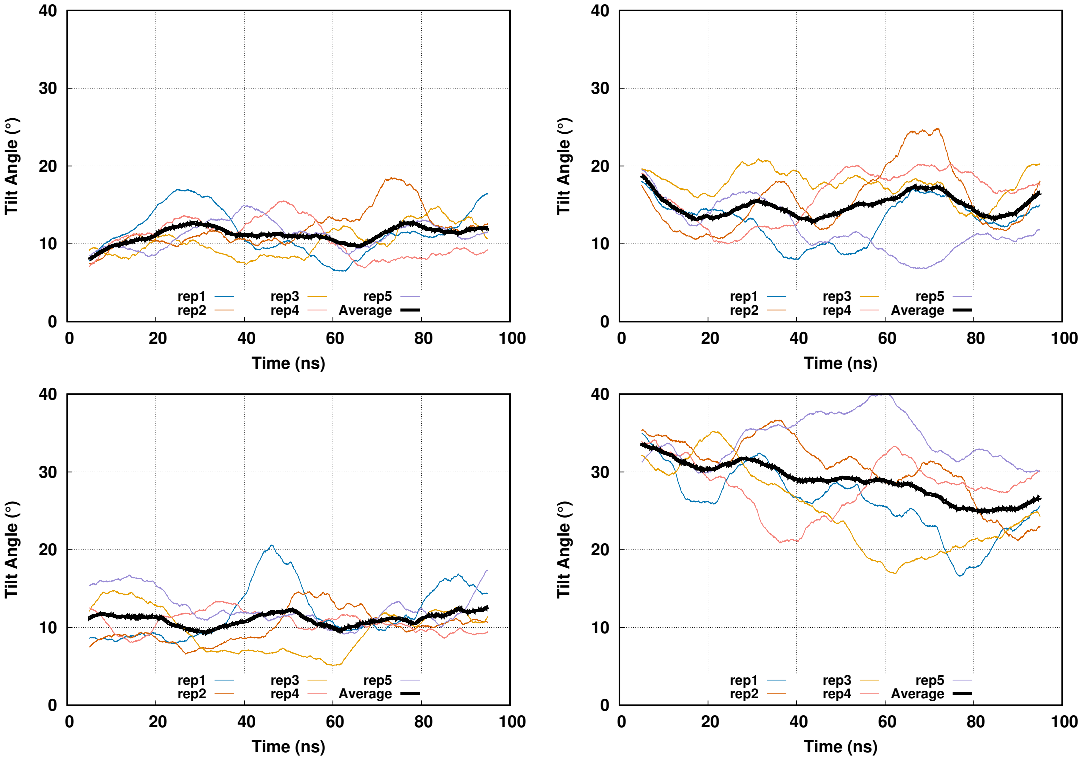


Figure C.11: Time series data of the peptide tilt angle (degrees °), between the H2 (22th to 30th) helical segment and the membrane normal, of each corresponding peptide variant at a representative pH value (**pH 5.75**). The helical segment was identified as the regions with higher helical content after the central Leu21 until the C-terminus. In each plot, the colored trends represent a different replicate, while the black trend computes the average of all replicates. All data shown was obtained from a smoothing procedure over the original data, using a moving window average (bin size of 1000) to reduce the data noise.

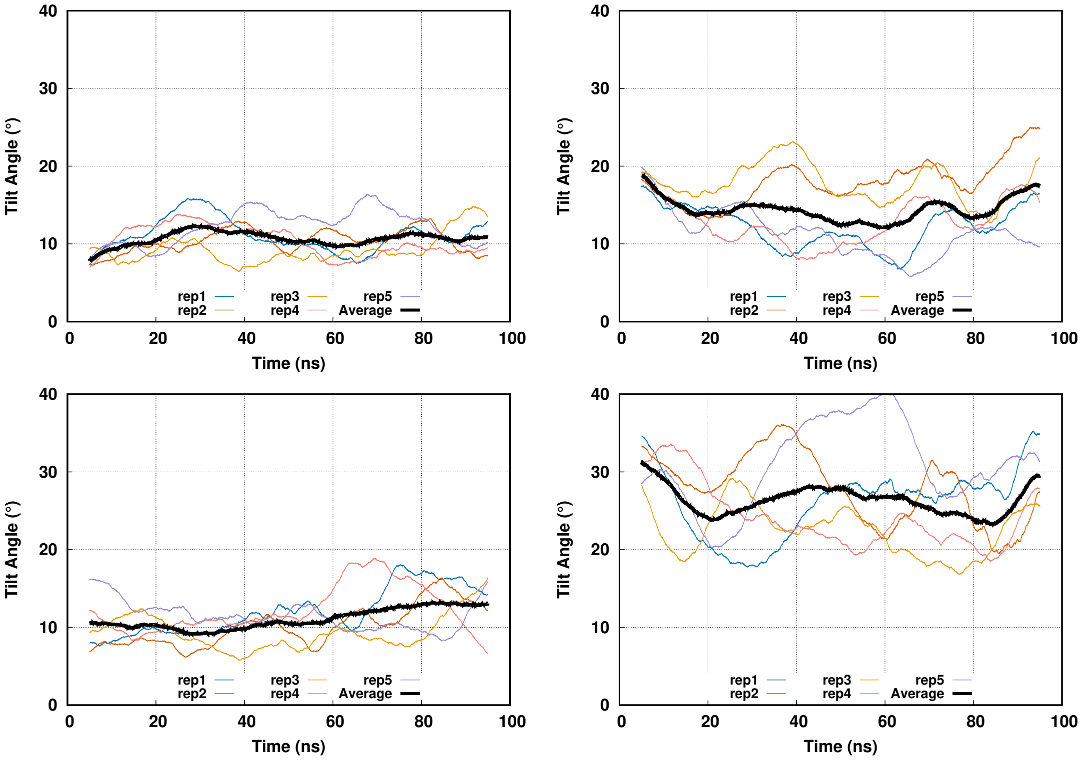


Figure C.12: Time series data of the peptide tilt angle (degrees °), between the H2 (22^{th} to 30^{th}) helical segment and the membrane normal, of each corresponding peptide variant at a representative pH value (**pH 6.50**). The helical segment was identified as the regions with higher helical content after the central Leu21 until the C-terminus. In each plot, the colored trends represent a different replicate, while the black trend computes the average of all replicates. All data shown was obtained from a smoothing procedure over the original data, using a moving window average (bin size of 1000) to reduce the data noise.

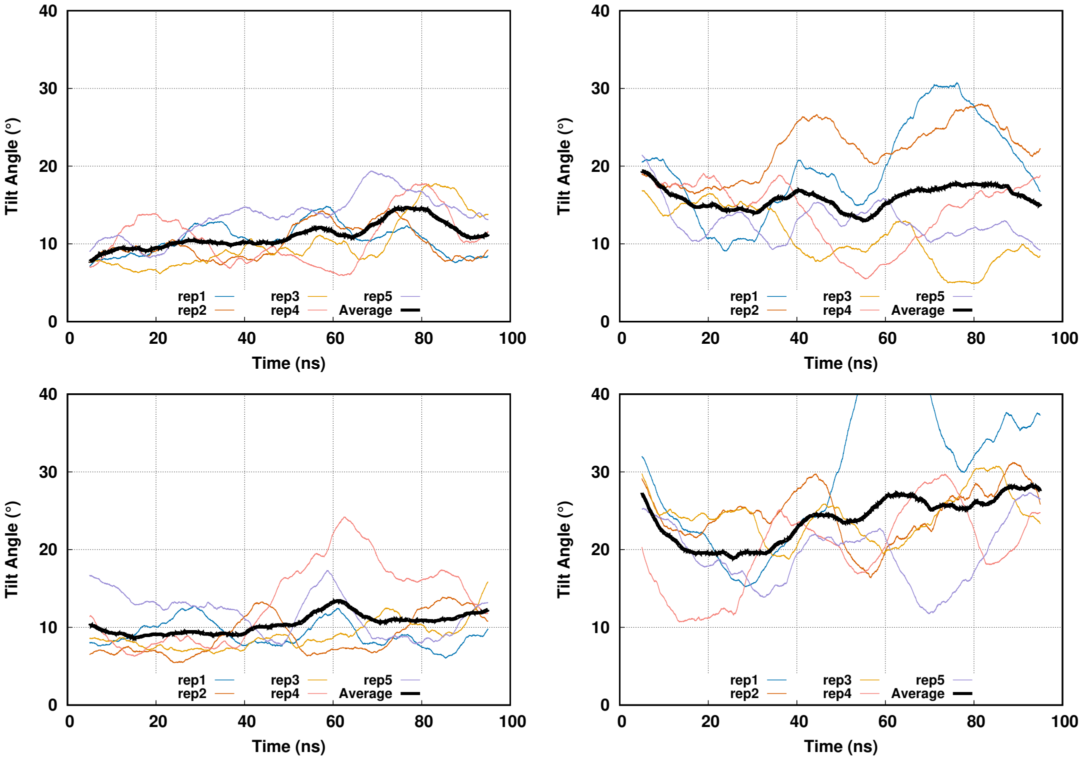


Figure C.13: Time series data of the peptide tilt angle (degrees °), between the H2 (22^{th} to 30^{th}) helical segment and the membrane normal, of each corresponding peptide variant at a representative pH value (**pH 7.25**). The helical segment was identified as the regions with higher helical content after the central Leu21 until the C-terminus. In each plot, the colored trends represent a different replicate, while the black trend computes the average of all replicates. All data shown was obtained from a smoothing procedure over the original data, using a moving window average (bin size of 1000) to reduce the data noise.

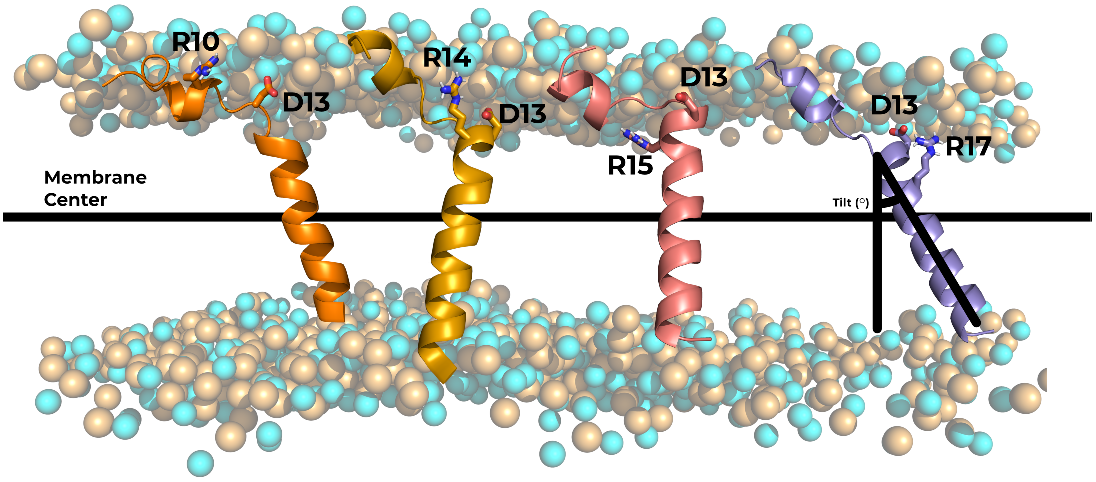


Figure C.14: Graphical representation of representative peptide conformations, obtained from the final of the pHRE simulation, in the POPC membrane bilayer. The secondary structure is represented as a cartoon, while the key residues are shown in sticks. The central black line represents the membrane center and the depicted angle measures the displacement of the 2^{nd} helix (22^{nd} to 30^{th} residue) relative to the membrane normal.

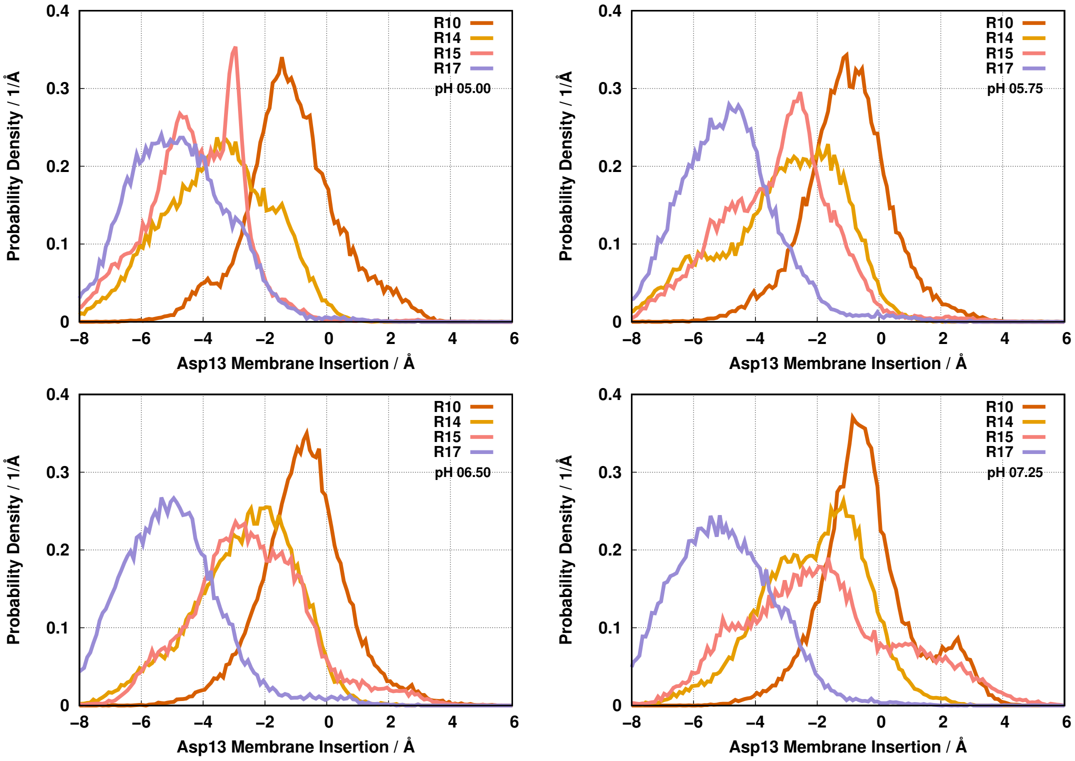
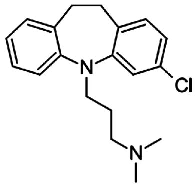


Figure C.15: Probability density function of the Asp13 populating a given insertion level, at all studied pH values. Each colored trend represents a different arginine peptide variant. Negative insertion values correspond to the membrane interior.

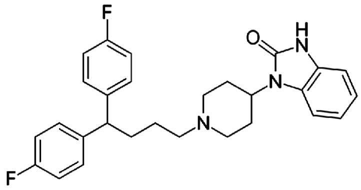
Appendix D

Supporting Information for Chapter 6

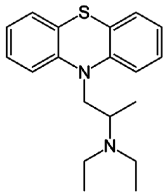
Clomipramine
log P 5.19
pKa 9.5



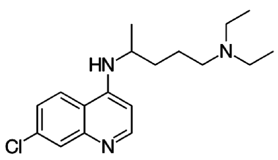
Pimozide
log P 6.3
pKa 8.63



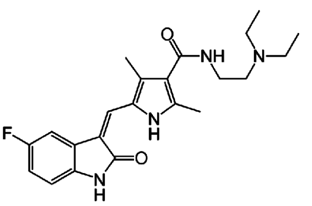
Ethopropazine
log P 4.77
pKa 9.64



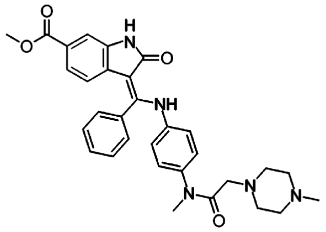
Chloroquine
log P 4.77
pKa 10.1



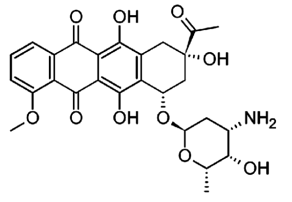
Sunitinib
log P 5.2
pKa 8.5



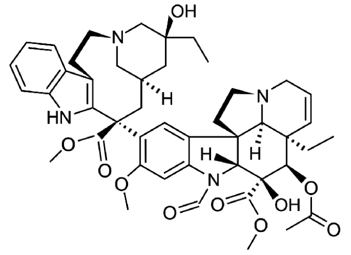
Nintedanib
log P 3
pKa 7.9



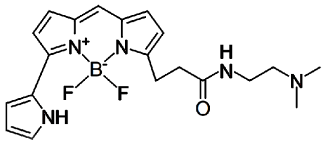
Daunorubicin
log P 1.83
pKa 8.46



Vincristine
log P 2.82
pKa 8.31



LysoTracker RED DND-99
log P 2.1
pKa 7.5



LysoTracker GREEN DND-26
log P U.P.
pKa U.P.

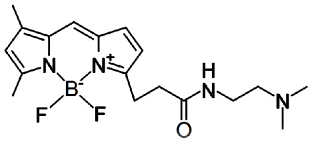


Figure D.1: Structure and physicochemical properties of various compounds employed in the current study. Experimental log P values were retrieved from the following publications; Pimo, Clomp, Ethop and CHQ [58]; SUN [53]; NTD [45]; DNR [59]; VCR [60]; and LTR [56]. Experimental pK_a values were retrieved from the following publications; Clomp [49]; Pimo [50]; Ethop [51]; CHQ [52]; SUN [53]; NTD [45]; DNR [54]; VCR [55]; and LTR [56]. Values for LTG were never reported (unpublished U.P.).

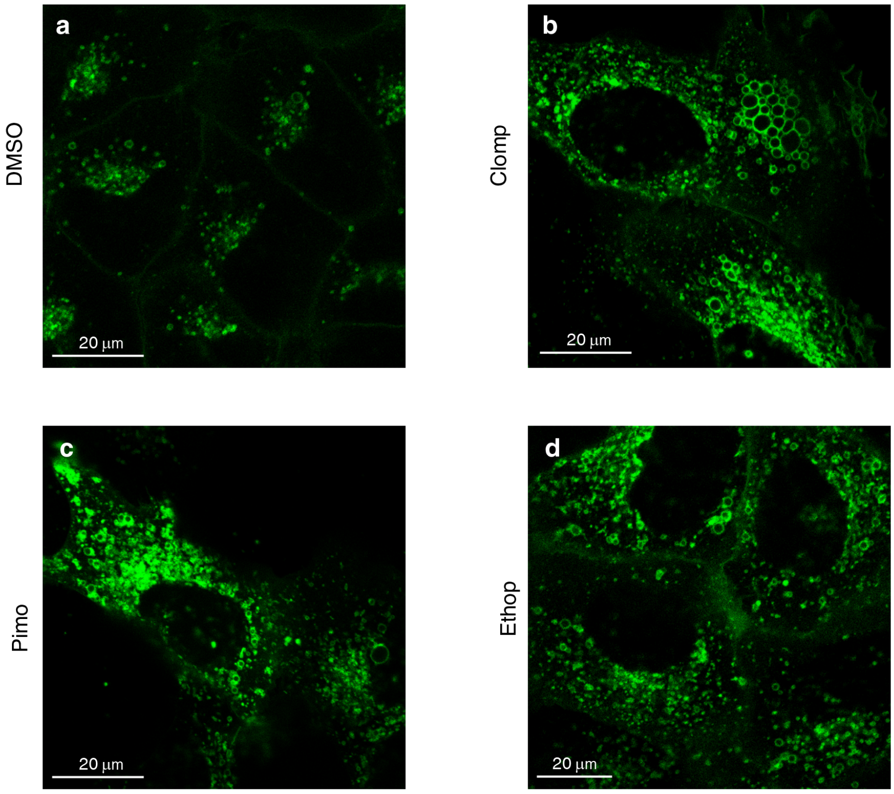


Figure D.2: Increased size and number of LAMP1-mGFP labeled lysosomes following treatment with CNSDs. U2OS cells stably expressing LAMP1-mGFP were treated for 72 h with either 10 μM Clomp, 10 μM Ethop, 30 μM Pimo or 0.1 % DMSO as a drug-free control. Fluorescence was recorded using a confocal Zeiss LSM 710 microscope (×63 magnification).

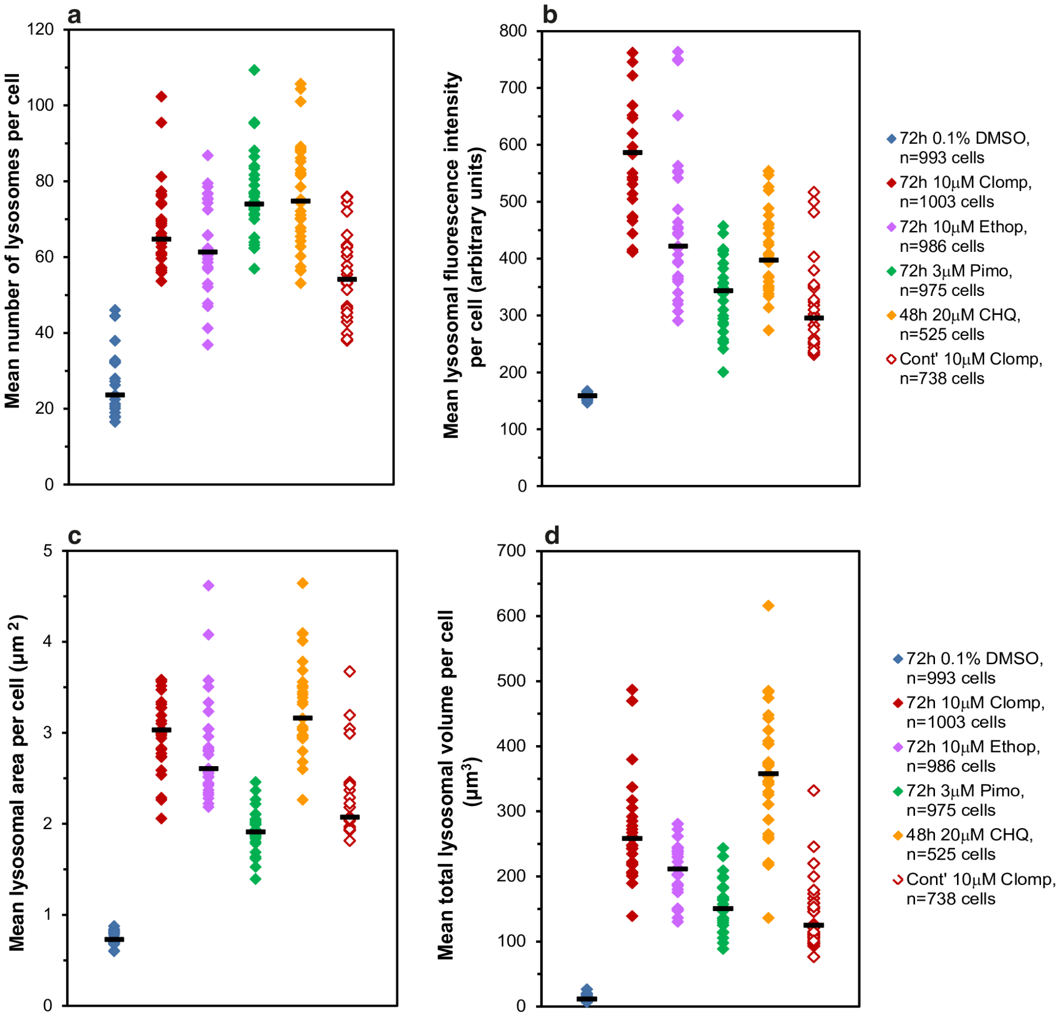


Figure D.3: Representative quantification of the increase in lysosomal parameters following CNSDs treatment. U2OS cells were seeded in black glass bottom plates and treated for 72 h with either 10 μM Clomp, 10 μM Ethop, 30 μM Pimo or 0.1 % DMSO. In addition, ClompR cells continuously grown in 10 μM Clomp were used, as well as U2OS cells treated for 48 h with 20 μM CHQ as a positive lysosomotropic control. Nuclei were stained with the DNA dye Hoechst 33342 and lysosomes with the lysosomal probe LTR. Fluorescence was captured using an InCell Analyzer 2000 microscope, and lysosomes were analyzed using the InCell Investigator software. Each diamond represents the mean result of a microscope field, containing ~10-50 cells, where 32 independent fields were captured for each treatment from two 24-wells. The black bars denote the median results for each parameter. All p values within an experiment were < 5x10-12.

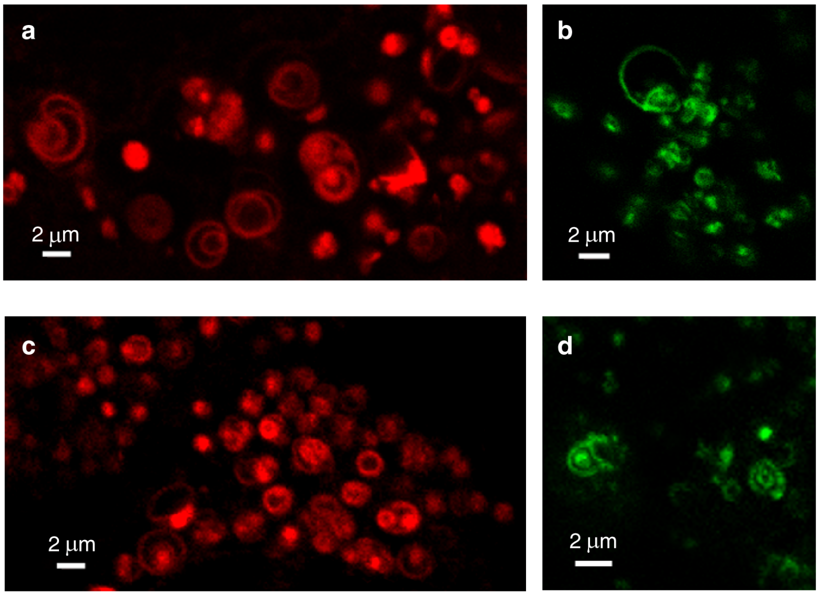


Figure D.4: ILV contain marked levels of the anticancer drugs DNR and NTD. Vacuolin-1-treated U2OS cells (a-b) and ClompR cells (c-d) were loaded for 45 min with DNR (a, c) or NTD (b, d) and captured with a confocal Zeiss LSM 710 microscope (×63 magnification).

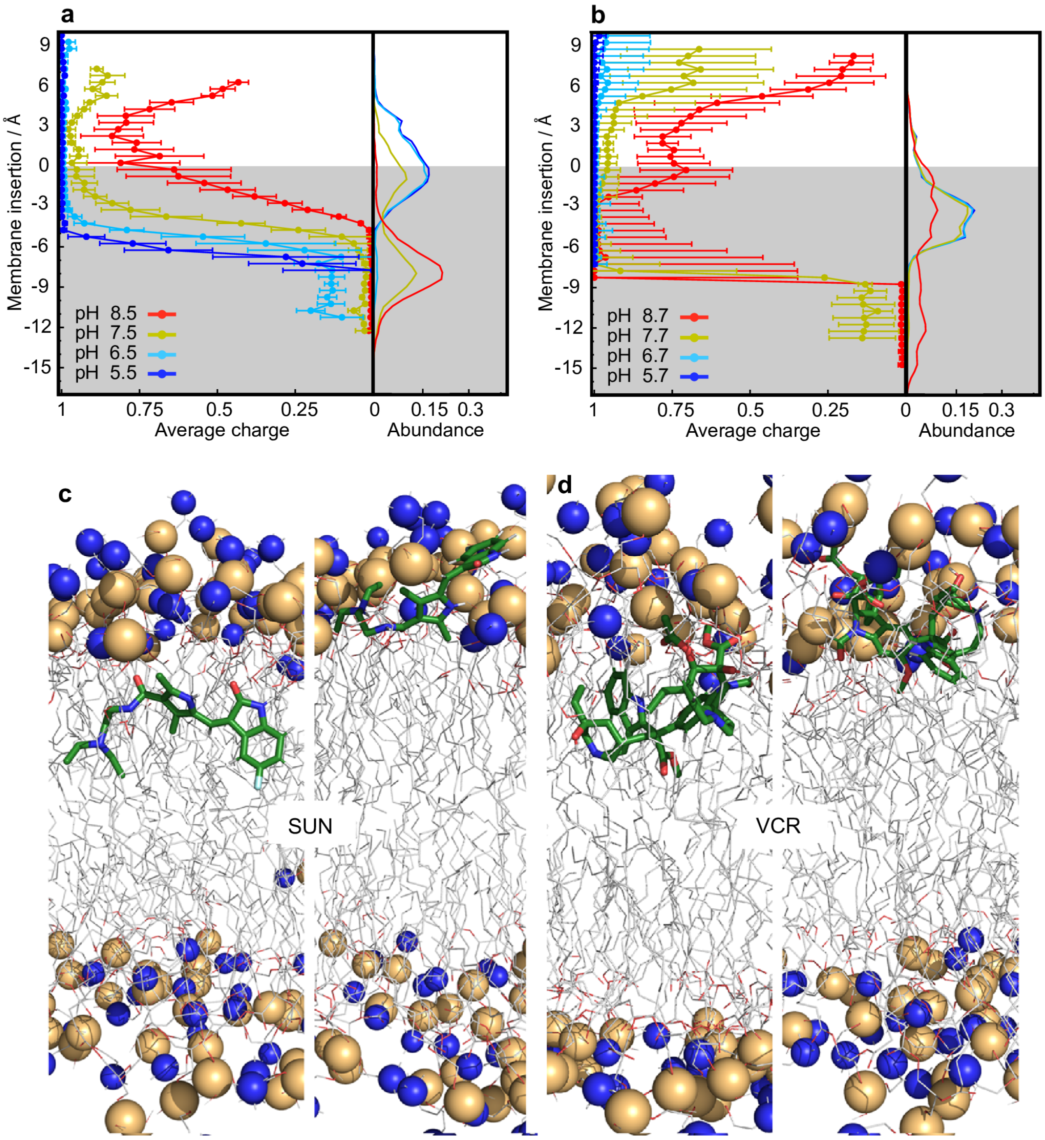


Figure D.5: CpHMD simulations of membrane insertion of lysosomotropic anticancer drugs. a-b, Average protonation (left subplots) and abundance histograms (right subplots) are depicted along the membrane insertion axis for Sun (a) and VCR (b). The grey-shaded area represents the membrane internal region below the phosphate groups, which are used as insertion references (see Materials and Methods). c-d, Selected conformations of SUN and VCR inserted into a DMPC membrane model are shown in two panels, with a more inserted location on the left panels and a shallower location on the right panels. The LDs are shown as sticks (green carbon atoms) and the phospholipids are shown in thin grey lines. The phosphorous and nitrogen atoms of DMPC are shown as orange and blue spheres, respectively, to highlight the polar region in the water/membrane interface.

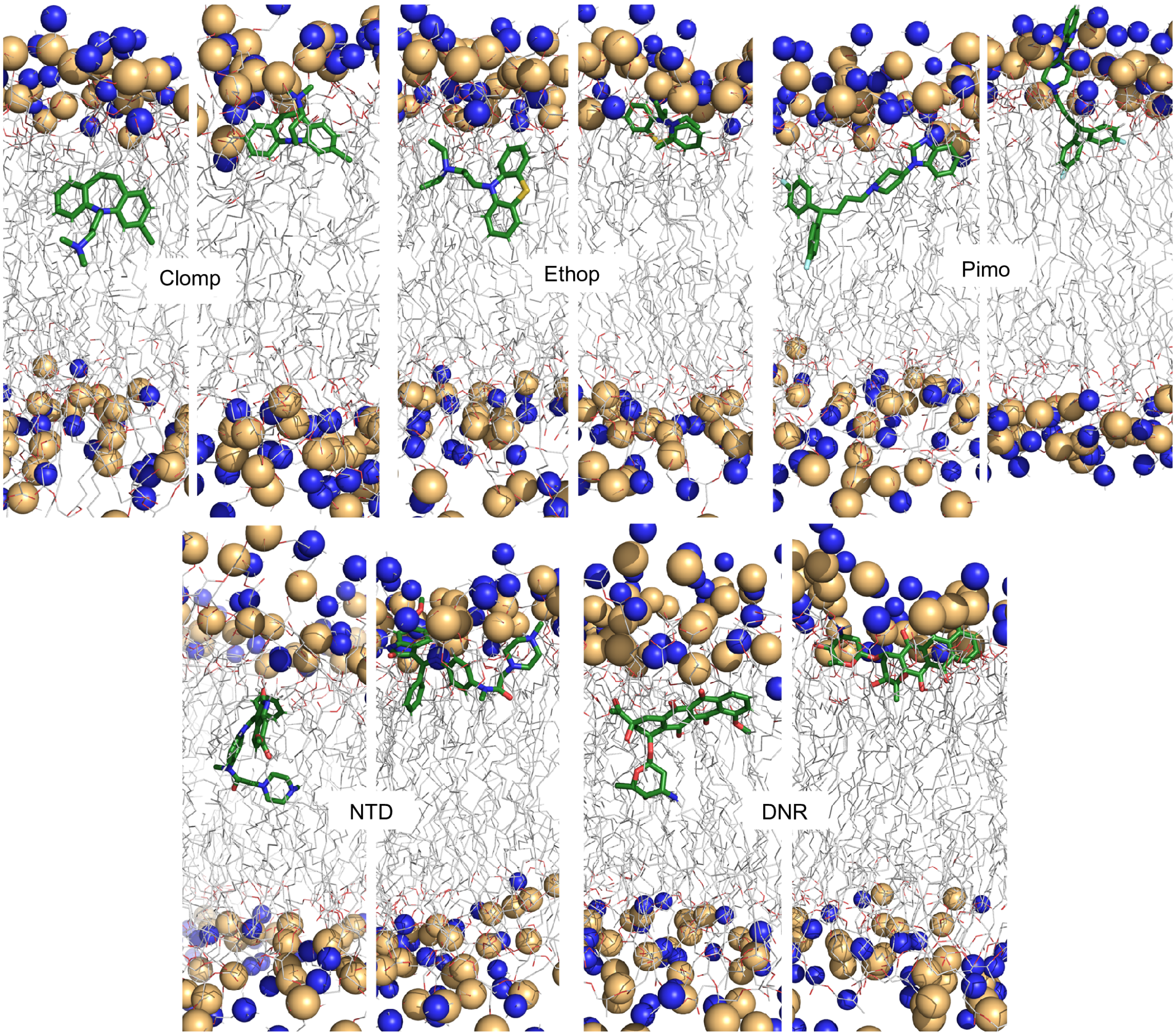


Figure D.6: Selected conformations of LDs inserted into a DMPC membrane model. Clomp, Ethop, Pimo, NTD and DNR are shown in two panels, with a more inserted location on the left panels and a shallower location on the right panels. The LDs are shown as sticks (green carbon atoms) and the phospholipids are shown in thin grey lines. The phosphorous and nitrogen atoms of DMPC are shown as orange and blue spheres, respectively, to highlight the polar region in the water/membrane interface.

References

[1] J. Ferlay et al. "Cancer statistics for the year 2020: An overview". In: *Int. J. Cancer* 149.4 (2021), pp. 778–789.

[2] D. Hanahan and R. A. Weinberg. "The hallmarks of cancer". In: *Cell* 100.1 (2000), pp. 57–70.

[3] D. Hanahan and R. A. Weinberg. "Hallmarks of cancer: the next generation". In: *Cell* 144.5 (2011), pp. 646–674.

[4] M. R. Junttila and F. J. De Sauvage. "Influence of tumour micro-environment heterogeneity on therapeutic response". In: *Nature* 501.7467 (2013), pp. 346–354.

[5] S. Brassart-Pasco et al. "Tumor microenvironment: extracellular matrix alterations influence tumor progression". In: *Front. Oncol.* 10 (2020), p. 397.

[6] F. Chen et al. "New horizons in tumor microenvironment biology: challenges and opportunities". In: *BMC Med.* 13.1 (2015), pp. 1–14.

[7] M. Wang et al. "Role of tumor microenvironment in tumorigenesis". In: *J. Cancer* (2017).

[8] P. Carmeliet and R. K. Jain. "Angiogenesis in cancer and other diseases". In: *Nature* 407.6801 (2000), pp. 249–257.

[9] O. Trédan et al. "Drug Resistance and the Solid Tumor Microenvironment". In: *J. Natl. Cancer Inst.* 99.19 (2007), pp. 1441–1454.

[10] O. Warburg, K. Posener, and E. Negelein. "The metabolism of cancer cells". In: *Biochem. Z* 152 (1924), pp. 319–44.

[11] O. Warburg. "On the origin of cancer cells". In: *Science* 123.3191 (1956), pp. 309–314.

[12] M. G. V. Heiden, L. C. Cantley, and C. B. Thompson. "Understanding the Warburg Effect: The Metabolic Requirements of Cell Proliferation". In: *Science* 324.5930 (2009), pp. 1029–1033.

[13] "The Warburg Effect: How Does it Benefit Cancer Cells?" In: *Trends Biochem. Sci.* 41.3 (2016). Special Issue: Mitochondria & Metabolism, pp. 211–218. ISSN: 0968-0004.

[14] C. Corbet et al. "Acidosis drives the reprogramming of fatty acid metabolism in cancer cells through changes in mitochondrial and histone acetylation". In: *Cell Metab.* 24.2 (2016), pp. 311–323.

[15] C. Corbet et al. “TGFβ2-induced formation of lipid droplets supports acidosis-driven EMT and the metastatic spreading of cancer cells”. In: *Nat. Commun.* 11.1 (2020), pp. 1–15.

[16] M. Anderson et al. “Probe for the measurement of cell surface pH in vivo and ex vivo”. In: *Proc. Natl. Acad. Sci. USA* 113.29 (2016), pp. 8177–8181.

[17] M. Logozzi et al. “Extracellular acidity and increased exosome release as key phenotypes of malignant tumors”. In: *Cancer Metastasis Rev.* 38.1 (2019), pp. 93–101.

[18] K. Kessenbrock, V. Plaks, and Z. Werb. “Matrix metalloproteinases: regulators of the tumor microenvironment”. In: *Cell* 141.1 (2010), pp. 52–67.

[19] E. Persi et al. “Systems analysis of intracellular pH vulnerabilities for cancer therapy”. In: *Nat. Commun.* 9.1 (2018), pp. 1–11.

[20] K. A. White, K. Kisor, and D. L. Barber. “Intracellular pH dynamics and charge-changing somatic mutations in cancer”. In: *Cancer Metastasis Rev.* 38.1 (2019), pp. 17–24.

[21] P. Vaupel and A. Mayer. “Tumor oxygenation status: facts and fallacies”. In: *Oxygen Transp.Tissue XXXIX* (2017), pp. 91–99.

[22] Y. G. Assaraf et al. “The multi-factorial nature of clinical multidrug resistance in cancer”. In: *Drug Resist. Updat.* 46 (2019), p. 100645.

[23] W. Aoi and Y. Marunaka. “Importance of pH homeostasis in metabolic health and diseases: crucial role of membrane proton transport”. In: *BioMed Res. Int.* 2014 (2014).

[24] R. W. Putnam. *Chapter 17 - Intracellular pH Regulation*. Ed. by N. Sperelakis. Fourth Edition. San Diego: Academic Press, 2012, pp. 303–321. ISBN: 978-0-12-387738-3.

[25] P. R. Magalhaes. “Inclusion of protonation effects in the simulation of proteins in membrane environments”. PhD thesis. Lisboa: Instituto de Tecnologia Química e Biológica, Universidade Nova de Lisboa, 2017.

[26] A. Warshel and S. T. Russell. “Calculations of electrostatic interactions in biological systems and in solutions”. In: *Q. Rev. Biophys.* 17.03 (1984), pp. 283–422.

[27] V. H. Teixeira et al. “pK_a Values of Titrable Amino Acids at the Water/Membrane Interface”. In: *J. Chem. Theory Comput.* 12.3 (2016), pp. 930–934.

[28] N. Gonen and Y. G. Assaraf. “Antifolates in cancer therapy: structure, activity and mechanisms of drug resistance”. In: *Drug Resist. Updat.* 15.4 (2012), pp. 183–210.

[29] C.-Y. Huang et al. “A review on the effects of current chemotherapy drugs and natural agents in treating non–small cell lung cancer”. In: *Biomedicine* 7.4 (2017).

[30] S. E. Denmark and G. L. Beutner. “Lewis base catalysis in organic synthesis”. In: *Angew. Chem. Int. Ed.* 47.9 (2008), pp. 1560–1638.

[31] E. P. Rock et al. “Food and Drug Administration drug approval summary: Sunitinib malate for the treatment of gastrointestinal stromal tumor and advanced renal cell carcinoma”. In: *The Oncologist* 12.1 (2007), pp. 107–113.

[32] R. J. Motzer et al. “Sunitinib: ten years of successful clinical use and study in advanced renal cell carcinoma”. In: *The Oncologist* 22.1 (2017), pp. 41–52.

[33] M. Rausch et al. “Molecular and functional analysis of sunitinib-resistance induction in human renal cell carcinoma cells”. In: *Int. J. Mol. Sci.* 22.12 (2021), p. 6467.

[34] K. J. Gotink et al. “Lysosomal Sequestration of Sunitinib: A Novel Mechanism of Drug Resistance Lysosomal Sequestration Is Involved in Sunitinib Resistance”. In: *Clin. Cancer Res.* 17.23 (2011), pp. 7337–7346.

[35] P. Nowak-Sliwinska et al. “Photoactivation of lysosomally sequestered sunitinib after angiostatic treatment causes vascular occlusion and enhances tumor growth inhibition”. In: *Cell Death Dis.* 6.2 (2015), e1641–e1641.

[36] B. Englinger et al. “Intrinsic fluorescence of the clinically approved multikinase inhibitor nintedanib reveals lysosomal sequestration as resistance mechanism in FGFR-driven lung cancer”. In: *J. Exp. Clin. Cancer Res.* 36.1 (2017), pp. 1–13.

[37] M. Stark et al. “The Lysosomotropic Activity of Hydrophobic Weak Base Drugs is Mediated via Their Intercalation into the Lysosomal Membrane”. In: *Cells* 9.5 (2020), p. 1082.

[38] M. M. Gottesman, T. Fojo, and S. E. Bates. “Multidrug resistance in cancer: role of ATP–dependent transporters”. In: *Nature Rev. Cancer* 2.1 (2002), pp. 48–58.

[39] B. Zhitomirsky and Y. G. Assaraf. “Lysosomes as mediators of drug resistance in cancer”. In: *Drug Resist. Updat.* 24 (2016), pp. 23–33.

[40] S. Dallavalle et al. “Improvement of conventional anti-cancer drugs as new tools against multidrug resistant tumors”. In: *Drug Resist. Updat.* 50 (2020), p. 100682.

[41] L. J. Goldstein et al. “Expression of multidrug resistance gene in human cancers”. In: *J. Natl. Cancer Inst.* 81.2 (1989), pp. 116–124.

[42] I. Ifergan, G. L. Scheffer, and Y. G. Assaraf. “Novel extracellular vesicles mediate an ABCG2-dependent anticancer drug sequestration and resistance”. In: *Cancer Res.* 65.23 (2005), pp. 10952–10958.

[43] W. Li et al. “Overcoming ABC transporter-mediated multidrug resistance: Molecular mechanisms and novel therapeutic drug strategies”. In: *Drug Resist. Updat.* 27 (2016), pp. 14–29.

[44] A. Palmeira et al. “Three decades of P-gp inhibitors: skimming through several generations and scaffolds”. In: *Curr. Med. Chem.* 19.13 (2012), pp. 1946–2025.

[45] V. Goler-Baron, I. Sladkevich, and Y. G. Assaraf. “Inhibition of the PI3K-Akt signaling pathway disrupts ABCG2-rich extracellular vesicles and overcomes multidrug resistance in breast cancer cells”. In: *Biochem. Pharmacol.* 83.10 (2012), pp. 1340–1348.

[46] B. Zhitomirsky and Y. G. Assaraf. “Lysosomal sequestration of hydrophobic weak base chemotherapeutics triggers lysosomal biogenesis and lysosome-dependent cancer multidrug resistance”. In: *Oncotarget* 6.2 (2015), p. 1143.

[47] D. Sousa, R. T. Lima, and M. H. Vasconcelos. “Intercellular transfer of cancer drug resistance traits by extracellular vesicles”. In: *Trends Mol. Med.* 21.10 (2015), pp. 595–608.

[48] N. Mc Namee and L. O’Driscoll. “Extracellular vesicles and anti-cancer drug resistance”. In: *Biochem. Biophys. Acta, Rev. Cancer* 1870.2 (2018), pp. 123–136.

[49] B. Zhitomirsky and Y. G. Assaraf. “Lysosomal accumulation of anticancer drugs triggers lysosomal exocytosis”. In: *Oncotarget* 8.28 (2017), p. 45117.

[50] B. Zhitomirsky et al. “Lysosomotropic drugs activate TFEB via lysosomal membrane fluidization and consequent inhibition of mTORC1 activity”. In: *Cell Death Dis.* 9.12 (2018), pp. 1–15.

[51] A. Sharma et al. “Hypoxia-targeted drug delivery”. In: *Chem. Soc. Rev.* 48.3 (2019), pp. 771–813.

[52] O. A. Andreev et al. “Mechanism and uses of a membrane peptide that targets tumors and other acidic tissues in vivo”. In: *Proc. Natl. Acad. Sci. USA* 104.19 (2007), pp. 7893–7898.

[53] R. Rajagopalan and J. V. Yakhmi. “Nanotechnological approaches toward cancer chemotherapy”. In: *Nanostructures for Cancer Therapy*. Elsevier, 2017, pp. 211–240.

[54] A. Swami et al. “Nanoparticles for targeted and temporally controlled drug delivery”. In: *Multifunctional nanoparticles for drug delivery applications*. Springer, 2012, pp. 9–29.

[55] C. Sun, J. S. Lee, and M. Zhang. “Magnetic nanoparticles in MR imaging and drug delivery”. In: *Adv. Drug Deliv. Rev.* 60.11 (2008), pp. 1252–1265.

[56] S. Macholl et al. “In vivo pH imaging with 99mTc-pHLIP”. In: *Mol. Imaging Biol.* 14.6 (2012), pp. 725–734.

[57] A. Beloqui et al. “pH-sensitive nanoparticles for colonic delivery of curcumin in inflammatory bowel disease”. In: *Int. J. Pharm.* 473.1-2 (2014), pp. 203–212.

[58] Y.-S. Zhu, K. Tang, and J. Lv. “Peptide–drug conjugate-based novel molecular drug delivery system in cancer”. In: *Trends Pharmacol. Sci.* 42.10 (2021), pp. 857–869.

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

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

[61] J. Katsaras, N. Kucerka, and M. Nieh. “Fluid phase lipid areas and bilayer thicknesses of commonly used phosphatidylcholines as a function of temperature”. In: *Biochem. Biophys. Acta* 1808 (2011), pp. 2761–2771.

[62] E. Wallin and G. V. Heijne. “Genome-wide analysis of integral membrane proteins from eubacterial, archaean, and eukaryotic organisms”. In: *Protein Sci.* 7.4 (1998), pp. 1029–1038.

[63] J. P. Overington, B. Al-Lazikani, and A. L. Hopkins. “How many drug targets are there?” In: *Nat. Rev. Drug Discov.* 5.12 (2006), pp. 993–996.

[64] T. J. Stevens and I. T. Arkin. “Do more complex organisms have a greater proportion of membrane proteins in their genomes?” In: *Proteins Struct. Funct. Bioinf.* 39.4 (2000), pp. 417–420.

[65] J. A. Killian. “Hydrophobic mismatch between proteins and lipids in membranes”. In: *Biochem. Biophys. Acta, Rev. Biomembr.* 1376.3 (1998), pp. 401–416.

[66] M. R. de Planque and J. A. Killian*. “Protein–lipid interactions studied with designed transmembrane peptides: role of hydrophobic matching and interfacial anchoring”. In: *Mol. Membr. Biol.* 20.4 (2003), pp. 271–284.

[67] F. Dumas, M. C. Lebrun, and J.-F. Tocanne. “Is the protein/lipid hydrophobic matching principle relevant to membrane organization and functions?” In: *FEBS Lett.* 458.3 (1999), pp. 271–277.

[68] K. R. Kampen. “Membrane proteins: the key players of a cancer cell”. In: *J. Membrane Biol.* 242.2 (2011), pp. 69–74.

[69] C. J. Mitchell, T. S. Johnson, and C. M. Deber. “Transmembrane peptide effects on bacterial membrane integrity and organization”. In: *Biophys. J.* (2022).

[70] K. R. MacKenzie, J. H. Prestegard, and D. M. Engelman. “A transmembrane helix dimer: structure and implications”. In: *Science* 276.5309 (1997), pp. 131–133.

[71] M. T. Duong et al. “Changes in apparent free energy of helix–helix dimerization in a biological membrane due to point mutations”. In: *J. Mol. Biol.* 371.2 (2007), pp. 422–434.

[72] M. A. Lemmon et al. “Sequence specificity in the dimerization of transmembrane. alpha.-helixes”. In: *Biochemistry-US* 31.51 (1992), pp. 12719–12725.

[73] F. X. Zhou et al. “Polar residues drive association of polyleucine transmembrane helices”. In: *Proc. Natl. Acad. Sci. USA* 98.5 (2001), pp. 2250–2255.

[74] C. Choma et al. “Asparagine-mediated self-association of a model transmembrane helix”. In: *Nat. Struct. Biol.* 7.2 (2000), pp. 161–166.

[75] J. P. Dawson, J. S. Weinger, and D. M. Engelman. “Motifs of serine and threonine can drive association of transmembrane helices”. In: *J. Mol. Biol.* 316.3 (2002), pp. 799–805.

[76] N. Sal-Man, D. Gerber, and Y. Shai. “The identification of a minimal dimerization motif QXXS that enables homo-and hetero-association of transmembrane helices in vivo”. In: *J. Biol. Chem.* 280.29 (2005), pp. 27449–27457.

[77] N. A. Noordeen et al. “A transmembrane leucine zipper is required for activation of the dimeric receptor tyrosine kinase DDR1”. In: *J. Biol. Chem.* 281.32 (2006), pp. 22744–22751.

[78] J. X. Qiu et al. “Toward β-amino acid proteins: a cooperatively folded β-peptide quaternary structure”. In: *J. Am. Chem. Soc.* 128.35 (2006), pp. 11338–11339.

[79] S. H. White et al. “How membranes shape protein structure”. In: *J. Biol. Chem.* 276.35 (2001), pp. 32395–32398.

[80] M. Orzáez et al. “Influence of hydrophobic matching on association of model transmembrane fragments containing a minimised glycophorin A dimerisation motif”. In: *FEBS Lett.* 579.7 (2005), pp. 1633–1638.

[81] N. Bordag and S. Keller. “α-Helical transmembrane peptides: a “divide and conquer” approach to membrane proteins”. In: *Chem. Phys. Lipids* 163.1 (2010), pp. 1–26.

[82] R. A. Melnyk et al. “Polar residue tagging of transmembrane peptides”. In: *Pept. Sci.* 71.6 (2003), pp. 675–685.

[83] M. R. de Planque et al. “The effects of hydrophobic mismatch between phosphatidylcholine bilayers and transmembrane α-helical peptides depend on the nature of interfacially exposed aromatic and charged residues”. In: *Biochemistry-US* 41.26 (2002), pp. 8396–8404.

[84] S. K. Kandasamy and R. G. Larson. “Molecular dynamics simulations of model trans-membrane peptides in lipid bilayers: a systematic investigation of hydrophobic mismatch”. In: *Biophys. J.* 90.7 (2006), pp. 2326–2343.

[85] Y. Liu and D. Bolen. “The peptide backbone plays a dominant role in protein stabilization by naturally occurring osmolytes”. In: *Biochemistry-US* 34.39 (1995), pp. 12884–12891.

[86] J. A. Killian et al. “Induction of nonbilayer structures in diacylphosphatidylcholine model membranes by transmembrane α-helical peptides: importance of hydrophobic mismatch and proposed role of tryptophans”. In: *Biochemistry-US* 35.3 (1996), pp. 1037–1045.

[87] A. Ridder et al. “Tryptophan supports interaction of transmembrane helices”. In: *J. Mol. Biol.* 354.4 (2005), pp. 894–902.

[88] O. Mouritsen and M. Bloom. “Models of lipid-protein interactions in membranes”. In: *Annu. Rev. Biophys. Biomol. Struct.* 22.1 (1993), pp. 145–171.

[89] L.-P. Liu et al. “Threshold hydrophobicity dictates helical conformations of peptides in membrane environments”. In: *Biopolymers* 39.3 (1996), pp. 465–470.

[90] L.-P. Liu and C. M. Deber. “Anionic phospholipids modulate peptide insertion into membranes”. In: *Biochemistry-US* 36.18 (1997), pp. 5476–5482.

[91] W. Im and C. L. Brooks III. “Interfacial folding and membrane insertion of designed peptides studied by molecular dynamics simulations”. In: *Proc. Natl. Acad. Sci. USA* 102.19 (2005), pp. 6771–6776.

[92] Y. Yano et al. “Topological stability and self-association of a completely hydrophobic model transmembrane helix in lipid bilayers”. In: *Biochemistry-US* 41.9 (2002), pp. 3073–3080.

[93] U. Harzer and B. Bechinger. “Alignment of lysine-anchored membrane peptides under conditions of hydrophobic mismatch: a CD, 15N and 31P solid-state NMR spectroscopy investigation”. In: *Biochemistry-US* 39.43 (2000), pp. 13106–13114.

[94] E. Strandberg et al. “Lipid dependence of membrane anchoring properties and snorkeling behavior of aromatic and charged residues in transmembrane peptides”. In: *Biochemistry-US* 41.23 (2002), pp. 7190–7198.

[95] E. Strandberg and J. A. Killian. “Snorkeling of lysine side chains in transmembrane helices: how easy can it get?” In: *FEBS Lett.* 544.1-3 (2003), pp. 69–73.

[96] V. V. Vostrikov et al. “Changes in transmembrane helix alignment by arginine residues revealed by solid-state NMR experiments and coarse-grained MD simulations”. In: *J. Am. Chem. Soc.* 132.16 (2010), pp. 5803–5811.

[97] A. Walrant et al. “Membrane interactions of two arginine-rich peptides with different cell internalization capacities”. In: *Biochem. Biophys. Acta, Biomembr.* 1818.7 (2012), pp. 1755–1763.

[98] J. K. Thibado et al. “Influence of high pH and cholesterol on single arginine-containing transmembrane peptide helices”. In: *Biochemistry-US* 55.45 (2016), pp. 6337–6343.

[99] T. M. Weiss et al. “Hydrophobic mismatch between helices and lipid bilayers”. In: *Biophys. J.* 84.1 (2003), pp. 379–385.

[100] M. B. Ulmschneider and J. P. Ulmschneider. “Folding peptides into lipid bilayer membranes”. In: *J. Chem. Theory Comput.* 4.11 (2008), pp. 1807–1809.

[101] M. Musial-Siwek et al. “Tuning the insertion properties of pHLIP”. In: *Biochem. Biophys. Acta, Biomembr.* 1798.6 (2010), pp. 1041–1046.

[102] F. N. Barrera et al. “Roles of carboxyl groups in the transmembrane insertion of peptides”. In: *J. Mol. Biol.* 413.2 (2011), pp. 359–371.

[103] D. Vila-Viçosa et al. “Membrane-Induced pK_a Shifts in wt-pHLIP and Its L16H Variant”. In: *J. Chem. Theory Comput.* 14.6 (2018), pp. 3289–3297.

[104] T. F. Silva, D. Vila-Viçosa, and M. Machuqueiro. “Improved Protocol to Tackle the pH Effects on Membrane-Inserting Peptides”. In: *J. Chem. Theory Comput.* 17.7 (2021), pp. 3830–3840.

[105] N. K. Subbarao et al. “The pH-dependent bilayer destabilization by an amphipathic peptide”. In: *Biochemistry-US* 26.11 (1987), pp. 2964–2972.

[106] E. Goormaghtigh et al. “Secondary structure and orientation of the amphipathic peptide GALA in lipid structures: An infrared-spectroscopic approach”. In: *Eur. J. Biochem.* 195.2 (1991), pp. 421–429.

[107] D. Haas and R. Murphy. “Design of a pH-sensitive pore-forming peptide with improved performance”. In: *J. Pept. Res.* 63.1 (2004), pp. 9–16.

[108] R. A. Parente et al. “Association of a pH-sensitive peptide with membrane vesicles: role of amino acid sequence”. In: *Biochemistry-US* 29.37 (1990), pp. 8713–8719.

[109] T. B. Wyman et al. “Design, synthesis, and characterization of a cationic peptide that binds to nucleic acids and permeabilizes bilayers”. In: *Biochemistry-US* 36.10 (1997), pp. 3008–3017.

[110] W. Li, F. Nicol, and F. C. Szoka Jr. “GALA: a designed synthetic pH-responsive amphipathic peptide with applications in drug and gene delivery”. In: *Adv. Drug Deliv. Rev.* 56.7 (2004), pp. 967–985.

[111] F. Nicol, S. Nir, and F. C. Szoka Jr. “Effect of phospholipid composition on an amphipathic peptide-mediated pore formation in bilayer vesicles”. In: *Biophys. J.* 78.2 (2000), pp. 818–829.

[112] W. C. Wimley and S. H. White. “Designing transmembrane α-helices that insert spontaneously”. In: *Biochemistry-US* 39.15 (2000), pp. 4432–4442.

[113] C. H. Chen et al. “Tuning of a membrane-perforating antimicrobial peptide to selectively target membranes of different lipid composition”. In: *J. Membr. Biol.* 254.1 (2021), pp. 75–96.

[114] S. D. Robinson et al. “Venom peptides as therapeutics: advances, challenges and the future of venom-peptide discovery”. In: *Expert Rev. Proteomics* 14.10 (2017), pp. 931–939.

[115] R. M. Epand. “Fusion peptides and the mechanism of viral fusion”. In: *Biochem. Biophys. Acta, Biomembr.* 1614.1 (2003), pp. 116–121.

[116] B. Apellániz et al. “The three lives of viral fusion peptides”. In: *Chem. Phys. Lipids* 181 (2014), pp. 40–55.

[117] Y. K. Reshetnyak et al. “A monomeric membrane peptide that lives in three worlds: in solution, attached to, and inserted across lipid bilayers”. In: *Biophys. J.* 93.7 (2007), pp. 2363–2372.

[118] M. E. Herbig, K. M. Weller, and H. P. Merkle. “Reviewing biophysical and cell biological methodologies in cell-penetrating peptide (CPP) research”. In: *Crit. Rev. Ther. Drug Carrier. Syst.* 24.3 (2007).

[119] S. Dissanayake et al. “Recent developments in anticancer drug delivery using cell penetrating and tumor targeting peptides”. In: *J. Control. Release* 250 (2017), pp. 62–76.

[120] Y. K. Reshetnyak et al. “Energetics of peptide (pHLIP) binding to and folding across a lipid bilayer membrane”. In: *Proc. Natl. Acad. Sci. USA* 105.40 (2008), pp. 15340–15345.

[121] J. F. Hunt et al. “Spontaneous, pH-dependent membrane insertion of a transbilayer α-helix”. In: *Biochemistry-US* 36.49 (1997), pp. 15177–15192.

[122] Y. K. Reshetnyak et al. “Measuring tumor aggressiveness and targeting metastatic lesions with fluorescent pHLIP”. In: *Mol. Imaging Biol.* 13.6 (2011), pp. 1146–1156.

[123] D. Weerakkody et al. “Family of pH (low) insertion peptides for tumor targeting”. In: *Proc. Natl. Acad. Sci. USA* 110.15 (2013), pp. 5834–5839.

[124] R.-C. Adochite et al. “Targeting breast tumors with pH (low) insertion peptides”. In: *Mol. Pharm.* 11.8 (2014), pp. 2896–2905.

[125] L. C. Wyatt et al. “Peptides of pHLIP family for targeted intracellular and extracellular delivery of cargo molecules to tumors”. In: *Proc. Natl. Acad. Sci. USA* 115.12 (2018), E2811–E2818.

[126] A. Moshnikova et al. “Antiproliferative effect of pHLIP-amanitin”. In: *Biochemistry-US* 52.7 (2013), pp. 1171–1178.

[127] Q. Song et al. “A smart tumor targeting peptide–drug conjugate, pHLIP-SS-DOX: synthesis and cellular uptake on MCF-7 and MCF-7/Adr cells”. In: *Drug Deliv.* 23.5 (2016), pp. 1734–1746.

[128] Y. K. Reshetnyak et al. “Translocation of molecules into cells by pH-dependent insertion of a transmembrane helix”. In: *Proc. Natl. Acad. Sci. USA* 103.17 (2006), pp. 6460–6465.

[129] M. Sahraei et al. “Suppressing miR-21 activity in tumor-associated macrophages promotes an antitumor immune response”. In: *J. Clin. Investig.* 129.12 (2019), pp. 5518–5536.

[130] A. A. Svoronos et al. “Tumor-targeted, cytoplasmic delivery of large, polar molecules using a pH-low insertion peptide”. In: *Mol. Pharm.* 17.2 (2019), pp. 461–471.

[131] L. Han et al. “pH-controlled delivery of nanoparticles into tumor cells”. In: *Adv. Healthc. Mater.* 2.11 (2013), pp. 1435–1439.

[132] M. Deskeuvre et al. “Targeting cancer cells in acidosis with conjugates between the carnitine palmitoyltransferase 1 inhibitor etomoxir and pH (low) Insertion Peptides”. In: *Int. J. Pharm.* (2022), p. 122041.

[133] K. E. Burns, T. P. McCleerey, and D. Thévenin. “pH-selective cytotoxicity of pHLIP-antimicrobial peptide conjugates”. In: *Sci. Rep.* 6.1 (2016), pp. 1–10.

[134] F. N. Barrera, J. Fendos, and D. M. Engelman. “Membrane physical properties influence transmembrane helix formation”. In: *Proc. Natl. Acad. Sci. USA* 109.36 (2012), pp. 14422–14427.

[135] V. Vasquez-Montes et al. “Comparison of lipid-dependent bilayer insertion of pHLIP and its P20G variant”. In: *Biochem. Biophys. Acta, Biomembr.* 1860.2 (2018), pp. 534–543.

[136] G. Slaybaugh et al. “Kinetics of pHLIP peptide insertion into and exit from a membrane”. In: *Proc. Natl. Acad. Sci. USA* 117.22 (2020), pp. 12095–12100.

[137] S. A. Otieno et al. “pH-dependent thermodynamic intermediates of pHLIP membrane insertion determined by solid-state NMR spectroscopy”. In: *Proc. Natl. Acad. Sci. USA* 115.48 (2018), pp. 12194–12199.

[138] A. G. Karabadzhak et al. “Bilayer thickness and curvature influence binding and insertion of a pHLIP peptide”. In: *Biophys. J.* 114.9 (2018), pp. 2107–2115.

[139] A. Kyrychenko et al. “Lipid Headgroups Modulate Membrane Insertion of pHLIP Peptide”. In: *Biophys. J.* 108.4 (2015), pp. 791–794.

[140] V. Burns and B. Mertz. “Using Simulation to Understand the Role of Titration on the Stability of a Peptide–Lipid Bilayer Complex”. In: *Langmuir* 36.41 (2020), pp. 12272–12280.

[141] A. G. Karabadzhak et al. “Modulation of the pHLIP transmembrane helix insertion pathway”. In: *Biophys. J.* 102.8 (2012), pp. 1846–1855.

[142] J. Westerfield et al. “Ions modulate key interactions between pHLIP and lipid membranes”. In: *Biophys. J.* 117.5 (2019), pp. 920–929.

[143] J. Fendos, F. N. Barrera, and D. M. Engelman. “Aspartate Embedding Depth Affects pHLIP’s Insertion pK_a”. In: *Biochemistry-US* 52.27 (2013), pp. 4595–4604.

[144] H. L. Scott, J. M. Westerfield, and F. N. Barrera. “Determination of the Membrane Translocation pK of the pH-Low Insertion Peptide”. In: *Biophys. J.* 113.4 (2017), pp. 869–879.

[145] J. Brito et al. “Ex-vivo imaging of upper tract urothelial carcinoma using novel pH low insertion peptide (variant 3), a Mol. Imaging probe”. In: *Urology* 139 (2020), pp. 134–140.

[146] M. Zhang et al. “In Vivo Distribution and Therapeutic Efficacy of Radioiodine-Labeled pH-Low Insertion Peptide Variant 3 in a Mouse Model of Breast Cancer”. In: *Mol. Imaging* 2022 (2022).

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

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

[149] S. Donnini et al. “Constant pH molecular dynamics in explicit solvent with λ-dynamics”. In: *J. Chem. Theory Comput.* 7.6 (2011), pp. 1962–1978.

[150] J. A. Wallace and J. K. Shen. “Charge-leveling and proper treatment of long-range electrostatics in all-atom molecular dynamics at constant pH”. In: *J. Chem. Phys.* 137.18 (2012), p. 184105.

[151] A. M. Baptista, V. H. Teixeira, and C. M. Soares. “Constant-pH molecular dynamics using stochastic titration”. In: *J. Chem. Phys.* 117 (2002), pp. 4184–4200.

[152] R. Burgi, P. A. Kollman, and W. F. van Gunsteren. “Simulating proteins at constant pH: An approach combining molecular dynamics and Monte Carlo simulation”. In: *Proteins Struct. Funct. Bioinf.* 47.4 (2002), pp. 469–480.

[153] M. Dlugosz and J. M. Antosiewicz. “Constant-pH molecular dynamics simulations: a test case of succinic acid”. In: *Chem. Phys.* 302.1-3 (2004), 161–170.

[154] J. Mongan, D. A. Case, and J. A. McCammon. “Constant pH molecular dynamics in generalized Born implicit solvent”. In: *J. Comput. Chem.* 25 (2004), pp. 2038–2048.

[155] H. A. Stern. “Molecular simulation with variable protonation states at constant pH”. In: *J. Chem. Phys.* 126 (2007), p. 164112.

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

[157] V. H. Teixeira et al. “Protonation of DMPC in a Bilayer Environment Using a Linear Response Approximation”. In: *J. Chem. Theory Comput.* 10 (2014), pp. 2176–2184.

[158] H. A. Santos et al. “Constant-pH MD simulations of DMPA/DMPC lipid bilayers”. In: *J. Chem. Theory Comput.* 11.12 (2015), pp. 5973–5979.

[159] M. Machuqueiro and A. M. Baptista. “Constant-pH Molecular Dynamics with Ionic Strength Effects: Protonation–Conformation Coupling in Decalysine”. In: *J. Phys. Chem. B* 110 (2006), pp. 2927–2933.

[160] P. T. Borges et al. “Unveiling molecular details behind improved activity at neutral to alkaline pH of an engineered DyP-type peroxidase”. In: *Comput. Struct. Biotechnol. J.* (2022).

[161] P. B. Reis et al. “Role of counterions in constant-pH molecular dynamics simulations of PAMAM dendrimers”. In: *ACS Omega* 3.2 (2018), pp. 2001–2009.

[162] P. R. Magalhães, M. Machuqueiro, and A. M. Baptista. “Constant-pH Molecular Dynamics Study of Kyotorphin in an Explicit Bilayer”. In: *Biophys. J.* 108.9 (2015), pp. 2282–2290.

[163] N. F. Oliveira and M. Machuqueiro. “Novel US-CpHMD Protocol to Study the Protonation-Dependent Mechanism of the ATP/ADP Carrier”. In: *J. Chem. Inf. Model.* 62.10 (2022), pp. 2550–2560.

[164] D. Vila-Viçosa et al. “Treatment of ionic strength in biomolecular simulations of charged lipid bilayers”. In: *J. Chem. Theory Comput.* 10 (2014), pp. 5483–5492.

[165] D. Vila-Viçosa et al. “Constant-pH MD simulations of an oleic acid bilayer”. In: *J. Chem. Theory Comput.* 11.5 (2015), pp. 2367–2376.

[166] D. Vila-Viçosa et al. “A pH replica exchange scheme in the stochastic titration constant-pH MD method”. In: *J. Chem. Theory Comput.* 15.5 (2019), pp. 3108–3116.

[167] Y. Sugita and Y. Okamoto. “Replica-exchange molecular dynamics method for protein folding”. In: *Chem. Phys. Lett.* 314.1 (1999), pp. 141–151.

[168] S. J. Marrink and H. J. Berendsen. “Permeation process of small molecules across lipid membranes studied by molecular dynamics simulations”. In: *J. Phys. Chem.* 100.41 (1996), pp. 16729–16738.

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

[170] M. P. Allen and D. J. Tildesley. *Computer Simulation of Liquids*. Oxford University Press, USA, 1987, pp. 191–198.

[171] W. F. van Gunsteren et al. *Biomolecular Simulation: the GROMOS96 Manual and User Guide*. Zürich: vdf Hochschulverlag AG an der ETH Zürich, 1996.

[172] N. Schmid et al. “Definition and testing of the GROMOS force-field versions 54A7 and 54B7”. In: *Eur. Biophys. J.* 40 (2011), pp. 843–856.

[173] K. B. Koziara et al. “Testing and validation of the Automated Topology Builder (ATB) version 2.0: prediction of hydration free enthalpies”. In: *J. Comput. Aided Mol. Des.* 28.3 (2014), pp. 221–233.

[174] M. Frisch and F. Clemente. “Gaussian 09, Revision A. 01, MJ Frisch, GW Trucks, HB Schlegel, GE Scuseria, MA Robb, JR Cheeseman, G”. In: *Scalmani, V. Barone, B. Mennucci, GA Petersson, H. Nakatsuji, M. Caricato, X. Li, HP Hratchian, AF Izmaylov, J. Bloino, G. Zhe* (2009).

[175] P. J. Stephens et al. “Ab initio calculation of vibrational absorption and circular dichroism spectra using density functional force fields”. In: *J. Phys. Chem.* 98.45 (1994), pp. 11623–11627.

[176] U. C. Singh and P. A. Kollman. “An approach to computing electrostatic charges for molecules”. In: *J. Comput. Chem.* 5.2 (1984), pp. 129–145.

[177] D. van der Spoel et al. “GROMACS: Fast, Flexible, and Free”. In: *J. Comput. Chem.* 26 (2005), pp. 1701–1718.

[178] R. W. Hockney, S. P. Goel, and J. W. Eastwood. “Quiet high-resolution computer models of a plasma”. In: *J. Comput. Phys.* 14.2 (1974), pp. 148–158.

[179] S. Reißer et al. “Real cost of speed: the effect of a time-saving multiple-time-stepping algorithm on the accuracy of molecular dynamics simulations”. In: *J. Chem. Theory Comput.* 13.6 (2017), pp. 2367–2372.

[180] T. F. Silva et al. “The impact of using single atomistic long-range cutoff schemes with the GROMOS 54A7 force field”. In: *J. Chem. Theory Comput.* 14.11 (2018), pp. 5823–5833.

[181] M. Diem and C. Oostenbrink. “The Effect of Using a Twin-Range Cutoff Scheme for Nonbonded Interactions: Implications for Force-Field Parametrization?” In: *J. Chem. Theory Comput.* 16.10 (2020), pp. 5985–5990.

[182] I. G. Tironi et al. “A generalized reaction field method for molecular dynamics simulations”. In: *J. Chem. Phys.* 102 (1995), pp. 5451–5459.

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

[184] L. Onsager. “Electric moments of molecules in liquids”. In: *J. Am. Chem. Soc.* 58.8 (1936), pp. 1486–1493.

[185] D. G. Luenberger, Y. Ye, et al. *Linear and nonlinear programming*. Vol. 2. Springer, 1984.

[186] S. Nosé. “A unified formulation of the constant temperature molecular dynamics methods”. In: *J. Chem. Phys.* 81.1 (1984), pp. 511–519.

[187] W. G. Hoover. “Canonical dynamics: Equilibrium phase-space distributions”. In: *Phys. Rev. A* 31.3 (1985), p. 1695.

[188] H. J. C. Berendsen et al. “Molecular dynamics with coupling to an external bath”. In: *J. Chem. Phys.* 81 (1984), pp. 3684–3690.

[189] G. Bussi, D. Donadio, and M. Parrinello. “Canonical sampling through velocity rescaling”. In: *J. Chem. Phys.* 126 (2007), p. 014101.

[190] W. Rocchia, E. Alexov, and B. Honig. “Extending the applicability of the nonlinear Poisson–Boltzmann equation: multiple dielectric constants and multivalent ions”. In: *J. Phys. Chem. B* 105.28 (2001), pp. 6507–6514.

[191] W. Rocchia et al. “Rapid grid-based construction of the molecular surface and the use of induced surface charge to calculate reaction field energies: Applications to the molecular systems and geometric objects”. In: *J. Comput. Chem.* 23 (2002), pp. 128–137.

[192] D. Bashford and M. Karplus. “pK_a’s of ionizable groups in proteins: atomic detail from a continuum electrostatic model”. In: *Biochemistry-US* 29.44 (1990), pp. 10219–10225.

[193] A. M. Baptista and C. M. Soares. “Some Theoretical and Computational Aspects of the Inclusion of Proton Isomerism in the Protonation Equilibrium of Proteins”. In: *J. Phys. Chem. B* 105 (2001), pp. 293–309.

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

[195] N. Metropolis et al. “Equation of state calculations by fast computing machines”. In: *J. Chem. Phys.* 21.6 (1953), pp. 1087–1092.

[196] P. Beroza et al. “Protonation of interacting residues in a protein by a Monte Carlo method: application to lysozyme and the photosynthetic reaction center of Rhodobacter sphaeroides”. In: *Proc. Natl. Acad. Sci. USA* 88.13 (1991), pp. 5804–5808.

[197] P. B. Reis et al. “PypKa: A Flexible Python Module for Poisson–Boltzmann-Based p K a Calculations”. In: *J. Chem. Inf. Model.* 60.10 (2020), pp. 4442–4448.

[198] N. Aho et al. “Scalable Constant pH Molecular Dynamics in GROMACS”. In: (2022).

[199] R. Anandakrishnan, B. Aguilar, and A. V. Onufriev. “H++ 3.0: automating pK prediction and the preparation of biomolecular structures for atomistic molecular modeling and simulations”. In: *Nucleic Acids Res.* 40.W1 (2012), W537–W541.

[200] L. Wang, M. Zhang, and E. Alexov. “DelPhiPKa web server: predicting p K a of proteins, RNAs and DNAs”. In: *Bioinformatics* 32.4 (2016), pp. 614–615.

[201] Y. Song, J. Mao, and M. Gunner. “MCCE2: improving protein pK_a calculations with extensive side chain rotamer sampling”. In: *J. Comput. Chem.* 30.14 (2009), pp. 2231–2247.

[202] D. E. Shaw et al. “Anton, a special-purpose machine for molecular dynamics simulation”. In: *Commun. ACM* 51.7 (2008), pp. 91–97.

[203] A. Laio and M. Parrinello. “Escaping free-energy minima”. In: *Proc. Natl. Acad. Sci. USA* 99.20 (2002), pp. 12562–12566.

[204] R. C. Bernardi, M. C. Melo, and K. Schulten. “Enhanced sampling techniques in molecular dynamics simulations of biological systems”. In: *Biochem. Biophys. Acta* 1850.5 (2015), pp. 872–877.

[205] M. Christen and W. F. Van Gunsteren. “On searching in, sampling of, and dynamically moving through conformational space of biomolecular systems: A review”. In: *J. Comput. Chem.* 29.2 (2008), pp. 157–166.

[206] S. G. Itoh, A. Damjanović, and B. R. Brooks. “pH replica-exchange method based on discrete protonation states”. In: *Proteins Struct. Funct. Bioinf.* 79.12 (2011), pp. 3420–3436.

[207] T. Yamaguchi et al. “Multi-pH Monte Carlo simulation of coil–globule transition of weak polyelectrolyte”. In: *Bull. Chem. Soc. Jpn.* 78.12 (2005), pp. 2098–2104.

[208] J. P. Jambeck and A. P. Lyubartsev. “Exploring the free energy landscape of solutes embedded in lipid bilayers”. In: *J. Phys. Chem Lett* 4.11 (2013), pp. 1781–1787.

[209] M. Paloncyova, K. Berka, and M. Otyepka. “Convergence of free energy profile of coumarin in lipid bilayer”. In: *J. Chem. Theory Comput.* 8.4 (2012), pp. 1200–1211.

[210] C. Neale et al. “Accelerating convergence in molecular dynamics simulations of solutes in lipid membranes by conducting a random walk along the bilayer normal”. In: *J. Chem. Theory Comput.* 9.8 (2013), pp. 3686–3703.

[211] I. Kabelka, R. Brozek, and R. Vácha. “Selecting collective variables and free-energy methods for peptide translocation across membranes”. In: *J. Chem. Inf. Model.* 61.2 (2021), pp. 819–830.

[212] G. Bussi and A. Laio. “Using metadynamics to explore complex free-energy landscapes”. In: *Nat. Rev. Phys.* 2.4 (2020), pp. 200–212.

[213] J. Kästner. “Umbrella sampling”. In: *Wiley Interdiscip. Rev. Comput. Mol. Sci.* 1.6 (2011), pp. 932–942.

[214] N. Kučerka, M.-. Nieh, and J. Katsaras. “Fluid phase lipid areas and bilayer thicknesses of commonly used phosphatidylcholines as a function of temperature”. In: *Biochem. Biophys. Acta, Biomembr.* 1808 (2011), pp. 2761–2771.

[215] J. M. Diamond and Y. Katz. “Interpretation of nonelectrolyte partition coefficients between dimyristoyl lecithin and water”. In: *J. Membr. Biol.* 17.1 (1974), pp. 121–154.

[216] E. Awoonor-Williams and C. N. Rowley. “Molecular simulation of nonfacilitated membrane permeation”. In: *Biochem. Biophys. Acta, Biomembr.* 1858.7 (2016), pp. 1672–1687.

[217] R. M. Venable, A. Kramer, and R. W. Pastor. “Molecular dynamics simulations of membrane permeability”. In: *Chem. Rev.* 119.9 (2019), pp. 5954–5997.

[218] J. P. Ulmschneider and M. B. Ulmschneider. “Molecular dynamics simulations are redefining our view of peptides interacting with biological membranes”. In: *Accounts Chem. Res.* 51.5 (2018), pp. 1106–1116.

[219] E. Deplazes et al. “Structural Characterization of a Cation-Selective, Self-Assembled Peptide Pore in Planar Phospholipid Bilayers”. In: *J. Phys. Chem Lett* 11.19 (2020), pp. 8152–8156.

[220] C. Albrecht et al. “Transmembrane Peptides as Inhibitors of Protein-Protein Interactions: An Efficient Strategy to Target Cancer Cells?” In: *Front. Oncol.* 10 (2020), p. 519.

[221] L. Yao et al. “pHLIP peptide targets nanogold particles to tumors”. In: *Proc. Natl. Acad. Sci. USA* 110.2 (2013), pp. 465–470.

[222] O. A. Andreev, D. M. Engelman, and Y. K. Reshetnyak. “Targeting diseased tissues by pHLIP insertion at low cell surface pH”. In: *Front. Physiol.* 5 (2014).

[223] S. Z. Hanz et al. “Protonation-Driven Membrane Insertion of a pH-Low Insertion Peptide”. In: *Angew. Chem. Int. Ed.* 55.40 (2016), pp. 12376–12381.

[224] O. A. Andreev et al. “pH (low) insertion peptide (pHLIP) inserts across a lipid bilayer as a helix and exits by a different path”. In: *Proc. Natl. Acad. Sci. USA* 107.9 (2010), pp. 4081–4086.

[225] M. Machuqueiro and A. M. Baptista. “Molecular Dynamics Constant-pH and Reduction Potential: Application to Cytochrome c_3”. In: *J. Am. Chem. Soc.* 131 (2009), pp. 12586–12594.

[226] Z. Yue et al. “Dynamic Protonation Dramatically Affects the Membrane Permeability of Drug-like Molecules”. In: *J. Am. Chem. Soc.* 141.34 (2019), pp. 13421–13433.

[227] B. Hess et al. “GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation”. In: *J. Chem. Theory Comput.* 4 (2008), pp. 435–447.

[228] P. E. Smith and W. F. van Gunsteren. “Consistent dielectric properties of the simple point charge and extended point charge water models at 277 and 300 K”. In: *J. Chem. Phys.* 100 (1994), pp. 3169–3174.

[229] B. Hess. “P-LINCS: A Parallel Linear Constraint Solver for Molecular Simulation”. In: *J. Chem. Theory Comput.* 4 (2008), pp. 116–122.

[230] J. Hermans et al. “A Consistent Empirical Potential for Water-Protein Interactions”. In: *Biopolymers* 23 (1984), pp. 1513–1518.

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

[232] M. Parrinello and A. Rahman. “Polymorphic transitions in single crystals: A new molecular dynamics method”. In: *J. Appl. Phys.* 52.12 (1981), pp. 7182–7190. ISSN: 0021-8979. DOI: `10.1063/1.328693`.

[233] V. H. Teixeira et al. “On the Use of Different Dielectric Constants for Computing Individual and Pairwise Terms in Poisson-Boltzmann Studies of Protein Ionization Equilibrium”. In: *J. Phys. Chem. B* 109 (2005), pp. 14691–14706.

[234] G. R. Grimsley, J. M. Scholtz, and C. N. Pace. “A summary of the measured pK values of the ionizable groups in folded proteins”. In: *Protein Sci.* 18.1 (2009), pp. 247–251.

[235] P. R. Magalhães et al. “Effect of a pH gradient on the protonation states of cytochrome c oxidase: A continuum electrostatics study”. In: *J. Chem. Inf. Model.* 57.2 (2017), pp. 256–266.

[236] K. Ataka et al. “pH-induced insertion of pHLIP into a lipid bilayer: In-situ SEIRAS characterization of a folding intermediate at neutral pH”. In: *Biochem. Biophys. Acta, Biomembr.* (2022), p. 183873.

[237] M. Zoonens, Y. K. Reshetnyak, and D. M. Engelman. “Bilayer interactions of pHLIP, a peptide that can deliver drugs and target tumors”. In: *Biophys. J.* 95.1 (2008), pp. 225–235.

[238] D. W. Demoin et al. “PET imaging of extracellular pH in tumors with 64Cu-and 18F-labeled pHLIP peptides: a structure–activity optimization study”. In: *Bioconjugate Chem.* 27.9 (2016), pp. 2014–2023.

[239] N. Frazee and B. Mertz. “Intramolecular interactions play key role in stabilization of pHLIP at acidic conditions”. In: *J. Comput. Chem.* 42.25 (2021), pp. 1809–1816.

[240] N. Calimet and G. M. Ullmann. “The influence of a transmembrane pH gradient on protonation probabilities of bacteriorhodopsin: the structural basis of the back-pressure effect”. In: *J. Mol. Biol.* 339.3 (2004), pp. 571–589.

[241] J. A. Killian and T. K. Nyholm. “Peptides in lipid bilayers: the power of simple models”. In: *Curr. Opinion Struct. Biol.* 16.4 (2006). Membranes / Engineering and design, pp. 473–479. ISSN: 0959-440X.

[242] “Cooperative Nonbonded Forces Control Membrane Binding of the pH-Low Insertion Peptide pHLIP”. In: *Biophys. J.* ().

[243] H. Herce et al. “Arginine-rich peptides destabilize the plasma membrane, consistent with a pore formation translocation mechanism of cell-penetrating peptides”. In: *Biophys. J.* 97.7 (2009), pp. 1917–1925.

[244] T. Bereau et al. “Folding and insertion thermodynamics of the transmembrane WALP peptide”. In: *J. Chem. Phys.* 143.24 (2015), p. 243127.

[245] C. Gupta and B. Mertz. “Protonation enhances the inherent helix-forming propensity of pHLIP”. In: *ACS Omega* 2.11 (2017), pp. 8536–8542.

[246] S. A. Otieno and W. Qiang. “Roles of key residues and lipid dynamics reveal pHLIP-membrane interactions at intermediate pH”. In: *Biophys. J.* 120.21 (2021), pp. 4649–4662. ISSN: 0006-3495.

[247] T. T. Tapmeier et al. “The pH low insertion peptide pHLIP Variant 3 as a novel marker of acidic malignant lesions”. In: *Proc. Natl. Acad. Sci. USA* 112.31 (2015), pp. 9710–9715.

[248] T. F. Silva, D. Vila-Viçosa, and M. Machuqueiro. “Increase the realism of *in silico* pHLIP peptide models with a novel pH gradient CpHMD method”. In: *Submitted* ().

[249] H. Hirose et al. “Transient focal membrane deformation induced by arginine-rich peptides leads to their direct penetration into cells”. In: *Mol. Therapy* 20.5 (2012), pp. 984–993.

[250] M. J. Abraham et al. “GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers”. In: *SoftwareX* 1-2 (2015), pp. 19–25. ISSN: 2352-7110. DOI: `https://doi.org/10.1016/j.softx.2015.06.001`. URL: `https://www.sciencedirect.com/science/article/pii/S2352711015000059`.

[251] P. Virtanen et al. “SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python”. In: *Nat. Methods* 17 (2020), pp. 261–272. DOI: `10.1038/s41592-019-0686-2`.

[252] F. Pedregosa et al. “Scikit-learn: Machine Learning in Python”. In: *J. Mac. Learn. Res.* 12 (2011), pp. 2825–2830.

[253] R. E. Lawrence and R. Zoncu. “The lysosome as a cellular centre for signalling, metabolism and quality control”. In: *Nat. Cell Biol.* 21.2 (2019), pp. 133–142.

[254] S. Inpanathan and R. J. Botelho. “The lysosome signaling platform: adapting with the times”. In: *Front. Cell Dev. Biol.* 7 (2019), p. 113.

[255] S. M. Davidson and M. G. Vander Heiden. “Critical functions of the lysosome in cancer biology”. In: *Annu. Rev. Pharmacol. Toxicol.* 57 (2017), pp. 481–507.

[256] C. Fennelly and R. K. Amaravadi. “Lysosomal biology in cancer”. In: *Lysosomes* (2017), pp. 293–308.

[257] A. Sun. “Lysosomal storage disease overview”. In: *Ann. Transl. Med.* 6.24 (2018).

[258] T. Woldemichael and G. R. Rosania. “The physiological determinants of drug-induced lysosomal stress resistance”. In: *PloS One* 12.11 (2017), e0187627.

[259] P. Schneider, T. A. Korolenko, and U. Busch. “A review of drug-induced lysosomal disorders of the liver in man and laboratory animals”. In: *Micros. Res. Tech.* 36.4 (1997), pp. 253–275.

[260] N. Anderson and J. Borlak. “Drug-induced phospholipidosis”. In: *FEBS Lett.* 580.23 (2006), pp. 5533–5540.

[261] S. Ohkuma and B. Poole. “Fluorescence probe measurement of the intralysosomal pH in living cells and the perturbation of pH by various agents.” In: *Proc. Natl. Acad. Sci. USA* 75.7 (1978), pp. 3327–3331.

[262] D. E. Johnson et al. “The position of lysosomes within the cell determines their luminal pH”. In: *J. Cell Biol.* 212.6 (2016), pp. 677–692.

[263] S. Pisonero-Vaquero and D. L. Medina. “Lysosomotropic drugs: pharmacological tools to study lysosomal function”. In: *Curr. Drug Met.* 18.12 (2017), pp. 1147–1158.

[264] J. Sironi et al. “Lysosome membrane permeabilization and disruption of the molecular target of rapamycin (mTOR)-lysosome interaction are associated with the inhibition of lung cancer cell proliferation by a chloroquinoline analog”. In: *Mol. Pharmacol.* 95.1 (2019), pp. 127–138.

[265] X. Zhang et al. “Trifluoperazine prolongs the survival of experimental brain metastases by STAT3-dependent lysosomal membrane permeabilization”. In: *Am. J. Cancer Res.* 10.2 (2020), p. 545.

[266] M. Circu et al. “Modulating lysosomal function through lysosome membrane permeabilization or autophagy suppression restores sensitivity to cisplatin in refractory non-small-cell lung cancer cells”. In: *PloS One* 12.9 (2017), e0184922.

[267] S. Noguchi et al. “Bosutinib, an SRC inhibitor, induces caspase-independent cell death associated with permeabilization of lysosomal membranes in melanoma cells”. In: *Vet. Comp. Oncol.* 16.1 (2018), pp. 69–76.

[268] M. S. Ostenfeld et al. “Effective tumor cell death by σ-2 receptor ligand siramesine involves lysosomal leakage and oxidative stress”. In: *Cancer Res.* 65.19 (2005), pp. 8975–8983.

[269] S. Drori, G. D. Eytan, and Y. G. Assaraf. “Potentiation of anticancer-drug cytotoxicity by multidrug-resistance chemosensitizers involves alterationsin membrane fluidity leading to increased membrane permeability”. In: *Eur. J. Biochem.* 228.3 (1995), pp. 1020–1029.

[270] T. Murata et al. “A comparative study of the plasma membrane permeabilization and fluidization induced by antipsychotic drugs in the rat brain”. In: *Int. J. Neuropsychopharmacol.* 10.5 (2007), pp. 683–689.

[271] G. Breuzard et al. “Changes in adsorption and permeability of mitoxantrone on plasma membrane of BCRP/MXR resistant cells”. In: *Biochem. Biophys. Res. Commun.* 329.1 (2005), pp. 64–70.

[272] Y.-C. Kim et al. “Synergistic enhancement of skin permeability by N-lauroylsarcosine and ethanol”. In: *Int. J. Pharm.* 352.1-2 (2008), pp. 129–138.

[273] A.-M. Ellegaard et al. “Sunitinib and SU11652 Inhibit Acid Sphingomyelinase, Destabilize Lysosomes, and Inhibit Multidrug ResistanceSunitinib and SU11652 Target Lysosomes”. In: *Mol. Cancer Ther.* 12.10 (2013), pp. 2018–2030.

[274] M. S. Ostenfeld et al. “Anti-cancer agent siramesine is a lysosomotropic detergent that induces cytoprotective autophagosome accumulation”. In: *Autophagy* 4.4 (2008), pp. 487–499.

[275] S. J. Hurwitz et al. “Vesicular anthracycline accumulation in doxorubicin-selected U-937 cells: participation of lysosomes”. In: *Blood, J. Am. Soc. Hemat.* 89.10 (1997), pp. 3745–3754.

[276] B. Zhao et al. “TFEB-mediated lysosomal biogenesis and lysosomal drug sequestration confer resistance to MEK inhibition in pancreatic cancer”. In: *Cell Death Dis.* 6.1 (2020), pp. 1–13.

[277] U. Hellwich and R. Schubert. “Concentration-dependent binding of the chiral β-blocker oxprenolol to isoelectric or negatively charged unilamellar vesicles”. In: *Biochem. Pharmacol.* 49.4 (1995), pp. 511–517.

[278] P. Barton et al. “Drug-phospholipid interactions. 2. Predicting the sites of drug distribution using n-octanol/water and membrane/water distribution coefficients”. In: *J. Pharm. Sci.* 86.9 (1997), pp. 1034–1039.

[279] A. Avdeef et al. “pH-Metric logP 10. Determination of liposomal membrane-water partition coefficients of lonizable drugs”. In: *Pharm. Res.* 15.2 (1998), pp. 209–215.

[280] C. A. Carvalheda, S. R. R. Campos, and A. M. Baptista. “The Effect of Membrane Environment on Surfactant Protein C Stability Studied by Constant-pH Molecular Dynamics”. In: *J. Chem. Inf. Model.* (2015).

[281] D. B. Rubin. “The bayesian bootstrap”. In: *Annals Of Statistics* (1981), pp. 130–134.

[282] H. Pajouhesh and G. R. Lenz. “Medicinal chemical properties of successful central nervous system drugs”. In: *NeuroRx* 2.4 (2005), pp. 541–553.

[283] G. Anderson, A. Linden, and K. Rabe. “Why are long-acting beta-adrenoceptor agonists long-acting?” In: *Eur. Respir. J.* 7.3 (1994), pp. 569–578.

[284] L. A. Opler and S. S. Feinberg. “The role of pimozide in clinical psychiatry: a review.” In: *J. Clin. Psychiatry* (1991).

[285] S. Zielke et al. “Loperamide, pimozide, and STF-62247 trigger autophagy-dependent cell death in glioblastoma cells”. In: *Cell Death Dis.* 9.10 (2018), pp. 1–16.

[286] J. Coleman et al. “A live-cell fluorescence microplate assay suitable for monitoring vacuolation arising from drug or toxic agent treatment”. In: *J. Biomol. Screen.* 15.4 (2010), pp. 398–405.

[287] M. Wilson and J. Tripp. “Clomipramine”. In: *StatPearls [Internet].* StatPearls Publishing, 2021.

[288] S. Nadanaciva et al. “A high content screening assay for identifying lysosomotropic compounds”. In: *Toxicol. In Vitro* 25.3 (2011), pp. 715–723.

[289] D. R. Brocks. “Anticholinergic drugs used in Parkinson’s disease: An overlooked class of drugs from a pharmacokinetic perspective”. In: *J. Pharm. Sci.* 2.2 (1999), pp. 39–46.

[290] J. Kornhuber et al. “Identification of novel functional inhibitors of acid sphingomyelinase”. In: *PloS One* 6.8 (2011), e23852.

[291] N. Beckmann et al. “Inhibition of acid sphingomyelinase by tricyclic antidepressants and analogons”. In: *Front. Physiol.* 5 (2014), p. 331.

[292] H. Zhao et al. “Chloroquine-mediated radiosensitization is due to the destabilization of the lysosomal membrane and subsequent induction of cell death by necrosis”. In: *Radiat. Res.* 164.3 (2005), pp. 250–257.

[293] Y. Adar et al. “Imidazoacridinone-dependent lysosomal photodestruction: a pharmacological Trojan horse approach to eradicate multidrug-resistant cancers”. In: *Cell Death Dis.* 3.4 (2012), e293–e293.

[294] Y. Lu et al. “Vacuolin-1 potently and reversibly inhibits autophagosome-lysosome fusion by activating RAB5A”. In: *Autophagy* 10.11 (2014), pp. 1895–1905.

[295] J. Cerny et al. “The small chemical vacuolin-1 inhibits Ca2+-dependent lysosomal exocytosis but not cell resealing”. In: *EMBO Rep.* 5.9 (2004), pp. 883–888.

[296] G. Aubel-Sadron and D. Londos-Gagliardi. “Daunorubicin and doxorubicin, anthracycline antibiotics, a physicochemical and biological review”. In: *Biochimie* 66.5 (1984), pp. 333–352.

[297] F. Varone et al. “Nintedanib for the treatment of idiopathic pulmonary fibrosis”. In: *Expert Opinion on Pharmacotherapy* 19.2 (2018), pp. 167–175.

[298] G. J. Roth et al. “Nintedanib: from discovery to the clinic”. In: *J. Med. Chem.* 58.3 (2015), pp. 1053–1063.

[299] A. M. Kaufmann and J. P. Krise. “Lysosomal sequestration of amine-containing drugs: analysis and therapeutic implications”. In: *J. Pharm. Sci.* 96.4 (2007), pp. 729–746.

[300] R. Logan, A. C. Kong, and J. P. Krise. “Time-Dependent Effects of Hydrophobic Amine-Containing Drugs on Lysosome Structure and Biogenesis in Cultured Human Fibroblasts”. In: *J. Pharm. Sci.* 103.10 (2014), pp. 3287–3296.

[301] J. Faigle and W. Dieterle. “The metabolism and pharmacokinetics of clomipramine (Anafranil)”. In: *Int. J. Med. Res.* 1.5 (1973), pp. 281–290.

[302] N. El Tayar, H. Van de Waterbeemd, and B. Testa. “Lipophilicity measurements of protonated basic compounds by reversed-phase high-performance liquid chromatography: II. Procedure for the determination of a lipophilic index measured by reversed-phase high-performance liquid chromatography”. In: *J. Chromatogr. A* 320.2 (1985), pp. 305–312.

[303] P. Marshall. “Some chemical and physical properties associated with histamine antagonism”. In: *Br. J. Pharmacol. Chemother.* 10.3 (1955), p. 270.

[304] N. Sun and A. Avdeef. “Biorelevant pKa (37° C) predicted from the 2D structure of the molecule and its pK_a at 25° C”. In: *J. Pharm. Biomed. Anal.* 56.2 (2011), pp. 173–182.

[305] M. J. O'Neil, A. Smith, P. E. Heckelman, et al. *The Merck Index: An Encyclopedia of Chemicals, Drugs, and Biologicals*. Inc.: Whitehouse Station, NJ, USA, 2006.

[306] A. Di Marco et al. "Changes of activity of daunorubicin, adriamycin and stereoisomers following the introduction or removal of hydroxyl groups in the amino sugar moiety". In: *Chem.–Biol. Interact.* 19.3 (1977), pp. 291–302.

[307] S. Sanli, Y. Altun, and G. Guven. "Solvent effects on pK_a values of some anticancer agents in acetonitrile–water binary mixtures". In: *J. of Chem. & Eng. Data* 59.12 (2014), pp. 4015–4020.

[308] M. Duvvuri et al. "Weak base permeability characteristics influence the intracellular sequestration site in the multidrug-resistant human leukemic cell line HL-60". In: *J. Biol. Chem.* 279.31 (2004), pp. 32367–32372.

[309] C. Hansch, A. Leo, and D. Hoekman. *Exploring QSAR.: Fundamentals and applications in chemistry and biology. Hydrophobic, electronic, and steric constants*. ACS professional reference book vol. 1-2. American Chemical Society, 1995. ISBN: 9780841229938.

[310] J. Sangster. "LOGKOW–a databank of evaluated octanol-water partition coefficients". In: *Sangster Res. Lab. Mont.* (1993), pp. 1–15.

[311] A. Leo, C. Hansch, and D. Elkins. "Partition coefficients and their uses". In: *Chem. Rev.* 71.6 (1971), pp. 525–616.

[312] F. Yang, Y. Zhang, and H. Liang. "Interactive association of drugs binding to human serum albumin". In: *Int. J. Mol. Sci.* 15.3 (2014), pp. 3580–3595.

[313] P. A. Shore, B. B. Brodie, and C. A. M. Hogben. "The gastric secretion of drugs: a pH partition hypothesis". In: *J. Pharmacol. Exp. Ther* 119.3 (1957), pp. 361–369.

[314] M. Paloncyova et al. "Amphiphilic drug-like molecules accumulate in a membrane below the head group region". In: *J. Phys. Chem. B* 118.4 (2014), pp. 1030–1039.

[315] K. J. Seu et al. "Influence of lipid chemistry on membrane fluidity: tail and headgroup interactions". In: *Biophys. J.* 91.10 (2006), pp. 3727–3735.

[316] D. Matyszewska. "Comparison of the interactions of daunorubicin in a free form and attached to single-walled carbon nanotubes with model lipid membranes". In: *Beilstein J. Nanotechnol.* 7.1 (2016), pp. 524–532.

[317] J. Kornhuber et al. "Lipophilic cationic drugs increase the permeability of lysosomal membranes in a cell culture system". In: *J. Cell. Physiol.* 224.1 (2010), pp. 152–164.

[318] A. Avdeef. "Physicochemical profiling (solubility, permeability and charge state)". In: *Curr. Top. Med. Chem.* 1.4 (2001), pp. 277–351.

[319] Y. Boulanger, S. Schreier, and I. C. Smith. "Molecular details of anesthetic-lipid interaction as seen by deuterium and phosphorus-31 nuclear magnetic resonance". In: *Biochemistry-US* 20.24 (1981), pp. 6824–6830.

[320] H. Schulze, T. Kolter, and K. Sandhoff. “Principles of lysosomal membrane degradation: Cellular topology and biochemistry of lysosomal lipid degradation”. In: *BBA-Mol. Cell Res.* 1793.4 (2009), pp. 674–683.

[321] T. Kolter and K. Sandhoff. “Principles of lysosomal membrane digestion: stimulation of sphingolipid degradation by sphingolipid activator proteins and anionic lysosomal lipids”. In: *Annu. Rev. Cell Dev. Biol.* 21 (2005), p. 81.

[322] A. M. Kaufmann and J. P. Krise. “Niemann-Pick C1 functions in regulating lysosomal amine content”. In: *J. Biol. Chem.* 283.36 (2008), pp. 24584–24593.

[323] Y. Gong et al. “Niemann-Pick C1 protein facilitates the efflux of the anticancer drug daunorubicin from cells according to a novel vesicle-mediated pathway”. In: *J. Pharmacol. Exp. Ther* 316.1 (2006), pp. 242–247.

[324] C. Sousa et al. “Solvent polarity and pH effects on the spectroscopic properties of neutral red: application to lysosomal microenvironment probing in living cells”. In: *J. Photochem. Photobiol. C* 63.5 (1996), pp. 601–607.

[325] M. V. Schmitt et al. “Quantitation of lysosomal trapping of basic lipophilic compounds using in vitro assays and in silico predictions based on the determination of the full pH profile of the endo-/lysosomal system in rat hepatocytes”. In: *Drug Metab. Dispos.* 47.1 (2019), pp. 49–57.

[326] B. Lemieux, M. D. Percival, and J.-P. Falgueyret. “Quantitation of the lysosomotropic character of cationic amphiphilic drugs using the fluorescent basic amine Red DND-99”. In: *Anal. Biochem.* 327.2 (2004), pp. 247–251.

[327] K. Yoshida et al. “Phospholipid membrane fluidity alters ligand binding activity of a G protein-coupled receptor by shifting the conformational equilibrium”. In: *Biochemistry-US* 58.6 (2019), pp. 504–508.

[328] G. Ge, J. Wu, and Q. Lin. “Effect of membrane fluidity on tyrosine kinase activity of reconstituted epidermal growth factor receptor”. In: *Biochem. Biophys. Res. Commun.* 282.2 (2001), pp. 511–514.

[329] R. Gowda et al. “Identifying the structure-activity relationship of leelamine necessary for inhibiting intracellular cholesterol transport”. In: *Oncotarget* 8.17 (2017), p. 28260.

[330] O. F. Kuzu et al. “Leelamine Mediates Cancer Cell Death through Inhibition of Intracellular Cholesterol TransportLeelamine Inhibits Intracellular Cholesterol Transport”. In: *Mol. Cancer Ther.* 13.7 (2014), pp. 1690–1703.

[331] S. D. Goldman et al. “Mechanisms of amine accumulation in, and egress from, lysosomes”. In: *Bioanalysis* 1.8 (2009), pp. 1445–1459.

[332] M. Duvvuri and J. P. Krise. “A novel assay reveals that weakly basic model compounds concentrate in lysosomes to an extent greater than pH-partitioning theory would predict”. In: *Mol. Pharm.* 2.6 (2005), pp. 440–448.

[333] W. Kopeć, J. Telenius, and H. Khandelia. “Molecular dynamics simulations of the interactions of medicinal plant extracts and drugs with lipid bilayer membranes”. In: *FEBS J.* 280.12 (2013), pp. 2785–2805.

[334] T. Róg and M. Pasenkiewicz-Gierula. “Non-polar interactions between cholesterol and phospholipids: a molecular dynamics simulation study”. In: *Biophys. Chem.* 107.2 (2004), pp. 151–164.

[335] J. M. Smaby et al. “Phosphatidylcholine acyl unsaturation modulates the decrease in interfacial elasticity induced by cholesterol”. In: *Biophys. J.* 73 (1997), pp. 1492–1505.

[336] A. C. Alves et al. “Influence of doxorubicin on model cell membrane properties: insights from in vitro and in silico studies”. In: *Sci. R.ep* 7.1 (2017), pp. 1–11.

[337] D. B. Goldstein. “The effects of drugs on membrane fluidity”. In: *Annu. Rev. Pharmacol. Toxicol.* 24.1 (1984), pp. 43–64.

[338] M. Kölzer, N. Werth, and K. Sandhoff. “Interactions of acid sphingomyelinase and lipid bilayers in the presence of the tricyclic antidepressant desipramine”. In: *FEBS Lett.* 559.1-3 (2004), pp. 96–98.

[339] S. Elojeimy et al. “New insights on the use of desipramine as an inhibitor for acid ceramidase”. In: *FEBS Lett.* 580.19 (2006), pp. 4751–4756.

[340] Y. Matsuzawa and K. Y. Hostetler. “Inhibition of lysosomal phospholipase A and phospholipase C by chloroquine and 4, 4’-bis (diethylaminoethoxy) alpha, beta-diethyldiphenylethane.” In: *J. Biol. Chem.* 255.11 (1980), pp. 5190–5194.

[341] J. A. Shayman et al. “Group XV phospholipase A2, a lysosomal phospholipase A2”. In: *Prog. Lipid Res.* 50.1 (2011), pp. 1–13.

[342] A. Gorelik et al. “Crystal structure of mammalian acid sphingomyelinase”. In: *Nat. Commun.* 7.1 (2016), pp. 1–9.

[343] A. Gebai et al. “Structural basis for the activation of acid ceramidase”. In: *Nat. Commun.* 9.1 (2018), pp. 1–11.

[344] A. Abe and J. A. Shayman. “The role of negatively charged lipids in lysosomal phospholipase A2 function”. In: *J. Lipid Res.* 50.10 (2009), pp. 2027–2035.

[345] T. Linke et al. “Interfacial regulation of acid ceramidase activity: stimulation of ceramide degradation by lysosomal lipids and sphingolipid activator proteins”. In: *J. Biol. Chem.* 276.8 (2001), pp. 5760–5768.

[346] R. Hurwitz, K. Ferlinz, and K. Sandhoff. “The tricyclic antidepressant desipramine causes proteolytic degradation of lysosomal sphingomyelinase in human fibroblasts”. In: *Biol. Chem. Hoppe-Seyler* 375.7 (1994), pp. 447–450.

[347] B. H. Ali et al. “Experimental gentamicin nephrotoxicity and agents that modify it: a mini-review of recent research”. In: *Basic Clin. Pharmacol. Toxicol.* 109.4 (2011), pp. 225–232.

[348] S. Kacew. “Cationic amphiphilic drug-induced renal cortical lysosomal phospholipidosis: an in vivo comparative study with gentamicin and chlorphentermine”. In: *Toxicol. Appl. Pharmacol.* 91.3 (1987), pp. 469–476.

[349] J. P. Davies and Y. A. Ioannou. “Topological analysis of Niemann-Pick C1 protein reveals that the membrane orientation of the putative sterol-sensing domain is identical to those of 3-hydroxy-3-methylglutaryl-CoA reductase and sterol regulatory element binding protein cleavage-activating protein”. In: *J. Biol. Chem.* 275.32 (2000), pp. 24367–24374.

[350] M. Kosicek et al. “N-glycome of the lysosomal glycocalyx is altered in niemann-pick type C disease (NPC) model cells”. In: *Mol. & Cell. Proteomics* 17.4 (2018), pp. 631–642.

[351] W. Neiss. “A coat of glycoconjugates on the inner surface of the lysosomal membrane in the rat kidney”. In: *Histochem.* 80.6 (1984), pp. 603–608.

[352] E.-L. Eskelinen, Y. Tanaka, and P. Saftig. “At the acidic edge: emerging functions for lysosomal membrane proteins”. In: *Trends Cell Biol.* 13.3 (2003), pp. 137–145.

[353] J. Li et al. “Glycosylation inhibition reduces cholesterol accumulation in NPC1 protein-deficient cells”. In: *Proc. Natl. Acad. Sci. USA* 112.48 (2015), pp. 14876–14881.

[354] J. A. Rodriguez-Navarro et al. “Inhibitory effect of dietary lipids on chaperone-mediated autophagy”. In: *Proc. Natl. Acad. Sci. USA* 109.12 (2012), E705–E714.

[355] J. E. Vance and R. Steenbergen. “Metabolism and functions of phosphatidylserine”. In: *Prog. Lipid Res.* 44.4 (2005), pp. 207–234.

[356] H. D. Gallala and K. Sandhoff. “Biological function of the cellular lipid BMP—BMP as a key activator for cholesterol sorting and membrane digestion”. In: *Neurochem. Res.* 36.9 (2011), pp. 1594–1600.

[357] J. R. Wherrett and S. Huterer. “Enrichment of bis-(monoacylglyceryl) phosphate in lysosomes from rat liver”. In: *J. Biol. Chem.* 247.13 (1972), pp. 4114–4120.

[358] W. Möbius et al. “Recycling compartments and the internal vesicles of multivesicular bodies harbor most of the cholesterol found in the endocytic pathway”. In: *Traffic* 4.4 (2003), pp. 222–231.

[359] H. M.-S. Monne et al. “Structure of glycocalyx”. In: *Biophys. J.* 104.2 (2013), 251a.

[360] Y.-P. Yi et al. “Phosphatidic Acid Osmotically Destabilizes Lysosomes through Increased Permeability to Kˆ+ and Hˆ+”. In: *Gen. Phys. and Biophys.* 25.2 (2006), p. 149.

[361] T. Kolter and K. Sandhoff. “Sphingolipid metabolism diseases”. In: *Biochem. Biophys. Acta, Biomembr.* 1758.12 (2006), pp. 2057–2079.

[362] P. M. Suzano. *Developing a computational method to calculate pH-dependent membrane permeabilities for anti-tumour drugs*. Lisboa, 2022.

[363] W. Kabsch and C. Sander. “Dictionary of Protein Secondary Structure: Pattern Recognition of Hydrogen-Bonded and Geometrical Features”. In: *Biopolymers* 22 (1983), pp. 2577–2637.