



Interaction between amphipathic helices and lipid membranes studied by computer simulations. Application to the LTX-315 anticancer peptide

Paula Milán Rodríguez

► To cite this version:

Paula Milán Rodríguez. Interaction between amphipathic helices and lipid membranes studied by computer simulations. Application to the LTX-315 anticancer peptide. Biological Physics [physics.bio-ph]. Sorbonne Université, 2022. English. NNT : 2022SORUS074 . tel-03961474

HAL Id: tel-03961474

<https://theses.hal.science/tel-03961474>

Submitted on 29 Jan 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Thèse de Doctorat

Domaine: Biophysique computationnelle

Interaction between amphipathic helices and lipid
membranes studied by computer simulations.
Application to the LTX-315 anticancer peptide.

Présentée par

Paula Milán Rodríguez

Pour obtenir le grade de docteur en sciences de Sorbonne Université sous la
direction de Patrick Fuchs

Thèse soutenue le 28 janvier 2022 devant le jury composé de:

Dr. Claire Loison	Université de Lyon	Rapportrice
Pr. James Sturgis	Université Aix-Marseille 2	Rapporteur
Dr. Nadège Jamin	CEA	Examinatrice
Dr. Jérôme Hénin	CNRS	Examineur
Dr. Jacques Chomilier	CNRS	Examineur
Dr. Patrick Fuchs	Université de Paris	Directeur de thèse



Acknowledgements

To Patrick, the best director I could have asked for. A great teacher, great boss and above all a great person. Thank you for your knowledge, support and understanding. It has been a pleasure to work with you and I hope we will get to work together again in the future.

To all the amazing researchers that I've been lucky enough to collaborate with, Oliver Lequin, Astrid Walrant, Emanuelle Sachon, Edward Chalouhi, Leila Betchella, Françoise Illien, Sandrine Sagan, Rodolphe Vuilleumier, Ari Seitsonen and David Taresté. Special thanks to Oliver and Sandrine for their work as LBM directors. Additional thanks to Jean-Jacques Lacapère, Nicolas Rodriguez and Claire Lacombe for their patience helping me out with experimental assays.

To the other team members that I did not mention yet, especially Emeric Miclet and Ludovic Carlier, for the coffee conversations and advice.

To all the members of the LBM, researchers and students, who make such a pleasant work space that will be hard to leave. Special mention to our secretary Eliane Moulinié for her efficiency, dedication and kindness.

To the members of the jury, the "rapporteurs" James Sturgis and Claire Loison and "examinateurs" Jérôme Hénin, Nadège Jamin, Jacques Chomilier, thank you for your participation in this project. A special mention to Jacques, who I met during my Erasmus and who made me discover structural bioinformatics. You have helped me on numerous occasions since and without you I probably wouldn't be here today.

To Elodie Anxolabéhère-Mallart, for her participation in my theses committee.

To the CINES, for computational time.

To LytixBiopharma and the CEO Øystein Reckdal, for supplies and funding.

To my "collègues thésards" Rachel, Edward, Soha, Raphaëlle and Stephanie, your company and support have made my Ph.D a very pleasant journey. A special mention for Rachel, who I miss and admire. Also to Yadira, for her support and salsa lessons.

To all the friends I met here, Raquel, Mélanie, Adhel, Charlie, Guillaume and all the others, for the time spent together.

To Chloé, for her support, her affection and her efforts helping me out with my french these

last months.

To friends from Aranjuez, who even though we are far away, always make me feel at home when we meet again.

Finally, to my family, to mamá y papá, even during hard times they are there for me, to provide support and affection.

Abstract

This manuscript presents a study of the interaction between amphipathic helices (AH) and phospholipid lipid membranes. Membranes surround the cell and the different organelles and can have a variety of lipid compositions. Amphipathic helices (AH) are key protein motifs that are capable of interacting with lipid membranes. They are normally unfolded in water and they fold into an α -helix in the water / bilayer interface. This α -helix is amphipathic: it has a polar side exposed to the solvent and a hydrophobic side which residues are inserted within the membrane. The affinity of the AH for the membrane will depend on the membrane composition and structure (notably flat or curved). AH dynamics and their interaction with membranes is the main focus of this research work. Our predilection methods are computational techniques, particularly molecular dynamics (MD). The first objective of this Ph.D was to determine the best suited force field (FF) to study AH / membrane interactions. We also aimed at designing a novel strategy to study AH dynamics. Specifically, we combined temperature-replica exchange molecular dynamics (T-REMD) and Markov State Models (MSM) to extract the structural, kinetic and thermodynamic properties of AH action mechanism. We first worked with a well-known model system, the mastoparan from wasp venom. It allowed us to optimize our protocols to apply them to other more complex system later, specifically, to the study of the oncolytic peptide LTX-315, a promising strategy in cancer research, that is being developed by the company LytixBiopharma. During this Ph.D, we have also studied how membrane composition affect its physico-chemical properties and, in consequence, their interaction with peptides. To sum up, this Ph.D work represents an advance in MD research for AH / membrane systems and we obtained novel information about their action mechanism. Our work will be very useful for designing new AHs to be used in the biomedical and pharmaceutical domains.

Résumé en français:

Étude des interactions entre des hélices amphipathiques et des membranes lipidiques par des simulations computationnelles. Application au peptide anticancéreux LTX-315

Ce manuscrit contient mon travail de doctorat de 3 ans que j'ai effectué sous la supervision de Patrick Fuchs. Patrick et moi nous intéressons aux hélices amphipathiques (AH) et à leur interaction avec les membranes, plus particulièrement avec la bicouche lipidique. Dans cette introduction générale, je vais présenter les aspects biologiques les plus importants de cette recherche, les principales techniques employées et la structure du manuscrit, avec un aperçu général de chaque chapitre.

Les membranes cellulaires sont des bicouches lipidiques qui entourent la cellule et les différents organites. Elles protègent la cellule et la divisent en compartiments ayant des fonctions différentes. Elles sont composées principalement de lipides et de protéines. Elles contiennent trois types principaux de lipides : les phospholipides, les glycolipides et les stérols. Dans la majorité des cas, les phospholipides sont les plus abondants, par exemple, dans la membrane plasmique ils représentent 50% de tous les lipides. Les glycolipides ne sont pas très abondants ($< 2\%$) et les 48% restants sont des stérols [148]. Le cholestérol est le principal stérol présent dans les membranes cellulaires. D'autre part, deux types de protéines sont présentes dans les membranes : les protéines intégrales, qui traversent la bicouche, et les protéines périphériques, qui sont situées dans l'un des feuillet. Les différentes compositions et ratios lipides/protéines dépendent du type d'organite et de cellule et sont liés à leurs fonctions. Par exemple, la composition de la membrane plasmique est d'environ 50/50 lipides/protéines en poids. Les membranes sont très dynamiques. À part leur rôle passif et protecteur, elles participent activement aux fonctions de la cellule, comme le transport des solutés et les voies de signalisation. Les membranes sont essentielles à l'équilibre et à la survie de la cellule. Des connaissances fondamentales sur leur composition, leur structure et leur propriétés sont incluses dans cette introduction. Les bicouches lipidiques sont présentes dans tous les projets réalisés au cours de cette thèse.

Les hélices amphipathiques (AH) sont des motifs protéiques clés capables d'interagir avec les membranes lipidiques. Leur séquence est courte et elles sont normalement dépliées dans l'eau. Lorsqu'elles interagissent avec une membrane, elles s'insèrent dans l'interface membrane/eau en se repliant en une hélice α [222]. Cette hélice α est amphipathique : elle possède un côté polaire exposé au solvant et un côté hydrophobe dont les résidus sont insérés dans la membrane. L'affinité

de l'AH pour la membrane dépend de la composition et de la structure de cette dernière [10]. La dynamique des AH et leur interaction avec les membranes est l'objet principal de notre recherche. Nous avons travaillé avec différents peptides et compositions membranaires qui seront présentés plus tard dans ce chapitre.

Nos méthodes de prédilection sont les techniques computationnelles, en particulier la dynamique moléculaire (MD). La majorité des résultats de cette thèse sont obtenus à partir de simulations MD. Elles permettent de décrire le comportement d'un système moléculaire au cours du temps. La MD consiste à générer des simulations qui permettent de prédire la dynamique d'un système moléculaire dans de conditions précises. Elles peuvent fournir des informations détaillées sur un modèle moléculaire qui ne sont pas observables par des tests expérimentaux. Cependant, les informations obtenues par MD doivent être validées. En général, cette validation se fait par comparaison avec des données expérimentales. Les applications de cette technique sont nombreuses : étude de la dynamique des protéines et des membranes, interactions membrane/protéine, interactions protéine/protéine, etc. Dans ce manuscrit, les simulations MD sont utilisées pour étudier le comportement des AH et les interactions AH/membrane.

Le modèle moléculaire détermine le niveau de détail dans lequel le système moléculaire est décrit. Le modèle tout-atome (AA) est le plus détaillé, il représente une particule pour chaque atome de la molécule. Dans un modèle atomes-unis (UA), chaque atome lourd est représenté avec ses atomes d'hydrogène dans une seule particule (par exemple, une particule CH₃). Le modèle gros-grains (CG) représente une particule pour quelques atomes lourds proches (avec leurs atomes d'hydrogène correspondants) ; par exemple, dans le modèle MARTINI, 4 atomes lourds sont généralement fusionnés en une seule particule [146, 109]. Le choix du modèle dépend de différents facteurs comme le niveau de détail choisi. Si l'on veut étudier les interactions au niveau atomique, on choisit un modèle AA. Cependant, le modèle CG diminue le coût computationnel des simulations et permet d'observer des processus plus lents. Au cours de ma thèse, j'ai utilisé de préférence des modèles AA, bien que j'aie du employer des modèles CG lorsque nous voulions observer des échelles de temps plus longues.

La Temperature-Exchange molecular dynamics (T-REMD) est une technique d'échantillonnage augmenté qui permet d'observer de longs processus moléculaires dont les échelles de temps sont trop élevées pour être observées par la MD classique. En général, dans une simulation MD classique, une réplique est simulée à une température constante, normalement physiologique. En revanche, dans les simulations T-REMD, un grand nombre de répliques identiques sont simulées en même temps. Cependant, chaque réplique est lancée à une température différente. La température la plus basse est normalement la température physiologique et la température la plus haute peut être extrêmement élevée (par exemple, environ 600 K). Tout au long de la simulation, les répliques voisines peuvent échanger leurs températures. Les échanges entre les températures donnent au système une hausse d'énergie qui l'aide à surmonter les barrières énergétiques plus élevées. Pour qu'un échange de température ait lieu, un critère de Metropolis est appliqué afin que le système

suive une distribution de Boltzmann aux deux températures. De cette façon, nous assurons la fiabilité des conformations observées. Enfin, chaque réplique suit une marche aléatoire dans l'espace des températures permettant de franchir les barrières d'énergie libre à haute température. La T-REMD est utilisée dans plusieurs projets de ce manuscrit car elle permet d'explorer un ensemble conformationnel plus large du système d'intérêt que les simulations MD classiques. Cependant, les changements de température de chaque réplique biaisent la dynamique observée, il est donc très difficile de déduire la cinétique du système parce qu'elle est dépendante de la température.

Une autre méthode de calcul importante que nous allons présenter dans ce manuscrit sont les modèles d'état de Markov (MSM). Un MSM est un formalisme mathématique qui permet de décrire la dynamique d'un système en moléculaire [94]. Il est possible de construire un MSM à partir de simulations MD. L'objectif est de regrouper les conformations observées dans les simulations en grands clusters structurels. Ensuite, la probabilité d'observer chaque cluster est calculée, ainsi que les probabilités de transition entre les clusters. De cette façon, en construisant un MSM, il est possible de déchiffrer les propriétés structurales, cinétiques et thermodynamiques du système. Normalement, les dynamiques les plus intéressantes pour l'activité moléculaire sont les processus lents. Le MSM facilite leur extraction à partir de simulations de MD complexes. Dans ce manuscrit, nous présenterons l'application des MSM à l'étude des interactions AH/membrane.

Bien que ce doctorat soit principalement axé sur les techniques computationnelles, les techniques expérimentales sont également présentes. Elles sont primordiales pour avoir une compréhension approfondie des systèmes moléculaires et ont également un rôle clé dans la validation des résultats de MD. Ainsi, au cours de ma première année, j'ai effectué des expériences de fluorescence et de dichroïsme circulaire (CD).

La spectroscopie de fluorescence est une technique expérimentale qui permet d'étudier les propriétés des molécules et son environnement grâce à sa capacité d'émettre de la fluorescence. La fluorescence est un type de luminescence causé quand un faisceau de lumière excite les électrons des molécules d'intérêt et les amène à émettre de la lumière. Cette lumière est dirigée vers un filtre et sur un détecteur pour être mesurée, donnant l'information sur le système moléculaire d'intérêt.

Le CD est une technique essentielle utilisée pour analyser les caractéristiques structurales des macromolécules. C'est une technique utile pour déterminer la structure secondaire des protéines. Elle est basée sur la différence d'absorption de la lumière polarisée circulairement à gauche et à droite dans les substances optiquement actives.

Grâce à la collaboration de nos collègues de laboratoire, le Dr Olivier Lequin et son doctorant Edward Chalouhi, des essais de résonance magnétique nucléaire (RMN) sont également présentés. La RMN est une technique en biophysique qui se base sur les caractéristiques magnétiques de la matière. Elle fournit des informations structurales et dynamiques sur des systèmes moléculaires, en particulier des protéines. L'échantillon est exposé à un champ magnétique puissant, ainsi les données générées pendant les expériences sont collectées, traitées et analysées afin d'obtenir des informations sur les positions relatives des atomes de la molécule.

Ce projet de doctorat est axé sur l'étude des AH et de son interaction avec les membranes. D'une part, nous avons effectué une recherche approfondie sur le comportement des AH au niveau moléculaire et atomique. Nous avons étudié leurs propriétés structurales dans le solvant et dans les membranes pour mieux comprendre leur paysage conformationnel. Nous avons également cherché à comprendre leur mécanisme d'insertion dans la membrane et à extraire la cinétique et la thermodynamique des processus. D'autre part, nous avons étudié l'impact des propriétés physico-chimiques de la membrane sur l'affinité des AH, comme les charges électrostatiques ou les interactions hydrophobes. Enfin, nous avons voulu mieux comprendre l'importance des propriétés structurales de la membrane dans l'interaction avec les AH, comme le degré de courbure, la compacité et la présence de défauts de packing.

Au total, ce manuscrit est divisé en 7 chapitres différents. Dans cette introduction, je présenterai les systèmes moléculaires étudiés pendant mon doctorat, ainsi que l'état de l'art de la recherche et l'importance des techniques employées et leur évolution. Dans le chapitre Matériaux et Méthodes, on explique les techniques computationnelles et *in vitro*, ainsi que les méthodes d'analyse qui ont été employées. Il y a cinq chapitres de résultats, un pour chacun des projets que nous avons menés pendant ce doctorat. Dans chaque chapitre de résultats, des informations spécifiques sur les protocoles sont données ainsi que les résultats les plus importants de nos travaux. Un aperçu de chaque projet est donné ci-dessous.

Dans le premier chapitre, nous avons comparé différents champs de forces (FF) pour déterminer leur fiabilité à simuler des systèmes de AH/membrane. Le FF est l'ensemble des équations et paramètres utilisés dans la MD pour prédire le comportement du système. Aujourd'hui, différents FF sont disponibles pour simuler des lipides et des protéines : CHARMM36m, OPLS, GROMOS et AMBER99SB-ILDN/Berger combinés pour les protéines/lipides. Le choix du FF est une étape délicate car la fiabilité des résultats dépend de la bonne paramétrisation des molécules. Dans ce chapitre, nous avons effectué des simulations de T-REMD avec quatre FF différents pour deux systèmes moléculaires. L'objectif est de déterminer le FF le plus précis pour les simulations AH/membrane. Les simulations avec OPLS, GROMOS et AMBER99SB-ILDN/Berger ont été réalisées précédemment par Patrick et ses anciens étudiants. J'ai effectué des simulations en utilisant CHARMM36m FF pour les lipides et les protéines [123, 93]. L'objectif était d'observer les propriétés structurales et dynamiques de deux AH dans une bicouche lipidique et de comparer les résultats avec des expériences *in vitro*. Le FF produisant les résultats les plus fiables est utilisé pour le reste des projets de cette thèse. Les deux systèmes employés étaient ALPS et le mastoparan dans une bicouche DOPC/DOG 60/10. ALPS est un domaine protéique présent dans différentes protéines périphériques, nécessaire pour cibler la membrane et s'attacher à la bicouche lipidique. ALPS est un senseur de courbure, cela signifie que son affinité pour la bicouche dépend du degré de courbure de la membrane [10]. Le côté hydrophobe de l'ALPS présente de grandes chaînes latérales (résidus aromatiques). Afin de s'insérer dans la membrane, elles ont besoin de gros défauts packing. Dans les membranes fortement courbées, les têtes polaires des lipides sont plus éloignées, ce qui génère

de plus grands défauts de packing. C'est pourquoi ALPS est un senseur de courbure. Le mastoparan est un peptide amphipatique provenant du venin de guêpe. Il possède une courte séquence de 14 résidus. Son structure dans la membrane a été résolue par des expériences de RMN [91]. Il perturbe l'activité cellulaire en interférant avec la voie de signalisation médiée par la protéine G. Le mastoparan est un peptide simple de chaîne courte. P. Fuchs étudie son comportement depuis des années, c'est donc un bon modèle pour étudier les interactions AH/membrane. Le mastoparan est présent dans 3 chapitres de cette thèse.

Dans le second chapitre, nous avons étudié l'insertion du mastoparan dans la membrane par des simulations de MD. Dans le précédent, quatre FF différents ont été utilisés et les résultats ont été comparés aux données expérimentales. Nous avons déterminé que les deux FF les plus fiables étaient AMBER99SB-ILDN/Berger et CHARMM36m. L'objectif maintenant est de vérifier s'ils parviennent également à prédire correctement le mécanisme d'insertion et de repliement de l'AH dans la membrane. Dans ce but, des simulations de T-REMD ont été lancées pour les deux FF. Le peptide est initialement placé déplié dans le solvant, dans un système qui contient une bicouche DOPC/DOG 60/10. Les simulations permettent d'observer le processus d'insertion et plissement à l'échelle atomique. Les résultats ont été comparés aux données de RMN et de CD afin de déterminer la fiabilité des simulations. CHARMM36m FF a été sélectionné pour le reste des projets.

Dans le troisième chapitre, nous présentons une nouvelle stratégie de recherche. Un protocole combinant la T-REMD avec les MSM est conçu pour étudier les propriétés structurales, cinétiques et thermodynamiques des AH. Le mastoparan est utilisé comme système modèle et CHARMM36m comme FF car il a été considéré comme le plus fiable dans les deux chapitres précédents. L'objectif est d'établir un protocole solide qui pourrait être appliqué à d'autres systèmes par la suite. Deux MSM ont été construits, un pour le mastoparan dans l'eau et un second pour le mastoparan dans la membrane. Dans cette stratégie, nous utilisons d'abord la T-REMD pour explorer le paysage conformationnel du système. Un clustering structurel est effectué pour toutes les conformations trouvées à 300 K. Les clusters sont utilisés pour sélectionner des structures de départ pour lancer des simulations de MD classiques non biaisées à 300 K. Plusieurs simulations sont donc lancées pour chaque conformation avec différentes vitesses initiales. De cette façon, nous maximisons l'exploration du paysage énergétique du système. Ces simulations sont ensuite utilisées pour construire le MSM. Le MSM vise à expliquer les propriétés dynamiques du système moléculaire. Deux MSM satisfaisants ont été construits : le mastoparan dans l'eau et dans une bicouche DOPC/DOG 60/10. L'utilisation du mastoparan comme système modèle a permis de préparer et valider le protocole de cette nouvelle stratégie qui servira à étudier des systèmes moléculaires plus complexes dans le futur.

Dans le quatrième chapitre de résultats, nous avons étudié le peptide oncolytique LTX-315 et son interaction avec des membranes grâce à des techniques expérimentales et computationnelles [204]. Il s'agit d'une AH développé par LytixBiopharma pour traiter le sarcome, un type de tumeur maligne rares qui se développe aux dépens des tissus de soutien. Le LTX-315 a une séquence de

9 résidus avec un acide aminé non standard, la diphénylalanine (di-Phe), qui est similaire à la phénylalanine mais avec deux cycles aromatiques [86]. Il est capable d’interagir avec la membrane plasmique en pénétrant dans la cellule. Ensuite, il cible la mitochondrie et perturbe sa membrane en induisant la mort cellulaire. Il a une affinité préférentielle pour les cellules cancéreuses bien qu’il interagisse également avec les cellules normales. Il est actuellement en phase II d’essais cliniques en Europe et aux États-Unis. Cependant, nous manquons d’informations sur son mécanisme d’action d’un point de vue moléculaire. Simuler un acide aminé non standard est un défi pour la MD et une clé pour comprendre son comportement. En premier lieu nous avons réalisé des expériences de fluorescence et de CD afin d’obtenir des informations préliminaires sur ses propriétés structurales et de liaison. Nos collaborateurs, le Dr Olivier Lequin et son étudiant en thèse Edward Chalouhi, ont effectué des expériences de RMN pour déterminer sa structure secondaire, son orientation dans la membrane ainsi que la position relative des cycles aromatiques de la di-Phe. Enfin, des simulations T-REMD du peptide dans l’eau et dans une bicouche lipidique ont été réalisées. La paramétrisation de la di-Phe fut un réel défi. Elle a été réalisée de trois manières différentes. Tout d’abord, nous avons utilisé des outils de prédiction pour obtenir des valeurs approximatives pour ensuite lancer des simulations d’essai du système dans le vide. Enfin, nos collaborateurs Rodolphe Vuilleumier et Ari Seitsonen à l’ENS ont effectué des calculs de mécanique quantique (MQ) dans l’objectif de paramétrer la di-Phe. Au même moment, les développeurs de CHARMM FF ont publié de nouveaux paramètres officiels sur une multitude d’acides aminés non standard, dont la di-Phe [54]. Ces derniers ont été utilisés pour lancer nos simulations de T-REMD. La prochaine étape consistera à appliquer le protocole présenté dans le chapitre précédent et d’établir un modèle thermodynamique du LTX-315.

Le dernier chapitre présente une collaboration qui a eu lieu au sein de notre laboratoire et à laquelle j’ai pu participer : la pénétratine et son interaction avec les lipides PIP₂. La pénétratine est un membre de la famille des peptides pénétrants (CPP). Ce sont de courtes séquences d’acides aminés capables de pénétrer dans les cellules et de délivrer des molécules biologiquement actives. La pénétratine est un peptide de 16 acides aminés responsable de la translocation de l’homéodomaine Antennapedia [108]. PI(4,5)P₂ est un lipide anionique portant entre 3 et 5 charges négatives au pH physiologique [147]. Le PI(4,5)P₂ a des fonctions multiples telles que précurseur de second messager, régulateur de la polymérisation de l’actine, trafic membranaire, etc [118]. Il est principalement localisé dans le feuillet interne de la membrane plasmique, où il pourrait être impliquée dans la régulation de l’adhésion cellulaire et de la motilité [231]. Cette collaboration est dirigée par le Dr Astrid Walrant et le Dr Emmanuelle Sachon. Nous avons travaillé ensemble pour présenter un modèle structurel de la pénétratine dans la membrane et ses interactions spécifiques avec les lipides PIP₂. Le point fort de ce projet est la combinaison de techniques expérimentales et computationnelles en biologie et biophysique qui génèrent des résultats solides et une compréhension plus profonde de son comportement. Tous les résultats ne sont pas présentés dans ce manuscrit, seulement les résultats de MD et RMN. Nous avons effectué des simulations CG-MD de la pénétratine avec deux compositions de bicouches : POPC/POPS/PI(4,5)P₂ (80/15/5) et POPC/POPS

(80/20), et deux avec la même composition membranaire mais sans pénétratine comme systèmes contrôle. Un back-mapping de ces simulations a été fait pour avoir un modèle AA et des simulations AA-MD ont été lancées. Edward Chalouhi, doctorant du LBM et son directeur de thèse Olivier Lequin ont réalisé des expériences de RMN-H avec trois échantillons différents : peptide dans un tampon, peptide avec des micelles et peptide avec des bicelles. L'ensemble de ces résultats permet de comprendre la structure et l'orientation de la pénétratine dans la membrane et son affinité spécifique pour les lipides PIP₂. Une publication du projet est sur le point d'être soumise.

Cette thèse a été centrée dans l'étude des interactions AH/membranes. Nous avons commencé par une étude comparative de la capacité des différents FF de reproduire les propriétés structurales des AH dans la membrane. Les FF les plus fiables, CHARMM36m et AMBER99SB-ILDN/Berger, ont été sélectionnés pour réaliser un projet sur l'insertion du mastoparan dans la membrane, dont on a pu observer son mécanisme d'interaction avec des détails à niveau atomique. Ensuite, CHARMM36m a été sélectionné pour poursuivre les travaux de thèse, pour sa fiabilité et le grand ensemble de paramètres qu'il contient pour tous les types de biomolécules. Après cela, une nouvelle stratégie très prometteuse a été développée pour étudier le comportement des AH. Un protocole combinant la T-REMD et les MSM a été mis au point pour étudier la thermodynamique et la cinétique des AH. Deux modèles ont été générés : le mastoparan dans la membrane et le mastoparan dans le solvant. Après cette première étape de la thèse centrée dans la méthodologie, nous avons travaillé sur le sujet principal de ce doctorat, le peptide oncolytique LTX-315. Grâce à la combinaison des techniques computationnelles et expérimentales, nous avons réussi à proposer un premier modèle d'interaction entre le peptide et la bicouche lipidique, dont on a pu observer son interaction directe avec les lipides et son affinité pour les membranes chargées négativement. Enfin, nous avons utilisé nos compétences en MD pour participer à une collaboration au sein de notre laboratoire. On a étudié la penetratine, un peptide pénétrant capable d'interagir avec les bicouches lipidiques. Dans cette collaboration, nous avons approfondi nos connaissances sur son interaction avec les lipides, concrètement avec PIP₂. Ces travaux de thèse ont permis de participer à l'évolution des méthodes computationnelle utilisées dans l'étude d'interactions AH-membrane, et ont approfondi nos connaissances sur la dynamique des à niveau atomique.

Acronyms

- **AA** Amino Acid
- **AH** Amphipathic Helices
- **ALPS** Amphipathic Lipid Packing Sensor
- **AMP** Antimicrobial Peptide
- **APL** Area per lipid
- **Arf1** ADP-ribosylation factor 1
- **ArfGAP1** Arf GTPase-Activating Protein 1
- **ATP** Adenosine Triphosphate
- **BB** Backbone
- **CD** Circular Dichroism
- **CO** Cut Off
- **COM** Center of mass
- **COPI** Coat Protein I
- **COPII** Coat Protein II
- **CPP** Cell Penetrating Peptide
- **DAG** Diacylglycerol
- **DHPC** 1,2-diheptanoyl-sn-glycero-3-phosphocholine
- **DLS** Dynamic Light Scattering
- **DMPC** 1,2-dimyristoyl-sn-glycero-3-phosphocholine
- **DMPG** 1 2-dimyristoyl-sn-glycero-3-phosphoglycerol
- **DOG** 1,2-dioleoyl-sn-glycerol

- **DOPC** 1,2-dioleoyl-sn-glycero-3-phosphocholine
- **DPPC** 1,2-dipalmitoyl-sn-glycero-3-phosphocholine
- **DSSP** Define Secondary Structures of Proteins
- **EM** Electron microscopy
- **ER** Endoplasmic Reticulum
- **FF** Force Field
- **H-REMD** Hamiltonian-Replica Exchange Molecular Dynamics
- **LINCS** Linear Constraint Solver
- **LUV** Large Unilamellar Vesicle
- **MD** Molecular Dynamics
- **MLV** Multilamellar Vesicles
- **MRE** Mean Residue molar Ellipticity
- **MSM** Markov State Model
- **NMR** Nuclear Magnetic Resonance
- **PBC** Periodic Boundary Conditions
- **PC** phosphatidylcholine
- **PDB** Protein Data Bank
- **PIP** phosphoinositide
- **PME** Particle Mesh Ewald
- **PMF** Potential of Mean Force
- **POPC** 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine
- **POPS** 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphoserine
- **RMSD** Root Mean Square Deviation
- **SC** Side Chain
- **SDS** Sodium Dodecyl Sulfate
- **SS** Secondary structure
- **SUV** Small Unilamellar Vesicle
- **T-REMD** T-REMD Temperature-Replica Exchange Molecular Dynamics
- **US** Umbrella Sampling

Contents

1	Introduction	33
1.1	General introduction	33
1.2	The Eukaryotic cell and its membranes	38
1.3	Membranes	41
1.3.1	Historical context	41
1.3.2	Membrane composition	43
1.3.3	Membrane Structure and dynamics	48
1.3.4	Membrane systems in research	50
1.4	Amphipathic Helices	52
1.5	AH/membrane interactions studied by MD	64
2	Materials and Methods	71
2.1	Computational Methods	71
2.1.1	Molecular Dynamics	71
2.1.2	MD Analysis	83
2.1.3	Temperature-Replica Exchange Molecular Dynamics	90
2.1.4	Markov State Models	93
2.2	Experimental methods	104
2.2.1	Circular Dichroism	104
2.2.2	Fluorescence	106
2.2.3	NMR	107
3	Force Field Comparison for Amphipathic Helices	111
3.1	Introduction	111

3.2	Materials and Methods	116
3.2.1	Molecular Systems	116
3.2.2	Force fields	116
3.2.3	REMD simulations	116
3.2.4	MD simulations of the membrane only	117
3.3	Experimental information	117
3.4	Computational results	120
3.5	Membrane properties in CHARMM36m simulations	123
3.5.1	Area Per Lipid	124
3.6	Discussion	125
4	Computational study on mastoparan insertion in the membrane	129
4.1	Introduction	129
4.2	Materials and Methods	131
4.2.1	Simulations in water	131
4.2.2	Simulations in the membrane	131
4.3	Mastoparan in water	132
4.3.1	Helicity content	132
4.3.2	Structural clustering	133
4.4	Mastoparan partitioning into the membrane	135
4.4.1	Insertion in the membrane	135
4.4.2	Partitioning/Folding coupled mechanism	136
4.4.3	Helix content	137
4.4.4	Orientation in the membrane	138
4.5	Discussion	139
5	Mastoparan folding in the membrane: Markov State Models to decipher the mechanism	143
5.1	Introduction	143
5.2	Mastoparan in the membrane	146
5.2.1	Simulation details	146
5.2.2	Model Construction	147

5.3	Mastoparan in water MSM	152
5.3.1	Simulation details	152
5.3.2	Model Construction	154
5.4	Constructing a model without TICA	157
5.5	Mastoparan free energy landscapes comparison	157
5.6	Discussion	158
6	Multidisciplinary study on LTX-315 interaction with membrane	163
6.1	Introduction	163
6.2	Materials and Methods	165
6.2.1	Liposome preparation	165
6.2.2	Lipid quantification: the Rouser Technique	166
6.2.3	Peptide quantification method	167
6.2.4	Tryptophan-Fluorescence	167
6.2.5	Circular Dichroism	168
6.2.6	NMR	169
6.2.7	Quantification assays by fluorimetry	169
6.2.8	Molecular Dynamics	171
6.3	Experimental Results	172
6.3.1	LTX-315 interacts with lipid membranes	172
6.3.2	Structural properties	174
6.3.3	Orientation in the membrane	177
6.4	Computational results	178
6.5	Penetration into the cell	181
6.6	Discussion	183
7	Involvement of PIP₂ in the internalization of the cell-penetrating peptide Penetratin	187
7.1	Penetratin	187
7.1.1	Introduction	187
7.1.2	Materials and Methods	189
7.1.3	Membrane Structural properties	190

7.1.4	Penetratin structure and partitioning into the membrane	192
7.1.5	PIP ₂ recruitment	195
7.1.6	Discussion	197
8	General conclusions	201
	Bibliography	204

List of Figures

1.1	Eukaryotic cell diagram. Image from: https://www.w3spoint.com/	40
1.2	Relevant membrane models in history.	42
1.3	Modern cell membrane model. By LadyofHats Mariana Ruiz - Own work. Image renamed from File:Cell membrane detailed diagram.svg, Public Domain, https://commons.wikimedia.org/w/index.php?curid=6027169	43
1.4	Lipid diversity in the Eukaryotic cell. The barplots indicate the lipid composition of each organelle in mammals (dark blue) and yeast (light blue). Cholesterol/phospholipid and ergosterol/phospholipid ratio is indicated for mammals and yeast respectively. Blue circles mark synthesis sites and red circles the lipid related to signaling pathways (less than the 1%) Image from [148]	44
1.5	Aliphatic chains with an insaturation in (A) <i>trans</i> and (B) <i>cis</i> . Image from Wikipedia.	45
1.6	. Phospholipid schematic representation and bilayer formation by its aggregation. Image from: https://www.creative-proteomics.com	46
1.7	Cholesterol's structure	47
1.8	Peripheral proteins' mechanisms to bind lipid membranes. Image [10]	48
1.9	Lipid shapes and the structures they can form. (a) Inverted conical lipids can aggregate in micelles, (b) cylindrical lipids form bilayers and (c) conical lipids can generate reverse hexagonal phase. Image from: [102]	49
1.10	Packing defects' schematic representation. The red arrows point to the defects. Image by Patrick Fuchs.	50
1.11	Phase transition in lipid bilayers. As the temperature increases, the membrane "melts", going from an ordered solid-like bilayer (left) to a disordered, fluid one (right). Around the melting temperature T_c , the two phases coexist. Image from [202].	50

1.12	Micelle and bicelle structure. In orange and purple, long chain lipids of different nature. In blue, detergent molecules. Bicelles are formed by a central planar bilayer composed by long chain lipids with the edges stabilized by either short-chain lipids or detergents. Different lipids can be used. The q value represents the extension of the bilayer. The larger q , the larger the bilayer. Micelles can be described as a bicelle with a q value equal to 0. Image from [45]	51
1.13	First helical wheel ever published. It corresponds to one segment of the whale sperm myoglobin, described by M. Schiffer in 1967 [191]. The residue number is written next to each amino acid and the hydrophobic amino acids are surrounded by a circle.	53
1.14	ALPS' sensitivity for membrane curvature. Image from [10]	55
1.15	ALPS and α -synucleine structure and membrane affinity. ALPS insertion is promoted by hydrophobic interactions because it presents big hydrophobic lateral chains while α -synucleine insertion is lead by electrostatic interactions due to its positive charges. Image from [10]	56
1.16	Proposed AMPs' insertion mechanisms. (a) Barrel-stave pore: peptides insert perpendicularly in the bilayer and parallel to the aliphatic chains and form a pore. (b) Carpet mechanism: peptides adsorb parallel to the bilayer, they aggregate and produce a detergent-like effect that disrupts the membrane. Specific peptide-peptide interactions are not required. (c) Toroidal pore: peptides insert perpendicularly in the bilayer, induce a local membrane curvature and the pore lumen is lined partly by peptides and partly by phospholipid headgroups. A continuity between inner and outer leaflets is established. (d) Disordered toroidal pore: modification to the toroidal pore where the peptide conformations and orientations are less-rigid. The pore lumen is lined by the phospholipid head groups. Image from [149]	57
1.17	Examples of internalization mechanisms of CPPs and its cargo.(A) Spontaneous membrane translocation across the plasma membrane. The membrane is not disrupted. (B) Transient plasma membrane permeabilization. (C) Endocytosis of membrane-bound peptide-cargo complex, along with unattached small and large molecule cargoes. (D) Endosomal membrane lysis and CPP-cargo and co-encapsulated cargo release. (E) Translocation across the endosomal membrane delivers CPP and attached cargo, but not co-encapsulated cargo. (F) CPP and cargo are rapidly degraded if release does not take place. The mechanisms depicted are not mutually exclusive; they can happen concurrently. Image from [119]	58
1.18	Amphipathic helix/membrane interaction phases described by Seelig [195].	60

1.19	Ac-18A-NH2 partitioning in a DOPC bilayer. The first image is the peptide helical wheel. The colors correspond to the ones established in Wimley and White's hydrophobicity scale (see next section). The dotted line shows the helix central axis that separates the polar and non-polar sides. The bottom figure is the electron density profile where the density of the main elements of the system are plotted on the z (vertical) axis of the membrane. This figure shows that the peptide is located at the glycerol level. Image from [92]	61
1.20	Hydrophobicity scale determined by Wimley and White in 1996 [223].	62
1.21	Thermodynamics of an Amphipathic Helix partition and folding by S. White [222]	63
1.22	Thermodynamic cycle of AH partitioning and folding in the membrane [3]. (A) Unfolded structure in water. (B) Unfolded structure in the membrane. (C) Ensemble of folded and unfolded structures in water. (D) Folded structure in the membrane. States A and B are transition conformations not stable enough to be observed by experimental assays. ΔG_{AB} is computed using Wimley and White Hydrophobic scale. The other free-energy differences are determined experimentally.	64
1.23	All-atom and coarse-grained model comparison. First, two lipid structures are showed: AA DMPC and CG DPPC. The fd phage coat protein is also shown in AT and CG format. In AA, the colors are relative to the atom type. In the CG models the particles are coloured according to the following scheme: green, mixed polar/non-polar bead; cyan, hydrophobic bead; red/blue, negative/positive, charged bead; and pink, polar bead. Image from [32]	66
1.24	Snapshots of WALPS simulations at 80 and 200°C. On the left, results for W16 peptide. On the right, results for W23 peptide. These simulations show the acceleration of the processes when the temperature increases. Bigger time scales are observed at high temperatures. Image adapted from [212]	68
1.25	Study of AH sensing membrane curvature by metadynamics simulations. The number of hydrogen bonds and the Alpha-beta similarity were chosen as reaction coordinates [55].	69
2.1	Spatio-temporal resolution of various biophysical techniques. MD time scales are compared to those from experimental techniques. Time scales achievable by MD simulations are much smaller than the ones from experimental assays. Image from [64]	72
2.2	Molecular Models used in MD simulations. Credit image: http://2015.igem.org/ .	72
2.3	Force Field terms illustration.	73

2.4	Cut-off application for long-range electrostatic interactions. Shift (red) progressively modifies the potential to arrive to a null value at the CO distance. Switch (blue) establish a cut-on distance to make the interruption of the potential smoother. PME (green) is a way to take into account the neglected part beyond the CO.	75
2.5	Leap Frog algorithm. Image adapted from: http://www.physics.drexel.edu	76
2.6	Periodic Boundary Conditions (PBCs). The pink particle moving to the neighbour box on the left will re-enter the box by the right side. Image from [81]	80
2.7	Illustrative scheme of the different phases of an MD protocol.	80
2.8	Box shapes: cubic on the left and rhombic dodecahedron on the right. Credit images: CC BY-SA 3.0, https://commons.wikimedia.org/w/index.php?curid=613268 ; CC BY-SA 3.0, https://commons.wikimedia.org/w/index.php?curid=38689	81
2.9	Energy minimization important concepts. (A) Example of a 2D energy landscape with four local minima and one global minimum. Credit image: https://vitalflux.com (B) Line Search example (C) Steepest Descent example. Credit image: http://trond.hjorteland.com	82
2.10	Euclidean voronoi diagram by Balu Ertl - Own work, CC BY-SA 4.0, https://commons.wikimedia.org/w/index.php?curid=1111111	
2.11	Model APL graph: APL values are plotted against the simulation time. The data was obtained from a pure POPC simulation with CHARMM36 FF at 303,15 K . . .	85
2.12	Model density graph. The densities in the z axis are plotted for each molecule type. On the left, the two lipid types are plotted separately (gray and red). The water is represented in blue and the ions in yellow. On the right, the different lipid types were merged in a single group (black) to visualize the water/membrane interface. The data was obtained from a POPC/POPS 50/50 simulation with CHARMM36 FF at 303,15 K	86
2.13	(A) In red: reference particle. In purple: particles at a distance r from the reference that are taken into account to calculate $\rho(r)$. In yellow: particles at a distance other than r from the reference. Both yellow and purple particles will be taken into account to calculate $\rho(bulk)$. The brown circle represents the density shell. (B) Model graph from RDF analysis. If the peak is higher than 1.0 means that the particle density of the shell is higher than the density of the bulk. Credit images: https://commons.wikimedia.org/w/index.php?curid=1111111	87
2.14	Example of partitioning plot taken from the RW16 CPP (taken from Jobin et al. [107]).	89
2.15	Illustrative image of the T-REMD technique. A number n of replicas is simulated. During trajectory calculation there are several attempts of temperature exchanging that can be accepted as the one between replica 0 and replica 1, or rejected as the one between replica 3 and replica 4.	92

2.16	Potential energy distributions of the molecular system at each temperature. Each curve belongs to a temperature used in the REMD simulations, from the bottom temperature (300 K, curve on the left) to the upper temperature (400K, curve on the right)	93
2.17	Example of a simple MSM process with two states: A and E. Each number represents the probability of changing from one state to another. Credit image: By Joxemai4 - Own work, CC BY-SA 3.0, https://commons.wikimedia.org	93
2.18	MSM principle at a glance. Image obtained from [95]	95
2.19	MSM clusterization process	96
2.20	MSM protocol: featurization and dimensionality reduction with TICA.(a) Pentapeptide 3D structure. (b) VAMP2 scores of the backbone torsions, positions and distances at a lag time of 0.5 <i>ns</i> .(c) Pentapeptide conformational landscape projected onto the first two independant components (ICs). The color is related to the structure free energy. (d) Values of the two first ICs throughout the trajectory. Images obtained from [221]	97
2.21	MSM protocol: discretization. (a) VAMP2 Score analysis of the number of microstates. The score is saturated at 75 states. (b) Trajectory conformations projected onto the two first ICs. The darkness of the color is proportional to the density of the region. The centroids of the microstates are marked in orange. Image obtained from [221]	99
2.22	MSM protocol: model validation.(a) ITS validation through convergence at increasing lag times. (b) A Passing Chapman-Kolmogorov (CK) test. When doing a CK test, PyEMMA automatically estimates a new MSM transition matrix at lag time $k\tau$ and propagates the original transition matrix by the k -th power. In the image, the diagonal values of these two matrices are plotted for 5 different values of k . 1->1 is the transition from macrocluster 1 to macrocluster 1 and so on. As the estimated and the predicted curve coincide, the test is passed. Image obtained from [221]	100
2.23	PCCA algorithm. r_2 , r_3 and r_4 are the second, third and fourth dominant eigenvectors respectively. Image adapted from: http://docs.markovmodel.org/	102
2.24	MSM Analysis. (a) Energy landscape. (b) Coarsed-grain model into 5 macrostates. (c) Second right eigenvector of the model . It shows the second slowest process of the model. The first slowest process is not shown as it correspond to the system total dynamics. (d) TPT: how the system transitions from state 2 to state 4. Image obtained from [221]	103

2.25	MSM resulting model. An ensemble of representative structures are shown for each macrostate. The transition path to evolve from state 2 to state 4 is marked with the arrows. The width of the arrow is proportional to the probability of the state to change. The probability is written in float numbers over the corresponding arrow. Image adapted from [221].	104
2.26	CD principle. The PPW is drawn in cyan and its to components, the right CPW and the left CPL are drawn in red and green respectively. The image on the left shows the effect on the three waves of traversing a material (orange box) with CD. On the right, a representative scheme of the three vectors and the resulting ellipticity.	105
2.27	Model of a CD spectra applicable to peptides and proteins: in yellow, the resulting CD curve of an α -helix, in blue the CD curve of a β -sheet and in red the CD curve of a random coil. Image adapted from: https://www.creative-proteomics.com/ . . .	106
2.28	Illustrative image of the fluorescence phenomenon	107
2.29	NMR Spectroscopy. (A) NMR principle. A magnetic field is applied to a hydrogen nucleus. The magnetic moment of the nucleus is aligned with the magnetic field in an α -spin state. The spin of the nuclei presents a precession motion. (B) NMR spectrum model. The number of chemical shifts (peaks) is equal to the number of equivalent proton groups. The signal integration is proportional to the number of protons in the group. The multiplicity of peaks within each massif is due to the spin-spin scalar coupling and it is related to through-bond interactions between nuclei in the molecule. Image adapted from: https://www.jeol.co.jp	109
3.1	ALPS23 and mastoparan helical wheels predicted with heliquist [77].	115
3.2	T-REMD initial structures for (a) ALPS23 and (b) mastoparan systems. The images show a front and an upper view. Peptides are shown in purple, DOPC lipids in blue and DOG in green. DOPC phosphor atoms are represented with a bead. The images are extracted from CHARMM36m simulations.	117
3.3	NMR resolved structures.	118
3.4	CD assays of the mastoparan in DOPC/DOG 85/15 LUVs. (a) CD spectra at different lipid compositions. (b) Helicity plot with respect to the lipid/peptide ratio in terms of concentration. Unpublished data.	118
3.5	ALPS23 CD spectra in DOPC/DOG 85/15 liposomes. Lipid concentration increases from 0 to 7.5 mM.	119
3.6	ALPS35 CD assays in SDS micelles. First, an scheme of SUV and LUV relative size is shown. (a) CD spectra of ALPS35 in buffer and pure DOPC SUVs. (b) CD spectra of ALPS35 in buffer, in pure DOPC LUVs and DOPC/DOG LUVs at 85/15 molar ratio. Image from [216]	120

3.7	Secondary structure analysis. The peptide secondary structure is plotted through time by residue. For these analyses, T-REMD conformations at 300K are used. We do not follow a continuous replica over the temperatures. On the left, results for the mastoparan; on the right, results for ALPS23. Analysis for the four FF are shown: CHARMM36m, AMBER99SB-ILDN, OPLS and GROMOS54A7. Color white is for random coil, black for β -bridges, green for bends, yellow for turns, blue for α -helices and grey for 3_{10} -helices.	122
3.8	(a) Mastoparan and (b) ALPS23 helicity per temperature. The conformations found in CHARMM36m (red) and AMBER99SB-ILDN (blue) simulations were classified by temperature. The mean helicity values are given in the plots with points, the lines represent the standard deviations.	124
3.9	Convergence of ALPS23 helicity with AMBER99SB-ILDN at the highest temperature (400 K). Each black symbol represents the helicity for each frame at 400 K. Because it can only take discrete values, it is hard to judge of its relative convergence. We thus added a smoothed curve in red using the lowess() function in R with a smoother span of 0.01.	128
4.1	Molecular system at time 0 <i>ns</i> for (a) CHARMM36m and (b) AMBER99SB-ILDN T-REMD insertion simulations. For CHARMM36m simulations, two different conformations alternate between replicas.	132
4.2	T-REMD in water.(a) Initial structures used to launch the REMD simulations: the folded structure was employed with both FF. The unfolded one was only used with AMBER99SB-ILDN FF. (b) Mastoparan helicity content through time. The ensemble of conformations at 300 K is analyzed, no replica is followed over the temperatures. On the left, results for the CHARMM36m simulation; on the right, results for the two AMBER99SB-ILDN simulations. In the three simulations, the system takes several <i>ns</i> to get equilibrated. Only the fraction of the simulations where the system is equilibrated is used for the analysis and it is shown with the black dotted line.	133
4.3	Mastoparan in water most populated conformations found in CHARMM36m (blue) and AMBER99SB-ILDN (green) simulations.	134
4.4	Illustrative example of the partitioning / folding coupled mechanism. The information is extracted from one of the CHARMM36m T-REMD replicas. The plot shows mastoparan COM <i>z</i> coordinate through time (red), with respect to the mean <i>z</i> position of the phosphorous atoms of the upper layer (black). In the same graph, the evolution of the helicity through time is plotted (green).	137

4.5	Mastoparan's helicity per residue in CHARMM36m and AMBER99SB-ILDN REMD insertion simulations. AMBER99SB-ILDN (pink) predicts an α -helix from residue LEU3 to residue ALA10 approximately. CHARMM36m (blue) predicts an α -helix from residue ASN2 to residue LYS12	138
4.6	Analysis of mastoparan partitioning in the membrane. (a) Helical wheel model of the mastoparan. (b) Box plot showing the partitioning per residue. The black and green lines mark the phosphorous and glycerol levels respectively. (c) ASN2, ALA8 and LEU14 Z position through time in one of the REMD replicas.	139
5.1	Branching reaction pathway for vesicle fusion from an unfused starting state (a) to the fully fused state (d). Two paths are described: (I) canonical pathway through a stalk-like early intermediate (b) and a hemifused late intermediate (c) and (II) rapid fusion with just the stalk-like intermediate.	144
5.2	Late steps of barnase-barstar binding to loosely bound state. Two possible pathways are presented depending on the interacting side of barstar. The major pathway (b - d - e) represents the 95 % of the flux where barnase Arg 59 'anchors' into barstar via Asp 35, Glu 76 and Trp 38.	145
5.3	REMD simulations of mastoparan partitioning in the membrane	147
5.4	On the left, the histograms of the two IC obtained after TICA dimensionality reduction. On the right, the conformational landscape projected on the two ICs. The color is related to the structure's free energy.	148
5.5	Trajectory conformations projected on the two ICs. The darkness of the color is proportional to the density of the region. The structures were divided in 30 microstates. The centroids are marked in black.	149
5.6	Model validation	150
5.7	eigenvectors comparison throughout the ITS. On the left, direct ITS comparison. On the right, ITS differences between consecutive vectors.	151
5.8	Trajectory conformations projected on the two ICs. The color is related to the value of the conformation on the first eigenvector.	151
5.9	. Fuzzy assignment. On the left, probability of the conformations of belonging to the first macrostate. On the right, probability of the conformation of belonging to the second macrostate.	152
5.10	Discrete assignment. Trajectory conformations are projected on the two ICs. The model was coarse-grained into 2 macrostates. The purple zone marks the first macrostate and the yellow zone the second one.	152
5.11	MSM of the mastoparan in the membrane	153

5.12	IC values through time for the first MD simulation. Two different ensembles of conformations are easily distinguished (bimodal distribution).	154
5.13	Model validation	155
5.14	Energy landscape. Conformations are projected on the single IC of the model. Two macrostates were created: state 1 in yellow and state 2 in purple.	156
5.15	MSM of the mastoparan in water	156
5.16	MSM of the mastoparan in membrane using the RMSD and the Rgyr instead of TICA ICs. (a) Histograms of features. (b) Energy landscape. (c) ITS validation. (d) CK test. (e) Conformational landscape with the center of microclusters plotted as black dots. The color is related to the density. (f) Conformational landscape divided in two macrostates, purple and yellow.	161
5.17	Ten representative structures of set 1 (purple) and set 2 (yellow) obtained from the model constructed without applying TICA.	162
5.18	Landscape comparison. Mastoparan landscape in the membrane (left) is compared to the landscape in water (right). The ICs are shared between the two systems. The conformations of the trajectories are plotted onto the ICs.	162
6.1	Amino acid sequence of the oncolytic peptide LTX-315. The sequence is mostly composed by Lysines and Tryptophans. Notice the presence of a non-standard amino acid, the Di-Phenylalanine in position 8. Image from [201]	164
6.2	Mitochondrial Outer Membrane permeabilization (MOMP). LTX-315 penetrates into the cell and targets the mitochondrion. The disruption of the mitochondrial outer membrane induces necrosis. Image from [201]	165
6.3	LTX-315 putative structure as an amphipathic α -helix. (A) Helical wheel scheme. (B) 3D representation of the secondary structure. Image adapted from [86]	166
6.4	Pacific Blue probe structure scheme. This probe was attached to the LTX-315 C-terminus to carry out quantification assays by fluorimetry. Image by Ed (Edgar181) - Own work, Public Domain, https://commons.wikimedia.org/w/index.php?curid=66991761170	
6.5	Protocol of the quantification by fluorimetry of the internalization of LTX-315 peptide.	171
6.6	Dip structure generated with MolView. Image generated with pymol.	172
6.7	Initial structure for LTX-315 T-REMD simulations of insertion. The peptide backbone is represented in pink cartoon, DMPC lipids in silver lines and DMPS lipids in blue lines. The phosphorous atoms of the polar heads are shown in spheres. . .	173

6.8	Fluorescence spectra of LTX-315 in absence or presence of LUVs of (a) POPC, (b) POPS and (c) at different concentrations (from 200 to 4000 μM). Each cross represents the maximum of the spectrum. The dotted line is pure buffer, the red line corresponds to the peptide in buffer, then each other color line shows the peptide with lipids at various concentrations. The values in the legends indicates the lipid concentration in units of μM	174
6.9	CD spectra of LTX315 in the UV region in absence or presence of POPC, POPS and POPC/POPS (50:50) LUVs. An smoothed curve of each spectra is shown to ease the view.	175
6.10	CSD of LTX-315 C_{α} nuclei in buffer, micelles or bicelles. CSD was not calculated for Dip8 residue.	176
6.11	Difference between LTX-315 chemical shifts in buffer and in (a) micelles or (b) bicelles.	177
6.12	177
6.13	Residual intensity calculated from the paramagnetic relaxation enhancement experiments in anionic lipids. In black, results for the backbone; in blue, results for the side chains.	178
6.14	LTX-315 T-REMD simulations in water. Only conformations at 300 K are considered.	179
6.15	LTX-315 insertion T-REMD simulation. Only conformations at 300 K are considered.	180
6.16	z positions of (a) LTX-315 COM (pink) and (b) Dip SC COM (blue). Phosphorous atoms mean position is plotted in black. For this analysis, T-REMD conformations at 300K are used. We do not follow a continuous replica over the temperatures. . .	181
6.17	Snapshot of LTX-315 T-REMD simulation of insertion at 144.5 ns and 300 K. Dip is the most buried residue.	181
6.18	Internalization assays	182
7.1	NMR results.	193
7.2	Study of Penetratin's orientation in the bilayer by computational techniques. . . .	194
7.3	Penetratin/bilayer interactions studied by AA-MD simulations.	195
7.4	Snapshots of the MD simulation (top view of the bilayer). Penetratin is showed in purple spheres, PC in blue, PS in magenta and PI(4,5)P ₂ in green.	196
7.5	Study on penetratin effect on PI(4,5) ₂ relative position in the membrane by CG-MD simulations.	197

List of Tables

1.1	Diversity of polar head's structures present in phospholipids	44
1.2	Most common aliphatic chains present in phospholipids.	45
3.1	Helix content for mastoparan and ALPS23 peptide in the membrane simulated with different FF. Only the conformations found at 300K are taken into account.	123
3.2	Area Per Lipid values for DOPC and DOG in MD and T-REMD simulations. In T-REMD simulations, the mastoparan is present, inserted in the membrane/water interface.	125
3.3	Thickness values for DOPC and DOG in MD and T-REMD simulations. In T-REMD simulations, the mastoparan is present, inserted in the membrane/water interface	125
4.1	Mastoparan mean helicity in water. Information extracted from "mastoparan in water" T-REMD conformations at 300 K.	134
4.2	Statistics of mastoparan partitioning times in the membrane. The position of the z coordinate of each α -carbon is taken for the analysis. Results for two different FF (CHARMM36m and AMBER99SB-ILDN) are compared.	136
6.1	Ensemble of samples measured by fluorescence. The table contains the lipid and peptide concentrations used and the lipid:peptide ratios. These 6 different samples were prepared for each of the LUVs compositions (POPC; POPS; 50/50 POPC/POPS). The samples corresponding to the buffer and the peptide in buffer were measured once.	168
6.2	Summary of the samples analyzed by Circular Dichroism.	168
6.3	Calculation of the quantity of LTX-315 internalized by translocation and endocytosis	183
7.1	Membrane APL values calculated from CG-MD simulations.	191
7.2	Membrane thickness values calculated from CG-MD simulations.	191

Chapter 1

Introduction

1.1 General introduction

This manuscript contains my 3 years Ph.D work that I carried out under Patrick Fuchs' supervision. Patrick and I are interested in amphipathic helices (AH) and their interaction with membranes, specifically with the lipid bilayer. In this general introduction, I will present the most important biological aspects of this research, the main techniques employed and the structure of the manuscript, with a general overview of each chapter.

Cell membranes are lipid bilayers that surround the cell and the different organelles. They protect the cell and divide it in compartments with different functions. They are composed mainly by lipids and proteins. Three main lipid types are present: phospholipids, glycolipids, and sterols. In the majority of cases, phospholipids are the most abundant, for example they generally contribute for over 50% of all lipids in the plasma membrane. Glycolipids are not very abundant ($< 2\%$) and the remaining 48 % are sterols [148]. Cholesterol is the main sterol present in cell membranes. On the other hand, there are two protein types present in membranes: integral proteins, which cross the bilayer; and peripheral proteins, which are located in one of the leaflets. The different lipid and protein compositions and ratios depend on the organelle and cell type and is related to their functions. For example, the composition of the plasma membrane is about half lipids and half proteins by weigh [29]. Membranes are highly dynamic. Apart from the passive, protective role, they actively participate to cell's functions as solute transport and in signaling pathways. Membranes are essential to cell's equilibrium and survival. Fundamental knowledge about membrane composition, structure and properties are included in this introduction and lipid bilayers are present in all the projects carried out during this Ph.D thesis.

Amphipathic helices (AH) are key protein motifs that are capable of interacting with lipid membranes. Their sequence is short and they are normally unfolded in water. When they interact with membranes, they insert into the membrane/water interface while folding into an α -helix [222]. This α -helix is amphipathic: it has a polar side exposed to the solvent and a hydrophobic side

which residues are inserted within the membrane. The affinity of the AH for the membrane will depend on the membrane composition and structure [10]. AH dynamics and their interaction with membranes is the main focus of this Ph.D research. We have worked with different peptides and membrane compositions that will be presented later on this chapter.

Our predilection methods are computational techniques, particularly molecular dynamics (MD). The major part of the results of this work are obtained from MD simulations. They allow to describe the behaviour of a molecular system through time. MD consist in computer simulations which allow the prediction of the dynamics of a molecular system within certain conditions. They can provide detailed information about a molecular model not observable with experimental assays. However, the information obtained computationally must be validated. In general, this validation is done by comparison to experimental data. The number of applications of this technique are countless: study of protein or membrane dynamics, membrane/protein interactions, protein/protein interactions, etc. In this manuscript, MD simulations are used to study AH behaviour and AH/membrane interactions.

The molecular model determines the degree of detail in which the molecular system is described. The all-atom (AA) model is the more detailed one, it represents one particle for each atom of the molecule. In a united-atom model, each heavy atom is represented with its apolar hydrogen atoms in a single particle (e.g. a CH₃ particle). A coarsed-grained (CG) model represents one bead for a few heavy nearby atoms (together with their corresponding hydrogen atoms); for example, in the MARTINI model, 4 heavy atoms are generally merged into a single bead [146, 109]. The choice of the model depends on different factors as the degree of detail needed. If we aim to study interactions at atomic level, an AA model is chosen. However, coarse-graining the model decreases the computational cost of the simulations and allows to observe slower processes and larger systems. During my Ph.D I preferentially used AA models although I employed CG models when we wanted to observe longer time scales.

One central technique that we used in different subprojects of myt thesis is Temperature replica-exchange molecular dynamics (T-REMD) [151]. T-REMD is an enhanced sampling technique that allows to reduce the computational cost of simulating long molecular processes whose time scales are too elevated to be observed by classical MD. Generally, in a classical MD simulation one replica is simulated at a constant physiological temperature. However, in T-REMD simulations, a high number of identical replicas are simulated at the same time. However, each replica is launched at a different temperature. The bottom temperature is normally the physiological one and the upper temperature can be extremely high (e.g. around 600 K). Throughout the simulation, neighbour replicas can exchange temperatures. The exchanges between the temperatures give the system boost of energy that help to overcome higher energetic barriers. For a temperature exchange to take place, a Metropolis criterion is used so that the system follows a Boltzmann distribution at both temperatures. This way we ensure the reliability of the conformations observed. Finally, each replica follows a random walk in temperature space allowing to cross free energy barriers at

high temperatures. T-REMD is used in several projects of this manuscript as it helps to explore a wider conformational ensemble of the system of interest than classical MD simulations. However, the changes in the temperature of each replica bias the observed dynamics, so it is very difficult to infer the kinetics of the system (i.e. the kinetics depends on temperature).

Another important computational method that we are going to present in this manuscript are the Markov State Models (MSM). A MSM is a mathematical formalism that allows to describe the dynamics of a changing system [94]. It is possible to construct an MSM from MD simulations. The aim is to group the conformations observed in the simulations in structural clusters. Then, the probability of observing each cluster is calculated, and also the probabilities of transitioning between the clusters. This way, by constructing an MSM it is possible to decipher the system structural, kinetic and thermodynamic properties. Normally, the most interesting dynamics for molecular activity are the slow processes. MSM facilitates their extraction from complex MD simulations. In this manuscript, we present the novel application of MSM to study AH/membrane interactions.

Although we mainly use computational techniques, experimental assays are also present. They are very important to have a deep understanding of the molecular systems and they also have a key role validating the computational results. During my first year, I carried out fluorescence and circular dichroism (CD) assays. Also, thanks to collaborations, I will present some results from NMR experiments as well as internalization assays performed by some colleagues from our lab.

This Ph.D project is based on the study of AH and their interactions with membranes. First, we focus on the methodology. The first objective is to determine the best suited FF to study AH / membrane interactions. We also aim to design and test novel protocols to study AH dynamics. This way, we will have optimized computational techniques to study our systems of interest. From a biological point of view, we would like to address several questions. First, we want to have a better picture of AH conformational landscape in buffer and in the membrane. We also aim to understand their mechanism of insertion into the membrane at atomic level and to extract the kinetics and thermodynamics of the process. Membranes also have a very important role in this research, as the affinity of AH for the bilayers depends on their composition. During this Ph.D thesis, we have studied how membrane composition affect its physico-chemical properties and, in consequence, their interaction with peptides. To sum up, with this Ph.D work will optimize MD research for AH / membrane systems and we will obtain novel information about their action mechanism. Our long term goal is to be able to computationally predict if a given AH can interact with a given membrane with a specific lipid composition under a certain state (flat, curved, under tension, etc.). As more and more medicines/drugs are based on peptides, our research has some perspectives in biomedicine and drug design where AH are involved.

In total, this manuscript is divided into 7 different chapters. In this introduction, the molecular systems that have been studied during my Ph.D are presented, as well as the state of the art of the research and the evolution and importance of the employed techniques. Chapter 2, Materials and

Methods, explains the computational and *in vitro* techniques as well as the methods that have been employed. Specific information about the protocols are given in each chapter of results. There are five chapters of results, one for each of the projects we carried out during this Ph.D. The main results are presented and discussed. An overview of each project is given below.

In chapter 3, we did a Force Field (FF) comparison for AH/membrane systems. The FF is the ensemble of equations and parameters used in MD to predict the behaviour of the system. Nowadays different FF are available to simulate lipids and proteins: CHARMM36m, OPLS, GROMOS and AMBER99SB-ILDN/Berger combined for proteins/lipids. The choice of the FF is a delicate step when doing MD as the reliability of the results depends on the correct parametrization of the molecules. In this chapter we carried out T-REMD simulations with four different FF for two molecular systems in order to determine the most accurate FF(s) for AH/membrane simulations. The two AHs employed were ALPS (a membrane curvature sensor) and the mastoparan (from wasp venom) in a DOPC/DOG 85/15 bilayer. The aim was to observe the structural and dynamic properties of two AH in a lipid bilayer and compare the results with *in vitro* assays to classify the FF by accuracy.

In chapter 4, we studied mastoparan insertion in the membrane by MD simulations. In the previous chapter four different FF were compared to study AH behaviour in the membrane and results were compared to experimental data. We determined that the two most reliable FF were AMBER99SB-ILDN/Berger and CHARMM36m. The objective of this chapter was to check if they also manage to correctly predict AH partition and folding mechanism. T-REMD simulations of mastoparan insertion were launched for the two FFs. The aim was to observe the structural and dynamic properties of two AH in a lipid bilayer and compare the results with *in vitro* assays.

In chapter 5, we present a novel research strategy on AH folding within a membrane. A protocol combining T-REMD with MSM was designed to study AH structural, kinetic and thermodynamic properties. In this strategy, we first used T-REMD to explore the conformational landscape of the system. From the T-REMD simulations, we selected an ensemble of conformations that were used as starting structure for non-bias classical MD simulations. The MD simulations were then used for MSM construction. This strategy was first employed with the mastoparan as model system to validate and optimize the protocol that will serve to study more complex molecular systems in the future.

In chapter 6, we studied the LTX-315 oncolytic peptide and his interaction with membranes by experimental and computational assays. It is an AH developed by the company LytixBiopharma to treat sarcoma disease. We have started a collaboration with Øystein Rekdal the CEO of Lytix-Biopharma who sent us some peptide allowing us to perform some experiments. LTX-315 has a 9 residues sequence with one non-standard amino acid, the diphenylalanine (DiP), which is similar to the phenylalanine but with two aromatic rings [86]. We lack information about LTX-315 action mechanism with membrane from a molecular point of view. The fact that it presents a non-standard amino acid is a challenge for computational research and a key to understand its be-

haviour. I first carried out fluorescence and CD assays myself as a first approach to its structural and binding properties. I also performed T-REMD simulations of the peptide in water and in a lipid bilayer. Our collaborators, Olivier Lequin and his Ph.D student Edward Chalouhi performed NMR assays to resolve its membrane-bounded structure, its orientation in the membrane as well as the relative position of the DiP aromatic rings. Finally, another collaborators from our Lab, Françoise Illien and Sandrine Sagan, applied a novel protocol to carry out internalization assays, where she studied the LTX-315 penetration capacity and preferred pathways.

Chapter 7, the last one, presents a collaboration that took place in our lab where I had the opportunity to participate: the penetratin and its interaction with PIP₂ lipids. The penetratin is a member of the cell-penetrating peptides (CPPs) family, able to enter cells and to deliver biologically active molecules. PI(4,5)P₂ is an anionic lipid with multiple functions such as second-messenger precursor, regulator of actin polymerization, membrane trafficking, etc [118]. This collaboration is headed by two other collaborators in the lab Astrid Walrant and Emmanuelle Sachon. We have worked together to present an structural model of the penetratin in the membrane and its specific interactions with PIP₂ lipids. The strength of this project is the combination of experimental and computational techniques in biology and biophysics that generate solid results and a deeper understanding of its behaviour. We carried out CG and AA-MD simulations of the penetratin with different bilayer compositions. Edward Chalouhi, Ph.D student of the LBM and his PhD Director Olivier Lequin carried out H-NMR experiments of the peptide in buffer and in presence of lipids. All together, these results help to understand penetratin structure and orientation in the membrane and its specific affinity for PIP₂ lipids. A publication of the whole project is about to be submitted.

At the end of the manuscript, a general conclusion chapter is included, where the main results are summed up and discussed. We highlight how we answered the main questions raised in this introduction and propose new perspectives for AH/membrane interaction research.

1.2 The Eukaryotic cell and its membranes

An Eukaryotic cell is any cell or unicellular organism that possesses a clearly defined nucleus. They are isolated from the external medium thanks to the plasma membrane. Internally, they present different organelles. Each one has its own membrane and particular function and they are illustrated in figure 1.1.

The plasma membrane

The plasma membrane separates the interior of the cell from the outside environment. It is composed by lipids and proteins. It provides protection for the cell and fixed environment inside the cell. It actively participates in the regulation of the transport of materials entering and exiting the cell: nutrients are internalized into the cell and toxic substances are transported out of the cell. They are also important for the cell communication: they have attached proteins to interact with other molecules and cells. The plasma membrane has different compositions depending on the cell type and organism.

The nucleus

It contains and protects the DNA, the hereditary information of the cell. The nucleus controls the activities of the cell by regulating the protein synthesis: the information encoded in the DNA is transcribed into messenger RNA (mRNA) in the nucleus and then the mRNA is translated into proteins in the cytoplasm. The nucleus is surrounded by the nuclear envelope, a double membrane that isolates the nucleus from the cytoplasm. The outer membrane is continuous with the endoplasmic reticulum. Both membranes contain pores which control the movement of substances in and out of the nucleus: the mRNA is transported into the cytoplasm, and proteins are selectively transported into the nucleus.

The endoplasmatic reticulum and the Golgi

The endoplasmatic reticulum (ER) and the Golgi apparatus are two organelles implied in molecule synthesis and intracellular transport. Their functions are combined to produce proteins and lipids, and subsequently transport them into vesicles to their target organelle. This process is called vesicle trafficking.

The synthesis of proteins and lipids takes place in the endoplasmatic reticulum (ER), which is located next to the nucleus. There are two types of ER, the smooth and the rough. The rough ER has attached ribosomes, which are small, round organelles which synthesize the proteins from the mRNA that comes from the nucleus. The smooth ER helps produce and concentrate other substances needed by the cell. Once the molecules are synthesized, the ER actively participate in molecule transport across the cell. Molecules are packed into transport vesicles and moved

along the cytoskeleton toward their destination. If they need modifications, as post-translational maturation in proteins, the vesicles are delivered to the Golgi apparatus. Secretory proteins and glycoproteins, cell membrane proteins, lysosomal proteins, and some glycolipids all pass through the Golgi apparatus at some point in their maturation.

In the Golgi, protein maturation and sorting takes place. The Golgi apparatus is a membrane-bound organelle made up of flattened, stacked pouches called cisternae (4 to 8 per Golgi). The cisternae are held together by matrix proteins, and the whole of the Golgi apparatus is supported by cytoplasmic microtubules. It is divided in three compartments with different functions. The "cis" cisternae are the closest to the ER, where the vesicles deliver the proteins and lipids. Then, these molecules pass to the "medial" zone, where they are modified into functional molecules and marked to be delivered in a specific intracellular or extracellular locations. There are different type of modifications, as glycosilation or fatty acid addition. Finally, in the "trans" cisternae (the farthest from the ER), proteins and lipids are sorted and packaged again into vesicles. The vesicles are sent into their final destination, that can be another organelle, the plasma membrane or the extracellular medium. In the last case, they are released by exocytosis. The vesicle traffic is an essential mechanism for the cell growth and maintenance and it is based on membrane fusion, fission and vesicle formation.

The mitochondrion

The mitochondrion is the powerhouse of the cell: it produces energy in form of ATP. In 1967, Lynn Margulis published the Endosymbiosis theory. She postulated that originally, the mitochondria was a prokaryotic organism in symbiosis in an eukaryotic cell [183]. This prokaryotic organisms progressively lost its autonomy becoming an organelle of the host cell. The main evidences for this theory is that the mitochondria possesses its own circular genome, it divides independently from the cell, it produces some proteins on its own and it has a double membrane system.

This double membrane system is fundamental for the energy production and the survivor of the cell. The outer membrane has a thickness of 60-75 Å and recovers all the organelle. It is highly permeable to ions and small molecules. The inner membrane presents a big surface because it bends in many folds called cristae. It is not permeable as the outer membrane and it has higher protein concentration. In the inner membrane energy is produced in form of ATP [84]. Between the two membranes there is a zone called the intermembrane space. The matrix is in the interior of the inner membrane where the Krebs cycle takes place. It is a series of chemical reactions that produces NADH. The NADH is fed into the oxidative phosphorylation (electron transport) pathway, which takes place in the inner membrane. In the electron transport pathway, an electron jumps between membrane's protein complexes until its ultimate acceptor, the oxygen, creating water. During the electron transport, the protein complexes push protons from the matrix to the intermembrane space, creating an electron gradient. This gradient is used by the enzyme ATP synthase to create energy in form of ATP. If the double membrane system is damaged, the mitochondrion is not

capable to produce energy and cell death is induced.

Lysosomes

Lysosomes are cell organelles present in the cytoplasm of eukaryotic cells. They are in charge of the intracellular digestion. They eliminate non-useful or toxic molecules thanks to their high content in digestive enzymes as lipases, peptidases and nucleases. The lysosomal membrane contains transport proteins (permeases), ionic channels for chloride ions, and proton pumps. The ion channels and proton pumps maintain a low pH inside of the lysosomes (between 4.5 and 5), which is essential for the enzymatic activity. The membrane assures the functioning of the lysosomes and protects the cell from its low pH and the digestive enzymes at the same time.

Lipid droplets

Lipid droplets are the main place for storing cellular fats in the form of triglycerides or esters of cholesterol. Lipid droplets are peculiar organelles since their interior is an oil while the exterior is an aqueous environment (cytoplasm). Therefore they are surrounded by a monolayer of phospholipids, which is unique among all organelles. A monolayer has very different properties compared to a regular bilayer, e.g. it can be pulled without disrupting. Here surface tension play a key role in the property of lipid droplets, which depend on many factors but mainly the degree of exposure of hydrophobic parts. Depending on this, Lipid droplets are highly fusogenic, but on the other hand they are wrapped by proteins, notably long amphipathic helices such as perilipins, which may moderate their fusogenicity.

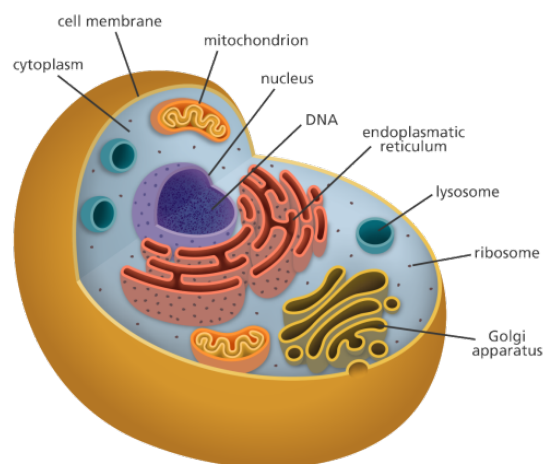


Figure 1.1 – Eukaryotic cell diagram. Image from: <https://www.w3spoint.com/>.

Conclusions

Membranes fundamental for the survivor of the cell. They divide the cell in compartments, so multiple chemical reactions can take place simultaneously. This increases the biochemical efficiency and prevents the dissemination of reaction products while the cell is functioning like a single unit. Apart from the passive, protective role, membranes are highly dynamic. They have active roles in energy production, intracellular membrane trafficking, fission and fusion, cell division, etc. Because of their importance in cells, they are becoming a target in biomedical research. There are more and more drug design projects based on membrane-interacting molecules.

Fundamental knowledge in membrane mechanism and functionalities is essential to understand cell processes and to understand their response to external agents. Membranes are one of the main focus of this Ph.D research. In this manuscript we are going to talk about their different compositions, physico-chemical properties and dynamics, and how these characteristics impact their interactions with other molecules, specifically with peptides.

1.3 Membranes

Cellular membranes are lipid bilayers that surround the cell and the different organelles. They protect the cell and divide it in compartments with different functions. They are composed by lipids and proteins, which can be integral or peripheral. Membranes are highly dynamic and they actively participate to cell's functions as solute transport and signaling pathways. Membranes are essential to cell's equilibrium and survival.

1.3.1 Historical context

Robert Hooke discovered cells in 1665. He proposed the Cell Theory where he stated that all cells contain hard cell wall since only plant cells could be observed at the time. This theory was maintained for more than 150 years until advances in microscopy were made. In the early 19th century, cells were recognized as being separate entities and studies were extended to include animal cells to suggest a universal mechanism for cell protection and development. At the end of the century, the existence of a cell membrane in animal cells, different to the plant wall was inferred but they were not considered vital to the cell survival, being just a secondary structure. Later discoveries on osmosis and permeability made the cell membranes gain more recognition. In 1895, Ernest Overton proposed that cell membranes were composed by lipids.

The lipid bilayer hypothesis was proposed in 1925 by Gorter and Grendel based on crystallographic studies and soap bubble observations. They extracted lipids from blood cells and measure its area. Then, the lipids were compressed on a water surface to create a monolayer. The surface was reduced a 50 %. The single possible explanation for this was that the lipids of the initial membrane were organised in a bilayer as it was drawn in their first model 1.2a.

In 1935, the paucimolecular model of Davson and Danielli became popular and it dominated cell membrane studies for the following 30 years. This model stands that membranes are composed by a central region of hydrophobic lipids with variable thickness, surrounded by a phospholipid bilayer which polar heads are orientated towards the outside, and an external layer composed by globular proteins [56]. A scheme of this model is shown in figure 1.2b

This model became rivaled by the fluid mosaic model of Singer and Nicolson in 1972. They published a membrane model where they described the structure of the plasma membrane as a mosaic of components in a fluid state [200]. The bilayer was modeled as a two dimensional pseudo liquid where proteins and lipids associate with enough freedom of movement to be functional as in figure 1.2c.

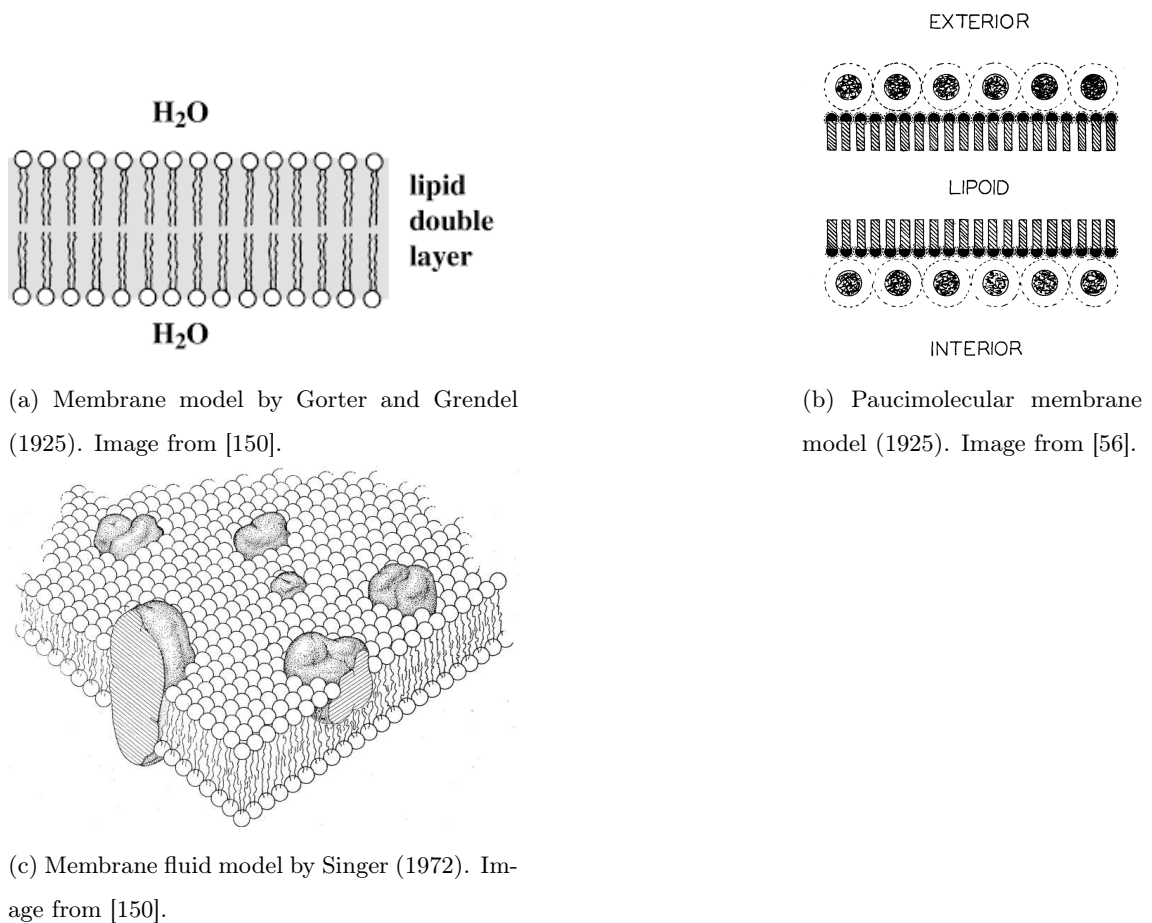


Figure 1.2 – Relevant membrane models in history.

The importance of Singer's model is that it is the first model to describe the membrane as a dynamic system. It has evolved through the years to adapt to more recent discoveries and to the specific properties of each cell type and organelle. In 1978, Israelachvili proposed a non-homogeneous model where the protein concentration is not constant throughout the surface. The protein types and concentration also depend on the membrane localisation and are related to the organelle function. Also membrane's structure is variable: some important features are the curvature and the thickness. In 1995, Sackmann, described the presence of a cytoskeleton and

glycocalyx. A modern membrane model is showed in figure 1.3.

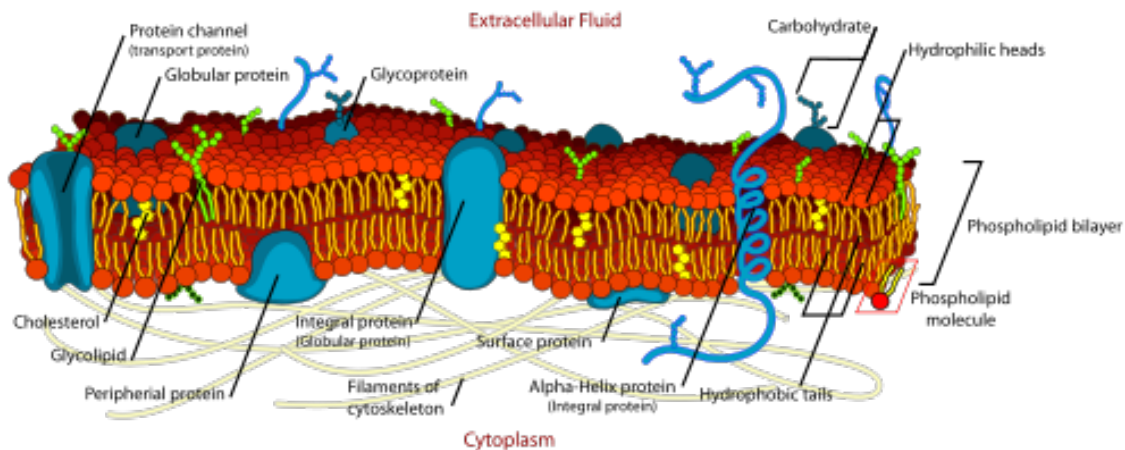


Figure 1.3 – Modern cell membrane model. By LadyofHats Mariana Ruiz - Own work. Image renamed from File:Cell membrane detailed diagram.svg, Public Domain, <https://commons.wikimedia.org/w/index.php?curid=6027169>

In the past decades, biochemical and biophysical findings have provided a detailed model of the composition and structure of membranes. In 2009, van Meer et al. published a very interesting and detailed review of the lipid diversity within the cell [148]. The different lipid compositions and ratios depend on the organelle and cell type. The scheme they published reporting on the composition of cell membranes is shown in figure 1.4.

1.3.2 Membrane composition

The cell membrane is composed by lipids and proteins. Three lipid types are present: phospholipids, glycolipids, and sterols. As it is shown in figure 1.4, the lipid composition depends on the membrane type. In the majority of cases phospholipids are the most abundant, often contributing for over 50% of all lipids in plasma membranes. Glycolipids are not very abundant (< 2%) and the remaining 50 % are sterols. There are two protein types present in membranes: integral, which cross the bilayer; and peripheral, which are located in one of the leaflets. The protein/lipid ratio depends on the cell type and organelle. For example, the composition of plasma membranes is about half lipids and half proteins by weigh [29]. The main types of molecules composing cell membranes will be discussed in this section.

Phospholipids

Phospholipids are the principal component of the lipid bilayer. They are composed by a glycerol molecule, with two aliphatic chains attached in positions sn-1 and sn-2 and a phosphoric acid in sn-3 position. Phospholipids are amphipathic: the two aliphatic chains constitute the non-polar tails and the phosphoric acid the polar head. A scheme of the structure is shown in figure 1.6 The

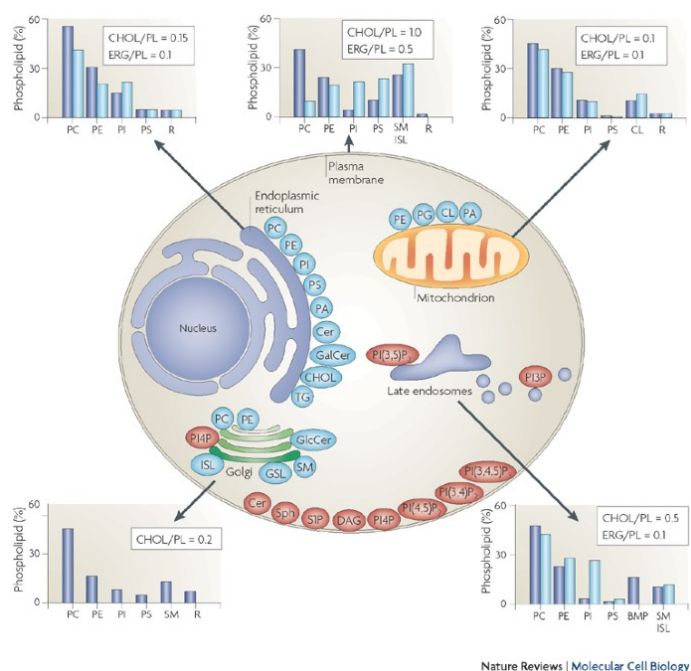


Figure 1.4 – Lipid diversity in the Eukaryotic cell. The barplots indicate the lipid composition of each organelle in mammals (dark blue) and yeast (light blue). Cholesterol/phospholipid and ergosterol/phospholipid ratio is indicated for mammals and yeast respectively. Blue circles mark synthesis sites and red circles the lipid related to signaling pathways (less than the 1%) Image from [148]

phosphoric acid can be bonded or not to a hydroxyl group which determines the different types of polar heads that exist. Depending on the polar head, it is possible to classify the phospholipids in two major groups: charged or neutral (zwitterionic). The principal polar head structures are shown in table 1.1 [138].

Hidroxyl group	Structure	Phospholipid type	Abbreviation	Charge
Hydrogen	H	Phosphatidic acid	PA	Negative
Choline	<chem>C[N+](C)(C)CCO</chem>	Phosphatidylcholine	PC	Zwitterionic
Ethanolamine	<chem>NCCO</chem>	Phosphatidylethanolamine	PE	Zwitterionic
Serine	<chem>N[C@@H](CO)C(=O)O</chem>	Phosphatidylserine	PS	Negative
Glycerol	<chem>OCC(O)CO</chem>	Phosphatidylglycerol	PG	Negative
Inositol	<chem>OCC1(O)CC(O)C(O)C(O)C1O</chem>	Phosphatidylinositol	PI	Negative

Table 1.1 – Diversity of polar head's structures present in phospholipids

Regarding the aliphatic chains, they are a sequence of enchainned carbons and their correspond-ent hydrogen atoms attached to a carboxyl group (-COOH). The length and number of insatur-

ations is variable. Most of the phospholipids present one saturated and one insaturated chain. Table 1.2 shows a selection of the most frequent ones and the nomenclature used to describe them. The unsaturations can be in *cis*, when the hydrogen atoms are located on the same side of the aliphatic chain or in *trans* when they are in opposite sides. In nature, the major part of the lipids have *cis* insaturations because they contribute positively to membrane's fluidity while *trans* insaturations decrease lipid mobility rigidifying the membrane. The two structures are show in figure 1.5) [179].

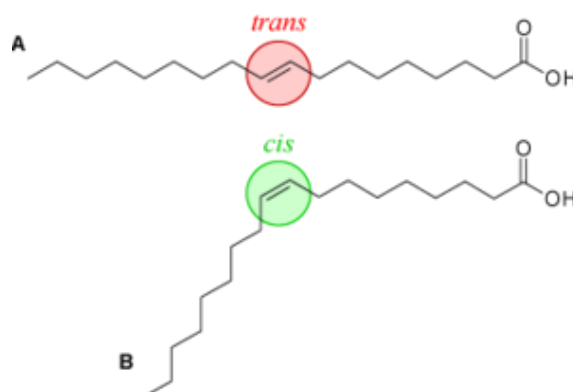


Figure 1.5 – Aliphatic chains with an insaturation in (A) *trans* and (B) *cis*. Image from Wikipedia.

Fatty acid	Length:Insaturations (Position-Conformation)
Lauric	12:0
Myristic	14:0
Palmitic	16:0
Palmitoleic	16:1 (9-cis)
Stearic	18:0
Oleic	18:1 (9-cis)
Vaccenic	18:1 (11-cis)
Linoleic	18:2 (9-cis, 12-cis)
γ -Linoleic	18:3 (6-cis, 9-cis, 12-cis)
α -Linoleic	18:3 (9-cis, 12-cis, 14-cis)
Arachidic	20:0
Behenic	22:0
Arachidonic	20:4 (5-cis, 8-cis, 11-cis, 14-cis)

Table 1.2 – Most common aliphatic chains present in phospholipids.

Different polar heads and aliphatic chains are combined to form different phospholipid types. There is a great diversity although not all the combinations are present in nature. The variation in headgroups and aliphatic chains allows the existence of > 1000 different lipid structures in eukaryotic organisms.

The polar head phosphatidylcholine (PC) is the most abundant one and is present in all membranes of all eukaryotic organisms, as described in van Meer's model (Fig. 1.4). POPC accounts for > 50% of the phospholipids in most eukaryotic membranes. It self-organizes spontaneously as a planar bilayer with the lipid tails facing each other and the polar headgroups interfacing with the aqueous phase as it is shown in figure 1.6. The impact of lipid's physico-chemical properties in membrane's general structure will be discussed in the next section.

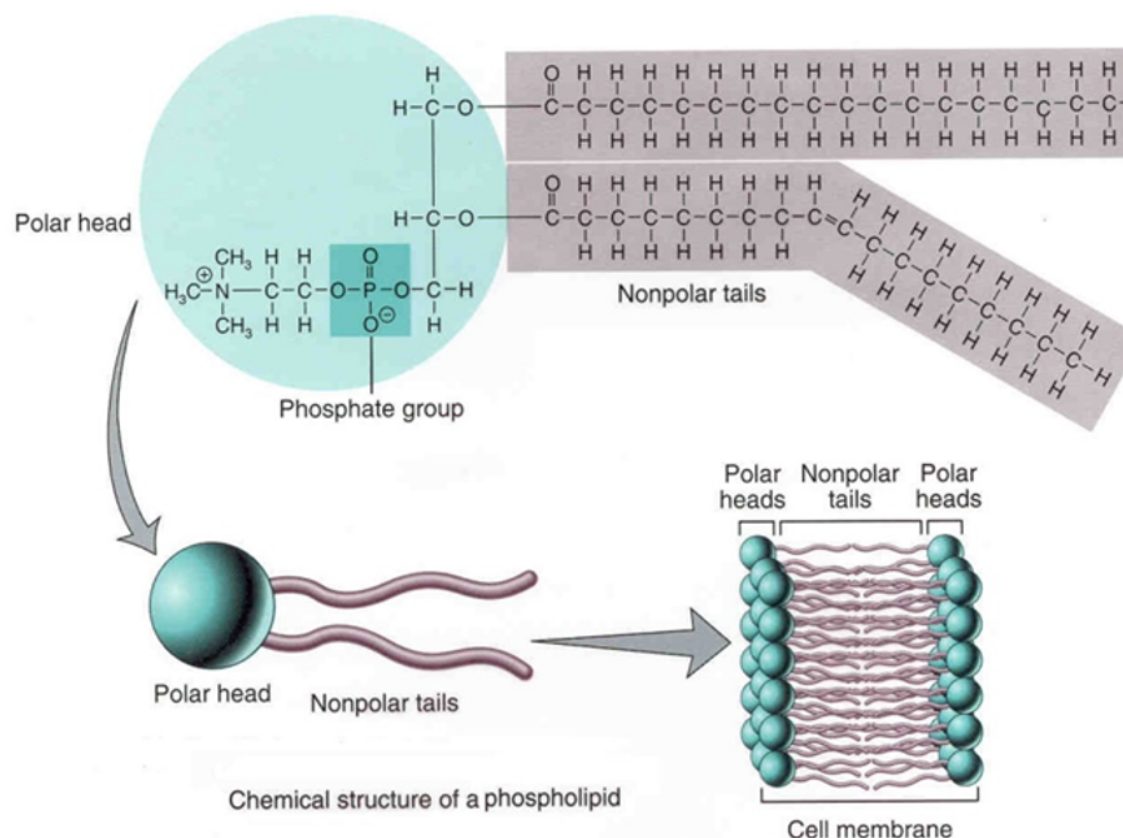


Figure 1.6 – . Phospholipid schematic representation and bilayer formation by its aggregation.
Image from: <https://www.creative-proteomics.com>.

Cholesterol

Sterols are another lipid type present in cellular membranes. Its concentration in the membrane is variable depending on the cellular type and organelle. The most important sterol in lipid bilayers is cholesterol. It is composed by four carbon cycles with a small carbon chain attached on one side and a polar head on the other side (Fig.1.7). The carbon chain is composed by 8 carbons and their respective hydrogen atoms. The polar head is only composed by a hydroxyl group (-OH) although sometimes it is esterified by a fatty acid forming a cholesteryl ester (this last form is the way of storing cholesterol within lipid droplets). Cholesterol increases membranes' lipid diversity. As the polar head is small and can be esterified, the cholesterol is a very hydrophobic molecule. In mammals, mitochondrial membranes present less than 5% of cholesterol, the Golgi apparatus a

8% and the plasma membrane between a 20% and a 50%, depending on the cell type (Fig. 1.4) [138].

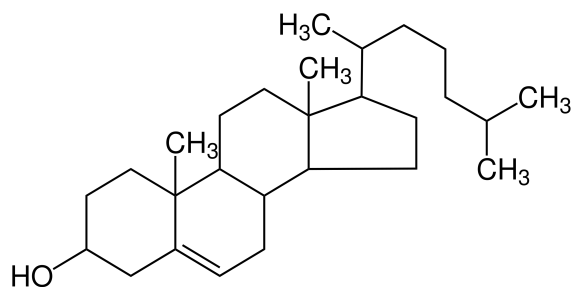


Figure 1.7 – Cholesterol's structure

The cholesterol concentration has a big impact in membrane's structure and dynamics. It has a role as natural regulator of membrane's fluidity. It has a cohesive effect on phospholipids: it increases the rigidity of the membrane general structure, forming highly ordered membranes. Membranes presenting high cholesterol concentration will be more condensed, less fluid and more difficult to cross by small molecules. This is the reason why high concentrations are present in the plasma membrane, it protects the cell against the entry of small molecules.

Integral proteins

Integral membrane proteins span across the bilayer. They have one domain exposed to the outside (extracellular space or cytosol), another domain on the inside (cytosol or lumen) and a transmembrane domain highly hydrophobic that interact with the lipid aliphatic chains. There are multiple types with various functions: connection between the interior and exterior of the cell, functions related to cell communication and signalisation pathways, ion channel transport, protein G coupled receptor, etc. They are of great interest in pharmaceuticals because they serve as receptors for drugs [90]. Their presence and concentration differ depending on the organelle and also on the cell type. For example, 18 % of membrane's components are integral proteins in neuron plasma membrane but 75 % in mitochondrial internal membrane. This is because in the inner mitochondrial membrane they have a key role in energy production.

Peripheral Proteins

Peripheral proteins are just attached to one of the layers and do not have a transmembrane domain crossing the bilayer. The interaction with the membrane is normally transient and the attachment is regulated by hydrophobic or electrostatic interactions. These proteins target specific membranes to interact with them. The affinity depends on membrane physico-chemical properties. Four different interaction mechanisms are showed in figure 1.8, but some other exist [120]. The first one is the specific recognition of a lipid's polar head as it was described by Lemmon for globular protein interaction with phospholipids [132]. Examples include the conserved region-1 (C1) domains, which

specifically recognize diacylglycerol. Membrane binding by these domains is typically dictated simply by the presence or absence of the target lipid in membranes. Some proteins insert into the membrane upon lipid modification that penetrates the membrane (esterification of an aliphatic chain over a residue), as the proteins from the myristoylated family [164]. A very common type of interaction are the electrostatic-driven interactions. Protein cationic chains insert into the membrane surface because it can present negative charges [223, 230]. Finally, the partition of an amphipathic helix (AH) into the membrane can also serve to target peripheral proteins [26].

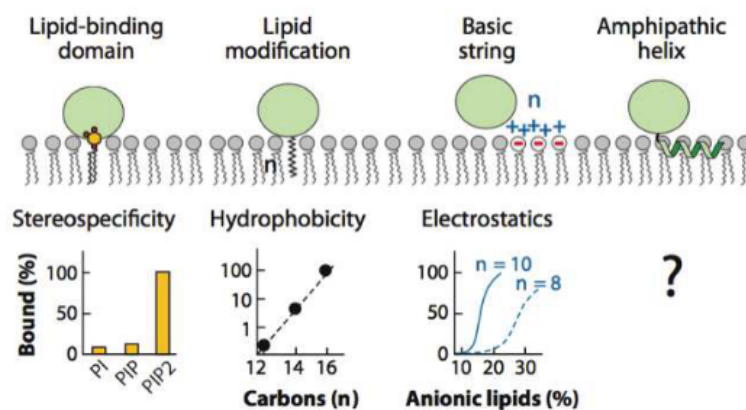


Figure 1.8 – Peripheral proteins' mechanisms to bind lipid membranes. Image [10]

All these strategies are principally driven by stereospecificity, hydrophobicity and electrostatic attraction. However, AH insertion into the membrane is a more complex process where the insertion is coupled to helix folding. AH and its insertion mechanism is one of the main focus of this research and it will be treated more extensively in the next section.

1.3.3 Membrane Structure and dynamics

The general structure of the lipid bilayer is directly impacted by the lipid composition, water to lipid ratio and temperature. The lipid shape depends on the relationship between the size of the polar head and the size of the aliphatic chain. There are three possibilities: cylinder, cone and inverted cone 1.9. Lipids have a cylindrical structure when the head and the chains have approximately the same width, as it is the case of the POPC. A cone structure is formed when the polar head is smaller than the chains width and the inverted cone when the polar head is bigger [102]. Accordingly, the packing parameter p describes the lipid shape with a single number which basically relates the volume of the polar head and the aliphatic tails: $p = 1$ for cylindrical lipids, $p < 1$ for conical lipids and $p > 1$ for inverted conical lipids. The abundance on those three classes of lipids can lead to the following shapes: flat, positively curved (convex), negatively curved (concave) respectively.

All this properties can be applied to another group of compounds with an amphiphilic structure, the detergents. They also have a hydrophilic (polar) head and a long hydrophobic (non-polar) tail.

The properties of detergents are dependent on the molecular structure of the monomer. Some of them aggregate in water to form micelles, which have a hydrophobic core. The hydrophobic core can enclose different molecules, as grease, protein or soiling particles. Thanks to this property they have a key role in protein purification assays, so they are necessary for *in vitro* protein synthesis.

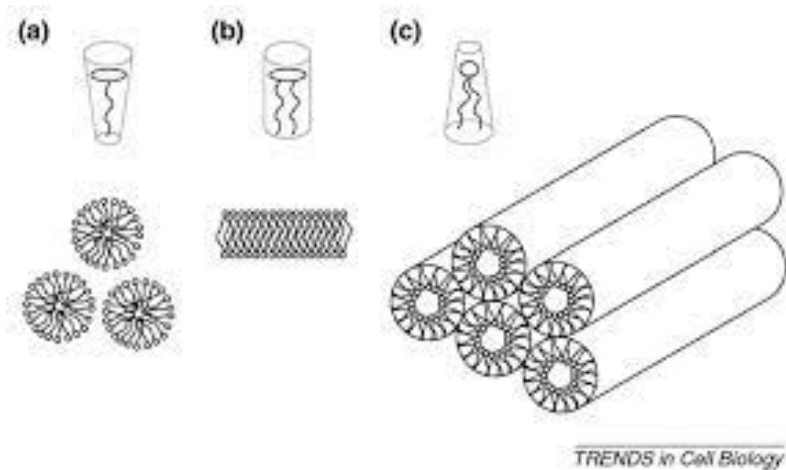


Figure 1.9 – Lipid shapes and the structures they can form. (a) Inverted conical lipids can aggregate in micelles, (b) cylindrical lipids form bilayers and (c) conical lipids can generate reverse hexagonal phase. Image from: [102]

Phospholipids are amphipathic molecules. In water they form emulsions where the hydrophobic tails aggregate to minimize exposition to water while the hydrophobic heads keep exposed to water. This mechanism helps to decrease surface tension which means that the phospholipids are surface-active molecules (i.e. surfactants). The way they self-assemble (i.e. the phase) critically depends on the lipid shape: cylindrical lipids generally form bilayers, conical lipids may form reverse hexagonal phases and inverted conical lipids can form micelles (Fig. 1.6 and 1.9). In Biology, the various cellular lipids form most of the time bilayers, although some other phases are possible during specific events (e.g. membrane fusion). Importantly, the relative abundance of the various lipid shapes in a biological membrane has a direct impact on their curvature, which plays a key role in its interaction with proteins. High membrane curvature is encountered for example in vesicular trafficking, neurotransmitter vesicles, etc.

The fluidity of the membrane also depends on lipid dynamics and is critical for its permeability. The lipids are able to diffuse in the membrane plane. The structure and dynamics of the aliphatic chains depend on their length and number of unsaturations. Unsaturation rigidify the structure and increase the volume they occupy. In consequence, unsaturations decrease the fluidity. In addition, the increase of the volume of the chains separate the polar heads, creating small cavities that expose the hydrophobic chains to the solvent. These cavities are called Packing Defects, a scheme is shown in figure 1.10. Packing Defects have a key role in membrane interactions with peptides as they need hydrophobic cavities to insert their bulky hydrophobic amino-acids into the membrane [216]. Packing defects are also critical for membrane fusion [73], especially for the first

part of the process which is called hemifusion (i.e. when the two outer leaflets fuse).

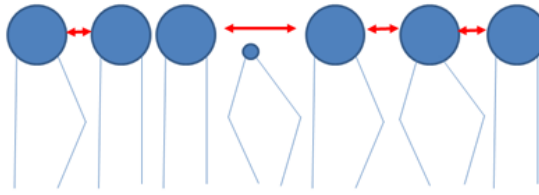


Figure 1.10 – Packing defects’ schematic representation. The red arrows point to the defects. Image by Patrick Fuchs.

Another factor that influences membrane dynamics is temperature. The phase transition temperature, is the temperature at which lipids change physical properties from gel phase L_β (rigid) to liquid crystalline L_α phase (fluid). If the temperature is higher, lipids are disorganized and the membrane gets more fluid. If the temperature is lower, lipids are highly organized and, in consequence, the membrane is more rigid. The transition temperature between the two, also called melting temperature, depends on the number of unsaturations: the higher the number of unsaturations, the more space between the lipids and so the lower the melting temperature. For example, the melting temperatures for DSPC, SOPC and DOPC are 55.6, 6.7 and -40.3°C , and they have 0, and 2 insaturations respectively.

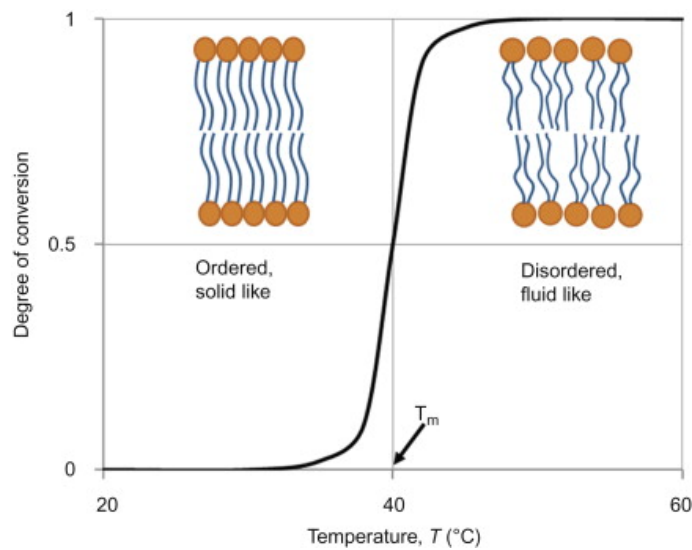


Figure 1.11 – Phase transition in lipid bilayers. As the temperature increases, the membrane “melts”, going from an ordered solid-like bilayer (left) to a disordered, fluid one (right). Around the melting temperature T_c , the two phases coexist. Image from [202].

1.3.4 Membrane systems in research

In nature, cell membranes are very complex system of various compositions and properties. Although research in this field have evolved a lot in the past decades, the different membrane compos-

itions, dynamics and functions are not fully yet understood. There are *in vivo* assays that allow to study cell membranes but they are normally limited in spatial resolution [10]. *In vitro* assays (e.g. electron microscopy (EM)) are commonly used to study membrane behaviour in a more detailed way. To study membrane containing systems, simplified bilayers are commonly used. The idea is to create a bilayer with a reduced number of lipid types that mimic the physico-chemical properties of the membrane of interest. This reduces the high costs and also the technical difficulties of producing lipid bilayers. Some of the most popular membrane structures used for experimental assays are micelles and bicelles, shown in figure 1.12. Bicelles are bilayer-mimetics composed by long phospholipids in the central part and detergents (phospholipids with short chains such as DHPC) in the extremes, where the structure needs to be curved. The q factor describes the size of the bicelle. When there are only detergents present, there is no bilayer formation, only an spherical structure called micelle with $q = 0$. Both structures are commonly used for experimental assays, especially in NMR experiments because the sample need to tumble fast.

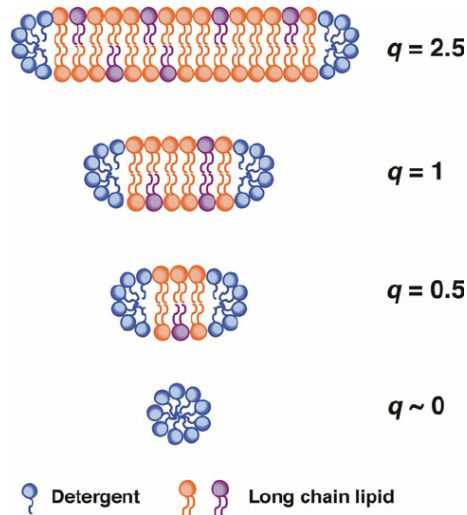


Figure 1.12 – Micelle and bicelle structure. In orange and purple, long chain lipids of different nature. In blue, detergent molecules. Bicelles are formed by a central planar bilayer composed by long chain lipids with the edges stabilized by either short-chain lipids or detergents. Different lipids can be used. The q value represents the extension of the bilayer. The larger q , the larger the bilayer. Micelles can be described as a bicelle with a q value equal to 0. Image from [45]

Normally, lipid bilayers contain one to four lipid types. This way, it is possible to construct *in vitro* bilayers to study their behaviour and interactions with peptides and proteins, solute transport, etc. For example, in 1999 S. White deduced the density profile of a pure DOPC lipid bilayer by X-Ray diffraction assays [92]. The density profile gives information about atom groups relative positions and membrane thickness. Results are shown in figure 1.19 and will be discussed deeply in the next section. Solid state NMR spectroscopy also provides valuable information regarding the structural fluctuations of lipid bilayers, including both the equilibrium properties and dynamics. From the NMR data, it is possible to extract the order parameters (OP) of the lipids. OP are a measure of the degree of order of the aliphatic chains. OP are then used to deduce structural

parameters of the lipid bilayer, such as the thickness or area per lipid (APL) [166, 167]. Fluorescence Correlation Spectroscopy (FCS) and Fluorescence recovery after photobleaching (FRAP) are a standard method used to study the dynamics of lipids and proteins, giving information about lipid diffusion [169].

In silico research is also concerned with membrane studies. The evolution of computational techniques in the last years have also positively impact this field. There are more and more computational research teams interested in membranes because of their importance for cellular activity. Molecular dynamics (MD) is one of the main computational techniques employed to study membranes [211]. For MD simulations, simplifications of the membrane are generally used, as *in vitro* assays. A set of lipid structures are chosen and a precise ratio to better represent the membrane characteristics of interest, such as the charge or packing defects for example. The simplification of the model reduces the computational cost of the simulation and allows the research to be focused in an small number of variables. The results from the simulations are compared to *in vitro* assays to validate the reliability of the data.

A good example of the symbiosis between *in vitro* and *in silico* research is the open-science collaborative project called NMRlipids [<http://nmrlipids.blogspot.com/>]. The aim of NMRlipids is to carry out MD simulations of lipid bilayers with different compositions and conditions and then compare them to experimental results. This way, the authors get to determine the reliability of the Force Field (FF) (ensemble of equations and parameters used to compute the simulations). In their first project, they simulated phosphatidylcholine (PC) lipid bilayers with 13 different all-atom models, and compared simulations to NMR experiments in terms of the highly structurally sensitive C-H bond vector order parameters (OP). Focusing on the glycerol backbone and choline headgroups, they showed that the OP comparison can be used to judge the structural accuracy of the FF [34]. I had the opportunity of participating on the last project with Patrick, where we worked with pure POPE and mixed POPC/POPE bilayers. The manuscript has been published recently [14]. MD is the principal technique employed during this Ph.D so we will be talking extensively about membrane MD simulations in this manuscript.

1.4 Amphipathic Helices

Amphipathic helices (AH) are key protein motifs that are capable of interacting with lipid membranes. Their sequence is short and they are normally unfolded in water. When they interact with the membrane, they fold into an α -helix and get inserted in the membrane/water interface. They are called amphipathic because they have two sides, a polar one exposed to the solvent and a hydrophobic one which residues are inserted within the membrane. The hydrophobic side chains establish hydrophobic contacts with the lipid aliphatic chains. The affinity of the AH for the membrane will depend on membrane physico-chemical properties as the degree of curvature, the charge and the packing defects. AH dynamics and their interaction with membranes is the main focus of

this Ph.D research. We have worked with different peptides and membrane compositions that will be presented later on this chapter.

Physico-Chemical Properties

The first representation of an AH was done in 1967 by Shiffer, he design the helical wheel representation [191]. He drew a circle where the lateral chains of the residues are projected each 100° because there are 3.6 residues per helix turn. This representation allows to see the relative positions of the residues in the helix and determine which side is polar and which side is non-polar. Most often, the division is not perfect but in all cases it is easy to distinguish the two sides. The first helical wheel ever published is shown in figure 1.13

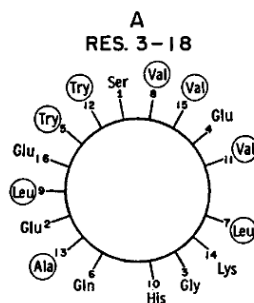


Figure 1.13 – First helical wheel ever published. It corresponds to one segment of the whale sperm myoglobin, described by M. Schiffer in 1967 [191]. The residue number is written next to each amino acid and the hydrophobic amino acids are surrounded by a circle.

In 1982, Eisenberg proposed to quantify the amphipathicity of the helix by calculating the helical hydrophobic moment [66] :

$$\langle \mu_H \rangle = \frac{|\sum_{i=1}^N \mathbf{H}_i|}{N} \quad (1.1)$$

Where \mathbf{H}_i is the vector which goes from the center of the helix in the direction of the lateral side chain projection; $|\mathbf{H}_i|$ corresponds to the hydrophobic value in Eisenberg scale; and N is the total number of hydrophobic residues. A high $\langle \mu_H \rangle$ value indicates high amphipathicity, which means that the polar and non-polar residues are well divided in two different sides of the helix.

Although useful to get a first glance, we know nowadays that the mechanism of action of the AH is more complex than a simple helical wheel description. This latter and hydrophobic moment are not enough to describe the complex dynamics of AH: the free energy of partitioning does not depend upon the hydrophobic moment [3]. To have a real comprehension of AH mechanism of action, they need to be observed with atomic detail. The key is to understand how the shape, the volume and the exposed area of each face of the helix match the complex chemical environment made by the membrane. The affinity of the AH for the membrane depends on the complementary

structural and chemical characteristics between the peptide and the bilayer. For example, big hydrophobic lateral chains need more space in the membrane, which is related to higher curvature and more packing defects. In their paper [62], Drin and Antonny point to computational methods such as MD simulations as the best fitted methods to study AH dynamics in a detailed atomic way.

Functions

Among others, AH can be classified in four important families: peripheral protein targeting to membranes, Antimicrobial Peptides (AMP), Cell-Penetrating Peptides (CPPs) and venom peptides. This classification is based in AH different functions and we are going to use it in this manuscript to talk about AH. However, other classifications are possible, as physical-chemical and structural based ones [196].

AH targeting peripheral proteins to membranes When targeting peripheral membranes, the AH will be a domain of a bigger sequence. This AH domains have specific affinity for membranes leading the protein to interact with a specific organelle. One example is the protein SAR involved in membrane trafficking. It is a monomeric small GTPase found in COPII vesicles. It regulates the assembly and disassembly of COPII coats. When it is interacting with GTP (Sar1-GTP complex), it bounds to the Golgi's membrane thanks to an AH motif. Its insertion provokes an area difference between the two layers creating a local curvature which is the first step in vesicle secretion [232].

Some AH that target peripheral membranes are also curvature sensors. This means that they have different affinities for membranes depending on its degree of curvature. Membrane curvature sensors participate in very diverse reactions, such as lipid transfer between membranes, the tethering of vesicles at the Golgi apparatus, and the assembly-disassembly cycle of protein coats [10]. One well-know curvature sensor is ALPS, a protein motif present in multiple proteins. ALPS motifs are supple sequences of 20 to 40 amino acids with no intrinsically defined structure. In ArfGAP1, it sensors the membrane to participate to the disassemble of COP1. [26]. In the presence of large liposomes ($R > 100 \text{ nm}$) containing cylindrical lipids, ALPS motifs are mostly soluble but the liposomes have a radius below 50 nm , ALPS peptides bind onto them by folding into an α -helix as illustrated in figure 1.14. It also has high affinity for liposomes containing lipids with small polar heads and/or high number of mono-unsaturated acyl chain. It seems that due to the presence of big hydrophobic chains and few charged residues in its sequence, it senses the defects in lipid packing that arise from the mismatch between the actual curvature of the bilayer and the shape of the lipids. The interaction between ALPS motif and the bilayer is driven by hydrophobic interactions rather than electrostatic attraction. ALPS motifs are present in proteins that function in the early secretory pathway and the nuclear envelope [63, 61]. These membranes are characterized by a low surface charge, low levels of cholesterol, and phospholipids with largely monounsaturated fatty acid side chains. This suggest that membrane curvature is not only the consequence of the

mechanical or enzymatic work of proteins, but also an effector and spatial information used by the cell [172, 10]. ALPS is one of the peptides studied during this Ph.D and it will be studied in one of the results chapters.

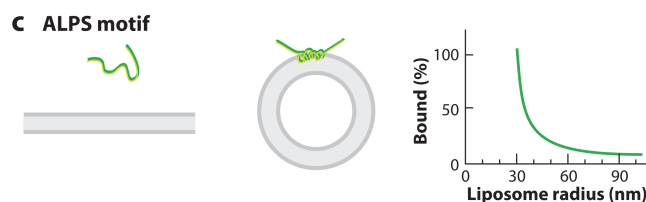


Figure 1.14 – ALPS' sensitivity for membrane curvature. Image from [10]

Membrane curvature sensors have diverse structures and chemistries, suggesting that they might have the intrinsic capacity to discriminate between different types of vesicles in cells. Unlike ALPS motifs, α -synuclein has a poorly developed hydrophobic face so it uses electrostatic interactions to overcome the minimal contribution of its hydrophobic face. This explains its dual sensitivity to negatively charged lipids and to membrane curvature [10]. Its precise function is not known, but it is expressed primarily in neurons, in which it localizes to synaptic vesicles and it is related to Parkinsons disease [157, 159, 11]. Comparison of ALPS and α -synuclein lipid-binding properties *in vitro* confirms that their differences in chemistry translate into specific binding to liposomes of different composition, complementary to the chemical properties of each curvature sensor (Fig. 1.15). This underlines the importance of properly mimicking the lipid composition and ratio of the bilayers in *in vitro* and *in silico* assays.

Another interesting model of AH targeting peripheral membranes are the sensors of membrane composition. These AH are able to detect changes in the lipid ratio, enhancing or decreasing their affinity for the membrane. A very interesting example is the AH of the regulatory enzyme CCT. This enzyme is in charge of maintaining the PC density of the bilayer, having a key role on the homeostatic control of the membrane. CCT binds bilayers that are deficient in PC, catalysing its synthesis. The CCT AH recognises the physico-chemical properties of cell membranes deficient in PC, specifically the high number of packing defects and the increase of negative charge density. These two properties increase the affinity of the AH for the membrane, that partitions and folds into the bilayer promoting PC synthesis. When PC content increases, the number of packing defects and the negative charge density decreases leading to CCT dissociation and a lower rate of PC synthesis [52].

Antimicrobial peptides (AMP) Apart from being a protein motif used by peripheral membranes, AH are also peptides that interact with membranes independently. The Antimicrobial peptides (AMP) are cationic peptides present in multi cellular organisms that are important for the immune defense of the organism. They are present in both animals and plants [75, 131, 37]. Not all AMPs are AH, they can present other structures as β -sheet, extended and looped although

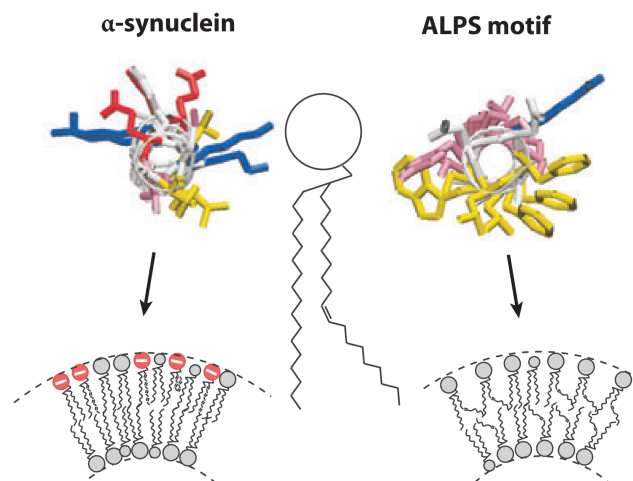


Figure 1.15 – ALPS and α -synuclein structure and membrane affinity. ALPS insertion is promoted by hydrophobic interactions because it presents big hydrophobic lateral chains while α -synuclein insertion is lead by electrostatic interactions due to its positive charges. Image from [10]

they are all cationic and amphipathic [229]. They are capable of interacting with pathogens' membranes, disrupting the bilayer and inducing the cell death [134, 236]. The exact mechanism is not yet known but the principal hypothesis propose that the peptides aggregate forming pores that increases the permeability of the membrane until it is disrupted [228]. Different disruption mechanisms are presented in figure 1.16. Independently of the chosen membrane perturbation model, an implicit concentration threshold is always required for disruption. Many peptides in this class are being intensively researched not only as antibiotics, but also as antivirals templates for cell-penetrating peptides, immunomodulators and antitumoural drugs [149].

Cell penetrating peptides (CPP) CPP are also short cationic peptides that are capable of crossing the bilayer without damaging the cell membrane [124]. They have a median length of 14 residues and a median charge of +5. Cell membranes have selective permeability that limits the access of exogenous compounds to the interior of cell. CPPs are able to cross the membrane without energy or a receptor, so they represent an interesting strategy in drug design for drug delivery [119]. In general, they have low hydrophobicity and low amphipathicity, which is not favorable for spontaneous partitioning into zwitterionic membranes [182]. For most CPPs, membrane binding requires electrostatic interactions. Its internalization depends mainly fixed factors that include peptide sequence and physico-chemical properties but also on variable factors such as local peptide concentration, local lipid composition and the properties of the cargo. [119]. The internalisation mechanisms are variable and can be passive (energy-independent) or active (energy-dependent). The direct translocation occurs when the peptide passively passes across membranes, at low concentration significant membrane disruption. The CPP can also be internalised by transient plasma membrane disruption, where peptide and cargo are delivered, but the disruption is not significant

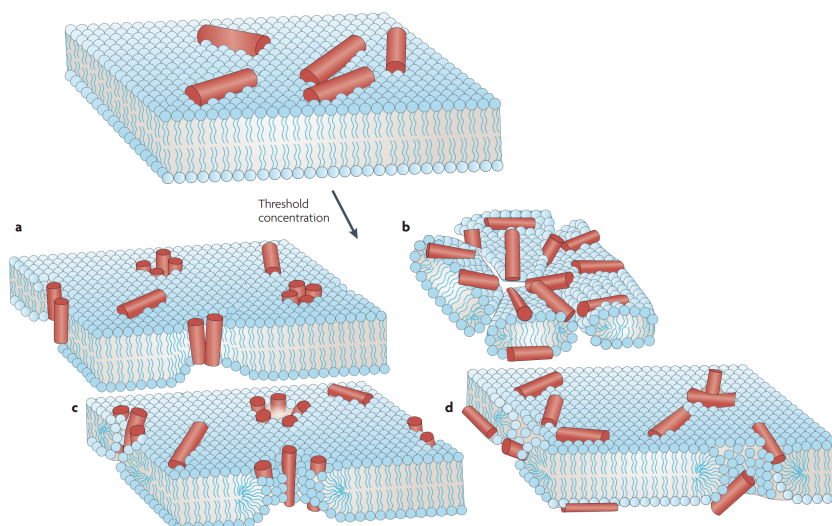


Figure 1.16 – Proposed AMPs' insertion mechanisms. (a) Barrel-stave pore: peptides insert perpendicularly in the bilayer and parallel to the aliphatic chains and form a pore. (b) Carpet mechanism: peptides adsorb parallel to the bilayer, they aggregate and produce a detergent-like effect that disrupts the membrane. Specific peptide–peptide interactions are not required. (c) Toroidal pore: peptides insert perpendicularly in the bilayer, induce a local membrane curvature and the pore lumen is lined partly by peptides and partly by phospholipid headgroups. A continuity between inner and outer leaflets is established. (d) Disordered toroidal pore: modification to the toroidal pore where the peptide conformations and orientations are less-rigid. The pore lumen is lined by the phospholipid head groups. Image from [149]

enough to cause large-scale cytotoxicity. Finally, CPP can be actively internalized by endocytosis: the cargo is targeted to membrane-bound endosomes that are taken up into the cytosol. An schematic representation of these mechanisms is showed in figure 1.17 [178].

CPPs have several applications in biomedicine. They are use for drug delivery, for example, in cancer treatment. CPPs have been effectively conjugated to siRNAs, increasing the ability of the siRNA to enter the cell and modulate selectivity for cancer cells over healthy cells [33]. Also, it is believed that some CPP have therapeutic properties themselves: Chang et al. published a study that demonstrated that optimised stapled α -helical peptides can work as dual inhibitors of the MDM2 and MDMX p53 regulatory proteins resulting in suppression of tumour growth *in vivo* [41, 178].

Penetratin is one important peptide of the CPP family. It comes from the antennapedia homeo-protein (pAntp) of *Drosophila*. When this helical sequence is mutated, the homeoprotein can not be internalized *in vitro*. It has 16 residues and it has been proven to be the minimal motif for internalization. The mechanisms of entry are variable although it has been shown to enter cells via direct translocation and also by endocytosis. Internalization also depends on experimental conditions, especially peptide concentration, and on the presence of glycosaminoglycans (GAGs)

[104] and on membrane domains created by either cholesterol depletion or ceramide formation in cell membranes [119]. At low concentrations, below $1\mu M$, the peptide enters by direct translocation (disruption of the plasma membrane). At high concentrations of the peptide, endocytosis is preferred. There is one chapter dedicated exclusively to Penetratin where we will be talking extensively about its interaction with lipid bilayers.

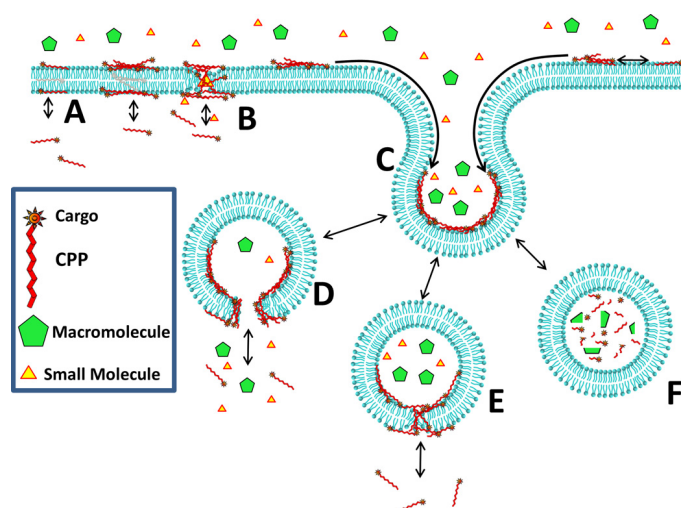


Figure 1.17 – Examples of internalization mechanisms of CPPs and its cargo. (A) Spontaneous membrane translocation across the plasma membrane. The membrane is not disrupted. (B) Transient plasma membrane permeabilization. (C) Endocytosis of membrane-bound peptide–cargo complex, along with unattached small and large molecule cargoes. (D) Endosomal membrane lysis and CPP–cargo and co-encapsulated cargo release. (E) Translocation across the endosomal membrane delivers CPP and attached cargo, but not co-encapsulated cargo. (F) CPP and cargo are rapidly degraded if release does not take place. The mechanisms depicted are not mutually exclusive; they can happen concurrently. Image from [119]

Venom peptides Venom peptides are AH present in insects venom, upon biting this compounds are injected into the host. In this AH family, there are some peptides that penetrate the host cell without damaging it and others that disrupt cell's membranes. Two well-known examples are the Melittin, from bee venom, and the mastoparan from wasp venom. Melittin is the main component of bee venom (40–60% of the dry weight). It is a pain producer and it is responsible of the tissue damage in the host. It has a 26 amino-acid sequence that forms a tetramer in water but it also can spontaneously integrate itself into cell membranes as an AH [208]. The mastoparan is a toxin from wasp venom that has a short sequence of 14 amino acids. Its action mechanism implies interactions with proteins from the host cell: the mastoparan interferes with G protein activity. It stimulates the GTPase activity, shortening the lifetime of the active G protein. It also promotes dissociation of any bound GDP from the protein, enhancing GTP binding: the GTP turnover of G proteins is greatly increased by mastoparan. The resultant G protein-mediated signaling cascade leads to intracellular IP₃ release and the resultant influx of Ca²⁺, leading to cell death. Its secondary

structure has been resolved by NMR assays. In micelles, it forms a 3 turns α -helix. By solid-NMR assays its orientation in the bilayer was also determined: the mastoparan binds mainly in the membrane/water interface as expected in AH behaviour [91]. During my Ph.D, the mastoparan has been used as a model system to study AH dynamics.

Other AH There exist AH that do not fit into this four-groups classification. Nowadays, new AH peptides are being developed with biomedical purposes. Because of their simplicity and their capability of interacting with membranes, they are interesting models for drug design. LTX-315 from Lytix-Biopharma is an oncolytic peptide that is being developed to treat sarcoma disease. In its 9 residues sequence, it presents one non-natural amino-acid, the diphenylalanine, which is similar to the phenylalanine but presents two aromatic rings [86]. This small peptide is capable of penetrating the cell membrane, with a slight affinity for cancer cells. Once entered in the cell, it targets the mitochondrial membrane and permeabilizes it, disrupting the electron gradient and inducing the necrosis of the cell [204]. This mechanism not only kills the cancer cells but also stimulates the immune system response generating immune memory. LTX-315 represents a very promising strategy to treat cancer and is actually in phase II of clinical assays in Europe and US [<https://www.lytixbiopharma.com/>]. We have dedicated a part of this research work to study its action mechanism with a combination of experimental and computational techniques in collaboration with a Norwegian company LytixBiopharma and its CEO Oystein Rekdall. MD simulations are a key to the understanding of the peptide action mechanism. This technique can shed light on LTX-315 structural and dynamic characteristics with atomic detail. This information is crucial to understand its biomedical properties, and in the future it can serve to design new drugs of the same family.

There are still a lot to know about AH behaviour and functions: the specific contribution of the lipids to the match or mismatch between a sensor and a membrane; the “footprints” of the peptides in the membrane; the atomic interactions between the lateral chains and the lipids, etc. Neither a detailed NMR structure of a peptide in a detergent micelle nor a CD spectrum of the same peptide with a relevant lipid bilayer will give the answer. MD simulation are one of the most promising approaches to give detailed information at an atomic level of these processes.

AH/membrane interaction

When talking about AH, the mechanism of interaction and insertion into the membrane is called partitioning. The concept was established by Stephen White, a researcher who studied in depth AH thermodynamics using experimental methods [222]. Partitioning means that the thermodynamics of the system is not described such as a ligand (AH) which binds to a protein (membrane), but rather the AH can reside either in the aqueous phase or the membrane (such as in two phase systems, e.g. water / octanol). S. White described the thermodynamics of AH partitioning and folding in lipid bilayers. In solution AH, are normally unfolded [26, 222, 2]. S. White described the

partitioning and folding into the membrane as a coupled mechanism. In other words, they take place simultaneously.

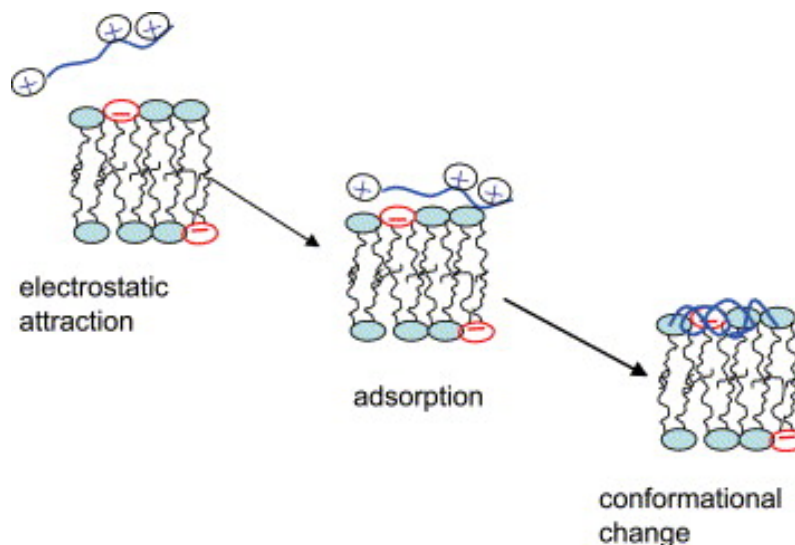


Figure 1.18 – Amphipathic helix/membrane interaction phases described by Seelig [195].

The general mechanism of AH/membrane interaction can be described in three steps: attraction, adsorption and folding (Fig. 1.18) [195]. The first step is the attraction, where the peptide approaches the membrane. This step is driven by electrostatic interactions mainly: the peptide is positively charged and the membrane generally presents negatively charged lipids. If the peptide sequence is not charged, the attraction will be driven by hydrophobic forces: the non-polar residues will tend to insert their hydrophobic chains between the aliphatic chains of the bilayer to avoid interacting with water. The closer the peptide to the membrane, the stronger the attraction.

The second step is the adsorption, where the peptide locates parallel to the membranes surface. The first interactions take place, but the residues are not inserted yet. As the peptide is still in the solvent, it presents an extended conformation.

Finally, the peptide will partition into the membrane while folding at the same time. Depending on the peptide properties, the insertion begins from one of the terminal ends or the center of the peptide. The residues are then inserted sequentially into the membrane. The hydrophobic ones will insert their side chains into deeper into the membrane while the polar ones will orientate their side chains to the solvent, trying to continue the interactions with water. This phenomenon will drive the folding of the peptide into an α -helix, so this is why the folding is coupled to the partitioning [223]. A model called nucleation propagation process was proposed to describe this mechanism: one hydrophobic residue gets inserted into the membrane and serves as a nucleation point for the helix folding. If the helix displays a repetitive pattern (similar composition of each turn), all of the folding steps should be associated with the same free energy [10].

Once it is partitioned and folded, the helix is positioned parallel to the membrane surface, on the lipid/water interface, at the glycerol level. The position and orientation of an AH was first

described in 1999 by experimental methods, specifically by X-Ray scattering [92]. The results are shown in figure 1.19.

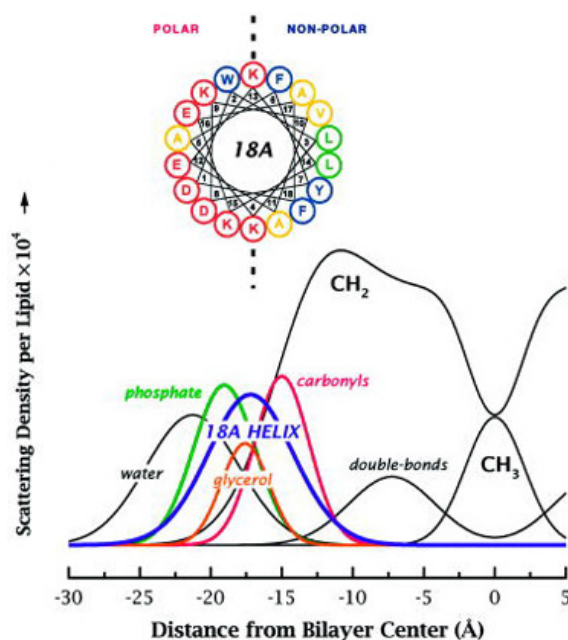


Figure 1.19 – Ac-18A-NH₂ partitioning in a DOPC bilayer. The first image is the peptide helical wheel. The colors correspond to the ones established in Wimley and White’s hydrophobicity scale (see next section). The dotted line shows the helix central axis that separates the polar and non-polar sides. The bottom figure is the electron density profile where the density of the main elements of the system are plotted on the *z* (vertical) axis of the membrane. This figure shows that the peptide is located at the glycerol level. Image from [92]

The thermodynamics of partitioning and folding

Describing the thermodynamics of the partitioning and folding of AH into the membranes is a key step in the understanding of its dynamics and its mechanism of action. In 1996, Wimley and White constructed a so-called hydrophobicity scale based on experimental assays on short peptides [223]. This scale describes the partitioning of the 20 amino-acids (AA) into the bilayer and it is based on the calculation of the free energy of binding of peptides whose sequence is Ac-WLXLL (where X is any of the 20 natural AA). Importantly, these peptides being very short, they can form any specific regular secondary structure, they are thus always in a coil conformation. In the scale, a value is attributed to each AA (Fig. 1.20). Lower values indicate higher tendency to partition into the membrane. In other words, the lower the value the more hydrophobic the residue. These values were determined from the partitioning of two series of small model peptides into the interfaces of neutral (zwitterionic) phospholipid membranes. Partition coefficients were measured using equilibrium dialysis and HPLC. Pure POPC LUVs were used as lipid bilayers. Nowadays, the White and Wimley hydrophobic scale is very much used and is a reference as important as other scales such as Kyte and Doolittle.

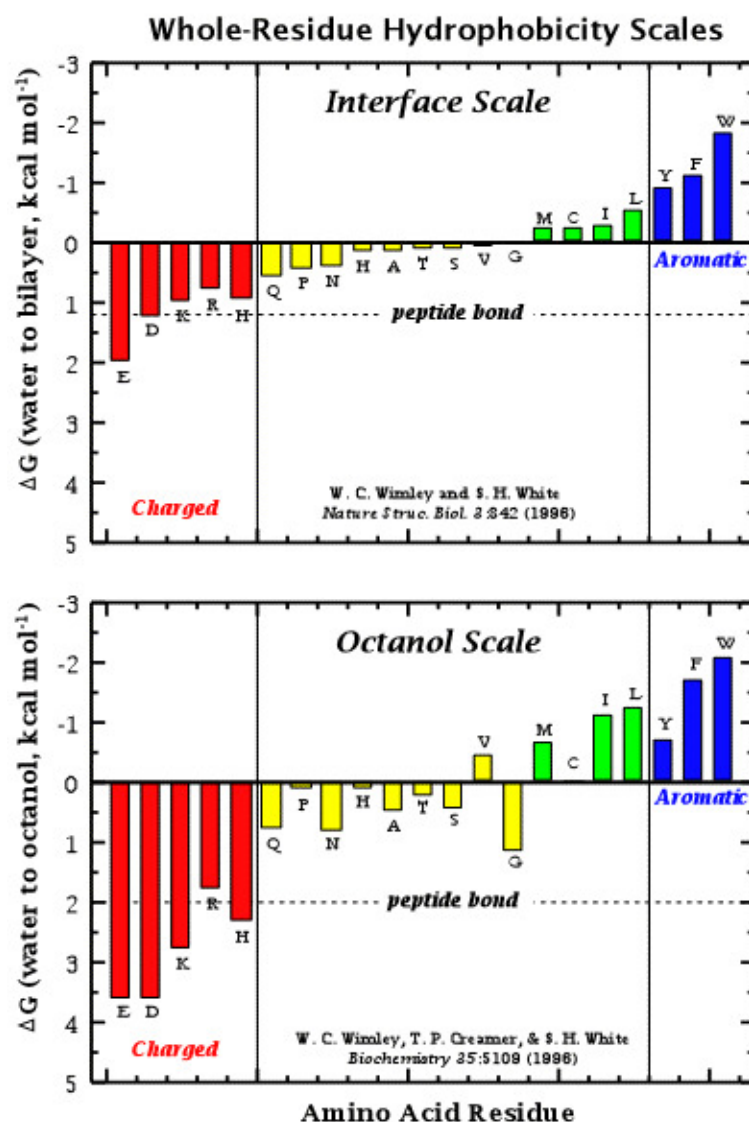


Figure 1.20 – Hydrophobicity scale determined by Wimley and White in 1996 [223].

The results shown in figure 1.20 indicate that the aromatic residues (tryptophan, phenylalanine and tyrosine) are highly hydrophobic and tend to insert their lateral chains into the membrane. However, the charged residues glutamate, aspartate, lysine and arginine present a positive free energy value, which means that they prefer to keep their lateral chains exposed to the solvent. As the free energy of partitioning is described for the 20 amino acids, this scale allows to calculate the free energy of partitioning of a peptide independently from the folding, and it has been used in literature to study the thermodynamics of binding of numerous peptides [2]

Two years later, S. White and co-workers proposed a way to evaluate, using experimental information, the free energy of folding of an AH in the membrane [222]. As the partitioning and folding is coupled, he proposed an innovative strategy to calculate the free energy of folding separately. First, he designed a thermodynamic cycle to describe the peptide thermodynamics (Fig. 1.21). The peptide presents three states: unfolded in water, unfolded in the membrane and

folded in the membrane. Using the Melittin as a model system, he calculated the thermodynamics of transitioning between the states. He first calculated the free energy of partitioning/folding coupling using titration CD assays (-8 kcal/mol). Then, he used the hydrophobic scale described above to calculate the free energy of partitioning independently (-3 to -2 kcal/mol). As stated above, the scale was determined for residues in a coil conformation, thus using the scale describes the process of inserting the peptide in a random coil conformation from water to the membrane. We know that the peptide is unfolded in water (from CD experiments), but up to this point the peptide is in a hypothetical state unfolded in the membrane. Thus it needs to fold to close the thermodynamic cycle. Therefore, we can deduce the difference between the free energy of partitioning and the free energy of the partitioning/folding coupling process, corresponding to the free energy of folding in the membrane: -6 to -11 kcal/mol .

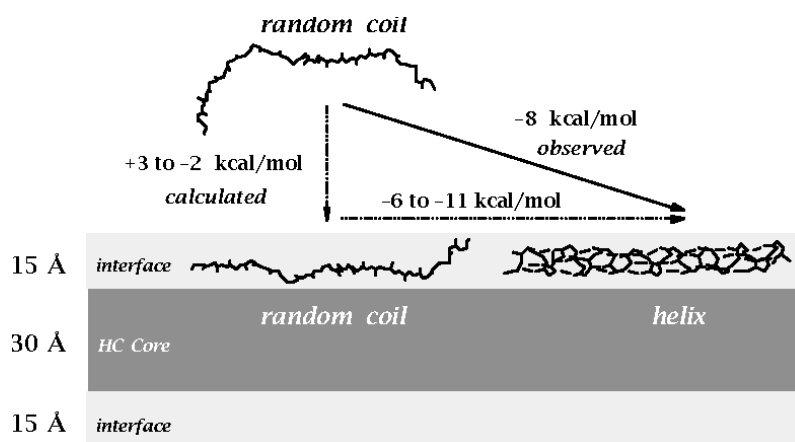


Figure 1.21 – Thermodynamics of an Amphipathic Helix partition and folding by S. White [222]

In 2012 the model was updated by Almeida *et al.*, using the Melittin from bee venom as model system. They described a four-states thermodynamic cycle shown in figure 1.22. The two reference states are the unfolded structure in water (A) and the folded structure in the membrane (D). State A is a hypothetical state, because peptides in aqueous solution usually have some, even if small, amounts helical structure. They also described two intermediate states: an ensemble of extended and folded structures in water (C), where the extended ones are predominant and an unfolded structure in the membrane (B). ΔG_{AB} is calculated using the Wimley-White hydrophobicity scale explained above, which gives a free energy value for the partitioning of each peptide into the water / membrane interface (assuming unfolded conformations). By adding the free energies of each residues of the peptide, the free energy of partitioning (ΔG_{AB}) is calculated. The energies ΔG_{CD} and ΔG_{AC} are accessible experimentally, for example by circular dichroism (CD). ΔG_{CD} represents the insertion/folding coupling of the peptide and can be evaluated by CD titration (recording CD spectra with increasing amount of liposomes). ΔG_{AC} is the energy required to constrain the peptide in state A from the ensemble C, and it is calculated by doing titration experiments of the peptide in water with various amounts of TFE (trifluoroethanol). By fitting the final curve, we can get the equilibrium constant K_α and the final free energy from A to C is

computed with $\Delta G_{AC} = -RT \ln(1 + K_\alpha)$. Finally, once we have these three free energies, we can deduce the one we are interested, that is, the free energy of folding in the membrane, thanks to the thermodynamic cycle: $\Delta G_{BD} = \Delta G_{AC} + \Delta G_{CD} - \Delta G_{AB}$.

From this model, Almeida *et al.* determined a free energy value for the folding of any single amino acid in the membrane: $\Delta G_{res} = -0.37 \pm 0.02$ kcal/mol [3]. ΔG_{res} does not depend upon on the hydrophobic moment ($\mu(H)$), as it is ascribed to the formation of backbone hydrogen bonds. This is a very relevant information for AH research as it provides thermodynamic information calculated experimentally that can be used as reference for computational research, as it will be shown later in this manuscript.

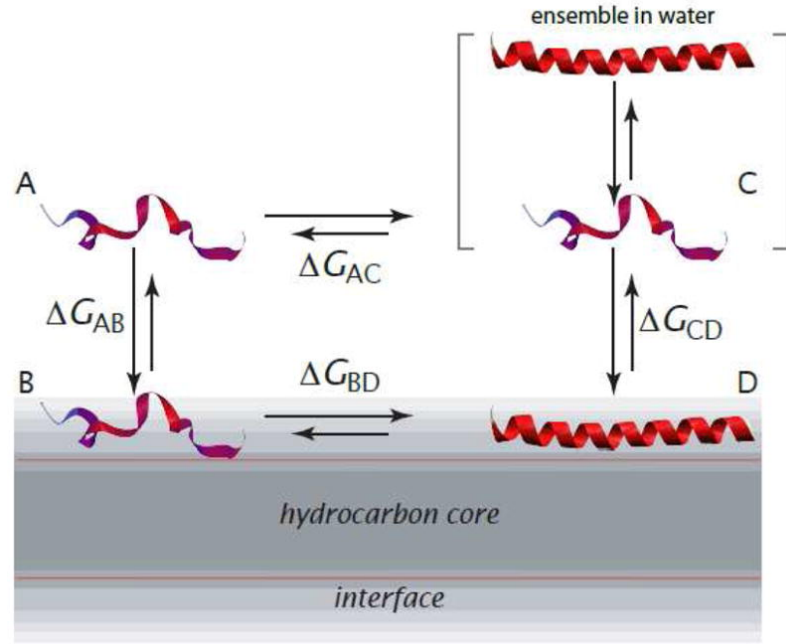


Figure 1.22 – Thermodynamic cycle of AH partitioning and folding in the membrane [3]. (A) Unfolded structure in water. (B) Unfolded structure in the membrane. (C) Ensemble of folded and unfolded structures in water. (D) Folded structure in the membrane. States A and B are transition conformations not stable enough to be observed by experimental assays. ΔG_{AB} is computed using Wimley and White Hydrophobic scale. The other free-energy differences are determined experimentally.

1.5 AH/membrane interactions studied by MD

Molecular dynamics (MD) is a computational technique that consists in performing computer simulations of a molecular system to predict its structural and dynamic properties. The first step is to construct a 3D model of the molecular system of interest. Then, we use a Force Field (FF) to compute the potential energy of the system. FF is an ensemble of parameters and equations that allow to predict many properties of our system. By applying the algorithm of MD (solving

Newton’s equation of motion), we get access to the movement of the system particles. This way, we get to observe the behaviour of the molecules through time in certain conditions. Lots of different conditions and analyses can be applied and the quantity of information that we get about the system is huge. This technique is having a fast evolution, becoming a key tool in biophysical research. One talk now of computational biophysics. Studying all complex biological phenomena *in silico* is a dream, bypassing the enormous experimental challenges and their associated costs. MD already achieved some great successes such as protein folding, protein-protein interactions, peptide/membrane interactions, membrane fusion, etc., all at atomic resolution. For example, MD simulations have succeeded in capturing the partitioning, folding and pore formation of membrane-active peptides (MAPs) into a lipid bilayer [12]. The advance in computational power, the increase in the efficiency of algorithms and techniques and the clever accelerated sampling schemes decrease the cost of MD simulations and allow to study more complex processes and longer time scales. Also the more accurate parametrization of the chemical interactions improved the reliability of the results [211].

MD simulations have succeeded in accurately capturing the process of peptide binding, folding, and partitioning into lipid bilayers as well as revealing how channels form spontaneously from polypeptide fragments and conduct ionic and other cargo across membranes, all at atomic resolution, in microsecond-timescales [43].

Experimental methods have been used to study membrane dynamics. However, these methods can reveal only stable structures that persist over long time scales while membranes dynamics often involve transient structures because of the fluid nature of the lipid bilayer. For example, very stable structures have been resolved by crystallographic techniques but no transient structures formed by peptides and proteins that inhabit both aqueous and membrane domains [211]. Also, experimental techniques are often limited by spatial and temporal resolution [43].

The first bilayer simulations were done using implicit membrane models (IMMs) [69, 213]. It consists in introducing a planar hydrophobic (i.e., solvent-excluded) zone. This reduces the number of atomic interactions needed to be evaluated, decreasing the computational cost and vastly increasing the simulation time scales. This method allowed to simulate peptide/membrane interactions with successful applications to, e.g., the M2 helices of influenza A and the nicotinic acetylcholine receptor, virus protein U, sarcolipin, glycophorin A, and individual helices of bacteriorhodopsin. However, important atomic details of protein/lipid and protein/water hydrogen bonding are missing so it is difficult to mimic different membrane compositions with these kind of models. Despite these difficulties, Lazaridis and co-workers could design smart methods that take the effect of lateral pressure/curvature stress in implicit membrane models [237].

One popular approach is the application of models that simplifies the molecule representation as Coarse-Grained (CG), hybrid, and multiscale methods to study membrane phenomena (Fig. 1.23). In these methods, the reduced representation allows for much-extended sampling time scale [32].

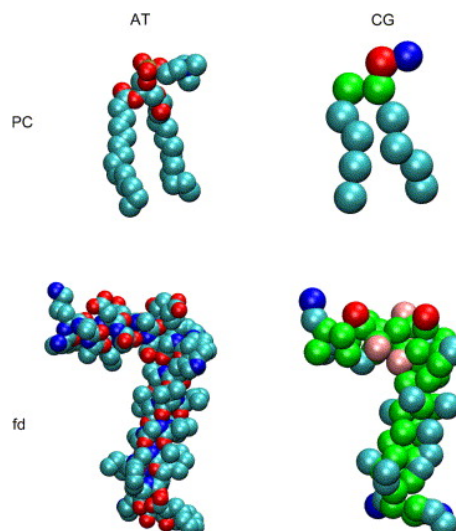


Figure 1.23 – All-atom and coarse-grained model comparison. First, two lipid structures are showed: AA DMPC and CG DPPC. The fd phage coat protein is also shown in AT and CG format. In AA, the colors are relative to the atom type. In the CG models the particles are coloured according to the following scheme: green, mixed polar/non-polar bead; cyan, hydrophobic bead; red/blue, negative/positive, charged bead; and pink, polar bead. Image from [32]

CG models, in which small groups of atoms are merged into larger particles (called beads), enable extended timescales to be studied. They have been used to characterize a number of peptides and proteins interactions with lipid bilayers [211]. For example, CG-MD was used to simulate the insertion of synthetic model membrane peptides (WALPs and LS3) into a lipid (PC) bilayer. WALP peptides insert in a transmembrane orientation, while the LS3 peptide adopts an interfacial location, both in agreement with experimental biophysical data. This approach was extended to a transmembrane fragment of the Vpu protein from HIV-1, and to the coat protein from fd phage showed in figure 1.23 [32]. Again, simulated protein/membrane interactions are in good agreement with solid state NMR data for these proteins. Taken together, these results demonstrate the utility of CG-MD simulations for studies of membrane/protein interactions.

Even if CG-MD simulations have been proven useful to study protein/membrane interactions, understanding some mechanism require atomic representation of the molecules. For these reasons, enhanced-sampling techniques for all-atom (AA) simulations were developed. In 2005, van der Spoel research group used Temperature-Replica Exchange Molecular Dynamics (T-REMD) to study the folding of a polypeptide [197] (This technique is explained in depth in Materials and Methods chapter). They used both, classical MD simulations and enhanced T-REMD simulations in explicit solvent. From the classical simulations the folding time was estimated to 1–2 μ s. The REMD simulations allow enough sampling to deduce the folding free energy. The global minimum of the energy landscape corresponds to the structure resolved by NMR assays. Starting from an extended state it takes about 50 ns before the native structure appears in the REMD simulations, about an order of magnitude faster than conventional MD.

The same year, a T-REMD simulation of an α -helical peptide within an explicit lipid bilayer was carried out [158]. They studied the interactions between a model peptide (WALP-16) and an pure DPPC membrane bilayer. They observed the insertion and folding of the peptide with a transbilayer orientation. The free energy surface suggests that the insertion of the peptide precedes secondary structure formation. They suggested that membrane peptides may have a diversity of insertion/folding behaviors depending on the exact membrane composition.

Unfortunately, T-REMD simulations go through temperature changes which render difficult to evaluate the kinetics. To overcome this problem, a trivial strategy was used, the elevated-temperature MD folding-partitioning simulation method, a useful tool for revealing the atomic-detail mechanisms for thermostable peptides. The advantage of this method is that it allows precise quantification of folding-partitioning kinetics as well as membrane transfer free energies while enhancing the sampling. Elevated-temperature simulations can be applied to membrane peptides because the bilayer protects peptides secondary structure, preventing denaturation. It acts as a conformational restraint on the peptide structure that can be exploited to vastly speed up the kinetics. Another important fact is that in MD simulations, the solvent retains liquid-state conditions above the boiling point of water because phase transitions are not possible using conventional pressure and temperature coupling algorithms. The solvent acts then as a super heated liquid, allowing reliable simulations up to 200°C and the membrane remains intact at these temperatures [212, 44, 211].

Martin and Jakob Ulmschneider have been pioneers in the use of this approach [212, 44]. They applied elevated-temperature simulations to two cases of study: synthetic peptides from WALP family (W16 and W23) and melittin. With WALP peptides, they observed the whole process of insertion and folding in the membrane (Fig. 1.24) at 80°C and 200°C. The structural results were validated with CD assays at high temperature (45-90°C). This first set of experiments proved the thermostability of the peptides in DCCP LUVs [212]. Melittin simulations were carried out at 120°C to observe its interaction and folding in the membrane/water interface. CD assays were again carried out at high temperature (25-95°C) in POPC LUVs. Results showed the spectra of a thermostable α -helix. Its positioning in the simulations (density profile) also agrees with X-ray scattering data [44].

Thermodynamic information was calculated from the simulations. They assumed that the kinetics follows the Arrhenius law to extrapolate the transition rates at high temperature to 30°C. When a process follows Arrhenius kinetics, the log of the rate has a linear dependence with the inverse of the temperature. Kinetics of WALP peptides insertion were deduced to last for more than 100 *ms* at physiological temperature, a time scale which is impossible to reach using conventional MD simulations at present. Similar time scales were deduced for Melittin insertion. This enhanced sampling technique allowed to observe slow process at an atomic level with low computational cost.

Although interesting and easy to apply, we believe this approach has some important drawbacks. First, it is applicable to thermostable peptides only. By definition, if the peptide is not

thermostable, high temperature simulations will not stabilize the most stable conformation at room temperature. Second, the behavior of water and (even more) lipids can be questioned at such elevated temperatures for which the FF has not necessarily be designed for. Ideally, a method that works at room temperature or that would relax the lipids to the room temperature would avoid this issue. Last, assuming the kinetics follows an Arrhenius law is absolutely not guaranteed for complex systems such as AH/lipids. For all these reasons, new biased or non-biased techniques are needed to be developed to obtain thermodynamic and kinetic data of AH partitioning/folding.

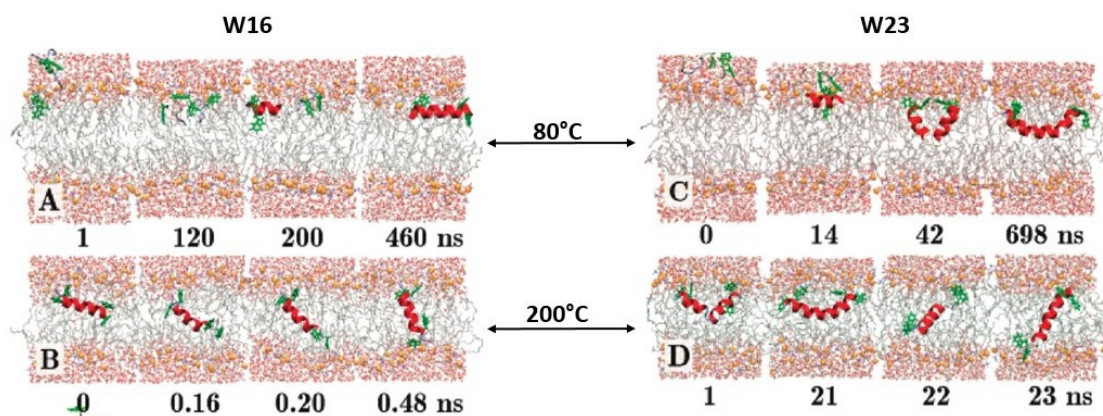
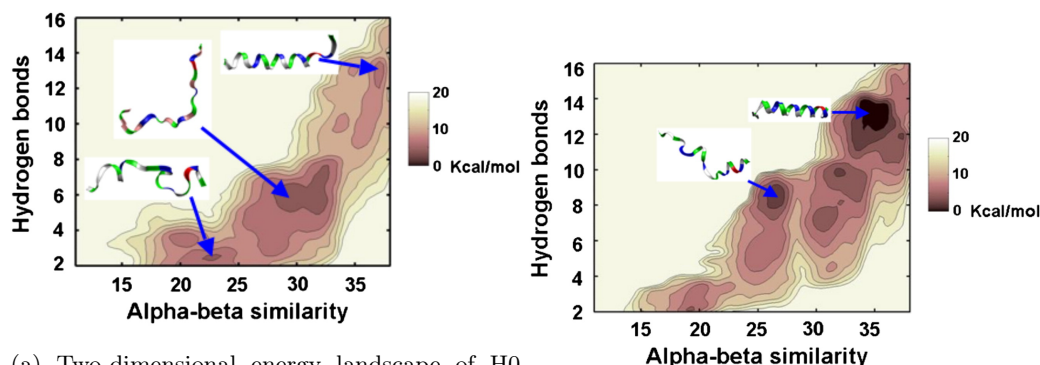


Figure 1.24 – Snapshots of WALPS simulations at 80 and 200°C. On the left, results for W16 peptide. On the right, results for W23 peptide. These simulations show the acceleration of the processes when the temperature increases. Bigger time scales are observed at high temperatures. Image adapted from [212]

The number of enhanced sampling techniques for MD simulations are numerous. Greg Voth employed in 2011 metadynamics to study the mechanism of membrane curvature sensing by AH [55]. Metadynamics is another approach to speed up slow processes where a history-dependent bias potential is imposed on the system during the simulation, encouraging the system to visit configurations which have not already been sampled along one (or more) collective variable(s). This way, the system is forced to sample all available conformations and generate a complete free energy landscape for the chosen collective variables [135, 127]. This method was applied to study the amphipathic N-terminal helix of endophilin (H0), which targets curved membranes by binding to packing defects which increase in number with increasing membrane curvature. It was found that in the presence of large defects, folding of H0 is favored by 3 *kcal/mol* approximately, while in bulk solvent H0 folding is disfavored by 2 *kcal/mol*. When large defects are rare like in flat membranes, H0 binds the membrane in an unfolded configuration, becoming kinetically trapped and unable to find the folded state [55].

Some other works used umbrella sampling and the MARTINI coarse-grained force field. The idea was to calculate a PMF (potential of mean force) for inserting the helix (TM or AH) from the bulk to the membrane [31, 46, 81]. Self-assembly simulations were used as well [85]. This approach may be appealing in terms of sampling, coarse-grained PMFs are easier to converge, but

it presents a major drawback: the helix in MARTINI has to be maintained rigid. Thus the free energy obtained is between a hypothetical fully helical state in water, and the inserted fully helical state in the membrane. A more interesting approach used H-REMD (Hamiltonian REMD) and another CG force field (named PLUM), which allow the secondary structure to change, on the insertion and folding of WALP (TM) peptides [22]. The free energy of folding (from coil to helix) of WALP peptides could be evaluated in water and in the membrane. In the membrane, the helix is as expected preferred (between 15 to 30 kcal/mol). More surprising is the case in water: the helix is preferred as well (between 10 to 23 kcal/mol), however the authors note the need to be cautious with this result partly confirmed by all-atom simulations. Unfortunately, this approach with the PLUM CG FF has never been applied to AHs.



(a) Two-dimensional energy landscape of H0 peptide in bulk solvent. The most stable state is the center of the three basins, the extended peptide in a curved membrane. Large packing defects promote folding of H0 into an α -helix.

Figure 1.25 – Study of AH sensing membrane curvature by metadynamics simulations. The number of hydrogen bonds and the Alpha-beta similarity were chosen as reaction coordinates [55].

To conclude with the enhanced sampling techniques, Kalathingal *et al.* have recently published a study of the membrane insertion process of pTB explored in detail combining non-biased AA-MD simulations with replica exchange umbrella sampling simulations (REUS) [115]. Umbrella sampling (US) simulations use harmonic biasing potentials, which is referred to as umbrella potentials. Adding the umbrella potential forces a chosen reaction coordinate to adopt a value on a certain range so the system gets to explore a region of the energy landscape which may not be visited during regular MD simulations. REUS simulations are similar to T-REMD but the temperature of each replica has the same value and different umbrella potentials are exchanged [99]. In the publication, they used REUS simulations to calculate the free energy of the insertion and they showed that there is a small barrier of 4.3 kcal/mol for the membrane insertion of pTB from bulk water [115].

Apart from the enhanced sampling techniques, there are other methods that help exploit information obtained from MD simulations. Markov State Models (MSM) are a mathematical formalism that can be applied to classical non-biased MD simulations to extract kinetic, thermodynamical

and structural information. They have been applied to study numerous systems. In 2017, the first protein-protein interaction model was build [170], describing barnase-barstar interactions from non-biased AA simulations. In 2018, Gerhard Stock designed a novel strategy to study protein dynamics by combining metadynamics with MSM [28]. He used metadynamics simulations to obtain well-distributed initial structures that served to launch classical MD simulations for MSM construction. This way, he optimized MSM construction. This approach was inspiring and a similar strategy will be presented in this manuscript.

MD simulations techniques are constantly evolving and improving. Providing reliable predictions of peptide/membrane phenomena is now a reality and is transforming the way peptide sequences are designed, selected, and optimized for biomedical applications. Also simple model membranes are increasingly being replaced by natural membranes containing the full plethora of different lipid types, proteins, and auxiliary molecules thanks to the computational cost decrease of the simulations. Validating the simulations with data obtained *in vitro* is a key step in computational research. However, matching *in vivo* measurements is much more challenging and will be the ultimate test of the utility of MD as a general research tool [211].

Chapter 2

Materials and Methods

2.1 Computational Methods

2.1.1 Molecular Dynamics

Introduction

Molecular Dynamics (MD) is the main technique employed in this PhD project. It allows to describe the behaviour of a molecular system through time. MD consist in performing computer simulations that predict the dynamics of molecules within certain conditions. The simulations allow the study of the dynamical behaviour of molecules with an atomic description and short time scales. MD can provide detailed and useful information about a molecular model not reachable with experimental assays as shown in figure 2.1. However, the information is obtained computationally and it must be validated. Normally, the validation is done by comparison with experimentally obtained information. The number of application of this techniques are countless: study of protein or membrane dynamics, membrane / protein interactions, protein / protein interactions, etc.

Molecular Models

The molecular model determines the degree of detail in which the molecular system is described. Figure 2.2 shows the different possibilities. The All-Atom (AA) model is the more detailed one. It represents one bead for each atom of the molecule. In a United-Atom model, each heavy atom is represented with its apolar hydrogen atoms merged in a single bead. Polar hydrogen atoms are still explicit, which is useful for simulating hydrogen-bonds. One way to simplify even more is to use Coarse-Grained (CG) models. Here, a few (e.g. four) heavy atoms with their corresponding hydrogen atoms are represented as one bead. One can go even further in the simplification with a Super Coarse-Grain Model, where many heavy atoms and their hydrogen atoms are merged into a single bead. There are many different other ways to coarse-grain a system (e.g. using a lattice,

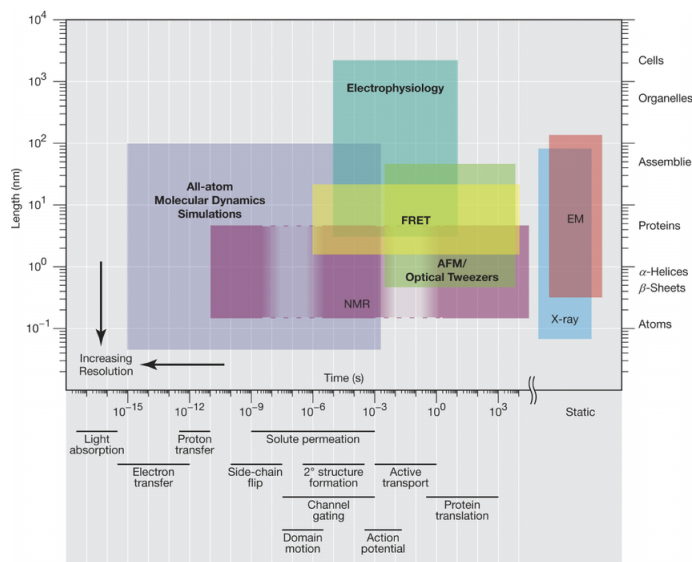


Figure 2.1 – Spatio-temporal resolution of various biophysical techniques. MD time scales are compared to those from experimental techniques. Time scales achievable by MD simulations are much smaller than the ones from experimental assays. Image from [64]

etc.). The choice of the model depends on the detail level required for the study, not all the properties need to be studied at an atomic level. In addition to this, by simplifying the system the computational cost decreases and so the simulation time and system size can be increased.

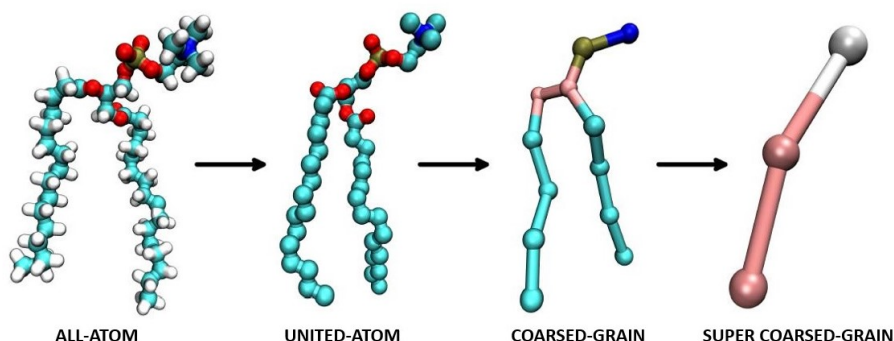


Figure 2.2 – Molecular Models used in MD simulations. Credit image: <http://2015.igem.org/>

The Force Field

The Force Field (FF) used for performing an MD simulation is the ensemble of parameters and equations that serve to evaluate the potential energy of the system. The potential energy is very important as it dictates the distribution of the conformations of the molecular system and allows to predict structural, thermodynamic and kinetic properties.

In MD, we use empirical FFs that allow to calculate the potential energy from the particle's coordinates of the molecular system. They are composed of two classes of terms: bonded and non-bonded interactions (Fig. 2.3). There are four bonded terms: the bonds between two consecutive

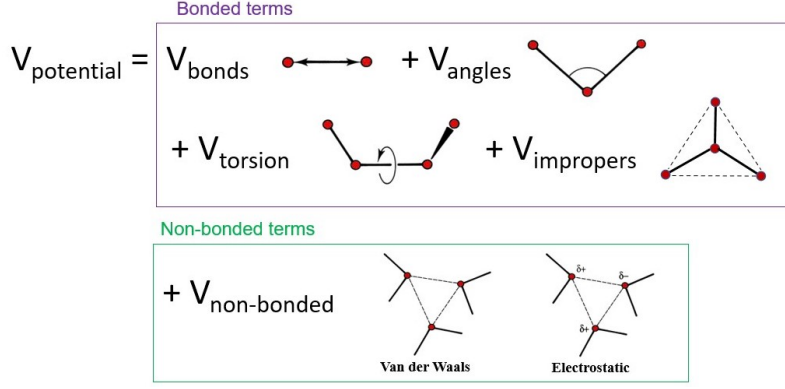


Figure 2.3 – Force Field terms illustration.

atoms, the angles between three consecutive atoms, the dihedral angles between four consecutive atoms and the improper dihedrals. The improper dihedrals are angles between four atoms that are meant to keep the planarity of planar groups as aromatic rings or prevent molecules from flipping over to their mirror images. The non-bonded terms are the electrostatic and van der Waals (VdW) interactions. The sum of the bonded and non-bonded terms gives the potential energy of the system:

$$\begin{aligned}
 V_{\text{potential}} = & \sum_{\text{bonds}} \frac{k_l}{2} (l - l_0)^2 + \sum_{\text{angles}} \frac{k_\theta}{2} (\theta - \theta_0)^2 \\
 & + \sum_{\text{dihedrals}} \sum_{m=1}^M \frac{V_m}{2} [1 + \cos(m\omega - \gamma)] + \sum_{\text{impropers}} \frac{k_\Omega}{2} (\Omega - \Omega_0)^2 \\
 & + \sum_{i=1}^{N-1} \sum_{j=i+1}^N 4\epsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^6 \right] + \sum_{i=1}^{N-1} \sum_{j=i+1}^N \frac{q_i q_j}{4\pi\epsilon_0 r_{ij}}
 \end{aligned} \tag{2.1}$$

where:

- (Bonds) l is the actual distance between the atoms, l_0 is the equilibrium bond length and k_l is the force constant.
- (Angles) θ is the angle between the atoms, θ_0 the equilibrium angle and k_θ the force constant.
- (Dihedrals) V_m is the dihedral amplitude, m is the dihedral multiplicity ω is the dihedral angle and γ is a phase factor .
- (Impropers) k_ω is the force constant, ω is the improper angle value and ω_0 is the value at the equilibrium.
- (VdW) i and j are the two atoms , ϵ_{ij} the Lennard-Jones well-depth, r_{ij} is the the distance between two atoms and σ the Lennard-Jones radius.
- (Electrostatic) q_i and q_j are the partial atomic charges and ϵ_0 the dielectric constant.

It is important to know that non-bonded interactions are not calculated between atoms that are directly bonded or separated by two apart. For pairs of atoms that are three bonds apart, the non-bonded terms are attenuated. These are generally called 1-4 interactions. These 1-4 terms are usually here to supplement the dihedral angles in order to get the right dihedral distributions. The way the 1-4 interactions are attenuated with respect to normal non-bonded interactions is force field dependant. For example, in OPLS [112], 1-4 interactions are merely divided by 2 compared to regular electrostatic and van der Waals interactions.

In principle, we should calculate the non-bonded interactions between all possible atom pairs. However, because these interactions decay quite rapidly with interatomic distance, one usually uses a cut-off (CO) distance. The idea is to avoid the calculation of interactions that are further apart the CO. This allows to decrease the computational cost and the use of Periodic Boundary Conditions (see section 2.1.1) or PBCs. PBCs are very convenient for the simulations of liquids. In order to avoid a step function, which is problematic when deriving the potential energy for force calculations (because it is not continuous), one usually uses shift or switch functions (see figure 2.4).

However, the use of a finite CO produces severe artefacts on highly charged systems [9] (e.g. nucleic acids, lipids, ionic liquids, etc.). Pairs of charged molecules tend to accumulate at the CO distances, while they should not. One accepted solution which is commonly used nowadays is the Particle Mesh Ewald (PME) algorithm [57, 67]. The idea dates back to the so-called Ewald summation [68] with some further optimizations to reduce the computational cost. The principle is to compute the contribution beyond the cutoff (see figure 2.4) using some computational tricks. The simulation box is assumed to be replicated in all directions of space like in a crystal. The total electrostatic interactions are then divided in two parts: a first part which is calculated in real space using the classical coulomb law (Eq. 2.1) up to a rather short cutoff (usually 10 Å), and a second part for all atoms beyond the CO which is calculated in Fourier space (as in a crystal). This last part is calculated using a grid (hence the term Mesh in PME) allowing the use of the fast Fourier transform (FFT) leading to an $N \log N$ complexity (where N is the number of atoms). In this project, PME is used for electrostatic interactions and a switch function for vdW interactions.

No existing FF possesses all the required parameters for simulating any possible molecules. There exist different FF specialized in some ensemble of molecules / macromolecules. For example, AMBER [53] is specialized in nucleic acids and proteins (among others). These FF are usually called "empirical" because they rely on simple functional form (see equation above) generally fitted against experimental data and/or more sophisticated quantum calculations. Two FFs are used to carry out the MD simulations presented in this manuscript. CHARMM36 [123] FF is employed for All-Atom (AA) simulations as is one of the most reliable FF existing today and it contains the wider ensemble of molecules. Martini22 [109] is used for Coarse-Grained (CG) simulations as it allows to simplify the systems and increase the timescales (Fig. 2.2). The difference between AA and CG simulations is explained in "Molecular Models" paragraph.

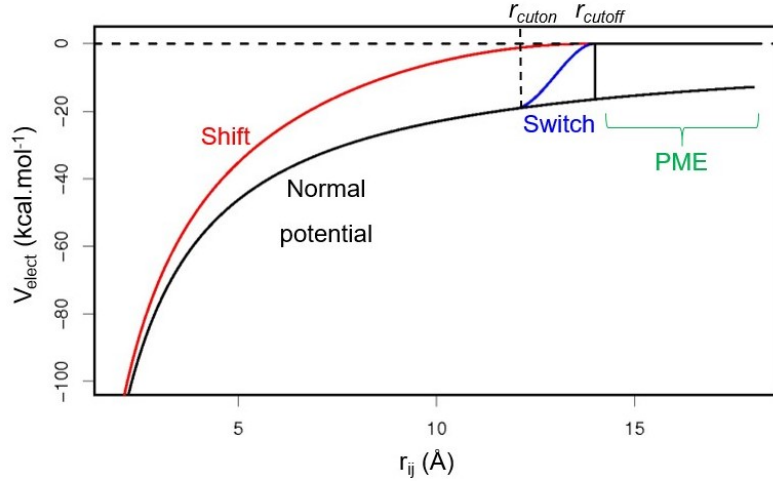


Figure 2.4 – Cut-off application for long-range electrostatic interactions. Shift (red) progressively modifies the potential to arrive to a null value at the CO distance. Switch (blue) establish a cut-on distance to make the interruption of the potential smoother. PME (green) is a way to take into account the neglected part beyond the CO.

The molecular dynamics principle

The FF allows to calculate the potential energy ($V_{potential}$) of the system, which we can simply call V in the following. It is possible to compute the force, $\mathbf{F}_i = (\mathbf{F}_{ix}, \mathbf{F}_{iy}, \mathbf{F}_{iz})$, acting on any atom i in the system by deriving the potential energy with respect to its position $\mathbf{r}_i = (x_i, y_i, z_i)$:

$$\mathbf{F}_i = -\frac{\partial V}{\partial \mathbf{r}_i} \quad (2.2)$$

This force can then be injected in Newton's equation (second law):

$$\mathbf{F}_i = m_i \cdot \mathbf{a}_i \quad (2.3)$$

We can rewrite Newton's equation knowing that the position of a particle i is the second derivative of its acceleration \mathbf{a}_i :

$$\frac{\mathbf{F}_i}{m_i} = \frac{\partial^2 \mathbf{r}_i}{\partial t^2} \quad (2.4)$$

This means that knowing the mass m_i of a particle and the force \mathbf{F}_i acting on it, it is possible to obtain its new position \mathbf{r}_i after a small fraction of time δt (i.e. the time step, see below). The Newton equation is a differential equation which needs to be solved. This can be done analytically for simple systems (such as the harmonic oscillator), but most of the time is solved numerically with specific algorithms (see below). By successively iterating between equation 2.4 which give new positions, and equation 2.2 which compute forces on a given system conformation, one thus obtains a molecular dynamics trajectory, that is, a succession of frames allowing the observation

of the system dynamics. There are several algorithms to integrate numerically equation 2.4. In this project, the Leap Frog algorithm [88] is applied (Fig. 2.5). It calculates the positions of the particles at a time t , and the velocities at time $t + \delta t$:

$$\begin{aligned} \mathbf{v}_i(t + \frac{1}{2}\delta t) &= \mathbf{v}_i(t - \frac{1}{2}\delta t) + \delta t \mathbf{a}_i(t) \\ \mathbf{r}_i(t + \delta t) &= \mathbf{r}_i(t) + \mathbf{v}_i(t + \frac{1}{2}\delta t)\delta t \end{aligned} \quad (2.5)$$

δt is called the time step. It is a critical parameter in MD and needs to be chosen carefully. It must be higher than the fastest movement but not too high to avoid abrupt results. Normally, the fastest movements are the bond vibrations. In the simulations showed in this manuscript they have been "frozen" to increase the time step and decrease the computational cost of the simulation. The algorithm employed here is called LINCS [87] as implemented in the MD software GROMACS [215, 1] used throughout this thesis. The time step used in this project is 2 *fs* and 20 *fs*, for all-atom and coarse-grained molecular models respectively.

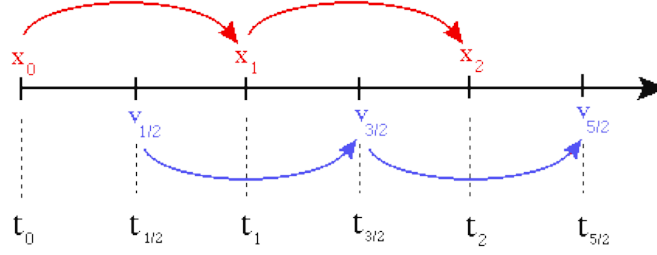


Figure 2.5 – Leap Frog algorithm. Image adapted from: <http://www.physics.drexel.edu>

Energies of the system

The total energy of the system is the sum of the kinetic energy and the potential energy:

$$E_{tot} = E_{kin} + E_{pot} \quad (2.6)$$

The kinetic energy comes from the particles movement and depends on their masses and velocities:

$$E_{kin} = \frac{1}{2} \sum_{i=1}^N m_i v_i^2 \quad (2.7)$$

where m_i is the mass of particle i , and v_i its speed (i.e. the magnitude of the velocity vector). According to the theorem of equipartition of energy, E_{kin} can also be written:

$$\begin{aligned} E_{kin} &= \frac{1}{2} N_{df} k_B T \\ N_{df} &= 3N - N_c \end{aligned} \quad (2.8)$$

where N_{df} is the number of degrees of freedom, k_B the Boltzmann constant, T the absolute temperature, N the number of particles and N_c the number of constraints. A constraint is a degree of freedom that has been frozen by forcing the variable to adopt a fixed value preset by the modeler (e.g. constraints on bonds of fixed length or removing the global motion of the system center of mass). For example, in the CHARMM36 force field, all bonds involving an a heavy atom / hydrogen atom are constrained. In this work, the LINCS algorithm [87] has been used for that. Equation 2.8 shows how the energy of the system is related to the temperature. The total energy E_{tot} is a central quantity in statistical mechanics. When the system we simulate is completely closed, no exchange of energy with the exterior (see microcanonical ensemble in the next paragraph below), E_{total} should be constant. By default, solving Newton's equation as shown above produces such an ensemble. When developing new algorithms, such as integration schemes, this energy conservation is an absolutely required test. In the so-called canonical ensemble (see next paragraph below), E_{tot} is directly related to the probability (p_i) of observing the system in a certain state i at a temperature T and defines the Boltzmann distribution:

$$p_i = \frac{e^{-\frac{E_i}{k_B T}}}{Z} \quad (2.9)$$

where k_B is the Boltzmann constant and Z is a normalization constant called the partition function, E_i is the total energy of conformation i . This central equation in statistical mechanics clearly shows that the probability of conformation is directly related to its energy: the lower E_i the higher p_i . Unfortunately, the partition function Z is extremely difficult to calculate on real systems. Some special techniques have been developed though. As for a given molecular system Z is constant, one can immediately see that we can get an accurate probability ratio p_i/p_j between two conformations i and j since the Z cancel out (assuming the thermodynamics parameters are identical).

Thermodynamic Ensembles

A thermodynamic ensemble is a group of macroscopic conditions that should be maintained during an MD simulation. Newton's dynamics defines a closed system where there is no energy exchange with the exterior, neither in form of heat or work, giving thus a constant total Energy (see previous paragraph). In this case, simulations are carried out in so-called NVE conditions, other way called the microcanonical ensemble. The number of particles (N), the volume of the box (V) and the (total) energy of the system (E) remain strictly constant throughout the simulation. Newton's conditions are interesting at the theoretical point of view, however, one would like to recreate conditions closer to the physiological ones or those of a test tube, more generally of the experimental methods we want to compare the simulations with. To recreate these conditions, we have to control temperature and pressure. Normally, the pressure is set at 1 bar and the temperature at 300 K approximately. The thermodynamic ensembles determine the conditions that are going to be controlled. In this project two thermodynamic ensembles are used: the canonical and the

isothermal / isobaric ones. The canonical ensemble corresponds to NVT conditions where the temperature (T) is maintained constant instead of the total energy. Therefore, the system has to exchange heat with a reservoir. The isothermal / isobaric corresponds to NPT conditions, where the volume varies but the pressure (P) is maintained constant instead. Here the system exchanges work with a reservoir. It is important to keep in mind that in NPT conditions, maintaining temperature and pressure (of temperature alone for NVT) still leads to fluctuations of both of them (of temperature for NVT). The use of a thermostat and barostat is useful when one wants to impose some specific conditions of an experiment, e.g. heating or cooling a system.

The Temperature Temperature is related to kinetic energy as shown in equations 2.7 and 2.8. In consequence, thermostats scale velocities to control the temperature. Scaling up the velocities corresponds to heating, scaling down to cooling. In this project two different algorithms have been used for the thermostat. Berendsen thermostat [23] adjusts the temperature at each step of the simulation of a ratio proportional to the difference between the current temperature and the reference temperature. It achieves quickly the desired temperature but does not reproduce properly the real temperature fluctuations. V-rescale [38] is the other thermostat that has been employed. It is slower at relaxing to the wanted temperature than Berendsen, but it uses an additional stochastic term in the equation of motion to assure reliable temperature fluctuations.

The Pressure The pressure (\mathbf{P}) of the system is related to the volume (V) of the box and is calculated as a tensor product (\otimes):

$$\mathbf{P} = \frac{1}{V} \sum_{i=1}^N m_i \mathbf{v}_i \otimes \mathbf{v}_i + \sum_{i<j} \mathbf{r}_{ij} \otimes \mathbf{F}_{ij} \quad (2.10)$$

where the first term corresponds to the kinetic energy and the second to the virial (contribution due to cohesive forces between particles), \mathbf{v}_i is the velocity vector of particle i , \mathbf{r}_{ij} the distance vector between particles i and j , \mathbf{F}_{ij} is the force acting on i due to j . In consequence, the pressure is a 3x3 tensor that represents the pressure in all the directions space.

$$\mathbf{P} = \begin{bmatrix} P_{xx} & P_{xy} & P_{xz} \\ P_{yx} & P_{yy} & P_{yz} \\ P_{zx} & P_{zy} & P_{zz} \end{bmatrix} \quad (2.11)$$

One can define a scalar pressure $P = \text{trace}(\mathbf{P})/3$. In general, the off-diagonal values are maintained to zero to avoid deformations of the simulation box such as in shear flows (the box no longer stays a cube or parallelepiped).

To control the pressure, the volume and the relative positions of the particles are re-adjusted. There are different ways to do this. If the volume adjustment is equal in the three directions of space, the coupling is isotropic. The isotropic coupling is generally used to simulate isotropic

liquids such as water. In this project it was used for systems containing a peptide in a water box. If the volume adjustment is coupled in the x, y directions but scaled independently in z , the coupling is semiisotropic. It is more appropriated for membrane simulations as it allows to impose a zero surface tension on the membrane, corresponding to equilibrium conditions. Under these conditions, the membrane structure should relax its area to its equilibrium value. This is a good test for lipid force fields when simulating a single lipid composition, as the area per lipid should relax to the experimental value (see section 2.1.2). When the volume scaling is independent in the three directions of the space the coupling is anisotropic. Anisotropic coupling is not used in this project.

Two different barostats are employed for this research. Berendsen barostat [23] uses the same principle as Berendsen thermostat but adjusts the volume. It relaxes quickly to the desired pressure but it has wrong fluctuations. Parrinello-Rahman barostat [160] introduces an additional term to the equation of motion to control pressure. As v-rescale, it is slower at relaxing to the desired pressure than Berendsen algorithm, but produces more reliable fluctuations, maintaining the isothermal / isobaric ensemble.

Water Models

Water is a difficult molecule to model. Solvent atoms are the most numerous in the simulation box and computing their interactions represent a big part of the computational cost of the simulation. For this reason, it is crucial to chose a water model that behaves in a reliable way with the smallest computational cost possible. There are two classes of solutions for simulating solvent: implicit or explicit. When modeling an implicit solvent, an extra term is be added to the potential energy calculation but no particles are represented. In fact, the effect of the solvent is sum not its atoms themselves. The computational cost decreases but also the reliability of the simulations. With explicit solvent, one represents each water molecule (or any other solvent, e.g. methanol) in the simulation box. In this project only explicit solvents are employed, specifically the TIP3 model [113], which represents the water molecule as three particles in AA simulations. Importantly, TIP3 is a rigid water model. In GROMACS [1], the rigidity is enforced with the SETTLE algorithm [153].

Periodic Boundary Conditions

Another issue in MD simulations is how to manage the interactions with the boundaries. If the simulation box has solid borders, the molecules of the system will interact with them, biasing the results and increasing the computational cost. The solution to this problem is to establish Periodic Boundary Conditions (PBCs). The simulation box is replicated in all the directions of space as shown in Fig. 2.6. In a molecule leaves the box, it re-enters by the opposite side. Interactions can be computed between particles of neighbours boxes, so the simulation box should be big enough to

prevent the molecules to interact with themselves. PBCs require the use of a CO which has to be shorter than the smallest half-box length. PBCs are especially well suited for simulating molecular liquids. One thus get a kind of infinite systems, even though the box size can have artefacts on some properties.

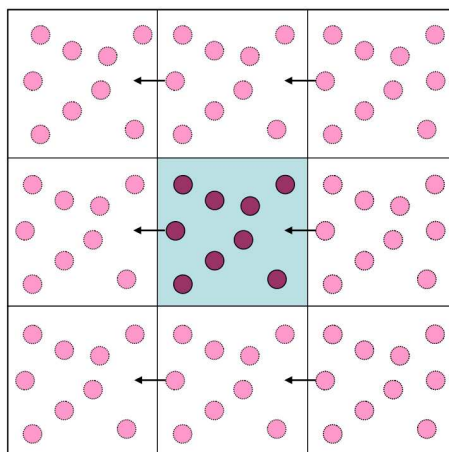


Figure 2.6 – Periodic Boundary Conditions (PBCs). The pink particle moving to the neighbour box on the left will re-enter the box by the right side. Image from [81]

The Protocol

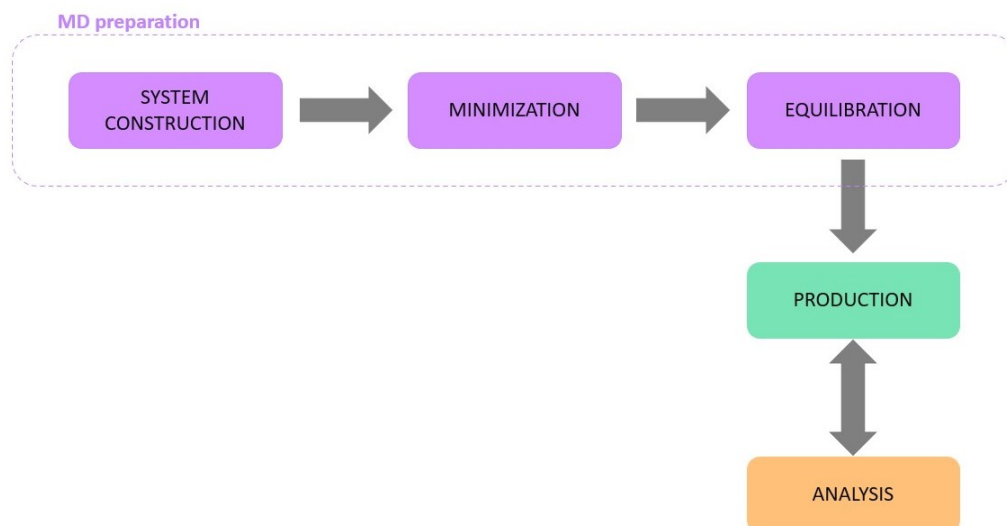


Figure 2.7 – Illustrative scheme of the different phases of an MD protocol.

System Construction Before launching an MD simulation, the system needs to be prepared (Fig. 2.7). The first step is to construct the system. One can use, among others, the GROMACS software [215] or the CHARMM-GUI web server [105]. During my PhD I used CHARMM-GUI as a general rule. The first thing to do is to choose the molecules to include in the system. It can be a small system, like a single peptide in a water box or a bigger one like a lipid bilayer, a peptide partitioned into a lipid bilayer, etc. To create a lipid bilayer, we first need to chose the number of

lipids and lipid types of the system. They will be placed randomly in the bilayer. When including a peptide, the position and orientation respect to the membrane can be chosen. The shape of the box should be chosen too. Due to bilayers shape, a simulation box containing a bilayer will be always cubic. However, if the system only contains a peptide, a rhombic dodecahedron box suits better, as it minimizes the number of solvent particles needed to fill the space (Fig. 2.8). Parameters can be also adjusted to get the desired number of water molecules and ion concentration. The output is adapted to the software and FF employed and it contains all the files needed to carry out the MD simulation.

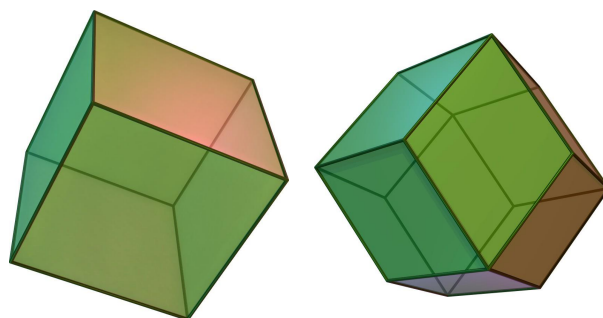


Figure 2.8 – Box shapes: cubic on the left and rhombic dodecahedron on the right. Credit images: CC BY-SA 3.0, <https://commons.wikimedia.org/w/index.php?curid=613268>; CC BY-SA 3.0, <https://commons.wikimedia.org/w/index.php?curid=38689>

Energy Minimization Once the system is constructed and all the necessary files are prepared, the system needs to be relaxed to carry out a reliable MD simulation. First, the energy minimization takes place, where the bonds, angles, dihedrals, and atoms distances are relaxed. This is done by using a mathematical algorithm that minimizes the potential energy of the system. Restraints can be employed to force the system to adopt specific relative positions. It is important to know that minimization does not use an MD algorithm. There are several algorithms to carry out an energy minimization. In this project, Steepest Descent (SD) algorithm is used [173], which searches for the local minimum of the current conformation (Fig. 2.9.A) using the first derivative of the energy function. Two parameters are needed to stop the minimization: the maximum force of the whole gradient (i.e. the derivative of V_{pot} with respect to the coordinates of each atom) and the maximum number of steps for the algorithm (number of times the algorithm proceeds). Once one of these two conditions is accomplished the minimization stops. At each step, SD will modify the structure in order to decrease the potential energy following the opposite of the gradient. To this aim, it needs the step length, λ which is usually done with a "line search" procedure [173]. An example is showed in figure 2.9.B. The new conformation (b) must have a lower energy than the previous conformation (a) and also lower energy than the conformation that will be obtained if λ is increased (c). If line search is employed, the molecular system will move in successive orthogonal directions on the energy landscape (Fig. 2.9.C). SD is a good energy minimization method for bad

structures (with steric clashes) as it assures the decrease of the potential energy. The downside is that it has problems of converge and sometimes it keeps oscillating round a minimum due to the orthogonality of the successive directions. However, SD is largely sufficient before an MD. Properly minimizing before an MD is required to avoid high forces which would make the algorithm crash.

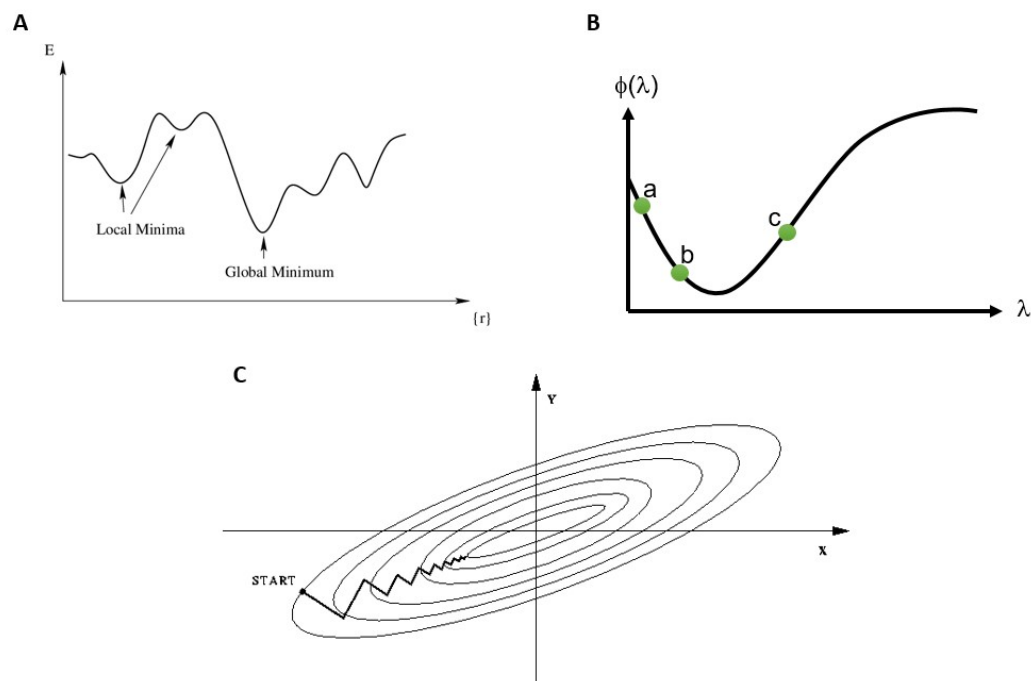


Figure 2.9 – Energy minimization important concepts. (A) Example of a 2D energy landscape with four local minima and one global minimum. Credit image: <https://vitalflux.com> (B) Line Search example (C) Steepest Descent example. Credit image: <http://trond.hjorteland.com>

Equilibration Phase Once the energy minimization is accomplished, the equilibration phase takes place. A short MD trajectory of a few nanoseconds is carried out to heat the system and allow the molecules to move in order to adopt more reliable conformations and relative positions. The generation of the temperature is done by assigning random velocities to the particles following a Maxwell-Boltzmann distribution. Position restraints can also be settled to direct the movements of the molecules. For example, one can restrain the peptide / protein in order to let the water molecules relax around them. Equilibration can be started by an NVT phase followed by a NPT phase, although is not always the case. Berendsen thermostat and barostat are chosen for this phase as they arrive really fast to the desired temperature and pressure. The fact that they do not produce reliable fluctuations can be neglected as the equilibration phase is never used for result analysis.

Production Phase Finally, the production phase takes place. The actual trajectory that is employed for the analyses is computed. It is done in NPT conditions, with v-rescale thermostat

and Parinello-Rahman barostat because they reproduce the correct fluctuations. The first frame of this trajectory is considered as the "time 0" of the simulation. Normally, during the first *ns* the system is still equilibrating so they are discarded for the analyses.

GROMACS 2018.6 [1] was employed to compute all the MD simulations that are showed in this manuscript. For most of the all-atom simulations carried out in this work, we employed the machine Occigen from the french national facility CINES. For coarse-grained simulations, we employed our local computers equipped with 40 cores and one GPU.

2.1.2 MD Analysis

Analysis for lipid membranes

Area Per Lipid The Area Per Lipid (APL) is one of the most common analyses applied to lipid membrane MD simulations. To calculate the APL, the layers are analysed separately. For simple monocomposition, one can just divide the *xy* area per the number of lipids per leaflet. However, it gets more difficult with more complex lipid compositions. In this thesis, we used instead an approach based on Voronoi tessellations which is convenient as it allows to compute the APL of all the lipids in case of complex membranes. Each individual lipid is represented like a single point. Normally, the coordinates of the point correspond to the phosphorous atom position (P). These positions are projected into the X, Y plane. Then, a Voronoi tessellations as the one shown in Fig. 2.10 is calculated to evaluate the APL. The Voronoi diagrams are one of the simplest interpolation methods and they use the euclidean distance. The first step to draw the diagram is to connect the adjacent points in the plane. Then, the bisector of the union is traced. The intersection of the bisectors determine an ensemble of polygons in the bi-dimensional space, surrounding the points. The area of each polygon correspond to the area occupied by the lipid. Then, the mean area of each lipid type is computed. This process is carried out for each frame and the results are normally plotted in a graph like the one showed in 2.11. This graph shows the APL fluctuations throughout the simulation, helping to determine how many time the system takes to reach equilibrium and which interval of time should be used to forward analyses.

Once the APL is calculated for each frame, the mean for the whole trajectory can be computed, obtaining a global APL value for each lipid structure. This value is useful for validating the simulations and the FF used. The APL can be determined by experimental methods (by X-ray scattering [166]) and these quantitative results are easy to compare to experimental ones. The agreement between them is one of the methods commonly used to check the reliability of MD simulations.

All the APL results showed in this manuscripts have been calculated using FatsLim[36] software.

Density Profiles Computing the density profiles is also one of the classical analysis that you do routinely when analysing a MD simulation containing a lipid bilayer. The density of each type

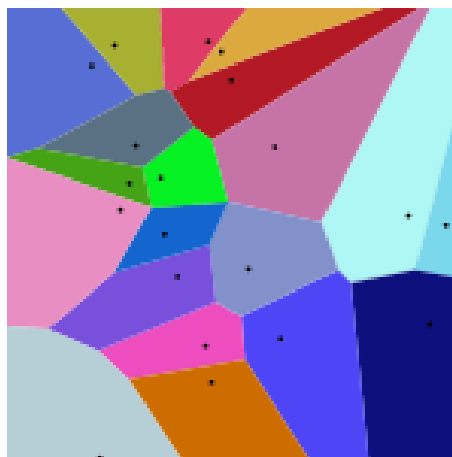


Figure 2.10 – Euclidean voronoi diagram by Balu Ertl - Own work, CC BY-SA 4.0, <https://commons.wikimedia.org/w/index.php?curid=38534275>

of molecule is computed in every Z position of the simulation box. The X and Y positions are not taken into account. As a result, a density plot is obtained where one can observe the distribution of the molecules along the Z axis of the simulation box (Fig. 2.12) .

The interpretation of the density plots is quite easy. When the two monolayers have the same lipid composition, the plot is symmetrical. Regarding the lipids, the first peak normally corresponds to the phosphorous groups and the density reduction in the middle of the plot corresponds to the intermembrane space. The density profile of water is also shown and allows to find the position of the mean interface water/membrane. It is considered that the interface is located where the water and lipid curves cross each other. The density of the ions can also be computed and it helps to understand the behaviour of the ions toward the membrane. When simulating a peptide in the membrane, this plot is useful to see its mean partitioning with respect to the bilayer (see section 2.1.2 below). Last, the plotted densities are generally mass densities (mass per unit of volume), but it is also possible to plot number densities (number of particles per unit of volume). The densities showed in this manuscript are computed with GROMACS software (program *gmx density*), version 2018.5.

Thickness The membrane thickness is calculated from the density profiles, considering that the space occupied by the membrane starts and ends at the interfaces with water. All the lipids of the system are merged into a single group and its density along the z axis is computed, obtaining the membrane density (Fig. 2.12). Then, we search for the z positions of the two membrane/water interfaces. The distance between them is the thickness of the membrane.

The study of the membrane thickness is helpful to understand the influence of a specific lipid type on the membrane structure. It is also important to know that the APL is closely related to the thickness. Generally, small APL values are related to highly ordered aliphatic chains, so the corresponding thickness is normally high. On the contrary, high APL values are generally related

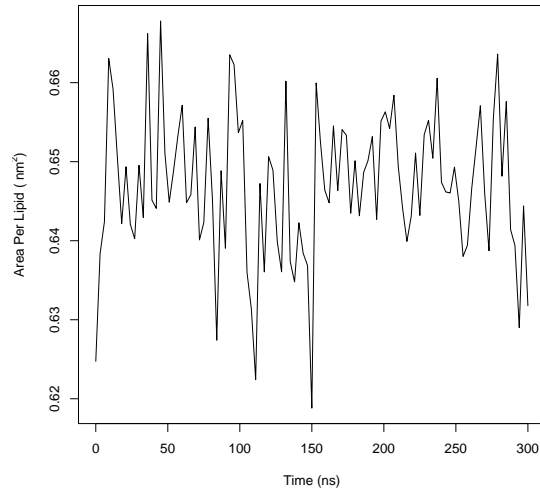


Figure 2.11 – Model APL graph: APL values are plotted against the simulation time. The data was obtained from a pure POPC simulation with CHARMM36 FF at 303,15 K

to disordered aliphatic chains and this decreases the thickness value. The study of the thickness is another tool to validate the simulations.

Radial Distribution Function The Radial Distribution Function (RDF), also noted $g(r)$, is a classical tool for describing the "structure" of a liquid. Here we use it in 2D in the plane of each monolayer. RDF allows to evaluate the arrangement of lipids along the layer, describing how lipids are laterally distributed. For a given pair of particles A and B (POPC phosphorous and POPS phosphorous, for example), $g(r)$ is calculated as the ratio between the local B density ρ at a distance r from any particle A with respect to the global B density in the whole membrane (Eq. 2.12). Note that here the number density is considered, that is, number of particles per unit volume. The RDF value varies typically around 1 and the obtained peaks give the position of the different shells of particles B around A (Fig. 2.13). The height of the peaks is directly proportional to the amount of particles in the corresponding shell. In this manuscript 2D RDF results are presented (that means all distances between lipids are calculate as 2D euclidean distances, i.e. within the plane of the monolayer), not taking into account the distances with respect to the z axis.

$$g(r) = \rho(r)/\rho(bulk) \quad (2.12)$$

Analysis for amphipathic helices

Secondary Structure and Helicity content The first thing to study when simulation protein systems is the secondary structure (SS) the peptide or protein can adopt. A program commonly known as DSSP (Define Secondary Structure of Proteins) [114, 110] implements a method to check

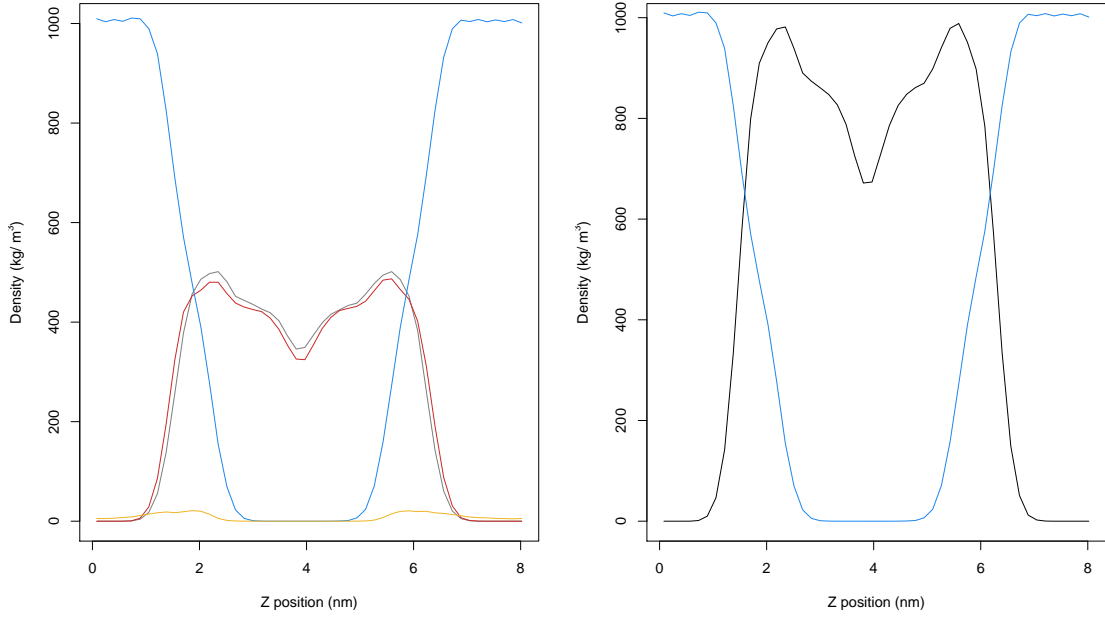


Figure 2.12 – Model density graph. The densities in the z axis are plotted for each molecule type. On the left, the two lipid types are plotted separately (gray and red). The water is represented in blue and the ions in yellow. On the right, the different lipid types were merged in a single group (black) to visualize the water/membrane interface. The data was obtained from a POPC/POPS 50/50 simulation with CHARMM36 FF at 303,15 K

the secondary structures adopted during an MD simulation. It first identifies the intra backbone H-bonds. It uses an electrostatic definition of the H-bonds, where charges of -0.42 , $+0.20$, $+0.42$ and -0.20 are attributed to the carbonyl oxygen, amide hydrogen, carbonyl carbon and amide nitrogen respectively. Then, equation 2.13 is applied, where the different r are the distances between the backbone atoms. If the energy obtained is bigger than 0.5 Kcal/mol , an H-bond is assigned.

$$E = 0.084 \left(\frac{1}{r_{ON}} + \frac{1}{r_{CH}} - \frac{1}{r_{OH}} - \frac{1}{r_{CN}} \right) \cdot 332 \quad (2.13)$$

Once all backbone hydrogen bonds are calculated, DSSP assigns eight different SS states depending on the H-bond pattern found. The SS can be grouped in three ensembles: helices, β -sheets and loops. The helices present a repetitive sequence of H-bonds where residues are 3, 4 or 5 times apart depending on the type of helix: 3_{10} -helix, α -helix or π -helix respectively. Parallel or antiparallel β -sheet structures present a set of H-bonds that grow in parallel or in the reverse direction respectively. All the remaining states belong to loops. If there is a unique H-bond between two strands, a β -bridge is assigned. Finally, three different types of unfolded structures can be identified: turns, if a unique H-bond is found with an helical pattern; a high backbone curvature region, if large angles are detected between consecutive residues; generic coil, if no H-bond or backbone curvature is observed.

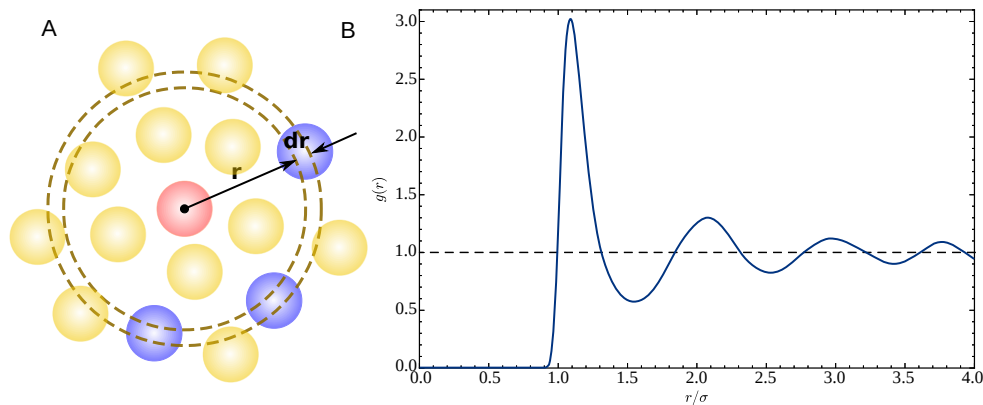


Figure 2.13 – (A) In red: reference particle. In purple: particles at a distance r from the reference that are taken into account to calculate $\rho(r)$. In yellow: particles at a distance other than r from the reference. Both yellow and purple particles will be taken into account to calculate $\rho(bulk)$. The brown circle represents the density shell. (B) Model graph from RDF analysis. If the peak is higher than 1.0 means that the particle density of the shell is higher than the density of the bulk. Credit images: <https://commons.wikimedia.org>

DSSP is implemented in GROMACS and it is very useful for simulation analysis. The SS adopted by each residue can be checked through time. In addition to this, in this research DSSP is employed to calculate the peptide helix content. The helix content (or helicity) is the percentage of residues forming an α -helix within the sequence. It can be plotted through time to observe the SS evolution during the simulation. It is also possible to obtain the mean helicity of a simulation for the whole peptide or per residue.

Root-Mean-Square deviation The Root-Mean-Square deviation (RMSD) is one of the most common analyses used to compare molecular structures, specially proteins. It is a measure of the structural distance between two conformations of the same molecule. First, the two structures are superposed. If working with proteins or peptides, only the $C\alpha$ or the backbone atoms are generally taken into account for the alignment. Then, the distances (δ) between identical atoms of different conformations of the molecule(s) are calculated (for example, the distance between the $C\alpha$ of residue 1 in structure 1 and the $C\alpha$ of residue 1 in structure 2). Having these distances, the RMSD can be calculated using the following equation:

$$RMSD = \sqrt{\frac{1}{N} \sum_{i=1}^N \delta_i^2} \quad (2.14)$$

Where N is the number of atoms participating in the RMSD calculation. This value gives

an idea of the structural similarity between two conformations: the lower, the more similar the structures. Normally, the initial superposition is chosen in order to minimize the RMSD value. There are different programs that are able to calculate the RMSD. GROMACS implements the RMSD calculation (*gmx rms*), and it can be done between a single structure and a reference or between a whole simulation and a reference. In the second case, a plot of the evolution of the RMSD trough time is obtained.

Radius of Gyration The Radius of Gyration (R_G) of a protein or a peptide, is a measure of the compactness of the structure. Mathematically it is calculated as the root-mean-square deviation between the particles of the molecule and the Center Of Mass (COM):

$$R_G = \sqrt{\frac{1}{M} \sum_{i=1}^N m_i r_i^2} \quad (2.15)$$

where r_i is the distance between the particle i and the COM, m_i the mass of particle i and M the sum of the masses of all the particles considered for the calculation. Normally, only the C α or the backbone atoms are taken into account. The lower the R_G , the more compact the structure. R_G is especially useful for monitoring folding / unfolding events, where it grows or decreases respectively. R_G calculation is implemented in GROMACS software (*gmx gyr*).

Analyses for Amphipathic helices in membranes

Residue partitioning Generally, an amphipathic helix (AH) that is inserted into a membrane has its principal axis parallel to the lipid / water interface (x, y plane). However, the exact position of the helix, determined by the degree of insertion and its tilt angle, depends on different factors, such as the amino acid sequence composition or the lipid types present in the membrane. Studying the details of the partitioning helps understand the behaviour of the AH in the membrane. To this aim, we have developed a R script that analyses the partitioning of an AH into the membrane from MD-CG simulations.

The script aims to generate a boxplot reporting on the partitioning as the one showed in figure 2.14, where the position of the α -helix is plotted against some membrane reference positions, determining the degree of insertion and the tilt of the helix. The script first calculates the reference positions. An ensemble of lipid beads are selected, as the ones containing the Phosphorous atom. The mean z position of these beads is calculated throughout the trajectory, giving a description of the membrane position in z . Thus no matter the bilayer moves in z , we always get the relative position of each residue with respect to the bilayer center. Then, for each amino acid of the peptide, the z position of the backbone bead is monitored throughout the simulation. As a final result, a boxplot is generated. For each residue z position, a box is drawn corresponding to the distribution of z between the first and third quartile, while the whiskers correspond to the values beyond the quartiles. Outliers are not represented for clarity. The position of the lipid beads are added to the

plot to have the reference positions of the membrane (Fig. 2.14).

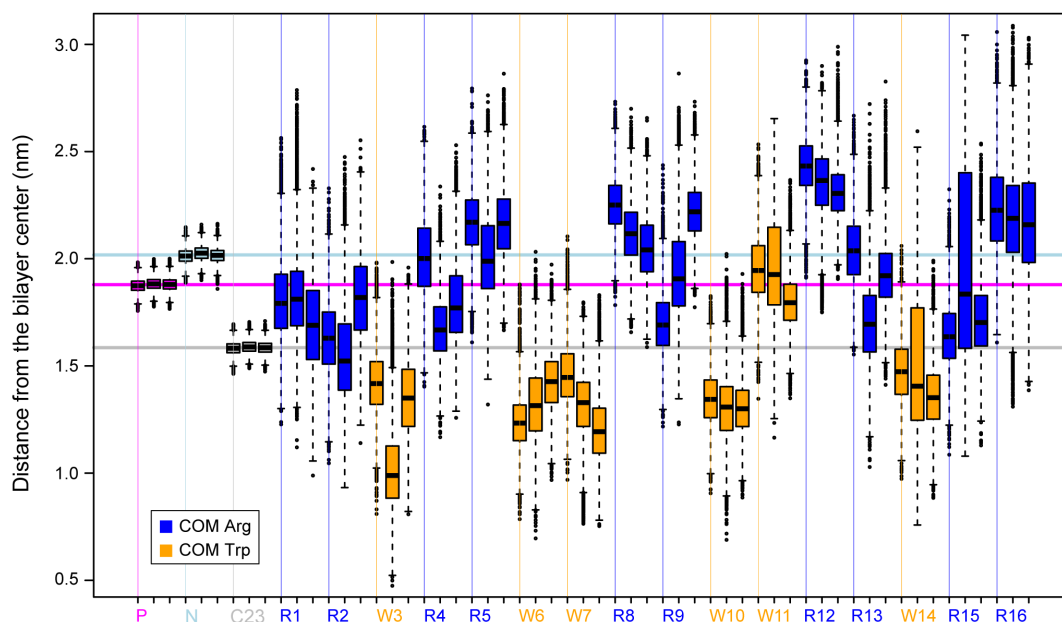


Figure 2.14 – Example of partitioning plot taken from the RW16 CPP (taken from Jobin et al. [107]).

This analysis helps to better understand the peptide behaviour in the membrane and also to validate the simulations. We can judge if the peptide is correctly positioned if the hydrophobic residues tend to be embedded and the polar ones exposed. In addition to this, the results are comparable to NMR experiments performed in presence of a paramagnetic probe. The positions are computed from the trajectory with GROMACS software 2018.5 (program *gmx traj*) version and the script is written in R 3.4.4 version.

Lipid recruitment Membrane structure can be altered by the partitioning of an AH. One of the characteristics that can be altered is the lipid lateral distribution. The peptide sequence can contain charged amino acids that can attract or repel some of the lipids composing the membrane.

We have created a python script that allows to analyse the lipid recruitment by an inserted peptide from MD simulations. The program first calculates the relative distance between the backbone beads of the residues and the phosphorous beads of all the lipids. It draws an histogram to determine at which cut-off distance from the peptide the first shell of lipids is located. We consider that the lipids located in this first shell are interacting with the peptide.

Once the distance cut-off (CO) is determined, the program examines the trajectory and for each frame it counts the lipids that are located closer to the peptide than the CO. The mean number for each lipid type is computed. Then, we can compute the ratio of each lipid type surrounding the peptide. This ratio is compared to the ratio of the whole box to determine if the peptide recruits preferentially one type of lipid.

This analysis is complementary to the RDF (see "Radial Distribution Function" section) and it helps understand the impact of the AH on the global membrane structure and dynamics.

The script is written in Python 3.8.6 version and the MDAnalysis module is employed to analyse the trajectory.

2.1.3 Temperature-Replica Exchange Molecular Dynamics

The Temperature-Replica Exchange Molecular Dynamics (T-REMD) [151] is an enhanced sampling technique. In certain cases, it allows the observation of long molecular processes whose time scales are too elevated to be observed by classical MD at room temperature. Normally, in a classical MD simulation one replica is simulated for a certain period of time at a constant temperature. However, in T-REMD simulations, a high number of replicas is simulated at the same time. Each replica is launched at a different temperature. The bottom temperature is normally the physiological one and the upper temperature can be extremely high (e.g. 400 K) but on the same system. It is possible to give different starting structures at each temperature. The random seed of each replica is different for a better effectiveness. Throughout the simulation, the replicas can exchange temperatures. GROMACS periodically attempts to exchange temperatures between neighboring replicas (i.e. with consecutive temperatures) and accept or reject the exchange with a defined probability. One fundamental criterion for the exchange is that the resulting ensemble of conformations at each temperature should follow a Boltzmann distribution at equilibrium (Eq. 2.9). To ensure this, the Metropolis criterion is imposed for each exchange:

$$P_{i \leftrightarrow j} = \min(1, \exp[(\frac{1}{kT_i} - \frac{1}{kT_j})(E_i - E_j)]) \quad (2.16)$$

where $P_{i \leftrightarrow j}$ is the probability of exchanging the temperatures T_i and T_j between replicas i and j ; E_i and E_j are the corresponding potential energies and k_B the Boltzmann's constant. To understand the meaning of equation 2.16, consider $T_j > T_i$. If $E_i > E_j$, the exponential in the equation above is higher than 1, thus the exchange is always accepted: $P_{i \leftrightarrow j} = 1$. If $E_i < E_j$, which is more often the case (higher temperatures correspond to higher potential energies, see figure 2.16), the exponential will be comprised between 0 and 1 corresponding thus to the probability $P_{i \leftrightarrow j}$. Its value will be compared to a random number taken between 0 and 1 in a uniform distribution. If the random number is lower than $P_{i \leftrightarrow j}$, the exchange is accepted, if not it is rejected. The higher the difference of energy and / or the higher the difference of temperature, the closer $P_{i \leftrightarrow j}$ to 0. One immediately understands that to get a chance to have the exchange accepted, the Gaussian distributions of energy need to overlap (see figure 2.16).

The choice of the Metropolis criterion in equation 2.16 leads to a proper equilibrium Boltzmann distribution because it fulfils the so-called "detailed balance" condition[151]:

$$P_{i \leftrightarrow j} P(X) = P_{j \leftrightarrow i} P(Y) \quad (2.17)$$

where $P(X)$ is the probability of finding the system in state X (X represents the state of all replicas, it can be seen as a vector of states $X = (i, j)$) before the exchange $i \leftrightarrow j$ has proceeded, $P(Y)$ is the probability of finding the system in state Y after the exchange $i \leftrightarrow j$ has proceeded. The detailed balance simply means that we should have the same probability of exchanging $i \leftrightarrow j$ starting from state X than exchanging $j \leftrightarrow i$ starting from state Y .

Equation 2.16 is used when simulating in the canonical ensemble (NVT). In the isothermal-isobaric ensemble (NPT), we also have to take the pressure and volume of the replicas into account [156]. For convenience, we define inverse temperatures $\beta_i = \frac{1}{k_B T_i}$ and $\beta_j = \frac{1}{k_B T_j}$. The probability to accept an exchange then becomes:

$$P_{i \leftrightarrow j} = \min(1, \exp[(\beta_i - \beta_j)(E_i - E_j) + (\beta_i P_i - \beta_j P_j)(V_i - V_j)]) \quad (2.18)$$

where P_i and P_j are the pressure of replicas i and j respectively, and V_i and V_j the corresponding system volumes. In general, the term due to volume fluctuations does not contribute much to $P_{i \leftrightarrow j}$ at low pressure [161].

Finally, if the exchange is accepted, velocities are adjusted to match the new temperature. The REMD principle is shown in figure 2.15.

Simulating at different temperatures allows the system to explore multiple conformational ensembles while maintaining a physical distribution. In addition to this, the increase of temperature gives a boost of kinetic energy to the molecular system, enabling it to cross higher free energy barriers more effectively than at physiological temperature. Furthermore, these free energy barriers may also be lowered at high temperature (for example, above the melting temperature of a peptide) giving even more efficiency to the process. In consequence, we get to explore a richer conformational ensemble by increasing the sampling. The downside of this technique is that the system kinetics is altered by the temperature modifications so it is generally difficult to extract kinetic information.

When preparing an REMD simulation, we have to make important choices. The total number of replicas needed and their temperature depend on the molecular system of interest. Patriksson and van der Spoel proposed a method to predict these [161]. They also provide a useful web-page : Temperature Generator. To determine the replica temperatures, the server proceeds like this. First, the user determines the extreme temperatures, here we chose between 300 and 400 K . Some RE[MD] of peptides in water have been performed at way higher temperatures than 400 K [233]. Here we chose 400 K as a compromise, high enough to enhance sampling, but not too high to

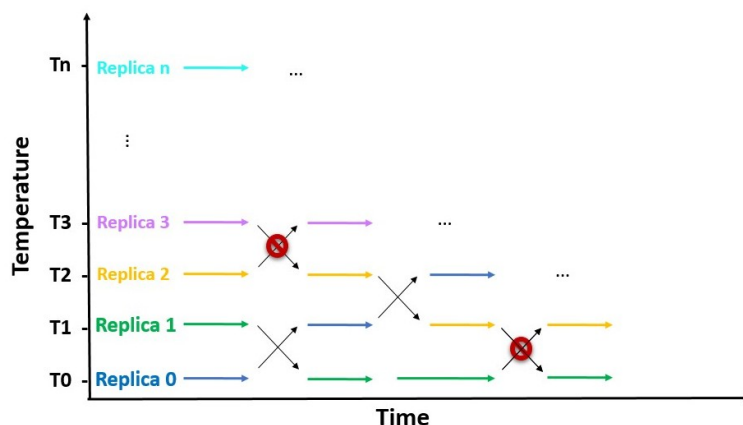


Figure 2.15 – Illustrative image of the T-REM technique. A number n of replicas is simulated. During trajectory calculation there are several attempts of temperature exchanging that can be accepted as the one between replica 0 and replica 1, or rejected as the one between replica 3 and replica 4.

keep the bilayer intact. In literature, it was shown that temperatures above 500 K made a DMPC bilayer explode. To avoid this explosion some authors added restraints on lipids [158]. We decided not to reach these extreme solutions. One may argue that 400 K is already high. However, the bilayer remains intact at this temperature and the time it takes for a replica at 400 K to reach the bottom temperature (a few tens of ns), the bilayer has time to relax and to resemble a real bilayer at room temperature. To sum up, 400 K was a good compromise between enhanced sampling and bilayer integrity. Then the user has to enter the number of atoms and constraints (rigid bonds, rigid water molecules). Then the mean probability of acceptance of each exchange has to be chosen. Here we use 0.20 which is a standard value used in literature [151]. The server then predicts the different temperatures by computing a predicted (Gaussian) distribution of the potential energy, and ensuring that two consecutive temperatures present some overlap. The degree of overlap is directly connected to the final probability of acceptance of exchanges. In figure 2.16 is shown an example of the potential energy distributions after an REMD we have performed. The final probabilities of exchange in this REMD are between 0.26 and 0.30, though we requested 0.20. This shows that the prediction made by the web-server model is not exact, it remains a prediction, but the final probabilities are in an acceptable range.

GROMACS 2018.6 was employed to carry out all the T-REM simulations that are showed in this manuscript.

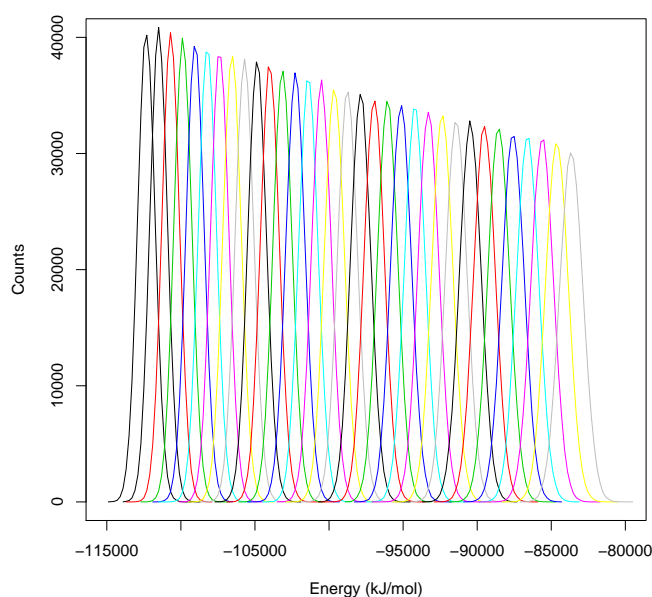


Figure 2.16 – Potential energy distributions of the molecular system at each temperature. Each curve belongs to a temperature used in the REMD simulations, from the bottom temperature (300 K, curve on the left) to the upper temperature (400K, curve on the right)

2.1.4 Markov State Models

Introduction

A Markov Chain is a stochastic model that describes a sequence of events where the probability of each event depends only on the previous state. This principle is applicable to studies of many different domains: physics, engineering, biophysics, etc. If a changing system behaves like a Markov Chain, it is possible to design a Markov State Model (MSM) to describe its behaviour. The MSM describes the different states of the system and the changing mechanisms [94].

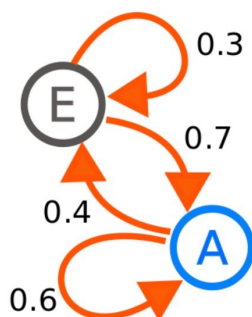


Figure 2.17 – Example of a simple MSM process with two states: A and E. Each number represents the probability of changing from one state to another. Credit image: By Joxemai4 - Own work, CC BY-SA 3.0, <https://commons.wikimedia.org>

Figure 2.17 shows a simple example of an imaginary system with two states: A and E, each state representing a structural ensemble. If the system is in state A, it has a probability of 0.6 of staying in state A, and 0.4 probability of changing to state E. If the system is in state E, it has a probability of 0.3 of staying in state E, and 0.7 probability of changing to state A. Interestingly, these probabilities are directly connected to the kinetic rates of transitioning from one state to the other. This way, the MSM helps understand the system dynamics. The MSM principle is applicable to MD simulations when they fulfil two conditions: behave like a Markov Chain and present reversible dynamics (any state can be attained independently of the starting point). The aim is to decipher the system structural, kinetic and thermodynamic properties. Normally, the dynamical information we are the most interested in are the slow processes. MSM facilitates their extraction from complex MD simulations.

The principle at a glance

An MSM can be constructed from one single long MD simulation or from several short MD trajectories (Fig. 2.18.a). The second option is the most popular as we can generally get more sampling (each trajectory has a different starting point and can be launched on different computers in parallel). In many cases, MSM are employed to study systems at equilibrium, although it is possible to apply them to non-equilibrium systems. In this work we used MSM only for systems at equilibrium. In such a case, the system has to fulfil detailed balance condition:

$$p_{ij}\pi_i = p_{ji}\pi_j \quad (2.19)$$

where p_{ij} and p_{ji} are the probabilities to have a transition between states i to j or j to i respectively, π_i and π_j are the stationary probabilities of states i and j respectively. Assuming detailed balance has some great advantages when constructing an MSM (see below).

The conformations observed in the trajectories are classified in different so-called "microstates" (Fig. 2.18.b). This classification is normally based on the structural similarity of the conformations. By doing this, the number of possible states is reduced and it becomes easier to study the molecular system. The number of microstates depend on the system, although the most common values are between 50 and 150 [170].

Once the system is decomposed into discrete (non overlapping) microstates, the transition count matrix (Fig. 2.18.c) is computed. This matrix stores the number of transitions between each pair of states occurring after a certain lag time τ . For example: the system went 11 times from green to green and 1 time from green to blue between times t and $t + \tau$. Note that some transitions may never occur. τ is a critical parameter of an MSM and has to be carefully chosen and validated. The transition probability matrix, usually called $\mathbf{P}(\tau)$, can then be computed from the transition count matrix (Fig. 2.18.d). It contains the same information but in terms of probability (probabilities of the system changing between the states). The adjusted populations (Fig. 2.18.e) can also be

obtained from the state discretization. They express the quantity of time that the system stays in a certain conformation.

At this point, lot of information about the system dynamics has been deduced: the structural properties (discrete states), the kinetics (probability matrix) and the thermodynamics (adjusted populations). Normally, the most interesting processes are the slowest ones (higher time scales). The information about the time scales of the processes are contained in the eigenvectors of the transition probability matrix. If we rank the eigenvalues of the matrix by descending order, the first one (λ_1) corresponds to the equilibrium distribution. The second one (λ_2) corresponds to the slowest process observed. (λ_3) corresponds to the second slowest process observed, etc. When we talk about slow process, we mean that each eigenvector describes a conformational change of the system, ordered by "slowness". Each eigenvector contains n values corresponding to the n microstates. The magnitudes and signs of these elements explain which states are contributing to the process identified by the eigenvalue. In general, we are only interested in a few of these processes, the slowest ones.

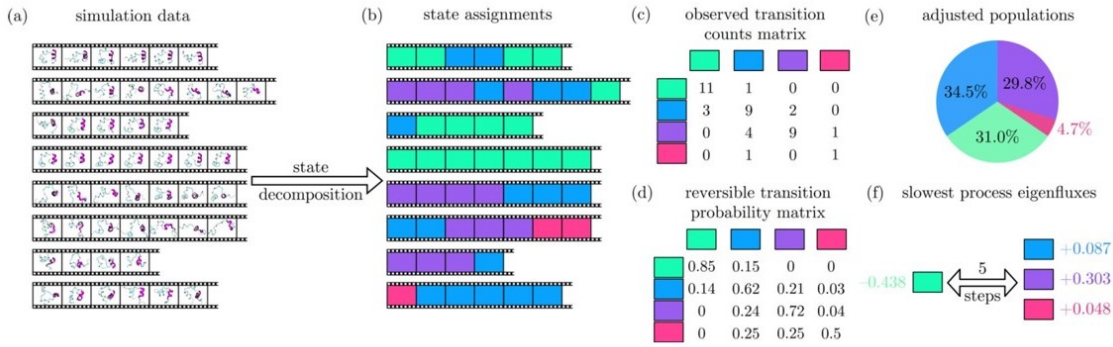


Figure 2.18 – MSM principle at a glance. Image obtained from [95]

As said above, the number of microstates is generally high and it is not always easy to disentangle what is going on. To deal with that, a second clusterization is often carried out (Fig. 2.19). After the clustering into microstates from raw data, one does another clustering called PCCP (for "Perron Cluster Clustering Analysis") [180]. The idea is to group the microstates into metastable states, that we call macrostates. The slow processes we are interested in are then described by the transitions between the macrostates.

It is also important to notice that MSM modeling is an iterative process. The MSM construction is repeated recursively to increase the sampling or to adjust the parameters. The model is progressively refined to obtain the most reliable result possible.

The protocol

In this section the protocol followed for MSM modeling is explained in depth. MSM are constructed using PYEMMA [190] and DeepTime [89] software. Figures used as examples are obtained

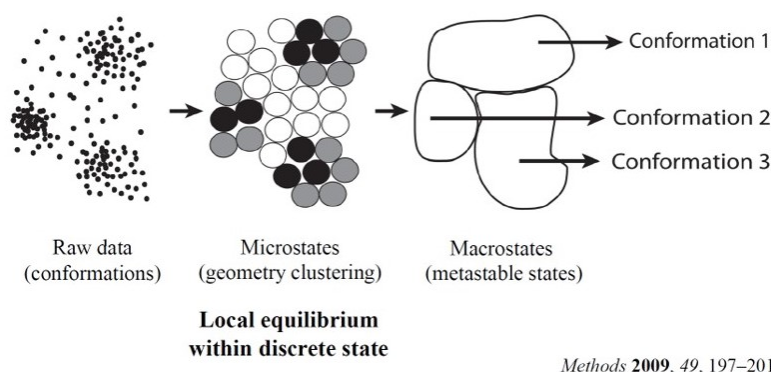


Figure 2.19 – MSM clusterization process

from PYEMMA software tutorial [221]. In the tutorial, 25 trajectories of a system containing a pentapeptide were simulated with CHARMM36 FF and implicit solvent and the trajectories were saved every 0.1 *ns* (Fig 2.20.a). Some of the theoretical parts shown in this section are extracted from: <http://docs.markovmodel.org/>

Step 1 - Featurization First, the ensemble of geometrical features that will be used for the clustering are chosen. The choice is done intuitively at first, applying the previous knowledge about the system dynamics. Some common ones are the dihedral angles, C_α positions, C_α distances, etc. It is also necessary to make an initial choice for the lag time τ that will be applied for dimensionality reduction with TICA (see next step).

Then, the features are ranked by means of the VAMP2 score which measures the kinetic variance contained in these features [224]. The minimum value of this score is 1, which corresponds to the equilibrium. The higher the value, the higher the variance contained in the selected features. Figure 2.20.b shows a VAMP2 score ranking for the pentapeptide system. With a 0.5 *ns* lag time, backbone torsions (dihedral angles) represent higher kinetic variance than backbone positions or distances. In consequence, backbone torsions are the first features employed for the MSM construction. This choice can be refined latter if necessary. Notice that the features employed in MSM construction can also be a combination of various geometrical features: backbone torsions + backbone distances, for example. Other popular features for peptide systems are the RMSD and the Radius of gyration explained before.

Step 2 - Dimensionality Reduction The feature space is normally of high dimensionality. To simplify the model construction and reduce its computational cost, the number of dimensions should be reduced while avoiding variance loss. In MSM construction, an algorithm called Time-lagged Independent Component Analysis (TICA) [165] is often used. It performs a linear transformation of some (usually high-dimensional) set of input coordinates to some (usually low-dimensional) set of output coordinates, maximizing the autocorrelation of the given coordinates. The final number of dimensions to keep should be as small as possible but high enough to capture

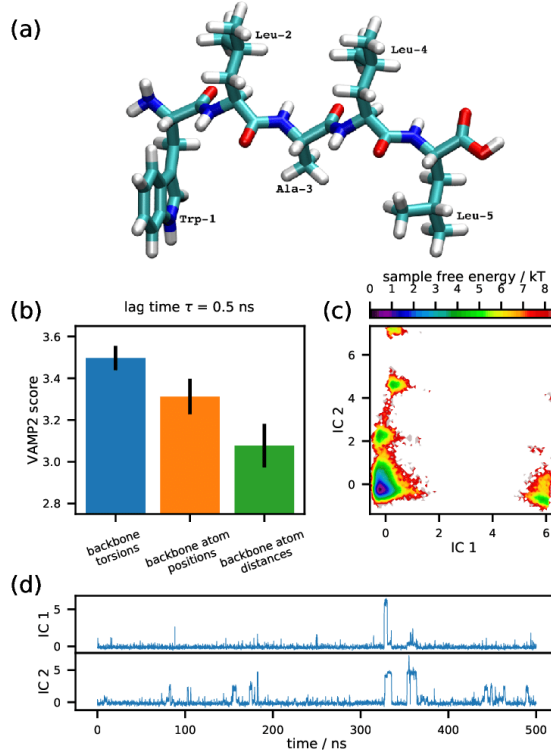


Figure 2.20 – MSM protocol: featurization and dimensionality reduction with TICA. (a) Pentapeptide 3D structure. (b) VAMP2 scores of the backbone torsions, positions and distances at a lag time of 0.5 ns. (c) Pentapeptide conformational landscape projected onto the first two independent components (ICs). The color is related to the structure free energy. (d) Values of the two first ICs throughout the trajectory. Images obtained from [221]

the important events. Normally, the protocols chose to keep the 95% variability. The fundamental TICA distinctive characteristic compared to other algorithms is that it takes into account the variation of the features in time: it attempts to detect the "slowest coordinates", the ones that variate at larger time scales. It first computes two covariance matrices, the instantaneous $\mathbf{C}(0)$ and the time-lagged $\mathbf{C}(\tau)$ ones:

$$\begin{aligned}
 c_{ij}(0) &= \frac{1}{N-1} \sum_{t=1}^N (x_i(t) - \mu_i)(x_j(t) - \mu_j) \\
 c_{ij}(\tau) &= \frac{1}{N-1-\tau} \sum_{t=1}^{N-\tau} (x_i(t) - \mu_i)(x_j(t+\tau) - \mu_j)
 \end{aligned} \tag{2.20}$$

where i and j are two different features; μ the mean value of the feature during the simulation; N the total number of frames; t the time and τ the lag time. Then, the time-lagged matrix obtained from the data $\mathbf{C}_d(\tau)$ must be diagonalized (the instantaneous matrix is already diagonal):

$$\mathbf{C}(\tau) = \frac{1}{2}(\mathbf{C}_d(\tau) + \mathbf{C}_d^T(\tau)) \tag{2.21}$$

Having the two covariance matrices, the eigenvalue problem can be solve directly by an appro-

priate generalized eigenvalue solver:

$$\mathbf{C}(\tau)\mathbf{U} = \mathbf{C}(0)\mathbf{U}\mathbf{\Lambda} \quad (2.22)$$

where \mathbf{U} is the eigenvector-matrix containing the Independent Components (ICs) and $\mathbf{\Lambda}$ is the diagonal eigenvalue matrix. Now, the data can be projected onto the TICA space:

$$\mathbf{z}^T(t) = \mathbf{r}^T(t)\mathbf{U} \quad (2.23)$$

where r is the coordinate vector and z the independent components. In this step, the dimensionality reduction is performed by selecting only a sub-matrix \mathbf{U} consisting of the first m columns of the full-rank \mathbf{U} , transforming the features into a set of slow coordinates. An example of the result of this process is shown in figure 2.20. Figure 2.20.d shows the values of the two first ICs of the pentapeptide MSM model. The first ICs are the slowest coordinates and they contain the biggest part of the variance of the system. Figure 2.20.c is a projection of the conformations of the trajectory in a 2D space generated by the two first ICs. Each point of the graph corresponds to one frame of the trajectory and the color is related to the sample free energy.

Step 3 - Discretization The aim of this step is to cluster conformations found in the trajectory into a set of discrete microstates, structural clusters generated with the IC information obtained from TICA. A geometric clustering is carried out based on the TICA projection (Eq. 2.23). The first thing to do is to choose the number of microstates. The standard number used in literature is 100. To improve the reliability of the results it is better to find a number of microstates adapted to the system of interest. For a first guess, VAMP2 score algorithm can be applied [224]. This algorithm calculates the quantity of variance that is preserved depending on the number of microstates used. Increasing the number of microstates increases the variance of the model until a plateau is attained (Fig. 2.21.a). The result presenting the smallest number of microstates and while maximizing the VAMP2 score is chosen. This method allows to find a balance between low computational effort (minimize the number of microstates) and the preservation of the dynamic information content.

Once the number of microstates (k) is set, a K-means algorithm is carried out [143]. It partitions the ensemble of conformations in k clusters. It first assigns randomly the centroids: the conformations that are going to be the centers of the clusters. The grouping of the conformations resembles Voronoi's partition (Fig. 2.10): each conformation is assigned to the cluster it resembles the most (minor difference between its projected features and the centroid projected features). Then, the centroids are re-assigned: the center of each cluster becomes the new centroid (the conformation with the minor difference respect to all the other observations of the same cluster). This process is repeated in an iterative way until the assignment of the centroids stop changing. In figure 2.21.b the centroids are highlighted in orange in the conformational landscape.

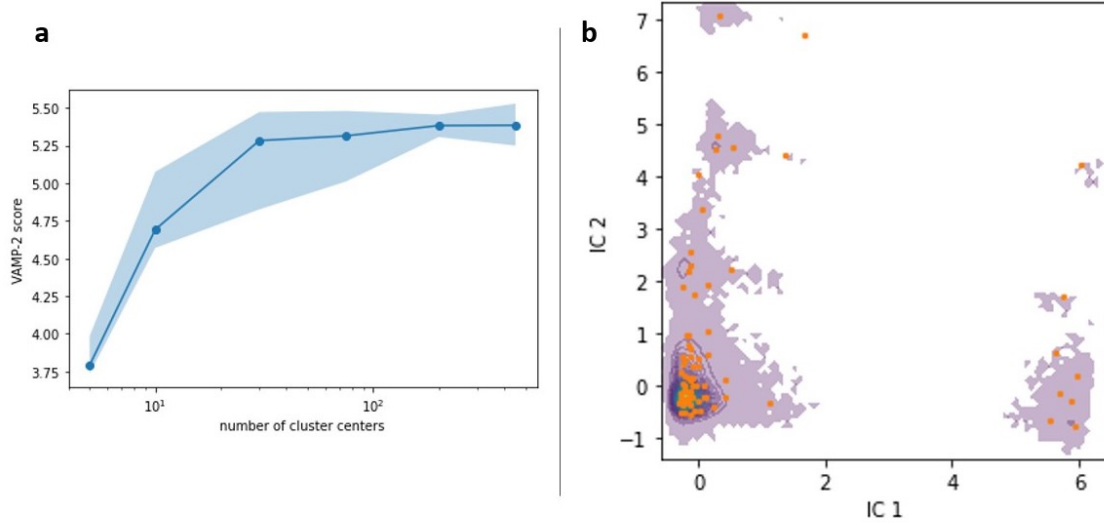


Figure 2.21 – MSM protocol: discretization. (a) VAMP2 Score analysis of the number of microstates. The score is saturated at 75 states. (b) Trajectory conformations projected onto the two first ICs. The darkness of the color is proportional to the density of the region. The centroids of the microstates are marked in orange. Image obtained from [221]

Step 4 - Transition Matrix Now that the conformations are divided into microstates, the MSM can be generated: the state populations and pairwise transition probabilities are determined. Remember that the dynamics must be reversible and the system memory less at the chosen lag time (the pathway by which the system enters any state does not affect the transition probabilities). The transition probability matrix is usually constructed with a maximum likelihood approach or a Bayesian procedure:

$$P_{ij}(\tau) = \frac{N_{ij}(\tau)}{\sum_k N_{ik}(\tau)} \quad (2.24)$$

where $P_{ij}(\tau)$ is the probability of changing from microstate i to microstate j , N_{ij} the number of transitions from i to j and N_{ik} the number of transitions from i to any other microstate. From the transition matrix $\mathbf{P}(\tau)$ the eigenvectors (ψ_i) and eigenvalues (λ_i) are obtained:

$$\mathbf{P}(\tau)\psi_i = \lambda_i(\tau)\psi_i \quad (2.25)$$

If the system is at equilibrium, the matrix is symmetric. Usually the procedures to compute $\mathbf{P}(\tau)$ enforces its symetrization which in the end fulfils the detailed balance condition. In such a case, λ_1 is equal to 1 and the corresponding (first) eigenvector describes the Boltzmann distribution. λ_2, λ_3 , etc., have decreasing values lower than 1, each one describing a kinetic process. The implied time scale (ITS) of each process depends on the eigenvalue and the lag time. They represent the relaxation time scales of the dynamic processes and are calculated as follow:

$$ITS_i = -\frac{\tau}{\ln |\lambda_i(\tau)|} \quad (2.26)$$

The larger the eigenvalue, the slower the time scale and so the process is more interesting. The magnitude and the sign of each element of the eigenvectors explains which states contribute to the process. Eigenvalues can be used for a first estimation of the number of macrostates: include the first eigenvectors until there is a large gap between the value of two consecutive eigenvalues. For a number N of eigenvectors chosen, $N + 1$ macrostates are created. Each macrostate is a metastable state that has the property of capturing the dynamics for long times before jumping to another set. It should be noticed that this is just an helpful trick for a first model construction. The number of macrostates must be adapted later on based on the previous knowledge of the system and the MSM results obtained.

Step 5 - Validation of the Lag Time The lag time (τ) is a central quantity in MSMs, so it must be validated. The validation is done through the estimation of Implied Time Scales (ITS) convergence. They are computed from the eigenvalues of the transition matrix as shown in equation 2.26. Thus, we look for ITS convergence and choose τ accordingly. Normally, there is a range of τ where the ITS are invariant.

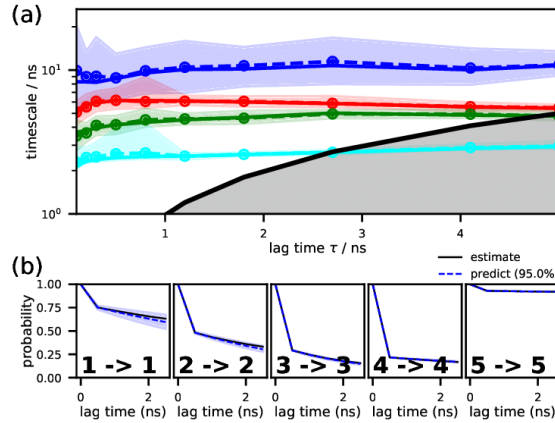


Figure 2.22 – MSM protocol: model validation. (a) ITS validation through convergence at increasing lag times. (b) A Passing Chapman-Kolmogorov (CK) test. When doing a CK test, PyEMMA automatically estimates a new MSM transition matrix at lag time $k\tau$ and propagates the original transition matrix by the k -th power. In the image, the diagonal values of these two matrices are plotted for 5 different values of k . 1 → 1 is the transition from macrocluster 1 to macrocluster 1 and so on. As the estimated and the predicted curve coincide, the test is passed. Image obtained from [221]

Figure 2.22.a shows an example of ITS validation for a pentapeptide MSM, where the ITS of the four slowest processes are studied. Each solid line correspond to one ITS calculated by a maximum likelihood approach. The sample means are given by dashed lines. They converged quickly, a lag time between 0.5 and 2 ns can be chosen. The confidence intervals of the implied timescales are

quantified according to a Bayesian scheme and it is shown with the colored shaded areas around the curves. The lag time also represents the time resolution limit of the estimated Markov model, shown by the grey zone of the graph.

Step 6 - Validation of the Markovianity of the Model The markovianity of the model must also be checked. This means that the systems has to behave like a Markov Chain. The validation is done through a Chapman-Kolmogorov (CK) test. It compares the right and the left side of the Chapman-Kolmogorov equation:

$$\mathbf{P}(k\tau) = \mathbf{P}^k(\tau) \quad (2.27)$$

with $\mathbf{P}(k\tau)$ being the transition matrix at lag time τ . The original transition matrix is propagated by the k^{th} power estimating consecutive transition matrices at lag time $k\tau$.

An example is shown in figure 2.22.b where the CK test is applied to pentapeptide's MSM. The CK test is carried out at 5 different values of k . The transition probability matrix has a size of 5x5 because the system was coarse-grained into 5 macrostates (see below how this is done). In the figure, the results of the CK test for the diagonal of the probability matrix are shown. Each square shows the probability to transition from on state to another (1 to 1, 2 to 2, etc.) at different lag times. The estimated and predicted values (right and left side of the equation) coincide, so it is a passing the CK test. The shaded areas show the confidence interval calculated according to a Bayesian scheme.

Step 7 - Coarse-Graining into Macrostates Now the lag time and the markovianity of the model for a certain number of macrostates is validated, it is time to coarse-grain the macrostates into macrostates to sum up the essential properties of the system kinetics. There are two ways for this coarse-graining: the fuzzy assignment or the discrete assignment. The fuzzy assignment computes the probability of the microstates to belong to each macrostate. The discrete assignment assigns each microstate to one single macrostate. Normally, both are carried out and if they agree, the discrete assignment is kept.

The discrete assignment is done using an algorithm called the Perron Cluster Analysis (PCCA) [175]. It uses the eigenvectors of the transition matrix to define the long-lived metastable sets. The dominant eigenvectors (higher associated eigenvalues) transport the probabilities of the slow kinetics so they can be used to decompose the system into macrostates.

The eigenvectors of the transition matrix work as a step function, changing from positive to negative values at a saddle point. The values of the first eigenvector can be plotted along one axis (Fig. 2.23.a). Notice that they are organized depending on the vector value, not the number of microstate. By partitioning the line in the middle, the two most metastable sets are generated. They exchange at the slowest time scale. It is possible to differentiate between smaller substates by

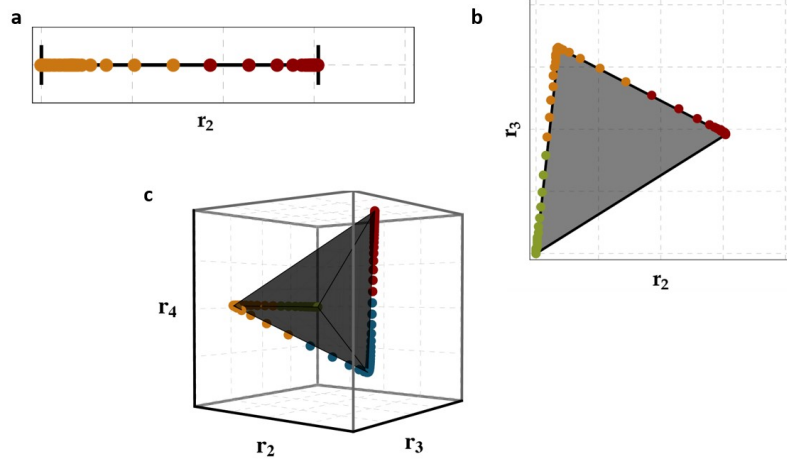


Figure 2.23 – PCCA algorithm. r_2 , r_3 and r_4 are the second, third and fourth dominant eigenvectors respectively. Image adapted from: <http://docs.markovmodel.org/>

considering more eigenvectors. In figure 2.23.b two eigenvectors are used to divide the microstates into three macrostates and in figure 2.23.c three eigenvectors are used to divide the microstates into four macrostates. In this way, the number of macrostates can be increased progressively.

The original PCCA method was introduced in [194] although today, PCCA+ [59] or PCCA++ [181] algorithms are used.

Using the PCCA algorithm, the clusterization into macrostates is obtained. An example is shown in figure 2.24 for the pentapeptide MSM. First, the energy landscape of the pentapeptide is shown in figure 2.24.a. Then, the landscape is divided into 5 macrostates to visualize their location (Fig. 2.24.b). It is also possible to plot the eigenvectors values to show how they affect the landscape clusterization (Fig. 2.24.c). Figure 2.24.d shows the transition path from state 2 to state 4 (Transition Path Theory (TPT) is explained in next section).

At this point, it is important to keep in mind that the MSM construction is an iterative process. This can be after incorporating new simulations or by changing some critical parameters of the MSM (e.g. choice of features, etc.). Normally, the model is re-constructed several times by applying the new knowledge obtained. The feature selection and different parameters are optimized to improve the reliability of the model. The final pentapeptide model is shown in figure 2.25

Step 8 - Compute various properties There are several information that can be extracted from an MSM. The free energy of the macrostates (G_i) is calculated from the stationary weights of the microstates (j) within the macrostate (i):

$$G_i = -k_B T \ln \sum_{j \in S_i} \pi_j \quad (2.28)$$

where S_i is the ensemble of microstates that belong to an specific macrostate i , p_j is the

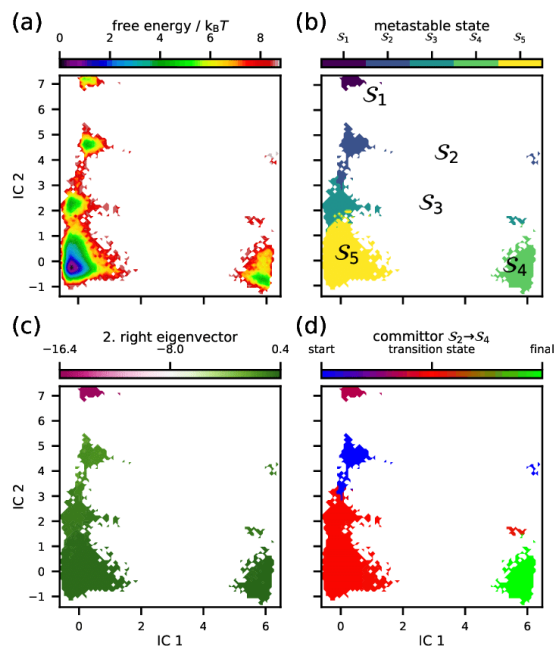


Figure 2.24 – MSM Analysis. (a) Energy landscape. (b) Coarsed-grain model into 5 macrostates. (c) Second right eigenvector of the model . It shows the second slowest process of the model. The first slowest process is not shown as it correspond to the system total dynamics. (d) TPT: how the system transitions from state 2 to state 4. Image obtained from [221]

stationary distribution of microstate j , k_B the Boltzmann's constant and T the temperature. Macrostates with the lowest free energy are the most stable and so the preferred for the system.

Mean first passage time (MFPT) between each pair of macrostates is also computed. It allows to deduce the kinetics between metastable states. Comparing the MFPT between states it is possible to calculate the life times of each of them.

The Transition Path Theory (TPT) analyzes the most probable transition paths between two macrostates. It allows to have a better understanding of how the system behaves and the structures evolves. An example is shown in figure 2.25, where the transition path from macrostate 2 to macrostate 4 in a pentapeptide MSM is described.

It is also possible to compute expectation values for experimental observables: helix ratio, radius of gyration, atom relative positions, etc. The structural and dynamic information explained with the model is compared to the information obtained by experimental assays. It is at the same time a way of validating the reliability of the model and a way to explain in depth experimental information that is sometimes difficult to interpret.

The MSM of this project were constructed using PYEMMA [190] and DeepTime (<https://github.com/deeptime-ml/deeptime.git>) software.

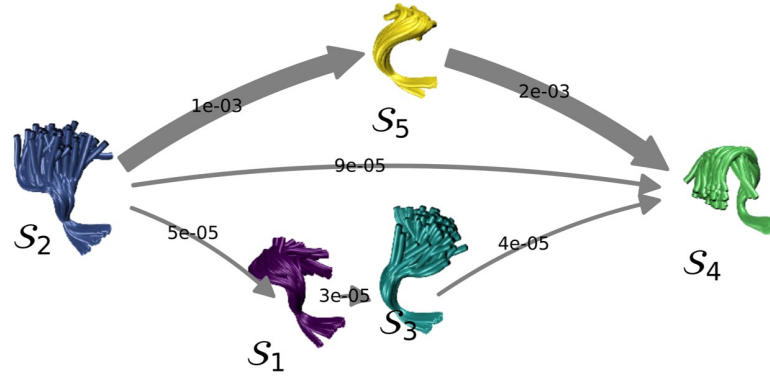


Figure 2.25 – MSM resulting model. An ensemble of representative structures are shown for each macrostate. The transition path to evolve from state 2 to state 4 is marked with the arrows. The width of the arrow is proportional to the probability of the state to change. The probability is written in float numbers over the corresponding arrow. Image adapted from [221].

2.2 Experimental methods

2.2.1 Circular Dichroism

The oscillations of electric and magnetic fields produce electromagnetic waves, commonly known as electromagnetic radiation or light. When the vector of the electric field oscillates in a straight line we talk about plane polarised or linearly polarised waves (PPW). The superposition in a different phase of two PPW that oscillates in perpendicular planes and that have the same amplitude and wave length, produces a third type of wave: Circular Polarised Wave (CPW). This wave can be left or right polarised. Curiously, when a right and a left CPW of the same amplitude and wave length are superposed, a PPW is obtained.

Light interacts with matter. This interaction can change light properties and give information about matter structure. The two basic phenomena of this interaction are the absorption and the decrease of the velocity of the light. The absorption is the reduction of the wave amplitude because matter absorbs part of the light. The decrease of the light velocity happens because of the refraction index ratio between the velocity in vacuum and the velocity in the specific material.

There are two special properties that materials can have regarding the interaction with light: Circular Dichroism (CD) and Circular Birefringence (CB). CD is observed when matter absorbs differently right and left CPW. As it was said before, a PPW is formed by the superposition of a right and a left CPW of the same amplitude. When a PPW traverses a medium that possesses CD, its properties will be modified because its components are absorbed differently. The resulting vector does not oscillate in a plane but in an ellipsoid path. This wave is called elliptically polarised light (Fig. 2.26).

Another special property matter can present is the Circular Birefringence (CB). CB happens

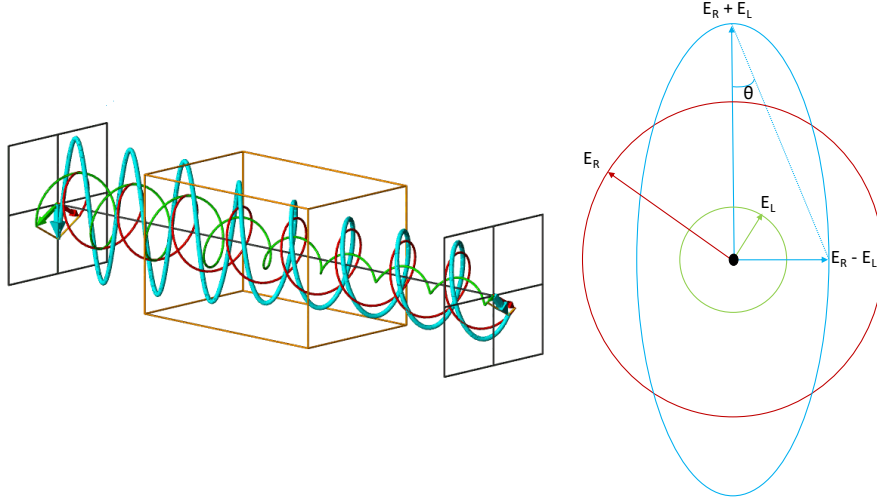


Figure 2.26 – CD principle. The PPW is drawn in cyan and its two components, the right CPW and the left CPL are drawn in red and green respectively. The image on the left shows the effect on the three waves of traversing a material (orange box) with CD. On the right, a representative scheme of the three vectors and the resulting ellipticity.

when matter has different refractive index for right and left CPW. In consequence, CB rotates the plane of polarisation of the PPW.

This two phenomenon normally occur together, it is rare to find materials that just present one of both. The combination of these two phenomenon transform a PPW into an elliptical polarised light with its big axis rotated. They are both caused by the molecular asymmetry of matter and they serve to extract important information about the structural properties of macromolecules.

The differential absorption measured is usually very small, from a few $1/100^{ths}$ to a few $1/10^{ths}$ of a percent. However, it can be determined accurately. Equation 2.29 serves to calculate the ellipticity of the sample from the raw data:

$$\theta_d = \frac{2.303}{4} \cdot (A_L - A_R) \cdot \frac{180}{\pi} \cdot [deg] \quad (2.29)$$

However, we need to normalize the data to be able to compare the ellipticity values. Equation 2.30 calculates the mean molar ellipticity per residue (MRE or $[\theta]$), that considers the path length (l), the concentration (C), the molecular mass (M) and the number of peptide bonds n_r . Accordingly, we can compare the MRE between protein of different lengths.

$$[\theta] = \theta_d \cdot \frac{M}{C \cdot l \cdot n_r} \quad (2.30)$$

In this manuscript, CD assays are carried out to study the Secondary Structure (SS) of peptides. Peptide bonds present CD and the intensity of the transitions depend on the ϕ and ψ angles. This

means that the absorbance obtained from a CD spectra will vary depending on the peptide SS: random coil, α -helix or β -sheet (Fig. 2.27). The problem faced with CD assays is that the spectra can be altered by some factors as the presence of aromatic residues or disulphide bonds. Phe, Trp and Tyr present aromatic rings in their lateral chains with absorption bands between 250 and 320 nm. This residues can complicate the interpretation of CD results in the far UV spectra. Of note, we perform CD experiments on samples that contain liposomes which diffuse light. Accordingly, our Jasco J-815 spectrometer was unable to go below 200 nm because of too much noise.

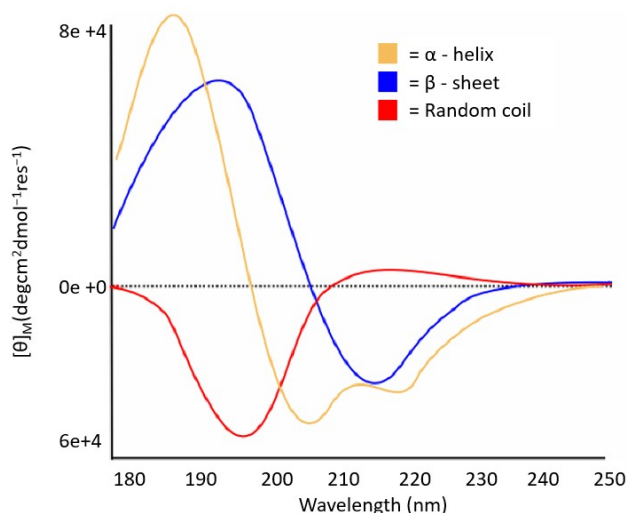


Figure 2.27 – Model of a CD spectra applicable to peptides and proteins: in yellow, the resulting CD curve of an α -helix, in blue the CD curve of a β -sheet and in red the CD curve of a random coil. Image adapted from: <https://www.creative-proteomics.com/>

2.2.2 Fluorescence

Fluorescence is a phenomenon caused by the capability of certain molecular systems to absorb and emit light. A molecular system that presents fluorescence is called a fluorophore. A fluorophore can absorb light of high energy (short wave length). This light excites the electrons and makes them transit from the grounded state to the excited state. When the electrons relax, they loose a small quantity of energy and then they release the stored energy, emitting a photon at a different wavelength. The lag time between the absorption and the emission last several nanoseconds and it is called fluorescence lifetime. The emitted light has lower energy (larger wave length) than the absorbed light. The difference between the two is called Stokes shift. This phenomenon is illustrated in figure 2.28

Some proteins have fluorescence properties, being able to emit light on a range of a few 10's of nm wideness. The reasons a protein can be a fluorophore are diverse. One of the most common

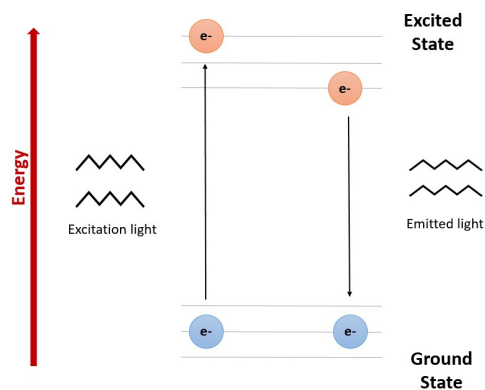


Figure 2.28 – Illustrative image of the fluorescence phenomenon

reason is the presence of Trp in its sequence. Protein fluorescence is a very interesting property because the obtained spectra varies depending of the fluorophore environment. This enables to design assays to determine protein interactions with other molecules like lipid membranes.

2.2.3 NMR

Nuclear Magnetic Resonance (NMR) is a biophysical technique used to study structural and dynamic properties of matter. Atomic nuclei with an odd number of neutrons or protons like ^1H or ^{13}C , present a property called spin that is used in NMR spectroscopy. In the case of hydrogen isotope ^1H , its nucleus has one proton and no neutrons. Hydrogen presents a positive charge and a charge rotating in space creates its own magnetic field. This phenomenon is called spin magnetic moment. In a sample of protons, each nucleus has its magnetic moment oriented in a different direction. If an external strong magnetic field is applied, the magnetic moments will align either parallel or antiparallel to the magnetic field direction and they will precess like a spinning top (Fig. 2.29.A). If they are aligned with the sense of the magnetic field they are in an α -spin state and if they align against the sense of the magnetic field they are in a β -spin state. The α -spin state has slightly lower energy and, in consequence, it is more stable. The difference in energy (ΔE) between the α state and the β state is proportional to the precession frequency of the magnetic moment, also called Larmor frequency (ν_0):

$$\Delta E = h\nu_0 \quad (2.31)$$

where h is Plank's constant. This frequency depends on the strength of the applied magnetic field (B_0) and it is specific of each atom nucleus type:

$$\nu_0 = \frac{\gamma}{2\pi} B_0 \quad (2.32)$$

where γ is the gyro-magnetic ratio of the particular nucleus. It is possible to force a nucleus to

switch from α state to β state. To do this, the sample needs to be irradiated with radio-frequency magnetic waves (B_1 field). The frequency of the pulse should be equal (or close) to the precession frequency of the spin. When the spin falls back to the α state, it emits radio frequency energy. If the spins are constantly flipping back and forward between the two states, a phenomenon called resonance is produced, giving the name to this technique. The signal is detected by a receiver coil as a time domain signal, called Free Induction decay (FID), which yields an NMR spectrum after Fourier transformation. The resonance frequency of a specific nucleus relative to a reference frequency of a standard compound is called chemical shift. The chemical shift is independent on the applied field B_0 and depends only on the nucleus type and its molecular environment. Only chemically equivalent atoms with the same molecular environment will present identical chemical shifts. Irradiation pulses can have different nature and duration. Depending on the sequence of pulses applied, different properties of spins can be studied. Figure 2.29.B shows a one-dimensional ^1H -NMR spectrum of ethanol, where the chemical shifts for the three equivalent proton groups are observed. In practice, one-dimensional NMR spectra of macromolecules cannot be accurately analysed because of the crowding of spectra. To overcome this problem, 2D NMR experiments have been designed. NOESY, COSY and TOCSY are different pulse sequences that generate 2D NMR spectra. NOESY experiments give information about the distances between the nuclei within the molecule. This information is called dipolar coupling and it helps to calculate the 3D structure of the molecule. It is also used in MD simulations to validate the reliability of the results. It is also possible to deduce the sequence specific assignment from the dipolar coupling as the spectra record interactions between H atoms from adjacent residues. COSY and TOCSY experiments give information about covalently connected hydrogens. This connection is called scalar coupling J and is typically transmitted via one, two or three covalent bonds between coupled nuclei. Three-bond J couplings allow to calculate the values for the dihedral angles. These values can be used to generate restraints for MD simulations or to validate the MD FF.

The applications of NMR spectroscopy are numerous: analysis of molecular structures, identification of unknown molecular substances, analysis of mixtures, etc. In this manuscript computational results are frequently compared to NMR data to validate the simulation protocols.

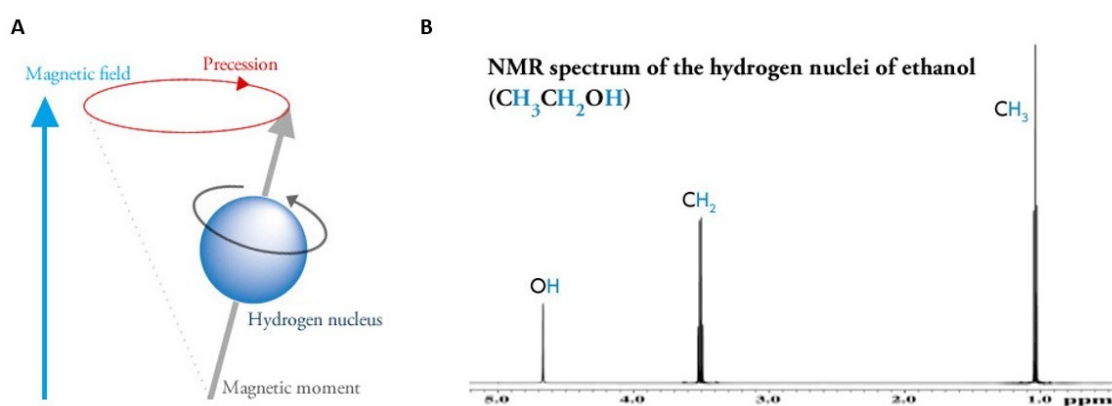


Figure 2.29 – NMR Spectroscopy. (A) NMR principle. A magnetic field is applied to a hydrogen nucleus. The magnetic moment of the nucleus is aligned with the magnetic field in an α -spin state. The spin of the nuclei presents a precession motion. (B) NMR spectrum model. The number of chemical shifts (peaks) is equal to the number of equivalent proton groups. The signal integration is proportional to the number of protons in the group. The multiplicity of peaks within each massif is due to the spin-spin scalar coupling and it is related to through-bond interactions between nuclei in the molecule. Image adapted from: <https://www.jeol.co.jp>

Chapter 3

Force Field Comparison for Amphipathic Helices

3.1 Introduction

Molecular dynamics (MD) simulations need to have reliable force fields (FF) to accurately reproduce the structural and dynamic properties of the system of interest. For the scientific community it is very important to have FF comparison studies to choose the one that better fits to describe a specific molecular system. To this aim, we have carried out a project of FF comparison on our system of interest for AH/membrane simulations. The objective is to perform simulations of amphipathic helices (AH)/membrane interactions with different FF, and then compare the results with experimental data to determine the accuracy of the FF. Comparing simulations results with experimental data is an excellent approach to validate simulation models, interpret experimental results at the atomic level or for FF calibration. Citing ref [188] of Sapay and Tieleman in 2008, they were warning the community for "*the urgent need for experimental data that allows a critical validation of simulation approaches*" on lipid protein interactions.

In literature, we can find many papers assessing the quality of force fields for proteins / peptides on one hand, or lipids on the other hand, but the combination of both at the same time has not been tested so much. We can still mention these following works. The most complete was published in 2017 by the team of Böckmann compared 4 different FF (or combination of FF), namely GROMOS54a7, CHARMM36 and the two force field combinations Amber14sb/Slipids and Amber14sb/Lipid14 [185]. They studied several different properties: the conservation of the secondary structure of transmembrane proteins, the orientation of transmembrane peptides in the lipid bilayer, the insertion depth of unfolded peptides at the membrane interface, and the free energy of insertion energies of Wimley-White peptides [223]. The conclusion shows that CHARMM36 and Amber14sb/Lipid14 give overall the best results, while GROMOS54a7 present more important

deviations. This work was very interesting because it was one of the first to test some properties on membrane proteins and peptides. Two other works assessed the AH melittin (an antimicrobial peptide). In 2013 Anderson *et al.* showed that melittin in DOPC was better represented by CHARMM36 than OPLS [8]. Another work from Fox *et al.* in 2018 tested many combinations of force fields and water models (namely using the same FF acronyms than the authors, ff99/tip3p, ff14SB/tip3p, c22/tip3p, c22/tips3p, c36/tip3p, c36/tips3p, c36m/tip3p, c36m/tips3p, and g53a6/spc using the same FF acronyms than the authors, tips3p is the tip3p specific to CHARMM [142] on melittin in water or in a POPC membrane [72]. They utilized H-REMD (Hamiltonian Replica-Exchange Molecular Dynamics) and also performed some CD experiments. Their conclusions show that the combinations c22/tips3p and c36/tips3p give the best results. Amber99SB and GROMOS53a6 appear to be biased toward beta-sheets structures. Some oddities appear in this work though: the very low content of helix in POPC (16 %), the lipid composition which is different between the CD and the H-REMD, the FF that are not properly cited, etc.

Since we are interested in AH, we wanted to go further in the testing of the different FFs peptides / lipids. We were especially interested to test the Berger lipid force field, which is a united-atom (UA) FF, and the different way to combine it with other protein FF. To explain the context, this project was started by my Ph.D director, Patrick Fuchs, around 10 years ago. At that moment, the Berger FF [24] was one of the most used FF for lipids. At that time, there were not as many lipid FF as now. For example, CHARMM36 [123] had just been published and before that there was CHARMM27 [70, 122] for which one had to impose a surface tension to get a fluid bilayer. One great advantage of Berger was/is the fact that it is a united-atom (UA) FF. This means that CH3 and CH2 groups are represented with a single bead, a simplification of the system compared to all-atom (AA) models. UA models decrease the computational cost of the simulation while in general keeping the reliability [199]. Since phospholipids possess many aliphatic hydrogen atoms, this can divide up to almost 3 the number of lipid particles to simulate. Berger lipids were parametrized independently but were in fact a mixture of different FF: GROMOS87 bonded parameters [49], Lennard-Jones parameters from OPLS-UA [162], partial charges from Chiu *et al.* [47]. The aliphatic tails were improved in the Berger paper using simulations of long alkanes with the Ryckaert-Bellemans potential [24], that is, a cosine power series in which there is no 1-4 interactions (see Materials and Methods chapter for definitions). The improvement was achieved by targeting the alkane experimental density and heat of vaporization. Although this mixture of different FF, Berger lipids were working quite well and were massively used at that time.

Even though Berger initially contained parameters for lipids only, some authors showed that it was possible to combine them with other FF for proteins. In the late 90s, Tieleman and Berendsen made a great contribution to the community of membrane proteins. They combined Berger lipids with the GROMOS family of force fields [210]. Initially, the protein FF was GROMOS96 [83] but other successor (notably GROMOS54A7 [192, 121]) are also compatible and were used

in literature (e.g. ref [235]). One key for the success of Berger lipids was the fact it had been made available on Peter Tieleman's web page, which prompted many people to use them! Later on, Monticelli and Tieleman [209] showed that Berger lipids were also compatible with all-atom OPLS FF [111] for proteins. One can note here that the protein is AA and the lipids UA, which is perfectly doable as long as the interactions between the two are properly parametrized. Sapay and Tieleman [189] made some tests in 2010 between the so-called "Calgary lipids", UA lipids largely based on Berger but with some improvements, and CHARMM22 force field for proteins [142]. They found it was possible to mix both but some further tests were necessary to fully validate this approach. It offered a way to use CHARMM FF for proteins with UA lipids simulable at zero surface tension. Although this combination was very attractive to our community, it has not been used so much in literature since a year later the complete reparametrization of CHARMM lipids, coined CHARMM36 and usable at zero surface tension, went out [123]. Later on, Cordomi et al. [51] showed that Berger lipids were also compatible with the AMBER family of force fields such as AMBER99SB-ILDN [139]. Some serious and convincing tests were done on the free energy of partitioning of residues as well as on some membrane proteins. Furthermore, it offered a way to simulate AMBER membrane proteins / peptides with UA lipids (there are AMBER FF for lipids but they are all-atom [60]). For both reasons, the work of Cordomi *et al.* was another attractive combination for our community.

The other important alternative to Berger for united-atom (UA) lipids is the CHARMM-UA parameter set. Hénin *et al.* developed in 2008 a UA version of acyl-chains [234] usable with CHARMM22 proteins [142] making them an interesting alternative to CHARMM27 all-atom lipids [70, 122]. Consistent with CHARMM27 all-atom lipids, it was necessary to use a non-zero surface tension to get the right area per lipid (around 30 mN/m) or to simulate at constant area. In 2010, CHARMM36 all-atom lipids were released [123]. It was a real change of paradigm since it was then possible to simulate bilayers at zero surface tension compared to CHARMM27 lipids. In 2013, CHARMM-UA lipids were updated (coined C36UA) which made them usable under zero surface tension [129]. They were further refined very recently in the set C36UAr [234].

We see in this short literature survey that many efforts were put in the development of UA lipids. In the years 2010, Berger were by far the most used UA lipids. Possible reasons for that can be attributed to i) the availability of parameters on Peter Tieleman's web page at that time (at that time CHARMM-GUI [105] did not exist!), ii) the fact they were usable at zero surface tension, iii) despite some possible issues they were giving overall reasonable results. In this context, Patrick was interested in testing different combination of protein force fields together with Berger lipids: GROMOS54A7/Berger, OPLS/Berger and AMBER99SB-ILDN/Berger to see which one(s) reproduced more accurately experimental results on AH/membrane interactions. The first REMD simulations were done with his previous Ph.D students, Dr. Lydie Vamparys and Dr. Amélie Bacle.

In 2021, the context has evolved a lot. The computer power has progressed enormously as well as the MD software. Simulating all-atom force fields on rather small systems such as small AH / membrane is less of an issue than 10 years ago. Furthermore, the CHARMM family of FF is one of the, or *the* most popular FF. It contains the widest ensemble of parameters for biomolecules among all FF, including, proteins [123] and lipids [234]. The community is very active and regularly some new lipids are published enabling the simulation of complex membranes [154, 133]. CHARMM-GUI [105] allows now to very easily construct membrane system with any lipid composition by simple mouse clicks. Very recently, some new parameters for non-natural amino-acids attracted our interest even more on the CHARMM family of FF [54] (see LTX315 Chapter 6). Last but not least, CHARMM36 stands out for its capability to accurately reproduce lipid structural properties. An open collaboration project called NMRipids, compares lipids simulations of different FF with NMR experimental data to classify them by reliability. In all the publications, CHARMM36 has been classified as one of the most reliable FF of our days [34, 14]. For all these reasons, Patrick and I decided to also test CHARMM36 regarding its performances on AH/membrane systems we are interested in.

Regarding the simulations, we decided to use an enhanced sampling technique to carry out this project. The objective was to explore the conformational space of the system finding a good trade-off between the computational cost / efficacy of sampling. We believe Temperature-Replica exchange molecular dynamics (T-REMD) was the most appropriate technique for this project. It increases the sampling by simulating several replicas at the same time (the number of replicas depends on the system) at different temperatures. Neighbour replicas exchange temperatures throughout the simulation to boost the kinetic energy of the system and increase the sampling. Furthermore, it allows to assess the thermal response of the AH, i.e. whether it is thermostable. The temperature range depends on the time-scales of the physical process that we want to observe, normally the bottom temperature is physiological and the upper temperature is very high (400-600 K). T-REMD was used in 2005 to describe a peptide insertion in a membrane for the first time [158]. WALP-16 was used as a model peptide and the bilayer was composed of DPPC lipids. They observed the insertion and folding of the peptide with a transbilayer orientation. Using T-REMD simulations we make sure the simulations perform an in depth exploration of the conformational space of the AH in the membrane. To have a detailed description of this method and our justification of choice 300-400 K for the temperature range, please consult the Materials and Methods chapter.

To perform the FF comparison, two different peptides were simulated in a membrane: the mastoparan and the ALPS23 from ArfGap1. Mastoparan (form L studied here) is a small peptide of 14 residues derived from wasp venom, its sequence is INLKALAALAKKIL-NH₂ (uncharged C-term which is amidated). It is a well-known peptide commonly used as a model for AH studies [62]. Moreover, we know its secondary structure [91]: its small size α -helix. Because of its high helix

content, we consider it as a "strong" α -helix. The helical wheel in figure 3.1.b, shows a beautiful AH with a hydrophobic side presenting small residues, and a polar side with charged residues. The mastoparan presents a well-balanced amino acid (aa) distribution that facilitates its partition and folding in the membrane.

ALPS (Amphipathic Lipid Packing Sensor) is a general motif found in many peripheral proteins. It has been shown to bind only to curved membranes [63] or more generally to bilayers enriched in packing defects [216, 141]. We simulated the ALPS23 peptide, which is an ALPS motif derived from the ArfGAP1 protein involved in vesicular trafficking. It is 23 residues long, its sequence is DFLNSAMSSLYSGWSSFTTGASK. Its secondary structure in SDS micelles has been resolved by NMR assays and it is available in the PDB with the code 2M0W (results unpublished). Some CD experiments showed it is moderately helical [REPORT the perc of helix in Vanni 2013] when bound to a lipid bilayer [216]. Previous studies [26, 27, 63] have shown that ALPS23 is a curvature sensor: it specifically binds to highly curved membranes. This sensitivity was ascribed to his particular amino-acid (AA) sequence [63]. This can be seen on its helical wheel presented in figure 3.1.a. The hydrophobic side presents residues with large side chains (Ile, Leu, Phe) that need a lot of free space to partition in the membrane. The polar side contain many small residues (Ser, Thr) and it is barely charged. This particular AA distribution makes ALPS23 a weak helix, with difficulties to bind and fold in membranes. It needs highly curved membranes, or more generally large and numerous packing defects. As its polar side chains are small and mostly non-charged, ALPS23 electrostatic interactions with the membrane are scarce. For these reasons, ALPS23 (or any ALPS motif) acts as a curvature sensor. Because it does not bind to all lipid compositions and its low helix content when bound to liposomes, we consider it as a "weak" helix. Because of this "weakness", ALPS23 represents a real challenge for lipid / protein FFs.

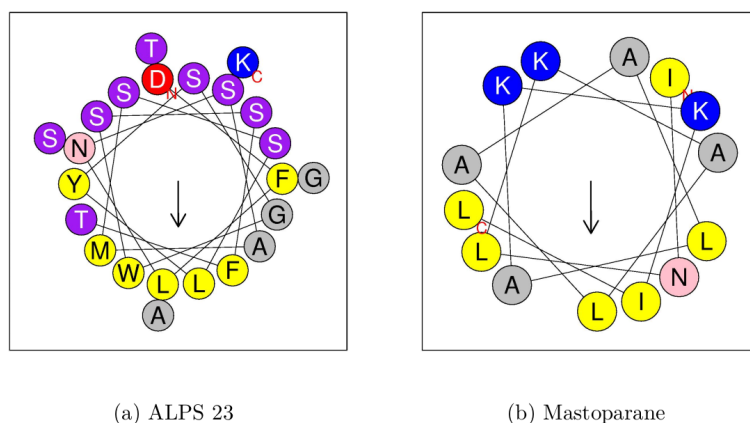


Figure 3.1 – ALPS23 and mastoparan helical wheels predicted with heliquest [77].

The mastoparan and ALPS23 are AH with very different properties: the mastoparan is a short and strong helix with high binding capacity and ALPS23 motif is a longer and weak helix, difficult to insert in bilayers. The use of these two very different systems for the FF comparison will try out the FF reliability for AH/membrane systems.

We used the same membrane composition for both peptides. It was composed of 1,2-dioleoyl-sn-glycero-3-phosphocholine (DOPC) and 1,3-dioleoyl-glycerol (DOG) lipids at 85/15 molar ratio. A common problem of membrane systems is that simulated bilayers are flat. It is very difficult to reproduce the interactions with peptides that bind preferably to curved membranes like ALPS. Including DOG lipids in a flat membrane has been proven to mimic the properties of a curved membrane: DOG is a conical lipid with a small head that increases the number of the packing defects in the bilayer, as if the membrane was curved [141]. In consequence, DOG lipids facilitate the insertion of curvature sensors. For the initial conformation, the peptides were placed folded (full α -helix) in the membrane/water interface, with the hydrophobic side towards the aliphatic chains. This way, we can focus our efforts on analysing the FF capacity to reproduce the peptide structural properties in the membrane by comparison to experimental data.

3.2 Materials and Methods

3.2.1 Molecular Systems

Two system were simulated: mastoparan and ALPS23 both in the membrane. Mastoparan system was composed by one peptide, 60 DOPC lipids, 10 DOG, 40 water molecules per lipid and ions to neutralize the system (4 Cl⁻). ALPS23 system was composed by 86 DOPC lipids, 14 DOG, 40 water molecules per lipid and ions to neutralize the system (10 Cl⁻). ALPS23 system contains more lipids because the helix is longer than the mastoparan, as you can see in figure 3.3.

3.2.2 Force fields

In the following I describe the simulations I performed using the CHARMM family of FF, the other ones (OPLS[111]/Berger, AMBER99SB-ILDN[139]/Berger, GROMOS54A7[192, 121]/Berger) were done by previous Ph.D students Lydie Vamparys [214] and Amélie Bacle [13]. The CHARMM36m for the peptide [93] and the CHARMM36 lipids [234] were used. For DOG, we used the parameters described in the PackMem paper [78]. The modified TIP3P water model for CHARMM was used for water [142].

3.2.3 REMD simulations

Two T-REMD simulations were carried out, one on Mastoparan and one on ALPS23. To be comparable, the conditions in these simulations are equal to those performed by Lydie and Amélie with the other FF combinations: temperature range between 300 and 400 K, 0.2 probability of exchange and pressure set at 1 bar. For CHARMM36 simulations, 33 replicas were needed. For the starting structure, the peptides were initially placed already folded in the membrane/water interface, as it is shown in figure 3.2.

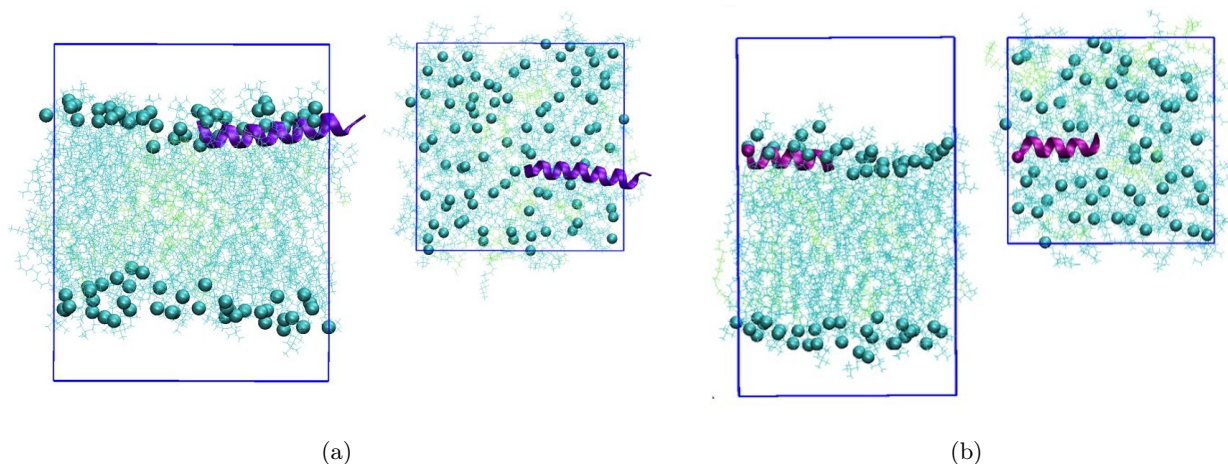


Figure 3.2 – T-REMD initial structures for (a) ALPS23 and (b) mastoparan systems. The images show a front and an upper view. Peptides are shown in purple, DOPC lipids in blue and DOG in green. DOPC phosphor atoms are represented with a bead. The images are extracted from CHARMM36m simulations.

3.2.4 MD simulations of the membrane only

To compare the effect of the peptide on lipids, one classical MD simulation of a system containing only the membrane was launched using CHARMM36 FF [123]. The composition of the membrane was the same used for the simulations with the mastoparan: 60 DOPC, 10 DOG, and 40 water molecules per lipid (TIP3P model [142]). Temperature was maintained constant at 300 K and pressure was set at 1 bar. The simulation is 500 ns long.

3.3 Experimental information

Mastoparan secondary structure was solved by NMR assays in the presence of detergent and a DMPC/DMPG membrane (code PDB 1D7N [91]). The structure is shown in figure 3.3b. In the membrane, the mastoparan is folded in an α -helix with unfolded termini, with an helicity value of 71%. Circular dichroism (CD) titration assays of the mastoparan were also performed in the presence of DOPC/DOG LUV (85/15 molar ratio) by our collaborator Guillaume Drin (Nice, Sophia-Antipolis). The results are shown in figure 3.4.a. The spectra have two minimum located at 208 and 222 nm, which is characteristic of an α -helix. The minimum presents lower values with increasing lipid concentration. This means that the peptide is more helical when peptide concentration increases. From the ellipticity value obtained at 222 nm, it is possible to extract the helicity. Figure 3.4.b shows the relationship the lipid/peptide ratio in terms of concentration and mastoparan helicity. As it was said before, mastoparan helicity increases with lipid concentration: at the highest lipid concentration, 7.5 mM, the mastoparan is folded at 80%, which is slightly overestimated compared to the NMR structure. This probable overestimation comes from the

difficulty of determining its concentration since it contains no Trp, it is thus required to do some other assays (BCA test). All in all, we see that mastoparan helicity is high ($>70\%$).

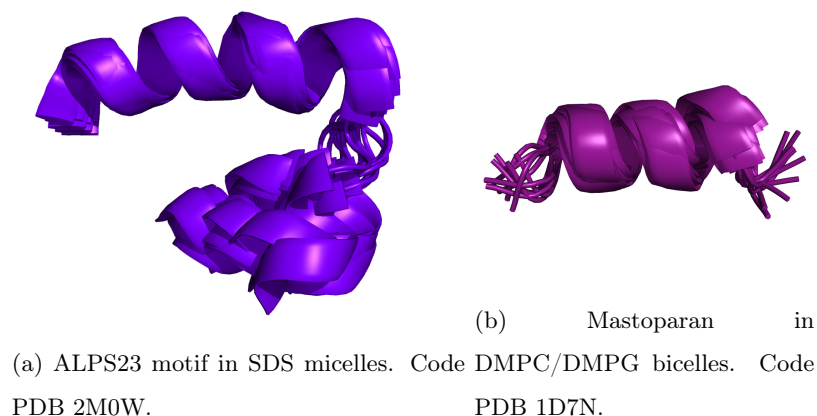


Figure 3.3 – NMR resolved structures.

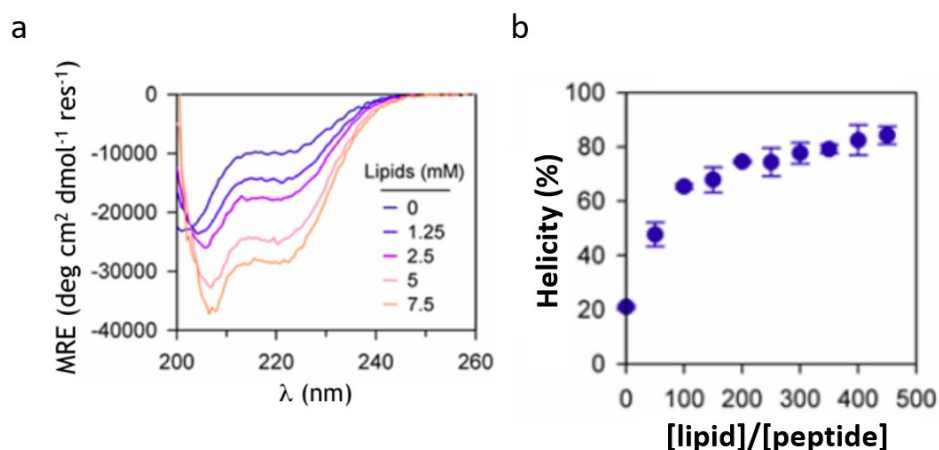


Figure 3.4 – CD assays of the mastoparan in DOPC/DOG 85/15 LUVs. (a) CD spectra at different lipid compositions. (b) Helicity plot with respect to the lipid/peptide ratio in terms of concentration. Unpublished data.

There is also experimental information available for ALPS23. Its structure has been resolved by NMR assays in SDS micelles and it is published with the PDB code 2M0W (publication not available yet) by a collaborator from Rennes (Liza Mouret). The structure is shown in figure 3.3a. Micelles are highly curved structures with only one lipid layer where a curvature sensor like ALPS23 can easily get inserted thanks to the big packing defects. Thanks to micelles curved structure, ALPS23 helicity is 73.9%. ALPS23 is a weak helix, it really needs favorable conditions to get folded. Unfortunately, no structure has been resolved in bicelles, which are structures

much similar to the membranes we simulate in MD. CD assays were also done for ALPS23 in DOPC/DOG 85/15 large unilamellar vesicles (LUV) by L. Vamparys and G. Drin. The results are shown in figure 3.5. When there are no lipid present, the spectra shows an unfolded peptide in buffer. ALPS23 helicity increases when lipids are added to 27 %. At a lipid concentration of 7.5 mM, ALPS23 CD spectra corresponds to an α -helical structure with the two negative peaks at 210 and 222 nm.

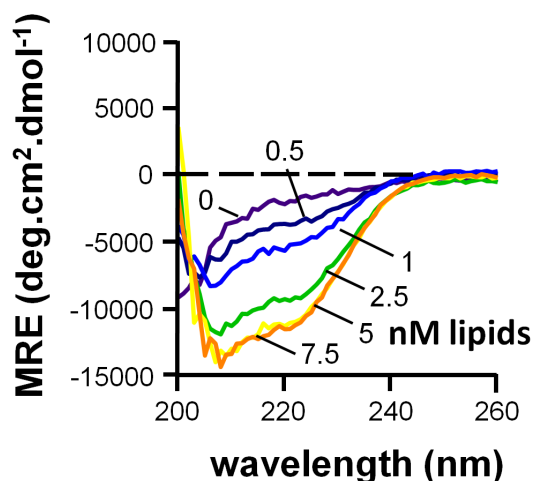


Figure 3.5 – ALPS23 CD spectra in DOPC/DOG 85/15 liposomes. Lipid concentration increases from 0 to 7.5 mM.

There is also CD data available for ALPS35 in membranes, which is a longer sequence of the same peptide. Assays were carried out by Dr. Lydie Vamparys, previous PhD student of Patrick Fuchs. They performed CD spectra on small unilamellar vesicles (SUVs), composed of DOPC 3.6.a); and large unilamellar vesicles (LUVs), composed of pure DOPC or DOPC/DOG at 85/15 molar ratio 3.6.b) [216]. The SUVs used had an average radius of 20 nm, measured by DLS (Dynamic Light Scattering). The curvature of the membrane surface is very high. In this configuration, we see that ALPS35 binds to the vesicles and folds into an α -helix with a helix rate of 38 %, as shown by the two negative peaks at 208 and 222 nm in the spectrum. In the second CD experiment performed, LUV vesicles, with an average radius of 100 nm were used. At this size, the membrane surface is locally perceived as flat for the peptide. When the vesicle is composed of DOPC only, ALPS35 barely binds and folds. However, including DOG molecules in the composition increases ALPS35 binding and helicity to 29 %. This is because DOG is a conical lipid with a small head which increases the number of packing defects of the membrane, facilitating the insertion of ALPS35 big lateral chains [216]. However, the helicity is still quite low, showing again ALPS35 "weak helix" nature. Comparing results in SUVs and LUVs, the value at 222 nm is larger for SUVs of pure DOPC, showing that ALPS35 is less helical at the LUV surface of DOPC/DOG than at the SUV surface of pure DOPC.

Thanks to these experimental data, we have information of mastoparan and ALPS23 structural properties in the membrane. The mastoparan is a strong helix that easily binds membranes of

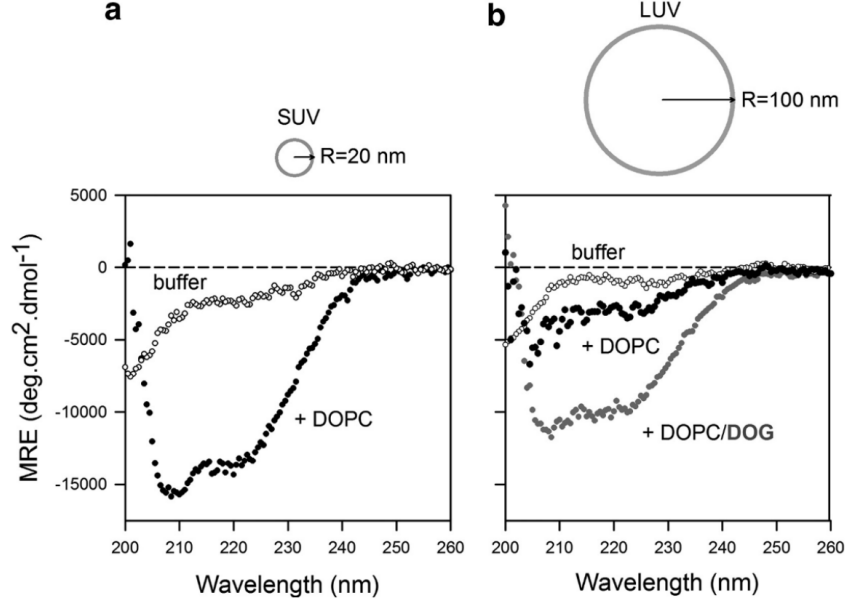


Figure 3.6 – ALPS35 CD assays in SDS micelles. First, an scheme of SUV and LUV relative size is shown. (a) CD spectra of ALPS35 in buffer and pure DOPC SUVs. (b) CD spectra of ALPS35 in buffer, in pure DOPC LUVs and DOPC/DOG LUVs at 85/15 molar ratio. Image from [216]

any composition. In contrast, ALPS23 is a weak helix. For ALPS23 to bind flat membranes, conical lipids are needed to increase the number of packing defects. For this reason, the membrane composition chosen for all T-REMD simulations was DOPC/DOG at 85/15 molar ratio.

3.4 Computational results

In this section, we are going to compare the results of the four T-REMD simulations carried out with the different FF. The comparison is focused on AH secondary structure, to check whether the FF are able to reproduce the peptide helicity in the membrane or not. For this analysis, we are going to use the conformations simulated at 300 K. This means that we do not follow continuous replicas over the temperatures, we only use the ensemble of conformations that have been simulated at the bottom temperature. This is because we want the peptide to follow the same Boltzmann distribution than in the experimental assays: the probability (p_i) of observing the peptide in a certain state i , with a total energy (E_i), depends on the temperature T . Figure 3.7 shows the secondary structure analysis for all the FF and for the two molecular systems at 300 K. On the left, the results for the mastoparan are showed and, on the right, for ALPS23. On the graph, the secondary structures are shown with different colors per residue and through time. The blue colour represents the α -helix. Therefore, the more present the blue in the graph, the more the FF retains the helical structure of the peptide. Helicity mean values are shown in table 3.1.

For the mastoparan, the NMR structure shows that it is 71% helical. When using GROMOS54A7

FF, the α -helix quickly disappears. The mean helicity value is 3%, very far from experimental values. On the contrary, the OPLS, AMBER99SB-ILDN and CHARMM36m FF, preserve the α -helix along the simulation with 67 and 71, and 76% of helix content. They all reproduce the experimental value very well.

We performed the same simulations and analyses for the ALPS23 peptide. Once again, GROMOS54A7 FF is not capable to maintain the initial helix during the simulation, retaining only 6% of helicity. However, the process is slower. Surprisingly, OPLS FF also fails to reproduce ALPS23 secondary structure, retaining also 6% of the helicity. However AMBER99SB-ILDN and CHARMM36m FF succeed in maintaining ALPS23 helicity, with 67 and 77.05 % of helix content respectively. AMBER99SB-ILDN and CHARMM36m FF give the more reliable results compared to NMR and CD data.

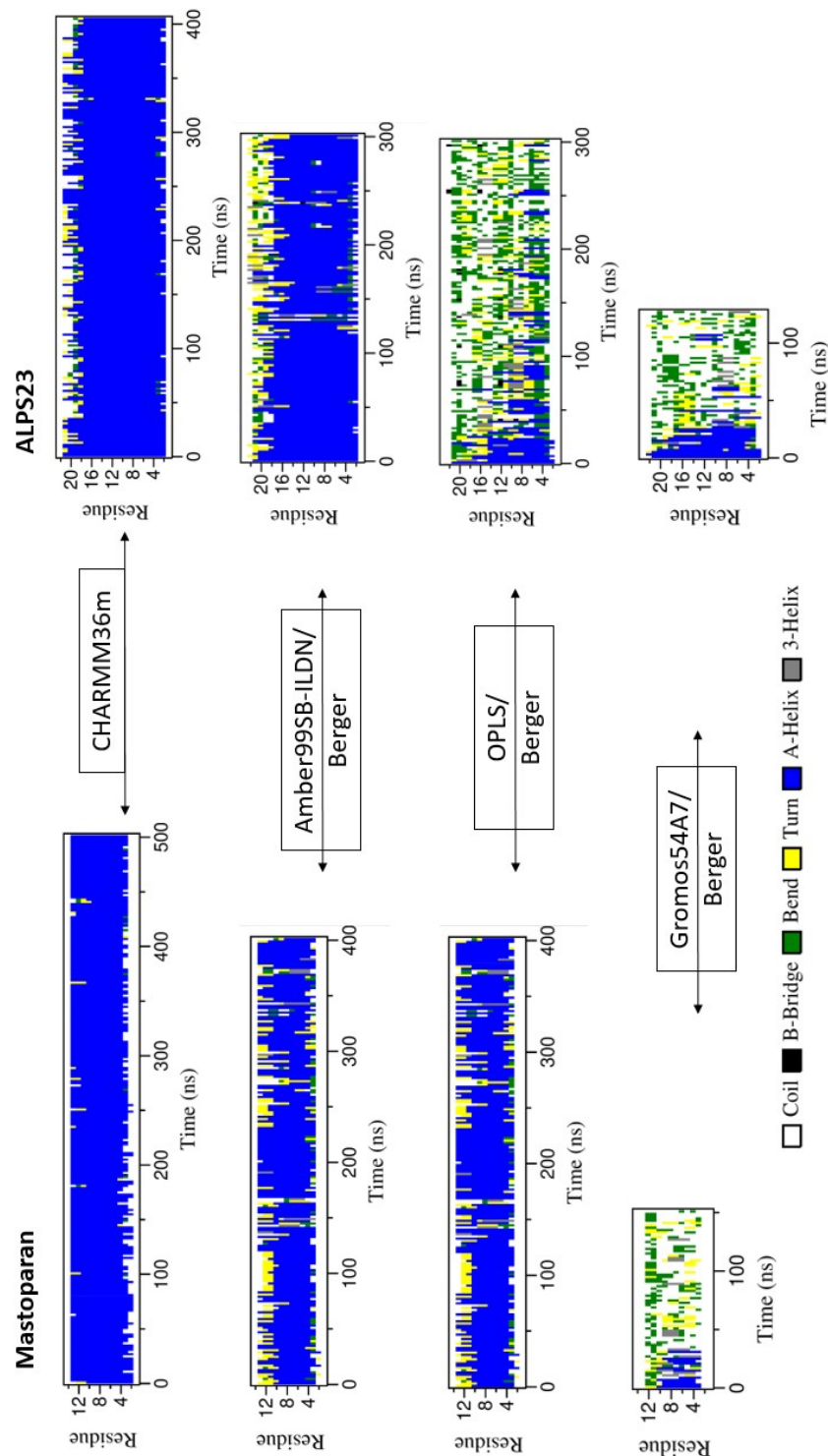


Figure 3.7 – Secondary structure analysis. The peptide secondary structure is plotted through time by residue. For these analyses, T-REMD conformations at 300K are used. We do not follow a continuous replica over the temperatures. On the left, results for the mastoparan; on the right, results for ALPS23. Analysis for the four FF are shown: CHARMM36m, AMBER99SB-ILDN, OPLS and GROMOS54A7. Color white is for random coil, black for β -bridges, green for bends, yellow for turns, blue for α -helices and grey for 3_{10} -helices.

	ALPS		Mastoparan	
Force Field	Helicity (%)	STDV	Helicity (%)	STDV
CHARMM36m	76	7	76	5
AMBER99SB-ILDN	67	11	71	21
OLPS	6	10	67	19
GROMOS54A7	6	9	3	9

Table 3.1 – Helix content for mastoparan and ALPS23 peptide in the membrane simulated with different FF. Only the conformations found at 300K are taken into account.

Interestingly, CHARMM36m FF seems to overestimate the helicity compared to experimental results. For both peptides, CHARMM36m FF reproduces a strong and stable α -helix with high helix content, which is surprising for ALPS23. To have more information about this, the helicity was calculated at all the temperatures for CHARMM36m and AMBER99SB-ILDN simulations. The mean values are shown in figure 3.8. CHARMM36m predicts higher helicity for both peptides at all temperatures. The helix content is maintained constant although temperature increases, variations are minimum. Regarding AMBER99SB-ILDN for both peptides, the mean helicity values is slightly lower and matches better the experimental results. Helicity decreases slightly with the increase of the temperature, although a quite high helix content is maintained. Interestingly, we observe a few conformations completely unfolded at high temperature (400 K), explaining the decrease of the mean helix content with increasing temperature, but they remain quite rare. The presence of these fully unfolded conformations explains the important standard deviations compared to CHARMM36m. To sum up, CHARMM36m predicts a very thermostable peptide for both mastoparan and ALPS23. For AMBER99SB-ILDN, we observe an effect of temperature, but not a complete melting as if we had two populations folded and unfolded. The accuracy of these results is discussed in the Discussion section below.

3.5 Membrane properties in CHARMM36m simulations

I carried out CHARMM36m/CHARMM36 T-REMD simulations for this project. I have been using this FF frequently during my Ph.D. For this reason, I have special interest in validating the reliability of this FF. In the previous section, we have checked the FF capability of reproducing AH structural properties in the membrane. In this section, we study CHARMM36m/CHARMM36 capability of reproducing membrane structural properties, specifically the two most common, the area per lipid (APL) and the thickness. To this aim, we carried out one classical MD simulation containing the membrane exclusively. The results are compared to experimental data and also to the values found in mastoparan T-REMD simulations at 300 and 400 K.

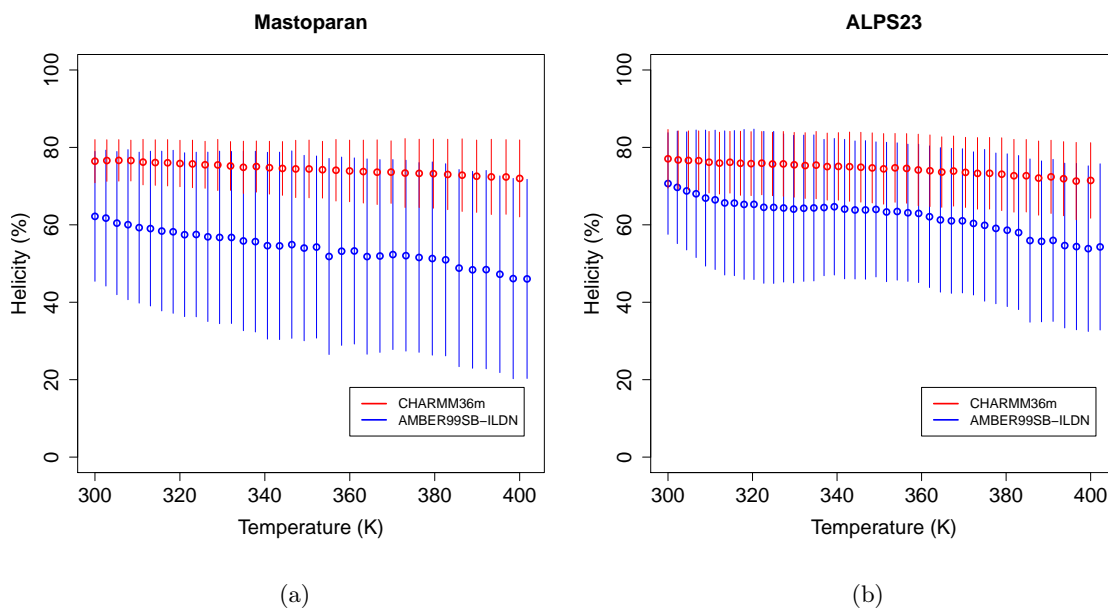


Figure 3.8 – (a) Mastoparan and (b) ALPS23 helicity per temperature. The conformations found in CHARMM36m (red) and AMBER99SB-ILDN (blue) simulations were classified by temperature. The mean helicity values are given in the plots with points, the lines represent the standard deviations.

3.5.1 Area Per Lipid

The area per lipid (APL) is one of the most common descriptors of a lipid bilayer structural properties. It can be measured by experimental assays and is also a good estimator of the FF reliability. X-ray scattering data of DOPC bilayers at 30°C determined a mean APL of 67.4 \AA^2 [125]. Table 3.2, shows the APL values found in the MD and the T-REMD simulations. The MD simulation of the membrane without peptide at 300 K, reproduces exactly the APL value found experimentally for pure DOPC membranes. Unfortunately there is no experimental data available for DOPC:DOG membranes. In T-REMD simulations, mastoparan is present. Conformations at 300 K show a DOPS APL of 0.62 \AA^2 , a bit smaller than the simulations without peptide. Mastoparan seems to slightly pack the lipids decreasing their area. However, at 400 K the APL increases notably. DOPC mean APL is 0.76 \AA^2 . Increasing the temperature increases lipid motility and separates them away in the membrane. DOG follows the same tendency than the DOPC.

Simulation	Temperature (K)	Lipids	APL (nm^2)	STDV (nm^2)
REMD with peptie	300	DOPC	0.62	0.01
		DOG	0.48	0.04
	400	DOPC	0.76	0.04
		DOG	0.63	0.08
MD without peptide	300	DOPC	0.67	0.01
		DOG	0.52	0.04

Table 3.2 – Area Per Lipid values for DOPC and DOG in MD and T-REMD simulations. In T-REMD simulations, the mastoparan is present, inserted in the membrane/water interface.

Thickness

The thickness is another important membrane structural property, closely related to the area: when a membrane reduces its area, the thickness increases and vice-versa. MD simulations without peptide show a membrane thickness of 4.22 nm . The results are consistent with the APL values discussed before. The thickness decreases when the peptide is present: the lipid are more compact so they need to order their aliphatic chains, increasing the bilayer global thickness. However, in the simulations at 400 K, the thickness decreases to 3.80 nm . As the lipids occupy more lateral space, their aliphatic chains are less ordered and the thickness decreases.

Simulation	Temperature (K)	Thickness (nm)	STDV (nm)
REMD	300	4.24	0.13
	400	3.80	0.09
MD	300	4.22	0.07

Table 3.3 – Thickness values for DOPC and DOG in MD and T-REMD simulations. In T-REMD simulations, the mastoparan is present, inserted in the membrane/water interface

3.6 Discussion

In this chapter, we have presented a FF comparison project for AH/membrane simulations. The reliability of reproducing AH structural properties in the membrane was assessed for four different FF combinations: GROMOS54A7/Berger, OPLS/Berger and AMBER99SB-ILDN/Berger and CHARMM36m/CHARMM36 for peptide/lipids. The results were conclusive, GROMOS54A7/Berger combination fails to reproduce neither mastoparan or ALPS23 experimental data. OPLS/Berger simulated mastoparan structural properties correctly but failed with a weak helix like ALPS23. In literature, the Ulmschneider brothers already observed the failure of OPLS to reproduce the helicity of melittin in DOPC [8]. The two FF that reproduced better the structural properties of

both peptides were AMBER99SB-ILDN/Berger combination and CHARMM36m. They predicted correctly the helicity values observed by NMR and CD assays. However, CHARMM36m slightly overestimates the helix content, especially for ALPS23. Regarding the secondary structure graph in figure 3.7, it seems that CHARMM36m predicts the folding of residues that are normally unfolded even in the membrane, as the two termini or the Gly in the case of ALPS23. This should be taken into account for future projects when extracting structural information from CHARMM36m simulations. Strangely, a recent work on melittin in DOPC found that CHARMM36 predicts an helicity slightly lower than those measured by CD, while CHARMM36m predicts a very low content of helix [72]. However in this work, the case is very peculiar since the helix content measured in their CD experiment is only 16 %. Furthermore the sources of the force field they used seem untraceable.

CHARMM36 reproduces correctly membrane structural properties in terms of APL and thickness as shown in the initial paper [123]. It also reproduces the decrease of APL when mastoparan is embedded in the membrane, a phenomenon that has been observed many times in literature such as in a recent ref. [240]. Its accuracy in reproducing lipid bilayer structural properties has also been checked in other projects such as NMRlipids [34, 14], where the order parameters of the polar heads (on the glycerol and choline carbon atoms) predicted by different FF were compared to solid state NMR data. CHARMM36m out-stands for its capacity to reproduce membrane structural properties and it is currently one of the most reliable FF for lipids.

In this project we have also used the ability of DOG-containing membranes of mimicking curved membranes structural properties [141]. CD assays showed that ALPS35 was capable of binding slightly curved membranes containing DOG but with less affinity than highly curved membranes [141]. In the simulations, ALPS23 stayed inserted into the membrane / water interface in an α -helical conformation during the simulations. As it is a weak helix, it cannot fold in flat bilayers. Including DOG in the membrane composition allow to carry out membrane simulation with properties similar to curved membranes facilitating curvature sensors binding and folding.

All the simulations were carried out using the enhanced sampling technique T-REMD. Simulations at high temperature showed that the peptide is thermostable in the membrane. Apparently, some AH are thermostable in the membrane because the bilayer acts as a conformational restraint, maintaining the peptide folded. The Ulmshneider brothers were the first to describe AH thermostability in MD simulations. They carried out elevated-temperature simulations of peptide / membrane systems and they were able to reproduce correctly the peptide structural properties in the membrane [212, 44]. They showed that melittin (an antimicrobial peptide) in POPC resists up to 95°C and keeps its helicity intact [44]! Regarding mastoparan, its thermostability was tested a few months ago up to 54°C by NMR assays [152]. The authors carried out studies on conformation and stability of mastoparan in methanol. Methanol was considered to be appropriate as a membrane mimetic for mastoparan. They observed that at low temperatures, mastoparan folds into an α -helix from residue 5 to residue 12 and the helix is maintained even at 54°C. Since Mastopa-

ran is an α -helix close to Melittin (their sequence has a related amino-acid composition), we can speculate it is probably thermostable up to higher temperatures than 54 °C. Mastoparan thermostability is also observed in this project up to 400 K with CHARMM36m. Figure 3.8 shows that the mean helix content barely varies when temperature increases. However, the helicity decreases a bit with AMBER99SB-ILDN. Thus, we can conclude that the thermostability of mastoparan is better reproduced by CHARMM36m than AMBER99SB-ILDN.

Regarding ALPS23, the peptide remains folded at all temperatures for CHARMM36m. For AMBER99SB-ILDN, the helicity decreases a bit when the temperature increases. As said above, predicting a thermostable peptide is expected for mastoparan. However, it is not for ALPS23. CD assays of ALPS35 in presence of DOPC/DOG (85/15) liposomes at increasing temperatures showed that this peptide is progressively denatured between 25 and 80°C [216]. At 80°C (353 K), there is almost no helicity left. Here we have a shorter version ALPS23, but we can expect it is not thermostable as well in liposomes. These results question the reliability of CHARMM36m to reproduce ALPS23 SS at high temperatures. We need to be cautious with structural information obtained at high temperatures and mainly focus on the results obtained at the physiological conditions. AMBER99SB-ILDN would be more in phase with experiments regarding its thermal response, even if it is probably still too stable at high temperatures. One might question whether we simulated long enough. The convergence of ALPS23 helicity with AMBER99SB-ILDN at the highest temperature (400 K) is presented in figure 3.9. We observe a relative convergence of the smoothed curve, although it is difficult to be really sure it has converged. It would be very interesting to prolong this REMD to at least 400 ns.

All in all, we showed in this chapter that REMD is definitely interesting for validating AH / membrane FF. It is a much more stringent test to pass than simply launching regular MD at room temperature. If the FF does not predict a stable helix, the ensemble of structures at the bottom temperature will show very rapid helicity loss. If on the other hand, the helix is maintained for a few hundreds of ns, its is a solid indication that the helix is predicted as stable by the FF. In literature, there have been many tests of FF on soluble proteins / peptides [100, 192], or pure lipids [133]. But the combination of both proteins and lipids are less tested [186, 98]. Here, the AH test is very challenging as it involves subtle details in the protein FF and in its interactions with the lipid FF. Reproduce the folding of AH is already difficult, but it is probably a greater challenge to properly reproduce AH sensitivity to temperature. Some delicate balance between the amino-acid composition, the AH fine interactions with lipids and most likely water, are all involved in a tricky way. The choice of T-REMD among all the enhanced sampling techniques is discussed further in next chapter.

To come back to the beginning of the chapter, Sapay and Tieleman were asking for more experimental data and more tests on protein / lipid systems [188]. We think AH / lipid systems assessed by REMD is a good step forward in that direction and it should be generalized for protein/lipid FF developments.

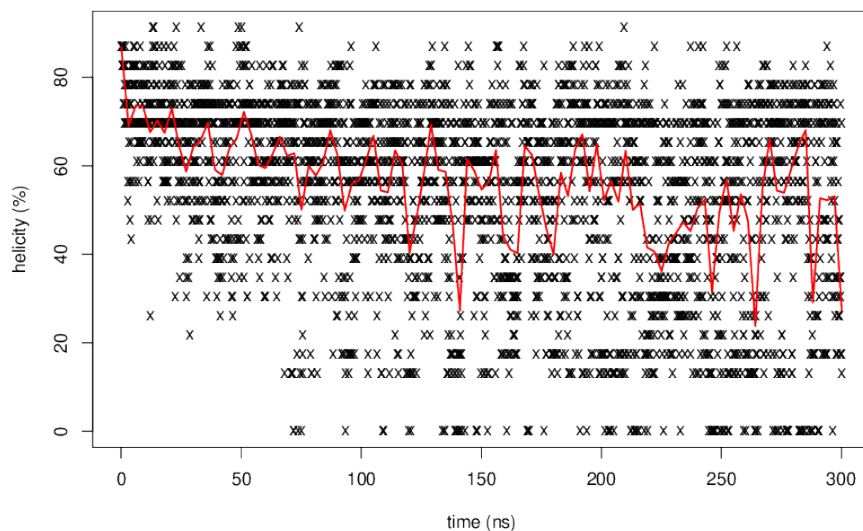


Figure 3.9 – Convergence of ALPS23 helicity with AMBER99SB-ILDN at the highest temperature (400 K). Each black symbol represents the helicity for each frame at 400 K. Because it can only take discrete values, it is hard to judge of its relative convergence. We thus added a smoothed curve in red using the `lowestess()` function in R with a smoother span of 0.01.

Next chapter is a continuation of this project. The two more reliable FF found in this project, AMBER99SB-ILDN/Berger and CHARMM36m, will be tested to check if they are able to reproduce correctly AH insertion and folding in membranes. With these two projects, we will be able to determine which is the best existing FF for AH/membrane simulations.

Chapter 4

Computational study on mastoparan insertion in the membrane

4.1 Introduction

Amphipathic Helices (AH) are generally extended in the solvent and they fold into an α -helix upon binding to the membrane. When talking about AH, the insertion mechanism is called partitioning [222] (For a further description of this process, please consult introduction section 1.4). The partitioning mechanism was described by Seelig in three phases: attraction, adsorption and folding [195]. During the attraction, the peptide approaches the membrane driven by electrostatic or hydrophobic interactions. The adsorption occurs when the peptide locates parallel to the membranes surface. The first interactions take place, but the residues are not inserted yet, so the peptide is still unfolded. Finally, the AH partitions in the membrane while folding at the same time. Depending on the peptide sequence, the insertion begins from one of the terminal ends or the center of the peptide and then the rest of the sequence partitions sequentially. The folding is coupled to the partitioning [223], they take place at the same time. The nucleation propagation model describes this mechanism: one hydrophobic residue gets inserted and the folding starts. Folding propagates through the sequence while the rest of the residues partition [136, 10].

Understanding the partitioning and folding process is a fundamental basis to decipher AH action mechanism. It has numerous applications, as, for example, in biomedicine. AH peptides are becoming more and more popular in drug design because of their simplicity and capacity of crossing cell membranes and deliver cargos. For example, designed AH peptides are being used as lead compounds to inhibit NUPR1 interactions, a multifunctional intrinsically disordered protein (IDP) involved in the development of pancreatic ductal adenocarcinoma (PDAC) [187]. Another example is the oncolytic peptide LTX-315, which is being developed by LytixBiopharma as treatment for sarcoma disease [204].

Numerous research groups have studied peptide/membrane interactions by computational techniques. Molecular Dynamics (MD) has enormously evolved in the past years, increasing the reliability of the results and decreasing the computational costs [211]. Within MD techniques, enhanced sampling methods have been developed to increase the amount of information obtained from the simulations. These methods allow to study long time scale processes and have been applied to peptide/membrane interaction research. Temperature-Replica exchange molecular dynamics (REMD) was used in 2005 to describe a peptide insertion in a membrane for the first time. WALP-16 was used as model peptide and the bilayer was composed of DPPC lipids. They observed the insertion and folding of the peptide with a transbilayer orientation. The results suggest that the insertion of the peptide precedes secondary structure formation [158]. The Ulmschneider brothers applied elevated-temperature simulations to study the interactions of two different peptide types with the membrane: synthetic transmembrane peptides from WALP family (W16 and W23) and melittin (interfacial peptide). Elevated-temperature simulations can be applied to membrane peptides because the bilayer protects peptides secondary structure, preventing denaturation. From these simulations they succeeded in observing the partitioning and folding of the peptides at atomic level, a process that is not possible to simulate yet with classical MD simulations due to the long time scales involved (microsecond-timescales) [212, 44, 43]. In 2011 metadynamics were used to study the mechanism of membrane curvature sensing by amphipathic helices [55], specifically, the amphipathic N-terminal helix of endophilin (H0), which targets curved membranes rich in packing defects. Last but not least, a recent article published in 2021, showed the novel application of replica-exchange umbrella sampling (REUS) simulations to study the insertion of a transmembrane peptide called pTB in the membrane. They succeeded in calculating the free energy of insertion, showing that there is a small barrier of 4.3 *kcal/mol* for the insertion from bulk water [115]. For further information about these researches and techniques, please consult the introduction chapter.

MD simulations have succeeded in accurately capturing the process of peptide binding, folding, and partitioning into lipid bilayers. In this chapter, the AH mastoparan from wasp venom, which has been presented in the previous chapter, is used as model system to study the insertion mechanism of AH in membranes. T-REMD simulations are carried out to observe the partitioning/folding process. First, T-REMD simulations of the peptide in water are launched to explore its conformational space in water. This simulations give reliable starting structures that we use afterwards to launch the simulations of the insertion. For the T-REMD of insertion, the peptide is initially placed in the solvent, a few *nm* away from the membrane. The peptide is free to move in the simulation box and to approach and bind the membrane. These enhanced sampling simulations are a good approach to explore the insertion mechanism and the secondary structures involved at atomic level. All simulations were done twice with two different force fields (FF): CHARMM36m, AMBER99SB-ILDN when the peptide is in a water box and AMBER99SB-ILDN/Berger combined for peptide/lipids where lipids are present. These FF were determined to be the most accurate for reproducing AH structural properties in the membrane in the previous chapter. Now, we want to test which one reproduces better the dynamics of the insertion. AMBER99SB-ILDN/Berger

simulations were done by Dr. A. Bacle under Patrick Fuchs supervision and CHARMM36m ones were done during this Ph.D by myself.

4.2 Materials and Methods

4.2.1 Simulations in water

In water, one T-REMD was launched with CHARMM36m FF, being the starting conformation the NMR structure 1D7N, which is folded into an α -helix. Two T-REMD were launched with AMBER99SB-ILDN FF, one also with 1D7N model as starting structure and the other one with an unfolded conformation. The starting structures are shown in figure 4.2.a. The three T-REMD were carried out with the same conditions for the results to be comparable: minimum temperature of 300 K, maximum temperature of 400 K, pressure set at 1 *bar* and 20 % probability of exchange between replicas. The systems are composed by one peptide, 2455 water molecules for CHARMM36m simulation, 3357 for AMBER99SB-ILDN's ones, and 4 chloride ions to neutralize the system. TIP3P water model was used for both FF [113, 142]. CHARMM36m simulation last for 500 *ns* and AMBER99SB-ILDN's for 100 *ns*. This is because CHARMM36m simulation takes longer to stabilize as seen in figure 4.2. For CHARMM36m system, 27 replicas were needed, 31 for AMBER99SB-ILDN. This difference was due to the difference in the number of water molecules. In both cases exchanges are attempted every 2 *ps*.

4.2.2 Simulations in the membrane

Two T-REMD of mastoparan insertion are presented in this chapter. One was carried out by myself using CHARMM36m FF. The other one was done by A. Bacle and combines AMBER99SB-ILDN FF for the peptide, TIP3P [145] for the solvent molecules and Berger for the lipids. CHARMM36m and AMBER99SB-ILDN are AA FF and Berger a UA FF. The two REMD were done with the same conditions for the results to be comparable: minimum temperature of 300 K, maximal temperature of 400 K, pressure set at 1 *bar* and 0.2 probability of exchange between replicas. The systems are composed by one peptide, 60 DOPC and 10 DOG lipids, 2800 water molecules for CHARMM36m simulation, 2796 for AMBER99SB-ILDN/Berger one, and 4 chloride ions. TIP3P water model was used for both FF [113, 142]. The bilayer composition was the same used in the previous chapter (Force Field comparsion Chp. 3.1) as it was demonstrated that it mimics curved-bilayer properties [216]. The simulations last for 300 *ns*. Exchanges were attempted every 2 *ps*. 33 replicas are needed for CHARMM36 simulation and 36 for AMBER99SB-ILDN/Berger. There is a difference between the number of replicas because Berger FF is a UA FF. Also, the number and size of the particles employed vary so do the number of replicas needed. The initial structures are extracted from the "mastoparan in water" simulations. AMBER99SB-ILDN's initial structure corresponds to the most populated cluster in water (Fig. 4.1.b). For CHARMM36m's T-REMD, two structures

are used, alternating between replicas. They correspond to the second and third most populated conformations and they are very similar: unfolded with one turn (Fig. 4.1.a).

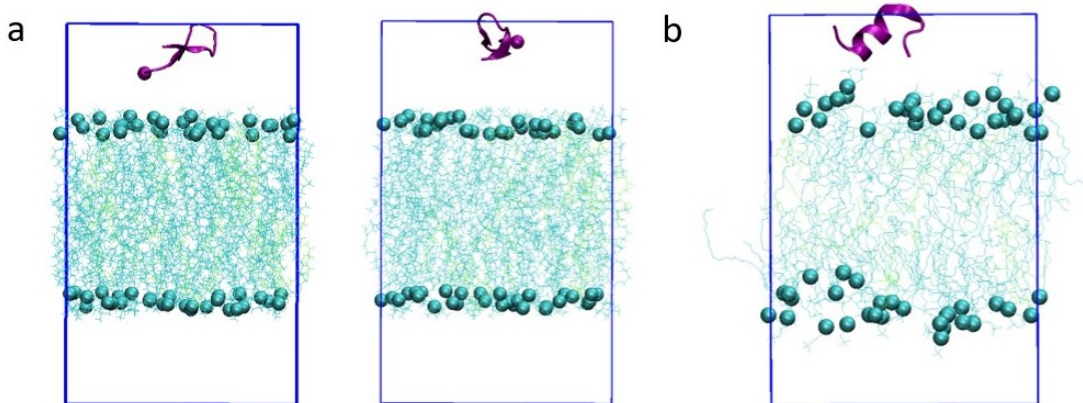


Figure 4.1 – Molecular system at time 0 *ns* for (a) CHARMM36m and (b) AMBER99SB-ILDN T-REMD insertion simulations. For CHARMM36m simulations, two different conformations alternate between replicas.

4.3 Mastoparan in water

4.3.1 Helicity content

The first thing to check on the simulations is the helicity content. Figure 4.2.b shows the evolution of the helicity through time at the bottom temperature for CHARMM36m and AMBER99SB-ILDN simulations. Notice that we are going to use the conformations simulated at 300 K. This means that we do not follow continuous replicas over the temperatures, we only use the ensemble of conformations that have been simulated at the bottom temperature. Comparing the two T-REMD with a folded initial structure, the behaviour is similar with both FF. The percentage of helicity decreases at the beginning of the simulation until it stabilizes around 20 % with high fluctuations. In AMBER99SB-ILDN's simulation with the unfolded starting structure, the helix content increases until it is stabilized at the same values than the previous two curves. Results are consistent: using different FF and starting structures, the mastoparan finally presents the same helicity content. Table 4.1 shows the statistics. Only the fraction of the simulations where the systems are equilibrated is considered for the analysis. Results are again consistent, they are very similar between the FF and independent from the starting structures. The standard deviation are very high, which means that the system is very dynamic and the peptide explores a wide range of conformations. CD assays carried out by our collaborations (Guillaume Drin, CNRS, Université de Nice) showed an helicity content of 20 % for the mastoparan in water (data not published yet), so our structural results are validated by experimental data, supporting the reliability of the FF employed.

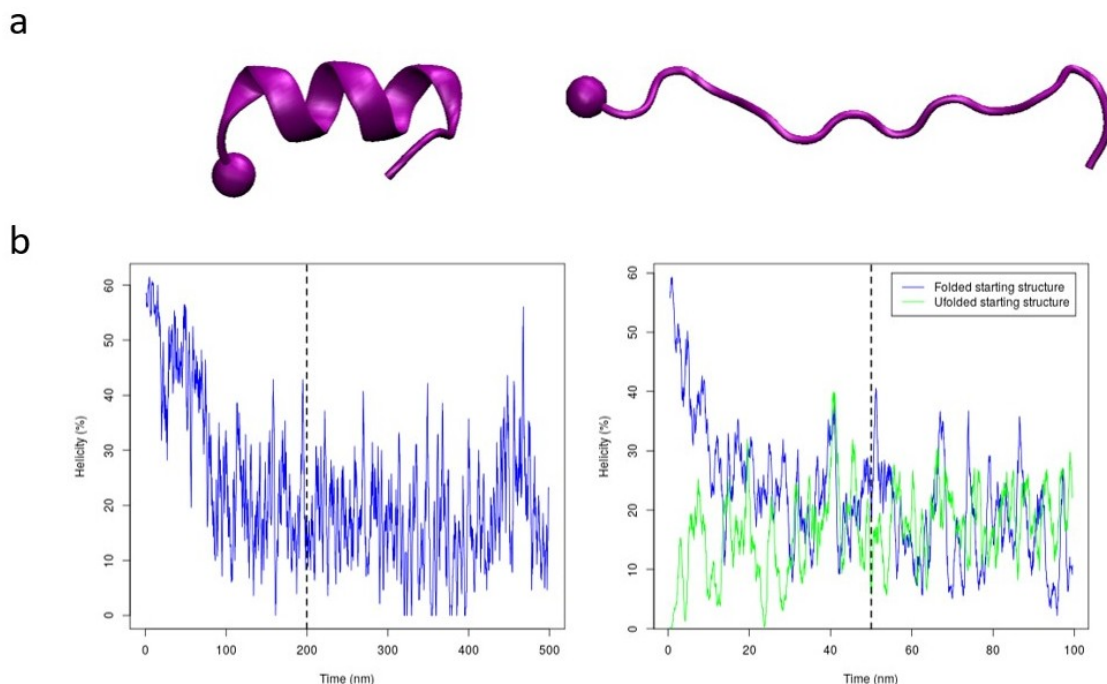


Figure 4.2 – T-REMD in water. (a) Initial structures used to launch the REMD simulations: the folded structure was employed with both FF. The unfolded one was only used with AMBER99SB-ILDN FF. (b) Mastoparan helicity content through time. The ensemble of conformations at 300 K is analyzed, no replica is followed over the temperatures. On the left, results for the CHARMM36m simulation; on the right, results for the two AMBER99SB-ILDN simulations. In the three simulations, the system takes several *ns* to get equilibrated. Only the fraction of the simulations where the system is equilibrated is used for the analysis and it is shown with the black dotted line.

4.3.2 Structural clustering

To get a deeper knowledge about the conformational ensemble of mastoparan in water, a structural clustering is carried out for the three simulations. The clusterization is done using GROMACS software and uses the RMSD value between the conformations as reaction coordinate. The RMSD only takes into account the positions of the α -carbons. Only conformations founded at 300 K are considered. For CHARMM36m simulations, the Cut-Off (CO) for the RMSD value was set at 0.5 *nm*. The conformations were divided in 5 different clusters. The most populated one contains 80.5 % of the conformations and the central structure is shown in figure 4.3. For AMBER99SB-ILDN simulations, a CO of 0.22 *nm* was used. More than 40 clusters were found for both simulations. The most populated clusters contain the same conformation in both simulations and it is also shown in figure 4.3, where it is aligned with the structure extracted from CHARMM36m simulation. This structure represents 46.3 % of the conformations in the simulation with the extended starting structure and 39 % in the simulation with the folded starting structure. Notice that the CO used for AMBER99SB-ILDN and for CHARMM36m differ, so we can not compare the number of different clusters that we obtain or the populations of the clusters. However, the conformations

that stand out for the simulations of the two FF are very similar. The RMSD value between them is 2.37 Å. The conformations are partially folded with a helix content of 57.1 % for CHARMM36M and 42.86 % for AMBER99SB-ILDN. This means that the peptide can also be found forming a partial α -helix in water.

Force Field	Helicity (%)	STDV
CHARMM36m	18.0	24.8
AMBER99SB-ILDN	19.0	7.6

Table 4.1 – Mastoparan mean helicity in water. Information extracted from "mastoparan in water" T-REMD conformations at 300 K.

Overall, CHARMM36m and AMBER99SB-ILDN results agree and are independent of the starting structure employed. Regarding the helicity content, the computational results also match experimental ones. This validates the reliability of both FF employed and confirms that they are a good choice to perform further simulations to observe mastoparan insertion in the membrane. AMBER99SB-ILDN major conformation shown in figure 4.3 is employed to launch the REMD simulation of the insertion. However, CHARMM36's major structure presents higher helicity than AMBER99SB-ILDN major structure. As seen in the previous chapter, CHARMM36 tends to slightly overestimate the helicity content. For this reason, the starting structures of CHARMM36m T-REMD of insertion are unfolded. Two conformations are employed that alternate between the replicas: the central structures of the second and third most populated clusters. They are both similar, unfolded with one turn (Fig. 4.1). This strategy is used so that the "helix overestimation" effect is counterblanced.

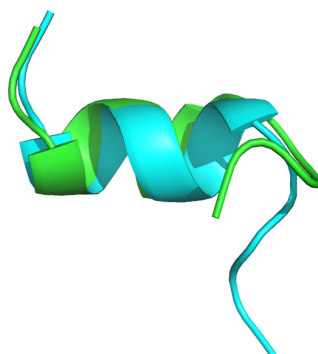


Figure 4.3 – Mastoparan in water most populated conformations found in CHARMM36m (blue) and AMBER99SB-ILDN (green) simulations.

4.4 Mastoparan partitioning into the membrane

4.4.1 Insertion in the membrane

In this section, a detailed analysis of mastoparan partition in the membrane is presented. The analysis is based on the results showed in table 4.2. Table 4.2 presents the detailed results of the partitioning process. The partitioning is studied per residue and results for both FF are compared. The first column shows the time that each residue takes to insert into the membrane for the first time. We consider that a residue is inserted when the z position of its C_α stays for more than 15 ns under the phosphorous level. For CHARMM36m FF, the minimum time is 16.51 ns and the maximal 30.89 ns . The N-terminus shows the lowest values, and it increases along the sequence. This means that the N-terminus gets inserted before the central residues, and the central residues get inserted before the C-terminus. This was expected as the N-terminus presents positive charges that are attracted towards the negative charges of the membrane, as the phosphor atoms. A similar tendency is observed for AMBER99SB-ILDN/Berger REMD. However, its time scales are shorter than CHARMM36m ones, ranging between 11.42 ns and 21.6 ns . May be, this is an artifact caused by Berger FF. As the lipid model is simplified (united-atom), the dynamics are accelerated.

The second information that is shown on table 4.2 is the number of times that each residue gets inserted. Once they partition into the membrane, the residues are able to get out back to the solvent and then get inserted again. In the simulations, if the z position of the C_α of a residue spends more than 15 ns over the phosphor level, we consider that the residue is no longer inserted. The less stable is the interaction, the more times it can get in and out of the membrane. Both simulations show that the residues close to the N-termini have less number of insertions, which means that the insertion is more stable, as expected because of its positive charges. Residues of the C-termini present higher number of insertions which means that the interaction is less stable.

To finish the analysis, table 4.2 shows the percentage of time that each residue spends inserted into the membrane. In general, the percentages are very high for both FF and for all the residues, which confirms the affinity of the peptide for the membrane. In both simulations, the N-termini presents the higher values and the residues having the lower percentage of time inserted are the two terminal ones: ILE13 and the LEU14.

Overall, the results show that the insertion is initialized by the N-termini, followed by the central residues and then by the C-termini. The N-termini has more affinity for the membrane and its insertion is more stable and last longer compared to C-termini insertion.

	CHARMM36m			AMBER99SB-ILDN		
Residue	First insertion (<i>ns</i>)	Number of insertions	Time inserted (%)	First insertion (<i>ns</i>)	Number of insertions	Time inserted (%)
ILE1	16.5	1.9	91.2	12.4	1.5	90.3
ASN2	19.8	2.1	90.0	11.4	1.2	95.2
LEU3	23.2	1.7	90.1	15.7	1.8	88.7
LYS4	23.6	2.6	89.1	13.1	1.7	94.1
ALA5	25.8	2.3	88.8	1.0	1.9	91.0
LEU6	25.5	1.8	89.7	18.9	1.4	88.2
ALA7	25.7	2.3	89.4	18.2	2.2	89.5
ALA8	23.5	2.4	88.7	19.	2.1	90.0
LEU9	28.2	2.3	88.0	21.6	2.0	87.6
ALA10	23.0	4.0	84.8	17.0	3.8	83.1
LYS11	23.6	5.3	82.4	15.0	2.3	92.1
LYS12	28.8	4.0	81.7	12.1	2.5	92.6
ILE13	30.8	3.9	77.8	20.3	2.8	80.9
LEU14	29.8	3.5	77.8	19.8	3.2	80.0

Table 4.2 – Statistics of mastoparan partitioning times in the membrane. The position of the z coordinate of each α -carbon is taken for the analysis. Results for two different FF (CHARMM36m and AMBER99SB-ILDN) are compared.

4.4.2 Partitioning/Folding coupled mechanism

Having a look at the simulations, one easily realize that the partitioning and the folding are related. Indeed, the partitioning/folding mechanism is coupled and it is observable in the T-REMD simulations. Of note, the results shown in this section are just an illustrative example of mastoparan behaviour observed in our simulations. One of the CHARMM36m replicas has been selected to study the partitioning / folding in the membrane. Figure 4.4 shows mastoparan insertion along time. The z coordinate of mastoparan center of mass (COM) is plotted with respect to the mean z position of the phosphorous atoms of the upper layer. The peptide approaches the membrane, and starts partitioning around 100 *ns*. Then, it goes out from the membrane for a moment and partitions again at 250 *ns* approximately. In the graph, the helicity content of the peptide is showed too. There is a direct relationship between the partitioning and the helicity content. This content increases when the peptide gets inserted and decreases when it is out of the membrane. At the end of the simulation, there is a 30 % of helix, which is quite low. This means that in the 50 last *ns* of simulation the peptide has not succeeded in realizing the full process of partitioning / folding into the membrane, so the replica need to be prolonged to observe the whole process.

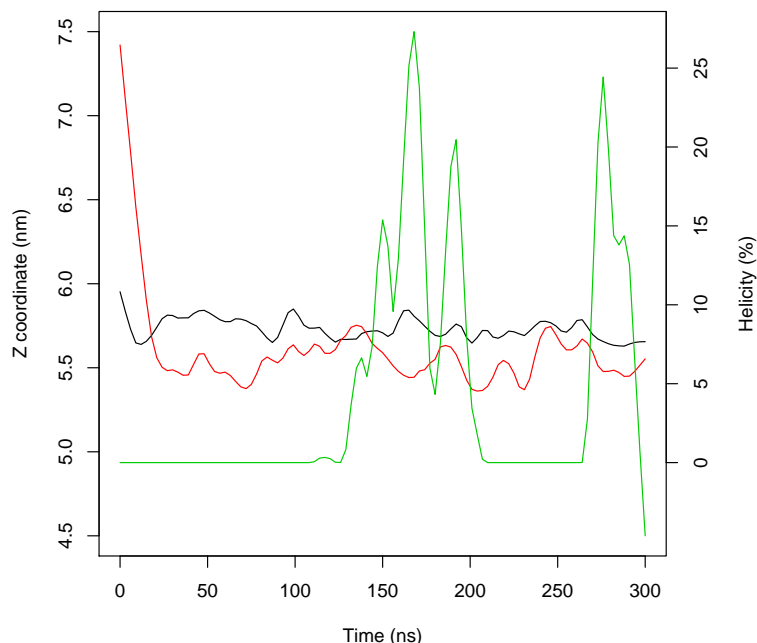


Figure 4.4 – Illustrative example of the partitioning / folding coupled mechanism. The information is extracted from one of the CHARMM36m T-REMD replicas. The plot shows mastoparan COM z coordinate through time (red), with respect to the mean z position of the phosphorous atoms of the upper layer (black). In the same graph, the evolution of the helicity through time is plotted (green).

4.4.3 Helix content

The objective now is to check if there is a relationship between the dynamic information extracted from table 4.2 and the secondary structure of the peptide. Figure 4.5 shows mastoparan helicity per residue for both FF. In both cases, the extreme residues have lower helicity values than the central residues. For both FF, the C-terminus is the less-structured part of the peptide. At the same time, the C-terminus presented the less-stable binding with the membrane as saw in the previous section. Together, these results suggest relationship between the insertion and the capacity of folding: the more stable the partition, the more stable the folding. These results also agree with the NMR-resolved structure [91], where mastoparan also presents unfolded termini. CHARMM3m FF seems to overestimate the helicity content, as discussed in the previous chapter: it predicts the folding of residues that are normally random-coil.

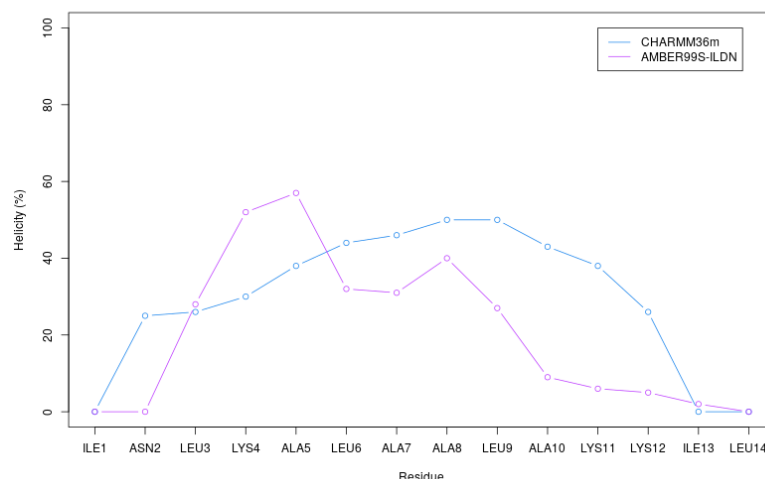


Figure 4.5 – Mastoparan’s helicity per residue in CHARMM36m and AMBER99SB-ILDN REMD insertion simulations. AMBER99SB-ILDN (pink) predicts an α -helix from residue LEU3 to residue ALA10 approximately. CHARMM36m (blue) predicts an α -helix from residue ASN2 to residue LYS12

4.4.4 Orientation in the membrane

To end up with mastoparan insertion study, partitioning analyses are carried out to better understand the orientation of the peptide in the membrane. This analyses are only done for CHARMM36m T-REMD. Figure 4.6.a shows mastoparan helical wheel predicted by Heliquist [77]. The predicted amphipathic helix is not perfect but still the two sides can be distinguished. Figure 4.6.b presents the partitioning box plot: the z positions of the residues are plotted relative to the lipid phosphorous mean positions. It is important to note that only inserted and folded conformations at 300 K are used for the construction of this plot. Overall, the partition results agree with the predicted helical wheel: the hydrophobic residues are located deeper in the membrane than the lysines which are charged and able to interact with the solvent. The only residue that do not behave as predicted is the ASN2, which is located higher than expected because it belongs to the unfolded N-terminus, not following the helical structure.

Figure 4.6.c shows the process of insertion observed in one of the T-REMD replicas as an example. Three residues are studied: one N-terminal residue, the ASN2, one central residue, the ALA8, and one C-terminal residue, the LEU14. Their z coordinate is plotted through time. During the first 50 ns, the insertion process takes place as deduced from table 4.2: the N-terminal residues initiate the insertion, followed by the central residues and then by the C-termini. In this replica, the insertion is very stable for the whole peptide. Once they are partitioned, the LEU14 is positioned much deeper than the other two residues, as illustrated in Fig.4.6.b. During this simulation, the peptide folds very quickly into a stable α -helix, as reflected by the mean helicity content (61.8% \pm 6.7). As the partitioning/folding is a couple process, the secondary structure

results are coherent with the insertion.

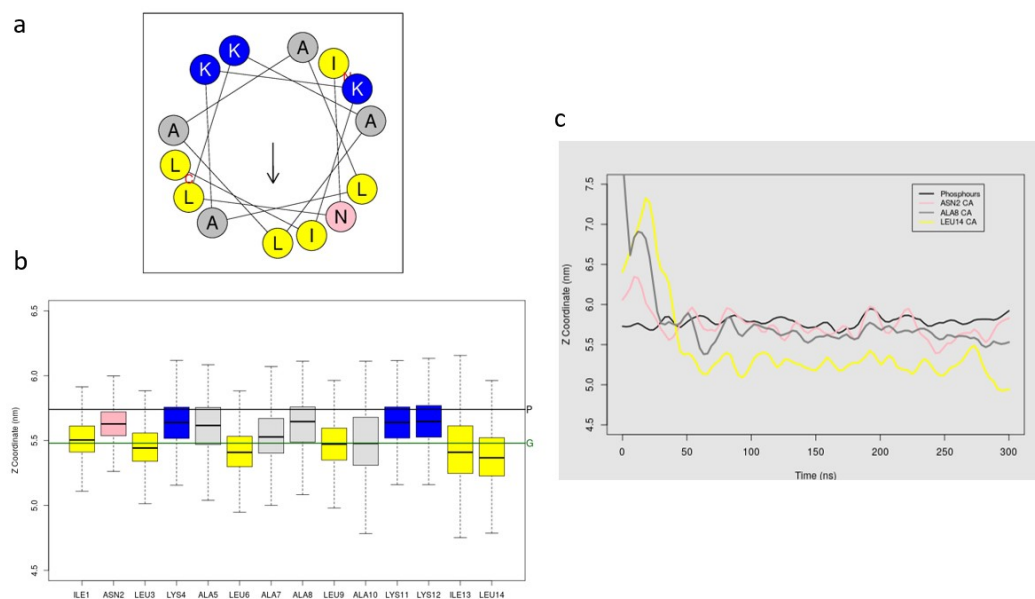


Figure 4.6 – Analysis of mastoparan partitioning in the membrane. (a) Helical wheel model of the mastoparan. (b) Box plot showing the partitioning per residue. The black and green lines mark the phosphorous and glycerol levels respectively. (c) ASN2, ALA8 and LEU14 Z position through time in one of the REMD replicas.

4.5 Discussion

In this chapter, T-REMD technique has been employed to study AH insertion in the membrane. Using the mastoparan as model system, this enhanced sampling technique allowed us to perform an in depth exploration of mastoparan insertion mechanism with atomic details. Compared to the elevated-temperature technique used by the Ulmschneider brothers [212, 44], T-REMD provides higher sampling because several replicas are launched at the same time. In addition to that, in T-REMD the temperature is exchanged between replicas. The temperature jumps act as some kinetic energy booster on the system, accelerating thereby the kinetics. Compared to REUS and metadynamics simulations, which are biased along some reaction coordinates of the system, in T-REMD simulations the system is free to explore any part of the energy landscape. Its dynamics is accelerated thanks to the high temperatures, but no reaction coordinate is biased, which is preferred to observe the partitioning/folding mechanism to assure that we explore in depth all the zones of the conformational landscape without neglecting possible orthogonal processes. Unfortunately, individual replicas of a T-REMD simulation have its temperature changed along time, accordingly the kinetics is modified as well. Thus it is generally difficult to deduce kinetic information from these simulations even though some methods have been proposed [225, 35, 48]. In the next chapter, another strategy is presented to study the kinetics of mastoparan folding.

Another important valuable information presented in this chapter is the FF comparison. To have accurate, reliable FF is essential for the evolution of MD techniques. The FF is the basis for the quality of the results. The scientific community needs to be aware which FF suits better their molecular systems of interest. Specific projects are carried out to this aim, as the NMRlipids project for lipids [34, 14], where CHARMM36m outstands for its capacity in reproducing lipid structural and dynamic properties (Please consult the introduction for further information). In this chapter, we successfully compared CHARMM36m and AMBER99SB-ILDN/Berger FF for AH insertion in lipid bilayers. In general terms, results from both FF agree. CHARMM36m seems to slightly overestimate the helix content of the α -helices compared to experimental results. However, its evolution in the past years has been amazing, it grows very fast and nowadays presents the widest ensemble of parameters among all FF. In addition to that, the developers have created CHARMM-GUI, a web server where you can easily construct your molecular system of interest and obtain the necessary files to launch MD simulations. CHARMM-GUI has simplified enormously the work of generating MD simulations and its an advantage of using CHARMM36m FF. For all these reasons, CHARMM36m is the FF that I have used the most during my Ph.D, and it will be present in most of the chapters.

The T-REMD simulations in water describe a major extended conformation. In water, the peptide is free to move as it has no restraints and explore a wide number of conformations. α -helical structures are observed for both FF, although they are less common (approximately 20%) and less stable, as we can deduce from the standard deviation values. Unpublished CD data from our collaborators, validate the helicity observed computationally. From the structural clustering carried out, we obtain a first big cluster of partially-folded helices. The main structure matches between the two FF, a two-turn helix with an unfolded C-terminus part. The agreement between the two FF supports the reliability of the results. From these first set of simulations, we can conclude that both FF accurately predict mastoparan behaviour in water.

The T-REMD simulations of the insertion showed the partitioning and folding mechanism of the mastoparan in the membrane. The peptide is mainly unfolded in water and folds into a completely-folded α -helix while partitioning in the membrane. This insertion mechanism agrees with the model proposed by S. White, who described the partitioning/folding process as a coupled mechanism [222, 223]. The insertion is lead by the N-terminus part of the peptide. Mastoparan N-terminus contains positive charges that are responsible of the attraction towards the phosphate atoms of the membrane. This attraction is lead by electrostatic interactions, a mechanism largely described in literature as, for example, in the case of the *alpha*-synucleine [10]. This phenomenon is more obvious in CHARMM36m simulations than in AMBER99SB-ILDN/Berger simulations.

Once inserted, it folds into a stable α -helix. Structural results are similar for both FF and match the NMR-resolved structure [91], as it was showed in the previous chapter. Peptide orientation in the membrane was analyzed for CHARMM36m simulations. The orientation seems reasonable and matches the relative positions predicted by the helical wheel. The hydrophobic residues partition

deeper in the membrane to establish hydrophobic interactions with the aliphatic chains. At the same time, polar residues keep the lateral chains towards the solvent to interact with water. The AH is positioned under the phosphate group, at the glycerol level, as it was described in 1999 by X-Ray diffraction [92]. During all the simulations, the N and C-termini stay unfolded. The N-terminus partitioning is stable and long-lasting thanks to the positively-charged residues. However, the C-termini is unfolded and has no charges, so its partitioning inside the membrane is weaker. It keeps changing its structure and moving in and out of the membrane in most of the replicas. Interestingly, when the C-terminus is partitioned in the membrane, it locates deeper than the rest of the sequence.

To sum up, using T-REMD simulations we successfully achieved the prediction of mastoparan insertion in the membrane. CHARMM36m and AMBER99SB-ILDN/Berger simulations reproduce the AH structural properties and the partition/folded mechanism as previously described in literature. CHARMM36m FF slightly overestimates the helicity content but is the most reliable FF for lipid simulations and contains the biggest ensemble of molecules. From now, CHARMM36m will be used as a preferred FF to study AH/membrane molecular systems.

Chapter 5

Mastoparan folding in the membrane: Markov State Models to decipher the mechanism

5.1 Introduction

A Markov State Model (MSM) is a mathematical formalism that describes the dynamics of a changing system. To build a MSM, the system needs to be markovian. This means that the probability of each state of the system depends only on the present state, not on the events that occurred before. This principle is very powerful and can be applied to molecular processes. Nowadays, it is possible to construct an MSM from MD simulations. It helps describe the structural, kinetic and thermodynamic properties of the system. The method is very powerful as it allows to study long processes whose time scales are not reachable by simple MD simulations. Two research teams have developed outstanding contributions to the theory and software for MSM construction from MD simulations. Dr. Vijay Pande research team from the University of Stanford developed a python library called MSMBuilder to provide statistical tools for MSM construction of long timescale dynamics of biomolecular systems [18]. Notable MSMs were constructed by this team, such as the study of membrane fusion [117]. The study is focused in the fusion of neural and viral highly curved membranes. The time scales and atomic details achieved were not previously reachable. They described membrane fusion as a branched pathway with stalk-like and hemifusion intermediates on a 6- to 9-*ns* time scale (Fig 5.1). The results were in good agreement with experimental ones and they discovered intermediate states too unstable to be seen by experimental assays. Unfortunately MSMBuilder is not maintained anymore and it has fallen in disuse.

Dr. Frank Noé, from the University of Berlin and his research team, have developed another program to apply MSM theory to MD analyses. The program is called PYEMMA [190] and it is

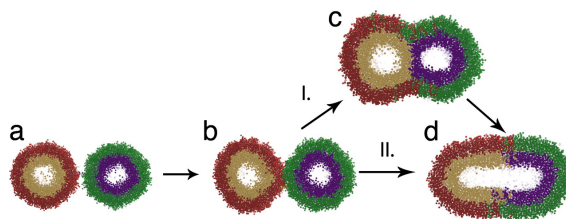


Figure 5.1 – Branching reaction pathway for vesicle fusion from an unfused starting state (a) to the fully fused state (d). Two paths are described: (I) canonical pathway through a stalk-like early intermediate (b) and a hemifused late intermediate (c) and (II) rapid fusion with just the stalk-like intermediate.

developed as a python library. Easy to use, it is currently the main software for MSM construction from MD simulations. Since its publication, it has served to study numerous molecular systems. In 2017, the first protein-protein interaction model was built [170]. Constructing an AA model of protein association and dissociation was a challenge for MD protocols because of the long time scales of the process, especially the protein-protein dissociation. Dr. Noé's team achieved the construction of an MSM of the interaction between the ribonuclease barnase and its inhibitor barstar from MD simulations. They produced a detailed model of the mechanics that describe the intermediate structures, energies and kinetics. Figure 5.2 shows the last intermediates of the model. The tightly bound model agrees with crystallographic observations. The publication of this model is not only important for the description of the barnase-barstar interaction, but it also demonstrates that by combining MD simulations with MSM building it is possible to describe molecular interactions with atomic precision at the millisecond time scales. Numerous models have been created since then to describe different types of interaction such as protein-ligand binding. The dynamics of the oncoprotein fragment Mdm2 interaction with its nanomolar inhibitor, the peptide PMI was described and for the first time direct estimates of the kinetics beyond the second time scale were reported [163]. The model describes with high accuracy and precision the mechanism of interaction by the strong binding of a variety of conformations with different hydrophobic contact surfaces that interconvert at the millisecond timescale. This project represented a double challenge: peptide exploration until reaching the active site and peptide folding within the binding pocket.

In this chapter, we present the novel application of MSM to the study of Amphipathic Helices (AH) and their interaction with membranes. Dr. S. White studied the dynamics of folding peptides in membranes [222], describing the partitioning and folding of the peptide as a coupled process (called "partitioning/folding coupling"). He calculated Melittin partitioning-folding thermodynamics by CD assays. However, to understand the insights of the process he wanted to determine the thermodynamics of the partitioning and the folding independently. To this aim, he constructed a hydrophobicity scale [223] and described the two processes separately (For more details, check section 1.4). This experimental work is very useful for comparing computational results to experimentally quantified thermodynamic and kinetic data.

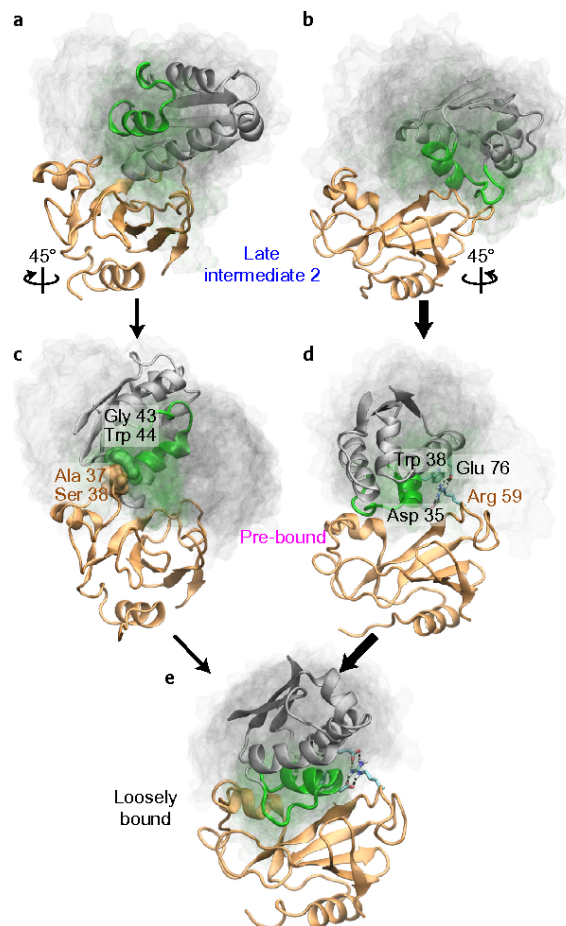


Figure 5.2 – Late steps of barnase-barstar binding to loosely bound state. Two possible pathways are presented depending on the interacting side of barstar. The major pathway (b - d - e) represents the 95 % of the flux where barnase Arg 59 ‘anchors’ into barstar via Asp 35, Glu 76 and Trp 38.

The aim of this chapter is to use MSMs to describe AH behaviour based on S. White description of the dynamics, that is the specific part dealing with the peptide folding within the membrane (described by ΔG_{AC} in figure 1.22 in section 1.4). A well-known peptide, the mastoparan from wasp venom, is used as model system. Our strategy was designed so that it may be served to study more complex systems in the future. In this strategy, REMD and MSM techniques are combined. We got inspired from Dr. Gherard Stock, who developed a novel strategy by combining metadynamics with MSM [28] (please check chapter 1.5 for further information). First, some REMD simulations are launched to explore the conformational space of the system. From the REMD conformations obtained at 300 K (bottom temperature), a set of conformations are selected and used as starting structures to launch classical MD simulations for MSM construction. This allows to enhance the sampling without biasing the final model and to obtain quantitative thermodynamic values of the process that could be compared to White’s experiences. However, the time scales of the whole partitioning/folding dynamics appeared too long to be simply solved like this. In consequence, we divided the dynamics in sub-processes. Two models are presented. The first one, the mastoparan in the membrane, shows the folding and unfolding of the already membrane inserted peptide. The

second one is the mastoparan in water, where the peptide has a greater structural freedom and gets to explore a higher number of conformations that would lead to its partitioning into the membrane.

5.2 Mastoparan in the membrane

5.2.1 Simulation details

The system

The system contains the mastoparan in the solvent, a few *nm* away from the membrane. Only one molecule of the peptide is present. The membrane is composed by 60 DOPC molecules and 10 DOG molecules: 70 lipids in total. This composition was chosen to increase the number of packing defects and mimic membrane curvature, as explained in the previous chapters (Chp. 3.1). The solvent contains 2800 water molecules and 4 chloride ions to neutralize the system. The total number of atoms in the simulation box is 18083. The system is contained in a cubic box.

REMD

The first step of the strategy is to carry out REMD simulations of the system to explore in depth its conformational space. From the 300 *K* simulation, starting structures are selected to launch classical MD simulations that will serve for MSM construction. REMD initial structure is shown in figure 5.3a. Although the peptide is in the solvent, it partitions very quickly into the membrane (less than 150 *ns*). The minimum temperature is set at 300 *K* and the maximum at 400 *K*. 33 replicas are simulated for 1 μ s. Figure 5.3b shows the potential energy distribution of the system at every temperature. Consecutive temperatures have an overlapping region. Temperature exchanges between replicas are only allowed when the structure potential energies are in the overlapping region. It ensures the two exchanged conformations are valid in terms of Boltzmann distribution at the new temperatures.

For the MSM MD simulations, only conformations where the peptide is inserted are considered. First, a structural clustering is done using GROMACS software. The RMSD between the structures is used for clustering. Only C_α are considered. This method is employed to have a first overview of the structural ensemble. Some of the cluster centers are picked as starting structures for regular MDs which will serve for the MSM construction. Other starting structures are chosen just by exploring visually the simulations. The aim is to have very different starting points to maximize the sampling. 10 different structures were chosen in total, from totally unfolded to perfect α -helices.

MSM simulations

40 MD classical simulations are used to construct the MSM, with 10 different starting points. Each simulation is launched with random initial velocities. The simulations are 500 *ns* long, 20 μ s are

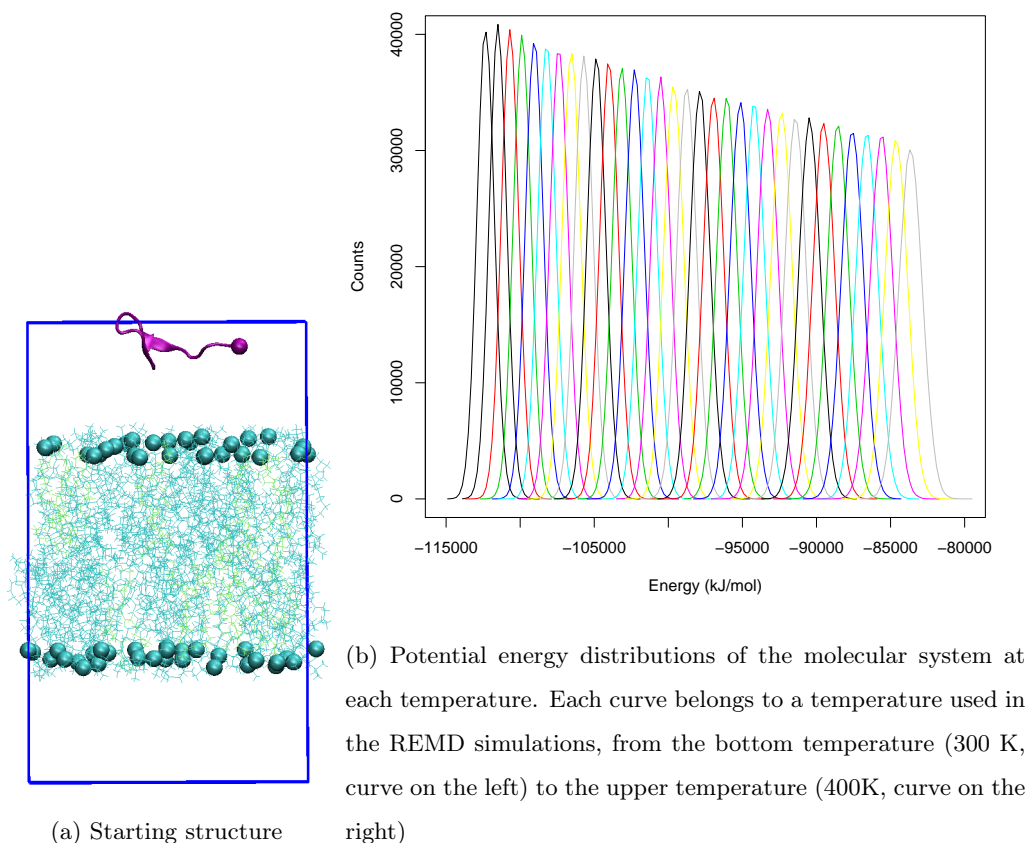


Figure 5.3 – REMD simulations of mastoparan partitioning in the membrane

simulated in total. CHARMM36m is used as FF, the temperature is set at 300 K and the pressure at 1 bar. The simulations are done in NPT conditions.

5.2.2 Model Construction

Featurization and Dimensionality reduction

The features that better describe the system are the RMSD with respect to the NMR structure (PDB code 1D7N) and the Radius of gyration (2 features are selected in total). By testing different features, we realized that these two gave the more reliable results in MSM construction for mastoparan system. Although this number is already low, TICA is still used to extract the slow coordinates with a lag time of 50 *ns*. The chose of this lag time is justified later, with the ITS validation. After TICA application two ICs were obtained so there was not really a reduction of dimensionality. However, TICA was still useful to extract slow ICs. Figure 5.4 shows the histogram of the ICs values. It seems that the values are homogeneously distributed along the conformational space. In the same figure, the first image of the conformational landscape is also shown where the conformations of the simulations are projected onto the two ICs. The color of the map is related to the free energy of the conformations. The conformational space seems well sampled, which is a

good sign regarding the convergence of the simulations. It is possible to distinguish about 3 energy basins, one of them very deep on the left.

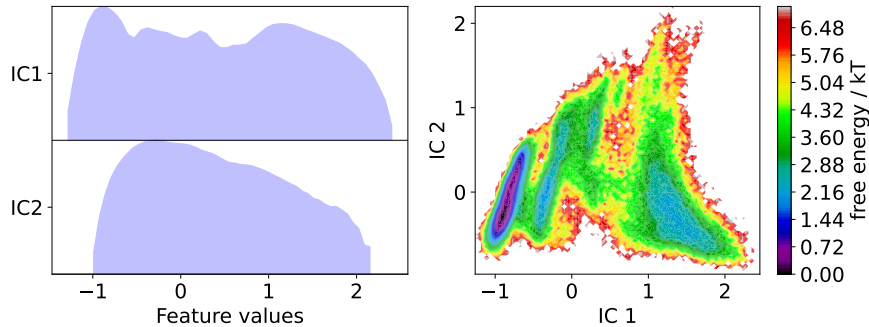


Figure 5.4 – On the left, the histograms of the two IC obtained after TICA dimensionality reduction. On the right, the conformational landscape projected on the two ICs. The color is related to the structure's free energy.

Discretization into microstates

VAMP2 score test is employed to choose the number of microstates. The score converges very quickly at a low number of clusters. For this model, 30 microstates are created using k-means algorithm. Figure 5.5 shows the conformational landscape with the center of the microstates marked in black. They are well scattered along the conformational landscape. The background color is related to the sample density. The high density regions correspond to the low energy regions that stand out in figure 5.4. From this discretization, the transition probability matrix is constructed. The number of jumps between microstates between time t and $t + \tau$ (i.e. where τ is the lag time) are counted, creating a transition matrix. The values of this first transition matrix are then transformed into probabilities.

Validation

The transition probability matrix is calculated after discretization into microstates. This matrix is employed for the model validation. First, the ITS convergence is checked from the eigenvectors of the matrix. The ITS of each process depends on the eigenvalue and the lag time, and it is calculated as follows:

$$ITS_i = -\frac{\tau}{\ln |\lambda_i(\tau)|} \quad (5.1)$$

The first eigenvectors represent the slowest dynamics of the system. The results of the test are showed in figure 5.6a. It shows the timescale values for the first 10 eigenvectors at different lag times. The gray region is the "non resolved area" because it corresponds to processes which are faster than the lag time. For lag times higher than 50 ns, the ITS stabilize and the values keep

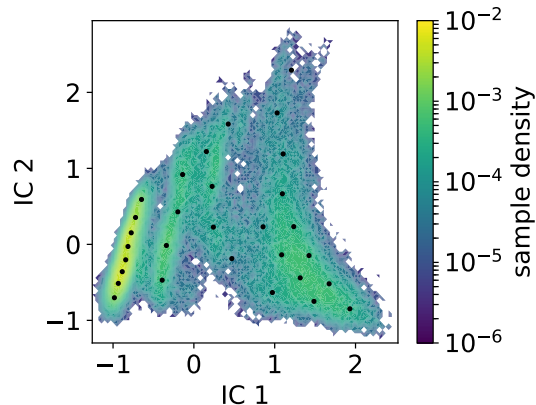


Figure 5.5 – Trajectory conformations projected on the two ICs. The darkness of the color is proportional to the density of the region. The structures were divided in 30 microstates. The centroids are marked in black.

somehow constant with increasing lag times. This means that 50 *ns* was a good choice for the lag time and that the ITS have converged.

The second part of the validation protocol is the Chapman-Kolmogorov (CK) test, that serves to validate the markovianity of the model. It compares the right and the left side of the Chapman-Kolmogorov equation:

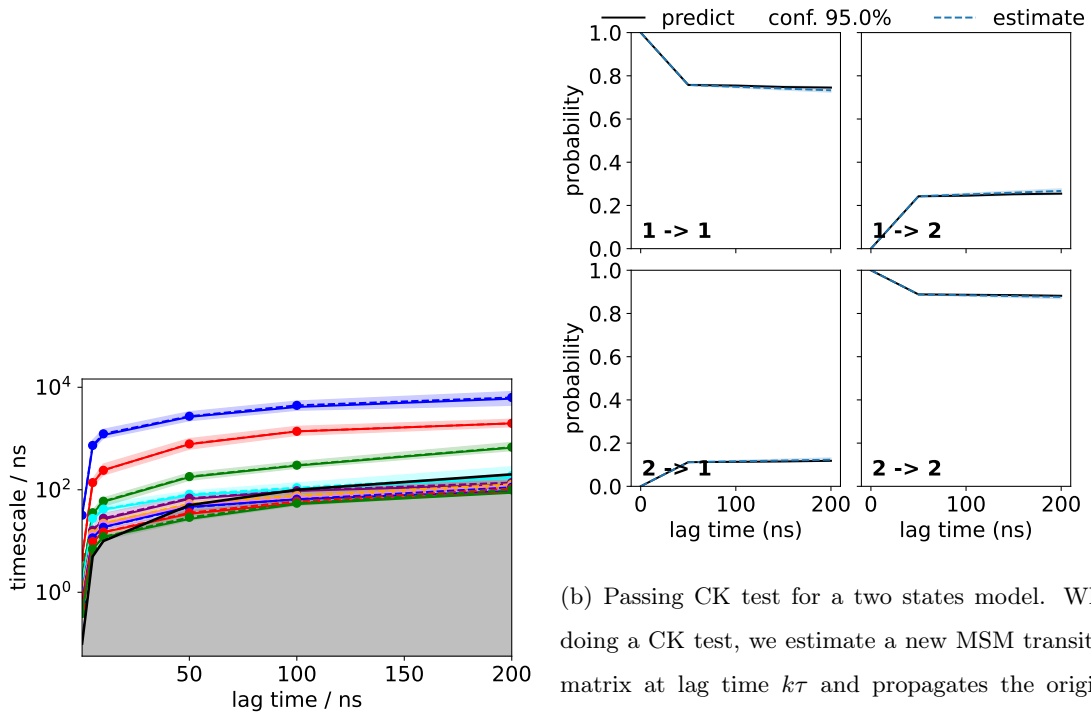
$$\mathbf{P}(k\tau) = \mathbf{P}^k(\tau) \quad (5.2)$$

The test is done for a theoretical model presenting two macrostates (macrostates will be presented in next section). As shown in figure 5.6b, the two sides of the CK equation coincide (the curves overlap). In consequence, this is a passing CK test and the dynamics of the system are proved to be markovian.

Coarse Graining into macrostates

Although conformations were clustered into microstates, the number of states is not small enough to construct a simple model. For these reason, a second clusterization is carried out, to assemble the microstates into big macrostates. The selection of the number of macrostates of the system is based on the ITS values of the transition probability matrix eigenvectors. The comparison of the ITS are shown in figure 5.7. Clearly the first eigenvector plotted (which corresponds to the second eigenvector of the matrix) stands out among the others so it is the only one employed for the coarse-graining. In consequence, the model only shows two macrostates.

Now, we search for better understanding of the process described by the first eigenvector. To do this, the conformations are projected onto the two ICs as done before, but now the color is related to the value of of the first eigenvector (Fig. 5.8). Recall, each eigenvector component corresponds to a microstate, thus the colors in Fig. 5.8 tell us to which microstate each conformation belongs



(a) ITS validation through convergence at increasing different values for k were chosen. For each trans-lag times. Each eigenvector is plotted with a different color. The grey zone is the "unresolved area", processes that are faster than the lag time. From 50 ns, together. The curves overlap, meaning that the system is markovian and the CK test positive.

Figure 5.6 – Model validation

to. In this figure it is easy to distinguish the extreme values of the eigenvector: the purple cluster on the bottom left and the green cluster on the top center. The rest of the conformations may belong to transition states between the two clusters.

Using the information contained in the first eigenvector we can proceed to the coarse-graining. First, a "fuzzy assignment" is computed, where each point of the conformational space is given a probability of belonging to macrostate 1 and macrostate 2. Results are presented in figure 5.9. They show a big central cluster that also contains the right side of the conformational space and a small cluster on the left. This division is a bit surprising because the two extremes of the first eigenvector observed in figure 5.8 have a high probability of belonging to the first macrostate.

To have concrete results, the discrete assignment is also carried out (Fig. 5.10), where each conformation is assigned to one single cluster. This division matches better the eigenvector information and it is not very different from the fuzzy assignment. It divides the conformational space in a big cluster containing the central and the right side and a small cluster containing the left side. This division is satisfactory and better matches the eigenvector information saw in figure 5.8 so it is kept for the model construction.

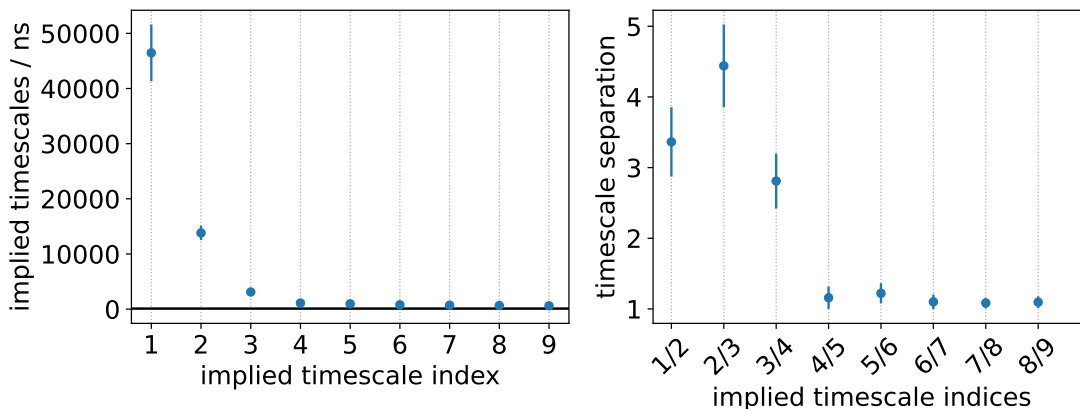


Figure 5.7 – eigenvectors comparison throughout the ITS. On the left, direct ITS comparison. On the right, ITS differences between consecutive vectors.

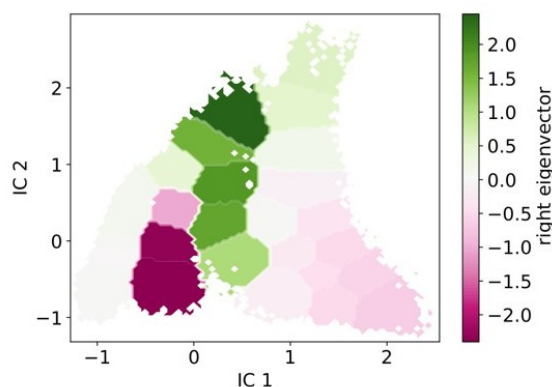


Figure 5.8 – Trajectory conformations projected on the two ICs. The color is related to the value of the conformation on the first eigenvector.

The model

So far, the model has been validated and coarse-grained, so it is time to compute the model properties. The first step is to study the structural characteristics of the two macrostates. Figure 5.11 shows the superposition of 10 representative structures of each cluster. State 1 contains unfolded structures and some partially folded while state 2 contains perfectly folded α -helices, similar to the experimentally-resolved structure. The first eigenvector (slowest process) corresponds to the folding of the peptide.

It is also possible to calculate the stationary probability of the states p_{folded} and p_{unfolded} . The folded structure is the most frequent in the membrane, 65% vs 35 % for the unfolded one. The numbers next to the arrows correspond to the MFPT between the structures, being the folding more than 2 times faster than the unfolding. Finally, converting these stationary probabilities into a free energy of folding $\Delta G_{\text{folding}} = -RT \ln(p_{\text{folded}}/p_{\text{unfolded}})$ gives a value of -0.37 kcal/mol. Although the sign is right, this very small value was somewhat of a surprise. This questions whether

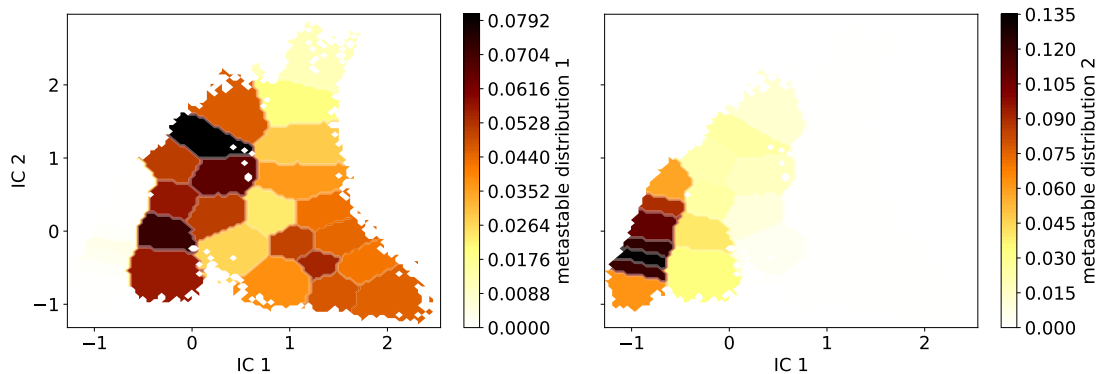


Figure 5.9 – . Fuzzy assignment. On the left, probability of the conformations of belonging to the first macrostate. On the right, probability of the conformation of belonging to the second macrostate.

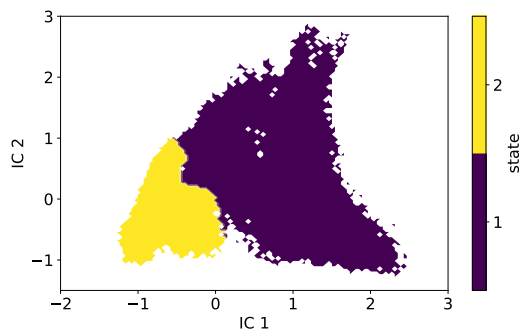


Figure 5.10 – Discrete assignment. Trajectory conformations are projected on the two ICs. The model was coarse-grained into 2 macrostates. The purple zone marks the first macrostate and the yellow zone the second one.

there could be an error in the evaluation of the stationary probabilities. We comment more on this very small $\Delta G_{\text{folding}}$ value in the section discussion below.

5.3 Mastoparan in water MSM

5.3.1 Simulation details

The system

The system contains the mastoparan in water. Only one molecule of the peptide is present. The solvent contains 2455 water molecules and 4 chloride ions to neutralize the system. The total number of atoms in the simulation box is 7608. The shape of the simulation box is a rhombic dodecahedron.

The final Model

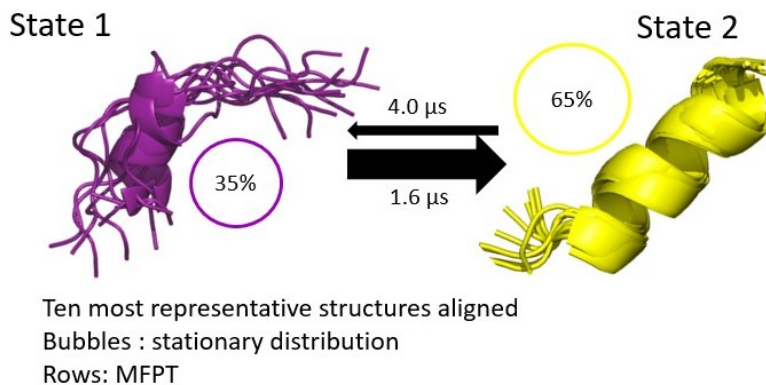


Figure 5.11 – MSM of the mastoparan in the membrane

REMD

As in the previous model, an REMD is done first to explore the system conformational space and to obtain the initial structures to launch MD classical simulations for MSM construction. REMD initial structure is the first model of the NMR structure with PDB code 1D7N. The minimum temperature is set at 300 K and the maximum at 400 K . 27 replicas are simulated for 500 ns . The potential energy distribution of the system at different temperatures also overlap as shown in figure 5.3b. Consecutive temperatures have an overlapping region and exchanges are only allowed when the structure potential energies are in the overlapping region. The conformations at 300 K are used to obtain the starting structures for MSM simulations. A clustering based on the RMSD values is used to classify the structures using GROMACS software. The RMSD value used as cut-off is 0.45 nm . 9 clusters were created and the center of each cluster was used to launch the simulations.

MSM simulations

30 classical MD simulations are used to construct the MSM, with 15 different starting points. The 9 first starting points are obtained from the REMD simulation and 18 MD simulations of 500 ns each were launched. With this simulations a first MSM was constructed, which contained 6 macrostates. One structure was selected for each of the macrostates and they were used to launch 12 new simulations of 1 μ s. The 18 first simulations were also extended to 1 μ s.

Each simulation is launched with random initial velocities. 30 μ s are simulated in total. CHARMM36m is used as FF, the temperature is set at 300 K at the pressure at 1 bar. The simulations are done in NPT conditions.

5.3.2 Model Construction

Featurization and Dimensionality reduction

The MSM of the mastoparan in water is constructed using the same features than the model in the membrane: the RMSD with respect to the experimental structure and the Radius of gyration. They also seem the best features to describe the peptide in water and it makes the model comparison easier. The lag time of this model is higher compared to the model in membrane: 300 *ns*. TICA was applied and the number of dimensions was reduced to 1, so the model is constructed with a single IC. For these reason, the 2D landscapes showed in the previous model are replaced by 1D graphs this time. The values of the IC for the first MD replica are showed in figure 5.12. Two different set of conformations are easily distinguished. The system takes long time scales to switch from one state to the other. Just by looking at the IC, we can already obtain interesting information about the system dynamics.

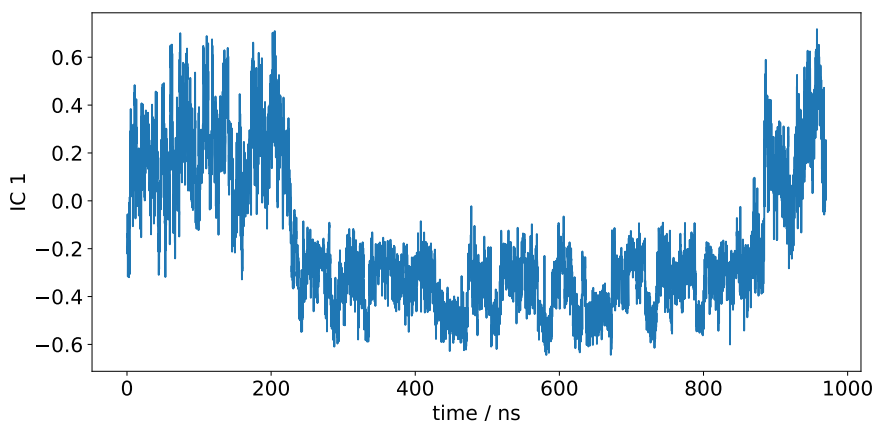


Figure 5.12 – IC values through time for the first MD simulation. Two different ensembles of conformations are easily distinguished (bimodal distribution).

Discretization into microstates

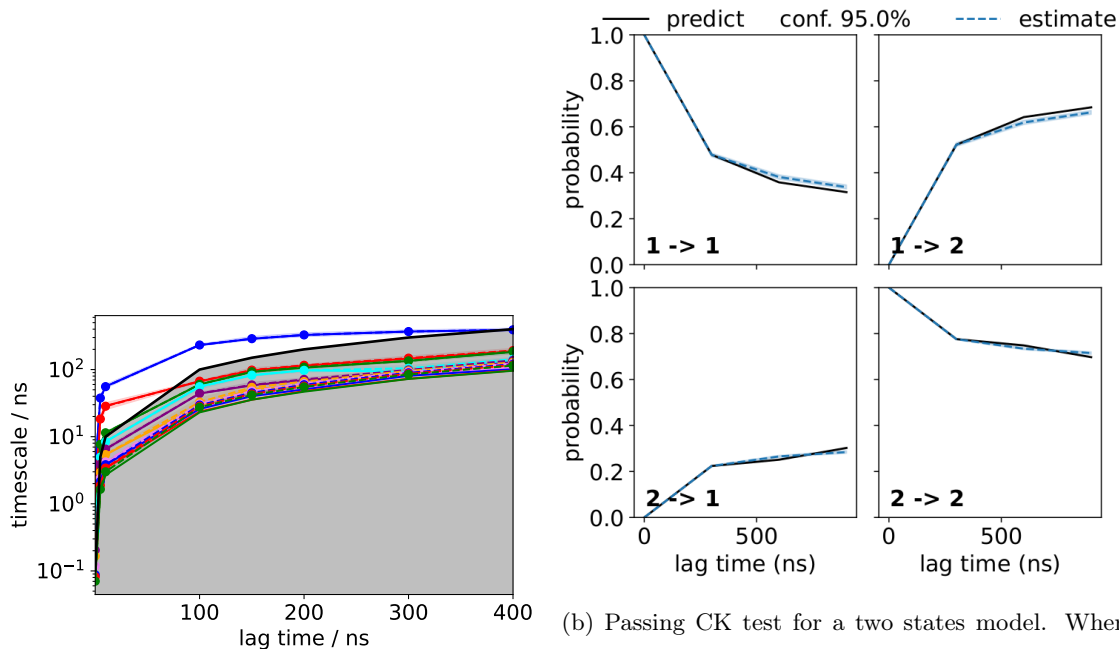
VAMP2 score test is employed to choose the number of microstates. The score converges very quickly, at a low number of clusters. For this model, 30 microstates are created using k-means algorithm. The transition probability matrix is calculated from this discretization.

Validation

The transition probability matrix needs to be validated before finishing the model construction. First, the ITS convergence is checked from the eigenvectors of the matrix. Results are shown in figure 5.13a. After multiple tests, we needed to go to quite long lag times. In these conditions, the grey area (processes which are faster than the lag time) expands a lot implying that only

one process can be resolved at the chosen lag time. This means that the model will only have 2 macrostates. Overall, the ITS seem to be converged.

The CK test is showed in figure 5.13b. The two curves (predicted and estimated) overlap so the system can be considered as markovian.



(a) ITS validation through convergence at increasing matrix at lag time $k\tau$ and propagates the original lag times. Each eigenvector is plotted with a different color. The grey zone is the "unresolved area", different values for k were chosen. For each transition between states (1->1, 1->2, 2->1 and 2->2), lag time of 300 K, only one eigenvector is out of the estimated and the predicted values are plotted grey zone, meaning that only one slow process can be together. The curves overlap, meaning that the system is markovian and the CK test positive.

Figure 5.13 – Model validation

Coarse Graining into macrostates

The system is coarse-grained into 2 macrostates by discrete assignment. Figure 5.14 shows the energy landscape of the system where the IC values are plotted against their corresponding free energy. 4 energy basins are observed and it is interesting to see the different energy barriers between them. The colors are used to show the division into the two macrostates: state 1 in yellow and state 2 in purple. Each macrostate contains 2 energy basins, which means that there is a lot of structural diversity within the macrostates.

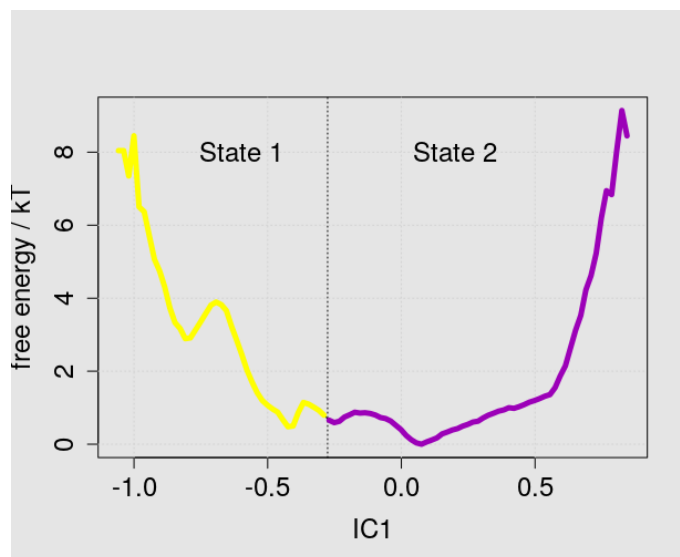


Figure 5.14 – Energy landscape. Conformations are projected on the single IC of the model. Two macrostates were created: state 1 in yellow and state 2 in purple.

The model

The final model is shown in figure 5.15. In this model, the state 1 corresponds to folded conformations and the state 2 to unfolded structures. The unfolded structures are preferred in water (71%). The presence of helices in water is minimal and the structures are only partially folded (1 or 2 turns). The MFPT show that the fastest processes is the unfolding, which represents the major flux of the dynamics. We converted the stationary probabilities into a free energy of folding $\Delta G_{\text{folding}}$ in water and obtained a value of +0.53 kcal/mol. Again, the sign is right, which means the folding is not favored in water, but the value remains very small as for the membrane case presented above. We comment on it in the section discussion below.

The final Model

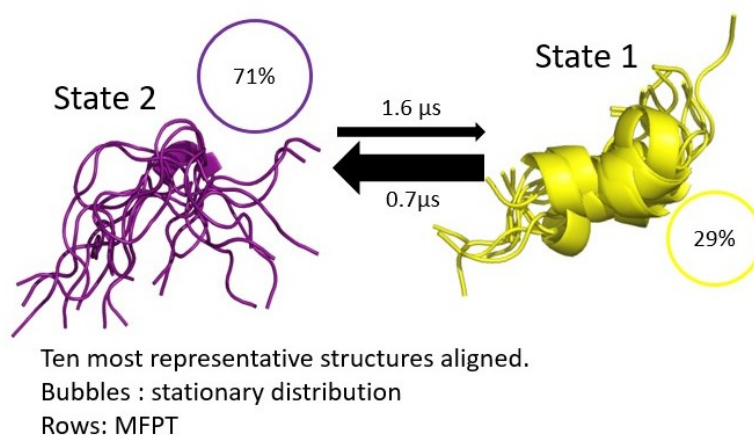


Figure 5.15 – MSM of the mastoparan in water

5.4 Constructing a model without TICA

The features used for the construction of the two models are the RMSD respect to the experimental structure and the Radius of Gyration. The total number of coordinates is two. Working with two dimensions is easy and no dimensionality reduction is needed to construct the model. To study the importance of TICA in MSM construction, a second model of the mastoparan in the membrane is constructed without it, using the two features directly as coordinates. The parameters are the exact same ones as those used in the previous model: lag time of 50 *ns*, 30 microstates, 2 macrostates, etc. This way the two models are comparable.

The values of the two features are first normalized by centering them at 0. The histogram of the values are shown in figure 5.16.a. Then, the conformations are projected into the two coordinates to visualize the energy landscape (Fig. 5.16.b) The shape of the landscape is very similar to the one obtained with the ICs (5.4) but it seems to be turned by 90°. This is because even though the dimensionality is not reduced in the model of the mastoparan in the membrane, TICA does a linear combination of the two coordinates to extract the slowest dynamics. To better understand the effect of this linear combination in the model, the construction of the model is continued hereafter. The validation protocol is applied and the model passes the two test (Fig.5.16.c and d) although the ITS seem not to be as well converged as in the previous model. Figures 5.16.e and f show the landscape clusterization in microstates and macrostates respectively. The number of microstates is 30 as in the previous model and the centers of clusters are well spread along the conformational landscape. Regarding the division into macrostates, the clusterization seems similar as the one obtained in the first model. In addition to this, the adjusted populations are the same: 35 % for state 1, the unfolded structures in the membrane; and 65 % for state 2, the folded structures. However, taking a look at the structures that belong to each of the clusters in figure 5.17, one realizes very fast that the structural clusterization of the conformations is poorly done: extended and folded structures are mixed. Regarding these results we can conclude that the exclusion of TICA algorithm of the model construction has decreased the quality of the model.

5.5 Mastoparan free energy landscapes comparison

To conclude with this research, we aimed to carry out a comparison between the model in water and the model in the membrane. To do this, the data of the two sets of simulations (the trajectories in water and in the membrane) are combined in a single set and TICA is applied. The algorithm takes into account the information of the two systems to perform the linear combination of the features. The lag time is set at 50 *ns* and two ICs are obtained. Then, the conformations of the trajectories are projected onto the two ICs to design the conformational landscape and the free energies are calculated (Fig. 5.18). The shapes of the two landscapes are very similar but there are three interesting observations:

- First, the mastoparan explores a wider conformational landscape in water than in the membrane. Peptide conformational changes in cell membranes are more complicated because of the presence of three chemically distinct environments, where folding occurs, namely, the aqueous phase, the chemically heterogeneous region of the lipid headgroups, and the hydrophobic environment of the hydrocarbon lipid tails. The membrane interfaces and the lipid tails represent spatially and chemically non-trivial fluid environments that react to the presence of a peptide. Partitioning between these environments is typically dependent and intricately linked to the conformation of the peptide. In consequence, LTX-315 in water presents a wider conformational ensemble than in the membrane.
- Second, the major energy basin of the model in the membrane is not present in water. This basin corresponds to the folded structures. As it was already shown, they are the most frequent in the membrane but they are barely present in water.
- Third and last, the energy basins corresponding to the unfolded structures are deeper in water where the folding is not favored. There is even an energy basin in water that is not even explored in the membrane.

5.6 Discussion

The major conformation of the model in membrane is a perfectly folded α -helix similar to the NMR structure 1D7N. This validates the structural aspect of this 2 states model. However, it would be interesting to construct a 3 states model of the system, where we could observe intermediate structures and better understand the folding pathway. This has not been possible yet using MSM because the 3 states model did not pass the CK test. Increasing the sampling could help the construction of this new model.

Regarding the MSM in water, the structural information matches CD assays done by our collaborators (see figure 3.4). This results showed a presence of $\sim 20\%$ of helix in water. Even though the structural information is experimentally validated, the lag time employed for the model construction is very high compared to the values that are normally seen in literature [170]. An increase of the sampling could allow to use smaller lag time values.

Comparing the two models, the dynamics of the peptide in water is much faster than in the membrane. This is normal because when the peptide is inserted in the membrane its movements are restricted by the contact with lipids and its dynamics are slowed down. The free energy barrier in the membrane for the peptide folding seems to be very high, whereas in water all the areas of the conformational landscape are easily accessible with small free energy barriers, the free energy landscape is smoother. In consequence, the peptide has greater freedom of movement in water, it gets to explore a wider conformational ensemble, and the kinetics is accelerated.

Apart of comparing the structural aspects of the model with experimental results, we need to

compare and validate the predicted thermodynamic values to experiments. The calculated free energy of folding we found in the membrane is amazingly small (-0.37 kcal/mol). If we compare this value to the model of Almeida and White [3], these authors determined a value for the free energy of folding per residue (ΔG_{res}) of -0.37 kcal/mol (check the introduction section 1.4 for a detailed explanation). Thus we see that our value of -0.37 makes no sense as it would correspond to the contribution of one residue only. According to the model of Almeida and White, taking 70% of helicity for the Mastoparan according to our CD experiments and ΔG_{res}) of -0.37 kcal/mol, the free energy of folding of Mastoparan in the membrane would be ~ -3.6 kcal/mol. It is possible that using our computational models the free energy of folding may be underestimated (e.g. -2 kcal/mol), but not this much. The value we find in water is too low as well (+0.53 kcal/mol). By turning things over in our head, we probably think we made an error in the extraction of the stationary probabilities p_{folded} and p_{unfolded} . Unfortunately we did not succeed in finding the culprit before finishing this manuscript, but we are actively searching.

It is also interesting to compare the kinetics obtained with our MSM to experiments and other computational works. Experimentally, a first work used stopped-flow fluorescence to study the kinetics of insertion and folding of mastoparan-X [206]. They could fit their data to a double exponential and obtain rates on the order of $\sim 160\text{s}^{-1}$ and $\sim 10\text{s}^{-1}$. However they attribute the first rate to the binding of the peptide on the membrane, they propose then that the folding is very fast (a few hundreds of ns) and then the second rate would correspond to some more buried insertion of the folded peptide. It is thus difficult to compare our results to theirs. However it is interesting to see that they propose that peptide binding to the membrane occurs first followed by its folding within the membrane. A second interesting study used T-jump experiments on mastoparan-X binding to DPPC vesicles [193]. T-jump is interesting as it can go down to faster timescales and does not have the issue of "dead-time" mixing of stopped-flow. In this study, the authors took advantage of the transition temperature of DPPC from gel to fluid ($T_m = 37^\circ\text{C}$). The authors fit their data with a double exponential with two lifetimes. A fast lifetime around $\sim 10\mu\text{s}$ which is attributed to the binding to the lipids (in the gel phase) and folding but with minimal membrane insertion, and a slower lifetime of a few hundreds of μs which would correspond to the insertion of the whole helix more deeply into the membrane (in the fluid phase). Although interesting, this is not easy to compare our data to theirs since their experimental setup probed different phenomena than those we simulated (notably a lipid phase change).

Regarding computational works, Chen *et al.* performed some high temperature simulations of melittin in POPC. Using a temperature of 120°C , they could observe the full insertion and folding of the peptide. They did not calculate kinetic rates as they did not observe sufficient events, but the order of magnitude of the full folding was $\sim 1.3\mu\text{s}$ at 120°C . Melittin is 26 residues long thus it takes longer to fold within the membrane than mastoparan. However, using our value of $1.6\mu\text{s}$ for mastoparan to fold, it seems we predict faster folding kinetics.

At the computational point of view, another important conclusion from this research is the

importance of TICA algorithm in MSM construction. TICA is a dimensionality reduction algorithm that reduces the number of features' coordinates to simplify the model construction. But most importantly, it does a linear combination of the coordinates to extract the slow dynamics of the system. Even if the number of coordinates of the system is small, TICA must be applied when looking for long time scale processes. In this chapter, it has been proved that TICA has a key role in MSM construction. For the model of the mastoparan in water, the dimensionality is not reduced after TICA is applied but the two obtained ICs are a linear combination of the RMSD and the Rgyr more appropriate to do a structural clustering of the system. This way, TICA helps to improve the quality of the structural clustering, and, in consequence, the quality of the model.

In this project, Patrick and I have designed an effective strategy to study AH dynamics. It is the first time a MSM has been constructed for this type of molecular systems. First, the REMD simulations served to explore in depth the conformational landscape of the system. Some structures were selected to launch classical MD simulations at 300 K to construct an MSM. This way, we constructed a non-biased dynamic model of the system from which we can obtain thermodynamical information at physiological temperature. This is an evolution compared to the elevated-temperature MD strategy [211], where thermodynamic values at physiological temperature needed to be inferred from values at high temperature. The inferred values are less reliable and the physics of the lipids and the solvent are unnatural, so we consider that combining REMD with MSM construction is a better strategy to study the kinetics of molecular systems. Moreover, it can be applied to non-thermostable systems unlike the high temperature strategy.

In the future the aim will be to construct a model of the whole partitioning process, where we get to see the peptide approaching the membrane, getting inserted and folded. The computational cost of doing this with classical MD simulation is enormous. Especially, the peptide desorption from the membrane is a so slow process that it is highly unlikely to be observed within a classical MD trajectory at room temperature. One possible solution would be to use an adaptive strategy where a first set of simulations is launched, then we select the one that has progressed the most to launch another set of simulations. Using a strategy like that Noé and collaborators have been able to reach processes on the millisecond time scale such as protein/protein dissociation [170]. However, we believe that a more efficient and promising strategy would be to use a technique called transition-based reweighting analysis method (TRAM) developed by Noé's group [226]. TRAM is a statistically optimal approach that combines large ensembles of short MD classical simulations with enhanced sampling techniques. As the enhanced sampling techniques are biased, the unbiased thermodynamics are calculated with reweighting estimators. This technique allows to construct an MSM from a combination of classical MD simulations and REMD simulations. This way, the sampling could be enhanced with a reduced computational cost.

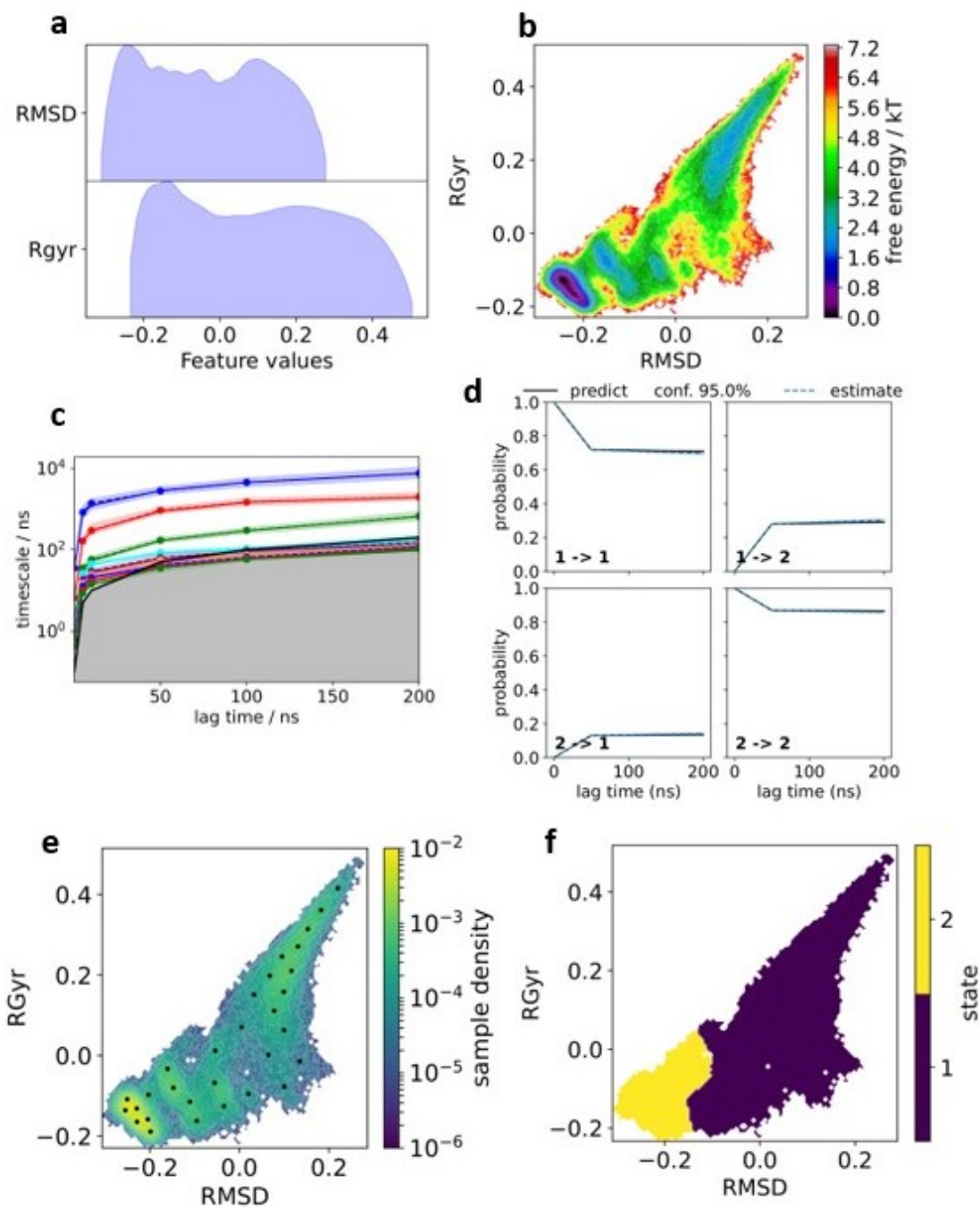


Figure 5.16 – MSM of the mastoparan in membrane using the RMSD and the Rgyr instead of TICA ICs. (a) Histograms of features. (b) Energy landscape. (c) ITS validation. (d) CK test. (e) Conformational landscape with the center of microclusters plotted as black dots. The color is related to the density. (f) Conformational landscape divided in two macrostates, purple and yellow.

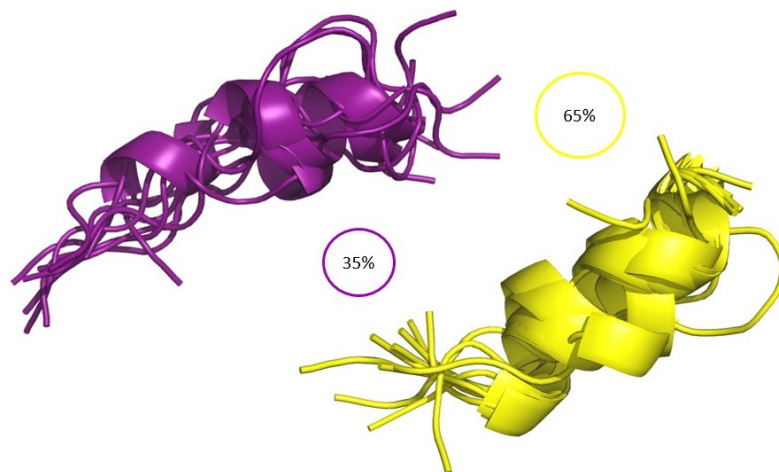


Figure 5.17 – Ten representative structures of set 1 (purple) and set 2 (yellow) obtained from the model constructed without applying TICA.

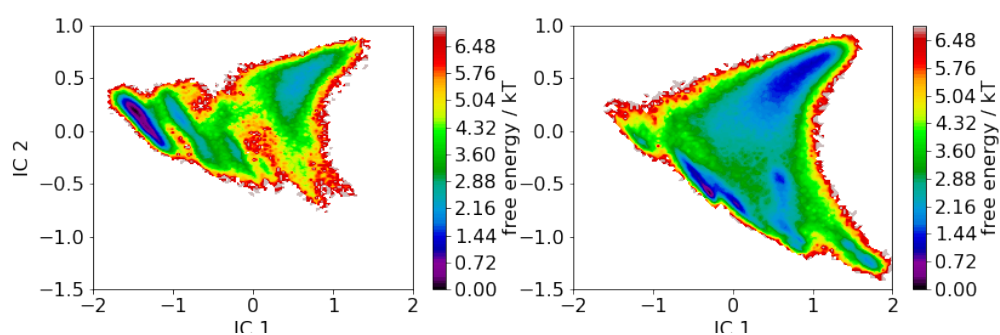


Figure 5.18 – Landscape comparison. Mastoparan landscape in the membrane (left) is compared to the landscape in water (right). The ICs are shared between the two systems. The conformations of the trajectories are plotted onto the ICs.

Chapter 6

Multidisciplinary study on LTX-315 interaction with membrane

6.1 Introduction

Cancer is the second leading cause of death globally. It caused nearly 10 million deaths in 2020 [71]. According to the World Health Organization, about 1 in 6 deaths is due to cancer. For this reason, scientific research is devoted to study cancer, create new and more advanced therapies and design novel drugs as treatment. In 2018, the Nobel prize of medicine was attributed to James P. Allison and Tasuku Honjo for their research in new immunotherapies against cancer.

The Norwegian company LytixBiopharma is developing proprietary oncolytic peptides as a novel immune therapy to fight cancer. Their lead compound is a 9-residues oncolytic peptide called LTX-315 (KKWWKKW-Dip-K). In its sequence, it presents a non-standard aminoacid, the Di-Phenylalanine (Dip), which is similar to the Phenylalanine but with two aromatic rings (Fig. 6.1). LTX-315 represents a very promising strategy to treat cancer and is actually in phase II of clinical assays in Europe and US [<https://www.lytixbiopharma.com/>].

LTX-315 presents pro-inflammatory and pro-immunogenic properties. It is able to cross the plasma membrane to get into the cytosol. Once there, it is enriched in mitochondria, where it produces the Mitochondrial Outer Membrane permeabilization (MOMP) (Fig. 6.2) [204]. LTX-315 disrupts the mitochondrial network and dissipates inner transmembrane potential. The mitochondrial intermembrane proteins are then released into the cytosol and necrosis is induced. There are evidences of BAX and BAK implication in this mechanism [238]. Unregulated necrosis favors the regression of solid tumors and the release of Danger Associated Molecules (DAMPs) [201]. This mean that a systemic tumor specific immune response is produced and that it is possible to create specific immune memory for cancer cells.

Despite the advances in LTX-315 development and clinical research, we lack of knowledge

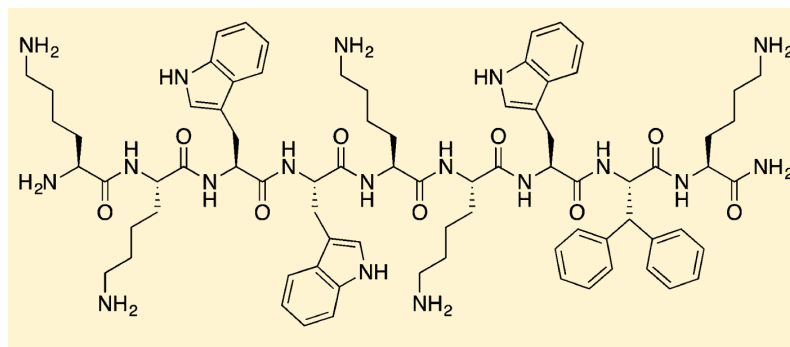


Figure 6.1 – Amino acid sequence of the oncolytic peptide LTX-315. The sequence is mostly composed by Lysines and Tryptophans. Notice the presence of a non-standard amino acid, the Di-Phenylalanine in position 8. Image from [201]

about its action mechanism from a molecular point of view. Regarding its structure, Haug *et al.* predicted it as an amphipathic helix (AH) by Haug (Fig. 6.3) [86]. As it is capable of getting internalized into the cell, it may be part of the Cell Penetrating Peptides (CPP) family [124]. However, there is no detailed information about LTX-315 interaction with membranes, like how it binds the plasma membrane or how it specifically targets the mitochondrial membrane. We have dedicated a part of this research work to study its action mechanism with a combination of experimental and computational techniques in collaboration with LytixBiopharma and its CEO Øystein Rekdal.

Øystein Rekdal could send us some mg of pure LTX-315 at the beginning of my thesis. With the help of our colleague Nicolas Rodriguez, I could carry out myself some fluorescence affinity assays to study LTX-315 interaction with Large Unilamellar Vesicles (LUVs). The aim was to study the impact of charged lipids in LTX-315 affinity for membranes. Then, I performed circular dichroism (CD) assays as a first approach LTX-315 structural properties in the solvent and in presence of LUVs. CD assays can shed light on LTX-315 AH behaviour. T-REMD simulations were also performed to observe the peptide in water and in the presence of neutral and charged lipid bilayers. MD simulations are a key to the understanding of the peptide action mechanism and technique can shed light on LTX-315 structural and dynamic characteristics with atomic detail. The diP parametrization was a challenge. It was done in three different ways. First, I used predictive tools to get approximate values and I launched in-void test simulations of the system. Then, our collaborators Dr. Rodolphe Vuilleumier and Dr. Ari Seitsonen from the ENS (département de Chimie) carried out some Quantum Mechanics (QM) calculations to help us parametrize the di-Phe. Unfortunately, we did not have time to assemble and present these results within this manuscript. However, at the same time the CHARMM FF developers published some new official CHARMM36 parameters on a wealth of non-standard amino acids, including di-Phe [54]. The last one was used to launch our T-REMD simulations.

We had the chance establish two collaborations with colleagues of our laboratory that have provided valuable information to this project. Our colleague Olivier Lequin and his Ph.D student

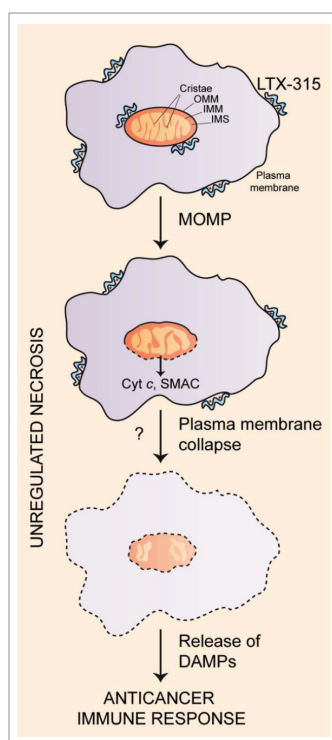


Figure 6.2 – Mitochondrial Outer Membrane permeabilization (MOMP). LTX-315 penetrates into the cell and targets the mitochondrion. The disruption of the mitochondrial outer membrane induces necrosis. Image from [201]

Edward Chalouhi carried out NMR experiments to study LTX-315 structural properties in buffer and in the presence of lipids. They aimed to describe the peptide secondary structure at residue level and give specific information about DiP behaviour and LTX-315 orientation in the membrane. Further in time, our collaborator Øystein Rekdal (CEO of Lytix) could send us some LTX-315 labeled with a fluorescent probe (Pacific blue). Our colleagues Françoise Illien and Sandrine Sagan could then carry out internalization assays, where she studied the LTX-315-Pacific blue penetration capacity and preferred pathways. By combining all these *in vitro* and *in silico* techniques, we aimed to construct a model of LTX-315 interaction with membranes to start deciphering its action mechanism. This information is crucial to understand its biomedical properties, and in the future it can serve to design new drugs of the same family. Also, this is the first study on Dip behaviour at a molecular and atomic level, giving light to the understanding of its intramolecular interactions and its contribution to the general structure of the peptide.

6.2 Materials and Methods

6.2.1 Liposome preparation

Large Unilamellar Vesicles (LUVs) are used in this project to perform fluorescence and Circular Dichroism (CD) assays. They are prepared as follows. An appropriate volume of the lipid of

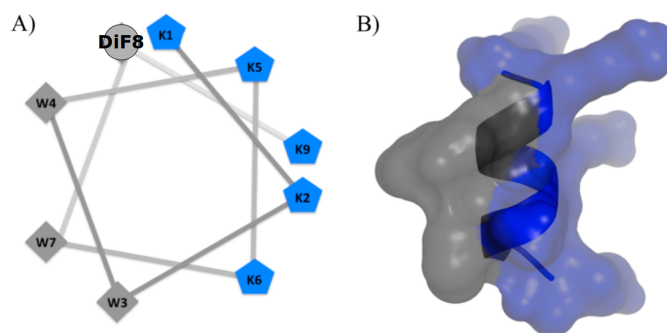


Figure 6.3 – LTX-315 putative structure as an amphipathic α -helix. (A) Helical wheel scheme. (B) 3D representation of the secondary structure. Image adapted from [86]

interest is dissolved in chloroform are dried in glass tubes under a N_2 flow. Final traces of solvent are removed in a vacuum chamber during 30 min. The obtained films are then hydrated using an appropriate amount of buffer (PBS buffer, $NaCl=140$ mM, $pH=7.4$) and vortexed extensively. The amount of buffer added depends on the final concentration wanted. In this case, we searched for a lipid concentration of 0.5 mM. The Multilamellar Vesicles (MLVs) thus obtained are then submitted to ten freeze/thaw cycles and the homogeneous lipid suspension is passed 15 times through a mini extruder equipped with a polycarbonate membrane perforated by multiple $200\ \mu m$ width pores. LUVs of approximately $200\ \mu m$ of diameter are obtained (the real size distribution was not checked though). Finally, the concentration of the samples is controlled with the Rouser technique detailed below.

6.2.2 Lipid quantification: the Rouser Technique

The Rouser technique allows to calculate the lipid concentration of a sample by measuring the concentration of the phosphorous atoms present in the lipids polar heads. This experience is performed after the generation of the LUVs. It is essential to know the exact LUV concentration of the samples to be able to measure the fluorescence and the CD at specific lipid:peptide ratios

The first step is to prepare samples of the LUVs by triplicate. The buffer used to generate the LUVs also contains phosphate ions. In consequence, samples of the buffer should also be measured to subtract the buffer contribution from that of the LUVs. Furthermore, twelve samples of a 0.5 mM phosphate solution are prepared to produce a calibration curve. Each sample contains a different volume of the stock phosphate solution, producing the different concentrations that will generate the calibration curve.

Once the samples are prepared, they are exposed to high temperatures to evaporate the solvent. Then, they are treated with $300\ \mu L$ of perchloric acid to break the covalent bonds that bind the phosphorous atom to the rest of the lipid structure, liberating the phosphorous ions. A sand bath is used to accelerate this reaction. Then, $1\ mL$ of water, $0.4\ mL$ of ammonium molybdate and $0.4\ mL$ of ascorbic acid is added to the samples in this order. The ammonium molybdate

forms a complex with the phosphorous ions called phosphomolybdic acid. This complex interacts then with the ascorbic acid generating a reduced complex of phosphomolybdate which presents a blue color. The intensity of the color (optical density) is directly proportional to the phosphorous concentration and its absorbance can be measured at 797 nm.

Once the absorbance of the calibration samples is measured, the calibration curve can be traced. Knowing the absorbance of the LUV and buffer samples its concentrations can be inferred using the calibration curve. Then, the resulting concentration of the buffer is subtracted from the LUVs concentration to get the real lipid concentration of the samples. This quantification technique allows to calculate the exact lipid concentration of the sample so the ratio peptide:lipid can be precisely controlled.

An UNICAM UV300 spectrometer has been used for this assay.

6.2.3 Peptide quantification method

The NanoDrop 2000C spectrophotometer from Thermo Scientific is a device that allows to precisely quantify the peptide concentration of a sample. It works for amino acids and nucleic acids, and it allow to quantify the concentration of the sample from really small volumes of just a few microliters (from 0,5 to 2,0 μL). This is achieved by placing the sample directly in the top of the detection surface, using the surface tension to create a column between the ends of the optical fibers, which forms the measurement optical path. The spectral range of the device goes from 190 to 840 nm and it is possible to scan all the wave lengths. Its speed, simplicity and robust performance make this device adequate to measure the concentrations of the peptides used in this manuscript. LTX-315 contains 3 tryptophan residues in its sequence and these latter have a very specific absorption at 280 nm. The intensity of the absorbance at 280 nm corresponded to a peptide concentration of 41 μM .

6.2.4 Tryptophan-Fluorescence

Emission fluorescence spectra of LTX315 in phosphate buffer (150 mM of NaCl, pH=7.4) were measured at room temperature from 300 to 450 nm (excitation at 280 nm) in a 200- μL quartz cell with or without LUVs of POPC, POPS or POPC/POPS mixture 50:50 at various concentrations (from 2 to 1000 μM). The different concentrations and ratios are shown in table 6.1 . We started to record the first spectrum of the peptide at 20 μM in pure buffer. For the next spectra, we added progressively some lipids (from some stock lipid solutions) to obtain the wanted (increasing) lipid concentrations. Conversely, the peptide concentration decreases (from 20 to 15 mM) when the lipid concentration increases (from 0 to 1000 μM). Since our goal was to get a first overview of LTX315 interaction with lipids, this way of proceeding had the advantage of consuming less peptide. But it is important to have this in mind when interpreting the spectra: we shall reason on wavelength variations rather than on intensity variations. Last, we did not subtract the spectrum

containing the solution without peptide since the intensity was systematically very low (below 0.01 a.u.).

Samples	[Lipid] (μM)	[Peptide] (μM)	Lipid:Peptide ratio
PBS	0	0	0:0
0	0	20	0:1
1	200	20	10:1
2	400	20	20:1
3	1000	20	50:1
4	2000	20	100:1
5	3000	20	150:1
6	4000	20	200:1

Table 6.1 – Ensemble of samples measured by fluorescence. The table contains the lipid and peptide concentrations used and the lipid:peptide ratios. These 6 different samples were prepared for each of the LUVs compositions (POPC; POPS; 50/50 POPC/POPS). The samples corresponding to the buffer and the peptide in buffer were measured once.

6.2.5 Circular Dichroism

CD experiments presented in this manuscript were done on a Jasco J-815 spectrometer at room temperature with a quartz cell of 0.1 *cm* path length. Each spectrum is the average of 5 scans recorded from 200 to 250 nm with a bandwidth of 1 nm, a step size of 0.5 nm and a scan speed of 50 nm min⁻¹. Spectra of LTX315 at 50 μM were recorded with and without LUVs of POPC, POPS and POPC/POPS (50:50) at 2500 μM (Tab. 6.2). The ratio peptide to lipid ratio was thus 1:50. Control spectra of buffer with or without LUVs were systematically subtracted from the peptide+lipids spectra. The final spectra are shown in units of molar ellipticity per residue $[\theta]_M$ (deg cm²dmol⁻¹res⁻¹) as a function of the wavelength λ (nm).

	[Lipid] (μM)	[Peptide] (μM)
Blank	0	0
Peptide	0	50
POPC	2500	0
Peptide + POPC	2500	50
POPS	2500	0
Peptide + POPS	2500	50
Mixture POPC/POPS (50:50)	2500	0
Peptide + Mixture POPC/POPS (50:50)	2500	50

Table 6.2 – Summary of the samples analyzed by Circular Dichroism.

6.2.6 NMR

LTX-315 was studied in three different conditions by NMR assays: buffer, micelles and bicelles. Samples in buffer contained 1 mM of LTX-315 and were of a volume of 300 μ M. We used 50 mM of sodium succinate buffer at pH 5.6 dissolved in H₂O/D₂O (90:10, v/v).

Experiments in micelles were carried out in 50 mM sodium succinate buffer at pH 5.6. with 300 μ L of H₂O/D₂O (90:10, v/v) in Shigemi tubes also containing 1 mM LTX-315 peptide and in the presence of 80 mM deuterated dodecyl phosphocholine DPC-d38 (Avanti Polar Lipids).

Bicelles samples were prepared in a volume of 300 μ L in Shigemi tubes and contained 1 mM LTX-315 peptide, 75 mM dihexanoyl phosphatidylcholine (DHPC) and 25 mM DMPC and/or DMPS in 50 mM sodium phosphate, pH 6.05, 100 mM sodium chloride, 10% D₂O, 0.02% (w/v) NaN₃.

NMR experiments were recorded on a Bruker Avance III 500 MHz spectrometer equipped with a 5 mM 1H/ 13C / 15N / 2H TCI cryoprobe. 1H resonances were assigned using 2D TOCSY (DIPSI-2 isotropic sequence of 40 and 60 ms) and 2D NOESY spectra (75 and 150 ms mixing times) recorded at 35°C. Because of the use of non-deuterated lipids, experiments in bicelles were acquired using a double pulsed field gradient spin echo with band-selective pulses (90° read pulse of 4 ms duration and G4 shape and 180 °REBURP pulses of 3 ms duration) centered on the amide / aromatic region. For the paramagnetic experiments, 1-palmitoyl-2-stearoyl-(5-doxyl)-*sn*-glycero-3-phosphocholine (5-doxylPC, Avanti Polar Lipids) was added at a final concentration of 0.25 mM.

6.2.7 Quantification assays by fluorimetry

LTX-315-Pacific blue internalization in Chinese hamster ovary cells (CHO) was quantified by fluorimetry. A fluorescent probe called Pacific Blue was attached to LTX-315 sequence to as the peptide is not fluorescent itself. The structure is shown in figure 6.4. Assays were carried out by our collaborators Françoise Illien and Sandrine Sagan. The protocol she used is detailed bellow and a descriptive scheme is shown in figure 6.5.

Cytotoxicity assay (CCK8 assay) Cytotoxicity was determined with the Cell Counting Kit 8 (CCK-8) from Dojindo Molecular Technologies (Sigma aldrich). This colorimetric assay allows to measure the viability of cells. 96-well plates were inoculated with 100 μ L/*well* of a suspension of CHO cells (2.100 *cells/well*). After 24 hours of incubation (37°C, 5 % CO₂) different concentration of peptide (0.1, 0.25, 0.5; 1, 2.5; 5; 10 and 20 μ M final) were added and the plate was further incubated for 3 hours at 37°C. After incubation, cells were washed and were incubated with 100 μ L of 10% CCK-8 in DMEM for 3 hours at 37°C. The absorbance was measured at 450 nm with a microplate reader (Polarstar Optima). Controls corresponded to untreated cells (negative control, 100% viability) and cells treated with 0.2% of Triton X-100 (positive control, 0% viability)

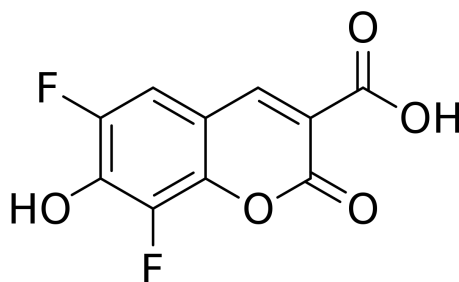


Figure 6.4 – Pacific Blue probe structure scheme. This probe was attached to the LTX-315 C-terminus to carry out quantification assays by fluorimetry. Image by Ed (Edgar181) - Own work, Public Domain, <https://commons.wikimedia.org/w/index.php?curid=66991761>

Quantification assays in adherents cells 12 well plates were inoculated with 1 ml/ well of suspension of CHO cell (500 000 cells/well). After 24 h incubation (5% CO₂ at 37°C), adherents and confluent cells (10⁶ cells /well) were incubated one hour at 37°C or 4 °C with different concentrations of the peptide (1. 2.5; 5 and 10 μ M) in 1 mL DMEM. To access internalized peptides, after washing cells with HBSS, (0.05%) 500 μ L trypsin/EDTA (37°C, 5 min) or (0.05%) pronase (4°C, 10 min) was added for 5 min to hydrolyze the remaining extracellular and the membrane-bound peptide and to detach cells. After addition of enzyme inhibitors (100 μ L at 37°C of trypsin inhibitor at 5 mg/ml; and at 4°C Complete Mini tablet, Roche, in 2.5 mL PBS) mixed with 100 μ L bovine serum albumin (1 mg/mL), cells were transferred into a microtube, centrifuged, washed with 1 mL 50 mM Tris buffer pH 7.4, 0.1% BSA, and lysed in 200 μ L 50 mM Tris pH 7.4, 1 M NaCl, Triton 1%. The samples were then sonicated for 30 min and centrifuged 10 min at 16,000 g.

To obtain the value of the total cell-associated peptide (internalized and membrane-bound species), after washing cells with HBSS, cells were directly lysed in 50 mM Tris pH 7.4, 1 M NaCl, triton 1%. The samples were then sonicated for 30 min and centrifuged 10 min at 16,000 g. Fluorescence intensity in the supernatants was monitored with a MOS 200 M fluorimeter (Biologic SAS, France). The samples were excited at 414 nm and the intensity fluorescence was measured between 425 et 500 nm (0.2 sec/0.5 nm), and the maximal intensity was detected around 460 nm. The maximal intensity around 460 nm was retained for the calibration curve and for quantification of samples. The amounts of total or internalized peptide were calculated by comparing the fluorescence intensity of the sample with a calibration curve.

Samples for the calibration curve were prepared in parallel. For this, 10 different amounts (from 2 to 500 pmoles) of LTX-315 were added to one million cells suspended in 200 μ L lysis buffer. We prepared a range of peptide amounts in the lysis buffer (50 mM Tris pH 7.4, 1 M NaCl, triton 1%). The samples were sonicated 30 min and centrifuged at 16,000 g for 10 min. Fluorescence was then measured in supernatants. The amounts of total or internalized peptide were calculated by comparing the fluorescence intensity of the sample with the calibration curve. For each experimental condition we used duplicate wells, and the experiments were all repeated independently at least two times, as indicated.

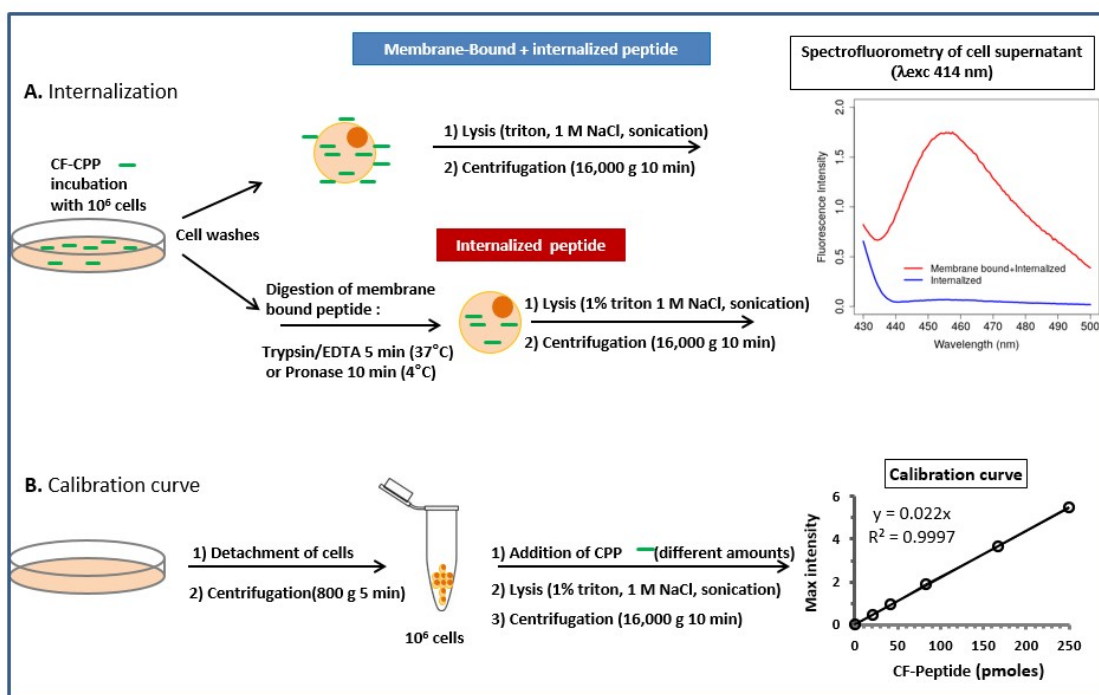


Figure 6.5 – Protocol of the quantification by fluorimetry of the internalization of LTX-315 peptide.

6.2.8 Molecular Dynamics

Generation of LTX-315 PDB file

First, we modeled the Dip structure from scratch. There are several programs and web servers that allow to do this. We used MolView [25] [<http://molview.org/>], a web server where you can easily draw any small molecule and download the structure in a .mol file. Then, we opened DIP.mol file with pymol and generated a .pdb file for the structure (Fig. 6.6).

Then, using pymol we constructed a PDB file for the whole LTX-315 sequence but with a Phe instead of the Dip, as non-standard amino acids are not available in pymol. Then, we proceed to replace the Phe for the Dip structure we generated previously with MolView. The first step is to align the LTX-315.pdb structure with the DIP.pdb structure to move the DIP atoms to the right coordinates: the aim is to align the DIP backbone and one of the benzene rings with the Phe residue. The easiest way to do this, is to use the program ProFit, which is specifically created for the alignment of proteins. You can download and install the program from this link: <http://acrmwww.biochem.ucl.ac.uk/programs/profit/index.html>. Having the DIP coordinates in the correct position, we substituted the Phe for the Dip in a copy of the LTX-315.pdb file. The Phe coordinates for the backbone were maintained. Dip atom positions were included from the C_β . This way, we constructed a PDB file for LTX-315 structure from scratch.

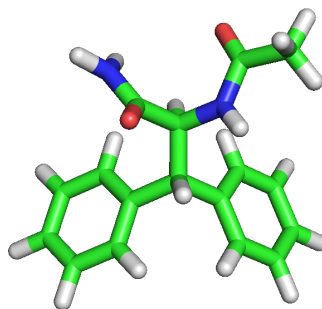


Figure 6.6 – Dip structure generated with MolView. Image generated with pymol.

T-REMD in water

In water, one T-REMD was launched with CHARMM36m FF [93]. For Dip we used the recent parameters of ref [54]. 33 replicas were simulated for 170 ns (the first 50 ns were discarded for the analysis). The starting structure was constructed from scratch as explained before. The minimum temperature was set at 300 K and the maximum at 400 K. Pressure was set at 1 bar and the probability of exchange between replicas was 20 %. The system was composed by one peptide, 3693 water molecules for CHARMM36m simulation, and an ion concentration of 50 mM to match NMR experimental conditions.

T-REMD of insertion

T-REMD simulations of LTX-315 insertion were also carried out. Initially, the peptide was placed in the solvent, a few nm away from the bilayer as shown in figure 6.7. The bilayer was composed by DMPC/DMPS 50:50. There were 72 lipids in total, 2829 water molecules and an ion concentration of 50 mM. This composition was chosen to match NMR experimental conditions. CHARMM36 was the selected FF for lipids [234]. The minimum temperature was set at 300 K and the maximum at 400 K. Pressure was set at 1 bar and the probability of exchange between replicas was 20 %.

6.3 Experimental Results

6.3.1 LTX-315 interacts with lipid membranes

We carried out fluorescence assays of LTX-315 in POPC, POPS or POPC/POPS LUVs to observe its interaction with bilayers. The spectra are shown in figure 6.8. In POPC LUVs, the fluorescence spectra slightly shift to the blue with the increase of lipid concentration (Fig. 6.8a). This variation of Trp fluorescence clearly indicates a change in their environment as a result of the binding of LTX-315 to the LUVs. The peptide reaches its maximum binding capacity when POPC concentration arrives to 3000 μM . This means that at a POPC:peptide ratio of 150:1, all the LTX-315 molecules are bound to the LUVs. In POPS LUVs we observe the same shift to the blue (Fig. 6.8b). However,

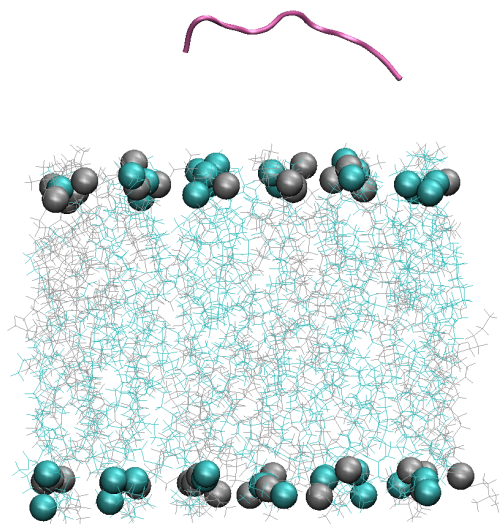


Figure 6.7 – Initial structure for LTX-315 T-REMD simulations of insertion. The peptide backbone is represented in pink cartoon, DMPC lipids in silver lines and DMPS lipids in blue lines. The phosphorous atoms of the polar heads are shown in spheres.

the maximum binding capacity is reached with at a POPS concentration of $400\ \mu M$ (POPS:peptide ratio of 20:1). When using POPS LUVs, the saturation is reached at a lipid concentration 7 times lower compared to POPC. This means that LTX-315 has higher affinity for POPS as the peptide is rich in lysines (net charge of +6) and POPS is a negatively charged lipid. The third fluorescence spectra correspond to the samples of LTX-315 with POPC/POPS (50/50) LUVs and the same shift to the blue is observed (Fig. 6.8c). In this case, the maximum binding capacity is reached with $1000\ \mu M$ of LUVs, corresponding to a LUVs:peptide ratio of 50:1. LTX-315 also presents higher affinity for POPC/POPS LUVs than for POPC, showing again the impact of the electrostatic attraction on its binding capacity.

Figure 6.8d presents the maximum intensity of each spectrum as a function of lipid concentration. This plot gives a simple visual representation of the affinity of LTX-315 for the three LUV compositions. As it was shown previously, saturation for POPC, POPS and POPC/POPS LUVs is reached at $3000\ \mu M$, $400\ \mu M$ and $1000\ \mu M$, respectively. LTX-315 presents higher affinity for negative lipids as it is rich in lysines. Interestingly, the curve for the mixture POPC/POPS is closer to that for POPS than that for POPC. It means that LTX-315 affinity for the membrane is not linearly correlated to the amount of charged lipids. A small fraction of negatively charged lipid is probably sufficient to get an affinity increase.

These fluorescence assays of LTX-315 interaction with LUVs are the first *in vitro* evidence of the peptide direct interaction with lipid bilayers and its electrostatic-driven attraction for membranes.

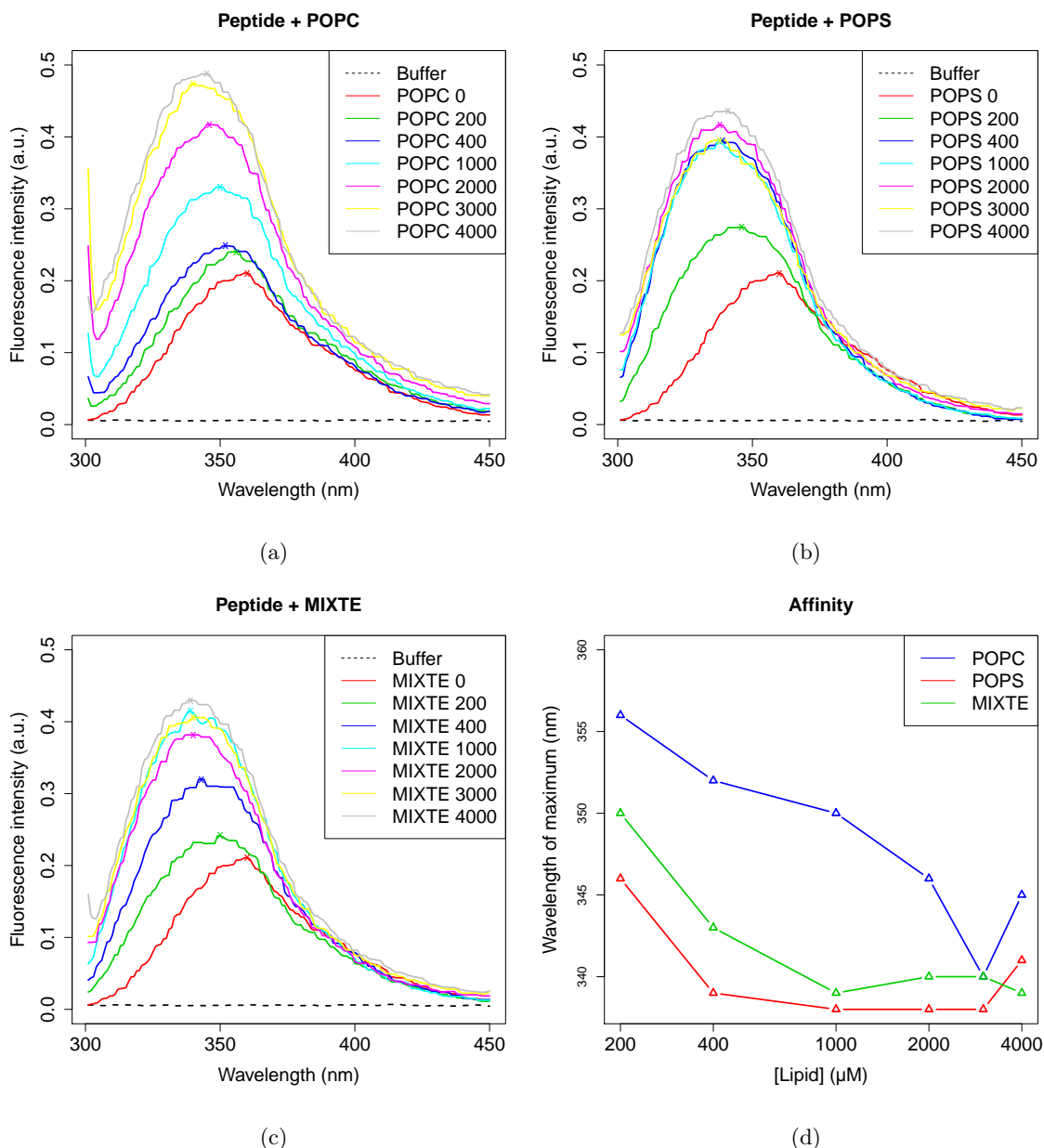


Figure 6.8 – Fluorescence spectra of LTX-315 in absence or presence of LUVs of (a) POPC, (b) POPS and (c) at different concentrations (from 200 to 4000 μM). Each cross represents the maximum of the spectrum. The dotted line is pure buffer, the red line corresponds to the peptide in buffer, then each other color line shows the peptide with lipids at various concentrations. The values in the legends indicates the lipid concentration in units of μM .

6.3.2 Structural properties

Figure 6.9 shows the CD spectra of LTX-315 in absence or presence of LUVs in the UV region. The UV region in CD usually reports on the secondary structure content of the peptide. However, LTX-315 is particularly rich in Trp (3 amino-acids out of 9, representing 33% of the whole sequence). Trp also has a contribution in the UV region, which gives rise to the unusual band centered between

225 and 230 nm. Already in buffer without lipids, this band seems to exist, and it increases when LUVs are added to the samples. This is a clear sign that the environment of Trp changes, which is fully consistent with the fluorescence experiments shown in the previous section, where we shown that LTX-315 is capable of binding to lipid bilayers.

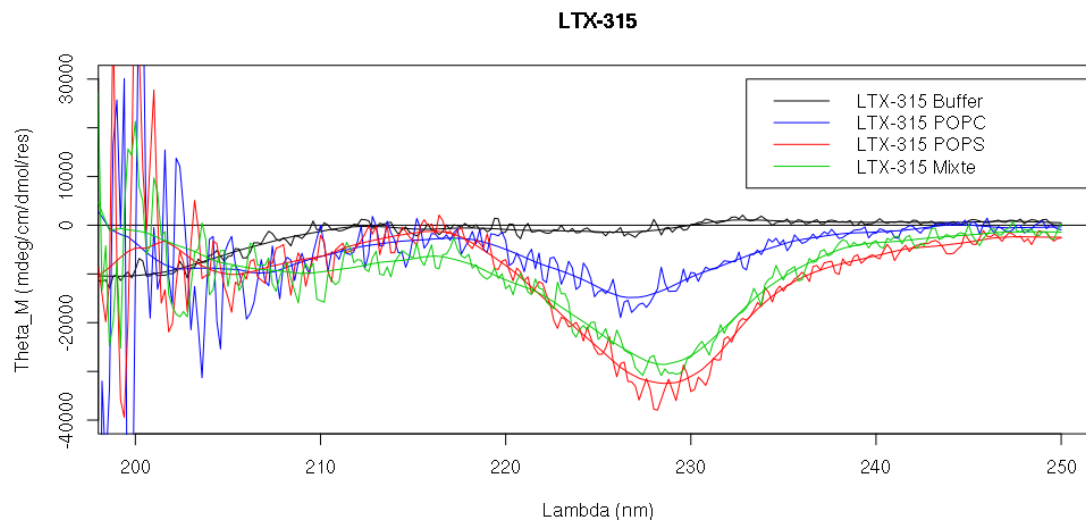


Figure 6.9 – CD spectra of LTX315 in the UV region in absence or presence of POPC, POPS and POPC/POPS (50:50) LUVs. An smoothed curve of each spectra is shown to ease the view.

In buffer (black curve in figure 6.9), LTX-315 presents the typical spectra of a random coil, with the ellipticity close to 0 from 210 nm to 250 nm, and a negative band centered on 200 nm. The noise is extremely high below 200 nm, we thus decided not to include this region in the figure.

Unfortunately, it is hard to infer any increase of helical content based on the blue, red and green spectra given the band due to Trp content as well as the important noise caused by the double bond of the oleoyl aliphatic chain at low wavelengths. Thus, we cannot definitely conclude on the secondary structure of the peptide. However, from the peak caused by the presence of the Trp it is possible to extract some interesting information. The peak of the samples containing POPS LUVs is more intense than the peak of the samples containing POPC LUVs. This is due to the higher affinity of LTX-315 for POPS LUVs than for POPC LUVs, as seen in fluorescence assays. But which is more interesting is that the spectra of the POPC/POPS mixture is closer to the POPS one than to the POPC one. The LTX-315 affinity for charged membranes is not linear: a small concentration of negative charges is sufficient to increase the affinity, as it was also seen with fluorescence assays.

To have more information about LTX-315 structural properties, NMR experiments were carried out by our collaborators Olivier Lequin and his Ph.D student Edward Chalouhi. They studied LTX-315 secondary structure in buffer, in micelles and in zwitterionic bicelles (a description of these structures is given in the introduction chapter 1.3.4).

They assigned the chemical shifts of the C_{α} of the peptide. A chemical shift is the offset in

resonance frequency of an specific nucleus relative to a reference frequency, divided by the Larmor frequency. It depends only on the nucleus type and its molecular environment. From the chemical shifts, they obtained the chemical shift deviations (CSD), which are calculated as the difference between the experimental chemical shift observed on the spectra and the chemical shift of the same residue in a random coil structure. CSDs were used to infer the secondary structures adopted by the peptide. The values used as reference were the Wishart set obtained from GGXAGG peptides. Results of CSD calculations of the peptide in buffer, micelles and zwitterionic bicelles are shown in figure 6.10. For a given residue, if the CSD values for the C_α are higher than +1 ppm, it tends to adopt an α -helical structure. Values lower than -1 ppm correspond to β -sheets and values close to 0 ppm correspond to unfolded conformations. Results show that, in buffer, the peptide is unfolded, showing mostly random coil chemical shifts. In micelles and bicelles, the peptide seems to fold into an α -helix mostly towards the C-terminal. Unfortunately, CSD cannot be calculated for Dip residue because there are no available data to be used as reference of its chemical shift in random coil.

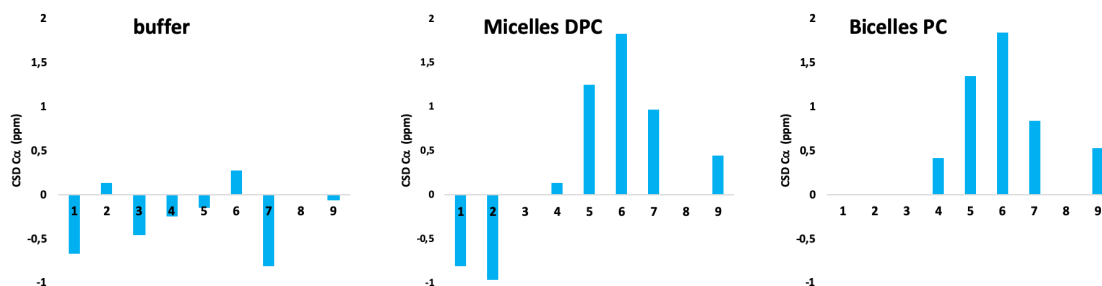


Figure 6.10 – CSD of LTX-315 C_α nuclei in buffer, micelles or bicelles. CSD was not calculated for Dip8 residue.

In order to understand the impact of the environment in the Dip conformation, they calculated the difference between the chemical shift of the residue in buffer and in micelles or bicelles. Results are shown in figure 6.11. The Dip shows a negative chemical shift difference between values in buffer and in a lipid environment. It has an opposite behaviour compared to the residues forming an α -helix. Results suggest that the Dip does not fold into an α -helix neither in the solvent or interacting with lipids. This could be due to a steric effect of the two big aromatic rings of its side chain.

Fluorescence and CD experiments showed that LTX-315 presents higher affinity for negatively-charged bilayers than for non-charged bilayers. For these reasons, NMR experiments on the peptide in anionic bilayers were recorded, specifically in bicelles containing 50% DMPS. Unfortunately, due to some experimental difficulties, CSD could not be obtained for C_α . Instead, CSD for H_α are presented (Fig. 6.12). When using H_α nuclei, the propensity is determined to be towards an α -helical structure when the CSD value is lower than -0.1 ppm. LTX-315 seems to fold into an α -helix in anionic bilayers. Compared to the results in zwitterionic bilayers, residues K2, W3 and W4 present a more stabilized helix in anionic conditions.

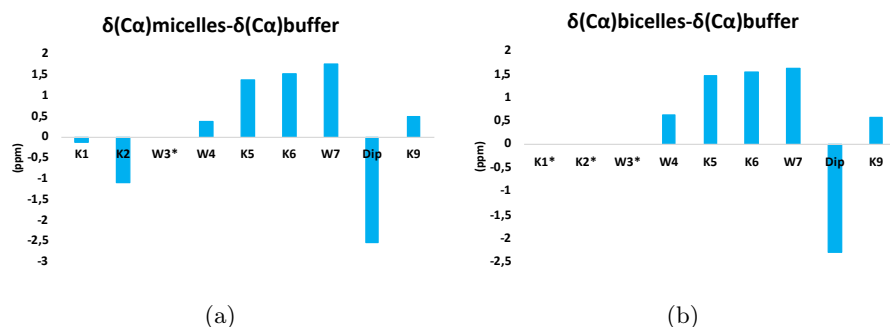


Figure 6.11 – Difference between LTX-315 chemical shifts in buffer and in (a) micelles or (b) bicelles.

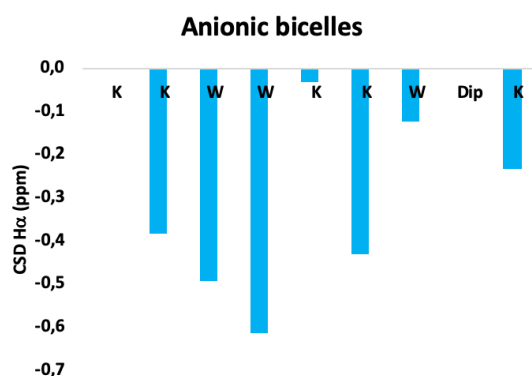


Figure 6.12

6.3.3 Orientation in the membrane

Apart from chemical shifts, other data that can be obtained from NMR experiments. The paramagnetic relaxation enhancement experiments provide information about the peptide position with respect to the membrane surface. A paramagnetic probe is attached to a modified phospholipid. The strong magnetic interaction between the radical electron spin and nearby nuclear spins causes enhanced nuclear spin relaxation, giving rise to signal broadening and intensity decrease. This paramagnetic relaxation enhancement effect (PRE) is monitored by calculating residual intensity and gives information about the peptide orientation and proximity to the paramagnetic lipid. This information is a key in the understanding of AH/membrane interactions. O. Lequin and E. Chalouhi carried out this experiment to study LTX-315 orientation in anionic bicelles. First, they inserted a paramagnetic probe at carbon 5 of the fatty acid chain of the doxyl-PC lipid. Atoms were divided in two groups to interpret the signal: backbone and side chain. The higher the intensity of the signal, the lower the degree of insertion. Results are shown in figure 6.13. As we can see, residues W3, W4, W7 and Dip are more affected by the presence of the paramagnetic probe than residues K5 and K6, the least affected. This means that the hydrophobic residues are inserted deeper in the membrane than the two lysines. In addition to that, the hydrophobic residues have their side chains more affected than their backbone, so the aromatic rings are the group of atoms

more inserted of the peptide. However K6 side chain presents the higher signal, meaning that it is the least inserted of all the side chains of the peptide. Interestingly, K9 backbone presents a lower signal, so it is deeply embedded. Unfortunately, because of experimental conditions, no signal was obtained for the first two residues. From these results we can conclude that the peptide is positioned parallel to the membrane/water interphase with an α -helical secondary structure as the sequence of intensity is well correlated with the periodicity of an α -helix. The peptide seems to be slightly tilted with his C-terminal more embedded than the N-terminal.

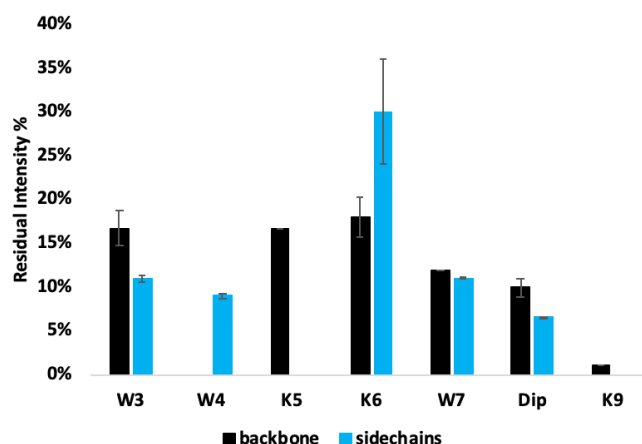
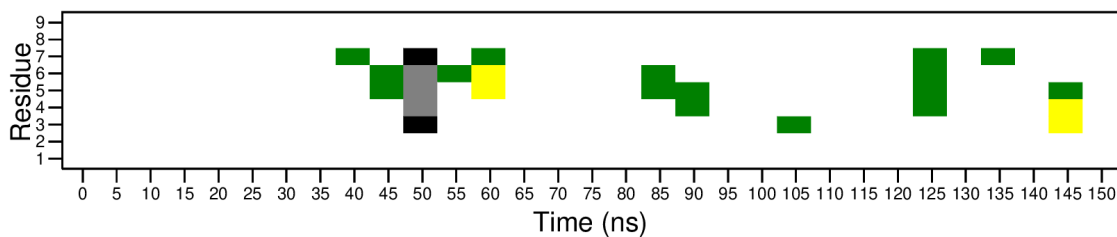


Figure 6.13 – Residual intensity calculated from the paramagnetic relaxation enhancement experiments in anionic lipids. In black, results for the backbone; in blue, results for the side chains.

6.4 Computational results

Simulations in water A first T-REMD simulation of LTX-315 was carried out to observe its behaviour in water. In this section, we analyze the replica at 300 K. The first 50 ns of the simulation were discarded. Regarding the secondary structure, LTX-315 in the solvent stays unfolded, with a mean helicity of $0.2\% \pm 3.0$. This value is so low that we can tell that the peptide is not helical in water. A deeper analysis is presented in figure 6.14a, where the secondary structure is determined by residue through time. The peptide is mostly in random coil. It sometimes presents small turns or loops but never an α -helical structure. The two extremes, including the Dip residue, are always in random coil.

We used the conformation shown at 300 K to perform an structural clustering. The RMSD between the C_α were compare and the cutoff was set at 0.2 nm. 40 structural clusters were obtained. The biggest one contained 35.3% of the conformations. The central structure is shown in figure 6.14b. It is a completely extended structure, as expected regarding the secondary structure analysis. Results match the NMR results in buffer shown in the previous section, validating the reliability of our structural analysis. For this reason, the structure shown in figure 6.14b was used then as initial structure for the simulations of the insertion in the membrane.



(a) LTX-315 secondary structure throughout time by residue in T-REMD simulations in water. Color green is for the bends, black for β -bridges, grey for 3_{10} -helices, yellow for turns and white for random coil. No α -helix is present. For this analysis, T-REMD conformations at 300 K are used. We do not follow a continuous replica over the temperatures.

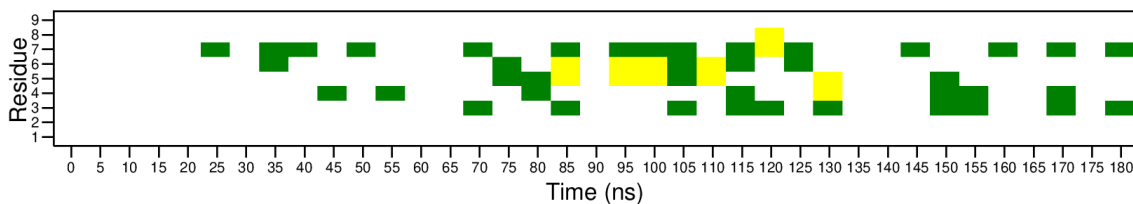


(b) Major conformation found in water at 300 K. The image corresponds to the central structure of the biggest cluster which contains 35.3 % of the conformations.

Figure 6.14 – LTX-315 T-REMD simulations in water. Only conformations at 300 K are considered.

Insertion simulations LTX-315 was initially placed in the solvent, a few nm away from the bilayer, composed by DMPC / DMPS 50 / 50. The T-REMD simulations last for 180 ns. Unfortunately, we haven't had the time to make them longer. Although we are going to extend them, preliminary results are shown in this chapter. Only the conformations at 300 K have been taken into account. In the simulation, the peptide takes about 50 ns to partition into the membrane. This process is very fast. For this reason, the first 50 ns of simulation were discarded for the analysis. The secondary structure analysis shows an unfolded peptide in the membrane. Although it partitions very fast into the membrane, it presents a helicity of 0% during the whole simulation. The secondary structure graph is shown in figure 6.15a, and we can see that the peptide only presents some bends and turns but there is no helical content. We carried out a structural clustering using the RMSD between the conformations and a cutoff of 0.2 nm. 19 clusters were found. The first one contains 23.4% of the conformations. The central structure is shown in figure 6.15b. As we can see it is an unfolded structure, very similar to the one found in water (Fig. 6.14b). Apparently, even if LTX-315 approaches very quickly to the membrane, simulations are not long enough to capture the peptide folding.

Even though the peptide does not fold in the simulated time, we wanted to have further information about the partitioning. Figure 6.16 shows the mean z position of LTX-315 center of mass (COM) (Fig. 6.16a) and of Dip side chain COM (Fig. 6.16b). In the first plot, one can see that LTX-315 insertion is fast and stable. For all the simulations, it takes about 50 ns to get



(a) LTX-315 secondary structure thought time by residue in T-REMD simulations of insertion. Color green is for the bends, yellow for turns and white for random coil. No α -helix is present. For this analysis, T-REMD conformations at 300K are used. We do not follow a continuous replica over the temperatures.



(b) Major conformation found in the membrane. The image corresponds to the central structure of the biggest cluster which contains 23.4 % of the conformations.

Figure 6.15 – LTX-315 insertion T-REMD simulation. Only conformations at 300 K are considered.

inserted. Once inserted, it does not go out from the membrane. Notice that there are points at both layers. This is because we are analyzing conformations at 300 K, so structures from all the replicas are being plotted. Some replicas bind the upper leaflet, some others the lower one. When looking at the Dip z positions, one realizes that it partitions very deep in the membrane. Figure 6.17 shows a snapshot at 144.5 ns of simulation. The Dip is the most embedded residue of all and its position is very deep towards the center of the membrane, even the BB is below the phosphorous level.

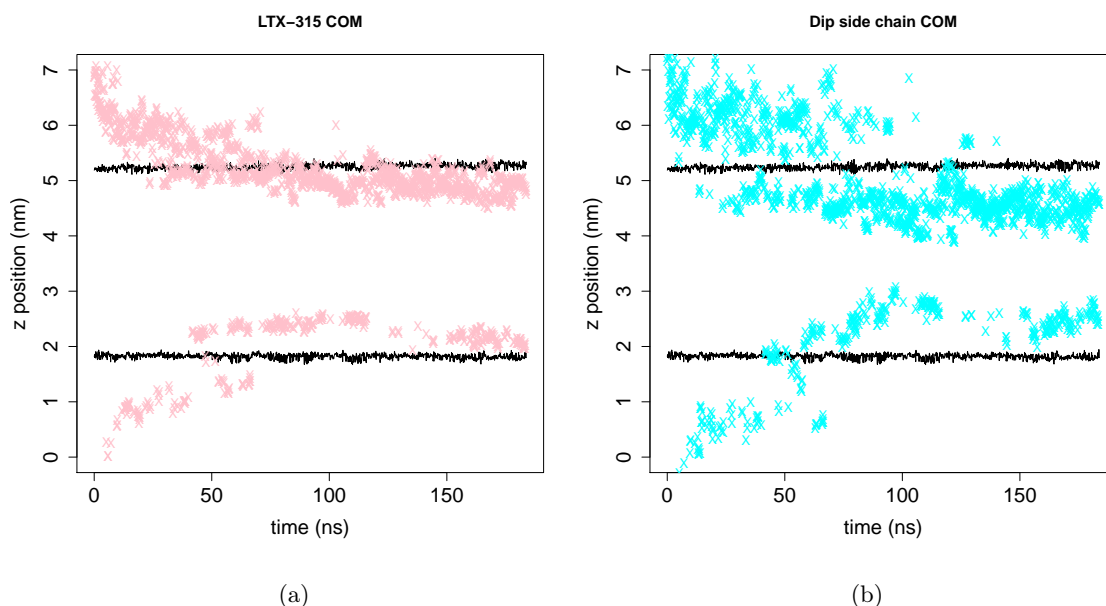


Figure 6.16 – z positions of (a) LTX-315 COM (pink) and (b) Dip SC COM (blue). Phosphorous atoms mean position is plotted in black. For this analysis, T-REMD conformations at 300K are used. We do not follow a continuous replica over the temperatures.

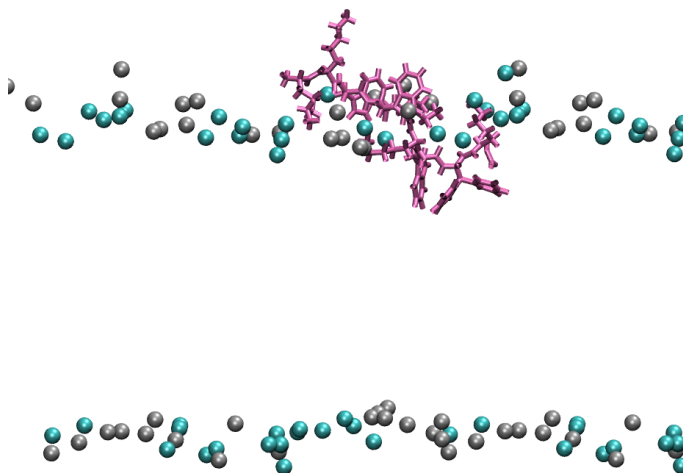
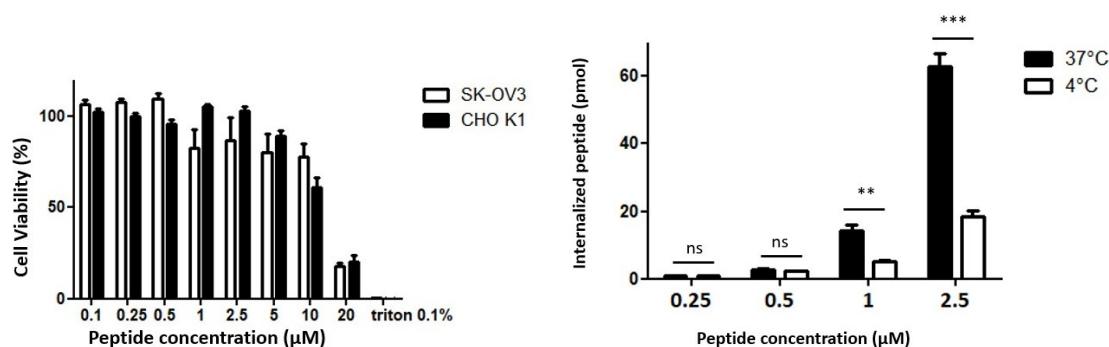


Figure 6.17 – Snapshot of LTX-315 T-REMD simulation of insertion at 144.5 ns and 300 K. Dip is the most buried residue.

6.5 Penetration into the cell

Internalization assays were carried by our collaborator Françoise Illien to observe LTX-315-Pacific blue penetration capacity. She has designed a novel protocol relying on fluorometry to quantify fluorescent peptide inside cells. This robust method performs an absolute quantification of the peptide internalized, and it is possible to distinguish endocytosis from direct membrane translocation [96]. Part of this protocol has been applied to study LTX-315-Pacific blue internalization

mechanism. First, she carried out cytotoxicity assays to check the range of concentrations at which the peptide kills cells. Cytotoxicity assays were done with two cell lines: SK-OV3 (human ovarian cancer cell line) and CHO-K1 (Chinese hamster ovary cells). Results are shown in figure 6.18a. Between 0.25 and 0,5 μM , cell viability is over the 50 %. LTX-315-Pacific blue starts being cytotoxic between 10 and 20 μM . It does not seem to be more cytotoxic for cancer cells compared to normal cells.



(a) Cytotoxicity test of LTX-315 peptide in SK-OV3 (human ovarian cancer cell line) and CHO-K1 (Chinese hamster ovary cells). (b) Quantitative fluorescence spectroscopy of LTX-315 internalization at two different temperatures: 4°C and 37°C.

Figure 6.18 – Internalization assays

Considering cytotoxicity results, internalization assays were done for four different concentrations of the peptide: 0.25, 0.5, 1 and 2.5 μM in CHO cells. LTX-315-Pacific blue concentration was not increased over 2,5 μM because cells need to be viable to observe the internalization. Quantification was done by quantitative fluorescence spectroscopy at 4 and 37°C. Internalization results are shown in figure 6.18b. At low concentrations of peptide (0.25 and 0.5 μM), we observe a small quantity of internalized peptide at both temperatures. At 1 μM , internalization slightly increases, especially at 37°C. The highest internalization amounts are obtained at a peptide concentration of 2.5 μM : about 20 *pmol* at 4°C and more than 60 *pmol* at 37°C. The quantity of internalized peptide is much higher at physiological temperature than at low temperature. This is because at 4°C, only direct translocation through the membrane takes place. There is no ATP production in the mitochondrion at low temperature, so no active (energy consuming) internalization can take place [218]. However, at 37°C the peptide can penetrate the cell by endocytosis, which consumes energy, and by translocation at the same time. Quantitative results are shown in table 6.3. They show that LTX-315 optimal concentration for internalization is 2.5 μM , as it was seen by cytotoxicity assays. Also, the peptide is clearly more internalized at physiological temperature as it can use both pathways, direct translocation and endocytosis. Finally, results show that the peptide uses active endocytosis as a preferred internalization pathway. This study allowed us to quantify LTX 315-Pacific blue peptide in the cell at 37 and 4°C. The results showed that it behaves like a CPP: it is able to enter the cell by direct interaction with the lipid bilayer and preferably by endocytosis

when the peptide concentration and the temperature are high enough.

[Peptide] (μM)	37°C Endocytosis + Translocation (<i>pmoles</i>)	4°C Translocation (<i>pmoles</i>)	Endocytosis (<i>pmoles</i>)
0.25	0.9 +/- 0.05	0.9 +/- 0.05	/
0.5	2.8 +/- 0.03	2.2 +/- 0.01	/
1.0	14.3 +/- 1.7	5.1 +/- 0.5	9.2 +/- 1.8
2.5	62.7 +/- 4.1	18.4 +/- 1.7	44.3 +/- 4.5

Table 6.3 – Calculation of the quantity of LTX-315 internalized by translocation and endocytosis

6.6 Discussion

The LTX-315 peptide represents a promising strategy to treat sarcoma disease. It is part of a family of oncolytic peptides developed by LytixBioPharma and they all share similarities in their sequence. Despite all the clinical assays that have been carried out, there is a lack of information about his action mechanism from a molecular point of view. In this research project, we have combined *in vitro* and *in silico* assays to construct a model of LTX-315 interaction with membranes.

We presented the first evidences of LTX-315 direct interaction with lipid bilayers. Fluorescence, CD and NMR assays proved the interaction between the peptide and lipid membranes without intermediates. They also showed LTX-315 preferred affinity for negatively-charged membranes due to its high content in Lys. Electrostatic forces seem to have a major role in this interaction with membrane, as it has been seeing by other AH as the α -synucleine [10]. However, hydrophobic interactions are also present and are important as it is capable of binding neutral membranes.

This project represents an advance on the insights of LTX-315 structural properties and behaviour. Unfortunately, no structural information was obtained from CD assays but NMR assays has proven that LTX-315 behaves as an amphipathic helix as predicted by Haug [86]. Unfolded in the solvent, it folds into an alpha helix when it partitions into the membrane. In the solvent, it is completely unfolded without any α -helical structure, as it was also seen by T-REMD simulations. When it partitions into the membrane, it folds into an α -helix with unfolded termini. This is the case of most of the AH, as for example, the mastoparan [91]. The secondary structure seems to be stabilized by when the membrane contains negative charges as seen by NMR assays. The Dip always stays in a random coil configuration. This can be because it is located in the C-terminal but may be the presence of two big aromatic rings can have an steric effect that prevents the Dip from folding. Unfortunately, T-REMD simulation of the insertion in DMPC/DMPs membrane was not long enough to observe its folding. LTX-315 partitions very quickly in the membrane but stays unfolded. The presence of large hydrophobic side chains (Trp and Dip) in its sequence makes the folding in the membrane more slow. For example, in chapter 4 we observed that the

mastoparan was folded after 100 ns of T-REMD simulation of insertion approximately (Fig. 4.2). The mastoparan has no aromatic rings in its sequence, so it has less steric restrictions and thus folds more quickly.

Thanks to NMR assays we also studied LTX-315 orientation in the membrane. In fact, the C-terminal partitions deeper in the bilayer than the N-terminal. In consequence, the helix is tilted. In T-REMD simulations we observed that the Dip partitions very deep into the membrane, even when the peptide is unfolded. Maybe, the presence of Dip large hydrophobic side chain of the C-terminal provokes the helix tilt. The two aromatic rings establish strong hydrophobic interactions with the aliphatic chains burying the C-terminal deeper. In literature, other AH present a tilted insertion in the membrane due to its non-homogeneous sequence as, for example, the penetratin [137]. Anyhow, these first REMD data seem to show that Dip is probably a key residue in the insertion mechanism of LTX-315.

LTX-315 penetration in cells was also studied by a novel strategy developed by our collaborators Françoise Illien and Sandrine Sagan [96]. She first carried out a cytotoxicity test and determined that LTX-315 is capable of killing cells at low concentration (from 0.5 μM) at establish range of concentrations adequate for the peptides efficacy (from 0.5 to 2.5 μM). This concentration is lower than the one they needed to study the penetratin [96]. In these assays, the peptide was not more cytotoxic for cancer than for normal cells.

Then, she carried out quantification assays by fluorimetry to study LTX-315 internalization pathways at 4 and 37 °. At 4°C, only direct translocation takes place; as it does not consume energy, it is independent from temperature [19]. With a peptide concentration of 2.5 μM , 18.3 pmoles of the peptide are internalized by direct translocation. Notice that this pathway may be underestimated as it was measured at 4 °C, where the plasma membrane is less dynamic and more rigid. These results agree with previous observations published by LM. Eike *et al.*. They observed that LTX-315 was able to disintegrate the plasma membrane at low temperature (4 °C), pointing toward a direct interaction with the plasma membrane [65]. Even though it is able to penetrate the cell by translocation, the peptide is mainly internalized by endocytosis. With a peptide concentration of 2.5 μM , 62.7 pmoles of the peptide are internalized by endocytosis. Endocytosis takes place in two steps: endocytic entry followed by endosomal escape. In the endosomal vesicle the peptide risks to be degraded by enzymes, so the escape crucial step for the peptide to be efficient. The favorable conditions for endosomal escape, as the influence of the peptide physico-chemical properties, are not yet known, giving an additional reason for studying AH/membrane interactions [19]. From this internalizations results, we could determine that LTX-315 is a CPP, a family of short cationic peptides that are capable of crossing the plasma membrane [124]. CPPs are more and more used in biomedicine, usually because they can efficiently transport biologically active molecules inside living cell, being promising devices for medical and biotechnological developments [19]. LTX-315 is an example of CPP used in biomedicine not as a transporter but because of his own oncolytic properties.

The presence of a non-natural amino acid, the Dip, have been a challenge for this project. There was no previous information about its behaviour, the relative positioning of the two aromatic rings, its intramolecular interactions or the way it impacts the peptide secondary structure. Thanks to a collaboration with Rodolphe Vuilleumier and Ari Seitsonen we obtained a parametrization for the Dip that we used to launch preliminary simulations. Then, the publication of the new parameters for non-standard amino acids in June 2021 allowed us to perform LTX-315 simulations using exclusively CHARMM36m FF [54]. It is also important to mention that LytixBioPharma, provided samples of the peptide that allowed us to carry out the experimental assays.

Summing up, thanks to the combination of experimental and computational techniques we have constructed a first model of LTX-315 behaviour: LTX-315 is an AH, unfolded in water, folded in the membrane/water interface, part of the CPP family and able to interact with lipid bilayers with preferred affinity for anionic membranes. The non-standard residue, the Dip, stays unfolded independently from the environment. Unfortunately, we didn't have time to finish and fully analyze the T-REMD simulations. They are going to be extended to increase the sampling and new analyses concerning SC interactions will be included. We hope to observe the folding of the peptide. We also plan to perform REMD simulations of LTX-315 insertion in pure DMPC. It will be interesting to compare these results to DMPC/DMPS.

In a longer term, it would be interesting to construct an MSM of LTX-315 action mechanism applying the protocol presented in the previous chapter (Chapter 5. Mastoparan folding in the membrane: Markov State Models to decipher the mechanism).

Chapter 7

Involvement of PIP₂ in the internalization of the cell-penetrating peptide Penetratin

7.1 Penetratin

7.1.1 Introduction

Cell-penetrating peptides (CPPs) are defined as short amino acid sequences able to enter cells and to deliver biologically active molecules such as nucleic acids, proteins, nanoparticles, drugs or diagnostic agents, in cells and tissues. In 1991, evidence of the translocation of the 60 amino acid Antennapedia homeodomain was reported [108]. Thereafter, the short 16 residue peptide sequence, pAntp(43-58), called Penetratin (sequence 43RQIKIWFQNRRMKWKK58-NH₂), was proved to be responsible of this translocation property [58]. Penetratin, together with the Tat peptide, responsible of the translocation of the HIV Tat protein [217], were the first described members of the so-called CPP family, which have been widely studied.

CPPs have been found to follow two distinct pathways for their internalization in cells: endocytosis and direct translocation [119]. An illustrative scheme is shown in the Introduction chapter (section 1.4). They can use these two routes concomitantly, in different proportions, depending on their nature, cell type and various environmental factors. Aside from the well described endocytic routes, many models have been proposed to explain how CPPs could cross the hydrophobic core of the plasma membrane. The variability in internalization efficiency according to the cell type and peptide sequence implies that CPPs associate with various interaction partners at the membrane surface. The negatively charged lipids and carbohydrates (Glycosylaminoglycans or GAGs) are considered to be the first molecular partners to be encountered by positively charged CPPs [219].

Many studies have investigated the interaction of Penetratin with negatively charged model membranes through various methods. It is now clearly admitted that Penetratin strongly interacts with anionic lipids [5, 239], and that they are involved in Penetratin translocation through membranes [205, 207, 79].

PI(4,5)P₂ is an anionic lipid carrying 3 to 5 negative charges at physiological pH [147]. PI(4,5)P₂ has multiple functions such as second-messenger precursor, regulator of actin polymerization, membrane trafficking, etc [118]. It is mostly localized in the inner leaflet of the plasma membrane, even though up to 2% of PI(4,5)P₂ can be found in the outer leaflet [39, 76], where it could be involved in the regulation of cell adhesion and motility [231].

Several studies have brought the focus on the unique role of PI(4,5)P₂ in unconventional secretion pathways for the HIV-1 Tat Protein [177], the Fibroblast Growth Factor (FGF2) [203] and more recently the homeoprotein Engrailed-2 (En-2) [7]. PI(4,5)P₂ has been shown to be the target of cytotoxic membrane-active peptides such as the plant defensins NaD1 [171, 16], TPP3 [17] and NoD173 [128] or the chimeric peptide TAT-RasGAP317-326 [198]. PI(4,5)P₂ has also been shown to be a binding partner for cyclic CPPs and has been proposed to play a role in their uptake mechanism [40]. Interestingly, another phosphoinositide, PI(3)P, has been shown to mediate the entry of pathogen effector proteins containing RXLR-like motifs into plant and animal cells [116].

Previous work carried out in our laboratory, has suggested a role for PI(4,5)P₂ in the kinetics of Penetratin internalization [104]. Very recently, a direct involvement of PI(4,5)P₂ in the internalization of En-2 was demonstrated [7]. This is of high significance as Penetratin is derived from a homeoprotein, Antennapedia, highly related to En-2. The cell-penetrating sequence derived from En-2 shares 81% similarity with Penetratin and has very close behavior in terms of internalization and interactions in a membrane-mimicking environment [15].

This present work aims to analyze the role of the negatively charged lipid PI(4,5)P₂ as a binding partner of the CPP Penetratin and as a potential effector of Penetratin internalization, combining experimental and computational techniques. It is part of a bigger collaboration, headed by Dr. Astrid Walrant and Dr. Emmanuelle Sachon. Penetratin structure, orientation and insertion in PI(4,5)P₂-containing model membranes was studied by NMR assays performed by our collaborations Dr. O. Lequin and E. Chalouhi. Then, MD-CG simulations were performed to observe the binding of Penetratin to PI(4,5)P₂, and its effect on the lateral organization of the lipids. Finally, from the conformations observed in the CG simulations, a back-map of the system was carried out to launch MD All-Atom simulations that allowed to study the PI(4,5)P₂-Penetratin interaction with atomic detail. Other *in vitro* assays that are not shown in this chapter were carried out. A publication containing the whole ensemble of results will be published soon.

7.1.2 Materials and Methods

MD CG Simulations

Coarse-grained (CG) molecular dynamics (MD) simulations were carried out using the MARTINI force field for lipids [146] and proteins [155] (version 2). Both electrostatic and Lennard-Jones interactions were computed using a 1.2 *nm* cut-off with a switch function and the relative dielectric constant for the medium was set to 15. The temperature for each group (peptide, lipid, water) was maintained at 310.15 *K* using the v-rescale temperature coupling algorithm [38] with a time constant of 1 *ps*. The pressure was kept constant using the Parrinello-Rahman algorithm [160] with a semi-isotropic pressure coupling (*x* and *y* dimensions, the bilayer plane, coupled independently from the *z* dimension) and a time constant of 12 *ps*. The integration time step was 20 *fs* and structures were saved every 100 *ps* for analysis. Four different systems were constructed using the CHARMM-GUI web server [105]: two Penetratin containing systems with lipid compositions POPC/POPS/PI(4,5)P₂ (80/15/5) and POPC/POPS (80/20), and two with the same membrane composition but without Penetratin as control systems. Each system had 200 lipids (100 per leaflet) and 13-14 water beads per lipid. Sodium and chloride beads were added to get the system neutral and reach a salt concentration of 150 *mM*. For the simulations with Penetratin, we put one peptide per leaflet. This enabled to get twice the sampling at almost no computational cost. The initial structure was the fully-helical conformation determined by NMR in the presence of PI(4,5)P₂ by our collaborators. The α -helix was inserted into the lipid membrane at the glycerol level with the hydrophobic face oriented towards the bilayer interior and the polar face towards water. Since the MARTINI CG force field has no specific term describing hydrogen-bonding, the helical conformation was maintained throughout the whole simulations using a specific dihedral term on all four consecutive backbone beads which were part of the α -helix. Each system was first energy minimized and equilibrated with position restraints on the peptide for 10 *ns*. Production simulations were then run for 10 μ s for each system. The first μ s was systematically discarded for the analyses. All simulations were carried out using GROMACS 2018.5 [1].

MD AA Simulations

All-Atom (AA) MD simulations were performed with GROMACS 2018.5 [1] using the CHARMM36m FF [123, 234] for lipids and CHARMM36m [93] for Penetratin. For constructing the system, we took the final frame (at 10 μ s) of the first trajectory of our CG MD simulations. This CG system was back-mapped to AA using the CHARMM-GUI web server [105].

Since we used a back-mapping strategy, the AA system is identical to the CG one in terms of composition, that is, 2 Penetratines (one per leaflet), 160 POPC, 10 PI(4,5)P₂ and 30 POPS. The system had 10704 water TIP3P molecules and an NaCl concentration of 150 *mM*. The system was then equilibrated using the CHARMM-GUI protocol, which consists in an energy minimization followed by several short MD simulations with position restraints on lipids that are progressively

released. Two trajectories of 100 *ns* with different starting velocities were then performed in the NPT ensemble. An integration time step of 2 *fs* was used. Electrostatic interactions were calculated with the particle-mesh-Ewald (PME) method [57, 67], with a real-space cutoff of 1 *nm*. Van der Waals interactions were computed using a Lennard-Jones force-switching function over 10 to 12 Å. Bond lengths were constrained using the LINCS algorithm [87]. Water molecules were kept rigid with the SETTLE algorithm [153]. The system was coupled to a Bussi thermostat [38] at a temperature of 313.15 *K* and to a semi-isotropic Parrinello–Rahman barostat [160] at a pressure of 1 *bar*. MD frames were saved every 10 *ps*. The first 10 *ns* of each trajectory were systematically discarded for the different analyses. A partitioning plot, showing the position of each side-chain center of mass with respect to the bilayer center, was computed by averaging each position over the two trajectories.

NMR assays

A collaboration was established with Edward Chalouhi, PhD student of the LBM and his PhD Director Olivier Lequin. They carried out H-NMR experiments to study Penetratin’s interaction with lipids with three different samples: peptide in buffer, peptide with micelles and peptide with bicelles.

NMR samples were prepared in 5 *mm* Shigemi tubes using a volume of 300 μ L. Peptide and peptide/micelles samples contained 1 *mM* peptide in 50 *mM* sodium buffer, pH = 5.6, 0.02 % (w/v) NaN₃, 0.11 *mM* sodium 2,2-dimethyl-2-silapentane-5-sulfonate-d6, in 90/10 (v/v) H₂O/D₂O. Micelles were composed by DPC and its concentration was 80 *mM*. Peptide/bicelles samples contained 1 *mM* peptide in 50 *mM* sodium phosphate buffer, pH 6.05, 0.02 % (w/v) NaN₃, 0.11 *mM* sodium 2,2-dimethyl-2-silapentane-5-sulfonate-d6, in 90/10 (v/v) H₂O/D₂O. Bicelles samples contained 75 *mM* DHPC and 25 *mM* DMPC.

Two different experiments were carried out : 2D ¹H-¹H TOCSY and 2D ¹H-¹H NOESY. They were recorded at 35°C on a Bruker Avance III 500 *MHz* spectrometer equipped with a TCI cryoprobe. Band-selective pulses were used in both experiments to improve the sensitivity [176]. They were applied to selectively observe the amide/aromatic region in the acquisition dimension. NMR data were processed with TopSpin 3.6 program and analyzed with NMRFAM-SPARKY [130].

7.1.3 Membrane Structural properties

The MD-CG simulations allow to study the structural properties of the membrane. We studied the effect of including PI(4,5)P₂ lipid in the membrane, the impact of penetratin insertion and the two of them combined.

The first analysis carried out was the Area Per Lipid (APL) calculation. The results are shown in table 7.1. First, we checked POPC and DOPS values in the control simulations without

penetratin. Both lipids present normal values already seen in the literature [126, 168]. This a way of validating the reliability of the FF employed as the APL values obtained match the experimental information. The insertion of penetratin in the membrane does not seem to have a big impact in membrane APL, neither the inclusion of PI(4,5)P₂. POPC and DOPS APL values do not variate despite the changes in the system composition.

PI(4,5)P₂ APL values give more interesting information with respect to the membrane structural properties. PI(4,5)P₂ shows the higher APL values among the three lipid structures. This could be due to the high number of unsaturations present in one of its aliphatic chains (18:0/20:4). Unsaturations make the aliphatic chains less ordered and, in consequence, they occupy more lateral space. The inclusion of penetratin also makes a remarkable difference. PI(4,5)P₂ APL increases notably when the peptide is inserted. This could be a consequence of penetratin PI(4,5)P₂ recruitment. PI(4,5)P₂ lipids enrich around the peptide, concentrating in the same zone of the membrane and increasing the area they occupy. This increase of APL is counter-intuitive, since most of the time amphipathic helices (AH) lower the APL of the neighboring lipids [240]. Here, PI(4,5)P₂ contain many negative charges and probably repel each other explaining this increase of APL.

Lipid	PIP ₂ Simulations		Control Simulations	
	Without Penetratin	With Penetratin	Without Penetratin	With Penetratin
POPC	0.64 +/- 0.0001	0.64 +/- 0.01	0.64 +/- 0.01	0.65 +/- 0.01
DOPS	0.66 +/- 0.0003	0.66 +/- 0.02	0.66 +/- 0.02	0.68 +/- 0.02
PIP ₂	0.68 +/- 0.0005	0.79 +/- 0.05	-	-

Table 7.1 – Membrane APL values calculated from CG-MD simulations.

Studying the thickness of the membrane also gives valuable information about its structure. Membrane thickness is directly related to the APL. Table 7.2 shows that the insertion of penetratin has no effect on membrane thickness. However, PI(4,5)P₂ has a great impact as the membrane thickens when it is included. This effect seems logical as it has longer aliphatic chains (18:0/20:4) compared to POPC (16:0/18:1) or DOPS (18:1) ones. The long aliphatic chain needs more space in the *s* dimension of the membrane, increasing the global thickness. In literature, it has been suggested that that PIP₂ induces a rearrangement and reorganization of the lipids around because it has long, ordered aliphatic chains and because its big polar head interacts with the other lipids, changing the global membrane conformation [140].

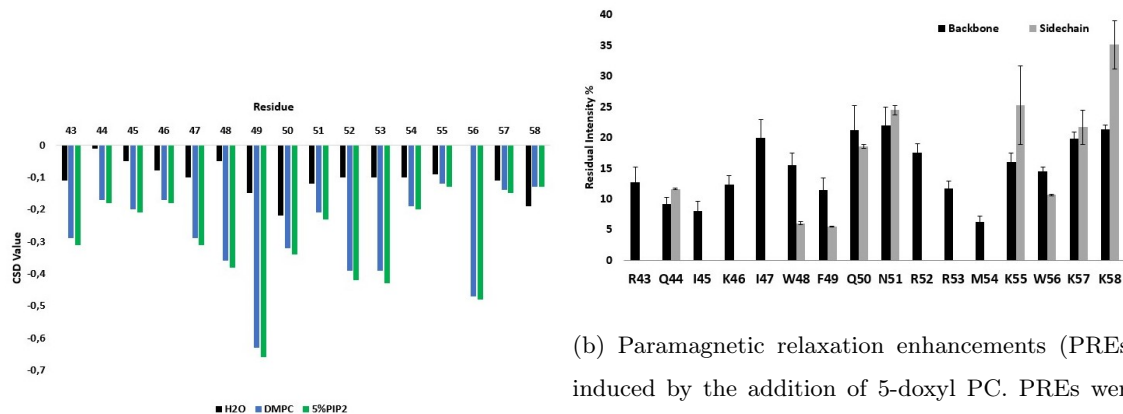
PIP ₂ Simulations		Control Simulations	
Without Penetratin	With Penetratin	Without Penetratin	With Penetratin
4.29 +/- 0.01	4.29 +/- 0.1	4.17 +/- 0.08	4.16 +/- 0.1

Table 7.2 – Membrane thickness values calculated from CG-MD simulations.

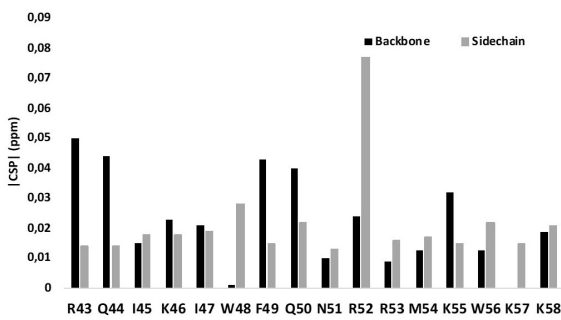
7.1.4 Penetratin structure and partitioning into the membrane

Penetratin behaviour in solution and in membranes was first studied by NMR assays by our collaborators Dr. Olivier Lequin and his Ph.D student Edward Chalouhi. Bicelles were used as membrane mimetics. The ratio between long chain and short chain lipids was set to 0.3 in order to get small isotropically tumbling bicelles compatible with high resolution liquid state NMR studies. Two bicelles compositions were used: zwitterionic bicelles made of 75 *mM* DHPC and 25 *mM* DMPC and anionic bicelles with the same DHPC/DMPC composition and incorporating 5% PI(4,5)P₂. The secondary structure of Penetratin was probed at the residue level by measuring the chemical shift deviations (CSDs) of H α protons. Random coil values were compared to the results obtained for the penetratin in the bicelles. (Fig. 7.1a). Penetratin residues display small negative H α CSDs in aqueous solution (average value of -0.10 ppm), indicative of a weak helical propensity in water. In the presence of zwitterionic bicelles, the CSDs become strongly negative throughout the sequence (average of -0.28 ppm), showing that Penetratin interacts with zwitterionic bicelles and adopts a stable helical structure. The CSDs get even more negative upon the addition of 5% PI(4,5)P₂ (average of -0.30 ppm), indicating that the interaction with PI(4,5)P₂ further stabilizes the helical conformation. NMR structures were calculated using ϕ and ψ dihedral angle restraints based on H α SDs analysis and distance restraints inferred from ¹H-¹H NOEs. Penetratin adopts a well-defined helical structure with numerous van der Waals interactions between *i*/*i*+3 and *i*/*i*+4 residues, as inferred from the NOEs. In order to get information on the position of Penetratin with respect to the lipid bilayer surface, NMR assays with a paramagnetic probe, 5-doxyl-PC were carried out. These assays allow to measure the paramagnetic relaxation enhancements (PREs), which give information about the relative position between the peptide and the bilayer. The paramagnetic tag is buried inside the hydrophobic core of the membrane, on the C5 carbon of the acyl chain, ensuring a positioning close to the bilayer surface. PREs were measured by comparing the intensities of selected cross-peaks on 2D TOCSY spectra in the absence and in the presence of 5-doxyl-PC. The residual intensities are shown for each residue in Figure 7.1b, by selecting protons close to the backbone or protons lying further in side chains. PREs are characterized by a periodic variation delineating two faces within the helix, residues I45, W48, F49, R53 and M54 being strongly affected while residues K46, I47, Q50, N51, R52, K55, K57, K58 are less affected. This periodic profile is in accordance with the regular helical secondary structure of Penetratin and indicates that the helix is positioned parallel to the membrane surface. A small tilt in the helix orientation can be deduced from the larger residual intensities observed in the C-terminal part of Penetratin, suggesting that the N-terminus is buried deeper in the membrane whereas the C-terminal residues are closer to the membrane surface. Accordingly, W56 turns out to be less buried than the central W48 and F49 residues. Finally, interproton NOEs were observed between the aromatic residues of Penetratin and the lipid acyl chains, indicating that the aromatic side chains are embedded within the hydrocarbon core of the membrane.

MD simulations can also provide information on the partitioning of the peptide within the



(a) NMR $^1\text{H}\alpha$ chemical shift deviation (CSD) of Pen- cross-peaks, before and after addition of the paramagnetic probe. Selected cross-peaks involve protons zwitterionic DHPC/DMPC bicelles (blue) and bi-close to the backbone (HN- $\text{H}\alpha$ and HN- $\text{H}\beta$ correlations) or protons lying further in side chains.



(b) Paramagnetic relaxation enhancements (PREs) induced by the addition of 5-doxyl PC. PREs were calculated by comparing the intensities of 2D TOCSY

(c) Chemical shift perturbation (CSP) induced by 5% PI(4,5)P₂ of backbone HN protons (black) or side chain protons (grey). In the case of side chain protons, the maximal observed CSP is reported

Figure 7.1 – NMR results.

membrane (Fig. 7.2a). Regarding the global partitioning of the peptide, we also observe the slight tilt of the α -helix, with the N-terminal residues more buried than the C-terminal ones. Comparing the results in presence or absence of PI(4,5)P₂ lipids, the relative position of the residues in both simulations are very similar. Since the secondary structures are maintained throughout the whole simulations, this indicates that Penetratin has a similar mean partitioning in both conditions. In contrast, we can see higher vertical fluctuations in the presence of PI(4,5)P₂, suggesting a possible influence of this lipid on the partitioning dynamics.

The analysis of side chain partitioning shows that Penetratin helix is far from being perfectly amphipathic. In particular, it can be clearly seen on a helical wheel representation (Fig. 7.2b) that R52 projects its side chain on the hydrophobic face of the helix. Therefore, Penetratin immersion within the membrane leads to a positioning of R52 side chain below the membrane surface. As

In order to get a better description of the interactions of the Penetratin residues with PI(4,5)P₂ lipids at an atomic resolution, we performed a back-mapping of the coarse-grained POPC/POPS/PI(4,5)P₂/Penetratin systems obtained after 1 μ s dynamics into all atoms models. Then AA-MD simulations were further run for a duration of 100 ns. Results of these simulations show that Penetratin adopts a stable helical conformation and a parallel orientation with respect to the membrane, which is similar to CG-simulations (Fig. 7.3a). A slight tilt of the α -helix is also observed, with the N-terminal residues being more buried than the C-terminal ones. These results are in good agreement with NMR PRE data. Penetratin is positioned within the membrane at the level of glycerol groups (Figure 7.3b). The most buried residues are I45, W48 and F49, which interact with the first methylene groups of lipid acyl chains. Most Lys and Arg side chains are found at the level of lipid polar head groups. Interestingly, R52 side chain partitions deeper in the membrane, at the level of glycerol atoms.

The analysis of distances between PI(4,5)P₂ lipids and Penetratin residues shows that hydrogen bonds and electrostatic interactions are favored between the PI(4,5)P₂ phosphate groups and Lys and Arg cationic groups (data not shown). One PI(4,5)P₂ lipid can interact simultaneously with several residues far apart in the peptide sequence, as seen in figure 7.1c. The different partitioning of Lys and Arg residues on the two faces of the peptide leads to specific interactions with PI(4,5)P₂ phosphate groups (Fig. 7.1c). Notably, most Lys and Arg residues interact with the surface exposed phosphate groups in positions 4 and 5 of PI(4,5)P₂. In contrast, R52 interacts exclusively with the more buried phosphate group in position 1 (linked to the glycerol group). The conformational space of R53 enables this residue to interact with phosphate groups of inositol at all positions.

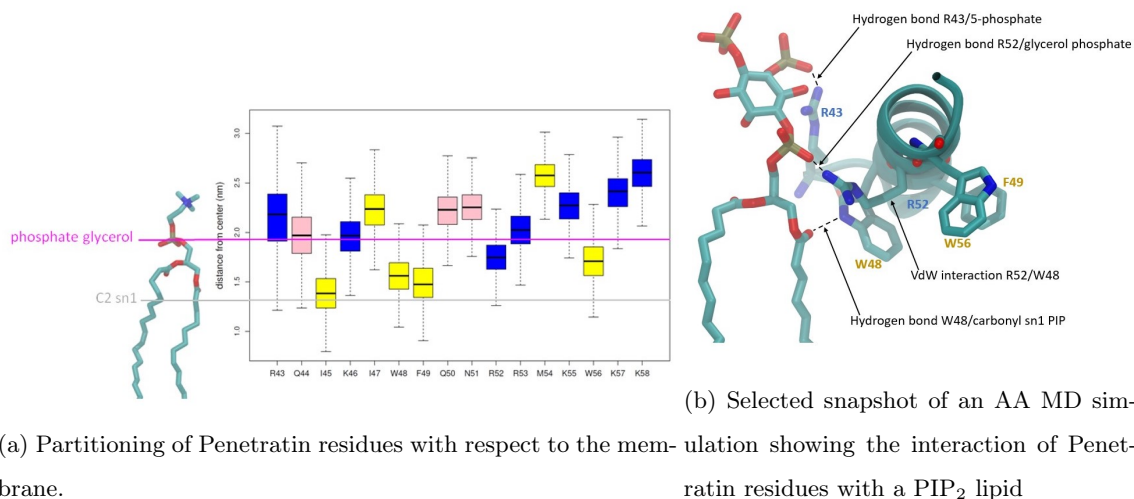


Figure 7.3 – Penetratin/bilayer interactions studied by AA-MD simulations.

7.1.5 PIP₂ recruitment

In this section, we used CG-MD simulations to study PIP₂ recruitment. CG simulations are better suited for this analysis as dynamics are faster, allowing to observe the whole recruitment

process. Two snapshots of the POPC/POPS/PI(4,5)P₂/Penetratin simulation are presented in figure 7.4. They correspond to an upper view of the bilayer taken at 0 and 5416 *ns* of the production phase, showing the initial and final placement of the peptide relative to the lipids. In the system construction, all lipids were placed randomly so that there was no bias in the relative position of the lipids within the system. At 0 *ns* of the production phase, two PI(4,5)P₂ molecules were already located in the Penetratin surroundings. In fact, these two molecules already approached Penetratin during the equilibration phase of 10 *ns*, prior to the production phase. Once the production proceeded, all five PI(4,5)P₂ lipids approached Penetratin, and eventually stayed within its immediate surroundings. This phenomenon happened very quickly (within the first 100 *ns*) and the enrichment was maintained until the end of the simulation. The second snapshot shows the system at 5416 *ns*, where the enrichment of PI(4,5)P₂ around Penetratin can clearly be observed.

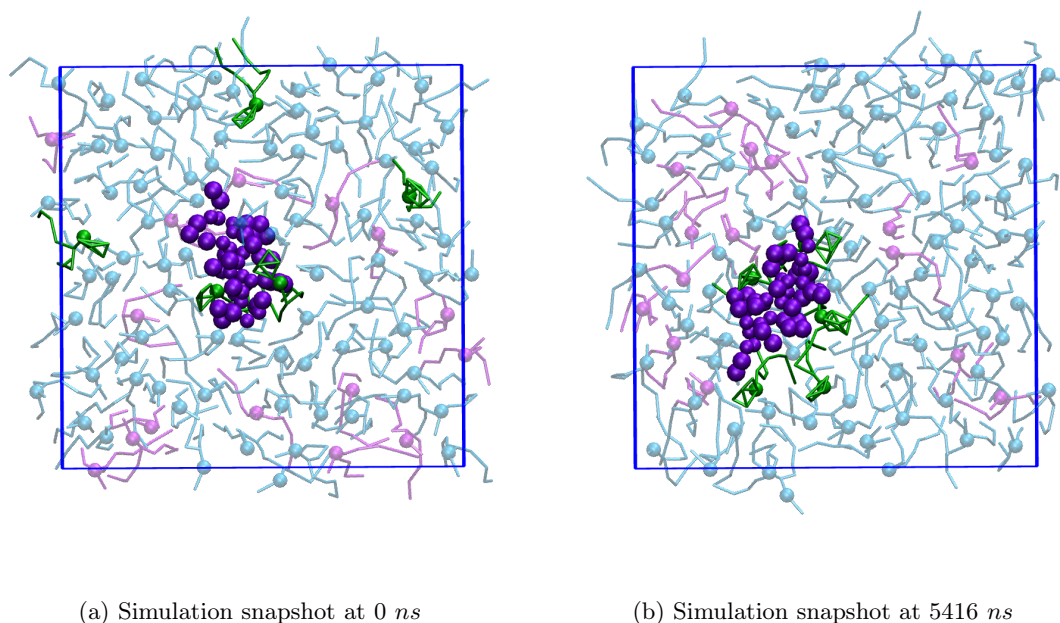
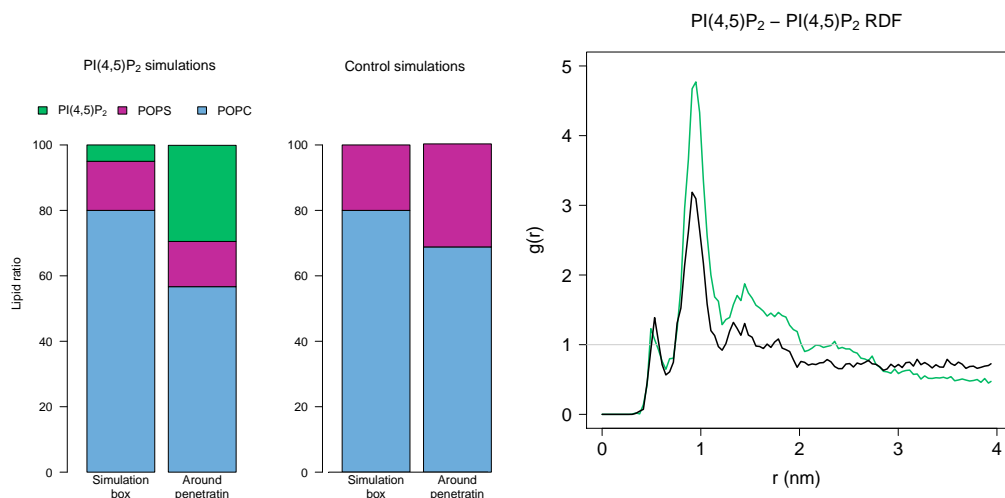


Figure 7.4 – Snapshots of the MD simulation (top view of the bilayer). Penetratin is showed in purple spheres, PC in blue, PS in magenta and PI(4,5)P₂ in green.

To have a quantitative measurement of this clustering, the ratio of each lipid type in the box at the end of the simulation is compared to its ratio around Penetratin in Figure 7.5b. Lipids located at a distance of 6.5 Å or smaller are considered as interacting with Penetratin. The plot on the left shows lipid ratios in POPC/POPS/PI(4,5)P₂/Penetratin simulation. The ratio of PI(4,5)P₂ around Penetratin is 5.9 times higher than the global ratio of PI(4,5)P₂ in the simulation box. The plot on the right shows the same analyses applied to the control POPC/POPS simulation. In the absence of PI(4,5)P₂, Penetratin interacts slightly more with POPS due to its negative charge, though no clear enrichment of PS in its immediate environment is observed.

It is also interesting to analyze the effect of Penetratin on PI(4,5)P₂ to PI(4,5)P₂ relative position. Intuitively, the negative charges carried by PI(4,5)P₂ should prevent these lipids to cluster

together, but this is not the case. Figure 7.5b shows the radial distribution function (RDF) between PI(4,5)P₂ molecules in the presence or absence of Penetratin. This RDF represents the probability of finding another PI(4,5)P₂ molecule as a function of the distance relative to a PI(4,5)P₂ molecule. In the absence of Penetratin, a peak around 1 nm that reaches an RDF value of around 3 can already be observed. It reflects the intrinsic tendency of PI(4,5)P₂ to cluster. This phenomenon has already been observed previously in MARTINI simulations [97] as well as experimentally on giant unilamellar vesicles [184]. Interestingly, this peak at 1 nm is 1.5 times higher when Penetratin is added. This means that PI(4,5)P₂ molecules will be on average closer to each other thanks to the positive charges of Penetratin. All these results show that PI(4,5)P₂ is a privileged interaction partner for Penetratin, and that Penetratin is able to reorganize a lipid bilayer containing low amounts of PI(4,5)P₂ by creating PI(4,5)P₂ clusters.



(b) Radial distribution functions of PI(4,5)P₂
(a) Lipid ratios (PC/PS/PI(4,5)₂ left, PC/PS showing the density of PI(4,5)P₂ as a function control right) at the end of the simulations, in of distance between molecules in the presence the whole simulation box and around Penetratin (green) or absence (black) of Penetratin

Figure 7.5 – Study on penetratin effect on PI(4,5)₂ relative position in the membrane by CG-MD simulations.

7.1.6 Discussion

Penetratin and its interactions with negatively charged membrane partners of diverse molecular nature have been studied for more than 20 years [219]. In particular, negatively charged lipids have been the focus of numerous studies. PG is usually used as a model negatively charged lipid for such studies despite its poor biological relevance. PG is indeed essentially found in bacterial membranes. It is cheap, easy to handle, available with a broad variety of fatty acid chains, and forms stable lamellar structures, which explains its popularity as a simple model lipid. PG-based models allowed important advances in the molecular characterization of Penetratin and

its interactions with membranes. Penetratin interacts preferentially with PG over PC [5, 103, 6], it adopts an α -helical structure in the presence of PG, though an α to β transition can be observed at high charge density or peptide concentration [144]. Penetratin also tends to form peptide-rich regions at the surface of PG liposomes [5]. Finally, PG appears as a permissive lipid for direct translocation across pure lipid membranes [207, 79, 205]. As opposed to PG, PS is a biologically relevant lipid when investigating internalization mechanisms of CPPs in mammalian cells, as it is a major lipid of the inner leaflet of the plasma membrane but it has been more scarcely used in model membranes studies. Penetratin has a higher affinity for PS-containing membranes than just PC, and adopts an α -helical structure in its presence [50]. Regarding direct translocation, the role of PS is not clear, as it was shown to allow direct translocation [207] or not [79], depending on the experimental setup.

Interactions of Penetratin with model membranes incorporating PI(4,5)P₂ have never been studied, though it was shown that direct translocation could occur in liposomes composed of PC, PI and PI-phosphates [207]. In the present study, we sought to characterize the molecular interactions of Penetratin with model membranes incorporating either PI(4,5)P₂ or PS. We showed that Penetratin has a more stable secondary structure when there is PI(4,5)P₂ present in the membrane. The preference for PI(4,5)P₂ over PS could be due to a charge-density effect, or the chemical nature of the moiety carrying the negative charge (phosphate or carboxylate), or the presence of a positive charge on the amine of serine, most probably a combination of these factors. Interestingly, a charge density effect has already been reported as an important factor for binding when studying CPP/GAGs interactions [239, 20, 174].

NMR experiments and MD simulations gave us more detailed molecular insight on the interaction between Penetratin and PI(4,5)P₂-containing membranes. In the presence of PI(4,5)P₂, Penetratin adopts a stabilized α -helical structure. The helix is oriented parallel to the lipid bilayer and is slightly tilted, leading to a more peripheral positioning of the highly cationic C-terminal part (K55, K57, K58). The hydrophobic residues I45, W48, F49 and W56 are anchored within the membrane, as shown by PREs experiments, NOEs and MD simulations. However, Penetratin is not deeply inserted in the membrane as its backbone is positioned at the level of lipid glycerol atoms. This mode of interaction appears to be quite similar to that observed by NMR for a Penetratin/anionic PG bicelle system [137]. A rather shallow insertion of Penetratin had previously been suggested by photocrosslinking experiments using PG model membranes [21, 103].

Such shallow insertion and parallel orientation has also been reported for (R/W)9, a CPP derived from Penetratin [220]. As revealed by MD simulations, R52 is indeed the most buried cationic residue. The localisation of R52 within the hydrophobic sector of Penetratin helix may explain the small depth of Penetratin insertion as the burial of a charged group in the hydrophobic interior of the membrane would be energetically unfavorable. The analysis of all atoms simulations also reveals a peculiar mode of interaction of R52 with PI(4,5)P₂ lipids, in contrast to other cationic residues. Indeed, R52 side chain interacts exclusively with the phosphate group of inositol

in position 1 whereas the other Lys and Arg interact with the phosphate groups in positions 4 and 5 of inositol, which lie closer to the surface. Trp has been recognized as an amino-acid playing a key role in the cell uptake of homeodomain-derived peptides as well as Arg and Trp-rich CPPs. Beside its favorable partitioning properties at water/lipid interfaces, Trp may establish cation- π interactions, as proposed for Penetratin [15] and for RW9 peptide [4]. More recently, Trp has been involved in ion pair- π interactions, which consist in energetically favorable tripartite interactions between the aromatic indole ring, a cationic guanidinium group and a negatively charged moiety of the ligand [80]. The NMR structures reveal that R52 side chain makes van der Waals interactions with W48, as evidenced by several $i/i+4$ interproton NOEs. The strong shielding observed for the methylenic $\beta\gamma\delta$ protons of R52 can be ascribed to an aromatic ring current effect and also supports a favored interaction with W48 indole group. Nevertheless, the residue contact analysis in MD trajectories indicates that cation- π or ion pair- π interactions may form transiently but are not persistent over the whole simulations. Figure 7.3b shows that W48 indole group may stabilize the interaction between R52 and PI(4,5)P₂ not only by establishing favorable van der Waals interactions with R52 side chain but also by donating a hydrogen bond to the carbonyl group of lipid acyl chain.

When considering the effect of Penetratin on the organization of the membrane, our MD simulations show that it can modify the lateral organization of PI(4,5)P₂ -containing membranes by recruiting PI(4,5)P₂ (Fig. 7.4 and 7.5). Previous DSC-based studies showed that Penetratin could recruit cardiolipin and that it induced lateral partitioning in liposomes composed of DPPC and cardiolipin [106]. Cationic peptides have been found to induce PI(4,5)P₂ clustering in model membranes [82]. Experiments using model membranes and analysis by fluorescence spectroscopy suggest that basic peptides are able to sequester multivalent PI(4,5)P₂ but not monovalent PS, supposing that the density of charge of the PI(4,5)P₂ polar heads is a key factor for such interactions and PI(4,5)P₂ clustering.

Recruitment of negatively charged lipids at the cell surface and CPP accumulation in these domains could be a key factor for CPP internalization. The propensity of PI(4,5)P₂ to form clusters in the plasma membrane has been evidenced through binding with specific proteins [42, 74, 30]. These proteins usually contain clusters of basic residues that serve as binding site to interact electrostatically with acidic lipids of the plasma membrane, and in particular with PI(4,5)P₂ polar head groups, recruiting several PI(4,5)P₂ to form clusters. One example of PI(4,5)P₂ lateral segregation is the one induced by the HIV-1 Gag protein, responsible for the assembly of HIV-1 particles, through its matrix domain (MA). Coarse-grained simulations of the HIV-1 myristoylated MA interaction with the membrane revealed that PI(4,5)P₂ head group binds with a highly basic region (HBR) of the domain via interactions with arginine and lysine residues [42] and suggests that specific interaction between the PI(4,5)P₂ polar head and the HBR motif leads to lateral segregation of PI(4,5)P₂ in the membrane. It was latter shown that Gag protein self-assembly would drive the formation of PI(4,5)P₂ /cholesterol clusters, rather than assembling on pre-existing domains

[227]. Altogether, these results strongly suggest that CPPs, as cationic sequences, could be able to recruit PI(4,5)P₂ through electrostatic interactions and induce PI(4,5)P₂ clustering, or amplify its natural tendency to cluster on its own [184].

In conclusion, our study clearly identifies PI(4,5)P₂ as a major effector of Penetratin internalization. Our results suggest different contributions of the intra- and extra-cellular pools of PI(4,5)P₂. Penetratin can drive the formation of PI(4,5)P₂ clusters but how PI(4,5)P₂ contributes to Penetratin internalization remains to be determined.

Chapter 8

General conclusions

This Ph.D project is focused on the study of AH and their interaction mechanism with membranes. It has two main axes, the methodology development and biological research. We first focused on MD techniques, we aimed to study different FF to determine their capacity in reproducing AH / membrane properties and to optimize new protocols to study their action mechanism. From a biological point of view, we wanted to answer several questions about AH. We wanted to study their conformational landscape, their partitioning / folding mechanism in lipid bilayers and to extract kinetic and thermodynamical information of the process. We were also interested in studying the membrane properties that impact AH affinity for the bilayer, such as the electrostatic charges or the curvature. In general terms, this work represent an advance for MD techniques to study AH / membrane systems and we have obtained novel information about their action mechanism that could be used in biomedicine and in drug design.

In chapter 3 dealing with FF comparison, we showed that REMD is an appropriate technique for validating AH / membrane FF. FF accuracy is essential for MD reliability. Having FF tests is essential for the evolution of the field. The capacity of reproducing AH structural properties in the membrane was assessed for four different FF combinations: GROMOS54A7/Berger, OPLS/Berger and AMBER99SB-ILDN/Berger and CHARMM36m/CHARMM36 for peptide/lipids (in the following we will simplify CHARMM36m/CHARMM36 to CHARMM36). The results were conclusive, GROMOS54A7/Berger combination fails to reproduce neither mastoparan or ALPS23 experimental data. OPLS/Berger simulated mastoparan structural properties correctly but failed with a weak helix like ALPS. We determined that CHARMM36m and AMBER99SB-ILDN are the two best FF to reproduce AH structural properties in the membrane. However, CHARMM36m slightly overestimates the helix content, even at high temperatures, where the two peptides maintain the helicity content. Mastoparan thermostability has already been proved by experimental assays [152], but ALPS motif is denatured at high temperatures [216]. This should be taken into account for future projects when extracting structural information from CHARMM36m simulations.

In chapter 4, we presented the mastoparan insertion. We used T-REMD simulations to success-

fully insert mastoparan in the membrane with atomic details. CHARMM36m and AMBER99SB-ILDN/Berger simulations reproduce accurately the AH structural properties and the partition/folded mechanism: the peptide is mainly unfolded in water and folds into a completely-folded α -helix while partitioning in the membrane. The results agree with AH model proposed by S. White [222, 223]. Again, CHARMM36m FF slightly overestimates the helicity content. However, it is the most reliable FF for lipid simulations [34] and contains the largest ensemble of molecules. For these reasons, we decided to use CHARMM36m as a preferred FF to study AH / membrane molecular systems.

In chapter 5, we presented a novel strategy that Patrick and I designed to study AH dynamics by combining T-REMD simulations with MSM construction. First, the REMD simulations served to explore in depth the conformational landscape of the system. Some structures were selected to launch classical MD simulations at 300 *K* to construct an MSM. This way, we constructed a non-biased dynamic model of the system from which we can obtain thermodynamical information at physiological temperature. Two models were constructed: the mastoparan in water and the mastoparan in the membrane. As expected, the mastoparan in water is mainly unfolded, with a 20% of helix in water, as saw by CD assays. The free energy of folding in water is positive, meaning that the folding is disfavored. The major conformation of the model in membrane is a perfectly folded α -helix similar to the NMR-resolved structure [91]. The free energy of folding is negative, meaning that the folding is favored. The estimation of the exact value of these free energies is ongoing. Comparing the kinetics between the two models, the dynamics of the peptide in water is much faster than in the membrane. In the membrane, peptide movement is restricted by the complicated the lipid polar heads and the aliphatic chains creating some friction, where there is the aqueous phase peptide motion is more free. The conformational energy barriers in the membrane are higher than in water. Eventually, this novel strategy will allow us to extract structural, kinetic and thermodynamic information of a AH / membrane system that we expect in agreement with experimental assays.

In chapter 6, we combined experimental and computational assays to construct a first model of LTX-315 dynamics with atomic details. CD and fluorescence assays and MD simulations carried out by myself, NMR assays performed by O. Lequin and E. Chalouhi, and internalization assays done by F. Illien were presented. Experimental and computational assays agree, LTX-315 behaves like an AH, it is unfolded in water and folds into an α -helix when interacting with the lipid bilayer. It has a higher affinity for anionic membranes. This project allowed us to classify LTX-315 as part of the CPP family, a group of peptides capable of interacting directly with the bilayer and penetrating into the cell, preferably, by endocytosis but also by direct translocation. Internalisation assays allowed us to quantify how much the two pathways are used at low and physiological temperatures. The presence of a non-standard amino acid, the Dip, has been a challenge for this project. Some REMD simulations are still ongoing and new results will be shortly accumulated before the defense. To sum up, experimental and computational results agree, the Dip stays unfolded independently from the environment.

Chapter 7 presents a collaboration that took place in our laboratory, headed by A. Warrant and E. Sanchon, where we studied Penetratin behaviour in the membrane and its interaction with PIP₂ lipids. Only the MD simulations that I carried out and the NMR spectra done by O. Lequin and E. Chalouhi were presented. The results demonstrate that Penetratin can drive the formation of PI(4,5)P₂ clusters and clearly identifies PI(4,5)P₂ as a major effector of Penetratin internalization. These two last chapters highlight the importance of combining experimental and computational assays. They are complementary and by combining them we get to explore in depth the molecular system of interest.

In the future, we would like to continue developing our novel strategy and construct MSM of the whole partitioning / folding mechanism. The most efficient and promising strategy would be to use a technique called transition-based re-weighting analysis method (TRAM) developed by F. Noé's group [226]. This technique allows to construct an MSM from a combination of classical MD simulations and T-REMD simulations, helping to enhance the sampling. This new protocol could be used to study LTX-315 to better understand its action mechanism and use this knowledge to improve its development as drug to treat sarcoma disease or to study similar oncolytic peptides of the same family.

Bibliography

- [1] Mark James 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>.
- [2] Paulo Almeida and Antje Pokorny. ‘Mechanisms of Antimicrobial, Cytolytic, and Cell-Penetrating Peptides: From Kinetics to Thermodynamics’. In: *Biochemistry* 48 (Sept. 2009), pp. 8083–93. DOI: 10.1021/bi900914g.
- [3] Paulo F. Almeida, Alexey S. Ladokhin and Stephen H. White. ‘Hydrogen-bond energetics drive helix formation in membrane interfaces’. In: *Biochimica et Biophysica Acta (BBA) - Biomembranes* 1818.2 (2012). Membrane protein structure and function, pp. 178–182. ISSN: 0005-2736. DOI: <https://doi.org/10.1016/j.bbamem.2011.07.019>. URL: <https://www.sciencedirect.com/science/article/pii/S0005273611002240>.
- [4] Isabel Alves et al. ‘Relationships between Membrane Binding, Affinity and Cell Internalization Efficacy of a Cell-Penetrating Peptide: Penetratin as a Case Study’. In: *PloS one* 6 (Sept. 2011), e24096. DOI: 10.1371/journal.pone.0024096.
- [5] Isabel D. Alves et al. ‘Membrane interaction and perturbation mechanisms induced by two cationic cell penetrating peptides with distinct charge distribution’. In: *Biochimica et Biophysica Acta (BBA) - General Subjects* 1780.7 (2008), pp. 948–959. ISSN: 0304-4165. DOI: <https://doi.org/10.1016/j.bbagen.2008.04.004>. URL: <https://www.sciencedirect.com/science/article/pii/S0304416508000998>.
- [6] Isabel D. Alves et al. ‘Relationships between Membrane Binding, Affinity and Cell Internalization Efficacy of a Cell-Penetrating Peptide: Penetratin as a Case Study’. In: *PLOS ONE* 6.9 (Sept. 2011), pp. 1–10. DOI: 10.1371/journal.pone.0024096. URL: <https://doi.org/10.1371/journal.pone.0024096>.
- [7] Irène Amblard et al. ‘Bidirectional transfer of Engrailed homeoprotein across the plasma membrane requires PIP 2’. In: *Journal of Cell Science* 133 (May 2020), jcs.244327. DOI: 10.1242/jcs.244327.
- [8] Magnus Andersson et al. ‘Conformational States of Melittin at a Bilayer Interface’. In: *Biophysical journal* 104 (Mar. 2013), pp. L12–4. DOI: 10.1016/j.bpj.2013.02.006.

- [9] Celine Anezo et al. ‘Methodological Issues in Lipid Bilayer Simulations’. In: *The Journal of Physical Chemistry B* 107 (Sept. 2003). DOI: 10.1021/jp0348981.
- [10] Bruno Antonny. ‘Mechanisms of Membrane Curvature Sensing’. In: *Annual Review of Biochemistry* 80.1 (2011). PMID: 21438688, pp. 101–123. DOI: 10.1146/annurev-biochem-052809-155121. eprint: <https://doi.org/10.1146/annurev-biochem-052809-155121>. URL: <https://doi.org/10.1146/annurev-biochem-052809-155121>.
- [11] Pavan Auluck, Gabriela Caraveo and Susan Lindquist. ‘Synuclein: Membrane Interactions and Toxicity in Parkinson’s Disease’. In: *Annual review of cell and developmental biology* 26 (Nov. 2010), pp. 211–33. DOI: 10.1146/annurev.cellbio.042308.113313.
- [12] Fatma Avcı, Berna Sariyar Akbulut and Elif Ozkirimli. ‘Membrane Active Peptides and Their Biophysical Characterization’. In: *Biomolecules* 8 (Aug. 2018), p. 77. DOI: 10.3390/biom8030077.
- [13] A. Bacle. ‘Etude in silico des gouttelettes lipidiques et de leur interaction avec des protéines périphériques via des hélices amphipathiques’. In: 2016.
- [14] Amélie Bacle et al. ‘Inverse Conformational Selection in Lipid–Protein Binding’. In: *Journal of the American Chemical Society* XXXX (Aug. 2021). DOI: 10.1021/jacs.1c05549.
- [15] Stéphane Balayssac et al. ‘Comparison of Penetratin and Other Homeodomain-Derived Cell-Penetrating Peptides: Interaction in a Membrane-Mimicking Environment and Cellular Uptake Efficiency’. In: *Biochemistry* 45.5 (2006). PMID: 16445283, pp. 1408–1420. DOI: 10.1021/bi0518390. eprint: <https://doi.org/10.1021/bi0518390>. URL: <https://doi.org/10.1021/bi0518390>.
- [16] Amy Baxter, Ivan Poon and Mark Hulett. ‘The plant defensin NaD1 induces tumor cell death via a non-apoptotic, membranolytic process’. In: *Cell Death Discovery* 3 (Jan. 2017), p. 16102. DOI: 10.1038/cddiscovery.2016.102.
- [17] Amy Baxter et al. ‘The Tomato Defensin TPP3 Binds Phosphatidylinositol (4,5)-Bisphosphate via a Conserved Dimeric Cationic Grip Conformation To Mediate Cell Lysis’. In: *Molecular and Cellular Biology* 35 (June 2015), pp. 1964–1978. DOI: 10.1128/MCB.00282-15.
- [18] Kyle A. Beauchamp et al. ‘MSMBuilder2: Modeling Conformational Dynamics on the Pico-second to Millisecond Scale’. In: *Journal of Chemical Theory and Computation* 7.10 (2011). PMID: 22125474, pp. 3412–3419. DOI: 10.1021/ct200463m. eprint: <https://doi.org/10.1021/ct200463m>. URL: <https://doi.org/10.1021/ct200463m>.
- [19] Cherine Bechara and Sandrine Sagan. ‘Cell-penetrating peptides: 20years later, where do we stand?’ In: *FEBS letters* 587 (May 2013). DOI: 10.1016/j.febslet.2013.04.031.
- [20] Chérine Bechara et al. ‘Tryptophan within basic peptide sequences triggers glycosaminoglycan-dependent endocytosis’. In: *The FASEB Journal* 27.2 (2013), pp. 738–749. DOI: <https://doi.org/10.1096/fj.12-216176>. eprint: <https://faseb.onlinelibrary.wiley.com/>

- doi/pdf/10.1096/fj.12-216176. URL: <https://faseb.onlinelibrary.wiley.com/doi/abs/10.1096/fj.12-216176>.
- [21] Leïla Bechtella et al. ‘Benzophenone Photoreactivity in a Lipid Bilayer To Probe Peptide/Membrane Interactions: Simple System, Complex Information’. In: *Analytical Chemistry* 91.14 (2019). PMID: 31251038, pp. 9102–9110. DOI: 10.1021/acs.analchem.9b01584. eprint: <https://doi.org/10.1021/acs.analchem.9b01584>. URL: <https://doi.org/10.1021/acs.analchem.9b01584>.
 - [22] Tristan Bereau et al. ‘Folding and insertion thermodynamics of the transmembrane WALP peptide’. In: *The Journal of Chemical Physics* 143.24 (2015), p. 243127. DOI: 10.1063/1.4935487. eprint: <https://doi.org/10.1063/1.4935487>. URL: <https://doi.org/10.1063/1.4935487>.
 - [23] Herman Berendsen et al. ‘Molecular-Dynamics with Coupling to An External Bath’. In: *The Journal of Chemical Physics* 81 (Oct. 1984), p. 3684. DOI: 10.1063/1.448118.
 - [24] O. Berger, O. Edholm and F. Jähnig. ‘Molecular dynamics simulations of a fluid bilayer of dipalmitoylphosphatidylcholine at full hydration, constant pressure, and constant temperature’. In: *Biophysical Journal* 72.5 (1997), pp. 2002–2013. ISSN: 0006-3495. DOI: [https://doi.org/10.1016/S0006-3495\(97\)78845-3](https://doi.org/10.1016/S0006-3495(97)78845-3). URL: <https://www.sciencedirect.com/science/article/pii/S0006349597788453>.
 - [25] Herman Bergwerf. ‘MolView: an attempt to get the cloud into chemistry classrooms’. In: Dec. 2015.
 - [26] J. Bigay et al. ‘ArfGAP1 responds to membrane curvature through the folding of a lipid packing sensor motif.’ In: *EMBO Journal* 24 (2005), pp. 2244–2253. URL: <https://hal.archives-ouvertes.fr/hal-00090227>.
 - [27] Joelle Bigay et al. ‘Lipid packing sensed by ArfGAP1 couples COPI coat disassembly to membrane bilayer curvature’. In: *Nature* 426 (Jan. 2004), pp. 563–6. DOI: 10.1038/nature02108.
 - [28] Mithun Biswas, Benjamin Lickert and Gerhard Stock. ‘Metadynamics Enhanced Markov Modeling of Protein Dynamics.’ In: *The journal of physical chemistry. B* 122 21 (2018), pp. 5508–5514.
 - [29] Neil W Blackstone. ‘The Cell: A Molecular Approach. Fourth Edition. By Geoffrey M Cooper and Robert E Hausman.’ In: *The Quarterly Review of Biology* 82.1 (2007), pp. 44–44. DOI: 10.1086/513338. eprint: <https://doi.org/10.1086/513338>. URL: <https://doi.org/10.1086/513338>.
 - [30] Geert Bogaart et al. ‘Membrane protein sequestering by ionic protein-lipid interactions’. In: *Nature* 479 (Nov. 2011), pp. 552–5. DOI: 10.1038/nature10545.

- [31] Peter Bond, Chze Wee and Mark Sansom. ‘Coarse-Grained Molecular Dynamics Simulations of the Energetics of Helix Insertion into a Lipid Bilayer’. In: *Biochemistry* 47 (Nov. 2008), pp. 11321–31. DOI: 10.1021/bi800642m.
- [32] Peter J. Bond et al. ‘Coarse-grained molecular dynamics simulations of membrane proteins and peptides’. In: *Journal of Structural Biology* 157.3 (2007). Advances in Molecular Dynamics Simulations, pp. 593–605. ISSN: 1047-8477. DOI: <https://doi.org/10.1016/j.jsb.2006.10.004>. URL: <https://www.sciencedirect.com/science/article/pii/S1047847706003108>.
- [33] Antonella Borrelli et al. ‘Cell Penetrating Peptides as Molecular Carriers for Anti-Cancer Agents’. In: *Molecules* 23.2 (2018). ISSN: 1420-3049. DOI: 10.3390/molecules23020295. URL: <https://www.mdpi.com/1420-3049/23/2/295>.
- [34] Alexandru Botan et al. ‘Toward Atomistic Resolution Structure of Phosphatidylcholine Headgroup and Glycerol Backbone at Different Ambient Conditions’. In: *The journal of physical chemistry. B* 119 (Oct. 2015). DOI: 10.1021/acs.jpcb.5b04878.
- [35] Nicolae-Viorel Buchete and Gerhard Hummer. ‘Peptide Folding Kinetics from Replica Exchange Molecular Dynamics’. In: *Physical review. E, Statistical, nonlinear, and soft matter physics* 77 (Apr. 2008), p. 030902. DOI: 10.1103/PhysRevE.77.030902.
- [36] Sébastien Buchoux. ‘FATSLiM: a fast and robust software to analyze MD simulations of membranes’. In: *Bioinformatics* 33.1 (Aug. 2016), pp. 133–134. ISSN: 1367-4803. DOI: 10.1093/bioinformatics/btw563. eprint: https://academic.oup.com/bioinformatics/article-pdf/33/1/133/25140303/btw563_supp.pdf. URL: <https://doi.org/10.1093/bioinformatics/btw563>.
- [37] Phillipe Bulet et al. ‘Antimicrobial peptides in insects; structure and function’. In: *Developmental and Comparative Immunology* 23.4 (1999). Invertebrate Immunity, pp. 329–344. ISSN: 0145-305X. DOI: [https://doi.org/10.1016/S0145-305X\(99\)00015-4](https://doi.org/10.1016/S0145-305X(99)00015-4). URL: <https://www.sciencedirect.com/science/article/pii/S0145305X99000154>.
- [38] Giovanni Bussi, Davide Donadio and Michele Parrinello. ‘Canonical sampling through velocity rescaling’. In: *The Journal of Chemical Physics* 126.1 (2007), p. 014101. DOI: 10.1063/1.2408420. eprint: <https://doi.org/10.1063/1.2408420>. URL: <https://doi.org/10.1063/1.2408420>.
- [39] Peter Butikofer et al. ‘Transbilayer distribution and mobility of phosphatidylinositol in human red blood cells’. In: *The Journal of biological chemistry* 265 (Oct. 1990), pp. 16035–8. DOI: 10.1016/S0021-9258(17)46183-7.
- [40] Laura Cascales et al. ‘Identification and Characterization of a New Family of Cell-penetrating Peptides’. In: *Journal of Biological Chemistry* 286 (Oct. 2011), pp. 36932–36943.

- [41] Yong S. Chang et al. ‘Stapled α -helical peptide drug development: A potent dual inhibitor of MDM2 and MDMX for p53-dependent cancer therapy’. In: *Proceedings of the National Academy of Sciences* 110.36 (2013), E3445–E3454. ISSN: 0027-8424. DOI: 10.1073/pnas.1303002110. eprint: <https://www.pnas.org/content/110/36/E3445.full.pdf>. URL: <https://www.pnas.org/content/110/36/E3445>.
- [42] Landry Charlier et al. ‘Coarse-Grained Simulations of the HIV-1 Matrix Protein Anchoring: Revisiting Its Assembly on Membrane Domains’. In: *Biophysical Journal* 106 (Feb. 2014), pp. 577–585. DOI: 10.1016/j.bpj.2013.12.019.
- [43] Charles H Chen et al. ‘Understanding and modelling the interactions of peptides with membranes: from partitioning to self-assembly’. In: *Current Opinion in Structural Biology* 61 (2020). Theory and Simulation Macromolecular Assemblies, pp. 160–166. ISSN: 0959-440X. DOI: <https://doi.org/10.1016/j.sbi.2019.12.021>. URL: <https://www.sciencedirect.com/science/article/pii/S0959440X19301605>.
- [44] Charles H. Chen et al. ‘Absorption and folding of melittin onto lipid bilayer membranes via unbiased atomic detail microsecond molecular dynamics simulation’. In: *Biochimica et Biophysica Acta (BBA) - Biomembranes* 1838.9 (2014). Interfacially active peptides and proteins, pp. 2243–2249. ISSN: 0005-2736. DOI: <https://doi.org/10.1016/j.bbamem.2014.04.012>. URL: <https://www.sciencedirect.com/science/article/pii/S0005273614001473>.
- [45] Qiuyan Chen et al. ‘The Rhodopsin-Arrestin-1 Interaction in Bicelles’. In: *Methods in molecular biology (Clifton, N.J.)* 1271 (Feb. 2015), pp. 77–95. DOI: 10.1007/978-1-4939-2330-4_6.
- [46] Alan Chetwynd et al. ‘The Energetics of Transmembrane Helix Insertion into a Lipid Bilayer’. In: *Biophysical Journal* 99.8 (2010), pp. 2534–2540. ISSN: 0006-3495. DOI: <https://doi.org/10.1016/j.bpj.2010.08.002>. URL: <https://www.sciencedirect.com/science/article/pii/S0006349510009689>.
- [47] See-Wing Chiu et al. ‘Incorporation of surface tension into molecular dynamics simulation of an interface: A fluid phase lipid bilayer membrane’. In: *Biophysical journal* 69 (Nov. 1995), pp. 1230–45. DOI: 10.1016/S0006-3495(95)80005-6.
- [48] John D. Chodera et al. ‘Dynamical reweighting: Improved estimates of dynamical properties from simulations at multiple temperatures’. In: *The Journal of Chemical Physics* 134.24 (2011), p. 244107. DOI: 10.1063/1.3592152. eprint: <https://doi.org/10.1063/1.3592152>. URL: <https://doi.org/10.1063/1.3592152>.
- [49] Markus Christen et al. ‘The GROMOS software for biomolecular simulation: GROMOS05’. In: *Journal of Computational Chemistry* 26.16 (2005), pp. 1719–1751. DOI: <https://doi.org/10.1002/jcc.20303>. eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1002/jcc.20303>. URL: <https://onlinelibrary.wiley.com/doi/abs/10.1002/jcc.20303>.

- [50] Bart Christiaens et al. ‘Tryptophan fluorescence study of the interaction of penetratin peptides with model membranes’. In: *European Journal of Biochemistry* 269.12 (2002), pp. 2918–2926. DOI: <https://doi.org/10.1046/j.1432-1033.2002.02963.x>. eprint: <https://febs.onlinelibrary.wiley.com/doi/pdf/10.1046/j.1432-1033.2002.02963.x>. URL: <https://febs.onlinelibrary.wiley.com/doi/abs/10.1046/j.1432-1033.2002.02963.x>.
- [51] Arnau Cordomi, Gianluigi Caltabiano and Leonardo Pardo. ‘Membrane Protein Simulations Using AMBER Force Field and Berger Lipid Parameters’. In: *Journal of Chemical Theory and Computation* 8 (Feb. 2012), pp. 948–958. DOI: 10.1021/ct200491c.
- [52] Rosemary B. Cornell. ‘Membrane lipid compositional sensing by the inducible amphipathic helix of CCT’. In: *Biochimica et Biophysica Acta (BBA) - Molecular and Cell Biology of Lipids* 1861.8, Part B (2016). The Cellular Lipid Landscape, pp. 847–861. ISSN: 1388-1981. DOI: <https://doi.org/10.1016/j.bbalip.2015.12.022>. URL: <https://www.sciencedirect.com/science/article/pii/S1388198115002449>.
- [53] Wendy D. Cornell et al. ‘A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules’. In: *Journal of the American Chemical Society* 117 (1995), pp. 5179–5197.
- [54] Anastasia Croitoru et al. ‘Additive CHARMM36 Force Field for Nonstandard Amino Acids’. In: *Journal of Chemical Theory and Computation* 17.6 (June 2021), pp. 3554–3570. DOI: 10.1021/acs.jctc.1c00254. URL: <https://hal.archives-ouvertes.fr/hal-03363116>.
- [55] Haosheng Cui, Edward Lyman and Gregory A. Voth. ‘Mechanism of Membrane Curvature Sensing by Amphipathic Helix Containing Proteins’. In: *Biophysical Journal* 100.5 (2011), pp. 1271–1279. ISSN: 0006-3495. DOI: <https://doi.org/10.1016/j.bpj.2011.01.036>. URL: <https://www.sciencedirect.com/science/article/pii/S0006349511001196>.
- [56] James Danielli and Hugh Davson. ‘A contribution to the theory of permeability of thin films’. In: *Journal of Cellular and Comparative Physiology* 5 (Feb. 2005), pp. 495–508. DOI: 10.1002/jcp.1030050409.
- [57] Tom Darden, Darrin York and Lee Pedersen. ‘Particle mesh Ewald: An Nlog(N) method for Ewald sums in large systems’. In: *The Journal of Chemical Physics* 98.12 (1993), pp. 10089–10092. DOI: 10.1063/1.464397. eprint: <https://doi.org/10.1063/1.464397>. URL: <https://doi.org/10.1063/1.464397>.
- [58] D Derossi et al. ‘Derossi, D., Joliot, A. H., Chassaing, G. and Prochiantz, A. The third helix of Antennapedia homeodomain translocates through biological membranes. J. Biol. Chem. 269, 10444-10450’. In: *The Journal of biological chemistry* 269 (May 1994), pp. 10444–50.
- [59] Peter Deuffhard and Marcus Weber. ‘Robust Perron cluster analysis in conformation dynamics’. In: *Linear Algebra and its Applications* 398 (2005). Special Issue on Matrices and Mathematical Biology, pp. 161–184. ISSN: 0024-3795. DOI: <https://doi.org/10.1016/>

- j.laa.2004.10.026. URL: <https://www.sciencedirect.com/science/article/pii/S0024379504004689>.
- [60] Callum Dickson et al. ‘Lipid14: The Amber Lipid Force Field’. In: *Journal of chemical theory and computation* 10 (Feb. 2014), pp. 865–879. DOI: 10.1021/ct4010307.
 - [61] Christine Doucet and Martin Hetzer. ‘Nuclear pore biogenesis into an intact nuclear envelope’. In: *Chromosoma* 119 (Oct. 2010), pp. 469–77. DOI: 10.1007/s00412-010-0289-2.
 - [62] Guillaume Drin and Bruno Antonny. ‘Amphipathic helices and membrane curvature’. In: *FEBS Letters* 584.9 (2010), pp. 1840–1847. DOI: <https://doi.org/10.1016/j.febslet.2009.10.022>. eprint: <https://febs.onlinelibrary.wiley.com/doi/pdf/10.1016/j.febslet.2009.10.022>. URL: <https://febs.onlinelibrary.wiley.com/doi/abs/10.1016/j.febslet.2009.10.022>.
 - [63] Guillaume Drin et al. ‘A general amphipathic -helical motif for sensing membrane curvature’. In: *Nature structural and molecular biology* 14 (Mar. 2007), pp. 138–46. DOI: 10.1038/nsmb1194.
 - [64] Ron Dror et al. ‘Perspectives on: Molecular dynamics and computational methods: Exploring atomic resolution physiology on a femtosecond to millisecond timescale using molecular dynamics simulations’. In: *The Journal of general physiology* 135 (June 2010), pp. 555–62. DOI: 10.1085/jgp.200910373.
 - [65] Liv-Marie Eike et al. ‘The oncolytic peptide LTX-315 induces cell death and DAMP release by mitochondria distortion in human melanoma cells’. In: *Oncotarget* 6 (2015), pp. 34910–34923.
 - [66] David Eisenberg, Robert Weiss and Thomas Terwilliger. ‘The helical hydrophobic moment: A measure of the amphiphilicity of a helix’. In: *Nature* 299 (Oct. 1982), pp. 371–4. DOI: 10.1038/299371a0.
 - [67] Ulrich Essmann et al. ‘A smooth particle mesh Ewald method’. In: *The Journal of Chemical Physics* 103.19 (1995), pp. 8577–8593. DOI: 10.1063/1.470117. eprint: <https://doi.org/10.1063/1.470117>. URL: <https://doi.org/10.1063/1.470117>.
 - [68] P. P. Ewald. *Die Berechnung optischer und elektrostatischer Gitterpotentiale*. Jan. 1921. DOI: 10.1002/andp.19213690304. URL: <https://doi.org/10.1002/andp.19213690304>.
 - [69] Michael Feig, Wonpil Im and Charles L. Brooks. ‘Implicit solvation based on generalized Born theory in different dielectric environments’. In: *The Journal of Chemical Physics* 120.2 (2004), pp. 903–911. DOI: 10.1063/1.1631258. eprint: <https://doi.org/10.1063/1.1631258>. URL: <https://doi.org/10.1063/1.1631258>.
 - [70] S.E. Feller and Alexander MacKerell. ‘An improved empirical potential energy function for molecular simulations of phospholipids’. In: *Journal of Physical Chemistry B* 104 (Aug. 2000), pp. 7510–7515.

- [71] Jacques Ferlay et al. 'Cancer statistics for the year 2020: An overview'. In: *International Journal of Cancer* 149.4 (2021), pp. 778–789. DOI: <https://doi.org/10.1002/ijc.33588>. eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1002/ijc.33588>. URL: <https://onlinelibrary.wiley.com/doi/abs/10.1002/ijc.33588>.
- [72] Stephen Fox et al. 'Conformational Transitions of Melittin between Aqueous and Lipid Phases: Comparison of Simulations with Experiments'. In: *The Journal of Physical Chemistry B* 122 (Aug. 2018). DOI: 10.1021/acs.jpcc.8b06781.
- [73] Claire François-Martin et al. 'Cooperation of Conical and Polyunsaturated Lipids to Regulate Initiation and Processing of Membrane Fusion'. In: *Frontiers in Molecular Biosciences* 8 (2021), p. 967. ISSN: 2296-889X. DOI: 10.3389/fmolb.2021.763115. URL: <https://www.frontiersin.org/article/10.3389/fmolb.2021.763115>.
- [74] Alok Gambhir et al. 'Electrostatic Sequestration of PIP2 on Phospholipid Membranes by Basic/Aromatic Regions of Proteins'. In: *Biophysical journal* 86 (May 2004), pp. 2188–207. DOI: 10.1016/S0006-3495(04)74278-2.
- [75] Tomas Ganz and Robert I Lehrer. 'Antimicrobial peptides of vertebrates'. In: *Current Opinion in Immunology* 10.1 (1998), pp. 41–44. ISSN: 0952-7915. DOI: [https://doi.org/10.1016/S0952-7915\(98\)80029-0](https://doi.org/10.1016/S0952-7915(98)80029-0). URL: <https://www.sciencedirect.com/science/article/pii/S0952791598800290>.
- [76] Philippe Gascard et al. 'Asymmetric distribution of phosphoinositides and phosphatidic acid in the human erythrocyte membrane'. In: *Biochimica et Biophysica Acta (BBA) - Biomembranes* 1069.1 (1991), pp. 27–36. ISSN: 0005-2736. DOI: [https://doi.org/10.1016/0005-2736\(91\)90100-M](https://doi.org/10.1016/0005-2736(91)90100-M). URL: <https://www.sciencedirect.com/science/article/pii/000527369190100M>.
- [77] Romain Gautier et al. 'HELIQUEST: a web server to screen sequences with specific α -helical properties'. In: *Bioinformatics* 24.18 (July 2008), pp. 2101–2102. ISSN: 1367-4803. DOI: 10.1093/bioinformatics/btn392. eprint: <https://academic.oup.com/bioinformatics/article-pdf/24/18/2101/478292/btn392.pdf>. URL: <https://doi.org/10.1093/bioinformatics/btn392>.
- [78] Romain Gautier et al. 'PackMem: A Versatile Tool to Compute and Visualize Interfacial Packing Defects in Lipid Bilayers'. In: *Biophysical Journal* 115.3 (2018), pp. 436–444. ISSN: 0006-3495. DOI: <https://doi.org/10.1016/j.bpj.2018.06.025>. URL: <https://www.sciencedirect.com/science/article/pii/S0006349518307628>.
- [79] P. Gehan et al. 'Penetratin translocation mechanism through asymmetric droplet interface bilayers'. In: *Biochimica et Biophysica Acta (BBA) - Biomembranes* 1862.11 (2020), p. 183415. ISSN: 0005-2736. DOI: <https://doi.org/10.1016/j.bbamem.2020.183415>. URL: <https://www.sciencedirect.com/science/article/pii/S0005273620302583>.

- [80] P. Gehan et al. ‘Penetratin translocation mechanism through asymmetric droplet interface bilayers’. In: *Biochimica et Biophysica Acta (BBA) - Biomembranes* 1862 (July 2020), p. 183415. DOI: [10.1016/j.bbamem.2020.183415](https://doi.org/10.1016/j.bbamem.2020.183415).
- [81] Paraskevi Gkeka. ‘Molecular dynamics studies of peptide-membrane interactions: insights from coarse-grained models’. PhD thesis. Jan. 2010.
- [82] Urszula Golebiewska et al. ‘Membrane-Bound Basic Peptides Sequester Multivalent (PIP₂), but Not Monovalent (PS), Acidic Lipids’. In: *Biophysical journal* 91 (Aug. 2006), pp. 588–99. DOI: [10.1529/biophysj.106.081562](https://doi.org/10.1529/biophysj.106.081562).
- [83] Wilfred F van Gunsteren et al. ‘Biomolecular simulation: the GROMOS96 manual and user guide’. In: *Vdf Hochschulverlag AG an der ETH Zürich, Zürich* 86 (1996), pp. 1–1044.
- [84] Karen G. Hales. ‘The machinery of mitochondrial fusion, division, and distribution, and emerging connections to apoptosis’. In: *Mitochondrion* 4.4 (2004), pp. 285–308. ISSN: 1567-7249. DOI: <https://doi.org/10.1016/j.mito.2004.05.007>. URL: <https://www.sciencedirect.com/science/article/pii/S1567724904000832>.
- [85] Benjamin A. Hall, Alan P. Chetwynd and Mark S.P. Sansom. ‘Exploring Peptide-Membrane Interactions with Coarse-Grained MD Simulations’. In: *Biophysical Journal* 100.8 (2011), pp. 1940–1948. ISSN: 0006-3495. DOI: <https://doi.org/10.1016/j.bpj.2011.02.041>. URL: <https://www.sciencedirect.com/science/article/pii/S000634951100261X>.
- [86] Bengt Haug et al. ‘Discovery of a 9-mer Cationic Peptide (LTX-315) as a Potential First in Class Oncolytic Peptide’. In: *Journal of Medicinal Chemistry* 59 (Mar. 2016). DOI: [10.1021/acs.jmedchem.5b02025](https://doi.org/10.1021/acs.jmedchem.5b02025).
- [87] Berk Hess. ‘P-LINCS: A Parallel Linear Constraint Solver for Molecular Simulation’. In: *Journal of Chemical Theory and Computation* 4.1 (2008). PMID: 26619985, pp. 116–122. DOI: [10.1021/ct700200b](https://doi.org/10.1021/ct700200b). eprint: <https://doi.org/10.1021/ct700200b>. URL: <https://doi.org/10.1021/ct700200b>.
- [88] R.W Hockney, S.P Goel and J.W Eastwood. ‘Quiet high-resolution computer models of a plasma’. In: *Journal of Computational Physics* 14.2 (1974), pp. 148–158. ISSN: 0021-9991. DOI: [https://doi.org/10.1016/0021-9991\(74\)90010-2](https://doi.org/10.1016/0021-9991(74)90010-2). URL: <https://www.sciencedirect.com/science/article/pii/0021999174900102>.
- [89] Moritz Hoffmann et al. ‘Deeptime: a Python library for machine learning dynamical models from time series data’. In: (Oct. 2021).
- [90] Andrew Hopkins and Colin Groom. ‘The Druggable Genome’. In: *Nature reviews. Drug discovery* 1 (Oct. 2002), pp. 727–30. DOI: [10.1038/nrd892](https://doi.org/10.1038/nrd892).
- [91] Yumiko Hori et al. ‘Interaction of mastoparan with membranes studied by 1H-NMR spectroscopy in detergent micelles and by solid-state 2H-NMR and 15N-NMR spectroscopy in oriented lipid bilayers’. In: *European Journal of Biochemistry* 268.2 (2001), pp. 302–309. DOI: <https://doi.org/10.1046/j.1432-1033.2001.01880.x>. eprint: <https://febs.org>.

- onlinelibrary.wiley.com/doi/pdf/10.1046/j.1432-1033.2001.01880.x. URL: <https://febs.onlinelibrary.wiley.com/doi/abs/10.1046/j.1432-1033.2001.01880.x>.
- [92] Kalina Hristova et al. 'An amphipathic alpha-helix at a membrane interface: a structural study using a novel X-ray diffraction method' Edited by D. C. Rees'. In: *Journal of Molecular Biology* 290.1 (1999), pp. 99–117. ISSN: 0022-2836. DOI: <https://doi.org/10.1006/jmbi.1999.2840>. URL: <https://www.sciencedirect.com/science/article/pii/S0022283699928401>.
 - [93] Jing Huang et al. 'CHARMM36m: an improved force field for folded and intrinsically disordered proteins'. eng. In: *Nature methods* 14.1 (2017), pp. 71–73. ISSN: 1548-7091.
 - [94] Brooke Husic and Vijay Pande. 'Markov State Models: From an Art to a Science'. In: *Journal of the American Chemical Society* 140 (Jan. 2018). DOI: 10.1021/jacs.7b12191.
 - [95] Brooke E. Husic and Vijay S. Pande. 'Markov State Models: From an Art to a Science'. In: *Journal of the American Chemical Society* 140.7 (2018). PMID: 29323881, pp. 2386–2396. DOI: 10.1021/jacs.7b12191. eprint: <https://doi.org/10.1021/jacs.7b12191>. URL: <https://doi.org/10.1021/jacs.7b12191>.
 - [96] Françoise Illien et al. 'Quantitative fluorescence spectroscopy and flow cytometry analyses of cell-penetrating peptides internalization pathways: Optimization, pitfalls, comparison with mass spectrometry quantification'. In: *Scientific Reports* 6 (Nov. 2016). DOI: 10.1038/srep36938.
 - [97] Helgi I. Ingólfsson et al. 'Lipid Organization of the Plasma Membrane'. In: *Journal of the American Chemical Society* 136.41 (2014). PMID: 25229711, pp. 14554–14559. DOI: 10.1021/ja507832e. eprint: <https://doi.org/10.1021/ja507832e>. URL: <https://doi.org/10.1021/ja507832e>.
 - [98] Mehtap Isik et al. *Assessing the accuracy of octanol-water partition coefficient predictions in the SAMPL6 Part II log P Challenge*. Jan. 2020. DOI: 10.1101/2020.01.20.913178.
 - [99] Shingo Ito et al. 'Implementation of replica-exchange umbrella sampling in GAMESS'. In: *Computer Physics Communications* 228 (2018), pp. 152–162. ISSN: 0010-4655. DOI: <https://doi.org/10.1016/j.cpc.2018.01.014>. URL: <https://www.sciencedirect.com/science/article/pii/S0010465518300274>.
 - [100] Matthew Jacobson, George Kaminski and Richard Friesner. 'Force Field Validation Using Protein Side Chain Prediction'. In: *Journal of Physical Chemistry B - J PHYS CHEM B* 106 (Oct. 2002). DOI: 10.1021/jp021564n.
 - [101] Majid Jafari, Faramarz Mehrnejad and Farahnoosh Doustdar. 'Insight into the interactions, residue snorkeling, and membrane disordering potency of a single antimicrobial peptide into different lipid bilayers'. In: *PLOS ONE* 12 (Nov. 2017), e0187216. DOI: 10.1371/journal.pone.0187216.

- [102] Paul Janmey and Paavo Kinnunen. ‘Biophysical Properties of Lipids and Dynamic Membranes’. In: *Trends in cell biology* 16 (Nov. 2006), pp. 538–46. DOI: 10.1016/j.tcb.2006.08.009.
- [103] Chen-Yu Jiao et al. ‘Exploiting Benzophenone Photoreactivity To Probe the Phospholipid Environment and Insertion Depth of the Cell-Penetrating Peptide Penetratin in Model Membranes’. In: *Angewandte Chemie International Edition* 56.28 (2017), pp. 8226–8230. DOI: <https://doi.org/10.1002/anie.201703465>. eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1002/anie.201703465>. URL: <https://onlinelibrary.wiley.com/doi/abs/10.1002/anie.201703465>.
- [104] Chen-Yu Jiao et al. ‘Translocation and Endocytosis for Cell-penetrating Peptide Internalization’. In: *The Journal of biological chemistry* 284 (Oct. 2009), pp. 33957–65. DOI: 10.1074/jbc.M109.056309.
- [105] Sunhwan Jo et al. ‘CHARMM-GUI: A web-based graphical user interface for CHARMM’. In: *Journal of Computational Chemistry* 29.11 (2008), pp. 1859–1865. DOI: <https://doi.org/10.1002/jcc.20945>. eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1002/jcc.20945>. URL: <https://onlinelibrary.wiley.com/doi/abs/10.1002/jcc.20945>.
- [106] Pierre Joanne et al. ‘Lipid reorganization induced by membrane-active peptides probed using differential scanning calorimetry’. In: *Biochimica et Biophysica Acta (BBA) - Biomembranes* 1788.9 (2009), pp. 1772–1781. ISSN: 0005-2736. DOI: <https://doi.org/10.1016/j.bbamem.2009.05.001>. URL: <https://www.sciencedirect.com/science/article/pii/S000527360900145X>.
- [107] Marie-lise Jobin et al. ‘Biophysical Insight on the Membrane Insertion of an Arginine-Rich Cell-Penetrating Peptide’. In: *International Journal of Molecular Sciences* 20 (Sept. 2019), p. 4441. DOI: 10.3390/ijms20184441.
- [108] A Joliot et al. ‘Antennapedia homeobox peptide regulates neural morphogenesis’. In: *Proceedings of the National Academy of Sciences* 88.5 (1991), pp. 1864–1868. ISSN: 0027-8424. DOI: 10.1073/pnas.88.5.1864. eprint: <https://www.pnas.org/content/88/5/1864.full.pdf>. URL: <https://www.pnas.org/content/88/5/1864>.
- [109] Djurre H. de Jong et al. ‘Improved Parameters for the Martini Coarse-Grained Protein Force Field’. In: *Journal of Chemical Theory and Computation* 9.1 (2013). PMID: 26589065, pp. 687–697. DOI: 10.1021/ct300646g. eprint: <https://doi.org/10.1021/ct300646g>. URL: <https://doi.org/10.1021/ct300646g>.
- [110] Robbie Joosten et al. ‘A series of PDB related databases for everyday needs’. In: *Nucleic acids research* 39 (Nov. 2010), pp. D411–9. DOI: 10.1093/nar/gkq1105.
- [111] William L. Jorgensen, David S. Maxwell and Julian Tirado-Rives. ‘Development and Testing of the OPLS All-Atom Force Field on Conformational Energetics and Properties of Organic Liquids’. In: *Journal of the American Chemical Society* 118 (1996), pp. 11225–11236.

- [112] William L. Jorgensen and Julian Tirado-Rives. ‘The OPLS [optimized potentials for liquid simulations] potential functions for proteins, energy minimizations for crystals of cyclic peptides and crambin.’ In: *Journal of the American Chemical Society* 110 6 (1988), pp. 1657–66.
- [113] William L. Jorgensen et al. ‘Comparison of simple potential functions for simulating liquid water’. In: *The Journal of Chemical Physics* 79.2 (1983), pp. 926–935. DOI: 10.1063/1.445869. eprint: <https://doi.org/10.1063/1.445869>. URL: <https://doi.org/10.1063/1.445869>.
- [114] Wolfgang Kabsch and Chris Sander. ‘Dictionary of protein secondary structure’. In: *Biopolymers* 22 (Dec. 1983), pp. 2577–2637. DOI: 10.1002/bip.360221211.
- [115] Mahroof Kalathingal et al. ‘Vectorial insertion of a beta-helical peptide into membrane: a theoretical study on polytheonamide B’. In: *Biophysical Journal* 120.21 (2021), pp. 4786–4797. ISSN: 0006-3495. DOI: <https://doi.org/10.1016/j.bpj.2021.09.028>. URL: <https://www.sciencedirect.com/science/article/pii/S0006349521007906>.
- [116] Shiv Kale et al. ‘External Lipid PI3P Mediates Entry of Eukaryotic Pathogen Effectors into Plant and Animal Host Cells’. In: *Cell* 142 (July 2010), pp. 284–95. DOI: 10.1016/j.cell.2010.06.008.
- [117] Peter M. Kasson et al. ‘Ensemble molecular dynamics yields submillisecond kinetics and intermediates of membrane fusion’. In: *Proceedings of the National Academy of Sciences* 103.32 (2006), pp. 11916–11921. ISSN: 0027-8424. DOI: 10.1073/pnas.0601597103. eprint: <https://www.pnas.org/content/103/32/11916.full.pdf>. URL: <https://www.pnas.org/content/103/32/11916>.
- [118] Matilda Katan and Shamshad Cockcroft. ‘Phosphatidylinositol(4,5)bispophosphate: Diverse functions at the plasma membrane’. In: *Essays in Biochemistry* 64 (Aug. 2020). DOI: 10.1042/EBC20200041.
- [119] William Kauffman et al. ‘Mechanism Matters: A Taxonomy of Cell Penetrating Peptides’. In: *Trends in Biochemical Sciences* 40 (Nov. 2015). DOI: 10.1016/j.tibs.2015.10.004.
- [120] Muhammad Khan et al. ‘Capturing Choline-Aromatics Cation-pi Interactions in the MARTINI Force Field’. In: *Journal of Chemical Theory and Computation* XXXX (Feb. 2020). DOI: 10.1021/acs.jctc.9b01194.
- [121] Adree Khondker et al. ‘Carbapenems and Lipid Bilayers: Localization, Partitioning, and Energetics’. In: *ACS Infectious Diseases* 4 (Mar. 2018). DOI: 10.1021/acsinfecdis.7b00156.
- [122] Jeffery B. Klauda et al. ‘An ab initio study on the torsional surface of alkanes and its effect on molecular simulations of alkanes and a DPPC bilayer.’ In: *The journal of physical chemistry. B* 109 11 (2005), pp. 5300–11.

- [123] Jeffery B. Klauda et al. ‘Update of the CHARMM All-Atom Additive Force Field for Lipids: Validation on Six Lipid Types’. In: *The Journal of Physical Chemistry B* 114.23 (2010). PMID: 20496934, pp. 7830–7843. DOI: 10.1021/jp101759q. eprint: <https://doi.org/10.1021/jp101759q>. URL: <https://doi.org/10.1021/jp101759q>.
- [124] Erez Koren and Vladimir Torchilin. ‘Cell-penetrating peptides: Breaking through to the other side’. In: *Trends in molecular medicine* 18 (June 2012), pp. 385–93. DOI: 10.1016/j.molmed.2012.04.012.
- [125] Norbert Kucerka et al. ‘Lipid Bilayer Structure Determined by the Simultaneous Analysis of Neutron and X-Ray Scattering Data’. In: *Biophysical journal* 95 (June 2008), pp. 2356–67. DOI: 10.1529/biophysj.108.132662.
- [126] Norbert Kučerka, Mu-Ping Nieh and John Katsaras. ‘Fluid phase lipid areas and bilayer thicknesses of commonly used phosphatidylcholines as a function of temperature’. In: *Biochimica et Biophysica Acta (BBA) - Biomembranes* 1808.11 (2011), pp. 2761–2771. ISSN: 0005-2736. DOI: <https://doi.org/10.1016/j.bbamem.2011.07.022>. URL: <https://www.sciencedirect.com/science/article/pii/S0005273611002276>.
- [127] Alessandro Laio and Francesco Gervasio. ‘Metadynamics: A method to simulate rare events and reconstruct the free energy in biophysics, chemistry and material science’. In: *Reports on Progress in Physics* 71 (Dec. 2008), p. 126601. DOI: 10.1088/0034-4885/71/12/126601.
- [128] Fung T. Lay et al. ‘Structural and functional characterization of the membrane-permeabilizing activity of *Nicotiana occidentalis* defensin NoD173 and protein engineering to enhance oncolysis’. In: *The FASEB Journal* 33.5 (2019), pp. 6470–6482. DOI: <https://doi.org/10.1096/fj.201802540R>. eprint: <https://faseb.onlinelibrary.wiley.com/doi/pdf/10.1096/fj.201802540R>. URL: <https://faseb.onlinelibrary.wiley.com/doi/abs/10.1096/fj.201802540R>.
- [129] Sarah Lee et al. ‘CHARMM36 united atom chain model for lipids and surfactants’. In: *The journal of physical chemistry. B* 118 (Dec. 2013). DOI: 10.1021/jp410344g.
- [130] Woonghee Lee, Marco Tonelli and John L. Markley. ‘NMRFAM-SPARKY: enhanced software for biomolecular NMR spectroscopy’. In: *Bioinformatics* 31.8 (Dec. 2014), pp. 1325–1327. ISSN: 1367-4803. DOI: 10.1093/bioinformatics/btu830. eprint: <https://academic.oup.com/bioinformatics/article-pdf/31/8/1325/17125651/btu830.pdf>. URL: <https://doi.org/10.1093/bioinformatics/btu830>.
- [131] Robert I Lehrer and Tomas Ganz. ‘Antimicrobial peptides in mammalian and insect host defence’. In: *Current Opinion in Immunology* 11.1 (1999), pp. 23–27. ISSN: 0952-7915. DOI: [https://doi.org/10.1016/S0952-7915\(99\)80005-3](https://doi.org/10.1016/S0952-7915(99)80005-3). URL: <https://www.sciencedirect.com/science/article/pii/S0952791599800053>.
- [132] Mark Lemmon. ‘Membrane recognition by phospholipid-binding domains’. In: *Nature reviews. Molecular cell biology* 9 (Mar. 2008), pp. 99–111. DOI: 10.1038/nrm2328.

- [133] Alison Leonard et al. ‘Developing and Testing of Lipid Force Fields with Applications to Modeling Cellular Membranes’. In: *Chemical Reviews* 119 (Feb. 2019). DOI: 10.1021/acs.chemrev.8b00384.
- [134] Hari Leontiadou, Alan E. Mark and Siewert J. Marrink. ‘Antimicrobial Peptides in Action’. In: *Journal of the American Chemical Society* 128.37 (2006). PMID: 16967965, pp. 12156–12161. DOI: 10.1021/ja062927q. eprint: <https://doi.org/10.1021/ja062927q>. URL: <https://doi.org/10.1021/ja062927q>.
- [135] Qinghua Liao. ‘Chapter Four - Enhanced sampling and free energy calculations for protein simulations’. In: *Computational Approaches for Understanding Dynamical Systems: Protein Folding and Assembly*. Ed. by Birgit Strodel and Bogdan Barz. Vol. 170. Progress in Molecular Biology and Translational Science. Academic Press, 2020, pp. 177–213. DOI: <https://doi.org/10.1016/bs.pmbts.2020.01.006>. URL: <https://www.sciencedirect.com/science/article/pii/S187711732030017X>.
- [136] Shneior Lifson and A. Roig. ‘On the Theory of Helix—Coil Transition in Polypeptides’. In: *The Journal of Chemical Physics* 34.6 (1961), pp. 1963–1974. DOI: 10.1063/1.1731802. eprint: <https://doi.org/10.1063/1.1731802>. URL: <https://doi.org/10.1063/1.1731802>.
- [137] Mattias Lindberg et al. ‘Structure and positioning comparison of two variants of penetratin in two different membrane mimicking systems by NMR’. In: *European Journal of Biochemistry* 270.14 (2003), pp. 3055–3063. DOI: <https://doi.org/10.1046/j.1432-1033.2003.03685.x>. eprint: <https://febs.onlinelibrary.wiley.com/doi/pdf/10.1046/j.1432-1033.2003.03685.x>. URL: <https://febs.onlinelibrary.wiley.com/doi/abs/10.1046/j.1432-1033.2003.03685.x>.
- [138] Göran Lindblom. ‘Ole G. Mouritsen. ‘Life – as a matter of fat’. The emerging science of lipidomics. Berlin: Springer-Verlag; Heidelberg: GmbH and Co. K. ISBN: 3-540-23248-6’. In: *Chemistry and Physics of Lipids - CHEM PHYS LIPIDS* 135 (May 2005), pp. 105–106. DOI: 10.1016/j.chemphyslip.2005.02.005.
- [139] Kresten Lindorff-Larsen et al. ‘Improved side-chain torsion potentials for the Amber ff99SB protein force field’. In: *Proteins: Structure, Function, and Bioinformatics* 78.8 (2010), pp. 1950–1958. DOI: <https://doi.org/10.1002/prot.22711>. eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1002/prot.22711>. URL: <https://onlinelibrary.wiley.com/doi/abs/10.1002/prot.22711>.
- [140] Dmitry Lupyan et al. ‘A Molecular Dynamics Investigation of Lipid Bilayer Perturbation by PIP2’. In: *Biophysical journal* 98 (Jan. 2010), pp. 240–7. DOI: 10.1016/j.bpj.2009.09.063.
- [141] Vamparys Lydie et al. ‘Conical Lipids in Flat Bilayers Induce Packing Defects Similar to that Induced by Positive Curvature’. In: *Biophysical journal* 104 (Feb. 2013), pp. 585–93. DOI: 10.1016/j.bpj.2012.11.3836.

- [142] Alexander D. MacKerell et al. ‘All-atom empirical potential for molecular modeling and dynamics studies of proteins.’ In: *The journal of physical chemistry. B* 102 18 (1998), pp. 3586–616.
- [143] J. B. MacQueen. ‘Some Methods for Classification and Analysis of MultiVariate Observations’. In: *Proc. of the fifth Berkeley Symposium on Mathematical Statistics and Probability*. Ed. by L. M. Le Cam and J. Neyman. Vol. 1. University of California Press, 1967, pp. 281–297.
- [144] Mazin Magzoub, L.E.Göran Eriksson and Astrid Gräslund. ‘Conformational states of the cell-penetrating peptide penetratin when interacting with phospholipid vesicles: effects of surface charge and peptide concentration’. In: *Biochimica et Biophysica Acta (BBA) - Biomembranes* 1563.1 (2002), pp. 53–63. ISSN: 0005-2736. DOI: [https://doi.org/10.1016/S0005-2736\(02\)00373-5](https://doi.org/10.1016/S0005-2736(02)00373-5). URL: <https://www.sciencedirect.com/science/article/pii/S0005273602003735>.
- [145] Pekka Mark and Lennart Nilsson. ‘Structure and dynamics of the TIP3P, SPC, and SPC/E water models at 298 K’. In: *Journal of Physical Chemistry A - J PHYS CHEM A* 105 (Nov. 2001). DOI: 10.1021/jp003020w.
- [146] Siewert J. Marrink et al. ‘The MARTINI Force Field: Coarse Grained Model for Biomolecular Simulations’. In: *The Journal of Physical Chemistry B* 111.27 (2007). PMID: 17569554, pp. 7812–7824. DOI: 10.1021/jp071097f. eprint: <https://doi.org/10.1021/jp071097f>. URL: <https://doi.org/10.1021/jp071097f>.
- [147] Stuart Mclaughlin et al. ‘PIP 2 and Proteins: Interactions, Organization, and Information Flow’. In: *Annual review of biophysics and biomolecular structure* 31 (Feb. 2002), pp. 151–75. DOI: 10.1146/annurev.biophys.31.082901.134259.
- [148] Gerrit Meer, Dennis Voelker and Gerald Feigenson. ‘Membrane lipids: Where they are and how they behave’. In: *Nature reviews. Molecular cell biology* 9 (Mar. 2008), pp. 112–24. DOI: 10.1038/nrm2330.
- [149] Manuel Melo, Rafael Ferre and Miguel Castanho. ‘Antimicrobial peptides: Linking partition, activity and high membrane-bound concentrations’. In: *Nature reviews. Microbiology* 7 (Apr. 2009), pp. 245–50. DOI: 10.1038/nrmicro2095.
- [150] ‘Membranes—An Introduction’. In: *Thermal Biophysics of Membranes*. John Wiley and Sons, Ltd, 2007. Chap. 1, pp. 1–13. ISBN: 9783527611591. DOI: <https://doi.org/10.1002/9783527611591.ch1>. eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1002/9783527611591.ch1>. URL: <https://onlinelibrary.wiley.com/doi/abs/10.1002/9783527611591.ch1>.
- [151] Ayori Mitsutake, Yuji Sugita and Yuko Okamoto. ‘Replica-exchange multicanonical and multicanonical replica-exchange Monte Carlo simulations of peptides. I. Formulation and benchmark test’. In: *The Journal of Chemical Physics* 118.14 (2003), pp. 6664–6675. DOI:

- 10.1063/1.1555847. eprint: <https://doi.org/10.1063/1.1555847>. URL: <https://doi.org/10.1063/1.1555847>.
- [152] Yoshinori Miura. ‘Nuclear magnetic resonance studies on conformation and stability of mastoparan in methanol’. In: *Journal of Peptide Science* 27.9 (2021), e3338. DOI: <https://doi.org/10.1002/psc.3338>. eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1002/psc.3338>. URL: <https://onlinelibrary.wiley.com/doi/abs/10.1002/psc.3338>.
 - [153] Shuichi Miyamoto and Peter A. Kollman. ‘Settle: An analytical version of the SHAKE and RATTLE algorithm for rigid water models’. In: *Journal of Computational Chemistry* 13.8 (1992), pp. 952–962. DOI: <https://doi.org/10.1002/jcc.540130805>. eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1002/jcc.540130805>. URL: <https://onlinelibrary.wiley.com/doi/abs/10.1002/jcc.540130805>.
 - [154] Viviana Monje-Galvan and Jeffery Klauda. ‘Modeling Yeast Organelle Membranes and How Lipid Diversity Influences Bilayer Properties’. In: *Biochemistry* 54 (Oct. 2015). DOI: 10.1021/acs.biochem.5b00718.
 - [155] Luca Monticelli et al. ‘The MARTINI Coarse-Grained Force Field: Extension to Proteins’. In: *Journal of Chemical Theory and Computation* 4.5 (2008). PMID: 26621095, pp. 819–834. DOI: 10.1021/ct700324x. eprint: <https://doi.org/10.1021/ct700324x>. URL: <https://doi.org/10.1021/ct700324x>.
 - [156] Yoshiharu Mori and Yuko Okamoto. ‘Generalized-Ensemble Algorithms for the Isobaric–Isothermal Ensemble’. In: *Journal of the Physical Society of Japan* 79.7 (2010), p. 074003. DOI: 10.1143/JPSJ.79.074003. eprint: <https://doi.org/10.1143/JPSJ.79.074003>. URL: <https://doi.org/10.1143/JPSJ.79.074003>.
 - [157] Brigitte Nuscher et al. ‘ α -Synuclein Has a High Affinity for Packing Defects in a Bilayer Membrane’. In: *The Journal of biological chemistry* 279 (June 2004), pp. 21966–75. DOI: 10.1074/jbc.M401076200.
 - [158] Hugh Nymeyer, Thomas B. Woolf and Angel E. Garcia. ‘Folding is not required for bilayer insertion: Replica exchange simulations of an alpha-helical peptide with an explicit lipid bilayer’. In: *Proteins: Structure, Function, and Bioinformatics* 59.4 (2005), pp. 783–790. DOI: <https://doi.org/10.1002/prot.20460>. eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1002/prot.20460>. URL: <https://onlinelibrary.wiley.com/doi/abs/10.1002/prot.20460>.
 - [159] Myriam Oubrai et al. ‘alpha-synuclein senses lipid packing defects and induces lateral expansion of lipids leading to membrane remodeling’. In: *Journal of Biological Chemistry* (June 2013). DOI: 10.1074/jbc.M111.478297.
 - [160] M. Parrinello and A. Rahman. ‘Polymorphic transitions in single crystals: A new molecular dynamics method’. In: *Journal of Applied Physics* 52.12 (1981), pp. 7182–7190. DOI: 10.

- 1063/1.328693. eprint: <https://doi.org/10.1063/1.328693>. URL: <https://doi.org/10.1063/1.328693>.
- [161] Alexandra Patriksson and David van der Spoel. ‘A temperature predictor for parallel tempering simulations’. In: *Physical chemistry chemical physics : PCCP* 10 (May 2008), pp. 2073–7. DOI: 10.1039/b716554d.
 - [162] Taras Patsahan, J. Ilnytskyi and Orest Pizio. ‘On the properties of a single OPLS-UA model curcumin molecule in water, methanol and dimethyl sulfoxide. Molecular dynamics computer simulation results’. In: *Condensed Matter Physics* 20 (June 2017), p. 23003. DOI: 10.5488/CMP.20.23003.
 - [163] Fabian Paul et al. ‘Protein-peptide association kinetics beyond the seconds timescale from atomistic simulations’. In: *Nature Communications* 8 (Oct. 2017). DOI: 10.1038/s41467-017-01163-6.
 - [164] Robert M. Peitzsch and Stuart McLaughlin. ‘Binding of acylated peptides and fatty acids to phospholipid vesicles: Pertinence to myristoylated proteins’. In: *Biochemistry* 32.39 (1993). PMID: 8399188, pp. 10436–10443. DOI: 10.1021/bi00090a020. eprint: <https://doi.org/10.1021/bi00090a020>. URL: <https://doi.org/10.1021/bi00090a020>.
 - [165] Guillermo Pérez-Hernández et al. ‘Identification of slow molecular order parameters for Markov model construction’. In: *The Journal of Chemical Physics* 139.1 (2013), p. 015102. DOI: 10.1063/1.4811489. eprint: <https://doi.org/10.1063/1.4811489>. URL: <https://doi.org/10.1063/1.4811489>.
 - [166] Horia I. Petrache, Steven W. Dodd and Michael F. Brown. ‘Area per Lipid and Acyl Length Distributions in Fluid Phosphatidylcholines Determined by 2H NMR Spectroscopy’. In: *Biophysical Journal* 79.6 (2000), pp. 3172–3192. ISSN: 0006-3495. DOI: [https://doi.org/10.1016/S0006-3495\(00\)76551-9](https://doi.org/10.1016/S0006-3495(00)76551-9). URL: <https://www.sciencedirect.com/science/article/pii/S0006349500765519>.
 - [167] Horia I. Petrache, Kechuan Tu and John F. Nagle. ‘Analysis of Simulated NMR Order Parameters for Lipid Bilayer Structure Determination’. In: *Biophysical Journal* 76.5 (1999), pp. 2479–2487. ISSN: 0006-3495. DOI: [https://doi.org/10.1016/S0006-3495\(99\)77403-5](https://doi.org/10.1016/S0006-3495(99)77403-5). URL: <https://www.sciencedirect.com/science/article/pii/S0006349599774035>.
 - [168] Horia I. Petrache et al. ‘Structure and Fluctuations of Charged Phosphatidylserine Bilayers in the Absence of Salt’. In: *Biophysical Journal* 86.3 (2004), pp. 1574–1586. ISSN: 0006-3495. DOI: [https://doi.org/10.1016/S0006-3495\(04\)74225-3](https://doi.org/10.1016/S0006-3495(04)74225-3). URL: <https://www.sciencedirect.com/science/article/pii/S0006349504742253>.
 - [169] Frédéric Pincet et al. ‘FRAP to Characterize Molecular Diffusion and Interaction in Various Membrane Environments’. In: *PLOS ONE* 11 (July 2016), e0158457. DOI: 10.1371/journal.pone.0158457.

- [170] Nuria Plattner et al. ‘Complete protein–protein association kinetics in atomic detail revealed by molecular dynamics simulations and Markov modelling’. In: *Nature Chemistry* 9 (June 2017). DOI: 10.1038/nchem.2785.
- [171] Ivan Poon et al. ‘Phosphoinositide-mediated oligomerization of a defensin induces cell lysis’. In: *eLife* 3 (Apr. 2014), e01808. DOI: 10.7554/eLife.01808.
- [172] Iwona M. Pranke et al. ‘alpha-Synuclein and ALPS motifs are membrane curvature sensors whose contrasting chemistry mediates selective vesicle binding’. In: *Journal of Cell Biology* 194.1 (July 2011), pp. 89–103. ISSN: 0021-9525. DOI: 10.1083/jcb.201011118. eprint: <https://rupress.org/jcb/article-pdf/194/1/89/1351826/jcb\201011118.pdf>. URL: <https://doi.org/10.1083/jcb.201011118>.
- [173] William H. Press et al. *Numerical Recipes 3rd Edition: The Art of Scientific Computing*. 3rd ed. USA: Cambridge University Press, 2007. ISBN: 0521880688.
- [174] Lisa Prevett et al. ‘Cell-penetrating compounds preferentially bind glycosaminoglycans over plasma membrane lipids in a charge density- and stereochemistry-dependent manner’. In: *Biophysical chemistry* 207 (Sept. 2015), pp. 40–50. DOI: 10.1016/j.bpc.2015.08.003.
- [175] Jan-Hendrik Prinz, Bettina Keller and Frank Noé. ‘Probing molecular kinetics with Markov models: Metastable states, transition pathways and spectroscopic observables’. In: *Physical chemistry chemical physics : PCCP* 13 (Aug. 2011), pp. 16912–27. DOI: 10.1039/c1cp21258c.
- [176] Zahid Raja et al. ‘Insight into the mechanism of action of temporin-SHa, a new broad-spectrum antiparasitic and antibacterial agent’. In: *PLOS ONE* 12.3 (Mar. 2017), pp. 1–41. DOI: 10.1371/journal.pone.0174024. URL: <https://doi.org/10.1371/journal.pone.0174024>.
- [177] Fabienne Rayne et al. ‘Phosphatidylinositol-(4,5)-bisphosphate enables efficient secretion of HIV-1 Tat by infected T-cells’. In: *The EMBO journal* 29 (Mar. 2010), pp. 1348–62. DOI: 10.1038/emboj.2010.32.
- [178] Lauren M. Reid, Chandra S. Verma and Jonathan W. Essex. ‘The role of molecular simulations in understanding the mechanisms of cell-penetrating peptides’. In: *Drug Discovery Today* 24.9 (2019), pp. 1821–1835. ISSN: 1359-6446. DOI: <https://doi.org/10.1016/j.drudis.2019.06.013>. URL: <https://www.sciencedirect.com/science/article/pii/S1359644618305294>.
- [179] Charles Roach et al. ‘Comparison of Cis and Trans Fatty Acid Containing Phosphatidylcholines on Membrane Properties †’. In: *Biochemistry* 43 (June 2004), pp. 6344–51. DOI: 10.1021/bi049917r.
- [180] Susanna Roblitz and Marcus Weber. ‘Fuzzy spectral clustering by PCCA+: Application to Markov state models and data classification’. In: *Advances in Data Analysis and Classification* 7 (June 2013). DOI: 10.1007/s11634-013-0134-6.

- [181] Susanna Röblitz and Marcus Weber. ‘Fuzzy spectral clustering by PCCA+: Application to Markov state models and data classification’. In: *Advances in Data Analysis and Classification* 7 (June 2013). DOI: 10.1007/s11634-013-0134-6.
- [182] Hanna Rydberg et al. ‘Effects of Tryptophan Content and Backbone Spacing on Uptake Efficiency of Cell-Penetrating Peptides’. In: *Biochemistry* 51 (June 2012), pp. 5531–9. DOI: 10.1021/bi300454k.
- [183] Lynn Sagan. ‘On the origin of mitosing cells’. In: *Journal of Theoretical Biology* 14.3 (1967), 225–IN6. ISSN: 0022-5193. DOI: [https://doi.org/10.1016/0022-5193\(67\)90079-3](https://doi.org/10.1016/0022-5193(67)90079-3). URL: <https://www.sciencedirect.com/science/article/pii/0022519367900793>.
- [184] Iyri L. Salvemini et al. ‘Low PIP2 molar fractions induce nanometer size clustering in giant unilamellar vesicles’. In: *Chemistry and Physics of Lipids* 177 (2014), pp. 51–63. ISSN: 0009-3084. DOI: <https://doi.org/10.1016/j.chemphyslip.2013.11.003>. URL: <https://www.sciencedirect.com/science/article/pii/S0009308413001552>.
- [185] Angelica Sandoval, Kristyna Pluhackova and Rainer Böckmann. ‘A Critical Comparison of Biomembrane Force Fields: Protein-Lipid Interactions at the Membrane Interface’. In: *Journal of chemical theory and computation* 13 (Apr. 2017). DOI: 10.1021/acs.jctc.7b00001.
- [186] Angelica Sandoval, Kristyna Pluhackova and Rainer Böckmann. ‘A Critical Comparison of Biomembrane Force Fields: Protein-Lipid Interactions at the Membrane Interface’. In: *Journal of chemical theory and computation* 13 (Apr. 2017). DOI: 10.1021/acs.jctc.7b00001.
- [187] Patricia Santofimia-Castaño et al. ‘Amphipathic helical peptides hamper protein-protein interactions of the intrinsically disordered chromatin nuclear protein 1 (NUPR1)’. In: *Biochimica et Biophysica Acta (BBA) - General Subjects* 1862.6 (2018), pp. 1283–1295. ISSN: 0304-4165. DOI: <https://doi.org/10.1016/j.bbagen.2018.03.009>. URL: <https://www.sciencedirect.com/science/article/pii/S0304416518300710>.
- [188] Nicolas Sapay and D. Tieleman. ‘Chapter 4 Molecular Dynamics Simulation of Lipid-Protein Interactions’. In: *Current Topics in Membranes* 60 (Dec. 2008), pp. 111–130. DOI: 10.1016/S1063-5823(08)00004-5.
- [189] Nicolas Sapay and D. Peter Tieleman. ‘Combination of the CHARMM27 force field with united-atom lipid force fields’. In: *Journal of Computational Chemistry* 32.7 (2011), pp. 1400–1410. DOI: <https://doi.org/10.1002/jcc.21726>. eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1002/jcc.21726>. URL: <https://onlinelibrary.wiley.com/doi/abs/10.1002/jcc.21726>.
- [190] Martin K. Scherer et al. ‘PyEMMA 2: A Software Package for Estimation, Validation, and Analysis of Markov Models’. In: *Journal of Chemical Theory and Computation* 11.11 (2015).

- PMID: 26574340, pp. 5525–5542. DOI: 10.1021/acs.jctc.5b00743. eprint: <https://doi.org/10.1021/acs.jctc.5b00743>. URL: <https://doi.org/10.1021/acs.jctc.5b00743>.
- [191] Marianne Schiffer and Allen Edmundson. ‘Use of Helical Wheels to Represent the Structures of Proteins and to Identify Segments with Helical Potential’. In: *Biophysical journal* 7 (Apr. 1967), pp. 121–35. DOI: 10.1016/S0006-3495(67)86579-2.
 - [192] Nathan Schmid et al. ‘Definition and testing of the GROMOS force-field versions 54A7 and 54B7’. In: *European biophysics journal : EBJ* 40 (July 2011), pp. 843–56. DOI: 10.1007/s00249-011-0700-9.
 - [193] Erin Schuler, Sureshbabu Nagarajan and Richard Dyer. ‘Submillisecond Dynamics of Mastoparan X Insertion into Lipid Membranes’. In: *The Journal of Physical Chemistry Letters* 7 (Aug. 2016). DOI: 10.1021/acs.jpclett.6b01512.
 - [194] Ch Schütte et al. ‘A Direct Approach to Conformational Dynamics Based on Hybrid Monte Carlo’. In: *Journal of Computational Physics* 151.1 (1999), pp. 146–168. ISSN: 0021-9991. DOI: <https://doi.org/10.1006/jcph.1999.6231>. URL: <https://www.sciencedirect.com/science/article/pii/S0021999199962319>.
 - [195] Joachim Seelig. ‘Thermodynamics of lipid–peptide interactions’. In: *Biochimica et Biophysica Acta (BBA) - Biomembranes* 1666.1 (2004). Lipid-Protein Interactions, pp. 40–50. ISSN: 0005-2736. DOI: <https://doi.org/10.1016/j.bbamem.2004.08.004>. URL: <https://www.sciencedirect.com/science/article/pii/S0005273604002056>.
 - [196] Jere P. Segrest et al. ‘Amphipathic helix motif: Classes and properties’. In: *Proteins: Structure, Function, and Bioinformatics* 8.2 (1990), pp. 103–117. DOI: <https://doi.org/10.1002/prot.340080202>. eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1002/prot.340080202>. URL: <https://onlinelibrary.wiley.com/doi/abs/10.1002/prot.340080202>.
 - [197] M. Marvin Seibert et al. ‘Reproducible Polypeptide Folding and Structure Prediction using Molecular Dynamics Simulations’. In: *Journal of Molecular Biology* 354.1 (2005), pp. 173–183. ISSN: 0022-2836. DOI: <https://doi.org/10.1016/j.jmb.2005.09.030>. URL: <https://www.sciencedirect.com/science/article/pii/S0022283605010958>.
 - [198] Marc Serulla et al. ‘TAT-RasGAP317-326 kills cells by targeting inner-leaflet-enriched phospholipids’. In: *Proceedings of the National Academy of Sciences* 117 (Nov. 2020). DOI: 10.1073/pnas.2014108117.
 - [199] Guilherme da Silva et al. ‘Are all-atom any better than united-atom force fields for the description of liquid properties of alkanes?’ In: *Journal of Molecular Modeling* 26 (Oct. 2020), p. 296. DOI: 10.1007/s00894-020-04548-5.
 - [200] S.J. Singer and Garth Nicolson. ‘The Fluid Mosaic Model of the Structure of Cell Membranes’. In: *Science (New York, N.Y.)* 175 (Mar. 1972), pp. 720–31. DOI: 10.1126/science.175.4023.720.

- [201] Antonella Sistigu, Gwenola Manic and Ilio Vitale. ‘LTX-315, CAPtivating immunity with necrosis’. In: *Cell Cycle* 15.9 (2016). PMID: 26986689, pp. 1176–1177. DOI: 10.1080/15384101.2016.1160609. eprint: <https://doi.org/10.1080/15384101.2016.1160609>. URL: <https://doi.org/10.1080/15384101.2016.1160609>.
- [202] Wilfred D. Stein and Thomas Litman. ‘Chapter 1 - Structural Basis of Movement Across Cell Membranes’. In: *Channels, Carriers, and Pumps (Second Edition)*. Ed. by Wilfred D. Stein and Thomas Litman. Second Edition. London: Elsevier, 2015, pp. 1–36. ISBN: 978-0-12-416579-3. DOI: <https://doi.org/10.1016/B978-0-12-416579-3.00001-0>. URL: <https://www.sciencedirect.com/science/article/pii/B9780124165793000010>.
- [203] Julia P Steringer et al. ‘Key steps in unconventional secretion of fibroblast growth factor 2 reconstituted with purified components’. In: *eLife* 6 (2017).
- [204] Baldur Sveinbjornsson et al. ‘LTX-315: A first-in-class oncolytic peptide that reprograms the tumor microenvironment’. In: *Future Medicinal Chemistry* 9 (May 2017). DOI: 10.4155/fmc-2017-0088.
- [205] Jean-Marie Swiecicki et al. ‘The Efficacies of Cell-Penetrating Peptides in Accumulating in Large Unilamellar Vesicles Depend on their Ability To Form Inverted Micelles’. In: *Chem-BioChem* 15.6 (2014), pp. 884–891. DOI: <https://doi.org/10.1002/cbic.201300742>. eprint: <https://chemistry-europe.onlinelibrary.wiley.com/doi/pdf/10.1002/cbic.201300742>. URL: <https://chemistry-europe.onlinelibrary.wiley.com/doi/abs/10.1002/cbic.201300742>.
- [206] Jia Tang et al. ‘Role of helix nucleation in the kinetics of binding of mastoparan X to phospholipid bilayers.’ In: *Biochemistry* 46 48 (2007), pp. 13856–63.
- [207] Donato Terrone et al. ‘Penetratin and Related Cell-Penetrating Cationic Peptides Can Translocate Across Lipid Bilayers in the Presence of a Transbilayer Potential’. In: *Biochemistry* 42.47 (2003). PMID: 14636045, pp. 13787–13799. DOI: 10.1021/bi035293y. eprint: <https://doi.org/10.1021/bi035293y>. URL: <https://doi.org/10.1021/bi035293y>.
- [208] T.C. Terwilliger, L. Weissman and D. Eisenberg. ‘Research Article’. English. In: *Biophysical Journal* 37.1 (1982), pp. 353–361. DOI: 10.1016/S0006-3495(82)84683-3.
- [209] D Tieleman et al. ‘Membrane protein simulations with a united-atom lipid and all-atom protein model: Lipid-protein interactions, side chain transfer free energies and model proteins’. In: *Journal of physics. Condensed matter : an Institute of Physics journal* 18 (July 2006), S1221–34. DOI: 10.1088/0953-8984/18/28/S07.
- [210] D.P. Tieleman and H.J.C. Berendsen. ‘A Molecular Dynamics Study of the Pores Formed by Escherichia coli OmpF Porin in a Fully Hydrated Palmitoylphosphatidylcholine Bilayer’. In: *Biophysical Journal* 74.6 (1998), pp. 2786–2801. ISSN: 0006-3495. DOI: [https://doi.org/10.1016/S0006-3495\(98\)77986-X](https://doi.org/10.1016/S0006-3495(98)77986-X). URL: <https://www.sciencedirect.com/science/article/pii/S000634959877986X>.

- [211] Jakob Ulmschneider and Martin Ulmschneider. ‘Molecular Dynamics Simulations Are Redefining Our View of Peptides Interacting with Biological Membranes’. In: *Accounts of chemical research* 51 (Apr. 2018). DOI: 10.1021/acs.accounts.7b00613.
- [212] Martin Ulmschneider et al. ‘Mechanism and Kinetics of Peptide Partitioning into Membranes from All-Atom Simulations of Thermostable Peptides’. In: *Journal of the American Chemical Society* 132 (Feb. 2010), pp. 3452–60. DOI: 10.1021/ja909347x.
- [213] Martin B. Ulmschneider et al. ‘A Generalized Born Implicit-Membrane Representation Compared to Experimental Insertion Free Energies’. In: *Biophysical Journal* 92.7 (2007), pp. 2338–2349. ISSN: 0006-3495. DOI: <https://doi.org/10.1529/biophysj.106.081810>. URL: <https://www.sciencedirect.com/science/article/pii/S0006349507710408>.
- [214] Lydie Vamparys. ‘Exploration de la reconnaissance de la courbure membranaire par le motif ALPS. (Exploring the recognition of membrane curvature by the ALPS motif)’. In: 2013.
- [215] David Van Der Spoel et al. ‘GROMACS: Fast, flexible, and free’. In: *Journal of Computational Chemistry* 26.16 (2005), pp. 1701–1718. DOI: <https://doi.org/10.1002/jcc.20291>. eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1002/jcc.20291>. URL: <https://onlinelibrary.wiley.com/doi/abs/10.1002/jcc.20291>.
- [216] Stefano Vanni et al. ‘Amphipathic Lipid Packing Sensor Motifs: Probing Bilayer Defects with Hydrophobic Residues’. In: *Biophysical journal* 104 (Feb. 2013), pp. 575–84. DOI: 10.1016/j.bpj.2012.11.3837.
- [217] Eric Vivès, Priscille Brodin and Bernard Lebleu. ‘A Truncated HIV-1 Tat Protein Basic Domain Rapidly Translocates through the Plasma Membrane and Accumulates in the Cell Nucleus’. In: *The Journal of biological chemistry* 272 (July 1997), pp. 16010–7. DOI: 10.1074/jbc.272.25.16010.
- [218] Astrid Walrant et al. ‘Membrane Crossing and Membranotropic Activity of Cell-Penetrating Peptides: Dangerous Liaisons ?’ In: *Accounts of Chemical Research* 50 (Nov. 2017). DOI: 10.1021/acs.accounts.7b00455.
- [219] Astrid Walrant et al. ‘Membrane Crossing and Membranotropic Activity of Cell-Penetrating Peptides: Dangerous Liaisons?’ In: *Accounts of Chemical Research* 50.12 (2017). PMID: 29172443, pp. 2968–2975. DOI: 10.1021/acs.accounts.7b00455. eprint: <https://doi.org/10.1021/acs.accounts.7b00455>. URL: <https://doi.org/10.1021/acs.accounts.7b00455>.
- [220] Astrid Walrant et al. ‘Membrane interactions of two arginine-rich peptides with different cell internalization capacities’. In: *Biochimica et Biophysica Acta (BBA) - Biomembranes* 1818.7 (2012), pp. 1755–1763. ISSN: 0005-2736. DOI: <https://doi.org/10.1016/j.bbamem.2012.02.024>. URL: <https://www.sciencedirect.com/science/article/pii/S0005273612000752>.

- [221] Christoph Wehmeyer et al. ‘Introduction to Markov state modeling with the PyEMMA software [Article v1.0]’. In: *Living Journal of Computational Molecular Science* 1 (Jan. 2018). DOI: 10.33011/livecoms.1.1.5965.
- [222] Stephen H. White et al. ‘[4] Protein folding in membranes: Determining energetics of peptide-bilayer interactions’. In: *Energetics of Biological Macromolecules Part B*. Vol. 295. Methods in Enzymology. Academic Press, 1998, pp. 62–87. DOI: [https://doi.org/10.1016/S0076-6879\(98\)95035-2](https://doi.org/10.1016/S0076-6879(98)95035-2). URL: <https://www.sciencedirect.com/science/article/pii/S0076687998950352>.
- [223] William C. Wimley and Stephen H. White. ‘Experimentally determined hydrophobicity scale for proteins at membrane interfaces’. In: *Nature Structural Biology* 3.10 (Oct. 1996), pp. 842–848. ISSN: 1545-9985. DOI: 10.1038/nsb1096-842. URL: <https://doi.org/10.1038/nsb1096-842>.
- [224] Hao Wu and Frank Noé. ‘Variational Approach for Learning Markov Processes from Time Series Data’. In: *Journal of Nonlinear Science* 30 (Feb. 2020). DOI: 10.1007/s00332-019-09567-y.
- [225] Hao Wu et al. ‘Multiensemble Markov models of molecular thermodynamics and kinetics’. In: *Proceedings of the National Academy of Sciences* 113.23 (2016), E3221–E3230. ISSN: 0027-8424. DOI: 10.1073/pnas.1525092113. eprint: <https://www.pnas.org/content/113/23/E3221.full.pdf>. URL: <https://www.pnas.org/content/113/23/E3221>.
- [226] Hao Wu et al. ‘Multiensemble Markov models of molecular thermodynamics and kinetics’. In: *Proceedings of the National Academy of Sciences* 113.23 (2016), E3221–E3230. ISSN: 0027-8424. DOI: 10.1073/pnas.1525092113. eprint: <https://www.pnas.org/content/113/23/E3221.full.pdf>. URL: <https://www.pnas.org/content/113/23/E3221>.
- [227] Naresh Yandrapalli et al. ‘Self assembly of HIV-1 Gag protein on lipid membranes generates PI(4,5)P₂/Cholesterol nanoclusters’. In: *Scientific Reports* (Dec. 2016).
- [228] Lin Yang et al. ‘Crystallization of Antimicrobial Pores in Membranes: Magainin and Pro-tegrin’. In: *Biophysical Journal* 79.4 (2000), pp. 2002–2009. ISSN: 0006-3495. DOI: [https://doi.org/10.1016/S0006-3495\(00\)76448-4](https://doi.org/10.1016/S0006-3495(00)76448-4). URL: <https://www.sciencedirect.com/science/article/pii/S0006349500764484>.
- [229] Michael R. Yeaman and Nannette Y. Yount. ‘Mechanisms of Antimicrobial Peptide Action and Resistance’. In: *Pharmacological Reviews* 55.1 (2003), pp. 27–55. ISSN: 0031-6997. DOI: 10.1124/pr.55.1.2. eprint: <https://pharmrev.aspetjournals.org/content/55/1/27.full.pdf>. URL: <https://pharmrev.aspetjournals.org/content/55/1/27>.
- [230] Tony Yeung et al. ‘Receptor Activation Alters Inner Surface Potential During Phagocytosis’. In: *Science (New York, N.Y.)* 313 (Aug. 2006), pp. 347–51. DOI: 10.1126/science.1129551.

- [231] Atsuko Yoneda et al. ‘Phosphatidylinositol 4,5-bisphosphate is localized in the plasma membrane outer leaflet and regulates cell adhesion and motility’. In: *Biochemical and Biophysical Research Communications* 527.4 (2020), pp. 1050–1056. ISSN: 0006-291X. DOI: <https://doi.org/10.1016/j.bbrc.2020.05.040>. URL: <https://www.sciencedirect.com/science/article/pii/S0006291X20309487>.
- [232] Tomohiro Yorimitsu, Ken Sato and Masaki Takeuchi. ‘Molecular mechanisms of Sar/Arf GTPases in vesicular trafficking in yeast and plants’. In: *Frontiers in plant science* 5 (Aug. 2014), p. 411. DOI: 10.3389/fpls.2014.00411.
- [233] Koji Yoshida, Toshio Yamaguchi and Yuko Okamoto. ‘Replica-exchange molecular dynamics simulation of small peptide in water and in ethanol’. In: *Chemical Physics Letters* 412.4 (2005), pp. 280–284. ISSN: 0009-2614. DOI: <https://doi.org/10.1016/j.cplett.2005.06.114>. URL: <https://www.sciencedirect.com/science/article/pii/S0009261405009632>.
- [234] Yalun Yu and Jeffery Klauda. ‘Update of the CHARMM36 United Atom Chain Model for Hydrocarbons and Phospholipids’. In: *The Journal of Physical Chemistry B* XXXX (July 2020). DOI: 10.1021/acs.jpcb.0c04795.
- [235] Tongtao Yue et al. ‘How transmembrane peptides insert and orientate in biomembranes: A combined experimental and simulation study’. In: *Phys. Chem. Chem. Phys.* 18 (June 2016). DOI: 10.1039/C6CP01133K.
- [236] Michael Zasloff. ‘Antimicrobial peptides of multicellular organisms’. In: *Nature* 415 (Feb. 2002), pp. 389–95. DOI: 10.1038/415389a.
- [237] Huan Zhan and Themis Lazaridis. ‘Inclusion of Lateral Pressure/Curvature Stress Effects in Implicit Membrane Models’. In: *Biophysical journal* 104 (Feb. 2013), pp. 643–54. DOI: 10.1016/j.bpj.2012.12.022.
- [238] Heng Zhou et al. ‘The oncolytic peptide LTX-315 kills cancer cells through Bax/Bak-regulated mitochondrial membrane permeabilization’. In: *Oncotarget* 6 (Sept. 2015). DOI: 10.18632/oncotarget.5613.
- [239] André Ziegler and Joachim Seelig. ‘Binding and Clustering of Glycosaminoglycans: A Common Property of Mono- and Multivalent Cell-Penetrating Compounds’. In: *Biophysical journal* 94 (Apr. 2008), pp. 2142–9. DOI: 10.1529/biophysj.107.113472.
- [240] Bogdan Zorilă et al. ‘Melittin Induces Local Order Changes in Artificial and Biological Membranes as Revealed by Spectral Analysis of Laurdan Fluorescence’. In: *Toxins* 12 (2020).