

Curtin Medical School

**Characterization of molecular interactions of small
molecules with model cell membranes**

Carlo Martinotti

0000-0001-5255-0555

**This thesis is presented for the Degree of
Phd in Biomedical Sciences
of
Curtin University**

November 2021

Declaration

To the best of my knowledge and belief this thesis contains no material previously published by any other person except where due acknowledgment has been made. This thesis contains no material which has been accepted for the award of any other degree or diploma in any university.

Signature:

Date: 30/06/2021.

Contents

| | | |
|-------|---|----|
| 1 | Chapter 1- Literature review..... | 13 |
| 1.1 | Introduction | 13 |
| 1.1.1 | Umbrella sampling | 13 |
| 1.1.2 | Biological Membranes and their interactions with small molecules..... | 15 |
| 1.1.3 | Model membrane systems | 16 |
| 1.1.4 | Challenges in understanding SMMs Molecular dynamics | 22 |
| 1.1.5 | Force fields..... | 23 |
| 1.2 | Free energy Calculations..... | 25 |
| 1.3 | Sampling of SMMs in MD simulations | 27 |
| 1.3.1 | Replica Exchange methods | 28 |
| 1.3.2 | Metadynamics..... | 31 |
| 1.4 | Test systems..... | 34 |
| 1.4.1 | Voltage gated ionic channels | 34 |
| 1.4.2 | ICK peptides | 35 |
| 1.4.3 | Choice of Protx-1 as a model system | 38 |
| 2 | Chapter 2 – Conventional US of Protx-1 binding to a POPC membrane | 40 |
| 2.1 | Introduction | 40 |
| 2.2 | Methods..... | 41 |
| 2.2.1 | Umbrella sampling | 41 |
| 2.2.2 | Branched umbrella sampling | 43 |
| 2.2.3 | Definition of orientation angles of Protx-1 | 44 |
| 2.2.4 | Orientation angle heat maps | 45 |
| 2.2.5 | Correction for angle probability heat maps..... | 45 |
| 2.2.6 | Calculation of the potential of mean force (PMF) | 46 |
| 2.2.7 | Calculation of the area per lipid (APL) | 47 |
| 2.3 | Results..... | 47 |
| 3 | Chapter 3 - Selective replica-exchange with solute tempering (REST3) for the prediction of the interactions of small molecules with lipid bilayers..... | 55 |
| 3.1 | Introduction | 55 |
| 3.1.1 | Theoretical derivation..... | 57 |
| 3.2 | Methods..... | 58 |
| 3.2.1 | Interaction of terpenes with a DMPC bilayer | 59 |
| 3.2.2 | Umbrella sampling | 59 |

| | | |
|-------|--|-----|
| 3.2.3 | US-REST3 | 60 |
| 3.2.4 | Potential of mean force calculations | 60 |
| 3.2.5 | Water-terpene interactions | 61 |
| 3.3 | Results and discussion | 61 |
| 3.3.1 | Free energies of binding | 61 |
| 3.3.2 | Potential of mean force profiles | 62 |
| 3.3.3 | Angle maps of the orientation of terpenes..... | 65 |
| 3.3.4 | Mechanism of binding of terpenes to phospholipid membranes | 71 |
| 3.3.5 | Assessing the efficiency of US-REST3 simulations | 73 |
| 3.4 | Summary and conclusions | 74 |
| 3.5 | Supporting Information | 74 |
| 4 | Chapter 4 – Application of US-REST3 to the Protx-1 – POPC membrane system..... | 84 |
| 4.1 | Intro..... | 84 |
| 4.2 | Methods..... | 85 |
| 4.2.1 | REST3 enhanced umbrella sampling simulation of Protx-1 | 85 |
| 4.2.2 | Potential of mean force calculations | 86 |
| 4.2.3 | Radius of gyration | 86 |
| 4.2.4 | Hydrogen bond analysis..... | 86 |
| 4.2.5 | Water layer analysis..... | 86 |
| 4.3 | Results..... | 87 |
| 4.3.1 | Other scaling regimes | 94 |
| 5 | Chapter 5 - Development of a US-Metadynamics protocol with a small test system | 97 |
| 5.1 | Intro..... | 97 |
| 5.2 | Methods..... | 99 |
| 5.2.1 | WTMetad | 99 |
| 5.2.2 | Systems setup | 100 |
| 5.2.3 | Removal of thermostat outliers..... | 101 |
| 5.2.4 | Reweighting of the simulations | 102 |
| 5.2.5 | Histogram calculation | 102 |
| 5.3 | Results and discussion | 103 |
| 6 | Chapter 6 – Development of a CV for semi-rigid objects rotation | 108 |
| 6.1 | Intro..... | 108 |
| 6.2 | Methods..... | 110 |
| 6.2.1 | Definition of Euler angles..... | 110 |
| 6.2.2 | Parallel bias metadynamics (PBMetad) | 112 |
| 6.2.3 | PBMetaD Simulations | 114 |

| | | |
|-------|--|-----|
| 6.3 | Results and discussion | 115 |
| 7 | Chapter 7 – Application of US-WT-MetaD and PB-MetaD protocols to the interaction of ProTx-I with a POPC membrane | 120 |
| 7.1 | Introduction | 120 |
| 7.2 | Methods..... | 121 |
| 7.2.1 | WT-MetaD..... | 121 |
| 7.2.2 | PB-MetaD | 121 |
| 7.2.3 | Metadynamics with walkers | 122 |
| 7.2.4 | Euler angles..... | 122 |
| 7.2.5 | US-WT-MetaD simulations..... | 122 |
| 7.2.6 | PB-MetaD simulation | 124 |
| 7.2.7 | Reconstruction of the PMF | 124 |
| 7.2.8 | Rotational Autocorrelation Functions..... | 124 |
| 7.3 | Results and Discussion | 125 |
| 7.3.1 | Rotational sampling in US-WT-MetaD | 125 |
| 7.3.2 | PMFs from US-WT-MetaD..... | 128 |
| 7.3.3 | Free energy surface from PB-MetaD | 131 |
| 8 | Conclusions | 135 |
| 8.1 | Summary of methods implemented and tested ‘ | 135 |
| 8.1.1 | Replica exchange with solute tempering for peptide-membrane systems..... | 135 |
| 8.1.2 | Metadynamics for semi-rigid molecules..... | 136 |
| 8.2 | Challenges | 137 |
| 8.2.1 | Dimensionality reduction in membrane systems | 137 |
| 8.2.2 | Replica exchange with solute tempering simulations of membrane systems..... | 139 |
| 8.2.3 | Metadynamics simulations of protein-membrane systems | 140 |
| 9 | Appendix | 143 |

List of Figures

| | |
|---|----|
| Figure 1. Schematic representation of a cell membrane composed of a lipid bilayer with embedded membrane proteins. Reprinted from ⁵ , Copyright 2017 with permission from Elsevier..... | 15 |
| Figure 2. Schematic representation of various types of model membrane systems. Supported bilayers and monolayers can have different kinds of linkers (polymeric, charged, and covalent) connecting the support with the actual bilayer. Licensed under a Creative Commons Attribution (CC BY) license. | 17 |
| Figure 3. Structures of common lipids used to build model membranes to mimic the mammalian plasma cell membrane..... | 18 |
| Figure 4. Lipid phase transition from gel (L_{β}) to fluid phase (L_{α}) induced by an increase in temperature. The gel state is characterised by a smaller area per lipid, larger bilayer thickness and higher lipid tail order compared to the fluid phase. Adapted from ²⁴ . Copyright 2005, American Chemical Society. | 19 |
| Figure 5. On the left, schematic representation of an umbrella sampling simulation of the interaction of a small peptide with a lipid bilayer. The reaction coordinate is the centre-of-mass distance between the molecule in solution (state A) and the molecule at the membrane (state B). On the right, sample plot of the free energy profile, also known as potential of mean force (PMF) for the binding of the peptide to the membrane, providing an estimate of the change in free energy between states A and B. | 13 |
| Figure 6. Schematic representation of the potential energy distribution in classical parallel tempering and solute tempering. Top panel: Hamiltonian tempering can achieve a wider distribution of the energy, such that a good overlap is obtained with a smaller number of replicas. The distribution is larger because the solvent-solvent term, which has a very narrow distribution around its mean value due to the large number of molecules, is cancelled out. Bottom panel: In parallel tempering, because of the narrow distribution of the solvent-solvent energy, the distributions of all replicas are much narrower because of the solvent-solvent term; therefore, a greater number of replicas is required to achieve a relatively small overlap in the energy distributions. | 31 |
| Figure 7. Schematic representation of metadynamics. From left to right repulsive Gaussian energy functions are gradually deposited on previously explored configurations, gradually filling energy wells and enabling exploration of other neighbouring energy wells. A) In the beginning the system is unbiased and the transition between the two wells is difficult because of the barriers. B) Gaussian potential functions start to be accumulated on the free energy surface. C) In the end, the resulting energy landscape is flat and sampling in the system becomes random in the collective variable space. | 32 |
| Figure 8 A: General structure of a VGIC., 4 voltage sensing domains typical of one specific ion channel surrounds the pore domain. The pore allows for the passage of ions through the hydrophobic membrane. B: A close up of the structure of the pore domain. | 35 |
| Figure 9 A) Schematic 2D representation of ICK peptide and their secondary structure. In green, the 2 of the disulfide bonds that form the ring, in red the threading one. B) 3D representation of the previous structure..... | 36 |
| Figure 10 Illustration of the membrane accumulation mechanism. In the first phase the peptide goes from the solution on the membrane, shifting from a 3D diffusion to a 2D diffusion. Once on the membrane it can reorient in the proper direction and bind to the channel at an enhanced rate. | 37 |
| Figure 11. Snapshot of the simulation cell prior to solvation..... | 42 |

| | |
|--|----|
| Figure 12 Setup of the branched umbrella sampling simulation. At 4.0 nm the main branch diverges from the top most simulation (described in section 2.1) and three new simulations with different orientations of ProTx-1 with respect to the membrane were performed. | 44 |
| Figure 13 Vectors that define θ and ψ . Vector u in red, vector v in purple and vector k in green | 45 |
| Figure 14. Time evolution of the protein RMSD during the US simulations. The values of RMSD are on average well under 0.2 nm ² , indicating that a stable structure of the protein was achieved after the NPT equilibration. A: RMSD for window at 5.0 nm. B: RMSD for window at 2.4 nm | 48 |
| Figure 15. Predicted PMF along the centre of mass z distance for the interaction of ProtX-1 with a POPC bilayer..... | 49 |
| Figure 16 Orientation angle heat maps for ProtX-1 for windows centred at 2.4 nm and at 5.0 nm ... | 50 |
| Figure 17. Predicted PMFs along the centre of mass z distance for the interaction of ProtX-1 with a POPC bilayer for the four branched US simulations. | 51 |
| Figure 18. Time evolution of the thickness and APL of a POPC membrane during the simulation of a pure POPC bilayer in water. | 52 |
| Figure 19 Orientational angle heat maps for the four different US branched simulations calculated at COM distances $2.4 < dz < 2.8$ | 53 |
| Figure 20. A) Chemical structures of the terpenes: limonene (LIM), peril aldehyde (PALD) and perillidic acid (DPAC). B) The angle ϕ and θ are defined between the two continuous line vectors (blue continuous vector along the principal axis of the terpene molecule and green continuous vector perpendicular to the plane of the terpene ring) and the z-axis corresponding to the normal vector to the membrane. | 59 |
| Figure 21. PMF for the interaction of the three terpenes PALD (orange), LIM (blue) and DPAC (green) with a DMPC bilayer obtained from conventional US simulations (panel A) and US-REST3 simulations (panel B). The regions of the simulation cell are indicated starting from the bulk solution (d), the membrane surface extending from 1.6 to 1.9 nm (c), the high-density hydrophobic tail region extending from 1.6 nm to 0.25 nm (b) and the low density tail region extending to 0.25 nm in the center of the membrane (a). | 63 |
| Figure 22. Mass density profile of a fully hydrated DMPC bilayer. The thickness d of the bilayer (3.54 nm) was calculated as the distance between the peaks of the nitrogen atoms of the DMPC molecules. The surface of the membrane is centered at around d/2 (1.77 nm) from the center of the bilayer. The choline group can extend up to 1.0 nm away from this surface. | 64 |
| Figure 23. Orientation angle heat maps of the terpene molecules in conventional US simulations. Orientation angle maps are 2D probability plots of the θ and ϕ angles that define the orientation of the terpene molecules with respect to the vector normal to the membrane surface. Angle θ is reported on the y-axes and angle ϕ on the x-axis, as represented in the bottom right panel. These angle maps were used to assess the sampling of orientations visited by the terpene molecule in representative windows at three different COM distances to the bilayer. Orientation angle maps are shown for the inner minimum (left), the membrane surface (middle) and the bulk solution (right). The COM distance to the membrane surface was assumed to be around 1.8 nm for all the three terpene molecules, whereas the COM distance to the inner minimum is different for each terpene (0.6 nm for LIM, 0.6 nm for PALD and 1.4 nm for DPAC). The terpene molecules in the bulk solution all exhibited very similar orientation angle maps (SI Figure S1), so only the angle map for LIM at a distance of 3.6 nm is shown. The maximum probabilities are of the same order of magnitude but are slightly different for each heat map. | 69 |
| Figure 24. Orientation angle heat maps of the terpene molecules in US-REST3 simulations. Orientation angle maps are 2D probability plots of the θ and ϕ angles that define the orientation of the terpene molecules with respect to the vector normal to the membrane surface. Angle θ is reported on the y-axes and angle ϕ on the x-axes, as represented in the bottom right panel. These | |

angle maps were used to assess the sampling of orientations visited by the terpene molecule in representative windows at three different COM distances to the bilayer. Orientation angle maps are shown for the inner minimum (left), the membrane surface (middle) and the bulk solution (right). The COM distance to the membrane surface was assumed to be around 1.8 nm for all three terpene molecules. In contrast, the COM distance to the inner minimum is different for each terpene (0.6 nm for LIM, 0.8 nm for PALD and 1.4 nm for DPAC). In the bulk solution, all three terpene molecules exhibited similar orientation angle maps (Figure S1 in the SI), so only the angle map for LIM at a distance of 3.6 nm is shown. The maximum probabilities are of the same order of magnitude but are slightly different for each heat map. 70

Figure 25. Changes in the minimum distance between the terpene molecules and water (blue) and number (N) of hydrogen bonds (red) between terpene and water as a function of terpene-membrane COM distance in the US-REST3 simulations for each terpene. 71

Figure 26. US-REST3 simulation of LIM with four and eleven replicas. A) Comparison of the predicted PMF using four (blue) and eleven replicas (red), showing convergence to the same free energy of binding. B) Angle maps for the window at a COM distance of 1.8 nm (surface) for the simulations with four (left) and eleven replicas (right). C) Angle maps for the window at a COM distance of 0.6 nm (internal minimum) for the simulations with four (left) and eleven replicas (right). 73

Figure 34. Diagram of the simulation setup of the REST3-enhanced umbrella sampling method. For every window at a distance dz , 11 replicas spaced from $\lambda = 0.2$ to $\lambda = 1.0$ were conducted in parallel and allowing exchange of configurations every 50 steps. 85

Figure 35 Illustration of the water layer selections. 87

Figure 36. Potential of mean force of the binding of ProTx-1 to a POPC bilayer predicted using conventional US and US-REST3 simulations. 88

Figure 37. Radius of gyration measured at three different peptide-membrane COM distances (2.6, 3.0 and 4.0 nm) obtained from the conventional US simulation of the interaction of ProTx-1 with a POPC bilayer and the corresponding windows from the ground replica in the US-REST3 simulation. 89

Figure 38. Number of hydrogen bonds in replicas with $\lambda = 0.2$ (most tempered), 0.5 (middle) and 1.0 (untempered) for the US-REST3 simulation as well as the conventional US simulation at three different peptide-membrane COM distances (2.6, 3.0 and 4.0 nm). 90

Figure 39. A: Number of intramolecular hydrogen bonds in ProTx-1 in a vacuum simulation conducted over 40 ns. B: Plot of RMSD vs time obtained from the last structure at a peptide-membrane COM distance of 2.6 nm in the US-REST3 simulation and extended using classical MD. The RMSD of ProTx-1 with respect to its initial structure in the extended simulation is shown in blue, and the RMSD with respect to a structure obtained in the conventional US is shown in red, (original branch described in Chapter 2, section 2.3). 92

Figure 40 A: Average number of water molecules located between the peptide and the POPC membrane at different peptide-membrane COM distance in both the US and US-REST3 simulations. B: Average number of hbonds between ProTx-1 and the POPC membrane vs time. 94

Figure 41. Representation of the simulation system. Xe atoms (pink) constitute the surface, Arg atoms (transparent blue) are used as background gas, and a single CO₂ molecule (light blue and red) that interacts with the atomic surface. 101

Figure 42 Distribution of quantities of interest for system A. A: Biased distribution of $\cos(\alpha)$. B: Biased distribution of dz . C: Reweighted distribution of $\cos(\alpha)$. D: Reweighted distribution of dz 104

Figure 43 Distribution of quantities of interest for system B. A: Biased distribution of $\cos(\alpha)$. B: Biased distribution of dz . C: Reweighted distribution of $\cos(\alpha)$. D: Reweighted distribution of dz . . 105

Figure 44 Distribution of quantity of interest for system C. A: Biased distribution of $\cos(\alpha)$. B: Biased distribution of dz . C: Reweighted distribution of $\cos(\alpha)$. D: Reweighted distribution of dz 106

Figure 45 FES of the three system along dz calculated using the wham method. 107

| | |
|--|-----|
| Figure 46. Two angle distributions with very different widths leading to the same overall conformational ensemble when considering the protein. In blue: conventional distribution of an angle between an atoms defined vector and the z axis. In red: Enhanced distribution of the same angle using MetaD. | 110 |
| Figure 47. Graphical representation of the procedure to obtain the rotated set of axis (xt,yt,zt) | 111 |
| Figure 48 Cosine of angle β as defined in section 2.1 for a classical MD simulation of Protx1 in water and a PBMetaD enhanced simulation using α and β as CVs. In blue: distribution obtained from the conventional MD. In orange: Distribution from the MetaD simulation. Dark red is the overlap between the two distributions. | 116 |
| Figure 49. A: Radius of gyration (Rg) of the main chain protein structure in water. B: RMSD of the main chain protein structure in water. | 117 |
| Figure 50 A: Comparison of the $\cos(\beta)$ distribution of simulations restrained at a protein-membrane COM distance of 2.6 nm using US and US with PBMetaD-US . B: RMSD of the protein C α from 100 ns of US-PBMetaD simulation. | 118 |
| Figure 51. Distribution of the Euler angle β for all the close proximity ($d_z < 3.0$ nm) trajectory frames for the US and US-WT-MetaD simulations. | 126 |
| Figure 52. Rotational auto correlation functions (RACFs) for every umbrella window of the simulations. A: Conventional US, original branch described in Chapter 2, section 2.3. B: US-WT-MetaD run1. C: US-WT-MetaD run2. | 127 |
| Figure 53. PMFs obtained by the two US-WT-MetaD simulations through the WHAM procedure compared with the FES obtained using conventional US described in Chapter 2, section 2.3. | 129 |
| Figure 54. RMSDs for windows with $d_z < 3.0$ nm. All the remaining windows had RMSD values falling below 0.2 nm, and were consequently deemed to be conformationally stable and, therefore, not reported. A: US-WT-MetaD run1. B: US-WT-MetaD run2. C: RMSD distribution in the conventional MD (top panel), US-WT-MetaD run1 (middle panel) and US-WT-MetaD run2 (bottom panel) simulations. | 130 |
| Figure 55. Time evolution of the RMSD of the protein for every walker in the PB-MetaD simulation. | 131 |
| Figure 56. A: Change in the predicted FES using the last 100,000 Gaussians in increments of 20,000. B: Change in the predicted PMF during the simulation for larger strides of Gaussians. | 132 |
| Figure 57. Time evolution of the protein-membrane COM distance. A: All the walkers B: Walkers 1, 4 and 6 have significant transitions in the COM distance, indicating some degree of translational freedom. | 133 |
| Figure 58. Distribution of Euler angle β of all the frames in close proximity to the membrane ($d_z < 3.0$ nm) for the US and PB-WT-MetaD simulations. | 134 |

Synopsis

In this thesis I will explore the application of a number of well-known enhanced sampling methods for the prediction of free energy changes in complex biological systems. These methods are usually devised and tested on small systems, such as the Ala-Ala dipeptide in vacuum, but the computation of the free energy in a biological system is much more sporadically attempted. Therefore, in this research I sought to assess the effectiveness of the most widely used enhanced sampling methods, namely umbrella sampling (US), replica exchange (RE) and metadynamics (MetaD). In particular, the aim was to use the US approach and enhance every simulation umbrella window with an additional enhanced sampling method. In Chapter 1 I will introduce the biological context of the problem, as well as the concepts of a molecular dynamics simulation, free energy calculations and the specific knowledge needed to understand this work.

The systems chosen for investigation were small drug-like molecules and peptides interacting with model biological membranes. Biological membranes allow the compartmentalisation and regulation of every living cell. They mediate the intake of nutrients, the permeation of drugs and a vast array of cellular recognition processes. Despite their far-reaching biological roles, a robust computational molecular simulation approach capable of predicting the interaction between different types of molecules with these lipid membranes remains to be developed. This is due partly due to the complexity of the composition of lipid membranes, partly because of the lack of optimally parameterized force fields that can describe their structure and interactions, and partly because of their plasticity, which exponentially multiplies the number of possible modes of interaction with peptides and small molecules.

Peptides themselves can be very flexible and, depending on their size, it can be difficult to predict their conformational ensembles in aqueous solution. To avoid this problem, a peptide was chosen such that it is stable to thermochemical changes and has a well-defined, ordered and rigid structure: Protx-1, a toxin found in the venom of red-back spiders as well as various other arthropod and cone snails species. Protx-1 is part of a family of peptides that contains an inhibitor cystein knot (ICK) motif, formed by three disulfide bonds that link the backbone together, giving them the exceptional aforementioned thermochemical stability. Conventional US simulations of the interaction of this peptide with a lipid bilayer were performed and it was found that the sampled orientations of the peptide with respect to the membrane strongly depended on the initial conformation of first approach

of the peptide to the membrane. After the peptide and membrane meet, the peptide spends the rest of the simulation (hundreds of nanoseconds) in the same orientation. Since a complete exploration of the available orientational space is required for an accurate calculation of the free energy, it was decided to enhance the sampling of each umbrella window using enhanced sampling methods. The first approach investigated was a purpose-designed replica exchange method, whereby replicas with reduced forces between the membrane and the peptide and the membrane and water were used to increase the free rotation of the peptide at the membrane surface. To my knowledge, there is no other method capable of specifically tuning the interactions of the membrane with the protein without also influencing other parts of the system Hamiltonian. Consequently I developed an in-house method called replica exchange with solute tempering 3 (REST3) to address this problem. To validate the method, a set of small molecules for which experimental partition free energy data is available. The molecules chosen were limonene, perillaldehyde and perillic acid, three terpenoid compounds with increasing polarity that have been thoroughly studied in recent years. The predicted free energy of partition of these three molecules was observed to match experimental data well and the trend in binding free energies is thus well predicted. When applied to Protx-1 though, this method failed to reproduce the correct binding free energy. The protein was found to experience an unexpected increase in its self-interaction energy, making it fall into deep energy minima which are difficult to escape and which change its average interaction with water molecules. This is probably due to the non-inclusion of the protein-protein self-interaction terms in the replica exchange probability. This technique will require future additional development to integrate these energy terms before it can be used with proteins.

In the light of these results, the performance of MetaD was tested for enhancing the orientational sampling of Protx-1 in combination with US as the underlying main method to obtain the binding free energy. The interaction of a US potential with a MetaD potential is all but obvious, so to develop a correct protocol to unbiased these simulations a test system was used and composed of a static xenon surface (emulating the presence of a membrane) and a CO₂ particle immersed in argon gas (emulating the protein immersed in water). With this test system the appropriate reweighting schemes that would be used with Protx-1 simulations were defined. The results shows that this reweighting protocol can correctly reproduce the underlying free energy landscape of the system.

Once the reweighting protocol was developed, the influence of the use of various collective variables (CVs) on the orientational sampling of Protx-1 was tested. The obvious approach of choosing three angles as CVs defined on some atoms of the protein (reminiscent of for the approach used with the terpenes and CO₂) revealed an unexpected problem. The distribution of orientational angles appeared to be enhanced substantially by MetaD, but the orientation of the protein did not in fact change at all.

This was found to be due to the size of the protein and its inertia, which makes it easier for an applied biasing potential to cause the local, abnormal displacement of the two atoms that define an orientation vector rather than rotate the entire protein structure. For this reason, I developed a new type of angles that are defined on the entire protein structure rather than on an arbitrary number of atoms. This was based notionally on quaternion-defined rotations but through the development of an Euler angle-based method that could work with the current implementation of PLUMED, which does not support quaternion operations. To test these new angular orientation CVs, a range of variations of conventional and fast-converging MetaD methods were used, including well-tempered MetaD (WT-MetaD), bias exchange MetaD (BE-MetaD) and parallel bias MetaD (PB-MetaD). After comparing the performance of these methods for enhancing the orientational sampling of ProTx-I, the PB-MetaD approach was chosen due to its lower computational effort and shorter convergence time. Use of these simulation approaches confirmed that the Euler angle CVs developed are suitable for enhancing the orientational space exploration of semi-rigid molecules.

Finally, the reweighting approach developed with the above described CO₂ system in combination with the new Euler angle CVs was used to predict the free energy of binding of ProTx-I to a POPC membrane. US-WT-MetaD and PB-MetaD approaches were used. The interaction of the US potential with the WT-MetaD potential was found to induce a deformation of the protein when it reaches the membrane. In contrast, the PB-MetaD approach was observed to avoid this problem, providing a more promising method to predict free energies of binding of proteins with lipid membranes.

1 Chapter 1- Literature review

1.1 Introduction

1.1.1 Umbrella sampling

Umbrella sampling (US) is one of the most common enhanced sampling methods to characterise small molecule-membrane interactions (SMMI) and obtain related free energy profiles. This approach has been vastly used in literature in many different contexts, cementing the robustness of its theoretical approach and protocols of application. In US simulations of SMMI sampling along the reaction coordinate (RC) is enhanced by applying an additional energy term (the biasing potential) that restrains the location of the small molecule at a given point along a RC of choice. The typical approach to US consists in setting up a number of simulations (referred to as 'windows'), each window will sample states of the system in the vicinity of the discrete value of RC on which the biasing potential is centred (**Figure 1**). The restraining potential forces the window to explore values close to the centre of the umbrella, energetically disfavouring values that are further apart. From these umbrella windows a series of N partially overlapping histograms of the collective variable can be extracted, where N is the number of windows. Each histogram provides a biased probability distribution of the RC. An alternative is to keep the biasing potential constant and slowly vary the value of the RC, in a protocol referred to as steered MD or pulling simulations ¹.

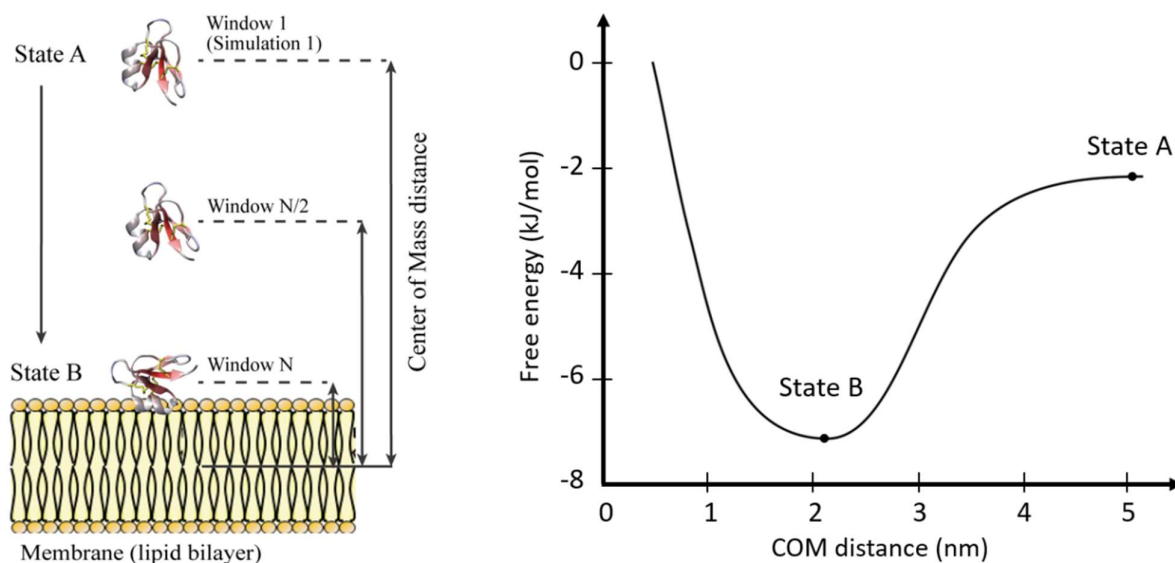


Figure 1. On the left, schematic representation of an umbrella sampling simulation of the interaction of a small peptide with a lipid bilayer. The reaction coordinate is the centre-of-mass distance between the molecule in solution (state A) and the molecule at the membrane (state B). On the right, sample

plot of the free energy profile, also known as potential of mean force (PMF) for the binding of the peptide to the membrane, providing an estimate of the change in free energy between states A and B.

To obtain the free energy, $A_{i(\xi)}$, where ξ is the reaction coordinate, the unbiased probability distribution of the system, $P_{i(\xi)}^u$, is calculated as ¹

$$P_{i(\xi)}^u = \frac{\int e^{-\beta E(r)} \delta[\xi'_{(r)} - \xi] d^N r}{\int e^{-\beta E(r)} d^N r} \quad (1)$$

Where $E(r)$ is the energy of the system with coordinates r , $\beta=1/kT$ and $\delta[\xi'_{(r)} - \xi]$ is the number of states in the simulation that have a value of ξ' . Assuming ergodicity in the system, the biased probability distribution, $P_{i(\xi)}^b$, obtained from the simulation will be equal to

$$P_{i(\xi)}^b = \frac{\int e^{-\beta E(r) + \omega_i(\xi'_{(r)})} \delta[\xi'_{(r)} - \xi] d^N r}{\int e^{-\beta E(r) + \omega_i(\xi'_{(r)})} d^N r} \quad (2)$$

With $\omega_i(\xi'_{(r)})$ being the biasing potential applied on the system. Since integration of the numerator is performed over all degrees of freedom except for ξ and the bias depends only on ξ , using Eq. 1 and Eq. 2 we can derive

$$P_{i(\xi)}^u = P_{i(\xi)}^b e^{-\beta \omega_i(\xi)} \langle e^{-\beta \omega_i(\xi)} \rangle \quad (3)$$

A more thorough derivation of Eq. 3 can be found in the in-depth review of US by Kästner ¹. This derivation is analytical and only assumes that enough sampling is provided for every value of ξ . As $P_{i(\xi)}^b$ is obtained from the US simulation and the biasing potential is given analytically, the free energy, $A_{i(\xi)}$ can be derived as

$$A_{i(\xi)} = -\frac{1}{\beta} \ln P_{i(\xi)}^b - \omega_i(\xi) + F_i \quad (4)$$

Here F_i is a constant that can be estimated by approaches such as the weighted histogram analysis method (WHAM) ² when combining results from different windows. However, WHAM requires the windows to have sufficient overlap in their distributions ³. An alternative to WHAM is umbrella integration, which instead relies on the calculation of the derivative $\partial \ln P_{i(\xi)}^b / \partial \xi$ and does not require an overlap in the energy distributions ⁴.

The most common RC used in simulations of SMMIs is the COM distance between the molecule of interest and the membrane, as illustrated in **Figure 1**. This focuses the sampling effort to only one RC and thus simplifies the N-dimensional landscape to one dimension. However, use of this RC will fail to

capture the effects of relevant degrees of freedom (DOFs) if these have slower relaxation times than the time dedicated to sample each window. The lack of sampling of these DOFs could affect the accuracy of the free energy calculation, especially if they are responsible for configurational transitions of the small molecule or the reorganisation of the lipids around it ⁵. These slow-varying variables are usually referred to as ‘hidden variables’ that govern the exploration of the configurational landscape ⁶. In any case, if the configurational ensemble is incomplete, i.e. it is insufficiently sampled, then the calculation of the free energy will be inaccurate.

1.1.2 Biological Membranes and their interactions with small molecules

Biological membranes are semi-permeable structures present in all living organisms. Membranes protect the cell from the external environment, compartmentalise cells and their organelles and control the selective transport of molecules in and out of these compartments. Biological membranes, and cell membranes in particular, are commonly described using the fluid mosaic model where the membrane consists of a lipid bilayer into which membrane proteins are embedded (**Figure 2**). Biological membranes fulfil a wide range of functions and are highly complex structures ⁷. A typical bacterial or mammalian membrane can consist of several hundreds of different types of lipids and include hundreds of different proteins ^{8,9}.

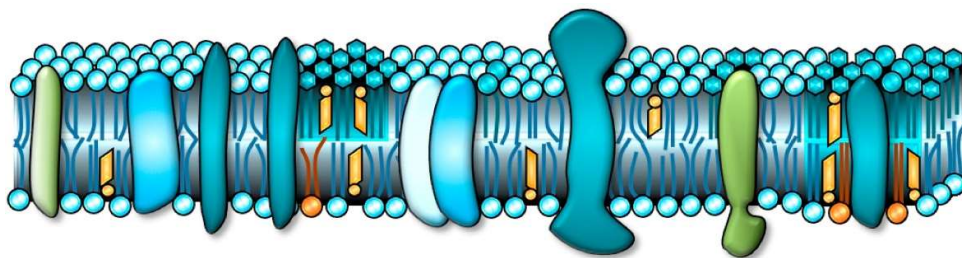


Figure 2. Schematic representation of a cell membrane composed of a lipid bilayer with embedded membrane proteins. Reprinted from ¹⁰, Copyright 2017 with permission from Elsevier.

The lipid bilayer of cell membranes is primarily made up of glycerophospholipids, sphingolipids and sterols, varying in proportion depending on the type of cell or the specific function of the membrane ⁹. The lipid bilayer has two main functions. Firstly, lipids are critical for the function and structural stability of membrane proteins, which make up nearly a third of the human cell membrane proteome and account for approximately 60% of known drug targets ¹¹⁻¹³. Secondly, the bilayer controls the permeation of small molecules across the membrane and thus controls what enters the cell or cell organelles. In addition, the membrane mediates the lateral diffusion of small molecules bound to the

membrane and can thus affect the mechanism of action of molecules that act as ligands for membrane proteins. It is thus not surprising that the interaction of small molecules with membranes affects the pharmacokinetics, bio-availability and mechanism of action of endogenous substances such as neurotransmitters as well as drugs.¹⁴⁻¹⁶ The permeation of small molecules is also critical for assessing the toxicity of exogenous particles such as diesel soot or silica dust¹⁷, the use of nanoparticles for imaging, biosensing and therapeutic applications¹⁸ and the use of organic molecules in the cryopreservation of plant germplasm^{19,20}. Consequently, understanding small molecule-membrane interactions (SMMIs) and, in particular, the ability to predict the binding affinity and permeation coefficients of small molecules is an active area of research and aids our understanding of physiological processes and facilitates pharmaceutical development and a diverse range of biotechnological applications.

SMMIs can be classified mainly into surface binding (or adsorption) and permeation across the membrane. Surface binding describes the interactions of the small molecule at the water-lipid interface and the binding affinity associated with the molecule moving from a fully solvated state to a membrane-bound state. In contrast, permeation describes the flux of molecules from a solution environment on one side of the membrane, through the hydrophobic core to the other side of the membrane. Both surface binding and permeation are initiated by the diffusion of the small molecule from the bulk solvent to the membrane surface. This process can be driven by electrostatic and/or hydrophobic interactions. The specific mechanism of interaction depends on the size, shape and physico-chemical properties of the small molecule, the composition of the membrane and environmental factors such as temperature, pH and ionic strengths. In this thesis the focus will be mainly set on surface binding and partition of the small molecules, rather than on their permeation capabilities. Seelig has described in detail the driving forces that govern the adsorption and insertion of peptides to membranes²¹. This theoretical framework can be used to model the adsorption of small molecules that do not penetrate membrane. For neutral molecules, adsorption is described using 'hydrophobic partitioning', where the concentration of a molecule in the bulk and at the membrane is considered to be at equilibrium, and the binding affinity is obtained from a simple partitioning principle.

1.1.3 Model membrane systems

Due to the high complexity of biological membranes, their interactions with other molecules are commonly studied using model membranes. They are usually composed one or more of the most abundant lipids found in the membrane of interest such that they reproduce the physico-chemical properties critical the process one aims to study (e.g. fluidity or surface charge). Besides reducing the

complexity, model membranes provide the ability to systematically investigate the effect of different environmental factors on SMMIs (e.g. lipid composition, pH, ionic strength, level of hydration or temperature). By controlling lipid composition and environmental factors, model membranes also enable a more direct comparison of the surface binding or permeation of a set of small molecules with varying physicochemical properties. The most commonly used model membrane systems include micelles, liposomes, supported monolayers and bilayers (**Figure 3**)^{22,23}. The choice of model system and its composition depends on the set of relevant physicochemical properties that are to be reproduced, the process that is under study (binding or permeation) and the experimental technique(s) used to monitor the progress of said process^{7,23–25}.

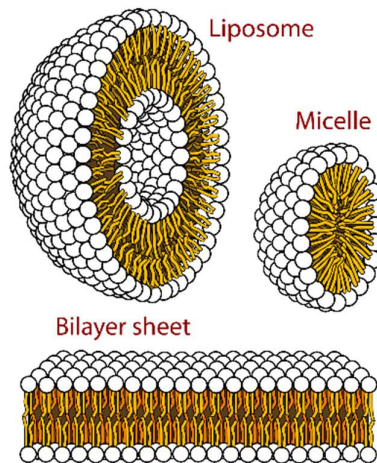


Figure 3. Schematic representation of various types of model membrane systems. Supported bilayers and monolayers can have different kinds of linkers (polymeric, charged, and covalent) connecting the support with the actual bilayer. Licensed under a Creative Commons Attribution (CC BY) license.

For plasma membranes, the most commonly used lipids are neutral (zwitterionic) phospholipids like 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine (POPC), 1,2-dipalmitoyl-sn-glycero-3-phosphocholine (DPPC) or 1,2-dioleoyl-sn-glycero-3-phosphocholine (DOPC)²⁵. To add surface charge, lipids like 1-palmitoyl-2-oleoyl-sn-glycero-3-phospho-L-serine (POPS, mainly used in mammalian membranes) or 1-palmitoyl-2-oleoyl-sn-glycero-3-phospho-(1'-rac-glycerol) (POPG, mainly used in bacterial membranes)^{24,26} are added. Other lipids commonly added to mimic the plasma membrane include sphingolipids, phosphatidylinositol or cholesterol.

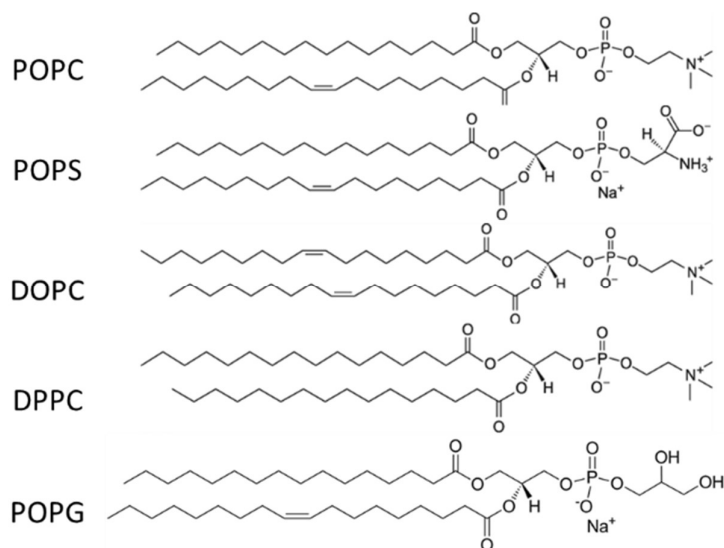


Figure 4. Structures of common lipids used to build model membranes to mimic the mammalian plasma cell membrane.

The thermophysical phase of the lipid bilayer used as a model system is also of importance. Lipid bilayers can be found in a variety of phases including sub-gel, gel, rippled and liquid crystalline (also called fluid phase). Phase transitions represent rearrangements in membrane structure leading to changes in the stability of bilayers (**Figure 5**). Each phase has a characteristic molecular arrangement described best with a set of structural properties like the area per lipid (APL), membrane thickness and lipid tail order parameter (S). For example, in the fluid phase lipid tails have high mobility and low lipid tail order. In contrast, in the gel phase lipid tail mobility is lower and is accompanied by an increase in lipid tail order. Phase transitions can be induced by changes in environmental conditions such as temperature, pH, ionic strength and hydration. In model membranes, transition from the fluid to the gel phase can be induced with either a reduction in temperature (thermotropic phase transition), an increase in pressure (barotropic phase transition), a reduction in hydration or a decrease in pH^{24,27,28}.

The interaction of small molecules with membranes can both induce phase transitions in model membranes as well as shift in the fluid-to-gel transition (T_{melt}). For example, in the context of cryobiology, a reduction in the T_{melt} reflects stabilisation of the liquid crystalline phase of plasma membranes and the retention of their biological function at lower temperatures. A reduction in hydration upon desiccation leads to the opposite effect: an increase in the T_{melt} and the stabilisation of the gel phase. During cryopreservation, where both desiccation and liquid-nitrogen temperatures

lead to severe membrane damage, addition of sugars and other non-penetrating cryoprotective agents is aimed at stabilising the fluid phase of cell membranes ²⁰.

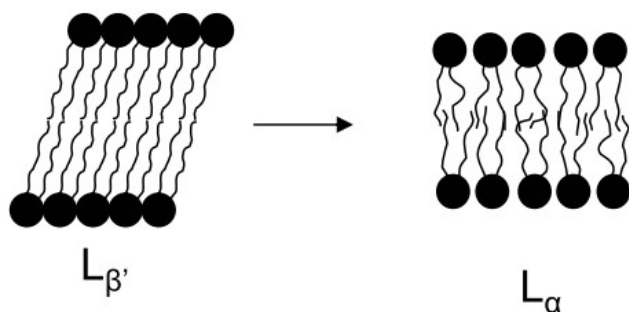


Figure 5. Lipid phase transition from gel ($L_{\beta'}$) to fluid phase (L_{α}) induced by an increase in temperature. The gel state is characterised by a smaller area per lipid, larger bilayer thickness and higher lipid tail order compared to the fluid phase. Adapted from ²⁹. Copyright 2005, American Chemical Society.

Various experimental techniques are used for characterising SMMIs. These techniques provide information about the changes in structure, dynamics and stability of membranes upon their interaction with small molecules, as well the orientation and location of the molecule. In addition, these techniques are used to investigate the thermodynamics and kinetics of SMMIs.

One of the most commonly used methods to characterise the structural and dynamical properties of lipid bilayers ^{10,30} is Nuclear magnetic resonance (NMR) spectroscopy. Solid-state NMR (ssNMR) can be used to study the dynamics of lipids in the fluid-phase over a wide range of time scales, providing information about the conformation and orientation of individual lipids and their diffusion, as well as collective motions such as membrane deformations. By comparing these data in the absence and presence of a membrane-active molecule, ssNMR can be used to describe the effect of SSMIs on structure or dynamics of lipids and membranes. NMR can also provide information to construct phase transition diagrams of model cell membranes ³¹, which can subsequently be used to study the effect of SMMIs on phase transition temperatures ³². NMR has also been used to determine the preferred position and orientation of drug-like molecules, water, ethanol and flavonoids in lipid bilayers ³³⁻³⁵ or to determine the high-resolution, 3D structure of small molecule-lipid complexes ³⁶.

Other spectroscopy techniques used to study membranes and SMMIs include Förster resonance energy transfer (FRET), electron paramagnetic resonance (EPR), fluorescence correlation spectroscopy (FCS) and Fourier transform infrared (FTIR) spectroscopy. FRET is particularly useful to monitor the spatial organisation and distribution of lipids including the formation of micro domains in membranes

³⁷, and to detect the intercalation of small molecules into lipid bilayers. Whilst FCS and EPR are less commonly used to directly study SSMMs, both techniques provide information on lipid dynamics and can thus be used to study effect of SSMMs on the membrane. For example, FCS can be used to characterise diffusion processes in membranes by 'tracking' individual lipids both in 'simple' model membranes as well as complex lipid mixtures or membranes with high heterogeneity ³⁸. EPR is used to determine the mobility and order of lipids and thus provides a measure of membrane fluidity, or to provide information on the water accessibility of the membrane as a function of membrane depth ^{39,40}.

In addition to spectroscopy techniques, popular techniques used to study membrane systems are neutron and x-ray scattering methods. These methods provide dispersion patterns that can differentiate the lateral packing arrangement of lipids characteristic to a specific phase, determine bilayer thickness ^{41,42}, lipid tilt angles and diffusion. Scattering methods can also be used to determine the position of sugars, alcohols and drug molecules in membranes ^{26,43-49}.

Besides the structural information, characterising of the thermodynamics and kinetics of SSMMs is critical to understand surface binding or permeation phenomena. These properties have been studied using a range of techniques including differential scanning calorimetry, isothermal titration calorimetry (ITC) and surface plasmon resonance (SPR). Differential scanning calorimetry is commonly used to determine the phase transition temperatures of membranes and associated activation energies of model membranes including any changes to these properties upon interaction with small molecules ^{26,27,40,50}. ITC can be used to measure the amount of thermal energy absorbed or released as a result of small molecules binding to membranes, thus allowing the determination of binding constants (free energy of binding), enthalpies and stoichiometry of SSMMs ⁵⁰⁻⁵². SPR is used to determine the kinetics of binding and unbinding of small molecules to lipid bilayers, yielding corresponding association and dissociation rate constants, as well as the overall binding affinity constant ⁵³⁻⁵⁵.

1.1.3.1 SPR methodology

SPR is a technique which measures electron oscillations that forms at the interface between a thin layer of a conductive material and a dielectric medium. These electron oscillations are referred to as Surface Plasmons, and they can be induced by shining a laser on the thin layer of metal with a particular angle⁵⁶. If the angle is the distinct angle of incidence required for the setup, a drop in the reflectivity of incident light can be observed, and the optical energy is dissipated into a guided electromagnetic wave along the metal-dielectric interface. The extent of this energy transfer is sensitive to the mass deposited on the metal layer, as well as the local index of refraction near the metal-dielectric interface. Changes in any of these parameters will change the angle of reflectivity, which can be measured by the SPR instrument⁵⁷. The film is usually a very thin gold film, in virtue of the malleability

of gold and the strength of the SPR signal generated by a gold surface. The usefulness of SPR is limited to materials that can form high quality uniform films in the nanoscale size (<100 nm) on the metal layer. Many materials like ceramics and polymers can be not suitable, but biological membrane can be made according to these parameters. To determine the free energy of adsorption with SPR a sensorgram is obtained, which is presented in resonance units or response units (RU, 1RU=1 pg/mm²) against time. The data are then used to generate an isotherm curve, by plotting the SPR signal as a function of the concentration⁵⁸. Various problems must be addressed when using SPR to measure a free energy of binding. Firstly, because SPR measures the refractive index change of the medium in the 300 nm around the metal surface it is sensitive to the amount of molecules adsorbed on the surface as well as the amount of molecules dissolved in the medium in the 300 nm around the surface. This is commonly referred to as the bulk-shift contribution and must be accounted for by subtracting it from the raw SPR signal⁵⁹. Secondly, analyte-analyte interactions can greatly affect the shape of the adsorption isotherm, leading to erroneous calculation of the free energy of adsorption. For this reason it is preferred to run SPR experiments at very low analyte concentration, to avoid analyte-analyte interactions. A solution commonly applied is to use only the initial slope of the isotherm, where the concentration of the analyte in bulk is negligible, in the assumption that in this regime analyte-analyte interactions are minimized. In this case the free energy of adsorption can be calculated as:

$$\Delta G_{ads}^0 = -RT \ln \left(\frac{C_s}{C_b} \right) = -RT \ln \left(\frac{QK}{\delta C^0} + 1 \right) \quad (5)$$

Where C_s is the concentration of the analyte solution, C_b is the bulk concentration, which has to be close to zero, Q and K are two fitted parameters, δ is the thickness of the layer and C^0 is the concentration of the analyte solution in standard state.

1.1.3.2 ITC methodology

ITC is a technique that measures the exchange of heat associated with the formation of a complex between a ligand and an analyte, or any phenomena that can lead to a change of heat. At its base an ITC instrument consists in a reference cell and a sample cell enclosed in an adiabatic shield jacket. The reference cell contains the same solvent of the sample cell and some dissolved analyte, but no ligand. To the sample cell, a solution of ligand dissolved in the solvent is injected through a very precise automated pipette. Differences in temperature between the sample cell and the reference are monitored using a sensitive thermocouple, and the temperature of the cell is kept the same by providing or subtracting heat from the sample cell. Depending on the exothermic or endothermic nature of the reaction taking place the instrument will have to add or subtract heat to keep the cell at the same temperature. The power required to return the sample cell to thermal equilibrium is then integrated over time. This integral is proportional to the amount of interaction that is occurring in the

sample cell⁶⁰. After the sample cell becomes saturated, the only changes to the heat are due to the dilution of the sample cell contents. The heat signal (or power provided) can be then described as an apparent ΔH of the reaction happening in the cell. The term apparent is used because the total ΔH is due to the sum of many phenomena such as electrostatic, hydrophobic, Van der Waals interactions, hydrogen bonding, protonation/deprotonations, conformational changes and many more. The equilibrium constant K can be measured from the precise knowledge of the concentration of the free and the bound ligand. ITC can measure K very precisely between milli molar and picomolar range. The free energy of adsorption then might be calculated from the relation:

$$\Delta G = \Delta H - TS = -RT \ln K \quad (6)$$

The high sensitivity of the technique combined with the fact that the sample doesn't have to be immobilized or artificially labelled (through radioactive compounds for example) makes it a very powerful technique. The usage of ITC with membranes involves some theoretical problems though. Firstly, as previously stated, to calculate the K and therefore the ΔG requires precise information of the initial concentration of the peptide and of the binding sites of the membranes. However clearly quantifying the binding site for a surface and specifically for a biological membrane is a very challenging task that has no clear solution to date. In second instance when dealing with bulky molecules as peptides might be, the possibility of a non-negligible fraction of the heat exchange being due to solvation entropy, structural changes of the peptide and peptide-peptide binding cannot be ignored⁶¹.

1.1.4 Challenges in understanding SMMIs Molecular dynamics

Despite the vast array of techniques available, obtaining a complete picture of the structure, kinetics and thermodynamics of a process involving SMMIs is still a very challenging task. Most of these techniques provide a single information on the system in very specific experimental conditions, which can differ from the physiological conditions at the process of interest occurs. With a few exceptions, these studies provide an averaged signal over conformations or time, which usually lacks the resolution to make mechanistic deduction on the process under study.

Molecular dynamics (MD) is a computational method that can help rationalising and complementing experiments by providing both atomistic level detail and high temporal resolution. MD simulations have been used extensively to study SMMIs, both for rationalising experimental data and for predicting changes to membrane properties such as APL, membrane thickness, order parameters and lipid lateral diffusion coefficients. Simulations have also provided insight into the effect of lipid composition on these properties⁶²⁻⁶⁶, particularly when the properties of interest are inaccessible (or

difficult to access) using other techniques⁶⁷. In MD simulations, a molecular model of the system of interest is built representing atoms as a particle and subject to a carefully built potential energy function that represents the bonded and non-bonded interactions between all particles. The initial position and velocities of the particles are given and Newtonian equations of motion are solved and propagated in time according to the potential energy function under specified temperature and pressure conditions. At every point in time the energy of the system is then dependent of the position of every atom, i.e. its state at time t. The result is a trajectory, time series of frames containing position and velocities for every atom in the system at every time, from which all manner of thermodynamic, structural and dynamic information can be in principle obtained. The use of MD simulations to study detailed models able to describe specific lipid–protein interactions is an established method that can be employed to rationalize and guide experiments^{68–76}.

1.1.5 Force fields

The accuracy of the prediction coming from a molecular dynamic simulation is first and foremost dependent on the representation of the potential energy functions. In a molecular dynamics system the atoms are represented by spheres bound by springs. The springs represent the bonded forces, like the direct covalent bonds between atoms, or additional angular and dihedral contributions that shape a molecule in its characteristic conformation. Additionally, non-bonded interactions like electrostatic and Van der Waals interactions are modelled respectively through the standard electrostatic potential and the Lennard-Jones (LJ) function. The resulting Hamiltonian for the potential energy of a conformation of atoms usually takes the form of :

$$\begin{aligned}
 H(x) = & \sum_{bonds} k_{bond}(r(x) - r_{eq})^2 + \sum_{angle} k_{angle}(\theta(x) - r_{eq})^2 + \sum_{dihedral} \frac{k_{dih}}{2} [1 + \cos(\eta\phi(x) - \gamma)] \\
 & + \sum_{i,j} \frac{A_{i,j}}{r_{i,j}(x)^{12}} - \frac{B_{i,j}}{r_{i,j}(x)^6} + \frac{q_i q_j}{\epsilon_r \epsilon_0 r_{i,j}(x)}
 \end{aligned} \tag{7}$$

The collection of these functions and parameters used to is generally called a force field (FF). A FF is a collection of functions to describe the interactions between atoms and their parameters used to describe these functions. Usually, these parameters are obtained from quanta-mechanical calculations and experimental data fitting, parametrized to reproduce the behaviour of the molecule. The FF is then generally validated against some experimental data sets that ensures that the set of parameters used can reproduce some measured experimental values. As an example, for membrane systems the set of parameters obtained to describe a phospholipid molecule would then be validated against experimental values of membrane thickness, area per lipid, lipid order and more⁷⁷. If the FF is

successful, simulation of a membrane of the given phospholipid will reproduce the experimentally measured APL for the membrane. A set of parameters that well describes a phospholipid molecule won't necessarily describe the interaction of that lipid molecule with a different species. In fact if during the development phase of the FF the data hasn't been expressly parametrized to reproduce protein interactions (or any other small molecule interactions) it is very unlikely that the FF will be able to correctly reproduce them. In the case of this work a force field that can not only reproduce protein-membrane interactions, but also small drug-like molecules interaction is needed.

There are several protein FFs with available parameters for a variety of lipids, among the most used ones we can find AMBER LIPID 14⁷⁸, GROMOS 54A7⁷⁹, CHARMM36⁸⁰ and SLIPIDS⁸¹. Each of these FF presents its advantages and disadvantages. A first distinction to make is between all atoms FFs and united atoms FFs. All atom FFs describe every single atoms in the system as a sphere with its own parameters, comprising all of the hydrogen atoms. Being the atoms with the smallest mass, hydrogens are the fastest oscillating atoms in the systems and furthermore they are usually among the most numerous atom components of organic molecules. The dynamics of these atoms are not always that interesting though, especially in the case of aliphatic hydrogens. For this reason some FF are developed to be united atoms. United atoms FF like GROMOS 54A7 apply a light degree of coarse graining to the system to reduce the number of atoms to simulate, trying to minimize the loss in accuracy. The aliphatic hydrogen and the carbon they are bound to are parametrized as a single bead, reducing considerably the number of particles in the system and allowing longer simulation time steps to be used.

Several works in literature that compare the performances of these FFs and outline their abilities to reproduce experimental data can be found ⁷¹⁻⁷³. From this comparison it is difficult to pick a clear best choice among all the FF for general small molecule-membrane interactions. On one hand all of the force fields manage to describe the properties of the phospholipid to a satisfying extent ^{84,85}, on the other hand there is not much data assessing the interaction between a membranes and small molecules, particularly on non-peptide small molecules. GROMOS 54a7 provides access to less computationally intensive simulations due to its united atoms nature in exchange for some resolution. In the light of its increased computational performances arising from its united atom nature, its extensive use in literature and to be consistent with previous work in the group we chose GROMOS 54a7 to be our FF of choice. This FFs has been shown to reproduce the APL, thickness and lipid order parameters of DMPC and POPC bilayers suitably ⁸⁵, is protein compatible and comes with an easy method of obtaining general small molecule parameters in the Automatic Topology Builder suite ⁸⁶. GROMOS 54A7 general force field was parametrized against the density, and Enthalpy of vaporization and Solvation Enthalpy in water and cyclohexane of 28 small molecules. The parameters for the lipids

instead has been parametrized to reproduce APL, Lipid order parameters and Gel-Liquid transition temperature. Due to the nature of this study being fundamentally similar to the measurement of a partition coefficient between a polar environment (water) and an apolar environment (membrane), this parametrization strategy, especially in the Enthalpy of solvation in a polar and apolar solvent seemed fitting. Despite the overall good performances in simulating the lipid dynamics, GROMOS 54a7 comes with a known overestimation of the interaction of proteins with the charged heads of the lipids^{83,87}. Recently a new correction for the lipid heads parameters accounting for this effect has been published by Marzuoli et al⁸⁷. In this work the partial charges and Lennard Jones parameters for the choline, the phosphate and the glycerol moieties of the lipids were tweaked to better reflect the real interactions that a protein would have with a model bilayer. For the relevant protein – membrane simulations then, this corrected version of the FF called GROMOS 54a8 was used instead of its previous version.

1.2 Free energy Calculations

By linking ensemble averages of molecular properties to bulk thermodynamic properties, MD simulations can be used to calculate binding constants, as well as many other physical quantities. Changes in free energy are the most important thermodynamical quantity that can be extracted from a simulation. This is because the difference in free energy between two states determines the direction and the speed of any chemical reaction. From the knowledge of the free energy it is possible to extract every other thermodynamic quantity from a simulation in a statistically correct way. However, obtaining these properties from MD simulations relies on sampling representative molecular configurations to calculate free energy differences. The free energy F of a system is related to its partition function Z by:

$$F(N, V, T) = -\beta \ln [h^{-3N} \int \int e^{-\beta H(p,r)} dp dr] \quad (8)$$

$$F = \frac{1}{\beta} \ln Z \quad (9)$$

Here $\beta = 1/k_B T$, k_B is the Boltzmann constant, T is the absolute temperature, h is the Plank constant, $H(p,r)$ is the energy of the system in configuration r with momentums p , and N is the number of degrees of freedom of the system. The integral in Eq 1 is over all the possible configurations r and all possible momentums p . This means that theoretically, the MD simulation used to calculate the free energy should sample all combinations of configurations and momentums available to the system in the given macroscopical state. As the degrees of freedom in a molecular system scales to the power

of $3N$, sampling all possible configurations is only feasible for systems of a few dozens of particles. For biological systems composed of thousands of particles this is no longer feasible and only a subset of all configurations can be sampled. The challenge then becomes sampling all configurations relevant to the changes in free energy associated with the process of interest.

In simulations of SMMI, one is usually interested in the difference in the free energy between two well defined states. For small molecules binding to the surface of a membrane, this usually corresponds to the free energy difference between the small molecule in solution (unbound state) and the small molecule bound to the water-lipid interface. Calculating free energies differences instead of absolute free energies means that the contribution of all the numerous degrees of freedom that are very slightly perturbed or unperturbed at all by the process of going from one state to another cancels out. e.g. it makes it very little difference in terms of absolute free energy of the system if the molecule is bound or unbound to the membrane, as the biggest contribution to the absolute free energy will be given by the vast number of atoms of solvent and of the membrane involved. The free energy contribution from all these atoms though is roughly equal in the unbound state and in the bound state, so the difference between the two states, only will account for the change of free energy resulting from the binding of the molecule, giving us therefore a quantity of interest.

Usually, to simplify the description of a process of binding, or any other process, we resort to the choice of one or more parameters on which we can project the free energy multidimensional surface (the free energy depending from all the possible degrees of freedom of the system). These parameters are usually referred to as a reaction coordinates (RCs) or a collective variables (CVs) and are usually indicated as ξ . The change in free energy as function of a ξ is called a free energy surface (FES).

Along the FES minima, maxima and other sorts of critical points can be found. Critical points are points with first derivative equal to zero. Minima are considered stable points, and they are defined as critical points with all second partial derivatives that are positive. Maxima on the other hand are critical points with all second partial derivatives that are negative. Saddle points are instead critical points where not all the second derivatives have the same signs. In particular a saddle point is defined as the point of highest energy (relative maximum) along the reaction path and is also the point of lowest energy in the direction perpendicular to the reaction path.

For small molecules binding to membranes, the energy barriers and minima along the path that connects the unbound to the bound state are usually not known a priori. This means that the simulations from which the FES is calculate have to sample all possible configurations accessible to the system on that path. In SMMIs, the CV or ξ is usually the centre-of-mass (COM) distance between the membrane and the interacting molecule, eventually supported by additional CVs like orientational

descriptors or system specific CVs to better describe the process. The probability distribution along ξ integrated over every degree of freedom except ξ is given as ¹:

$$Z(\xi) = \frac{\int e^{-\beta E(r)} \delta(\xi(r) - \xi) d^N r}{\int e^{-\beta E(r)}} \quad (10)$$

Substituting equation 10 in 9 expressed as a function of the reaction coordinate ξ enables the calculation of $F(\xi)$, which is also commonly referred to as the potential of mean force (PMF), or the FES along the degree of freedom of choice. Once the free energy is known, the binding constant (and thus binding affinity) can be obtained upon integration on a arbitrarily chosen interval of ξ of the PMF using the relation:

$$K(\xi) = -RT \ln F(\xi) \quad (11)$$

1.3 Sampling of SMMIs in MD simulations

Obtaining a correct thermodynamical ensemble from which to extract a correct FES from an MD simulation is not trivial, requiring the user a prudent choice of parameters and an extensive knowledge of the challenges involved. The description of many of the common problems of MD simulations has been extensively covered in other works ⁸⁸ and will not be discussed in detail in this present document . In this document I will tackle one specific problem that affects MD simulations of complex systems. This is what it is commonly referred to as the sampling problem, namely the inability of a MD simulation to sample all the relevant molecular configurations effectively in a finite amount of time ⁸⁸.

This problem affects any ensemble averages extracted from the simulation, such as thermodynamical quantities, structural properties and indeed the free energy. In the case of SMMIs, the energy barriers associated with the roto-translational motion of the solute molecules on the surface or within the membrane can limit effective sampling. The energy associated with the rearrangement of lipid molecules in the membrane as a result of the interaction with the solute molecules necessary for the solute tumbling usually exceeds thermal (kT) energy. This can result in the system being kinetically trapped in a local energy minimum for a long period of time, leading to a state of quasi-ergodicity in the simulation, which does impair the accuracy of the physical properties that can be obtained from the simulation ⁸⁹.

Adequate sampling of the configurational space in small molecule - membrane systems remains a significant challenge, which can only be achieved using enhanced sampling methods⁹⁰. A large number of enhanced sampling methods have been developed for biomolecular simulations. These methods mostly fall in two categories: the first category of methods enhance non-specifically the exploration of all degrees of freedoms (DOFs) ^{91,92} , while the second one, including methods like umbrella

sampling (US) or metadynamics (MetaD) only enhance the sampling along one or more RC or CV. The appeal of the former type of methods, like temperature replica exchange MD (T-REMD) and other REMD-based methods, is that they can be used without any a priori knowledge of the system^{91,93}. However, for efficient sampling the number of required replicas scales as $\text{DOF}^{1/2}$ and, as the number of DOFs is related to the number of particles in a system, REMD-based methods are less suitable to study large systems such as membranes⁸⁹. The latter type of methods instead requires the a priori selection of RCs or CVs that can provide an adequate projection of the multi-dimensional FES of the transition between an initial and a final state, e.g one must choose the CVs that can best approximate the real binding process of the molecule to the membrane^{1,94}. In the following sections, the most commonly used enhanced sampling methods are described.

1.3.1 Replica Exchange methods

Replica exchange MD (REMD) is an enhanced sampling⁹⁵ method that consists of simulating several copies (replicas) of the same system in different perturbed conditions. A replica where the system is unperturbed acts as the reference system from which the properties of interest are calculated. The conditions in the reference replica are usually the biologically or chemically relevant ones. The perturbed conditions are chosen such that each replica can explore a larger volume of phase space than the unperturbed ones. During the simulations, neighbouring replicas are allowed to swap their configurations in a manner that satisfies the detailed balance condition. The probability of exchange between replica m and a replica n is given by:

$$p_m(X_m)p_n(X_n) p(X_m \rightarrow X_n) = p_m(X_n)p_n(X_m)p(X_n \rightarrow X_m) \quad (12)$$

Here $p_m(X_m)$ is the probability of finding a state with coordinates X_m in replica m , and $p(X_m \rightarrow X_n)$ is the transition probability. The most common approach is to allow the system to exchange configurations with a certain probability according to the classic Metropolis criterion, such that $\Delta = (\beta_n - \beta_m)(E_m - E_n)$ with $\beta_m = 1/k_B T_m$ and E_m being the energy of the system m at the moment of exchange⁹⁶ with probabilities

$$p(X_m \rightarrow X_n) = \begin{cases} 1 & \text{if } \Delta < 0 \\ e^{-\Delta} & \text{if } \Delta > 0 \end{cases} \quad (13)$$

In this way configurations derived from regions of phase space usually not accessible by unbiased MD simulations can gradually make their way to the reference replica. The exploration of high-energy configurational states is usually achieved in one of two ways: i) by simulating each replica at

increasingly higher temperatures, in which case the method is referred to as ‘parallel tempering’ or temperature-REMD (T-REMD); or ii) the Hamiltonian of the system is perturbed in a similar incremental fashion, a class of techniques generally known as ‘Hamiltonian tempering’.

1.3.1.1 Parallel tempering

In parallel tempering the variable used to perturb the system is the temperature. Low temperature replicas allow for fine sampling of an energy minima at physically relevant condition but may become kinetically trapped in such minima during the timescale of a typical simulation⁹⁶. Higher temperature replicas have higher kinetic and potential energies and can thus sample larger volumes of phase space more sparsely. The main drawback of parallel tempering lies in its high computational cost due to the large number of replicas needed. The large number of replicas are required because of the exchange acceptance ratio described in Eq. 13. In parallel tempering the probability of exchange is exponentially related to the difference in the total energy of the system (including solvent molecules). For this probability of exchange to be acceptable (i.e. not too small) the two replicas need to exhibit a non-zero overlap in their potential energy distributions⁹⁶. Small molecule – membrane systems have a large number of solvent (water) molecules, which usually translates into a very narrow potential energy distribution. Consequently, many replicas are required to sample a large span of temperatures. As a result, parallel tempering has seen little application in SMMI simulations. In addition to the high computational cost, the range of temperatures used is limited in SMMI simulations as membranes are generally sensitive to thermal disruption. This means that high temperatures would threaten a disruption of the structure of the membrane to an extent where the configurational ensemble of the higher replicas would be sampling unwanted areas of the configurational space.

1.3.1.2 Hamiltonian tempering

Hamiltonian tempering simulations constitute a general class of enhanced sampling methods that include a variety of different techniques, each differing in the way the Hamiltonian of the system is perturbed^{92,97–102}. The general principle of these methods is that each replica uses a perturbed Hamiltonian, which is usually a linear combination of the starting Hamiltonian H_A and the target Hamiltonian H_B . The aim of the Hamiltonian perturbation is the same as the scaling of the temperature in parallel tempering: to explore higher energy regions and thus sample larger volumes of phase space. H_B is usually chosen to flatten energy barriers on some or all degrees of freedom, allowing thermal fluctuations at the given simulation temperature to overcome the energy barriers that impede the sampling of phase space. Exchange between replicas is essentially the same as in eq 13, but with some

important differences that make Hamiltonian tempering more resource efficient than parallel tempering. The first one is that directly modifying the Hamiltonian of the system allows the selective perturbation of subsets of atoms or molecules in the system, such as solutes or a part of a protein, giving the user more control on which part of the system he wants to perturb (e.g. the user can choose to not perturb the thermosensitive membrane). The second advantage is that the exchange probability between neighbouring replicas can be modified such that it is based only on the difference in energy between the relevant portions of the system ⁹². This means for example, that in replica exchange with solute tempering (REST), leaving out the energy terms for the solvent from the functional form of the exchange probability reduces the computational effort by an order of magnitude and decreasing the number of replicas needed. This happens because the potential energy of the solvent is very narrowly distributed and predominant on the energy of the solutes or the membrane itself, so that cutting out of the exchange probability calculation this terms allows for more overlap between potential energy curve and therefore higher exchange probabilities. As a result REST can be used for simulations of much larger systems compared to parallel tempering ⁹⁸. **Figure 6** schematically illustrates the differences in the potential energy distribution across replicas between classical parallel tempering and REST. The efficiency of this class of methods in regard to SMMIs has not been fully explored, probably due to the lack of easily accessible and flexible implementations in MD simulation packages. A new implementation of REST that we called REST3 along with a more in depth discussion about the original REST paper can be found in Chapter 3 of this thesis.

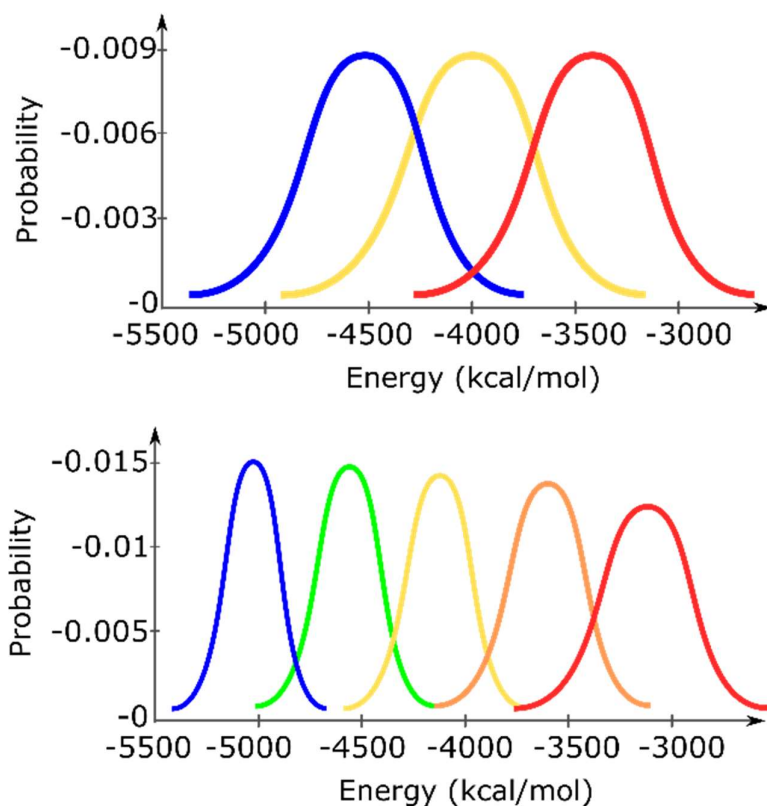


Figure 6. Schematic representation of the potential energy distribution in classical parallel tempering and solute tempering. Top panel: Hamiltonian tempering can achieve a wider distribution of the energy, such that a good overlap is obtained with a smaller number of replicas. The distribution is larger because the solvent-solvent term, which has a very narrow distribution around its mean value due to the large number of molecules, is cancelled out. Bottom panel: In parallel tempering, because of the narrow distribution of the solvent-solvent energy, the distributions of all replicas are much narrower because of the solvent-solvent term; therefore, a greater number of replicas is required to achieve a relatively small overlap in the energy distributions.

1.3.2 Metadynamics

Besides US, metadynamics (MetaD) is one of the most used enhanced sampling method for biomolecular simulations including simulations to describe SMMI. In MetaD, a history-dependent potential along one or more CVs is used to prevent the system from re-sampling previously visited configurations. The bias usually takes the form of a sum of repulsive Gaussian energy functions. Like for RCs, the CVs are assumed to describe the largest degrees of freedom of the process of interest such that all significant energy barriers in the FES can be sampled. The biasing potential V_G as a function of the CVs is defined as

$$V_G(S_{(x)}, t) = w \sum_{\substack{t=\tau G, 2\tau G, \dots \\ t' < t}} e^{-\frac{(S_{(x)} - S_{(t')})^2}{2\delta s^2}} \quad (14)$$

Here $S_{(x)}$ describes the CV as a function of the coordinates of the system, $s_{(t)} = S_{(x(t))}$ is the value of the CV at time t , w is the Gaussian height, δs is the Gaussian width and τG is the frequency with which Gaussian functions are added. For sufficiently long simulation times, the biasing potentials added enable the system to exit every energy minima and sample neighbouring ones, leading eventually to the free diffusion of the system in configurational space (**Figure 7**). From the resulting biasing potential V_G the underlying, unbiased free energy surface can be obtained as $F(S) = -V_G(S) + C$, where C is an arbitrary additive constant¹⁰³.

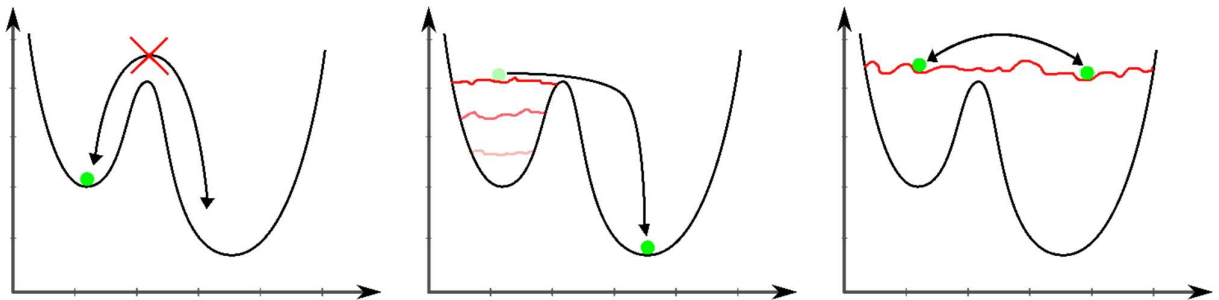


Figure 7. Schematic representation of metadynamics. From left to right repulsive Gaussian energy functions are gradually deposited on previously explored configurations, gradually filling energy wells and enabling exploration of other neighbouring energy wells. A) In the beginning the system is unbiased and the transition between the two wells is difficult because of the barriers. B) Gaussian potential functions start to be accumulated on the free energy surface. C) In the end, the resulting energy landscape is flat and sampling in the system becomes random in the collective variable space.

There are two main limitations to this technique. The first problem concerns the convergence of the FES. In a single MetaD simulation, the FES does not converge to specific values, but rather fluctuates around the real values, leading to a statistical error proportional to the square root of the Gaussian deposition rate¹⁰⁴. Because of this, it can be difficult to decide when to end a simulation. On the other hand, if the simulation is run for too long, the continued addition of the biasing potential can push the system into exploring non-relevant parts of the FES. This issue is mostly circumvented by using well-tempered metadynamics (WT-MetaD). Here, the biasing potential is modified such that the height of the repulsive Gaussian functions is decreased over time in a manner that is proportional to the overall

biasing potential present on the current value of the system CV ¹⁰⁴. The equation for the derivative of the potential then becomes:

$$\dot{V}_G(S(x), t) = w e^{-\frac{V(s,t)}{\Delta T}} e^{-\frac{(S(x)-s(t'))^2}{2\delta s^2}} \quad (15)$$

Where ΔT has the value of a temperature and it's a parameter that can be used to tune the behaviour of the metadynamics from a normal MD ($\Delta T=0$) to a standard metadynamics ($\Delta T= \infty$). This means practically that using a finite non-zero ΔT , the more gaussians are deposited on one point of the CV space, the smaller the gaussian gets, until it gets so small that one can approximate that even for big number of additional gaussians, that value of potential will not change. At this point the free energy corresponds to the inverse of the potential rescaled by a factor¹⁰⁴ :

$$F(s, t) = - \frac{T + \Delta T}{\Delta T} V(s, t) \quad (16)$$

We call the factor preceding the potential γ or bias factor. This factor regulates the virtual maximum amount of potential that can be added during a WT-MetaD simulation, which also correspond to the highest value of free energy barrier along the CV that can be crossed . It follows that a WT-MetaD is considered converged when the gaussian heights along all of the CV values are close to zero.

The second problem of MetaD lies in the choice of the CVs and is similar to the problem of choosing the RC in US simulations. All the relevant, slow-varying CVs should be considered in a MetaD simulation. Failure to do so can lead to non-physical dynamics because of the possible presence of hidden energy barriers on other slow-varying degrees of freedom, thus resulting in a FES that is not representative of the process of interest. An extra constraint on this problem is that the aforementioned requirement has to be satisfied whilst keeping the number of CV small enough (a rule of thumb is number of CVs<4) because the time required to converge an N-dimensional space of CVs scales exponentially with the number of CVs. To solve this problem, we decided to use a Parallel Bias MetaD (PBMetaD) ¹⁰⁵, a method capable of applying multiple mono-dimensional potential simultaneously, which are therefore easier to converge. This method which will be more extensively discussed in Chapter 3, basically allows multiple mono dimensional WT-MetaD potentials to deposited

simultaneously, each with a certain conditional weight that depends on the value of the potential along that CV at time t .

1.4 Test systems

As mentioned in the synopsis the test system that we chose to analyse is ProtX-1, a protein responsible for the inhibition of Nav channels, a type of membrane proteins with ion transportation properties and involved in the pain signalling network. This protein is part of a wide family of peptides called Inhibitor Cystein Knot peptides, which members have been found to have inhibition activity toward some selected Voltage gated ionic channels. In this section I will try to provide the biological context in which these peptide work while exploring the current state of the art knowledge for ProtX-1.

1.4.1 Voltage gated ionic channels

Voltage gated ionic channels (VGICs) are a class of membrane proteins responsible for the voltage-dependent and selective transport of ions across cell membranes^{106–109}. VGICs are involved in a wide range of physiological processes in mammals including muscle and nerve relaxation, regulation of blood pressure, and sensory transduction^{110–113}. More than 60 diseases caused by the malfunction of ion channels (channelopathies), have been reported. This makes VGICs an important target for the development of drugs to treat a range of conditions, including neuropathic and inflammatory pain, epilepsy, multiple sclerosis and prevention of migraine^{109,114}.

The majority of VGICs share a common architecture that consists of a large central pore domain, through which the ions flow across the membrane, and four voltage-sensing domains (VSDs) that control the gating of the channel by detecting changes in membrane electric potential. The pore domain contains a selectivity filter that determines which metal ion can permeate through the pore. Depending on their ion selectivity VGICs are divided into potassium (K_v), sodium (Na_v), calcium (Ca_v) and chlorine channels (ClC)¹⁰⁹. The pore domain is highly conserved among types of channels and through different species. In contrast, the VSDs are more specific to each channel and show variations even among closely related subtypes of VGIC¹¹⁵.

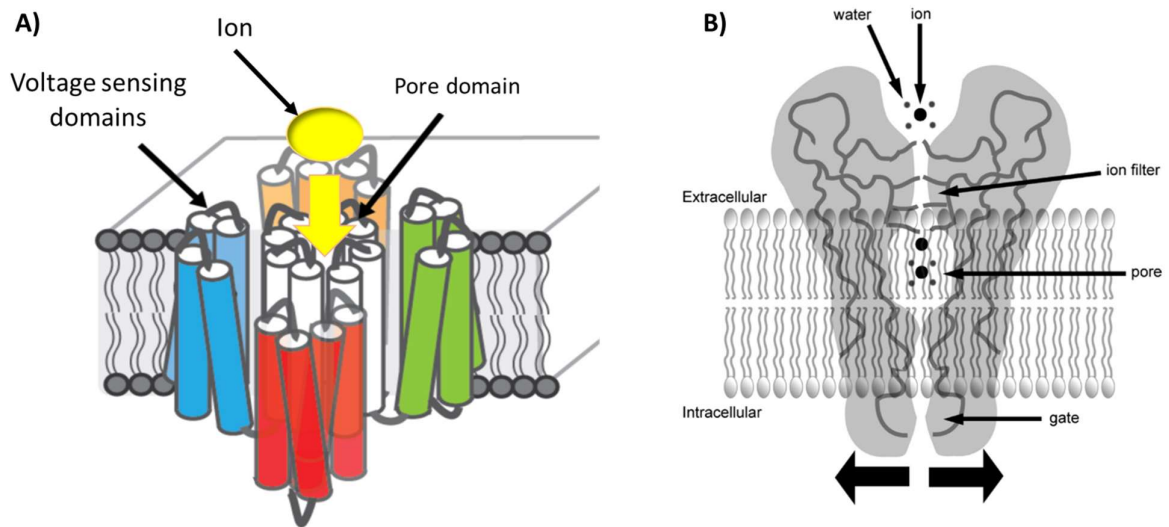


Figure 8 A: General structure of a VGIC., 4 voltage sensing domains typical of one specific ion channel surrounds the pore domain. The pore allows for the passage of ions through the hydrophobic membrane. B: A close up of the structure of the pore domain.

Influencing the channel response to an electrostatic potential would therefore likely require reengineering of the channel itself and thus not be a feasible drug designing approach. Studying the response of the channels to an inhibitor is much easier and has been the mainstream approach to modulate the activity of these channel until now. There are two main ways an inhibitor can act on VGICs; as a pore blocker (PB) or as a gating modifier (GM). In general, the mechanism of action is better characterised for PBs compared to GMs¹¹⁶. PBs act on the channel by binding to the top part of the channel, close to the selectivity filter, thus occluding the pore and preventing ions to flow through¹¹⁷⁻¹¹⁹. From a drug design point of view, PBs present the problem that the pore domain is very conserved among subtypes of ion channels. A drug that belong to this class of inhibitors would therefore be not selective enough to target one subclass of ion channels. These would turn out to be a considerable problem considering that some of these ion channels are responsible for cardiac activity and other essential physiological processes such as muscle contraction. GMs bind to the VSDs of the VGIC instead, altering the kinetics of gating¹²⁰. As VSDs are more specific to each channel, GMs can thus be very selective for a specific subtype of channel. In contrast with PBs, much less is known about the interactions between GMs and VSDs¹²¹.

1.4.2 ICK peptides

Spiders, scorpions, cone-snails and other venomous animals rely on venom as their primary threat mechanism. Their venom needs to immobilize or to kill the prey and it needs to do so rapidly to prevent retaliation or the escape of the prey. It is unsurprising then that many of these animals

developed toxins able to inhibit VGICs action in a highly specific and highly effective way. Some of the most potent inhibitors of VGICs are Inhibitory Cysteine knot (ICK) peptides^{106,107}. ICK peptides are mostly found in the venoms of spiders, scorpion and marine molluscs, but can also be found in snakes as well as plants and fungi¹²². The ICK is a structural motif in which an embedded ring formed by two disulphide bonds and their connecting backbone segments is threaded by a third disulphide bond¹²³ (Figure 9). These knotted disulphide bonds provides them with high structural stability¹²⁴. ICK peptides behave like structured mini-proteins and present exceptional thermochemical stability¹⁰⁸. ICKs are able to withstand high temperatures, extreme pH, organic solvents and proteases^{123,124}. These characteristics combined with their ability to inhibit ion channels or receptors with high specificity and selectivity makes them valuable lead molecules for the development of pharmacological tools and drugs to treat neurological disorders^{106,107,120,125}.

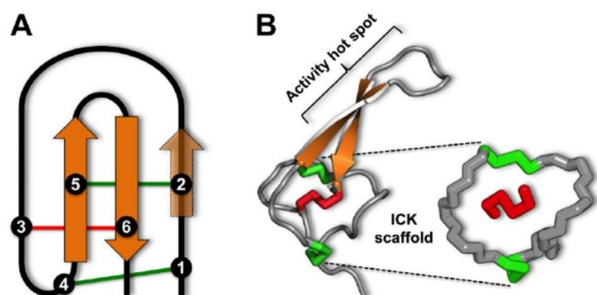


Figure 9 A) Schematic 2D representation of ICK peptide and their secondary structure. In green, the 2 of the disulfide bonds that form the ring, in red the threading one. B) 3D representation of the previous structure.

Some ICKs are known to inhibit VGICs with high potency and sub-type selectivity. For example, the conotoxin MVIIA isolated from marine cone snails inhibits calcium channel activity with high potency and was found to have analgesic activity, leading to its clinical development as a drug¹²⁶. It has been suggested by various studies that for a peptide to act as a GM it must be able to interact with the cell membrane^{127,128}. While membrane binding is not a prerequisite for the inhibition of VGIC via a gating-modifying mechanism¹²⁸, for some GMs the mechanism of action is directly related to their ability to bind to membranes. For example, studies by Henriques et al showed that increasing the membrane binding of the ICK peptides ProTxII and HwTx-IV also increases their potency for inhibiting the Nav1.7, a target for the treatment of neuropathic pain^{53,129}. Similarly, the membrane binding of the ICK peptide Pn3a correlates with its potency to inhibit Nav1.4. It is likely that for these peptides the initial step in inhibition of the channel is driven by a membrane activation mechanism, such that the peptide first binds to the membrane and subsequently to the VSD of the channel¹³⁰.

This mechanism ensures enhanced kinetics of the process, as it increases the local concentration of the peptide at the surface and also reduces the dimensionality of the diffusion from 3D to pseudo 2D (along the surface of the membrane), making it more likely for a peptide to find its target¹³¹ (Figure 10)

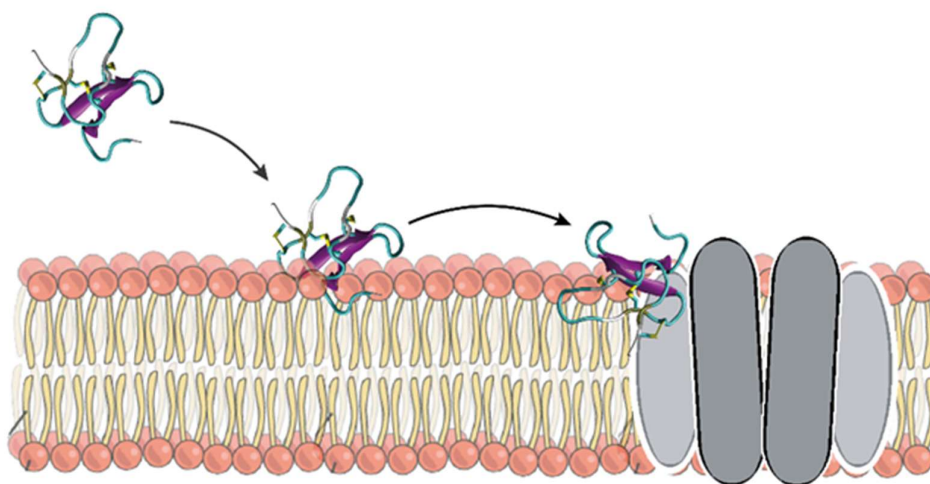


Figure 10 Illustration of the membrane accumulation mechanism. In the first phase the peptide goes from the solution on the membrane, shifting from a 3D diffusion to a 2D diffusion. Once on the membrane it can reorient in the proper direction and bind to the channel at an enhanced rate.

Our ability to predict the membrane binding properties of a generic ICK peptide is, however, limited and so is our understanding of the mechanisms that drives these interactions. Various attempts have been made to find some physical descriptors which could predict the binding affinity of ICKs with PC membranes, among them presence of a hydrophobic cluster, residual dipole of the molecule, surface exposed cationic and anionic areas and hydrophobic to hydrophilic area ratio. Still experiment shows that despite being very similar in these descriptors, molecules like Hd1a and Protx-1 show vastly different membrane binding properties^{128,132}.

Modulating the membrane binding potential of a peptide that binds to the channel through a first membrane binding step would provide extra space for design, especially on amino acids that are not crucial to the primary biological activity. This is a very precious feature in drug design, and therefore membrane binding ICK peptide were chosen to be investigated in this study. We will now restrict this dissertation only to peptides that has been shown to bind PC membranes and contextualize why Protx-1 among all of them was chosen to be the subject to this work.

1.4.3 Choice of Protx-1 as a model system

The ability to bind zwitterionic membranes is we discussed a characteristic of high importance in the study of ICKs. Despite the considerable number of independent studies on ICK toxins, the free energies of partition and/or binding to biological membranes are still unknown for many of these peptides. Nevertheless, the ICK peptide selected for this study should fulfil the criteria for selection as a good model system. In first place its NMR or crystallographic structure must be available and well resolved, without disordered zones and possibly rigid and stable in time. This is to avoid the problem of conformational sampling, which would multiply the difficulty in the study by adding a conformational sampling part to the already difficult orientational sampling required. In second place the ideal peptide must show strong binding for both neutral and negatively charged membranes. A successful simulation protocol should in fact be able to reproduce the binding affinity with different membranes and different amount of charged lipid heads. Another desirable feature for the ideal peptide is that the mechanism of binding to the membrane has to be simple, like a superficial binding rather than involving more complex permeation mechanism, like carpeting effects or pore formation. This last feature is ideal in the first phase of the development of the protocol, as all of these more complex mechanisms are only the subsequent steps of a first membrane binding phase and can be modelled in successive steps of higher order protocols. More complex mechanisms will be left to investigate for those future protocols that can build upon the information that this one will provide.

Among the many possible peptides we chose to study Protx-1. ProTx-I is a 35-residue peptide with sequence ECRYWLGGCSAGQTCKHLVCSRRHGWCVWDGTFS, isolated from *Thrixopelma pruriens* responsible for the inhibition of all Na_v subtype channels^{133–135}. The three disulfide bonds that constitute the ICK motif located between residues C2-C16, C9-C21 and C15-C28 provide the aforementioned structural rigidity which will allow us to partially discount the intramolecular conformational sampling. The peptide was first characterized by solution NMR using a 0.5 mM concentration of peptide in a 90% H₂O, 10% D₂O at 298 K and pH 5¹³⁵. SPR measurements showed that ProTx-I binds to both neutral (POPC only) and anionic (POPC/POPS 4:1) lipid membranes with an increased affinity for negatively charged ones¹²⁸. These experiments were conducted using a Bioacore 3000 with a L1 biosensor chip in a HEPES buffer with a peptide concentration that ranged from 1 to 64 μM. These experiment do not provide a fitting for the data to extrapolate a value of k_a and k_d for the peptides that they measure, but it is shown in the paper that in the same condition, on the same instrument with the same chip, Protx-1 yield an RU response much higher than other known membrane binding peptide such as Protx-2 and Hd1a. Trp fluorescence quenching experiments

showed no significant change in the signal of the Trp, suggesting that the peptide does not penetrate deeply into the lipid membrane but only binds to the membrane on the surface¹²⁸. This, combined with the absence of quenching of the signal by the addition of aqueous acrylamide quencher indicate a superficial binding of the Trp into the membrane with no permeation¹²⁸. A study by Gui et al¹³⁵ measuring two electrode voltage-clamp currents in oocyte expressing Nav 1.2 also provides information on the protein active residues through alanine scanning. This tests measure the current carried by Na ions from outside the oocyte to inside the oocyte. This current is abolished by the action of Protx-1 on the channel, which causes the channel to close and exclude the ions from passing the oocyte membrane. This study reports that replacing residues W5, W27 and W30 with alanine abolishes the peptide's ability to inhibit Nav 1.2. suggesting that these three residues might be involved in the biological activity of the protein. Another more recent voltage clamp study¹³⁶ provides insight on the activity of Protx-1 and its mutants on various sub-types of Nav channels, including Nav 1.2, 1.5, 1.6 and 1.7. The study confirmed the importance of the Trp residues in previously noted, with some of these residues being vital to the interactions with some specific channels but not others. Interestingly for the current study some residues are found to non-specifically reduce the binding affinity to every channel, which could suggest their role to be involved in the membrane association part rather than in the actual interaction with the channel. For example, the mutant Protx-1 K17E shows reduced inhibition activity toward all tested Nav subspecies.

There is evidence that upon titration with POPC micelles the only residue exhibiting an NMR shift are superficial hydrophobic residues like W5 and W27, with the rest of the peptide showing no significant shift, which implies the overall structure of the peptides does not change substantially upon binding¹²⁸. This is consistent with the same behaviour observed for other ICK peptides such as omega-agatoxin IVA and Purotoxin-2, for which the ICK structures part of the peptides exhibit very little NMR shifts upon contact with membranes¹³⁷. This fact will be an assumption that underlies the way that the data coming from the simulation are interpreted.

Given all of this characteristic Protx-1 was chosen as a model peptide to test out our enhanced sampling method on membrane system with big molecules.

2 Chapter 2 – Conventional US of Protx-1 binding to a POPC membrane

2.1 Introduction

As discussed in Chapter 1 MD simulations have been found to be effective in studying the dynamics of peptides on flat surfaces and even on membranes^{100,138}. We also saw in Chapter 1 that in ergodic systems the ensemble average can be approximated by its average over time. In the case that the system is ergodic, given enough time it will explore every possible available state. Once we obtain a long enough trajectory and therefore a big enough statistical sample we then can assume that we can obtain the average property of a quantity $A(r)$ simply measuring the average over all frames of $A(r)$. Even more interestingly we can obtain the relative probability of our system to be found in a state where A assumes an arbitrary value, which may be of biological or physical interest. The latter is an assumption that in general has to be demonstrated, especially for systems that include big energy barriers separating their stable energy minima. Let's take for example a protein-membrane system with a large energy barrier that separates state a state with coordinates r_1 from another state with coordinate r_2 , both accessible at room temperature and of similar free energy. Let's say that for the first N steps the system oscillate around the conformation r_1 with a property $A(r_1)$, not managing to cross the barrier. If the ensemble average is taken at the end of these N steps all the states will show $A=A(r_1)$, and everything on the other side of the barrier will be ignored. If we had given the system another few steps it would have explored state $A(r_2)$, changing the value of the average. The average extracted will therefore be inaccurate, and it will be inaccurate until r_1 and r_2 are visited with a probability of $e^{-\frac{F(r)}{kT}}$, where F is the free energy. For complex systems N can be very high, in the order of hundreds of nanoseconds. This is generally called the sampling problem. The free energy of the system is therefore the most important quantity that can be extracted from a simulation, as once the probability that every frame of a simulation carries any property can then be extracted as a weighted ensemble average. Free energies can be a function of the atomic coordinates r , or functions of functions of atomic coordinates, like distances $d(r)$ or more complex collective variables $CV(r)$. The most used CVs to describe the binding of small molecules and proteins to biological membranes is generally the centre of mass (COM) distance between the molecule and the membrane. Free energies for SMMIs are usually provided along the COM distance and they are called either Potential of Mean Force curves (PMF). To calculate the free energy accurately though, as we mentioned, the protein must be able to cross all the relevant energy barriers that limit the available conformational space at every COM distance from the membrane. Generally speaking, these barriers can be of many kinds, ranging from rotational barriers when near the membrane, to dihedral barriers that regulates the

opening of a beta hairpin, to desolvation barriers and many more. In this work I will particularly focus on rotational barriers because, as stated in Chapter 1, inhibitor cysteine knot (ICK) peptides are very rigid structures in virtue of the ICK knot motif and are less subject to the lack of internal conformational sampling. As described in Chapter 1, ProTx-1 was chosen as a model system to calculate binding affinity to phospholipid bilayers. The reproduction of a configurational ensemble with an appropriate exploration of the orientational states which yields a free energy of interaction consistent with experiment will be the aim of this work and that described in subsequent chapters. To generate this kind of ensemble we test and compare a number of enhanced sampling techniques. The first method chosen is conventional Umbrella Sampling (US), the most commonly used technique in the field. This method is not expected to overcome the sampling difficulties intrinsic to this kind of system, but will lay the foundation for more complex method and provide a base line comparison for subsequent chapters. A “branched US” approach was also used, which consisted of using four different starting structures with differing orientation with respect to the membrane to assess the influence of the starting relative orientation of the protein on the shape of the potential of mean force (PMF).

2.2 Methods

2.2.1 Umbrella sampling

The system (Figure 11) was assembled using GROMACS tools (GROMACS 4.6.8 version) , and consisted of a rectangular box of 6.0 x 6.0 x 11.0 nm containing one ProTx-1 peptide (PDB structure 2MNL.pdb¹³⁹) and a pre-equilibrated (250 ns of simulation at 303 K) POPC bilayer containing 128 lipid molecules (64 per layer). The pre-equilibrated membrane structure was downloaded from the Automated Topology Builder (ATB, molecule ID 31⁸⁶). The membrane centre of mass (COM) was situated around 3 nm from the edge of the simulation cell whereas the COM of the protein was situated at 8.4 nm, which gives a distance between the two COMs of 5.4 nm.

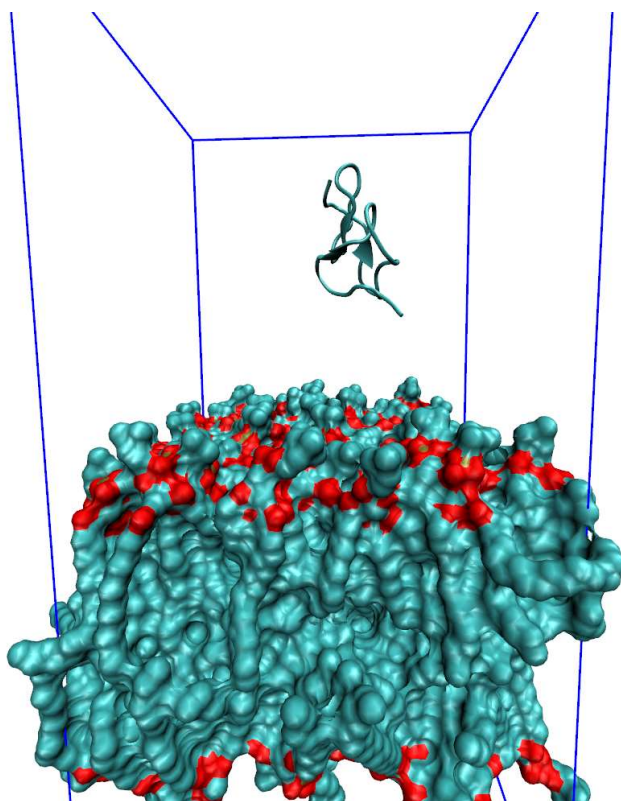


Figure 11. Snapshot of the simulation cell prior to solvation.

This system, represented in Figure 11, was then solvated with 9850 SPC waters¹⁴⁰, 30 sodium ions and 34 chlorine ions. The ion concentration was chosen to emulate a physiological concentration of 0.1 M of NaCl, with the additional four chlorines used to neutralize the charge of the protein in the system. The topology for the system was generated using the GROMOS54a7 FF¹⁴¹ and GROMACS 4.6.8.

An energy minimization was run using a steepest descents algorithm until a maximum force of less than $250 \text{ kJ/mol}^{-1}\text{nm}^{-1}$ was achieved. The system was equilibrated for 5.0 ns in an NPT ensemble at 298 K and 1 bar with a timestep of 2 fs, using the Berendsen thermostat with $\tau=0.1 \text{ ps}$ and a Berendsen semi-isotropic barostat with $\tau=1 \text{ ps}$ ¹⁴². The electrostatic interactions were treated using a Reaction-Field approach with $r_{\text{coul}}=1.4 \text{ nm}$ and an $\epsilon=62 \text{ F/m}$. These parameters were used to be consistent with the parametrization of the force field in use. The US simulation was then constructed with 18 umbrellas spaced 0.2 nm apart between 2.0 and 5.4 nm of protein-membrane COM distance, the spacing was chosen so that the overlap between the histograms of the z-coordinate was adequate for the wham procedure to yield an accurate PMF.

To produce starting configurations for the US windows, a pull simulation was used. For this, the peptide was moved using a shifting umbrella potential with a force constant of $k = 500 \text{ kJ/mole}$ and a pull rate of 0.01 nm/ps over 400 ps. From the resulting trajectory, 5 frames were extracted with Protx-

1-POPC membrane COM distances of 2.0, 2.8, 3.6, 4.4 and 5.2 nm. For each window, a 200-ns NPT equilibration simulation was carried out using the same conditions as the previous NPT equilibration. The peptide-membrane COM distance was restrained using a harmonic potential with a force constant of $k = 500$ kJ/mole. This additional equilibration ensures that any distortion to the membrane caused by the pulling can re-equilibrate.

The remaining neighbouring windows were generated from these 5 windows. The initial frames for the “offspring” windows were extracted from the equilibrated frames of the “originating” windows when the distance at which the potential for the offspring window would be centred occurred naturally in the originating window. Once the windows were generated, they were equilibrated for other 200 ns in the same conditions as the originating ones. This was done to allow the system to diffuse in roto-translational space to avoid correlations in neighbouring windows. The next generation of windows was then generated from the previous one.

2.2.2 Branched umbrella sampling

The final structure derived from the trajectory of the umbrella window centred around 4.0 nm was used as a starting point for three other simulations. This distance was chosen because this is the distance that was later found as the zero of the PMF. The water molecules were removed, and the protein was manually rotated around the x-axis to generate three different starting structures. In structure A the protein was rotated by 90° , in structure B by 180° , and in structure C by 270° . These systems were then solvated again, adding respectively 9936, 9955 and 10015 water molecules and 28 Na and 32 Cl ions. The systems were equilibrated for 10 ns using the same parameters as described in section 2.1.

From each of the equilibrated structures a pulling simulation was performed to rapidly bring the protein to the POPC membrane, with the same parameters described in section 2.1. The rapid process of pulling preserved the starting orientation of Protx-1 with respect to the membrane. This was done to obtain all the intermediate states between bulk and protein membrane given the initial conformation. From these simulations, conformations with a COM distance of 4.0, 2.8 and 1.8 nm were restrained with a harmonic potential, equilibrated in an NVT ensemble and simulated for 200 ns in an NPT ensemble, in the same manner as described in section 2.1. The remainder of the windows were then generated as previously described, until three separate “umbrella branches” were filled with equi-spaced 0.2 nm umbrella windows, as shown in Figure 12.

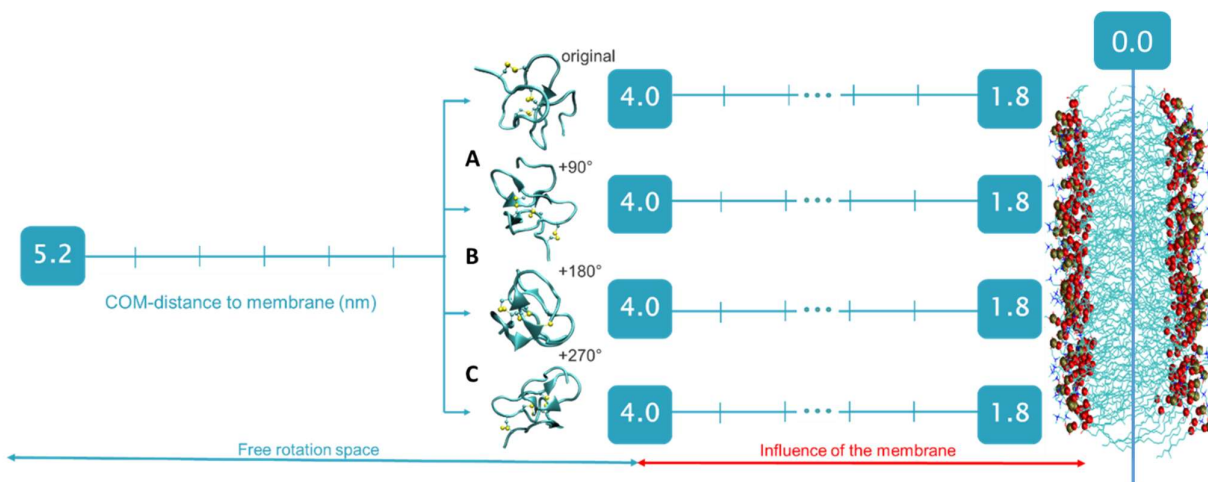


Figure 12 Setup of the branched umbrella sampling simulation. At 4.0 nm the main branch diverges from the top most simulation (described in section 2.1) and three new simulations with different orientations of ProTx-1 with respect to the membrane were performed.

2.2.3 Definition of orientation angles of Protx-1

To quantify the relative orientation of Protx-1 with respect to the membrane surface, two angles were used (ψ and θ). These angles were defined by choosing two orthogonal vectors on the protein, and measuring their angle with a vector that runs along the z-axis of the system (normal to the membrane surface) (Fig 3). To define the orthogonal vectors along the protein, three atoms were chosen within the structure of the protein. To aid in the choice of atoms, the root mean square fluctuation (RMSF) of Protx-1 backbone atoms extracted from the window centred at 4.0 nm (in bulk water) of the original US branch was measured. The atoms needed to define the vectors were chosen to be the least fluctuating among the protein atoms, so that their movement is the most correlated with the rigid rotation of the protein. The atoms chosen were SG17, CA97 and SG99, which unsurprisingly belong to the ICK motif, which showed particularly small fluctuations. Two vectors were then defined: vector \mathbf{v} connecting the first two and vector \mathbf{u} connecting the second two. The cross product between \mathbf{v} and \mathbf{u} was then taken and called vector \mathbf{k} . \mathbf{u} and \mathbf{k} angles with the z-normal to the membrane then become ψ and θ .

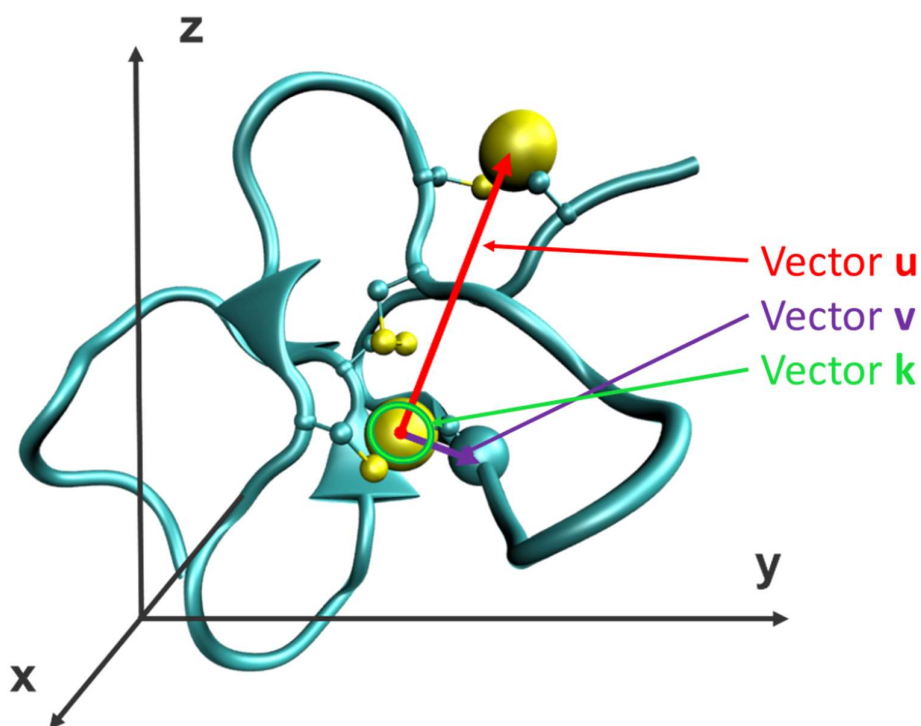


Figure 13 Vectors that define θ and ψ . Vector **u** in red, vector **v** in purple and vector **k** in green

2.2.4 Orientation angle heat maps

The angles ψ and θ were then measured for each configuration of the protein molecule sampled during the simulations and plotted as a heatmap. The probability of each pair of angles was calculated, collected into a matrix and reweighted to account for the non-uniform distribution of the angles in 3D (described in section 2.2.5). In the reweighted distributions, a perfectly uniform probability distribution of angles would appear evenly coloured on a heat map. Any deviation from this distribution reflects an increased probability of sampling a particular set of angles more often than others. A Gaussian 2D filter of 0th order with $\sigma = 2$ obtained from the standard scientific Python library *ndimage*¹⁴³ was then applied to smooth the resulting matrices to yield heat maps that are less rugged, due to the finite amount of simulation trajectory frames at our disposal.

2.2.5 Correction for angle probability heat maps

To account for the non-uniform distribution of the angles in 3D, the probability heat maps needed to be re-weighted. This implies that the real probability that we measure when constructing the matrix as described in the previous section has to be weighted by a factor $F(\psi, \theta)$ to account for the non-uniform distribution of angles that a couple of orthogonal vector forms in 3D.

Let us consider the COM of a molecule as fixed in space and the reference axis for θ and ψ as freely rotating. In other words, let us treat the angles θ and ψ as the angles that two fixed orthogonal vectors form with a uniformly random direction. Let the fixed x and y axes point along the two orthogonal vectors. A uniformly random unit direction vector has independent uniform distributions of z over $[-1, 1]$ and of ϕ (the angle about the z axis) over $[0, 2\pi]$. Here the z axis is not the random reference axis, but the fixed coordinate axis orthogonal to the x and y axes. The transformation from z, ϕ to x, y is

$$x = \sqrt{1 - z^2} \cos \phi$$

$$y = \sqrt{1 - z^2} \sin \phi$$

Thus, taking the Jacobian of the transformation

$$\frac{\partial(x, y)}{\partial(z, \phi)} = \begin{vmatrix} \frac{\partial x}{\partial z} & \frac{\partial x}{\partial \phi} \\ \frac{\partial y}{\partial z} & \frac{\partial y}{\partial \phi} \end{vmatrix} = \sqrt{1 - x^2 - y^2}$$

The further transform according to $x = \theta$ and $y = \psi$ has the Jacobian

$$\frac{\partial(x, y)}{\partial(z, \phi)} = \sin \theta \sin \psi$$

So overall we have

$$\frac{\partial(x, y)}{\partial(z, \phi)} = \frac{\sin \theta \sin \psi}{\sqrt{1 - \cos^2 \theta - \cos^2 \psi}} = 1/F(\psi, \theta)$$

This is the factor that is needed to be divided by to normalise the angle distribution, since the density for z and ϕ is flat (as these variables are independently uniformly distributed).

2.2.6 Calculation of the potential of mean force (PMF)

The PMF along the reaction coordinate (i.e. the z coordinate) was reconstructed using the WHAM algorithm^{3,144} implemented in the GROMACS g_wham tool³ on the full length of the simulation. The error was calculated using a Bayesian bootstrapping analysis. In this technique, the complete histograms of the collective variable are taken as independent points. Every histogram is composed by N samples (e.g. the reading of the collective variable) with mean m_k . Then a random probability is assigned to every sample k belonging to the initial histogram, and the mean is recalculated. This

procedure is repeated $n_bootstraps$ number of times. The $n_bootstrap$ means constitute a new set of samples that is used to obtain an improved standard deviation and mean.

The Bayesian bootstrapping for the calculation of errors in this work was performed using $n_bootstraps = 200$.

2.2.7 Calculation of the area per lipid (APL)

The APL was calculated using the equation

$$APL = \frac{(x * y)}{N}$$

where x and y are the respective dimensions of the box, and N is the number of lipids per leaflet (in this case 64). This will yield the approximate area per lipid for each lipid.

2.3 Results

The aim of this work was to investigate the fitness of classical US simulation to obtain a representative PMF in a protein-membrane system. In an ideal case the simulation should be able to comprehensively explore all the available degrees of freedom to a given macroscopic state and not spend overly lengthy amount of time stuck in local energy minima. This has to happen while conserving the structural features of ProtX-1, which is known not to undergo major secondary structure rearrangements. As the structure of the protein is pretty stable in time, in this work a particular interest is given to the orientational sampling of the protein in reference to the bilayer surface. If the orientation of the molecule are exhaustively sampled, conversion between different states are observed and the structure of the protein is constant in time we can expect a PMF that is representative and therefore a good estimation of the free energy of binding to a POPC bilayer.

In this work we used the difference between the free energy of the protein in bulk and the free energy of the protein at the minimum of its interaction well to estimate the free energy of binding. This methods does not map exactly to an binding free energy, but it provides a useful and quick proxy for it, especially when the PMF is monotonically decreasing from bulk to the membrane surface.

Figure 14 shows the RMSD values for ProtX-1 as function of time for a window far from the membrane (at a COM distance of 5.0 nm) and a window close to the membrane (at a COM distance of 2.4 nm) using the initial frame as a reference structure. The window at 5.0 nm can be considered converged since the beginning, whereas the window at 2.4 nm has a slow drift that stabilises at around 120 ns. The predicted PMF for the interaction of ProTx-1 with a POPC bilayer is shown in Figure 15.

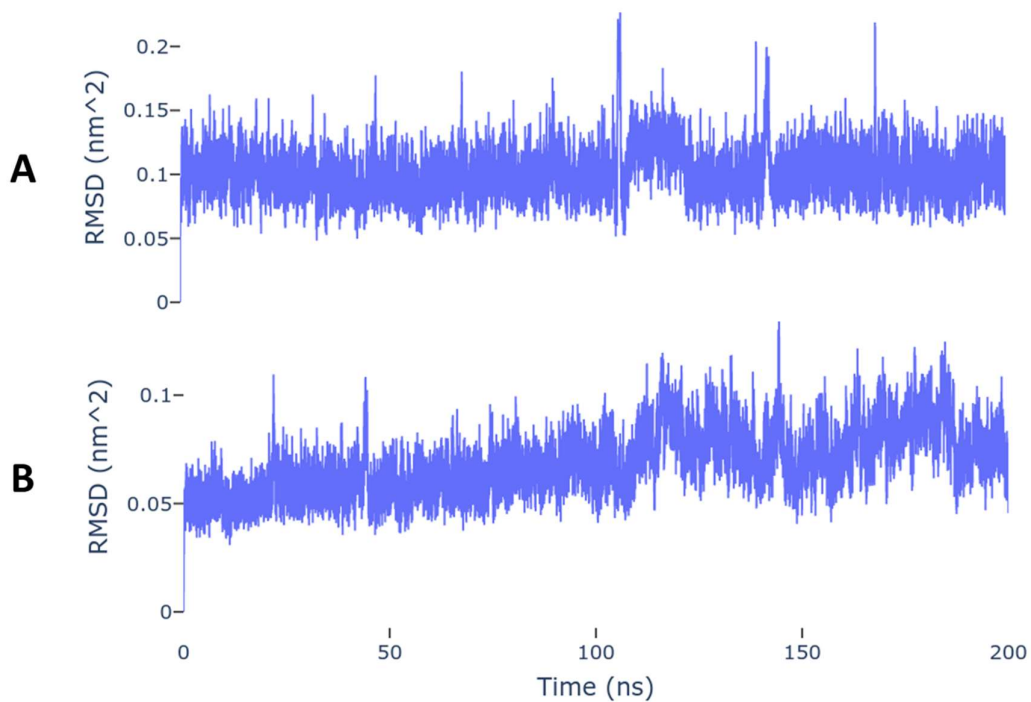


Figure 14. Time evolution of the protein RMSD during the US simulations using the initial frame as a reference structure. The values of RMSD are on average well under 0.2 nm^2 , indicating that a stable structure of the protein was achieved after the NPT equilibration. A: RMSD for window at 5.0 nm . B: RMSD for window at 2.4 nm . The predicted PMF for the interaction of ProTx-1 with a POPC bilayer is shown in Figure 15.

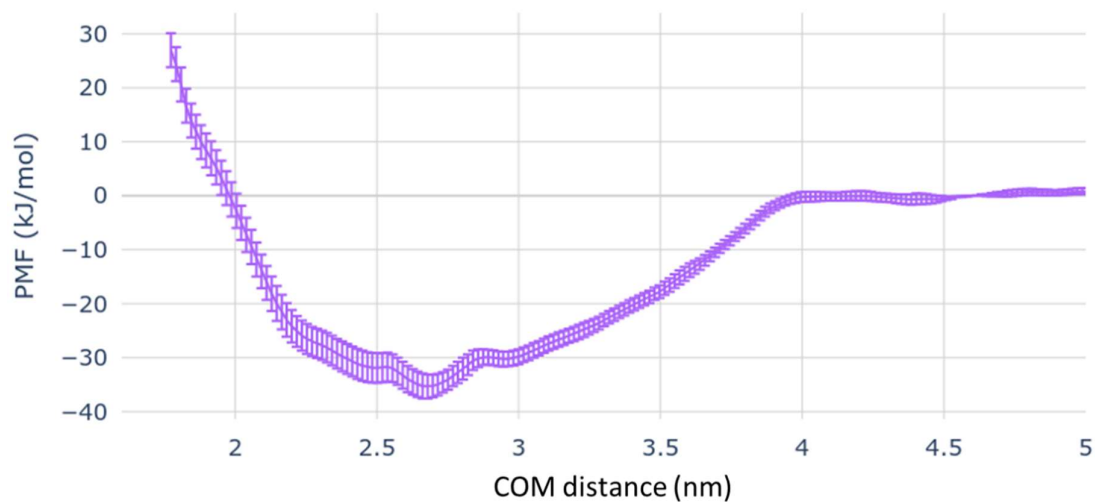


Figure 15. Predicted PMF along the centre of mass z distance for the interaction of Protx-1 with a POPC bilayer.

The PMF shows the overall shape expected for a membrane surface binding peptide. Far away from the membrane, at COM-distances >4.0 nm, the PMF is flat, indicating that the protein does not interact with the membrane. At COM distances < 4.0 nm, the interaction between the peptide and the membrane becomes more favourable, which is reflected in a drop in free energy. The minimum in the PMF is situated around 2.7 nm, with a relative free energy of -35 ± 2 kJ/mole, after which the free energy rises rapidly, reaching positive values at 2.0 nm. This is consistent with Trp fluorescence experiments, which suggest that Protx-1 does not penetrate the membrane but binds only superficially¹²⁸. As noted in Chapter 1, the binding affinity of ProTx-1 is not experimentally available, but it is known from SPR experiments that Protx-1 binds to POPC membranes stronger than Protx-2 does, whose binding affinity is in the micromolar range¹²⁸. This corresponds to a binding affinity of <-32 kJ/mol. If Protx-1 is to be a stronger binder it should be expected to exhibit a free energy of binding that is more negative than the aforementioned. The predicted value therefore falls in the expected range for this peptide.

Upon visual inspection of the simulation trajectory, it became obvious that once the protein encounters the POPC – water interface it “locks in” its relative orientation state at the time of first contact, i.e. it does not sample any other relative orientations, even transiently. To better quantify this lack of orientational sampling the orientation angle probability heat maps were computed. The orientation angle heat maps for every window of the simulation were obtained. Figure 16 reports the angle heat maps for the last 50 ns of windows for z COM distances of 2.6 nm (i.e. at the membrane-water interface) and 5.0 nm (in the bulk solution).

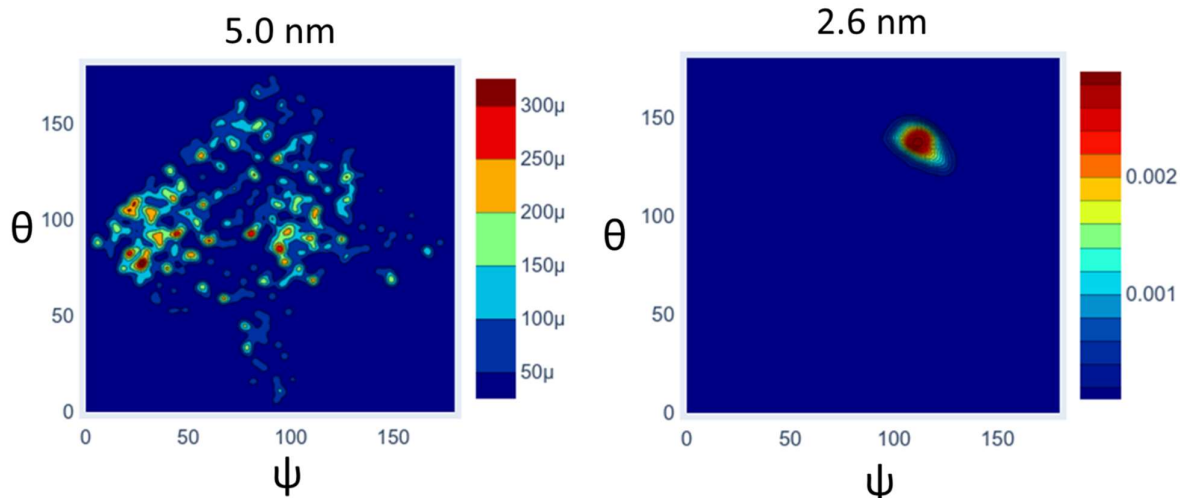


Figure 16 Orientation angle heat maps for Protx-1 for windows centred at 2.4 nm and at 5.0 nm

The orientational space sample is distributed in a diamond shape with the corners touching 90° on the four sides of the blue squares. This reflects the fact that a pair of fixed orthogonal vectors in 2D cannot explore every possible point in space because part of the orientational angle space is precluded to them. The distribution in the angle map for the window at 5.0 nm indicates that far from the membrane the protein is free to rotate and explore most of the orientational space available. In contrast, when the protein is near the surface (i.e. at a distance of 2.4 nm), a lack of sampling is indicated by the explored region on the angle heat map being limited to a narrow region around $\theta = 145^\circ$, $\psi = 115^\circ$. For a free energy estimate to be accurate the system must be able to explore all the relevant configurational space accessible to the system at the given macroscopic state, including high-energy configurations. Figure 16 suggests instead that the protein becomes locked in one relative energy minimum which is never left during the entire 200 ns simulation of this umbrella window in this particular example.

The configuration explored in this particular simulation might have been the result of the initial conditions used to build the system despite the effort to create umbrella windows incrementally. To assess if this was the case, a branched US simulation with 3 more branches was set up as described in the Methods section. These simulations were set up to exacerbate the initial condition bias in every branch, such that the initial orientations of the protein for each of the umbrella windows were taken from the respective pull simulation. Figure 17 shows the resulting predicted PMFs.

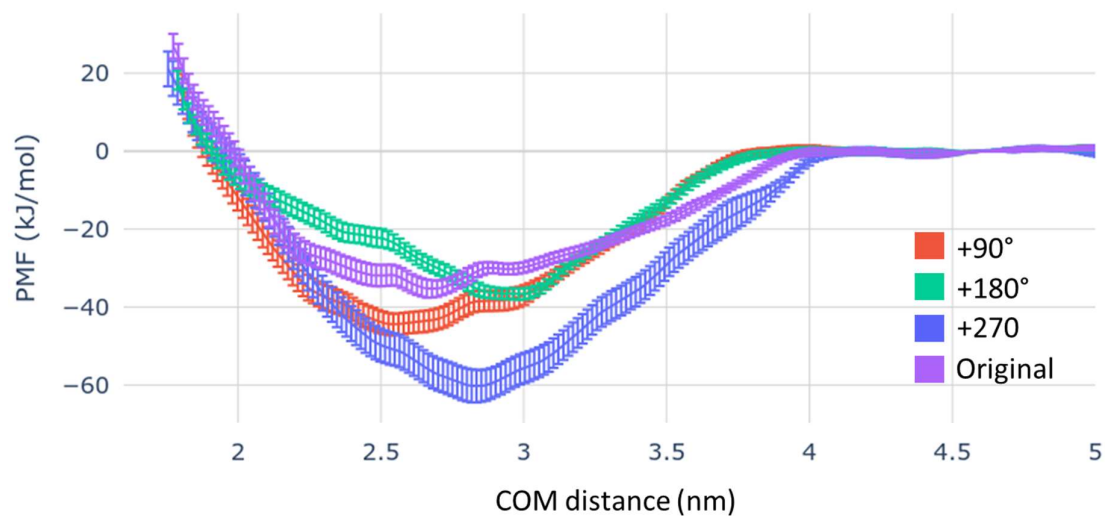


Figure 17. Predicted PMFs along the centre of mass z distance for the interaction of ProtX-1 with a POPC bilayer for the four branched US simulations.

Comparison of the predicted PMFs from the four US branched simulations shows that they are all different in shape and differ in the location of the energy minimum along the COM-distance, as well as the value of the relative free energy of binding. The location of the free energy minimum range from 2.5 nm to 3.0 nm in the COM distance. This is due in part to the plasticity of the membrane, whose thickness and APL oscillate locally significantly during the course of the simulation, as shown in Figure 18, and in part because of the different orientation of the peptide in different windows.

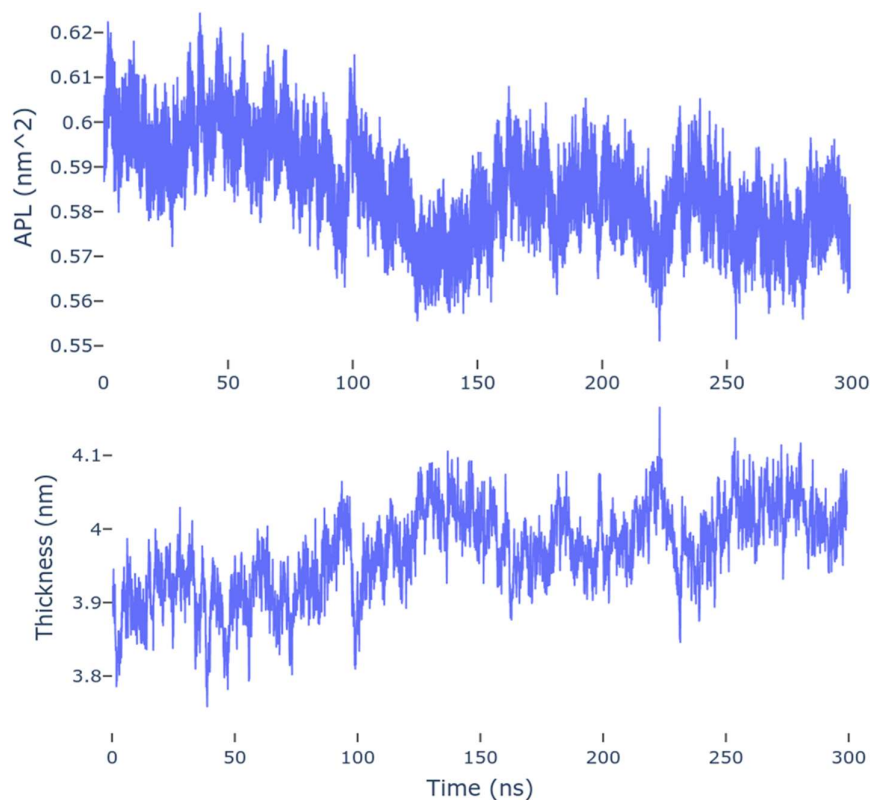


Figure 18. Time evolution of the thickness and APL of a POPC membrane during the simulation of a pure POPC bilayer in water.

The predicted relative free energy of binding values for the minimum are -35 ± 2.2 kJ/mol for the original branch, -60 ± 4 kJ/mol for system A, -36 ± 1.4 kJ/mol for system B and -44 ± 2.8 kJ/mol for system C. To determine if the protein explored new energy minima in its orientational space or fell back to the same minimum as observed in the original simulation, the orientational angle heat maps for windows at COM distances of 2.4, 2.6 and 2.8 nm for each branch of the umbrella branches were extracted. These distances correspond to the approximate location in which the free energy minimum can be found across the different branches of the umbrella. The joined angle maps are reported in Figure 19. The original branch samples 3 different minima with (θ, ψ) values of (125,60)(130,110) and (120,110). System A samples 2 minima at (120,50) and (120,145). System B explores 2 minima at (75,35) and (150,110). System C explores one deep minimum at (78,30) and 2 smaller minima at (78,45) and (105,42). From Figure 19 is then clear that although there are no clear correlations between the starting structure from where the initial conformations were dragged onto the membrane, the orientational space explored by the three simulation branches is substantially different. This suggests that that the orientation of the peptide with respect to the membrane plays a critical role in determining the magnitude of its interaction with the membrane and the distance at which the optimum interaction occurs. This also suggests that it is very likely that there are other

binding poses with different orientations of the peptide that were not sampled in these four simulations.

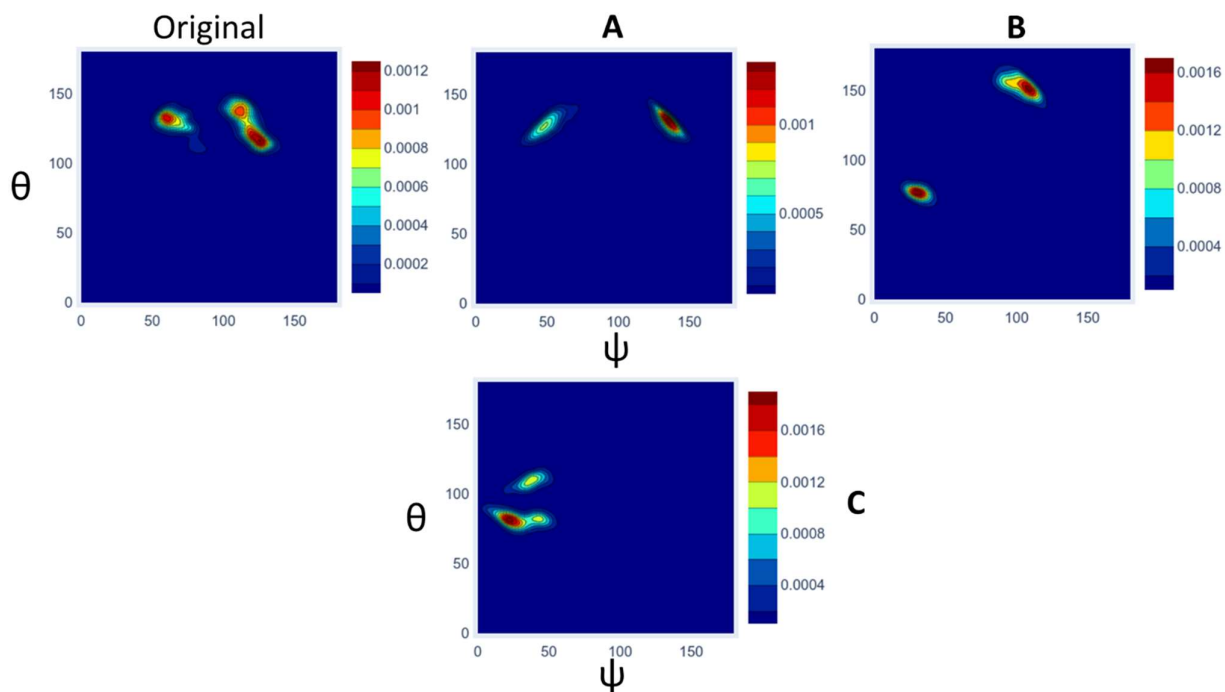


Figure 19 Orientational angle heat maps for the four different US branched simulations calculated at COM distances $2.4 < dz < 2.8$

If this is the case, the free energy estimates cannot be accurate because the system did not access all the available orientational space at each distance dz .

This suggests that classical one-dimensional US simulations are not suitable to obtain accurate predictions of binding affinity for small proteins such as Protx-1 interacting with a membrane. Given that this is a small and rigid peptide that only shows surface binding to a mono component membrane, making it likely that for larger, more flexible peptides/proteins and more complex membranes the sampling problem might be more severe.

The simplicity and robustness of US as an enhanced sampling approach is nonetheless still desirable. In the next chapters we describe a new approach aimed at solving this problem by combining US with a different enhanced sampling approach to enhance the sampling across umbrella windows. The

influence of improved orientational sampling provided by these enhanced sampling methods will be tested against the predictions made by the branched US approach described in this Chapter.

3 Chapter 3 - Selective replica-exchange with solute tempering (REST3) for the prediction of the interactions of small molecules with lipid bilayers

3.1 Introduction

Despite their ubiquitous presence and importance, characterizing small molecule-membrane interactions (SMMIs) remains challenging. This is because the diversity in lipid composition and fluidity of biological membranes gives rise to complex and dynamic supra-molecular structures^{9,145–147}. Biological membranes also feature diversity in their lateral and lamellar arrangement, as well as asymmetry between the inner and outer leaflet. The interaction of small molecules with membranes adds an extra layer of complexity and poses its own unique challenges. These interactions can range from simple surface binding to permeation or passive diffusion to more complex mechanisms such as membrane destabilization involving pore formation or carpeting effects. To reduce the complexity of the system, biomolecular simulation and biophysical characterization of SMMIs often use simple lipid bilayer systems to mimic the complexity of biological membranes.

Molecular dynamics (MD) simulations have been used extensively to predict structural, dynamical and morphological properties of lipid membrane systems^{62–66}. This technique is particularly useful when the properties of interest are inaccessible or difficult to access by experimental techniques⁶⁷. MD simulations are also used to calculate the free energy of binding, diffusion and permeation of small molecules across lipid membranes, as outlined in several recent reviews^{7,90,148–151}. Insufficient sampling affects the accuracy of dynamical and thermodynamic properties including the free energy. For SMMIs, the energy barriers associated with the roto-translational motion of solute molecules at the surface or within the membrane can limit sampling. Another problem is the energy associated with the rearrangement of membrane structure upon the interaction with solute molecules, which usually exceeds thermal (kT) energy. As a result, systems can be kinetically trapped in a local energy minimum for long periods of time, leading to a state of quasi-ergodicity, thus affecting the accuracy of the calculations⁹².

In most simulation systems, sufficient sampling of configurations can only be achieved using enhanced sampling methods^{90,152}. Replica exchange with solute tempering (REST) is one such method that is suitable for simulations of SMMIs. REST is a Hamiltonian exchange technique where the scaling of the Hamiltonian of the system is done for the solute interactions while keeping the solvent self-interactions unscaled⁹⁸. The reasoning behind REST resides in defining and enhancing sampling of a portion of the phase space that includes all of the degrees of freedom (DOFs) with a relevant weight

on the free energy. Typically, enhancing the sampling of all the DOFs associated with a solute will improve the calculation of the free energy such that it is closer to the actual free energy of the entire system. REST is thus significantly more computationally efficient than other replica exchange methods that increase the sampling of all DOFs.

REST2 is a variant of REST where changes to the scaling factors further increase the efficiency of the method⁹². Initially, REST2 was developed for simulations of protein folding, where the main sampling difficulties arise from the energy barriers associated with the rotation of the dihedral angles in the protein backbone. Scaling of the self-interaction terms of the protein (the solute) increases the sampling of the relevant portion of phase space at a small computational cost⁹⁸. In practice, this consists of altering the dihedral angles, LJ and charge parameters of the atoms of the protein only^{98,153}. In this way, the protein self-interactions are scaled, while the interactions with the other components in the simulation box are only indirectly affected to a much-reduced extent⁹². This method works well for simulations involving binary systems involving one “solute” group and one “solvent” group.

In contrast, in an SMMI system, at least three components in the simulation system need to be considered separately: the solvent (usually water and any ions present), the lipids making up the membrane, and the small molecule (the solute). The long-lived, favourable interactions between the solute and the solvent and/or the membrane can affect the sampling of the roto-translational space of the solute. As described above, scaling only the solute terms (dihedrals, LJ parameters and charges) would have a marginal effect on the sampling of the interactions between the solute and other system components. Furthermore, for peptides with a well-defined secondary structure, which do not undergo conformational changes during their interactions with a lipid bilayer, it is not desirable to scale their self-interaction terms in the Hamiltonian. Consequently, scaling solute terms is unlikely to be suitable for increasing the sampling of SMMIs.

Sampling efforts are thus required to be focused on the interactions between the solute and the membrane. However, separate scaling of the solute-solvent and solute-bilayer interactions is not possible within the REST2 formulation. To address this, we have developed a modified formulation of REST2 that enables the selective scaling of the non-bonded terms in the Hamiltonian (electrostatic and van der Waals interactions). Specifically, these modifications allow scaling of the interactions for specific pairs of components while leaving the Hamiltonian terms of the other pairs of components unperturbed. This approach is henceforth referred to as selective REST (REST3).

We tested REST3 using a set of three terpene molecules binding to a phospholipid bilayer. The three terpenes are limonene (LIM), perillaldehyde (PALD) and deprotonated perillic acid (DPAC).

3.1.1 Theoretical derivation

In REST2, the potential energy (E) of a replica is divided into three distinct contributions, each scaled by a separate scaling factor^{92,98} defined by

$$E(X) = \frac{\beta_i}{\beta_0} E_{mm}(X) + \sqrt{\frac{\beta_i}{\beta_0}} E_{ms}(X) + E_{ss}(X) \quad (17)$$

where X indicates the position coordinates of the system, E_{mm} is the small molecule self-interaction energy, E_{ms} is the small molecule-solvent intermolecular interaction energy, E_{ss} is the solvent-solvent intermolecular interaction energy and β_i is $1/k_b T_i$, where i refers to the replica number. For the ground (unscaled) replica $i = 0$ and all other replicas, the scaling factor for each term is related to the virtual temperature (β_i) of replica i . Using the extended detailed balance condition¹⁵⁴ given by

$$p_i(X_i)p_j(X_j)p(X_i \rightarrow X_j) = p_i(X_j)p_j(X_i)p(X_j \rightarrow X_i). \quad (18)$$

Considering that the probability of finding the system with Hamiltonian i and coordinates X_i is given by

$$p_i(X_i) = e^{-\beta_i E_i(X_i)} \quad (19)$$

equations 19 and 17 can be substituted into equation 18 to obtain the transition probability ratio

$$\frac{p(X_i \rightarrow X_j)}{p(X_j \rightarrow X_i)} = e^{-\Delta_{ij}} \quad (20)$$

where

$$\Delta_{ij} = (\beta_i - \beta_j) [E_{mm}(X_j) - E_{mm}(X_i)] + \frac{\sqrt{\beta_0}}{\sqrt{\beta_i} + \sqrt{\beta_j}} (E_{ms}(X_j) - E_{ms}(X_i)) \quad (21)$$

In REST3, the solvent term can be divided into multiple components as required (e.g., membrane, water, ions, etc.), with each cross-term scaled independently. Furthermore, each of the energy terms in equation 17 is further divided into its underlying energy contributions (bonded, electrostatic and vdW). This allows controlling the strength of the different types of interactions. For example, membrane-membrane and membrane-solvent interactions can be left unperturbed, while electrostatic interactions of the molecule-membrane component may be selectively scaled. In membrane-molecule systems such control is critical because a small change in membrane-membrane interactions may become amplified by the large magnitude of these interactions, leading to destabilization of the entire membrane. At the same time, full membrane-peptide interactions cannot be retained in classical REST2 implementation without reducing peptide-peptide interactions to almost zero. It is then crucial to decouple the scaling of these two contributions flexibly.

Let us take one of the terms of the Hamiltonian in the case of a small molecule-solvent system, the intermolecular interaction energy E_{ms} , which is defined as:

$$E_{ms}^{REST3} = \lambda_{ms,bond} E_{ms,bond} + \lambda_{ms,elec} E_{ms,elec} + \lambda_{ms,vdW} E_{ms,vdW} \quad (22)$$

where each $\lambda_i = \beta_i/\beta_0$ is the associated scaling factor for the corresponding energy contribution. It follows that the Δ_{ij} for REST3 does not depend on the terms that are left unscaled in the equation. For example, if all of the λ parameters are set to 1 except for the small molecule-solvent intermolecular terms, the resulting Δ_{ij} simply reduces to:

$$\Delta_{ij}^{REST3} = (\beta_i - \beta_j)[E_{ms(x_j)} - E_{ms(x_i)}] \quad (23)$$

As a result, the terms of the equation that are to be ignored by the replica-exchange method (in this case, all terms except the small molecule-solvent terms) are left unscaled with $\lambda = 1$, enhancing the exchange probability between two replicas several-fold in the case of systems where the energy is dominated by the solvent. An outline of the algorithmic and practical implementation of this method in GROMACS 4.6.7 can be found in Section 1 of the Supporting Information (SI).

3.2 Methods

All conventional umbrella sampling (US) simulations were performed using GROMACS 4.6.7¹⁵⁵, whereas all the US-REST3 simulations were performed using an in-house version of GROMACS 4.6.7. For the US-REST3 simulations, the exchange attempt frequency was chosen to be 0.1 ps, in line with the findings by Sindhikara et al.¹⁵⁶. Several different exchange attempt intervals, ranging from 0.1 ps to 500 ps. The replica exchange probabilities slightly changed for the worst with higher exchange attempt times, but overall, there was no significant difference in the final ground replica ensembles. The GROMOS 54A7 compatible lipid parameters^{85,157} and the SPC model of water¹⁵⁸ were used. All simulations were performed using rectangular periodic boundary conditions and a time step of 2 fs, removing the COM motion of the system at every step. The LINCS algorithm¹⁵⁹ was used to constrain the lengths of all bonds. Non-bonded interactions were evaluated using a single-range cut-off scheme, whereby interactions within a 1.4 nm cut-off were calculated at every step, and the pair list was updated every five steps. To account for electrostatic interactions beyond the cut-off, a reaction field¹⁶⁰ was applied with a relative dielectric constant (ϵ_r) of 62. Note that both GROMOS 54a7 parameters and the GROMOS-compatible lipid FF were developed with this scheme and have been shown to reproduce the structure and dynamics of a wide range of phospholipid bilayers^{84,85}. Trajectories were visualized using VMD 1.9.2¹⁶¹ and analysed using standard GROMACS tools¹⁶² and custom Python scripts based on MDAnalysis tools¹⁶³.

3.2.1 Interaction of terpenes with a DMPC bilayer

Topologies for the three terpenes were obtained from the ATB⁸⁶ (LIM Molid: 100030, PALD Molid: 87548, DPAC Molid: 229709). For each simulation system, a single terpene molecule was added ~ 3.6 nm away from a pre-equilibrated DMPC bilayer consisting of 128 lipid molecules. The DMPC bilayer, equilibrated for 235 ns at 303 K was obtained from the ATB (taken from the ATB Molecule box ID 29⁸⁶). The systems were solvated with 9841 SPC water molecules. For the DMPC-DPAC system, the net charge was neutralized by adding a single Na⁺ ion. Each system was equilibrated in the NVT ensemble for 10 ns at 298 K using the Berendsen thermostat¹⁶⁴ with $\tau_t = 0.1$ ps, followed by another 10 ns simulation in the NPT ensemble at 1 bar using the Berendsen barostat¹⁶⁴ with $\tau_p = 0.5$ ps.

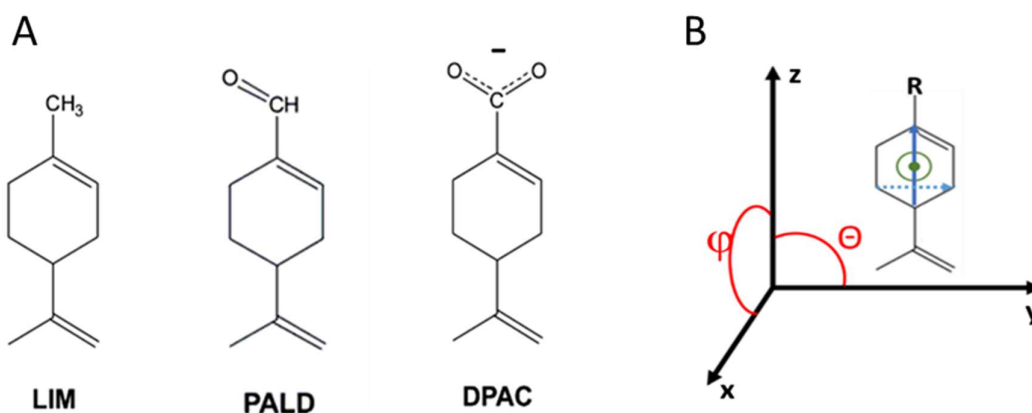


Figure 20. A) Chemical structures of the terpenes: limonene (LIM), peril aldehyde (PALD) and perillic acid (DPAC). B) The angle ϕ and θ are defined between the two continuous line vectors (blue continuous vector along the principal axis of the terpene molecule and green continuous vector perpendicular to the plane of the terpene ring) and the z-axis corresponding to the normal vector to the membrane.

3.2.2 Umbrella sampling

The z-axis component of the distance between the centers of mass (COM) of the bilayer and the terpenes was used as the reaction coordinate (RC) for the US simulations. The bilayer was situated along the x-y plane of the simulation cell, while the terpene molecule was positioned on the z-axis that passes through the center of the x-y plane at a specific distance for each umbrella window. US windows were spaced 0.2 nm from the membrane COM (0.0 nm) to the bulk solution (3.6 nm for PALD and LIM, 4.2 nm for DPAC). For PALD and LIM, a total of 18 windows were used. Since DPAC is a charged molecule, the influence of the proximity of the membrane on the interaction energies remained significant at larger distances, so three additional windows were used for a total of 21 windows. Each window was equilibrated for 40 ns at 298 K and 1 bar using the Nosé-Hoover thermostat¹⁶⁵ with $\tau_t = 0.5$ ps and the semi-isotropic Parrinello-Rahman barostat¹⁶⁶ with $\tau_p = 1.0$ ps

and compressibility of $4.6 \times 10^{-5} \text{ bar}^{-1}$. The position of each one of the terpene molecules along the RC was restrained using a harmonic potential with a force constant $k = 500 \text{ kJ mol}^{-1} \text{ nm}^{-2}$. The coordinates of the system were saved every 10 ps. After equilibration, each window was simulated until convergence was reached using the same simulation parameters described above. The convergence of simulations was assessed by plotting the PMF vs time and splitting the data in increments of 10 ns until consecutive PMFs were determined to be within a standard deviation. The combined simulation time for all the windows was 11 μs for LIM, 22 μs for PALD and 23.8 μs for DPAC (the total simulation time for each window is reported in Table S1 in the SI).

3.2.3 US-REST3

The final configurations obtained from the trajectories of each window in the US simulations were used as the starting point for the US-REST3 simulations. Every window of the US simulation is replicated 11 times and run at a different value of λ . The upper end (most scaled) values of $\lambda_{ms,elec}$ (λ_e) and $\lambda_{ms,vdW}$ (λ_v) were chosen so that all the angles defined by the orientation vectors are adequately sampled (see section 3.1.5 for details of the orientation angle maps). This ensured that every possible relative configuration would be sampled in the highest virtual temperature replica, while the replica-exchange routine ensured that these configurations “trickled down” to the ground replica. The λ values used were $\lambda_e = \lambda_v = 0.2$ (corresponding to a virtual temperature of 1490 K). For each window, the number of replicas needed to interpolate between $\lambda = 1$ and $\lambda = 0.2$ was empirically determined running a short (1 ns) test simulation, starting with an equally spaced replica regime and with $\lambda_e = \lambda_v$. The values of λ chosen to interpolate between the unperturbed Hamiltonian and $\lambda = 0.2$ were 0.2, 0.27, 0.35, 0.4, 0.5, 0.6, 0.7, 0.77, 0.84, 0.92 and 1.0. These values were chosen arbitrarily to obtain an exchange probability between replicas i and j higher than 15%. In the case of DPAC, 11 replicas were required to reach the threshold of >15%, whereas for PALD and LIM, only four replicas were required. Nevertheless, as this was the first test of the method and for consistency, 11 replicas were used for all three terpene molecules to allow a direct comparison between systems. To test the performance of the method, a simulation of LIM using only four replicas per window ($\lambda = 0.2, 0.4, 0.6$ and 1.0) was also performed. The convergence of simulations was assessed by plotting the PMF vs time and splitting the data in increments of 10 ns until consecutive PMFs were determined to be within a standard deviation. The combined simulation time for all windows was 20.6 μs for LIM, 18.4 μs for PALD and 11 μs for DPAC (the total simulation time per window is reported in Table S2 in the SI).

3.2.4 Potential of mean force calculations

The potential of mean force (PMF) along the RC was reconstructed using the WHAM algorithm^{3,144} implemented in the GROMACS `g_wham` tool³. Bayesian bootstrapping for the calculation of errors was performed with `n_bootstraps = 200`. In this work we used the difference between the free energy

of the protein in bulk and the free energy of the protein at the minimum of its interaction well to estimate the free energy of binding. This method does not map exactly to a binding free energy, but it provides a useful and quick proxy for it, especially when the PMF is monotonically decreasing from bulk to the membrane surface.

3.2.5 Water-terpene interactions

The interactions of terpenes with water were determined by analyzing the minimum distance between the water molecule atoms and the terpene atoms using the `g_mindist` and `g_hbond` tools in GROMACS 4.6.7.

3.3 Results and discussion

This section is divided into two subsections. In the first part, we compare the predicted binding affinities to experimental values, followed by a comparative presentation of the PMFs of the three terpene molecules predicted using the conventional US and US-REST3 simulations. In the second part, an analysis of roto-translational sampling as the molecules move from the bulk solution to the membrane interior is presented. The section concludes with a discussion of insights gained into the mechanism of binding of terpenes to lipid bilayers.

3.3.1 Free energies of binding

Table 1 lists ΔG_b values predicted by the conventional US (ΔG_{b-US}) and US-REST3 simulations ($\Delta G_{b-US-REST3}$), as well as the binding affinity of the three terpenes for DMPC bilayers measured by isothermal titration calorimetry (ITC), ΔG_{b-ITC} ¹⁶⁷. For DPAC, only a very small amount was observed to partition into DMPC vesicles. The partition coefficient for DPAC ($\leq 5 \text{ M}^{-1}$) was thus much smaller than for LIM ($10,870 \text{ M}^{-1}$) and PALD (305 M^{-1}). Consequently, for DPAC the ΔG_{b-ITC} could not be reliably determined. This suggests that DPAC exhibits very weak or no binding. Overall, comparison of ΔG_{b-ITC} and ΔG_b calculated by both conventional US and US-REST3 simulations reveals that the values predicted by US-REST3 are in better agreement with experiment. For LIM, ΔG_{b-ITC} is -34.17 kJ/mol . ΔG_{b-US} is $-24.0 \pm 0.64 \text{ kJ/mol}$, giving a difference of $+10.17 \text{ kJ/mol}$. $\Delta G_{b-US-REST3}$ is instead -29.7 kJ/mol , with a smaller deviation from experiment of $+4.47 \text{ kJ/mol}$. PALD ΔG_{b-ITC} is -26.01 kJ/mol and ΔG_{b-US} is $-21.4 \pm 0.6 \text{ kJ/mol}$, giving a difference of $+4.61 \text{ kJ/mol}$. $\Delta G_{b-US-REST3}$ is -28.3 ± 0.8 , differing by only -2.28 kJ/mol from the experimental value. Finally, for DPAC ΔG_{b-US} and $\Delta G_{b-US-REST3}$ are $-3.4 \pm 0.4 \text{ kJ/mol}$ and $+1.3 \pm 0.4 \text{ kJ/mol}$, respectively, which are values consistent with binding below reliable ITC detection. Experimentally, these terpenes are predicted to have relative binding affinities for DMPC bilayers that decrease as LIM

> PALD >> DPAC. The rankings obtained by conventional US simulations and by US-REST3 simulations are in agreement with experiment. The ΔG_{b-ITC} values of the two terpenes differ by 8.2 kJ/mol, while the $\Delta G_{b-US-REST3}$ values differ by only 1.4 kJ/mol. In the next section, we discuss how enhanced sampling in REST3 alters the PMF profiles and explains the improved predictions by REST3 compared to US.

Table 1. Predicted and experimental free energies of binding of terpenes to DMPC bilayers.

| | ΔG_{b-US} (kJ/mol) | $\Delta G_{b-US-REST3}$ (kJ/mol) | ΔG_{b-ITC} (kJ/mol) |
|------|----------------------------|----------------------------------|-----------------------------|
| LIM | -24.0 ± 0.6 | -29.7 ± 0.6 | -34.17 |
| PALD | -21.6 ± 0.6 | -28.3 ± 0.8 | -26.01 |
| DPAC | -3.4 ± 0.4 | $+1.3 \pm 0.4$ | - |

3.3.2 Potential of mean force profiles

Figure 21A shows the PMFs for the three terpene molecules obtained from the conventional US. The main difference between the PMFs of LIM and PALD, compared with DPAC, is the location of the global minima. Once again in this chapter I refer to the binding energy of a terpene as the difference between the bulk value of the free energy with the absolute minimum of the PMF. This quantity does not map exactly to the experimental partition free energy, but it is a useful proxy to make comparisons between the two, especially when the PMF is monotonically decreasing from bulk to the minimum. For LIM and PALD, the global minima are at COM distances < 1.0 nm, corresponding to the membrane interior. In contrast, for DPAC, there is a broad global minimum around a COM distance 2.8 nm, which corresponds to a location outside of the membrane. This indicates that for LIM and PALD, the preferred location is inside the membrane, while for DPAC, the preferred location is in the bulk solution. This suggests that LIM and PALD exhibit much stronger binding to DMPC than DPAC, which is consistent with experiment, as discussed in more detail further below. Overall, the shapes of the PMF for all the terpenes using the two methods are very similar.

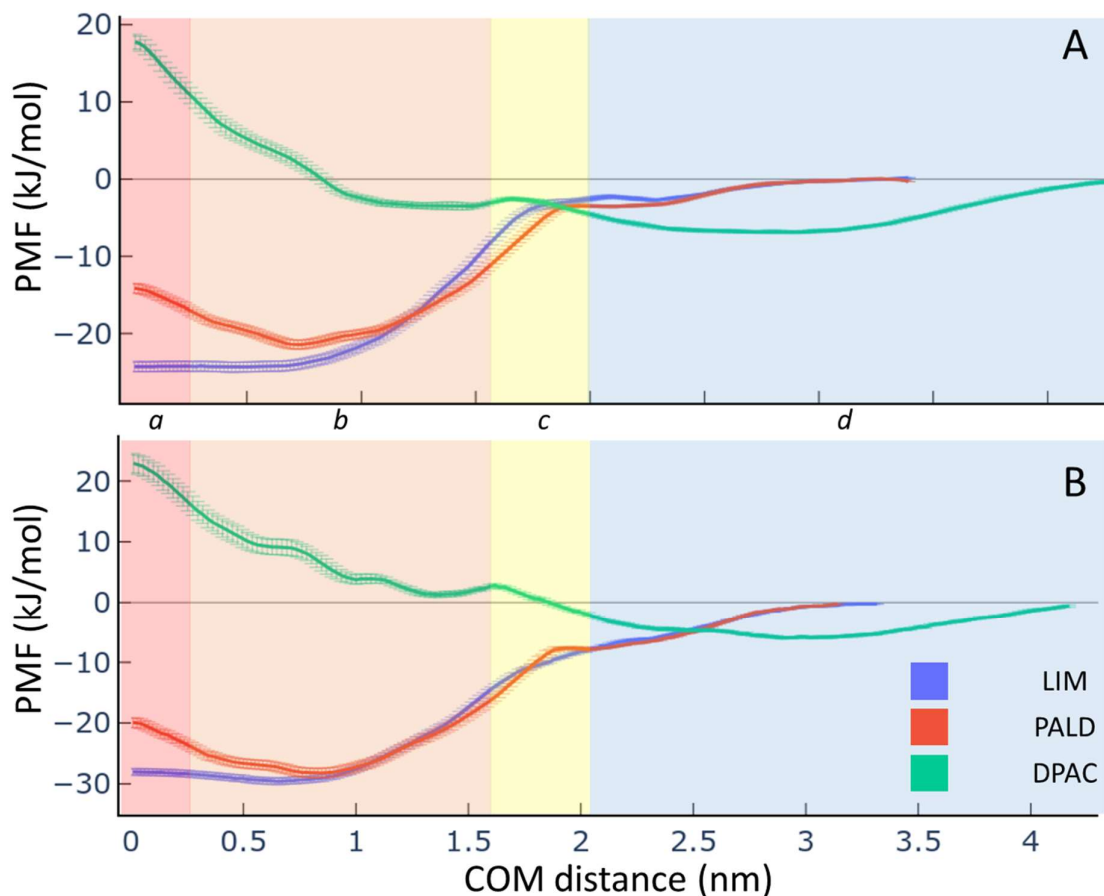


Figure 21. PMF for the interaction of the three terpenes PALD (orange), LIM (blue) and DPAC (green) with a DMPC bilayer obtained from conventional US simulations (panel A) and US-REST3 simulations (panel B). The regions of the simulation cell are indicated starting from the bulk solution (d), the membrane surface extending from 1.6 to 1.9 nm (c), the high-density hydrophobic tail region extending from 1.6 nm to 0.25 nm (b) and the low density tail region extending to 0.25 nm in the center of the membrane (a).

The shape of both PMFs for LIM reveal that the global minimum is broad and flat, extending out from the center of the bilayer (COM distance = 0) to about 1 nm. This indicates that LIM experiences favorable interactions within the hydrophobic core of the bilayer and suggests that it can penetrate deep into the membrane and eventually into the other leaflet. The large basin of this minimum is in agreement with the predictions of Witzke et al.¹⁶⁷, who observed that the majority of the terpene molecules are found situated in the high-density tail region.

For PALD, the global free energy minimum is situated at a COM distance of 0.8 nm, which corresponds to a location where its carbonyl moiety sits directly below the lipid head groups at the top of the hydrophobic core. This is consistent with predictions from MD simulations reported by Witzke et al.¹⁶⁷, which showed that the electron density profile of PALD in the membrane has a maximum at COM distances between 0.8 and 1.0 nm. This minimum is not flat as in LIM, and it shifts towards the center of the membrane (region 'a' in **Figure 21**), where the free energy rises to less negative values. Consequently, compared to LIM, PALD is less likely to penetrate through region 'a' and into the other leaflet, which is consistent with the more hydrophilic nature of PALD.

As noted above, the PMF for DPAC exhibits a global minimum that is very broad, extending from a COM distance of 2.0 to 4.0 nm from the center of the bilayer. The lowest point is at a COM distance of ~ 2.8 nm, which corresponds to the molecule being in solution. The well depth for this minimum is -6.4 ± 0.3 kJ/mol. Another local minimum is located inside the membrane at a COM distance of 1.2 nm. The well depth for this minimum is -4.4 ± 0.5 kJ/mol and is thus less energetically favourable than the main, broad minimum. There is a small free energy barrier separating these two minima. The mass density profile extracted from an independent simulation of a full hydrated DPMC bilayer (**Figure 22**) reveals that the position of this barrier corresponds to the COM distance where the choline moiety of the lipid head groups is situated (region c). For COM distances < 1.0 nm, the free energy increases sharply, indicating that penetration of DPAC into the hydrophobic core of the bilayer is highly unfavourable.

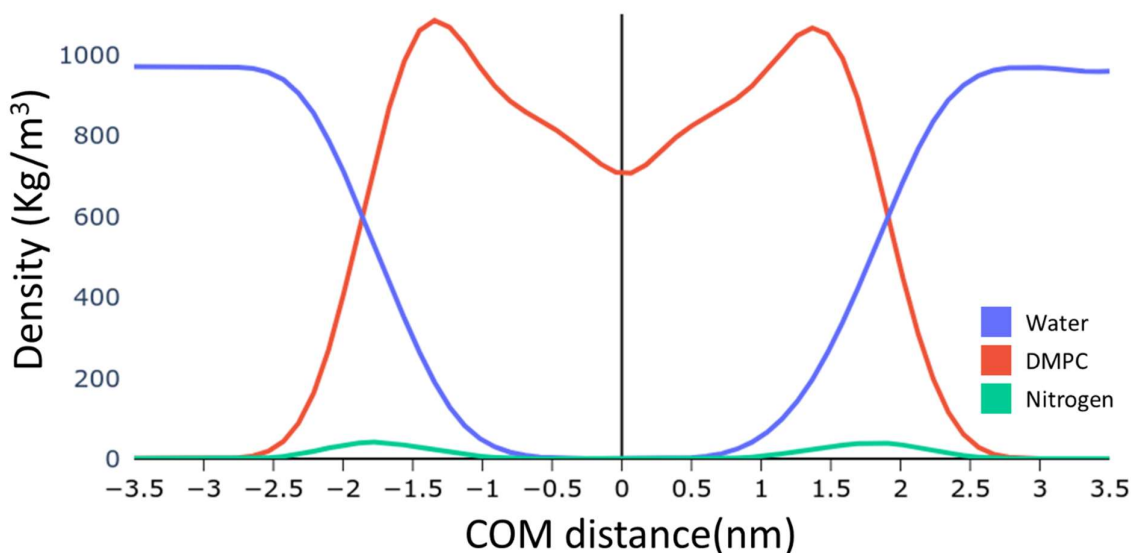


Figure 22. Mass density profile from a 200 ns simulation of a fully hydrated DMPC bilayer without terpenes. The thickness d of the bilayer (3.54 nm) was calculated as the distance between the peaks

of the nitrogen atoms of the DMPC molecules. The surface of the membrane is centered at around $d/2$ (1.77 nm) from the center of the bilayer. The choline group can extend up to 1.0 nm away from this surface.

The time required to reach convergence in each window can be found in Table S1 of the SI.

We now compare the conventional US PMFs with those obtained with US-REST3. **Figure 21B** shows the PMFs for the three terpenes calculated using US-REST3 simulations. Overall, the shape of each PMF is very similar to the ones predicted by the conventional US simulations. In the case of LIM and PALD, the location of the global minimum remains mostly unchanged, but the predicted ΔG_b differs by several energy units. In the case of LIM, the ΔG_b is -29.7 ± 0.6 kJ/mol (compared to -24.0 ± 0.6 kJ/mol with conventional US), while in the case of PALD the predicted ΔG_b is -28.3 ± 0.8 kJ/mol (compared to -21.4 ± 0.6 kJ/mol). In the case of DPAC, the PMF predicted by the US-REST3 simulations exhibits the same broad global minimum located at around a COM distance of 2.8 nm, as is also predicted by the conventional US simulations. The lowest free energy for this broad minimum is -6.3 ± 0.2 kJ/mol with US-REST3 compared to -6.4 ± 0.3 kJ/mol with conventional US. However, compared to the PMF predicted by conventional US, in the PMF predicted by US-REST3, the small local minimum around a COM distance of 1.2 nm shifts from a favourable value of -3.4 ± 0.4 kJ/mol to an unfavourable value of $+1.0$ kJ/mol. Consequently, the PMF for DPAC shows an almost consistent increase in value from COM distances of around 3.0 nm toward the center of the bilayer. Consequently, US-REST3 predicts the penetration of DPAC into the bilayer as being more unfavourable than that predicted by conventional US simulations.

3.3.3 Angle maps of the orientation of terpenes

Data from the various windows of both the US and US-REST3 simulations were also used to analyse the extent of orientational sampling for the terpenes as they approach and penetrate the bilayer. The orientation of the terpene molecules was defined by two angles (θ and ϕ) subtended by each of perpendicular vectors running along with the terpene molecules and the vector normal to the membrane surface (x-y plane), as illustrated in **Figure 20** and described in the Methods section. Orientation angle maps showing the 2D probability distribution plots of the θ and ϕ angles were constructed for each window. These angle maps thus capture the extent of sampling of the orientations of the terpenes at the membrane surface. By comparing angle maps for a given terpene molecule across different windows, the presence of preferred configurations (orientations) along the RC was determined. This allowed assessing whether the terpene molecules preferentially interact through their polar group (in the case of PALD and DPAC) or their non-polar isopropenyl tail.

Analyses of these orientation angle maps are organized as follows. First, angle maps for the conventional US simulations are presented. For each terpene molecule, analyses of three different windows are presented, representing key stages in the binding to the DMPC bilayer as revealed by the predicted PMFs: i) the terpene molecule in the bulk solution (a COM-distance of 3.6 nm for LIM and PALD, and 4.0 nm for DPAC); ii) the terpene molecule at the interface interacting with the lipid head groups (a COM-distance of 1.8 nm for all terpenes was chosen); and iii) the terpene molecule at the free energy minimum inside the membrane (a COM-distance of 0.6 nm was chosen for LIM, 0.8 nm for PALD, and 1.4 nm for DPAC). Where relevant, other interesting regions in the PMF for a given terpene molecule were also analyzed. In the second part of this section, orientation angle maps computed from the US-REST3 simulations are shown in the same order and compared with the corresponding angle maps computed from the conventional US simulations.

It is also worth to discuss the difference in PMFs predicted by WHAM in the light of the data that will be presented. The differences in the PMF between classical and REST3 simulations might be due to differences in the tails of the histograms of the COM distance. Inspection of the COM distance histograms (Figure S6 in the SI) overlaps between US and US-REST3 windows, suggests that these differences might be due to differences in statistical sampling of the number of frames with values that fall far from the umbrella center, e.g. $\sqrt{(x-\bar{x})^2} > 0.1$ nm. The stability of these states is influenced by the local environment that the terpenes experience, which in turns depend on the correspondent thickness and APL of the membrane. These quantities are known to fluctuate slowly, usually in the range of hundreds of nanoseconds for a fluid state membrane. With the shorter nature of each window in US-REST3 simulations, the amount of time that the system spends in fringe values of APL and thickness is intrinsically smaller than in an equivalent conventional US simulation. The statistical sampling of these fringe states may therefore be slightly different from that in US simulations just by virtue of the given fluctuations of the membrane sampled. Nevertheless, as it will be seen in this section, the orientations of the terpenes converge very rapidly to the same orientation maps that long classical US simulations converge to. Conventional US simulations take time ranges ranging from 200 to 600 ns to converge to their final orientation distributions and which US-REST3 converges to.

Figure 23 and **Figure 24** show the orientation angle maps for the above three representative COM-distances for each of the terpene molecules, respectively, in the conventional US and US-REST3 simulations. Both methods provide similar results for the orientation of the terpenes. For all three terpenes, it is clear that all possible orientations are uniformly sampled in the bulk solution. This is expected since all three terpene molecules can rotate freely when they are located far from the membrane surface. Once the terpene molecules reach the membrane surface, the interactions with the lipid head groups hinder this free rotation, and, as a result, certain orientations are sampled more

than others. This is evident at a COM-distance of 1.8 nm for LIM, where θ angles around 175° show a higher probability. A θ angle of around 90° corresponds to an orientation where the LIM molecule lies sideways with both its methyl and isopropenyl groups interacting with the head groups. This orientation is sampled in the first 160 ns of the simulation (see Fig S1a in the SI). A θ angle closer to 0° or 180° instead corresponds to an orientation where the main axis along the molecule is parallel to the z-direction (normal of the membrane). The sideways orientation is unexpected because the angle maps computed from all other windows where the LIM molecule can interact with the lipids (window < 2.0 nm) show that the most likely θ angle is either around 15° or $\sim 175^\circ$. These angles correspond to an orientation whereby the axis of the molecule aligns with the bilayer normal, exposing to the solvent either the isopropenyl or the methyl moieties.

Interestingly, for the first 200 ns of the US simulation, the orientations around $\theta = 15^\circ$ and $\theta = 175^\circ$ seem to be almost equiprobable, whereas extending the simulation to 500 ns results in a net preference for the orientation at $\theta = 175^\circ$. This behavior is conversely immediately captured by the US-REST3 simulation, suggesting that tempering the terpene-membrane interaction in a region of space where rotation is sterically impeded can help the interconversion between different orientations. In summary, the behavior of LIM on the surface of the membrane preferentially orients the methyl group toward the center of the membrane.

PALD exhibits strong orientational preferences at the membrane surface, with a clear preference for θ angle values around 10° . This corresponds to an orientation whereby the polar aldehyde ($\text{CH}=\text{O}$) group points towards the bulk solution, and the isopropenyl group points towards the membrane. This is expected because the aldehyde group is polar and preferentially solvated by water, while the isopropenyl group can form hydrophobic interactions with the hydrophobic phospholipid tails of the membrane.

DPAC shows similar behavior to PALD close to the membrane surface, exhibiting a broad distribution that peaks at θ values of around 10° . This angle corresponds to an orientation where the charged carboxyl group is exposed to water, and the isopropenyl group is oriented towards the membrane. However, in contrast to PALD, for DPAC, the distribution of θ values is much broader, ranging from 0° to 100° , rapidly collapsing to a narrower peaked distribution when the COM distance reduces from 1.8 to 1.6 nm, as shown in the angle map for this shorter COM distance (Figure S1f in the SI). This suggests that when DPAC reaches the surface of the membrane, where it can still interact with water, its orientation is less constrained. When DPAC inserts into the membrane and crosses a small free energy barrier, the molecule aligns such that the molecular axis is perpendicular to the membrane

normal. In this orientation, the charged carboxylate group is exposed to water, and the isopropenyl group and ring on DPAC interact with the phospholipid tails.

Once inside the membrane, the orientation angle maps at the specified COM distances correspond to a minimum in the PMF for the given terpene molecule. For LIM, this corresponds to the global free energy minimum at a COM distance < 0.8 nm, in which it is located in the hydrophobic region of the membrane. The orientation angle map shows a bimodal distribution for θ at around 20° and 160° , indicating that the principal axis of the molecule prefers to stay aligned with the lipid tails to minimize lipid dislocation. Similarly, for PALD, the global free energy minimum is at a COM distance of 0.8 nm in the hydrophobic core of the membrane. The orientation angle map shows a narrow distribution for angle θ centered around 20° , indicative of the principal axis of the terpene molecule in alignment with the normal of the membrane. Compared to the angle map of LIM at the membrane surface, the peak is broader. In the case of DPAC, the orientation angle map inside the membrane was computed at a COM-distance of 1.4 nm, corresponding to a local free energy minimum in the PMF (the global minimum is outside of the membrane). A narrow peak for angle θ at 10° , similar to PALD, reveals that the principal axis of the terpene is aligned with the normal of the membrane because its charged carboxylate group remains solvated by water molecules that infiltrate the membrane.

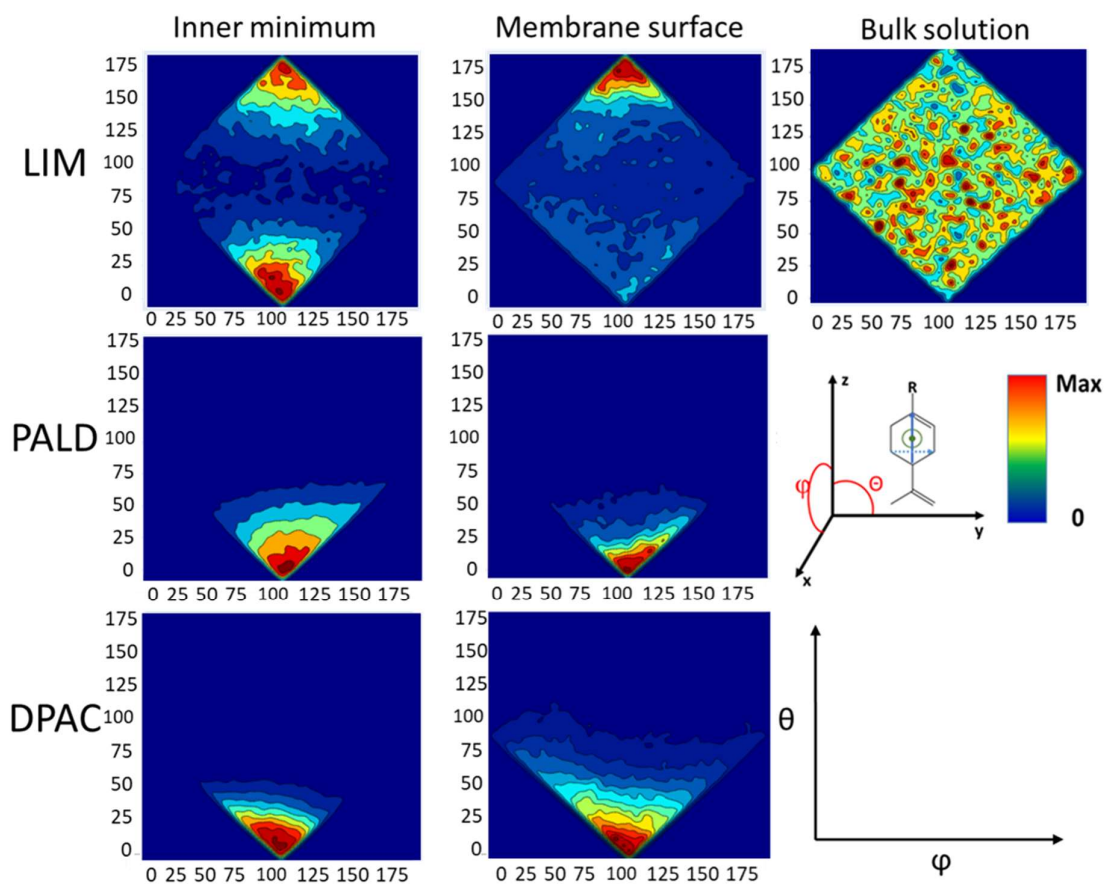


Figure 23. Orientation angle heat maps of the terpene molecules in conventional US simulations. Orientation angle maps are 2D probability plots of the θ and ϕ angles that define the orientation of the terpene molecules with respect to the vector normal to the membrane surface. Angle θ is reported on the y-axis and angle ϕ on the x-axis, as represented in the bottom right panel. These angle maps were used to assess the sampling of orientations visited by the terpene molecule in representative windows at three different COM distances to the bilayer. Orientation angle maps are shown for the inner minimum (left), the membrane surface (middle) and the bulk solution (right). The COM distance to the membrane surface was assumed to be around 1.8 nm for all the three terpene molecules, whereas the COM distance to the inner minimum is different for each terpene (0.6 nm for LIM, 0.6 nm for PALD and 1.4 nm for DPAC). The terpene molecules in the bulk solution all exhibited very similar orientation angle maps (SI Figure S1), so only the angle map for LIM at a distance of 3.6 nm is shown. The maximum probabilities are of the same order of magnitude but are slightly different for each heat map.

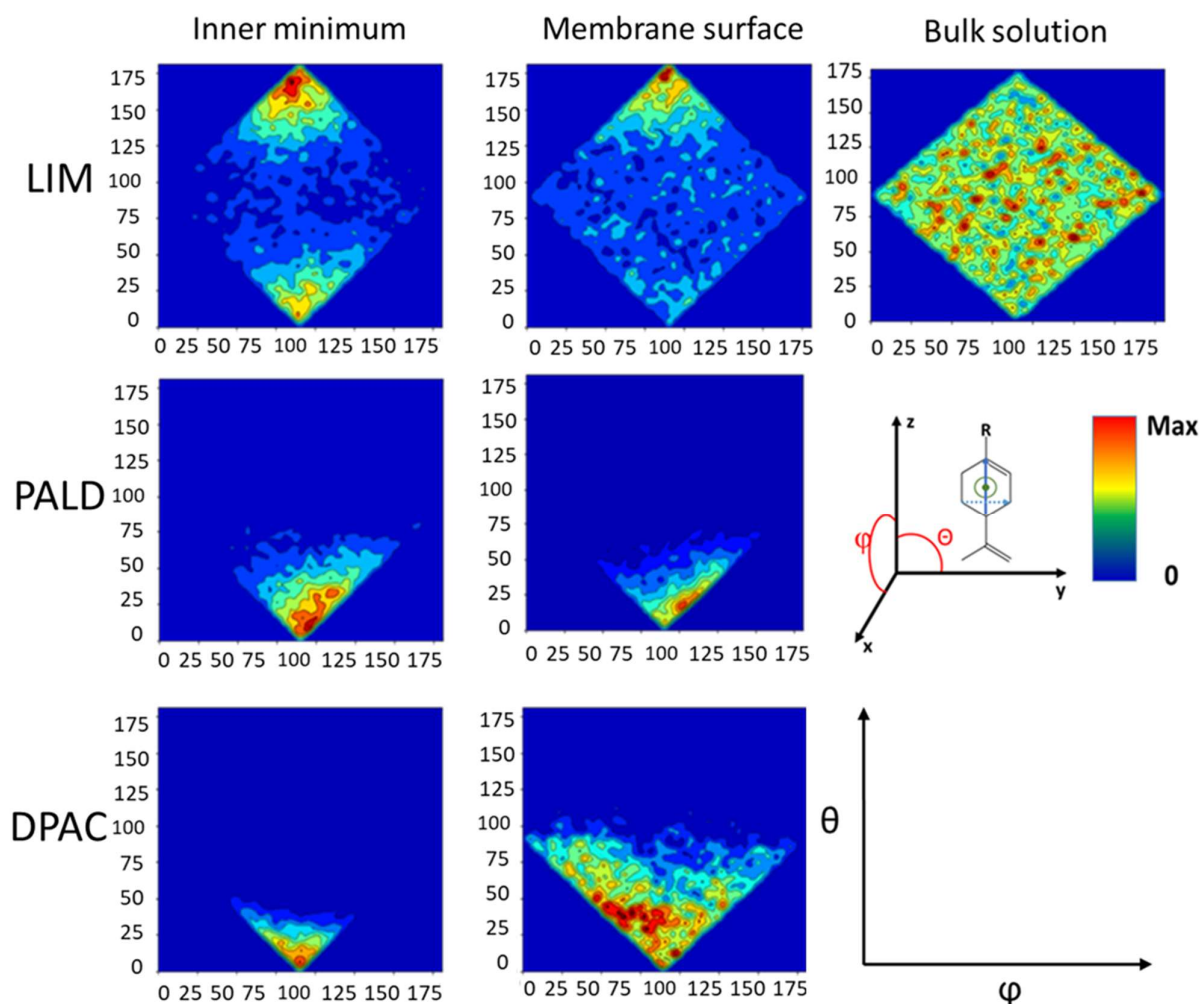


Figure 24. Orientation angle heat maps of the terpene molecules in US-REST3 simulations. Orientation angle maps are 2D probability plots of the θ and ϕ angles that define the orientation of the terpene molecules with respect to the vector normal to the membrane surface. Angle θ is reported on the y-axis and angle ϕ on the x-axis, as represented in the bottom right panel. These angle maps were used to assess the sampling of orientations visited by the terpene molecule in representative windows at three different COM distances to the bilayer. Orientation angle maps are shown for the inner minimum (left), the membrane surface (middle) and the bulk solution (right). The COM distance to the membrane surface was assumed to be around 1.8 nm for all three terpene molecules. In contrast, the COM distance to the inner minimum is different for each terpene (0.6 nm for LIM, 0.8 nm for PALD and 1.4 nm for DPAC). In the bulk solution, all three terpene molecules exhibited similar orientation angle maps (Figure S1 in the SI), so only the angle map for LIM at a distance of 3.6 nm is shown. The maximum probabilities are of the same order of magnitude but are slightly different for each heat map.

3.3.4 Mechanism of binding of terpenes to phospholipid membranes

The PMF and orientation angle maps obtained from US-REST3 simulations help to understand the mechanism of binding to the membrane of the three terpene molecules. We now consider only the US-REST3 simulations for subsequent analysis because of the better agreement with experiment of the predicted ΔG_b .

For LIM, the orientation angle maps for windows near the membrane surface reveal that LIM prefers to penetrate the membrane with its methyl group pointing toward the center of the membrane ($\theta = 160^\circ$) isopropenyl group interacting with water. Once LIM is located entirely below the membrane surface (at a COM-distance < 0.8 nm), the molecule no longer shows a preference for the methyl group to point toward the center of the membrane. Instead, the molecule can also exhibit an inverted orientation such that the methyl group can now point toward the bulk solution ($\theta = 20^\circ$). This reflects that once the isopropenyl group can no longer interact with interfacial water molecules and choline groups, both orientations become equally probable. Once LIM reaches the center of the membrane, it has complete rotational freedom, and the orientation angle map reverts to resembling the one in bulk solution (Figures S1b and S1c in the SI). The penetration of LIM into the membrane does not appear to drag any water molecules with it. This is reflected in the measured terpene-water minimum distance, which steadily increases from 0.22 ± 0.02 nm in the bulk solution to 0.88 ± 0.2 nm in the center of the membrane (at a COM distance of 0) (**Figure 25**). This mechanism is consistent with the strongly non-polar (hydrophobic) character of this molecule.

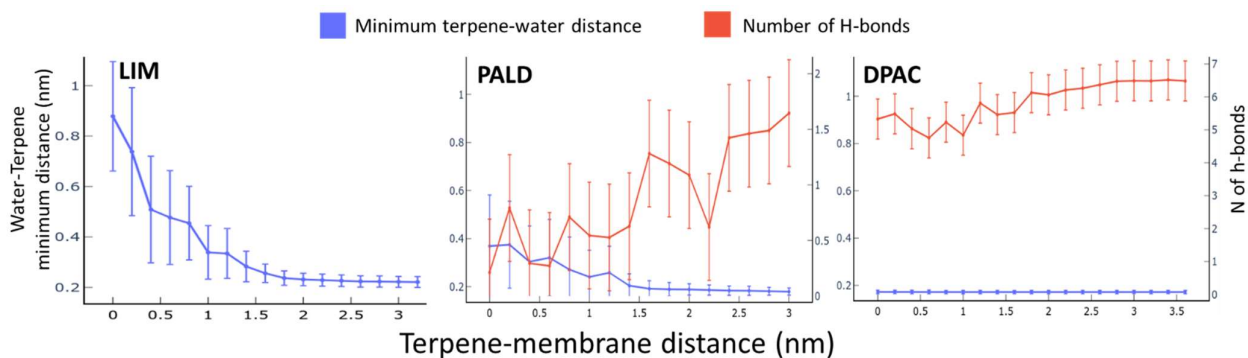


Figure 25. Changes in the minimum distance between the terpene molecules and water (blue) and number (N) of hydrogen bonds (red) between terpene and water as a function of terpene-membrane COM distance in the US-REST3 simulations for each terpene. The vertical bars represent standard deviations. The hbonds were defined using `g_hbond` using a cutoff radius of 0.35 nm and a cutoff angle of 30 degrees.

For PALD, analysis of the orientation angle maps for windows near the membrane surface reveals that this molecule prefers to penetrate the membrane with the isopropenyl group pointing toward the center of the membrane and the polar aldehyde group oriented toward the aqueous environment. Due to the hydrogen bonding ability of its polar group, PALD is expected to have a stronger interaction with water than LIM. The polar aldehyde group enables water molecules to be dragged along with the PALD molecule when it enters the membrane. This is reflected in the terpene-water minimum distance inside the membrane (**Figure 25**), which steadily increases from 0.18 ± 0.014 nm in the bulk solution (at COM distances > 2.8 nm) to 0.37 ± 0.2 nm in the center of the membrane (at a COM distance of 0). The latter value is significantly lower than for LIM, whose minimum terpene-water distance was computed to be 0.88 nm at the same COM distance to the membrane. The hydrogen bond analysis reported in **Figure 25** for PALD shows that in solution, PALD forms an average of 1.71 ± 0.64 H-bonds, whereas it forms only 0.4 ± 0.6 H-bonds in the membrane core, substantially reducing but not losing all interactions with water. This suggests that PALD has a soft solvation shell, meaning that the aldehyde group interacts with water as best as possible as the molecule penetrates the membrane, partially dragging water molecules with it. Combined with the information provided by the orientation angle maps, this suggests that as PALD loses interactions with water upon insertion into the membrane, it becomes orientationally restrained. In this preferred orientation, the polar aldehyde group is directed away from the center of the membrane and is thus more exposed to water. By contrast, once PALD reaches the central region of the membrane (at a COM-distance < 0.2 nm), it can freely change orientation but remaining aligned along the normal of the membrane (Figures S1d and S1e in the SI). This is facilitated by the local symmetry in the water density of the bilayer.

DPAC enters the membrane in the same way as PALD, with its isopropenyl group oriented toward the center of the membrane. Due to the charged nature of DPAC, its solvation shell is expected to be substantially stronger than that of PALD. This is confirmed by the terpene-water minimum distances reported in **Figure 25**, which show that DPAC exhibits no significant changes as it penetrates the bilayer, with an average distance of 0.17 ± 0.007 nm even at the center of the bilayer. The hydrogen bond analysis reported in **Figure 25** shows that the carbonyl group of DPAC forms 6.5 ± 0.93 H-bonds in the bulk solution, whereas this figure reduces to 5.3 ± 1.0 upon insertion into the center of the membrane. This suggests that DPAC has a tight solvation shell that is not lost when it buries itself into the membrane, dragging it into the hydrophobic core of the bilayer. This is the likely reason for the large free energy barrier to the permeation of DPAC. It is interesting to note that when the COM-distance reaches values < 0.6 nm, the charged carboxylate group of DPAC can find and interact with water molecules that have permeated from the other side of the membrane, allowing it to invert its orientation to interact with them (Figures S1f and S1g in the SI).

3.3.5 Assessing the efficiency of US-REST3 simulations

As noted in the methods section, a larger number of replicas than actually needed was used for LIM and PALD to keep the simulation setup consistent for all of the three terpenes. Nevertheless, simulations with LIM and PALD only required four replica per windows to obtain an exchange rate >20%. To ensure that using a different number of replicas did not affect the final configurational ensemble obtained for the ground replica, a simulation for LIM was run using $\lambda = 0.2, 0.4, 0.6$ and 1.0 for 80 ns per replica. These simulations were started from the same configurations as used in the simulations with eleven replicas.

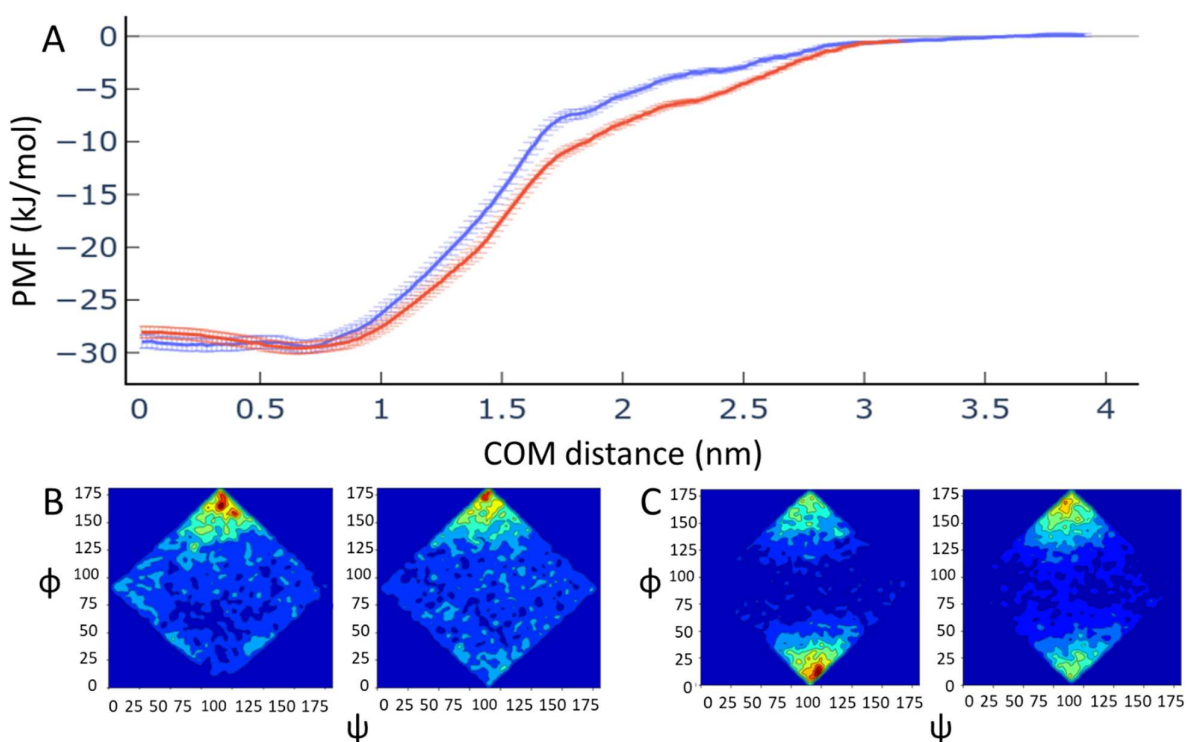


Figure 26. US-REST3 simulation of LIM with four and eleven replicas. A) Comparison of the predicted PMF using four (blue) and eleven replicas (red), showing convergence to the same free energy of binding. B) Angle maps for the window at a COM distance of 1.8 nm (surface) for the simulations with four (left) and eleven replicas (right). C) Angle maps for the window at a COM distance of 0.6 nm (internal minimum) for the simulations with four (left) and eleven replicas (right).

Comparison of the PMFs predicted with the simulations using four and eleven replicas (**Figure 26A**) shows that the shape of the PMF is very similar and that the predicted values of the ΔG_b are statistically indistinguishable between the simulation using four replicas (-29.4 ± 0.6 kJ/mol) the one using eleven replicas (-29.7 ± 0.6 kJ/mol). Comparison of the orientation angle maps for the surface region (**Figure**

26B) reveals almost no differences between the simulations using four and 11 replicas, with the preferred orientation of LIM being one with the methyl group pointing toward the membrane. In the case of the membrane-internal minimum, there is only a small difference in the relative intensity of the preferred orientation peaks and the two orientation angle distributions. Nonetheless, the overall shape of the bimodal distribution of orientations is maintained, and the two orientations of the LIM aligned with the normal of the membrane are both visited in the simulation using four replicas. This suggests that a simulation with a smaller number of replicas converges to similar results, reducing computation time. In the next test case of the US-REST3 method, different scaling schemes were indeed used for replicas at different COM distances from the center of the membrane.

3.4 Summary and conclusions

We have validated the efficacy of the use of selective scaling in REST3 in the context of small molecule-membrane interactions (SMMIs). REST3 was developed to overcome long-lived interactions that hamper sampling for free energy calculations of SMMIs. This was achieved by scaling down van der Waals and electrostatic interactions between a small molecule and its surroundings (i.e. solvent and lipid membrane). The effectiveness of REST3 was investigated in terms of its ability to: i) accurately predict the binding affinity (ΔG_b) of three terpene molecules to a DMPC bilayer.

In the case of the terpenes, US-REST3 was observed to allow the system to rapidly lose the memory of its initial configuration and to rapidly converge the shape of the PMF, resulting in accurate predictions of ΔG_b .

In spite of the increased computational cost of running US-REST3, the replica-exchange scheme could be optimized to balance fast convergence of the PMF with efficient replica-exchange rates for a given small molecule at particular COM distances. This was achieved, for example, by reducing the number of replicas to almost a third (as discussed in section 4.1.5) without significantly affecting the final PMF in the case of LIM. With these replica schemes, it was thus possible to gain confidence in the estimates from US-REST3 with a significantly reduced computational effort.

These findings demonstrate that REST3 is a viable method to study SMMIs when the Hamiltonian of the system is well understood. The enhancement of sampling achieved by REST3 can substantially improve predictions made using conventional US simulations of typical time scales. REST3 is thus an effective method for the study of SMMIs in explicit solvation conditions.

3.5 Supporting Information

1. Implementation of REST3 in GROMACS 4.6.7

Several GROMACS modules were modified, including non-bonded routines specific to SSE 4.2 and AVX 256 architectures, user definitions and distributed communication. Routines specific to replica exchange were modified to include specification of scaling coefficients related to pair-wise interactions, energy rules that may lead to a successful replica exchange, and modules that update the calculation of scaled electrostatics and van der Waals interactions. The current implementation only supports the reaction field method for non-bonded interactions. The input requires the user to specify the lower triangular matrix of possible interaction λ for both electrostatic and van der Waals energies. These matrices are broadcast to all nodes participating in the simulation. Below is a sample specification of related user input:

```
; Compound #mols
Protein_X      1
POPC           128
SOL            9955
NA             28
CL             32

[ scale_vdw ]
; Specify this section after 'molecules' section`
1.0
0.8 1.0
1.0 1.0 0.2
1.0 1.0 1.0 0.0
1.0 1.0 1.0 1.0 1.0

[ scale_q ]
; Lower triangular matrix includes the diagonal
1.0
0.8 0.2
0.2 0.0 1.0
1.0 1.0 1.0 1.0
1.0 1.0 1.0 1.0 1.0
```

GROMACS involves a master-slave communication protocol, wherein decisions related to exchange criterion and broadcasting replica exchange coordinates rest with the master. Before the exchange, GROMACS follows an odd-even check between replicas and designates all possible replica pairs that fall in the criterion for exchange. Owing to costs involved in the gathering of the energy of each replica at the end of each time step, the routine only computes and exchanges replicas at every exchange

attempt (Nreplex). Routines that are involved in the calculation of non-bonded interactions update both the scaled forces as well as the energies. With respect to the calculation of non-bonded electrostatics, due to possible complications that could arise due to charge imbalance, only energies and forces were modified with scaled electrostatic interactions. A full version of the modified GROMACS source is available at <https://github.com/curtinic/gromacs>

2. Total simulation time for terpenes

Table S1. Conventional US simulations

| COM dist (nm) | LIM (ps) | PALD (ps) | DPAC (ps) |
|---------------|----------|-----------|-----------|
| 0.0 | 550000 | 1100000 | 850000 |
| 0.2 | 550000 | 1100000 | 850000 |
| 0.4 | 550000 | 1100000 | 850000 |
| 0.6 | 550000 | 1100000 | 850000 |
| 0.8 | 550000 | 1100000 | 850000 |
| 1.0 | 550000 | 1100000 | 850000 |
| 1.2 | 550000 | 1100000 | 850000 |
| 1.4 | 550000 | 1100000 | 850000 |
| 1.6 | 550000 | 1100000 | 850000 |
| 1.8 | 550000 | 1100000 | 850000 |
| 2.0 | 550000 | 1100000 | 850000 |
| 2.2 | 550000 | 1100000 | 850000 |
| 2.4 | 550000 | 1100000 | 850000 |
| 2.6 | 550000 | 1100000 | 850000 |
| 2.8 | 550000 | 1100000 | 850000 |
| 3.0+ | 550000 | 1100000 | 850000 |

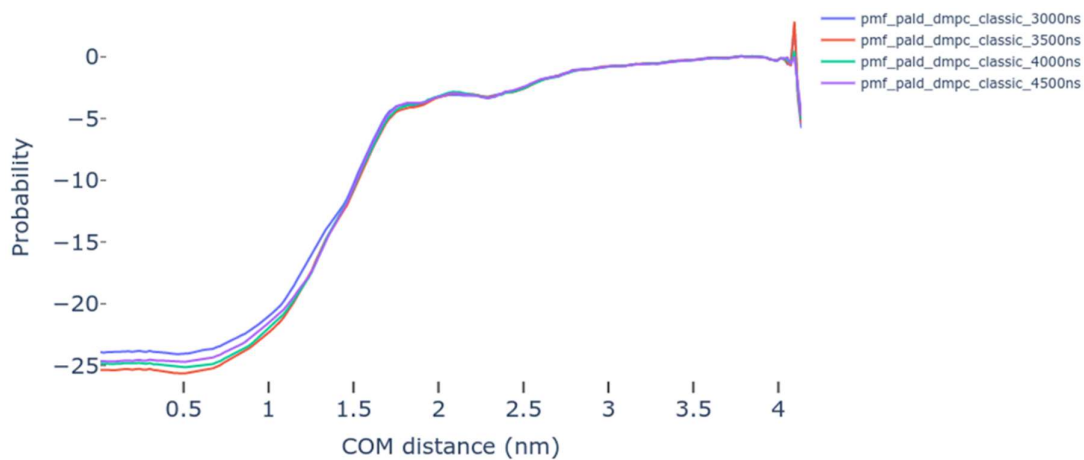
Table S2. US-REST3 simulations

| COMdist (nm) | LIM (ps) | PALD (ps) | DPAC (ps) |
|--------------|----------|-----------|-----------|
| 0.0 | 120000 | 80000 | 50000 |
| 0.2 | 120000 | 80000 | 50000 |
| 0.4 | 120000 | 80000 | 50000 |
| 0.6 | 120000 | 120000 | 50000 |
| 0.8 | 120000 | 120000 | 50000 |
| 1.0 | 120000 | 120000 | 50000 |
| 1.2 | 120000 | 120000 | 50000 |
| 1.4 | 80000 | 80000 | 50000 |
| 1.6 | 80000 | 80000 | 50000 |
| 1.8 | 80000 | 120000 | 50000 |
| 2.0 | 80000 | 120000 | 50000 |
| 2.2 | 80000 | 120000 | 50000 |
| 2.4 | 80000 | 120000 | 50000 |

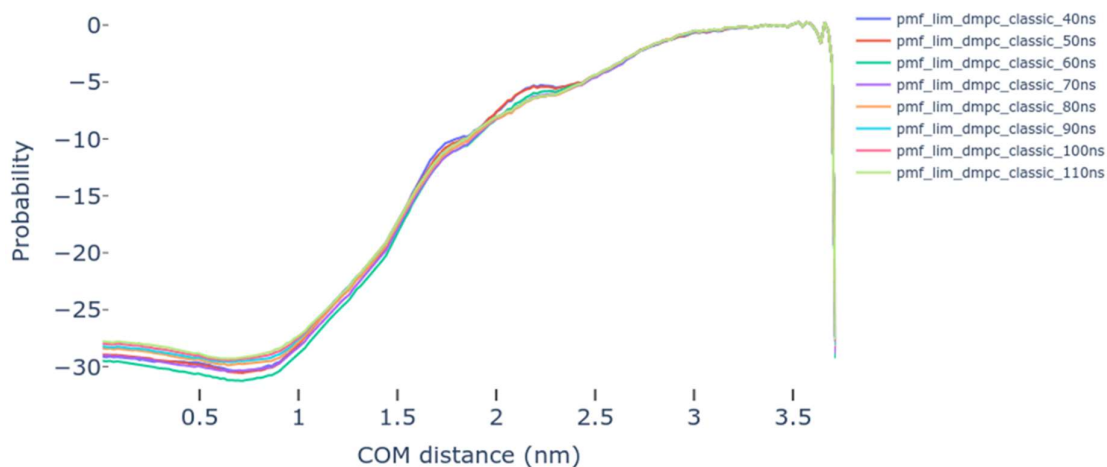
| | | | |
|------|-------|-------|-------|
| 2.6 | 80000 | 80000 | 50000 |
| 2.8 | 80000 | 80000 | 50000 |
| 3.0+ | 80000 | 40000 | 50000 |

Figure S1: Convergence of the PMFs for each terpene.

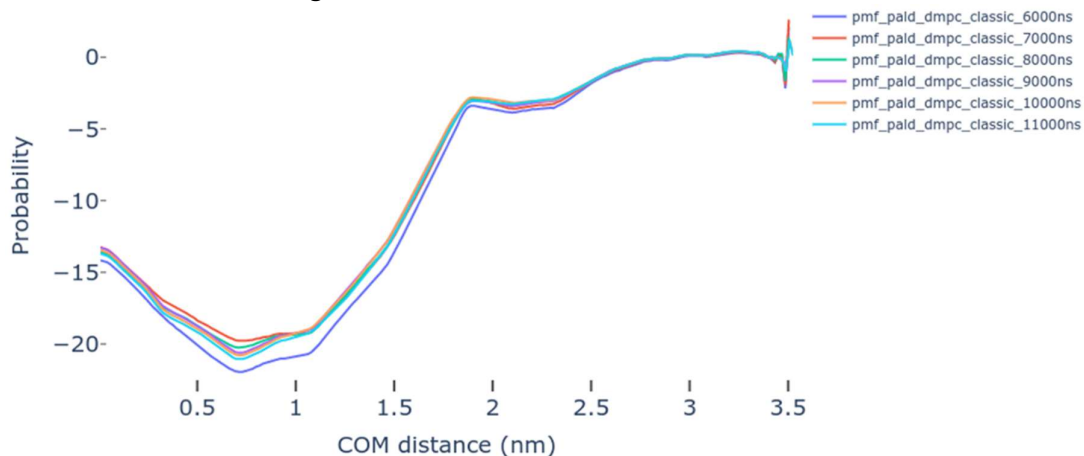
PMF vs Time for LIM using US



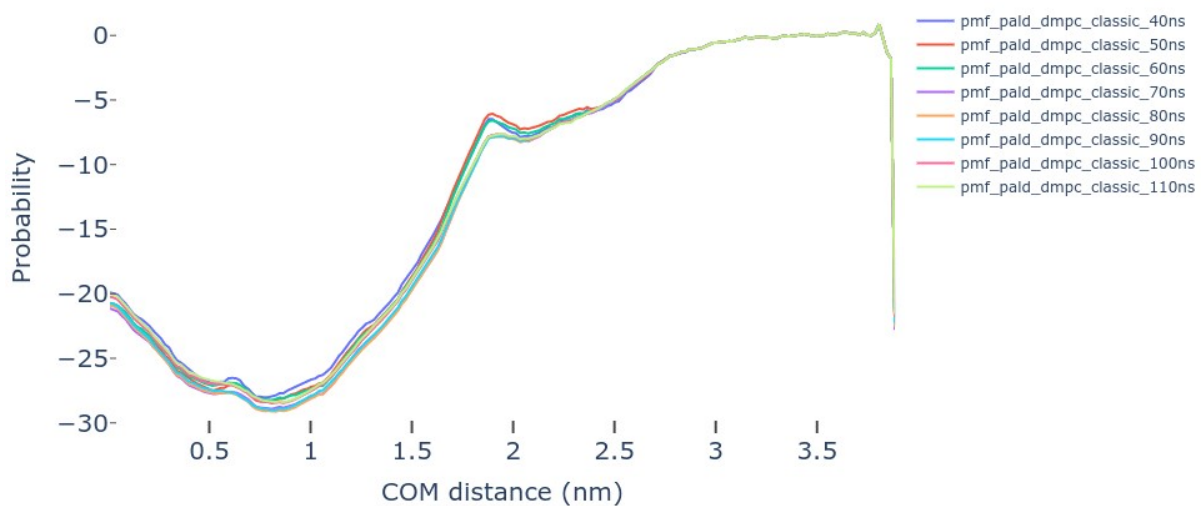
PMF vs Time for LIM using US-REST3



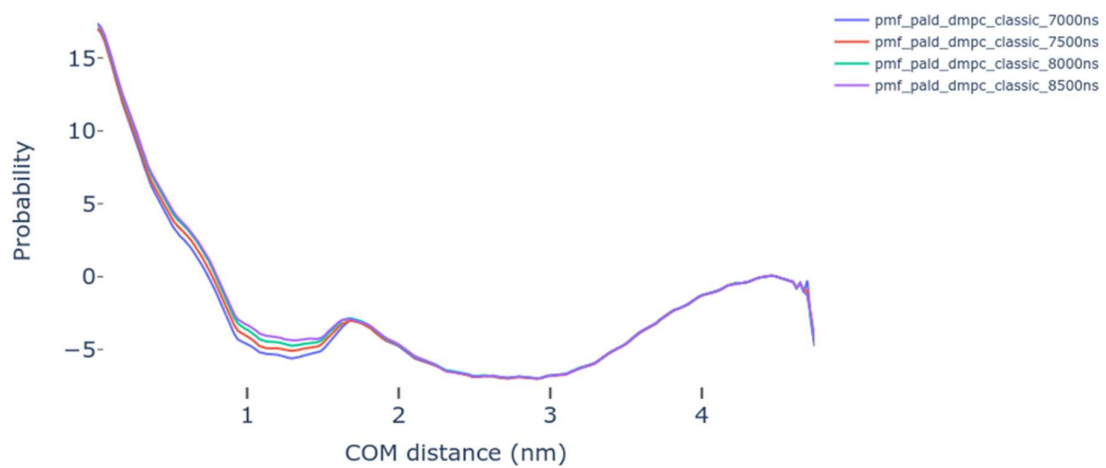
PMF vs Time for PALD using US



PMF vs Time for PALD using US-REST3



PMF vs Time for DPAC using US



PMF vs Time for DPAC using US-REST3

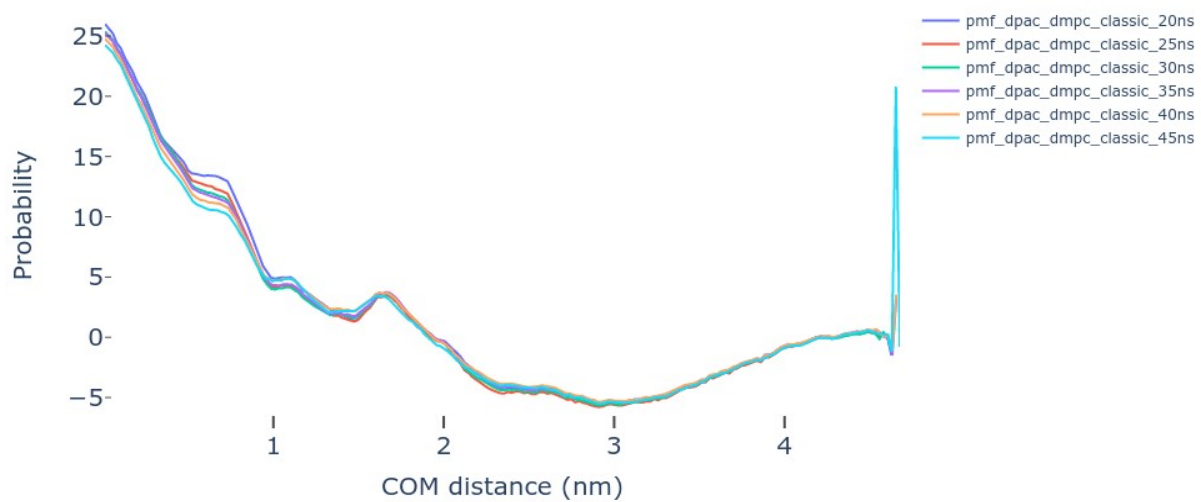
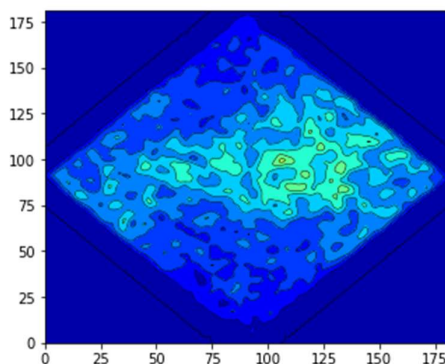
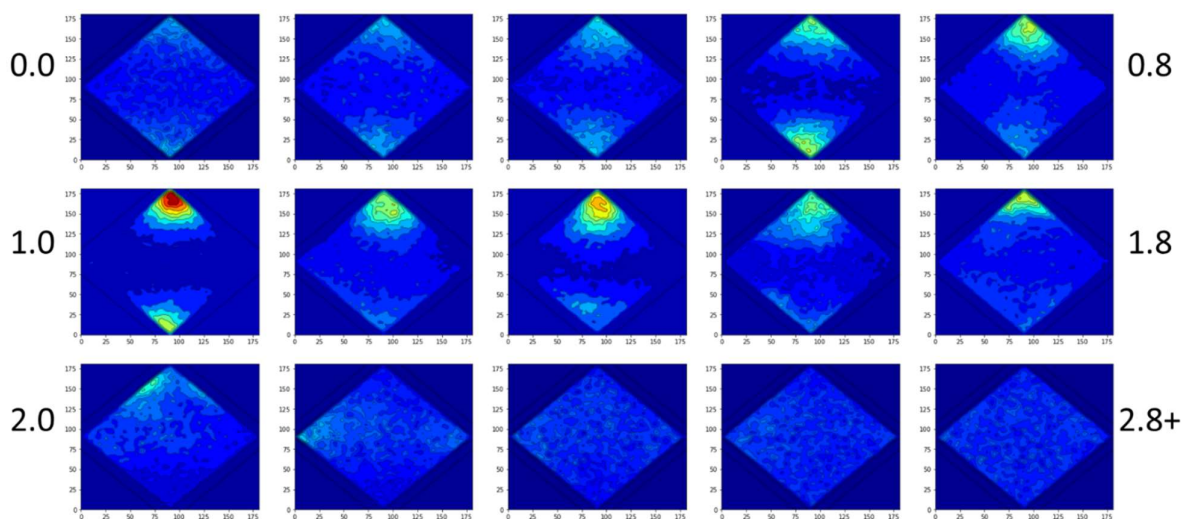


Figure S2: Angle maps for all the simulations of terpenes

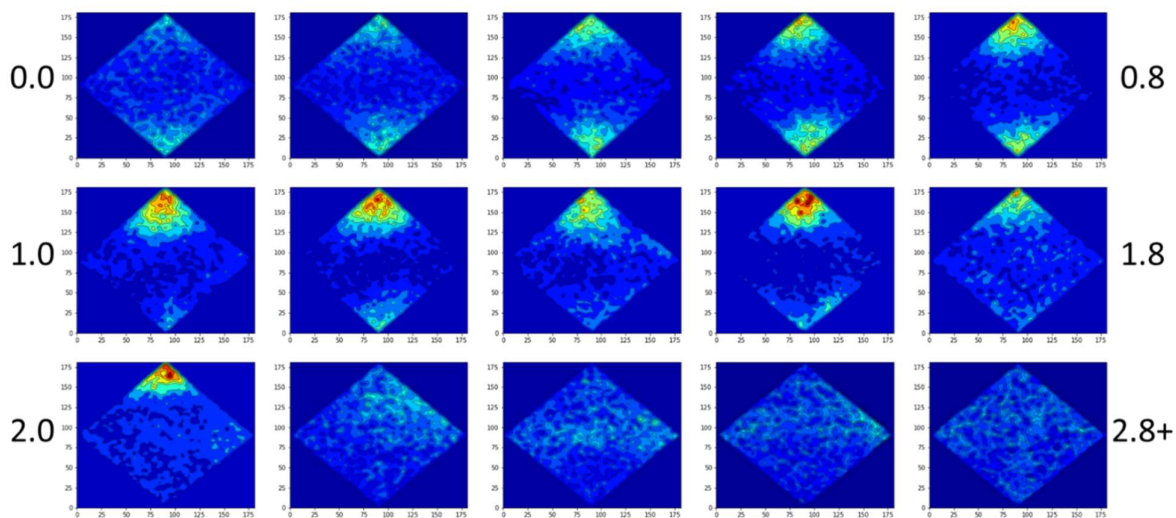
a. LIM US, window 1.8 nm, angle map for the first 160 ns.



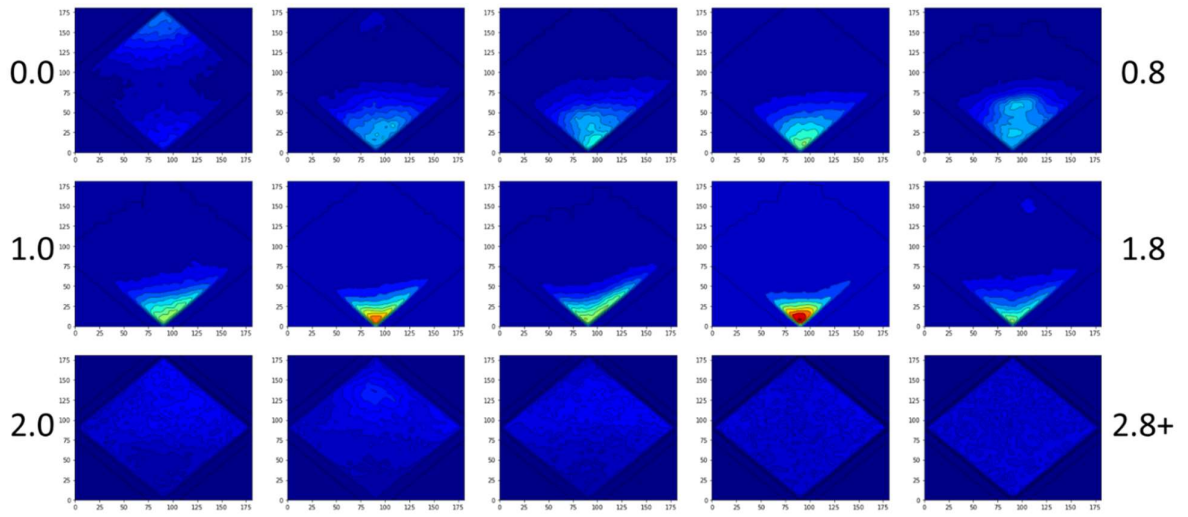
b. LIM US



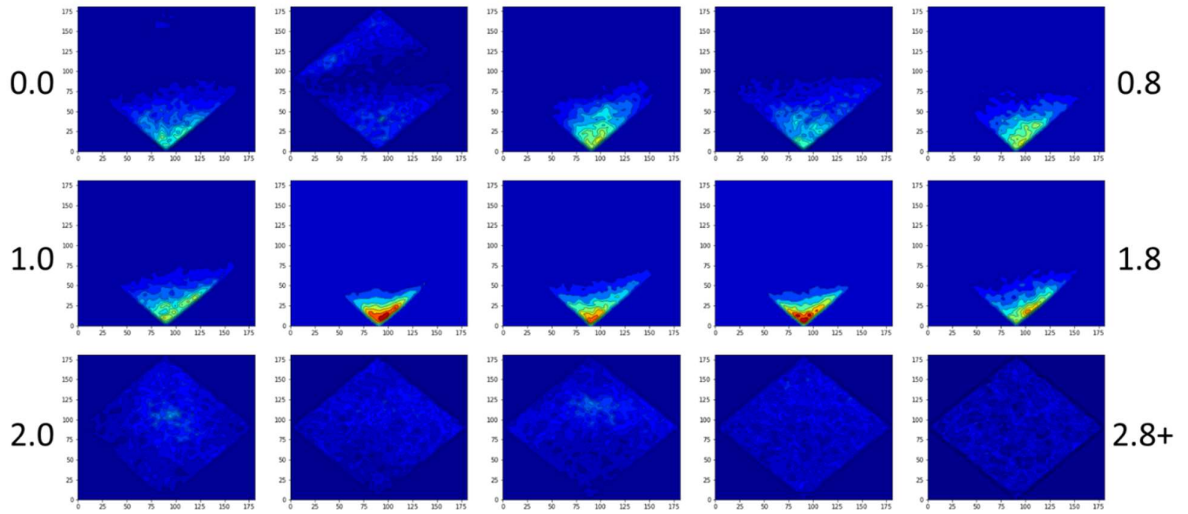
c. LIM US-REST3



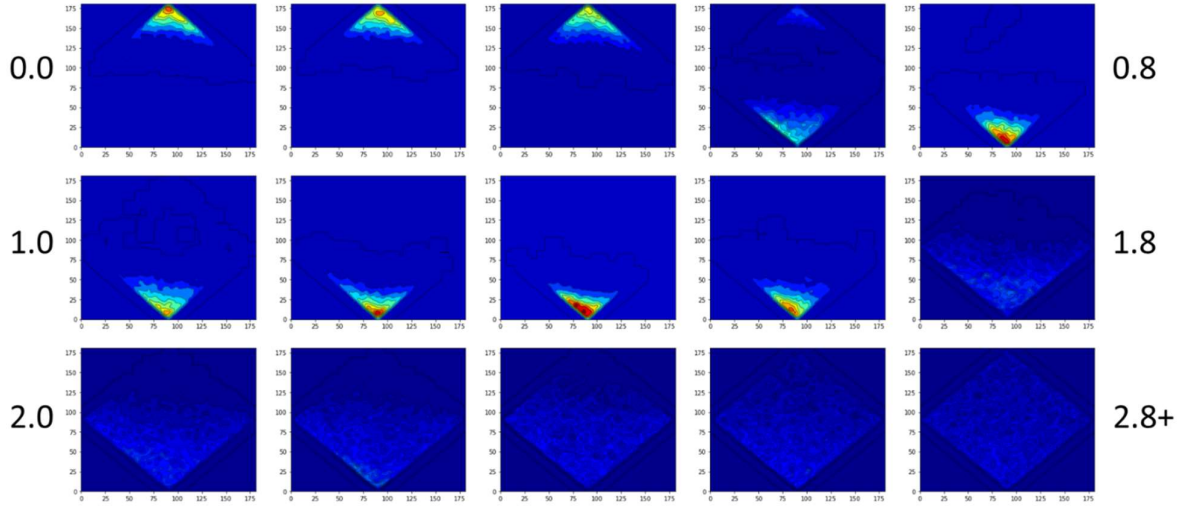
d. PALD US



e. PALD US-REST3



f. DPAC US



g. DPAC US-REST3

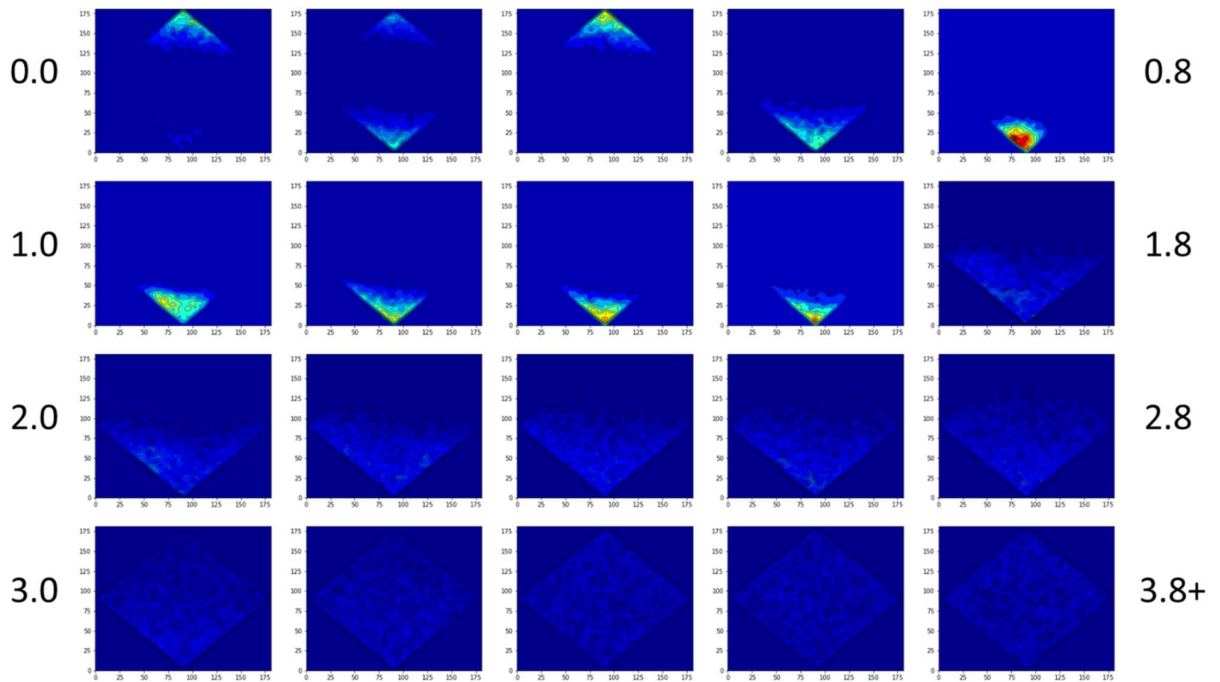
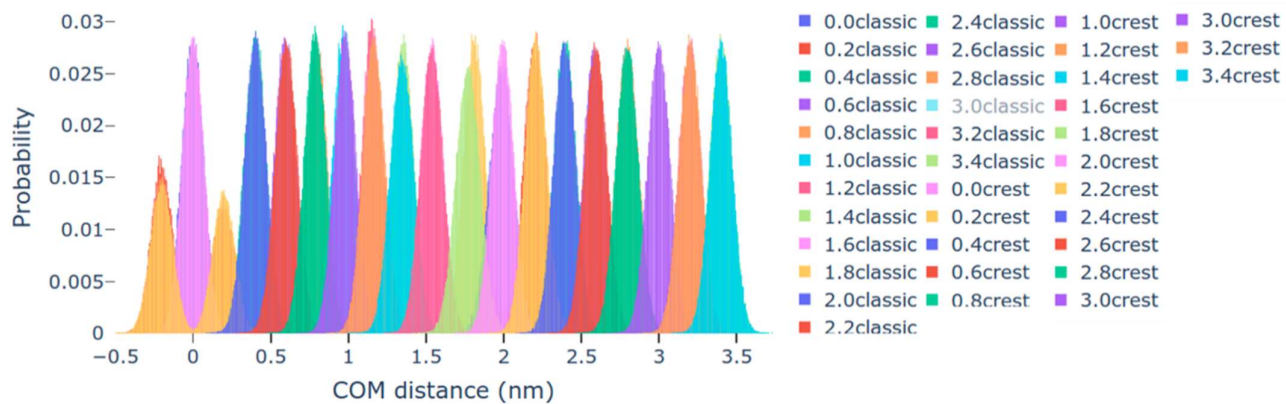


Figure S3: Histograms of COM distance for every window for LIM



4 Chapter 4 – Application of US-REST3 to the Protx-1 – POPC membrane system

4.1 Introduction

After validating the newly developed US-REST3 methodology with small molecule – membrane interactions (Chapter 3), the subsequent aim was to test the method on larger and more complex systems. The objective is to enhance umbrella sampling (US) simulations with improved orientational sampling such as is described in Chapter 2, Section 2.1, by applying the REST3 method to each one of the umbrella windows. As described in Chapter 2, US simulations of the interaction of Protx-1 with a POPC bilayer exhibited potentially inadequate sampling of orientational configurations of the peptide with respect to the membrane surface. While the peptide shows free rotation in solution, upon binding to the membrane surface the peptide tends to remain “locked” in a specific orientation and reaches a deep energy minimum from which it does not escape for hundreds of nanoseconds. This lack of orientational sampling impedes obtaining an accurate PMF, which is necessary for estimating the membrane binding affinity of the peptide.

The work described in this chapter aimed to determine the enhancement of orientational sampling of the peptide on a membrane surface without the need for a collective variable. As discussed in Chapter 3, Section 2, REST3 increases orientational sampling by scaling down peptide-membrane interactions whilst leaving the membrane-membrane, peptide-membrane, membrane-water and water-water parts of the Hamiltonian unchanged. This alteration of specific interactions retains the properties of the membrane even in highly tempered replicas, thus overcoming the problem encountered in other replica exchange methods like parallel tempering. To apply REST3 to a peptide-membrane system, the same approach as described in Chapter 3 was used. The water-peptide and membrane-peptide interactions were scaled down in tempered replicas to decrease the energy barriers between different orientational configurations, thus enhancing the exploration of orientational states in the ground replica. It was observed that the scaling used in REST3 causes the peptide to fall into highly self-interacting conformations, which are then difficult for the system to escape from. These self-interacting conformations lead to considerable conformational distortion in the peptide and reduced interactions with the surrounding water and the membrane. As a result, the peptide was observed to experience an unfavourable energy landscape in the proximity of the membrane-water surface. The result is a PMF that does not resemble the expected profile for the binding of a medium or strong binder to a lipid membrane. As outlined in this chapter, the same issues were encountered in simulations of ProTx1 binding to a POPC membrane.

4.2 Methods

4.2.1 REST3 enhanced umbrella sampling simulation of Protx-1

The setup for the simulation of Protx-1 is similar to that described in Chapter 2, Section 2.2. Specifically, US simulations with 20 equally-spaced windows from $dz = 2.0$ nm to 4.0 nm were used, where dz is the distance between the centre of mass (COM) of the peptide and the COM of the lipid membrane. In each window a harmonic potential with a force constant k of 500 kJ/mol was used. The difference with respect to the US simulations described in Chapter 2 is that the sampling in each window in US-REST3 is further enhanced with the REST3 method described in Chapter 3, Section 3.1.2. In each window 11 replicas with different scaling factors ranging from $\lambda = 0.2$ to $\lambda = 1.0$ were used (Figure 27).

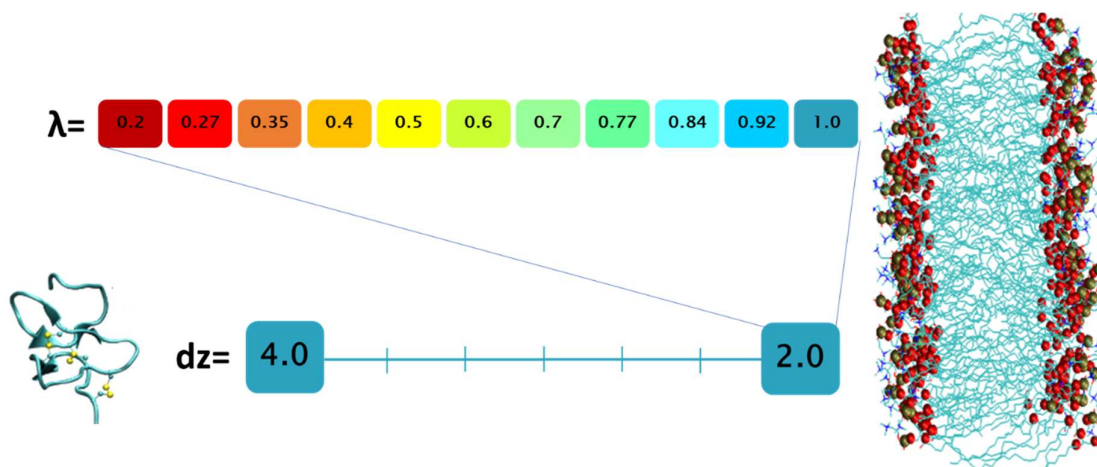


Figure 27. Diagram of the simulation setup of the REST3-enhanced umbrella sampling method. For every window at a distance dz , 11 replicas spaced from $\lambda = 0.2$ to $\lambda = 1.0$ were conducted in parallel and allowing exchange of configurations every 50 steps.

To generate the simulation system for each replica, the last frame of the original branch of the classical US simulation described in Chapter 2, Section 2.1 was used. A Nosé-Hoover thermostat with $\tau_t = 0.5$ ps was used to keep the temperature of the system at 298 K and a semi-isotropic Parrinello-Rahman barostat with $\tau_p = 0.2$ ps was used to enforce a pressure of 1 bar. All other simulation parameters were the same as those reported in Chapter 2.

Concerning the REST3 part of the simulations, both the van der Waals and electrostatic interactions in the peptide-membrane, peptide-water and peptide-ions terms were equally tempered in replicas with perturbed Hamiltonians. For simplicity, when referring to peptide-water interactions, the peptide-water and peptide-ion interactions are also included unless differently stated. The replica

scaling factors chosen to interpolate between the unperturbed Hamiltonian and the Hamiltonian with $\lambda = 0.2$ are 0.2, 0.27, 0.35, 0.4, 0.5, 0.6, 0.7, 0.77, 0.84, 0.92, 1.0. Testing in previous studies¹⁶⁸ showed that this scaling scheme yielded a consistent exchange probability between 13 and 25% across all the replicas and windows. All simulations were run for 40 ns per replica, giving a total simulation time of 440 ns per window and a total of 4.84 μ s for the entire simulation. At this point analysis of the simulations indicated the problems that will be illustrated and the simulations were stopped.

All analyses were performed on the ground replica ($\lambda = 1.0$), where the Hamiltonian was left untempered.

4.2.2 Potential of mean force calculations

The potential of mean force (PMF) along the dz reaction coordinate was reconstructed using the WHAM methods^{3,144} as implemented in the GROMACS `g_wham` module³. Bayesian bootstrapping for error calculation was performed with `n_bootstraps = 200`. All the PMF calculations were performed on the ground replica using the last 20 ns of the simulation in each window.

4.2.3 Radius of gyration

The radius of gyration of the peptide was obtained with the `g_gyrate` module in GROMACS using the C α in the protein backbone.

4.2.4 Hydrogen bond analysis

The hydrogen bonds (H-bonds) between the various species in the simulation were identified using the `g_hbond` module in GROMACS with a cutoff of 0.35 nm for the H-bond distance and 30° for the donor-hydrogen-acceptor angle.

4.2.5 Water layer analysis

Near the surface of the membrane the peptide can interact with three species, the membrane, the solvent and itself. An equilibrium within this three types of interaction is to be expected at physiological condition. If REST3 induces an imbalance of these three contributions, changes in the relative amount of these interactions will show in the relative amount of interactions that the protein experience. To quantify the extent of the imbalance in interactions that occurs between the peptide and the lipid membrane at given COM distance, the amount of water that accumulates between the two at a given dz was quantified.

The number of water molecules was estimated using the MDAnalysis Python library¹⁶³. First, the bottom half of the peptide relative to the membrane was selected excluding all the atoms with a z-coordinate higher than the z-coordinate of centre of geometry of the peptide. Then all the water molecules within 10 Å of this atom selection were identified. Of these water molecules, only the ones

within a rectangular region delimited by the minimum and maximum x and y coordinate of the previously selected bottom half of the peptide were retained as shown in **Figure 28**. This selection was carried out for each frame in the last 40 ns of the ground replica simulation, and then averaged over time.

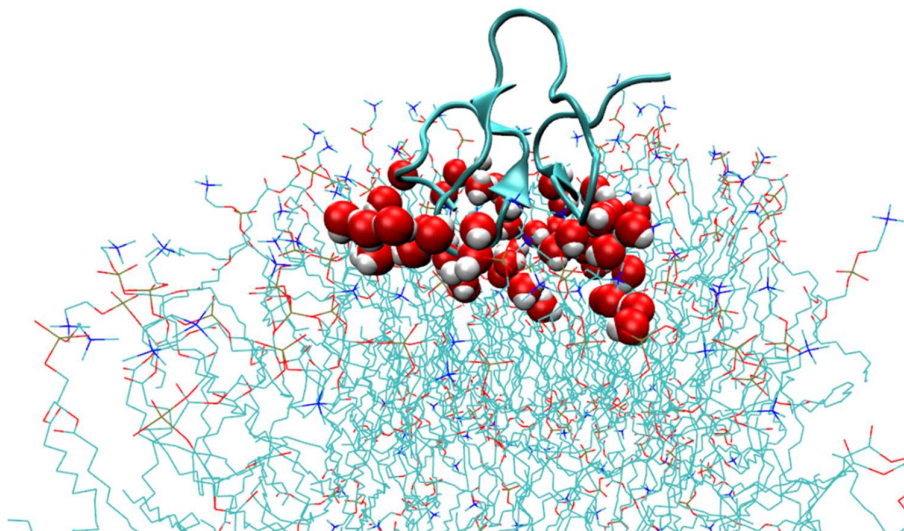


Figure 28 Illustration of the water layer selections

4.3 Results

As noted in Chapter 2, a simulation would ideally yield sufficient orientational sampling such that the various orientations of the peptide on the membrane surface are sampled and the peptide can interchange between different energy minima in the orientational space. At the same time, as stated in Chapter 1 ICK peptides are rigid, so it is expected that Protx-1 should overall retain its secondary and tertiary structure without huge variations. If these conditions are met, one would expect to compute a PMF that accurately determines the free energy of the binding of the peptide to a lipid membrane. The resulting PMF reconstructed from the US-REST3 simulations using the WHAM method is shown in Figure 29.

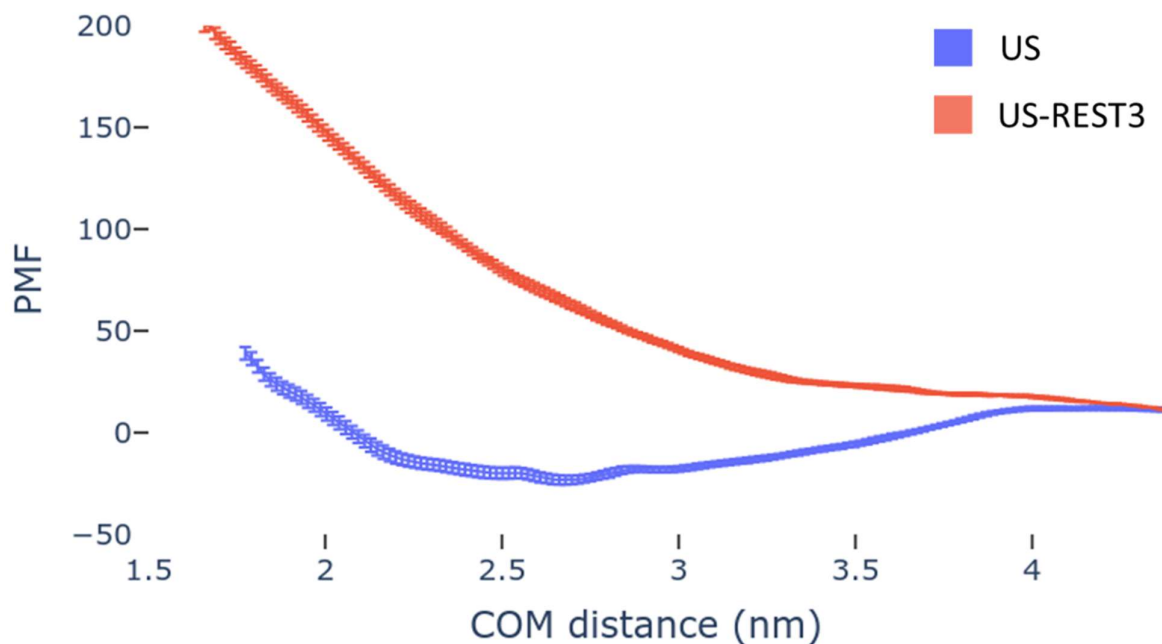


Figure 29. Potential of mean force of the binding of ProTx-1 to a POPC bilayer predicted using conventional US and US-REST3 simulations.

As it can be seen, the shape of the two PMFs is very different. The expected minimum around 2.7 nm, previously predicted with conventional US simulations, is no longer predicted by the US-REST3 simulations. Instead, the PMF monotonically increases from bulk water to the centre of the membrane, indicating that the interaction between the peptide and the membrane is predicted to be always unfavourable. Previously discussed SPR experiments have shown that ProTx1 has a binding affinity in the sub-micromolar range for POPC membranes ¹²⁸, and that the peptide likely resides at the membrane surface and does not penetrate into the membrane centre. This should be reflected in the PMF by an energy minimum at dz 2.4 to 3.0, followed by an increase in free energy for distances $dz < 2.4$. The PMF predicted with US-REST3 is thus not consistent with these experimental observations.

To assess whether distortions in the peptide structure could be the cause of this incorrect PMF, the radius of gyration of the peptide in both the US and US-REST3 simulation was compared. If the peptide undergoes significant structural changes, this would likely be reflected in the radius of gyration. The distributions of R_g values sampled in the two simulations, at three different COM-distances, are reported in Figure 30. These distances represent the protein on the membrane surface (2.6 nm), in the vicinity of the membrane (3.0 nm) and out in bulk solution (4.0 nm). Analysis at other COM distances resulted in similar behaviour.

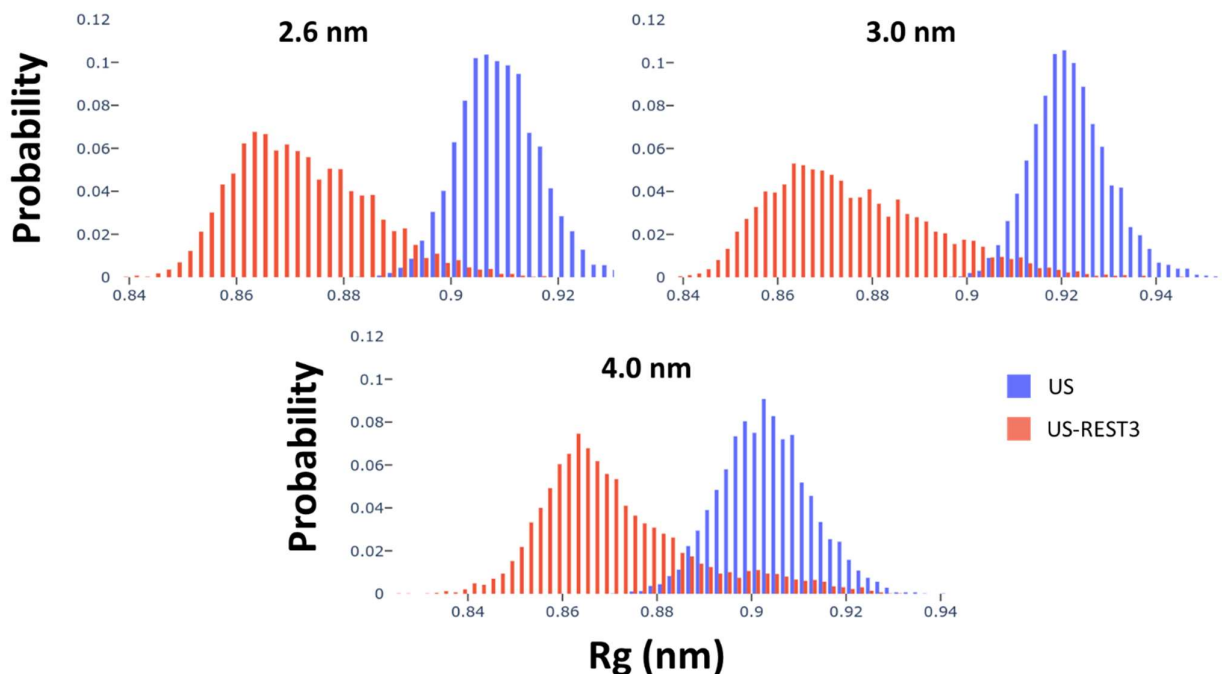


Figure 30. Radius of gyration measured at three different peptide-membrane COM distances (2.6, 3.0 and 4.0 nm) obtained from the conventional US simulation of the interaction of ProTx-1 with a POPC bilayer and the corresponding windows from the ground replica in the US-REST3 simulation.

The Rg distributions computed in the US-REST3 simulation show significantly lower values than the corresponding distributions in the US simulation. The average Rg for the US-REST3 simulations are 0.87 ± 0.012 nm at a COM distance of 2.6 nm, 0.88 ± 0.017 nm at a COM distance of 3.0 nm and 0.87 ± 0.016 nm at a COM distance of 4.0 nm. In the same windows the classical US simulation gave, respectively, average Rg values of 0.906 ± 0.007 , 0.922 ± 0.008 nm, and 0.90 ± 0.01 nm. On average, the radii of gyration are lower in all three windows in the US-REST3 simulation compared to the corresponding US simulation, indicating a consistent compaction of the peptide structure.

To investigate the reason for this structural change, the average number of H-bonds that the peptide formed with itself was computed. The number of these intra-molecular H-bonds gives an indication of the extent of the self-interactions that the peptide experiences in a given replica with a certain interaction scaling. Since this effect is likely to arise from scaling the Hamiltonian, it is expected that the effect would be more pronounced in replicas with a high scaling factor λ . To understand the trend of intramolecular H-bonds as a function of the scaling factor λ , the distribution of the number of H-bonds of three of the windows belonging to the US-REST3 simulation with respective λ values of 0.2 (very tempered), 0.5 (middle tempered), and 1.0 (no tempering) was analysed. Figure 31 shows these

intra-peptide H-bonds for the above three classical US windows compared to the equivalent windows in the US-REST3 simulation.

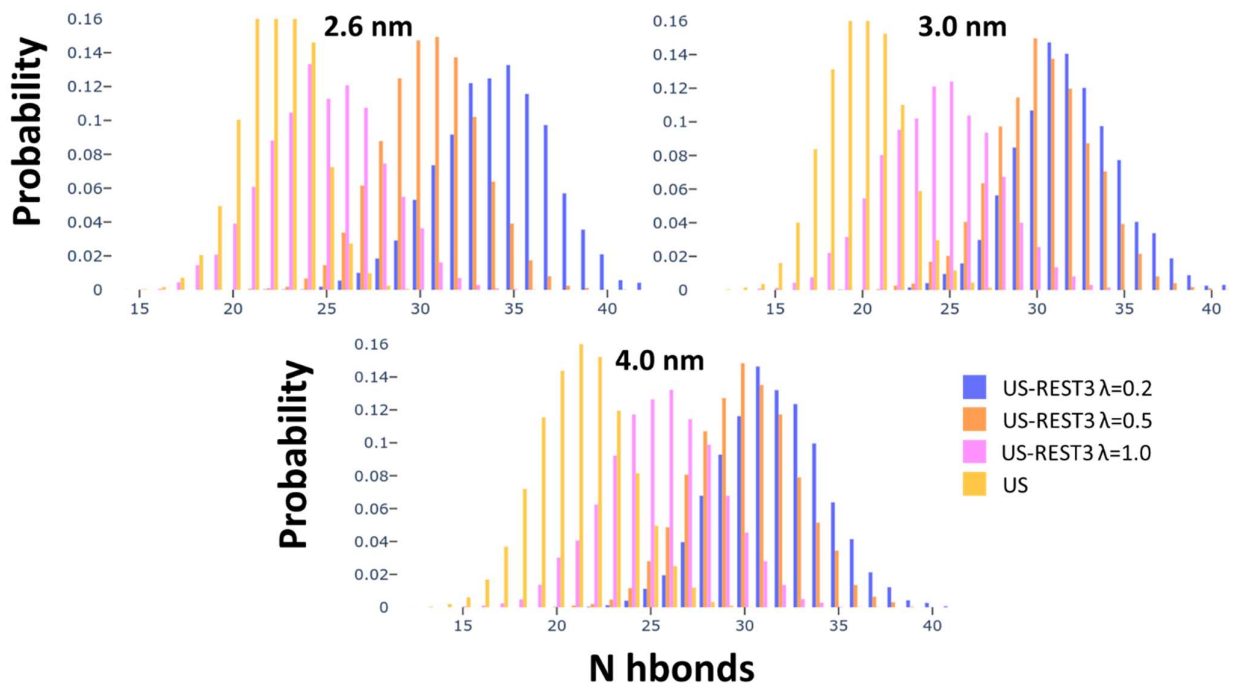


Figure 31. Number of hydrogen bonds in replicas with $\lambda = 0.2$ (most tempered), 0.5 (middle) and 1.0 (untempered) for the US-REST3 simulation as well as the conventional US simulation at three different peptide-membrane COM distances (2.6, 3.0 and 4.0 nm).

The average number of H-bonds found in the ground (untempered) replica in the US-REST3 simulation is 24.8 ± 3.1 for a COM distance of 2.6 nm, $24.3.2 \pm 3.2$ for a COM distance of 3.0 nm and 25.5 ± 3.0 for a COM distance of 4.0 nm. The average number of H-bonds for the corresponding conventional US simulations at the same COM distances are 22.2 ± 1.9 , 19.8 ± 2.2 and 21.2 ± 2.5 . It is evident that in the ground replica of the US-REST3 simulation the peptide exhibits 3 to 4 intra-molecular H-bonds more than in the respective conventional US windows. With higher tempered replicas, the difference in H-bonds is even larger, with average values reaching 34.1 ± 3.0 H-bonds in the case of the replica with $\lambda = 0.2$ at a COM distance of 2.6 nm, which is 12 H-bonds more than in the corresponding conventional US window. This increased number of intra-molecular H-bonds is probably caused by the reduced interactions that the peptide has with water and, therefore, it behaves as if it were in a quasi-vacuum. The effect of this quasi-vacuum appears to force the peptide to increase the number of self-interactions to minimise its potential energy to counteract the loss of interactions with the solvent.

The energy minima that the peptide reaches due to these increased self-interactions might be difficult to escape from, as the original distribution of H-bonds is never fully recovered, even in the untempered replica ($\lambda = 1.0$). To further reinforce this hypothesis, Figure 32A shows the number of intramolecular H-bonds computed from a conventional MD simulation of ProTx-1 in vacuum. The average number of intramolecular H-bonds of the peptide in vacuum is 39.5 ± 2.1 , which corresponds to only 5 H-bonds more than the value obtained for replica $\lambda = 0.2$ at a peptide-membrane COM distance of 2.6 nm in the US-REST3 simulation. This confirms that the scaling down of the interactions of the peptide with its surrounding environment has an effect that is similar to the peptide being in a vacuum. This would not necessarily be a problem if the structural deformation that the peptide undergoes in these regimes were to be rapidly eliminated in the ground replica. To assess the extent of the persistence of these interactions, the final configuration of windows at a COM distance of 2.6 nm in the US-REST3 simulation was taken and simulated using conventional US for 100 ns. Figure 32B shows a plot of the RMSD of the peptide vs time for this simulation. Two RMSDs are reported, the first one computed with respect to the starting structure of this conventional US extension (in blue). The second is the RMSD computed with respect to a structure taken from the window centred at 2.6 nm of the conventional US simulation.

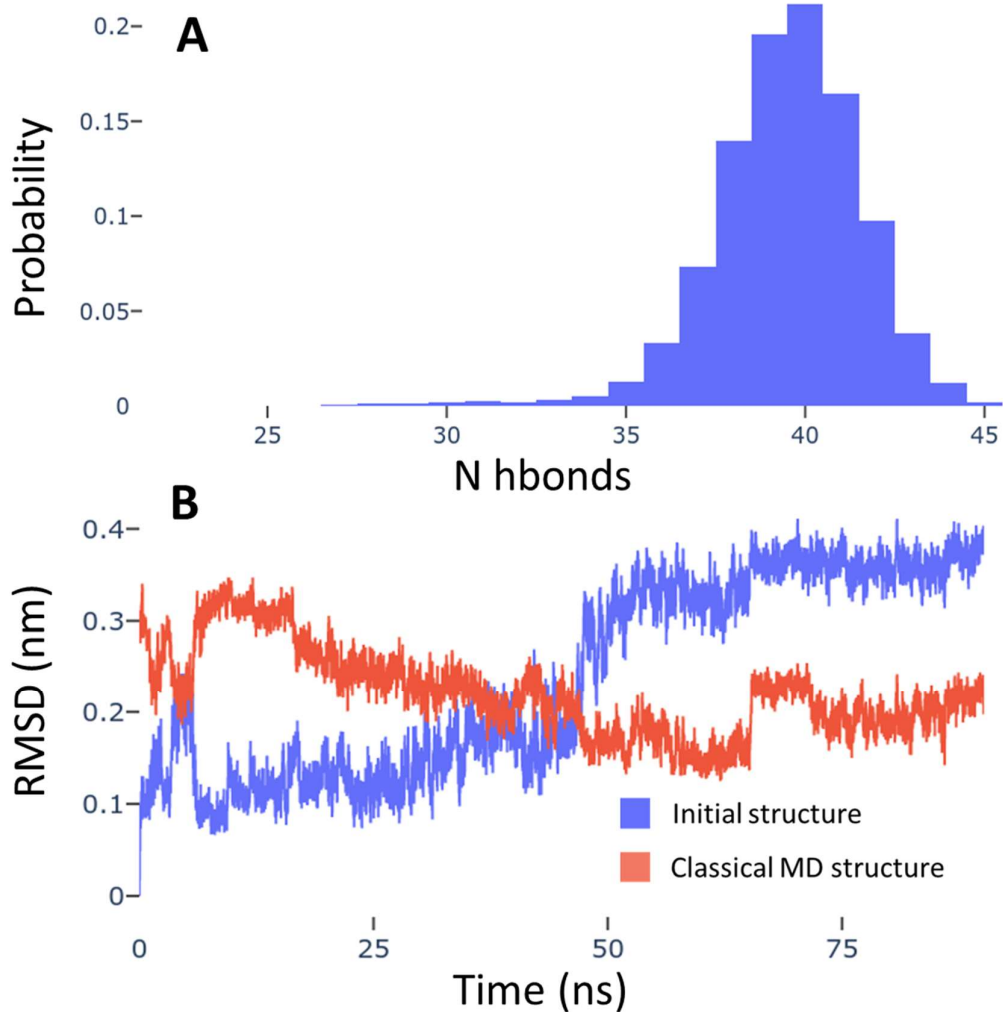


Figure 32. A: Number of intramolecular hydrogen bonds in ProTx-1 in a vacuum simulation conducted over 40 ns. B: Plot of RMSD vs time obtained from the last structure at a peptide-membrane COM distance of 2.6 nm in the US-REST3 simulation and extended using classical MD. The RMSD of ProTx-1 with respect to its initial structure in the extended simulation is shown in blue, and the RMSD with respect to a structure obtained in the conventional US is shown in red, (original branch described in Chapter 2, section 2.3).

The RMSD of the peptide computed with respect to the initial structure fluctuates around a value of 0.1 nm for the first 45 ns before increasing to a stable state with a value of 0.34 nm. On the other hand, the RMSD of the peptide computed with respect to a structure taken from the conventional US simulation has an initial value of 0.3 nm and settles to values around 0.2 nm after 50 ns. These plots reveal that the deformed structure observed in the US-REST simulation slowly returns to a structure

similar to the one observed in the conventional US simulation over the course of 50 ns. This is much longer than any reasonable replica exchange attempt frequency that would allow efficient sampling. As the relaxation of the structure of the peptide cannot be completed rapidly enough, the direct result of this effect is the effective compaction of the peptide previously discussed. The effects of the increase in intramolecular H-bonds and the compaction of the peptide also affect its ability to interact with the membrane.

The increased self interaction of the protein has also found to be correlated with an accumulation of water molecules between the peptide and the membrane. Taking the data from a conventional US with the data from a US-REST3 simulation on a per window base, the amount of water found in between the protein and the membrane can be compared. Figure 33A reports the average number of water molecules found between the peptide and the POPC membrane during the last 40 ns of the US and US-REST3 simulations. It can be seen that there is a higher number of water molecules consistently across all windows in the US-REST simulation, leading to reduced interactions between the peptide and membrane (Figure 33B). These observations agree with previous findings in the simulation of the interaction of the peptide Gomesin with a POPC membrane using the US-REST3 method ¹⁶⁸. It is evident that the tempering of the interaction of the peptide with its surrounding environment, in the same manner as was done successfully with terpenes (Chapter 3), is not applicable to larger molecules such as peptides/proteins. Systems like proteins with extensive intramolecular van der Waals and electrostatic interactions can exhibit unphysical conformations that do not represent normal solvated conformations. These abnormal conformations can persist for several tenths of nanoseconds, which would therefore require a similar amount of time between exchanges in REST3 to yield a realistic ensemble of conformations. Such long times between exchanges would make the replica exchange approach very inefficient and, therefore, not feasible for large systems.

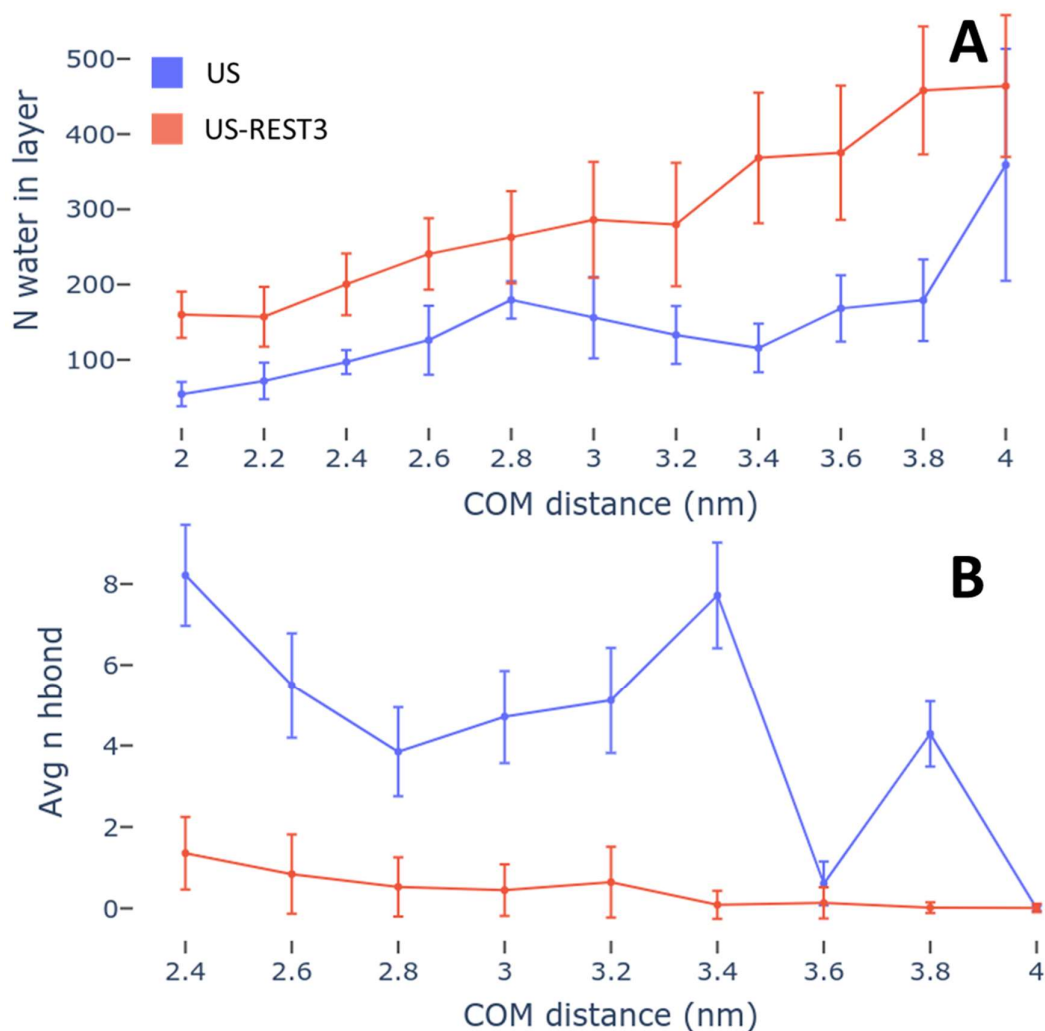


Figure 33 A: Average number of water molecules located between the peptide and the POPC membrane at different peptide-membrane COM distance in both the US and US-REST3 simulations. B: Average number of hbonds between Protx-1 and the POPC membrane vs time.

4.3.1 Other scaling regimes

Other scaling regimes were attempted previously¹⁶⁸ to try to address the problem outlined in the previous section. These simulations were carried out with another membrane-binding peptide, Gomesin (Gm), a 18-residue peptide of sequence QCRRLCYQRVCVTYCRGRX, originally isolated from the immune cells of the tarantula *Acanthoscurria gomesiana*¹⁶⁹. This peptide is characterised by a rigid elongated, β -hairpin-like fold stabilized by two disulfide bonds¹⁷⁰. Simulations of Gomesin on a POPC bilayer exhibited the same problems in the exploration of its orientational space as those observed here in simulations of Protx-1. The main findings of these tests are summarised below.

If the exchange with tempered ensembles when the protein is in a vacuum-like conformation is the reason for the above described problems, a possible solution could be to temper the peptide-membrane interactions only, without tempering the peptide-water interactions. This might reduce the likelihood of abnormal conformations of the peptide from appearing. Simulations were attempted by only tempering the Gomesin-POPC interactions and leaving all the other interactions untempered. The ability of the water molecules and ions to interact with the membrane as well as the peptide resulted in these molecules accumulating between the peptide and the membrane surface. The water molecules and ions remained between the peptide and the membrane “bridging” the interaction between the two, with very long residence times in the case of ions. This accumulation of ions at the peptide-membrane contact surface also induced local deformations of the membrane structure, in some cases leading to the extraction of lipids from the membrane in highly tempered replicas¹⁶⁸. This approach was deemed unfit to enhance the sampling of the peptide-membrane system.

Another solution attempted was to partially temper the non-bonded (LJ and electrostatic) peptide-peptide interactions to counteract the formation of the intra-molecular H-bonds. In these simulations, peptide-peptide interactions were tempered concurrently with peptide-water and peptide-membrane interactions. The average number of H-bonds was found to be lower than in the equivalent US-REST3 simulation with the peptide-peptide interaction untempered. This shift in the number of H-bonds to values closer to those observed in classical US simulation was accompanied, however, by a considerable loss of secondary structure of the peptide.

Gomesin was observed to become deformed and α -helical segments replaced β sheets. This solution was thus also deemed unfit for ProTx-1. Other solutions that were attempted included upscaling the peptide-membrane interactions and various *ad hoc* choices of the values of λ ¹⁶⁸, but none of them yielded an ensemble that was physically representative of the expected stable conformation of the protein.

4.4 Conclusions

The US-REST3 methodology developed and tested for small terpenoid molecules was here tested in the context of a protein-membrane interaction using ProTx-I interacting with a pure POPC membrane.

The results suggest that in bigger systems like this one, unphysical conformations accompanied by a higher number of water molecules and ions intercalating between the membrane and the protein can occur even in lower tempered replicas. These abnormal conformations with significant loss of secondary structure can persist for several tenths of nanoseconds, which would therefore require a similar amount of time between exchanges in REST3 to yield a realistic ensemble of conformations. Such long times between exchanges would make the replica exchange approach very inefficient and,

therefore, not feasible for large systems. The inability to replicate a physically representative ensemble for Protx-1, in conjunction with data previously obtained for Gomesin, suggests that in the current state REST3 is a method that cannot be used effectively with highly self-interacting systems like proteins to simulate their interaction with membranes. The method remains, however, a valid approach to study small molecules where the intramolecular network of interactions is not as extensive.

5 Chapter 5 - Development of a US-Metadynamics protocol with a small test system

5.1 Introduction

Molecular dynamics (MD) simulations of molecules interacting with membranes, as already stated in Chapter 1, are affected by the sampling problem. In particular one has to extensively sample all the possible relative orientations of a molecule in respect to the membrane for a given macroscopic state in order to obtain an accurate free energy of binding estimate. Rotational diffusion in these kind of system is generally very slow, either due to an overestimation of the rotational barriers by the FF, because of the capacity of the membrane to partially envelop the molecule under study or just due to the sheer size of the solutes. To overcome this problem a good idea is to enhance the rotational motion by the use of an enhanced sampling method. There are generally two main ways to approach this problem, the first one is to enhance the exploration of every degree of freedom of the system, or of one subset of the system. This is the approach that was attempted in the previous Chapter through the use of REST3.

Chapter 4 assessed the artefacts in REST3 enhanced simulations of proteins, showing that tempering the interaction of the protein with the surrounding of big molecules with extensive self-interactions can lead the system into highly self-interacting conformation which are difficult to escape. The second approach is instead to target specifically only a couple of arbitrary chosen CV, so that the general Hamiltonian of the system have less possibilities for misbehaviour. The CV chosen in this way have to be the chosen so that they are descriptive of the process under study and also the slowest relaxing CVs of the system. For this reason in an attempt to enhance only the rotational motion of the protein without perturbing its structure, we decided to target the rotational motion of the protein through a selection of a limited number of collective variables (CVs). One option would be to use metadynamics (MetaD) to accelerate the sampling of the rotational motion of the protein by applying a potential energy function along a set of angles defined between the protein and the normal vector to the membrane. In this case it is generally advisable to use WT-MetaD as this allows for better estimates of the free energy surface (FES) than standard MetaD¹⁰⁴. The previously defined angles could be used to define an orientation, so that MetaD would deposit a history-dependent potential to disfavour previously explored orientations. An additional CV could be chosen to sample the COM distance in the z-dimension, perpendicular to the membrane. However, WT-MetaD scales poorly with the number of CVs and it is thus not recommended to use more than two CVs in a system with slow diffusion in CV space, such as protein-membrane interactions¹⁷¹. This is due to the increase in volume of the phase space that the Gaussians have to fill with the increase in the number of CVs (see Chapter 1, section

1.9). Based on these factors, it is probably best to instead delegate the exploration of the CV along the z-direction of the centre-of-mass (COM) distance between the molecule and the membrane to US. Ideally the aim would be to apply a MetaD history-dependent bias on the system to cross the barriers that impede the system to explore all its relevant conformations, leading to the removal of the effect of the bias to retrieve the underlying unbiased free energy. The last part of this procedure is called reweighting and can serve to recover the unbiased distribution of quantities even if the simulation was run under the effect of a biasing potential¹⁰⁴. This approach involves some unknowns, for example the kind of interactions that can occur between a static US potential with a history-dependent MetaD potential, potentially producing artefacts or unexpected behaviours.

As the underlying free energy of a complex system like the interaction of a protein with a lipid bilayer is unknown, we first need to characterise a system for which the analytical solution of the free energy is known. Once this system is available all the relevant biased and unbiased ensemble averages can be obtained and, therefore, the actual effectiveness of the bias applied can be assessed. We can first consider the simple case of a system consisting of two atoms in vacuum. A LJ interaction potential is used to describe the interaction between the atoms within a simulation where a MetaD potential is applied on the distance between the atoms. During the reweighting phase the underlying free energy of interaction will of course correspond to a LJ potential energy if the reweighting algorithm used is correct.

Following this reasoning, a simple model system composed of a stationary surface of Xe atoms interacting with a CO₂ molecule was constructed. CO₂ was used because it is a small, linear molecule and it has a centre of symmetry located on the carbon atom. Xe was used just for simplicity, because the interaction of Xe with other species can be represented as a LJ function. In such a simple system it is possible to easily isolate the effect of all the possible variables of the system (e.g. temperature, pressure, COM distance, height of gaussians, etc.) and facilitate the reproduction of the expected shape of the PMF.

Three systems were setup with increasing potential energy complexity to test the reweighting procedures described by Bussi et al.¹⁷². The first system, A, has the simplest potential energy surface, featuring only a harmonic potential on the COM distance between the CO₂ and the Xenon surface (referred to as the COM-distance). In the second system, B, a harmonic potential on the angle α defined between the surface normal and the CO₂ molecular vector is added. In the third system, C, a MetaD potential is applied to the angle α . For each system, the distributions of the COM-distance and of angle α are calculated. Once the distributions of the angle α and COM-distance (dz) are obtained and, if they fit the analytical solution that is expected for every system, the aforementioned

reweighting scheme will be applied to systems B and C to replicate the results obtained from system A.

5.2 Methods

5.2.1 WTMetad

Well-tempered Metadynamics is a variant of classical Metadynamics in which the hill height decreases with time as a function of the bias potential previously deposited on a given value of collective variable. The expression for the bias potential becomes⁹⁴:

$$V(S, t) = k_B \Delta T \ln \left(1 + \frac{\omega N(s, t)}{k_B \Delta T} \right) \quad (24)$$

$$\dot{V}(S, t) = \omega e^{\frac{-V(s, t)}{k_B \Delta T}} \delta_{S, S(t)} \quad (25)$$

Where k_B is the Boltzman constant, T is the temperature, ΔT is an input parameter with the dimension of a temperature, ω is the deposition rate and $N(s, t)$ is the histogram as a function of time and collective variable s . \dot{V} is the time derivative and $\delta_{S, S(t)}$ is the kernel function. ΔT is an important parameter that regulates the virtual temperature of your CV and therefore the amount of final bias that will be deposited along it. We define the bias factor as $\gamma = T + \Delta T / T$ as a functional way to quantify how “hotter” the CV of choice will be compared to the rest of the system. In practice these set of equations is smoothly implemented in the metadynamics algorithm by rescaling the Gaussian height with:

$$W = \omega \tau_G e^{\frac{V_G(s, t)}{k_B \Delta T}} \quad (26)$$

Where τ_G is the deposition stride. By doing so the potential deposited will converge to :

$$V_G(S, t \rightarrow \infty) = -\frac{\Delta T}{T + \Delta T} F(S) + C \quad (27)$$

Where C is an immaterial constant. The bias converges so to a limiting value that is depending of the bias factor used. This method avoids the possible overfilling that can happen with conventional metadynamics and converges to its final value with a lower error¹⁰³.

5.2.2 Systems setup

In system A, a surface comprising 200 Xe atoms arranged in a hexagonal close packed square grid of 50×50 Å, which formed the base for the rectangular box of $50 \times 50 \times 70$ Å. For these atoms, which comprise a surface, the equations of motion were not integrated so that they do not move. A total of 518 Ar atoms were then added to the box to act as background gas to facilitate the thermostating of the system with the target CO_2 molecule. Periodic boundary conditions were applied on the x and y axis, but not on the z axis, to avoid long range interactions between periodic images of the surface. Umbrella windows were then generated by placing a CO_2 molecule at increments of 1.0 Å away from the Xe surface, starting at a COM distance of 8.0 Å from the surface and finishing at 32.0 Å. In each umbrella window, the CO_2 – surface COM distance dz was restrained by a harmonic potential (V_{US}) with a force constant $k = 10$ kJ/mole acting on the carbon atom. A LJ potential (V_{LJ}) was defined starting from a plane parallel to the Xe surface and situated 1.0 Å away from it. This potential only acts on the carbon atom of the CO_2 molecule and was used to avoid complex multi-body convolutions of the potentials (resulting from the sum of the LJ potential at every point in space on the three atoms). The V_{LJ} was defined with $\epsilon = 28.95$ kJ/mole, $\sigma = 10.0$ Å, $r_{\text{cut-off}} = 50.0$ Å, with ϵ being the depth of the energy well and σ being the zero-energy crossing point. Finally, a harmonic repulsive potential with a $k=190$ kJ/mol, starting at zero with a cut off at 4 Å in the z direction was applied to every atom of the system, to keep the Ar atoms from crossing the periodic boundaries. A representation of the system and its components is depicted in Figure 34 .

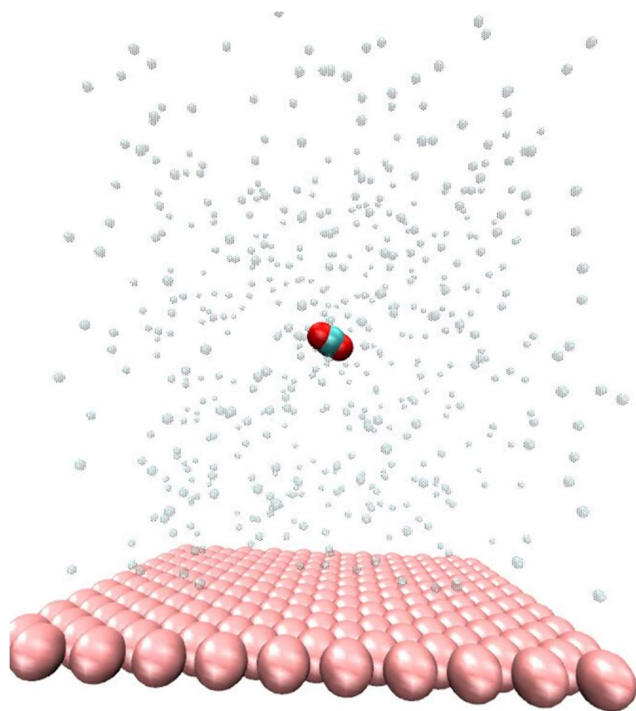


Figure 34. Representation of the simulation system. Xe atoms (pink) constitute the surface, Arg atoms (transparent blue) are used as background gas, and a single CO₂ molecule (light blue and red) that interacts with the atomic surface.

In system B, an additional restraining potential V_α (with a force constant $k = 5$ kJ/mole, $x_0 = 0^\circ$) was placed on the angle α , defined as the angle between the normal vector to the Xe surface and the vector defined by the two oxygen atoms of the CO₂ molecule. In system C, a well-tempered MetaD potential was applied on the angle α , $V_G(\alpha,t)$. $V_G(\alpha,t)$ consists of gaussians of height 1 kJ/mole, a width of 0.03 and with a bias factor of 5.0, deposited every 100 steps on the system. The choice of the width of the Gaussian was made to make the width of a size comparable to the average fluctuation of the system in 500 steps. The choice for the height was arbitrary to obtain a rapid filling of the FES for the expected size of the energy barrier (which by design was known).

All simulations were performed using LAMMPS 29 Oct 2020¹⁷³ patched with PLUMED 2.6.1¹⁷⁴ in a NVT ensemble at a temperature of 298 K controlled using thea CSV thermostat with a relaxation time of 0.1 ps. A short equilibration simulation was run for 1.0 ns, followed by a production run of 40.0 ns. Equilibration of the simulations of systems A and B was assessed by monitoring the temperature of the system. Being small systems and under the effect of a static potential they can be considered to be converged for the entirety of the 40.0 ns production runs. System C is instead under the influence of a history-dependent MetaD potential and hence convergence was assessed by monitoring the decrease in hill height over time. The simulation was considered to have converged when the hill height was lower than 0.01 kJ/mole for each α value. The simulation reached convergence in the first 20.0 ns, so only the last 20.0 ns of simulation C were used for the analysis. The potential energy in this part of the simulation was considered to be quasi-stationary for reweighting purposes.

5.2.3 Removal of thermostat outliers

These systems are very small and, therefore, are very sensitive to fluctuations in energy and forces due to random collision events. In such a small system, the background gas (Ar) density is far lower than the real density of gaseous Ar (1.784 g/L). Approximately 4800 atoms of Ar would be required in the simulation cell for the density to be close to the real density. Consequently, due to the much lower density of the system, collisions are much less frequent than they should be. As a result, the CO₂ molecule occasionally collides with a particularly energetic Ar atom (belonging to the tail-end of the kinetic energy distribution) and is knocked into a ballistic trajectory, not experiencing a second collision for a few time steps. If the energy of the collision is high enough the displacement over dz in

the following few time steps can be very high, which causes an artificially disproportionate weight on the ensemble for the corresponding frames. To address this problem, we assumed a normal distribution around the mean value μ_{dz} of each umbrella and discarded every frame that had a displacement $\mu_{dz}+3\sigma$ or $< \mu_{dz}-3\sigma$. Following this criterion, more than 99.5% of the simulation trajectory frames were retained in all systems and the outlying frames were discarded. This has resulted in the removal of 180 of 40000 frames of the simulation, which was considered a number low enough not to be of concern for the accuracy.

5.2.4 Reweighting of the simulations

The probability of a microstate with some position q and velocities p , under the effect of no bias, is

$$P(q, p) \propto e^{-\frac{H(q,p)}{kbT}} \quad (28)$$

where $H(q,p)$ is the Hamiltonian of the system. If the system is under the influence of a static or quasi-stationary potential $V(q)$, its biased probability $P'(q)$ becomes instead

$$P'(q, p) \propto e^{-\frac{H(q,p)}{kbT}} e^{-\frac{V(q)}{kbT}} \quad (29)$$

Substituting 4 into 5 and rearranging we obtain

$$P(q, p) \propto P'(q, p) e^{+\frac{V(q)}{kbT}} \quad (30)$$

These equations are equally valid when taking into account the probabilities in CV space, such that

$$P(s) \propto P'(s) e^{+\frac{V(s)}{kbT}} \quad (31)$$

Equation 5 links the unbiased probability of a system with its biased probability weighted by the bias applied to that conformation. Using the ergodicity theorem we can approximate the probability $P(q,p)$ as the histogram $\bar{H}(s)$ gathered from the simulation when the bias is quasi-stationary and, therefore, obtain the free energy as

$$F(s) = -KT \ln P(s) = -KT \ln \bar{H}(s) = -KT \ln \bar{H}'(s) e^{+\frac{V(s)}{KT}} \quad (32)$$

Since the free energy is obtained as the natural logarithm of $\bar{H}(s)$ multiplied by KT and the natural logarithm is a monotonic function, the shape of the two are the same, so only the latter is reported rather than converting it to a FES.

5.2.5 Histogram calculation

The free energy surface along dz was calculated using a similar method to the aforementioned free energy calculation in α . In particular the joint information from all the umbrellas windows is desired

to reconstruct the FES along dz . This means that the weight that a given simulation trajectory frame of the entire “concatenated” umbrella sampling trajectory has is calculated accounting for all the harmonic potential in that given frame. To obtain these weights first all the singular umbrella trajectories were concatenated into a single macro trajectory. The bias exerted by every external potential applied on the system, when the bias is static or quasi-stationary, was then obtained. This operation is repeated for the potential of every window at every point in time along the concatenated simulation. That is achieved using the PLUMED driver on the concatenated trajectory for every umbrella window and combining all the external potentials acting on that window. This yields a $N \times k$ matrix where N is the number of umbrellas and k is the number of frames of the concatenated trajectory. Using these potentials WHAM was used to obtain the weights for every trajectory frame, which in turn can be used to recalculate unbiased distributions of any property in the system. A thorough description of the mathematical derivation of this reweighting algorithm has been provided by Bonomi et al.¹⁷²

It's worth noting the in the case of system C, only the last 20.0 ns of the trajectory had to be used to perform this calculation, as this theoretical framework works only when the potential can be considered quasi-stationary.

5.3 Results and discussion

The simulation described in system A is used to obtain the target distributions of the angle α and of dz . These values are reported in Figure 35A and Figure 35B and should be considered to be the expected values of the reweighted distribution for systems B and C. The distribution of the cosine of the angles is expected to be flat because in the absence of any directional potential the molecule should experience free rotation. Figure 35A reports the distribution for $\cos(\alpha)$, showing that it oscillates steadily around 0.5 as expected by the absence of biasing potentials on the angle α . The distributions of dz are each peaked around the value at which the harmonic potential is centred around, with a slight displacement in windows near the minimum of the LJ potential. This is to be expected because at small values of displacement from the centre of the harmonic potential the LJ potential manages to overcome the harmonic potential, such that the resulting potential will be slightly off-centred.

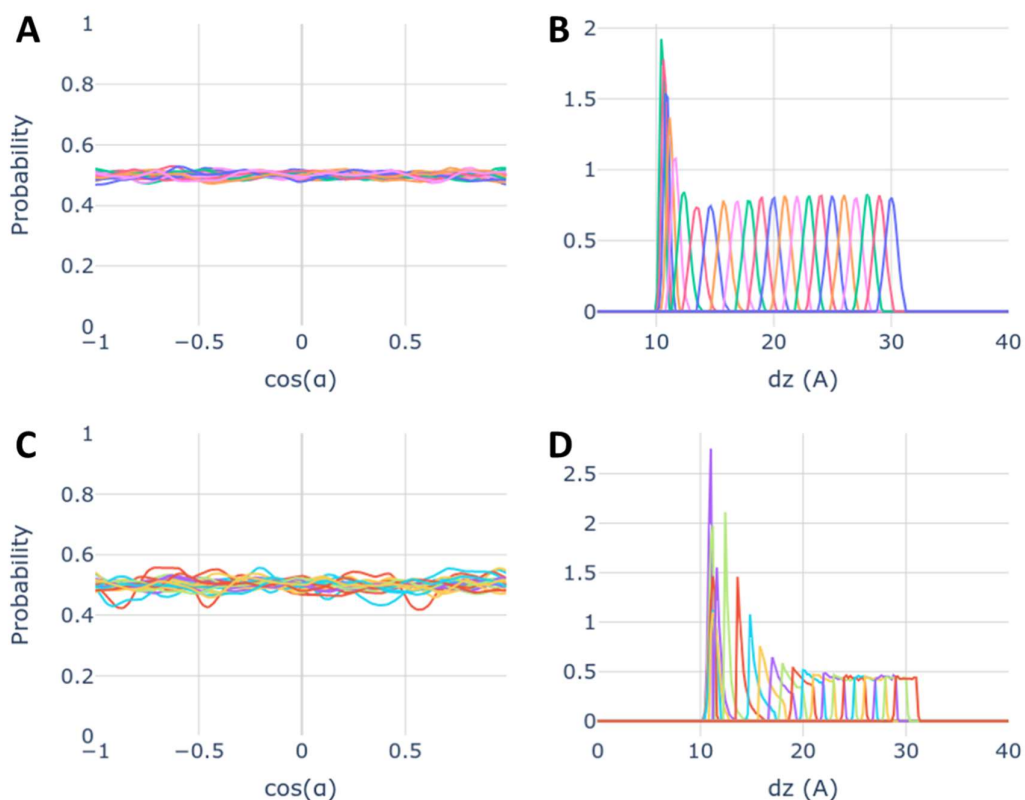


Figure 35 Distribution of quantities of interest for system A. A: Biased distribution of $\cos(\alpha)$. B: Biased distribution of dz . C: Reweighted distribution of $\cos(\alpha)$. D: Reweighted distribution of dz .

The reweighting protocol was first tested on system A to reweight the distributions of $\cos(\alpha)$ over the US potential that was applied along dz . Figure 35C and D report the distributions after reweighting. As expected the distributions for $\cos(\alpha)$ remain largely unchanged due to the absence of a potential that acts on α . The reweighted distribution of dz instead indicates nearly flat curves for far away windows and tilted distributions following the slope of the underlying LJ free energy well for windows closer to the surface. From these data it appears that the reweighting protocol can correctly unbiased static potentials in this test system.

In system B the restraint around 0° is arbitrarily chosen to mimic a specific orientation restriction that a surface could impose on a molecule when they come close enough to interact. In these conditions we expect the angle α to sample preferentially angles around 0° , but the shape of the free energy along z should still resemble a LJ potential. Figure 36 reports the biased and reweighted distributions for dz and $\cos(\alpha)$ for system B.

It can be seen that the dz plots remain unchanged compared to system A, which should be expected since no extra potential was applied on dz . The biased distribution of $\cos(\alpha)$ shown in Figure 36A has

the shape of the negative of the harmonic potential applied on it, with a peak around 0° which corresponds to the umbrella centre. Figure 36C instead shows that after reweighting using both V_L and V_{US} the distribution of $\cos(\alpha)$ returns flat. This means that the reweighting algorithm can remove effectively the influence of two static biases acting on this system.

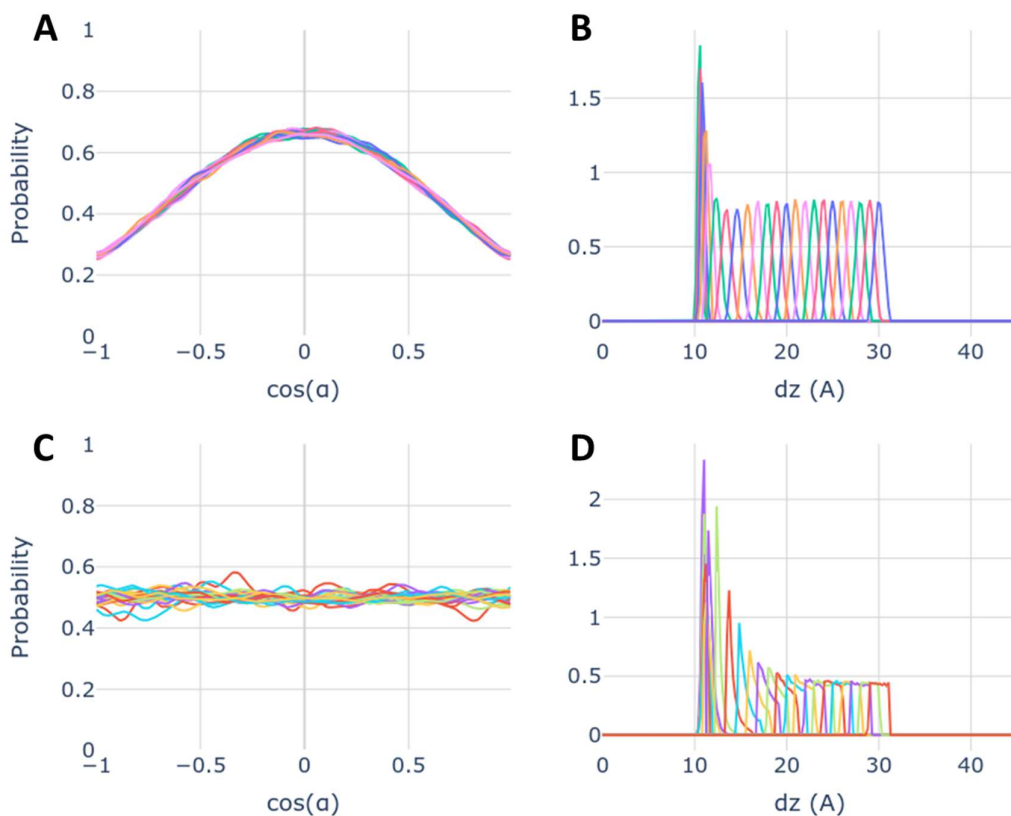


Figure 36 Distribution of quantities of interest for system B. A: Biased distribution of $\cos(\alpha)$. B: Biased distribution of dz . C: Reweighted distribution of $\cos(\alpha)$. D: Reweighted distribution of dz .

Finally, in system C MetaD was used on the angle α to overcome the restraining effect of the potential V_α . The idea behind this system is to determine if MetaD can overcome the conformational restriction that the restraint on the angle α imposes on the system and return the correct distribution of the angle probability using standard reweighting techniques¹⁷⁵ the underlying free energy on $\cos(\alpha)$ and on dz . Once again the distributions after the reweighting should be expected to be similar to the ones obtained in system A, even if a history-dependent potential is applied to the angle α . Figure 37 shows the distribution before and after reweighting of.

As it can be seen from Figure 37A, the shape of the biased distribution resembles the one in Figure 36A, but the depth of the well has been filled by the WT-MetaD potential. After reweighting the probability appears flat, as expected, indicating that the reweighting method is working as intended

for the two static potentials in conjunction with a quasi-stationary WT-MetaD potential. WT-MetaD can overcome the orientational restraint given by the harmonic potential and, after reweighting, the simulation yields the correct free energy surface.

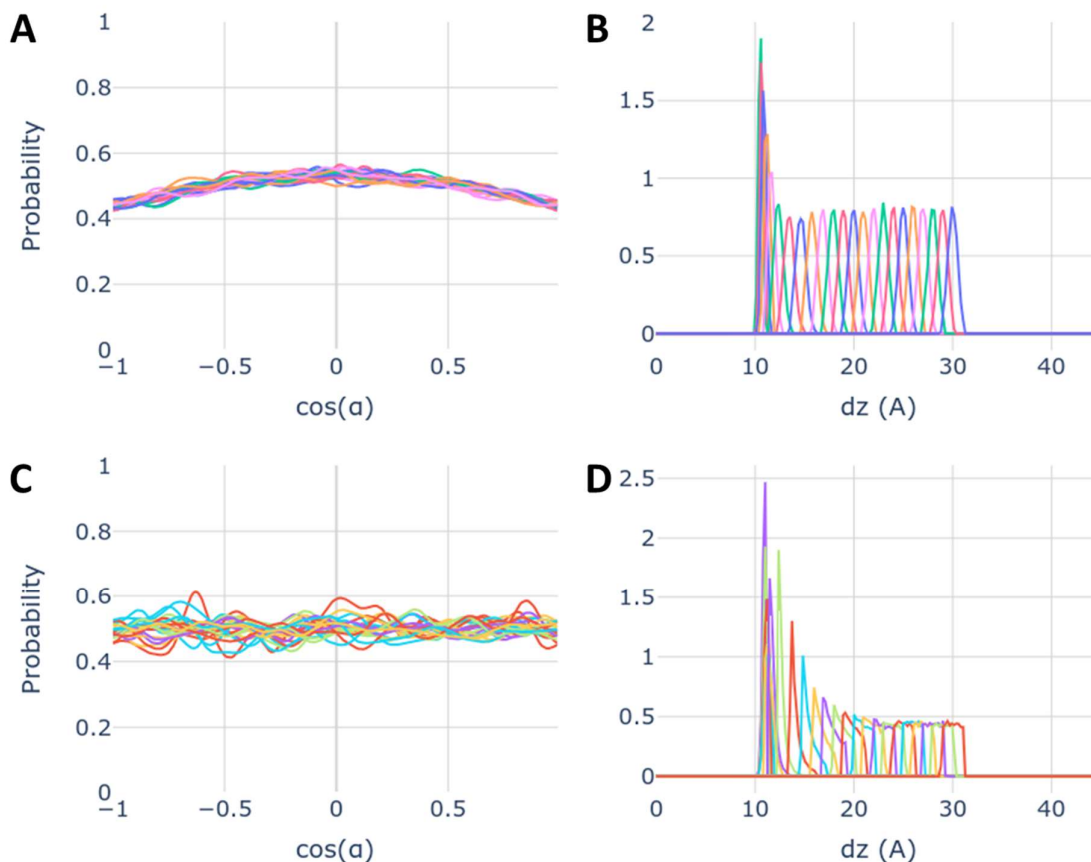


Figure 37 Distribution of quantity of interest for system C. A: Biased distribution of $\cos(\alpha)$. B: Biased distribution of dz . C: Reweighted distribution of $\cos(\alpha)$. D: Reweighted distribution of dz .

Figure 38 shows the FES along dz obtained for the three reweighted ensembles using the WHAM protocol described in the Methods section. As it can be seen from the plot the three PMF, despite having different potential energy functions match perfectly after the reweighting. It can be concluded that metadynamics can be used to overcome orientational restraint in an umbrella sampling simulation and that the reweighting protocol here described is able to correctly remove the effect of the bias on the system.

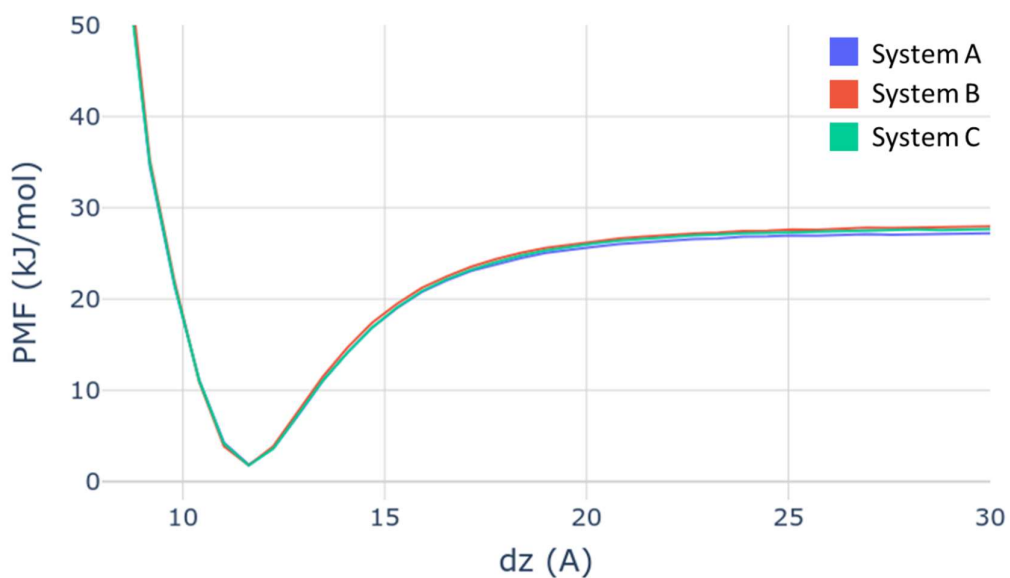


Figure 38 FES of the three system along dz calculated using the wham method.

5.4 Conclusion

In this chapter a protocol to use a combined US-MetaD was developed and validated. The underlying LJ free energy curve that was expected by construction of the system was reproduced correctly through reweighting of the simulation A, B and C independently of the bias deposited on the system. The results presented suggest that this approach could be suitable in principle for more complex systems, such as those involving proteins and lipid membranes. This would be true if the protocol uses an orientation angle as CV and as long as the potential acting on that angle is quasi-stationary.

6 Chapter 6 – Development of a CV for semi-rigid objects rotation

6.1 Introduction

Assessing sampling of the rotational motion of a peptide or small molecule on a membrane surface requires the definition of angles to describe its relative orientation with respect to the membrane. In the previous chapter we used the angle α between a vector and the z axis to measure the relative orientation of the CO₂ molecule in respect to a Xe surface. CO₂ is a small, linear molecule with internal symmetry and its orientation in space can be approximated by a single vector passing through its two apical oxygen atoms. Furthermore, applying a potential on the angle α means applying a force to two out of the three atoms in the CO₂ molecule. This angle can thus be used to bias the rotation of the CO₂ molecule on the surface without distorting the shape of the molecule itself. In the case of proteins, which are much larger and more flexible than small drug-like molecules, the definition of an orientation and the application of a potential to bias its rotation on the surface is more difficult. In the absence of any geometric symmetries, a 3D rigid object requires at least three angles to represent its orientation completely. In addition, if the protein is flexible, the time-dependent fluctuation of its protein means that the physical meaning of an orientation can get lost or change concurrently. As an example, if a vector were to be defined between two atoms on an elongated protein hairpin it could easily lose its meaning if the hairpin opens up or changes conformation substantially during the simulation. In the case of ProTx-1, the rigid structure of the ICK motif means that the secondary and tertiary structure of the protein will remain mostly unchanged even during a long simulation, even if atoms rearrange locally. As a result, the protein can be defined as a 3D semi-rigid object and its orientation on the membrane surface can be defined by three vectors. This property is an important prerequisite for the method described in this Chapter.

Even if the orientation of a protein can be defined by three vectors, the definition of these vectors by two atoms can cause problems when a force is applied in MetaD. To illustrate this, let us consider the example of a protein rotating on a membrane surface. The system is set up so that the membrane is parallel to the xy plane of the simulation cell, and its surface orientation can be conveniently defined by the unit vector along the z axis (normal to the x-y plane). The orientation of the protein in space can be defined by a vector connecting any two atoms at opposite ends of the protein. The CV, θ , is then defined as the angle formed by this vector in the protein and the z axis normal to the plane of the membrane.

Applying an external biasing potential on θ in MetaD means applying a force on these two atoms to disfavour any previously visited orientations. In a small and rigid molecule like CO₂, where the vector involves the majority of atoms in the molecule, this force will enhance the rotation of the entire

molecule since its flexibility is limited. The same is not true for larger and more flexible molecules. For proteins, it is usually the case that the energy barriers for the movement of two atoms that define the vector to locally shift their positions to obtain a change in θ are substantially lower than the energy barriers for the rotation of the whole protein to rotate. In a plot of θ vs time it would appear that the orientation of the protein is changing, whereas only the respective positions of the two atoms are actually changing, leaving the orientation of the protein on the membrane surface almost unperturbed. In Figure 39 two distributions of θ , one from an unbiased distribution (blue) and one from a MetaD enhanced distribution (red) are shown. The MetaD distribution is much wider than the unbiased one, but the protein does not rotate its orientation with respect to the membrane at all. For this reason, some CVs that depend on the positions of all of the atoms of the protein rather than on a finite set is preferable. For such a CV, the force exerted by the MetaD potential will be applied to all atoms, causing a collective rotation of the entire protein.

Various approaches are possible to address this problem¹⁷⁶. One very efficient solution would be to consider the protein as a rigid body and use quaternions¹⁷⁷ to define an absolute orientation and make rotations to it. The problem with this solution is that to deposit a biasing potential on an orientation defined by quaternions the MD engine has to be able to use quaternion algebra. Most MD engines currently lack this capacity and, therefore, implementation of quaternions-based orientation routines becomes a laborious task. A second, less sophisticated approach but more easy to implement is to use Euler angles to define the orientation of a molecule¹⁷⁸. Euler angles are a set of three angles used to represent the orientation of a rigid body with respect to a fixed reference coordinate system. This representation has been widely used to represent rotations in many other fields, such as mechanical engineering, aviation and computer games. In this chapter a new CV based on Euler angles to define and bias the orientation of a semi-rigid macromolecule is presented. This set of Euler angles will be used to define the orientation of the protein on the membrane surface using all atoms in the protein. In practice, MetaD simulations using Euler angles to define an orientation CVs are easily achieved using GROMACS 5.1.4 patched with PLUMED 2.4.4 through the use of the PLUMED routine FIT_TO_TEMPLATE. This new Euler angle CV was tested to enhance the rotational sampling of ProTx-1 on a POPC membrane surface. The aim was not to compute a free energy profile (PMF), but to define a set of appropriate CVs to enhance the orientational space sampled by the protein without artificially creating local distortions to the protein structure. Conventional and WT-MetaD is renowned to be slow converging for simulations with more than two CVs. Therefore, a MetaD variant that could make use of multiple CVs and quickly converge despite the number of CVs was used. The method used was parallel bias MetaD, by virtue of its ability to handle multiple uni-dimensional potentials and, therefore, converge much faster than a regular multi-dimensional MetaD simulation. In this method

the information given by the singular uni-dimensional potentials can be then combined into a multi-dimensional potential that approximate the corresponding potential that would be obtained using a multi-dimensional Gaussian.

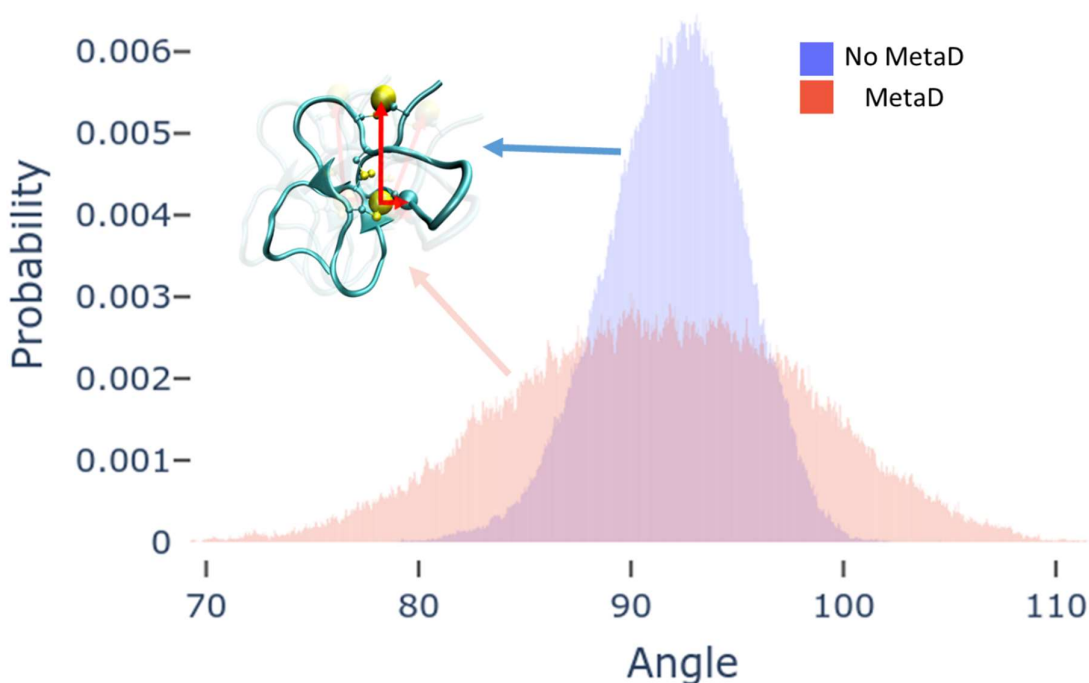


Figure 39. Two angle distributions with very different widths leading to the same overall conformational ensemble when considering the protein. In blue: conventional distribution of an angle between an atoms defined vector and the z axis. In red: Enhanced distribution of the same angle using MetaD.

6.2 Methods

6.2.1 Definition of Euler angles

The concept behind the definition of these three angles is to define a reference structure that includes only selected atoms of the molecule whose rotational diffusion needs to be enhanced. A set of axes defining its simulation cell is associated with this structure. A structure at a successive point in time t will have the same set of simulation cell axes but a different orientation of the molecule. To determine the angle of rotation, the FIT_TO_TEMPLATE routine in PLUMED is used to perform rigid roto-translation of the structure at time t onto the reference structure. The FIT_TO_TEMPLATE routine makes use of an implementation of the Kabsch algorithm¹⁷⁹ to find the optimal rotation to

superimpose two sets of vectors. The Euler angles are then defined between the reference set of axes and the rotated ones, so that they depend on all of the atoms of the initial selection.

As an example, to enhance the rotation of a solvated protein interacting with a surface, the reference structure is stripped of all the atoms that do not belong to the protein. Let us call this reference structure R_0 defined by the position $[x, y, z]$ of every atom of the protein at time 0 and the vectors defining the axes of the simulation cell are \mathbf{x}_0 , \mathbf{y}_0 and \mathbf{z}_0 (noted in bold as they are vectors). At any given time t , the system will be in a different orientation R_t , with its simulation cell axes unchanged. We will call these axes \mathbf{x}_0' , \mathbf{y}_0' and \mathbf{z}_0' to distinguish them from \mathbf{x}_0 , \mathbf{y}_0 and \mathbf{z}_0 . We now perform an alignment of R_t onto R_0 , which can be done via the FIT_TO_TEMPLATE function but generally it can be done by finding the optimal roto-translation that minimises the RMSD, using standard alignment methods¹⁸⁰.

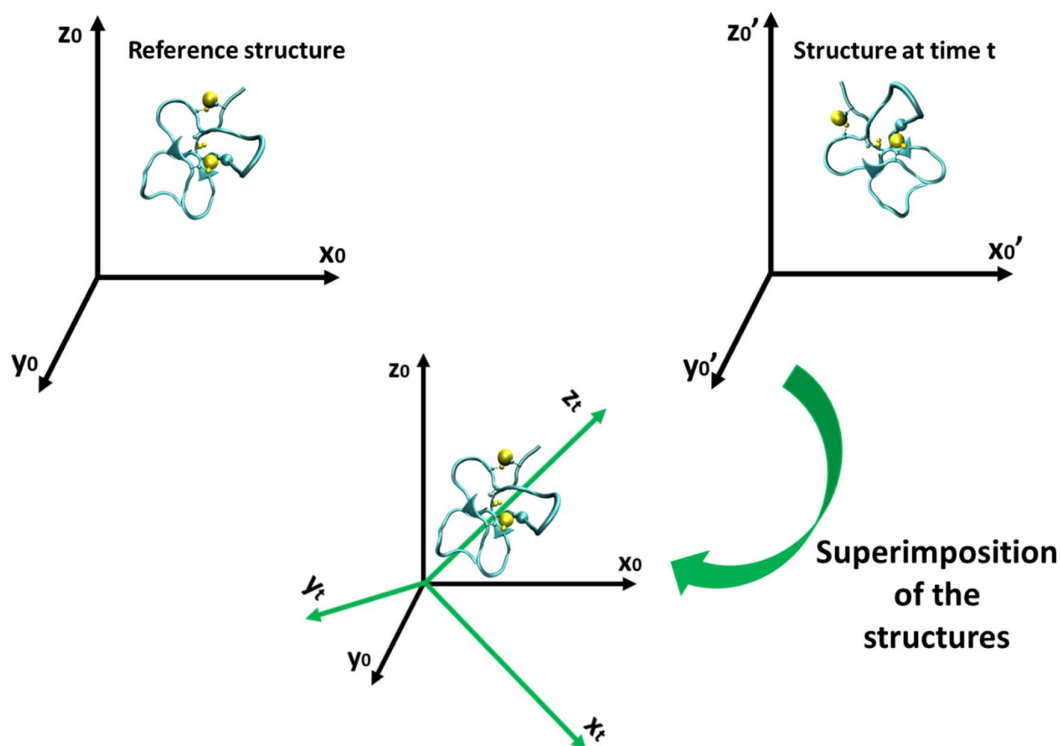


Figure 40. Graphical representation of the procedure to obtain the rotated set of axis (x_t, y_t, z_t)

Note that this rotation is calculated using all atoms of the protein but no other atoms in the system. For a protein-membrane system, this alignment therefore does not involve the membrane. As a result of this alignment operation, the axes \mathbf{x}_0' , \mathbf{y}_0' and \mathbf{z}_0' will be rotated to a new set of axes that we call \mathbf{x}_t , \mathbf{y}_t and \mathbf{z}_t as presented in Figure 40. Following this rotation, the three Euler angles between \mathbf{x}_0 , \mathbf{y}_0 , \mathbf{z}_0 and \mathbf{x}_t , \mathbf{y}_t , \mathbf{z}_t are calculated. This is done by first calculating the vector perpendicular to z and z_t :

$$\mathbf{N} = \mathbf{z} \times \mathbf{z}_t \quad (33)$$

The three angles α , β and γ are then calculated:

$$\cos \alpha = \frac{\mathbf{x}_0 \cdot \mathbf{N}}{\|\mathbf{x}_0\| \cdot \|\mathbf{N}\|} \quad (34)$$

$$\cos \beta = \frac{\mathbf{z}_0 \cdot \mathbf{z}_t}{\|\mathbf{z}_0\| \cdot \|\mathbf{z}_t\|} \quad (35)$$

$$\cos \gamma = \frac{\mathbf{N} \cdot \mathbf{x}_t}{\|\mathbf{N}\| \cdot \|\mathbf{x}_t\|} \quad (36)$$

As these angles are continuous and differentiable functions of all the atoms in the protein, they can be used in conjunction with MetaD to enhance the rotation of the protein on the membrane surface.

Note that this process requires an alignment obtained through the optimal roto-translation that minimises the RMSD, so this process is applicable as long as the protein does not undergo major structural changes during the simulation, i.e. a protein whose structure can be approximated by a semi-rigid body. If it is the case that some domains of the protein undergo structural changes, while others retain their structure, the same alignment procedure can be applied only to those domains that retain their structure. In this case, a selection of atoms from the domain that retains its structure during the simulation can be made (e.g. the main chain of the domain of interest) and the vectors of the simulation cell and their rotation is then defined only on this selection.

6.2.2 Parallel bias metadynamics (PBMetaD)

Classical MetaD is a well-established method to enhance sampling in MD simulations. It makes use of a history-dependent potential in the form of Gaussian potential deposited on a collective variable (CV) of choice to move the system away from previously visited states⁹⁴. Use of a Gaussian potential of fixed height (magnitude), though, results in an estimation error of the calculated free energy. This problem is addressed by using well-tempered MetaD (WTMetaD), a variant of classical MetaD in which the Gaussian potential decreases in height as a function of simulation time¹⁰⁴.

Another limitation of classical MetaD is that using multi-dimensional Gaussian potentials to enhance the exploration of more than one CV makes it much more difficult for the simulation to converge because of the growth in size of phase space that needs to be explored¹⁸¹. As a result, WTMetaD with a N-dimensional Gaussian where $N > 3$ becomes computationally prohibitive for most simulation systems. To address this issue several approaches have been developed^{89,90,176}. A commonly used approach is parallel bias MetaD (PBMetaD), a variation of classical MetaD in which multiple, one-

dimensional bias potentials are applied in parallel simulations instead of a single multi-dimensional one¹⁷¹. These one-dimensional bias potentials are much easier to converge than a multidimensional one. During post-processing, the uni-dimensional potentials are then recombined with standard reweighting techniques into an approximate, multi-dimensional free energy surface¹⁸².

In WTMetaD the potential is expressed as:

$$V_G(S, t) = k_B \Delta T \log \left[1 + \frac{\omega N(S, t)}{k_B \Delta T} \right] \quad (37)$$

Where k_B is the Boltzmann constant and ω and ΔT are input parameters with dimensions of energy rate and a temperature, respectively. At the same time, $V_G(S, t)$ can be read as a sum of Gaussian potentials as expressed in the original MetaD method:

$$V_G(S, t) = \int_0^t \omega(t') \exp \left(- \sum_{i=1}^d \frac{(S_i(R) - S_i(R(t')))^2}{2\sigma_i^2} \right) \quad (38)$$

with

$$\omega(t) = \omega \exp \left(- \frac{V_G(S, t)}{k_B \Delta T} \right) \quad (39)$$

Here σ_i is the width of the Gaussian potential on the i -th CV, $\omega(t)$ is the deposition rate and $S_i(R(t'))$ is the value of the i -th CV, S_i , at time t' . If one wants to deposit multiple, uni-dimensional potentials in parallel, each with its own ω_i , the fact that CVs are generally correlated has to be considered. The correlation of CVs can lead to each of the potentials to converge to the wrong free energy. For example, for two CVs, the biasing potential of CV1 can be “felt” by the biasing potential on CV2.

To solve this issue a weight $P(\eta/R)$ is introduced, with η being a discrete variable with two possible values: (0,1) and (1,0). Every time biasing potentials are updated during the simulation, these variables are used to decide which potential should be updated; (1,0) for V_{G1} or (0,1) for V_{G2} . A more extensive discussion of η and the derivation of $P(\eta/R)$ is available in the original paper on PBMetaD by Pfaendtner et al¹⁷¹. This weight $P(\eta/R)$ is applied to eq. 39 for both of the CVs, so that the deposition rates now take the form of

$$\omega_1(t) = \omega_1 \exp \left[- \frac{V_G(S_1, t)}{k_B \Delta T_1} \right] P(\eta = (1,0)|R) \quad (40)$$

$$\omega_2(t) = \omega_2 \exp \left[-\frac{V_G(S_2, t)}{k_B \Delta T_2} \right] P(\eta = (0,1)|R) \quad (41)$$

with

$$P(\eta = (1,0)|R) = \frac{\exp[-V_G(S_1, t)]}{\exp[-V_G(S_1, t)] + \exp[-V_G(S_2, t)]} \quad (42)$$

$$P(\eta = (0,1)|R) = \frac{\exp[-V_G(S_2, t)]}{\exp[-V_G(S_1, t)] + \exp[-V_G(S_2, t)]} \quad (43)$$

In practice at every step some amount of biasing potential is deposited on all the CVs, but $P(\eta/R)$ assigns more weight to the potential deposited on the CVs with lower value of the bias potential. Doing so it can be shown that the two biases converge to the exact potentials in the same way that conventional WTMetaD does¹⁷¹. As a note, the biasing potential used can be a conventional Gaussian potential or a WTMetaD potential with the height of the Gaussian decreasing in time. In this work we opted for the second option, to avoid problems of overfilling of the FES.

6.2.3 PBMetaD Simulations

The Euler angle CVs were tested using two simulation systems: i) ProTx-1 in solution and ii) ProTx-1 on the surface of a POPC membrane. The first system is composed of a single ProtX-1 protein in a 6.4 x 6.4 x 5.5 nm simulation cells and solvated with 7329 SPC water molecules. A total of 14 Na⁺ ions and 18 Cl⁻ ions were added to neutralize the +4 charge of the protein and set the concentration of NaCl at 0.1 M. The FF chosen was Gromos 54A8, a newly refined version of the classic Gromos 54a7 FF with special treatment for the lipid head groups, which were previously known to overestimate lipid-protein interactions⁸⁷. A reaction field with $\epsilon = 62$ was used for the electrostatic interaction treatment and a straight cut-off of 1.4 nm was used for vdW forces, following the original parametrization of Gromos 54a7 parameters¹⁸³. The system was then equilibrated in an NVT ensemble at 298 K using the Berendsen thermostat with $\tau_t = 0.1$ ps for 10 ns and a time step of 2 fs. After that a further simulation for 10 ns in the NPT ensemble at a pressure of 1 bar was conducted using the Berendsen barostat with $\tau_p = 0.5$ ps. The final configuration from this equilibration was used to run a PBMetaD simulation using the same parameters as before, and using angles α and β as CVs for 100 ns with a deposition rate of 500 steps and a Gaussian potential height of 2.0 kJ/mol.

As this work studied the interaction of ProtX-1 with a POPC bilayer and a lipid bilayer is symmetrical to rotations around the z-axis, it was only necessary to enhance the sampling along two of the angles

with respect to the bilayer surface. The use of γ as CV was therefore not necessary. The Gaussian potential height $h(s)$ decreased over time and this was used as the criterion to determine if the bias potential had converged when $h(s)$ fell under 0.05 kJ/mol for each of α and β values. This threshold was chosen as indicating an arbitrarily small change in the height of the Gaussian after which the difference in the long run for that point of the CV would have been negligible. This was observed to occur after 10 ns in the simulation; therefore, the first 10 ns were discarded and the analysis performed on the last 90 ns.

The second system is composed of a lipid bilayer containing 128 POPC molecules, with a molecule of ProTx-1 initially located at a centre-of-mass (COM) distance of 2.6 nm in the z axis from the COM of the bilayer. This configuration was extracted from the converged region of the umbrella window at a COM distance of 2.6 nm described in Chapter 2, section 2.2. This position approximately corresponds to the molecule lying on the bilayer surface, which in previous simulations yielded the narrowest peaks in the orientation angle probability heat maps. In this configuration, with the same parameters described above, a WT MetaD simulation was performed for 100 ns using angles α and β as CVs with a deposition rate of 500 steps and a Gaussian potential height of 2.0 kJ/mol. A harmonic potential was applied with a force constant $k = 500$ kJ/mol centred at a COM-distance of 2.6 nm in the z-axis between the protein. The simulation was not run until convergence, but was used to assess the equivalent orientational space explored by a WMetaD simulation using the Euler angles as CVs compared to 100 ns of a conventional US simulation with no MetaD.

All the simulations were performed in GROMACS 5.1.4¹⁶² patched with PLUMED 2.1.4^{184,185} and the analysis was carried out using standard GROMACS tools in conjunction with MDAnalysis¹⁶³.

6.3 Results and discussion

To validate the use of Euler angles as CVs, we need to first assess if they can reproduce the free rotation of the protein in the absence of the membrane without artificially inducing preferential conformations. For this, two simulation of ProTx-1 in water were carried out. In the first simulation no bias was applied. In the second simulation PB-MetaD was used with two Euler angle CVs: α and β . To compare the sampling in the two simulations, the cosine of the angle β was measured in both of the simulations and the probability of $\cos(\beta)$ was calculated (Figure 3). The angle β was chosen because it represents the angle in the z axis between the reference conformation and the conformation at time t (as reported in eq. 3) and, therefore, is the most significant rotation. As seen in Chapter 5 the choice of measuring the cosine of the angle was done because for an ensemble of random 3D vectors, if the vectors are equally distributed on the surface of a sphere, the distribution of the cosine of their angles with another fixed vector will be approximately flat. After 100 ns of simulation, the distribution of

$\cos(\beta)$ shows that the orientational space sampled is qualitatively the same and there seems to be no particular preference for specific protein orientations. This means that both of the simulations exhibit a freely rotating protein molecule and, therefore, the new CVs do not induce artificial energy minima in the orientational space.

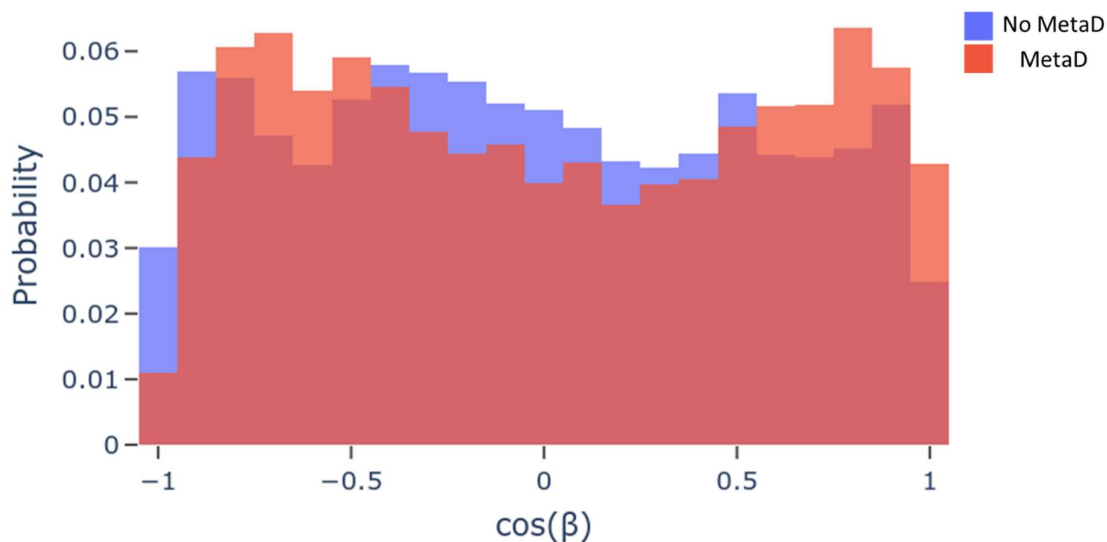


Figure 41 Cosine of angle β as defined in section 2.1 for a classical MD simulation of Protx1 in water and a PBMetaD enhanced simulation using α and β as CVs. In blue: distribution obtained from the conventional MD. In orange: Distribution from the MetaD simulation. Dark red is the overlap between the two distributions.

The next test was to determine if the biasing potential of the Euler angle CVs induces any local deformation in the structure of the protein. This needs to be avoided as inducing structural deformations while biasing protein rotation would likely compromise the validity of simulations aimed calculating the binding affinity of protein-membrane systems. Radius of gyration (R_g) and root mean square displacement (RMSD) of the alpha carbons with respect to the starting structure were computed for the protein in water with and without the use of PBMetaD.

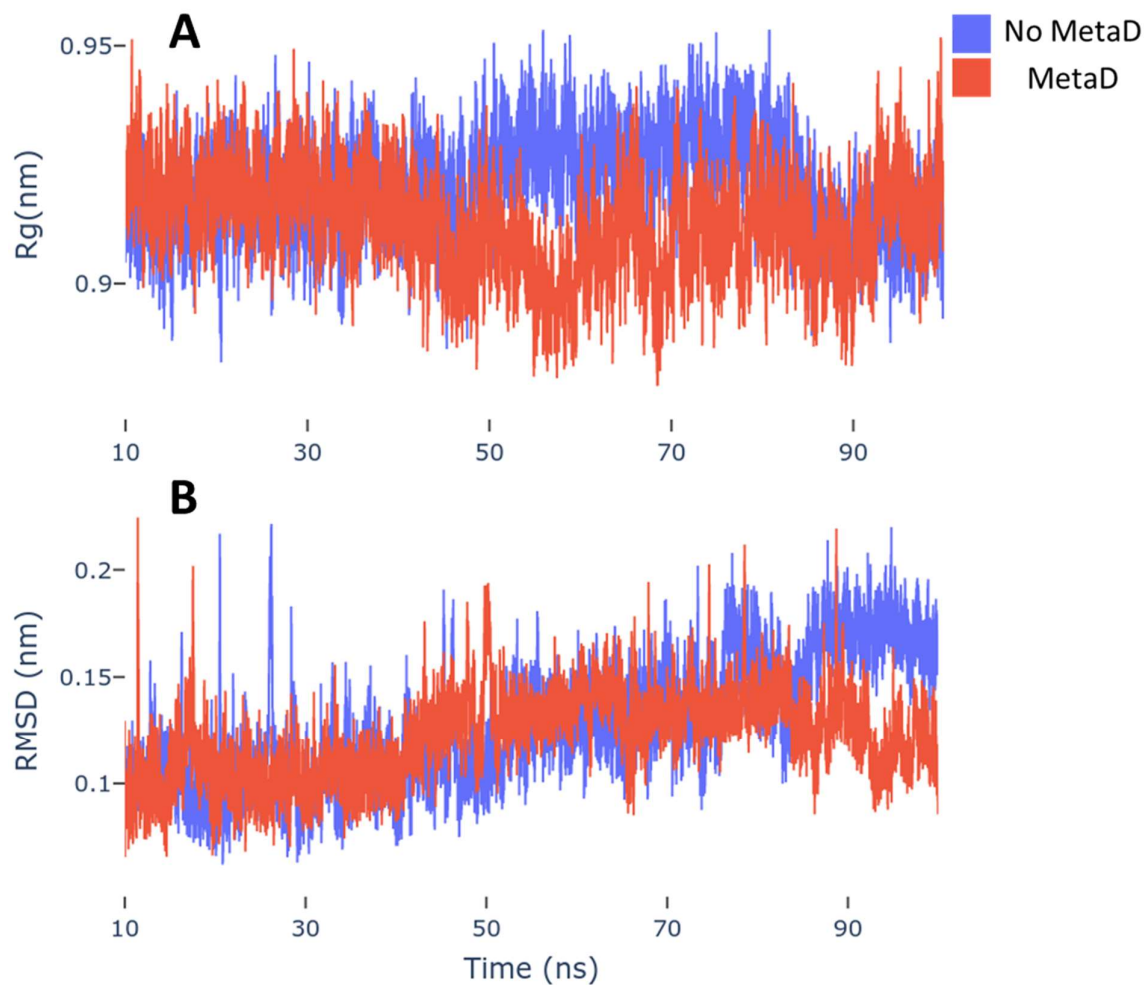


Figure 42. A: Radius of gyration (Rg) of the main chain protein structure in water. B: RMSD of the alpha carbon of the protein structure in water.

As it can be seen in Figure 42 the fluctuations of Rg and alpha carbon RMSD reflect expected fluctuations and do not exhibit any significant change in time. The average Rgs are 0.92 ± 0.01 nm for the classical MD simulation and 0.91 ± 0.01 nm for the PBMetaD simulation. The average alpha carbon RMSD the values are respectively 0.14 ± 0.03 nm and 0.12 ± 0.02 nm. These values are statistically indistinguishable, indicating minimal changes in the overall structure of the protein. This, in conjunction with the previous analysis that confirmed the absence of induced angular preferences, suggests that the Euler angle CVs can be safely used to enhance orientational sampling in a simulation of a protein in solution with PBMetaD without creating artefacts.

The second simulation system was used to assess the effectiveness of these orientation angle CVs in a protein-membrane system. For this a simulation of ProTx-1 with a POPC bilayer was carried out, using a starting configuration extracted from the converged region of an orientationally locked

simulation previously described in Chapter 2, Section 2.3 . The protein was restrained at a peptide-membrane COM distance of 2.6 nm with a harmonic potential and its orientational sampling is enhanced using PBMetaD to enhance sampling along α and β . As a comparison, a US simulation without PBMetaD was carried using the same starting configuration.

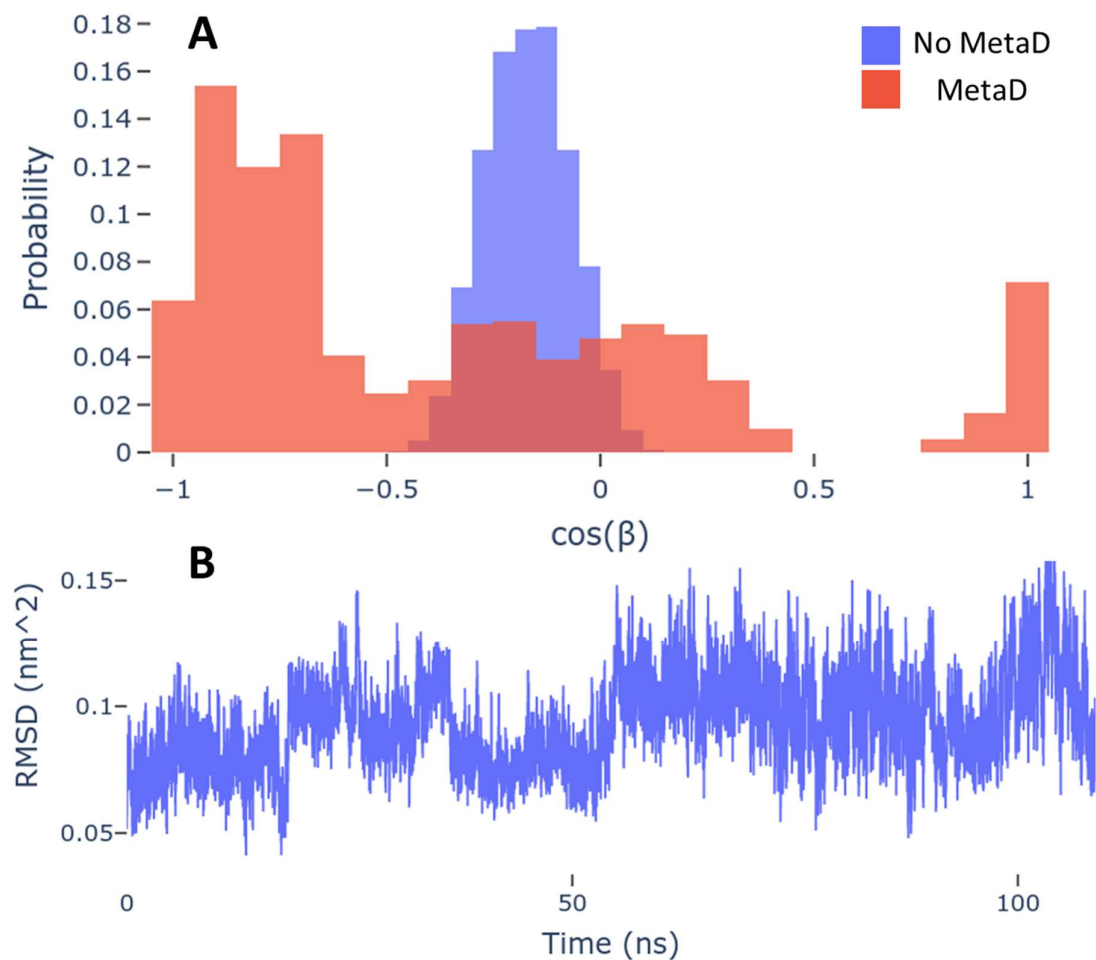


Figure 43 A: Comparison of the $\cos(\beta)$ distribution of simulations restrained at a protein-membrane COM distance of 2.6 nm using US and US with PBMetaD-US . B: RMSD of the protein C α from 100 ns of US-PBMetaD simulation.

Figure 43 compares the distribution of $\cos(\beta)$ for the conventional US simulation and the PBMetaD-US simulation. It can be seen that the distribution sampled during a 90-ns PBMetaD simulation is much wider than the equivalent distribution observed in a conventional US simulation.

6.4 Conclusions

In this chapter, an Euler angle based collective variable have been developed to describe relative rotation of semi-rigid objects. The protein ProTx-I and its rotation in respect to a lipid POPC membrane was used to validate the CV. The RMSD data in conjunction with the Rg data suggests that in a PBMetaD simulation using Euler angles as CVs the protein explores a wider range of orientations than in a conventional MD simulation without significantly perturbing its internal conformation. This makes these set of CVs fit to be used for analysis and biasing of Protx-1 simulations, which are described in the next chapter of this thesis.

7 Chapter 7 – Application of US-WT-MetaD and PB-MetaD protocols to the interaction of ProTx-I with a POPC membrane

7.1 Introduction

As discussed in the previous chapters, sufficiently sampling the rotational motion of a protein on a membrane surface is a problem with no known trivial solution. The surface binding event and subsequent roto-translational motion of a protein on a membrane surface is a complex process that involves the desolvation of specific membrane-interacting residues and concurrent resolution of residues on other parts of the protein surface. The residues that govern membrane binding are, however, not known a priori. At the same time, the process also involves the desolvation and structural response of the membrane, with accompanying local rearrangement of lipids. The research discussed in Chapter 2 showed that a conventional umbrella sampling (US) approach is not an effective solution for this sampling problem, yielding free energy profiles that are dependent on the initial conditions of the system, thus indicating a lack of convergence. In Chapter 4, the same free energy profiles (PMF) were calculated using a hybrid US - replica exchange approach utilizing an in-house developed REST3 algorithm to enhance the configurational sampling of every umbrella window. This method proved to be effective for enhancing the sampling of the roto-translational motion of the protein at the membrane surface, but introduced artefacts in the structural ensemble of the protein that invalidated the physical meaning of the simulation. In the research described in this chapter the aim was to test an approach where the roto-translational motion of the protein on the membrane surface is treated as a semi-rigid tumbling motion of an object on a surface. To achieve this, metadynamics in conjunction with the Euler collective variables (CVs) developed and described in Chapter 6 were used to enhance the roto-translational sampling of ProTx-1 on a POPC membrane. Two different metadynamics approaches were tested. In the first approach, well-tempered metadynamics (WT-MetaD) in conjunction with US was used (referred to as US-WT-MetaD). The WT-MetaD approach involved using a two-dimensional potential with Euler angle CVs to enhance the sampling of the roto-translational motion of the protein while US was used to restrain the protein at specific distances from the membrane surface along a reaction coordinate running perpendicular to the membrane surface. The free energy was then reconstructed using the reweighting protocol validated with the CO₂ system described in Chapter 5. The rationale of this first approach was to use US to explore the centre-of-mass (COM) distance between the protein and the membrane, whilst the Euler angle CVs enhance the rotational motion of the protein on the membrane surface. This ideally would yield a more complete configurational ensemble from which a more accurate and reliable potential of mean force (PMF) can be extracted. This approach, whilst allowing for extensive sampling

of the z-coordinate around the centres of restraining harmonic potentials centre, does not allow for free translation of the protein along the z-coordinate. Because the harmonic potential causes the potential energy to increase quadratically with the distance from the centre of the umbrella, it very quickly becomes impossible for the protein to explore distances far from the umbrella centre. It is possible that this lack of translation in the z-coordinate prevents the protein from changing orientation on the membrane surface. To investigate if this affects the resulting PMF, in the second approach sampling was enhanced using three, mono-dimensional CVs. These CVs are the COM distance between the protein and the membrane and two Euler angles for the rotational motion. Parallel bias metadynamics (PB-MetaD) was used for this purpose as this method provides much faster convergence for three CVs than WT-MetaD.

7.2 Methods

7.2.1 WT-MetaD

WT-MetaD was described in detail in Chapter 5, section 2.1. Briefly, WT-MetaD is a variant of classical metadynamics in which the Gaussian hill height decreases with time as a function of the bias potential previously deposited at a given value of a CV. The Gaussian height is rescaled according to equation ⁹⁴:

$$W = \omega \tau_G e^{-\frac{V_G(s,t)}{k_B \Delta T}} \quad (44)$$

By doing so the potential deposited will converge to

$$V_G(S, t \rightarrow \infty) = -\frac{\Delta T}{T + \Delta T} F(S) + C \quad (45)$$

where C is an arbitrary constant. The bias converges to a limiting value that is dependent of the bias factor γ used, where:

$$\gamma = \frac{T + \Delta T}{\Delta T} \quad (46)$$

WT-MetaD avoids the possible overfilling of the PMF that can occur with conventional metadynamics and converges to its final value with a lower error ¹⁰³.

7.2.2 PB-MetaD

PB-MetaD was described in detail in Chapter 6, section 2.2. Briefly, PB-MetaD is a variation of the conventional metadynamics algorithm that makes uses of multiple, mono-dimensional biasing

potentials to explore, and later reconstruct, a multi-dimensional PMF ¹⁷¹. This method circumvents the problems with convergence seen in conventional metadynamics, where multiple CVs are sampled using a single, multi-dimensional biasing potential. As a consequence, it is not recommended to use more than a two-dimensional Gaussian potential in conventional metadynamics as the amount of time to reach convergence becomes prohibitive. PB-MetaD thus offers a trade-off in the accuracy of the resulting N-dimensional PMF obtained from combining one dimensional PMFs, with the possibility of handling efficiently an arbitrary number of CVs.

7.2.3 Metadynamics with walkers

To increase the efficiency of a metadynamics simulations, an algorithm making use of 'walkers' has been developed such that the simulation can be parallelized on multiple nodes ¹⁸⁶. A single, long metadynamics simulation is split into N simulations and each one of these simulations is called a walker, and all walkers share and simultaneously sample the same PMF. The free energy $F(s, t)$ is given at any moment by the sum of the Gaussian potentials deposited by all the walkers. While all walkers sample the same PMF, they do not interact in any other way, therefore minimizing the communication required between nodes. This approach is particularly effective if the starting structures for each walker can be selected to be situated at different points in the PMF. Rather than allowing the system to diffuse naturally to all the available regions of the PMF, the simulation has many starting points that can diffuse simultaneously. In our protein-membrane systems, the starting structures of the 10 walkers were spaced along the z direction.

7.2.4 Euler angles

The collective variables used to enhance the rotational motion of ProTx-1 on the membrane surface were chosen to be the Euler angles α and β , as defined in Chapter 6, section 2.1. As before, the angle γ was not used, as the membrane surface is symmetric to rotation along the z axis of the peptide.

7.2.5 US-WT-MetaD simulations

Two replicas of this simulation setup were constructed in order to assess any starting condition bias after using WT-MetaD to enhance every window. The windows from 2.4 to 4.0 nm of the conventional US simulation described in Chapter 2, section 2.3, were used as starting points. Two sets of different starting structures were chosen from the original branch simulation and run until convergence of the MetaD potential. The starting structures for run 1 were chosen from the end of the simulation ($t = 200$ ns) and the ones for run 2 were chosen from $t = 130$ ns. The windows after 4.0 nm were taken directly from the conventional US windows described in Chapter 2, section 2.3, since for COM-distances > 4.0 the protein experiences free rotation in solution and no enhanced sampling is required.

The position of the protein in each window was restrained to its reference value with a force constant $k = 500$ kJ/mol. A WT-MetaD potential was deposited over two Euler angles α and β (as defined in

Chapter 6, section 2.1). The WT-MetaD potential was deposited every 1 ps steps with a Gaussian potential height of 2.0 kJ/mol, with a width of 0.03 and a bias factor of 10. The rate of decrease in hill height over time was used as a criterion for the convergence of the bias potential, which was considered to have been reached when $h(s)$ was less than 0.05 kJ/mol for α and β , after which $h(s)$ was considered quasi-stationary. All the analyses were carried out using the part of the simulation with a quasi-stationary potential.

Consistently with how GROMOS54A8 force field was parametrized⁸⁷ electrostatic interactions were treated using a reaction-field approach with a cut-off $r_{\text{coul}} = 1.4$ nm and $\epsilon = 62$ F/m. The temperature and pressure of the system were kept constant using the Nosé-Hoover thermostat¹⁶⁵ with $\tau_T = 0.5$ ps, and a semi-isotropic Parrinello-Rahman barostat¹⁶⁶ with $\tau_p = 1.0$ ps and compressibility of 4.6×10^{-5} bar⁻¹.

The total time of simulation for every window, along with the time needed to reach convergence are reported in Table 2.

Table 2 Simulation times for every window of the US-WT-MetaD simulations

| Window centre (nm) | Total time | Time to convergence | Total time | Time to convergence |
|-----------------------|--------------------------|--------------------------|--------------------------|--------------------------|
| | US-WT-Metad run1 (ns) | US-WT-MetaD run1 (ns) | US-WT-Metad run2 (ns) | US-WT-MetaD run2 (ns) |
| 2.4 | 1165 | 1000 | 966 | 800 |
| 2.6 | 1094 | 820 | 822 | 700 |
| 2.8 | 1042 | 700 | 986 | 700 |
| 3.0 | 1004 | 750 | 867 | 750 |
| 3.2 | 1019 | 700 | 908 | 600 |
| 3.4 | 1168 | 700 | 929 | 500 |
| 3.6 | 808 | 400 | 667 | 500 |
| 3.8 | 824 | 450 | 692 | 500 |
| 4.0 | 931 | 400 | 704 | 300 |
| 4.2 | 539 | 250 | 471 | 300 |
| 4.4 | 536 | 200 | 748 | 300 |

7.2.6 PB-MetaD simulation

A PB-MetaD simulation using 10 parallel walkers was performed using three CVs: the ProTx-1 – POPC membrane COM distance along the z-axis (dz), and the Euler angles α and β . The starting structures for the 10 walkers were obtained by extracting frames with varying dz values from the conventional US simulation described in Chapter 2, section 2.3. The 10 starting structures had dz values of 2.4, 2.6, 2.8, 3.0, 3.2, 3.4, 3.8, 4.0 and 4.6 nm. For each CV, a PB-MetaD potential was deposited every 1 ps with an initial Gaussian potential height of 2.0 kJ/mol, with a width of 0.03 and a bias factor of 15. The rate of decrease in the hill height over time was used as criterion for the convergence of the bias potential, which was considered to have been reached when $h(s)$ was less than 0.01 kJ/mol for each CV, after which the potential was considered quasi-stationary. All the analyses were carried over the part of the simulation with a quasi-stationary potential.

The electrostatic interactions, temperature and pressure were treated the same way described in paragraph 2.5.

7.2.7 Reconstruction of the PMF

The PMF in the US-WT-MetaD simulations was obtained through the reweighting process described in detail in Chapter 5, section 2.3, which makes use of WHAM to reweigh the combined trajectories from all windows. The weight of every frame in the ensemble was obtained using the quasi-stationary bias of every US-WT-MetaD window on the combined trajectory. The full PMF was then calculated from the histograms of dz , reweighted with the appropriate weight for every frame, using the relation $F(s) = -k_bT \ln H(s)$. The calculation was performed in two sections because the implementation of WHAM used to obtain the weight from every frame could not handle the exponential of the bias of windows too far from each other. First, the free energy from the windows starting from 2.4 to 3.8 nm (MetaD enhanced) was calculated using the combined trajectory. Second, windows from 4.0 to 4.4 nm from the conventional US original branch described in Chapter 2, section 2.3, were concatenated and the free energy was calculated in the same way. As the PMFs obtained from WHAM are offset by an energy constant F_i ², the two sections of the PMF were aligned visually in the region that they overlap.

In the case of the PB-MetaD simulations, the PMF was obtained through simple summation of the Gaussian hills using the standard `sum_hills` tool available in the PLUMED package¹⁸⁵.

7.2.8 Rotational Autocorrelation Functions

Rotational autocorrelation functions were calculated as the autocorrelation function of the cross product of two ij and jk vectors, where j and k are atoms (or group of atoms) in the molecule, as

implemented in the GROMACS tool `gmx rotafc`. The atoms S17, C α 97 and S99 of ProTx-I were used to define the vectors **ij** and **jk**.

7.3 Results and Discussion

The aim of this part of the research was to test if the roto-translational motion of the peptide on a membrane surface could be described by a tumbling semi-rigid object. Two different approaches were tested: the first one combined US and WT-MetaD (US-WT-MetaD) and the second one used PB-MetaD. The results from the US-WT-MetaD are presented first, followed by the PB-MetaD results, and a discussion of the overall findings from these two approaches.

7.3.1 Rotational sampling in US-WT-MetaD

To assess the effect of the metadynamics potential on the rotational sampling of ProTx-1 on the membrane surface, the trajectories from the different windows were combined. The protein was considered to be in close proximity to the membrane surface at $dz < 3.0$ and, thus, all corresponding frames were extracted from the combined trajectory. As seen in Chapter 6, the Euler angle β is the most difficult one to sample, being the angle defined between the z-axis of the current frame and the reference frame z-axis. The distribution of the Euler angle β was thus calculated for all of the frames of the combined trajectory (Fig 1). As a comparison, the data from the conventional US simulations are also shown. A flat distribution of $\cos \beta$ means that every angle is sampled with equal probability. Therefore, a flat distribution of $\cos \beta$ would indicate that all possible angles are sampled and the rotational motion of the protein on the surface is uniformly sampled. **Figure 44** shows that in the conventional US the protein exhibits three different preferred orientations, with $\cos \beta$ values of -0.35, 0.4 and 0.75. Other orientations are not sampled or sampled very poorly. In contrast, in the US-WT-MetaD the protein samples all available orientations, as indicated by an almost flat distribution of $\cos \beta$ in the case of US-WT-MetaD run1 and close to a flat distribution for US-WT-MetaD run2.

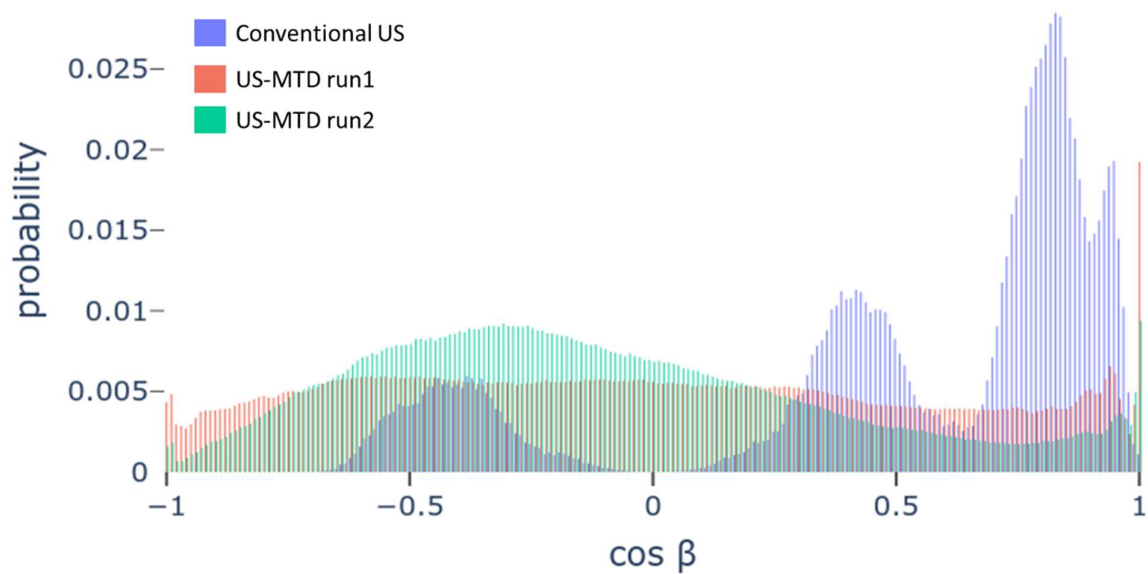


Figure 44. Distribution of the Euler angle β for all the close proximity ($dz < 3.0$ nm) trajectory frames for the US and US-WT-MetaD simulations.

As a second, independent assessment of rotational sampling, the rotational correlation autocorrelation function (RACF) was calculated for every window in the conventional US simulations and the corresponding windows in the two US-WT-MetaD simulations (**Figure 45**). It can be seen that in the classical US simulations, in most windows from $dz = 2.4$ to 3.8 nm, with the exception of windows $dz = 3.2$ and 3.6 nm, the RACF has values higher than 0.8 for the entire duration of the simulation. Such a high value and the complete lack of decay in the RACFs indicate the presence of a strong autocorrelation, such that the orientation at the end of the simulation is strongly correlated to the starting configuration of the protein in the simulation, suggesting that the protein does not diffuse rotationally at all. For windows $dz = 3.2$ and 3.6 nm there is some decay in the RACF but it takes more than 70 ns to reach a value of 0. This is much longer than for windows far from the membrane ($dz \geq 4.0$ nm), where the RACF decays to zero in approximately 5 ns. A decay time of less than 5 ns is consistent with the decorrelation time measured in a 200 ns simulation of the protein in fully hydrated in water.

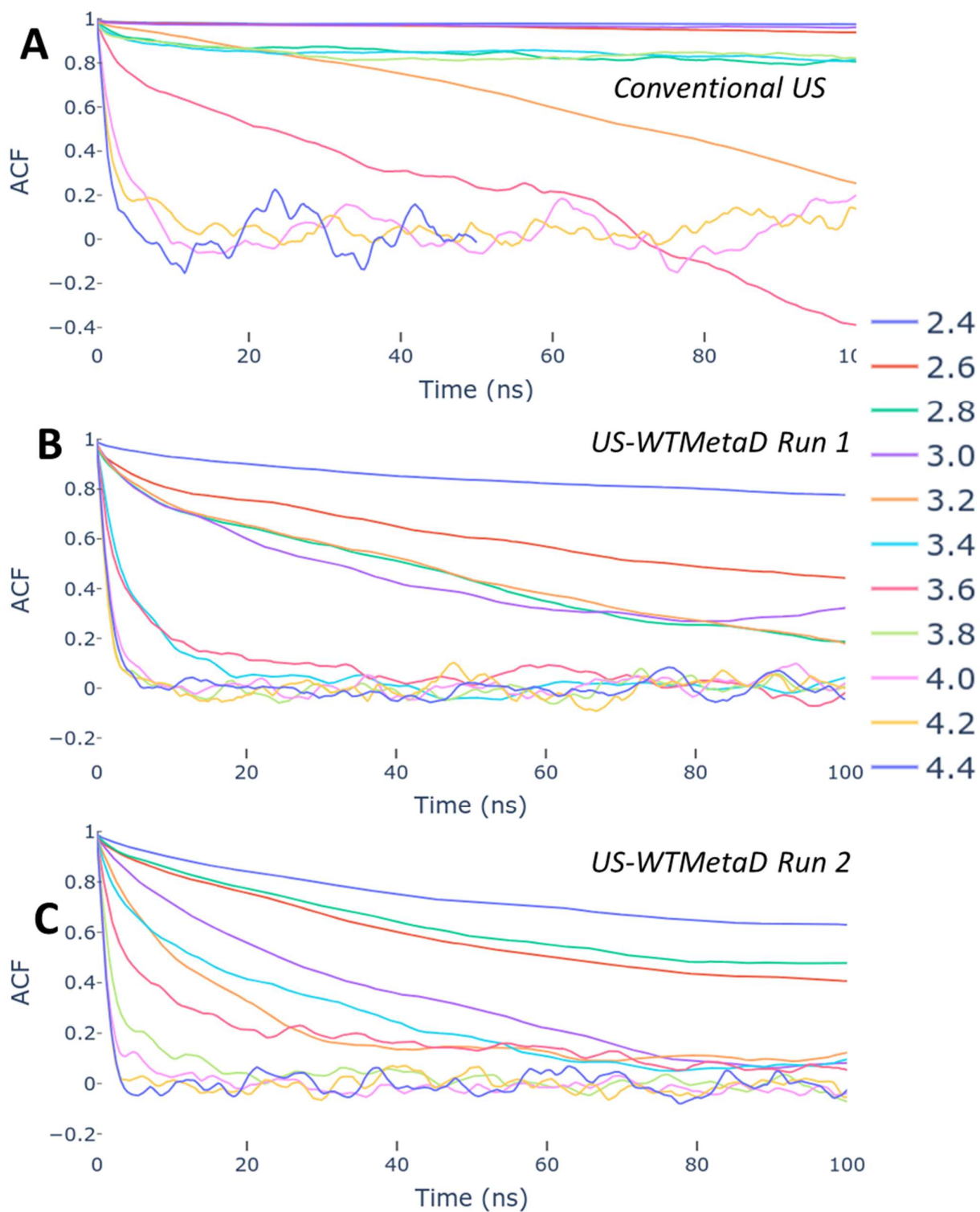


Figure 45. Rotational auto correlation functions (RACFs) for every umbrella window of the simulations. A: Conventional US, original branch described in Chapter 2, section 2.3. B: US-WT-MetaD run1. C: US-WT-MetaD run2.

In contrast to the conventional US simulation, the RACF in the US-WT-MetaD simulations indicate the presence of a higher degree of rotational freedom, consistent with the distribution of the Euler angles. Although window $dz = 2.4$ nm remains strongly correlated (RACF = 0.75 for run 1 and RACF = 0.65 for run 2), the RACF for other windows exhibits faster decay compared to their counterparts in the conventional US simulations.

For run 1, over 100 ns window $dz = 2.6$ nm decay to a value of RACF < 0.6, whereas windows with $dz > 2.6$ nm all drop to RACFs values < 0.4 and further decay with time, with the exception of window $dz = 3.0$ nm, which appears to have a slight deviation from the trend. For run 2, windows $dz = 2.6$ and 2.8 nm decay to values of RACF < 0.6 by 100 ns. Nevertheless, in both runs windows with $dz \geq 3.0$ nm become completely decorrelated or are close to (RACF ≈ 0) by 100 ns, with the exception of the aforementioned window $dz = 3.0$ in run 1, which keeps oscillating around a value of 0.4.

The combined results from the distribution of Euler angles (Figure 44) and the RACF (Figure 45) suggests that Euler angle CVs are effective at enhancing the rotational sampling of ProTx-1 close to the membrane surface.

7.3.2 PMFs from US-WT-MetaD

As discussed in Chapter 1, a simulation with improved rotational sampling is expected to yield a more accurate estimate of the PMF. **Figure 46** shows the PMFs obtained from US-WT-MetaD run 1 and US-WT-MetaD run 2 compared with the PMF obtained from the conventional US simulation. The PMFs from run 1 and run 2 agree on the overall shape, being flat far from the membrane (> 4.0 nm), gradually decreasing to a minimum (~ 3.2 nm), encountering a barrier with a maximum (~ 2.6 nm) and another minimum closer to the membrane interior (~ 2.3 nm). Beyond this minimum the protein seems to be unable to push further into the membrane, as revealed by a rapidly increasing free energy. Despite agreement on the position of the minima and the maxima of the PMF, the PMFs from run1 and run2 exhibit different relative values of the minima and the magnitude of the barrier that separate them. Both the PMFs differ in shape from the ones obtained through conventional US, as seen in Chapter 2, section 3. In the latter there is always only one minimum, which can vary in location from 2.5 to 3 nm, which correspond to the protein being in contact with the membrane surface and more or less pushing into the membrane.

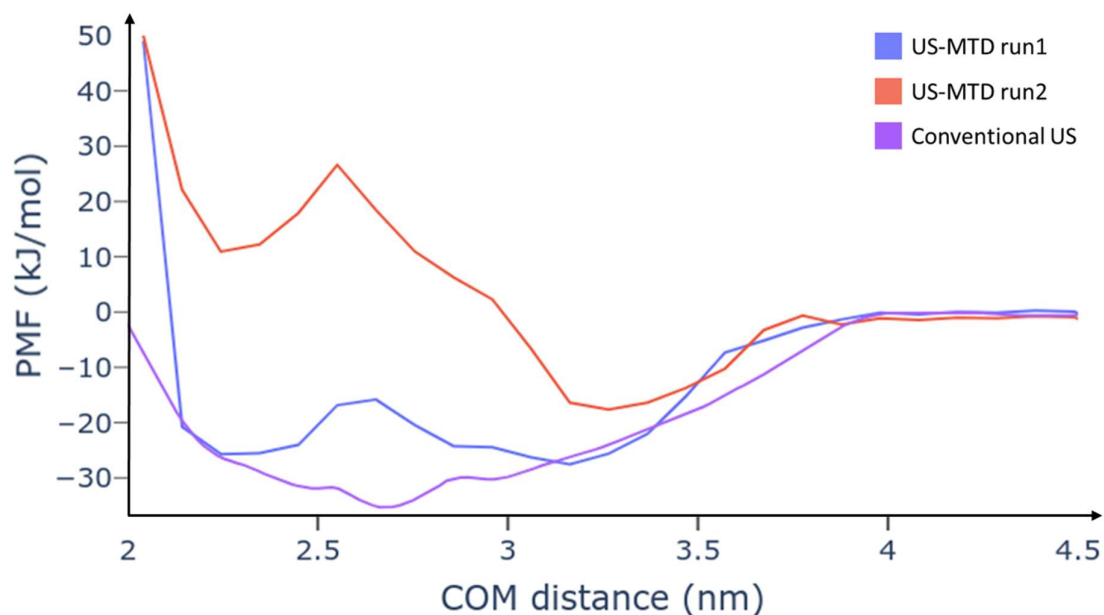


Figure 46. PMFs obtained by the two US-WT-MetaD simulations through the WHAM procedure compared with the FES obtained using conventional US described in Chapter 2, section 2.3.

The barrier seen in US-WT-MetaD PMFs around 2.6 nm has not been previously observed in any of the simulations described in previous chapters, and its origin is therefore unclear. Upon visual inspection of window $dz = 2.6$ nm of both run 1 and run 2 it is evident that the protein exhibits conformational changes that deform its structure. To quantify these structural changes in window $dz = 2.6$ nm and other windows, the change in RMSD over time was calculated for the two US-WT-MetaD runs. **Figure 47** shows plots of the RMSD as a function of time calculated using the $C\alpha$ atoms in the protein and the initial structure of the protein (at $t = 0$ ns) as a reference. It can be seen that for windows $dz = 2.4, 2.6$ and 2.8 in US-WT-MetaD run1 and windows $dz = 2.6$ and 2.8 nm in US-WT-MetaD run 2, the protein RMSD increases to values above 0.3 nm. The most significant change is shown in window $dz = 2.6$ nm, which exhibits an RMSD value of 0.6 nm in run 1 and 0.47 nm in run 2. **Figure 47C** compares the distribution of RMSDs values in the original branch of conventional US, run 1 and run 2. As it can be seen the fraction of frames with a value or RMSD > 0.3 nm is not negligible for either run 1 or run 2. This means that the protein undergoes substantial changes in structure during the simulation. Upon visual inspection this appears evident, especially around the free terminal regions not involved in the ICK motif.

Due to the above conformational changes in the protein, the distribution of Euler β angle values appears wider than it actually is. Furthermore, the meaning of the Euler α and β angles in these

conditions is partially lost, as the optimal roto-translation that superimposes the two structures as defined for the two CVs no longer superimpose two quasi-identical structures.

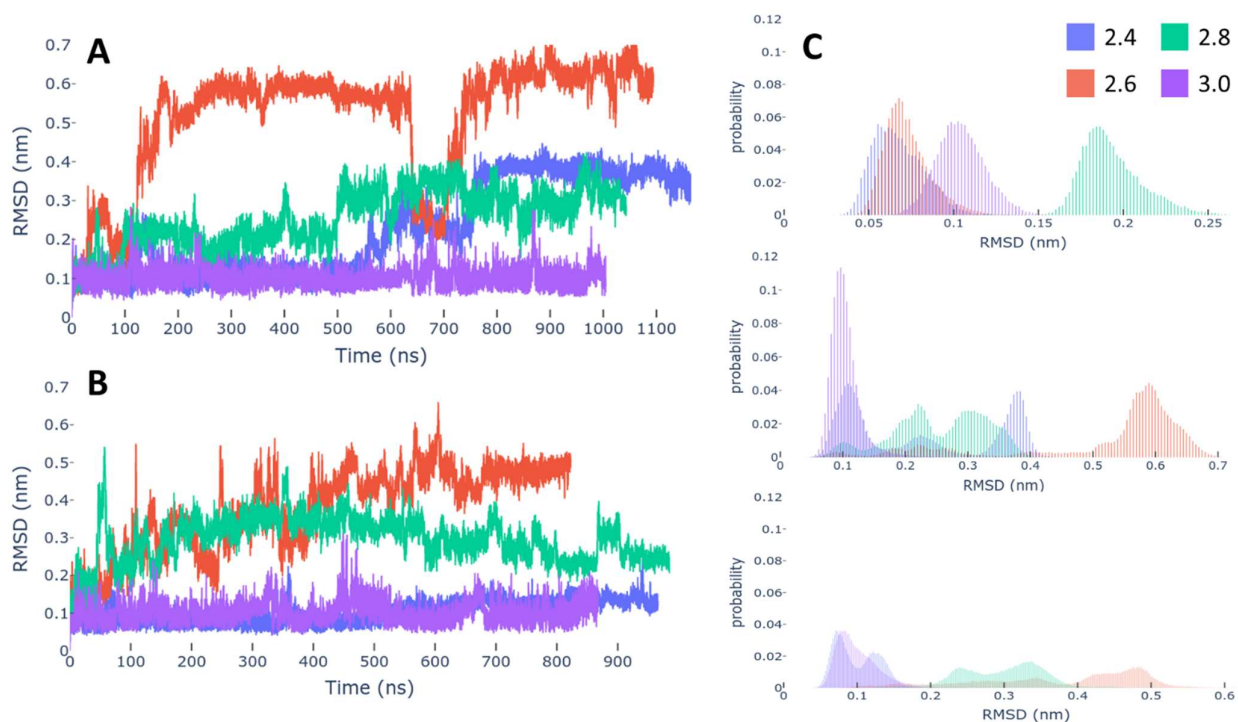


Figure 47. RMSDs for windows with $dz < 3.0$ nm. All the remaining windows had RMSD values falling below 0.2 nm, and were consequently deemed to be conformationally stable and, therefore, not reported. A: US-WT-MetaD run1. B: US-WT-MetaD run2. C: RMSD distribution in the conventional MD (top panel), US-WT-MetaD run1 (middle panel) and US-WT-MetaD run2 (bottom panel) simulations.

The reason for this behaviour, which manifests only in windows near the membrane surface, is likely to be due to the interplay between the US harmonic potential, the WT-MetaD biasing potential, and the shape of the underlying PMF. The US harmonic potential restrains the protein to specific distances from the membrane with a force constant of 500 kJ/mol. At the same time, the WT-MetaD biasing potential deposited in the Euler angle space disfavors previously visited rotational states. If the barriers for rotation of the protein at a certain distance dz from the membrane are high enough due to steric hindrance or electrostatic attractions with the membrane, it then becomes energetically favourable for the MetaD algorithm to perturb the structure of the protein itself, rather than forcing the protein to rotate.

The question therefore remains if the barrier that appears at 2.6 nm is a real barrier or is an artefact induced by the deformation of the protein because of the biasing potential deposited by WT-MetaD.

A possible solution to avoid this deformation in the protein is to use a softer US harmonic potential, allowing the protein the freedom to translate in dz , rather than enforce the rotation at every distance to the membrane. This might open up lower free energy pathways that would not have been accessible using a harmonic potential with a relatively high force constant. A PB-MetaD using Euler α, β and dz as mono dimensional CVs was conducted to assess this. If there are no or lower physical barriers on dz , the protein should be able to follow its preferred path of rotational motion across the different CVs. **Figure 48** reports the RMSDs for all of the 10 walkers in the PB-MetaD simulations.

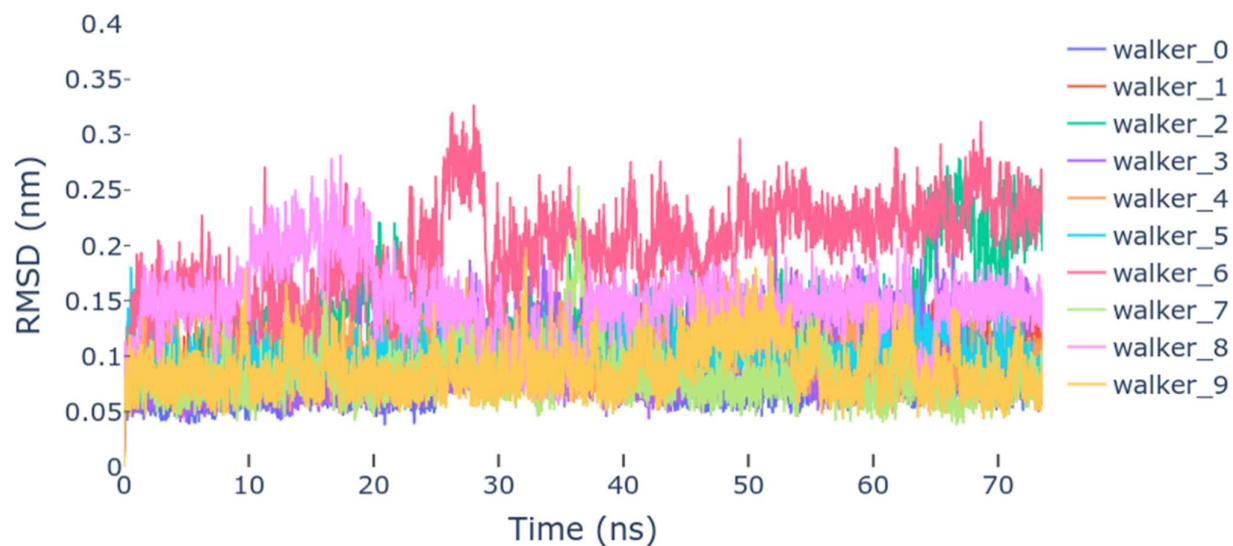


Figure 48. Time evolution of the RMSD of the protein for every walker in the PB-MetaD simulation.

It can be seen that most of the RMSD values fall under 0.2 nm, indicating that the protein exhibits a stable structure with no major changes in its overall conformation. This further suggests that forcing the system to remain at a given dz while concurrently applying a biasing potential to the rotation of the protein is likely to cause the above described deformation of the structure of the protein in the US-WT-MetaD simulations.

7.3.3 Free energy surface from PB-MetaD

To assess the convergence of the PMF, their computed values were plotted as a function of the number of Gaussians deposited (**Figure 49A**). Using blocks of 20,000 Gaussians each to calculate the PMF, we can see that during the addition of the last 100,000 Gaussians the shape of the PMF remains substantially unchanged and can be considered to be converged. **Figure 49B** is instead manually picked using bigger block size for the Gaussians, still toward the end of the simulation. As it can be seen despite the overall shape not changing substantially, over time the position of the minimum closer to the membrane interface shifts from 2.9 nm to 2.7 nm. This shift in the minimum over long times likely depends on local fluctuations in the thickness of the membrane, which leave the COM distance

between the protein and the membrane unchanged, but change the local environment that the protein experiences. These fluctuations occur over hundreds of nanoseconds (as pointed out in Chapter 1, section 3) and are, therefore, a limiting factor in the convergence of a simulation.

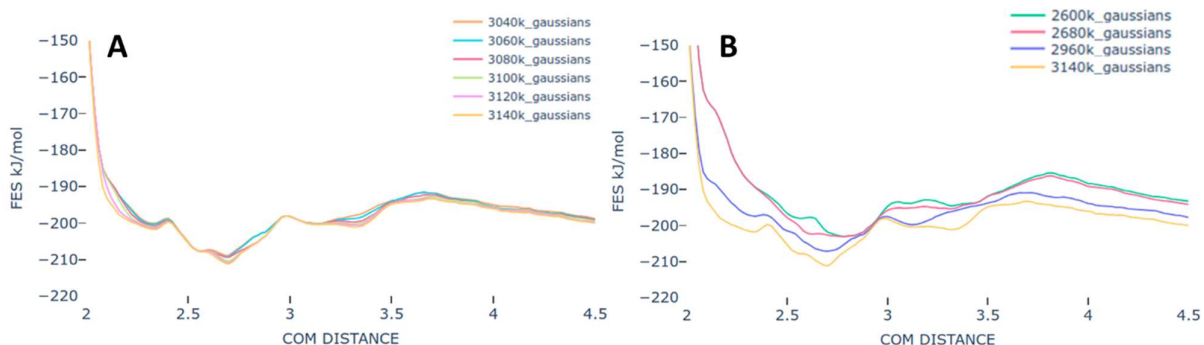


Figure 49. A: Change in the predicted FES using the last 100,000 Gaussians in increments of 20,000. B: Change in the predicted PMF during the simulation for larger strides of Gaussians.

Upon convergence of the PMF, if the selected CVs include all the slowest relaxing relevant fluctuations in the system, and the bias factor selected is appropriate, the energy barriers along the CVs should be easily overcome. This implies that the system should be able to freely diffuse from the membrane surface to the bulk and back along dz , similarly to the behaviour of a protein in pure water. In particular, every walker simulation should be capable of doing so. To assess this, the dz coordinate of the protein in every walker simulation as a function of time was plotted in **Figure 50**. As it can be seen that no walker diffuses along the entire dz space. **Figure 50B** reveals that walkers 3, 4 and 6 can diffuse far from their initial positions, whereas the rest of the walkers either fluctuate around their initial position or slowly drift away.

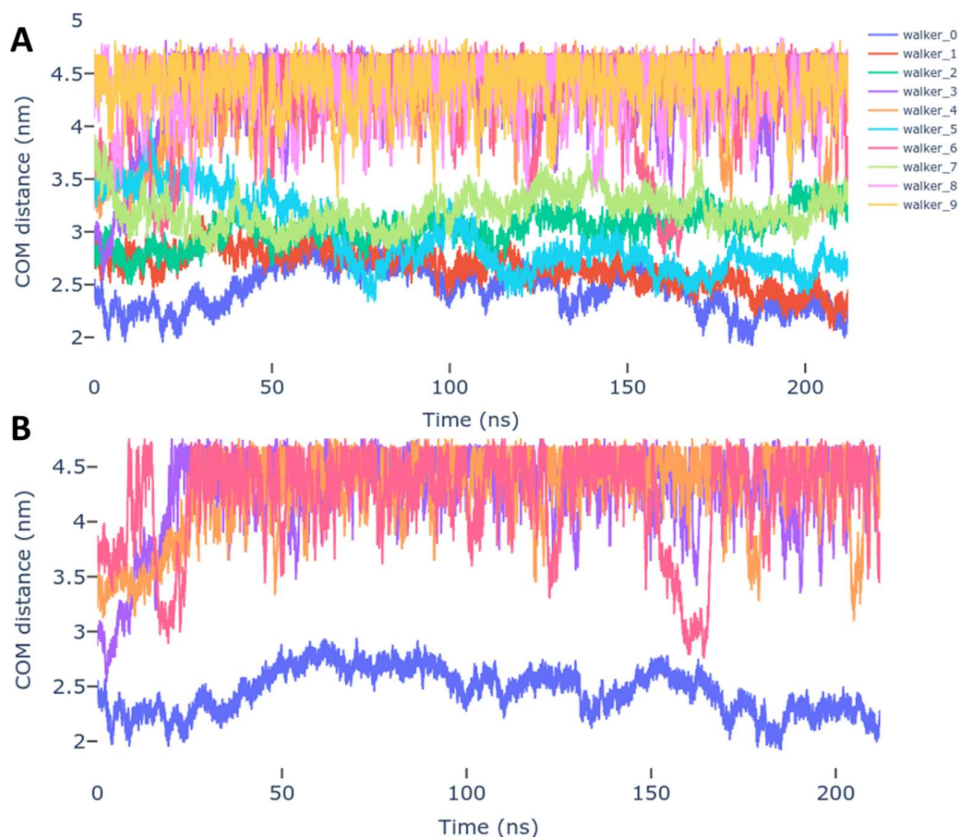


Figure 50. Time evolution of the protein-membrane COM distance. A: All the walkers B: Walkers 1, 4 and 6 have significant transitions in the COM distance, indicating some degree of translational freedom.

This signifies one of two things. Either the energy barriers on dz require a higher bias factor to be overcome, or that an important, slow-relaxing CV has not been included in the metadynamics simulation. If the latter is true, this would mean that the energy barriers are substantially higher than the current bias factor of 15 allows to explore. The effect of the bias factor γ on the energy barriers of the system is to reduce them by a factor of γ . Assuming that a barrier of 2 kT is expected to be overcome in an MD simulation, then the barrier that the system experiences on dz should be substantially higher than 2 kT after being affected by the MetaD potential. This would mean that the original height of the energy barrier would be substantially higher than $\gamma \cdot 2 \cdot kT$ (i.e. $15 \times 2 \times 2.479$ kJ/mol = 74.37 kJ/mol), which is an unlikely high order of magnitude for the energy barriers of a biological system, such as the one under investigation. The omission of an important CV is thus more likely to be the problem. In my opinion this missing CV is related to the slow membrane dynamics, which is not currently included in the MetaD. This missing CV could be, for example, local fluctuations in the APL in the proximity of the protein, or the rapid exposure of one or more phosphate groups in the vicinity of one of the protein amino acids and/or the desolvation of residues in the protein and some lipid headgroups.

To assess if the method still manages to enhance the orientational sampling of the protein, the distribution of all the values of $\cos \beta$ for frames with $dz < 3.0$ was measured and is reported in **Figure 51**.

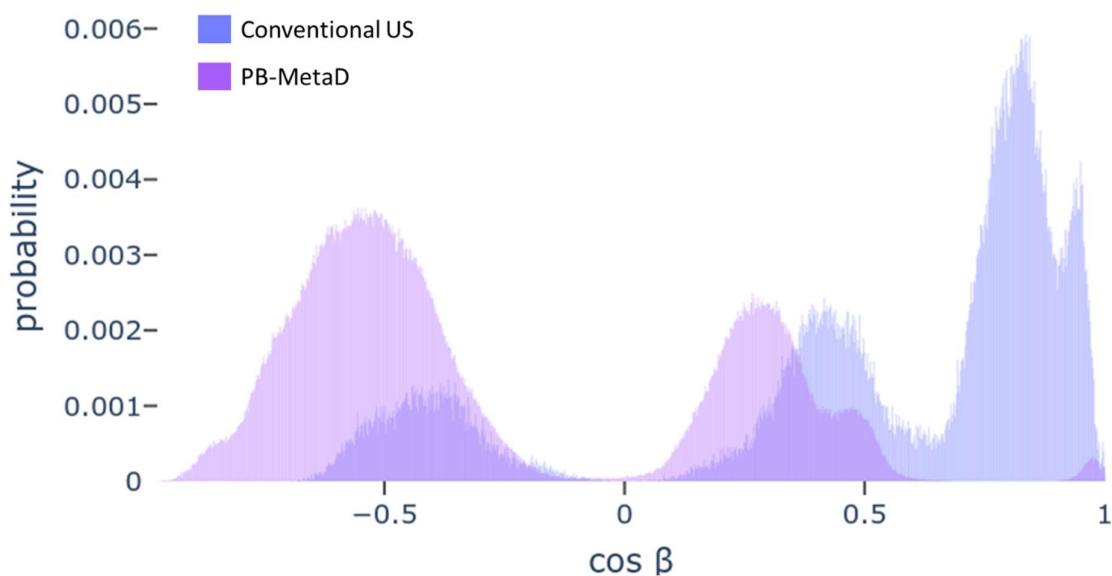


Figure 51. Distribution of Euler angle β of all the frames in close proximity to the membrane ($dz < 3.0$ nm) for the US and PB-WT-MetaD simulations.

The distribution of $\cos \beta$ appears somewhat flatter in the maximum values (0.006 for conventional US and 0.004 for PB-MetaD). Nevertheless, the enhancement of sampling is relatively minor and, despite the method not perturbing the structure of the protein, more research is required to determine a collection of CVs that allows the system to semi-freely diffuse along dz .

This approach is therefore promising because it does not seem to induce the same artefacts that were found in US-WT-MetaD, but it requires further investigation to identify the missing relevant CVs that can correctly describe the membrane dynamics.

8 Conclusions

The aim of the research described in this thesis was to develop new molecular dynamics simulation methods to obtain an estimate of the free energy of binding of small drug-like molecules and peptides to model biological membranes. To this end a number of enhanced sampling approaches were tested, including umbrella sampling (US), replica exchange with solute tempering (REST) and metadynamics (MetaD). In this chapter I summarise the achievements of this research and discuss the problems that are yet to be addressed in future work.

8.1 Summary of methods implemented and tested

8.1.1 Replica exchange with solute tempering for peptide-membrane systems

REST was originally developed as a technique to enhance the sample of dihedral angles of solvated peptides. This meant that the focus of the technique was on intramolecular interactions. The scaling of intermolecular interactions was only an indirect effect of the modification of the Hamiltonian parameters necessary to temper the intramolecular interactions. Furthermore, the technique was developed for systems with only two tempering groups: the solute and the solvent. The system component (or components) chosen as the solute would have all the intramolecular interactions tempered by the tempering factor λ , while the intermolecular interactions with the solvent would be tempered by $\sqrt{\lambda}$. The component (or components) chosen as the solvent would have instead their self-interactions left untempered. Being a stable and rigid structure ProTx-I is not expected to undergo conformational changes upon binding to a POPC membrane. The application of REST in its basic implementation is not ideal to enhance the sampling of this system, since to apply a scaling of $\sqrt{\lambda}$ to the intermolecular interactions, a scaling of λ is needed on the intramolecular interactions, which are known not to vary in a significant way.

In this research an alternative approach has been developed and implemented in a convenient and flexible way in GROMACS 4.6.7. This alternative allows the use of multiple tempering schemes for Lennard-Jones (LJ) and coulombic (coul) interactions. Different inter-molecular terms for the interaction between species i and j can be tempered each according to its own $\lambda_{\text{coul}(i,j)}$ and $\lambda_{\text{LJ}(i,j)}$. The method was validated on three terpenoid molecules interacting with a DMPC bilayer and on an Ala-Trp dipeptide (described in the PhD thesis of Lanie Ruiz-Perez) penetrating through a model of the stratum corneum. The free energies of interaction and permeation of these test systems were consistent with previously reported experimental values.

8.1.2 Metadynamics for semi-rigid molecules

There are limited examples of the use of MetaD to enhance the sampling of small molecules-membrane interactions, especially in the field of protein-membrane interactions. Even fewer studies combine US with MetaD. In the few studies that have been reported, usually the molecule under investigation is a small drug-like molecule, and the collective variables (CVs) used are simple angles between vectors defined on atom positions and the z-axis or centre-ofmass (COM) distances between the membrane and the molecule. This can, however, cause local deformations in the protein or other artefacts, as discussed in Chapter 6. These issues are further discussed in section 2.1.2 of this chapter.

In Chapter 5 of this thesis, I describe the development of a protocol to perform WMetaD-enhanced US simulations of small molecules interacting with a surface, including a reweighting protocol to extract the unbiased ensemble of any quantity of interest from a US-WMetaD trajectory. As a test system, the free energy of interaction of a CO₂ molecule with a LJ surface was predicted using this protocol. The predicted free energy of interaction matched well the theoretically predicted one, meaning that the simulation protocol is can reproduce the underlying free energy landscape.

Next, in Chapter 6, I discussed the development of a new CV based on Euler angles to describe the rotation of a semi-rigid molecule on a membrane surface that would avoid the aforementioned local deformation problems. This was applied to ProTx-I in water and on its interaction with a model POPC membrane. The use of PB-MetaD with these Euler angle CVs showed an increase in sampling of the orientational space without deformation of the protein at various COM-distances. In Chapter 7, I described the simulations conducted for the interaction of Protx-1 with a POPC bilayer using the US-WMetaD and PB-MetaD approaches developed to obtain the free energy profile of the system.

Simulations conducted with the US-WMetaD approach suggest that using a high force constant for the US potential in conjunction with a history-dependent WMetaD biasing potential can lead to structure deformation of the protein when it is in close proximity to the membrane. Nevertheless, in windows where the COM protein-membrane distance is larger than 2.8 nm, the protein retains its structure and there is enhanced sampling of its rotational motion with respect to the membrane surface. Future research into this approach could include the investigation of smaller force constants for the US potential or the use of replica exchange umbrella sampling (REUS)⁹⁹ instead of US to enhance the sampling of the COM distance CV.

Simulations conducted with the PB-MetaD approach suggest that using a soft potential can avoid the aforementioned problems of structural deformations in the protein when it is near the membrane surface. This method can be used with an arbitrary number of CVs due to the mono-dimensional nature of the Gaussian potentials applied. There is, however, a trade-off in accuracy in favour of the

number of CVs that can be biased switching from N-dimensional to mono-dimensional Gaussian functions, but this trade-off can be worthwhile, especially if all the relevant, slow-relaxing CVs associated with the process under study can be included. At present the method manages to partially enhance the orientational sampling of the protein near the surface of a membrane, but the diffusion of the system along the CV is not as free as it would be expected in a converged MetaD simulation. Future work could involve a thorough correlation analysis of membrane properties with the proximity of the protein, to find the possible hidden CVs that are currently missing in the MetaD simulations.

Furthermore, the current simulation approach could be extended under a converged static potential for a long enough amount of time such that adequate statistics for all the possible relevant configurational states of the membrane can be sampled. Once this is achieved, the ensemble from all the walker simulations could be combined and reweighted to remove the biasing potential effect so that the free energy profile can be calculated with the same protocol developed for US-WT-MetaD.

8.2 Challenges

The research described in this thesis assessed a range of enhanced sampling methods to address the problem of sampling the rotational motion of a peptide on a membrane surface. The original rationale behind this was that established enhanced sampling methods that have been tested extensively and applied to small molecules in the vacuum or in aqueous solution could be extended to larger and more complex systems, such as protein-membrane systems. As the findings discussed in the previous chapters demonstrated, this premise did not hold true to a sufficient extent.

In the following sections, I summarise the main unresolved challenges that were encountered. I will start with a general discussion on dimensionality reduction which is, in my opinion, a problem of the utmost importance. Any efforts aimed at improving the sampling of small molecule-membrane interactions should be based on a thorough understanding and awareness of this challenge. I then discuss the main challenges that were encountered with the specific enhanced sampling methods developed and used in this thesis, and I suggest some possible solutions.

8.2.1 Dimensionality reduction in membrane systems

The larger the number of non-solvent atoms (N) in a system, the more complex its free energy landscape will tend to be. Also, the larger the value of N , the more complex and multi-dimensional the paths from a state A to another state B will be on this landscape. Consequently, for $3N$ -dimensional free energy landscapes, dimensionality reduction becomes essential. In a small system one can

discount many of its degrees of freedom (DOFs). For example, a Ala-Ala dipeptide in vacuum is a routinely used test system in which the sampling of the free energy landscape can be reduced to the sampling of its dihedral angles ψ and ϕ .

The Ala-Ala dipeptide system contains only 23 atoms and reducing its dimensionality is relatively trivial. However, reducing the dimensionality of a system with several thousands of atoms is a much more challenging task. Consider, for example, a typical peptide-membrane system composed of a small protein 35 residues long and a bilayer with 128 lipids represented using a united atom force field. Discounting the solvent atoms, the membrane will have approximately 7000 atoms, interacting with approximately 375 atoms in the protein. The conformational space that these 7350 atoms can explore is extremely large, and no conformational parameter used to reduce its dimensionality can yield a full description of all the processes occurring at any given time in the system.

A possible approach to reduce this problem is to reduce the number of particles in the system using coarse graining^{70,187}. Coarse graining techniques describe groups of atoms as single beads reparametrized to match experimental data. By coarse graining you can reduce the number of particles of the system by a factor of between 3 to 8^{187} . This technique obviously reduces the dimensionality of the system and allows to run longer simulations that can be then back mapped to atomistic precision.

Nevertheless, even if using coarse graining techniques it is generally necessary to apply some kind of dimensionality reduction to assess convergence and interpret the trajectory of a complex system. A first obvious division of the system into the membrane and the protein is usually made. Some parameters are then used to describe each of the components, such as area per lipid (APL), membrane thickness or lipid tail order parameters in the case of the lipid bilayer to assess convergence to a stable state. Parameters like root mean square deviation (RMSD), radius of gyration (Rg) or secondary structure content (e.g. the percentage of β -sheet in the protein) are measures used to assess the convergence of the structure of a protein. These parameters are then also used to assess the effect of binding to the membrane surface on the structure of the membrane or the protein.

To illustrate the complex interplay that the parameters chosen to describe a system have between each other in protein-membrane systems, let us divide the system into protein and membrane and use typical parameters used to describe both of them as single components. APL and membrane thickness can be used to monitor the membrane and RMSD and Rg can be used to monitor the structure of the protein. We then want to use these parameters to measure the effect of the interaction between the membrane and the protein. This is a specific example, but the same kind of argument can be made for nearly any set of parameters chosen. All of these quantities undergo

characteristic, system-dependent fluctuations which, in the case of peptide-membrane systems, can take place over hundreds of nanoseconds. If all of these parameters were to be independent of each other as they are in a membrane in water or protein in water system, the convergence of the system can be assessed by monitoring the time-dependent evolution of these parameters. Instead all of these quantities are interdependent. For example, the APL of the membrane depends on how much the protein pushes into the membrane, which in turn depends on the instantaneous shape of the protein or the particular part of the protein that is exposed to the membrane surface. The instantaneous shape of the protein is described by the R_g and the RMSD, which again start to change when the protein interacts with the membrane. This interdependence make it difficult to separate or quantify the contribution that a change in APL has on protein structure or vice versa. Furthermore, assuming that all of these parameters will converge to some stationary value, they will do so in a system-dependent time frame, which is not guaranteed to be the same for every parameter. Assessing the overall state of convergence of a simulation then becomes strictly dependent on the set of parameters used to monitor the simulation. It is easy to underestimate these problems, which in my opinion are central to peptide-membrane simulations.

Before commencing the enhanced sampling simulation of a protein-membrane system, it is thus advisable that a substantial amount of effort is dedicated to gathering accurate and extensive statistics on the separate components of the systems, e.g. the phospholipid bilayer in water and the protein in water. For each of these separate components, multiple replicas of long, conventional MD simulations should be carried out to collect statistics on long-term fluctuations. Once this information has been gathered, a careful statistical analysis of the properties that will be monitored in the protein-membrane system should be conducted. Characterising the time-dependent fluctuations of parameters such as the diffusion coefficient or rotational autocorrelation functions of the protein in pure water, can help to rationalise any changes in these parameters obtained in simulations of protein-membrane systems.

8.2.2 Replica exchange with solute tempering simulations of membrane systems

Although the REST approach has many advantages, the scaling of the Hamiltonian in a complex system such as a protein-membrane system can easily lead to unexpected problems. Every time a new scaling regime is used, one needs to carefully test the system for the presence of any artefacts. For example, if the protein-membrane interaction is tempered down, but the protein-ion interaction is left untempered, one might expect to see abnormal behaviour of the ions with a lower interaction with

the membrane. This was indeed observed in the systems described in this thesis, where because the ions can still interact with the membrane and the protein with the same strength, the ions were observed to “bridge” the interactions between the protein and the water. These clearly anomalous behaviours can have substantial effects on the accuracy of a simulation.

In the light of the results from MetaD with Protx-1 and the associated difficulty in finding all the relevant CVs, the REST3 approach still seems the most suitable approach to enhance the sampling of the rotational motion of a protein at a membrane surface. REST3 can be used to enhance all of the degrees of freedom (DOFs) associated with the protein-membrane interactions simultaneously. Tests on Protx-1 and Gomesin ¹⁶⁸ showed that the method falls short when describing proteins with an extensive amount of self-interactions. This appears to be due to the presence of long-lived, self-interacting conformations that appear in higher tempered replicas. After “trickling” down to the ground replica, these conformations require times of the order of tens of nanoseconds to relax to a stable conformation. Initially it was assumed that the rigidity of the ICK motif in the protein would keep its secondary and tertiary structure stable if the intramolecular Hamiltonian was kept unchanged. It was not expected that the structure of the protein in the tempered replicas would undergo extensive deformations nor that the relaxation times of these deformed structures would fall in a computationally prohibitive time range for a replica exchange simulation. These long relaxation times are not practical for a replica exchange simulation, so currently REST3 is not a method that can be suggested for use with proteins. A solution to this problem might involve the inclusion of the protein-protein dihedral and non-bonded terms in the replica exchange probability between adjacent replicas, without tempering the two terms. In the current implementation of REST3 these terms cancel out by design and, therefore, do not penalize deformed protein structure exchange attempts. This allows the trickling of deformed protein structures down the ladder of replicas as long as the protein-membrane + protein-water contributions are of a similar magnitude. This effect is not significant in the case of small molecules like terpenes because the self-interaction contribution is almost inexistent, but it clearly becomes a problem with the increasing size of the solute.

Unfortunately, due to time and resource constraints it was not possible to implement and test the inclusion of these terms in the Hamiltonian exchange routine, but future work should explore this possibility to hopefully extend the applicability of the REST3 method to highly self-interacting systems.

8.2.3 Metadynamics simulations of protein-membrane systems

MetaD is a very effective and powerful method to enhance the sampling along predefined DOFs. However, the application of a biasing potential on the system requires a thorough

understanding of the dynamics of the process and of the mathematical features of MetaD. The experience gained in the research described in this thesis suggests that altering the Hamiltonian of the system, even in slight and apparently intuitive ways, can lead to unexpected results. This is the case even for small systems such as the test system composed of CO₂ and a LJ surface, described in Chapter 5. The slightest mistake in the application of a potential can lead to large deviations from the expected behaviour. For example, when applying a LJ potential to increase the rotation of the CO₂ molecule it matters to which atom the potential is applied to. If it is applied to one of the oxygen atoms, rather than the central carbon atom, it can lead to a torque acting on the molecule that forces it to adopt a specific conformation rather than rotate freely. For larger system with thousands of DOFs, it is important to remember that applying a Gaussian biasing potential on a CV means applying a force on the system through the derivative of the mathematical equation that defines the CV. This is important because the Gaussian function that is deposited on the potential energy surface only depends on the CVs used, which in turn are represented by mathematical functions defined on some atoms a_1, a_2, \dots, a_N . When the Gaussian function is deposited on the system an extra force is applied to all atoms involved in the definition of the CV. In my experience, for large systems the step where the force is applied to this subset of atoms can occur in a counterintuitive way.

In the first attempts to bias the rotation of the protein at the membrane surface, the rationale was to use as CVs the angle between the z-axis of the simulation system and a vector defined based on the coordinates of two atoms belonging to the protein. This approach not only seemed intuitive but was also extensively used in previously reported studies to define the orientation of a protein (or molecule) on a surface. A logical step is then to enhance the rotational motion of the protein on the membrane surface using a CV that is based on the angle formed by the protein vector and the z-axis. When this CV was used in MetaD simulations, rather than the protein reorienting itself, only the two atoms involved in defining the vector moved. The displacement of the two atoms defining the vector turned out to be more energetically favoured than the displacement of the large number of water molecules necessary to reorient the protein. A useful exercise when designing a simulation that modifies the Hamiltonian could then be to try to imagine all of the possible ways in which the simulation might yield unphysical results because of the mathematical definition of the CV. Once the limiting behaviour of the CV has been identified, extensive tests should be conducted to check for anomalous behaviours.

Another relevant problem is that for large systems the slowest relevant DOF is usually not known, and even if it is known conceptually it is usually difficult to design a CV to describe it effectively. The findings reported in this thesis show that in peptide-membrane systems the membrane thickness and APL fluctuation cycles are usually among the slowest DOFs in the system, with fluctuations taking over hundreds of nanoseconds. In this research, the focus was primarily on enhancing the sampling of the

rotational motion of the protein. This choice was made under the assumption that the orientation of a protein is easier to define in a CV than the complex equilibrium processes involved in the integrity of a membrane and its fluctuations. Furthermore, if one is interested in enhancing the sampling of the structural fluctuations of the membrane, there is, to the best of my knowledge, no prior studies that report use of a CV that can enhance the sampling of fluctuations in membrane thickness or APL. Nevertheless, without enhancing the sampling of these membrane fluctuations, long, continuous simulations are likely needed to attain the proper computation of the free energy landscape. This is because the effect of changes in membrane thickness and APL have to be averaged out over a statistically relevant sample of membrane fluctuations.

When performing any type of MD simulation using MetaD with a new CV or if using any non-standard reweighting technique, it is important to test the simulation conditions of the final system with a smaller system for which the theoretical solution is known or can be accurately approximated. This is analogous to what I have done with the CO₂ test system to validate the protocol that would have successively been used with Protx-1. Once the protocol can reproduce the theoretical solution, it is possible to proceed to apply it to larger, more complex systems.

Finally, when using MetaD with membranes systems, it is advisable not to fully rely on the free energy that can be computed from the summation of the Gaussian biasing potential functions. Instead, after the MetaD potential has converged, the simulation should be further extended for a few hundreds of nanoseconds using the converged potential as static. Later this same potential can be used to reweight the ensemble obtained from this extended simulation. The free energy can then be obtained from the reweighted histograms of any quantity of interest, if the simulation is long enough, the slow orthogonal DOFs can be averaged out.

9 Appendix

This thesis presents material from a published manuscript. I declare that I have obtained, where necessary, permission to use any of this material by the copyright holders.

Chapter 1: Publication 1)

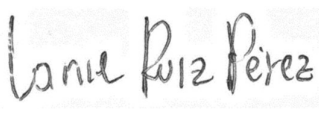
Martinotti, C., Ruiz-Perez, L., Deplazes, E. & Mancera, R. L. Molecular Dynamics Simulation of Small Molecules Interacting with Biological Membranes. ChemPhysChem **21**, 1486–1514 (2020)

I, Carlo Martinotti, author of this document declare that all the material that has been taken from this published paper has been conceptualized and written by me.

X 

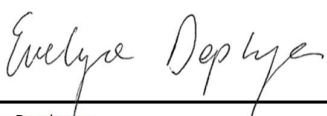
Carlo Martinotti

I, Lanie Ruiz Perez, as an author endorse that the above described sections of this paper and used by the PhD candidate in this thesis are an accurate reflection of his contribution to this paper.

X 

Lanie Ruiz Perez

I, Evelyne Deplazes, as a co-author endorse that the above described sections of this paper and used by the PhD candidate in this thesis are an accurate reflection of his contribution to this paper.

X 

Evelyne Deplazes

I, Ricardo Mancera, as a co-author endorse that the above described sections of this paper and used by the PhD candidate in this thesis are an accurate reflection of his contribution to this paper.

X 

Ricardo Mancera

JOHN WILEY AND SONS LICENSE
TERMS AND CONDITIONS

Format Print and electronic

Jun 29, 2021

This Agreement between carlo martinotti ("You") and John Wiley and Sons ("John Wiley and Sons") consists of your license details and the terms and conditions provided by John Wiley and Sons and Copyright Clearance Center.

Title

License Number 5098140048888

License date Jun 29, 2021

Licensed Content Publisher John Wiley and Sons

Licensed Content Publication ChemPhysChem

Licensed Content Title Molecular Dynamics Simulation of Small Molecules Interacting with Biological Membranes

Licensed Content Author Ricardo L. Mancera, Evelyne Deplazes, Lanie Ruiz-Perez, et al

Total

Licensed Content Date Jun 22, 2020

Licensed Content Volume 21

Licensed Content Issue 14

Licensed Content Pages 29

Type of use Dissertation/Thesis

Portion Full article

Will you be translating? No

Characterization of molecular interactions of small molecules with model cell membranes

Institution name Curtin University

Expected presentation date Jul 2021

carlo martinotti
56 andrews road, wilson

Requestor Location
Perth, 6107
Australia
Attn: carlo martinotti

Publisher Tax ID EU826007151

0.00 AUD

Terms and Conditions

TERMS AND CONDITIONS

This copyrighted material is owned by or exclusively licensed to John Wiley & Sons, Inc. or one of its group companies (each a "Wiley Company") or handled on behalf of a society with which a Wiley Company has exclusive publishing rights in relation to a particular work (collectively "WILEY"). By clicking "accept" in connection with completing this licensing transaction, you agree that the following terms and conditions apply to this transaction (along with the billing and payment terms and conditions established by the Copyright Clearance Center Inc., ("CCC's Billing and Payment terms and conditions"), at the time that you opened your RightsLink account (these are available at any time at <http://myaccount.copyright.com>).

Terms and Conditions

- The materials you have requested permission to reproduce or reuse (the "Wiley Materials") are protected by copyright.

- You are hereby granted a personal, non-exclusive, non-sub licensable (on a stand-alone basis), non-transferable, worldwide, limited license to reproduce the Wiley Materials for the purpose specified in the licensing process. This license, **and any CONTENT (PDF or image file) purchased as part of your order**, is for a one-time use only and limited to any maximum distribution number specified in the license. The first instance of republication or reuse granted by this license must be completed within two years of the date of the grant of this license (although copies prepared before the end date may be distributed thereafter). The Wiley Materials shall not be used in any other manner or for any other purpose, beyond what is granted in the license. Permission is granted subject to an appropriate acknowledgement given to the author, title of the material/book/journal and the publisher. You shall also duplicate the copyright notice that appears in the Wiley publication in your use of the Wiley Material. Permission is also granted on the understanding that nowhere in the text is a previously published source acknowledged for all or part of this Wiley Material. Any third party content is expressly excluded from this permission.
- With respect to the Wiley Materials, all rights are reserved. Except as expressly granted by the terms of the license, no part of the Wiley Materials may be copied, modified, adapted (except for minor reformatting required by the new Publication), translated, reproduced, transferred or distributed, in any form or by any means, and no derivative works may be made based on the Wiley Materials without the prior permission of the respective copyright owner. **For STM Signatory Publishers clearing permission under the terms of the [STM Permissions Guidelines](#) only, the terms of the license are extended to include subsequent editions and for editions in other languages, provided such editions are for the work as a whole in situ and does not involve the separate exploitation of the permitted figures or extracts**, You may not alter, remove or suppress in any manner any copyright, trademark or other notices displayed by the Wiley Materials. You may not license, rent, sell, loan, lease, pledge, offer as security, transfer or assign the Wiley Materials on a stand-alone basis, or any of the rights granted to you hereunder to any other person.
- The Wiley Materials and all of the intellectual property rights therein shall at all times remain the exclusive property of John Wiley & Sons Inc, the Wiley Companies, or their respective licensors, and your interest therein is only that of having possession of and the right to reproduce the Wiley Materials pursuant to Section 2 herein during the continuance of this Agreement. You agree that you own no right, title or interest in or to the Wiley Materials or any of the intellectual property rights therein. You shall have no rights hereunder other than the license as provided for above in Section 2. No right, license or interest to any trademark, trade name, service mark or other branding ("Marks") of WILEY or its licensors is granted hereunder, and you agree that you shall not assert any such right, license or interest with respect thereto
- NEITHER WILEY NOR ITS LICENSORS MAKES ANY WARRANTY OR REPRESENTATION OF ANY KIND TO YOU OR ANY THIRD PARTY, EXPRESS, IMPLIED OR STATUTORY, WITH RESPECT TO THE MATERIALS OR THE ACCURACY OF ANY INFORMATION CONTAINED IN THE MATERIALS, INCLUDING, WITHOUT LIMITATION, ANY IMPLIED WARRANTY OF MERCHANTABILITY, ACCURACY, SATISFACTORY QUALITY, FITNESS FOR A PARTICULAR PURPOSE, USABILITY, INTEGRATION OR NON-INFRINGEMENT AND ALL SUCH WARRANTIES ARE HEREBY EXCLUDED BY WILEY AND ITS LICENSORS AND WAIVED BY YOU.
- WILEY shall have the right to terminate this Agreement immediately upon breach of this Agreement by you.
- You shall indemnify, defend and hold harmless WILEY, its Licensors and their

threatened claims, demands, causes of action or proceedings arising from any breach of this Agreement by you.

- IN NO EVENT SHALL WILEY OR ITS LICENSORS BE LIABLE TO YOU OR ANY OTHER PARTY OR ANY OTHER PERSON OR ENTITY FOR ANY SPECIAL, CONSEQUENTIAL, INCIDENTAL, INDIRECT, EXEMPLARY OR PUNITIVE DAMAGES, HOWEVER CAUSED, ARISING OUT OF OR IN CONNECTION WITH THE DOWNLOADING, PROVISIONING, VIEWING OR USE OF THE MATERIALS REGARDLESS OF THE FORM OF ACTION, WHETHER FOR BREACH OF CONTRACT, BREACH OF WARRANTY, TORT, NEGLIGENCE, INFRINGEMENT OR OTHERWISE (INCLUDING, WITHOUT LIMITATION, DAMAGES BASED ON LOSS OF PROFITS, DATA, FILES, USE, BUSINESS OPPORTUNITY OR CLAIMS OF THIRD PARTIES), AND WHETHER OR NOT THE PARTY HAS BEEN ADVISED OF THE POSSIBILITY OF SUCH DAMAGES. THIS LIMITATION SHALL APPLY NOTWITHSTANDING ANY FAILURE OF ESSENTIAL PURPOSE OF ANY LIMITED REMEDY PROVIDED HEREIN.
- Should any provision of this Agreement be held by a court of competent jurisdiction to be illegal, invalid, or unenforceable, that provision shall be deemed amended to achieve as nearly as possible the same economic effect as the original provision, and the legality, validity and enforceability of the remaining provisions of this Agreement shall not be affected or impaired thereby.
- The failure of either party to enforce any term or condition of this Agreement shall not constitute a waiver of either party's right to enforce each and every term and condition of this Agreement. No breach under this agreement shall be deemed waived or excused by either party unless such waiver or consent is in writing signed by the party granting such waiver or consent. The waiver by or consent of a party to a breach of any provision of this Agreement shall not operate or be construed as a waiver of or consent to any other or subsequent breach by such other party.
- This Agreement may not be assigned (including by operation of law or otherwise) by you without WILEY's prior written consent.
- Any fee required for this permission shall be non-refundable after thirty (30) days from receipt by the CCC.
- These terms and conditions together with CCC's Billing and Payment terms and conditions (which are incorporated herein) form the entire agreement between you and WILEY concerning this licensing transaction and (in the absence of fraud) supersedes all prior agreements and representations of the parties, oral or written. This Agreement may not be amended except in writing signed by both parties. This Agreement shall be binding upon and inure to the benefit of the parties' successors, legal representatives, and authorized assigns.
- In the event of any conflict between your obligations established by these terms and conditions and those established by CCC's Billing and Payment terms and conditions, these terms and conditions shall prevail.
- WILEY expressly reserves all rights not specifically granted in the combination of (i) the license details provided by you and accepted in the course of this licensing transaction, (ii) these terms and conditions and (iii) CCC's Billing and Payment terms and conditions.
- This Agreement will be void if the Type of Use, Format, Circulation, or Requestor Type was misrepresented during the licensing process.

- This Agreement shall be governed by and construed in accordance with the laws of the State of New York, USA, without regards to such state's conflict of law rules. Any legal action, suit or proceeding arising out of or relating to these Terms and Conditions or the breach thereof shall be instituted in a court of competent jurisdiction in New York County in the State of New York in the United States of America and each party hereby consents and submits to the personal jurisdiction of such court, waives any objection to venue in such court and consents to service of process by registered or certified mail, return receipt requested, at the last known address of such party.

WILEY OPEN ACCESS TERMS AND CONDITIONS

Wiley Publishes Open Access Articles in fully Open Access Journals and in Subscription journals offering Online Open. Although most of the fully Open Access journals publish open access articles under the terms of the Creative Commons Attribution (CC BY) License only, the subscription journals and a few of the Open Access Journals offer a choice of Creative Commons Licenses. The license type is clearly identified on the article.

The Creative Commons Attribution License

The [Creative Commons Attribution License \(CC-BY\)](#) allows users to copy, distribute and transmit an article, adapt the article and make commercial use of the article. The CC-BY license permits commercial and non-

Creative Commons Attribution Non-Commercial License

The [Creative Commons Attribution Non-Commercial \(CC-BY-NC\) License](#) permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.(see below)

Creative Commons Attribution-Non-Commercial-NoDerivs License

The [Creative Commons Attribution Non-Commercial-NoDerivs License \(CC-BY-NC-ND\)](#) permits use, distribution and reproduction in any medium, provided the original work is properly cited, is not used for commercial purposes and no modifications or adaptations are made. (see below)

Use by commercial "for-profit" organizations

Use of Wiley Open Access articles for commercial, promotional, or marketing purposes requires further explicit permission from Wiley and will be subject to a fee.

Further details can be found on Wiley Online Library
<http://olabout.wiley.com/WileyCDA/Section/id-410895.html>

Other Terms and Conditions:

v1.10 Last updated September 2015

Questions? customercare@copyright.com or +1-855-239-3415 (toll free in the US) or +1-978-646-2777.

1. Kästner, J. Umbrella sampling. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **1**, 932–942 (2011).
2. Kumar, S., Rosenberg, J. M., Bouzida, D., Swendsen, R. H. & Kollman, P. A. The weighted histogram analysis method for free-energy calculations on biomolecules. I. The method. *J. Comput. Chem.* **13**, 1011–1021 (1992).
3. Hub, J. S., De Groot, B. L. & Van Der Spoel, D. G-whams-a free Weighted Histogram Analysis implementation including robust error and autocorrelation estimates. *J. Chem. Theory Comput.* **6**, 3713–3720 (2010).
4. Kästner, J. & Thiel, W. Bridging the gap between thermodynamic integration and umbrella sampling provides a novel analysis method: ‘umbrella integration’. *J. Chem. Phys.* **123**, (2005).
5. Neale, C., Bennett, W. F. D., Tieleman, D. P. & Pomès, R. Statistical convergence of equilibrium properties in simulations of molecular solutes embedded in lipid bilayers. *J. Chem. Theory Comput.* **7**, 4175–4188 (2011).
6. Neale, C. & Pomès, R. Sampling errors in free energy simulations of small molecules in lipid bilayers. *Biochim. Biophys. Acta - Biomembr.* **1858**, 2539–2548 (2016).
7. Lopes, D., Jakobtorweihen, S., Nunes, C., Sarmiento, B. & Reis, S. Shedding light on the puzzle of drug-membrane interactions: Experimental techniques and molecular dynamics simulations. *Progress in Lipid Research* **65**, 24–44 (2017).
8. Schulz, G. E. The structure of bacterial outer membrane proteins. *Biochim. Biophys. Acta - Biomembr.* **1565**, 308–317 (2002).
9. van Meer, G. & de Kroon, A. I. P. M. Lipid map of the mammalian cell. *J. Cell Sci.* **124**, 5–8 (2011).
10. Lautner, L. *et al.* Dynamic processes in biological membrane mimics revealed by quasielastic neutron scattering. *Chemistry and Physics of Lipids* **206**, 28–42 (2017).
11. Fagerberg, L., Jonasson, K., Von Heijne, G., Uhlén, M. & Berglund, L. Prediction of the human membrane proteome. *Proteomics* **10**, 1141–1149 (2010).
12. Brown, M. F. Soft Matter in Lipid–Protein Interactions. **46**, 379–410 (2017).
13. Klose, C., Surma, M. A. & Simons, K. Organellar lipidomics-background and perspectives. *Curr. Opin. Cell Biol.* **25**, 406–413 (2013).
14. Nagar, S. & Korzekwa, K. Drug Distribution. Part 1. Models to Predict Membrane Partitioning. *Pharm. Res.* **34**, 535–543 (2017).
15. Liu, X., Testa, B. & Fahr, A. Lipophilicity and Its Relationship with Passive Drug Permeation. *Pharm. Res.* **28**, 962–977 (2011).
16. Dahan, A. & Miller, J. M. The Solubility–Permeability Interplay and Its Implications in Formulation Design and Development for Poorly Soluble Drugs. *AAPS J.* **14**, 244–251 (2012).
17. Wallace, W. E. *et al.* Phospholipid lung surfactant and nanoparticle surface toxicity: Lessons from diesel soots and silicate dusts. in *Nanotechnology and Occupational Health* 23–38 (Springer Netherlands, 2006). doi:10.1007/978-1-4020-5859-2_4

18. Alkilany, A. M. & Murphy, C. J. Toxicity and cellular uptake of gold nanoparticles: what we have learned so far? *J. Nanoparticle Res.* **12**, 2313–2333 (2010).
19. Funnekotter, B., Mancera, R. L. & Bunn, E. Advances in understanding the fundamental aspects required for successful cryopreservation of Australian flora. *Vitr. Cell. Dev. Biol. - Plant* **53**, 289–298 (2017).
20. Kaczmarczyk, A. *et al.* Current Issues in Plant Cryopreservation. in *Current Frontiers in Cryobiology* (InTech, 2012). doi:10.5772/32860
21. Seelig, J. Thermodynamics of lipid-peptide interactions. *Biochim. Biophys. Acta - Biomembr.* **1666**, 40–50 (2004).
22. Andersson, J. & Köper, I. Tethered and polymer supported bilayer lipid membranes: Structure and function. *Membranes* **6**, (2016).
23. Chan, Y.-H. M. & Boxer, S. G. Model membrane systems and their applications. *Curr. Opin. Chem. Biol.* **11**, 581–7 (2007).
24. Heerklotz, H. The microcalorimetry of lipid membranes. *J. Physics-Condensed Matter* **16**, R441–R467 (2004).
25. Peetla, C., Stine, a & Labhasetwar, V. Biophysical interactions with model lipid membranes: applications in drug discovery and drug delivery. *Mol. Pharm.* **6**, 1264–1276 (2009).
26. Valerio J, Bernstorff S & Funari SS. Effect of urea and tmao on lipid bilayers. doi:10.1515/afpuc-2017-0003
27. Oka, T., Hasan, M., Islam, M. Z., Moniruzzaman, M. & Yamazaki, M. Low-pH-Induced Lamellar to Bicontinuous Primitive Cubic Phase Transition in Dioleoylphosphatidylserine/Monoolein Membranes. *Langmuir* **33**, 12487–12496 (2017).
28. Ichimori, H., Hata, T., Matsuki, H. & Kaneshina, S. Barotropic phase transitions and pressure-induced interdigitation on bilayer membranes of phospholipids with varying acyl chain lengths. *Biochim. Biophys. Acta - Biomembr.* **1414**, 165–174 (1998).
29. Almeida, C. *et al.* Membrane re-arrangements and rippled phase stabilisation by the cell penetrating peptide penetratin. *Biochim. Biophys. Acta - Biomembr.* **1858**, 2584–2591 (2016).
30. Leftin, A. & Brown, M. F. An NMR database for simulations of membrane dynamics. *Biochim. Biophys. Acta - Biomembr.* **1808**, 818–839 (2011).
31. Kot, E. F., Goncharuk, S. A., Arseniev, A. S. & Mineev, K. S. Phase Transitions in Small Isotropic Bicelles. *Langmuir* **34**, 3426–3437 (2018).
32. Pope, J. M. & Dubro, D. W. The interaction of n-alkanes and n-alcohols with lipid bilayer membranes: a 2H-NMR study. *Biochim. Biophys. Acta - Biomembr.* **858**, 243–253 (1986).
33. Ma, J., Domicевичa, L., Schnell, J. R. & Biggin, P. C. Position and orientational preferences of drug-like compounds in lipid membranes: a computational and NMR approach. *Phys. Chem. Chem. Phys.* **17**, 19766–19776 (2015).
34. Furlan, A. L. *et al.* Membrane lipids protected from oxidation by red wine tannins: A proton NMR study. *Biochimie* **107**, 82–90 (2014).
35. Scheidt, H. A. & Huster, D. The interaction of small molecules with phospholipid membranes studied by 1H NOESY NMR under magic-angle spinning. *Acta Pharmacol. Sin.* **29**, 35–49 (2008).

36. Matsumori, N. & Murata, M. 3D structures of membrane-associated small molecules as determined in isotropic bicelles. *Nat. Prod. Rep.* **27**, 1480 (2010).
37. Loura, L. M. S. & Prieto, M. FRET in Membrane Biophysics: An Overview. *Front. Physiol.* **2**, 82 (2011).
38. Chiantia, S., Ries, J. & Schwille, P. Fluorescence correlation spectroscopy in membrane structure elucidation. *Biochim. Biophys. Acta - Biomembr.* **1788**, 225–233 (2009).
39. Subczynski, W. K., Widomska, J. & Feix, J. B. Physical properties of lipid bilayers from EPR spin labeling and their influence on chemical reactions in a membrane environment. *Free Radic. Biol. Med.* **46**, 707–718 (2009).
40. Abboud, R., Charcosset, C. & Greige-Gerges, H. Biophysical methods: Complementary tools to study the influence of human steroid hormones on the liposome membrane properties. *Biochimie* (2018). doi:10.1016/j.biochi.2018.02.005
41. Pabst, G., Kučerka, N., Nieh, M. P., Rheinstädter, M. C. & Katsaras, J. Applications of neutron and X-ray scattering to the study of biologically relevant model membranes. *Chem. Phys. Lipids* **163**, 460–479 (2010).
42. Brun, A. L. P., Darwish, T. A. & James, M. Studies of Biomimetic Cellular Membranes Using Neutron Reflection. *J. Chem. Biol. Interfaces* **1**, 3–24 (2013).
43. Markiewicz, M. & Pasenkiewicz-Gierula, M. Comparative model studies of gastric toxicity of nonsteroidal anti-inflammatory drugs. *Langmuir* **27**, 6950–6961 (2011).
44. Arsov, Z., González-Ramírez, E. J., Goñi, F. M., Tristram-Nagle, S. & Nagle, J. F. Phase behavior of palmitoyl and egg sphingomyelin. *Chem. Phys. Lipids* **213**, 102–110 (2018).
45. Durand, E. *et al.* The nonlinear effect of alkyl chain length in the membrane interactions of phenolipids: Evidence by X-ray diffraction analysis. *Eur. J. Lipid Sci. Technol.* **119**, 1600397 (2017).
46. Toppozini, L. *et al.* Partitioning of ethanol into lipid membranes and its effect on fluidity and permeability as seen by X-ray and neutron scattering. *Soft Matter* **8**, 11839 (2012).
47. Nagle, J. F. Experimentally determined tilt and bending moduli of single-component lipid bilayers. *Chem. Phys. Lipids* **205**, 18–24 (2017).
48. Nagle, J. F., Jablin, M. S. & Tristram-Nagle, S. Sugar does not affect the bending and tilt moduli of simple lipid bilayers. *Chem. Phys. Lipids* **196**, 76–80 (2016).
49. Barrett, M. A. *et al.* Interaction of Aspirin (Acetylsalicylic Acid) with Lipid Membranes. *PLoS One* **7**, e34357 (2012).
50. Moreno, M. M., Garidel, P., Suwalsky, M., Howe, J. & Brandenburg, K. The membrane-activity of Ibuprofen, Diclofenac, and Naproxen: A physico-chemical study with lecithin phospholipids. *Biochim. Biophys. Acta - Biomembr.* **1788**, 1296–1303 (2009).
51. Bozelli, J. C., Hou, Y. H. & Eppard, R. M. Thermodynamics of Methyl- β -cyclodextrin-Induced Lipid Vesicle Solubilization: Effect of Lipid Headgroup and Backbone. *Langmuir* **33**, 13882–13891 (2017).
52. Contini, C., Schneemilch, M., Gaisford, S. & Quirke, N. Nanoparticle–membrane interactions. *J. Exp. Nanosci.* **13**, 62–81 (2018).
53. Agwa, A. J. *et al.* Spider peptide toxin HwTx-IV engineered to bind to lipid membranes has an

- increased inhibitory potency at human voltage-gated sodium channel hNa V 1.7. *Biochim. Biophys. Acta - Biomembr.* **1859**, 835–844 (2017).
54. Mashaghi, A. *et al.* Label-free characterization of biomembranes: from structure to dynamics. *Chem. Soc. Rev.* **43**, 887–900 (2014).
 55. Baird, C. L., Courtenay, E. S. & Myszka, D. G. Surface plasmon resonance characterization of drug/liposome interactions. *Anal. Biochem.* **310**, 93–99 (2002).
 56. Pitarke, J. M., Silkin, V. M., Chulkov, E. V & Echenique, P. M. Theory of surface plasmons and surface-plasmon polaritons. *Reports Prog. Phys.* **70**, 1 (2006).
 57. Homola, J., Yee, S. S. & Gauglitz, G. Surface plasmon resonance sensors: review. *Sensors Actuators B Chem.* **54**, 3–15 (1999).
 58. Stenberg, E., Persson, B., Roos, H. & Urbaniczky, C. Quantitative determination of surface concentration of protein with surface plasmon resonance using radiolabeled proteins. *J. Colloid Interface Sci.* **143**, 513–526 (1991).
 59. Wei, Y. & Latour, R. A. Determination of the Adsorption Free Energy for Peptide–Surface Interactions by SPR Spectroscopy. *Langmuir* **24**, 6721 (2008).
 60. Cliff, M. J., Gutierrez, A. & Ladbury, J. E. A survey of the year 2003 literature on applications of isothermal titration calorimetry. *J. Mol. Recognit.* **17**, 513–523 (2004).
 61. Marion J. Limo, Carole C. Perry, A.A. Thyparambl, Yang Wei, R. A. L. Experimental Characterization of Peptide-Surface Interaction. in *Experimental Characterization of Peptide-Surface Interaction* 1–314 (2013).
 62. Tieleman, D. P., Marrink, S. J. & Berendsen, H. J. C. A computer perspective of membranes : molecular dynamics studies of lipid bilayer systems. (1997).
 63. Ingólfsson, H. I. *et al.* Lipid organization of the plasma membrane. *J. Am. Chem. Soc.* **136**, 14554–14559 (2014).
 64. Jefferies, D. & Khalid, S. Molecular Simulations of Complex Membrane Models. in *Modeling of Microscale Transport in Biological Processes* (ed. Sid M. Becker) 1–18 (Academic Press, 2017). doi:10.1016/B978-0-12-804595-4.00001-8
 65. Ollila, O. H. S. & Pabst, G. Atomistic resolution structure and dynamics of lipid bilayers in simulations and experiments. *Biochim. Biophys. Acta - Biomembr.* **1858**, (2016).
 66. Piggot, T. J., Holdbrook, D. A. & Khalid, S. Electroporation of the E. coli and S. aureus membranes: Molecular dynamics simulations of complex bacterial membranes. *J. Phys. Chem. B* **115**, 13381–13388 (2011).
 67. Bemporad, D., Luttmann, C. & Essex, J. W. Computer Simulation of Small Molecule Permeation across a Lipid Bilayer: Dependence on Bilayer Properties and Solute Volume, Size, and Cross-Sectional Area. *Biophys. J.* **87**, 1–13 (2004).
 68. Lindahl, E. & Sansom, M. S. Membrane proteins: molecular dynamics simulations. *Curr. Opin. Struct. Biol.* **18**, 425–431 (2008).
 69. Ayton, G. S. & Voth, G. A. Systematic multiscale simulation of membrane protein systems. *Curr. Opin. Struct. Biol.* **19**, 138–144 (2009).
 70. Khalid, S. & Bond, P. J. Multiscale molecular dynamics simulations of membrane proteins. *Methods Mol. Biol.* **924**, 635–657 (2013).

71. Braun, A. R., Lacy, M. M., Ducas, V. C., Rhoades, E. & Sachs, J. N. α -Synuclein-Induced Membrane Remodeling Is Driven by Binding Affinity, Partition Depth, and Interleaflet Order Asymmetry. *J. Am. Chem. Soc.* **136**, 9962–9972 (2014).
72. Lai, C.-L., Landgraf, K. E., Voth, G. A. & Falke, J. J. Membrane Docking Geometry and Target Lipid Stoichiometry of Membrane-Bound PKC α C2 Domain: A Combined Molecular Dynamics and Experimental Study. *J. Mol. Biol.* **402**, 301–310 (2010).
73. Bucher, D., Hsu, Y.-H., Mouchlis, V. D., Dennis, E. A. & McCammon, J. A. Insertion of the Ca²⁺-independent phospholipase A₂ into a phospholipid bilayer via coarse-grained and atomistic molecular dynamics simulations. *PLoS Comput. Biol.* **9**, e1003156 (2013).
74. Ohkubo, Y. Z. & Tajkhorshid, E. Distinct Structural and Adhesive Roles of Ca²⁺ in Membrane Binding of Blood Coagulation Factors. *Structure* **16**, 72–81 (2008).
75. Vermaas, J. V. *et al.* Efficient Exploration of Membrane-Associated Phenomena at Atomic Resolution. *J. Membr. Biol.* **248**, 563–582 (2015).
76. Lai, C.-L. *et al.* Molecular Mechanism of Membrane Binding of the GRP1 PH Domain. *J. Mol. Biol.* **425**, 3073–3090 (2013).
77. Poger, D. & Mark, A. E. On the Validation of Molecular Dynamics Simulations of Saturated and *cis*-Monounsaturated Phosphatidylcholine Lipid Bilayers: A Comparison with Experiment. *J. Chem. Theory Comput.* **6**, 325–336 (2010).
78. Dickson, C. J. *et al.* Lipid14: The amber lipid force field. *J. Chem. Theory Comput.* **10**, 865–879 (2014).
79. Poger, D., Gunsteren, W. F. V. A. N., Mark, A. E. & Van Gunsteren, W. F. A New Force Field for Simulating Phosphatidylcholine Bilayers. *J. Comput. Chem.* **31**, 1117–25 (2009).
80. Klauda, J. B. *et al.* Update of the CHARMM All-Atom Additive Force Field for Lipids: Validation on Six Lipid Types. *J. Phys. Chem. B* **114**, 7830–7843 (2010).
81. Jämbeck, J. P. M. & Lyubartsev, A. P. Derivation and systematic validation of a refined all-atom force field for phosphatidylcholine lipids. *J. Phys. Chem. B* **116**, 3164–3179 (2012).
82. Paloncýová, M. *et al.* Benchmarking of force fields for molecule-membrane interactions. *J. Chem. Theory Comput.* **10**, 4143–4151 (2014).
83. Sandoval-Perez, A., Pluhackova, K. & Böckmann, R. A. Critical Comparison of Biomembrane Force Fields: Protein-Lipid Interactions at the Membrane Interface. *J. Chem. Theory Comput.* **13**, 2310–2321 (2017).
84. Poger, D., Caron, B. & Mark, A. E. Validating lipid force fields against experimental data: Progress, challenges and perspectives. *Biochim. Biophys. Acta - Biomembr.* **1858**, 1556–1565 (2016).
85. Poger, D. & Mark, A. E. On the validation of molecular dynamics simulations of saturated and *cis*-monounsaturated phosphatidylcholine lipid bilayers: A comparison with experiment. *J. Chem. Theory Comput.* **6**, 325–336 (2010).
86. Malde, A. K. *et al.* An Automated force field Topology Builder (ATB) and repository: Version 1.0. *J. Chem. Theory Comput.* **7**, 4026–4037 (2011).
87. Marzuoli, I., Margreitter, C. & Fraternali, F. Lipid Head Group Parameterization for GROMOS 54A8: A Consistent Approach with Protein Force Field Description. *J. Chem. Theory Comput.* **15**, 5175–5193 (2019).

88. Van Gunsteren, W. F. *et al.* Biomolecular modeling: Goals, problems, perspectives. *Angew. Chemie - Int. Ed.* **45**, 4064–4092 (2006).
89. Wang, L., Friesner, R. A. & Berne, B. J. Replica Exchange with Solute Scaling: A More Efficient Version of Replica Exchange with Solute Tempering (REST2). *J. Phys. Chem. B* **115**, 9431–9438 (2011).
90. Martinotti, C., Ruiz-Perez, L., Deplazes, E. & Mancera, R. L. Molecular Dynamics Simulation of Small Molecules Interacting with Biological Membranes. *ChemPhysChem* **21**, 1486–1514 (2020).
91. Earl, D. J. & Deem, M. W. Parallel tempering: Theory, applications, and new perspectives. (2005).
92. Wang, L., Friesner, R. A. & Berne, B. J. Replica exchange with solute scaling: A more efficient version of replica exchange with solute tempering (REST2). *J. Phys. Chem. B* **115**, 9431–9438 (2011).
93. Zheng, W., Andrec, M., Gallicchio, E. & Levy, R. M. Simulating replica exchange simulations of protein folding with a kinetic network model. *Proc. Natl. Acad. Sci.* **104**, 15340–15345 (2007).
94. Barducci, A., Bonomi, M. & Parrinello, M. Metadynamics. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **1**, 826–843 (2011).
95. Huang, K. & García, A. E. Acceleration of lateral equilibration in mixed lipid bilayers using replica exchange with solute tempering. *J. Chem. Theory Comput.* **10**, 4264–4272 (2014).
96. Zhou, R. Replica exchange molecular dynamics method for protein folding simulation. *Methods Mol. Biol.* **350**, 205–223 (2007).
97. Fukunishi, H., Watanabe, O. & Takada, S. On the Hamiltonian replica exchange method for efficient sampling of biomolecular systems: Application to protein structure prediction. *J. Chem. Phys.* **116**, 9058–9067 (2002).
98. Liu, P., Kim, B., Friesner, R. A. & Berne, B. J. Replica exchange with solute tempering: a method for sampling biological systems in explicit water. *Proc. Natl. Acad. Sci. U. S. A.* **102**, 13749–54 (2005).
99. Sugita, Y., Kitao, A. & Okamoto, Y. Multidimensional replica-exchange method for free-energy calculations. *J. Chem. Phys.* **113**, 6042–6051 (2000).
100. Wright, L. B., Palafox-Hernandez, J. P., Rodger, P. M., Corni, S. & Walsh, T. R. Facet selectivity in gold binding peptides: exploiting interfacial water structure. *Chem. Sci.* **6**, 5204–5214 (2015).
101. Yang, M., Huang, J. & MacKerell, A. D. Enhanced Conformational Sampling Using Replica Exchange with Concurrent Solute Scaling and Hamiltonian Biasing Realized in One Dimension. *J. Chem. Theory Comput.* **11**, 2855–2867 (2015).
102. Jiang, W., Luo, Y., Maragliano, L. & Roux, B. Calculation of free energy landscape in multi-dimensions with hamiltonian-exchange umbrella sampling on petascale supercomputer. *J. Chem. Theory Comput.* **8**, 4672–4680 (2012).
103. Laio, A. & Gervasio, F. L. Metadynamics: a method to simulate rare events and reconstruct the free energy in biophysics, chemistry and material science. *Reports Prog. Phys.* **71**, 126601 (2008).
104. Barducci, A., Bussi, G. & Parrinello, M. Well-tempered metadynamics: A smoothly converging

- and tunable free-energy method. *Phys. Rev. Lett.* **100**, 1–4 (2008).
105. Prakash, A., Fu, C. D., Bonomi, M. & Pfendner, J. Biasing Smarter, Not Harder, by Partitioning Collective Variables into Families in Parallel Bias Metadynamics. *J. Chem. Theory Comput.* **14**, 4985–4990 (2018).
 106. Knapp, O., McArthur, J. R. & Adams, D. J. Conotoxins targeting neuronal voltage-gated sodium channel subtypes: Potential analgesics? *Toxins (Basel)*. **4**, 1236–1260 (2012).
 107. Priest, B. T., Blumenthal, K. M., Smith, J. J., Warren, V. A. & Smith, M. M. ProTx-I and ProTx-II: Gating modifiers of voltage-gated sodium channels. *Toxicon* **49**, 194–201 (2007).
 108. Herzig, V. & King, G. The Cystine Knot Is Responsible for the Exceptional Stability of the Insecticidal Spider Toxin ω -Hexatoxin-Hv1a. *Toxins (Basel)*. **7**, 4366–4380 (2015).
 109. Camerino, D. C., Tricarico, D. & Desaphy, J.-F. Ion Channel Pharmacology CHANNELS. *Society* **4**, 184–198 (2007).
 110. McGivern, J. G. Advantages of voltage-gated ion channels as drug targets. *Expert Opin. Ther. Targets* **11**, 265–271 (2007).
 111. Bagal, S. K. *et al.* Ion Channels as Therapeutic Targets: A Drug Discovery Perspective. *J. Med. Chem.* **56**, 593–624 (2013).
 112. Wulff, H., Castle, N. A. & Pardo, L. A. Voltage-gated potassium channels as therapeutic targets. *Nat. Rev. Drug Discov.* **8**, 982–1001 (2009).
 113. Catterall, W. A. Voltage-gated sodium channels at 60: structure, function and pathophysiology. *J. Physiol.* **590**, 2577–2589 (2012).
 114. Waszkielewicz, A. M. *et al.* Ion channels as drug targets in central nervous system disorders. *Curr. Med. Chem.* **20**, 1241–85 (2013).
 115. Tikhonov, D. B. & Zhorov, B. S. Possible roles of exceptionally conserved residues around the selectivity filters of sodium and calcium channels. *J. Biol. Chem.* **286**, 2998–3006 (2011).
 116. Miller, C. The charybdotoxin family of K⁺ channel-blocking peptides. *Neuron* **15**, 5–10 (1995).
 117. Swartz, K. J. Tarantula toxins interacting with voltage sensors in potassium channels. *Toxicon* **49**, 213–230 (2007).
 118. Milescu, M. *et al.* Tarantula toxins interact with voltage sensors within lipid membranes. *J. Gen. Physiol.* **130**, 497–511 (2007).
 119. Bosmans, F. & Swartz, K. J. Targeting voltage sensors in sodium channels with spider toxins. *Trends Pharmacol. Sci.* **31**, 175–182 (2010).
 120. Klint, J. K. *et al.* Spider-venom peptides that target voltage-gated sodium channels: Pharmacological tools and potential therapeutic leads. *Toxicon* **60**, 478–491 (2012).
 121. Xiao, Y., Blumenthal, K., Jackson, J. O., Liang, S. & Cummins, T. R. The Tarantula Toxins ProTx-II and Huwentoxin-IV Differentially Interact with Human Nav1.7 Voltage Sensors to Inhibit Channel Activation and Inactivation. *Mol. Pharmacol.* **78**, 1124–1134 (2010).
 122. Zhu, S., Darbon, H., Dyason, K., Verdonck, F. & Tytgat, J. Evolutionary origin of inhibitor cystine knot peptides. *FASEB J.* **17**, 1765–1767 (2003).
 123. Craik, D. J., Daly, N. L. & Waite, C. The cystine knot motif in toxins and implications for drug design. *Toxicon* **39**, 43–60 (2001).

124. Shah, A. D. *et al.* Spider toxin (Hvt) gene cloned under phloem specific RSs1 and RoIC promoters provides resistance against American bollworm (*Heliothis armigera*). *Biotechnol. Lett.* **33**, 1457–1463 (2011).
125. Klint, J. K. *et al.* Seven novel modulators of the analgesic target Na_v 1.7 uncovered using a high-throughput venom-based discovery approach. *Br. J. Pharmacol.* **172**, 2445–2458 (2015).
126. Miljanich, G. P. Ziconotide: neuronal calcium channel blocker for treating severe chronic pain. *Curr. Med. Chem.* **11**, 3029–40 (2004).
127. Lee, S.-Y. & MacKinnon, R. A membrane-access mechanism of ion channel inhibition by voltage sensor toxins from spider venom. *Nature* **430**, 232–235 (2004).
128. Deplazes, E. *et al.* Membrane-binding properties of gating modifier and pore-blocking toxins: Membrane interaction is not a prerequisite for modification of channel gating. *Biochim. Biophys. Acta - Biomembr.* **1858**, 872–882 (2016).
129. Henriques, S. T. S. T. S. T. S. T. *et al.* Interaction of tarantula venom peptide ProTx-II with lipid membranes is a prerequisite for its inhibition of human voltage-gated sodium channel NaV 1.7. *J. Biol. Chem.* **291**, 17049–17065 (2016).
130. Sargent, D. F. & Schwyzer, R. Membrane lipid phase as catalyst for peptide-receptor interactions. *Proc. Natl. Acad. Sci. U. S. A.* **83**, 5774–8 (1986).
131. Langelaan, D. N. & Rainey, J. K. Membrane catalysis of peptide-receptor binding. *Biochem. Cell Biol.* **88**, 203–10 (2010).
132. Posokhov, Y. O., Gottlieb, P. A., Morales, M. J., Sachs, F. & Ladokhin, A. S. Is lipid bilayer binding a common property of inhibitor cysteine knot ion-channel blockers? *Biophys J* **93**, L20-2 (2007).
133. Catterall, W. A. *et al.* International Union of Pharmacology. XLVIII. Nomenclature and Structure-Function Relationships of Voltage-Gated Calcium Channels. *Pharmacol. Rev.* **57**, 411–425 (2005).
134. Richard E. Middleton, ‡, § *et al.* Two Tarantula Peptides Inhibit Activation of Multiple Sodium Channels†. (2002). doi:10.1021/BI026546A
135. Gui, J. *et al.* A tarantula-venom peptide antagonizes the TRPA1 nociceptor Ion channel by binding to the S1-S4 gating domain. *Curr. Biol.* **24**, 473–483 (2014).
136. Rupasinghe, D. B. *et al.* Mutational analysis of ProTx-I and the novel venom peptide Pe1b provide insight into residues responsible for selective inhibition of the analgesic drug target NaV1.7. *Biochem. Pharmacol.* 114080 (2020). doi:10.1016/j.bcp.2020.114080
137. Oparin, P. B. *et al.* Structure of purotoxin-2 from wolf spider: modular design and membrane-assisted mode of action in arachnid toxins. *Biochem. J.* **473**, 3113–3126 (2016).
138. Wolf, M. G., Jongejan, J. A., Laman, J. D. & De Leeuw, S. W. Rapid free energy calculation of peptide self-assembly by REMD umbrella sampling. *J. Phys. Chem. B* **112**, 13493–13498 (2008).
139. Berman, H. M. *et al.* The protein data bank. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **58**, 899–907 (2002).
140. Wu, Y., Tepper, H. L. & Voth, G. A. Flexible simple point-charge water model with improved liquid-state properties. *J. Chem. Phys.* **124**, (2006).

141. Pluhackova, K. *et al.* A Critical Comparison of Biomembrane Force Fields: Structure and Dynamics of Model DMPC, POPC, and POPE Bilayers. *J. Phys. Chem. B* **120**, 3888–3903 (2016).
142. Berendsen, H. J. C., Postma, J. P. M., Van Gunsteren, W. F., Dinola, A. & Haak, J. R. Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* **81**, 3684–3690 (1984).
143. Oliphant, T. E. Python for scientific computing. *Comput. Sci. Eng.* **9**, 10–20 (2007).
144. Kumar, S., Rosenberg, J. M., Bouzida, D., Swendsen, R. H. & Kollman, P. A. THE weighted histogram analysis method for free-energy calculations on biomolecules. I. The method. *J. Comput. Chem.* **13**, 1011–1021 (1992).
145. Monje-Galvan, V. & Klauda, J. B. Peripheral membrane proteins: Tying the knot between experiment and computation. *Biochim. Biophys. Acta - Biomembr.* **1858**, 1584–1593 (2016).
146. Berkowitz, M. L. & Vácha, R. Aqueous Solutions at the Interface with Phospholipid Bilayers. *Acc. Chem. Res.* **45**, 74–82 (2012).
147. Disalvo, E. A. Membrane Hydration: A Hint to a New Model for Biomembranes. in *Sub-cellular biochemistry* **71**, 1–16 (2015).
148. Shinoda, W. Permeability across lipid membranes. *Biochim. Biophys. Acta - Biomembr.* **1858**, 2254–2265 (2016).
149. Venable, R. M., Krämer, A., Pastor, R. W., Krä, A. & Pastor, R. W. Molecular Dynamics Simulations of Membrane Permeability. *Chem. Rev.* **119**, acs.chemrev.8b00486 (2019).
150. Di Meo, F. *et al.* In silico pharmacology: Drug membrane partitioning and crossing. *Pharmacological Research* **111**, 471–486 (2016).
151. Awoonor-Williams, E. & Rowley, C. N. Molecular simulation of nonfacilitated membrane permeation. *Biochim. Biophys. Acta - Biomembr.* **1858**, 1672–1687 (2016).
152. Pokhrel, N. & Maibaum, L. Free Energy Calculations of Membrane Permeation: Challenges Due to Strong Headgroup-Solute Interactions. *J. Chem. Theory Comput.* **14**, 1762–1771 (2018).
153. Bussi, G. Hamiltonian replica exchange in GROMACS: a flexible implementation. *Mol. Phys.* **112**, 379–384 (2014).
154. Terakawa, T., Kameda, T. & Takada, S. On easy implementation of a variant of the replica exchange with solute tempering in GROMACS. *J. Comput. Chem.* **32**, 1228–1234 (2011).
155. Van Der Spoel, D. *et al.* GROMACS: fast, flexible, and free. *J. Comput. Chem.* **26**, 1701–1718 (2005).
156. Sindhikara, D. J., Emerson, D. J. & Roitberg, A. E. Exchange Often and Properly in Replica Exchange Molecular Dynamics. *J. Chem. Theory Comput.* **6**, 2804–2808 (2010).
157. Poger, D., Van Gunsteren, W. F. & Mark, A. E. A new force field for simulating phosphatidylcholine bilayers. *J. Comput. Chem.* **31**, 1117–1125 (2010).
158. Berendsen, H. J. C., Postma, J. P. M., van Gunsteren, W. F. & Hermans, J. Interaction Models for Water in Relation to Protein Hydration. in 331–342 (Springer, Dordrecht, 1981). doi:10.1007/978-94-015-7658-1_21
159. Berk Hess, Henk Bekker, Herman J. C. Berendsen & Johannes G. E. M. Fraaije. LINCS: A linear constraint solver for molecular simulations. *J. Comput. Chem.* **18**, 1463–1472 (1997).

160. Tironi, I. G., Sperb, R., Smith, P. E. & Van Gunsteren, W. F. A generalized reaction field method for molecular dynamics simulations. *J. Chem. Phys.* **102**, 5451–5459 (1995).
161. Humphrey, W., Dalke, A. & Schulten, K. VMD: visual molecular dynamics. *J. Mol. Graph.* **14**, 33–38 (1996).
162. Berendsen, H. J. C., van der Spoel, D. & van Drunen, R. GROMACS: A message-passing parallel molecular dynamics implementation. *Comput. Phys. Commun.* **91**, 43–56 (1995).
163. Gowers, R. *et al.* MDAnalysis: A Python Package for the Rapid Analysis of Molecular Dynamics Simulations. in *Proceedings of the 15th Python in Science Conference* 98–105 (SciPy, 2016). doi:10.25080/majora-629e541a-00e
164. Berendsen, H. J. C., Postma, J. P. M., Van Gunsteren, W. F., DiNola, A. & Haak, J. R. Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* **81**, 3684–3690 (1984).
165. Hoover, W. G. Canonical dynamics: Equilibrium phase-space distributions. *Phys. Rev. A* **31**, 1695–1697 (1985).
166. Parrinello, M. & Rahman, A. Polymorphic transitions in single crystals: A new molecular dynamics method. *J. Appl. Phys.* **52**, 7182–7190 (1981).
167. Witzke, S. *et al.* Inclusion of terpenoid plant extracts in lipid bilayers investigated by molecular dynamics simulations. *J. Phys. Chem. B* **114**, 15825–15831 (2010).
168. Tanner, J. Assessment of a Novel Enhanced Sampling Method for Molecular Simulations of Peptide-membrane Interactions. (Curtin University, 2020).
169. Silva, P. I., Daffre, S. & Bulet, P. Isolation and characterization of gomesin, an 18-residue cysteine-rich defense peptide from the spider *Acanthoscurria gomesiana* hemocytes with sequence similarities to horseshoe crab antimicrobial peptides of the tachyplesin family. *J. Biol. Chem.* **275**, 33464–33470 (2000).
170. Mandard, N., Bulet, P., Caille, A., Daffre, S. & Vovelle, F. The solution structure of gomesin, an antimicrobial cysteine-rich peptide from the spider. *Eur. J. Biochem.* **269**, 1190–1198 (2002).
171. Pfaendtner, J. & Bonomi, M. Efficient Sampling of High-Dimensional Free-Energy Landscapes with Parallel Bias Metadynamics. *J. Chem. Theory Comput.* **11**, 5062–5067 (2015).
172. Bonomi, M. & Camilloni, C. *Biomolecular simulations*. (2019). doi:10.1007/978-1-4939-9608-7_21
173. Plimpton, S. Fast parallel algorithms for short-range molecular dynamics. *Journal of Computational Physics* **117**, 1–19 (1995).
174. Tribello, G. A., Bonomi, M., Branduardi, D., Camilloni, C. & Bussi, G. PLUMED 2: New feathers for an old bird. (2013). doi:10.1016/j.cpc.2013.09.018
175. Distributions, S., Free-energy, C. & Valleau, J. P. Valleau JP, Torrie GM. *J. Comput. Phys.* **23**, 187–199 (1977).
176. Fiorin, G., Klein, M. L. & Hémin, J. Using collective variables to drive molecular dynamics simulations. *Mol. Phys.* **111**, 3345–3362 (2013).
177. Diamond, R. On the multiple simultaneous superposition. 1279–1287 (1992).
178. Weisstein, E. W. Euler Angles. *MathWorld-A Wolfram Web Resour.*
179. Kabsch, W. A solution for the best rotation to relate two sets of vectors. *Acta Crystallogr.*

- Sect. A* **32**, 922–923 (1976).
180. Kabsch, W. A discussion of the solution for the best rotation to relate two sets of vectors. *Acta Crystallogr. Sect. A* **34**, 827–828 (1978).
 181. Cao, Z. *et al.* Bias-exchange metadynamics simulation of membrane permeation of 20 amino acids. *Int. J. Mol. Sci.* **19**, (2018).
 182. Pfafndtner, J. & Bonomi, M. Efficient Sampling of High-Dimensional Free-Energy Landscapes with Parallel Bias Metadynamics. *J. Chem. Theory Comput.* **11**, 5062–5067 (2015).
 183. Schmid, N. *et al.* Definition and testing of the GROMOS force-field versions 54A7 and 54B7. *Eur. Biophys. J.* **40**, 843–856 (2011).
 184. Bonomi, M. *et al.* PLUMED: A portable plugin for free-energy calculations with molecular dynamics. *Comput. Phys. Commun.* **180**, 1961–1972 (2009).
 185. Tribello, G. A., Bonomi, M., Branduardi, D., Camilloni, C. & Bussi, G. PLUMED 2: New feathers for an old bird. *Comput. Phys. Commun.* **185**, 604–613 (2014).
 186. Raiteri, P., Laio, A., Gervasio, F. L., Micheletti, C. & Parrinello, M. Efficient Reconstruction of Complex Free Energy Landscapes by Multiple Walkers Metadynamics †. 3533–3539 (2006). doi:10.1021/jp054359r
 187. Marrink, S. J. & Tieleman, D. P. Perspective on the Martini model. *Chem. Soc. Rev.* **42**, 6801 (2013).

