

Optimization of an *in silico* protocol to characterize membrane PAINS

Pedro R. Magalhães,[†] Pedro B. P. S. Reis,[†] Diogo Vila-Viçosa,[‡] Miguel

Machuqueiro,^{*,†} and Bruno L. Victor^{*,†}

[†]*BioISI - Biosystems & Integrative Sciences Institute, Faculty of Sciences, University of
Lisboa, Campo Grande, C8 bdg, 1749-016 Lisboa, Portugal*

[‡]*Kinetikos, 3030-199 Coimbra, Portugal*

E-mail: machuque@ciencias.ulisboa.pt; blvictor@ciencias.ulisboa.pt

Phone: +351-21-7500112; +351-21-7500000

Abstract

Membrane Pan-Assay INterference compoundS (PAINS) are a class of molecules that interact non-specifically with lipid bilayers and alter their physicochemical properties. An early identification of these compounds avoids chasing false leads and the needless waste of time and resources in drug discovery campaigns. In this work, we optimized an *in silico* protocol based on umbrella sampling (US)/MD simulations to discriminate between compounds with different membrane PAINS behavior. We showed that the method is quite sensitive to membrane thickness fluctuations, which was mitigated by changing the US-reference position to the P-atoms of the closest interacting monolayer. The computational efficiency was improved further by decreasing the number of umbrellas and adjusting their strength and position in our US scheme. The ISDM-calculated membrane permeability coefficients confirmed that resveratrol and curcumin have distinct membrane PAINS characteristics and

indicate a misclassification of nothofagin in a previous work. Overall, we have presented here a promising *in silico* protocol which can be adopted as a future reference method to identify membrane PAINS.

1 Introduction

High-throughput screening is a commonly used approach in drug discovery campaigns to identify compounds showing activity to a specific therapeutic target.¹ Depending on the used test readout, certain molecules can emerge as hits without actually interacting with the desired target. In addition to this lack of specificity, such compounds can also be promiscuous and show activity in different independent assays.²⁻⁴ Such “frequent hitters”, or commonly known as false-positives, are impossible to optimize and consequently do not lead to a successful drug development process, wasting time and resources. Therefore, the ability to identify such compounds in the early steps of drug discovery campaigns is mandatory for small-to-large pharma and biotech companies.⁴

This class of promiscuous compounds was named in 2010 by Baell *et al.*² as Pan Assay INterference compoundS and gained more attention in this past decade.⁴ PAINS comprise a large variety of compounds with different sources of diverse behavior or assay interference. Compound fluorescence events,⁵ chelation,⁶ chemical aggregation,⁷ redox activity,⁸ membrane perturbation/disruption⁹ and non-selective compounds¹⁰ are just a few examples of characteristic interference’s chemicals. Although the selectivity problems associated to PAINS have been the main focus of the scientific community, there are other categories, like phytochemicals, that, due to the compound large abundance and perceived health benefits, gained a lot of recent attention.¹¹ Phytochemicals are one of the major components of plants and have been long used in traditional medicine to treat several different health problems.¹² The molecular mechanism associated with these compounds has been conventionally interpreted or theorized by effects on receptors, biological pathways, ion channels,

and transporters.¹³ However, their broad pharmacological activity spectra make it unfeasible for these compounds to target any specific protein. Furthermore, their activity modulation of apparently unrelated proteins led different authors to pinpoint the interaction of these compounds with cell membranes as the underlying mechanism behind their promiscuity.^{9,13} These special phytochemicals are currently known as membrane PAINS since they affect membrane physicochemical properties, such as curvature, fluidity, viscosity, elasticity, and permeability.^{14,15} These membrane perturbations are more pronounced than common protein/membrane-binding phenomena and seem to have more points of contact with membrane-acting drugs, such as anesthetic, cholinergic, anti-inflammatory, adrenergic, and anti-tumor compounds.¹³

In recent years, there has been a large interest in developing computational methods to identify membrane PAINS and characterize their mode of action.^{9,16,17} The developed methodologies focused on quantifying the membrane deformations due to the presence of embedded potential membrane PAINS. Ingólfsson and co-workers⁹ explored the membrane perturbation effects of several phytochemicals in membranes and mechanosensitive membrane proteins through a combination of gramicidin-based assays (experimental) and coarse-grained Molecular Dynamics (MD) simulations. The developed protocol, which relied on computationally demanding umbrella sampling (US) simulations, worked as the basis for several subsequent methodologies.¹⁶⁻¹⁸

The new implementations of the US protocol increased the molecular detail by using atomistic MD and significantly improved the description of the membrane energy barriers.^{9,16,17} However, such methodologies are very computationally intensive, leading to limitations in their implementation, such as the use of single replicates and short MD simulations.^{16,17} In this work, we have implemented a new optimized computational protocol for the identification and characterization of membrane PAINS. Along the same lines as previous implementations, our approach uses atomistic MD simulations coupled to a US scheme to calculate the potential of mean force (PMF) energy profiles. It uses a Lennard-Jones probe

to evaluate the effects of different compounds with varying degrees of reported membrane PAINS character, namely curcumin, resveratrol, and nothofagin.^{9,16} We have tested the use of long MD simulations, multiple replicates, different US schemes, and different atoms as US reference groups.

2 Methods

2.1 System setup and MM/MD parameters

We started from a pre-equilibrated lipid bilayer system consisting of 128 1-palmitoyl-2-oleoyl-*sn*-glycero-3-phosphocholine (POPC) lipids solvated by ~ 6000 water molecules.¹⁶ This was also used as the template to build all membrane PAINS systems by adding curcumin (CUR), nothofagin (NOT), or resveratrol (RES) molecules. The compounds were evenly and randomly distributed between membrane leaflets in a 12:128 mol/mol ratio, resulting in different starting systems: pure POPC, POPC+CUR, POPC+NOT, and POPC+RES. An additional system, containing 24 CUR molecules (CUR24) was also built following the same protocol.

Molecular dynamics simulations were performed using GROMACS 2018.6^{19–21} and the GROMOS 54A7 force field.²² Topologies for the compounds were obtained using the automated topology builder (ATB),^{23–25} as previously described.^{16,18} The force field parameters used for POPC were the ones included in GROMOS 54A7.^{26,27}

Long-range electrostatic interactions were computed with the particle mesh Ewald (PME) method^{28,29} using a Fourier grid spacing of 0.12 nm and a cutoff of 0.9 nm for direct contributions. Lennard-Jones interactions were calculated using a nonbonded neighbor pair list with a cutoff of 0.9 nm, allowing the use of a cutoff scheme.³⁰ Lipid and PAINS bonds were constrained with the parallel linear constraint solver (P-LINCS),³¹ while water molecules were constrained using the SETTLE algorithm.³² The simple-point charge (SPC) water model was used.³³

The system was coupled to a temperature bath at 298.15 K using the v-rescale thermo-

stat³⁴ with a coupling constant of 0.1 ps. A semi-isotropic Parrinello–Rahman barostat^{35,36} was used in order to keep a constant pressure of 1 bar with a coupling constant of 2.0 ps and a compressibility of $4.5 \times 10^{-5} \text{ bar}^{-1}$.

The energy of each system was minimized using the steepest descent algorithm³⁷ in two steps: first, with no constraints and with a maximum step size of 0.0001 nm; and second, with all bonds constrained and a maximum step size of 0.001 nm. The tolerance was set to $0.0 \text{ kJ mol}^{-1} \text{ nm}^{-1}$ in both steps, meaning the algorithm stopped when reaching machine precision. The velocities for each system were then generated according to a Maxwell distribution at 298.15 K varying the initial seed. These system initializations were performed for 200 ps with a time step of 2 fs using the MD integrator.

Systems were pre-equilibrated in 200 ns-long unbiased MD simulations in order to assess how the presence of PAINS affected membrane bulk properties, such as the total x/y area (Figure S1 of Supporting Information). The systems were considered to be equilibrated after 100 ns. Using the final converged 100 ns of the unbiased MD simulations, we calculated the average insertion of the different PAINS compounds in the bilayer (Figure S2 of Supporting Information). The insertion relative to the near phosphate monolayer was calculated using the geometric center of each PAINS compound. The preferred insertion regions for each compound are also illustrated in the snapshots shown in the right panel of Figure S2 in Supporting Information.

2.2 Umbrella sampling

To prepare each system for the umbrella sampling simulations, we added a probe that is analogous to a benzene molecule, as described in reference 16. Different replicates were then built by replacing one of the bulk water molecules in each system at random. We ran a steered MD simulation for each system in which the probe is gradually pulled in the z coordinate across the membrane normal (with a force of $1000 \text{ kJ mol}^{-1} \text{ nm}^{-2}$ and a velocity of 1 nm/ns) while keeping the xy coordinates restrained. From these simulations,

we selected several initial conformations with the probe placed from the center of the bilayer to the bulk water every 0.1 nm. The force constant (K_f) used in these umbrellas was 1000 kJ mol⁻¹ nm⁻². For each of these initial conformations, we performed 600 ns-long umbrella sampling simulations, with the initial 300 ns being discarded for equilibration. This conservative approach was based on several structural properties, such as the local monolayer thickness and deformation, which proved harder to converge in the umbrellas near the membrane/water interface (Figure S3 of Supporting Information). In addition to the general protocol described above, some variations in the number and positions of umbrellas (as well as their corresponding K_f values) were introduced. The details of these variations are included and discussed in the results section.

2.3 Analyses and Error Calculations

The potential of mean force (PMF) profiles for each system were calculated using the weighted-histogram analysis method (WHAM)³⁸ implemented in GROMACS. Membrane permeabilities were calculated using the inhomogeneous solubility-diffusion model (ISDM),^{39,40} implemented in a software package developed by Vila-Viçosa and co-workers,⁴¹ based on the formalism described in references 42 and 43. The standard error values included in figures and tables were obtained using a modified jackknife resampling approach, as described in detail in reference 44.

The local membrane thickness was obtained by calculating the half thickness for each monolayer using all P-atoms within a radius cutoff (10 Å) in the xy plane centered on the probe, while the membrane center is calculated using all P-atoms outside of a secondary 15 Å radius, i.e. the P-atoms that are unperturbed by the probe.^{45,46} These calculations were performed using the *MembIT* tool (<https://github.com/mms-fcul/MembIT>).

Other analyses, were performed using either GROMACS or in-house tools, with all plots and figures being generated using Gnuplot,⁴⁷ PyMOL,⁴⁸ and GIMP.⁴⁹

3 Results and Discussion

3.1 Protocol Optimization

The umbrella sampling biases over the course of an entire 600 ns-long simulation of replicate 1 of a pure POPC system are shown in Figure S4A of Supporting Information. This system, which we designate as 37U^B, contains 37 umbrellas, each spaced 0.1 nm and uses the bilayer center as the reference, i.e. the umbrella at 0.0 is located in the center of the bilayer, while the umbrella at 3.6 is in the bulk water. As previously mentioned (see Methods), the umbrellas located near the membrane/water interface are the most perturbed, and thus, more difficult to converge (around 1.9 nm), while the ones located the furthest away from this interface are the least perturbed. It should be noted that discarding the initial 300 ns of these simulations does not seem to affect the proper sampling of the different umbrellas. When the equilibrated regions of all replicates are taken into account, we obtain the population histogram shown in Figure S4B of Supporting Information. Despite the observed small difficulty in the aforementioned membrane/water interface region, there is a high and consistent overlap between all neighboring umbrellas, which guarantees that a good sampling of the simulated system is achieved.

We computed the PMF profiles using the entire sampling for the 37U^B system (Figure S4B of Supporting Information) or by separating the individual replicates (Figure 1A). Although the system is composed only of pure POPC bilayer and the probe, there is substantial variability between some replicates. This illustrates the importance of using several replicates, as using only one¹⁶ could result in an underestimated or overestimated free energy. As previously remarked,⁵⁰ these differences are the consequence of the heterogeneity in the membrane, i.e. in different replicates, the probe can cross the headgroup layer of the membrane in different regions with different compactness, thus encountering different resistances. We observed a clear similarity between our PMF profile (Figure 1A) and the one obtained by Jesus *et al.*,¹⁶ in particular, the relative location of the energy maximum

and overall shape. These shared characteristics cannot be dissociated from the fact that they shared the same probe parameters and lipid force field. However, these PMF profiles differ significantly from the one published by Ingólfsson and co-workers,⁹ which has an extra energy maximum located at the bilayer center. This has been previously discussed¹⁶ and is related with the coarse-grained MARTINI force field used by the authors.

As observed in the PMF profile (Figure 1A), the main source of variability between replicates occurs at the peak of the energy barrier, which corresponds to the region where the probe encounters the most resistance, i.e. at the P-headgroup region. Depending on the sim-

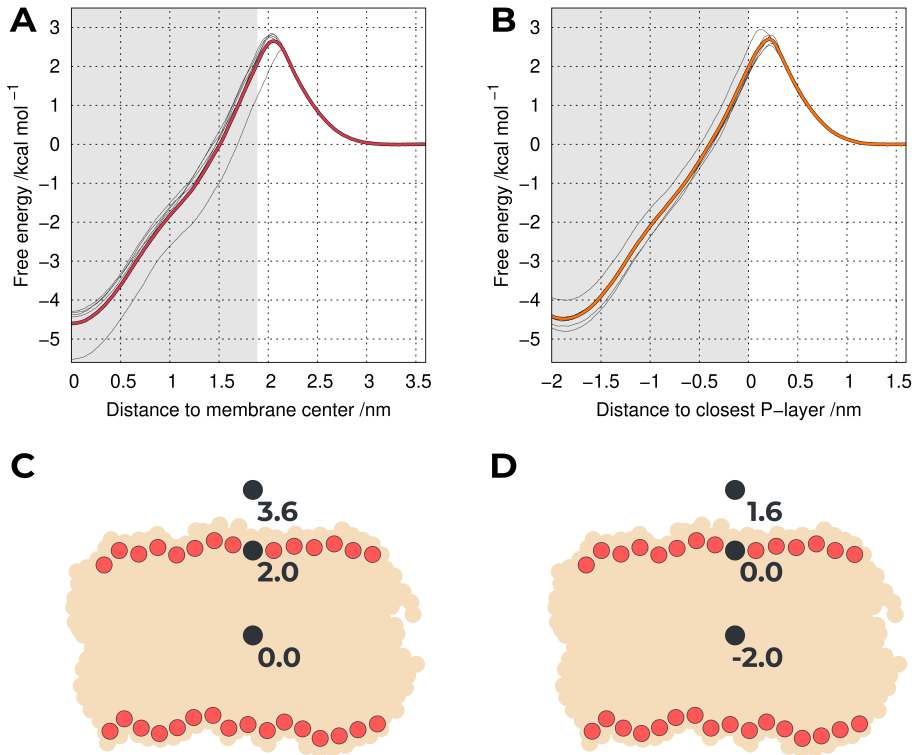


Figure 1: PMF of translocating a probe across a POPC bilayer using either the center of the membrane as the reference (A) or the closest P-layer (B). The thicker colored lines include all replicates, whereas the thinner grey lines correspond to individual replicates. The grey area is half the average bilayer thickness. The bottom panels show a cartoon of the relative positions of the probe in the z-axis using the membrane as a reference (C) or the closest P-layer (D).

ulated replicate, the probe interacts with heterogeneous P-headgroup packing environments, with different membrane thickness fluctuations, which can take hundreds of nanoseconds to

equilibrate. The observed differences are also likely related to the initial definition of the membrane center, which was calculated by the average position along the membrane normal of all P-atoms. In an attempt to reduce some of the variability between replicates, we used as a new US reference the average position of the closest monolayer P-atoms. This new reference axis becomes analogous to a measure of the probe monolayer insertion, where the new 0.0 nm umbrella corresponds to the average position of the closest P-atoms, while an umbrella at -2.0 nm corresponds to a deeper insertion, near the membrane center. This new system, which we termed $37U^M$ (Figure 1B), still does not account for all local deformations in the membrane, but most of the larger membrane thickness fluctuations are attenuated, reducing the impact of some individual outlier replicates. It is also worth noting that this change in the US reference did not impact the quality of the sampling (Figure S5 of Supporting Information). Although this is just a simple change in the reference position, it already leads to a better description of membrane perturbation by the probe. In fact, this will be particularly important to deal with increased system complexity, as in those in the presence of potential PAINS compounds.

We still observe some remaining heterogeneity present in the $37U^M$ system, which is evidenced in the peak of the energy barrier (Figure 1B). This is mainly due to local membrane deformation events, which are also difficult to equilibrate between replicates in our timescale. In an attempt to quantify this local deformation phenomenon and understand how the first coordination sphere of the P-atoms interacts with the probe, we calculated the local monolayer thickness (Figure S6 of Supporting Information). When the probe is located outside and near the P-headgroup region, its presence creates a local depression, i.e. the phosphate atoms are pushed down towards the center of the bilayer, reducing the monolayer thickness. In contrast, when the probe is right below the P-headgroup region, the opposite effect is observed, as the phosphate groups cover the probe, creating a protrusion and increasing the value of the thickness.

After significantly attenuating the membrane heterogeneity in the probe insertion ref-

erence, we need to decrease the computational load of using 5 replicates and 37 umbrellas along the membrane insertion pathway. We identified three approaches to accomplish this: reduce the number of replicates, reduce the length of simulations or reduce the number of individual umbrellas. All these approaches would lead to the desired affect, however, the first two are more likely to result in sampling issues. A decrease in the number of replicates would inevitably reduce our sensitivity in the error estimations, which could lead to serious limitations when trying to distinguish between systems containing different membrane-perturbing compounds. Reducing the length of the simulations also seems to be a precarious solution, especially since we need significantly long equilibration runs to eliminate all initial bias introduced in the probe/membrane setup (see Methods), a limitation that can be worse with the introduction of potential PAINS compounds.

We focused on the reduction in the number of individual umbrellas and, for that, we divided the probe insertion pathway into different regions according to the sampling difficulty. An estimation of this difficulty can easily be inferred from the data already presented (Figure S4 of Supporting Information) and can be roughly correlated to the proximity to the membrane/water interface. In the regions around the phosphate groups, the sampling was harder and we kept all 0.1 nm-spaced umbrellas, whereas in the easier regions away from the phosphate regions, we managed to increase this spacing distance, with a concomitant adjustment of the force constant values (K_f). After some trial-and-error, we decreased the number of umbrella windows to 22 (22U^M), representing a substantial decrease in the overall computational cost. The number and position of the umbrella windows can not be dissociated from the K_f values used and we increased or decreased these values in regions near the phosphate or away from the phosphate groups, respectively (Table S1 of Supporting Information).

Using this protocol, we managed to maintain the quality of the sampling (overlap between the neighbouring umbrellas, Figure 2A). As shown in Table S1 (Supporting Information) and Figure 2A, we increased the K_f value in the region near the P-layer (i.e. -0.4 to 0.4)

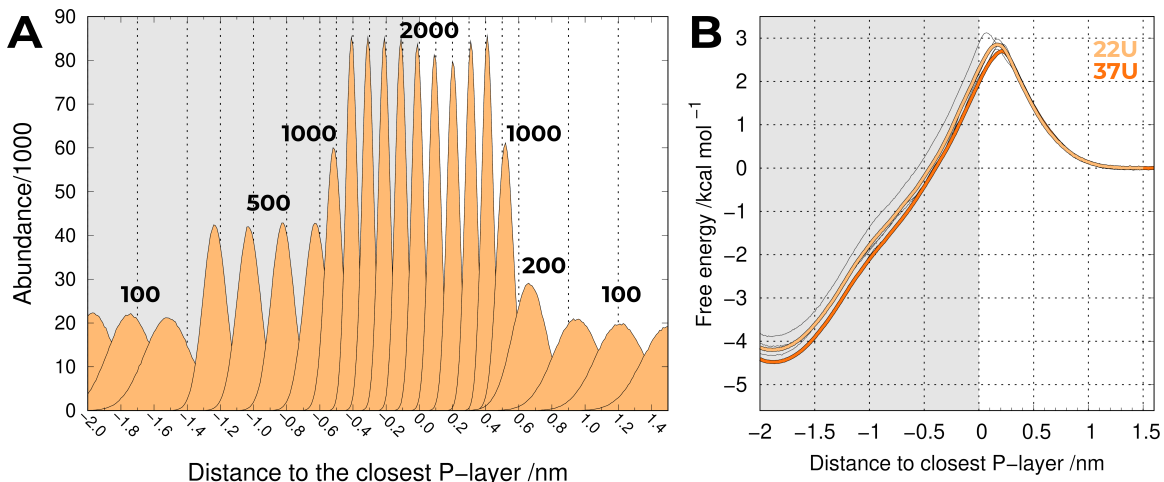


Figure 2: (A) Populations histograms for POPC 22U^M including all replicates and the corresponding K_f values (shown in $\text{kJ mol}^{-1} \text{nm}^{-2}$). (B) PMF of translocating a probe across a POPC bilayer using the closest P-layer as reference using either 22 umbrellas (yellow) or 37 umbrellas (orange). The thicker lines include all replicates, whereas the thinner ones correspond to individual replicates in the 22U^M system.

from 1000 to 2000 $\text{kJ mol}^{-1} \text{nm}^{-2}$, which was pivotal in attenuating some of the sampling heterogeneity observed in that region. As a result, we obtained more homogeneous PMF profiles (Figure 2B) that resemble very closely the one obtained using 37 umbrellas. Despite our efforts in tightening the constraints in those key umbrellas, some variability can still be observed between replicates, probably due to the inherent differences between the regions where the probe inserts in the membrane (local deformations), which seem very difficult to mitigate in our time scale. This reinforces our decision to keep using five replicates and not to reduce the simulation length. Notwithstanding, there was a significant computational gain from reducing the overall number of umbrellas from 37 to 22. Since the PMF profiles resulting from the three different US protocols used in this work (Figures 1 and 2) are very similar, we will focus on 22U^M and take advantage of its reduced computational cost.

The structure of the PMF profiles carries information on the physical state of the membrane and is usually very well-correlated with several other membrane properties.^{16,43,51} Nevertheless, it is not obvious how to assign deviation in specific parts of the profile with the membrane stability and possible deformation due to the presence of different membrane

PAINS. Previously, this has been done by looking at the entry/exit energy barriers (size and position)¹⁶ and at the eventual barrier at the center of the membrane, when present.⁹ However, the PMF profile also conveys information on the probe diffusion across the membrane, the resistance encountered, and, ultimately, a membrane permeability coefficient.^{39–43} This coefficient has many important membrane properties convoluted in one value and can be particularly advantageous when comparing between membrane systems with different compounds embedded. The permeability coefficient values are calculated using the ISDM method,^{39,40} as previously described.^{41–43} Although ISDM has been successfully used to estimate the membrane permeability to hydrophobic compounds,^{41,43} the method is highly sensitive to the convergence and overall sampling quality of the PMF profiles. However, in our system setup, since we use a hydrophobic sphere as a probe instead of an explicit hydrophobic molecule, the lack of rotational entropy might result in a better convergence.

We calculated the permeability coefficients for the three US protocols used in this work (37U^B, 37U^M and 22U^M) and observed only small differences between them, all within the error margin (Table 1). The permeability coefficient values seem to be correlated with

Table 1: Membrane permeabilities coefficients (cm s⁻¹) and energy barriers of entering and exiting the POPC bilayer (kcal mol⁻¹). The entry barrier (ΔG_{entry}) is the energy difference between the maximum value (at the membrane/water interface) and bulk water, while the exit barrier (ΔG_{exit}) is the difference between the global minimum at the membrane center and previous maximum. Errors were calculated using a jackknife approach.

System	Permeability	ΔG_{entry}	ΔG_{exit}
37U ^B	4.6 ± 0.8	2.6 ± 0.1	7.2 ± 0.1
37U ^M	4.3 ± 0.5	2.7 ± 0.1	7.2 ± 0.1
22U ^M	4.4 ± 0.4	2.9 ± 0.0	7.1 ± 0.1

the entry and exit energy barriers, which are also very similar between the three systems (Table 1). The quality of the PMF profiles obtained for the 22U^M system is also expressed in the smaller error observed for this setup. Using this protocol, we increased the complexity of our systems by adding compounds with different degrees of PAINS-like behavior, as described in the literature.^{9,16}

3.2 Protocol Application to Identify PAINS Compounds

In the previous section, we showed that the 22U^M approach was the least computationally-demanding option while still retaining the sampling quality and a high degree of homogeneity in the harder-to-sample regions of the membrane. The well-converged PMF profiles obtained allowed us to obtain permeability coefficient values with relatively small errors. We applied this protocol setup to several compounds, known to have different degrees of PAINS-like behavior, as described in the literature:^{9,16} resveratrol (RES), as a mild membrane PAINS, nothofagin (NOT) as a non-PAINS, and curcumin (CUR) as a strong membrane PAINS. The different complex systems were built by embedding each of the three PAINS compounds in the lipid bilayer in a 12:128 mol/mol ratio, evenly distributed between the two leaflets, and then pre-equilibrated using a relatively short unbiased MD simulation, as detailed in the methodology section. An additional system was also prepared using a 24:128 mol/mol ratio for curcumin (termed CUR24), as discussed below. All compounds had an impact on the final PMF profiles obtained when compared to the pure POPC membrane (Figure 3). In all cases, we observe significantly more variability between individual replicates than what was previously observed in the pure POPC systems (Figure 2B). The source for this variability is also the local membrane deformation triggered by the probe insertion in the water/membrane interface. However, the overall heterogeneity is now also impacted by the presence of the compounds, which are not uniformly distributed (Figure S2 of Supporting Information).

We observe that all compounds change the PMF profile, however, the differences between compounds are not easily clarified by visual inspection. Furthermore, it is difficult to evaluate how the specific PMF differences affect the membrane properties. To tackle this issue, we calculated the membrane permeability coefficients (Table 2), which provide a quantification of the membrane perturbation and overall stability. From the permeability coefficients, resveratrol showed the lowest perturbation to the pure POPC membrane (4.5 ± 0.3 vs 4.4 ± 0.4), which is in agreement with the previously reported description of resveratrol as a mild membrane PAINS.⁹ On the other hand, nothofagin has a more noticeable impact

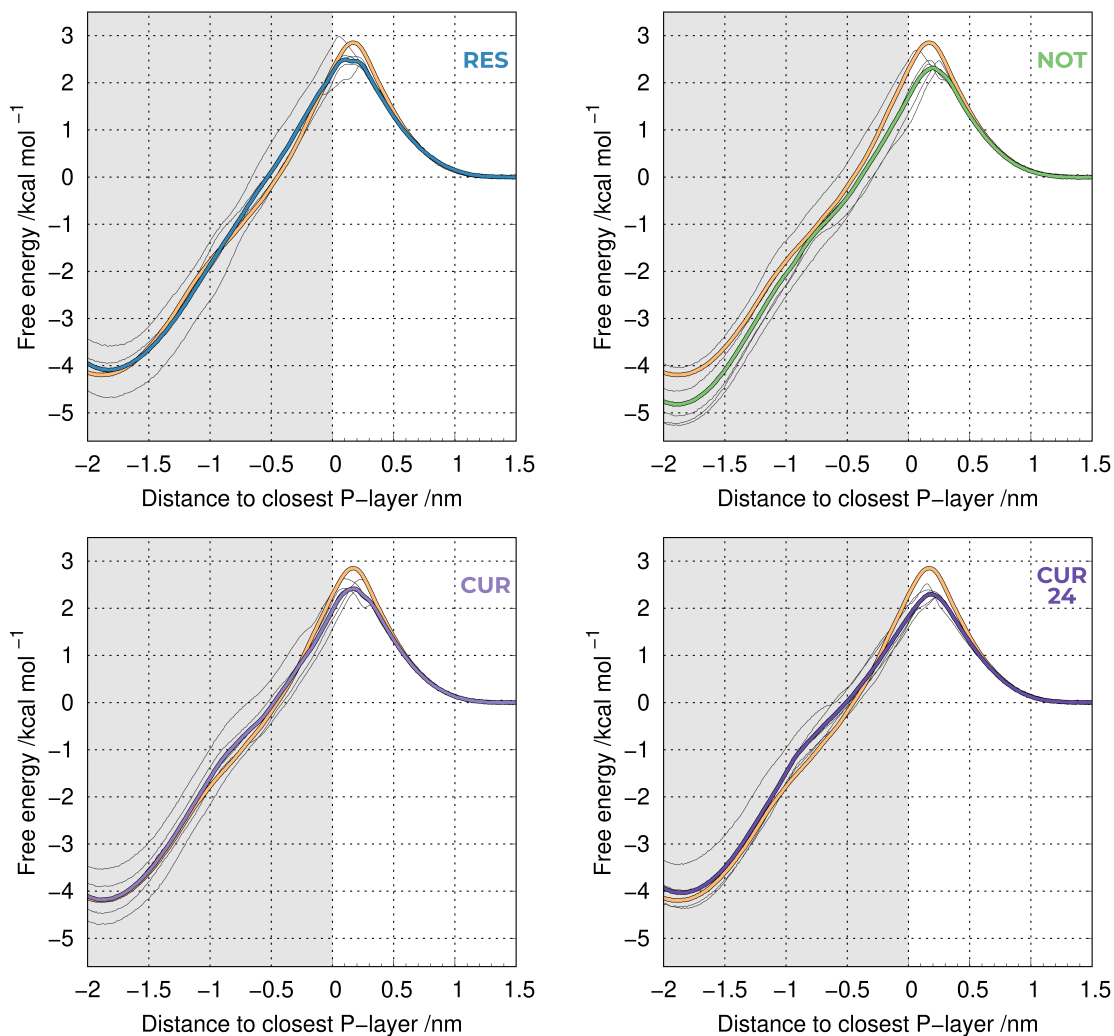


Figure 3: PMF profiles of translocating a probe across a POPC bilayer in the absence or presence of the tested compounds: resveratrol (RES), nothofagin (NOT), curcumin 10% (CUR) and curcumin 20% (CUR24). The thicker lines include all replicates (average), whereas the thinner ones correspond to individual replicates of the system with PAIRS compounds. The orange profile corresponds to the compound-free control.

Table 2: Energy barriers of entry (ΔG_{entry}) and exiting (ΔG_{exit}) the POPC bilayer, and membrane permeabilities. Energy values are shown in kcal mol⁻¹ and permeabilities are shown in cm s⁻¹. Errors were calculated using a leave-one-out (jackknife) approach.

System	Permeability	ΔG_{entry}	ΔG_{exit}
POPC	4.4 \pm 0.4	2.9 \pm 0.1	7.1 \pm 0.1
RES	4.5 \pm 0.3	2.5 \pm 0.1	6.6 \pm 0.1
NOT	5.6 \pm 0.6	2.3 \pm 0.0	7.1 \pm 0.2
CUR	5.0 \pm 0.5	2.4 \pm 0.0	6.6 \pm 0.2
CUR24	5.7 \pm 0.3	2.3 \pm 0.0	6.3 \pm 0.1

on the PMF (Figure 3), leading to a significant difference in the calculated permeability coefficient (5.6 \pm 0.6). This is in disagreement with previous observations that nothofagin is not a membrane PAINS.¹⁶ This discrepancy is most likely related to the lack of sampling (single short-length US simulation) in the work by the Jesus and co-workers.¹⁶ As we found in the previous section, both the length and the number of replicates for these simulations are key factors to achieve a good convergence. From our data, it is clear that nothofagin can, in fact, exhibit some membrane PAINS behavior.

We have also used curcumin in our study, which has been identified as a strong membrane PAINS.⁹ We did observe noticeable deviations in the overall PMF profile (Figure 3), leading to an apparent increase in the membrane permeability coefficient value (Table 2). However, the difference of its permeability coefficient, when compared with pure POPC, is within the error values (5.0 \pm 0.5 vs 4.4 \pm 0.4), which are significant due to the system heterogeneity among individual replicates. Since curcumin has been branded as a strong membrane PAINS,⁹ we expected a more significant membrane perturbation effect. This prompted us to design an extra system (CUR24) with twice the concentration of the compound (24:128 mol/mol ratio). The effect of doubling the concentration of curcumin appeared to be twofold: first, an increase in the perturbation effect when compared to pure POPC, as evidenced by the shape of the PMF profile (Figure 3) and corresponding permeability coefficient (5.7 \pm 0.3 vs 4.4 \pm 0.4); and second, a reduction in the heterogeneity between individual replicates (Figure 3) which is reflected in the smaller error value when

compared to several other systems. In overall, it seems that the addition of more curcumin molecules to the lipid bilayer does lead to a statistically significant perturbation effect, hence confirming its membrane PAINS character.

To complement the comparison between the different compounds, we also calculated the entry/exit barriers, similarly to the pure POPC systems using different methodologies (Table 1). From the data in Table 2, we calculated the correlation between the permeability coefficients and either the ΔG_{entry} or the ΔG_{exit} values. Similarly to the values for pure POPC (Table 1), we observed a strong negative correlation (-0.83) for ΔG_{entry} , and a significantly smaller value for ΔG_{exit} (-0.28). With such a high anti-correlation between the membrane permeability values and the ΔG_{entry} , we could argue that the estimation of this energy barrier will suffice to quickly gauge the membrane PAINS-like potential of a compound. It should be noted that the calculations required to estimate the entry barrier are only a modest fraction of the total needed for calculating the complete energy profile.

4 Conclusion

Membrane PAINS are promiscuous compounds that can alter membrane physicochemical properties and perturb the function of transmembrane mechanosensitive proteins. To avoid the waste of time and resources in drug discovery companies due to this class of compounds, we have devised a computational protocol based on umbrella-sampling MD simulations to discriminate between compounds with differentiated membrane PAINS behaviour. By coupling a molecular probe to this robust atomistic sampling scheme, we concluded that this method strongly depends on the specific environment sensed by the probe, especially at the membrane/water interface. To mitigate the impact of this heterogeneity, we changed the US reference position from the membrane center to the closest interacting monolayer P-atoms. This approach resulted in a higher homogeneity between replicates, in particular, the description of the energetic barrier at the water/membrane interface. Additionally, in order

to decrease the computational cost of such demanding simulations while retaining a high accuracy, we evaluated a reduction in the number of replicates, umbrellas, and simulation length. The final optimized scheme focused mainly on a reduced number of umbrellas, which significantly decreased the computational time spent without compromising the overall accuracy. This final optimized protocol was then applied to membrane systems in the presence of three compounds with different reported membrane PAINS behaviors: curcumin (membrane PAINS), resveratrol (mild membrane PAINS), and nothofagin (non-membrane PAINS). The membrane permeability coefficients calculated using the in-homogeneous solubility-diffusion model for the different systems confirmed that resveratrol has mild membrane PAINS characteristics and that curcumin exhibits a concentration-dependent membrane PAINS behaviour. However, our results indicate that nothofagin, which was previously identified as a non-membrane PAINS compound,¹⁶ presents a significant membrane perturbation effect, suggesting a misclassification of this compound. Interestingly, we have observed a very high anti-correlation between the probe entry energy barrier (ΔG_{entry}) and the membrane PAINS character. This can be very useful since this energy barrier can be calculated using a fraction of the computational cost of our protocol. Overall, our effective approach emerges as a reference *in silico* method to identify and discriminate among membrane PAINS compounds.

Acknowledgments

We thank to Tomás Silva for helping with the calculations using the *MemBIT* tool. We acknowledge financial support from Fundação para a Ciência e a Tecnologia through grants SFRH/BD/136226/2018 and CEECIND/02300/2017 and projects PTDC/BIA-BFS/28419/2017, UIDB/04046/2020 and UIDP/04046/2020.

Supporting Information Available

Time series of membrane total area, monolayer local thickness, and probe positions in the US scheme. Membrane insertion distributions of all compounds studied. Histogram distributions of the probe positions (pullx) for the different schemes. Local thickness per umbrella from the 37U^M system. Table with the positions and force constants used in the 22U^M system.

References

- (1) Macarron, R.; Banks, M. N.; Bojanic, D.; Burns, D. J.; Cirovic, D. A.; Garyantes, T.; Green, D. V. S.; Hertzberg, R. P.; Janzen, W. P.; Paslay, J. W., et al. Impact of high-throughput screening in biomedical research. *Nat. Rev. Drug Discov.* **2011**, *10*, 188–195.
- (2) Baell, J. B.; Holloway, G. A. New substructure filters for removal of pan assay interference compounds (PAINS) from screening libraries and for their exclusion in bioassays. *J. Med. Chem.* **2010**, *53*, 2719–2740.
- (3) Baell, J. Feeling nature’s PAINS: natural products, natural product drugs, and pan assay interference compounds (PAINS). *J. Nat. Prod.* **2016**, *79*, 616–628.
- (4) Baell, J. B.; Nissink, J. W. M. Seven Year Itch: Pan-Assay Interference Compounds (PAINS) in 2017 – Utility and Limitations. *ACS Chem. Biol.* **2018**, *13*, 36–44.
- (5) Simeonov, A.; Jadhav, A.; Thomas, C. J.; Wang, Y.; Huang, R.; Southall, N. T.; Shinn, P.; Smith, J.; Austin, C. P.; Auld, D. S., et al. Fluorescence spectroscopic profiling of compound libraries. *J. Med. Chem.* **2008**, *51*, 2363–2371.
- (6) Schorpp, K.; Rothenaigner, I.; Salmina, E.; Reinshagen, J.; Low, T.; Brenke, J. K.; Gopalakrishnan, J.; Tetko, I. V.; Gul, S.; Hadian, K. Identification of small-molecule

- frequent hitters from AlphaScreen high-throughput screens. *J. Biomol. Screen.* **2014**, *19*, 715–726.
- (7) Feng, B. Y.; Simeonov, A.; Jadhav, A.; Babaoglu, K.; Inglese, J.; Shoichet, B. K.; Austin, C. P. A high-throughput screen for aggregation-based inhibition in a large compound library. *J. Med. Chem.* **2007**, *50*, 2385–2390.
- (8) Soares, K. M.; Blackmon, N.; Shun, T. Y.; Shinde, S. N.; Takyi, H. K.; Wipf, P.; Lazo, J. S.; Johnston, P. A. Profiling the NIH Small Molecule Repository for compounds that generate H_2O_2 by redox cycling in reducing environments. *Assay Drug Dev. Technol.* **2010**, *8*, 152–174.
- (9) Ingólfsson, H. I.; Thakur, P.; Herold, K. F.; Hobart, E. A.; Ramsey, N. B.; Periole, X.; De Jong, D. H.; Zwama, M.; Yilmaz, D.; Hall, K., et al. Phytochemicals perturb membranes and promiscuously alter protein function. *ACS Chem. Biol.* **2014**, *9*, 1788–1798.
- (10) Huth, J. R.; Mendoza, R.; Olejniczak, E. T.; Johnson, R. W.; Cothron, D. A.; Liu, Y.; Lerner, C. G.; Chen, J.; Hajduk, P. J. ALARM NMR: a rapid and robust experimental method to detect reactive false positives in biochemical screens. *J. Am. Chem. Soc.* **2005**, *127*, 217–224.
- (11) Harrison, A. M.; Heritier, F.; Childs, B. G.; Bostwick, J. M.; Dziadzko, M. A. Systematic review of the use of phytochemicals for management of pain in cancer therapy. *Biomed. Res. Int.* **2015**, *2015*.
- (12) Palombo, E. A. Phytochemicals from traditional medicinal plants used in the treatment of diarrhoea: modes of action and effects on intestinal function. *Phytother Res.* **2006**, *20*, 717–724.
- (13) Tsuchiya, H. Membrane interactions of phytochemicals as their molecular mechanism applicable to the discovery of drug leads from plants. *Molecules* **2015**, *20*, 18923–18966.

- (14) Barnoud, J.; Rossi, G.; Marrink, S. J.; Monticelli, L. Hydrophobic compounds reshape membrane domains. *PLOS Comp. Biol.* **2014**, *10*, e1003873.
- (15) Menichetti, R.; Kremer, K.; Bereau, T. Efficient potential of mean force calculation from multiscale simulations: solute insertion in a lipid membrane. *Biochem. Biophys. Res. Commun.* **2018**, *498*, 282–287.
- (16) Jesus, A. R.; Vila-Viçosa, D.; Machuqueiro, M.; Marques, A. P.; Dore, T. M.; Rauter, A. P. Targeting type 2 diabetes with C-glucosyl dihydrochalcones as selective sodium glucose co-transporter 2 (SGLT2) inhibitors: synthesis and biological evaluation. *J. Med. Chem.* **2017**, *60*, 568–579.
- (17) Matos, A. M.; Blázquez-Sánchez, M. T.; Bento-Oliveira, A.; Almeida, R. F. M.; Nunes, R.; Lopes, P. E. M.; Machuqueiro, M.; Cristóvão, J. S.; Gomes, C. M.; Souza, C. S., et al. Glucosylpolyphenols as inhibitors of A β -induced Fyn kinase activation and Tau phosphorylation: synthesis, membrane permeability, and exploratory target assessment within the scope of type 2 diabetes and Alzheimer’s disease. *J. Med. Chem.* **2020**, *63*, 11663–11690.
- (18) Magalhães, P. R.; Reis, P. B.; Vila-Viçosa, D.; Machuqueiro, M.; Victor, B. L. *Computational Design of Membrane Proteins*; Springer, 2021; pp 263–271.
- (19) Berendsen, H. J. C.; van der Spoel, D.; van Drunen, R. GROMACS: a message-passing parallel molecular dynamics implementation. *Comput. Phys. Commun.* **1995**, *91*, 43–56.
- (20) van der Spoel, D.; Lindahl, E.; Hess, B.; Groenhof, G.; E., M. A.; Berendsen, H. J. C. GROMACS: Fast, Flexible, and Free. *J. Comput. Chem.* **2005**, *26*, 1701–1718.
- (21) Abraham, M. J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J. C.; Hess, B.; Lindahl, E. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **2015**, *1*, 19–25.

- (22) Schmid, N.; Eichenberger, A. P.; Choutko, A.; Riniker, S.; Winger, M.; Mark, A. E.; van Gunsteren, W. F. Definition and testing of the GROMOS force-field versions 54A7 and 54B7. *Eur. Biophys. J.* **2011**, *40*, 843–856.
- (23) Malde, A. K.; Zuo, L.; Breeze, M.; Stroet, M.; Poger, D.; Nair, P. C.; Oostenbrink, C.; Mark, A. E. An automated force field topology builder (ATB) and repository: version 1.0. *J. Chem. Theory Comput.* **2011**, *7*, 4026–4037.
- (24) Canzar, S.; El-Kebir, M.; Pool, R.; Elbassioni, K.; Malde, A. K.; Mark, A. E.; Geerke, D. P.; Stougie, L.; Klau, G. W. Charge group partitioning in biomolecular simulation. *J. Comput. Biol.* **2013**, *20*, 188–198.
- (25) Koziara, K. B.; Stroet, M.; Malde, A. K.; Mark, A. E. Testing and validation of the Automated Topology Builder (ATB) version 2.0: prediction of hydration free enthalpies. *J. Computer-Aided Mol. Design* **2014**, *28*, 221–233.
- (26) Poger, D.; van Gunsteren, W. F.; Mark, A. E. A new force field for simulating phosphatidylcholine bilayers. *J. Comput. Chem.* **2010**, *31*, 1117–1125.
- (27) 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.* **2010**, *6*, 325–336.
- (28) Darden, T.; York, D.; Pedersen, L. Particle mesh Ewald: An $N \cdot \log(N)$ method for Ewald sums in large systems. *J. Chem. Phys.* **1993**, *98*, 10089–10092.
- (29) Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. A smooth particle mesh Ewald method. *J. Chem. Phys.* **1995**, *103*, 8577–8593.
- (30) Páll, S.; Hess, B. A flexible algorithm for calculating pair interactions on SIMD architectures. *Comput. Phys. Commun.* **2013**, *184*, 2641–2650.

- (31) Hess, B. P-LINCS: A Parallel Linear Constraint Solver for Molecular Simulation. *J. Chem. Theory Comput.* **2008**, *4*, 116–122.
- (32) Miyamoto, S.; Kollman, P. A. SETTLE: An analytical version of the SHAKE and RATTLE algorithm for rigid water models. *J. Comput. Chem.* **1992**, *13*, 952–962.
- (33) Hermans, J.; Berendsen, H. J. C.; van Gunsteren, W. F.; Postma, J. P. M. A Consistent Empirical Potential for Water-Protein Interactions. *Biopolymers* **1984**, *23*, 1513–1518.
- (34) Bussi, G.; Donadio, D.; Parrinello, M. Canonical sampling through velocity rescaling. *J. Chem. Phys.* **2007**, *126*, 014101.
- (35) Nosé, S.; Klein, M. L. Constant pressure molecular dynamics for molecular systems. *Mol. Phys.* **1983**, *50*, 1055–1076.
- (36) Parrinello, M.; Rahman, A. Polymorphic transitions in single crystals: A new molecular dynamics method. *J. Appl. Phys.* **1981**, *52*, 7182–7190.
- (37) Luenberger, D. G.; Ye, Y., et al. *Linear and nonlinear programming*; Springer, 1984; Vol. 2.
- (38) Hub, J. S.; de Groot, B. L.; van der Spoel, D. g_wham – A Free Weighted Histogram Analysis Implementation Including Robust Error and Autocorrelation Estimates. *J. Chem. Theory Comput.* **2010**, *6*, 3713–3720.
- (39) Diamond, J. M.; Katz, Y. Interpretation of nonelectrolyte partition coefficients between dimyristoyl lecithin and water. *J. Membrane Biol.* **1974**, *17*, 121–154.
- (40) Marrink, S.-J.; Berendsen, H. J. C. Simulation of water transport through a lipid membrane. *J. Phys. Chem.* **1994**, *98*, 4155–4168.
- (41) Vila-Viçosa, D.; Victor, B. L.; Ramos, J.; Machado, D.; Viveiros, M.; Switala, J.; Loewen, P. C.; Leitão, R.; Martins, F.; Machuqueiro, M. Insights on the mechanism of action of INH-C₁₀ as an antitubercular prodrug. *Mol. Pharm.* **2017**, *14*, 4597–4605.

- (42) Hummer, G. Position-dependent diffusion coefficients and free energies from Bayesian analysis of equilibrium and replica molecular dynamics simulations. *New J. Phys.* **2005**, *7*, 34.
- (43) Dickson, C. J.; Hornak, V.; Pearlstein, R. A.; Duca, J. S. Structure–kinetic relationships of passive membrane permeation from multiscale modeling. *J. Am. Chem. Soc.* **2017**, *139*, 442–452.
- (44) Silva, T. F. D.; Vila-Viçosa, D.; Reis, P. B. P. S.; Victor, B. L.; Diem, M.; Oostenbrink, C.; Machuqueiro, M. The impact of using single atomistic long-range cutoff schemes with the GROMOS 54A7 force field. *J. Chem. Theory Comput.* **2018**, *14*, 5823–5833.
- (45) Vila-Viçosa, D.; Silva, T. F. D.; Slaybaugh, G.; Reshetnyak, Y. K.; Andreev, O. A.; Machuqueiro, M. Membrane-Induced pK_a Shifts in *wt*-pHLIP and its L16H Variant. *J. Chem. Theory Comput.* **2018**, *14*, 3289–3297.
- (46) Silva, T. F. D.; Vila-Viçosa, D.; Machuqueiro, M. Improved Protocol to Tackle the pH Effects on Membrane-Inserting Peptides. *J. Chem. Theory Comput.* **2021**, *17*, 3830–3840.
- (47) Williams, T.; Kelley, C. Gnuplot. 2004.
- (48) Schrödinger, LLC, The PyMOL Molecular Graphics System, Version 1.8. 2015.
- (49) GIMP 2.10.14, www.gimp.org. 1997-2021.
- (50) Magalhães, P. R.; Machuqueiro, M.; Baptista, A. M. Constant-pH Molecular Dynamics Study of Kyotorphin in an Explicit Bilayer. *Biophys. J.* **2015**, *108*, 2282–2290.
- (51) de Faria, C. F.; Moreira, T.; Lopes, P.; Costa, H.; Krewall, J. R.; Barton, C. M.; Santos, S.; Goodwin, D.; Machado, D.; Viveiros, M., et al. Designing new antituber-

cular isoniazid derivatives with improved reactivity and membrane trafficking abilities.

Biomed. Pharmacother. **2021**, *144*, 112362.