Optimization of an *in Silico* Protocol Using Probe Permeabilities to Identify Membrane Pan-Assay Interference Compounds

Pedro R. Magalhães, Pedro B. P. S. Reis, Diogo Vila-Viçosa, Miguel Machuqueiro,* and Bruno L. Victor*

**ABSTRACT:** Membrane pan-assay interference compounds (PAINS) are a class of molecules that interact nonspecifically 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 on the basis of umbrella sampling (US)/molecular dynamics (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 phosphate 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 inhomogeneous solubility-diffusion model (ISDM) used to calculate the membrane permeability coefficients confirmed that resveratrol and curcumin have distinct membrane PAINS characteristics and indicated a misclassification of nothofagin in a previous work. Overall, we have presented here a promising *in silico* protocol that can be adopted as a future reference method to identify membrane PAINS.

**INTRODUCTION**

High-throughput screening is a commonly used approach in drug discovery campaigns to identify compounds showing activity to a specific therapeutic target. 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. Such “frequent hitters,” 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.

This class of promiscuous compounds was named in 2010 by Baell and Holloway as pan-assay interference compounds (PAINS) and gained more attention in this past decade. PAINS comprise a large variety of compounds with different sources of diverse behavior or assay interference. Compound fluorescence events, chelation, chemical aggregation, redox activity, membrane perturbation/disruption, and nonselective compounds are just a few examples of characteristic interference chemicals. Although the selectivity problems associated with 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. Phytochemicals are one of the major components of plants and have long been used in traditional medicine to treat several different health problems. The molecular mechanism associated with these compounds has been conventionally interpreted or theorized by effects on receptors, biological pathways, ion channels, and transporters. 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. These special phytochemicals are currently known as membrane PAINS since they affect membrane physicochemical properties, such as curvature, fluidity, viscosity, elasticity, and permeability. 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 antitumor compounds.

Received: March 31, 2022
Published: June 13, 2022
In recent years, there has been a large interest in the development of computational methods to identify membrane PAINS and characterize their mode of action. The developed methodologies focused on quantifying the membrane deformations due to the presence of embedded potential membrane PAINS. Ingólfsson and co-workers exp10, and POPC+RES. An additional system containing an embedded potential membrane deformations due to the presence of embedded potential membrane PAINS. Ingólfsson and co-workers explored the membrane perturbation effects of several phytochemicals in membranes and mechanosensitive membrane proteins through a combination of gromicidin-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

The new implementations of the US protocol increased the molecular detail using atomistic MD and significantly improved the description of the membrane energy barrier.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 nothoagin.9,16 We have tested the use of long MD simulations, multiple replicates, different US schemes, and different atoms as US reference groups.

## METHODS

### 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), nothoagin (NOT), or resveratrol (RES) molecules. The compounds were evenly and randomly distributed between membrane leaflets and POPC in a 1:10 molar ratio (similarly to Ingólfsson et al.), resulting in different starting systems: pure POPC, POPC+CUR, POPC+NOT, and POPC+RES. An additional system containing a 2:10 molar ratio of POPC to CUR molecules (CUR24) was also built following the same protocol.

Molecular dynamics simulations were performed using GROMACS 2018.6.19–21 and the united-atom GROMOS 54A7 force field.22 Topologies for the compounds were obtained using the automated topology builder server (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) method28,29 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 thermostat34 with a coupling constant of 0.1 ps. A semi-isotropic Parrinello–Rahman barostat35,36 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 × 10⁻⁴ 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; second, with all bonds constrained and a maximum step size of 0.001 nm. The tolerance was set to 0.0 kJ mol⁻¹ 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). 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). 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.

### Umbrella Sampling

To prepare each system for the umbrella sampling simulations, we added a probe defined as a Lennard-Jones “sphere” with an overall size comparable to a benzene molecule, as described in ref 16. The main role of the probe particle is to detect membrane perturbations. If the probe is too small, it will lose sensitivity, while if it is too large, it may itself perturb the membrane. Therefore, all probe sizes between these two extremes should lead to similar results. 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 kJ mol⁻¹ nm⁻² and a velocity of 1 nm/ns) while keeping the x/y 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 (K₁) 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). According to our findings, the use of 300 ns of equilibration assured that any initial conformational bias was removed or, at least, well-mitigated. Since the most difficult part to converge is the phosphate region, only after applying such a long protocol, we managed to decrease the dispersion between replicates. In addition to the general protocol described above, some variations in the number and positions of the umbrellas (as well as their corresponding K₁ values) were introduced. The details of these variations are included and discussed in the Results and Discussion.

### Analyses and Error Calculations

The potential of mean force (PMF) profiles for each system were calculated using the
weighted-histogram analysis method (WHAM)\textsuperscript{38} implemented in GROMACS. Membrane permeabilities were calculated using the inhomogeneous solubility-diffusion model (ISDM),\textsuperscript{39,40} implemented in a software package developed by Vila-Vicosa and co-workers\textsuperscript{41} based on the formalism described in refs 42 and 43 and succinctly described in the Supporting Information. The position-dependent diffusion values of the probe are computed using the autocorrelation function of its $zz$-position at each umbrella window. Combining the diffusion and the PMF values, we can calculate the position-dependent resistance profile that can then be integrated to obtain the overall permeation coefficient (shown in Figure S4). The standard error values included in the figures and tables were obtained using a modified jackknife resampling approach. This method uses combinations of replicates of a given system in a leave-one-out strategy. Thus, using $n$ combinations of $n-1$ replicate subsamples, we can estimate the standard error values of the original sampling.\textsuperscript{44} This approach has the advantage of avoiding the error estimation of system properties from single replicate averages that may have convergence issues.

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.\textsuperscript{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,\textsuperscript{37} PyMOL,\textsuperscript{48} and GIMP.\textsuperscript{39}

\textbf{RESULTS AND DISCUSSION}

\textbf{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 S5A. This system, which we designate as 37UB, 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 the 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 S5B. 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 37UB system (Figure S5B) or by separating the individual replicates (Figure 1A). Although the system is composed only of the pure POPC bilayer and the probe, there is substantial variability between some replicates. This illustrates the importance of using long equilibration times with several replicates, since using a shorter single replicate\textsuperscript{16} can result in an underestimated or overestimated free energy. As previously remarked,\textsuperscript{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., in particular, the relative location of the energy maximum and the 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 et al. As previously discussed, the extra detail in the atomistic (united-atom) simulations, compared with coarse-grained MARTINI, seems to be very important to discriminate compounds with different membrane PAINS characteristics. The major differences observed between the resulting PMF profiles are found at (i) the center of the lipid bilayer with MARTINI creating an artificial energy barrier, which is corrected in the atomistic simulations; (ii) at the phosphate region, where the coarse-grained force field does not correctly model the expected energy barrier of a hydrophobic probe. This limitation is particularly important since the changes in the phosphate group region seem to be more impactful for membrane structural stability and permeability.

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 simulated 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 theory, the choice of the reference point should have no effect on the final PMF. However, in most systems, this is not true since there are always sampling limitations. 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 37UM (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 S6). 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 37UM 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 time scale. 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 S7). 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 toward 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 reference, 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 effect; 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. A reduction in 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 the Methods), a limitation that can be worsen 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 S5) 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 (22UM), representing a substantial decrease in the overall computational cost. The number and position of the umbrella windows cannot 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).

Using this protocol, we managed to maintain the quality of the sampling (overlap between the neighboring umbrellas, Figure 2A).

As shown in Table S1 and Figure 2A, we increased the $K_f$ value in the region near the P-layer (i.e., −0.4 to 0.4) from 1000 to 2000 kJ mol$^{-1}$ 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 very closely resemble 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 22UM 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. 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) and at the eventual barrier at the center of the membrane when present. However, the PMF profile also conveys information on the probe diffusion across the membrane, the resistance encountered and, ultimately, a membrane permeability coefficient. 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, as previously described. Although ISDM has been successfully used to estimate the membrane permeability to hydrophobic compounds, 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 (37UB, 37UM, and 22UM) and observed only small differences between them, all within the error margin (Table 1). The permeability coefficient values seem to be correlated with 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 22UM 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. Protocol Application to Identify PAINS Compounds. In the previous section, we showed that the 22UM 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 obtained well-converged PMF profiles allowed us to acquire 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: resveratrol (RES), as a mild membrane PAINS, noothofagin (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 1:10 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 2:10 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 S3).

We observe that all compounds change the PMF profile; however, the differences between the 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

Table 1. Membrane Permeability Coefficients (cm s$^{-1}$) and Energy Barriers of Entering and Exiting the POPC Bilayer (kcal mol$^{-1}$)

| system | permeability | $\Delta G_{entry}$ | $\Delta G_{exit}$ |
|--------|--------------|------------------|------------------|
| 37UB   | 4.6 ± 0.8    | 2.6 ± 0.1        | 7.2 ± 0.1        |
| 37UM   | 4.3 ± 0.5    | 2.7 ± 0.1        | 7.2 ± 0.1        |
| 22UM   | 4.4 ± 0.4    | 2.9 ± 0.0        | 7.1 ± 0.1        |

The entry barrier ($\Delta G_{entry}$) is the energy difference between the maximum value (at the membrane/water interface) and bulk water, while the exit barrier ($\Delta G_{exit}$) is the difference between the global minimum at the membrane center and the previous maximum. Errors were calculated using a jackknife approach.
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 on the PMF (Figure 3), leading to a signiﬁcant diﬀerence in the calculated permeability coeﬃcient (5.6 ± 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 in the work by the Jesus and co-workers,16 which used a single 50 ns replicate. In our work, we performed (5×) 300 ns of pre-equilibration MD simulations, followed by 300 ns production MD, to allow convergence in the nothofagin/membrane conﬁguration space. This will mitigate any initial system building bias and allow the correct equilibration of our system properties, such as the local membrane thickness and total area of the membrane patch. 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 ± 0.5 vs 4.4 ± 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 (2:10 mol/mol ratio). The eﬀect of doubling the concentration of curcumin appeared to be 2-fold: ﬁrst, an increase in the perturbation eﬀect when compared to pure POPC, as evidenced by the shape of the PMF proﬁle (Figure 3) and the corresponding permeability coeﬃcient (5.7 ± 0.3 vs 4.4 ± 0.4); 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. Overall, it seems that the addition of more curcumin molecules to the lipid bilayer does lead to a statistically signiﬁcant

Table 2. Energy Barriers of Entry \( (\Delta G_{\text{entry}}) \) and Exiting \( (\Delta G_{\text{exit}}) \) the POPC Bilayer and Membrane Permeabilities

| system | permeability | \( \Delta G_{\text{entry}} \) | \( \Delta G_{\text{exit}} \) |
|--------|--------------|-----------------|-----------------|
| POPC   | 4.4 ± 0.4    | 2.9 ± 0.1       | 7.1 ± 0.1       |
| RES    | 4.5 ± 0.3    | 2.5 ± 0.1       | 6.6 ± 0.1       |
| NOT    | 5.6 ± 0.6    | 2.3 ± 0.0       | 7.1 ± 0.2       |
| CUR    | 5.0 ± 0.5    | 2.4 ± 0.0       | 6.6 ± 0.2       |
| CUR24  | 5.7 ± 0.3    | 2.3 ± 0.0       | 6.3 ± 0.1       |

*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.*

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 on the PMF (Figure 3), leading to a signiﬁcant diﬀerence in the calculated permeability coeﬃcient (5.6 ± 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 in the work by the Jesus and co-workers,16 which used a single 50 ns replicate. In our work, we performed (5×) 300 ns of pre-equilibration MD simulations, followed by 300 ns production MD, to allow convergence in the nothofagin/membrane conﬁguration space. This will mitigate any initial system building bias and allow the correct equilibration of our system properties, such as the local membrane thickness and total area of the membrane patch. 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 ± 0.5 vs 4.4 ± 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 (2:10 mol/mol ratio). The eﬀect of doubling the concentration of curcumin appeared to be 2-fold: ﬁrst, an increase in the perturbation eﬀect when compared to pure POPC, as evidenced by the shape of the PMF proﬁle (Figure 3) and the corresponding permeability coeﬃcient (5.7 ± 0.3 vs 4.4 ± 0.4); 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. 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 $\Delta G_{\text{entry}}$ or the $\Delta G_{\text{exit}}$ values. Similarly to the values for pure POPC (Table 1), we observed a strong negative correlation ($-0.83$) for $\Delta G_{\text{entry}}$ and a significantly smaller value for $\Delta G_{\text{exit}}$ ($-0.28$). With such a high anticorrelation between the membrane permeability values and the $\Delta G_{\text{entry}}$ 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 to calculate the complete energy profile.

**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 on the basis of umbrella sampling MD simulations to discriminate between compounds with differentiated membrane PAINS behavior. 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 (nonmembrane PAINS). The membrane permeability coefficients calculated using the inhomogeneous solubility-diffusion model for the different systems confirmed that resveratrol has mild membrane PAINS characteristics and that curcumin exhibits a concentration-dependent membrane PAINS behavior. However, our results indicate that nothofagin, which was previously identified as a nonmembrane PAINS compound, presents a significant membrane perturbation effect, suggesting a misclassification of this compound. The high accuracy and robustness of our protocol comes at a significant computational cost; therefore, it should also be used as a reference control for future developments aiming at faster protocols. Interestingly, we have observed a very high anticorrelation between the probe entry energy barrier ($\Delta G_{\text{entry}}$) and the membrane PAINS character. This can be an important step toward faster methods since this energy barrier can be calculated using just a fraction of our protocol. Overall, our effective approach emerges as a reference in silico method to identify and discriminate among membrane PAINS compounds.

**DATA AND SOFTWARE AVAILABILITY**

The GROMACS package is a freely available software used to perform MD simulations and can be downloaded at https://manual.gromacs.org/documentation/2018.6/download.html. PyMOL v2.0 is also a free software for molecular visualization and the generation of high quality images. It can be downloaded from https://pymol.org/. Gnuplot 5.4 is a portable command-line driven graphing utility used to generate plots, and it is freely available from http://www.gnuplot.info/download.html. Gimp 2.10.30 is a free and open source software image editor that can be downloaded from https://www.gimp.org/downloads/. Additionally, as Supporting Information, we provide the system starting configurations and topologies.

**ASSOCIATED CONTENT**

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jcim.2c00372.

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 (pulx) for the different schemes; local thickness per umbrella from the 37UM system; table with the positions and force constants used in the 22UM system (PDF)

**AUTHOR INFORMATION**

Corresponding Authors

Miguel Machuqueiro — BioISI - Biosystems & Integrative Sciences Institute, Faculty of Sciences, University of Lisbon, 1749-016 Lisboa, Portugal; orcid.org/0000-0001-6923-8744; Phone: +351-21-7500112; Email: machuque@ciencias.ulisboa.pt

Bruno L. Victor — BioISI - Biosystems & Integrative Sciences Institute, Faculty of Sciences, University of Lisbon, 1749-016 Lisboa, Portugal; orcid.org/0000-0002-6472-5700; Phone: +351-21-7500000; Email: blvictor@ciencias.ulisboa.pt

Authors

Pedro R. Magalhães — BioISI - Biosystems & Integrative Sciences Institute, Faculty of Sciences, University of Lisbon, 1749-016 Lisboa, Portugal

Pedro B. P. S. Reis — BioISI - Biosystems & Integrative Sciences Institute, Faculty of Sciences, University of Lisbon, 1749-016 Lisboa, Portugal; orcid.org/0000-0003-3563-6239

Diogo Vila-Viçosa — BioISI - Biosystems & Integrative Sciences Institute, Faculty of Sciences, University of Lisbon, 1749-016 Lisboa, Portugal

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jcim.2c00372

Notes

The authors declare no competing financial interest.
ACKNOWLEDGMENTS

We thank 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.

REFERENCES

(1) Macarron, R.; Banks, M. N.; Bojanic, D.; Burns, D. J.; Ciovic, D. A.; Gáryantes, 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 Discovery 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) Schirpp, K.; Rothenaigner, I.; Salmina, E.; Reinschagen, 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, 715–725.
(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.; Takiy, H. K.; Wipf, P.; Lazo, J. S.; Johnston, P. A. Profiling the NIH Small Molecule Repository for compounds that generate H2O2 by redox cycling in reducing environments. Aasay Drug Dev. Technol. 2010, 8, 152–174.
(9) Ingólfsson, H. L.; Thakur, P.; Herold, K. F.; Hobart, E. A.; Ramsey, N. B.; Periolo, X.; De Jong, D. H.; Zwama, M.; Yilmaz, D.; Hall, K.; et al. Phychotoxins perturb membranes and promiscuously alter protein function. ACS Chem. Biol. 2014, 9, 1788–1798.
(10) Firth, J. B.; Mendoza, R.; Olejniczak, E. T.; Johnsen, R. W.; Cothron, D. A.; Liu, Y.; Lerner, C. G.; Chen, J.; Haiduk, 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.; Heritter, F.; Childs, B. G.; Bostwick, J. M.; Dziadzko, M. A. Systematic review of the use of phytochemicals for the management of pain in cancer therapy. J. Med. Chem. 2017, 60, 568–579.
(12) de 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. Glucosylpolypeh-nols as inhibitors of Aβ-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.
(13) Magalhães, P. B.; Reis, P. P.; Vila-Viçosa, D.; Machuqueiro, M.; Victor, B. L. Computational Design of Membrane Proteins; Springer, 2021; pp 263–271.
(14) 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.
(15) Van Der Spoel, D.; Lindahl, E.; Hess, B.; Groenhof, G.; Mark, A. E.; Berendsen, H. J. C. GROMACS: Fast, Flexible, and Free. J. Comput. Chem. 2005, 26, 1701–1718.
(16) 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 super-computers. SoftwateX 2015, 1, 19–25.
(17) 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.
(18) 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.
(19) 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.
(20) Kozik, 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.
(21) Poger, D.; van Gunsteren, W. F.; Mark, A. E. A new force field for simulating phosphatidylincholine bilayers. J. Comput. Chem. 2010, 31, 1117–1125.
(22) 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.
(23) Darden, T.; York, D.; Pedersen, L. Particle mesh Ewald: An N − log(N) method for Ewald sums in large systems. J. Chem. Phys. 1993, 98, 10089–10092.
(24) Eismann, 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.
(25) Páll, S.; Hess, B. A flexible algorithm for calculating pair interactions on SIMD architectures. Comput. Phys. Commun. 2013, 184, 2641–2650.
(26) Hess, B. P-LINCS: A Parallel Linear Constraint Solver for Molecular Simulation. J. Chem. Theory Comput. 2008, 4, 116–122.
(27) 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.
(28) 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.
(29) Bussi, G.; Donadio, D.; Parrinello, M. Canonical sampling through velocity rescaling. J. Chem. Phys. 2007, 126, 014101.
(30) Nosé, S.; Klein, M. L. Constant pressure molecular dynamics simulation. J. Chem. Phys. 1984, 80, 1555–1576.
(31) 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. Membr. 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. Pharmaceutics 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.; Slabyaugh, G.; Reshetnyak, Y. K.; Andreev, O. A.; Machuqueiro, M. Membrane-Induced pK\textsubscript{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; http://gnuplot.sourceforge.net/docs_4.2/.
(48) Schrödinger, Inc. The PyMOL Molecular Graphics System, Version 1.8; 2015.
(49) GIMP 2.10.14; 1997–2021; www.gimp.org.
(50) Magalhães, P. R.; Machuqueiro, M.; Baptista, A. M. Constant-pH Molecular Dynamics Study of Kytorphin 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 antitubercular isoniazid derivatives with improved reactivity and membrane trafficking abilities. Biomed. Pharmacother. 2021, 144, 112362.