Estimating the accuracy of the MARTINI model towards the investigation of Peripheral Protein-Membrane Interactions

Sriraksha Srinivasan[1], Valeria Zoni[1], Stefano Vanni[1,*]

[1] Department of Biology, University of Fribourg, Switzerland

ABSTRACT: Peripheral membrane proteins play a major role in numerous biological processes by transiently associating with cellular membranes, often with extreme membrane specificity. Because of the short-lived nature of these interactions, molecular dynamics (MD) simulations have emerged as an appealing tool to characterize at the structural level the molecular details of the protein-membrane interface. Transferable coarse-grained (CG) MD simulations, in particular, offer the possibility to investigate the spontaneous association of peripheral proteins to lipid bilayers of different compositions at limited computational cost, but they are hampered by the lack of a reliable *a priori* estimation of their accuracy and thus typically require *a posteriori* experimental validation. In this Discussion, we investigate the ability of the MARTINI CG force field, and specifically of its 3 open-beta version, to reproduce known experimental observations regarding the membrane binding behavior of 12 peripheral membrane proteins and peptides. Based on observations of multiple binding and unbinding events in several independent replicas, we found that, despite the presence of false positives and false negatives, this model is mostly able to correctly characterize the membrane binding behavior of peripheral proteins, and to identify key residues found to disrupt membrane binding in mutagenesis experiments. While preliminary, our investigations suggest that transferable chemical-specific CG force fields have enormous potential towards the characterization of the membrane binding process by peripheral proteins, and that the identification of negative results could help drive future force field development efforts.

INTRODUCTION

The interaction of peripheral proteins with cellular membranes plays a pivotal role in numerous biological processes including lipid metabolism and transport[1-3],membrane trafficking, and signal transduction[4, 5]. The interaction between proteins and the membrane surface is often reversible and transient[6], and it is tightly modulated by numerous factors, including protein-protein interactions, lipid post-translational modifications, specific binding to signaling lipids, or recognition of bulk membrane properties such as electrostatics or membrane packing[7].

However, due to the transient nature of protein-membrane interactions, determination of the three-dimensional structure of peripheral proteins is generally achieved in their soluble, membrane-free conformation. Furthermore, dynamic phenomena such as large-scale conformational changes upon membrane binding, and the molecular and energetic details of these interactions, are extremely difficult to elucidate via experimental methods. Molecular dynamics (MD) simulations provide an alternative route by allowing a detailed investigation of protein-membrane complexes, thus enabling the study of their dynamic behavior and the molecular characterization of the membrane binding process[8]. In particular, coarse-grained (CG) MD simulations have emerged in the last few years as a powerful tool to investigate interactions between proteins and membranes, and they have been shown to provide a powerful cost-effective alternative to fully atomistic MD simulations[9, 10].

In detail, one of the most widely used CG models in this context is the MARTINI model[11], which has fully compatible parameters available for several biomolecules such as proteins[12, 13], lipids[11, 14], nucleic acids[15, 16], and carbohydrates[17]. In the MARTINI model, four non-hydrogen atoms are generally represented as a single interaction site, called a *bead*. As a consequence, simulations can access much longer time- and size-scales, at a reasonable computational cost.

An intrinsic drawback of this CG strategy, however, is the almost complete lack of a reliable *a priori* estimation of its accuracy. This is related to the inevitable presence of systematic errors associated with the development of force field parameters, which generally takes place in simple molecular systems, and to the lack of a detailed understanding of whether these errors cancel or add up in more complex molecular systems[18], such as those involving protein-membrane interactions. Thus, corroboration

of the quality of a CG model is generally achieved *a posteriori*, based on comparison with experimental data.

For what pertains the binding of peripheral proteins to lipid bilayers, the MARTINI CG model has shown in the last few years to be a very promising tool[19], and especially with respect to strong or irreversible membrane binding, such as in the case of the specific recognition of signaling lipids such as phosphoinositides[20-24] or in the case of lipidated proteins[25, 26]. On the other hand, the ability of this model to accurately describe more transient ("on/off") membrane binding by peripheral proteins, as well as to reproduce the experimentally-determined ability of proteins to sense bulk membrane properties, such as electrostatics[4, 27] or lipid packing density[28-30], remains mostly unexplored.

A further complication of such investigations is the potential dependence of membrane binding to protein conformational changes or structural fluctuations. This is particularly challenging for the MARTINI force field as while the parameters for lipids and amino acids have been derived using a self-consistent strategy, a faithful representation of protein secondary structure in this model requires the *ad-hoc* use of elastic network models (ENMs)[31]. These approaches preserve the structure of the protein by generating additional harmonic bonds between the backbone beads of the protein and are thus characterized by supplementary parameters, such as the force constant of the harmonic bonds (fc) and the cut-off distance (R_C). Tuning of these parameters can affect the propensity of proteins to interact with lipid bilayers, and while recent studies suggest a potential role of the elastic network on protein-protein interactions and clustering[32], its influence on interactions between peripheral proteins and membranes has not yet been thoroughly investigated.

In this work, we investigate the ability of the MARTINI model (and specifically of the 3 open-beta version) to accurately describe and predict the transient binding of 12 peripheral proteins or peptides to lipid bilayers of different compositions. We design a simple protocol to extract information from membrane binding/unbinding time traces and we specifically focus on its ability to reproduce the sensing of bulk membrane properties displayed by certain protein families. Our data suggests that despite significant protein-to-protein variability and the presence of false positives and false negatives, this model provides, for the most part, meaningful structural information of the protein-membrane

interface. In addition, we find that elastic network parameters have marginal influence on membrane binding for globular proteins, but they affect *in silico* binding for those proteins that have been shown experimentally to undergo conformational transitions upon binding.

METHODS

Software details

All MD simulations of the protein-membrane systems were performed with the GROMACS (v 2018x)[33] package, using the open beta version of the Martini 3 force field[19]. All molecular images were rendered using Visual Molecular Dynamics (VMD)[34].

System setup

A total of 12 peripheral proteins were studied. They are listed in Table 1 along with their PDB IDs, the bilayer compositions they were set up with, and the duration of the simulations. The protein structures were obtained from RCSB PDB[35], with the exception of ArfGAP1 amphipathic lipid packing sensor (ALPS) helix (residues 199 to 223) that was modeled as an alpha helix. For the Arf1 protein, structures of two different conformations (GDP-bound and GTP-bound) are present in the RCSB PDB. Since the two conformations are markedly different, particularly in the N-terminal region which is responsible for membrane binding[36], we removed the ligands from the structures in our MD simulations. This approach is particularly justified for CG simulations since the elastic network employed retains the secondary structure of the protein. The atomistic structures were converted to CG models with an additional elastic network using the martinize[12] script. The lower and upper elastic bond cut-off were set to 0.5 nm and 0.9 nm respectively. For each protein system, independent simulations were performed with two elastic bond force constants: 500 and 1000 kJ mol^{-1} nm^{-2}. Side chain dihedral corrections were applied to the CG models using the addDihedral.tcl and bbsc.sh scripts, as necessitated for the 3 open beta version of the Martini force field.

Lipid bilayers of different compositions (as indicated in table 1) of 20 nm x 20 nm in lateral dimensions were constructed using the CHARMM-GUI Membrane Builder[37], and were equilibrated according to the standard six step equilibration protocol provided by CHARMM-GUI. The bilayers were then stripped of water molecules and ions, and the protein was placed such that the minimum distance between any bead of the protein and any bead of the lipid molecules was at a distance of at least 2.5 nm. The initial orientation of the protein was such that its principal axes were aligned with the *x*, *y*, and *z* directions

of the system, with the longer dimension of the protein aligned along the z direction. The systems were then solvated and ionized with 0.12 M of sodium and chloride ions.

Table 1. CG simulation details of proteins and peptides tested

Protein	PDB ID	fc (kJ mol^{-1} nm^{-2})	Bilayer composition	Simulation time (µs)	Number of replicas	Total simulation time (µs)
Ubiquitin	1UBQ	500,1000	DOPC	2	5	10
			DOPC/DOPS (80:20)	2	5	10
Hen egg white lysozyme	1AKI	500,1000	DOPC	2	5	10
			DOPC/DOPS (80:20)	2	5	10
PDK1 PH Domain (residues 407-549)	1W1D	500,1000	DOPC	4	6	24
			DOPC/DOPS (90:10)	4	6	24
			DOPC/DOPS (80:20)	4	6	24
			DOPC/DOPS (70:30)	4	6	24
Evectin-2 PH Domain (residues 1-112)	3VIA	500,1000	DOPC	2	6	12
			DOPC/DOPS (80:20)	2	6	12
			DOPC/DOPS (70:30)	2	6	12
Lactadherin C2 Domain (residues 1-158)	3BN6	500,1000	DOPC	2	6	12
			DOPC/DOPS (90:10)	2	6	12
			DOPC/DOPS (80:20)	2	6	12
FVa C2 Domain (residues 1-160)	1CZT	500,1000	DOPC	3	6	18
			DOPC/DOPS (80:20)	3	6	18
PLA$_2$	1POA	500,1000	DMPC	2	5	10
			DOPC	2	5	10
			DOPC/DOPS (80:20)	2	5	10
Arf1 - GTP	2KSQ	500,1000	DOPC	2	6	12
Arf1 - GDP	2K5U	500,1000	DOPC	2	6	12
Ricin	1BR5	500,1000	DOPC	4	6	24
			DOPC/DOPS (70:30)	4	6	24
Mastoparan	1D7N	500,1000	DMPC	1	6	6
			POPC	1	6	6
			DOPC	1	6	6

Osh4 ALPS peptide (residues 1-29)	1ZHY	500,1000	DMPC	2	5	10
			POPC	2	5	10
			DOPC	2	5	10
ArfGAP1 ALPS peptide (residues 199-223)	Helical model	500,1000	DMPC	2	6	12
			POPC	2	6	12
			DOPC	2	6	12
Total simulation time (µs)						860

Simulation details

Initial equilibration was carried out by performing energy minimization using the steepest descent algorithm, followed by a short MD run of 250 ps with the protein backbone beads restrained. Production runs were performed at 310K using a velocity-rescale thermostat[38], with separate temperature coupling for the protein, lipids, and solvent particles. The pressure was maintained at 1 bar using the Parrinello-Rahman barostat[39], along with a semi-isotropic pressure coupling scheme. The non-bonded interactions were calculated by generating a pair-list using the Verlet scheme with a buffer tolerance of 0.005. The Coulombic terms were calculated using reaction-field and a cut-off distance of 1.1 nm. A cutoff scheme was used for the vdW terms, with a cut-off distance of 1.1 nm and the Verlet cut-off scheme for the potential-shift[40]. The md integrator was used, with a time step of 20 fs. The first 100 ns of the production runs were not considered for analyses.

Analysis details

Membrane binding events were characterized by analyzing the time-trace of the minimum distance between the protein and the lipid bilayer (Figs. S1-S11). This distance was computed using the *gmx mindist* tool in GROMACS[33].

Membrane binding was subsequently assessed by generating probability density distributions from these time traces using the kernel density estimation (KDE) method, and bound states were defined as those instances when the minimum distance was lower than or equal to 0.7 nm. The percentage of binding, which is the ratio between the area below the curve in the distribution within 0.7 nm and the total area under the curve was

computed in each case. Statistical errors for these values are reported as standard errors from the computed values in each individual replica.

Membrane-interacting residues were determined using the following protocol: a residue was considered to interact with the membrane if the distance between any bead of the residue and any lipid-bead was lower than or equal to 0.5 nm. For each residue, the number of instances of its interaction with the membrane during the trajectory was counted and summed over all the replicas, and a corresponding normalized value was computed.

RESULTS

To investigate the binding of peripheral proteins to membranes, we initially positioned the proteins at a distance of at least 2.5 nm from equilibrated lipid bilayers of different compositions (Fig. 1A). After MD simulations of few microseconds, as indicated in Table 1, binding events were identified and quantified using the protocol depicted in Fig.1. First, the time-trace of the minimum distance between the protein and the lipid bilayer was computed for all replicas (Fig. 1B, only one replica is shown), and the probability density distribution of these values (Fig. 1C) was generated by averaging the corresponding probability density distribution for all individual replicas. Membrane-bound states were defined as those instances when the minimum distance was lower than or equal to 0.7 nm (Fig. 1C), roughly corresponding to the first minimum observed in the probability density distribution.
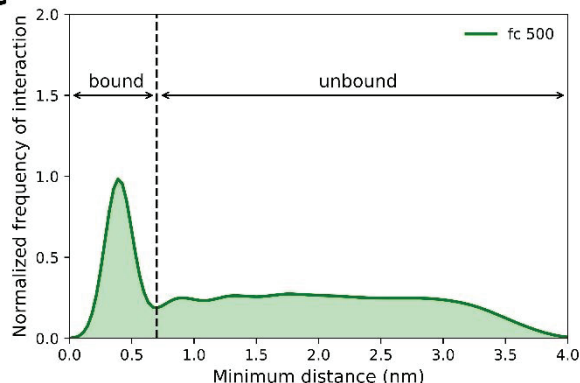
A
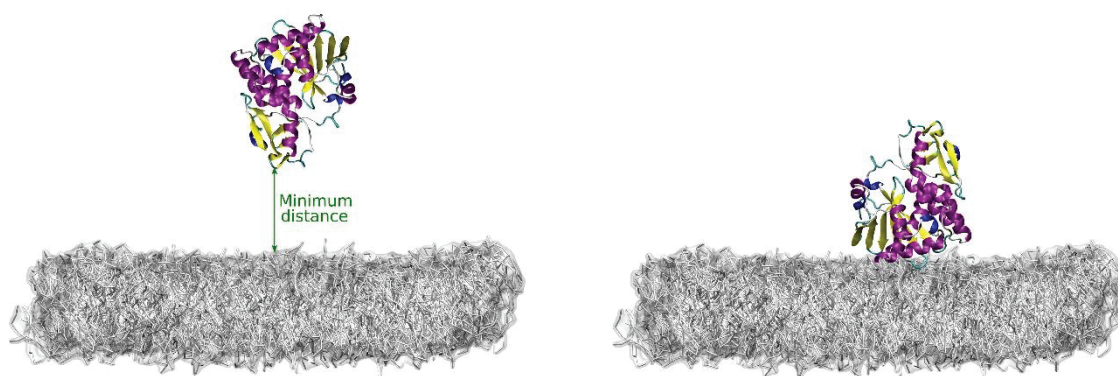
B
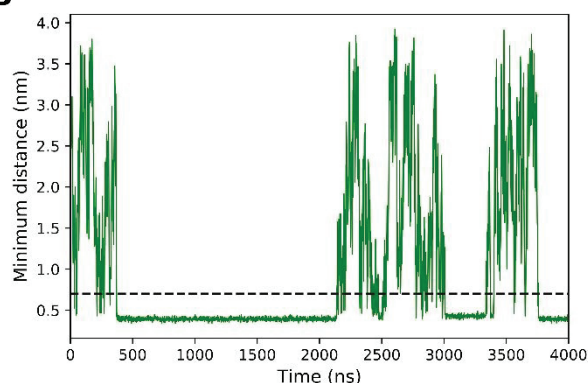
C

Figure 1: **Protocol used in this work to characterize membrane binding.** (A) Initial system set-up for a representative peripheral protein. The protein is shown in cartoon representation; the membrane is shown in gray. Water and ion beads are omitted for clarity. (B) Time trace of the minimum distance between the protein and the bilayer. (C) Probability density distribution of the time series in (B). Values are obtained by averaging over several independent replicas.

Soluble proteins.

As a first test to evaluate the ability of the MARTINI model to estimate membrane binding by peripheral proteins, we investigated, as a negative control, whether the model is able to accurately reproduce the lack of binding by proteins that are known to remain soluble in the cytosol and *in vitro* assays[41-43]. To this end, we tested the binding to pure DOPC lipid bilayers of two well-known soluble proteins, ubiquitin (Fig. 2A) and hen-egg white lysozyme (Fig. 2D), at two different force constants for the elastic network model, fc=500 kJ mol^{-1}nm^{-2} and fc=1000 kJ mol^{-1} nm^{-2}. In both cases, we observed no significant binding of the proteins to the bilayer, as indicated by the very low maximum at values below 0.7 nm in Fig. 2B and 2E. Rather, the two proteins adopted an unbound (protein-membrane

9

minimum distance > 0.7 nm) state in solution for most (90%) of the trajectory (Fig 2C,F and Fig. S1). To further stress-test the methodology, we next added 20% DOPS lipids in our membrane composition, to potentially promote binding by means of electrostatic interactions. Even in those conditions, no significant binding was observed (Fig. 2C,F). Of note, no influence of the elastic network force constant was observed (Fig. 2B,C,E,F).
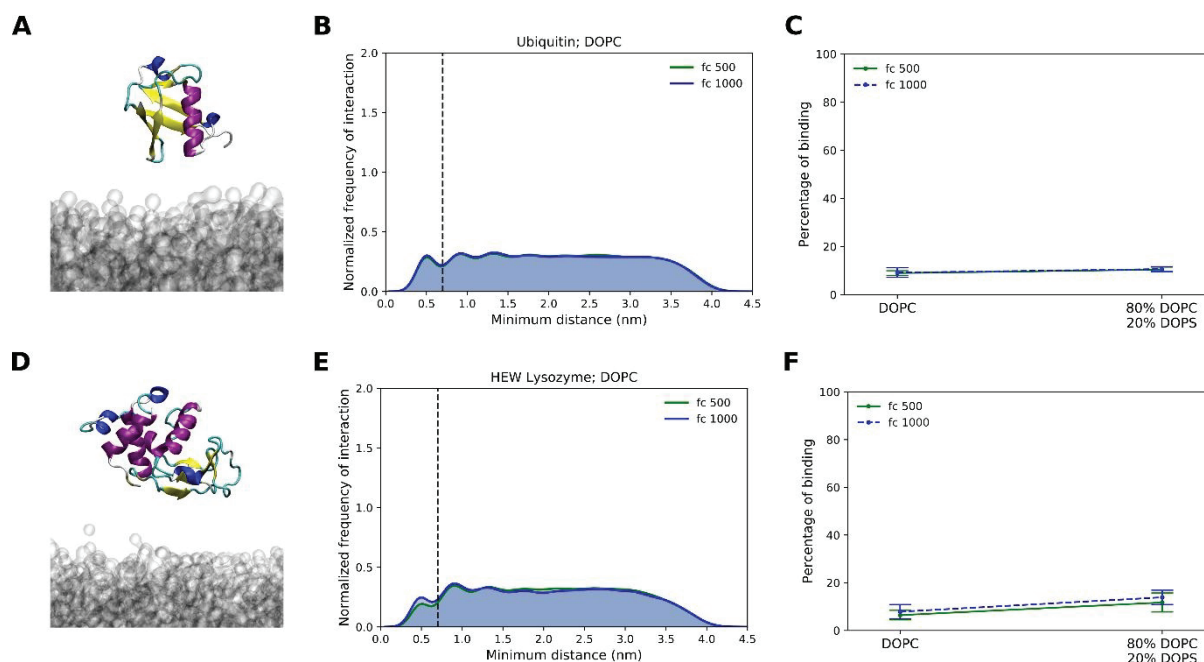

Figure 2: **Binding of soluble proteins to lipid bilayers.** (A,D) Representative structure of ubiquitin (A) and HEW Lysozyme (D). (B,E) Probability density distribution of minimum distance values for ubiquitin (B) and HEW Lysozyme (E), for two values of the force constant, in DOPC bilayers. No membrane binding was observed to DOPC bilayers for both proteins. (C,F) Percentage of binding at different membrane compositions. Addition of DOPS to the bilayer did not enhance binding significantly for both proteins.

Globular proteins.

Next, we evaluated whether the MARTINI model is able to reproduce the experimentally observed sensitivity to PS lipids that is a characteristic feature of specific protein families. To this end, we first tested two pleckstrin homology (PH) domains, belonging to 3-Phosphoinositide-dependent kinase-1 (PDK1) (Fig. 3A) and evectin-2 (Fig. 3D), as well as two C2 domains - those of lactadherin (Fig. 5A) and human coagulation factor V (FVa) (Fig. 5D), that have all been shown to bind to membranes in the presence of PS lipids[44-48].

In the case of the PH domain of PDK1, we observe significant membrane binding in the presence of PS lipids (Fig. 3B), in agreement with experimental observations[44]. When increasing the concentration of PS lipids, membrane binding increased almost monotonically, with the protein showing almost no binding in the absence of PS lipids and binding for more than 50% of the time during the trajectory in the presence of 30% PS (Fig. 3C). On the other hand, for the second PH domain we tested, that of evectin-2, we observed only marginal binding in the mixed PC/PS bilayers (Fig. 3E). In the absence of PS lipids, membrane binding further, albeit slightly, decreased (Fig. 3F), suggesting that the MARTINI model possibly underestimates membrane binding affinity for this protein. In both cases, no effect of the elastic network force constant was observed (Fig. 3C,F)
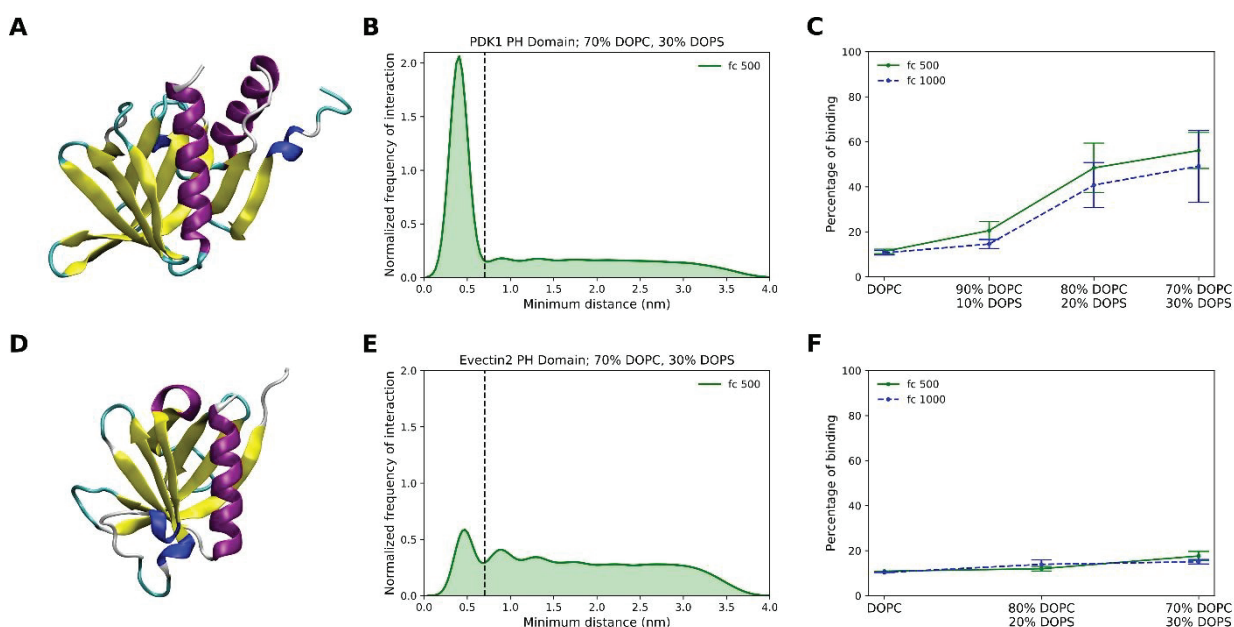
Figure 3: **Binding of PH domains to lipid bilayers.** (A,D) Representative structure of the PH domain of PDK1 (A) and evectin-2 (D). (B,E) Probability density distribution of protein-membrane minimum distances in PC/PS lipid bilayer. (C,F) Percentage of binding of the two proteins at different membrane compositions. Addition of PS to the bilayer enhanced binding significantly for the PH domain of PDK1 but not for that of evectin-2.

We next analyzed the membrane binding interface for both PH domains (Fig. 4A,C). To do so, we counted the instances of residue-membrane interactions in the simulations performed at high PS concentration and collected the corresponding frequencies (Fig. 4B,D). For PDK1, we found that residues R466 and K467 have the highest probability to bind the bilayer (Fig. 4B), in agreement with the binding mode that has been previously

characterized experimentally[44]. For Evectin-2, on the other hand, the protein is known to bind to PS via a pocket made by three basic residues (R11, R18, K20) and via the backbone nitrogen atoms of T14, I15 and L16[45, 49]. Analysis of membrane interacting residues, however, indicates no particular preference, and the model does not identify the correct binding interface (Fig. 4C,D).
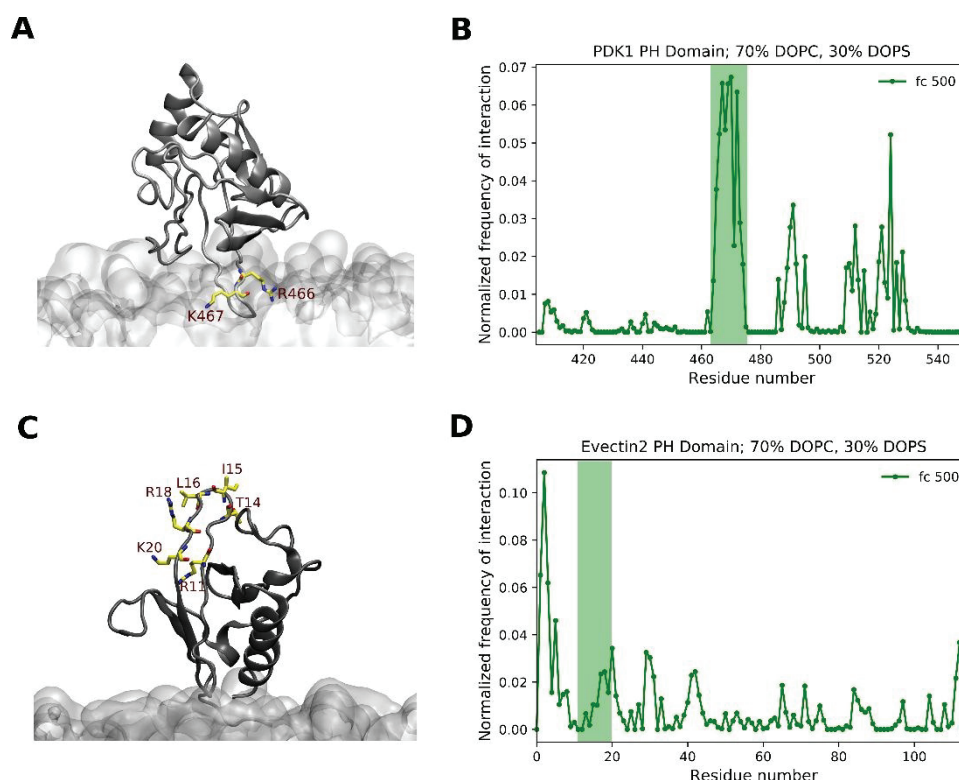
Figure 4: **Protein-membrane interface of PH domains.** (A,C) Representative membrane binding mode for the PH domain of PDK1 (A) and Evectin-2 (C). Experimentally determined residues responsible for membrane binding are shown explicitly in licorice representation. (B,D) Normalized frequency of contacts for protein residues for membranes with 30% PS. The shaded regions represent the experimentally observed binding regions.

Next, we performed similar analyses for the C2 domains of Lactadherin (Fig. 5A) and FVa (Fig. 5D). Unlike for PH domains, in our simulations, the C2 domains displayed membrane binding despite the absence of PS lipids (Fig. 5B,E), and PS lipids did not significantly increase membrane binding (Fig. 5C,F), unlike in experimental observations[45, 46].
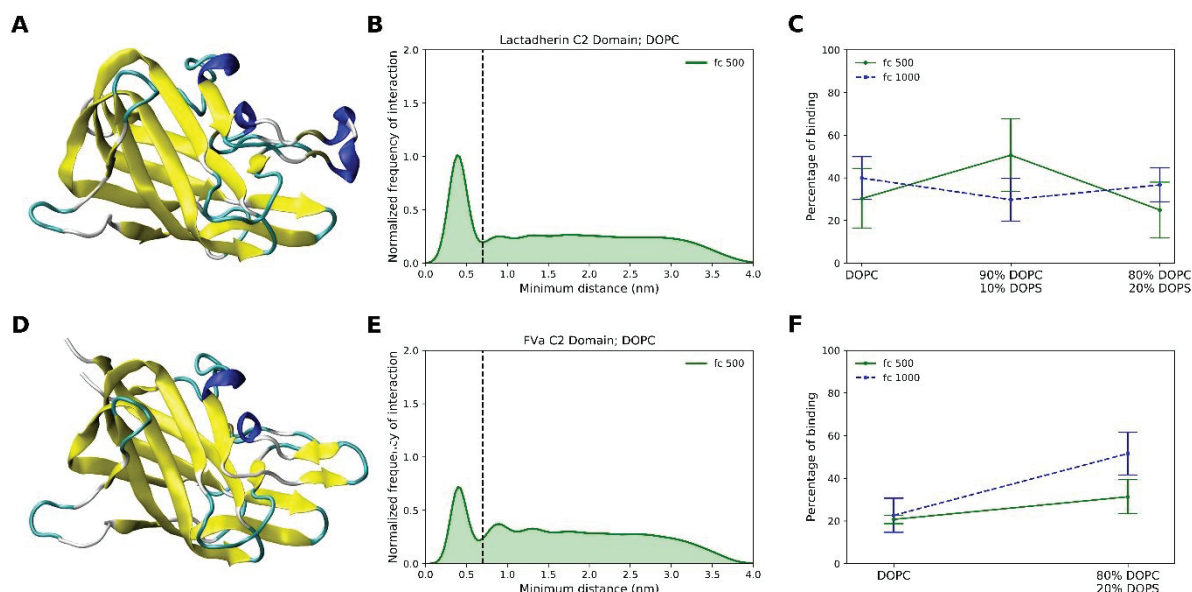
Figure 5: **Binding of C2 domains to lipid bilayers.** (A,D) Representative structure of the C2 domain of lactadherin (A) and FVa (D). (B,E) Probability density distribution of protein-membrane minimum distances in PC lipid bilayer. (C,F) Percentage of binding of the two proteins at different membrane compositions. Addition of PS to the bilayer did not enhance binding significantly for both the proteins.

In both cases, however, analysis of the membrane binding interface (Fig. 6A,C) , reveals that the model is capable of correctly identifying the experimentally known membrane binding interface for both the C2 domains (Fig. 6B,D). In detail, the lactadherin C2 domain is predicted to interact with membranes via its W26, G27, L28, F31 and F81 regions[47] while the FVa domain is predicted to do so via the residues K23, W26, W27, Q48, S78 and L79[48].
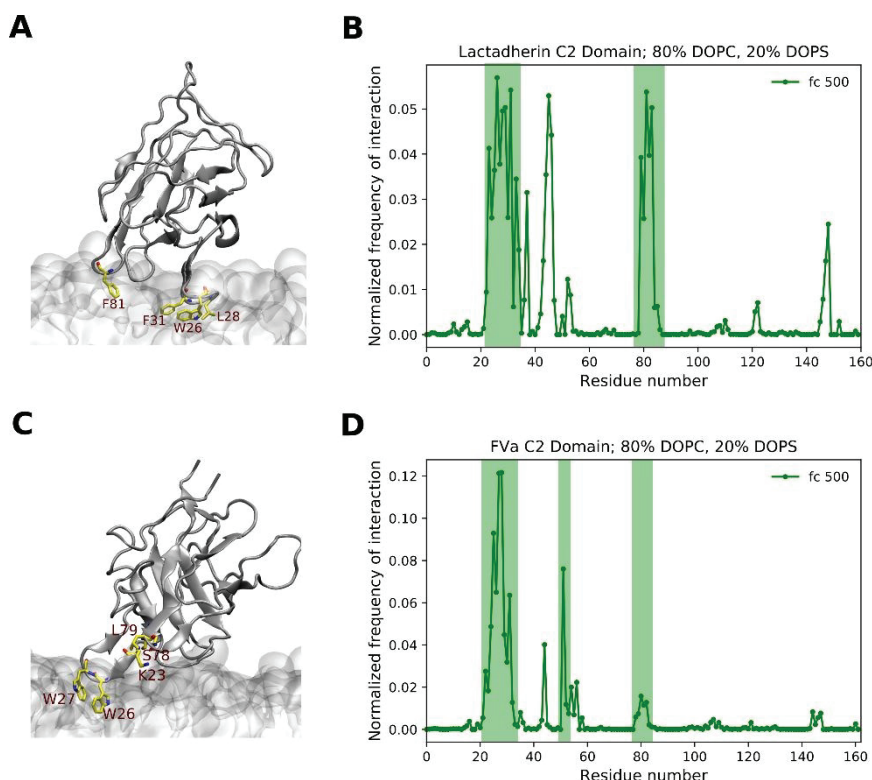
Figure 6: **Protein-membrane interface of C2 domains**. (A,C) Representative membrane binding mode for the C2 domain of lactadherin (A) and FVa (C). Experimentally determined residues responsible for membrane binding are shown explicitly in licorice representation. (B,D) Normalized frequency of contacts for protein residues with membranes with 20% PS. The shaded regions represent the experimentally observed binding regions.

Next, we investigated the binding of the enzyme Phospholipase A_2 (PLA$_2$) (Fig. 7A) to lipid bilayers consisting of PC-only or PC-PS mixtures (Fig. 7). The membrane binding of PLA$_2$ is a particularly challenging test case for the MARTINI force field, as it has been shown to depend on cation-pi interaction between lipid polar heads and protein aromatic residues[50, 51, 52]. We found that our model is not able to reproduce the experimentally observed binding to pure PC bilayers, regardless of acyl chain composition (Fig. 7B and S6). On the other hand, adding PS lipids significantly increases binding (Fig. 5C), particularly via the binding interface proposed via a combination of atomistic and experimental assays using single point mutations, involving several aromatic residues (Y3, W18, W19, W61, F64, Y110)[50, 53] (Fig. 7D,E).
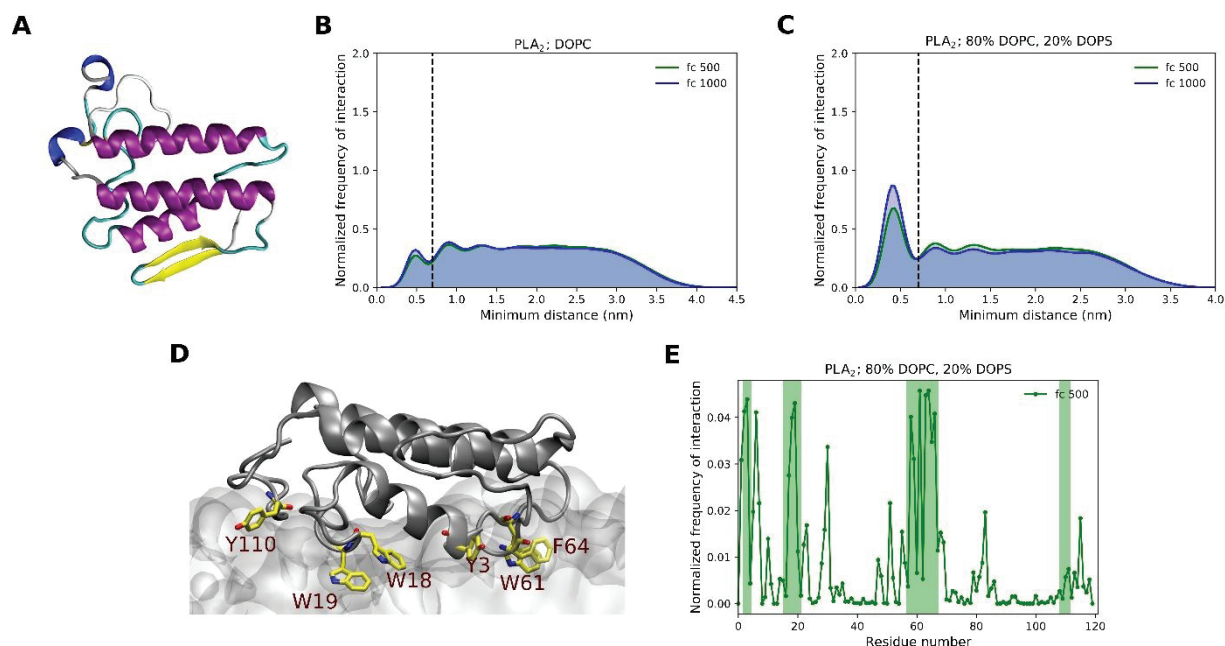
14

Figure 7: **Binding of PLA$_2$ to lipid bilayers.** (A) Representative structure of PLA$_2$. (B,C) Probability density distribution of protein-membrane minimum distances in PC, and PC/PS lipid bilayers . (D) Representative membrane binding mode for PLA$_2$. Experimentally determined residues responsible for membrane binding are shown explicitly in licorice representation. (E) Normalized frequency of contacts for protein residues for membranes with 20% PS. The shaded regions represent the experimentally observed binding regions.

In summary, our data suggest that the model is able, with the sole exception of the PH domain of Evectin-2, to identify most of the residues that have been experimentally identified to drive membrane association. On the other hand, the model is generally not well-suited to identify the correct membrane conditions for binding, including the experimentally observed dependency of membrane binding upon increase in membrane PS levels for the PH and C2 domains tested in this study, with the exception of that of PDK1. In all cases, varying elastic network parameters has no significant effect on the observed binding behavior of globular peripheral proteins.

Peripheral proteins adopting different conformations.

We next investigated whether this approach is able to discriminate between conformations that have been experimentally shown to promote or prevent membrane

15

binding. To test this approach, we first focused on the protein Arf1, which has been shown to bind to membranes upon GDP-to-GTP exchange via its N-terminal amphipathic helix (AH)[36]. As both GDP- and GTP- bound conformations are available in the Protein Data Bank (Fig. 8A,D)[54, 55], we tested their binding to pure DOPC bilayers. In agreement with experimental observations[55], we found that the GDP-bound conformer of the protein does not bind to DOPC bilayers (Fig. 8B), while the GTP-bound conformer does (Fig. 8E). Analysis of the protein residues that interact with the lipid bilayer in the GTP-bound form, correctly identified the N-terminal AH as the main culprit for membrane binding (Fig. 8F). Again, no major effect of the elastic network parameters was observed (Fig. 8B,E).
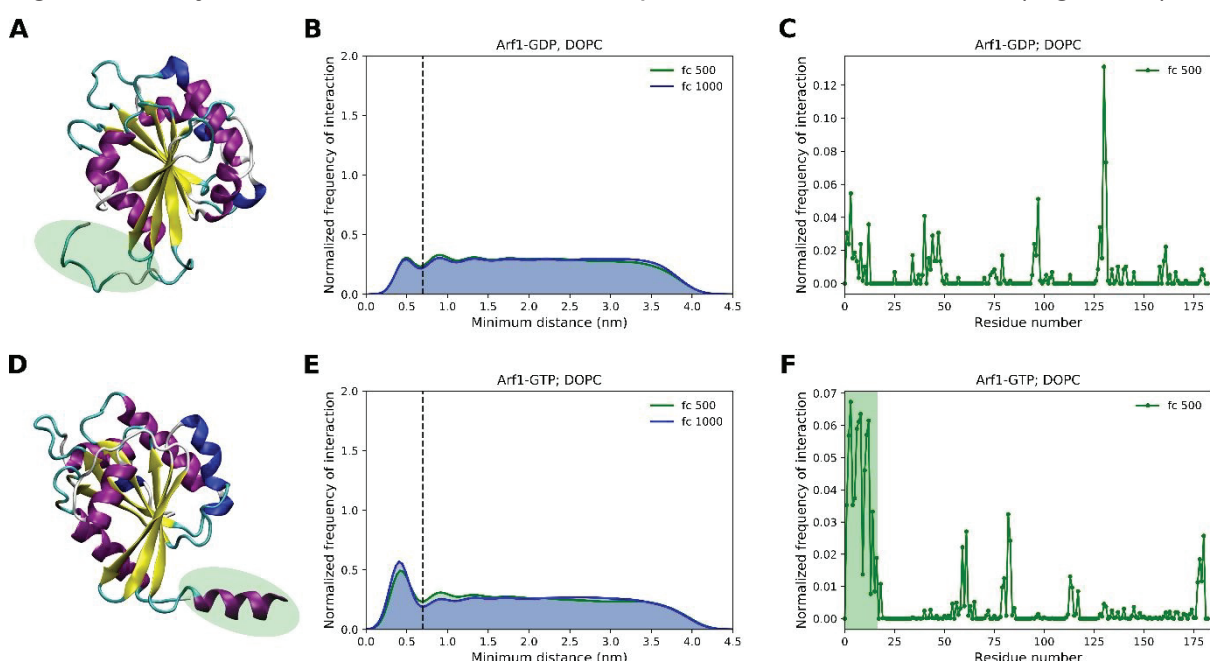

Figure 8: **Binding of Arf1 to lipid bilayers.** (A,D) Representative structure of the Arf1 protein when bound to GDP (A) and GTP (D). GDP and GTP are omitted for clarity. The shaded region indicates the difference in the two structures at the N-terminus. (B,E) Probability density distribution of protein-membrane minimum distances in PC lipid bilayer. (C,F) Normalized frequency of contacts for protein residues for membranes with PC lipid bilayer. The shaded region represents the experimentally observed binding region.

Next, we investigated whether our protocol is able to reproduce, at least at the qualitative level, membrane binding events in which protein conformational changes have been proposed to occur. As a test system, we studied the Ricin A-chain (RTA) (Fig. 9A), a cytotoxic protein known to bind to PS containing vesicles, possibly upon binding-driven conformational changes[56, 57]. In agreement with experimental observations[56], we

observed significant membrane binding in mixed PC/PS bilayers (Fig. 9B), but almost no binding of ricin to pure PC bilayers (Fig. 9C). Interestingly membrane binding in presence of PS lipids was larger when using a low fc of 500 kJ mol^{-1} nm^{-2} (Fig. 9B). On the other hand, when a high fc of 1000 kJ mol^{-1} nm^{-2} was used, the protein could initiate binding with the lipid bilayers, but this did not result into stable binding events (Fig. S8). Analysis of the membrane binding residues for the simulations with PS-containing bilayers revealed that the binding interface is mostly consistent with what has been experimentally determined[56] (Fig. 9D,E).
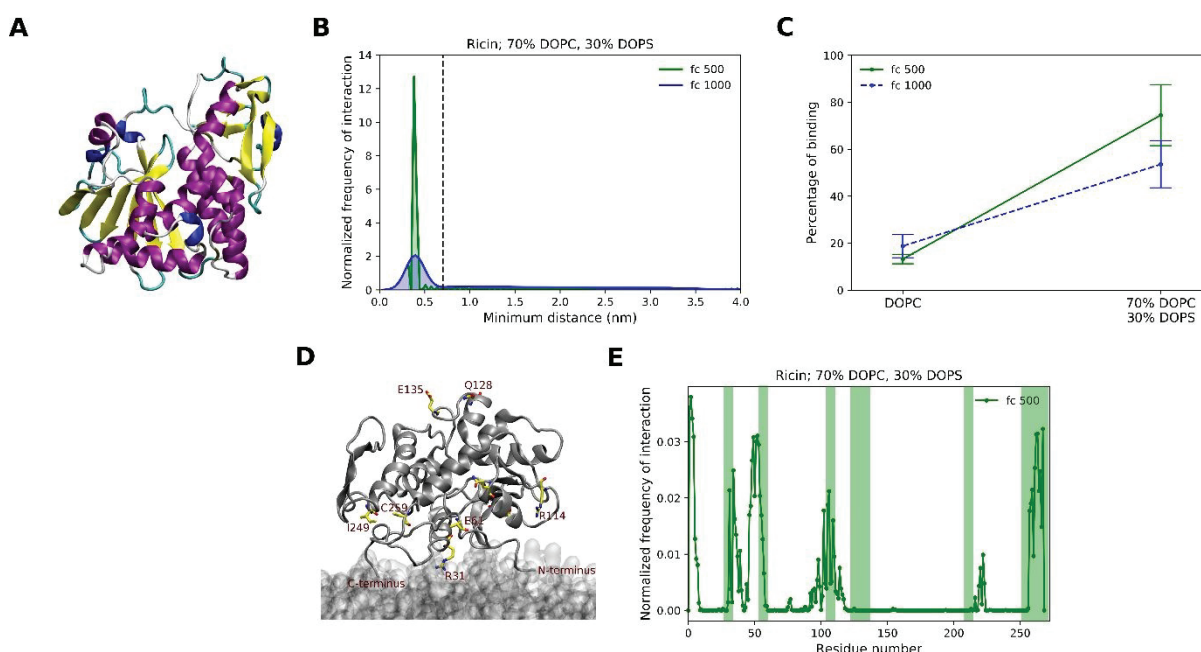

Figure 9: **Ricin binding to lipid bilayers.** (A) Representative structure of ricin. (B) Probability density distribution of protein-membrane minimum distances in PC/PS lipid bilayers at two values of fc. (C) Percentage of binding of ricin at different membrane compositions. Addition of PS to the bilayer enhanced binding significantly, especially for fc = 500 kJ mol^{-1} nm^{-2}. (D) Representative membrane binding mode of ricin. Experimentally determined residues responsible for membrane binding are shown explicitly in licorice representation. (E) Normalized frequency of contacts for protein residues for membranes with 30% PS. The shaded regions represent the experimentally observed binding region.

Amphipathic helices.

Finally, we investigated the ability of the model to reproduce the binding of AHs to lipid bilayers. AHs are unfolded in solution and adopt a helical conformation upon membrane binding, posing thus a unique challenge to our modeling strategy that requires the *a priori*

determination of the protein secondary structure. As a consequence, modeling them in their folded, membrane-bound, helical conformation could also intrinsically favor membrane binding.

To investigate membrane binding of AHs to lipid bilayers we studied the binding of three distinct peptides, that of mastoparan (Fig. 10A) - an antimicrobial peptide known to stimulate the release of histamine from mast cells[58], and the Amphipathic Lipid Packing Sensor (ALPS) motifs of two distinct proteins, the lipid transporter Osh4 (Fig. 10B) and the GTPase-activating protein (GAP) for Arf1 (ArfGAP1)[59] (Fig. 10C). We studied the binding of these three peptides to bilayers composed of PC lipids with different acyl chains, that were shown to mimic an increase in lipid-packing defects[60], a membrane property that ALPS motifs are sensitive to[30, 59], and that substantially increases going from saturated DMPC lipids to doubly unsaturated dioleoyl chains[30, 61].

Notably, the membrane binding behavior of mastoparan is affected by the choice of the elastic network parameters (Fig. 10D). In particular, using a fc of 500 kJ mol^{-1} nm^{-2}, the peptide shows a consistent binding to all lipid compositions, in agreement with experimental observations[62]. On the other hand, using a fc of 1000 kJ mol^{-1} nm^{-2}, the protein appears to exhibit sensitivity to lipid packing defects (Fig. 7A,D), even showing almost no binding in the presence of fully saturated DMPC lipids (Fig. 7A).

On the other hand, the ALPS motif of Osh4 displays the experimentally-observed sensitivity for lipid packing defects regardless of the force constant used (Fig. 10E). Of note, this behavior is enhanced when using an fc of 500 kJ mol^{-1} nm^{-2}.

The ALPS motif of ArfGAP1, on the other hand, displays a stable irreversible binding in all simulations, in disagreement with the experimentally observed sensitivity for lipid-packing defects[30]. We attribute this negative result, in part, to the incorrect modeling of the ArfGAP1 AH as a fully folded helix, in disagreement with circular dichroism experiments showing a helical content of approximately 45% for the membrane bound conformation[63]. This also highlights a limitation of our protocol, which is able to discriminate between different proteins and/or membrane conditions only when multiple binding/unbinding events are observed in the MD simulations.

Taken together, these results suggest that the model is able, at least in some instances, to reproduce important properties of the binding of AHs to model membranes, including

their sensitivity to lipid-packing. At the same time, special care must be taken, as the choice of the protein conformation and of the elastic network force constant have a strong influence on the final results. Using a low fc of 500 kJ mol^{-1} nm^{-2} appears to provide a better agreement with experimental observations.
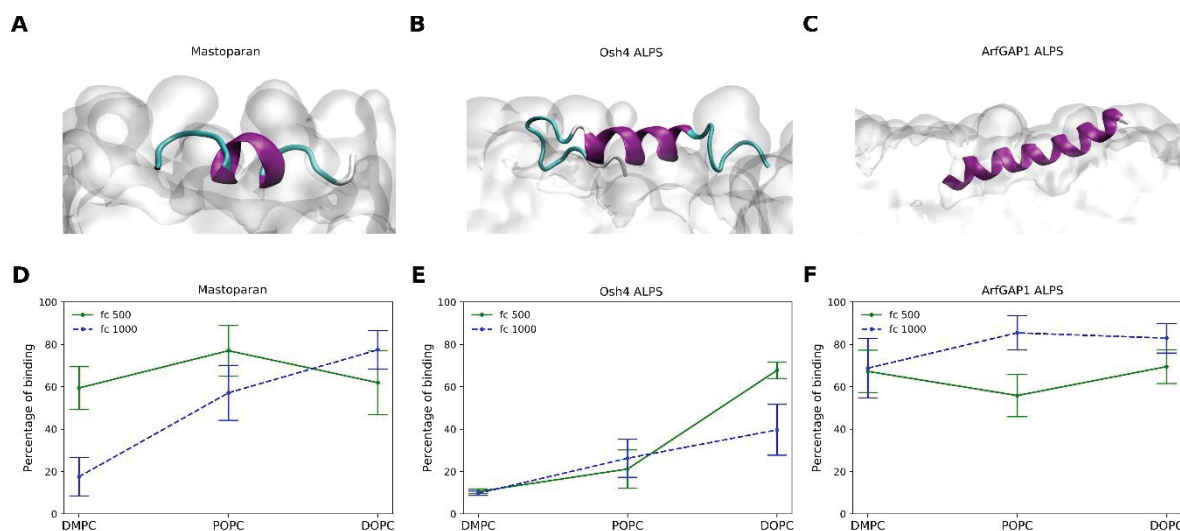

Figure 10: **AH binding to lipid bilayers**. (A-C) Representative structures of the three AHs investigated in this study – mastoparan (A), ALPS motif of Osh4 (B), and ALPS motif of ArfGAP1 (C).(D-F) Percentage of binding of the three AHs at different membrane compositions.

DISCUSSION

In this Faraday Discussion on Peptide-membrane interactions, we investigated the ability of a widely-used CG model, the MARTINI force field, and more precisely its 3 open-beta release, to accurately describe the binding of peripheral proteins to model membranes. To this end, we have developed a protocol to characterize membrane binding from multiple replicas of unbiased MD trajectories for proteins and peptides belonging to different families and displaying different membrane binding mechanisms. Despite our

limited dataset, our simulations allow us to draw some preliminary general conclusions about the performance of this model.

First, the model does not appear to overestimate protein-membrane binding affinity, as: 1. for most of the proteins in our dataset we observed multiple binding/unbinding events within the timescales of our simulations (microseconds), and 2. the model is able to correctly identify soluble proteins that do not bind to membranes, as we did not observe membrane binding for the two soluble proteins we selected as negative controls: hen egg white lysozyme and ubiquitin.

Second, the observed on/off behavior is paramount for the ability of our protocol to provide qualitative information on the binding process, and to quantify variations in membrane binding as a result of changes in membrane properties. For example, it allows for the investigation of protein sensitivity to specific membrane properties, such as electrostatics or lipid-packing defects. To this extent, we could correctly characterize the experimentally-observed sensitivity for membrane properties, for example that of the PH domains of PDK1 to PS lipids[44], and that of the ALPS motif of Osh4 for lipid-packing defects[59]. On the other hand, we could not accurately reproduce the membrane sensitivity of different proteins, for example that of C2 domains for PS lipids, or that of ArfGAP1 for lipid packing defects. Interestingly, the membrane-binding interface of the C2 domains tested in this study is largely composed of aromatic residues, while that of PH domains is largely composed of positively-charged residues. It is thus possible that the negative results for C2 domains might originate from the previously established underestimation of aromatic residues-lipid interactions[52] of the MARTINI force field, an observation that could also explain why we could not observe the binding of PLA_2 to PC lipid bilayers that has been reported experimentally[50]. Taken together, these observations suggest that slight under- or over-estimation of membrane binding affinity is sufficient to severely limit the ability of the model to correctly predict the ability to sense specific membrane properties displayed by some proteins.

Lastly, our protocol succeeds in correctly characterizing the protein-membrane interface for all but one protein in our dataset, as we were mostly able to identify the key residues found to disrupt membrane binding upon mutagenesis experiments. Since our approach lacks the ability to investigate protein conformational changes, it is possible that negative

results might stem from the use of the crystal structure as the starting point, which might not represent the correct membrane-interacting conformation. Overall, however, when multiple independent binding events are observed, our findings indicate that MARTINI 3 can be used as a predictive tool to infer the protein-membrane interface and to rationally prioritize mutagenesis experiments for validation of the *in silico* results.

From a technical point of view, we found that varying the force constant in the elastic network model used to restrain the secondary structure of the protein has only minimal impact for folded globular proteins, but it becomes a relevant parameter in those cases where protein conformational flexibility plays a key role in membrane binding, as is the case, for example, for amphipathic helices. As the extent of conformational plasticity during binding is not generally known *a priori*, we suggest, as a practical rule, to keep the force constant low, to prevent unintentional biasing of the membrane binding affinity.

In summary, our preliminary investigations suggest that transferable chemical-specific CG force fields have enormous potential towards the characterization of structural properties of the membrane binding process by peripheral proteins, even if room for improvement remains. We foresee that future investigations in this direction will shed further light into the capabilities and limitations of these approaches, possibly becoming part of the development and testing strategy for future releases of MARTINI or other CG force fields.

Conflicts of interest

There are no conflicts of interest to declare.

REFERENCES

1. A. Chiapparino, K. Maeda, D. Turei, J. Saez-Rodriguez and A.-C. Gavin, *Progress in Lipid Research*, 2016, **61**, 30-39.
2. L. H. Wong, A. Copic and T. P. Levine, *Trends in biochemical sciences*, 2017, **42**, 516-530.
3. L. H. Wong, A. T. Gatta and T. P. Levine, *Nature Reviews Molecular Cell Biology*, 2019, **20**, 85-101.
4. W. Cho and R. V. Stahelin, *Annual Review of Biophysics and Biomolecular Structure*, 2005, **34**, 119-151.
5. T. F. J. Martin, *Annual Review of Cell and Developmental Biology*, 1998, **14**, 231-264.
6. J. E. Johnson and R. B. Cornell, *Molecular Membrane Biology*, 1999, **16**, 217-235.
7. B. Antonny, *Annual Review of Biochemistry*, 2011, **80**, 101-123.
8. G. Enkavi, M. Javanainen, W. Kulig, T. Róg and I. Vattulainen, *Chemical Reviews*, 2019, **119**, 5607-5774.
9. H. I. Ingólfsson, C. Arnarez, X. Periole and S. J. Marrink, *Journal of Cell Science*, 2016, **129**, 257.
10. T. A. Soares, S. Vanni, G. Milano and M. Cascella, *The Journal of Physical Chemistry Letters*, 2017, **8**, 3586-3594.
11. S. J. Marrink, H. J. Risselada, S. Yefimov, D. P. Tieleman and A. H. de Vries, *The Journal of Physical Chemistry B*, 2007, **111**, 7812-7824.
12. D. H. de Jong, G. Singh, W. F. D. Bennett, C. Arnarez, T. A. Wassenaar, L. V. Schäfer, X. Periole, D. P. Tieleman and S. J. Marrink, *Journal of Chemical Theory and Computation*, 2013, **9**, 687-697.
13. L. Monticelli, S. K. Kandasamy, X. Periole, R. G. Larson, D. P. Tieleman and S.-J. Marrink, *Journal of Chemical Theory and Computation*, 2008, **4**, 819-834.
14. S. J. Marrink, A. H. de Vries and A. E. Mark, *The Journal of Physical Chemistry B*, 2004, **108**, 750-760.
15. J. J. Uusitalo, H. I. Ingólfsson, P. Akhshi, D. P. Tieleman and S. J. Marrink, *Journal of Chemical Theory and Computation*, 2015, **11**, 3932-3945.
16. J. J. Uusitalo, H. I. Ingólfsson, S. J. Marrink and I. Faustino, *Biophysical Journal*, 2017, **113**, 246-256.
17. C. A. López, A. J. Rzepiela, A. H. de Vries, L. Dijkhuizen, P. H. Hünenberger and S. J. Marrink, *Journal of Chemical Theory and Computation*, 2009, **5**, 3195-3210.
18. R. Friedman, S. Khalid, C. Aponte-Santamaría, E. Arutyunova, M. Becker, K. J. Boyd, M. Christensen, J. T. S. Coimbra, S. Concilio, C. Daday, F. J. van Eerden, P. A. Fernandes, F. Gräter, D. Hakobyan, A. Heuer, K. Karathanou, F. Keller, M. J. Lemieux, S. J. Marrink, E. R. May, A. Mazumdar, R. Naftalin, M. Pickholz, S. Piotto, P. Pohl, P. Quinn, M. J. Ramos, B. Schiøtt, D. Sengupta, L. Sessa, S. Vanni, T. Zeppelin, V. Zoni, A.-N. Bondar and C. Domene, *The Journal of Membrane Biology*, 2018, **251**, 609-631.
19. S. J. Marrink and D. P. Tieleman, *Chemical Society Reviews*, 2013, **42**, 6801-6822.
20. E. Yamamoto, A. C. Kalli, T. Akimoto, K. Yasuoka and M. S. P. Sansom, *Scientific Reports*, 2015, **5**, 18245.
21. E. Yamamoto, J. Domański, F. B. Naughton, R. B. Best, A. C. Kalli, P. J. Stansfeld and M. S. P. Sansom, *Science Advances*, 2020, **6**, eaay5736.
22. L. Picas, J. Viaud, K. Schauer, S. Vanni, K. Hnia, V. Fraisier, A. Roux, P. Bassereau, F. Gaits-Iacovoni, B. Payrastre, J. Laporte, J.-B. Manneville and B. Goud, *Nature Communications*, 2014, **5**, 5647.

23. F. B. Naughton, A. C. Kalli and M. S. P. Sansom, *The Journal of Physical Chemistry Letters*, 2016, **7**, 1219-1224.
24. A. Buyan, A. C. Kalli and M. S. P. Sansom, *PLOS Computational Biology*, 2016, **12**, e1005028.
25. H. Li and A. A. Gorfe, *PLoS One*, 2013, **8**, e71018.
26. E. Jefferys, M. S. P. Sansom and P. W. Fowler, *Faraday Discussions*, 2014, **169**, 209-223.
27. D. Murray, A. Arbuzova, B. Honig and S. McLaughlint, in *Current Topics in Membranes,* Academic Press, 2002, vol. 52, pp. 277-307.
28. B. Antonny, I. Huber, S. Paris, M. Chabre and D. Cassel, *Journal of Biological Chemistry*, 1997, **272**, 30848-30851.
29. J. Bigay and B. Antonny, *Developmental Cell*, 2012, **23**, 886-895.
30. S. Vanni, H. Hirose, H. Barelli, B. Antonny and R. Gautier, *Nature Communications*, 2014, **5**, 4916.
31. M. Cascella and S. Vanni, in *Chemical Modelling: Volume 12*, The Royal Society of Chemistry, 2016, vol. 12, pp. 1-52.
32. R. Alessandri, P. C. T. Souza, S. Thallmair, M. N. Melo, A. H. de Vries and S. J. Marrink, *Journal of Chemical Theory and Computation*, 2019, **15**, 5448-5460.
33. D. Van Der Spoel, E. Lindahl, B. Hess, G. Groenhof, A. E. Mark and H. J. C. Berendsen, *Journal of computational chemistry*, 2005, **26**, 1701-1718.
34. W. Humphrey, A. Dalke and K. Schulten, *Journal of Molecular Graphics*, 1996, **14**, 33-38.
35. H. M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov and P. E. Bourne, *Nucleic acids research*, 2000, **28**, 235-242.
36. B. Antonny, S. Beraud-Dufour, P. Chardin and M. Chabre, *Biochemistry*, 1997, **36**, 4675-4684.
37. E. L. Wu, X. Cheng, S. Jo, H. Rui, K. C. Song, E. M. Davila-Contreras, Y. Qi, J. Lee, V. Monje-Galvan, R. M. Venable, J. B. Klauda and W. Im, *Journal of computational chemistry*, 2014, **35**, 1997-2004.
38. G. Bussi, D. Donadio and M. Parrinello, *The Journal of Chemical Physics*, 2007, **126**, 014101.
39. M. Parrinello and A. Rahman, *Journal of Applied Physics*, 1981, **52**, 7182-7190.
40. D. H. de Jong, S. Baoukina, H. I. Ingólfsson and S. J. Marrink, *Computer Physics Communications*, 2016, **199**, 1-7.
41. B. P. Roscoe, K. M. Thayer, K. B. Zeldovich, D. Fushman and D. N. A. Bolon, *Journal of molecular biology*, 2013, **425**, 1363-1377.
42. T. R. Butt, S. Jonnalagadda, B. P. Monia, E. J. Sternberg, J. A. Marsh, J. M. Stadel, D. J. Ecker and S. T. Crooke, *Proceedings of the National Academy of Sciences*, 1989, **86**, 2540.
43. S. B. Howard, P. J. Twigg, J. K. Baird and E. J. Meehan, *Journal of Crystal Growth*, 1988, **90**, 94-104.
44. N. Lucas and W. Cho, *The Journal of biological chemistry*, 2011, **286**, 41265-41272.
45. Y. Uchida, J. Hasegawa, D. Chinnapen, T. Inoue, S. Okazaki, R. Kato, S. Wakatsuki, R. Misaki, M. Koike, Y. Uchiyama, S.-i. Iemura, T. Natsume, R. Kuwahara, T. Nakagawa, K. Nishikawa, K. Mukai, E. Miyoshi, N. Taniguchi, D. Sheff, W. I. Lencer, T. Taguchi and H. Arai, *Proc Natl Acad Sci U S A*, 2011, **108**, 15846-15851.
46. D. E. Otzen, K. Blans, H. Wang, G. E. Gilbert and J. T. Rasmussen, *Biochimica et Biophysica Acta (BBA) - Biomembranes*, 2012, **1818**, 1019-1027.
47. C. Shao, V. A. Novakovic, J. F. Head, B. A. Seaton and G. E. Gilbert, *The Journal of biological chemistry*, 2008, **283**, 7230-7241.

48. S. Macedo-Ribeiro, W. Bode, R. Huber, M. A. Quinn-Allen, S. W. Kim, T. L. Ortel, G. P. Bourenkov, H. D. Bartunik, M. T. Stubbs, W. H. Kane and P. Fuentes-Prior, *Nature*, 1999, **402**, 434-439.
49. S. Okazaki, R. Kato, Y. Uchida, T. Taguchi, H. Arai and S. Wakatsuki, *Acta Crystallographica Section D*, 2012, **68**, 117-123.
50. M. Sumandea, S. Das, C. Sumandea and W. Cho, *Biochemistry*, 2000, **39**, 4206-4206.
51. R. V. Stahelin and W. Cho, *Biochemistry*, 2001, **40**, 4672-4678.
52. H. M. Khan, P. C. T. Souza, S. Thallmair, J. Barnoud, A. H. de Vries, S. J. Marrink and N. Reuter, *Journal of Chemical Theory and Computation*, 2020, **16**, 2550-2560.
53. Q. Waheed, H. M. Khan, T. He, M. Roberts, A. Gershenson and N. Reuter, *The Journal of Physical Chemistry Letters*, 2019, **10**, 3972-3977.
54. Y. Liu, R. A. Kahn and J. H. Prestegard, *Nature Structural & Molecular Biology*, 2010, **17**, 876-881.
55. Y. Liu, R. A. Kahn and J. H. Prestegard, *Structure*, 2009, **17**, 79-87.
56. P. U. Mayerhofer, J. P. Cook, J. Wahlman, T. T. J. Pinheiro, K. A. H. Moore, J. M. Lord, A. E. Johnson and L. M. Roberts, *Journal of Biological Chemistry*, 2009, **284**, 10232-10242.
57. T. S. Ramalingam, P. K. Das and S. K. Podder, *Biochemistry*, 1994, **33**, 12247-12254.
58. A. P. Watt, *InflammoPharmacology*, 2001, **9**, 421-434.
59. G. Drin, J.-F. Casella, R. Gautier, T. Boehmer, T. U. Schwartz and B. Antonny, *Nature Structural & Molecular Biology*, 2007, **14**, 138-146.
60. S. Vanni, L. Riccardi, G. Palermo and M. De Vivo, *Accounts of Chemical Research*, 2019, **52**, 3087-3096.
61. S. Vanni, L. Vamparys, R. Gautier, G. Drin, C. Etchebest, Patrick F. J. Fuchs and B. Antonny, *Biophysical Journal*, 2013, **104**, 575-584.
62. J. A. Whiles, R. Brasseur, K. J. Glover, G. Melacini, E. A. Komives and R. R. Vold, *Biophysical Journal*, 2001, **80**, 280-293.
63. P. Gonzalez-Rubio, R. Gautier, C. Etchebest and P. F. Fuchs, *Biochimica et biophysica acta*, 2011, **1808**, 2119-2127.