Martini 3 force field parameters for protein lipidation post-translational modifications

Panagiotis I. Koukos[1], Sepehr Dehghani-Ghahnaviyeh[2], Camilo Velez-Vega[2], John Manchester[2], D. Peter Tieleman[3, 4], José S. Duca[2,], Paulo C. T. Souza[5,6,*], Zoe Cournia[1,*]*

1: Biomedical Research Foundation, Academy of Athens, 4 Soranou Ephessiou, 11527 Athens, Greece

2: Computer-Aided Drug Discovery, Global Discovery Chemistry, Novartis Institutes for BioMedical Research, 181 Massachusetts Avenue, Cambridge, Massachusetts 02139, United States

3: Department of Biological Sciences, University of Calgary, Calgary, Alberta, Canada

4: Centre for Molecular Simulation, University of Calgary, Calgary, Alberta, Canada

5: Molecular Microbiology and Structural Biochemistry, (MMSB, UMR 5086), CNRS & University of Lyon, Lyon, France

6: Laboratory of Biology and Modeling of the Cell, École Normale Supérieure de Lyon, Université Claude Bernard Lyon 1, CNRS UMR 5239 and Inserm U1293, 46 Allée d'Italie, 69007, Lyon, France

*: All correspondence should be directed to these authors

ABSTRACT

Protein lipidations are vital co/post-translational modifications that tether lipid tails to specific protein amino acids, allowing them to anchor to biological membranes, switch their subcellular localization, and modulate association with other proteins. Such lipidations are thus crucial for multiple biological processes including signal transduction, protein trafficking and membrane localization, and are implicated in various diseases as well. Examples of lipid-anchored proteins include the Ras family of proteins that undergo farnesylation; actin and gelsolin that are myristoylated; phospholipase D, which is palmitoylated; glycosylphosphatidylinositol-anchored proteins and others.

Here, we develop parameters for cysteine-targeting farnesylation, geranylgeranylation and palmitoylation, as well as glycine-targeting myristoylation for the latest version of the Martini 3 coarse-grained force field. The parameters are developed using the CHARMM36m all-atom force field parameters as reference. The behavior of the coarse-grained models is consistent with that of the all-atom force field for all lipidations and reproduces key dynamical and structural features of lipid-anchored peptides such as solvent-accessible surface area, bilayer penetration depth, and representative conformations of the anchors. The parameters are also validated in simulations of the lipid-anchored peripheral membrane proteins Rheb and Arf1, after comparison with independent all-atom simulations. The parameters, along with mapping schemes for the popular *martinize2* tool, are available for download at https://doi.org/10.5281/zenodo.7849262 and also as supporting information.

INTRODUCTION

Proteins that interact with cellular membranes are classified in two categories depending on the mechanism of interaction: integral and peripheral membrane proteins[1]. Integral membrane proteins embed partially or entirely in the lipid bilayer and can span its entire width once or multiple times, thus becoming an integral part of the bilayer itself.[2] Peripheral membrane proteins on the other hand, are proteins that transiently associate with the membrane through multiple mechanisms. One of the most common mechanisms is through favorable electrostatic interactions between a positively charged protein patch and the negatively charged lipid heads of the cytosolic part of the plasma membrane.[3–5] In some cases, instead of a positively charged protein patch (often called a polybasic domain), a cation interacts with the anionic lipid heads and tethers the protein to the membrane.[6] In addition to electrostatically mediated interactions, proteins also interact with the bilayer through hydrophobic segments such as loops or helices that partially or entirely embed themselves in the lipid core of the bilayer, thus anchoring the protein to the bilayer.[7–10] Yet another mechanism is through Post-Translational Modifications (PTMs), which covalently link a lipid tail to one or more protein residues, which are then embedded in the membrane as a result of their increased hydrophobicity.[1,11,12] These PTMs are more specifically labelled lipidations. Some of most prominent ones are the processes of prenylation, palmitoylation and myristoylation, which are the focus of the present work.

Prenylation is a process during which isoprenoid lipids are covalently attached to the side chain of a cysteine residue through a thioether bond. Depending on the length of the lipid tail being added to the cysteine residue, the prenylation process can be further subdivided into farnesylation (addition of a farnesyl group consisting of 15 carbons, three repeating units of 2-methyl-2-butene) or geranylgeranylation (addition of a geranylgeranyl group consisting of 20 carbons, 4 repeating units of 2-methyl-2-butene)[13]. The prenylation motif is well characterized

and consists of the sequence CAAX located at the C terminus of a protein, with the C corresponding to the cysteine, which is the prenylation target, the A to any aliphatic residue and residue X (mostly) determining whether the cysteine is farnesylated or geranylgeranylated with distinct residues shifting the probability toward one or the other PTM[14]. Palmitoylation is an S-acylation process during which palmitic acid is linked to a cysteine side chain through a thioester bond[15]. Unlike the prenylation process, there is no consensus target sequence for the process of palmitoylation. Myristoylation is also an acylation process although one that targets the nitrogen atom of N-terminal glycine residues (N-acylation) to which it adds a myristoyl group irreversibly.[16]

Expanding into details regarding lipidations is beyond the scope of this manuscript, as it has been the subject of multiple recent reviews,[17–19] but it is important to establish their relevance. These lipidation PTMs are all part of (or implicated in) many cellular processes of high significance. Prenylation is associated with multiple biological processes such as protein trafficking, activity, and stability[14,15]. For example, KRas-4B, the prevalent splice variant of the KRAS gene, and the kinetochore proteins CENP-E and CENP-F are farnesylated at their C-terminal cysteine.[20,21] Unlike KRas-4B, other members of the Ras family such as NRas and HRas are palmitoylated near the C-terminus.[22] The Ras family of proteins is a prime drug development target, due to the role that dysregulation of Ras proteins plays in cancer. A recent analysis of genomic data from samples originating in the United States concluded that almost one in five human cancers (15 %) can be traced back to a Ras-family mutation[23]. Palmitoylation and myristoylation are also associated with many of the same biological processes as prenylation such as membrane targeting or stability[15], but also additional ones such as signal transduction (for myristoylation)[16].

However, the study of membrane proteins, where these lipidations are encountered, remains fraught with difficulties for experimental approaches, especially with regard to the native cell

context.[24–26] Computational approaches, and in particular all-atom Molecular Dynamics (MD) simulations, present an alternative that enables the study of the structure and dynamics of large biomolecular assemblies in their native context, at a fraction of the cost required by experimental approaches. Often enough, however, even with access to High-Performance Computing (HPC) facilities the size and complexity of membrane-associated systems is such that the study of such systems remains well out of reach for most practitioners, at least for simulation timescales that would render robust and well-converged findings. In this context, a number of studies have used coarse-grained (CG) MD approaches. CG force fields speed up MD simulations by reducing the degrees of freedom of the system by grouping together atoms into beads, which then behave as one particle thus reducing the number of interactions and therefore the processing capabilities required to perform simulations. One of the most popular force fields for CG MD simulations of membrane-associated systems is the Martini force field.[27–34]

Atsmon-Raz and Tieleman[35] developed parameters for the CG Martini force field (v2.2) for protein lipidation post-translational modifications in 2017[27]. However, these parameters are not available in the latest version of the Martini force field (v3.0).[36] In Martini 3, significant improvements include a substantially expanded chemical universe with new bead types and a more balanced parameterization that is capable of reproducing a wide range of experimental data. Coarse-grained simulations have been used previously for the study of peripheral membrane proteins H- and N-Ras[37], for the dimerization of palmitoylated protein CD44 on lipid bilayers[38], the translocation and insertion of lipidated LC3 protein into lipid bilayers, the effect[39] of palmitoylation on the dimerization behavior of the human dopamine transporter[40] and the palmitoylated transmembrane peptide tLAT[41].

Here, we report new parameters for the four protein lipidation post-translational modifications of interest (farnesylation, geranylgeranylation, palmitoylation and myristoylation) for the

Martini 3 force field. The parameters have been developed based on a comparison with CHARMM36m[42] all-atom results, i.e., physicochemical, structural and dynamical properties that are crucial for the biophysics of lipid-anchored proteins. We have computed the solvent accessible surface area (SASA), isolated and analyzed cluster representative structures, and calculated the bilayer penetration depth for the PTM atoms/beads for both all-atom and CG simulations. Our results are in excellent agreement between the CG and atomistic simulations, with molecular volumes (based on SASA calculations) differing by less than 5%, representative structures of the PTMs by less than 2.5 Å (in terms of root mean square deviation) and bilayer penetration depth by less than 2.5 Å. We also validated our Martini 3 model using two lipid-anchored membrane proteins, Rheb and Arf1, which have been previously simulated in Ref.69 using atomistic force fields. We analyzed the orientational profile of the two proteins relative to the membrane bilayer, in which their lipidated side-chains are embedded, by computing the Euler angles of the protein with respect to the bilayer normal, and also compared the contact network between protein residues and bilayer lipids. The protein simulation results reveal that the lipidation parameters proposed herein are transferable to more complicated systems than the small peptides used for parameterization.

METHODS

Reference all-atom simulations

We carried out all-atom simulations for the four lipidations of interest using the protocol reported in Ref.[35] for Martini 2.2, but with GROMACS v2021.4[43,44] and the July 2021 release of CHARMM36m[42] for GROMACS. Briefly, we generated a 10 x 10 nm 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine (POPC) bilayer with 147 lipids per leaflet, using the CHARMM-GUI "Bilayer Builder" module[45–47]. We generated the starting penta-/hexa-peptide conformations ($MFC_{farn/ger/palm}IH$ / $G_{myri}MFCIH$) with PyMOL[48] (in extended conformation) and then added the different PTMs using the "PDB Reader and Manipulator" module of CHARMM-GUI[49]. We then replaced one POPC lipid from one bilayer leaflet with the side chain of the modified amino acid residue.

The systems were solvated using the CHARMM36m implementation of TIP3P[42,50] and minimized using the steepest descent algorithm for max 50000 steps. Position restraints were placed on the backbone and side chain atoms of the peptides, as well as lipid atoms, with force constants equal to 4000, 2000 and 1000 kJ/mol/nm^2, for the three atom groups, respectively. Systems were equilibrated for 1 ns in three stages: (1) 500 ps in the NVT ensemble using the same restraints as for the minimization, followed by (2) 250 ps in the NPT ensemble using the same restraints as in the previous step, but with a force constant equal to 1/10th of the previous value, and finally (3) 250 ps in the NPT ensemble without any restraints. The systems were then simulated in the NPT ensemble for production runs of 500 ns. The settings for each equilibration stage and production run are provided below.

For the NVT stage, we used a 1 fs time step, the Berendsen thermostat[51] with a coupling constant τ_T = 1 ps and temperature set to 310 K (single temperature group), which was also the temperature used for assigning velocities after minimization. For the NPT equilibration steps,

we additionally coupled the system to a semi-isotropic Berendsen barostat[51] with τ_P = 5 ps, reference pressure value set to 1 bar, and compressibility values equal to 4.5 x 10^{-5} bar^{-1}. The time step was also increased to 2 fs. For the production simulations we switched the temperature coupling scheme to the Nose-Hoover thermostat[52–54] with τ_T equal to 1 ps and temperature set to 310 K, and the pressure coupling scheme to the Parrinello-Rahman barostat[55] with τ_P equal to 5 ps, reference pressure value set to 1 bar and compressibility values equal to 4.5 x 10^{-5} bar^{-1}.

The GROMACS implementation of the Verlet[56] scheme was used for neighbor searching (updated every 20 steps) with short range electrostatic and van der Waals interactions switched off between 1 and 1.2 nm. Long range electrostatics were handled with the Particle Mesh Ewald[57,58] summation scheme with a force switch applied to the van der Waals interactions. All input files were retrieved directly from the CHARMM-GUI server. All input files are available at https://doi.org/10.5281/zenodo.7849262.

All-atom trajectories were processed with built-in GROMACS tools (`gmx trjconv`) for the removal of periodicity, centering, etc. and mapped to coarse grained coordinates and analyzed with the MDAnalysis toolkit. [59,60] All scripts can be found in the Zenodo dataset indicated above.

All-atom to coarse-grained mapping schemes

We chose to fully re-parameterize the lipidations as opposed to simply updating the parameter set previously released for Martini 2.2[35] to use the new bead types made available with the recent release of the third major version of the Martini force field[36]. The mapping schemes were designed with modularity and extensibility in mind, as seen in Figure 1. To better evaluate possible options for the branched lipid tails of the prenylation PMTs (farnesyl and geranylgeranyl), we also considered an alternative scheme for the farnesyl PTM (see Figure

S1). The main differentiating feature between the two schemes is the sharing of atoms across different beads in Figure S1. Specifically, carbon 1 of the lipid chain is shared between the sulfur-containing “C6” bead and the first lipid chain “SC4” bead. Also, carbons attached to the branching methyl groups and the subsequent ones are shared between the lipid chain “SC4” beads and the branching “TC2” beads. After evaluating this alternative mapping scheme (see “SASA” section in “Results”), we settled on the scheme of Figure 1, panel A, due to the similar performance of both schemes (see “SASA” section in “Results” and SI) and the conceptual simplicity of the mapping scheme in Figure 1 over the alternative in Figure S1.

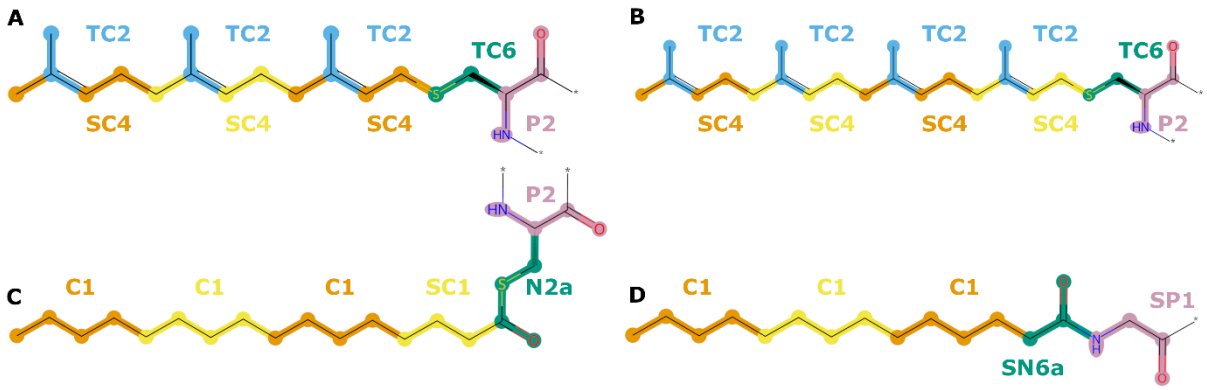


Figure 1. *Mapping schemes for the four PTMs. Unsaturated/branched prenylation PTMs are shown at the top: A. farnesylation, B. geranylgeranylation. The saturated PTMs are shown at the bottom: C. palmitoylation D. myristoylation). The colored segments indicate the beads to which the respective atoms were mapped; the type of the bead also indicated in a label of the same color.*

The bead types were chosen based on ideal (according to Martini 3 parameterization guidelines) bead assignments expected for atoms/chemical moieties grouped in each bead, although for non-obvious choices such as the thioester group of the palmitoyl PTM several tests were performed to select the optimal bead (see below). The bead sizes were also selected according the default mapping ratios for Martini 3, with four atoms mapping to a single “regular” bead (no prefix in the atom type), three atoms to a single “small” bead (any bead whose type is prefaced by “S”), and two atoms to a single “tiny” bead (any bead whose type starts with “T”),

with the intention of finding the best middle ground between resolution, accuracy (in terms of overall molecular volume and packing of aliphatic chains) and performance. For consistency with the protein model, we used the default Martini 3 bead types for backbone atoms ("P2" and "SP1", for the cysteine and glycine PTMs, respectively). We used "C1" beads for the fully saturated lipid tails of the palmitoyl and myristoyl PTMs to maintain consistency with the Martini 3 bead types for lipids.

The thioether bridges of the prenylations were modelled as "TC6" beads containing only the sulfur atom, the adjacent cysteine carbon atom (C_β) and its associated hydrogens. Doing so allowed us to isolate the repeating chemical moiety of 2-methyl-2-butene in distinct beads. The benefit of choosing this scheme becomes apparent when one examines the prenylation of the lipid tails (Figure 1). The lipid tails consist of repeating 2-methyl-2-butene groups, indicating that once parameters for one of these PTMs have been generated they can be easily transferred to other prenylation-based PTMs (or unsaturated and branched lipid tails in general), which only differ in the number of 2-methyl-2-butene units they carry in the lipid tail. We followed this strategy for the geranylgeranyl parameters, for which we transfer the parameters that we obtained for the farnesyl lipidation. Small and tiny beads were used to model the "main-" and "side-chain" carbons of the lipid tails, respectively. The presence of the branched segments and double carbon-carbon bonds led us to use higher polarity C-type beads to better reflect their chemical nature, with "SC4" and "TC2" beads used for the "main-chain" and "side-chain" atoms, respectively. "C2" type is usually the best bead assignment for branched aliphatic moieties.

The thioester group of the palmitoylation lipid tail was modelled as a "N2a" bead to better reflect the polar nature of the carbonyl and sulfur and to represent the ability of the carbonyl oxygen to act as a hydrogen bond acceptor. Because, to our knowledge, this grouping of atoms has never been mapped in the Martini 3 force field before, we investigated multiple bead types

before finalizing the bead choice. We were only able to identify computational predictions for the relevant physicochemical properties of this group[61], which guided the bead types we investigated. Specifically, we tested the "N1a/r", "N2a/r" and "N3a/r" bead types and concluded that N2a is the optimal bead choice (see "Comparison of bead types for palmitoylation PTM" section in "Results"). The carbonyl-containing bead of the myristoyl lipid tail was modelled as a SN6a to capture the polar nature of this group and to act as a hydrogen bond acceptor, similar to the palmitoylation carbonyl-containing bead. The myristoyl lipid tail was modelled as three "C1" beads because it only contains fully saturated, unbranched carbon atoms and their bonded hydrogens. For the longer palmitoyl lipid tail, we modelled carbons 1,2,3 with an "SC1" bead, to avoid creating a non-regular-sized terminal bead for the lipid tail, which could possibly lead to excessive interdigitation between the leaflets of a bilayer. This choice is also in line with the recent mapping schemes tested by Empereur-mot and co-workers.[62]

Coarse-grained simulations

Peptide simulations

We used version 3 of the Martini coarse-grained (CG) force field[36] and GROMACS v2021.4[43,44] for the CG simulations. We transformed the initial CHARMM-GUI-derived coordinates of the peptides with the modified cysteine/glycine residues to Martini 3 beads using the mappings described in the previous section (see "All-atom to coarse-grained mapping schemes" section above) using the *martinize2* tool.[63] The initial peptide structures were generated in extended conformation; we did not specify secondary structure assignments for the peptides during the coarse-graining nor did we make use of side-chain corrections, to avoid imposing restraints on the peptide motions. We generated a CG POPC bilayer with the same dimensions as the all-

atom system using the INSANE tool.[64] We then removed a single POPC lipid and replaced it with the lipid tail of the modified residue. The entire system was then solvated with INSANE.

Each system was energy-minimized for up to 1000 steps using the steepest descent algorithm, followed by a 100 ns-long NPT equilibration and a production run of 2 μs in the NPT ensemble. For the minimization and equilibration, 3D position restraints were placed on the backbone beads of the peptide, with a force constant of 1000 kJ/mol (for all dimensions). Velocities were generated and maintained at 310 K through the use of the v-rescale[65] thermostat for both equilibration and production runs. The reference pressure was set at 1 bar through the use of a semi-isotropic barostat (Berendsen with τ_P equal to 4 ps and compressibility values equal to 4.5 x 10^{-5} bar^{-1}, and Parrinello-Rahman with τ_P equal to 12 ps and compressibility values equal to 3 x 10^{-4} bar^{-1}, for the equilibration and production runs, respectively). We used the recommended settings for the Martini 3 force field non-bonded interactions, i.e., a 1.1 nm cut-off for short-range electrostatics and van der Waals interactions, a reaction field for the electrostatics with a screening constant of 15 (set to 0 beyond the short-range cut-off) and shifting the van der Waals potential to 0 at the cut-off (with the Verlet cut-off scheme).[66] A 20 fs time step was used for all CG simulations.

Protein simulations

To further validate the parameters proposed herein, we tested their performance on two peripheral membrane proteins that bear PTMs: Rheb and Arf1. Both are members of the RAS superfamily and are small GTPases that anchor to the membrane with lipidated residue side chains. Specifically, Rheb contains a C-terminal hyper variable region (HVR), which terminates at the CAAX prenylation motif with the protein being farnesylated at its terminal cysteine,[67] while Arf1 is the target of N-myristoylation at its N-terminal glycine[68]. Prakash and Gorfe recently studied[69] the membrane interactions of three proteins, namely Rheb, Arf1, and RhoA, belonging to the RAS superfamily using long-scale all-atom MD simulations. We

performed simulations for Rheb and Arf1 in the coarse-grained Martini 3 force field using our newly derived parameter set and compared the results to those of Ref.69 to determine whether findings from all-atom simulations can be reproduced using coarse-grained parameters.

We first generated an atomistic structural model for Rheb by modifying PDB entry 1XTS[70] with the HVR modelled in extended conformation with PyMOL[48]. The C-terminal cysteine was farnesylated in CHARMM-GUI, as described in Methods. No long-distance elastic network restraints were active on the Rheb HVR residues (residue indices 171-182), because they would impede the free movement of this flexible region by locking it into the starting conformation. To create an atomistic model for Arf1, we combined PDB entries 4HMY[71] for the G domain and 2KSQ[72] for the N-terminal helix. We then embedded the lipidated side chains of the respective residues for both proteins in the center of a pre-equilibrated (15 μs) POPC:POPS (80:20 % mol.) lipid bilayer, with initial dimensions equal to 12 * 12 * 12 nm. The systems were then solvated with INSANE and the system was neutralized by adding Na^+ ions. Additionally, Na^+ and Cl^- ions were added as a salt at physiological concentration (150mM) similar to Ref.69. The systems were then energy-minimized, equilibrated (1 μs) and simulated for 10 μs production runs using the same settings as for the coarse-grained runs of the peptides.

Structural analyses

Extraction of bonded parameters from all-atom simulations

To generate the initial parameters for the CG runs, we mapped the all-atom trajectories to CG by computing the Center of Geometry (CoG) for all atoms (including hydrogen atoms) belonging to each bead (see "All-atom to coarse-grained mapping schemes" section above), and then computed the distances, angles and torsions between the bonded CoGs. We used in-house code built on the MDAnalysis[59,60] toolkit for the all-atom to CG mapping, as well as the

calculation of the inter-bead bonded parameters. The mean values from the distance and angle distributions became the reference values for the respective bonded terms of the CG simulations and the width/standard deviation of each distribution determined the force constant, with narrower distributions corresponding to a higher force constant.

Historically, Martini parameters for lipid tails have only included bond length and angle terms. Since the addition of dihedral parameters for lipids is a potential addition in future versions of the Martini force field, to ensure transferability of the derived parameters, we also include dihedral terms in the CG models and perform the analyses both without and with dihedral terms (see "CG bonded parameter extraction and comparison to all-atom" section in "Results") to gauge their impact on the quality of the model as well as the stability of the simulations. It should be noted that for simplicity only one dihedral term (type 1) was included in the present parameter set, which could be revisited by including additional/alternative dihedral terms in the future, should the dihedral terms prove to be beneficial in reproducing experimental or atomistic simulation properties.

Clustering

For the clustering analysis of the all-atom trajectories, we used the GROMACS implementation of the GROMOS method[73] with a cut-off value of 0.25 nm, including only the main-chain carbon atoms and excluding methyl groups and backbone carbonyl, and amide atoms. For the clustering of CG trajectories, same method and cut-off values were used but this time only the "SC4" beads of the prenylations (Figure 1), were included. We isolated representative structures from all identified clusters as those closest to the centroid of the cluster. We carried out the clustering analysis every ten frames for an effective sampling frequency of 200 ps (except for the CG palmitoyl simulations). The sampling frequency for the CG palmitoyl simulations was

100 ps, because the initial analysis (with a sampling frequency of 200 ps) did not identify the minor population.

Backmapping CG trajectories to all-atom resolution

We benchmarked the performance of two recently-published protocols for backmapping CG structures/trajectories to all-atom resolution to facilitate comparisons with the representative all-atom structures: CG2AT2[74] and GLIMPS[75].

Briefly, the GLIMPS tool uses a machine-learning-based approach, which is trained on a dataset in which the reference all-atom structures are already known. The training of the machine-learning model is achieved by mapping a set of all-atom structures (we used 500 equally spaced frames of each PTM simulation from the all-atom production runs) to CG and then backmapping them to all-atom resolution. The resulting backmapping map can be used to transform any CG PTM structure to an all-atom one. To further improve the quality of the generated backmapped structures, we performed a brief energy minimization (up to 500 steps) with GROMACS using the steepest descent algorithm.

CG2AT2 is a geometric fragment-based approach which requires no training. It supports multiple protocols for the backmapping. We tested multiple variations of the default protocol: 1) we tested the default protocol, which consists of fitting the all-atom fragments to the CG beads and performs a brief energy minimization, followed by a brief NVT run, the last frame of which is morphed by steered MD to the initial mapped structure; 2) we stopped the process after the initial energy minimization; 3) we stopped the simulation after the energy minimization, followed by the same brief NVT run as for protocol (1); and finally 4) the energy minimization was followed by the steered MD simulation, to a provided reference structure.

Both methods (GLIMPS and CG2AT2) were tested on a set of 100 equally-spaced structures from the all-atom simulations, which were not part of the training set for GLIMPS. The

structures were then mapped to CG, before being backmapped to all-atom resolution. The Root Mean Square Deviation (RMSD) between each all-atom structure and the respective backmapped structure (using all heavy atoms) was then computed.

Solvent Accessible Surface Area

We calculated the SASA with GROMACS (`gmx sasa`) for both all-atom and CG systems, using a probe with radius of 0.191 nm for both systems. For the all-atom calculations, we substituted the default van der Waals radii values, which are part of the GROMACS distribution, with those listed in Table S1[76,77]. For the CG calculations, we used the values recommended by the Martini 3 developers, i.e., radii of 0.264, 0.23 and 0.191 nm for the "Regular", "Small" and "Tiny" beads, respectively.[36]

Bilayer penetration depth of PTMs

We calculated the coordinates of the CoG for the two leaflet lipid head groups, defined as the CoG for all phosphorus atoms and "PO4" beads for the all-atom and CG runs, respectively, over the production trajectories to compute the depth of penetration of each PTM into the bilayer. We also computed the coordinates of the CoG of the entire bilayer, defined as the CoG of upper and lower lipid head groups using the same atom/bead selections as before. We finally computed the coordinates of the CoG of all PTM atoms/beads. We then examined the Z-axis (parallel to the membrane normal) component of the coordinates and determined the penetration depth of the PTM atoms/beads relative to the proximal and opposite leaflet lipid head layers, as well as the bilayer center. Figure S2 contains a representation of these components, with the proximal (to the peptide) and distal lipid head bilayer shown as black- and grey-colored spheres, respectively, the CoG of the bilayer as an orange sphere, the peptide backbone and side chains shown as mauve-colored spheres and sticks, respectively, with the exception of the modified

residue, which is shown as a light blue-colored sphere and stick model. The simulation box dimensions are highlighted with green lines.

Analysis of the protein simulations and comparison with atomistic data

For the analysis, we examined the orientation of the proteins with respect to the bilayer using Euler angles (Figure 2) describing i) the rotation (Euler angle α), ii) the tilt (Euler angle β) and iii) the spin (Euler angle γ) of the G-domain of each protein with respect to the membrane normal. To estimate the uncertainty of the Euler angle distribution populations, we subsampled the original sample 1000 times, while randomly sampling a tenth of the original sample each time. We also determined the distance of the two lobes of the G domain of each protein from the lipid bilayer by splitting each protein in two lobes (residues 1-86 and 17-94 for the first lobe, and 87-170 and 95-180 for the second lobe, for the Rheb and Arf1 proteins, respectively). Then, we computed the CoG of all residues within each lobe using all backbone beads for the entire production trajectories. We then calculated the Z-axis distance between the CoGs of the two lobes and that of the bilayer (computed using all lipid beads). We projected the distribution of each lobe in a 2D grid to determine the high-density areas. We then carried out a clustering analysis of the 2D grid points using the scikit-learn[78] (v 1.3.0) implementation of the Hierarchical Density-Based Spatial Clustering of Applications with Noise (HDBSCAN) algorithm[79,80]. We set the minimum number of members for each cluster to 2500 species and left all other settings to their default values. Representative structures were then isolated from each cluster by extracting the frame of each cluster closest to the cluster centroid.

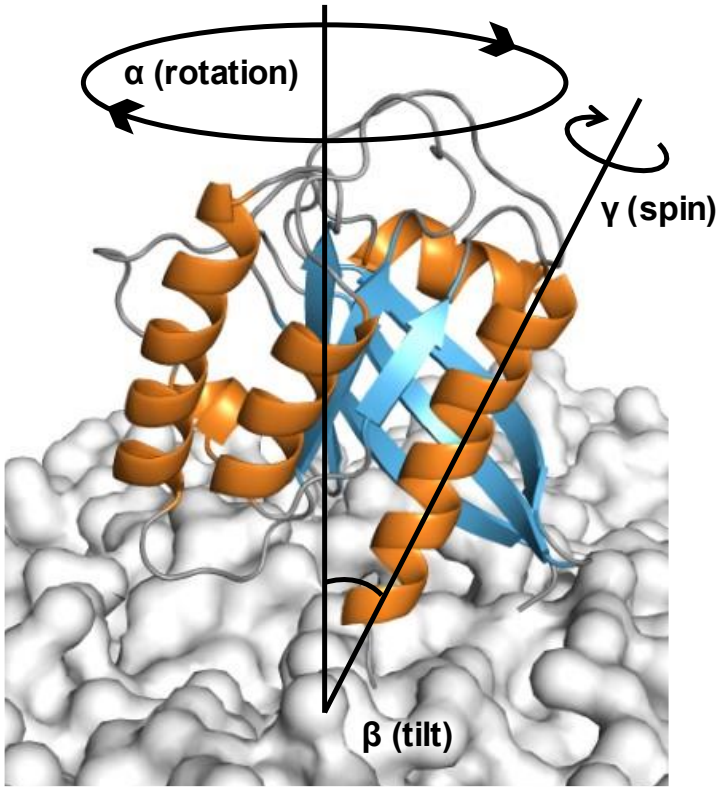


Figure 2: Schematic representation of the three Euler angles used to describe the orientation of the simulated proteins with respect to the bilayer normal. The protein structure is colored according to secondary structure with helices, sheets and loops/coil colored orange, light blue and light grey, respectively. Bilayer is shown as white surface.

We determined the contact network of all residues with respect to the bilayer lipids by calculating the minimum distance between each protein residue and any lipid as a function of simulation time. A contact was recorded if any protein-lipid inter-bead distance was smaller than or equal to 0.7 nm. The respective all-atom analysis used a cut-off value of 0.5 nm, however, for the coarse-grained analysis we used two distance cut-off values equal to 0.7 and 1 nm to accommodate the larger vdW radii of the coarse-grained force field. We have only used the contacts within 0.7 nm in the discussion of the results, but also highlight the ones within 1.0 nm. We used the minimum distance module of GROMACS for this analysis (`gmx mindist`).

RESULTS AND DISCUSSION

CG bonded parameter extraction and comparison to all-atom

First, we calculated the distributions of bond length, angle and torsion values during the CG simulations between the bonded CoGs of the mapped atoms (Figures S3-S6). The distributions of bond length values for the prenylation PTMs (Figures S3A and S4A) reveal two distribution types, depending on the size of the beads each bond connects. The bonds between the peripheral ("TC2") beads (in which the branched methyl groups and the "main-chain" carbon they are bonded to are grouped) and their bonded "main-chain" ("SC4") beads have a narrow range of observed values around a single peak below 3 Å. The bonds between the backbone bead and successive "main-chain" beads feature a broad range of values with a single peak above 3.5 Å. The bond length distributions for the saturated PTMs (Figures S5A and S6A) show bimodal distributions for the bonded beads. The angle plots (Figures S3B through S6B) show a plethora of distribution types from broad, almost uniform-like, to narrow single-peak and even multimodal distributions with as many as three peaks. The angles between the backbone, first and second "main-chain" beads of the prenylation PTMs (Figures S3B and S4B, dark gray color) peak close to 180° indicating an almost linear arrangement between the peptide backbone and first part of the lipidated side chain. The angles between the "main-chain" and branched beads feature multiple populations, with the first two and three angles for the farnesyl and geranylgeranyl PTMs, respectively, represented by bimodal distributions. The angle involving the terminal bead for both prenylation PTMs features a single dominant population with two minor populations on either side of it, forming a three-peak multimodal distribution. The remaining angles between the "main-chain" beads feature broad range of observed values. The angle values for the saturated PTMs (Figures S5B and S6B) show single-peak distributions (with some minor, secondary populations) with the majority of values over 140°, indicating a

mostly linear arrangement for the lipidated side chains. The torsion distributions (Figures S3C through S6C) feature examples of unimodal (both myristoyl torsions), bimodal (all torsions involving non-terminal backbone or lipidated side-chain beads for all PTMs except the myristoyl) as well as multimodal distributions with four distinct populations (torsions involving the terminal lipidated side chain beads of the prenylations). Figure S7 shows CG PTM structures superimposed on the atomistic representation from which they were mapped. As mentioned previously (see "Extraction of bonded parameters from all-atom simulations" section in "Methods"), the mean values of the all-atom distributions served as reference values for the CG simulations with the width of the distributions determining the force constant associated with each term. Tables S2 through S5 summarize atom to CG bead mappings for all PTMs.

Figure 3 shows a comparison of the all-atom and CG parameters for all PTMs, apart from the bonds of the farnesyl and geranylgeranyl PTMs that were modelled as constraints (bond IDs 3, 5, 7 and 3, 5, 7, 9, respectively) and which are shown separately in Figure S8. As is typical for CG force fields, the multimodal populations shown in the bond length and angle plots of the all-atom distributions cannot be reproduced in a CG force field.[27–29] This limitation exists because the groups of atoms, which are grouped into single beads in the CG force field, might fall into more than one population (multimodal distributions). However, the bonded terms, and specifically the bond lengths and angles, are represented with plain harmonic potentials in GROMACS, meaning they can only fluctuate around a single value. The end result is a CG bond length or angle potential that accurately captures the average behavior of the all-atom force field. The SI lists the functional forms we employed for all bonded terms presented in this work, along with a list of all the derived parameters (Tables S6 through S17).

It should be noted that a discrepancy can be observed between the mean values of the bond lengths between the "SC4" and "TC2" beads (see Figure S8) in the mapped all-atom trajectories

(0.28, 0.28 and 0.214 nm for each bond, respectively, starting from the one closest to the backbone and ending with the one furthest away from the backbone), and the CG parameters (0.33, 0.33 and 0.289 nm for the same bonds). We chose to increase the bond length between these beads by 0.05, 0.05 and 0.075 nm, for each bond, respectively, because our analysis revealed the initial parameters resulted in a CG model whose 3D molecular volume (as measured by SASA) was significantly smaller (difference of 12.6 %) than the equivalent all-atom structure. Figure S9 shows the Connolly surface area[81,82] (as calculated with the initial parameters) for the farnesyl PTM for both the all-atom and CG models, with the "missing" volume clearly visible for the CG surface area (orange mesh) when compared with the all-atom surface area (grey mesh). This procedure is in line with the current rules of Martini 3 parametrization, with bond length rescaling recently used for monosaccharides and phosphatidylinositide lipids.[83,84] Bonds between the "SC4" and "TC2" beads were modelled as GROMACS constraints (GROMACS constraint type 1) due to the narrow distributions observed in the all-atom simulations, which explains the extremely small range of values seen in the CG simulations. All other bonds were modelled as simple harmonic oscillations (GROMACS bond type 1). The angles were modeled with cosine-based terms (GROMACS angle type 2).

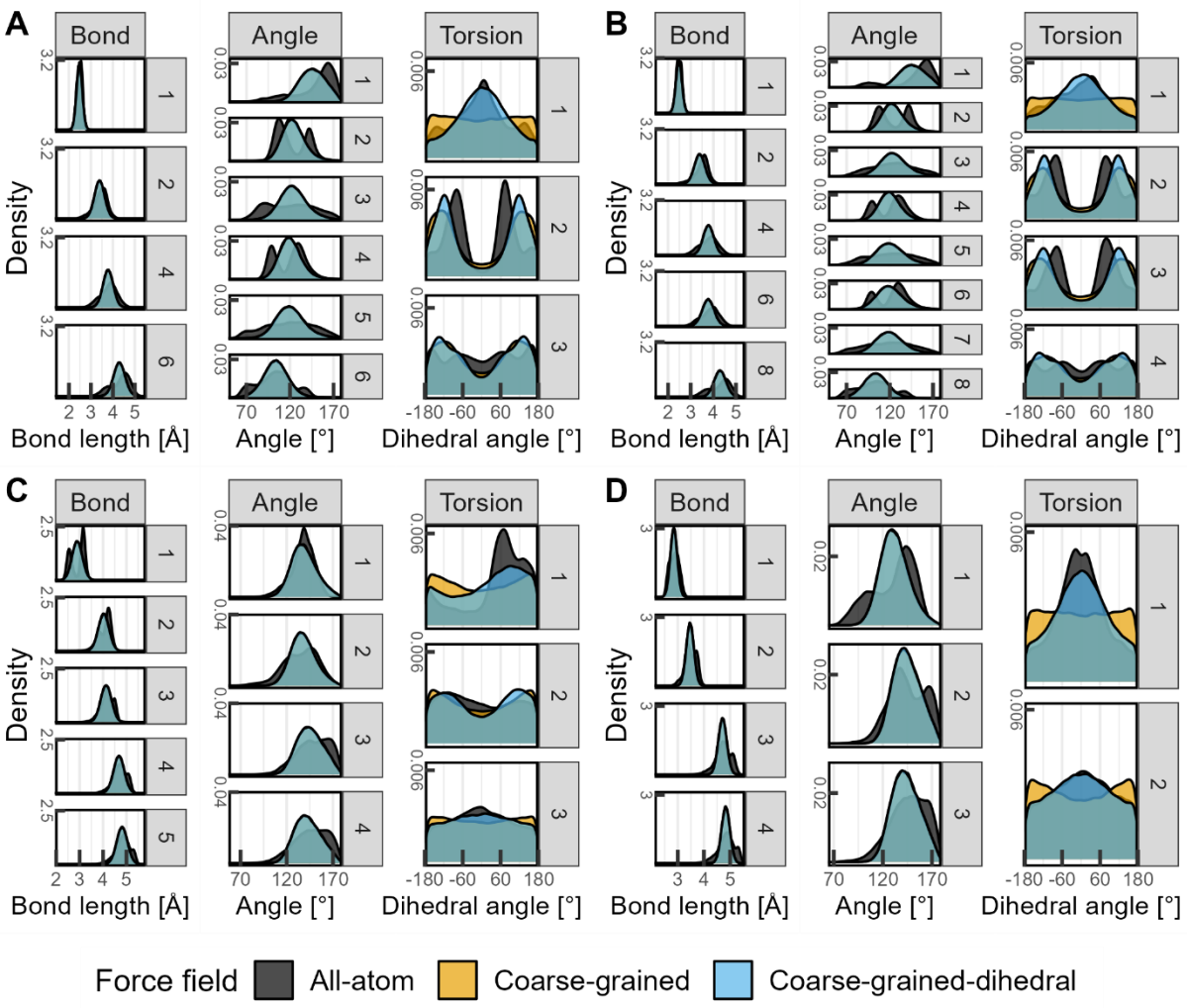


Figure 3. Comparison of all-atom and CG distributions of the bonded parameters for (A) the farnesylation, (B) geranylgeranylation, (C) palmitoylation, and (D) myristoylation PTMs. The bond/angle/torsional angle IDs are broken down in Tables S6 through S17. The dark grey, light orange, and light blue distributions correspond to the values obtained from the all-atom, the CG without dihedral terms, and the CG with dihedral terms simulations. Note that for most plots (except the torsional angle ones) the two CG distributions (with and without dihedrals) overlap with minor deviations. Bonds 3, 5 and 7 (panel A), and 3, 5, 7 and 9 (panel B) are omitted from this plot because they are modelled as constraints (see description in the text) due to the narrow range of observed values in the atomistic simulations. Figure S8 shows only these constraints with separate Y axes to further highlight the narrow and shifted (with respect to all-atom simulations) distributions.

It is almost impossible to discern the impact of the dihedral terms on the CG simulations if we examine only the bond and angle distributions as they fully overlap. However, the impact of the dihedral terms on the overall structures is significant and shifts the behavior of the CG simulation to more closely match that of the all-atom simulation. We opted to model these

dihedrals as simply as possible using only a single proper dihedral term (GROMACS dihedral type 1[85]). The all-atom dihedral distributions, which showed a single peak, were the simplest ones to model and we used a dihedral potential with an appropriate phase angle and a multiplicity of 1, with force constants determined from the width of the respective distributions (see tables S8, S11, S14 and S17 for the torsion parameters for all PTMs). The torsional angles that were modelled this way were the prenylation torsional angles with identifier (ID) 1, the palmitoyl torsional angles with IDs 1 and 3, and both myristoyl torsional angles (see aforementioned SI tables for the ID values). Farnesyl torsional angle ID 2, geranylgeranyl angles ID 2 and ID 3, and palmitoyl angle ID 2 were modelled with a multiplicity of 2 and an appropriate phase angle. The terminal torsional angles of the prenylations were modelled with a multiplicity of 4. Given the clear impact the dihedral terms had on shifting the torsional angle values of the CG force fields closer to the all-atom (see Figure 3), we decided to implement the torsional angle potentials using the simplest terms (GROMACS dihedral type 1) instead of investigating more complex implementations.

One of the aims of the design principles when developing these mapping schemes was their modularity and extensibility (see "All-atom to coarse-grained mapping schemes" section in "Methods"). We thus chose to use the farnesyl-derived parameters on the geranylgeranyl PTM instead of using the parameters derived from its all-atom simulation. Since the geranylgeranyl lipid tail is longer than the farnesyl one, with one additional 2-methyl-2-butene unit (corresponding to one additional "SC4" and "TC2" beads for the CG model), we replicated beads "C2" and "C5" and their associated parameters once (see Tables S9 through S11).

Validation of the CG parameters

In addition to comparing the CG bonded term distributions (both with and without dihedrals) to the all-atom ones, we carried out a series of additional analyses to further validate our parameters.

SASA

Figure 4 shows the SASA comparisons between the all-atom simulations, the all-atom simulations mapped to CG, and the CG simulations (without and with dihedral terms). Overall, there is excellent agreement between the all-atom and CG distributions with the largest discrepancy belonging to the myristoyl PTM, for which the CG models show a difference of 4.71 and 4.86 %, for the parameters without, and with dihedral terms, respectively.

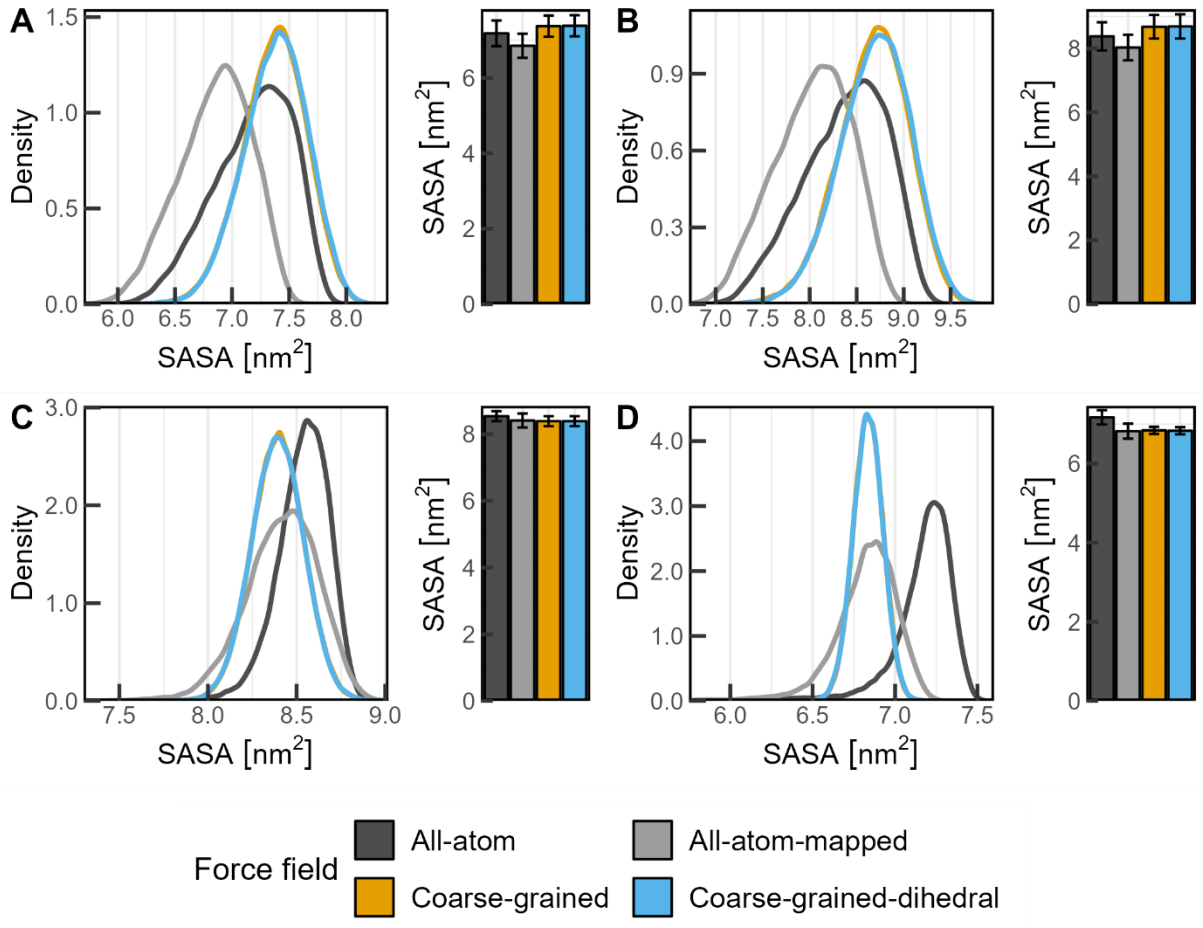


Figure 4. SASA distributions for all PTMs with the (A) farnesylation, (B) geranylgeranylation, (C) palmitoylation, and (D) myristoylation. Each panel shows the distribution of values to the left and bar plots of the mean SASA values (with error bars corresponding to standard deviations) to the right. The dark grey, light grey, orange and light blue distributions and bars correspond to the all-atom, mapped all-atom, CG without dihedral terms, and CG with dihedral terms simulations, respectively.

Tables 1 and 2 show the absolute and percent differences of the mean SASA values for all PTMs for the CG simulation without and with dihedral terms, respectively. As was evident from the respective plot, the SASA values are in excellent agreement between the all-atom and CG simulations for all PTMs, with mean value differences of less than 5%.

Table 1. Absolute and percent differences of mean SASA values between all-atom and CG (without dihedral terms) simulation. The absolute difference is the absolute value of the difference of the two mean values, the average difference is the sum of the two mean values divided by 2, and percent difference is the absolute difference over the average difference multiplied by 100.

PTM	Absolute difference [nm^2]	Average difference [nm^2]	Percent difference [%]
Farnesylation	0.19	7.26	2.61
Geranylgeranylation	0.3	8.53	3.52
Palmitoylation	0.15	8.47	1.77
Myristoylation	0.33	7.01	4.71

Table 2. Absolute and percent difference of mean SASA values between all-atom and CG (with dihedral terms) simulation. For an explanation of each column see the caption of Table 1.

PTM	Absolute difference [nm^2]	Average difference [nm^2]	Percent difference [%]
Farnesylation	0.2	7.28	2.75
Geranylgeranylation	0.31	8.54	3.63
Palmitoylation	0.15	8.47	1.77
Myristoylation	0.34	7.0	4.86

Figure S10 shows the comparison of SASA for the two mapping schemes proposed for the farnesyl PTM. The distribution and bar plots (left and right panel, respectively) show that the differences between the two mapping schemes and the all-atom values are not statistically significant. Specifically, the all-atom, CG (proposed mapping scheme) and CG (alternative mapping scheme) simulations have average SASA values of 7.18 ± 0.34, 7.37 ± 0.28 and 7.18 ± 0.24 nm^2, respectively. Given the insignificant differences between the two schemes, we favored the proposed mapping scheme (see Figure 1, panel A) over the alternative mapping scheme (Figure S1) due to the conceptual simplicity of its non-overlapping atom-to-bead mappings.

Clustering and comparison of representative structures

We then performed RMSD-based clustering on the production trajectories and extract representative structures from each cluster to compare how similar the CG-derived ones were to the ones extracted from the all-atom simulation.

Before carrying out the comparison, however, the structures derived from the CG simulations had to be backmapped to all-atom representation. As mentioned previously (see "Backmapping CG trajectories to all-atom resolution" section in "Methods"), we tested the performance of two tools on our dataset: GLIMPS[75] and CG2AT2[74]. Figure S11 shows the results of this performance comparison. Out of the protocols we tested for CG2AT2 the default one (energy minimization, followed by brief NVT simulation, followed by steered MD) yielded the best results over all PTMs. For GLIMPS, the additional energy minimization step did not significantly change the distribution of RMSD values, but it did improve the bonded interactions within the molecules (bond lengths, angles, etc.); therefore, we retained it over the default protocol. Comparisons of the default CG2AT2 protocol and GLIMPS default protocol followed by energy minimization highlighted a clear advantage of GLIMPS over CG2AT2, with the average (over all PTMs) RMSD equal to 0.8 ± 0.2 and 1.4 ± 0.2 Å for the two codes, respectively. The chirality of the alpha carbon for all modified residues (except for the glycine which does not have a chiral center) was correctly determined by the GLIMPS software, as can be seen in the representative structures shown in Figure 5. We did not examine the chiral status of the structures backmapped with CG2AT2. Therefore, all subsequent CG to all-atom transformations were carried out with GLIMPS.

Figure 5 shows comparisons between the representative structures extracted from the all-atom and CG simulations with and without dihedral terms. For the farnesyl PTM (see Figure 5A), a single cluster was identified (100% population) for the all-atom and both CG simulations. Both backmapped representative structures from the simulations with and without dihedral terms have the same RMSD from the all-atom representative structure (1.8 Å). The representative structure for the all-atom and CG simulations is characterized by a kink along the main chain, immediately following the first branched segment, resulting in a 90° angle between the part of the molecule extending from the backbone to the first branching segment and the part

immediately after the first branched segment through the lipid tail terminus. The same findings can be seen for the myristoyl PTM (see Figure 5D) as well, with a single cluster (100% population) for all simulations and even smaller structural differences with the representative structures from the CG simulations with and without dihedral terms deviating just 0.7 and 0.9 Å from the representative all-atom structures. The farnesyl lipid tail though extends in a relatively straight fashion, unlike the palmitoyl tail, perhaps due to the absence of branched segments and double carbon-carbon bonds and the planarity restrictions they impose.

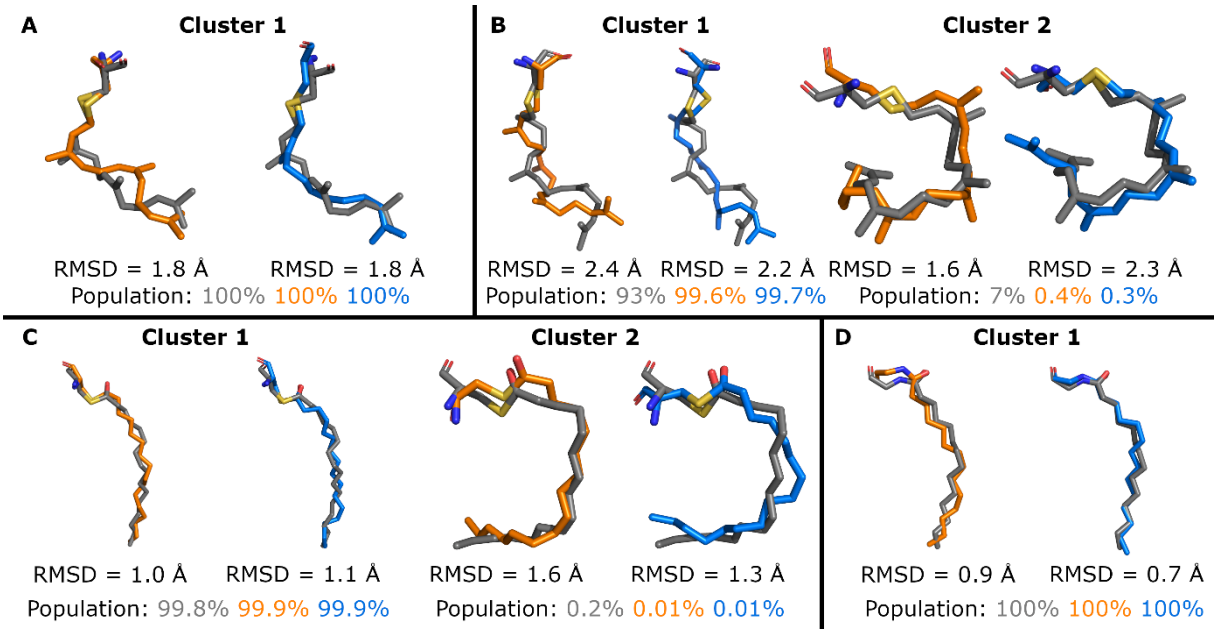


Figure 5. Comparison of the representative structures extracted from the clustering analysis for all simulations. Panels A through D show the results for farnesyl, geranylgeranyl, palmitoyl and myristoyl, respectively. The structures extracted from the CG simulations with and without dihedrals terms are colored blue and orange, respectively. The all-atom representative structures are colored grey. The RMSD (computed over all heavy atoms) for each comparison is reported below every structure. The percentage of the trajectory grouped in each cluster is shown immediately below, with each population percentile value colored according to its corresponding simulation. For the geranylgeranyl analysis, the minor cluster populations have been included in the respective major cluster. All structures are shown as sticks with the oxygen sulfur and nitrogen atoms colored red, yellow and dark blue, respectively.

More clusters were identified for the geranylgeranyl and palmitoyl PTMs compared to farnesyl and myristoyl PTMs, perhaps owing to the longer lipid tails of these PTMs.

For the geranylgeranyl (see Figure 5B), the all-atom simulation analysis reveals four clusters with all frames considered for the analysis ending up in one of the four clusters. The first two clusters are the major ones with 92.2 and 7% population, for the first and second cluster, respectively. The third and fourth clusters (0.8 and 0.08% population, respectively) were not considered for this analysis because their representative structures are very similar to those of the two major clusters, with the RMSD between the representative structures of the first and third clusters, and between those of the second and fourth equal to 2.5 and 3.5 Å, respectively. The CG simulations behave in a similar fashion, with the simulation without dihedrals terms also totaling four clusters, albeit with distinctly different population ratios (99.6, 0.3, 0.07 and 0.03% for the four clusters, respectively), and the simulation with dihedral terms totaling three clusters with population ratios similar to those seen for the simulation without the dihedral terms (99.7, 0.03, and 0.01% for the three clusters, respectively). As for the all-atom analysis, we have not considered clusters three and four of the CG simulation without dihedral terms because their representative structures are very similar to those of the two major clusters from the same simulation (RMSD between representative structures of the first and fourth clusters and between those of the second and third were equal to 3.1 and 2.3 Å, respectively). The third cluster of the CG simulation with dihedral terms only contains a single structure which is similar to the representative structure of the major cluster from the same simulation (2.8 Å RMSD), and has therefore also been excluded from the analysis.

The representative structures from the major cluster for all simulations exhibit the same kink observed for the farnesyl but in this instance the kink manifests immediately past the second (if we start traversing the molecule from the backbone end) branching segment. The minor population exhibits one additional kink to create this structure in which the terminal parts of the lipid tails have almost folded onto the preceding segments.

The structural comparisons between the top clusters indicate a good agreement between the representative structures of the CG simulations and the all-atom ones, with the representative structures of the major cluster of the CG simulations differing by 2.4 and 2.2 Å from the representative structures of the major cluster of the all-atom simulation, for the CG simulation without and with dihedral terms, respectively. The same findings apply to the minor cluster as well, with the structural differences between the representative structures of the CG simulations and the all-atom one amounting to 1.6 and 2.3 Å for the CG simulations without and with dihedral terms, respectively. However, we do note differences between the all-atom and CG simulations in the percentage of the trajectory that the major and minor clusters occupy. In the all-atom simulation (and after folding clusters three and four into the major/minor cluster they are most similar to) the major cluster occupies 93% of the trajectory and the minor one the remaining 7%. The differences may be related to the lower sampling of the atomistic trajectories, which may affect the convergence of the populations. In both CG simulations the major cluster occupies 99.6-99.7% of the trajectory and the minor one the remaining 0.3-0.4%. Importantly though, the relevant part of the landscape is sampled, as is indicated from the low RMSD values between the representative structures of the minor cluster of the CG simulations and the representative structure of the minor cluster of the all-atom simulation.

The palmitoyl all-atom simulations (see Figure 5C) also yield two clusters (99.8 and 0.2% population). The CG simulations behave in a similar fashion, with the simulation without dihedral terms yielding two clusters with population equal to 99.9 and 0.01%, respectively, and the simulation with dihedral terms yielding two clusters with population equal to 99.9 and 0.01%, respectively. Although the size of the minor clusters is small when compared with the minor clusters of the geranylgeranyl CG simulations, the difference in terms of population between the all-atom and CG simulations is smaller, since the minor all-atom cluster only occupies 0.2 % of the trajectory (compared to 7% for the geranylgeranyl).

The major cluster adopts a mostly extended conformation that penetrates into the bilayer in a straightforward fashion, the minor cluster adopts a horseshoe-shaped orientation (see Figure 5C). The differences between the all-atom and CG simulations are minimal with the two CG simulation representative structures of the major cluster only 1 and 1.1 Å (for the simulation without and with dihedral terms, respectively) apart from the respective all-atom representative structure. The small structural differences can also be seen for the minor cluster with RMSD values as low as 1.6 and 1.3 Å (for the CG simulation without and with dihedral terms, respectively). It is also worth noting that even though the populations of the minor clusters are very small, they reproduce the behavior of the all-atom system.

Overall, the clustering analysis reveals excellent agreement between the all-atom and CG models, with RMSD values between representative structures as low as 0.7 Å for the myristoyl PTM (see Figure 5 D). The impact of the dihedral terms is not as clear-cut. For most cases they marginally benefit the structural comparisons (geranylgeranyl major cluster, both palmitoyl clusters and single myristoyl cluster), in one case they have no impact (farnesyl) and in one case they do significantly worse (geranylgeranyl minor cluster). Overall, the representative structure for the single/major cluster for all systems highlights an orientation in which the lipidated side chain is embedded in the hydrophobic core almost linearly, with the two saturated systems showcasing more linear arrangements due to the absence of kinks introduced by the presence of double bonds in the prenylated side chains. No snorkeling was observed, even among the representative structures of the minor clusters for the geranylgeranyl and palmitoyl simulations, which feature models that tend to fold on themselves, but still remain firmly embedded in the hydrophobic core of the bilayer.

Bilayer penetration depth of the PTMs

We investigated how deeply the PTMs penetrate into the bilayer by computing the Z-axis (parallel to the membrane normal) coordinates of the CoG of the PTM atoms or beads

(including the backbone) relative to the CoG of the lipid head groups (represented by phosphorus atoms or “PO4” beads) of the leaflet into which the PTM was inserted as well as the lipid head layer of the opposite leaflet, and the CoG of the entire bilayer (represented by the CoG of phosphorus atoms or “PO4” beads from both leaflets). Figure S2 shows an illustration of the elements involved in these calculations. Figure 6 shows a comparison of the Z-axis distances between CoGs of the PTM and the lipid bilayer, the lipid head layer proximal to the peptide and the opposite one for all PTMs. Exhibiting the same trend as the prior two analyses, the behavior of the palmitoyl PTM shows almost no differences between the all-atom and either of the CG simulations. The bilayer penetration depth for the farnesyl, geranylgeranyl and myristoyl PTMs remains consistent between the all-atom and CG with all the differences seen in Figure 6 being well under the margin of statistical significance, as illustrated by the overlapping error bars. Figures S12 to S15 show the fluctuation of the PTM CoG coordinates relative to the leaflets and bilayer CoG for all PTMs. The first three columns show values from simulations using all-atom, CG without dihedral terms and CG with dihedral terms, respectively. The coordinate plots (top row, first three columns from the left) show the position of the CoG for the PTM (light blue), proximal (to peptide) and distal lipid heads (black and light grey, respectively) and that of the bilayer (orange). The plots immediately below them (bottom row, first three columns from the left) show the absolute value of the distance between the position of the PTM CoG and the bilayer, proximal lipid heads and distal lipid heads, colored orange, black and light grey, respectively. The distribution plots (fourth column) show the range of all absolute value for the CoG distances as black, light grey and orange lines for the all-atom, CG without dihedral terms, CG with dihedral terms, respectively. The top, middle and bottom panels show the distribution of distance values between PTM CoGs and the bilayer, proximal and distal lipid head CoGs, respectively.

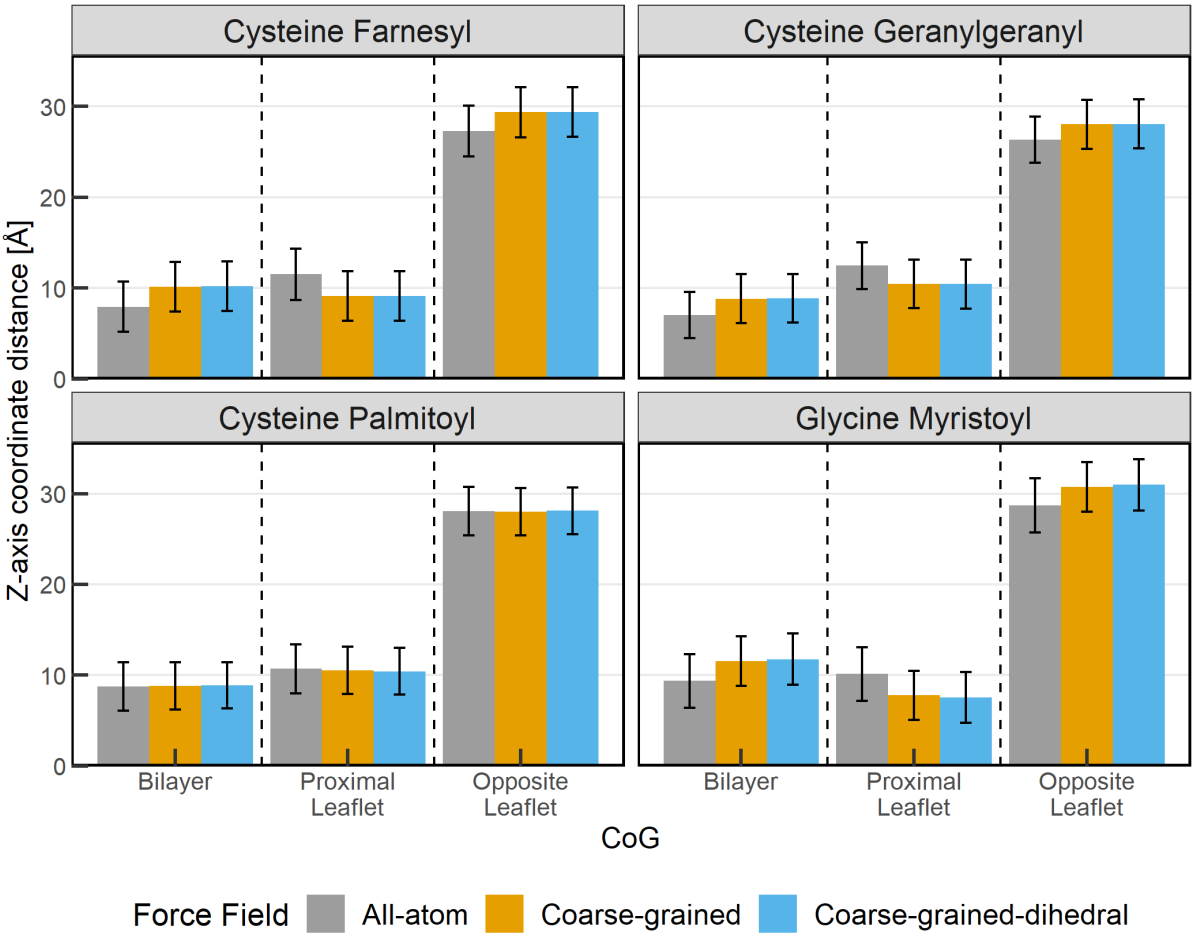


Figure 6. Comparison of the Z-axis distances between CoGs of the PTM and the lipid bilayer, the lipid head layer proximal to the peptide and the opposite one, for all PTMs. Each panel contains three sets of bar plots (separated by vertical dotted lines) showing the mean (and standard deviation) values of the Z-axis (parallel to the membrane normal) distance between the CoG of the PTM atoms or beads and the CoG of the bilayer (first set from the left), the CoG of the proximal leaflet lipid head layer (middle set) and the CoG of the opposite leaflet lipid head layer (first set from the right). The color of the bar indicates the simulation from which the values originate with the all-atom, CG without dihedral terms and CG with dihedral terms values shown as grey, orange and light blue colored bars, respectively.

Protein use cases and comparison with all-atom simulations

Figure 7 shows the distribution of values for the three Euler angles (see Figure 2 for a schematic representation of these angles). Comparing the obtained values for the Rheb simulation with

the equivalent all-atom ones (top plot of right panel of Figure 2 of Ref.69) reveals that the CG simulation distributions closely match the atomistic ones, with the rotation around the membrane normal and the spin around the tilted position (α and γ Euler angles, respectively) spanning the entire range between [-π, π].

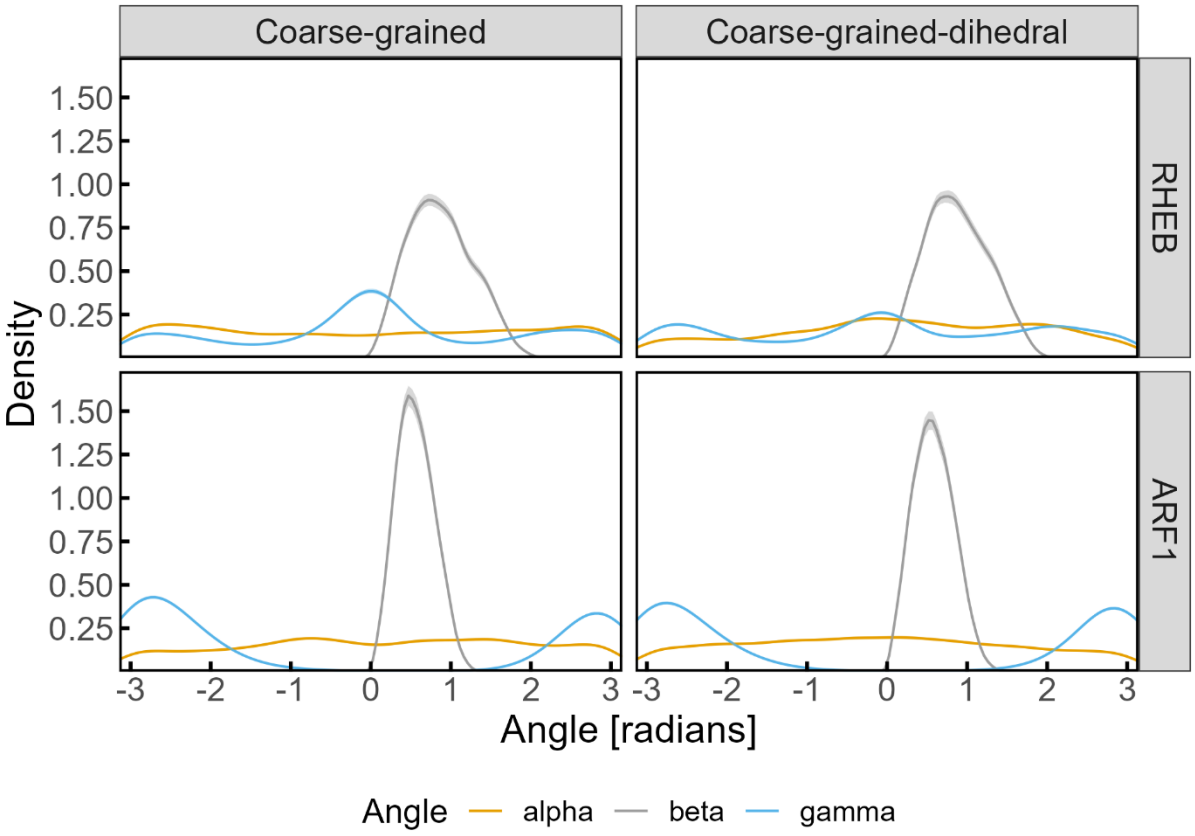


Figure 7: Distribution of Euler angles for the tested proteins. Each panel shows the distribution of the three angles with the rotation around the membrane normal (alpha) shown as an orange line, the tilt off of the membrane normal (beta) as a grey line and the spin around the tilted position (gamma) as a light blue line. The shaded areas (more clearly visible around the peak of the beta distribution), represent the uncertainty of each population.

The distribution of the γ angle in the simulation with active dihedral terms active is flattened compared to the one without dihedral terms, however the respective population in the latter is minor and poses no significant divergence from the atomistic equivalent. The population of the

tilt of the membrane normal angle (angle β) spans angles that range between 0 and 90°, with no significant differences between the simulations with or without dihedral terms or the equivalent atomistic values.

The Arf1 Euler angle distributions reveal the only significant difference between the atomistic and CG simulations. The atomistic simulations (bottom plot of right panel of Figure 2 of Ref.69) show a very narrow range of rotation angles (angle α) as opposed to the flat distributions observed in the CG simulations; also, there are no significant differences in the distributions of the α angle between CG simulations that use dihedral terms or not. The remaining two angles (tilt and spin, β and γ, respectively) demonstrate the same broad behavior as their atomistic counterparts. We believe the discrepancy observed for the rotation angle might be the result of the protein being unable to move away from its original conformation due to the presence of the elastic network constraints, which are placed on the starting coordinates, a limitation which is not present for atomistic force fields. In Ref.69 it is noted that the Arf1 protein undergoes significant conformational rearrangements during the simulation that might enable it to make additional stabilizing contacts with the bilayer lipids, which could prevent it from accessing the full 360° rotational space that the α angle occupies in the CG simulations.

Figure 8 shows the distance between the bilayer and the G domain of the proteins and the contacts made between the protein residues and the bilayer lipids. Specifically, Figure 8A and D show the Z-axis distance between the CoGs of the first lobe of the G domain of each protein and that of the bilayer, Figure 8B and E the contact network between protein residues and bilayer, and Figure 8C and F the normalized 2D distribution of the Z-axis distance of the CoGs of both lobes with respect to the bilayer. These plots are directly comparable with the bottom panels of Figure 3A-C and Figure 6D-E of Ref.69.

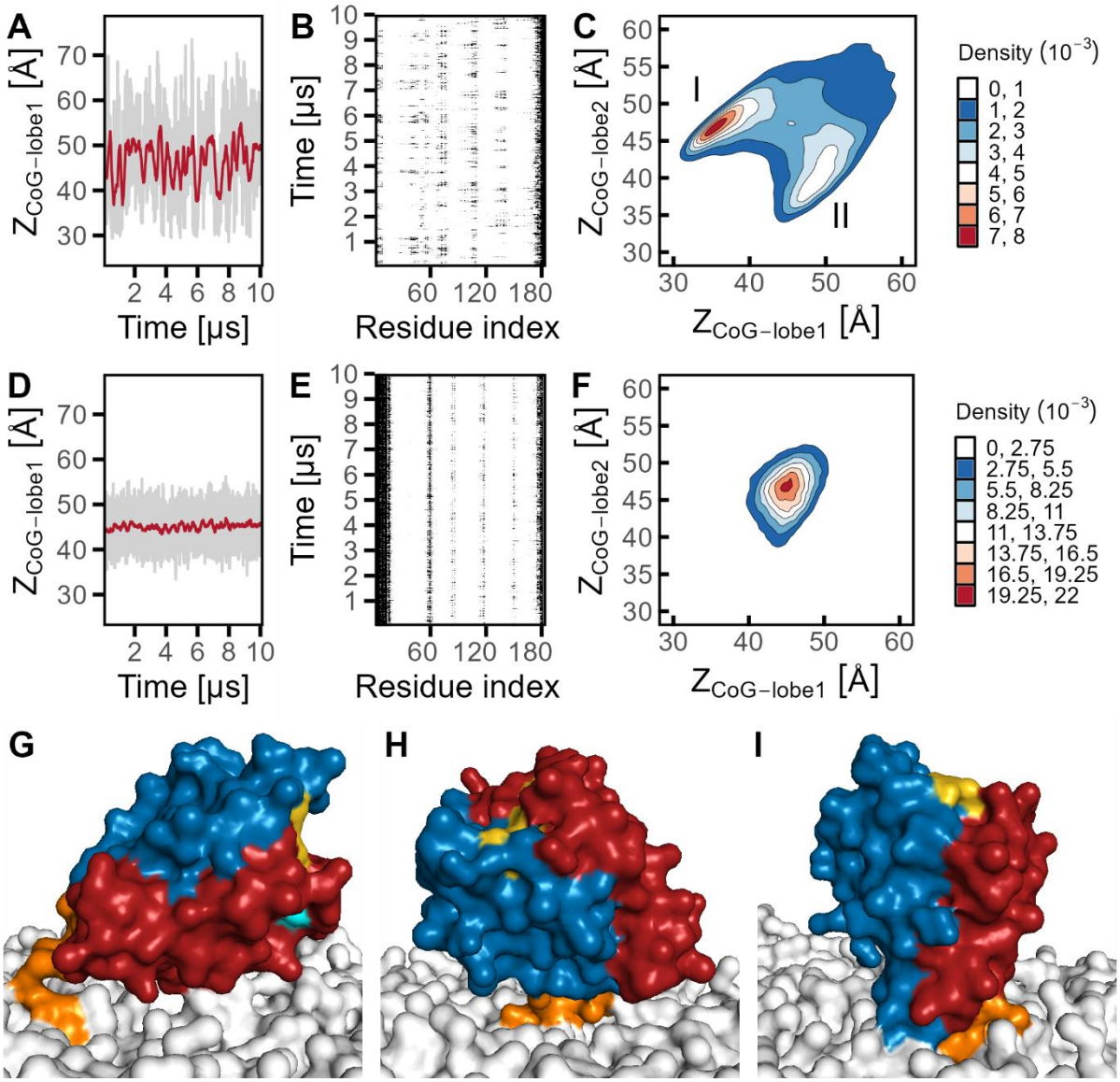


Figure 8: Distance and contact network between the proteins and bilayer for the simulations without dihedral terms. Panels A and D show the evolution of the Z-axis distance between the first lobe of proteins Rheb and Arf1, respectively, and the bilayer. The red line indicates a sliding window of 100 ns. Panels B and E show the contact network between Rheb and Arf1 residues, respectively, and any bilayer lipids as a function of time. Any cells shaded black and grey indicate a contact within 7 and 10 Å, respectively, while white areas indicate no contact. Panels C and F show the normalized 2D distribution of Z-axis distances between the CoGs of the two lobes of proteins Rheb and Arf1, respectively, and that of the bilayer. Areas shaded red indicate higher density, while areas shaded blue indicate lower density. The bottom panels show representative structures from the high-density areas of panels C and F. Panels G and H show structures from the major ("I") and minor ("II") density areas of the 2D distributions for Rheb, respectively, and panel I shows the single representative structure from the Arf1 simulation. Lobes 1 and 2 of the proteins are colored red and blue, respectively, while GTP and Mg^{2+} are colored yellow and cyan, respectively. The HVR of Rheb and first 16 N-terminal residues of Arf1 are colored orange with the bilayer is colored white. The distance and contact network between the proteins and the bilayer for the simulations with dihedral terms can be found in Figure S17.

Arf1 is in excellent agreement with the atomistic values with respect to the distance of the bilayer center because there are no significant differences between the CG and atomistic simulations, both in terms of contact-making residues (Figure 8E) or the Z-axis distance of the two lobes of Arf1 from the bilayer center (Figure 8D and F). Figure 8A and C show that the G domain of Rheb is more distant to the bilayer center in the CG simulation than it is in the atomistic simulation: the highest density (Figure 8C) area corresponds to the lobe 1 Z-axis height being ~35 Å, with the respective value in the atomistic simulation centering at 25 Å (difference of 10 Å). The lobe 2 Z-axis value is ~45 Å, with the respective atomistic region centering at 30 Å (difference of 15 Å). The normalized 2D distribution of Z-axis distance values for the two lobes (Figure 8C) shows that the shape of the distribution is the same as for the atomistic simulation. The results of the clustering analysis for the 2D density plots (see Figure 8C) can be seen in Figure S16, which clearly shows that the algorithm has successfully identified the major and minor high-density areas. The representative structures of the major and minor high-density areas (Figure 8G and H) reinforce the findings from the previous analyses, with lobe 1 (red) interacting with the bilayer in the major cluster and lobe 2 (blue) in the minor. We hypothesize that one likely reason for the greater observed (10-15 Å) distance between the G domain and the bilayer CoG for Rheb (see Figure 8C) in the CG simulation compared to the atomistic one, could be that HVR may be more prone to dissociate from the lipid bilayer to freely mix with the solvent instead. Indeed, an examination of the contact network for the flexible Rheb HVR residues (see Figure 8 B, residue indices 171-182) shows residues 171-175 hardly ever contact any lipids. In contrast, the myristoylated glycine at the N-terminus of the Arf1 protein, is not part of such a flexible region but it is a stable alpha helix instead, which is partially embedded in the bilayer for the majority of the simulation (see Figure 8 E), thus anchoring the protein to the bilayer at a constant distance.

Figure S17 shows the same plots as Figure 8, but for the simulations with active dihedral terms. Overall, the results are highly similar, the sole exception being the 2D distribution of Z-axis distances for the two lobes of Rheb (panel C), which features a flipped orientational profile with respect to the equivalent panel of Figure 8 and the equivalent atomistic one. Although we would need to perform more extensive simulations to be certain this flipping is not the result of insufficient sampling, we believe the reason for this inconsistency lies with the dihedral terms that might impose certain restrictions on the behavior of the molecules, thus limiting the transferability of the model.

Comparison of bead types for palmitoyl PTM

As noted in the Methods section, we tested multiple bead types for the thioester group of the palmitoyl PTM. Figure 9 shows a comparison of the metrics that were used for parameter validation (SASA, clustering analysis and bilayer penetration depth).

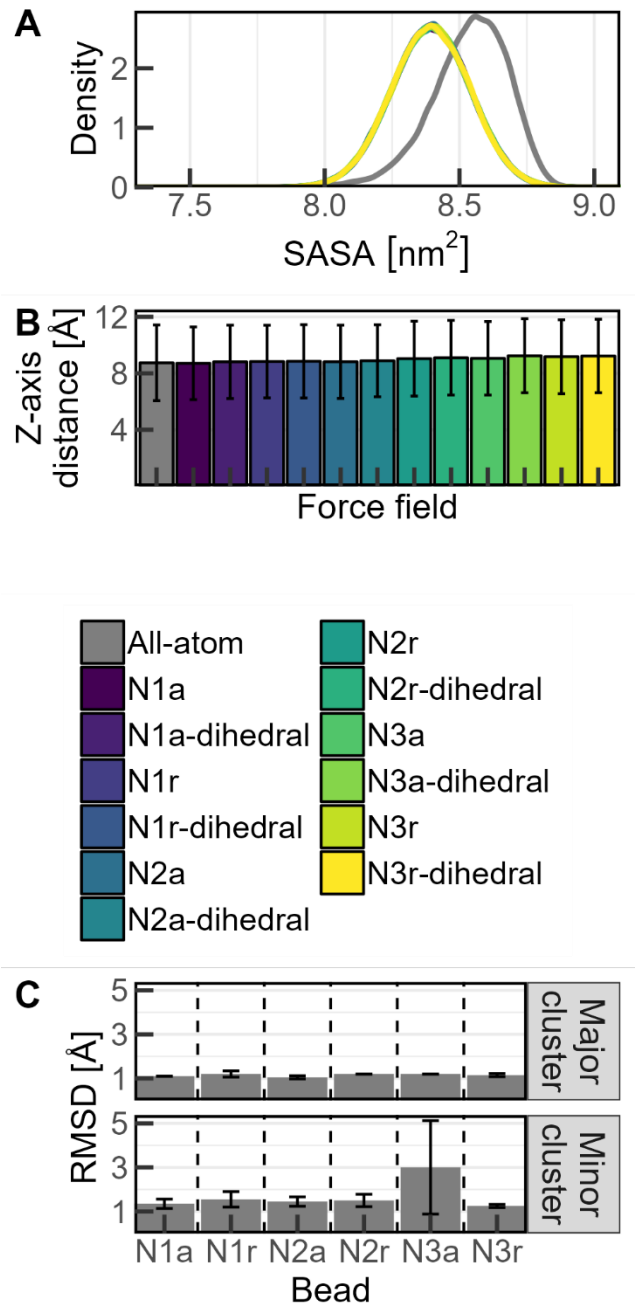


Figure 9: Comparison of bead types for the thio-ester group of the palmitoyl PTM. (A) distributions of the SASA values for the all-atom (grey) and CG (deep violet to yellow spectrum) for the palmitoyl PTM. (B) mean absolute distance between the bilayer and palmitoyl PTM CoGs with greater numbers indicating shallower insertion into the bilayer (same coloring as for panel A). (C) average RMSD values between the cluster representative structures of the all-atom and CG simulations. RMSD values have been averaged over the comparisons with and without dihedral terms.

The SASA analysis (Figure 9A) reveals no differences between the various comparisons with all CG simulations reporting an average value of 8.39 ± 0.15 nm^2, which represents less than 2

% (1.78 %) difference from the mean all-atom value of 8.54 ± 0.15 nm^2, indicating all bead types have excellent agreement with the all-atom data. The comparison of the bilayer penetration depth (Figure 9B) for the various bead selections also shows no statistically significant differences between the all-atom and CG simulations, as can be seen by the overlapping error bars. The clustering analysis (Figure 9C) shows the most significant differences between the various beads. While all beads resulted in the major cluster being identified successfully, with RMSD values between the respective all-atom structure and the ones backmapped from the CG simulations averaging a value close to 1 Å, the same is not true for the minor cluster. The simulations with the most polar beads ("N3"), and specifically "N3a", failed to identify the minor population (even when sampling every 100 ps) when making use of dihedral terms, resulting in the high average RMSD value shown in the graph. The simulation of the "N3r" bead did identify the minor population, however in the simulation without dihedral terms, it identified two minor populations, only one of which was a match for the equivalent all-atom one. Given the minor differences in the performance of the "N1a/r" and "N2a/r" beads, as well as the better-suited physicochemical characteristics of the "N2a" bead, led us to select the N2a bead for the thioester group.

CONCLUSIONS

We have developed Martini 3 parameters for the most common co-/post-translational lipidations, specifically farnesylation, geranylgeranylation, palmitoylation and myristoylation. The all-atom to CG mapping schemes we developed use the improved capabilities of Martini 3 to reproduce chemical properties, resolution, and modularity. Coarse-grained simulations with Martini 3 have recently demonstrated several realistic use cases[34,36,83,84,86–89]; adding parameters for protein lipidation PTMs expands the modeling capabilities of this model to important systems such as the K-Ras signalosome[90]. Moving to a coarse-grained force field also results in significant performance gains over an atomistic one with the Martini 3 simulations achieving an 800-fold speedup over the equivalent simulations in CHARMM36m (~ 0.5 vs 434 ns/day/core for the all-atom vs the coarse-grained simulations presented herein, respectively).

The comparison of the bonded interaction distributions of each PTM, in combination with the independent validation that we performed through the SASA, clustering and structural comparisons as well as bilayer penetration analyses for the peptide PTMs, show that the derived CG parameters accurately reproduce the behavior of the all-atom force field. The mean SASA value differences between the atomistic and CG simulation are 2.61, 3.52, 1.77 and 4.71 %, for the farnesyl, geranylgeranyl, palmitoyl and myristoyl PTMs, respectively, for the runs without dihedral parameters. The representative structures, that were extracted following the clustering analysis of all simulations, reveal the all-atom and CG simulations converge to the same structures/cluster representatives. Specifically, the RMSD values between respective clusters of the all-atom and CG (without dihedral terms) simulations are 1.8 Å for the farnesyl simulations, 2.4 and 1.6 Å for the major and minor cluster of the geranylgeranyl simulations, respectively, 1 and 1.6 Å for the major and minor cluster of the palmitoyl simulations, respectively, and 0.9 Å, for the myristoyl simulations. The bilayer depth penetration analysis

revealed no statistically significant differences between the atomistic and CG simulations, as can be seen by the overlapping error bars between the respective bar plots of Figure 6. The differences with the simulations that use dihedral parameters are minimal and all conclusions apply to those results as well. Importantly, and regardless of the minor differences in the SASA and cluster analysis comparisons, all CG simulations clearly show that the parameters that we are proposing for these lipidations, allow the lipid tails of the modified residues to stably tether peptides/proteins to a lipid bilayer, therefore achieving their primary purpose.

This stability was also validated in simulations of proteins bearing lipid PTMs with the parameters proposed here. Specifically, we simulated Rheb and Arf1, two small peripheral membrane GTPases, that were modified with a C-terminal farnesylated cysteine and an N-terminal glycine, respectively. These proteins were then anchored in a pre-equilibrated POPC:POPS (80:20 % mol.) lipid bilayer by embedding the lipid tail of the lipidated residues in the hydrophobic part of the bilayer. The orientation and distance (with respect to the bilayer) of the G-domain of both proteins were then analyzed and the results are compared with the equivalent all-atom simulation recently published by Prakash and Gorfe.[69] The proteins have a highly similar orientational profile compared to the atomistic simulations, with the exception of Arf1, which in our simulations was found to rotate freely around the membrane normal in a uniform manner, whereas in the atomistic simulation the rotation angle distribution spanned a very narrow range. The Z-axis distance of the G-domain lobes of Rheb is greater than the equivalent metric in the atomistic simulation by approximately 10-15 Å, with the all-atom lobes hovering 25 and 30 Å above the bilayer and their CG counterparts 35 and 45 Å, respectively. We believe this discrepancy is the result of the HVR residues of Rheb more frequently dissociating from the bilayer lipids, compared with the atomistic simulation, even though the farnesylated lipid tail remains embedded in the hydrophobic part of the bilayer, thus stably anchoring the protein to it. This is also supported by the reduced contact frequency between

HVR residues and bilayer lipids. In contrast, Arf1, the lipidation of which is part of a stable part of its structure – an amphipathic alpha helix, hovers above the bilayer at the same distance, 45 Å, as in the atomistic simulation. We repeated these simulations and analyses using dihedral terms, and observed that including the dihedral terms either has a neutral impact on the simulation or flipped the orientation of the G-domain of Rheb with respect to the bilayer, compared to the atomistic simulation.

In the derived parameter set we also included optional dihedral terms is to provide compatibility with future versions of the Martini 3 force field, in case lipid parameters use such terms in subsequent versions. However, adding dihedrals for fully flexible molecules/fragments may add a conformational bias towards the system used for the dihedral calibration. In addition, many of the energy differences in dihedrals involve barriers lower than kT (at 300 K), which also make their presence nonessential in many of the typical biological applications of Martini. Our calculations with and without the dihedral terms show that there are no significantly improved metrics arising from the use of dihedral terms on any of the analyses presented in this work, including the simulated protein systems. Therefore, at present we recommend the use of this parameter set without dihedral terms, as it is a simpler model that also matches the currently available Martini 3 lipid parameters. We will revisit this in the future if the dihedral terms are found to confer an important advantage or as part of the ongoing effort for the re-parameterization of the lipid parameters for the Martini 3 force field.

The extensibility of these parameters was showcased in this work when we successfully applied the parameters generated for the farnesyl lipid tail (the smaller of the two prenylation PTMs) to the geranylgeranyl (the larger of the two). We expect that future additions to this set should transfer with the same ease, provided they are made of (and can be modularized into) the same chemical building blocks. These parameters could even serve as the basis for any branched lipids that are parametrized for the Martini 3 force field. The building block approached used

in this parametrization will also allow future updates in the saturated tails, which are currently in development and may have minor changes in the future. In summary, these parameters enable the robust and efficient simulation of lipidation-modified proteins and protein complexes with the latest version of the popular Martini force field.

ASSOCIATED CONTENT

Supporting information

Supporting figures and Tables. Alternative mapping scheme for the farnesyl PTM, scheme for bilayer penetration analysis, Distributions of the bonded interactions for the four PTMs, Illustration of one structure for each PTM in all-atom and coarse-grained representation, Distributions for the constraints used for the farnesyl and geranylgeranyl PTMs, Comparison of the Connolly surface area for the farnesyl PTM between the all-atom and CG models, SASA comparisons, Performance comparison of the two backmapping tools, Bilayer penetration plots, Clustering analysis of the 2D grid of Z-axis values for lobes 1 and 2 for the Rheb simulations, Evolution of Z-axis coordinates of the lobe 1 of Rheb and Arf1, Table with van der Waals radii per element considered for the all-atom SASA calculation, Tables with Atom to bead mapping information for the four PTMs, Tables with bonded parameters for each PTM.

AUTHOR INFORMATION

Corresponding Author

Zoe Cournia – Biomedical Research Foundation, Academy of Athens, 4 Soranou Ephessiou, 11527 Athens, Greece; Phone: +30 210 6597 195; Email: zcournia@ bioacademy.gr

Paulo C. T. Souza – Molecular Microbiology and Structural Biochemistry, (MMSB, UMR 5086), CNRS & University of Lyon, Lyon, France, and Laboratory of Biology and Modeling of the Cell, École Normale Supérieure de Lyon, Université Claude Bernard Lyon 1, CNRS UMR 5239 and Inserm U1293, 46 Allée d'Italie, 69007, Lyon, France. Email: paulocts@gmail.com

José S. Duca – Computer-Aided Drug Discovery, Global Discovery Chemistry, Novartis Institutes for BioMedical Research, 181 Massachusetts Avenue, Cambridge, Massachusetts 02139, United States, Email: jose.duca@novartis.com

Authors

Panagiotis I. Koukos – Biomedical Research Foundation, Academy of Athens, 4 Soranou Ephessiou, 11527 Athens, Greece

Sepehr Dehghani-Ghahnaviyeh – Computer-Aided Drug Discovery, Global Discovery Chemistry, Novartis Institutes for BioMedical Research, 181 Massachusetts Avenue, Cambridge, Massachusetts 02139, United States

Camilo Velez-Vega – Computer-Aided Drug Discovery, Global Discovery Chemistry, Novartis Institutes for BioMedical Research, 181 Massachusetts Avenue, Cambridge, Massachusetts 02139, United States

John Manchester – Computer-Aided Drug Discovery, Global Discovery Chemistry, Novartis Institutes for BioMedical Research, 181 Massachusetts Avenue, Cambridge, Massachusetts 02139, United States

D. Peter Tieleman – Department of Biological Sciences, University of Calgary, Calgary, Alberta, Canada - and - Centre for Molecular Simulation, University of Calgary, Calgary, Alberta, Canada

DATA AVAILABILITY

All data associated with this publication have been deposited (and are publicly accessible) at https://zenodo.org/record/7849262

ABBREVIATIONS USED

Post-Translational Modification: PTM

All-atom: AA

Molecular Dynamics: MD

High-performance computing: HPC

Coarse-grained: CG

1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine: POPC

Number of particles, Volume, Temperature: NVT

Number of particles, Pressure, Temperature: NPT

Picoseconds nanoseconds microseconds: ps ns μs

Centre of Geometry: CoG

Hierarchical Density-Based Spatial Clustering of Applications with Noise: HDBSCAN

Root mean square deviation: RMSD

Solvent accessible surface area: SASA

Identifier: ID

ACKNOWLEDGMENTS

This work was supported by computational time granted from the Greek Research & Technology Network (GRNET) in the National HPC facility ARIS under Project ID pr012011/CG-KRAS, granted to ZC and PK. ZC and PK also acknowledge PRACE for granting computational time in Marconi100, CINECA, Italy. P.C.T.S. acknowledges the support of the French National Center for Scientific Research (CNRS). D.P.T. acknowledges support from the Natural Sciences and Engineering Research Council (Canada) and the Canada Research Chairs Program.

REFERENCES

(1) Boes, D. M.; Godoy-Hernandez, A.; McMillan, D. G. G. Peripheral Membrane Proteins: Promising Therapeutic Targets across Domains of Life. *Membranes (Basel)* **2021**, *11* (5), 346. https://doi.org/10.3390/membranes11050346.

(2) Allen, K. N.; Entova, S.; Ray, L. C.; Imperiali, B. Monotopic Membrane Proteins Join the Fold. *Trends Biochem Sci* **2019**, *44* (1), 7–20. https://doi.org/10.1016/j.tibs.2018.09.013.

(3) Dong, W.; Zhang, X.; Liu, W.; Chen, Y.; Huang, J.; Austin, E.; Celotto, A. M.; Jiang, W. Z.; Palladino, M. J.; Jiang, Y.; Hammond, G. R. V.; Hong, Y. A Conserved Polybasic Domain Mediates Plasma Membrane Targeting of Lgl and Its Regulation by Hypoxia. *J Cell Biol* **2015**, *211* (2), 273–286. https://doi.org/10.1083/jcb.201503067.

(4) Zhou, Y.; Hancock, J. F. A Novel Prenyl-Polybasic Domain Code Determines Lipid-Binding Specificity of the K-Ras Membrane Anchor. *Small GTPases* **2020**, *11* (3), 220–224. https://doi.org/10.1080/21541248.2017.1379583.

(5) Khan, H. M.; He, T.; Fuglebakk, E.; Grauffel, C.; Yang, B.; Roberts, M. F.; Gershenson, A.; Reuter, N. A Role for Weak Electrostatic Interactions in Peripheral Membrane Protein Binding. *Biophys J* **2016**, *110* (6), 1367–1378. https://doi.org/10.1016/j.bpj.2016.02.020.

(6) Corbalan-Garcia, S.; Gómez-Fernández, J. C. Signaling through C2 Domains: More than One Lipid Target. *Biochimica et Biophysica Acta (BBA) - Biomembranes* **2014**, *1838* (6), 1536–1547. https://doi.org/10.1016/j.bbamem.2014.01.008.

(7) Lomize, A. L.; Pogozheva, I. D.; Lomize, M. A.; Mosberg, H. I. The Role of Hydrophobic Interactions in Positioning of Peripheral Proteins in Membranes. *BMC Structural Biology* **2007**, *7* (1), 44. https://doi.org/10.1186/1472-6807-7-44.

(8) Campelo, F.; McMahon, H. T.; Kozlov, M. M. The Hydrophobic Insertion Mechanism of Membrane Curvature Generation by Proteins. *Biophys J* **2008**, *95* (5), 2325–2339. https://doi.org/10.1529/biophysj.108.133173.

(9) De Marothy, M. T.; Elofsson, A. Marginally Hydrophobic Transmembrane α-Helices Shaping Membrane Protein Folding. *Protein Science* **2015**, *24* (7), 1057–1074. https://doi.org/10.1002/pro.2698.

(10) Fuglebakk, E.; Reuter, N. A Model for Hydrophobic Protrusions on Peripheral Membrane Proteins. *PLOS Computational Biology* **2018**, *14* (7), e1006325. https://doi.org/10.1371/journal.pcbi.1006325.

(11) Larsen, A. H.; John, L. H.; Sansom, M. S. P.; Corey, R. A. Specific Interactions of Peripheral Membrane Proteins with Lipids: What Can Molecular Simulations Show Us? *Bioscience Reports* **2022**, *42* (4), BSR20211406. https://doi.org/10.1042/BSR20211406.

(12) Araya, M. K.; Gorfe, A. A. The Role of Conformational Dynamics in Intrinsically Disordered Lipid Anchors of Peripheral Membrane Proteins in Specifying Interactions with Membrane Lipids. *Biophysical Journal* **2023**, *122* (3), 449a. https://doi.org/10.1016/j.bpj.2022.11.2420.

(13) Blanden, M. J.; Suazo, K. F.; Hildebrandt, E. R.; Hardgrove, D. S.; Patel, M.; Saunders, W. P.; Distefano, M. D.; Schmidt, W. K.; Hougland, J. L. Efficient Farnesylation of an Extended C-Terminal C(x)3X Sequence Motif Expands the Scope of the Prenylated Proteome. *Journal of Biological Chemistry* **2018**, *293* (8), 2770–2785. https://doi.org/10.1074/jbc.M117.805770.

(14) Wang, M.; Casey, P. J. Protein Prenylation: Unique Fats Make Their Mark on Biology. *Nat Rev Mol Cell Biol* **2016**, *17* (2), 110–122. https://doi.org/10.1038/nrm.2015.11.

(15) Chen, J. J.; Fan, Y.; Boehning, D. Regulation of Dynamic Protein S-Acylation. *Frontiers in Molecular Biosciences* **2021**, *8*.

(16) Wang, B.; Dai, T.; Sun, W.; Wei, Y.; Ren, J.; Zhang, L.; Zhang, M.; Zhou, F. Protein N-Myristoylation: Functions and Mechanisms in Control of Innate Immunity. *Cell Mol Immunol* **2021**, *18* (4), 878–888. https://doi.org/10.1038/s41423-021-00663-2.
(17) Chen, B.; Sun, Y.; Niu, J.; Jarugumilli, G. K.; Wu, X. Protein Lipidation in Cell Signaling and Diseases: Function, Regulation and Therapeutic Opportunities. *Cell Chem Biol* **2018**, *25* (7), 817–831. https://doi.org/10.1016/j.chembiol.2018.05.003.
(18) Fhu, C. W.; Ali, A. Protein Lipidation by Palmitoylation and Myristoylation in Cancer. *Frontiers in Cell and Developmental Biology* **2021**, *9*.
(19) Wang, R.; Chen, Y. Q. Protein Lipidation Types: Current Strategies for Enrichment and Characterization. *International Journal of Molecular Sciences* **2022**, *23* (4), 2365. https://doi.org/10.3390/ijms23042365.
(20) Ambrogio, C.; Köhler, J.; Zhou, Z.-W.; Wang, H.; Paranal, R.; Li, J.; Capelletti, M.; Caffarra, C.; Li, S.; Lv, Q.; Gondi, S.; Hunter, J. C.; Lu, J.; Chiarle, R.; Santamaría, D.; Westover, K. D.; Jänne, P. A. KRAS Dimerization Impacts MEK Inhibitor Sensitivity and Oncogenic Activity of Mutant KRAS. *Cell* **2018**, *172* (4), 857-868.e15. https://doi.org/10.1016/j.cell.2017.12.020.
(21) Ciossani, G.; Overlack, K.; Petrovic, A.; Huis In 't Veld, P. J.; Koerner, C.; Wohlgemuth, S.; Maffini, S.; Musacchio, A. The Kinetochore Proteins CENP-E and CENP-F Directly and Specifically Interact with Distinct BUB Mitotic Checkpoint Ser/Thr Kinases. *J Biol Chem* **2018**, *293* (26), 10084–10101. https://doi.org/10.1074/jbc.RA118.003154.
(22) Lynch, S. J.; Snitkin, H.; Gumper, I.; Philips, M. R.; Sabatini, D.; Pellicer, A. The Differential Palmitoylation States of N-Ras and H-Ras Determine Their Distinct Golgi Sub-Compartment Localizations. *J Cell Physiol* **2015**, *230* (3), 610–619. https://doi.org/10.1002/jcp.24779.
(23) Mendiratta, G.; Ke, E.; Aziz, M.; Liarakos, D.; Tong, M.; Stites, E. C. Cancer Gene Mutation Frequencies for the U.S. Population. *Nat Commun* **2021**, *12* (1), 5961. https://doi.org/10.1038/s41467-021-26213-y.
(24) Almeida, J. G.; Preto, A. J.; Koukos, P. I.; Bonvin, A. M. J. J.; Moreira, I. S. Membrane Proteins Structures: A Review on Computational Modeling Tools. *Biochimica et Biophysica Acta (BBA) - Biomembranes* **2017**, *1859* (10), 2021–2039. https://doi.org/10.1016/j.bbamem.2017.07.008.
(25) Li, F.; Egea, P. F.; Vecchio, A. J.; Asial, I.; Gupta, M.; Paulino, J.; Bajaj, R.; Dickinson, M. S.; Ferguson-Miller, S.; Monk, B. C.; Stroud, R. M. Highlighting Membrane Protein Structure and Function: A Celebration of the Protein Data Bank. *Journal of Biological Chemistry* **2021**, *296*, 100557. https://doi.org/10.1016/j.jbc.2021.100557.
(26) Levental, I.; Lyman, E. Regulation of Membrane Protein Structure and Function by Their Lipid Nano-Environment. *Nat Rev Mol Cell Biol* **2023**, *24* (2), 107–122. https://doi.org/10.1038/s41580-022-00524-4.
(27) de Jong, D. H.; Singh, G.; Bennett, W. F. D.; Arnarez, C.; Wassenaar, T. A.; Schäfer, L. V.; Periole, X.; Tieleman, D. P.; Marrink, S. J. Improved Parameters for the Martini Coarse-Grained Protein Force Field. *J. Chem. Theory Comput.* **2013**, *9* (1), 687–697. https://doi.org/10.1021/ct300646g.
(28) Marrink, S. J.; Risselada, H. J.; Yefimov, S.; Tieleman, D. P.; de Vries, A. H. The MARTINI Force Field: Coarse Grained Model for Biomolecular Simulations. *J. Phys. Chem. B* **2007**, *111* (27), 7812–7824. https://doi.org/10.1021/jp071097f.
(29) Monticelli, L.; Kandasamy, S. K.; Periole, X.; Larson, R. G.; Tieleman, D. P.; Marrink, S.-J. The MARTINI Coarse-Grained Force Field: Extension to Proteins. *J. Chem. Theory Comput.* **2008**, *4* (5), 819–834. https://doi.org/10.1021/ct700324x.
(30) Gkeka, P.; Angelikopoulos, P.; Sarkisov, L.; Cournia, Z. Membrane Partitioning of Anionic, Ligand-Coated Nanoparticles Is Accompanied by Ligand Snorkeling, Local

Disordering, and Cholesterol Depletion. *PLOS Computational Biology* **2014**, *10* (12), e1003917. https://doi.org/10.1371/journal.pcbi.1003917.
(31) Angelikopoulos, P.; Sarkisov, L.; Cournia, Z.; Gkeka, P. Self-Assembly of Anionic, Ligand-Coated Nanoparticles in Lipid Membranes. *Nanoscale* **2017**, *9* (3), 1040–1048. https://doi.org/10.1039/C6NR05853A.
(32) Podewitz, M.; Wang, Y.; Gkeka, P.; von Grafenstein, S.; Liedl, K. R.; Cournia, Z. Phase Diagram of a Stratum Corneum Lipid Mixture. *J. Phys. Chem. B* **2018**, *122* (46), 10505–10521. https://doi.org/10.1021/acs.jpcb.8b07200.
(33) Wang, Y.; Gkeka, P.; Fuchs, J. E.; Liedl, K. R.; Cournia, Z. DPPC-Cholesterol Phase Diagram Using Coarse-Grained Molecular Dynamics Simulations. *Biochimica et Biophysica Acta (BBA) - Biomembranes* **2016**, *1858* (11), 2846–2857. https://doi.org/10.1016/j.bbamem.2016.08.005.
(34) Lamprakis, C.; Andreadelis, I.; Manchester, J.; Velez-Vega, C.; Duca, J. S.; Cournia, Z. Evaluating the Efficiency of the Martini Force Field to Study Protein Dimerization in Aqueous and Membrane Environments. *J. Chem. Theory Comput.* **2021**, *17* (5), 3088–3102. https://doi.org/10.1021/acs.jctc.0c00507.
(35) Atsmon-Raz, Y.; Tieleman, D. P. Parameterization of Palmitoylated Cysteine, Farnesylated Cysteine, Geranylgeranylated Cysteine, and Myristoylated Glycine for the Martini Force Field. *J. Phys. Chem. B* **2017**, *121* (49), 11132–11143. https://doi.org/10.1021/acs.jpcb.7b10175.
(36) Souza, P. C. T.; Alessandri, R.; Barnoud, J.; Thallmair, S.; Faustino, I.; Grünewald, F.; Patmanidis, I.; Abdizadeh, H.; Bruininks, B. M. H.; Wassenaar, T. A.; Kroon, P. C.; Melcr, J.; Nieto, V.; Corradi, V.; Khan, H. M.; Domański, J.; Javanainen, M.; Martinez-Seara, H.; Reuter, N.; Best, R. B.; Vattulainen, I.; Monticelli, L.; Periole, X.; Tieleman, D. P.; de Vries, A. H.; Marrink, S. J. Martini 3: A General Purpose Force Field for Coarse-Grained Molecular Dynamics. *Nat Methods* **2021**, *18* (4), 382–388. https://doi.org/10.1038/s41592-021-01098-3.
(37) Jong, D. H. de; Lopez, C. A.; Marrink, S. J. Molecular View on Protein Sorting into Liquid-Ordered Membrane Domains Mediated by Gangliosides and Lipid Anchors. *Faraday Discuss.* **2013**, *161* (0), 347–363. https://doi.org/10.1039/C2FD20086D.
(38) Ma, Z.; Shi, S.; Ren, M.; Pang, C.; Zhan, Y.; An, H.; Sun, F. Molecular Mechanism of CD44 Homodimerization Modulated by Palmitoylation and Membrane Environments. *Biophysical Journal* **2022**, *121* (14), 2671–2683. https://doi.org/10.1016/j.bpj.2022.06.021.
(39) Thukral, L.; Sengupta, D.; Ramkumar, A.; Murthy, D.; Agrawal, N.; Gokhale, R. S. The Molecular Mechanism Underlying Recruitment and Insertion of Lipid-Anchored LC3 Protein into Membranes. *Biophysical Journal* **2015**, *109* (10), 2067–2078. https://doi.org/10.1016/j.bpj.2015.09.022.
(40) Zeppelin, T.; Pedersen, K. B.; Berglund, N. A.; Periole, X.; Schiøtt, B. Effect of Palmitoylation on the Dimer Formation of the Human Dopamine Transporter. *Sci Rep* **2021**, *11* (1), 4164. https://doi.org/10.1038/s41598-021-83374-y.
(41) Lin, X.; Gorfe, A. A.; Levental, I. Protein Partitioning into Ordered Membrane Domains: Insights from Simulations. *Biophysical Journal* **2018**, *114* (8), 1936–1944. https://doi.org/10.1016/j.bpj.2018.03.020.
(42) Huang, J.; Rauscher, S.; Nawrocki, G.; Ran, T.; Feig, M.; de Groot, B. L.; Grubmüller, H.; MacKerell, A. D. CHARMM36m: An Improved Force Field for Folded and Intrinsically Disordered Proteins. *Nat Methods* **2017**, *14* (1), 71–73. https://doi.org/10.1038/nmeth.4067.

(43) Berendsen, H. J. C.; van der Spoel, D.; van Drunen, R. GROMACS: A Message-Passing Parallel Molecular Dynamics Implementation. *Computer Physics Communications* **1995**, *91* (1–3), 43–56. https://doi.org/10.1016/0010-4655(95)00042-E.
(44) Abraham, M. J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J. C.; Hess, B.; Lindahl, E. GROMACS: High Performance Molecular Simulations through Multi-Level Parallelism from Laptops to Supercomputers. *SoftwareX* **2015**, *1–2*, 19–25. https://doi.org/10.1016/j.softx.2015.06.001.
(45) Jo, S.; Kim, T.; Iyer, V. G.; Im, W. CHARMM-GUI: A Web-Based Graphical User Interface for CHARMM. *Journal of Computational Chemistry* **2008**, *29* (11), 1859–1865. https://doi.org/10.1002/jcc.20945.
(46) Brooks, B. R.; Brooks III, C. L.; Mackerell Jr., A. D.; Nilsson, L.; Petrella, R. J.; Roux, B.; Won, Y.; Archontis, G.; Bartels, C.; Boresch, S.; Caflisch, A.; Caves, L.; Cui, Q.; Dinner, A. R.; Feig, M.; Fischer, S.; Gao, J.; Hodoscek, M.; Im, W.; Kuczera, K.; Lazaridis, T.; Ma, J.; Ovchinnikov, V.; Paci, E.; Pastor, R. W.; Post, C. B.; Pu, J. Z.; Schaefer, M.; Tidor, B.; Venable, R. M.; Woodcock, H. L.; Wu, X.; Yang, W.; York, D. M.; Karplus, M. CHARMM: The Biomolecular Simulation Program. *Journal of Computational Chemistry* **2009**, *30* (10), 1545–1614. https://doi.org/10.1002/jcc.21287.
(47) Wu, E. L.; Cheng, X.; Jo, S.; Rui, H.; Song, K. C.; Dávila-Contreras, E. M.; Qi, Y.; Lee, J.; Monje-Galvan, V.; Venable, R. M.; Klauda, J. B.; Im, W. CHARMM-GUI Membrane Builder toward Realistic Biological Membrane Simulations. *Journal of Computational Chemistry* **2014**, *35* (27), 1997–2004. https://doi.org/10.1002/jcc.23702.
(48) Schrödinger, LLC. The PyMOL Molecular Graphics System, Version 2.3, 2019.
(49) Park, S.-J.; Kern, N.; Brown, T.; Lee, J.; Im, W. CHARMM-GUI PDB Manipulator: Various PDB Structural Modifications for Biomolecular Modeling and Simulation. *Journal of Molecular Biology* **2023**, 167995. https://doi.org/10.1016/j.jmb.2023.167995.
(50) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of Simple Potential Functions for Simulating Liquid Water. *J. Chem. Phys.* **1983**, *79* (2), 926–935. https://doi.org/10.1063/1.445869.
(51) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A.; Haak, J. R. Molecular Dynamics with Coupling to an External Bath. *J. Chem. Phys.* **1984**, *81* (8), 3684–3690. https://doi.org/10.1063/1.448118.
(52) Nosé, S. A Unified Formulation of the Constant Temperature Molecular Dynamics Methods. *J. Chem. Phys.* **1984**, *81* (1), 511–519. https://doi.org/10.1063/1.447334.
(53) Hoover, W. G. Canonical Dynamics: Equilibrium Phase-Space Distributions. *Phys. Rev. A* **1985**, *31* (3), 1695–1697. https://doi.org/10.1103/PhysRevA.31.1695.
(54) Hoover, W. G.; Holian, B. L. Kinetic Moments Method for the Canonical Ensemble Distribution. *Physics Letters A* **1996**, *211* (5), 253–257. https://doi.org/10.1016/0375-9601(95)00973-6.
(55) Parrinello, M.; Rahman, A. Polymorphic Transitions in Single Crystals: A New Molecular Dynamics Method. *Journal of Applied Physics* **1981**, *52* (12), 7182–7190. https://doi.org/10.1063/1.328693.
(56) Páll, S.; Hess, B. A Flexible Algorithm for Calculating Pair Interactions on SIMD Architectures. *Computer Physics Communications* **2013**, *184* (12), 2641–2650. https://doi.org/10.1016/j.cpc.2013.06.003.
(57) 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* (12), 10089–10092. https://doi.org/10.1063/1.464397.
(58) Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. A Smooth Particle Mesh Ewald Method. *J. Chem. Phys.* **1995**, *103* (19), 8577–8593. https://doi.org/10.1063/1.470117.

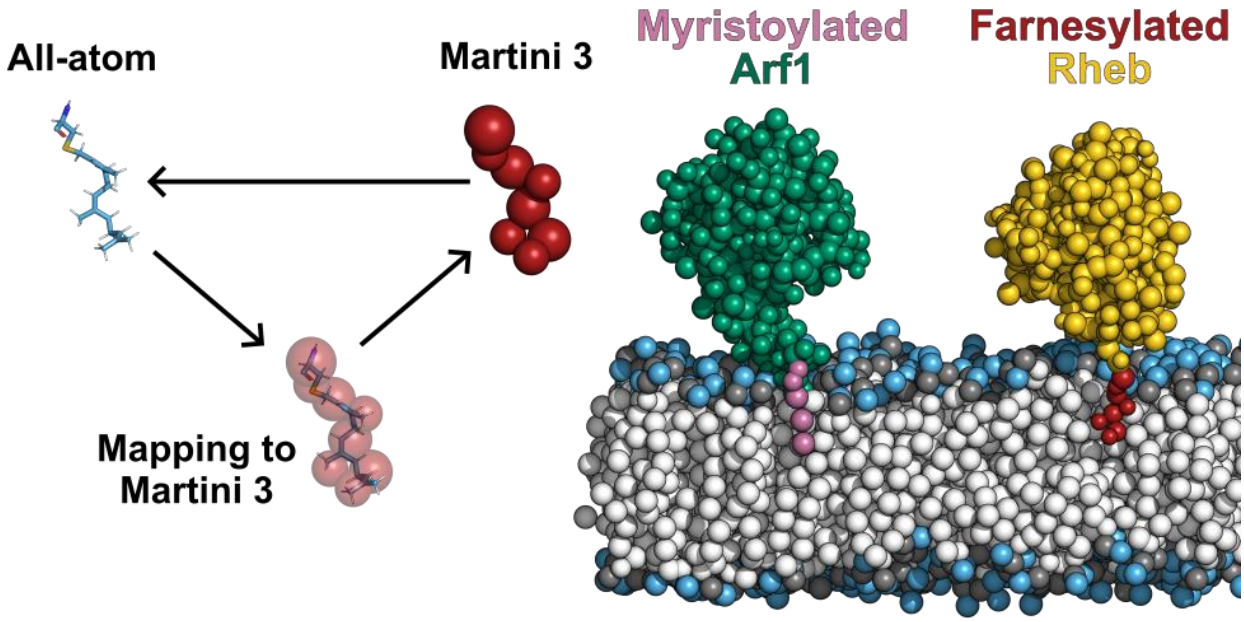
(59) Michaud-Agrawal, N.; Denning, E. J.; Woolf, T. B.; Beckstein, O. MDAnalysis: A Toolkit for the Analysis of Molecular Dynamics Simulations. *Journal of Computational Chemistry* **2011**, *32* (10), 2319–2327. https://doi.org/10.1002/jcc.21787.
(60) Gowers, R. J.; Linke, M.; Barnoud, J.; Reddy, T. J. E.; Melo, M. N.; Seyler, S. L.; Domański, J.; Dotson, D. L.; Buchoux, S.; Kenney, I. M.; Beckstein, O. MDAnalysis: A Python Package for the Rapid Analysis of Molecular Dynamics Simulations. *Proceedings of the 15th Python in Science Conference* **2016**, 98–105. https://doi.org/10.25080/Majora-629e541a-00e.
(61) PubChem. *S-Methyl thioacetate*. https://pubchem.ncbi.nlm.nih.gov/compound/73750 (accessed 2023-05-29).
(62) Empereur-mot, C.; Pedersen, K. B.; Capelli, R.; Crippa, M.; Caruso, C.; Perrone, M.; Souza, P. C. T.; Marrink, S. J.; Pavan, G. M. Automatic Optimization of Lipid Models in the Martini Force Field Using SwarmCG. *J. Chem. Inf. Model.* **2023**, *63* (12), 3827–3838. https://doi.org/10.1021/acs.jcim.3c00530.
(63) Kroon, P. C.; Grunewald, F.; Barnoud, J.; Tilburg, M. van; Souza, P. C. T.; Wassenaar, T. A.; Marrink, S. J. Martinize2 and Vermouth: Unified Framework for Topology Generation. *eLife* **2023**, *12*. https://doi.org/10.7554/eLife.90627.
(64) Wassenaar, T. A.; Ingólfsson, H. I.; Böckmann, R. A.; Tieleman, D. P.; Marrink, S. J. Computational Lipidomics with Insane: A Versatile Tool for Generating Custom Membranes for Molecular Simulations. *J. Chem. Theory Comput.* **2015**, *11* (5), 2144–2155. https://doi.org/10.1021/acs.jctc.5b00209.
(65) Bussi, G.; Donadio, D.; Parrinello, M. Canonical Sampling through Velocity Rescaling. *The Journal of Chemical Physics* **2007**, *126* (1), 014101. https://doi.org/10.1063/1.2408420.
(66) de Jong, D. H.; Baoukina, S.; Ingólfsson, H. I.; Marrink, S. J. Martini Straight: Boosting Performance Using a Shorter Cutoff and GPUs. *Computer Physics Communications* **2016**, *199*, 1–7. https://doi.org/10.1016/j.cpc.2015.09.014.
(67) Yamagata, K.; Sanders, L. K.; Kaufmann, W. E.; Yee, W.; Barnes, C. A.; Nathans, D.; Worley, P. F. Rheb, a Growth Factor- and Synaptic Activity-Regulated Gene, Encodes a Novel Ras-Related Protein. *J Biol Chem* **1994**, *269* (23), 16333–16339.
(68) Liu, Y.; Kahn, R. A.; Prestegard, J. H. Structure and Membrane Interaction of Myristoylated ARF1. *Structure* **2009**, *17* (1), 79–87. https://doi.org/10.1016/j.str.2008.10.020.
(69) Prakash, P.; Gorfe, A. A. Determinants of Membrane Orientation Dynamics in Lipid-Modified Small GTPases. *JACS Au* **2022**, *2* (1), 128–135. https://doi.org/10.1021/jacsau.1c00426.
(70) Yu, Y.; Li, S.; Xu, X.; Li, Y.; Guan, K.; Arnold, E.; Ding, J. Structural Basis for the Unique Biological Function of Small GTPase RHEB *. *Journal of Biological Chemistry* **2005**, *280* (17), 17093–17100. https://doi.org/10.1074/jbc.M501253200.
(71) Ren, X.; Farías, G. G.; Canagarajah, B. J.; Bonifacino, J. S.; Hurley, J. H. Structural Basis for Recruitment and Activation of the AP-1 Clathrin Adaptor Complex by Arf1. *Cell* **2013**, *152* (4), 755–767. https://doi.org/10.1016/j.cell.2012.12.042.
(72) Liu, Y.; Kahn, R. A.; Prestegard, J. H. Dynamic Structure of Membrane-Anchored Arf•GTP. *Nat Struct Mol Biol* **2010**, *17* (7), 876–881. https://doi.org/10.1038/nsmb.1853.
(73) Daura, X.; Gademann, K.; Jaun, B.; Seebach, D.; van Gunsteren, W. F.; Mark, A. E. Peptide Folding: When Simulation Meets Experiment. *Angewandte Chemie International Edition* **1999**, *38* (1–2), 236–240. https://doi.org/10.1002/(SICI)1521-3773(19990115)38:1/2<236::AID-ANIE236>3.0.CO;2-M.

(74) Vickery, O. N.; Stansfeld, P. J. CG2AT2: An Enhanced Fragment-Based Approach for Serial Multi-Scale Molecular Dynamics Simulations. *J. Chem. Theory Comput.* **2021**, *17* (10), 6472–6482. https://doi.org/10.1021/acs.jctc.1c00295.
(75) Louison, K. A.; Dryden, I. L.; Laughton, C. A. GLIMPS: A Machine Learning Approach to Resolution Transformation for Multiscale Modeling. *J. Chem. Theory Comput.* **2021**, *17* (12), 7930–7937. https://doi.org/10.1021/acs.jctc.1c00735.
(76) Rowland, R. S.; Taylor, R. Intermolecular Nonbonded Contact Distances in Organic Crystal Structures: Comparison with Distances Expected from van Der Waals Radii. *J. Phys. Chem.* **1996**, *100* (18), 7384–7391. https://doi.org/10.1021/jp953141+.
(77) Bondi, A. Van Der Waals Volumes and Radii. *J. Phys. Chem.* **1964**, *68* (3), 441–451. https://doi.org/10.1021/j100785a001.
(78) Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; Vanderplas, J.; Passos, A.; Cournapeau, D.; Brucher, M.; Perrot, M.; Duchesnay, É. Scikit-Learn: Machine Learning in Python. *Journal of Machine Learning Research* **2011**, *12* (85), 2825–2830.
(79) Campello, R. J. G. B.; Moulavi, D.; Sander, J. Density-Based Clustering Based on Hierarchical Density Estimates. In *Advances in Knowledge Discovery and Data Mining*; Pei, J., Tseng, V. S., Cao, L., Motoda, H., Xu, G., Eds.; Lecture Notes in Computer Science; Springer: Berlin, Heidelberg, 2013; pp 160–172. https://doi.org/10.1007/978-3-642-37456-2_14.
(80) Campello, R. J. G. B.; Moulavi, D.; Zimek, A.; Sander, J. Hierarchical Density Estimates for Data Clustering, Visualization, and Outlier Detection. *ACM Trans. Knowl. Discov. Data* **2015**, *10* (1), 5:1-5:51. https://doi.org/10.1145/2733381.
(81) Richards, F. M. Areas, Volumes, Packing and Protein Structure. *Annu Rev Biophys Bioeng* **1977**, *6*, 151–176. https://doi.org/10.1146/annurev.bb.06.060177.001055.
(82) Connolly, M. L. Analytical Molecular Surface Calculation. *Journal of Applied Crystallography* **1983**, *16* (5), 548–558. https://doi.org/10.1107/S0021889883010985.
(83) Grünewald, F.; Punt, M. H.; Jefferys, E. E.; Vainikka, P. A.; König, M.; Virtanen, V.; Meyer, T. A.; Pezeshkian, W.; Gormley, A. J.; Karonen, M.; Sansom, M. S. P.; Souza, P. C. T.; Marrink, S. J. Martini 3 Coarse-Grained Force Field for Carbohydrates. *J. Chem. Theory Comput.* **2022**, *18* (12), 7555–7569. https://doi.org/10.1021/acs.jctc.2c00757.
(84) Borges-Araújo, L.; Souza, P. C. T.; Fernandes, F.; Melo, M. N. Improved Parameterization of Phosphatidylinositide Lipid Headgroups for the Martini 3 Coarse-Grain Force Field. *J. Chem. Theory Comput.* **2022**, *18* (1), 357–373. https://doi.org/10.1021/acs.jctc.1c00615.
(85) *Bonded interactions - GROMACS 2023.1 documentation*. https://manual.gromacs.org/current/reference-manual/functions/bonded-interactions.html#proper-dihedrals-periodic-type (accessed 2023-04-27).
(86) Vainikka, P.; Thallmair, S.; Souza, P. C. T.; Marrink, S. J. Martini 3 Coarse-Grained Model for Type III Deep Eutectic Solvents: Thermodynamic, Structural, and Extraction Properties. *ACS Sustainable Chem. Eng.* **2021**, *9* (51), 17338–17350. https://doi.org/10.1021/acssuschemeng.1c06521.
(87) Alessandri, R.; Barnoud, J.; Gertsen, A. S.; Patmanidis, I.; de Vries, A. H.; Souza, P. C. T.; Marrink, S. J. Martini 3 Coarse-Grained Force Field: Small Molecules. *Advanced Theory and Simulations* **2022**, *5* (1), 2100391. https://doi.org/10.1002/adts.202100391.
(88) Brosz, M.; Michelarakis, N.; F. Bunz, U. H.; Aponte-Santamaría, C.; Gräter, F. Martini 3 Coarse-Grained Force Field for Poly(Para -Phenylene Ethynylene)s. *Physical Chemistry Chemical Physics* **2022**, *24* (17), 9998–10010. https://doi.org/10.1039/D1CP04237H.
(89) Borges-Araújo, L.; Borges-Araújo, A. C.; Ozturk, T. N.; Ramirez-Echemendia, D. P.; Fábián, B.; Carpenter, T. S.; Thallmair, S.; Barnoud, J.; Ingólfsson, H. I.; Hummer, G.;

Tieleman, D. P.; Marrink, S. J.; Souza, P. C. T.; Melo, M. N. Martini 3 Coarse-Grained Force Field for Cholesterol. *J. Chem. Theory Comput.* **2023**. https://doi.org/10.1021/acs.jctc.3c00547.

(90) Mysore, V. P.; Zhou, Z.-W.; Ambrogio, C.; Li, L.; Kapp, J. N.; Lu, C.; Wang, Q.; Tucker, M. R.; Okoro, J. J.; Nagy-Davidescu, G.; Bai, X.; Plückthun, A.; Jänne, P. A.; Westover, K. D.; Shan, Y.; Shaw, D. E. A Structural Model of a Ras–Raf Signalosome. *Nat Struct Mol Biol* **2021**, *28* (10), 847–857. https://doi.org/10.1038/s41594-021-00667-6.

TOC Graphic


All-atom
Martini 3
Myristoylated
Arf1
Farnesylated
Rheb
Mapping to
Martini 3