Scaling protein-water interactions in the Martini 3 coarse-grained force field to simulate transmembrane helix dimers in different lipid environments

Ainara Claveras Cabezudo[†,‡], *Christina Athanasiou*[†,‡,§], *Alexandros Tsengenes*[†,‡,§], *Rebecca C. Wade*[*,†,‡,∥,⊥]

† Molecular and Cellular Modeling Group, Heidelberg Institute for Theoretical Studies (HITS), Schloss-Wolfsbrunnenweg 35, 69118 Heidelberg, Germany

‡ Faculty of Biosciences, Heidelberg University, Im Neuenheimer Feld 234, 69120 Heidelberg, Germany

§ Heidelberg Biosciences International Graduate School, Heidelberg University, Im Neuenheimer Feld 501, 69120 Heidelberg, Germany

∥ Center for Molecular Biology (ZMBH), DKFZ-ZMBH Alliance, Heidelberg University, Im Neuenheimer Feld 282, 69120 Heidelberg, Germany

⊥ Interdisciplinary Center for Scientific Computing (IWR), Heidelberg University, Im Neuenheimer Feld 205, 69120 Heidelberg, Germany

ABSTRACT

Martini 3, the latest version of the widely used Martini force field for coarse-grained molecular dynamics simulations, is a promising tool to investigate proteins in phospholipid bilayers. However, simulating other lipid environments, such as detergent micelles, presents challenges due to the absence of validated parameters for their constituent molecules. Here, we propose parameters for the micelle-forming surfactant, dodecylphosphocholine (DPC). These result in micelle assembly with aggregation numbers in agreement with experimental values. However, we identified a lack of hydrophobic interactions between transmembrane helix protein dimers and the tails of DPC molecules, preventing insertion and stabilization of the protein in the micelles. This problem was also observed for protein insertion in self-assembled 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine (POPC) bilayers. We propose the reduction of the non-bonded interactions between protein and water beads by 10% as a simple and effective solution to this problem that does not alter protein dimerization or bilayer structure.

INTRODUCTION

Molecular dynamics (MD) simulation is a powerful approach to explore the behaviour of biomolecules at a spatio-temporal resolution that cannot be obtained with experimental techniques (Karplus & McCammon, 2002) (Bottaro & Lindorff-Larsen, 2018). This method can, however, be limited by a lack of sampling of the relevant configuration space, which thus compromises the plausibility of the results obtained. In atomistic MD simulations, the length of the simulations is limited by the high number of computationally expensive calculations required. Coarse-grained (CG) simulations alleviate this problem by grouping several atoms into "beads", thus reducing the number of degrees of freedom, and therefore, the number of calculations required. This, for a given amount of computation time, much longer simulation times can be achieved.

The Martini force field is one of the most widely used CG models due to its good performance and simplicity. Contrary to all-atom calculations, this method does not aim to provide a detailed structural view of the system, but to provide a broader picture of the processes taking place over a longer time scale (Voth, 2008). Martini beads usually represent groups of four non-hydrogen atoms and have different degrees of polarity. Parameters for specific beads are assigned against thermodynamic data, particularly oil/water partition coefficients. The Martini model has been shown to be able to reproduce a considerable number of experimental properties (Marrink, et al., 2007).

A new, improved version of the Martini force field has recently been released (Souza, et al., 2021). This improved version, referred to as Martini 3, overcomes some problems that have been reported for previous versions of the force field, such as an excessive aggregation of soluble and transmembrane proteins (Stark, et al., 2013) (Javanainen, et al., 2017), or an excessive stacking of nucleotides (Alessandri, et al., 2019) (Uusitalo, et al., 2015). One of the most important changes in Martini 3 is the introduction of a specific type of bead for water (W), which has been parameterized separately from the rest of the beads.

Since its release, Martini 3 has been reported to present a number of issues, such as numeric instability of simulations with large time steps and an underestimation of the dimensions of intrinsically disordered proteins (Fabian, et al., 2022) (Thomasen, et al., 2022). An increase of the bead size or the mass of ion beads has been proposed to overcome the first problem, while an up-scaling of non-bonded interactions between protein and water beads resolved the second one.

Detergent micelles are frequently used to experimentally determine the structure of transmembrane proteins, as they mimic the hydrophobic environment of phospholipid bilayers (Kallick, et al., 1995) (Warschawski, et al., 2011). Hence, many experimental structures obtained in micelles, are used as starting structures in molecular dynamics simulations. The fact that these simulations are usually run for the proteins in phospholipid bilayers hinders a direct comparison to experimental data. The simulation of these transmembrane proteins in detergent micelles is thus necessary to understand the influence of the lipid environment on their three-dimensional structure and dynamics. However, parameters for micelle-forming

lipids in Martini 3 are currently not available. To overcome this problem, the widely used surfactant dodecylphosphocholine (DPC) was parameterized in this study on the basis of the existing parameters for other phospholipids in Martini 3 and using the parameters of DPC in the previous version of the force field, Martini 2, as a reference. Simulations at different lipid concentrations showed that the newly parameterized DPC molecules were able to assemble into micelles with properties similar to those obtained by different experimental techniques. However, we found that transmembrane (TM) helix dimer proteins fail to insert into the DPC micelles in Martini 3: they interacted with water beads or with the polar head groups of surfactant molecules rather than with their hydrophobic tails. Moreover, the transmembrane helix proteins rarely inserted into 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine (POPC) bilayers during bilayer self-assembly. We here show that reducing protein-water interactions by introducing a small uniform scaling factor of the Lennard-Jones interactions in Martini 3 overcomes this aberrant behaviour for TM helix dimer proteins, leading to their correct insertion in micelles and self-assembled phospholipid bilayers.

MATERIAL AND METHODS

System preparation

The structure of the TM helix dimer of glycophorin A (GpA), which was obtained by NMR in DPC micelles, was used in most of the simulations in which protein was present (PDB ID: 1AFO (MacKenzie, et al., 1997)). The martinize2.py script (Souza, et al., 2021) was used with DSSP (Kabsch & Sander, 1983) to map atomistic structures to CG models. An elastic network was applied to the helical region of the TM monomers (residues 73 to 95) using a force constant of 700 kJ/mol nm^2 and distance cut-offs of 0.5 and 0.9 nm to preserve secondary structure, while retaining the intrinsic flexibility of the N-loop and C-loop regions.
Some test simulations were also run for the TM helix dimer of the tyrosine kinase neurotrophin receptor TrkA, whose structure was also determined by NMR in DPC micelles (PDB ID: 2N90 (Franco, et al., 2020)). The CG model of TrkA was obtained from the atomistic structure using an analogous procedure to that for GpA.
To explore micelle formation at different surfactant concentrations, simulations were run in boxes of different sizes and with different numbers of DPC molecules. In all cases, DPC molecules were placed at random positions at the beginning of the trajectory. Each system was solvated in water and neutralized in 0.15 M NaCl.
To investigate POPC bilayer self-assembly, 250 POPC molecules and the protein dimer were placed at random positions in a 10x10x10 nm^3 with subsequent solvation in water and neutralization in 0.15 M NaCl. Some simulations were initiated using the *INSANE* method to insert proteins into a preformed lipid bilayer (Wassenaar, et al., 2015).

CG-MD simulations

CG simulations were performed using the 2020.4 version of the GROMACS software package (Lindahl, et al., 2020) and the Martini 3 force field (Souza, et al., 2021). A series of parameters which have been previously shown to boost performance were used (De Jong, et al., 2016). These include the use of the Verlet neighbour search (VNS) algorithm and cut-offs of 1.1 nm for Lennard Jones and Coulomb potentials. The velocity rescale thermostat and the Parrinello-Rahman barostat (Bussi, et al., 2007) (Parrinello & Rahman, 1981) were used with coupling parameters of 1.0 and 12.0 ps, respectively. A compressibility of $3e^{-4}$ was used, as well as a temperature of 310 K and a pressure of 1 bar. A time step of 20 fs was set and the neighbour list was updated every 10 time steps.
For each system, different replicas were run using random initial velocities and positions (A summary of all simulations can be found in **Supplementary Table 1**). After an energy minimization step, equilibration and production runs were performed in the NPT ensemble. Position restraints with a force constant of 4000 kJ/mol nm^2 were applied to the backbone beads of the TM region of the protein during equilibration and released for production. Simulations of DPC micelles were run using isotropic pressure coupling. In simulations of POPC bilayer self-assembly, isotropic pressure coupling was used for the first 60 ns of equilibration. Bilayer formation was visually assessed afterwards, and, if required, a rotation of the entire system was applied to place the bilayer in the X-Y plane. The remainder of the simulation was then run using semi-isotropic pressure coupling.
Some simulations were run with Martini 2.2 in order to compare the two versions of the force field. In these Martini 2.2 simulations, cut-offs of 1.2 nm were used for the Lennard-Jones and Coulomb potentials. The non-polarizable water model was used in the Martini 2.2 simulations. All other simulation parameters were the same as for the Martini 3 simulations.

Scaling of non-bonded interactions in Martini 3

In order to scale the non-bonded interactions of the force field, a uniform scaling factor was applied on energy well depths ε_{ij} of the Lennard Jones potentials of all protein-water bead pairs, as has been described previously (Majumder & Straub, 2021).

$$\varepsilon_{ij}^{new} = \alpha\varepsilon_{ij}$$

Data analysis and visualisation

Some basic analysis tasks, such as the calculation of the root mean square deviation (RMSD) or temperature and pressure monitoring were performed using the tools provided in the GROMACS software package. More complex analyses, such as the calculation of dihedral angles or the distances between centers of geometry of the two helices were implemented in the Python programming language (Van Rossum & Drake, 2009) with extensive use of the MDAnalysis (Michaud-Agrawal, et al., 2011) (Gowers, et al., 2016), NumPy (Harris, et al., 2020) and Pandas (McKinney, 2010) packages.
In order to evaluate the exploration of the conformational landscape, kernel density heatmaps for different structural parameters were plotted using the Seaborn Python package (Waskom, 2021). Further data analysis was performed using the R language for statistical computing (R Core Team, 2022). PyMol (Schrödinger, 2020) and VMD (Humphrey, et al., 1996) were used for visualization of the trajectories, and the Tachyon ray-tracer was used for image generation (Stone, 1998).

Structural parameters of transmembrane helix dimers

To explore the conformational space of the helix dimers, three different dihedral angles were defined. These were calculated after centering on one of the two helices (referred to as the "first") and aligning each frame of the trajectory to this centred helix. The *crossing angle* was defined as the dihedral angle between the vectors that connect the centroid of each helix with the centroid of its upper half (Domanski, et al., 2018). Furthermore, rotation angles of the second helix with respect to the first helix and to itself, referred to as *position* and *phase*, respectively, were calculated. The position is represented by the dihedral angle between the vector connecting the centroid of each helix and the vector that connects the centroid of the upper half of the first helix with a residue located in the dimer interface in the same helix (Wassenaar, et al., 2015). To calculate the phase, the dihedral angle between the line that passes through the centroids of both helices and the one that connects the reference residue in the second helix and the centroid of its upper half was computed.

Contact analysis

To compute the percentage occupancies of contacts between TM helix dimers during the simulations, all replicas were merged into a single trajectory and the whole system was aligned to one TM helix. Broken molecules (due to the use of a periodic simulation box) were made whole using the *gmx trjconv* command in GROMACS. By TCL scripting, contacts between the beads of the two helices were defined using a cut-off distance of 6 Å and contact occupancies were calculated for each bead pair. The contact occupation of a pair of residues was defined as the highest contact percentage between any beads of the two residues.

Analysis of bilayer properties

The density of phosphate beads, which was used to determine bilayer thickness, was computed using the GROMACS tool *gmx density*. To obtain the lipid lateral diffusion, jumps over periodic boundaries were removed and *gmx msd* was used upon removal of the overall center of mass motion. The second rank lipid order parameter (P2) was calculated for each bond of POPC using the equation

$$\mathbf{P_2 = 1/2\ (3\ \cos^2\langle\theta\rangle - 1),}$$

where θ represents the angle between the bilayer normal and the given bond (Marrink, et al., 2004). This was calculated using the script *do-order-gmx5.py* (available at www.cgmartini.nl/index.php/downloads/tools/229-do-order). The area per lipid was calculated for each bilayer leaflet separately using the membrane analysis tool Gridmat-MD (Allen, et al., 2009), which makes use of a grid to consider the area occupied by the TM protein. A grid with 50x50 points was used, and the area per lipid of all POPC molecules was computed every 5th frame of the trajectory.

RESULTS

Parameterization of DPC and characterization of DPC micelles

Due to its similarity, the parameters for DPC were mostly derived from those of POPC. Hence, choline and phosphate groups were mapped to beads of type Q1 and Q5, respectively. The hydrocarbon chain, which contains 12 carbon atoms, was described using 3 beads of type C1 (**Figure 1A**). This bead type is used in the hydrophobic tails of all saturated phospholipids in Martini 3. A force constant of 7000 kJ/mol nm^2 was used to describe the bond between the phosphate and the choline beads, with a bond length of 0.40 nm. As this bond is present in several phospholipids, such as POPC, its force constant and length were directly obtained from the parameters for these lipids in Martini 3. For the rest of the bonds, force constants of 3800 kJ/mol nm^2 and lengths of 0.47 nm were used. These parameters are used to describe bonds between beads in the hydrocarbon chains of all saturated phospholipids in Martini 3. The bond between the phosphate bead and the first bead of the alkyl chain was assigned the same parameters as all the bonds between beads in the tail of the lipid, as done for DPC in Martini 2. To maintain the shape of the DPC monomers, an angle of 180° between the three beads of the hydrophobic chain was maintained with a force constant of 35 kJ/mol nm^2. Due to the similarity between the hydrocarbon chain of DPC and the hydrocarbon chains of bilayer-forming saturated phospholipids, this force constant was directly obtained from the POPC parameters.

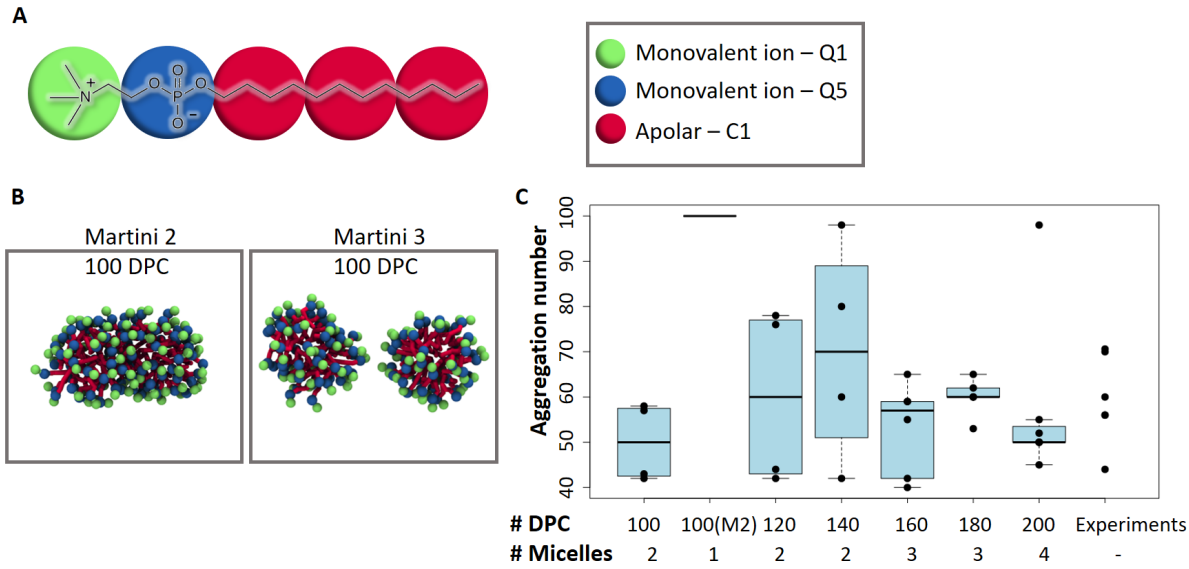

Figure 1: Parameterization of DPC in Martini 3. A: Beads of type Q1, Q5 and C1 were used to represent the choline, phosphate and hydrophobic tail groups, respectively. **B**: In simulations in a 10x10x10 nm^3 box with 100 DPC molecules (at a concentration of 166 mM) using Martini 2, a single micelle containing 100 lipids was observed. Using the same conditions and Martini 3, two smaller micelles of 45-55 molecules were obtained. **C**: Aggregation numbers obtained with Martini 3 at different DPC concentrations, compared to those obtained with Martini 2 (labelled M2) and experiments.

Table 1: Summary of micelle assembly simulations with different DPC concentrations. For each system, 2 replicas of 5 μs each were run at 310 K using isotropic pressure coupling.

Martini version	# DPC	[DPC] (mM)	Box size (nm)	# Micelles	Aggregation number
Martini 3	100	166	10 x 10 x 10	2	43-57
Martini 3	120	199	10 x 10 x 10	2	44-76
Martini 3	140	233	10 x 10 x 10	2	60-80
Martini 3	160	266	10 x 10 x 10	3	40-65
Martini 3	180	299	10 x 10 x 10	3	60
Martini 3	200	333	10 x 10 x 10	4	45-55
Martini 3	100	125	11 x 11 x 11	2	44-58
Martini 3	100	96	12 x 12 x 12	2	43-57
Martini 3	100	76	13 x 13 x 13	2	43-57
Martini 2	100	166	10 x 10 x 10	1	100
Martini 2	100	100	12 x 12 x 12	1	100

To investigate the ability of DPC molecules to assemble into micelles, simulations were run at different lipid concentrations (**Table 1**). Simulations using Martini 2 were computed as well in order to compare the behaviour of the preliminary DPC parameters to those of the previous version of the force field, which have been made publicly available by its developers (and were obtained from http://www.cgmartini.nl/index.php/force-field-parameters/lipids).

The most remarkable difference between DPC micelles in Martini 2 and Martini 3 was their aggregation number: whereas simulations with 100 DPC monomers (166 mM) showed the

formation of a single, big micelle in Martini 2, two smaller micelles were formed in Martini 3. The distribution of lipids between them was uneven, leading to one consisting of about 57 and the other about 43 molecules (**Figure 1B**).

In simulations with 120 DPC molecules in Martini 3, formation of two micelles was observed as well; in this case, however, one micelle was significantly bigger than the other, with ca. 76 lipids. When 140 DPC molecules were included in the simulation box, two micelles with around 60 and 80 lipids were formed. 160 DPC monomers assembled into 3 micelles with ~ 40, 55 and 65 molecules, and 180 lipids formed 3 more uniform micelles with ~ 60 monomers each. When the DPC concentration was further increased up to 333 mM (200 DPC molecules), 4 micelles were formed. The polar groups of adjacent micelles tended to interact with each other, eventually leading to the fusion of two micelles into elongated, tubular structures with ~ 98 monomers each (**Figure 1C**).

In general, the aggregation numbers obtained with the derived DPC parameters in Martini 3 show better agreement with experimental values than those obtained using Martini 2. Although the DPC aggregation numbers described in the literature vary considerably, which can be due to differences in the method, temperature or lipid concentration, they generally agree with the results obtained here. NMR spectroscopy experiments with DPC concentrations of 228 mM and 98 mM showed aggregation numbers of 44 ± 5 and 56, respectively (Kallick, et al., 1995) (Gao & Wong, 1998). Higher aggregation numbers have been observed with small-angle X-ray scattering (SAXS, 60-70 molecules) and small-angle neutron scattering (SANS, 70.6 ± 5 molecules) (Lipfert, et al., 2007) (Pambou, et al., 2015). Ultracentrifugation experiments showed aggregation numbers of 56 ± 5 at 20 mM and 294 K (Lauterwein, et al., 1979).

Similar aggregation numbers (~ 54 molecules) have been described in atomistic MD simulations using different force fields (Wymore, et al., 1999) (Marrink, et al., 2000) (Abel, et al., 2012). CG MD simulations using Martini 2 showed a high variability of aggregation numbers: at a temperature of 300 K and a concentration of 40 mM, micelles of 40 to 70 molecules have been reported (Marrink, et al., 2004), whereas at 300 K and a lipid concentration of 126 mM (175 molecules), micelles consisting of 45 monomers were formed (Sanders & Panagiotopoulos, 2010). A summary of all DPC aggregation numbers obtained in previous experimental and simulation studies can be found in **Table 2**.

Table 2: DPC aggregation numbers reported in the literature.

Method	Temperature (K)	Aggregation number	[DPC] (mM)	Reference
NMR[1]	310	44 ± 5	228	(Kallick, et al., 1995)
NMR[1]	298	56	98	(Gao & Wong, 1998)
SAXS[2]	298	60-70	77	(Lipfert, et al., 2007)
SANS[3]	295	70.6 ± 5	100	(Pambou, et al., 2015)
Ultracentrifugation	294	56 ± 5	20	(Lauterwein, et al., 1979)
AA-MD[4]	300	56	461	(Wymore, et al., 1999)
AA-MD[4]	300	54	120 and 460	(Marrink, et al., 2000)
AA-MD[4]	300	54	200	(Abel, et al., 2012)
CG-MD[5]	300	40-70	40	(Marrink, et al., 2004)
CG-MD[5]	370	45	126	(Sanders & Panagiotopoulos, 2010)

Systems of transmembrane helix dimer proteins embedded in DPC micelles or in self-assembled POPC bilayers are not achievable in Martini 3

TM helix dimers are components of many receptors and are thought to play an important role in transmitting extracellular signals to the cellular interior (Westerfield & Barrera, 2020). Experimental determination of the structure of TM helix dimers is often performed for TM helix dimers immersed in DPC micelles rather than a membrane bilayer. To investigate the insertion of TM helix dimers into DPC micelles using the previously derived DPC parameters, simulations starting with 100 DPC molecules at random positions and a position-restrained glycophorin A (GpA) TM helix dimer were run with Martini 3. This protein is considered to represent an archetypal TM helix dimer and is often used to assess computational methods for studying TM helix dimer proteins (Psachoulia, et al., 2010) (Domanski, et al., 2018) (Majumder & Straub, 2021) (Majumder, et al., 2022). Surprisingly, although the DPC lipids assembled into micelles successfully, with aggregation numbers similar to those obtained in simulations without protein, GpA was not encapsulated within a micelle. The TM helix dimer interacted with the polar head beads of the DPC lipids and with the water beads rather than with the hydrocarbon chains (**Figure 2A**).

To eliminate a possible sequence dependence of this effect, simulations in identical conditions were run for the TM helix dimers of the neurotrophin receptor TrkA. Similarly, a lack of insertion in micelles was observed. In simulations starting from a micelle-embedded TM helix dimer, immediate micelle dissociation and re-assembly excluding the protein from the micelles was observed (**Supplementary Figure 1**).

[1] Nuclear Magnetic Resonance
[2] Small-Angle X-ray Scattering
[3] Small-Angle Neutron Scattering
[4] All-Atom Molecular Dynamics simulation
[5] Coarse-Grained Molecular Dynamics

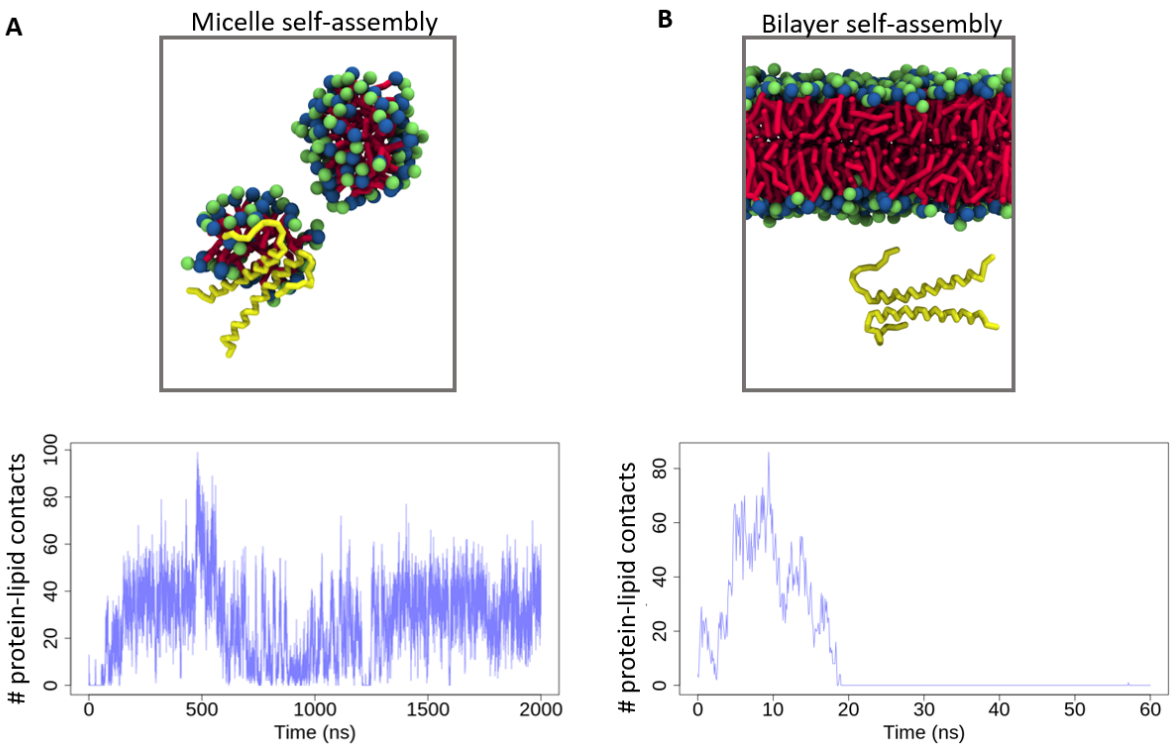

Figure 2: Aberrant behaviour of transmembrane helix dimer proteins in lipid environments in Martini 3. A: GpA TM helix dimers (yellow) do not interact with the hydrophobic tails (red) of micelle-forming DPC molecules, but rather with the water and with the polar head groups (blue and green) of the lipids. **B**, The POPC bilayer does not assemble around the position-restrained GpA.

We considered that a wrong choice of DPC parameters could be the reason for the lack of protein-lipid interactions in these simulations, given that these parameters have not been as extensively validated as those for other phospholipids that have been released by the Martini 3 developers. To test this hypothesis, simulations of the GpA dimer with and without position restraints and POPC lipids at random initial positions were run. The rationale behind these tests is that this lipid has been parameterized and validated by the developers of the Martini 3 force field; thus, protein-lipid interactions should be correctly reproduced in systems containing POPC. Strikingly, although the lipid molecules rapidly assembled to form a bilayer, GpA was not embedded inside it but rather interacted with the water beads (**Figure 2B**). In those simulations in which the protein dimer was not position-restrained during bilayer self-assembly, the two monomers showed a tendency to dissociate from one another (**Supplementary Figure 2**). On the other hand, in simulations that had been prepared using the INSANE method to insert transmembrane proteins into preassembled bilayers, the GpA TM helix dimers remained inside the membrane for the entire trajectory (20 μs).

Scaling down protein-water interactions in Martini 3 enables simulation of transmembrane helix dimer proteins in DPC micelles and in self-assembled POPC bilayers

Different strategies were tested in order to correct the lack of interaction between TM helix dimers and the hydrophobic tails of DPC and POPC molecules. The first group of strategies did not use any modification of the interaction parameters of the Martini 3 force field. These approaches mainly focused on the insertion of the transmembrane proteins into micelles, and were based on modifications of the previously described DPC parameters. A more detailed description of these tests, which were conducted using TrkA TM helix dimers, can be found in **Supplementary Table 1 and Supplementary Figure 3**. None of these strategies led to insertion of the protein into micelles.

The second group of strategies was based on the modification of the nonbonded parameters for interactions between given bead types (as described in the Methods section). In order to correct for the lack of protein-lipid interactions, non-bonded interactions between protein beads and lipid chains were scaled up as described in the Methods section, using α values greater than one. Although a slight increase in protein-lipid interactions was observed, this was not sufficient to fully embed the protein into the micelles. Furthermore, after release of the position restraints, the protein monomers dissociated from one another in all replicas and either ended in different micelles or on opposite sides of the same micelle (**Supplementary Figure 3**).

In view of the lack of success of an increase of protein-lipid non-bonded interactions, a different strategy was tested: decreasing protein-water interactions. Applying a scaling factor α of 0.9 to all interactions between protein and water beads, 10 replicas of 20 μs each were run for a system containing 100 DPC molecules and a GpA TM helix dimer. In all replicas, a large micelle with 100 lipids assembled around the protein (**Figure 3A**) and remained stable during the entire trajectory. Similar results were observed when a TrkA TM helix dimer was simulated under the same conditions.

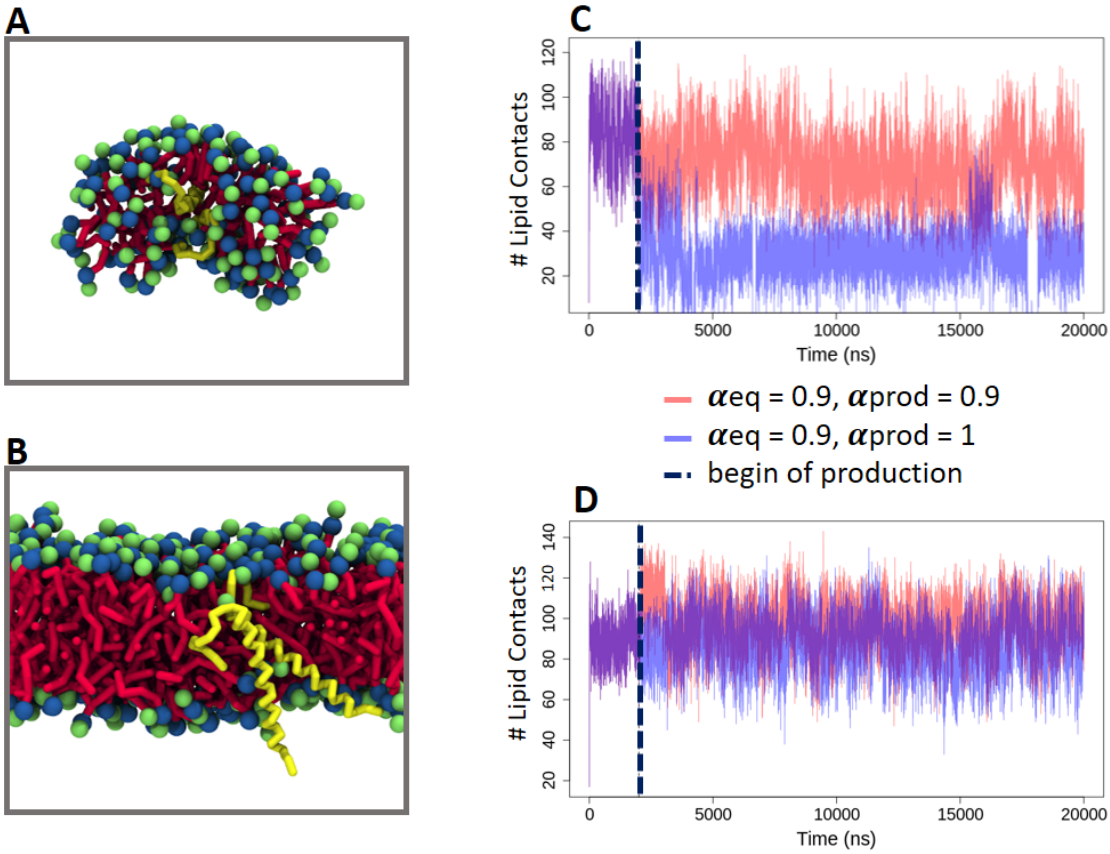

Figure 3: Protein-lipid systems simulated with Martini 3 after introduction of a scaling factor. **A, B**: Decreasing protein-water interactions by 10% (α = 0.9) leads to correct insertion of GpA dimers into a micelle (A) and a POPC bilayer (B). **C**: After removal of the scaling factor at the beginning of the production (dashed line), GpA leaves the micelle, which results in a reduction of protein-lipid contacts (blue line) compared to maintaining the scaling factor (red line). **D**: In simulations of GpA dimers in POPC bilayers, the protein dimer remains inside the membrane upon removal of the scaling factor after equilibration.

Removing the scaling after an equilibration of 2 μs led to immediate dissociation of the large micelle containing 100 DPC molecules into two smaller ones and interruption of the interactions with the transmembrane protein (**Figure 3C**). This suggests that the scaling factor proposed here needs to be maintained during the totality of the trajectory in order to simulate TM proteins in micelles.

Scaling down protein-water interactions by 10% also led to POPC bilayer self-assembly around TM helix dimers (**Figure 3B**). In order to assess the statistical robustness of these observations, 100 short (60 ns) simulations of bilayer self-assembly in the presence of GpA were performed with different scaling factors. Whereas assembly of the phospholipid bilayer around the TM helix dimer was observed in only one out of 100 simulations without scaling, it was observed in 30 trajectories when a scaling factor of $\alpha = 0.9$ was used (**Supplementary Figure 4**).

10 of the 100 short simulations with $\alpha = 0.9$ were equilibrated for 2 μs. Afterwards, each of them was used as input for two 20 μs production runs: one maintaining protein-water scaling, and the other having the default Martini 3 parameters. In half of the replicas, the protein had

acquired a correct orientation inside the lipid bilayer during equilibration. In all these cases, the protein remained inside the membrane during the production run regardless of whether it was run with or without down-scaling of the protein-water interactions (data from two trajectories are shown in **Figure 3D**). Thus, in contrast to simulations of TM helix dimers in micelles, where scaling seems to be crucial to obtain meaningful results, in the case of self-assembled bilayers, scaling can be removed as soon as the TM helix dimer acquires the correct orientation in the bilayer.

In the remaining half of the replicas, GpA did not orient correctly inside the membrane, acquiring either a diagonal or a horizontal arrangement, with one of the two monomers partially outside of the bilayer. In these cases, the protein exited the membrane after release of the position restraints in production runs. This behaviour was observed in simulations with and without scaling of protein-water interactions.

Scaling of protein-water interactions does not affect the dimerization behaviour of transmembrane helix dimers

To investigate whether the down-scaling of protein-water interactions affects the dimer arrangements adopted by GpA in a membrane, structural and contact information was extracted from the simulations of the GpA dimer in self-assembled POPC bilayers with and without scaling described in the previous section. Only those replicas (5 out of 10) in which the TM helix dimer remained inside the lipid bilayer during the entire trajectory were used for this analysis.

To analyse the conformational landscape of the GpA dimer, all five replicas were merged and the whole system was aligned to one of the two helices. A series of structural parameters was computed to characterize the different possible helix dimer arrangements. These parameters included the distance between the centres of geometry (COG), the crossing angle, and the rotation angles around the first and second helices (described in the Methods section). Kernel density estimation (KDE) was used to identify the most populated conformations as a two-dimensional function of the given structural parameters.

A greater exploration of the conformational landscape was observed in simulations in which protein-water interactions had been reduced by 10% (**Figure 4, Supplementary Figure 5**). However, the density maxima corresponding to the conformations of the NMR structure of GpA were observed in simulations both with and without scaling of protein-water interactions. These helical arrangements showed a crossing angle of about -34º, a distance between COGs of 8-9 Å, and position and phase angles around 90º with Gly 79 defined as the reference residue. A similar conformational landscape was observed in simulations of GpA dimers inserted into a preformed membrane with the INSANE method without protein-water scaling (**Supplementary Figure 6**). Hence, down-scaling protein-water interactions does not seem to affect the structural arrangements of the GpA TM helix dimers.

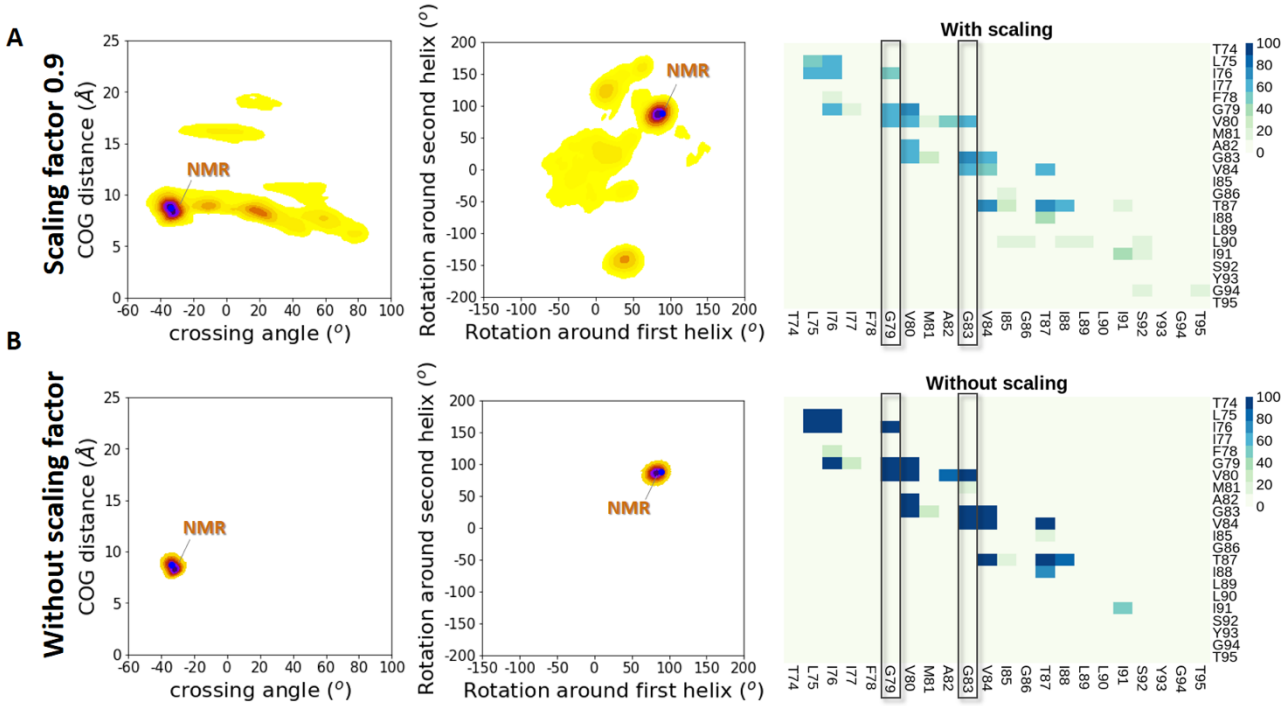

Figure 4: Transmembrane helix dimer proteins conserve their structural properties upon scaling protein-water interactions down by 10% in Martini 3. A: Conformational space and contact analysis of GpA TM helix dimers in POPC bilayers simulated with a uniform scaling factor of $\boldsymbol{\alpha}$ = 0.9. Kernel density estimation was used to determine the most populated conformations described by the crossing angle and the distance between centers of geometry (COG) (left panel) or the rotation dihedrals around the two helices (middle panel). This analysis was run on 5 merged replicas, each 20 µs long. Positive and negative crossing angles denote left- and right-handed dimers, respectively. The blue dot corresponds to the NMR structure, which is at a density maximum in both representations of the conformational landscape. The dimerization motif $^{79}GxxxG^{83}$, is part of the dimer interface during most of the simulation time (right panel). A cut-off distance of 6 Å was used to define contacts and their percentage occupation is given by color in the heat map. **B**: Conformational space and contact analysis of GpA TM helix dimers in POPC bilayers simulated without introducing a scaling factor. The explored portion of the conformational space is narrower than in simulations with a scaling factor, with a single density maximum corresponding to the NMR structure. The dimer interface contains the same residues as in simulations with a scaling factor, including the $^{79}GxxxG^{83}$ dimerization motif.

To further investigate the effect of the scaling of protein-water interactions on the dimerization behaviour of GpA, contact analysis was performed for simulations performed with and without the introduction of the scaling factor. In both cases, the same residues participated in the dimer interface, which contained the dimerization motif ^{79}GxxxG^{83}. This sequence has been proposed to be crucial for the dimerization of GpA. Residues L75, I76, V80, G83, V84 and T87 were part of the dimer interface as well during most of the trajectories (**Figure 4**). Thus, it can be concluded that introducing a factor to down-scale protein-water interactions by 10% does not disturb the dimer interface or other structural parameters of TM helix dimers embedded in lipid bilayers.

To evaluate the effect of the scaling on the POPC bilayers, structural parameters of the lipid bilayer were calculated for simulations with and without scaling. These include the lipid lateral diffusion, phosphate density (which gives an estimate of the bilayer thickness), area per lipid, and second rank order parameter (P2). No noticeable differences in these parameters were observed between simulations performed with and without scaling (**Supplementary Figure 7**), which indicates that the reduction of the protein-water interactions does not affect the structure of the phospholipid bilayer.

The conformational landscape of the GpA TM helix dimers in DPC micelles in simulations with a 10% reduction of protein-water interactions showed notable differences to that observed in POPC bilayers (**Figure 5**). This indicates a significant effect of the lipid environment on the structure of TM helix dimers in Martini 3.

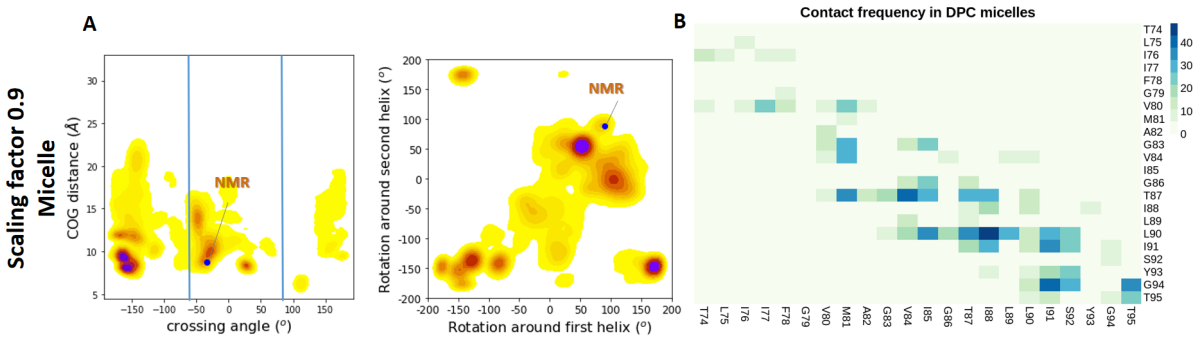

Figure 5: Structural features of GpA dimers in DPC micelles upon reduction of protein-water interactions by 10% in Martini 3. A: A significantly greater conformational space is explored compared to simulations in POPC bilayers (shown in figure 4). In many frames, the dimer adopts higher or lower crossing angles than observed in bilayers (beyond the blue vertical lines). This analysis was run on 10 merged replicas, each 20 µs long. The blue dot represents the NMR structure. **B**: Percentage occupancy of residue-residue contacts between the two TM helix monomers in DPC micelles. Dimers with crossing angles greater than 80° and less than -60° were not considered for this analysis due to their lack of relevance in biological membranes. A cut-off distance of 6 Å was used to define contacts.

DISCUSSION

Detergent micelles are commonly used to study the structural properties of TM proteins. MD simulations are usually carried out for such proteins in phospholipid bilayers, which better resemble their physiological environment. However, this means that it is not always clear whether differences between simulations and experiments arise from the different nature of these methods or from the differences in the lipid environments. Simulation of TM proteins in detergent micelles can thus help to clarify this matter, and if combined with simulations in bilayers, enable comparison of experimental and computational results.

In the present study, we show that TM helix dimers cannot be simulated in DPC micelles using the most recent version of the Martini force field, Martini 3, due to a lack of interaction of the protein with the hydrophobic tails of the surfactant molecules. We also demonstrate that self-assembled phospholipid bilayers rarely encapsulate TM helix proteins in Martini 3. These

observations suggest that the TM protein beads are relatively hydrophilic in Martini 3. Indeed, a comparison of the Lennard Jones potentials of beads of different residues with hydrophobic lipid tail beads and with water beads shows remarkable differences between hydrophilic and hydrophobic amino acids (**Supplementary Figure 8**).

When TM helix dimers are inserted in preformed membranes using the INSANE method, however, they behave as expected and remain inside the lipid bilayer. This is consistent with previously reported studies of other TM helix dimers using Martini 3, e.g. of Ephrin type-A receptor I (EphA1) dimerization (Lamprakis, et al., 2021) (Sahoo, et al., 2022). On the contrary, when simulations were initiated with a TM helix dimer inserted in a preformed DPC micelle, the micelle dissociated immediately, leaving the protein exposed to water. This difference between bilayer and micelle environments can arise because the former is more stable, while the latter is more dynamic and thus the protein has more degrees of freedom (Bordag & Keller, 2010) (Lind, et al., 2008).

Here, we show that introducing a scaling factor to reduce protein-water interactions by 10% in Martini 3 results in insertion of GpA TM helix dimers into phospholipid bilayers and detergent micelles. The fact that Martini 3 uses a specific bead type for water, which has been parameterised independently, makes this solution very attractive. Indeed, we find that this down-scaling does not affect the dimerization behaviour of the GpA TM helices. The structural and dynamic parameters of the lipid bilayers, such as bilayer thickness and lateral diffusion, also remain unaltered upon scaling.

Protein-water interactions have recently been scaled up in Martini 3 to reproduce the experimental dimensions of soluble and intrinsically disordered proteins (Thomasen, et al., 2022). The fact that opposite adjustments need to be made to reproduce the experimental behaviour of TM helix dimers reflects the challenging nature of CG force field parameterization: due to a reduction of the chemical space, system-dependent adjustments need to be made in order to obtain biologically meaningful results.

We observe remarkable differences in the conformational landscapes of GpA dimers in POPC bilayers and DPC micelles. We believe these differences are due to the more restraining environment of POPC bilayers as opposed to the more dynamic nature of micelles, which allow the helix dimers to adopt arrangements with much larger crossing angles. Indeed, it has been shown experimentally that the fluctuations in the crossing angles of GpA dimers in DPC micelles are twice as large as those in bilayers (Mineev, et al., 2011). Another study suggests that the stability of GpA dimers increases linearly with the aggregation number of the detergent (Stangl, et al., 2012). As the aggregation number of DPC micelles is significantly lower with our Martini 3 parameters than with Martini 2, one would expect the stability of the TM helix dimer to be lower as well. Other peptides have been shown to adopt differing conformations in DPC micelles and disk-like bicelles (Chou, et al., 2002) (Andersson & Mäler, 2002) (Lind, et al., 2008). We believe the differences in conformational landscape in micelles and bilayers found in this study are caused by the differences in lipid environment rather than by the scaling applied to the protein-water interactions. The fact that only minimal differences are observed in the helix dimer arrangements obtained in bilayers with and without scaling supports this hypothesis. Hence, we expect the scaling of the non-bonded protein-water interactions in

Martini 3 proposed here to provide a rapid, simple and effective correction of the aberrant behaviour of TM helix proteins in self-assembled lipid bilayers and detergent micelles.

ACKNOWLEDGEMENTS

This research was funded by the European Union's Horizon 2020 research and innovation programme "Euroneurotrophin" under the Marie Skłodowska-Curie grant agreement No 765704 and the Klaus Tschira Foundation. The authors acknowledge support by the state of Baden-Württemberg through bwHPC and the German Research Foundation (DFG) through grant INST 35/1134-1 FUGG. We thank Paulo Cesar Telles de Souza for helpful discussions and Stefan Richter for technical support.

BIBLIOGRAPHY

Abel, S., Dupradeau, F. Y. & Marchi, M., 2012. Molecular dynamics simulations of a characteristic DPC micelle in water. *Journal of Chemical Theory and Computation,* 8(11), pp. 4610-4623.

Alessandri, R. et al., 2019. Pitfalls of the Martini model. *Journal of chemical theory and computation,* 15(10), pp. 5448-5460.

Allen, W. J., Lemkul, J. A. & Bevan, D. R., 2009. GridMAT-MD: a grid-based membrane analysis tool for use with molecular dynamics. *Journal of computational chemistry,* 30(12), pp. 1952-1958.

Andersson, A. & Mäler, L., 2002. NMR solution structure and dynamics of motilin in isotropic phospholipid bicellar solution. *Journal of Biomolecular NMR,* 24(2), pp. 103-112.

Bordag, N. & Keller, S., 2010. α-Helical transmembrane peptides: a "divide and conquer" approach to membrane proteins. *Chemistry and physics of lipids,* 163(1), pp. 1-26.

Bottaro, S. & Lindorff-Larsen, K., 2018. Biophysical experiments and biomolecular simulations: A perfect match?. *Science,* 361(6400), pp. 355-360.

Bussi, G., Donadio, D. & Parrinello, M., 2007. Canonical sampling through velocity rescaling. *The Journal of chemical physics,* 126(1), p. 014101.

Chou, J. J. et al., 2002. Micelle-induced curvature in a water-insoluble HIV-1 Env peptide revealed by NMR dipolar coupling measurement in stretched polyacrylamide gel. *Journal of the American Chemical Society,* 124(11), pp. 2450-2451.

De Jong, D. H., Baoukina, S., Ingólfsson, H. I. & Marrink, S. J., 2016. Martini straight: Boosting performance using a shorter cutoff and GPUs. *Computer Physics Communications,* 199(1), pp. 1-7.

Domanski, J., Sansom, M. S., Stansfeld, P. J. & Best, R. B., 2018. Balancing force field protein-lipid interactions to capture transmembrane helix--helix association. *Journal of chemical theory and computation,* 14(3), pp. 1706-1715.

Fabian, B., Thallmair, S. & Hummer, G., 2022. Small ionic radii limit time step in Martini 3 molecular dynamics simulations. *arXiv preprint arXiv:2204.05121.*

Franco, M. L. et al., 2020. Structural basis of the transmembrane domain dimerization and rotation in the activation mechanism of the TRKA receptor by nerve growth factor. *Journal of Biological Chemistry,* 295(1), pp. 275-286.

Gao, X. & Wong, T. C., 1998. Studies of the binding and structure of adrenocorticotropin peptides in membrane mimics by NMR spectroscopy and pulsed-field gradient diffusion. *Biophysical journal,* 74(4), pp. 1871-1888.
Gowers, R. J. et al., 2016. MDAnalysis: a Python package for the rapid analysis of molecular dynamics simulations. In: e. S. Benthall and S. Rostrup, ed. *Proceedings of the 15th python in science conference.* Austin, TX: SciPy, pp. 98-105.
Harris, C. R. et al., 2020. Array programming with NumPy. *Nature,* 585(7825), pp. 357-362.
Humphrey, W., Dalke, A. & Schulten, K., 1996. VMD: visual molecular dynamics. *Journal of molecular graphics,* 14(1), pp. 33-38.
Javanainen, M., Martinez-Seara, H. & Vattulainen, I., 2017. Excessive aggregation of membrane proteins in the Martini model. *Javanainen, Matti; Martinez-Seara, Hector; Vattulainen, Ilpo,* 12(11), p. e0187936.
Kabsch, W. & Sander, C., 1983. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers: Original Research on Biomolecules,* 22(12), pp. 2577-2637.
Kallick, D. A., Tessmer, M. R., Watts, C. R. & Li, C.-Y., 1995. The use of dodecylphosphocholine micelles in solution NMR. *Journal of Magnetic Resonance, Series B,* 109(1), pp. 60-65.
Karplus, M. & McCammon, J. A., 2002. Molecular dynamics simulations of biomolecules. *Nature structural biology,* 9(9), pp. 646-652.
Lamprakis, C. et al., 2021. Evaluating the efficiency of the Martini force field to study protein dimerization in aqueous and membrane environments. *Journal of Chemical Theory and Computation,* 17(5), pp. 3088-3102.
Lauterwein, J., Bösch, C., Brown, L. R. & Wüthrich, K., 1979. Physicochemical studies of the protein-lipid interactions in melittin-containing micelles. *Biochimica et Biophysica Acta (BBA)-Biomembranes,* 556(2), pp. 244-264.
Lind, J. N. J. & Mäler, L., 2008. Lipid dynamics in fast-tumbling bicelles with varying bilayer thickness: effect of model transmembrane peptides. *Biochimica et Biophysica Acta (BBA)-Biomembranes,* 1778(11), pp. 2526-2534.
Lind, J., Nordin, J. & Mäler, L., 2008. Lipid dynamics in fast-tumbling bicelles with varying bilayer thickness: effect of model transmembrane peptides. *Biochimica et Biophysica Acta (BBA)-Biomembranes,* 1778(11), pp. 2526-2534.
Lindahl, Abraham, Hess & Spoel, v. d., 2020. *GROMACS 2020.4 Source code,* https://doi.org/10.5281/zenodo.4054979: Zenodo.
Lipfert, J. et al., 2007. Size and shape of detergent micelles determined by small-angle X-ray scattering. *The journal of physical chemistry B,* 111(43), pp. 12427-12438.
MacKenzie, K. R., Prestegard, J. H. & Engelman, D. M., 1997. A transmembrane helix dimer: structure and implications. *Science,* 276(5309), pp. 131-133.
Majumder, A. & Straub, J. E., 2021. Addressing the Excessive Aggregation of Membrane Proteins in the MARTINI Model. *Journal of chemical theory and computation,* 17(4), pp. 2513-2521.
Majumder, A., Kwon, S. & Straub, J. E., 2022. On Computing Equilibrium Binding Constants for Protein-Protein Association in Membranes. *Journal of Chemical Theory and Computation,* 18(6), p. 3961–3971.
Marrink, S. J. et al., 2007. The MARTINI force field: coarse grained model for biomolecular simulations. *The journal of physical chemistry B,* 111(27), pp. 7812-7824.
Marrink, S. J., De Vries, A. H. & Mark, A. E., 2004. Coarse grained model for semiquantitative lipid simulations. *The Journal of Physical Chemistry B,* 108(2), pp. 750-760.

Marrink, S. J., Tieleman, D. P. & Mark, A., 2000. Molecular dynamics simulation of the kinetics of spontaneous micelle formation. *The Journal of Physical Chemistry B,* 104(51), pp. 12165-12173.
McKinney, W., 2010. Data structures for statistical computing in python. *Proceedings of the 9th Python in Science Conference,* 445(1), pp. 51-56.
Michaud-Agrawal, N., Denning, E. J., Woolf, T. B. & Beckstein, O., 2011. MDAnalysis: a toolkit for the analysis of molecular dynamics simulations. *Journal of computational chemistry,* 32(10), pp. 2319-2327.
Mineev, K. et al., 2011. Dimeric structure of the transmembrane domain of glycophorin a in lipidic and detergent environments. *Acta Naturae,* 3(2), pp. 90-98.
Pambou, E. et al., 2015. Structural features of micelles of zwitterionic dodecyl-phosphocholine (C12PC) surfactants studied by small-angle neutron scattering. *Langmuir,* 31(36), pp. 9781-9789.
Parrinello, M. & Rahman, A., 1981. Polymorphic transitions in single crystals: A new molecular dynamics method. *Journal of Applied physics,* 52(12), pp. 7182-7190.
Psachoulia, E., Marshall, D. P. & Sansom, M. S., 2010. Molecular dynamics simulations of the dimerization of transmembrane α-helices. *Accounts of chemical research,* 43(3), pp. 388--396.
R Core Team, 2022. *R: A Language and Environment for Statistical Computing.* https://www.R-project.org/ ed. Vienna, Austria: R Foundation for Statistical Computing.
Sahoo, A. R., Souza, P. C. T., Meng, Z. & Buck, M., 2022. How transmembrane helices of type 1 receptors transmit information: prediction of dimers with the Martini 3 coarse grained model. *Biophysical Journal,* 121(3), p. 464a.
Sanders, S. A. & Panagiotopoulos, A. Z., 2010. Micellization behavior of coarse grained surfactant models. *The Journal of chemical physics,* 132(11), p. 114902.
Schrödinger, L., 2020. *The PyMOL Molecular Graphics System.* s.l.:PyMol.
Souza, P. C. et al., 2021. Martini 3: a general purpose force field for coarse-grained molecular dynamics. *Nature methods,* 18(4), pp. 382-388.
Stangl, M. et al., 2012. Detergent properties influence the stability of the glycophorin A transmembrane helix dimer in lysophosphatidylcholine micelles. *Biophysical journal,* 103(12), pp. 2455-2464.
Stark, A. C., Andrews, C. T. & Elcock, A. H., 2013. Toward optimized potential functions for protein--protein interactions in aqueous solutions: osmotic second virial coefficient calculations using the martini coarse-grained force field. *Journal of chemical theory and computation,* 9(9), pp. 4176-4185.
Stone, J. E., 1998. An efficient library for parallel ray tracing and animation. *University of Missouri, Rolla.*
Thomasen, F. E. et al., 2022. Improving Martini 3 for Disordered and Multidomain Proteins. *Journal of Chemical Theory and Computation,* 18(4), pp. 2033-2041.
Uusitalo, J. J. et al., 2015. Martini coarse-grained force field: extension to DNA. *Journal of chemical theory and computation,* 11(8), pp. 3932-3945.
Van Rossum, G. & Drake, F. L., 2009. *Python 3 Reference Manual.* Scotts Valley, CA: CreateSpace.
Voth, G. A., 2008. *Coarse-graining of condensed phase and biomolecular systems.* Boca Raton: CRC press.
Warschawski, D. E. et al., 2011. Choosing membrane mimetics for NMR structural studies of transmembrane proteins. *Biochimica et Biophysica Acta (BBA)-Biomembranes,* 1808(8), pp. 1957-1974.

Waskom, M. L., 2021. Seaborn: statistical data visualization. *Journal of Open Source Software,* 6(60), p. 3021.
Wassenaar, T. A. et al., 2015. Computational lipidomics with insane: a versatile tool for generating custom membranes for molecular simulations. *Journal of chemical theory and computation,* 11(5), pp. 2144-2155.
Wassenaar, T. A. et al., 2015. High-throughput simulations of dimer and trimer assembly of membrane proteins. The DAFT approach. *Journal of chemical theory and computation,* 11(5), pp. 2278-2291.
Westerfield, J. M. & Barrera, F. N., 2020. Membrane receptor activation mechanisms and transmembrane peptide tools to elucidate them. *Journal of Biological Chemistry,* 295(7), pp. 1792-1814.
Wymore, T., Gao, X. & Wong, T., 1999. Molecular dynamics simulation of the structure and dynamics of a dodecylphosphocholine micelle in aqueous solution. *Journal of Molecular Structure,* Volume 485, pp. 195-210.