Addressing the Excessive Aggregation of Membrane Proteins in the MARTINI Model

Ayan Majumder and John E. Straub*

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: The MARTINI model is a widely used coarse-grained force field popular for its capacity to represent a diverse array of complex biomolecules. However, efforts to simulate increasingly realistic models of membranes, involving complex lipid mixtures and multiple proteins, suggest that membrane protein aggregates are overstabilized by the MARTINI v2.2 force field. In this study, we address this shortcoming of the MARTINI model. We determined the free energy of dimerization of four trans-membrane protein systems using the nonpolarizable MARTINI model. Comparison with experimental FRET-based estimates of the dimerization free energy was used to quantify the significant overstabilization of each protein homodimer studied. To improve the agreement between simulation and experiment, a single uniform scaling factor, α, was used to enhance the protein–lipid Lennard–Jones interaction. A value of $\alpha = 1.04-1.045$ was found to provide the best fit to the dimerization free energies for the proteins studied while maintaining the specificity of contacts at the dimer interface. To further validate the modified force field, we performed a multiprotein simulation using both MARTINI v2.2 and the reparameterized MARTINI model. While the original MARTINI model predicts oligomerization of protein into a single aggregate, the reparameterized MARTINI model maintains a dynamic equilibrium between monomers and dimers as predicted by experimental studies. The proposed reparameterization is an alternative to the standard MARTINI model for use in simulations of realistic models of a biological membrane containing diverse lipids and proteins.
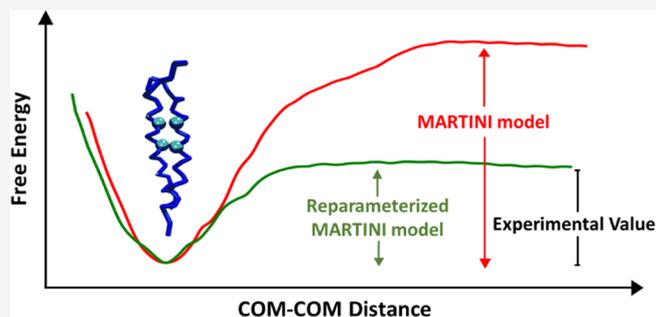
INTRODUCTION

Molecular dynamics simulation has been extensively used to explore the association of protein in membranes.[1−4] However, even the association of the transmembrane (TM) helices in membranes is challenging to model accurately using all-atom simulation. Coarse-grained (CG) molecular dynamics simulation can be used to probe time scales beyond the reach of all-atom simulation[5,6] while sacrificing structural details. While the MARTINI model[7,8] is most commonly used to model lipids and surfactants, proteins are also represented using a similar CG scheme. The MARTINI lipid force field has been validated by many studies with results in agreement with the experimental data.[9,10] Nonbonded parameters for proteins were optimized to reproduce individual amino acid dimerization energies.[6,8] However, severe restraints are placed on the protein backbone potential, and the protein–protein interactions are typically not validated against experimental results.[11] A shortcoming of the MARTINI protein potential was first noted by Elcock and coworkers, who observed precipitation of protein below the solubility limit.[12] It was subsequently noted that in the dimerization of protein in the membrane, the protein dimers were overstabilized.[13−16]

A related but qualitatively different shortcoming was observed in the all-atom CHARMM36 force field[17,18] in that TM protein homodimers lacked stability when compared with the dissociated monomeric state. Best and coworkers effectively addressed this issue by scaling the Lennard–Jones (LJ) interaction potential using a single scaling factor so as to make the protein–lipid interaction less favorable, thereby destabilizing the monomeric state relative to the homodimeric state.[19] A scaling factor value of 0.9−0.95 provided the best fit to the experimental data. The insight provided by that work, addressing the understabilization of the TM protein homodimers in the all-atom CHARMM36 potential, suggests that a similar approach might be used to address the overstabilization of TM protein homodimers in the MARTINI v2.2 potential.

The Förster resonance energy transfer (FRET) technique has been used to evaluate the dimerization free energy of protein in membranes.[20] Comparison of protein dimerization free energies derived from computer simulation with those

obtained from FRET studies provides a critical test of the simulation model. It has been observed that dimerization free energies for membrane proteins obtained using MARTINI model simulations show disagreement with FRET data, typically predicting higher dimerization free energies than observed in experiment. For example, the MARTINI simulation of TM protein ErbB1 predicts the free energy of dimerization to be −9.4 kcal/mol[21] in contrast to the experimental value of −2.5 kcal/mol.[22] For the TM protein EphA1, the FRET value was reported to be −3.7 kcal/mol in dimyristoylphosphatidylcholine (DMPC) bilayer,[23] while simulations using the MARTINI model led to an estimate of −14.3 kcal/mol in dipalmitoyl phosphatidylcholine (DPPC) bilayer.[15] An extensive comparison of several all-atom models and the MARTINI model also observed the propensity of the MARTINI model to overstabilize membrane protein dimers.

In this study, we address a shortcoming of the MARTINI v2.2 force field in the observed overstabilization of protein–protein interactions leading to the unphysical oligomerization of the protein in membranes. We demonstrate that scaling the protein–lipid interaction destabilizes protein aggregates while preserving the membrane properties. The standard and rescaled MARTINI models predict similar structures for TM homodimers, while the latter predicts homodimer thermodynamic stability comparable to that observed in experiment. The simulation of eight glycophorin A proteins in a POPC bilayer using the rescaled potential leads to dynamic equilibrium in which proteins exist as monomers or dimers.

METHODS

Four TM protein homodimers were simulated in this study, glycophorin A (GpA), EphA1, FGFR3, and APP C99. (1) The structure of GpA was taken from PDB entry 1AFO.[24] The residues 69 to 97 were taken to form a congener for our study. The GpA dimer was embedded in a lipid bilayer made with 408 POPC lipid molecules. The temperature of the system was maintained at 310 K throughout the simulation. (2) The EphA1$_{544-572}$ dimer (PDB entry 2K1L[25]) was embedded in a DLPC lipid bilayer. The bilayer consisted of 405 DLPC molecules, and the temperature of the system was maintained at 303 K. The experimental FRET study of EPHA1 was carried out in DMPC liposomes.[23] Each aliphatic tail of DMPC has 14 carbons. In applying the 4-to-1 CG mapping, the MARTINI model "rounds down" and assigns three interaction sites for each DMPC tail, making DMPC equivalent to DLPC in the MARTINI model. (3) Residues 367 to 399 of FGFR3 (PDB entry 2LZL,[26]) were placed in a POPC bilayer consisting of 407 lipid molecules. Residues 367 and 368 were mutated to Arginine (known as RR−FGFR3) to match the experimental sequence, and the simulation was performed at 298 K. (4) Finally, a congener of C99 formed by residues 22 to 55 (PDB entry 2LOH[27]) was simulated in a membrane bilayer prepared using 75% POPC and 25% POPG with 406 lipid molecules. The simulation was performed at 298 K. Relevant information is tabulated in characterizing each system (Supporting Information, Table S1).

Each lipid bilayer was solvated by 25 nonpolarizable water beads (100 water molecules) per two lipids, with 10% anti-freeze water beads. A Na$^+$ and Cl$^-$ ion concentration of 150 mmol was used. The MARTINI v2.2 force field[7] was used to simulate each protein-embedded bilayer system. The initial membrane structure was prepared using the insane.py program.[28] As has been observed previously, some simulated

systems include an unequal number of lipids in the two leaflets (see Supporting Information, Table S1). In a larger system, we might expect to find a roughly symmetric lipid distribution between leaflets. However, in the smaller system sizes used in our study, we feel the asymmetry is not inappropriate. The leap-frog integration method with a 20 fs time step was used for the *NPT* simulation. A velocity rescaling thermostat was used with a coupling time of 1 ps. A semi-isotropic Berendsen barostat with 2 ps coupling time and 3×10^{-4} bar^{-1} compressibility was used for the equilibration of each system. A semi-isotropic Parrinello−Rahman barostat with a coupling time of 12 ps and compressibility of 3×10^{-4} bar^{-1} was used in production runs to maintain the pressure at 1 bar. The remainder of the simulation parameters were set according to the MARTINI website and referred to as "common".[29]

A 100 ns simulation was performed to equilibrate each lipid bilayer system. Umbrella sampling simulation was performed to determine the potential of mean force (PMF) between the center-of-mass (COM) of the TM protein homodimers. A harmonic restraint of 1000 kJ mol^{-1} nm^{-2} was used to restrain the distance between the COM of the TM protein helices.[15] To ensure proper overlap between two adjacent umbrellas, we spaced the harmonic restraint potentials evenly at a distance of 0.15 nm. A 2 μs simulation run was performed for each umbrella window to calculate the PMF, where the first 100 ns were discarded to ensure the equilibration of each window.

Additional umbrella sampling simulations were carried out to calculate the PMF of partitioning an amino acid side chain between water and a dioleoylphosphatidylcholine (DOPC) lipid bilayer. Simulations were performed in a bilayer consisting of 72 DOPC lipids and 1250 water beads. A single amino acid side chain was placed in the membrane, and the PMF was calculated along the membrane normal. A harmonic restraint of 1000 kJ mol^{-1} nm^{-2} was used with a 0.15 nm spacing. A 200 ns simulation was performed for each umbrella window with a total of 7 μs simulation to calculate the partitioning PMF of each amino acid side chain. The weighted histogram analysis method (WHAM)[30] was used to unbias the umbrella windows. All simulations were performed with the GROMACS 2018.3 program.[31]

RESULTS AND DISCUSSION

Protein−protein interactions in membranes play an important role in cellular organization[32] and the function of TM proteins.[33] FRET studies are used to evaluate the free energy of dimerization of proteins in membranes and many computational studies have been inspired by and used to interpret FRET data.[22,23,34] In this study, we have used free energies of dimerization derived from FRET studies as a standard to guide the reparameterization of the MARTINI v2.2 force field. Experimental study of GpA homodimer formation using FRET reported the free energy of dimerization to range from 3.2 ± 0.2 to 4.0 ± 0.2 kcal/mol.[35] A recent all-atom simulation of GpA dimerization by Best *et al.* measured the free energy of dimerization to be 3.6 kcal/mol in a POPC bilayer.[19] The free energy of dimerization of EphA1 in DMPC and RR-FGFR3 in POPC was reported to be 3.7 ± 0.1[23] and 2.7 ± 0.1 kcal/mol,[36] respectively. The free energy of dimerization of full length C99 protein in a 75% POPC−25% POPG membrane was derived from FRET studies and reported to be 3.2 ± 0.1 kcal/mol.[37]

We have evaluated the dimerization free energy of GpA, EphA1, RR-FGFR3, and C99 dimer using a nonpolar

MARTINI CG model. We used the same protein sequence (with the exception of C99) and membrane composition as found in the corresponding FRET experiments, in the absence of the explicit treatment of the labels. The protein sequences are provided in Figure 2. We used roughly 400 lipids in the primary cell as suggested by previous studies.[15,21] In selecting this system size, we carried out simulations of larger systems in order to test the sensitivity of the dimerization free energy to finite size effects. The results obtained are shown in the Supporting Information, Figure S1. Based on these results, we concluded that a bilayer consisting of roughly 400 lipids is appropriate for the umbrella sampling simulations performed in this study.

In computing the dimerization free energy, we evaluated the PMF along the reaction coordinate defined by the relative COM distance of the proteins. The dimerization free energy obtained from the simulation is shown in Figure 1. The results
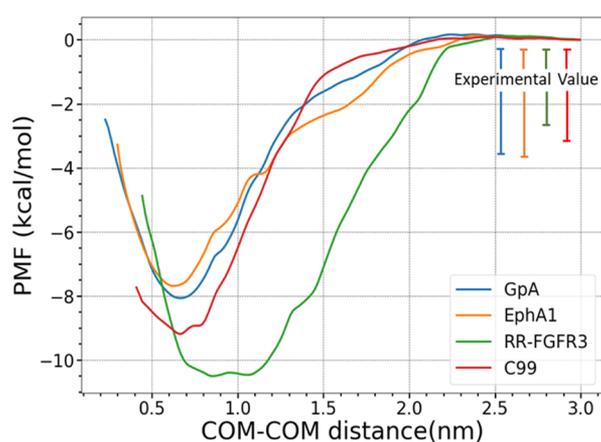

Figure 1. PMF for dimer association of GpA, EphA1, RR-FGFR3, and C99 as a function of distance between the COM of the TM helices. Experimental values of the free energy of dimerization are plotted as a scale bar. The blue, orange, green, and red lines represent the PMF of GpA, EphA1, RR-FGFR3, and C99, respectively. The simulations were performed using MARTINI v2.2. The maximum value of the PMF curve was used to set the zero and frame the comparison.

agree with previous studies that observed excessive protein aggregation resulting from the overstabilization of protein aggregates. The dimerization free energy derived from simulations using the MARTINI model is typically twice that derived from experiment. A previous study using the MARTINI v2.2P model reported the same artifact.[21] We have performed an equivalent simulation using MARTINI v2.2P on GpA, which also predicts the overstabilization of the dimer state (Supporting Information, Figure S2).

Many experimental and theoretical studies have been performed to characterize the dominant protein–protein interactions that stabilize the formation of homodimers of GpA, EphA1, FGFR3, and C99. Both NMR[24] and crystallographic studies[38] of the GpA dimer have revealed that the dimer is stabilized by the glycine zipper motif with a crossing angle of −20°. The EphA1 dimer is also stabilized by the glycine zipper motif[25] and forms a right-handed helical structure with a crossing angle of −21°.[15] NMR studies of the FGFR3 dimer indicate a strong interaction involving residues in the central region of the TM helices. The dimer is predominantly stabilized through the interaction of G380 and A391 and forms a left-handed helix. The favorable interaction

of the F384 residues of the two helices is another stabilizing factor.[26] A previous study of WT FGFR3 using the MARTINI model reported a distribution containing both left-handed and right-handed helix packing with a maximum crossing angle of 5°.[39] On the other hand, in the simulation of the $C99_{22-55}$ dimer, a number of competing homodimer structures were observed to contribute to the overall distribution. The $C99_{22-55}$ dimer formed both left-handed and right-handed structures with a preference for Gly-in structures stabilized by the GXXXG repeat motif.[40]

We have analyzed the homodimer structures characteristic of each minimum of the PMF for each protein by computing the contact map of each protein studied as shown in Figure 2. The complementary distribution of the helix crossing angle is shown in Figure 3. The contact fraction between identical residues of two TM helices is shown in the Supporting Information, Figure S3. The contact map of GpA demonstrates that the homodimer is stabilized by contacts mediated by the $G_{79}XXXG_{83}XXXT_{87}$ motif interaction. Similar analysis of EphA1 shows that the homodimer is stabilized by contacts mediated by the $G_{554}XXXG_{558}$ motif. Both GpA and EphA1 form a right-handed dimer with a crossing angle of −26 and −29°, respectively. While the RR-FGFR3 and $C99_{22-55}$ homodimers exhibit both left-handed and right-handed structures, RR-FGFR3 prefers to form a left-handed helical structure with a crossing angle of 5°. The TM helices form a strong interaction mediated by the $G_{380}X_2FF_{384}X_2ILX_2A_{391}$ motif, as observed in the experimental study.[41] The calculations above demonstrate that in each system the homodimer structure is in agreement with experiment, while the free energy of dimerization is exaggerated.

A number of approaches might be used to address the observed overstabilization of protein dimers and larger oligomers in simulations based on the MARTINI v2.2 force field. One intuitive approach would be to modify the protein–protein interactions to destabilize the homodimer. However, doing so would risk undermining the demonstrated specificity of contacts in the homodimer structures. Instead, we chose to uniformly scale the lipid–protein nonbonded interactions in order to enhance the stability of the monomeric state of each protein in line with the approach employed by Best and co-workers.[19] To accomplish this, we have scaled the well depth of the protein–lipid LJ interaction of the MARTINI v2.2 force field by a single uniform scaling factor, preserving all other interactions. Specifically, we scale the LJ potential well depth as

$$\varepsilon_{ij}^{new} = \alpha\varepsilon_{ij} \tag{1}$$

where ε and ε^{new} are the LJ potential well depth before and after reparameterization, $\alpha > 1$ is the scaling factor, and i and j are indices of protein and lipid beads, respectively. Simulation of the membrane using the reparameterized MARTINI model was performed following the same protocol used in the MARTINI v2.2 model simulations described above. The results obtained for all proteins studied are shown in Figure 4 and the dimerization free energies are tabulated in Table 1. The PMF data demonstrate that the MARTINI model is sensitive to small percentage changes in the LJ potential well depth. The dimerization free energy is observed to decrease monotonically with increasing protein–lipid interaction.

We studied the effects of different α scaling factors on the free energy of dimerization for each protein. To obtain the best α value for all proteins, we calculated the root mean square
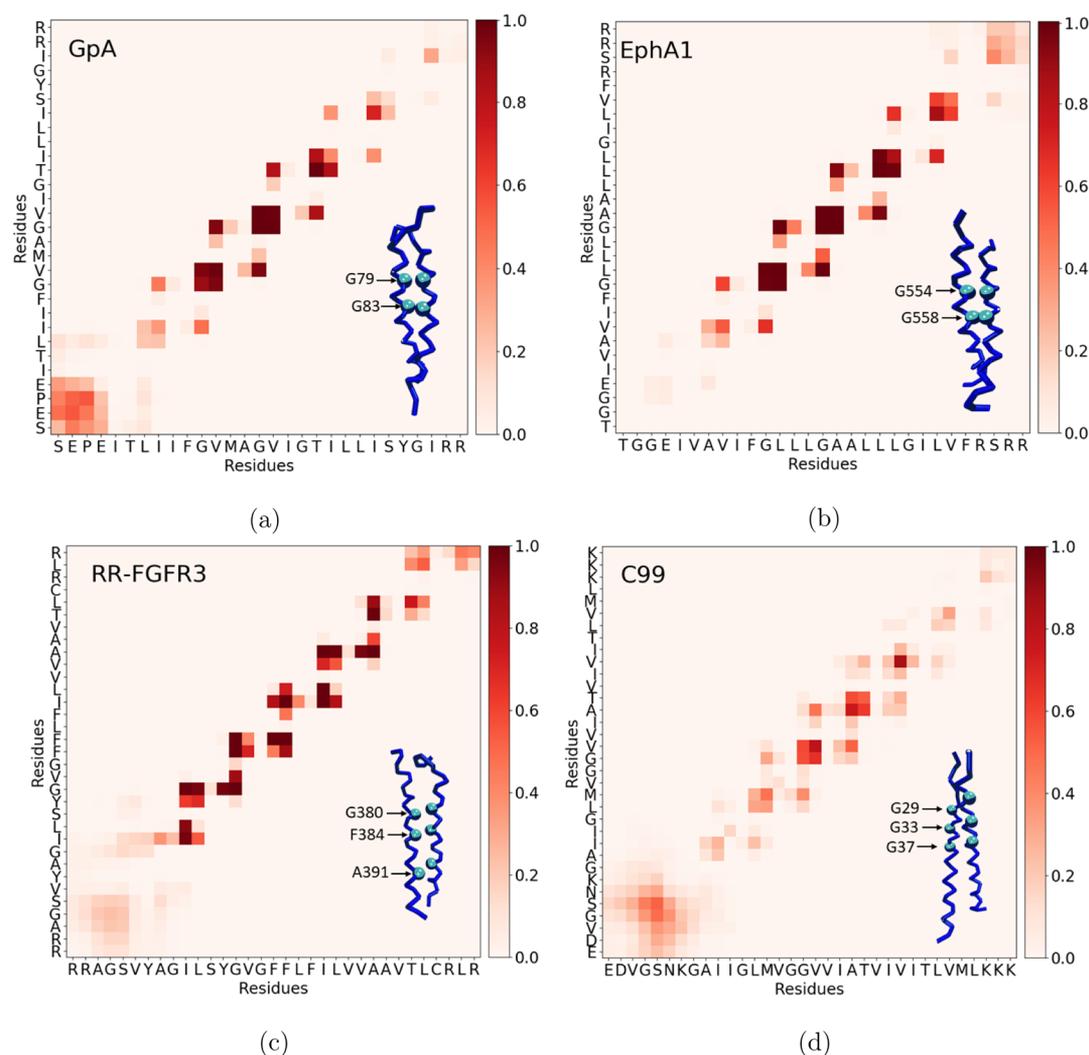
Figure 2. Contact map of (a) GpA, (b) EphA1, (c) RR-FGFR3, and (d) C99 dimers obtained by analyzing the minima of the PMF obtained using the MARTINI v2.2 model. A 0.7 nm cutoff was used to identify contacts. The color bars represent the contact probability, where the value 1 represents a contact maintained between the residues throughout the simulation trajectory.
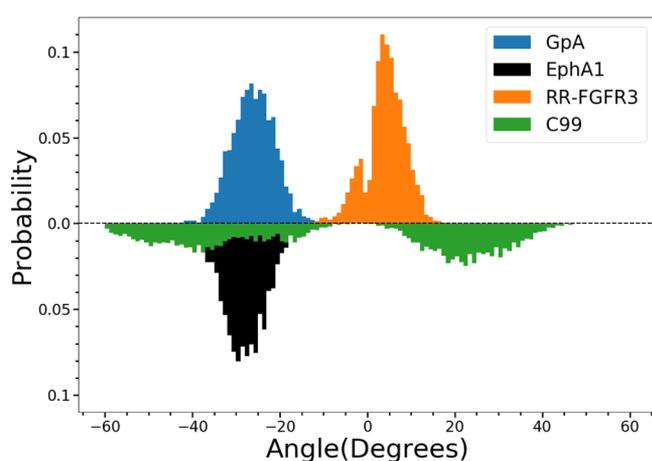

Figure 3. Crossing angle distribution of the GpA, EphA1, RR-FGFR3, and C99 TM helices represented by blue, black, orange, and green bars. The maximum of crossing angle distribution of GpA, EphA1, and RR-FGFR3 appears at −26°, −29°, and 5°, respectively. While C99 shows both left-handed and right-handed helical structures. The simulations were performed using MARTINI v2.2.

difference, σ^G, of the free energy of dimerization obtained from simulation and experiment (Table 1) defined as

$$\sigma^G = \sqrt{\frac{\sum_i \left(G_i^{exp} - G_i^{sim}\right)^2}{4}} \qquad (2)$$

where G^{exp} and G^{sim} are the dimerization free energies obtained from experiment and simulation, respectively. By this measure, an α value of 1.045 appears to provide the best overall fit to the experimental data for the four systems studied, while $\alpha = 1.04$ provides the best fit for GpA and EphA1. Scaling factors of $\alpha = 1.04$ or 1.045 lead to dimerization free energies in line with the experimental results (see Table 1).

We have analyzed the homodimer structures predicted by the reparameterized MARTINI model by characterizing the contacts between the TM helices of each homodimer in the reparameterized MARTINI simulation. The contact map of the protein dimer and the interaction map between the same residues of the TM helices are shown in the Supporting Information, Figures S3 and S4. The crossing angle distribution for all dimers is displayed in the Supporting Information, Figure S5. In the reparameterized MARTINI model, the GpA and EphA1 dimers are primarily stabilized by
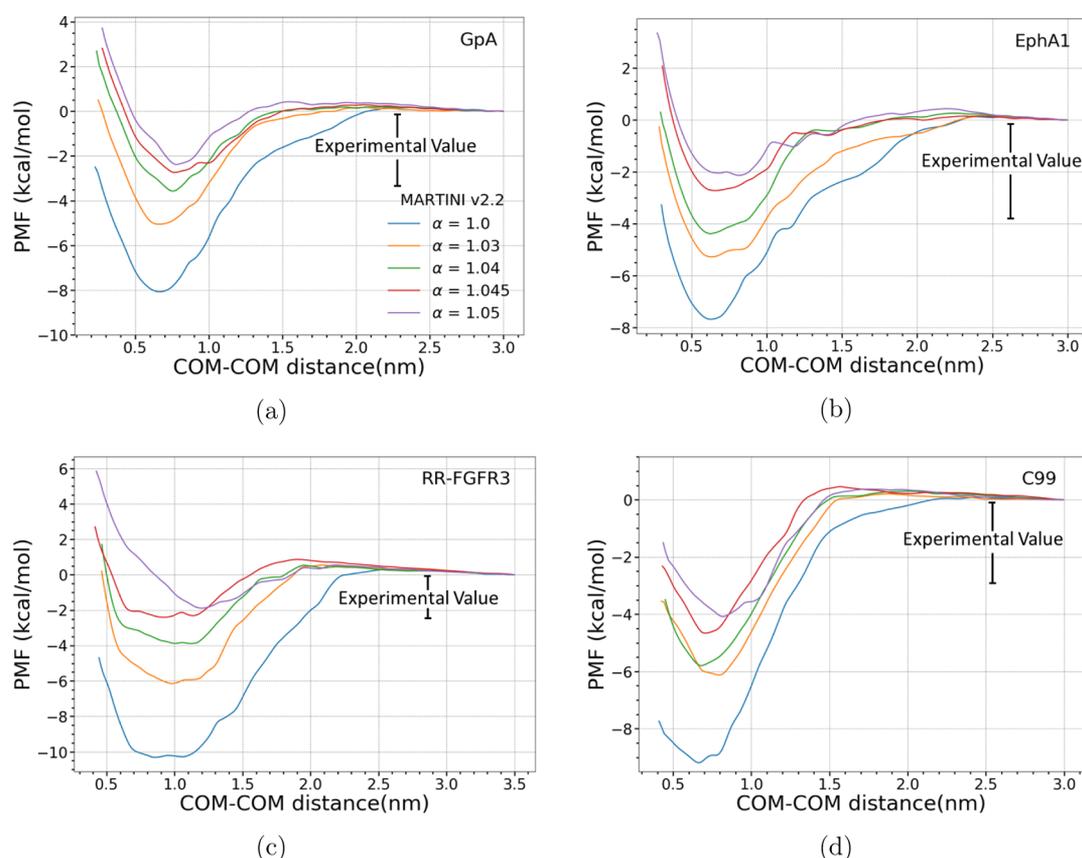
Figure 4. Dimerization free-energy profiles of (a) GpA, (b) EphA1, (c) RR-FGFR3, and (d) C99 using the reparameterized MARTINI model. The maximum value of the PMF curve was used to set the zero and frame the comparison.

Table 1. Dimerization Free Energies Obtained for the TM Proteins Studied Using the Standard MARTINI v2.2 Model and Reparameterized MARTINI Model

	GpA kcal/mol	EphA1 kcal/mol	RR-FGFR3 kcal/mol	C99 kcal/mol	σ^G
Experimental value	3.6[19,35]	3.7[23]	2.7[36]	3.2[37]	
MARTINI v2.2 ($\alpha = 1.0$)	8.0	7.7	10.3	9.1	5.65
$\alpha = 1.03$	5.0	5.3	6.2	6.1	2.51
$\alpha = 1.04$	3.6	4.4	3.9	5.7	1.43
$\alpha = 1.045$	2.8	2.8	2.4	4.6	0.93
$\alpha = 1.05$	2.4	2.1	1.9	4.0	1.15

interactions mediated by the *GXXXG* motif, as observed in the standard MARTINI model with crossing angles of −26° at α = 1.04 (−26° at α = 1.045) and −27° at α = 1.04 (−29° at α = 1.045), respectively. The RR-FGFR3 dimer is stabilized by interactions conveyed by residues composing the $G_{380}X_2FF_{384}X_2ILX_2A_{391}$ motif that act to stabilize a left-handed helical structure with a crossing angle of 4° at α = 1.04 (6° at α = 1.045). The observed interactions of the TM helices predicted by the reparameterized MARTINI model, as measured by residue−residue contacts and crossing angle between TM helices, are consistent with results derived from simulations using the standard MARTINI force field, as well as with experimental observations.

To further explore the impact of scaling the protein−lipid LJ interactions, we have also measured the average depth-of-insertion in the bilayer of the residues of each TM protein studied using both MARTINI v2.2 and the reparameterized MARTINI model. The results are shown in the Supporting Information, Figure S6. The insertion profile for all protein residues is largely unchanged by rescaling of the protein−lipid

interactions. Taken together, these results demonstrate that the reparameterized MARTINI model preserves the structural specificity of the homodimer studied while correcting the homodimer stability.

The proposed modification of the MARTINI model is designed to improve the modeling of TM protein−protein interactions in membranes. To further validate the reparameterization of the MARTINI model, we calculated the partitioning free energy of amino acid side chains in DOPC lipid bilayer using both MARTINI v2.2 and the reparameterized MARTINI model (Figure 5). This is an important validation step as the nonbonded parameters of the amino acid side chains of the MARTINI model are adjusted to reflect thermodynamic data describing the partition coefficient between water and organic solvent. As can be seen in the free-energy profile, the partitioning PMF remains mostly unchanged after the reweighting of the LJ interaction governing the lipid−protein interactions. Exceptions include Ile/Leu and Phe where differences are 9% and 23%, respectively.
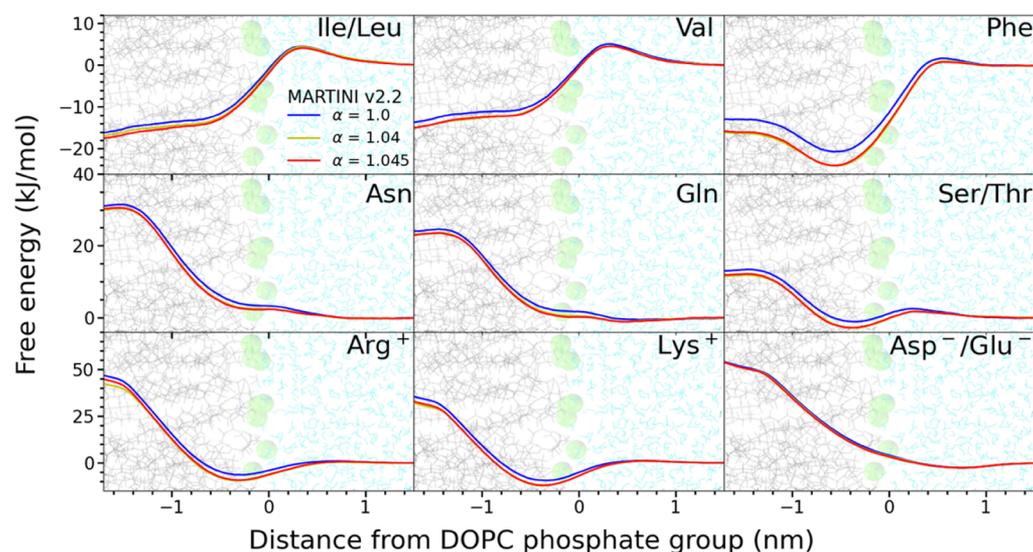
Figure 5. PMF for the transfer of an amino acid side chain from DOPC bilayer to water solvent. The value of the PMF on the water side (positive *x*-axis) was used to set the zero and frame the comparison. The peak density of the phosphate group (green balls) of the lipid was set to 0 nm. The positive and negative values of the distances refer to the water solvent and lipid bilayer, respectively. The blue, yellow, and red lines represent the results of simulations using MARTINI v2.2 with $\alpha = 1.0$, $\alpha = 1.04$, and $\alpha = 1.045$, respectively.
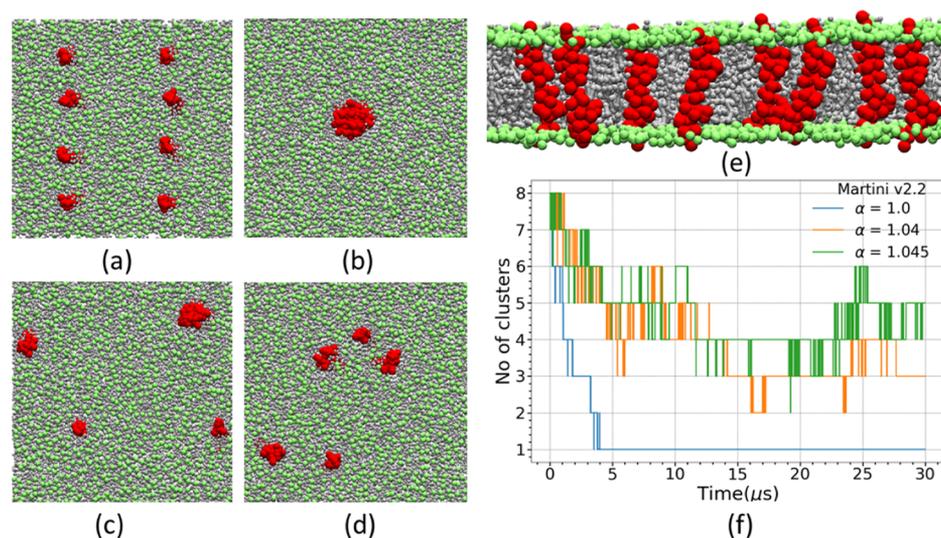

Figure 6. Top view of the (a) initial bilayer structure of the multiple protein simulation and equilibrated lipid bilayer structures obtained (b) using MARTINI v2.2 and for the reparameterized MARTINI model using (c) $\alpha = 1.04$ and (d) $\alpha = 1.045$. (e) Side view of the initial bilayer structure used in the multiple protein simulations. The first bead of the head group of POPC, the tail of POPC, or GpA is represented by green, grey, or red color, respectively. (f) Number of clusters in the membrane bilayer obtained by the simulation of eight GpA proteins in a POPC bilayer. The blue, orange, and green lines represent the results of simulations using MARTINI v2.2 with $\alpha = 1.0$, $\alpha = 1.04$, and $\alpha = 1.045$, respectively.

In order to further test the predictions of the reparameterized MARTINI model, we have simulated a membrane bilayer comparable to a realizable experimental composition consisting of eight GpA proteins in a POPC bilayer with a protein-to-lipid ratio of 1:200 using MARTINI v2.2 and the reparameterized MARTINI model with $\alpha = 1.04$ and 1.045. A 30 μs simulation was performed by initially placing eight GpA proteins evenly separated in a POPC bilayer with a protein-to-lipid ratio of 1:200. The results obtained from this simulation are presented in Figure 6. In the case of the MARTINI v2.2 model, all proteins are observed to aggregate to form an octameric protein oligomer within the early stage of the simulation. Once formed, the oligomer persists throughout the remainder of the simulation. In contrast, the reparameterized

MARTINI model shows a significant improvement in terms of the interprotein dynamics. In particular, while the MARTINI v2.2 model shows the formation of a single aggregate, the reparameterized model presents a dynamic equilibrium with a fluctuating population of monomers and dimers that is established and maintained. We have also calculated the total population of monomeric and dimeric clusters in the simulation (Supporting Information, Figure S7). In the simulation using $\alpha = 1.045$, the majority of the protein is observed to exist as monomer or dimer, consistent with our expectation based on experimental observation.

We have also analyzed the contact surface of the GpA dimers in this multiple protein simulation (Supporting Information, Figure S8). The contact maps demonstrate that

GpA proteins associate through favorable residue—residue interactions that are not limited to those residues composing the *GxxxG* motifs. In the CG MARTINI model, the parameters defining the residue—residue interaction are relatively uniform when compared with the parameters of an all-atom or more detailed CG model. Consider the case of GpA where the MARTINI model treats glycine and alanine using similar intermolecular interaction terms. Gly—Gly and Ala—Ala interactions are identical, while Ala—lipid and Gly—lipid are different. We observe that alanine residues can play a role similar to glycine in stabilizing interhelical interaction. The existence of these competing homodimer structures underlying the heterogeneity was observed in the homodimer ensemble when simulated using the reparameterized MARTINI model.

CONCLUSIONS

The MARTINI force field is a widely used CG model with the capacity to model a large library of molecules and complex molecular systems.[42,43] However, reliable simulation of multiple proteins in membranes has not been possible using the MARTINI v2.2 or 2.2P force field due to the exaggerated tendency of membrane proteins to aggregate. As cell membranes generally contain up to 50% protein,[44] it is critical to eliminate this artifact to enable the simulation of realistic models of cellular membranes using the MARTINI model. To address this shortcoming, we estimated the PMF for homodimer formation using the MARTINI v2.2 force field and compared the predictions with experimental results obtained for the dimerization free energy of GpA, EphA1, FGFR3, and C99 dimers. While predicted TM dimer structures were found to agree with experimental observations, the dimerization free energy of each TM protein predicted by the MARTINI model was found to be considerably more stable than indicated by FRET experiments.

In order to stabilize the separated monomers relative to the dimer state, we scaled the MARTINI protein—lipid LJ nonbonded interaction parameters keeping all other interactions unchanged. We explored the effect of different scaling factors on four TM proteins in order to obtain the best fit to the experimentally derived dimerization free energies. We concluded that a 4—4.5% upscaling of the LJ interaction is an optimum solution irrespective of the membrane composition or protein sequence. To test the reparameterization, a simulation of multiple TM proteins was performed using the MARTINI v2.2 model and the modified force field. While the standard MARTINI force field leads to the formation of a single protein aggregate, the modified parameters lead to a dynamic equilibrium of protein association and dissociation in agreement with experiment.

The reparameterization of the MARTINI v2.2 force field proposed in this work addresses shortcomings of this widely used CG model. Our results demonstrate that a simple rescaling of the nonbonded LJ interaction between protein and lipid leads to qualitative improvements in the behavior of the simulation involving multiproteins. While no single value of the scaling parameter α is best for all proteins studied, a value of α = 1.045 provides a significant improvement in the observed stability of the protein homodimers while preserving the specificity of binding. In addition to lipid bilayers, protein homodimers in aqueous environments were also reported to be overstabilized when simulated using the MARTINI model.[12] Given adequate experimental data for protein association constants, a similar strategy to the one employed here could be

used to reparameterize the protein—water interactions in the MARTINI force field with the goal of addressing the overstabilization of associated proteins in water and improving the association constants for proteins in aqueous solution.

In closing, we note that the MARTINI force field employs a variety of well depths, ε_{ij}, in the LJ nonbonded interactions. In this work, we have explored a single scaling factor for all ε_{ij} values. This approach was justified by the uniform over-stabilization of homodimer structures for four TM proteins. At this time, there exists limited experimental and all-atom simulation data for protein—protein association in membranes. Nevertheless, with expanded experimental data or data derived from extensive all-atom simulations, more than one scaling factor could be used to obtain more quantitative fits to TM homodimer stability across a variety of proteins.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jctc.0c01253.

> System compositions examined in this study; sensitivity of the dimerization free energy to finite size effects; dimerization free energy of GpA using the MARTINI v2.2 and MARTINI v2.2p force fields; contact fraction between residues of two TM helices; contact map of the protein dimer; crossing angle distribution of the protein dimer; depth-of-insertion of protein residues; total population of monomers and dimers obtained from the multiple protein simulation; and contact map of the GpA dimer obtained from the multiple protein simulation (PDF)

AUTHOR INFORMATION

Corresponding Author

John E. Straub — *Department of Chemistry, Boston University, Boston 02215, Massachusetts, United States;* orcid.org/0000-0002-2355-3316; Email: straub@bu.edu

Author

Ayan Majumder — *Department of Chemistry, Boston University, Boston 02215, Massachusetts, United States*

Complete contact information is available at:
https://pubs.acs.org/10.1021/acs.jctc.0c01253

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

REFERENCES

(1) Ghosh, A.; Sonavane, U.; Joshi, R. Multiscale modelling to understand the self-assembly mechanism of human β2-adrenergic receptor in lipid bilayer. *Comput. Biol. Chem.* **2014**, *48*, 29—39.

(2) Mondal, S.; Johnston, J. M.; Wang, H.; Khelashvili, G.; Filizola, M.; Weinstein, H. Membrane driven spatial organization of GPCRs. *Sci. Rep.* **2013**, *3*, 2909.

(3) Periole, X.; Huber, T.; Marrink, S.-J.; Sakmar, T. P. G Protein-Coupled Receptors Self-Assemble in Dynamics Simulations of Model Bilayers. *J. Am. Chem. Soc.* **2007**, *129*, 10126−10132.

(4) Koldsø, H.; Sansom, M. S. P. Organization and dynamics of receptor proteins in a plasma membrane. *J. Am. Chem. Soc.* **2015**, *137*, 14694−14704.

(5) Bond, P. J.; Sansom, M. S. P. Insertion and assembly of membrane proteins via simulation. *J. Am. Chem. Soc.* **2006**, *128*, 2697−2704.

(6) 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. Theor. Comput.* **2008**, *4*, 819−834.

(7) 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*, 7812−7824.

(8) 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. Theor. Comput.* **2013**, *9*, 687−697.

(9) Bandara, A.; Panahi, A.; Pantelopulos, G. A.; Nagai, T.; Straub, J. E. Exploring the impact of proteins on the line tension of a phase-separating ternary lipid mixture. *J. Chem. Phys.* **2019**, *150*, 204702.

(10) Pantelopulos, G. A.; Nagai, T.; Bandara, A.; Panahi, A.; Straub, J. E. Critical size dependence of domain formation observed in coarse-grained simulations of bilayers composed of ternary lipid mixtures. *J. Chem. Phys.* **2017**, *147*, 095101.

(11) Nishizawa, M.; Nishizawa, K. Potential of mean force analysis of the self-association of leucine-rich transmembrane α-helices: Difference between atomistic and coarse-grained simulations. *J. Chem. Phys.* **2014**, *141*, 075101.

(12) Stark, A. C.; Andrews, C. T.; Elcock, A. H. Toward optimized potential functions for protein−protein interactions in aqueous solutions: osmotic second virial coefficient calculations using the martini coarse-grained force field. *J. Chem. Theor. Comput.* **2013**, *9*, 4176−4185.

(13) Periole, X. Interplay of G protein-coupled receptors with the membrane: insights from supra-atomic coarse grain molecular dynamics simulations. *Chem. Rev.* **2017**, *117*, 156−185.

(14) Dunton, T. A.; Goose, J. E.; Gavaghan, D. J.; Sansom, M. S.; Osborne, J. M. The free energy landscape of dimerization of a membrane protein, NanC. *PLoS Comput. Biol.* **2014**, *10*, e1003417.

(15) Chavent, M.; Chetwynd, A. P.; Stansfeld, P. J.; Sansom, M. S. P. Dimerization of the EphA1 receptor tyrosine kinase transmembrane domain: insights into the mechanism of receptor activation. *Biochemistry* **2014**, *53*, 6641−6652.

(16) Johnston, J. M.; Wang, H.; Provasi, D.; Filizola, M. Assessing the relative stability of dimer interfaces in g protein-coupled receptors. *PLoS Comput. Biol.* **2012**, *8*, e1002649.

(17) Huang, J.; MacKerell, A. D., Jr CHARMM36 all-atom additive protein force field: Validation based on comparison to NMR data. *J. Comput. Chem.* **2013**, *34*, 2135−2145.

(18) MacKerell, A. D., Jr; Bashford, D.; Bellott, M.; Dunbrack, R. L., Jr; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; et al. All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J. Phys. Chem. B* **1998**, *102*, 3586−3616.

(19) Domański, J.; Sansom, M. S.; Stansfeld, P. J.; Best, R. B. Balancing force field protein−lipid interactions to capture transmembrane helix−helix association. *J. Chem. Theor. Comput.* **2018**, *14*, 1706−1715.

(20) You, M.; Li, E.; Wimley, W. C.; Hristova, K. Förster resonance energy transfer in liposomes: measurements of transmembrane helix dimerization in the native bilayer environment. *Anal. Biochem.* **2005**, *340*, 154−164.

(21) Javanainen, M.; Martinez-Seara, H.; Vattulainen, I. Excessive aggregation of membrane proteins in the Martini model. *PLoS One* **2017**, *12*, e0187936.

(22) Chen, L.; Merzlyakov, M.; Cohen, T.; Shai, Y.; Hristova, K. Energetics of ErbB1 transmembrane domain dimerization in lipid bilayers. *Biophys. J.* **2009**, *96*, 4622−4630.

(23) Artemenko, E. O.; Egorova, N. S.; Arseniev, A. S.; Feofanov, A. V. Transmembrane domain of EphA1 receptor forms dimers in membrane-like environment. *Biochim. Biophys. Acta* **2008**, *1778*, 2361−2367.

(24) MacKenzie, K. R.; Prestegard, J. H.; Engelman, D. M. A transmembrane helix dimer: structure and implications. *Science* **1997**, *276*, 131−133.

(25) Bocharov, E. V.; Mayzel, M. L.; Volynsky, P. E.; Goncharuk, M. V.; Ermolyuk, Y. S.; Schulga, A. A.; Artemenko, E. O.; Efremov, R. G.; Arseniev, A. S. Spatial structure and pH-dependent conformational diversity of dimeric transmembrane domain of the receptor tyrosine kinase EphA1. *J. Biol. Chem.* **2008**, *283*, 29385−29395.

(26) Bocharov, E. V.; Lesovoy, D. M.; Goncharuk, S. A.; Goncharuk, M. V.; Hristova, K.; Arseniev, A. S. Structure of FGFR3 transmembrane domain dimer: implications for signaling and human pathologies. *Structure* **2013**, *21*, 2087−2093.

(27) Nadezhdin, K. D.; Bocharova, O. V.; Bocharov, E. V.; Arseniev, A. S. Dimeric structure of transmembrane domain of amyloid precursor protein in micellar environment. *FEBS Lett.* **2012**, *586*, 1687−1692.

(28) 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. Theor. Comput.* **2015**, *11*, 2144−2155.

(29) De Jong, D. H.; Baoukina, S.; Ingólfsson, H. I.; Marrink, S. J. Martini straight: Boosting performance using a shorter cutoff and GPUs. *Comput. Phys. Commun.* **2016**, *199*, 1−7.

(30) Hub, J. S.; De Groot, B. L.; Van Der Spoel, D. g_wham-A Free Weighted Histogram Analysis Implementation Including Robust Error and Autocorrelation Estimates. *J. Chem. Theor. Comput.* **2010**, *6*, 3713−3720.

(31) Abraham, M. J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J. C.; Hess, B.; Lindahl, E. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to super-computers. *SoftwareX* **2015**, *1−2*, 19−25.

(32) Popot, J. L.; Engelman, D. M. Membrane protein folding and oligomerization: the two-stage model. *Biochemistry* **1990**, *29*, 4031−4037.

(33) Endres, N. F.; Das, R.; Smith, A. W.; Arkhipov, A.; Kovacs, E.; Huang, Y.; Pelton, J. G.; Shan, Y.; Shaw, D. E.; Wemmer, D. E.; et al. Conformational coupling across the plasma membrane in activation of the EGF receptor. *Cell* **2013**, *152*, 543−556.

(34) Prakash, A.; Janosi, L.; Doxastakis, M. Self-association of models of transmembrane domains of ErbB receptors in a lipid bilayer. *Biophys. J.* **2010**, *99*, 3657−3665.

(35) Sarabipour, S.; Hristova, K. Glycophorin A transmembrane domain dimerization in plasma membrane vesicles derived from CHO, HEK 293T, and A431 cells. *Biochim. Biophys. Acta* **2013**, *1828*, 1829−1833.

(36) You, M.; Li, E.; Hristova, K. The achondroplasia mutation does not alter the dimerization energetics of the fibroblast growth factor receptor 3 transmembrane domain. *Biochemistry* **2006**, *45*, 5551−5556.

(37) Song, Y.; Hustedt, E. J.; Brandon, S.; Sanders, C. R. Competition between homodimerization and cholesterol binding to the C99 domain of the amyloid precursor protein. *Biochemistry* **2013**, *52*, 5051−5064.

(38) Trenker, R.; Call, M. E.; Call, M. J. Crystal structure of the glycophorin A transmembrane dimer in lipidic cubic phase. *J. Am. Chem. Soc.* **2015**, *137*, 15676−15679.

(39) Reddy, T.; Manrique, S.; Buyan, A.; Hall, B. A.; Chetwynd, A.; Sansom, M. S. P. Primary and secondary dimer interfaces of the fibroblast growth factor receptor 3 transmembrane domain: character-

ization via multiscale molecular dynamics simulations. *Biochemistry* **2014**, *53*, 323−332.

(40) Dominguez, L.; Foster, L.; Straub, J. E.; Thirumalai, D. Impact of membrane lipid composition on the structure and stability of the transmembrane domain of amyloid precursor protein. *Proc. Natl. Acad. Sci. U. S. A.* **2016**, *113*, E5281−E5287.

(41) Moore, D. T.; Berger, B. W.; DeGrado, W. F. Protein-protein interactions in the membrane: sequence, structural, and biological motifs. *Structure* **2008**, *16*, 991−1001.

(42) Qi, Y.; Ingólfsson, H. I.; Cheng, X.; Lee, J.; Marrink, S. J.; Im, W. CHARMM-GUI Martini maker for coarse-grained simulations with the Martini force field. *J. Chem. Theor. Comput.* **2015**, *11*, 4486−4494.

(43) Wassenaar, T. A.; Pluhackova, K.; Moussatova, A.; Sengupta, D.; Marrink, S. J.; Tieleman, D. P.; Böckmann, R. A. High-throughput simulations of dimer and trimer assembly of membrane proteins. The DAFT approach. *J. Chem. Theor. Comput.* **2015**, *11*, 2278−2291.

(44) Dupuy, A. D.; Engelman, D. M. Protein area occupancy at the center of the red blood cell membrane. *Proc. Natl. Acad. Sci. U. S. A.* **2008**, *105*, 2848−2852.