JCTC Journal of Chemical Theory and Computation

pubs.acs.org/JCTC Article

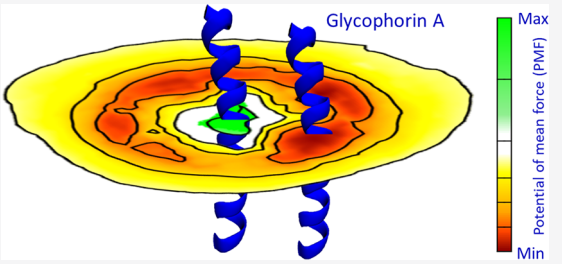
On Computing Equilibrium Binding Constants for Protein−Protein Association in Membranes

Ayan Majumder, Seulki Kwon, and John E. Straub*

Cite This: J. Chem. Theory Comput. 2022, 18, 3961−3971 Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: Protein association in lipid membranes is fundamental to membrane protein function and of great biomedical relevance. All-atom and coarse-grained models have been extensively used to understand the protein−protein interactions in the membrane and to compute equilibrium association constants. However, slow translational and rotational diffusion of protein in membrane presents challenges to the effective sampling of conformations defining the ensembles of free and bound states contributing to the association equilibrium and the free energy of dimerization. We revisit the homodimerization equilibrium of the TM region of glycophorin A. Conformational sampling is performed using umbrella sampling along previously proposed one-dimensional collective variables and compared with sampling over a two-dimensional collective variable space using the MARTINI v2.2 force field. We demonstrate that the one-dimensional collective variables suffer from restricted sampling of the native homodimer conformations leading to a biased free energy landscape. Conversely, simulations along the two-dimensional collective variable effectively characterize the thermodynamically relevant native and non-native interactions contributing to the association equilibrium. These results demonstrate the challenges associated with accurately characterizing binding equilibria when multiple poses contribute to the bound state ensemble.



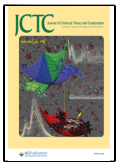
INTRODUCTION

The association of transmembrane (TM) proteins plays a key role in membrane protein structure and function,[1] with G-protein coupled receptors being one important example.[2] In addition, membrane proteins are known to play an important role in the biogenesis of amyloid proteins associated with a variety of diseases including Alzheimer's disease[3] and Parkinson's disease.[4] Classical computer simulations have been used to predict the structure of TM protein complexes in membranes.[5−7] Additionally, umbrella sampling (US) may be used to calculate free energy surfaces or potentials of mean force (PMF) that may be used to quantitatively compute equilibrium binding constants. A converged and reproducible PMF can be compared with an experimental data for quantitative comparison of different interactions and can be used to calibrate the molecular dynamics models.

The slow diffusion of molecules in membranes relative to water leads to significant challenges in effectively sampling conformations of protein and lipids,[8] a challenge that grows with the increasing complexity of the membrane model.[9] As such, TM protein association is most commonly studied using multiscale modeling approaches that rely on coarse-grained models.[10,11] To calculate the free energy surface of a TM protein homodimer, it is important to have a proper sampling of all thermodynamically relevant free energy basins characterizing the membrane protein conformation ensemble.[12,13]

Enhanced sampling methods that require the effective sampling of reaction coordinates or collective variables (CV), including umbrella sampling,[14,15] metadynamics,[16] and adaptive biasing force,[17] have been extensively used to address the sampling challenges in membrane protein simulation.[18] These methods have also been used to calculate protein−protein or protein−lipid free energy surfaces in complex membrane systems using both all-atom[19] or coarse-grained models.[11,20] To calculate the PMF relevant in describing protein homodimerization using umbrella sampling methods, we must define a reaction coordinate connecting the relevant bound and dissociated states. The distance between the center-of-mass of protein helices has been commonly used in defining the umbrella bias.[21,22] In this approach, the initial configurations of different umbrella windows were generated by a simple translation of TM helices along the reaction coordinate. After a sufficiently long simulation in each window, unbiasing can be performed by using the weighted histogram analysis method (WHAM) tool to obtain the final PMF.[23,24] While the alchemical decoupling method is effective for characterizing receptor−ligand interactions,[25] the PMF-based strategy is most

Received: January 31, 2022
Published: May 17, 2022



effective in characterizing protein−protein interactions and the associated binding constants.

The theoretical foundation for the rigorous calculation of binding free energies was developed over a period of 20 years.[26−37] Those studies were extended to calculations exploring binding free energies of proteins in membranes.[11,38−40] Recent perspective articles and textbooks have organized this collected wisdom in the form of guidance on best practices for free energy calculation.[41−45] While careful works were done to provide a rigorous theoretical foundation, the application of those ideas has proven to be more complex. There are three concerns most commonly raised. (1) When using a computed potential of mean force (PMF) to determine a binding constant, the binding free energy should be determined by an integral over the PMF rather than the surrogate minimum of the PMF. While the minimum of the PMF is often a reasonable approximation to the free energy difference, that is not always the case.[45] (2) Symmetry factors arising from molecular point group symmetries, or symmetry factors involved in homodimerization as opposed to heterodimerization, are straightforward in their formulation but challenging in their application.[37−39,41] (3) The choice of reaction coordinate or order parameter in the calculation of a potential of mean force can facilitate or frustrate sampling in the calculation of PMFs. The string method may be used to create curvilinear paths.[44] Such paths should connect initial and final states, which may be characterized by multiple substates.

Glycophorin A (GpA) homodimer formation has been extensively studied as a model of the TM protein to analyze the homodimer interactions in membranes. The small size and simple structure of GpA along with an abundance of available experimental data in different conditions make this homodimer system an appropriate choice in accessing the accuracy of the computational models and sampling methods.

The structure of GpA has been derived from solution NMR in detergent micelles,[46] solid-state NMR in dimyristoylphosphocholine (DMPC),[47] and an X-ray crystallographic structure obtained in the lipidic cubic phase (LCP),[48] providing a consistent picture of GpA dimer. Key interactions stabilizing the GpA homodimer are mediated by steric and C_α hydrogen bonding mediated by the GXXXG TM domain sequence motif. Mutagenesis experiments have been performed in which destabilizing the GXXXG motif[49,50] or mutating the T87 residue[51,52] has been reported to decrease the homodimerization propensity.

Several experimental studies have been performed to analyze the free energy of dimerization of the GpA dimer. Fleming et al. reported the free energy of dissociation of the GpA dimer to be 9.0 kcal/mol in pentaoxyethylene octyl ether (C_8E_5).[53] The mole fraction standard state free energy change for GpA dimerization was measured to be 7.0 kcal/mol in C_8E_5 and 5.5 to 4.5 kcal/mol in SDS at 25 °C.[54] The free energy of dissociation of the GpA in the *Escherichia coli* inner membrane was found to be 7.5 kcal/mol.[55] Hong et al. has reported the free energy of dissociation of the GpA dimer to be 12.1 kcal/mol using the steric trap method in a POPC bilayer.[56] Experiments on the plasma membrane reported the dimerization free energy between 3.4 and 4.0 kcal/mol.[57,58] Theoretical studies have been used to calculate the dissociation kinetics of the GpA dimer in different conditions. Sengupta et al. reported the free energy of dissociation to be 9.1 kcal/mol in a DPPC bilayer.[21] Janosi et al. reported the dimerization free energy of the GpA homodimer to be 6.6 and 6.7 kcal/mol in a DOPC and DLPC bilayer and 7.5 kcal/mol in a DPPC bilayer.[59] Using the CHARMM27 force field, Hénin et al. reported the free energy of dimerization to be 11.5 kcal/mol.[38] This data is summarized in the Supporting Information, Table 1. We note that the measurement of binding constants characterizing the reversible lateral association/dissociation defining the equilibrium state of proteins in a membrane is extremely challenging. As such, there is substantial hidden uncertainty reflected in the range of measured values.

The coarse-grained MARTINI force field[60,61] is a widely used model capable of representing a wide variety of materials and biomolecular systems. It has been used in many studies characterizing complex membrane bilayer systems.[62,63] However, previous studies have shown that in umbrella sampling simulations of protein homodimer formation in the membrane, the conformational sampling was restricted to the native contact region.[20,22] All possible conformations of the dimer were not properly sampled along a particular radial distance.[21,64]

In this work, we discuss the difficulties of calculating binding free energies in the context of the homodimerization of the GpA protein in a lipid bilayer. While the issues of the relationship between the PMF and the binding constant have been clarified,[11,38−40] the issue of the impact of the choice of reaction coordinate on the sampling of the free energy surface remains challenging. We find that recent studies employing reaction coordinates based on center-of-mass or native dimer contacts can fail to sample critical regions of the homodimer ensemble, leading to errors in the computed binding constants. We performed umbrella sampling over the plane defined by the x-projection and y-projection of the COM−COM distance between GpA monomers using the MARTINI force field to characterize the ensemble of conformations informing the GpA homodimer binding constant. We further elucidate this point by exploring the calculation of binding constants for two simple model systems, one having a single binding pose and one having multiple binding poses (Supporting Information, Appendix B). The results obtained using exhaustive sampling achieved through Brownian dynamics demonstrate that the time scale for sampling the multiple binding poses places a lower bound on the time scale required to determine accurate binding constants.

METHODS

The dimerization equilibrium of GpA_{69-97} (PDB entry 1AFO[46]) was simulated in this study. The GpA_{69-97} dimer ensemble was found to include contributions from multiple conformation substates. In addition, a simple model was constructed in order to gain insight into the dimerization equilibrium of GpA and the time scale required for complete sampling (Supporting Information, Appendix B).

Coarse-Grained Simulation of GpA Dimerization Equilibrium. The GpA dimer was placed in a POPC bilayer consisting of 408 lipid molecules using insane.py program.[65] The N-terminus of the protein was placed in the upper leaflet. The lipid bilayer was solvated by 25 nonpolarizable water beads per two lipid molecules with 10% being antifreeze water beads and a salt concentration of 0.15 mol. The MARTINI v2.2 force field[60,61] was used, and the simulation parameters were set according to the "common" parameter file referred to on the MARTINI Web site.[66] The leapfrog integration method was used with 20 fs time steps. A velocity rescaling thermostat

https://doi.org/10.1021/acs.jctc.2c00106
J. Chem. Theory Comput. 2022, 18, 3961−3971

was used to maintain the temperature of the simulation at 310 K. A 200 ns simulation was performed to equilibrate the lipid bilayer system. A semi-isotropic Berendsen barostat with a compressibility of 3×10^{-4} bar^{-1} was used during the equilibration steps, and a semi-isotropic Parrinello−Rahman barostat was used during the production run to maintain a pressure of 1 bar.

Two different reaction coordinates were used to calculate the one-dimensional potential of mean force (PMF) between two GpA helices. The (1) center-of-mass (COM) distance between two helices (D_{com}) and (2) distance root-mean-square displacement (D_{rmsd}) were used as reaction coordinates in the umbrella sampling simulations. In each case, data obtained while sampling over D_{com} or D_{rmsd} were used to determine the PMF as a function of D_{com}, and integration over D_{com} was used to determine the binding constant. A harmonic restraint of 1000 kJ mol^{-1} nm^{-2} was placed at a separation of 0.15 nm along D_{com} to ensure proper overlap between adjacent umbrella windows. A 2 μs production run was performed for each umbrella window. For the umbrella sampling simulation with D_{rmsd}, the experimental structure obtained from PDB entry 1AFO was used as a reference structure (X^{ref}). The collective variable D_{rmsd} is defined as

$$D_{rmsd}(X, X^{ref}) = \sqrt{\frac{1}{N(N-1)} \sum_{i \neq j} [d(x_i, x_j) - d(x_i^{ref}, x_j^{ref})]^2} \quad (1)$$

where $d(x_i, x_j)$ is the distance between atoms i and j in configuration X. Residues Glu72 to Ile95 were used in calculating the D_{rmsd}, where a lower cutoff of 0.1 nm and upper cutoff of 0.6 nm was used.[11] A harmonic potential with a force constant of 100 kJ mol^{-1} nm^{-2} was placed at a separation of 0.15 nm along the D_{rmsd} coordinate. A 5 μs production run was performed for each umbrella window.

Two-dimensional umbrella sampling was performed over the plane defined by the x-projection and y-projection of the COM−COM distance between the two helices. To restrict the rotational motion of one of the helices, an angular harmonic restraint of 2000 kJ mol^{-1} rad^{-2} was used on the angle between the origin(COM) to Gly79 vector of the helix and the positive x-axis. A harmonic restraint of 1000 kJ mol^{-1} nm^{-2} was placed at a separation of 0.2 nm along both dimensions. A 2 μs production run was performed for each umbrella window. The WHAM tool developed in the Grosfield lab[67] was used to unbias the umbrella sampling simulations. All simulations were performed using the GROMACS 2018.3 program,[68] and the simulations with D_{rmsd} were performed using the PLUMED version 2.6.3[69] patched with the GROMACS 2018.3 program.

Computing Equilibrium Binding Constants from a PMF. The equilibrium probability distribution function, p_{eq}, was determined as a function of radial distance $p_{eq}^{WHAM}(r)$ and a function of position in the xy-plane $p_{eq}^{WHAM}(x,y)$. From those probability distribution functions, the associated potentials of mean force and equilibrium binding constants were determined.

The 1D potential of mean force as a function of distance r is defined by $W(r) = -k_BT \ln g(r)$, where $g(r)$ is a radial distribution function. The 1D equilibrium probability distribution obtained from WHAM is related to $g(r)$ such that $p_{eq}^{WHAM}(r) \sim 2\pi r g(r)$, as the umbrella sampling was performed over the 2D xy-plane. As such, we find that $\Delta W(r) = W(r) - W_{bulk}$ can be expressed as $\Delta W(r) = -k_BT \ln p_{eq}^{WHAM}(r) + k_BT \ln(r)$. The term $k_BT \ln(r)$, sometimes referred to as an entropic factor, is included to account for the increasing area accessible to dimers with increasing r. When this factor is included, $\Delta W(r)$ converges to a plateau value at large r. Certain prior studies have explicitly included this factor.[19,24,70−72] The corresponding binding constant $K_{a,1D}$ is defined as

$$K_{a,1D} = \frac{1}{2} \times \frac{2\pi}{A_0} \int r e^{-\Delta W(r)/k_BT} \, dr \quad (2)$$

where $\Delta W(r)$ is the potential of mean force at a distance r along D_{com}. The value of $K_{a,1D}$ was also determined, using US simulation data along the D_{rmsd} coordinate. In that case, the calculation was performed using a 1D PMF as a function of D_{com}, where the 1D PMF was derived from the 2D PMF along D_{com} and D_{rmsd} (Supporting Information, Figure 1).

The 2D potential of mean force is defined $\Delta W(x,y) = -k_BT \ln p_{eq}^{WHAM}(x,y)$. In this case, no explicit entropic correction is required. The corresponding binding constant $K_{a,2D}$ is defined as

$$K_{a,2D} = \frac{1}{2} \times \frac{1}{A_0} \iint e^{-\Delta W(x,y)/k_BT} \, dx dy \quad (3)$$

In each expression above, A_0 is the standard reference area set to 1 nm^2. The factor of $\frac{1}{2}$ appearing in each expression is included to account for the symmetry of the homodimer.[41] Prior studies have[21,64] and have not[19] included this factor. In the case of exhaustive sampling, the exact calculation of $\Delta W(r)$, resulting in the binding constant $K_{a,1D}$, should agree with the value of $K_{a,2D}$, derived from $\Delta W(x, y)$. The equilibrium binding constant may be reported using a standard concentration, expressed in terms of the number of molecules per unit area, or the number of proteins per lipid. The latter accounts for variations in the average area per lipid.[54]

The free energy of association can be obtained from K_a as $\Delta G = -k_BT \ln K_a$. It is useful to consider the approximate formula for the binding constant using the area of binding states (A^{site}) as[45]

$$K_a \simeq \frac{A^{site}}{A_0} e^{-\Delta W_{min}/k_BT}$$

$$\Delta W_{min} - \Delta G \simeq k_BT \ln \frac{A^{site}}{A_0} \quad (4)$$

If binding occurs at a specific site, $A^{site} \ll A_0$ then the difference between ΔG and ΔW_{min} can be large.[45]

2D Brownian Dynamics Simulation of a Simple System for the Isotropic and Two-Binding Site Model. In order to explore the effects of multiple binding poses on binding equilibria, we performed 2D Brownian dynamics simulations on two simple binding models. The system consists of two disk particles having either (1) orientationally isotropic binding or (2) two specific binding sites having different binding strengths. The latter case was designed to mimic multiple binding poses observed for GpA transmembrane helices. Details of the simple model simulations are described in Supporting Information, Appendix B. The 1D or 2D potential of mean force for the isotropic and two-site binding models were calculated from the equilibrium probability distribution functions obtained from unbiased simulations as well as the umbrella sampling simulations using a one-dimensional reaction coordinate. The minimum in

https://doi.org/10.1021/acs.jctc.2c00106
J. Chem. Theory Comput. 2022, 18, 3961−3971

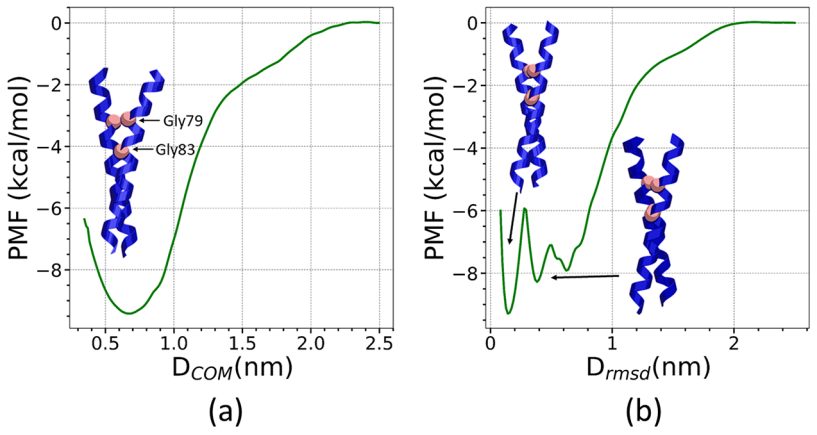


Figure 1. Potential of mean force for dimer association of GpA as a function of (a) distance between the COM of the TM helices (D_{com}) and (b) D_{rmsd} using the reference state defined by the interhelical distances present in the native GpA homodimer. The maximum value of the PMF curve was used to set the zero and guide the comparison.

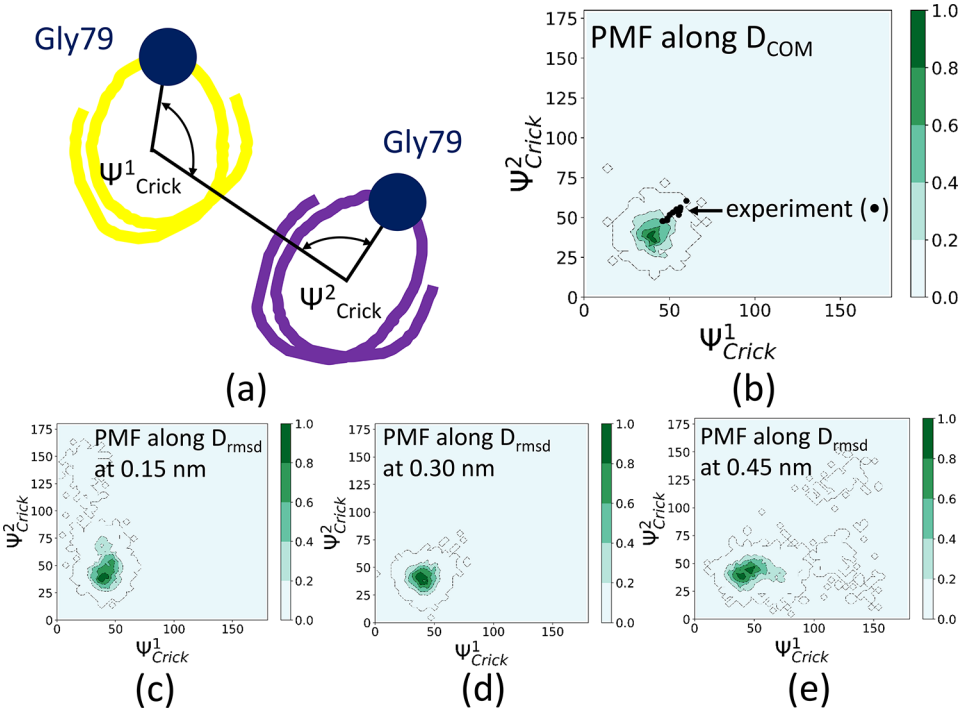


Figure 2. (a) Pictorial representation of the Crick angles of the GpA homodimer. Probability density projected on to the Crick angles of the GpA dimers obtained by analyzing the minima of the PMF along (b) D_{com} as well as D_{rmsd} at (c) 0.15 nm, (d) 0.30 nm, and (e) 0.45 nm.

the PMF (ΔW_{min}), the binding constant (K_a), and the free energy of association (ΔG) were calculated as described above.

RESULTS AND DISCUSSION

Interaction of TM protein domains in membrane plays an important role in cellular organization and signaling.[73,74] Structure and binding constituents of TM domain of GpA homodimer have been studied both experimentally and computationally.[11,19,22,38,49−52,75] Wild type and mutated GpA dimer interactions are well characterized at coarse-grained[20,21] and all-atom[19] resolution. The abundance of data available for this homodimer makes it an ideal system for studying the dimerization free energy landscape of protein−protein interactions and testing approaches for effective conformational sampling.

The effectiveness of US calculations used to define the PMF depends on a proper choice of sampling coordinates often defined in the form of collective variables (CVs). The proper choice of CV combined with enhanced sampling methods should facilitate the sampling of both native and non-native homodimer interactions and establish convergence. We demonstrate this concept by conducting simulations of two simple model systems for which exhaustive sampling can be achieved. The results for the simple model systems demonstrate that the estimation of the binding constant (K_a) and the free energy of association (ΔG) for a one-dimensional CV can be significantly impacted by the existence of multiple binding poses that contribute to the binding equilibrium. In this work, we have studied the free energy surface of GpA homodimer formation using the nonpolarizable MARTINI v2.2 model. A membrane bilayer consisting of approximately

https://doi.org/10.1021/acs.jctc.2c00106
J. Chem. Theory Comput. 2022, 18, 3961−3971

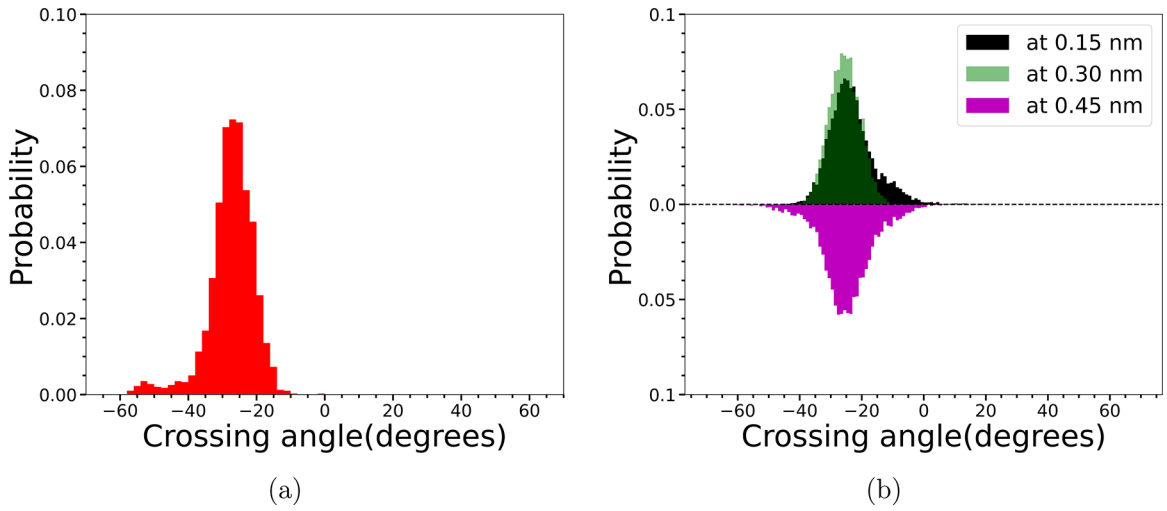


Figure 3. Probability densities of the crossing angles of the GpA homodimer obtained by analyzing the minima of the PMF of (a) D_{com} and (b) D_{rmsd} at 0.15, 0.30, and 0.45 nm represented by black, green, and magenta bars.

400 lipids was used for all simulations to avoid finite-size effects.[20]

Sampling of GpA Dimer Ensemble Depends on the Choice of Collective Variable. Three different CVs were used to calculate the dimerization free energy of GpA homodimer, and a comparison between these CVs was performed to understand the ability of the particular CV to distinguish and sample the native and non-native homodimer structures. First, we have evaluated the PMF of the homodimer interaction along the relative COM distance of the helices (D_{com}) in order to calculate the free energy of dimerization. Second, following Domański and co-workers,[11] we calculated the PMF of the homodimer interaction along the collective variable defined by the distance root-mean-square displacement (D_{rmsd}) from the native GpA homodimer reference structure (X^{ref}). The PMFs obtained from the simulations shown in Figure 1 agree with those published previously.[11,20,21]

The free energy landscape computed using D_{rmsd} as a collective variable has three minima as opposed to the single minimum observed with sampling along D_{com}. The minimum closest to the native structure on the D_{rmsd} axis has the lowest free energy of −9.3 kcal/mol. The second and third minima of the PMF are also characterized by low free energies. The two-dimensional projection of D_{rmsd} to D_{com} is shown in Supporting Information, Figure 1. As has been observed previously, sampling along D_{rmsd} allows for the identification of local free energy minima not resolved when sampling along D_{com}.[11] Determination of the binding constant from data acquired while sampling over D_{rmsd} involves calculating a 2D PMF along D_{com} and D_{rmsd} (Supporting Information, Figure 1a), projecting this data onto D_{com} (Supporting Information, Figure 1b), and integrating the 1D PMF over D_{com}.

Experimental and computational studies have been conducted to analyze the structure of the GpA homodimer. X-ray crystallography[48] and NMR studies[46] have shown that the GpA helices are stabilized by steric and C_α hydrogen bonding interactions facilitated by the glycine zipper (GXXXG) motif preferentially stabilizing a right-handed helix crossing angle of −20°. Previous computational studies have also predicted that the GpA homodimer forms a right-handed helical structure.[20,21]

We have analyzed the minimum of the PMFs obtained from sampling along both D_{rmsd} and D_{com} to characterize the homodimer structures. The Crick angle and crossing angle distributions for the homodimer are shown in Figures 2 and 3, respectively. Contact maps of the GpA homodimer obtained from the simulations are shown in Supporting Information, Figure 2. Both the Crick angle and contact map demonstrate that the sampling along D_{com} is restricted to the Gly-in homodimer structure. In contrast, sampling along D_{rmsd} captures the Gly-in structures but includes Gly-side structures as well (Supporting Information, Figure 3). When sampling along D_{com} or D_{rmsd}, we observe that GpA homodimer forms stable right-handed helical structures with a crossing angle of approximately −26°.

Enhanced Sampling over Two-Dimensional Collective Variable Space. To have a deeper understanding of the difference in sampling the ensemble along D_{com} and D_{rmsd}, we have projected the population density of the dimer onto the two-dimensional *xy*-plane. The center-of-mass of one helix was placed in the center of the *xy*-plane. In order to analyze the relative orientation of the second helix, the COM-to-Gly79 vector of the same helix was oriented along the positive *x*-axis. The result obtained from the analysis of the PMFs is shown in Figure 4.

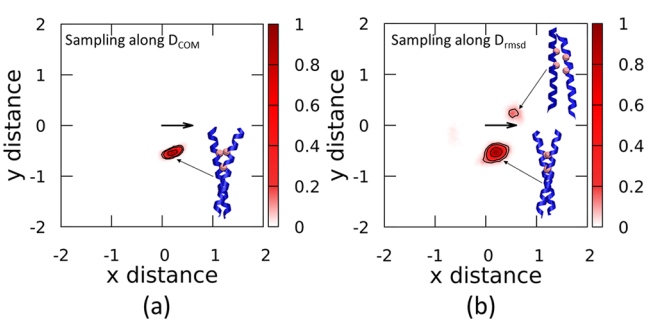


Figure 4. Population density of the dimer in the two-dimensional *xy*-plane (nm) obtained by analyzing the minima of the PMF of (a) D_{com} and (b) D_{rmsd}. The reference helix was centered on the *xy*-plane, and the COM-to-Gly79 vector was oriented along the positive *x*-axis, as represented by the black arrow.

https://doi.org/10.1021/acs.jctc.2c00106
J. Chem. Theory Comput. 2022, 18, 3961−3971

Finally, we evaluated the two-dimensional free energy landscape by considering the *x*- and *y*-projections of the relative COM distance of the GpA helices. The COM-to-Gly79 vector of one helix was restrained to the positive *x*-axis to sample all possible conformations of the GpA homodimer. The PMF obtained from the simulation is shown in Figure 5

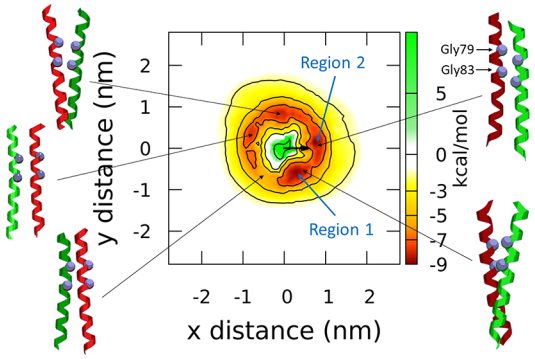


Figure 5. Two-dimensional potential of mean force for dimer association of GpA as a function of the *xy*-projection of the distance between the COM of the TM helices. The colorbar represents the value of PMF in kcal/mol. The COM-to-Gly79 vector of the helix, centered on the *xy*-plane, is represented by the black arrow. The maximum value of the PMF curve was used to set the zero and guide the comparison. GpA homodimer structures obtained from the last frame of the trajectory are shown where the restrained and free helices are represented in red and green, respectively.

and the corresponding probability distribution is shown in Supporting Information, Figure 4. The two-dimensional PMF presents a more diverse ensemble of homodimer structures including homodimers stabilized by native (glycine zipper interactions) or non-native motif interactions. Although the interaction with the GXXXG motif leads to the most stable homodimer conformation, with a free energy of −9.0 kcal/mol, other non-native structures are also thermodynamically relevant in the MARTINI v2.2 model.

The two-dimensional projection (Figure 4) shows restricted sampling with some advantage derived from sampling along D_{rmsd}. Comparing the distribution $p(x, y)$ derived from the two-dimensional PMF with the projection of the distribution derived from the one-dimensional umbrella sampling, we see that sampling along D_{com} or D_{rmsd} suffers from insufficient sampling of all possible native and non-native conformations of the homodimer. The incomplete sampling of the dimeric state ensemble leads to a biased probability distribution and biased free energy landscape. We have also calculated the free energy landscape of GpA homodimerization along D_{rmsd} using the replica exchange umbrella sampling (REUS) method, in which exchange between neighboring replica windows was enabled.[76] The results obtained from the simulations are shown in Supporting Information, Figure 5. The PMF along D_{rmsd} that was computed using the REUS method also suffers from insufficient sampling. As in the case of straightforward MD, umbrella sampling using replica exchange is overly biased toward the native state employed in the definition of D_{rmsd}. As such, it undersamples basins associated with competing homodimer structures. As a result, similar PMFs are obtained for sampling based on US and REUS.

Sampling obtained from the D_{com} or D_{rmsd} is largely restricted to regions 1 and 2 of the 2D PMF (Figure 5), near the native structure. Complete sampling of the relevant non-native homodimer structures was not obtained, resulting in a biased probability distribution. This point is demonstrated by a comparison of PMFs derived from data obtained from the US over the D_{com}, D_{rmsd}, or 2D collective variables, subsequently projected onto D_{com} (as shown in Supporting Information, Figure 6). Substantial differences in the three PMFs are observed, including shifts in the position and depth of the global minimum.

We have analyzed homodimer structures characteristic of regions 1 and 2 of the 2D PMF to compare the structural ensemble obtained from sampling along the collective variable D_{com} and D_{rmsd}. The angular distribution of the unrestrained helix with reference to the COM of the restrained helix is shown in Figure 6 and Supporting Information, Figure 7. The complementary crossing angle distribution of the GpA dimer is shown in Supporting Information, Figure 8. The results suggest that the unrestrained GpA helix interacts with the restrained helix via the glycine zipper motif forming a right-handed Gly-in helical structure in regions 1 and 2. This is also observed from the analysis of the free energy landscapes derived from the one-dimensional umbrella sampling simulations.

Computed Equilibrium Binding Constants Reflect the Extent of Sampling over Collective Variables. Several experimental and computational studies were performed to evaluate the free energy of dissociation of the GpA homodimer. Experimental values of dimerization free energies vary from 3.4 to 12.1 kcal/mol, depending on the methods and solvents used in the experiments.[53−58] Computational studies performed on GpA in various membrane bilayers reported a dimerization free energy varying from 3.0 to 11.5 kcal/mol. The free energy of the stable dimeric form of GpA compared

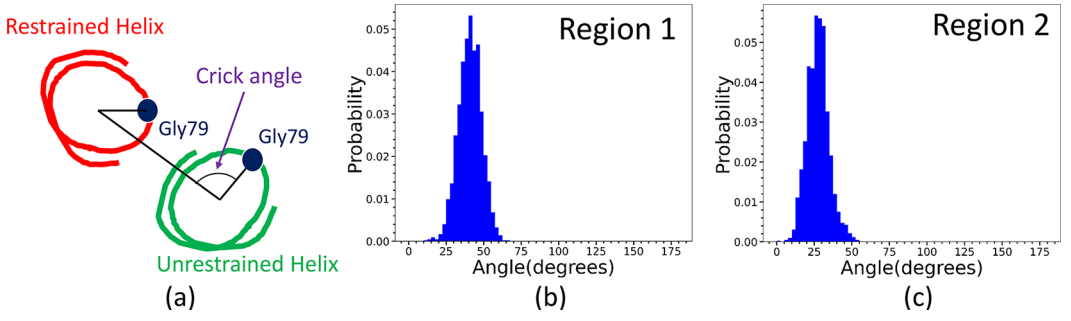


Figure 6. (a) Pictorial representation of the Crick angle distribution of the unrestrained GpA helix. Probability density of the Crick angle obtained by analyzing the two-dimensional PMF along *xy*-projection of the COM−COM distance between two GpA helices in (b) region 1 and (c) region 2.

https://doi.org/10.1021/acs.jctc.2c00106
J. Chem. Theory Comput. 2022, 18, 3961−3971

to the fully dissociated state is −9.4 kcal/mol and −9.7 kcal/mol obtained from the PMF along D_{com} and D_{rmsd} to D_{com} projection (Supporting Information, Figure 1). Direct comparison of the computational and experimental data can be obtained by integrating the PMF up to a cutoff distance differentiating the associated and dissociated states of the protein dimer. The standard reference area of 1 nm^2 was used to evaluate the association constant of the protein homodimer in the membrane.[21,38] The cutoff distance was chosen to be 2.3 nm, above which the $\Delta W(r)$ reached a plateau (Figure 1 and Supporting Information, Figure 1). The results are presented in Table 1. From the analysis of the two-dimensional PMF, we find that the ΔG value is −7.5 kcal/mol compared to the ΔW_{min} value of −9.0 kcal/mol.

Table 1. Dimerization Free Energy Obtained for the GpA Homodimer along Different Collective Variables Using the MARTINI v2.2 Model

PMF	ΔW_{min} (kcal/mol)	K_a	ΔG (kcal/mol)
D_{com}	−9.4	3.2×10^6	−9.1
D_{rmsd}	−9.7	4.3×10^6	−9.3
2D PMF	−9.0	2.2×10^5	−7.5

Model System for Exploring Contributions of Multiple Binding Poses to Binding Equilibrium. Results for the two model systems provide insight into the measured difference between ΔW_{min} and ΔG for GpA dimerization. Figure 7a shows the 2D PMF of the two-site binding model. Two binding poses are depicted as red-facing and yellow-facing disks. The red-facing configuration has a lower free energy of $-8k_BT$, which corresponds to the most stable native structure for the GpA homodimer. The yellow-facing configuration of $-4k_BT$ represents the less stable non-native GpA homodimer. If all possible dimeric states are fully sampled during the simulation, we obtain the well-averaged 1D PMF $\Delta W(r)$ along the disk separation distance r corresponding to D_{com} for GpA. The binding interactions are averaged around a given r, which contains both the local and global minima as well as the unbound states. We also describe two different 1D PMF curves, which are 1D radial slices through the global and local minima, respectively, denoted $\Delta W_g(r)$ and $\Delta W_l(r)$ (see Figure 7b) . If sampling is incomplete and restricted to only one state, the 1D PMF will be similar to $\Delta W_g(r)$ or $\Delta W_l(r)$ depending on which dimer conformation was sampled.

Table 2 shows the value of ΔW_{min}, K_a, and ΔG for the anisotropic two-site binding model with a fixed $\epsilon_l = 6k_BT$ and varying ϵ_g from $6k_BT$ to $10k_BT$. As ϵ_g increases, ΔW_{min} in the 2D PMF decreases, which means that the affinity toward global binding increases. In the case of the 2D PMF, where the exhaustive sampling is achieved, we observed noticeable differences between ΔG and ΔW_{min}, ranging from $1.5k_BT \sim 2.7k_BT$. This is expected from eq 4 as the absolute value of $\ln \frac{A_{site}}{A_0}$ is large when the area of binding (A_{site}) near the global minimum is much smaller than the standard area (A_0). The difference between ΔG and ΔW_{min} becomes larger as ϵ_g increases. As the binding becomes more specific, the area of the binding site is reduced as a result of the smaller thermal fluctuations in the dimer state. This behavior is consistent with results of GpA dimerization (Table 1) where ΔG and ΔW_{min} in the 2D PMF have a difference of 1.5 kcal/mol. This indicates that the existence of a preferential binding orientation is responsible for the observed difference between ΔG and ΔW_{min}. In the case of the isotropic binding without preferential binding orientation, no noticeable difference between ΔG and ΔW_{min} is observed (Supporting Information, Table 2).

Table 2 also shows the results for ΔW_{min} and ΔG derived from various 1D PMFs. For the well-averaged 1D PMF ($\Delta W(r)$), the binding free energy ΔG is consistent with that observed in the 2D PMF. On the other hand, $\Delta W_g(r)$ only samples dimer configurations at the global minimum and therefore overestimates the binding free energy. In contrast,

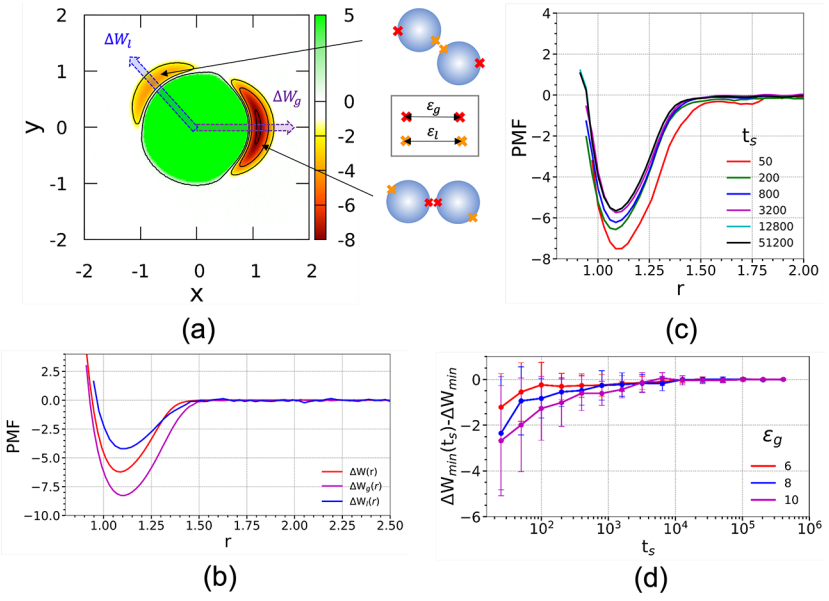


Figure 7. (a) 2D potential of mean force for $\epsilon_g = 10k_BT$ and $\epsilon_l = 6k_BT$ of two-site binding model. (b) 1D potential of mean force of two-site binding model: the well-averaged 1D PMF ($\Delta W(r)$) and the 1D PMF only samples configurations at the global minimum ($\Delta W_g(r)$) or the local minimum ($\Delta W_l(r)$). (c) Potential of mean force obtained using the US simulation starting from the native state with $\epsilon_g = 10k_BT$ and $\epsilon_l = 6k_BT$. (d) $\Delta W_{min}(t_s) - \Delta W_{min}$ values are plotted as a function of the sampling time (t_s).

https://doi.org/10.1021/acs.jctc.2c00106
J. Chem. Theory Comput. 2022, 18, 3961−3971

Table 2. Free Energies of Dimerization and Binding Constants Obtained by Analyzing the Anisotropic Two-Site Binding Model with $\epsilon_l = 6k_BT$

	2D PMF ($\Delta W(x, y)$)			1D PMF ($\Delta W(r)$)		1D PMF ($\Delta W_g(r)$)		1D PMF ($\Delta W_l(r)$)	
ϵ_g	ΔW_{min}	$K_{a,2D}$	ΔG	ΔW_{min}	ΔG	ΔW_{min}	ΔG	ΔW_{min}	ΔG
6	−4.20	12.6	−2.54	−3.03	−2.53	−4.20	−3.71	−4.19	−3.71
8	−6.22	43.6	−3.77	−4.41	−3.76	−6.22	−5.62	−4.17	−3.70
10	−8.21	248.7	−5.52	−6.23	−5.50	−8.21	−7.54	−4.18	−3.72

$\Delta W_l(r)$ depends on the binding strength in non-native dimer conformations and cannot accurately estimate the equilibrium ΔG.

GpA dimer in the Gly-in configuration rarely dissociates within the simulation time scale of a few microseconds, making it challenging to achieve enhanced sampling with conventional 1D US simulations. ΔG's calculated from the PMF along one-dimensional CV (D_{com} and D_{rmsd}) are −9.1 kcal/mol and −9.3 kcal/mol, respectively (Table 1). Both differ from the accurate ΔG = −7.5 kcal/mol derived from the 2D PMF (Table 1). This indicates that sampling performed along D_{com} or D_{rmsd} is insufficient to determine the accurate binding constant (Figure 2). The results of the 1D PMF of GpA are consistent with the case of $\Delta W_g(r)$ in the simple model case. The most stable native conformation is sampled, resulting in an overestimation of the binding free energy.

The umbrella sampling simulation of the simple model study provides further insights into how specific binding affects the sampling efficiency of binding equilibrium. We performed the US simulation starting from native dimer conformations only for a finite sampling time t_s. The US simulations at a short sampling time (t_s) overestimate the binding free energy, as it takes time to escape the free energy barrier of the native conformations and sample the non-native and unbound conformations (Figure 7c). Convergence to the equilibrium PMF was observed after $t_s \geq 3200$. The difference between the minimum in the PMF at each t_s ($\Delta W_{min}(t_s)$) and ΔW_{min} of the 1D PMF is shown in Figure 7d. As expected, the time required for the PMF to converge increases with increasing ϵ_g, and the time required for the convergence of $\Delta W_{min}(t_s)$ to ΔW_{min} is roughly comparable to the time required for a bound dimer to dissociate. Similarly, US simulations at a short t_s starting from the unbound conformation results in an underestimation of the binding free energy (Supporting Information, appendix B). The average residence time for a dimeric state was found to be 78, 460, and 3000 for $\epsilon_g = 6k_BT$, $8k_BT$, and $10k_BT$, respectively. These results demonstrate that a simulation time longer than the lifetime of the dimer is required to achieve an accurate estimation of the dimerization free energy. That calculation is challenging in the case of the GpA dimer which presents multiple stable binding poses.

Multiprotein Simulation Shows Both Native and Non-Native Interactions. To further characterize the contacts between GpA helices, we simulated eight GpA helices in a POPC bilayer consisting of 1600 lipids. A protein-to-lipid ratio of 1:200 was used to mimic the experimental conditions. Initially, all of the proteins were separated by equivalent distances to study the spontaneous formation of protein aggregates. The proteins were observed to form a single aggregate, as shown in a previous study (Figure 8).[20] As observed in the 2D PMF, a variety of binding poses, in addition to the Gly-in structure, were found to be thermodynamically relevant. To analyze the protein−protein interactions stabilizing the aggregates, in this unbiased simulation, we projected the population of the neighbor helices in the *xy*-plane. The COM-to-Gly79 vector of the reference helix was rotated along the positive *x*-axis and was placed at the center of the *xy*-plane. The resulting distribution is shown in Figure 8. In addition, we have performed 15 simulations of the GpA homodimer in a POPC bilayer starting from independent initial conformations. The proteins were initially well separated and subsequently allowed to spontaneously associate. A 5 μs production run was performed for each replica without restraint. The population density distribution obtained from the simulations is shown in Supporting Information, Figure 9. Simulations initiated from the dimer state explore a restricted non-native homodimer ensemble. On the other hand, those initiated from the monomeric state present a variety of non-native structures. Results obtained from the *xy*-plane projection of the relative COM−COM distance show the existence of non-native interactions along with the expected native interactions. Protein density in the unbiased simulations also follows the trend predicted by the 2D PMF. As has been observed previously, the MARTINI model lacks the specificity of binding in the experimentally observed Gly-in native state where the native state is stabilized by interactions facilitated by the GXXXG motif. As a result, helices in the many protein simulation readily form amorphous aggregates.

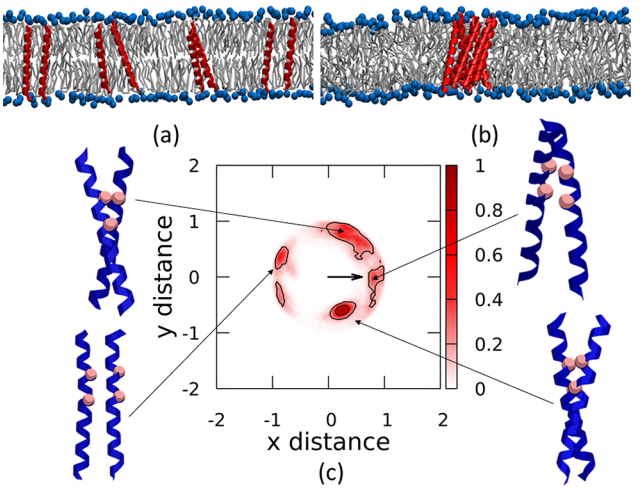


Figure 8. Side view of the (a) the initial and (b) equilibrated membrane bilayer. (c) Population density of the GpA helices in the two-dimensional *xy*-plane obtained from the many-protein simulation. The reference helix was centered on the *xy*-plane, and the COM-to-Gly79 vector was oriented along the positive *x*-axis, as represented by the black arrow.

CONCLUSIONS

It is well appreciated that the efficiency of sampling conformations, defining equilibrium constants for protein−protein association in aqueous solution or membrane, using

https://doi.org/10.1021/acs.jctc.2c00106
J. Chem. Theory Comput. 2022, 18, 3961−3971

US simulations depends on the choice of collective coordinates. For homodimer formation of the well-studied TM protein GpA, we demonstrate that sampling along the previously proposed one-dimensional collective variables D_{com} and D_{rmsd} is less effective than sampling over a two-dimensional collective variable space defined by the *xy*-projection of the relative COM distance. Biased sampling over two-dimensional collective variable space is shown to effectively explore the full range of relevant orientations of the TM helices. In contrast, one-dimensional CVs show limitations in sampling homodimer conformations needed to accurately characterize the dimer ensemble and accurately compute the association equilibrium constant.

To mimic the dimerization of transmembrane helices with multiple binding hotspots as observed for the GpA homodimer (Figure 5), we constructed a 2D binding model in which each model protein has two distinct interaction sites (Supporting Information, Appendix B). Our simple model study conceptually demonstrates how the use of a one-dimensional CV impacts the assessment of accurate PMFs and equilibrium constants. When the dimerization equilibrium is characterized by multiple specific binding poses, calculation of the association constant fails unless the complete sampling is achieved. In that case, in the absence of biasing schemes designed to effectively explore all relevant binding poses, the computed PMF depends on the sampling time. Considering results for the GpA homodimerization, sufficient time required to explore multiple binding conformations is challenging to achieve in high viscosity lipid membrane. Dimerization free energies obtained from 1D PMFs derived from sampling along D_{com} and D_{rmsd} differ from the free energy of dimerization derived from sampling over a two-dimensional CV space. Sampling along 1D CVs overestimates the association free energy due to oversampling the global minimum. In contrast, the 2D PMF includes significant contributions of non-native structures to the homodimer ensemble. The ΔG value obtained from the 2D PMF is in line with the ΔG value used in parametrizing the MARTINI v3 force field.

In previous work, we proposed one way to circumvent the excessive aggregation of TM proteins in lipid bilayers using the MARTINI v2.2 model by scaling the protein−lipid nonbonded interaction. An upscaling factor of 1.04 or 1.045 was shown to provide the best fit to experimental results for several TM proteins studied. In the case of the GpA homodimer, a free energy of dimerization value of 3.6 kcal/mol was used. Experimental values of free energies of dimerization of GpA vary from 3.4 to 12.1 kcal/mol.[53−58] A recent study of Souza et al. has identified a best reference free energy of dimerization of 6.6 kcal/mol to calibrate the MARTINI v3 protein model.[72] A study leading to a proposed reparameterization of the all-atom CHARMM36 force field, Domański et al. obtained a free energy of dimerization of 3.0−3.8 kcal/mol.[19] Considering the reference used in the MARTINI v3 model, an upscaling factor of 1.03 can be used in the case of GpA.

An efficiency not considered in this work is the use of restraining potentials to further reduce the range of conformational space that must be sampled to accurately characterize the association equilibrium. Approaches effectively employed in the past have explored the use of restraints based on Euler angles[35] or employing a set of images using the string method.[44] The use of either of these methods requires *a priori* knowledge of all relevant binding poses contributing to the binding equilibrium. However, for the case of GpA homodimerization, those relevant binding poses were only known once umbrella sampling over the two lateral dimensions was performed. Nevertheless, those alternative methods could be employed *post facto* to explore the relative efficiencies of those approaches.

Although the ΔG value derived from the 2D PMF is in close agreement with the recently established experimental standard ΔG value for GpA homodimerization, excessive nonspecific protein aggregation remains a shortcoming of the MARTINI v2.2 model. Probability densities describing the relative orientation of GpA helices obtained from the unbiased simulations match well with the probability densities observed in the 2D PMF. This demonstrates the thermodynamic relevance of non-native conformations to the overall dimer conformational ensemble. We conjecture that the uniform parametrization of protein beads in the MARTINI v2.2 model results in the less specific GpA homodimer interactions. The results obtained from the 2D PMF can inform future reparametrizations of the MARTINI model to provide the best agreement with experimental results.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jctc.2c00106.

Appendix A: (Table 1) Dimerization free energy of the GpA homodimer obtained from the previous studies; (Figure 1) post analysis of the PMF obtained along D_{rmsd}; (Figure 2) contact map of the GpA homodimer; (Figure 3) pictorial representation of Gly-in, Gly-side, and Gly-out structures; (Figure 4) unbiased probability density obtained from the 2D US; (Figure 5) post analysis of the PMF obtained along D_{rmsd} using REUS; (Figure 6) projection of the PMFs on D_{com}; (Figure 7) crick angles and (Figure 8) crossing angle distributions obtained from the 2D US; (Figure 9) density of GpA helices obtained from unbiased simulation; Appendix B: discussion of model systems (PDF)

■ AUTHOR INFORMATION

Corresponding Author

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

Authors

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

Seulki Kwon − *Department of Chemistry, Boston University, Boston, Massachusetts 02215, United States*

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jctc.2c00106

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

The authors gratefully acknowledge the generous support of the National Institutes of Health (Grant R01 GM107703), National Science Foundation (Grant CHE1900416), and the high-performance computing resources of the Boston University Shared Computing Cluster (SCC). We thank Conor B.

Abraham, George A. Pantelopulos, and Qiang Cui for scientific discussions and Benoît Roux for valuable comments and suggestions.

REFERENCES

(1) Popot, J.-L.; Engelman, D. M. Helical membrane protein folding, stability, and evolution. *Annu. Rev. Biochem.* **2000**, *69*, 881−922.
(2) Gomes, I.; Ayoub, M. A.; Fujita, W.; Jaeger, W. C.; Pfleger, K. D.; Devi, L. A. G protein−coupled receptor heteromers. *Annual review of pharmacology and toxicology* **2016**, *56*, 403−425.
(3) Haass, C.; Selkoe, D. J. Soluble protein oligomers in neurodegeneration: lessons from the Alzheimer's amyloid β-peptide. *Nat. Rev. Mol. Cell Biol.* **2007**, *8*, 101−112.
(4) Irvine, G. B.; El-Agnaf, O. M.; Shankar, G. M.; Walsh, D. M. Protein aggregation in the brain: the molecular basis for Alzheimer's and Parkinson's diseases. *Mol. Med.* **2008**, *14*, 451−464.
(5) 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.
(6) Dominguez, L.; Foster, L.; Meredith, S. C.; Straub, J. E.; Thirumalai, D. Structural heterogeneity in transmembrane amyloid precursor protein homodimer is a consequence of environmental selection. *J. Am. Chem. Soc.* **2014**, *136*, 9619−9626.
(7) Dominguez, L.; Meredith, S. C.; Straub, J. E.; Thirumalai, D. Transmembrane fragment structures of amyloid precursor protein depend on membrane surface curvature. *J. Am. Chem. Soc.* **2014**, *136*, 854−857.
(8) Filippov, A.; Orädd, G.; Lindblom, G. The effect of cholesterol on the lateral diffusion of phospholipids in oriented bilayers. *Biophys. J.* **2003**, *84*, 3079−3086.
(9) Javanainen, M.; Hammaren, H.; Monticelli, L.; Jeon, J.-H.; Miettinen, M. S.; Martinez-Seara, H.; Metzler, R.; Vattulainen, I. Anomalous and normal diffusion of proteins and lipids in crowded lipid membranes. *Faraday Discuss.* **2013**, *161*, 397−417.
(10) 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.
(11) Domański, J.; Hedger, G.; Best, R. B.; Stansfeld, P. J.; Sansom, M. S. Convergence and sampling in determining free energy landscapes for membrane protein association. *J. Phys. Chem. B* **2017**, *121*, 3364−3375.
(12) Chipot, C. Frontiers in free-energy calculations of biological systems. *Wiley Interdisciplinary Reviews: Computational Molecular Science* **2014**, *4*, 71−89.
(13) Gumbart, J. C.; Roux, B.; Chipot, C. Standard binding free energies from computer simulations: What is the best strategy? *J. Chem. Theory Comput.* **2013**, *9*, 794−802.
(14) Roux, B. The calculation of the potential of mean force using computer simulations. *Comput. Phys. Commun.* **1995**, *91*, 275−282.
(15) Zhu, F.; Hummer, G. Convergence and error estimation in free energy calculations using the weighted histogram analysis method. *J. Comput. Chem.* **2012**, *33*, 453−465.
(16) Leone, V.; Marinelli, F.; Carloni, P.; Parrinello, M. Targeting biomolecular flexibility with metadynamics. *Curr. Opin. Struct. Biol.* **2010**, *20*, 148−154.
(17) Comer, J.; Gumbart, J. C.; Hénin, J.; Lelièvre, T.; Pohorille, A.; Chipot, C. The adaptive biasing force method: Everything you always wanted to know but were afraid to ask. *J. Phys. Chem. B* **2015**, *119*, 1129−1151.
(18) Hénin, J.; Lelièvre, T.; Shirts, M. R.; Valsson, O.; Delemotte, L. Enhanced sampling methods for molecular dynamics simulations.arXiv preprint, *arXiv:2202.04164*, **2022**.
(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. Theory Comput.* **2018**, *14*, 1706−1715.
(20) Majumder, A.; Straub, J. E. Addressing the Excessive Aggregation of Membrane Proteins in the MARTINI Model. *J. Chem. Theory Comput.* **2021**, *17*, 2513−2521.
(21) Sengupta, D.; Marrink, S. J. Lipid-mediated interactions tune the association of glycophorin A helix and its disruptive mutants in membranes. *Phys. Chem. Chem. Phys.* **2010**, *12*, 12987−12996.
(22) Chavent, M.; Chetwynd, A. P.; Stansfeld, P. J.; Sansom, M. S. Dimerization of the EphA1 receptor tyrosine kinase transmembrane domain: insights into the mechanism of receptor activation. *Biochemistry* **2014**, *53*, 6641−6652.
(23) Souaille, M.; Roux, B. Extension to the weighted histogram analysis method: combining umbrella sampling with free energy calculations. *Comput. Phys. Commun.* **2001**, *135*, 40−57.
(24) Hub, J. S.; De Groot, B. L.; Van Der Spoel, D. g_wham-A Free Weighted Histogram Analysis Implementation Including Robust Error and Autocorrelation Estimates. *J. Chem. Theory Comput.* **2010**, *6*, 3713−3720.
(25) Cruz, J.; Wickstrom, L.; Yang, D.; Gallicchio, E.; Deng, N. Combining alchemical transformation with a physical pathway to accelerate absolute binding free energy calculations of charged ligands to enclosed binding sites. *J. Chem. Theory Comput.* **2020**, *16*, 2803−2813.
(26) Wu, J. Z.; Azimi, S.; Khuttan, S.; Deng, N.; Gallicchio, E. Alchemical Transfer Approach to Absolute Binding Free Energy Estimation. *J. Chem. Theory Comput.* **2021**, *17*, 3309−3319.
(27) Hermans, J.; Shankar, S. The free energy of xenon binding to myoglobin from molecular dynamics simulation. *Isr. J. Chem.* **1986**, *27*, 225−227.
(28) Miyamoto, S.; Kollman, P. A. Absolute and relative binding free energy calculations of the interaction of biotin and its analogs with streptavidin using molecular dynamics/free energy perturbation approaches. *Proteins: Struct., Funct., Bioinf.* **1993**, *16*, 226−245.
(29) Roux, B.; Nina, M.; Pomes, R.; Smith, J. C. Thermodynamic stability of water molecules in the bacteriorhodopsin proton channel: a molecular dynamics free energy perturbation study. *Biophysical journal* **1996**, *71*, 670−681.
(30) Gilson, M. K.; Given, J. A.; Bush, B. L.; McCammon, J. A. The statistical-thermodynamic basis for computation of binding affinities: a critical review. *Biophysical journal* **1997**, *72*, 1047−1069.
(31) Helms, V.; Wade, R. C. Hydration energy landscape of the active site cavity in cytochrome P450cam. *Proteins: Struct., Funct., Bioinf.* **1998**, *32*, 381−396.
(32) Luo, H.; Sharp, K. On the calculation of absolute macromolecular binding free energies. *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99*, 10399−10404.
(33) Boresch, S.; Tettinger, F.; Leitgeb, M.; Karplus, M. Absolute binding free energies: a quantitative approach for their calculation. *J. Phys. Chem. B* **2003**, *107*, 9535−9551.
(34) Mihailescu, M.; Gilson, M. K. On the theory of noncovalent binding. *Biophysical journal* **2004**, *87*, 23−36.
(35) Woo, H.-J.; Roux, B. Calculation of absolute protein−ligand binding free energy from computer simulations. *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *102*, 6825−6830.
(36) Deng, Y.; Roux, B. Calculation of standard binding free energies: Aromatic molecules in the T4 lysozyme L99A mutant. *J. Chem. Theory Comput.* **2006**, *2*, 1255−1273.
(37) Mobley, D. L.; Chodera, J. D.; Dill, K. A. On the use of orientational restraints and symmetry corrections in alchemical free energy calculations. *J. Chem. Phys.* **2006**, *125*, 084902.
(38) Hénin, J.; Pohorille, A.; Chipot, C. Insights into the recognition and association of transmembrane α-helices. The free energy of α-helix dimerization in glycophorin A. *J. Am. Chem. Soc.* **2005**, *127*, 8478−8484.
(39) Hénin, J.; Pohorille, A.; Chipot, C. Insights into the recognition and association of transmembrane α-helices. the free energy of α-helix dimerization in glycophorin A. *J. Am. Chem. Soc.* **2010**, *132*, 9510.
(40) Corey, R. A.; Vickery, O. N.; Sansom, M. S.; Stansfeld, P. J. Insights into membrane protein−lipid interactions from free energy calculations. *J. Chem. Theory Comput.* **2019**, *15*, 5727−5736.

https://doi.org/10.1021/acs.jctc.2c00106
J. Chem. Theory Comput. 2022, 18, 3961−3971

(41) Duboué-Dijon, E.; Hénin, J. Building intuition for binding free energy calculations: Bound state definition, restraints, and symmetry. *J. Chem. Phys.* **2021**, *154*, 204101.
(42) Mey, A. S.J.S.; Allen, B. K.; Bruce Macdonald, H. E.; Chodera, J. D.; Hahn, D. F.; Kuhn, M.; Michel, J.; Mobley, D. L.; Naden, L. N.; Prasad, S.; Rizzi, A.; Scheen, J.; Shirts, M. R.; Tresadern, G.; Xu, H. Best practices for alchemical free energy calculations. *LiveCoMS* 2020, *2*, 1, .
(43) Gumbart, J. C.; Roux, B.; Chipot, C. Efficient determination of protein−protein standard binding free energies from first principles. *J. Chem. Theory Comput.* **2013**, *9*, 3789−3798.
(44) Suh, D.; Jo, S.; Jiang, W.; Chipot, C.; Roux, B. String method for protein−protein binding free-energy calculations. *J. Chem. Theory Comput.* **2019**, *15*, 5829−5844.
(45) Roux, B. *Computational Modeling and Simulations of Biomolecular Systems*; World Scientific, 2021.
(46) MacKenzie, K. R.; Prestegard, J. H.; Engelman, D. M. A transmembrane helix dimer: structure and implications. *Science* **1997**, *276*, 131−133.
(47) Smith, S. O.; Eilers, M.; Song, D.; Crocker, E.; Ying, W.; Groesbeek, M.; Metz, G.; Ziliox, M.; Aimoto, S. Implications of threonine hydrogen bonding in the glycophorin A transmembrane helix dimer. *Biophys. J.* **2002**, *82*, 2476−2486.
(48) 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.
(49) Brosig, B.; Langosch, D. The dimerization motif of the glycophorin A transmembrane segment in membranes: importance of glycine residues. *Protein Sci.* **1998**, *7*, 1052−1056.
(50) Langosch, D.; Brosig, B.; Kolmar, H.; Fritz, H.-J. Dimerisation of the glycophorin A transmembrane segment in membranes probed with the ToxR transcription activator. *J. Mol. Biol.* **1996**, *263*, 525−530.
(51) Doura, A. K.; Kobus, F. J.; Dubrovsky, L.; Hibbard, E.; Fleming, K. G. Sequence context modulates the stability of a GxxxG-mediated transmembrane helix−helix dimer. *Journal of molecular biology* **2004**, *341*, 991−998.
(52) Lemmon, M. A.; Flanagan, J. M.; Treutlein, H. R.; Zhang, J.; Engelman, D. M. Sequence specificity in the dimerization of transmembrane. alpha.-helixes. *Biochemistry* **1992**, *31*, 12719−12725.
(53) Fleming, K. G.; Ackerman, A. L.; Engelman, D. M. The effect of point mutations on the free energy of transmembrane α-helix dimerization. *J. Mol. Biol.* **1997**, *272*, 266−275.
(54) Fleming, K. G. Standardizing the free energy change of transmembrane helix−helix interactions. *J. Mol. Biol.* **2002**, *323*, 563−571.
(55) Nash, A.; Notman, R.; Dixon, A. M. De novo design of transmembrane helix−helix interactions and measurement of stability in a biological membrane. *Biochimica et Biophysica Acta (BBA)-Biomembranes* **2015**, *1848*, 1248−1257.
(56) Hong, H.; Blois, T. M.; Cao, Z.; Bowie, J. U. Method to measure strong protein−protein interactions in lipid bilayers using a steric trap. *Proc. Natl. Acad. Sci. U. S. A.* **2010**, *107*, 19802−19807.
(57) Sarabipour, S.; Hristova, K. Glycophorin A transmembrane domain dimerization in plasma membrane vesicles derived from CHO, HEK 293T, and A431 cells. *Biochimica et Biophysica Acta (BBA)-Biomembranes* **2013**, *1828*, 1829−1833.
(58) Chen, L.; Novicky, L.; Merzlyakov, M.; Hristov, T.; Hristova, K. Measuring the energetics of membrane protein dimerization in mammalian membranes. *J. Am. Chem. Soc.* **2010**, *132*, 3628−3635.
(59) Janosi, L.; Prakash, A.; Doxastakis, M. Lipid-modulated sequence-specific association of glycophorin A in membranes. *Biophys. J.* **2010**, *99*, 284−292.
(60) 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.
(61) de Jong, D. H.; Singh, G.; Bennett, W. D.; Arnarez, C.; Wassenaar, T. A.; Schafer, L. V.; Periole, X.; Tieleman, D. P.; Marrink, S. J. Improved parameters for the martini coarse-grained protein force field. *J. Chem. Theory Comput.* **2013**, *9*, 687−697.
(62) 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.
(63) Pantelopulos, G. A.; Straub, J. E. Regimes of complex lipid bilayer phases induced by cholesterol concentration in MD simulation. *Biophys. J.* **2018**, *115*, 2167−2178.
(64) Javanainen, M.; Martinez-Seara, H.; Vattulainen, I. Excessive aggregation of membrane proteins in the Martini model. *PLoS One* **2017**, *12*, No. e0187936.
(65) Wassenaar, T. A.; Ingolfsson, H. I.; Bockmann, R. A.; Tieleman, D. P.; Marrink, S. J. Computational lipidomics with insane: a versatile tool for generating custom membranes for molecular simulations. *J. Chem. Theory Comput.* **2015**, *11*, 2144−2155.
(66) 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.
(67) Grossfield, A. *WHAM: an implementation of the weighted histogram analysis method*. http://membrane.urmc.rochester.edu/wordpress/?page_id=126 (accessed 2021−03−09).
(68) Abraham, M. J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J. C.; Hess, B.; Lindahl, E. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **2015**, *1*, 19−25.
(69) Tribello, G. A.; Bonomi, M.; Branduardi, D.; Camilloni, C.; Bussi, G. PLUMED 2: New feathers for an old bird. *Comput. Phys. Commun.* **2014**, *185*, 604−613.
(70) Trzesniak, D.; Kunz, A.-P. E.; van Gunsteren, W. F. A comparison of methods to compute the potential of mean force. *ChemPhysChem* **2007**, *8*, 162−169.
(71) Khavrutskii, I. V.; Dzubiella, J.; McCammon, J. A. Computing accurate potentials of mean force in electrolyte solutions with the generalized gradient-augmented harmonic Fourier beads method. *J. Chem. Phys.* **2008**, *128*, 044106.
(72) Souza, P. C.; Alessandri, R.; Barnoud, J.; Thallmair, S.; Faustino, I.; Grünewald, F.; Patmanidis, I.; Abdizadeh, H.; Bruininks, B. M.; Wassenaar, T. A.; Kroon, P. C.; Melcr, J.; Nieto, V.; Corradi, V.; Khan, H. M.; Domański, J.; Javanainen, M.; Martinez-Seara, H.; Reuter, N.; Best, R. B.; Vattulainen, I.; Monticelli, L.; Periole, X.; Tieleman, D. P.; de Vries, A. H.; Marrink, S. J. Martini 3: a general purpose force field for coarse-grained molecular dynamics. *Nat. Methods* **2021**, *18*, 382−388.
(73) Popot, J.-L.; Engelman, D. M. Membrane protein folding and oligomerization: the two-stage model. *Biochemistry* **1990**, *29*, 4031−4037.
(74) 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.; Groves, J. T.; Kuriyan, J. Conformational coupling across the plasma membrane in activation of the EGF receptor. *Cell* **2013**, *152*, 543−556.
(75) Domański, J.; Sansom, M. S.; Stansfeld, P. J.; Best, R. B. Atomistic mechanism of transmembrane helix association. *PLoS computational biology* **2020**, *16*, No. e1007919.
(76) Sugita, Y.; Kitao, A.; Okamoto, Y. Multidimensional replica-exchange method for free-energy calculations. *J. Chem. Phys.* **2000**, *113*, 6042.

https://doi.org/10.1021/acs.jctc.2c00106
J. Chem. Theory Comput. 2022, 18, 3961−3971