

Efficient Sampling of Free Energy Landscapes for the Calculation of Protein–Protein Binding Affinities in Membranes

Published as part of *The Journal of Physical Chemistry B* special issue “At the Cutting Edge of Theoretical and Computational Biophysics”.

Ayan Majumder and John E. Straub*



Cite This: *J. Phys. Chem. B* 2025, 129, 9031–9042



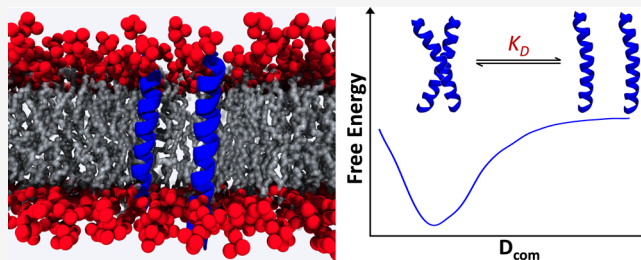
Read Online

ACCESS |

Metrics & More

Article Recommendations

ABSTRACT: The accurate simulation of realistic biomembranes is a long-term goal in the field of membrane biophysics. Efforts to simulate increasingly complex lipid bilayers, consisting of multiple lipid types and proteins, have been hindered by the shortcomings of current force fields, both coarse-grained and all-atom, in the modeling of protein–protein and protein–lipid interactions. Due to the fundamental importance of protein dimerization to cellular signaling and protein trafficking, the study of protein–protein association and the related dimerization free energies has received significant attention in both simulations and experiments. Detailed comparisons of simulation results with NMR, crystallography, and FRET studies served as a test of the accuracy of the simulation methods and provided insights into the underlying structural distributions and thermodynamic driving forces defining the interactions. These comparisons have led to the conclusion that existing state-of-the-art simulation methods have failed to effectively sample the equilibrium between associated and dissociated states, leading to inaccurate estimates of binding constants and misrepresentation of the associated structural ensembles. Here, we discuss the drawbacks of previously used protocols and review our systematic development of effective computational methods for enhanced sampling simulations that exhaustively sample the native and non-native dimer conformations and provide precise estimates of the associated equilibrium binding constants. We conclude by identifying the most important current challenges to the field that must be met in closing the gap between simulation and experiment in the study of protein–protein association in the membrane.



INTRODUCTION

Biomembranes play an essential role in defining cellular organization and function.^{1–5} A typical biomembrane is composed of a variety of lipid molecules, asymmetrically distributed between membrane leaflets, and contains a high concentration of proteins that occupy 15–30% of the membrane area.^{6–9} In such high concentrations, protein–protein interactions play a crucial role. Lipid phase separation and specific protein–protein and protein–lipid complexes result in a highly heterogeneous membrane structure with distinct nanodomains, supporting cellular signaling and protein trafficking.^{10–13} Membrane protein interactions also have considerable biomedical relevance, with G-protein-coupled receptors being an obvious example, accounting for 30–40% of drug targets.^{14,15} The interaction of membrane proteins has also been found to play an important role in amyloid protein biogenesis, crucial to Alzheimer’s disease.¹⁶ As such, the accurate quantitative assessment of protein–protein association in membranes, including characterization of the dimer ensembles and estimation of the related equilibrium association constants, is an essential goal for cellular biophysics.

While X-ray, NMR, and cryo-EM have contributed to our understanding of membrane protein structure, membrane protein complexes can exhibit significant structural fluctuation and depend sensitively on the local membrane environment.¹⁷ Changes in local lipid composition, membrane thickness, the presence of cholesterol, and pH can induce significant structural transformations. Moreover, the structures obtained from experiments may represent one of the many possible functional forms of a membrane protein complex.

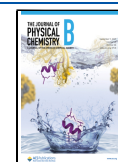
Informatics and physics-based approaches are routinely used to model and simulate membrane proteins in complex membrane environments.^{18–20} Recent advances in CPUs/GPUs^{21,22} and MD-specialized supercomputers,²³ as well as

Received: June 20, 2025

Revised: August 14, 2025

Accepted: August 18, 2025

Published: September 2, 2025



multiscale simulation approaches, combining all-atom and coarse-grained models,²⁴ have enabled the simulation of increasingly complex environments. However, simulation of protein–protein interactions in a complex biomembrane environment remains challenging.

The systematic study of membrane protein association using computation is typically performed using two approaches. (1) The self-assembly approach has been used to investigate spontaneous oligomerization of proteins by running multiple independent trajectories or using directed simulations to elucidate the thermodynamic driving forces that favor protein association by computing the potential of mean force (PMF). However, effective sampling of protein–protein interactions that define the free energy of dimerization remains challenging. (2) In most cases, umbrella sampling,²⁵ adaptive biasing force method,^{26,27} replica-exchange molecular dynamics,^{28–30} generalized-ensemble algorithms,³¹ or metadynamics^{32,33} methods are used to calculate free energy surfaces for membrane protein association.

The calculation of the binding free energy for transmembrane protein association is of special interest due to its biological significance. It also provides an excellent benchmark for validating computational models against association constant data obtained from experimental Förster resonance energy transfer (FRET) assays. However, significant challenges remain for existing enhanced sampling protocols designed to explore thermodynamically relevant dimer structural ensembles for comparison with experiments. (1) The slow translational and rotational diffusion³⁴ of membrane proteins leads to slow convergence in sampling all thermodynamically relevant structures characterizing the dimerization equilibrium. (2) Enhanced sampling methods based on increasing temperature or softening interactions can lead to instability of the membrane. (3) While FRET experiments can report on the absolute proximity of two TM proteins, the relative orientation of the TM proteins is uncertain. As such, it is unclear if the experiment reports all dimer structures observed in simulation or a subset of structures. (4) Finally, it is known that the labeled proteins, modified to include donor and acceptor FRET pairs, can exhibit different affinities for lipid–protein interactions compared with native proteins.

In this perspective, we survey current enhanced sampling protocols used to study transmembrane protein dimerization. We do not review the history of this field but rather focus on the most recent developments of efficient enhanced sampling protocols to exhaustively sample important dimer structures and explore the convergence of association free energy calculations. We conclude by highlighting the most important questions for future studies.

METHODS BASED ON A SINGLE COLLECTIVE VARIABLE

Due to the rugged nature of biomolecular energy landscapes, where multiple thermodynamically relevant substates may be separated by high-energy barriers, the efficacy of direct molecular dynamics is limited, and enhanced sampling is needed. The efficiency of enhanced sampling in the study of rare events typically relies on the proper choice of the reaction coordinate. In the study of the dimerization of transmembrane proteins, it is important to choose collective variables (CVs) that connect the monomeric and dimeric states of the dimer and facilitate the sampling of both native and non-native dimer configurations to establish convergence. A converged proba-

bility density along the CVs was then used to calculate the binding constant of the TM dimer.

Glycophorin A (GpA) transmembrane protein homodimer has been extensively studied using various enhanced sampling methods.^{35–41} The most obvious choice of a single CV for sampling TM protein dimerization is the center-of-mass distance between two helices (D_{com}). Henin et al. studied the dimerization of GpA in a dodecane medium using D_{com} as a reaction coordinate with the adaptive bias force method.⁴² They reported the free energy of dimerization for GpA to be -11.5 kcal/mol using the all-atom CHARMM27 force field.

An alternative single CV is the distance-root-mean-squared displacement (D_{rmsd}) defined as the change in dimer structure (X) from a reference dimer structure (X^{ref}) and has also been widely used in the study of TM helix dimerization.

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

where $d(x_i, x_j)$ is the distance between atoms i and j in dimer configuration X . In practice, an experimental structure obtained from the PDB is used as the reference configuration X^{ref} , and a D_{rmsd} value closer to zero indicates that the dimer configuration (X) is more similar to X^{ref} .

Best and co-workers investigated GpA homodimerization in a POPC bilayer using the CHARMM36 force field.³⁷ They performed replica-exchange umbrella sampling (REUS) along D_{rmsd} to study the dimerization equilibrium. The PMF along D_{rmsd} showed no stable dimeric state of GpA. They further proposed a reparametrization of the CHARMM36 force field by reducing the protein–lipid Lennard-Jones (LJ) interactions to match the experimental binding constant results.

Chipot and co-workers employed a so-called geometric route to simulate the dimerization of GpA and reported the dimerization free energy to be -10.7 kcal/mol in a POPC bilayer using the CHARMM36m force field.³⁶ In this approach, the protein helices were progressively separated by restraining their orientational and conformational degrees of freedom to accelerate convergence of the PMF. However, implementing this protocol can be challenging for systems with multiple thermodynamically relevant binding poses in the dimer conformational ensemble.

The coarse-grained MARTINI model^{44,45} has also been used to study GpA homodimerization. Marrink and Sengupta performed umbrella sampling (US) along D_{com} and reported a dimerization free energy of GpA to be -9.1 kcal/mol in a DPPC bilayer.⁴⁶ They also demonstrated the disruptive effect of the T87F mutant and triple mutants of the glycine zipper (GXXXG) motif by comparing the computed PMF for the dimerization of the mutants with that of wild-type GpA. Sansom and co-workers studied GpA homodimerization in a POPC bilayer by performing REUS along D_{rmsd} using the MARTINI force field.⁴⁷ They also calculated the association free energy of the EphA1 dimer by performing US along the center-of-mass distance between the glycine zipper motifs using the MARTINI force field.⁴⁸ The study showed that EphA1 forms a dimer stabilized by interactions mediated by glycine zipper motifs. Tieleman and co-workers studied the association of the WALP23 homodimer in DMPC, DPPC, and DOPC bilayers, computing the PMF for dimer association along D_{com} using US. They reported that the association of WALP23 has no free

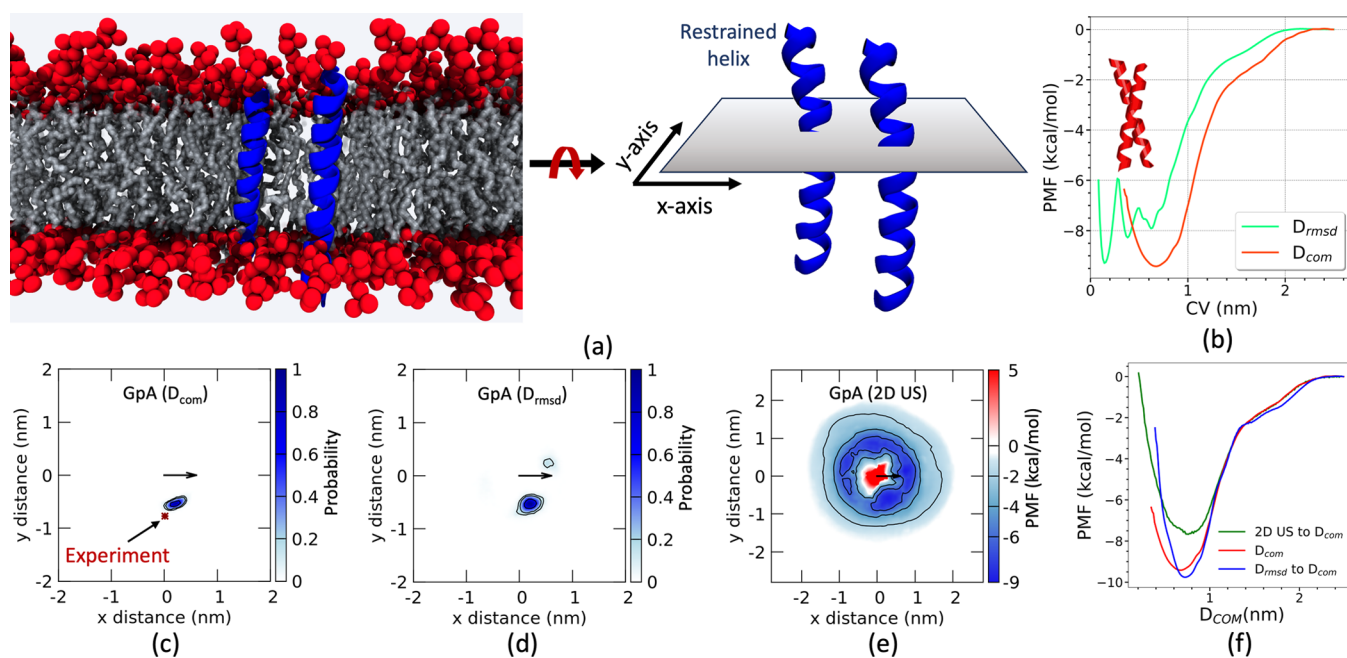


Figure 1. (a) Side view of the membrane bilayer containing a transmembrane protein homodimer and the representation of the homodimer onto the xy -plane. (b) Potential of mean force (PMF) for the association of GpA along D_{com} and D_{rmsd} . Density of dimer conformations on the xy -plane obtained from analyzing the minima of the PMF along (c) D_{com} and (d) D_{rmsd} . The experimental structure is represented by the red dot. (e) Two-dimensional (2D) PMF for the association of the GpA homodimer calculated using umbrella sampling along the xy -projection of the center-of-mass distance between two helices. The restrained helix is centered on the xy -plane, and the COM^{79} -Gly79 vector of the restrained helix was aligned to the positive x -axis and shown by the black arrow. (f) PMF along D_{com} and the projection of the PMFs obtained along D_{rmsd} and 2D US onto D_{com} . This figure is generated using data from a previously published study that employed the MARTINI v2.2 force field.⁵²

energy barrier and is enthalpically favorable while noting that the association is entropically favorable only in a bilayer with significant hydrophobic mismatch.⁴⁹

In an attempt to overcome the overstabilization of protein oligomers in the MARTINI v2 force field, Vattulainen and co-workers computed PMFs sampling over D_{com} for the homodimerization of EphA1 and ErbB1.⁵⁰ They suggested a reduction in the scale of the protein–protein LJ interactions. Using a complementary approach, we addressed the over-stabilization of protein dimers in the MARTINI v2 force field by enhancing the scale of the protein–lipid LJ interactions.⁵¹ We showed that the reparameterized MARTINI model can reproduce the experimental dimerization free energies of GpA, EphA1, C99, and RR-FGFR3 dimers while maintaining the original interaction modes between two helices unaffected.

In the following sections, we elaborate on the methodological developments conducted by our group in the study of transmembrane protein association. Due to the extensive study performed on GpA homodimerization, it is considered an ideal system for benchmarking new methodologies. Many of our proposed protocols for studying the TM homodimer were performed on GpA. However, we have also demonstrated the versatility of our proposed methods through their application to a variety of other TM homodimers.

■ ENHANCEMENTS THROUGH SAMPLING OVER TWO COLLECTIVE VARIABLES

We studied the dimerization of GpA using the one-dimensional reaction coordinates D_{com} and D_{rmsd} .⁵² The reference structure of GpA used to inform D_{rmsd} was derived from PDB ID 1AFO.⁵³ Umbrella sampling simulations were performed to obtain the PMFs. The coarse-grained MARTINI v2.2 force field was used

for the simulation of the protein-embedded lipid bilayer. A detailed description was provided in a previous study.⁵² The free energy landscape along D_{rmsd} contains three minima compared to one minimum observed along D_{com} (Figure 1b). The free energy difference of the most stable homodimer configuration compared to the dissociated state of the protein was found to be -9.4 and -9.3 kcal/mol along D_{com} and D_{rmsd} respectively. The characteristic and absolute values of free energy surfaces agree with the previously published results.⁴⁷

The lateral movement of the helices in a membrane bilayer can be visualized as the movement of the protein in a plane (xy -plane) perpendicular to the membrane normal (z -axis). A schematic representation of the TM homodimer projected onto the xy -plane is shown in Figure 1a. We analyzed the minimum of the PMF to have a better understanding of the conformational ensembles of dimers sampled during the simulation along D_{com} and D_{rmsd} .

To understand the interaction pattern of the homodimer, one of the helices was centered on the xy -plane. The COM^{79} -Gly79 vector of the centered helix was aligned to the positive x -axis, and the probability density of the second helix is shown as a colormap on the xy -plane, where COM^{79} is defined as the center-of-mass of 77–81 residues. Gly79 is the first glycine residue of the GpA glycine zipper motif. Therefore, the density observed along the positive x -axis indicates an interaction mediated by the glycine zipper motif. The probability density of dimer conformations obtained from simulation along D_{com} and D_{rmsd} showed that the sampling is restricted only to the native interaction (glycine zipper motif) of GpA (Figure 1). Given these observations, there are two possibilities to consider: (1) The glycine zipper motif interaction is the only favorable interaction predicted by the MARTINI v2.2 force field or (2)

the native interaction is one of the important structures; however, the other relevant structures are not sampled properly.

To further explore GpA homodimerization, we calculated a two-dimensional free energy surface along the xy -projection of the relative COM–COM distance of the TM helices.⁵² The COM⁷⁹-Gly79 vector of one helix was restrained to the positive x -axis, allowing the other helix to rotate without any restriction. A detailed description of the calculation of 2D PMF using US is provided in a previous study.⁵² The 2D PMF obtained for GpA using the MARTINI v2.2 force field is shown in Figure 1e, where the restrained helix was centered on the xy -plane.

The two-dimensional PMF identifies multiple thermodynamically relevant substates in the homodimer ensemble, including the native homodimer, stabilized by interaction facilitated by the glycine zipper motif, and non-native homodimer structures. Although the native homodimer conformation was found to represent the global free energy minimum, there exist other thermodynamically relevant non-native conformations. Comparison of the 2D PMF with the results derived from the 1D PMF demonstrates that incomplete sampling was achieved by using D_{com} or D_{rmsd} as CVs. The restricted sampling obtained from the 1D US results in a biased free energy surface, leading to an overestimation of the binding free energy (Figure 1f). Apparent convergence of the free energy surface may also be observed in the US simulations along D_{com} and D_{rmsd} due to the overly restricted sampling of the native dimer conformation.

We further characterized the interaction between GpA helices by studying the spontaneous aggregation of dimers and multiple helices in independent unbiased simulations.⁵² We observed similar interaction patterns as predicted by the 2D PMF. The GpA helices were found to associate by interacting through both native and non-native contacts in unbiased simulations.

NEED FOR SAMPLING OVER THREE COLLECTIVE VARIABLES

Building on the concept of TM helix movement in the xy -plane, we recently proposed a protocol using metadynamics (3D metadynamics).³⁵ We studied TM protein homodimerization using three collective variables, defined as the x -projection and y -projection of the center-of-mass distance between two helices and the crossing angle between the helix vectors (Figure 2). The 3D metadynamics protocol was used to study the dimerization

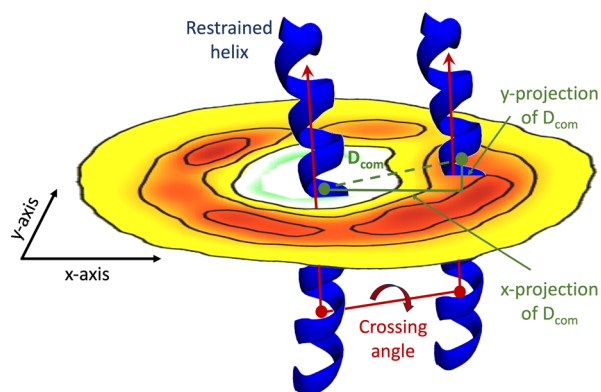


Figure 2. Schematic representation of three CVs used in the 3D metadynamics protocol: the x -projection and y -projection of the center-of-mass distance between two helices and the interhelical crossing angle.

of GpA in POPC, EphA1 in DLPC, and WALP23 in POPC and DOPC using the MARTINI v3 force field. Following a similar approach as 2D US, the angular motion of one helix was restrained by applying a harmonic potential between the COM^X-ResX vector and the positive x -axis, where COM^X is defined as the center of mass of $X - 2$ to $X + 2$ residues. Gly79, Gly554, and Ala11 were used as ResX in the case of GpA, EphA1, and WALP23, respectively. A detailed description of the 3D metadynamics method is provided in a previous study.³⁵

Projection of the 3D PMF onto the xy -plane is shown in Figure 3. The restrained helix was centered on the xy -plane, and the COM^X-ResX vector of the restrained helix was aligned to the positive x -axis and is shown as a black arrow. Multiple hotspots were observed on the xy -plane for GpA and EphA1, demonstrating multiple interaction sites on the helix, whereas one predominant hotspot observed for WALP23 suggests a more specific interaction. We also compared the 3D PMF results obtained from the metadynamics calculations with the 2D US calculations. The PMF obtained from the 2D US was found to capture similar interaction sites and predict the same dimerization free energies, Figure 4.

In a recent study, Ito and Sugita used four collective variables defined as the interhelical distance, crossing angle, and the two Crick angles between TM helices to study the dimerization of GpA and the C-terminal fragment of the amyloid precursor protein (APP-C99) in an implicit membrane using the CHARMM36m all-atom force field.⁴¹ They employed bias-exchange adaptively biased molecular dynamics (BE-ABMD) along four CVs and reported multiple important dimer configurations for both GpA and APP-C99. These findings validate our earlier observations regarding the role of multiple thermodynamically significant minima in the dimer-state ensemble.

We note that in the space defined by the two Crick angles, each homodimer conformation is represented twice, leading to redundant sampling. To reduce this redundancy and achieve faster convergence, the interhelical distance and two Crick angle CVs can be replaced with the xy -projection of the COM distance between two helices. This is the approach taken in our 3D metadynamics protocol, where one helix is restrained at the origin. It would, however, be of interest to investigate the performance of BE-ABMD as an alternative to metadynamics, as well as any relative advantage provided by the introduction of a fourth CV.

In addition to calculating dimerization free energy, another aspect of studying homodimerization of proteins using an enhanced sampling protocol is to identify the important dimer conformations. Several experimental studies have characterized the TM protein homodimer structures. X-ray crystallography^{53,55} and NMR spectroscopy^{53,55} experiments have established that the GpA homodimer is stabilized by the glycine zipper motif interaction, forming a right-handed helical structure. Arseniev and co-workers studied the EphA1 dimer in lipid bicelles using NMR spectroscopy.⁵⁶ The predominant EphA1 homodimer structures involved a glycine zipper motif interaction to form a right-handed helix. Several computational studies have also characterized the homodimer structural ensemble of GpA,^{47,51} EphA1,⁴⁸ WALP23,³⁵ and RR-FGFR3.⁵¹ However, in most cases, the simulations were found to be trapped in an energy minimum, resolving only one representative dimer conformation.

To identify the crossing angle distribution characterizing the homodimer ensembles, we projected the 3D PMF onto the

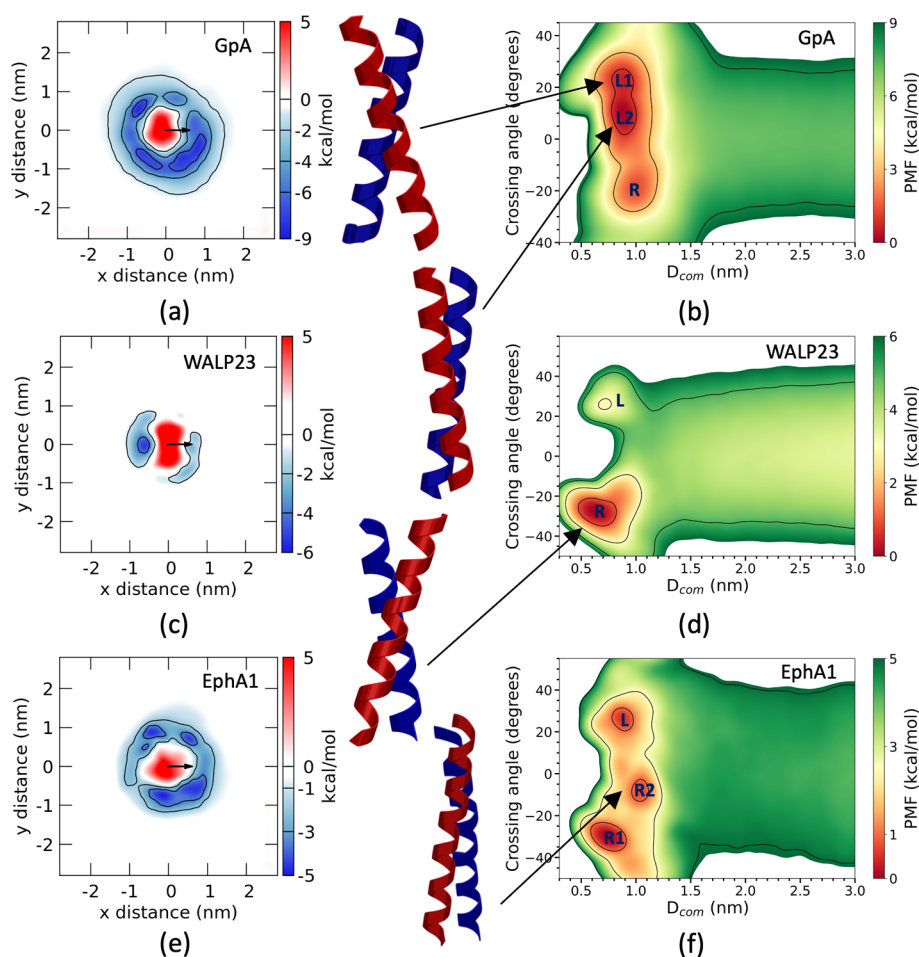


Figure 3. Projection of the 3D PMF obtained from the 3D metadynamics protocol onto the xy -plane and the crossing angle and D_{com} for the dimer association of (a,b) GpA in POPC, (c,d) WALP23 in POPC, and (e,f) EphA1 in DLPC. COM⁷⁹-Gly79 vector of GpA, COM¹¹-Ala11 vector of WALP23, and COM⁵⁴-Gly554 vector of EphA1 were centered on the xy -plane and are shown by the black arrow. This figure is generated using data from a previously published study using the MARTINI v3 force field.³⁵

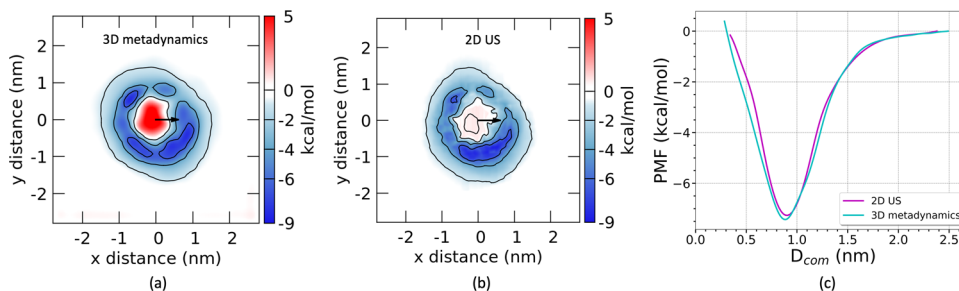


Figure 4. (a) Projection of the PMF obtained from the 3D metadynamics protocol onto the xy -plane and (b) 2D PMF over the xy -plane derived from umbrella sampling for the homodimerization of GpA using the MARTINI v3 force field. (c) Projection of the PMF onto D_{com} calculated using 2D US and 3D metadynamics. This figure is generated using data from a previously published study.³⁵

crossing angle and D_{com} (Figure 3). The energy minima on the free energy surface predicted that GpA and EphA1 form stable homodimer structures with three different average crossing angles, while WALP23 predominantly forms right-handed helical structures. Buck and co-workers also studied GpA and EphA1 homodimers using the MARTINI v3 force field and observed three conformational ensembles characterized by different average crossing angles.⁵⁷

The 3D metadynamics protocol also resolves the residue-level interactions that contribute to the specific helicity of the homodimer (Figure 5). Different hotspots on the xy -plane can

be linked to distinct residue-level interactions. To identify the interactions leading to a particular handedness, a density profile on the xy -plane can be obtained by averaging over the 3D PMF corresponding to that specific crossing angle region. We identified three different conformations leading to the L1 cluster of GpA, with one conformation in each L2 and R cluster (Figure 5). Additionally, we identified two conformations leading to the L cluster of EphA1 and one conformation in each R2 and R1 cluster. The glycine zipper motif was found to play an important role in the formation of right-handed helical structures in both GpA and EphA1, as predicted by experimental studies. WALP23

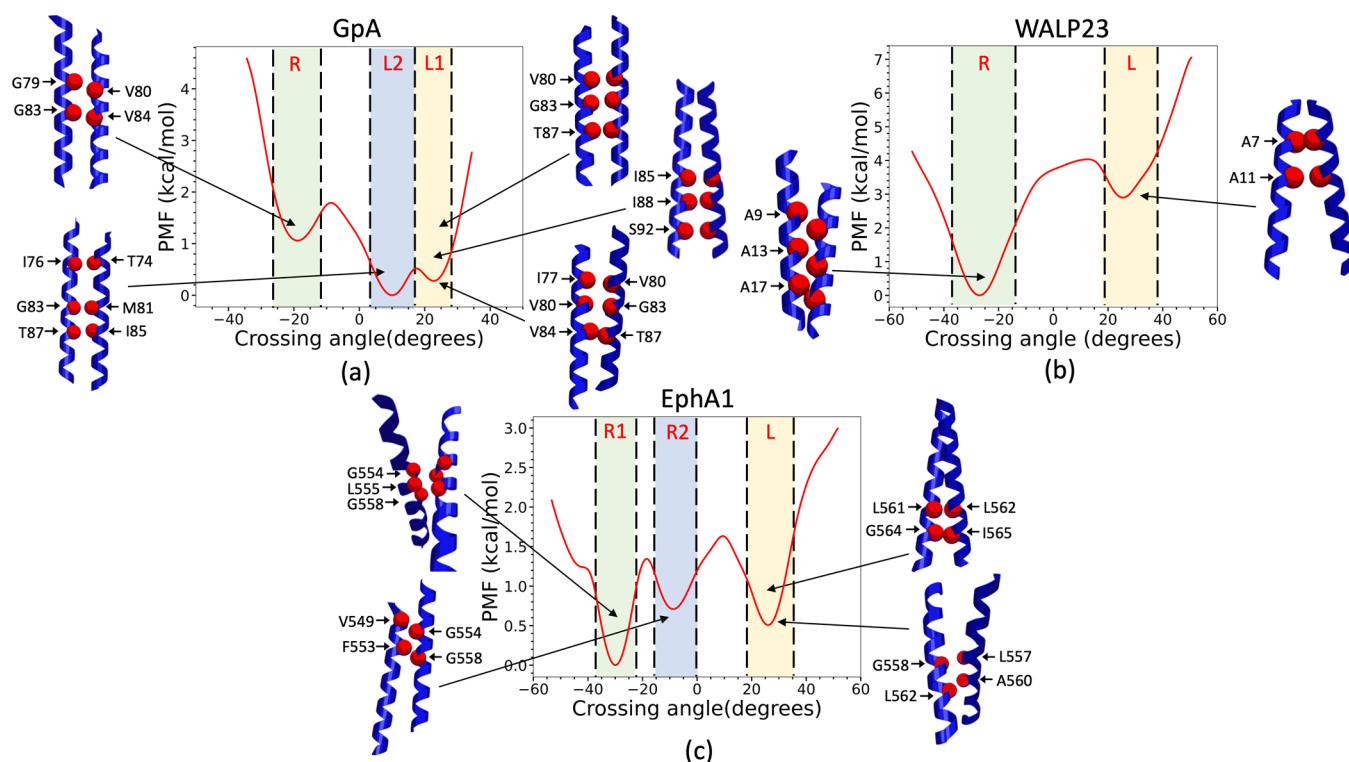


Figure 5. Projection of the 3D PMF obtained from the 3D metadynamics for the association of (a) GpA, (b) WALP23, and (c) EphA1 homodimers along the crossing angle using the MARTINI v3 force field. The names of the structural clusters (L or R) were assigned based on the average crossing angle distribution. The interhelical contacts observed in each cluster are shown as schematics. This figure is generated using data from a previously published study.³⁵

forms right-handed helical structures with interaction mediated by the Ala9-Ala13-Ala17 motif, as also indicated by the major hotspot on the *xy*-plane was observed along the negative direction of the director vector defined as the COM¹¹-Ala11 vector (Figure 3c).

The results obtained from the 3D metadynamic calculations were compared to unbiased simulations. The association of GpA, EphA1, and WALP23 was studied by using a self-assembly approach through multiple independent unbiased simulations. Although establishing equilibrium association–dissociation events from unbiased simulations proved challenging, stable homodimer conformations were identified. Similar intermolecular interactions were observed in the unbiased simulations, as predicted by the 3D metadynamics simulations. Taken together, the results demonstrate that the 3D metadynamics method exhaustively samples the homodimer structural ensemble and can identify thermodynamically relevant conformations contributing to the equilibrium association constant.

APPLICATION OF THE MACHINE-LEARNING-BASED PROTOCOLS

Recent developments in machine-learning-based protocols for identifying collective variables have significantly improved the study of biological rare events. Parrinello and co-workers have used deep linear discriminant analysis (DeepLDA) to study the aldol reaction and rotational transition of alanine dipeptide.⁵⁹ The DeepLDA protocol uses deep learning techniques to perform a nonlinear transformation of input descriptors that separate the data into different classes along the collective variable space, optimized by linear discriminant analysis.⁶⁰ Tiwary and co-workers have developed the state predictive information bottleneck (SPIB) protocol to identify collective

variables using the input descriptors obtained from the molecular dynamics trajectory.^{61,62} The SPIB protocol optimizes the representation of the important states of biological rare events onto the collective variable space using the time delay as a hyperparameter.

We used the DeepLDA and SPIB protocols to derive collective variables to effectively study the association/dissociation equilibrium of GpA and WALP23 homodimers in a POPC bilayer.⁵⁸ The collective variables using the DeepLDA and SPIB protocols were derived by using data obtained from extensive unbiased simulations. The free energies of dimerization of GpA and WALP23 obtained from the simulation along the DeepLDA- and SPIB-derived CVs were found to be similar to those obtained from the 3D metadynamics simulations (Figure 6). Additionally, we characterized the thermodynamically relevant homodimer conformational ensembles by projecting the data obtained from simulations along the CVs derived from DeepLDA and SPIB protocols onto the *xy*-plane and onto the crossing angle and D_{com} . Simulation along DeepLDA- and SPIB-derived CVs was found to sample similar homodimer conformational ensembles as predicted by the 3D metadynamics simulations. This study further validates the effectiveness of our intuition-based collective variables in the study of protein–protein association in membranes.

CONNECTING COMPUTATIONS WITH EXPERIMENTAL RESULTS

Several experiments have been performed to quantify the dimerization free energy of the TM homodimers. Fleming et al. reported the free energy of dimerization of GpA to be -9.0 kcal/mol in pentaerythritol octyl ether (C_8E_5) detergent.⁶⁴ They also reported the mole fraction standard state dimerization free

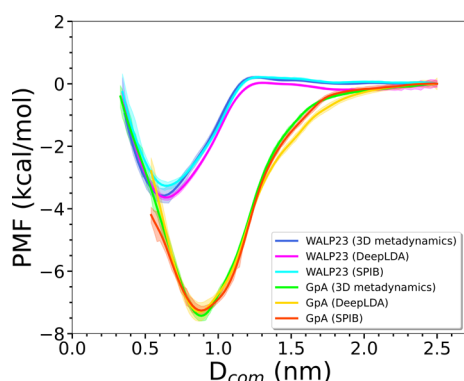


Figure 6. Potential of mean force (PMF) for the association of GpA and WALP23 homodimers along D_{com} obtained from the 3D metadynamics simulations and the simulations along the DeepLDA- and SPIB-derived CVs. This figure is generated using data from a previously published study using the MARTINI v3 force field.⁵⁸

energy of GpA to be -7.0 kcal/mol in the same detergent.⁶⁵ The free energy of dimerization in SDS detergent was found to be -4.5 to -5.5 kcal/mol.⁶⁵ Nash et al. studied a de novo design of TM helix with the GpA TM domain motif in *Escherichia coli* inner membrane and reported the dimerization free energy to be -7.5 kcal/mol.⁶⁶ Hong et al. used a steric trap method to study the dimerization of GpA in POPC and reported the free energy of dimerization to be -12.1 kcal/mol.⁶⁷ Hristova and co-workers studied residues 75–98 of the GpA TM domain in different plasma membranes and found the free energy of dimerization to be -3.4 to -4.0 kcal/mol using a FRET assay.^{68,69} Artemenko et al. reported the dimerization free energy of EphA1 to be -3.7 kcal/mol in a DMPC liposome using a FRET assay.⁷⁰ Sanders and co-workers studied the C99 domain of the amyloid precursor protein in a POPC/POPG lipid vesicle using electron paramagnetic resonance (EPR) and a FRET assay, reporting the dimerization free energy of -3.2 kcal/mol.⁷⁴ Hristova and co-workers studied the FGFR3 homodimer in POPC and reported the free energy of dimerization to be -2.7 kcal/mol⁷⁶ using a FRET assay. They also studied ErbB1 TM domain dimerization in DLPC using FRET and reported the dimerization free energy to be -2.5 kcal/mol.⁷⁸ The dimerization free energy values of different TM homodimers obtained from previous computational and experimental studies are tabulated in Table 1.

The free energy of dimerization (ΔG) can be calculated from a surface free energy defining the homodimer association. The PMF obtained from 2D US or 3D metadynamics calculations can be projected along the D_{com} coordinate. The resulting PMF along D_{com} can be integrated to obtain the dimerization free energy:^{42,79–81}

$$K_D = \frac{1}{2} \times \frac{2\pi}{A_0} \int_0^{r_c} r e^{-\Delta W(r)/k_B T} dr$$

$$\Delta G = -RT \ln K_D \quad (2)$$

where $\Delta W(r)$ is the PMF along D_{com} . An entropic factor correction, $k_B T \ln(r)$, is added to $\Delta W(r)$ to account for the increasing area accessible to dimers as r increases. A distance cutoff r_c is chosen, beyond which the PMF reaches a plateau. The standard reference area A_0 is typically set to 1 nm^2 . The factor $\frac{1}{2}$ is added as a symmetry factor in the case of a homodimer.

Following Fleming's approach,⁶⁵ Marrink and co-workers related the association constant to the mole fraction standard free energy⁴⁵

$$K_D = K_X \frac{a_L}{2A_0}$$

$$\Delta G_X = \Delta G + RT \ln \frac{a_L}{2A_0} \quad (3)$$

where ΔG_X is the mole fraction binding free energy and a_L is the area per lipid. The mole fraction standard free energy for the homodimerization of different protein homodimers is shown in Table 2. The correction typically leads to changes in ΔG on the order of $k_B T$.

A computational protocol for studying the TM homodimer should efficiently calculate dimerization free energy and identify important dimer contacts. 3D metadynamics is proven to be an ideal protocol as it exhaustively samples the conformational ensemble of TM homodimers, providing a converged estimation of the dimerization free energy. It is found to be more efficient than 2D umbrella sampling along the xy -projection of the relative COM distance between two helices. A converged estimation of the dimerization free energy obtained from the simulation can be compared to experimental results. Accurate knowledge of the free energy of dimerization is essential for parametrizing a force field and understanding the mechanism of membrane protein interactions.

We note that a minimum of $800 \mu\text{s}$ of simulation is required to obtain a converged two-dimensional PMF using US along the xy -plane ($2 \mu\text{s}/\text{umbrella}$). In contrast, an $80 \mu\text{s}$ simulation is sufficient to obtain a converged one-dimensional PMF along the D_{com} or D_{rmsd} CVs ($4 \mu\text{s}/\text{umbrella}$). However, only a restricted set of dimer conformations is sampled along D_{com} or D_{rmsd} . Compared with the US approach, a converged PMF can also be obtained after approximately $100 \mu\text{s}$ of simulation using the 3D metadynamics protocol. Similarly, metadynamics along machine learning-derived CVs can yield a converged PMF within $100 \mu\text{s}$ of simulation. However, a drawback of using machine learning-derived CVs is the computational cost associated with calculating an extensive set of descriptors at each time step, especially when compared to intuition-driven CVs.

FUTURE DIRECTIONS

Characterizing TM Homodimer Structures Using All-Atom Models. The study of TM protein dimerization has mainly been explored using coarse-grained models. The all-atom force field has not yet been effectively validated for the study of single-pass TM helix association. Some studies have attempted to use all-atom models to explore protein dimerization; however, there have been significant variations in reported dimerization free energy values for the same system using the same force field. Our proposed method, employing a 3D metadynamics approach, was validated using the MARTINI coarse-grained force field. However, it is straightforward to implement the method in future applications using all-atom model simulations. The successful implementation of the method could lead to the characterization of thermodynamically relevant homodimer structures and the evaluation of the dimerization free energies for TM homodimers. The results can be compared with available experimental data to either solidify or propose necessary calibrations for the all-atom force fields.

Bridging the Gap between Computational and Experimental Studies. Achieving the goal of establishing a

Table 1. Dimerization Free Energy of Different Protein Homodimers Reported in Previous Studies

protein/sequence	methods/force field	free energy of dimerization (kcal/mol)	medium
GpA ⁵³			
Computational			
residues 73–95	CHARMM27	−11.5 ⁴²	dodecane
residues 69–97	CHARMM36m	−10.7 ³⁶	POPC bilayer
residues 69–97	CHARMM36 (rescaled)	−3.0 to −3.8 ³⁷	POPC bilayer
residues 70–96	MARTINI	−7.5 ⁶³	DPPC bilayer
residues 70–96	MARTINI	−6.7 ⁶³	DLPC bilayer
residues 70–96	MARTINI	−6.6 ⁶³	DOPC bilayer
TM domain	MARTINI v2	−9.1 ⁴⁶	DPPC bilayer
residues 69–97	MARTINI v2	−8.4 ⁴⁷	POPC bilayer
residues 69–97	MARTINI v2	−9.3 ⁵²	POPC bilayer
residues 69–97	MARTINI v2	−7.5 ⁵²	POPC bilayer
residues 69–97	MARTINI v3	−5.9 ⁴⁵	POPC bilayer
TM domain	MARTINI v3	−3.1 ⁵⁷	DLPC bilayer
residues 69–97	MARTINI v3	−7.0 ³⁵	POPC bilayer
Experimental			
TM domain	ultracentrifugation	−9.0 ⁶⁴	pentaerythritol octyl ether (C ₈ E ₃)
TM domain	sedimentation equilibrium	−7.0 ⁶⁵	pentaerythritol octyl ether (C ₈ E ₃)
TM domain	sedimentation equilibrium	−4.5 to −5.5 ⁶⁵	sodium dodecyl sulfate (SDS)
De novo design of GpA TM domain motifs	GALLEX assay	−7.5 ⁶⁶	<i>Escherichia coli</i> inner membrane
TM domain	steric trap	−12.1 ⁶⁷	POPC bilayer
residues 75–98	FRET	−3.4 to −4.0 ^{68,69}	plasma membrane
EphA1 ⁵⁶			
Computational			
residues 542–572	MARTINI v2	−14.3 ⁴⁸	DPPC bilayer
residues 536–573	MARTINI v2	−7.1 ⁵⁰	DLPC bilayer
residues 544–572	MARTINI v2	−7.7 ⁵¹	DLPC bilayer
residues 536–573	MARTINI v3	−4.7 ⁴⁵	DLPC bilayer
residues 544–572	MARTINI v3	−3.8 ³⁵	DLPC bilayer
Experimental			
residues 544–572	FRET	−3.7 ⁷⁰	DMPC bilayer
WALP23			
Computational			
	MARTINI	−2.9 ⁷¹	Di-C18:2PC
	MARTINI v2	−4.8 ⁴⁹	DOPC bilayer
	MARTINI v3	−2.9 ⁴⁵	POPC bilayer
	MARTINI v3	−2.6 ⁴⁵	DOPC bilayer
	MARTINI v3	−3.2 ³⁵	POPC bilayer
	MARTINI v3	−2.6 ³⁵	DOPC bilayer
Experimental			
(AALALAA) ₃	FRET	−3.0 ⁷²	Di-C18:1PC
C99 ⁷³			
Computational			
residues 22–55	MARTINI v2	−9.1 ⁵¹	POPC–POPG bilayer
Experimental			
full length	FRET	−3.2 ⁷⁴	POPC–POPG bilayer
FGFR3 ⁷⁵			
Computational			
residues 367–399	MARTINI v2	−10.3 ⁵¹	POPC bilayer
Experimental			
residues 367–399	FRET	−2.7 ⁷⁶	POPC bilayer
ErbB1 ⁷⁷			
Computational			
residues 634–677	MARTINI v2	−9.4 ⁵⁰	DLPC bilayer
residues 634–677	MARTINI v3	−2.4 ⁴⁵	DLPC bilayer
Experimental			
KKKSIATGVVGALLLLVVALGIGLFMKKK	FRET	−2.5 ⁷⁸	DLPC bilayer

force field capable of predicting realistic interprotein contacts and interaction energies is a central goal for the field. A

membrane protein force field relies primarily on two types of experimental data: (1) Partition coefficients of amino acids in

Table 2. Dimerization Free Energy (ΔG) and Mole Fraction Standard Free Energy (ΔG_X) of Different Protein Homodimers Obtained from the 3D Metadynamics Protocol Using the MARTINI v3 Force Field

protein (lipid)	temperature (K)	a_L (nm ²)	ΔG (kcal/mol)	ΔG_X (kcal/mol)
GpA (POPC)	310	0.66	-7.0 ± 0.1	-7.7 ± 0.1
EphA1 (DLPC)	303	0.59	-3.8 ± 0.1	-4.5 ± 0.1
WALP23 (POPC)	310	0.66	-3.2 ± 0.2	-3.9 ± 0.2
WALP23 (DOPC)	310	0.69	-2.6 ± 0.1	-3.2 ± 0.1

water and membrane bilayers, which are relatively easy to reproduce in computer simulations. (2) Free energy of dimerization and contacts between TM helices in the membrane, which are challenging to interpret. The pattern of contacts between two membrane proteins can be understood from structures obtained from X-ray crystallography and NMR experiments. The free energy of TM protein association is primarily derived from experimental association constants obtained from FRET assays. Considering that for the dimerization of GpA, arguably one of the most extensively studied protein homodimers, the reported values of the free energy of dimerization vary from -3.4 to -12.1 kcal/mol.

Focusing solely on data obtained from FRET assays commonly used to study several TM homodimers, differences in results can be ascribed to variations in the length of the TM helix studied, variations in the membrane bilayer composition, and the use of different fluorescence labels. While some FRET studies were conducted in complex membrane environments, most of the studies were performed in simple bilayer environments composed of one or two types of lipids. In computer simulations, replicating the protein sequence used in experiments and modeling a single or bicomponent lipid bilayer are straightforward. However, the impact of the donor and acceptor labels used in FRET studies has not been quantified through MD simulations. These fluorescence labels may influence both native and non-native homodimer conformations, thereby affecting the dimerization free energy. A more precise comparison between computer simulations and experimental results can be achieved by explicitly modeling these fluorescence tags onto the protein helices (Figure 7), representing an important area for future study.

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

Author

Ayan Majumder – Department of Chemistry, Boston University, Boston, Massachusetts 02215, United States; orcid.org/0009-0005-4218-3276

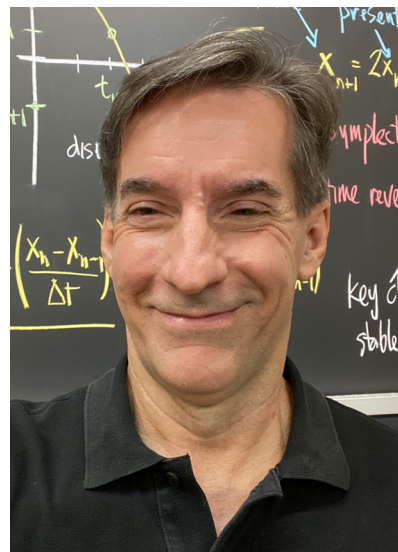
Complete contact information is available at:

<https://pubs.acs.org/10.1021/acs.jpcb.5c04286>

Notes

The authors declare no competing financial interest.

Biographies



John E. Straub is a Professor of Chemistry at Boston University. He received a B.S. in Chemistry in 1982 from the University of Maryland, where he worked with Millard H. Alexander, and a Ph.D. in Chemical Physics in 1987 from Columbia University, working with Bruce J. Berne. He subsequently conducted postdoctoral research with Professor Martin Karplus at Harvard University before moving to Boston University in 1990. He has held visiting positions at Hebrew University in Jerusalem, Montana State University in Bozeman, Nagoya

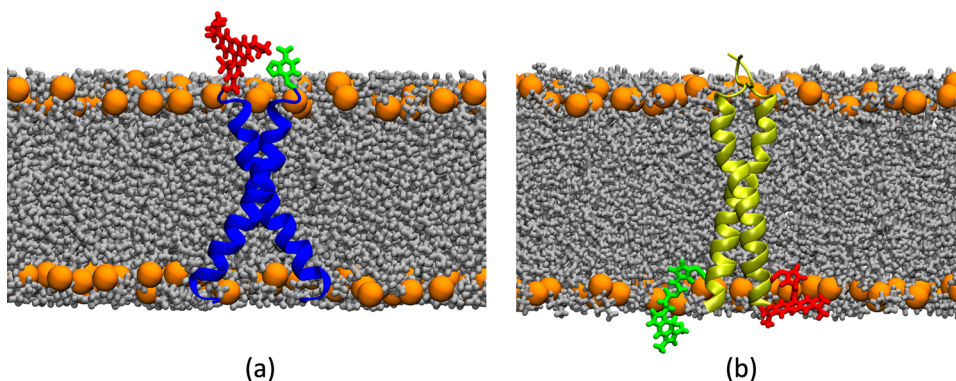
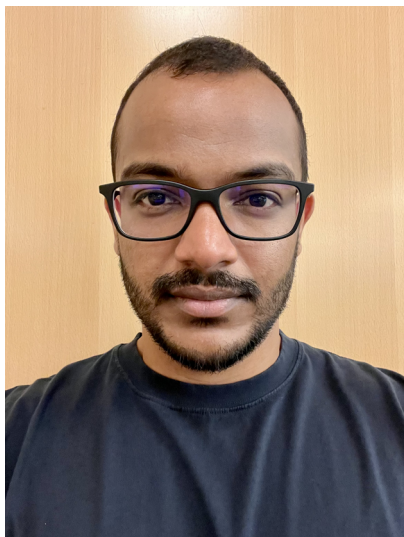


Figure 7. Pictorial representation of a membrane bilayer containing a fluorescence tag labeled transmembrane protein dimer, showing EphA1 (blue), FGFR3 (yellow), lipid (gray), lipid head (orange), and FRET donor (green) and acceptor (red). A FRET donor/acceptor pair, 4-chloro-7-nitrobenzofurazan/sulforhodamine B sulfonyl chloride, is attached to the N-terminal of the EphA1 dimer,⁷⁰ and a FRET donor/acceptor pair, BODIPY-fluorescein-maleimide/tetramethylrhodamine-maleimide, is attached to the Cys396 residue (near C-terminal) of the FGFR3 dimer.⁷⁶

University in Japan, and University Grenoble Alpes in France. His research interests include the development and application of computational methods for the study of equilibrium and dynamical properties of complex liquids and bimolecular systems. He is the author of the textbook "Mathematical Methods for Molecular Science" (MIT Press, 2022).



Ayan Majumder received his B.Sc. from Ramakrishna Mission Residential College, Narendrapur, and his M.Sc. from the Indian Institute of Technology, Kharagpur. He then pursued a Ph.D. in chemistry at Boston University, where he worked with Prof. John E. Straub. He is currently a postdoctoral associate at the University of Chicago, working with Prof. Gregory A. Voth. His research interests include molecular dynamics simulations of membrane proteins and multiscale modeling of biological systems.

ACKNOWLEDGMENTS

The authors thank Shingo Ito, Yuko Okamoto, Benoît Roux, and Yuji Sugita for helpful discussions and gratefully acknowledge the generous support of the National Science Foundation (CHE1900416) and the National Institutes of Health (R01 GM107703).

REFERENCES

- (1) Nicolson, G. L. The Fluid—Mosaic Model of Membrane Structure: Still relevant to understanding the structure, function and dynamics of biological membranes after more than 40 years. *Biochim. Biophys. Acta, Biomembr.* **2014**, 1451–1466.
- (2) Nagle, J. F.; Tristram-Nagle, S. Structure of lipid bilayers. *Biochimica et Biophysica Acta (BBA)-Reviews on Biomembranes* **2000**, 1469, 159–195.
- (3) Vinothkumar, K. R.; Henderson, R. Structures of membrane proteins. *Q. Rev. Biophys.* **2010**, 43, 65–158.
- (4) Majumder, A.; Gu, Y.; Chen, Y.-C.; An, X.; Reinhard, B. M.; Straub, J. E. Probing the Origins of the Disorder-to-Order Transition of a Modified Cholesterol in Ternary Lipid Bilayers. *J. Am. Chem. Soc.* **2024**, 146, 27725–27735.
- (5) Marrink, S. J.; Corradi, V.; Souza, P. C.; Ingolfsson, H. I.; Tieleman, D. P.; Sansom, M. S. Computational modeling of realistic cell membranes. *Chem. Rev.* **2019**, 119, 6184–6226.
- (6) Engelman, D. M. Membranes are more mosaic than fluid. *Nature* **2005**, 438, 578.
- (7) Yang, Y.; Lee, M.; Fairn, G. D. Phospholipid subcellular localization and dynamics. *J. Biol. Chem.* **2018**, 293, 6230–6240.
- (8) Harayama, T.; Riezman, H. Understanding the diversity of membrane lipid composition. *Nat. Rev. Mol. Cell Biol.* **2018**, 19, 281–296.
- (9) Van Meer, G.; Voelker, D. R.; Feigenson, G. W. Membrane lipids: where they are and how they behave. *Nat. Rev. Mol. Cell Biol.* **2008**, 9, 112–124.
- (10) Lingwood, D.; Simons, K. Lipid rafts as a membrane-organizing principle. *science* **2010**, 327, 46–50.
- (11) Jacobson, K.; Mouritsen, O. G.; Anderson, R. G. Lipid rafts: at a crossroad between cell biology and physics. *Nature cell biology* **2007**, 9, 7–14.
- (12) Phillips, R.; Ursell, T.; Wiggins, P.; Sens, P. Emerging roles for lipids in shaping membrane-protein function. *Nature* **2009**, 459, 379–385.
- (13) Kwon, S.; Majumder, A.; Straub, J. E. Exploring Free Energy Landscapes for Protein Partitioning into Membrane Domains in All-Atom and Coarse-Grained Simulations. *J. Chem. Theory Comput.* **2024**, 20, 9687–9698.
- (14) Sriram, K.; Insel, P. A. G protein-coupled receptors as targets for approved drugs: how many targets and how many drugs? *Molecular pharmacology* **2018**, 93, 251–258.
- (15) Laeremans, T.; Sands, Z. A.; Claes, P.; De Blieck, A.; De Cesco, S.; Triest, S.; Busch, A.; Felix, D.; Kumar, A.; Jaakola, V.-P.; et al. Accelerating GPCR drug discovery with conformation-stabilizing VHHs. *Frontiers in Molecular Biosciences* **2022**, 9, No. 863099.
- (16) Nguyen, P. H.; Ramamoorthy, A.; Sahoo, B. R.; Zheng, J.; Faller, P.; Straub, J. E.; Dominguez, L.; Shea, J.-E.; Dokholyan, N. V.; De Simone, A.; et al. Amyloid oligomers: A joint experimental/computational perspective on Alzheimer's disease, Parkinson's disease, type II diabetes, and amyotrophic lateral sclerosis. *Chem. Rev.* **2021**, 121, 2545–2647.
- (17) Toyoshima, C.; Cornelius, F. New crystal structures of PII-type ATPases: excitement continues. *Curr. Opin. Struct. Biol.* **2013**, 23, 507–514.
- (18) Psachoulia, E.; Nikolaidi, B.; Marshall, D.; Sansom, M. S. Molecular dynamics simulations of the dimerization of transmembrane α -helices. *Biophys. J.* **2010**, 98, 417a.
- (19) Mori, T.; Miyashita, N.; Im, W.; Feig, M.; Sugita, Y. Molecular dynamics simulations of biological membranes and membrane proteins using enhanced conformational sampling algorithms. *Biochimica et Biophysica Acta (BBA)-Biomembranes* **2016**, 1858, 1635–1651.
- (20) Majumder, A.; Vuksanovic, N.; Ray, L. C.; Bernstein, H. M.; Allen, K. N.; Imperiali, B.; Straub, J. E. Synergistic computational and experimental studies of a phosphoglycosyl transferase membrane/ligand ensemble. *J. Biol. Chem.* **2023**, 299, 105194.
- (21) Lee, T.-S.; Cerutti, D. S.; Mermelstein, D.; Lin, C.; LeGrand, S.; Giese, T. J.; Roitberg, A.; Case, D. A.; Walker, R. C.; York, D. M. GPU-accelerated molecular dynamics and free energy methods in Amber18: performance enhancements and new features. *J. Chem. Inf. Model.* **2018**, 58, 2043–2050.
- (22) Friedrichs, M. S.; Eastman, P.; Vaidyanathan, V.; Houston, M.; Legrand, S.; Beberg, A. L.; Ensign, D. L.; Bruns, C. M.; Pande, V. S. Accelerating molecular dynamic simulation on graphics processing units. *Journal of computational chemistry* **2009**, 30, 864–872.
- (23) Shaw, D. E.; Dror, R. O.; Salmon, J. K.; Grossman, J.; Mackenzie, K. M.; Bank, J. A.; Young, C.; Deneroff, M. M.; Batson, B.; Bowers, K. J.; *Millisecond-Scale Molecular Dynamics Simulations on Anton*. In Proceedings of the Conference on High Performance Computing Networking, Storage and Analysis, 2009; pp 1–11.
- (24) 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.
- (25) Roux, B. The calculation of the potential of mean force using computer simulations. *Computer physics communications* **1995**, 91, 275–282.
- (26) 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.
- (27) Chipot, C. Free energy methods for the description of molecular processes. *Annual Review of Biophysics* **2023**, 52, 113–138.

- (28) Sugita, Y.; Okamoto, Y. Replica-exchange molecular dynamics method for protein folding. *Chemical physics letters* **1999**, *314*, 141–151.
- (29) Miyashita, N.; Straub, J. E.; Thirumalai, D.; Sugita, Y. Transmembrane structures of amyloid precursor protein dimer predicted by replica-exchange molecular dynamics simulations. *J. Am. Chem. Soc.* **2009**, *131*, 3438–3439.
- (30) Pantelopulos, G. A.; Matsuoka, D.; Hutchison, J. M.; Sanders, C. R.; Sugita, Y.; Straub, J. E.; Thirumalai, D. Formation of extramembrane β -strands controls dimerization of transmembrane helices in amyloid precursor protein C99. *Proc. Natl. Acad. Sci. U. S. A.* **2022**, *119*, No. e2212207119.
- (31) Mitsuake, A.; Sugita, Y.; Okamoto, Y. Generalized-ensemble algorithms for molecular simulations of biopolymers. *Peptide Science: Original Research on Biomolecules* **2001**, *60*, 96–123.
- (32) Barducci, A.; Bussi, G.; Parrinello, M. Well-tempered metadynamics: a smoothly converging and tunable free-energy method. *Physical review letters* **2008**, *100*, No. 020603.
- (33) Leone, V.; Marinelli, F.; Carloni, P.; Parrinello, M. Targeting biomolecular flexibility with metadynamics. *Curr. Opin. Struct. Biol.* **2010**, *20*, 148–154.
- (34) Filippov, A.; Orådd, G.; Lindblom, G. The effect of cholesterol on the lateral diffusion of phospholipids in oriented bilayers. *Biophysical journal* **2003**, *84*, 3079–3086.
- (35) Majumder, A.; Straub, J. E. The role of structural heterogeneity in the homodimerization of transmembrane proteins. *J. Chem. Phys.* **2023**, *159*, 134101.
- (36) Blazhynska, M.; Gumbart, J. C.; Chen, H.; Tajkhorshid, E.; Roux, B.; Chipot, C. A Rigorous Framework for Calculating Protein–Protein Binding Affinities in Membranes. *J. Chem. Theory Comput.* **2023**, *19*, 9077–9092.
- (37) 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.
- (38) Kokubo, H.; Okamoto, Y. Prediction of transmembrane helix configurations by replica-exchange simulations. *Chemical physics letters* **2004**, *383*, 397–402.
- (39) Kokubo, H.; Okamoto, Y. Prediction of membrane protein structures by replica-exchange Monte Carlo simulations: case of two helices. *J. Chem. Phys.* **2004**, *120*, 10837–10847.
- (40) Urano, R.; Kokubo, H.; Okamoto, Y. Predictions of tertiary structures of α -helical membrane proteins by replica-exchange method with consideration of helix deformations. *J. Phys. Soc. Jpn.* **2015**, *84*, No. 084802.
- (41) Ito, S.; Sugita, Y. Free-energy landscapes of transmembrane homodimers by bias-exchange adaptively biased molecular dynamics. *Biophys. Chem.* **2024**, *307*, No. 107190.
- (42) 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.
- (43) Sugita, Y.; Kitao, A.; Okamoto, Y. Multidimensional replica-exchange method for free-energy calculations. *J. Chem. Phys.* **2000**, *113*, 6042–6051.
- (44) 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.
- (45) Souza, P. C.; Alessandri, R.; Barnoud, J.; Thallmair, S.; Faustino, I.; Grünewald, F.; Patmanidis, I.; Abdizadeh, H.; Bruininks, B. M.; Wassenaar, T. A.; et al. Martini 3: a general purpose force field for coarse-grained molecular dynamics. *Nat. Methods* **2021**, *18*, 382–388.
- (46) 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.
- (47) 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.
- (48) 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.
- (49) Castillo, N.; Monticelli, L.; Barnoud, J.; Tieleman, D. P. Free energy of WALP23 dimer association in DMPC, DPPC, and DOPC bilayers. *Chemistry and physics of lipids* **2013**, *169*, 95–105.
- (50) Javanainen, M.; Martinez-Seara, H.; Vattulainen, I. Excessive aggregation of membrane proteins in the Martini model. *PLoS one* **2017**, *12*, No. e0187936.
- (51) Majumder, A.; Straub, J. E. Addressing the excessive aggregation of membrane proteins in the MARTINI model. *J. Chem. Theory Comput.* **2021**, *17*, 2513–2521.
- (52) Majumder, A.; Kwon, S.; Straub, J. E. On Computing Equilibrium Binding Constants for Protein–Protein Association in Membranes. *J. Chem. Theory Comput.* **2022**, *18*, 3961–3971.
- (53) MacKenzie, K. R.; Prestegard, J. H.; Engelman, D. M. A transmembrane helix dimer: structure and implications. *Science* **1997**, *276*, 131–133.
- (54) 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.
- (55) 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. *Biophysical journal* **2002**, *82*, 2476–2486.
- (56) 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.
- (57) Sahoo, A. R.; Souza, P. C.; Meng, Z.; Buck, M. Transmembrane dimers of type 1 receptors sample alternate configurations: MD simulations using coarse grain Martini 3 versus AlphaFold2Multimer. *Structure* **2023**, *31*, 735–745.
- (58) Majumder, A.; Straub, J. E. Machine Learning Derived Collective Variables for the Study of Protein Homodimerization in Membrane. *J. Chem. Theory Comput.* **2024**, *20*, 5774–5783.
- (59) Bonati, L.; Rizzi, V.; Parrinello, M. Data-driven collective variables for enhanced sampling. *Journal of physical chemistry letters* **2020**, *11*, 2998–3004.
- (60) Dorfer, M.; Kelz, R.; Widmer, G. Deep linear discriminant analysis. *arXiv preprint arXiv:1511.04707*, 2015.
- (61) Wang, D.; Tiwary, P. State predictive information bottleneck. *J. Chem. Phys.* **2021**, *154*, 134111.
- (62) Mehdi, S.; Wang, D.; Pant, S.; Tiwary, P. Accelerating all-atom simulations and gaining mechanistic understanding of biophysical systems through state predictive information bottleneck. *J. Chem. Theory Comput.* **2022**, *18*, 3231–3238.
- (63) Janosi, L.; Prakash, A.; Doxastakis, M. Lipid-modulated sequence-specific association of glycophorin A in membranes. *Biophysical journal* **2010**, *99*, 284–292.
- (64) Fleming, K. G.; Ackerman, A. L.; Engelman, D. M. The effect of point mutations on the free energy of transmembrane α -helix dimerization. *Journal of molecular biology* **1997**, *272*, 266–275.
- (65) Fleming, K. G. Standardizing the free energy change of transmembrane helix–helix interactions. *Journal of molecular biology* **2002**, *323*, 563–571.
- (66) 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.
- (67) 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.
- (68) 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.

- (69) 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.
- (70) Artemenko, E. O.; Egorova, N. S.; Arseniev, A. S.; Feofanov, A. V. Transmembrane domain of EphA1 receptor forms dimers in membrane-like environment. *Biochimica et Biophysica Acta (BBA)-Biomembranes* **2008**, *1778*, 2361–2367.
- (71) Schäfer, L. V.; de Jong, D. H.; Holt, A.; Rzepiela, A. J.; de Vries, A. H.; Poolman, B.; Killian, J. A.; Marrink, S. J. Lipid packing drives the segregation of transmembrane helices into disordered lipid domains in model membranes. *Proc. Natl. Acad. Sci. U. S. A.* **2011**, *108*, 1343–1348.
- (72) Yano, Y.; Matsuzaki, K. Measurement of thermodynamic parameters for hydrophobic mismatch 1: self-association of a transmembrane helix. *Biochemistry* **2006**, *45*, 3370–3378.
- (73) 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 letters* **2012**, *586*, 1687–1692.
- (74) 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.
- (75) 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.
- (76) 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.
- (77) Bocharov, E. V.; Lesovoy, D. M.; Pavlov, K. V.; Pustovalova, Y. E.; Bocharova, O. V.; Arseniev, A. S. Alternative packing of EGFR transmembrane domain suggests that protein–lipid interactions underlie signal conduction across membrane. *Biochimica et Biophysica Acta (BBA)-Biomembranes* **2016**, *1858*, 1254–1261.
- (78) Chen, L.; Merzlyakov, M.; Cohen, T.; Shai, Y.; Hristova, K. Energetics of ErbB1 transmembrane domain dimerization in lipid bilayers. *Biophysical journal* **2009**, *96*, 4622–4630.
- (79) Duboué-Dijon, E.; Hénin, J. Building intuition for binding free energy calculations: Bound state definition, restraints, and symmetry. *J. Chem. Phys.* **2021**, *154*, 204101.
- (80) Luo, H.; Sharp, K. On the calculation of absolute macromolecular binding free energies. *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99*, 10399–10404.
- (81) Roux, B. *Computational Modeling and Simulations of Biomolecular Systems*; World Scientific, 2022.