

# Evaluating the efficiency of the Martini force field to study protein dimerization in aqueous and membrane environments

*Christos Lamprakis<sup>1</sup>, Ioannis Andreadelis<sup>1</sup>, John Manchester<sup>2</sup>, Camilo Velez-Vega<sup>2</sup>,*

*José S. Duca<sup>2</sup>, Zoe Cournia<sup>1,\*</sup>*

<sup>1</sup>Biomedical Research Foundation, Academy of Athens, 4 Soranou Ephessiou, 11527 Athens, Greece

<sup>2</sup>Computer-Aided Drug Discovery, Global Discovery Chemistry, Novartis Institutes for BioMedical Research, 100 Technology Square, Cambridge, Massachusetts 02139, United States

Email: [zcournia@bioacademy.gr](mailto:zcournia@bioacademy.gr)

Phone: +30 210 6597 195

## KEYWORDS

Coarse-grained simulations; MARTINI; force field; protein dimerization; MD simulations

## ABSTRACT

Protein-protein complex assembly is one of the major drivers of biological response. Understanding the mechanisms of protein oligomerization/dimerization would allow one to elucidate how these complexes participate in biological activities and could ultimately lead to new approaches in designing novel therapeutic agents. However, determining the exact association pathways and structures of such complexes remains a challenge. Here, we use parallel tempering metadynamics simulations in the well-tempered ensemble to evaluate the performance of Martini 2.2P and Martini open-beta 3 (Martini 3) force fields in reproducing the structure and energetics of the dimerization process of membrane proteins and proteins in an aqueous solution in reasonable accuracy and throughput. We find that Martini 2.2P systematically overestimates the free energy of association by estimating large barriers in distinct areas, which likely leads to overaggregation when multiple monomers are present. In comparison, the less viscous Martini 3 results in a systematic underestimation of the free energy of association for proteins in solution, while it performs well in describing the association of membrane proteins. In all cases the near-native dimer complexes are identified as minima in the free energy surface albeit not always as the lowest minima. In the case of Martini 3 we find that the spurious supramolecular protein aggregation present in Martini 2.2P multimer simulations is alleviated and thus this force field may be more suitable for the study of protein oligomerization. We propose that the use of enhanced sampling simulations with a refined coarse-grained force field and appropriately defined collective variables is a robust approach for studying the protein dimerization process, although one should be cautious of the ranking of energy minima.

## 1. INTRODUCTION

Protein-protein association plays a fundamental role in the regulation of protein function ranging from enzyme catalysis and inhibition to immune response by cytokines and DNA repair; hence, protein-protein association may control various cellular processes such as metabolism, signal transduction, and proliferation.<sup>1</sup> Understanding the mechanistic contributions of protein oligomerization/dimerization to regulating protein function would allow us to elucidate how these supramolecular complexes orchestrate biological responses and could ultimately lead to new approaches in designing new therapeutic agents.

Determining the association pathways and structures of protein complexes is the first step in unveiling their biological implications in the cellular machinery. The structure of protein complexes can be determined through biophysical methods, which provide either distinct microscopic snapshots of these systems or average properties of their interactions, kinetics, or energetics. Nonetheless, protein complex structural determination poses a significant number of challenges,<sup>2</sup> leaving a large number of known, functional protein-protein structures unresolved. Protein-association mechanisms are even more demanding to determine and require the use of specialized experimental approaches such as double-mutant cycles and paramagnetic relaxation enhancement to acquire potential transition states and intermediates.<sup>3,4</sup>

Although valuable, experimental studies aimed at determining the exact physicochemical interactions governing the process of formation of such protein complexes very often lack sufficient resolution for investigating the detailed underlying molecular interactions involved, such as non-bonded enthalpic and entropic effects between proteins and between protein and solvent, which occur after diffusion and spontaneous collision of the monomers and their subsequent association to the native complex. Thus, obtaining atomic-level information for

protein-association pathways is still an open problem. A wealth of computational approaches has attempted to underpin this information for protein-protein complexes.<sup>5</sup> Among these approaches, Molecular Dynamics (MD) simulations offer structural, dynamical, energetic, and ultimately also kinetic information of the relevant interactions at the atomic level. In practice, however, it has proven difficult to study protein-protein association and dissociation with MD simulations because the reversible protein association/dissociation is a slow process compared to the time scales accessible by conventional MD simulations. In fact, protein dimerization prerequisites the formation and destruction of interactions at the protein-protein desolvation interface, introducing slow degrees of freedom along the free energy landscape toward the native state; these slow processes may often trap protein monomers in non-native states with barriers that are difficult to overcome in time scales achievable by plain MD simulations.<sup>6,7</sup>

To overcome this problem, one may resort to enhanced sampling techniques such as long-time-scale MD simulations in combination with tempered binding,<sup>8</sup> calculations along a predefined reaction pathway,<sup>6,9,10</sup> or performing multiple short simulations and Markov State modeling.<sup>11</sup> Recent studies have succeeded in obtaining dimerization pathways in atomistic detail for the ribonuclease barnase and its inhibitor barstar<sup>11</sup> and for five different protein complexes (ribonuclease barnase and its inhibitor barstar, insulin dimer, H-Ras/Raf[RBD] complex, RNase HI-SSB-Ct complex, and the TYK2-Pseudokinase dimer).<sup>8</sup> Yet, these studies required extensive computational time to sample the configurational space of such transitions in atomic-level detail and even specialized computer architectures in some cases.<sup>8</sup> Therefore, full configurational sampling of protein dimerization including all metastable states in atomistic description is difficult to accomplish routinely. To tackle this issue, Coarse Grained (CG) representations of protein degrees of freedom can be employed to accelerate the simulations and

smoothen the free energy landscape in combination with enhanced sampling techniques to study protein association. The combination of these techniques may provide a reasonable approximation of the output of a computationally demanding atomistic simulation, given the time scales typically related with protein association processes (milliseconds or longer), which occur due to the conformational changes of the encounter complex.<sup>12</sup>

The Martini force field is the most widely used for the CG representation of biomolecular systems.<sup>13, 14</sup> Martini has been successfully applied in a variety of studies including membrane reorganization, transmembrane oligomerization, nanoparticle self-assembly and many others.<sup>15-19</sup> Until recently, however, only a few applications of Martini force field to investigate protein-protein interactions in aqueous solution appeared in the literature,<sup>20</sup> perhaps owing to the fact that Martini may overestimate protein-protein associations in water<sup>21</sup> and in some cases also in membranes.<sup>22</sup> Toward improving non-selective protein aggregation, the Martini development group released the open Martini 3.0.beta.3.2 (henceforth referred to as “Martini 3”) force field version,<sup>20, 23, 24</sup> encouraging the scientific community to provide feedback prior to the official release of Martini 3. The Martini 3 beta version has been used in recent studies in the literature. In the first study, a protein allosteric pathway in combination with a Go-like model was investigated.<sup>25</sup> In another study, Martini 3 was used to examine the binding of the S-component of the folate-specific energy coupling factor (ECF) transporter to its ECF module.<sup>26</sup> More recently, Martini 3 was used to simulate the transport of a cysteine-lipid moiety mediated by the LolA and LolB lipoprotein carrier proteins from the bacterial inner to the outer membrane across the periplasm.<sup>27</sup> The associated energetics for the extraction of the moiety from the LolA protein that were calculated in this study were along the same line as in an atomistic study, where a different lipoprotein was extracted from the LolA, although there was no direct comparison to

any structural or thermodynamic data from previous experimental or atomistic simulations.<sup>28</sup> In another study, the conformations of the three-domain TIA-1 protein in solution were analyzed with small-angle X-ray (SAXS) and neutron scattering (SANS) and with MD simulations using Martini 3 CG simulations.<sup>29</sup> The obtained results displayed poor agreement between experiment and simulation, which, however, was improved by a uniform increase (6% of the original interaction strength) of the protein-water interactions described in the Martini 3 force field. Vazquez-Salazar *et al.*<sup>30</sup> parameterized and consequently studied ionic liquids with the Martini 3 open beta scheme. Finally, Srinivasan *et al.*<sup>31</sup> evaluated Martini 3 efficiency in studying peripheral protein-membrane interactions. Nevertheless, Martini 3 has not been widely tested yet against a broad set of structurally diverse protein-protein systems in aqueous and membrane environments.

In this paper, we evaluate the Martini 2.2P<sup>32, 33</sup> and Martini open-beta 3<sup>20, 23, 24</sup> force fields to reproduce the formation of three different dimer complexes, i.e., barnase-barstar, insulin dimer, and the H-Ras/Raf[RBD] dimer, to assess the ability of the two force fields to reproduce protein dimerization in aqueous solution with reasonable accuracy and throughput. We study protein association and dissociation processes in water by investigating the free energy surface using Parallel Tempering MetaDynamics in the Well-Tempered Ensemble (PT-MetaD-WTE) calculations with respect to two collective variables (CVs). Moreover, unbiased MD simulations were performed in order to investigate the association and dissociation processes of the insulin dimerization. To validate our simulations, we also study the dimerization of two transmembrane proteins, which have been previously studied both experimentally and computationally, and compare to the available experimental and CG simulation results. Our results show that the Martini 2.2P force field systematically overestimates protein dimerization barriers both in the

membrane environment and in solution. Martini 3 predicts the structures and protein dimerization barriers for the two transmembrane systems that we studied with reasonable accuracy compared to previous computational and experimental results, but underestimates dimerization barriers for solvated proteins. Furthermore, we find that the PT-MetaD-WTE scheme in conjunction with CG simulations and an appropriate choice of collective variables is an efficient method to sample the free energy landscape of protein dimerization both in membrane and in aqueous environments achieving a balance between computational cost and accuracy.

## 2. RESULTS

To calculate the dimer association free energy landscape for the insulin, barnase-barstar, and H-Ras-Raf[RBD] dimers, as well as for the transmembrane Ephrin type-A receptor I (EphA1) and rhodopsin dimers with Martini 2.2P and Martini 3, we use PT-MetaD-WTE with two CV sets as a) the distance between the center of mass of the  $C\alpha$  atoms of each monomer and b) the root mean squared deviation (RMSD) of the  $C\alpha$  atoms from the experimentally-determined structure. Potential of mean force (PMF) calculations are employed to estimate the free energy surface of protein dimerization (see the Methods section). Protein structures residing at minima across this profile are back-mapped to atomistic resolution to analyze the dimer contact interfaces. Performed simulations are summarized in Table 1 and free energy of association values in Table 2. Convergence of the simulations was assessed by monitoring the diffusion of the CVs with respect to time (see Figures S2, S11, S12) as well as from the evolution of the 1D free energy profiles across the distance reaction coordinate with varying time increments (see Figures S1, S3, S5, S9, S10).

**Table 1.** Total simulation time for the unbiased and enhanced sampling MD simulations. Insulin dimer simulations were performed with Martini 3 with side chain dihedral corrections and without side chain dihedral corrections (see Methods). All other simulations with Martini 3 were run with side chain dihedral corrections. The Gaussian height (kcal/mol) and the biasing factor used for the PT-MetaD-WTE in each system is mentioned.

MD Simulations										
	Rhodopsin dimer		EphA1		Insulin dimer		H-Ras/Raf[RBD]		Barnase / Barstar	
	Time	Gaussian height (kcal/mol) / biasing factor	Time	Gaussian height (kcal/mol) / biasing factor	Time	Gaussian height (kcal/mol) / biasing factor	Time	Gaussian height (kcal/mol) / biasing factor	Time	Gaussian height (kcal/mol) / biasing factor
Martini 2.2P PT-MetaD-WTE	13.5 $\mu$ s	0.4/6	11 $\mu$ s	0.12/6	14.5 $\mu$ s	0.19/6	11 $\mu$ s	2.39/12	13.5 $\mu$ s	4.78/12
Martini 3 PT-MetaD-WTE	15.5 $\mu$ s	0.4/6	11 $\mu$ s	0.07/12	20 $\mu$ s	0.0012/2	11.5 $\mu$ s	0.024/2	17 $\mu$ s	0.24/2
Martini 3 Unbiased	-	-	-	-	15 $\mu$ s	-	-	-	-	-



## **Protein Dimerization in Membranes**

First, to validate our setup, we performed protein-association simulations of rhodopsin and the transmembrane segments of receptor tyrosine kinase EphA1 as these systems have been previously simulated with Martini 2.1/2.2/2.2P and have experimentally- or computationally-determined free energies of binding and structures.

### **Rhodopsin Dimerization**

The G protein-coupled receptor rhodopsin self-assembles into supramolecular structures in native bilayers. In a 2012 study, the dimerization of rhodopsin embedded in a lipid bilayer was studied with Martini 2.1.<sup>34</sup> The study identified the H1/H8-H1/H8 interface as the most probable interface, with a free energy of binding at -13 kcal/mol; additionally, four other protein interfaces were assessed for their free energy of binding: the H5-H5 interface was computed to have a binding free energy of -9 kcal/mol, the H4/H5-H4/H5 interface a binding free energy of -8 kcal/mol, and the H4-H6 and H4-H4 interfaces were estimated to have binding free energies of -1.2 and -2.4 kcal/mol, respectively.<sup>34</sup> The H1/H8-H1/H8 interface orientation of the rhodopsin dimer is consistent with early rhodopsin structural studies,<sup>35</sup> and also with the 2019 cryo-EM structures of the reconstituted rhodopsin dimer into nanodiscs, which confirmed that the H1/H8-H1/H8 is the physiologically relevant dimer interface.<sup>36</sup> Here, we reconstruct the free energy profile of rhodopsin dimerization using as CVs the RMSD from the experimentally-determined structure<sup>36</sup>(PDB ID: 6OFJ) and the distance of the center of mass of two monomers using Martini 2.2P and Martini 3.

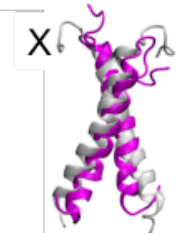
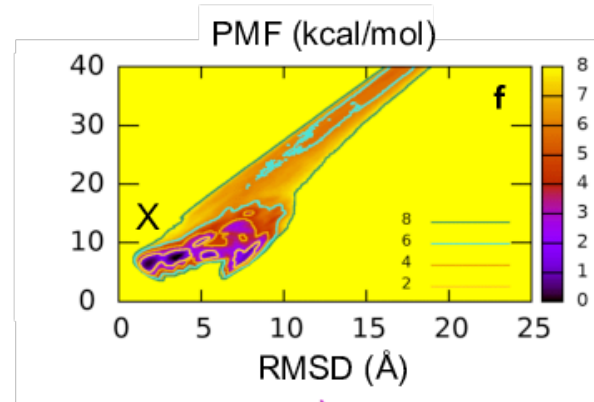
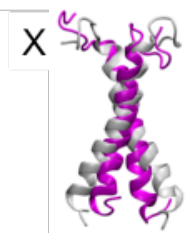
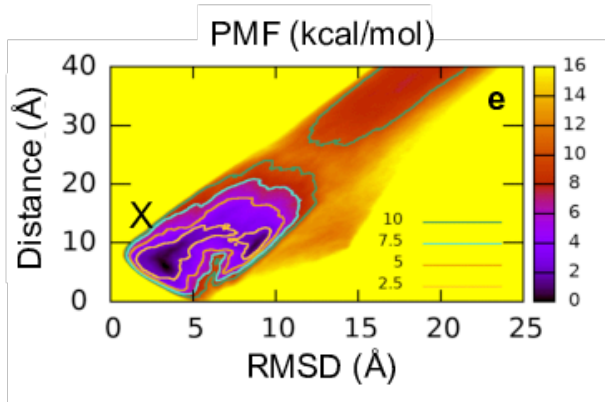
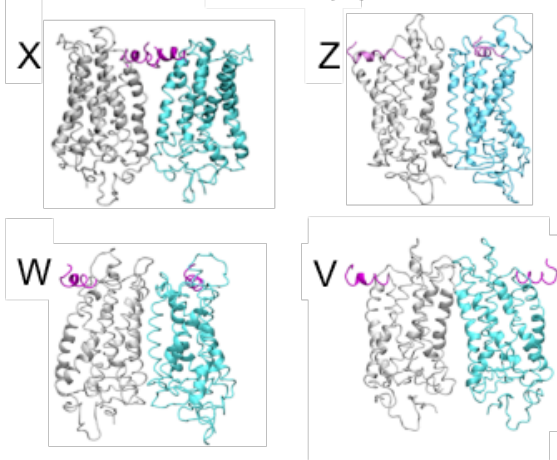
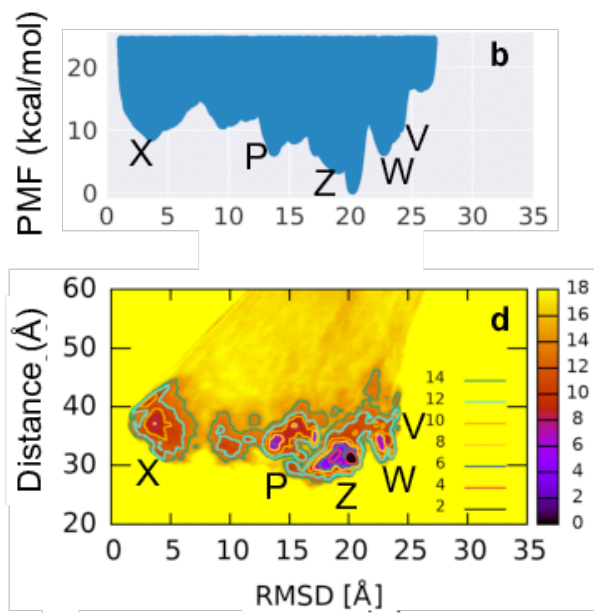
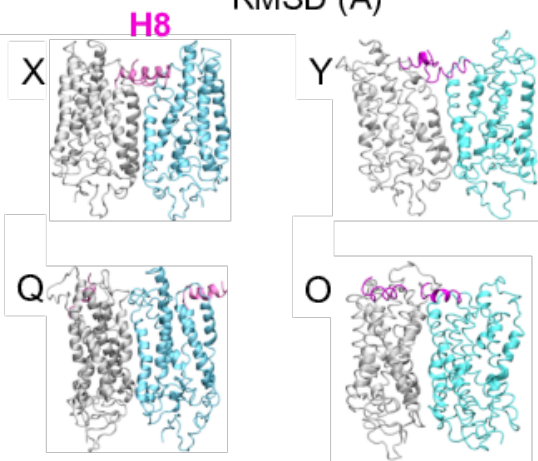
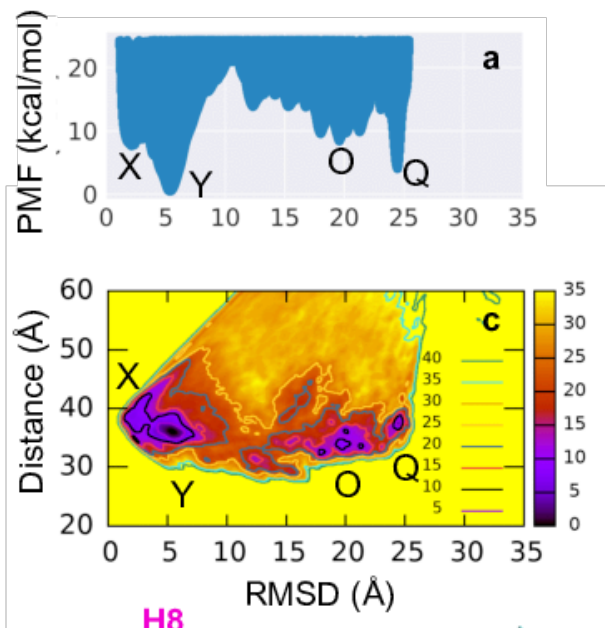
The 1D free energy plot as a function of RMSD for the rhodopsin dimerization calculated with Martini 2.2P is shown in Figure 1a and the 2D free energy surface in Figure 1c. The

minimum closest to the cryo-EM structure (structure X) is found at an RMSD of 2.1 Å from the cryo-EM structure and the center-of-mass distance between the two monomers at 38.8 Å. The interface for this structure is the experimentally-determined H1/H8-H1/H8 interface (Figure 1c, structure X), and the free energy of binding is estimated at -18.4 kcal/mol. The difference of 5.4 kcal/mol compared to the study of Ref.<sup>34</sup> could be attributed to the different force field version (here we use Martini 2.2P and Ref.<sup>34</sup> uses Martini 2.1); nonetheless, both force fields are able to identify the experimental structure correctly, although in Martini 2.2P this structure is not the global minimum. The dimerization energy for rhodopsin has not been yet experimentally measured. Another minimum at RMSD 5.3 Å (structure Y) corresponds to the global free energy minimum in this study (-25.6 kcal/mol) adopting the interface of H1/H8-H1/H8 albeit with different orientations of the H8 helices and a closer intermonomer distance of 36.1 Å. The structure at point O (RMSD = 19.6 Å and intermonomer distance of 34.2 Å) corresponds to a structure resembling the H4-H6 interface with an estimated energy of -17.7 kcal/mol, ~8 kcal/mol higher than the global minimum at the H1/H8-H1/H8 interface. The structure at point Q (RMSD = 24.2 Å and intermonomer distance = 37.5 Å) appears to be an intermediate structure between the H5-H5 and H4-H6 rhodopsin interfaces as characterized in Ref.<sup>34</sup>, having a relative orientation of the receptors of  $\varphi_1/\varphi_3 = 164/82$  degrees.

The 1D free energy plot with respect to RMSD (Figure 1b) and the 2D free energy surface of rhodopsin dimerization calculated with Martini 3 (Figure 1d) also reveals multiple minima. Structure X in Figures 1b,d corresponds to the experimentally resolved structure with RMSD = 3.6 Å and distance between the monomers 37.1 Å, although this structure is again not the global minimum on the free energy surface (-7 kcal/mol). Martini 3 estimates the global free energy minimum of dimerization at -15.4 kcal/mol, albeit for a structure with RMSD = 20.1 Å and distance 31.3 Å (Figure 1d, point Z). This structure has a relative orientation of the H8

helices of  $\phi_1/\phi_3 = -160/82$  degrees, and is an intermediate structure between the H4/H5-H4/H5 and H4-H6 rhodopsin interfaces as characterized in Ref.<sup>34</sup> Another minimum is located at point W, having a relative orientation of the H8 helices of  $\phi_1/\phi_3 = 164/82$  degrees and is an intermediate structure between the H5-H5 and H4-H6 rhodopsin interfaces as characterized in Ref.<sup>34</sup> and an RMSD of 22.7 Å from the PDB ID 6OFJ.<sup>36</sup> Next to W, a minimum is located at structure V, which relates to the H5-H5 interface<sup>34</sup> (-6.1 kcal/mol), which was also identified in the previous study using the Martini 2.1 force field.<sup>34</sup> Finally, a minimum at point P with RMSD = 13.8 Å, distance = 33.7 Å, and free energy = -9.4 kcal/mol is also identified without, however, resembling any of the interfaces as they were described in the previous study of rhodopsin association.<sup>34</sup> Parameters for all characterized minima are described in Table 3.

Convergence of these simulations was assessed by monitoring the evolution of the reaction coordinates with varying time increments (see Figure S1a,b) and the diffusion of the reaction coordinates with time (Figure S2a,b). The PMF with respect to the RMSD CV is shown in Figures S1c,d.



**Figure 1.** 1D free energy plots as a function of RMSD and 2D free energy surfaces for the association of transmembrane proteins along the two reaction coordinates (RMSD from the experimentally determined structure and center of mass distance between the two monomers). Structures that correspond to minima in the PMF are marked with capital letters and depicted in atomistic detail after backmapping. a) 1D free energy plot across the reaction coordinate of RMSD for the dimerization rhodopsin after applying entropic corrections (see equation (1)) using Martini 2.2P, and b) using Martini 3. c) 2D free energy surface of rhodopsin dimer association using Martini 2.2P. H8 is colored in purple and the letter X corresponds to the lowest RMSD structure with respect to the cryoEM structure. d) 2D free energy surface of rhodopsin dimer association using Martini 3. H8 is colored in purple. e) 2D free energy surface of EphA1 dimer association using Martini 2.2P. The backmapped atomistic structure is depicted in purple and is aligned with the crystallographic structure, which is depicted in silver. f) 2D free energy surface of EphA1 dimer association using Martini 3. The backmapped atomistic structure is depicted in purple and is aligned with the crystallographic structure, which is depicted in silver. The contour lines in the lower right of the graphs represent the energy barriers between the enclosed areas with the same corresponding energy. The barriers can be inferred from the contour lines as well as the 1D plots in the SI following entropic corrections and integration of the PMF as discussed in the Methods section. The final calculated barriers appear in Tables 2 and 3.

## EphA1 dimer

The transmembrane segments of the receptor tyrosine kinase EphA1 associate in a right-handed parallel  $\alpha$ -helical bundle through the N-terminal glycine zipper motif as determined by NMR (PDB ID: 2K1L).<sup>37</sup> The dimerization free energy for the EphA1 dimer was experimentally measured at -3.68 kcal/mol in a DMPC lipid bilayer using Förster resonance energy transfer (FRET).<sup>38</sup> Computational studies using the Martini 2.2 and Martini 2.2P force fields in a DMPC bilayer estimated the dimer binding free energies to be -7.14 kcal/mol and -8 kcal/mol, respectively, using umbrella sampling calculations across the distance of the center of mass of the two monomers.<sup>22</sup> Another study for the same system in a DPPC bilayer estimated the association free energy at -14.3 kcal/mol using the Martini 2.2 force field.<sup>39</sup> The difference in the free energy of association between the two studies may have resulted from the different acyl chain length of the chosen phospholipid, i.e., DPPC compared to DMPC.

Here, to study the dimerization pathway of EphA1, we employ the same conditions as in Ref.<sup>22, 38</sup> As CVs, we use the center-of-mass distance between the two monomers, also used in Ref.<sup>22</sup> and the RMSD from the NMR structure (PDB ID: 2K1L<sup>37</sup>). As seen in Figures 1e and S3a,b, Martini 2.2P converges to a free energy value of -7.2 kcal/mol in agreement with previous computational studies. Martini 3 converges to a free energy value of -4.9 kcal/mol, which is close to the experimentally determined value of -3.7 kcal/mol (Figure 1f). Martini 2.2P estimates the free energy minimum with the lowest RMSD from the NMR structure value at 3.3 Å (Figure 1e and S3c, structure X) and an intermonomer distance of 6 Å, while Martini 3 has an RMSD = 1.9-3.4 Å (Figure 1f and S3d, structure X) and an intermonomer distance of 6.6-7.5 Å, showing that both force fields are able to identify the experimentally-determined interface also in agreement with Ref.<sup>22</sup>, where both Martini 2.2 and Martini 2.2P force field identify the minimum

structure at an intermonomer distance of 7 Å for the two transmembrane monomers. The extensive broadness of the free energy well apparent in Martini 2.2P simulation compared to Martini 3 (Figure 1e) could be the result of excessive non-specific aggregation as the two monomers tilt against each other maintaining their interactions even above 25 Å, thus leading to profiles that are not fully reaching a plateau (Figure S3a). However, for Martini 3 the 1D free energy profile of the association across the distance reaction coordinate (Figure S3b) is flat when the distance is larger than 18 Å, compared to Martini 2.2P (Figure S3a). This result suggests that the long-range artificial overbinding between the two monomers is not present in the case of Martini 3.

**Table 2.** Estimated free energy of association for rhodopsin, EphA1, insulin, H-Ras/Raf[RBD] and barnase/barstar dimers. For the insulin simulations with Martini 3 two values are reported, (a) corresponds to Martini 3 with dihedral corrections and (b) to Martini 3 without dihedral corrections.

Estimated free energy of dimerization (kcal/mol) corresponding to the global minimum					
	Rhodopsin	EphA1	Insulin	H-Ras/ Raf[RBD]	Barnase / Barstar
Experimental	-	-3.68 from Ref. <sup>38</sup>	-7.2 from Ref. <sup>40</sup>	-9.6 from Ref. <sup>41</sup>	-19.2 from Ref. <sup>42</sup>
Previous simulations	-13.0 (Martini 2.1) from Ref. <sup>34</sup>	-7.14 (Martini 2.2) from Ref. <sup>22</sup> -8.0 (Martini 2.2P) from Ref. <sup>22</sup> -14.3 (Martini 2.1, modified) from Ref. <sup>39</sup>	-11.9 (atomistic simulation) from Ref. <sup>43</sup> -10.8 (atomistic simulation) from Ref. <sup>44</sup>	-	-
Martini 2.2P from PMF minima	-24.5 ± 0.3	-7.0 ± 0.2	-12.2 ± 0.2	-38.4 ± 0.2	-31.8 ± 0.2
Martini 2.2P from PMF integration	-25.6	-7.2	-13.1	-39.5	-32.3
Martini 3 from PMF minima	-14.6 ± 0.2	-5.4 ± 0.1	-2.5 ± 0.1 <sup>(a)</sup> -2.1 ± 0.1 <sup>(b)</sup>	-6.8 ± 0.1	-4.1 ± 0.1
Martini 3 from PMF integration	-15.4	-5.1	-4.1 <sup>(a)</sup> -6.0 <sup>(b)</sup>	-7.9	-5.6



**Table 3.** Estimated free energy of protein dimer complex, RMSD between the experimentally-determined structure, and inter-monomer distance for all characterized minima for the studied systems. Estimated free energy of association refers to values calculated after PMF integration (see the Methods section).

RMSD, Distance, and Estimated free energy of protein dimer complex for identified minima				
Dimer System	Structure	RMSD (Å)	Distance (Å)	Estimated free energy (kcal/mol)
Rhodopsin Martini 2.2P	X	2.1	38.8	-18.4
Rhodopsin Martini 2.2P	Y	5.3	36.1	-25.6
Rhodopsin Martini 2.2P	O	19.6	34.2	-17.7
Rhodopsin Martini 2.2P	Q	24.2	37.5	-23.9
Rhodopsin Martini 3	X	3.6	37.1	-7.0
Rhodopsin Martini 3	P	13.8	33.7	-9.4
Rhodopsin Martini 3	Z	20.1	31.3	-15.4
Rhodopsin Martini 3	W	22.7	33.7	-9.3
Rhodopsin Martini 3	V	23.7	32.3	-6.1
EphA1 Martini 2.2P	X	3.3	6	-7.2

EphA1 Martini 3	X	1.9-3.4	6.6-7.5	-4.9
Insulin Martini 2.2P	A	2.7	18.6	-8.7
Insulin Martini 2.2P	B	7.1	19.0	-13.1
Insulin Martini 2.2P	C	10.3-11.4	19.5-20.1	-12.4
Insulin Martini 2.2P	D	14.0	21.2	-10.1
Insulin Martini 3 without dihedral corrections	A	2.6	19.3	-4.6
Insulin Martini 3 without dihedral corrections	E	4.6	18.0	-4.8
Insulin Martini 3 without dihedral corrections	F	11.0	17.7	-6.0
Insulin Martini 3 with dihedral corrections	A	2.3	18.6	-4.1
Insulin Martini 3 with dihedral corrections	G	4.3	18.0	-3.6
Insulin Martini 3 with dihedral corrections	H	5.4	17.0	-3.4
Insulin Martini 3 with dihedral corrections	I	10.7	17.4	-3.1
H-Ras/Raf[RBD] Martini 2.2P	A	1.7-4.1	26.2-26.5	-35.8
H-Ras/Raf[RBD] Martini 2.2P	B	11.3	27.1	-39.5
H-Ras/Raf[RBD]	C	13.7	24.0	-39.1

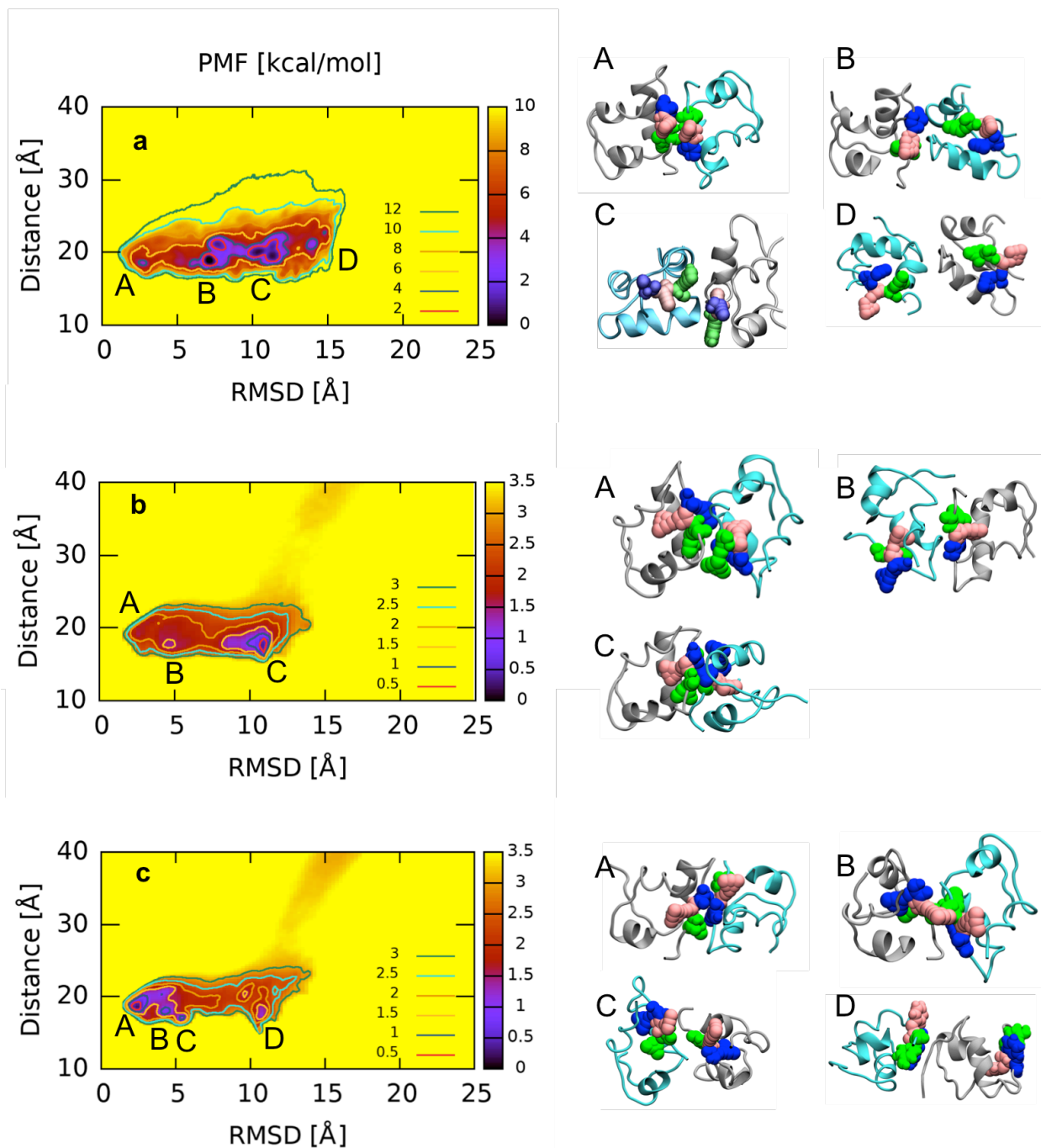
Martini 2.2P				
H-Ras/Raf[RBD] Martini 3	A	5.5	26.5	-7.1
H-Ras/Raf[RBD] Martini 3	D	18.4	24.1	-7.9
Barnase-Barstar Martini 2.2P	A	2.7	22.7	-32.3
Barnase-Barstar Martini 2.2P	B	10.3	20.8	-28.9
Barnase-Barstar Martini 2.2P	C	17.7	26.5	-25.4
Barnase-Barstar Martini 3	D	11.5	22.4	-5.6

## Protein Dimerization in Aqueous Solution

### Insulin Dimer

The bioactive forms of insulin include monomeric, dimeric, and hexameric states. The dimerization process has been studied both experimentally<sup>40, 45</sup> and computationally<sup>43, 44, 46</sup> because the dimer structure plays an important role as an intermediate step in insulin metabolism.<sup>47</sup> The insulin monomer consists of 51 amino acids organized in two chains, A and B. The B chains of two monomers associate into a dimer through an antiparallel intermolecular  $\beta$ -sheet formation, which is stabilized by four intermolecular hydrogen bonds. Moreover, hydrophobic interactions among Phe24, Phe25, and Tyr26 stabilize the dimeric structure excluding water from the interface (Figure S4). The binding free energy of the dimer has been experimentally determined at -7.2 kcal/mol<sup>40</sup> and computational studies using atomistic models have estimated this barrier to be -11.9 kcal/mol<sup>43</sup> using MM-GBSA calculations, and -10.8 kcal/mol using well-tempered metadynamics with parallel tempering.<sup>44</sup>

The free energy surface of insulin dimerization across the two reaction coordinates using the Martini 2.2P force field is shown in Figure 2a. The estimated free energy of association at -8.7 kcal/mol for structure A, which is the experimentally-determined structure (PDB ID: 4INS<sup>48</sup>) with an RMSD=2.7 Å is close to the experimental free energy of association of -7.2 kcal/mol, albeit it does not correspond to the global free energy minimum on the PMF. The global minimum, structure B, is observed when the center-of-mass distance of the two monomers is at 19.0 Å. Two other minima with higher free energies of association are estimated at RMSDs of 10.3-11.4 Å (point C, Figure 2a) and 14 Å (point D, Figure 2a), all at a intermonomer center-of-mass distance of ~19-21.5 Å. The free energy as a function of the distance CV was calculated (Figure S5a) and as a function of the RMSD CV (Figure S5b) in order to monitor the thermodynamic relevant regions and the convergence of the system.



**Figure 2.** Free energy surfaces for insulin dimer association along the two reaction coordinates (RMSD from the experimentally determined structure and center-of-mass distance between the two monomers). Structures that correspond to main minima in the graphs are marked with capital letters and depicted in atomistic detail after backmapping. The two insulin monomers are shown in cartoon (gray and cyan) and

the aromatic residues of the antiparallel  $\beta$ -sheet are depicted in van der Waals representations: Phe25 (pink), Phe24 (blue), Tyr26 (green). a) Calculation with Martini 2.2P and the four identified minima A, B, C, D. b) Calculation with Martini 3 without using side chain dihedral corrections and the three identified minima A, E, F. c) Calculation with Martini 3 using side chain dihedral corrections and the four identified minima A, G, H, I. The contour lines in the lower right of the graphs represent the energy barriers between the enclosed areas with the same corresponding energy. The barriers can be inferred from the contour lines as well as the 1D plots in the SI following entropic corrections and integration of the PMF as discussed in the Methods section. The final calculated barriers appear in Tables 2 and 3.

To study the insulin dimer association with Martini 3, we used two schemes implemented in Martini 3. In the first scheme, additional dihedral potential terms that are fitted to the atomistic crystallographic structure are added to Martini 3 as side chain dihedral corrections (see also Methods section). In the second scheme, these dihedral corrections are not present. The Martini 3 force field using the dihedral corrections results in a global minimum of -4.1 kcal/mol for the insulin dimerization (Figure 2c) for structure A, which corresponds to the experimentally determined structure with an RMSD = 2.3 Å and an intermonomer distance of 18.6 Å. Another three minima (G, H, I) are identified with slightly higher energies (See Table 3), albeit with higher RMSD values ranging from 4.3-10.7 Å. Although the interface interactions are not the ones identified in the experimental structure, the intermonomer distance between these structures are closer (17-18 Å) than the experimental structure (18.5 Å). The Martini 3 calculation without the dihedral corrections results into a global minimum of -6 kcal/mol, although this structure (F, Figure 2b) is not the experimental structure. The experimental structure is identified as structure A (RMSD = 2.6 Å) with a free energy of -4.6 kcal/mol, close to another identified minimum

(structure E) with -4.8 kcal/mol (RMSD = 4.6 Å). In both calculations with Martini 3, the free energies of insulin association are lower compared to the experimentally determined free energy of binding of -7.2 kcal/mol.<sup>40</sup> The multiple identified minima could represent the intermolecular conformational transitions during the dimerization process. However, the free energy changes between these states are estimated to be very low in the order of 2.5 kcal/mol, which means that at room temperature these transitions will happen at very fast timescales. The discretization of the local minima observed on the free energy surface at RMSD < 6 Å is not so evident when using the Martini 3 without the side chain dihedral corrections. To assess the convergence of the free energy estimates with Martini 3, the potential of mean force (PMF) is plotted as a function of time across each of the reaction coordinates (Figures S5c,d,e,f).

The local minima on the PMF surface obtained using Martini 3 with side chain dihedral corrections resemble those described by Banerjee *et al.*,<sup>44</sup> where the association and dissociation mechanisms are driven by microscopic structural rearrangements. These rearrangements are based on the change of distance between the aromatic residues at the interface of the intermolecular antiparallel  $\beta$ -sheet (Figure S4) along the dimerization pathway and correspond to intermediate states, where the three aromatic residues at the interface rearrange their conformations interacting with a different partner in each state. Indeed, structure A in Figure 2c resembles the initial state of Ref.<sup>44</sup>, where the Tyr26-Tyr26 distance is the longest and the Tyr26-Phe24 distance is the shortest. In addition, structures G and H shown in Figure 2c resemble states where the two Tyr-26 are in close proximity, with the Phe25-Phe25 distance populating two states: one at 7.4 Å and the other at 11 Å. In all cases, however, the transition free energies here (~ -0.5-1 kcal/mol) are underestimated with respect to the aforementioned atomistic study (~ -3.6-7 kcal/mol). Two other small basins populated by structure I in Figure 2c and F in Figure 2b

reveal two different insulin dimer conformations, where the interface of the dimer is smaller and the interactions are not mediated by intermolecular aromatic contacts.

To obtain a better correspondence to the study of Ref.<sup>44</sup>, we ran a validation calculation of the free energy of the insulin dimer association using the same collective variables as in Ref.<sup>44</sup>, which are a) the number of contacts between the C $\alpha$  of the two monomers ( $N_{MM}$ ) and b) the distance of the center of mass of the two monomers using our PT-MetaD-WTE settings (see Methods). The obtained free energy surfaces (Figures S6a,b) show a minimum at a distance of  $\sim 18$  Å and a  $N_{MM} \sim 40$  in agreement with Banerjee et al.<sup>44</sup>, who calculated a distance of  $\sim 18$  Å and a  $N_{MM} \sim 50$ . Regarding its shape, the 2D free energy profile shows an excellent correspondence to the atomistic study, however the free energy of association that Martini 3 estimates at  $\sim -3.2$  kcal/mol is lower compared to Ref.<sup>44</sup>, which calculates a value of  $-10.8$  kcal/mol.

To obtain an unbiased rough description of the association/dissociation procedure, we also performed unbiased MD simulations using Martini 3. The dimer conformations obtained after 15  $\mu$ s of simulations were aligned to the crystallographic structure (PDB ID: 4INS<sup>48</sup>), and their corresponding RMSD values were measured together with the probability density of these RMSD values for both simulations. Figure S7 shows that the probability density of the RMSD for the unbiased MD simulations with Martini 3 exhibits different distributions depending on whether dihedral corrections are added to the potential energy function of the amino acid side chains. Martini 3 with the dihedral side chain corrections results in lower RMSD values (RMSD  $< 4$  Å) compared to Martini 3 without dihedral side chain corrections. This implies that in order to capture conformations close to the crystallographic structure, the dihedral potentials, which bias the side chain orientations to be similar to the crystallographic side chain orientations, are



needed. The addition of these dihedral corrections results in the exploration of more relevant conformational space, which is related to intermediate structures of the association pathway. For a more extended discussion on the unbiased simulations with Martini 3, see the SI.

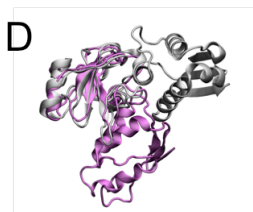
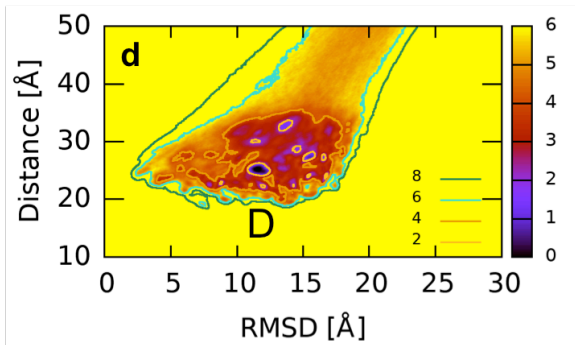
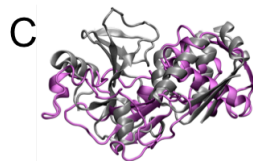
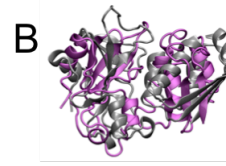
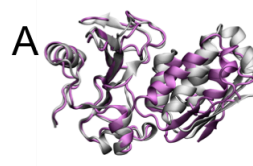
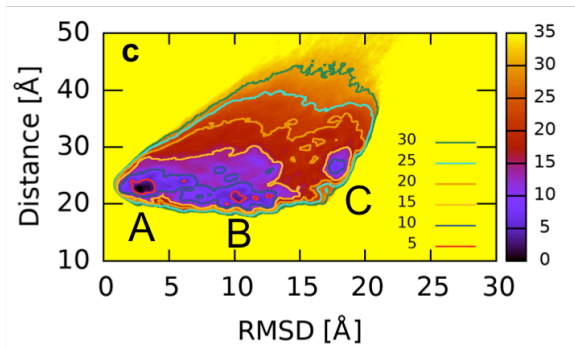
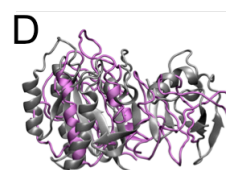
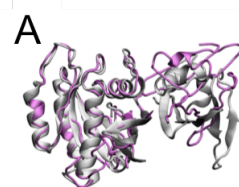
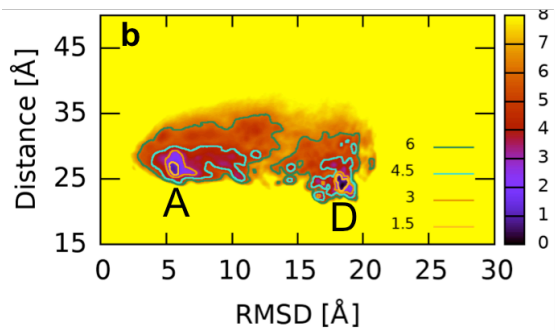
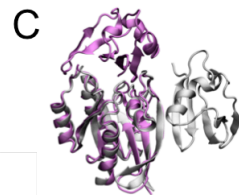
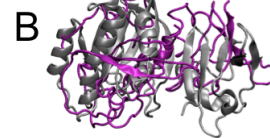
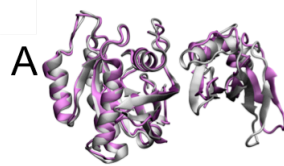
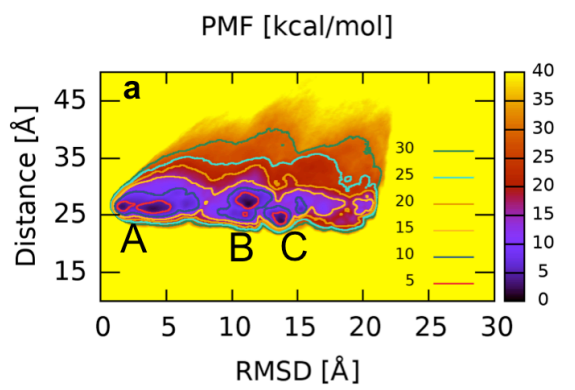
Martini 3 simulations without the side chain dihedral corrections result in a global minimum at RMSD 11 Å (structure F, Figure 2b), which coincides with the peak of the distribution from the unbiased simulations (Figure S7). In the absence of the side chain dihedral corrections, Martini 3 fails to estimate the global minimum at the crystallographic structure because it engages in non-specific aromatic packing that result in structure F. The two local minima of the unbiased simulation correspond to structures A and E from Figure 2b, where the interface is occupied by aromatic residue packing (Figure S8).

### **H-Ras/Raf[RBD] dimer**

Next, we examine the association pathways for the human H-Ras and the H-Ras-binding domain (RBD) of the C-Raf1 complex. H-Ras is a GTP-hydrolyzing protein that acts as central switch in the regulation of cell proliferation and differentiation.<sup>49</sup> After its activation, H-Ras activates a cascade of protein kinases the first of which is the downstream effector Raf. The Raf[RBD] domain assumes the classic ubiquitin fold, with five  $\beta$ -strands and two  $\alpha$ -helices connected by loops. In its complex with H-Ras, the backbone hydrogen bonds among  $\beta$ 2 of H-Ras and  $\alpha$ 2 of Raf-RBD form an extended intermolecular  $\beta$  sheet. In addition to the  $\beta$  sheet interactions that form the interface in the H-Ras/Raf[RBD] complex, there are significant interactions between Raf-RBD  $\alpha$ 1 residues Lys84, Val88, and Arg89 with switch I residues of H-Ras. The free energy of the complex was experimentally determined to be -9.6 kcal/mol<sup>41</sup> and multiple structures have been resolved (here, we use PDB ID: 4G0N).<sup>50</sup>

The free energy surface across the two CVs used in this study was calculated using Martini 2.2P and Martini 3 using PT-MetaD-WTE (Figures 3a,b). The one-dimensional free energy profiles for Martini 2.2P and Martini 3 along the two CVs are shown in Figures S9a,b,c,d. As can be seen, the free energy is converged after 9.5  $\mu$ s at -38.4 kcal/mol using the PMF minimum and -39.5 kcal/mol using PMF integration for Martini 2.2P. The global minimum of the PMF corresponds to structure B at an RMSD = 11.3 Å, close also to structure C with an RMSD = 13.7 Å and free energy = -39.1 kcal/mol (Figure 3a). Martini 2.2P identifies the experimental structure as a wide minimum in point A of the PMF with an RMSD = 1.7-4.1 Å and intermonomer distance of 26.2-26.5 Å. The estimated free energy of -35.8 kcal/mol is significantly higher than the experimental value (-9.6 kcal/mol). The free energy surface of the H-Ras/Raf[RBD] association using the Martini 3 force field is shown in Figure 3b showing two minima. Structure A has an estimated free energy of association equal to -7.9 kcal/mol, which is close to the experimentally-reported value, although the structure has an RMSD = 5.5 Å compared to the experimental one. The D structure of the Martini 3 Ras-Raf[RBD] exhibits the same interface of structure B in Martini 2.2P ( $\alpha$ 2 helix of Raf[RBD] and  $\alpha$ 5- $\alpha$ 4 helices of H-Ras) albeit the orientation of  $\alpha$ 3- $\alpha$ 4- $\alpha$ 5 H-Ras helices is different from the crystallographic structure, thus having an RMSD 18.4 Å.

Overall, Martini 2.2P estimates the global minimum at a structure, which is very close to the experimental interface; however, it overestimates the free energy of association possibly due to an increased hydrogen bonding strength between the  $\beta$ 2 strands of RBD and H-Ras. On the other hand, Martini 3 does not favor the formation of this interface and overall the estimated energy levels are lower than the experimental values.



**Figure 3.** Free energy surfaces for H-Ras/Raf[RBD] and the barnase-barstar complex association along the two CVs (RMSD compared to the experimental structure and center-of-mass distance between the two monomers). Structures that correspond to main minima in the graphs are marked with capital letters and depicted in atomistic detail after backmapping. The crystallographic structure is colored with silver and the calculated one with purple. a) Calculation of H-Ras/Raf[RBD] association with Martini 2.2P. b) Calculation of H-Ras/Raf[RBD] association with Martini 3. Structure A corresponds to the experimental structure. In structure B, the dimer interactions are mediated mainly through the  $\alpha 2$  helix of Raf[RBD] and  $\alpha 5$ - $\alpha 4$  helices of H-Ras. Interestingly, in structure C although Raf is flipped, the beta sheet of each monomer contributes to the intermonomer antiparallel beta sheet as in the crystallographic structure. c) Calculation of barnase-barstar association with Martini 2.2P. d) Calculation of barnase-barstar association with Martini 3. The contour lines in the lower right of the graphs represent the energy barriers between the enclosed areas with the same corresponding energy. The barriers can be inferred from the contour lines as well as the 1D plots in the SI following entropic corrections and integration of the PMF as discussed in the Methods section. The final calculated barriers appear in Tables 2 and 3.

### **Barnase-Barstar dimer**

The interaction of barnase, an extracellular RNase of *bacillus amyloliquefaciens* with its intracellular polypeptide inhibitor barstar has been widely studied as a paradigm for protein-protein interactions. The complex formed by barnase and barstar is one of the tightest protein complexes known, with an experimentally-determined free energy of association at -19.2 kcal/mol.<sup>42</sup> Martini 2.2P force field estimates a free energy of association for the barnase-barstar dimer at -32.3 kcal/mol, which is 13.1 kcal/mol higher compared to the experimental value (-19.2 kcal/mol). The global minimum on the free energy surface (Figure 3c) corresponds to the crystallographic structure (structure A; RMSD = 2.7 Å, free energy = 32.3 kcal/mol) and another

minimum at structure B has an RMSD = 10.3 Å and free energy = -28.9 kcal/mol. The free energy profile along the two CVs with Martini 3 (Figure 3d) estimates a free energy of binding at -5.6 kcal/mol, which is 13.6 kcal/mol lower compared to the experimental one. Another minimum, structure C, is located at RMSD = 17.7 Å with an estimated free energy of -25.4 kcal/mol. The Martini 3 force field in this case is estimating the global minimum on the dimerization at structure D on the free energy profile (Figure 3d) with RMSD at 11.5 Å and an intermonomer distance of 22.4 Å. While Martini 3 underestimates the barnase-barstar binding free energy and does not approximate the crystallographic structure, Martini 2.2P was successful in estimating the experimental interface, albeit with a much higher estimated free energy of association. The one-dimensional free energy profiles for Martini 2.2P and Martini 3 along the two CVs are shown in Figures S10,a,b,c,d.

### 3. DISCUSSION AND CONCLUSIONS

In the realm of the computational ‘microscope’, atomistic resolution simulations provide a rigorous way to study biological processes at the molecular level. However, the spatiotemporal range of application of MD simulations at this resolution is limited from the underlying most computationally intensive algorithmic requirements. Ongoing algorithmic advances coupled with the design of novel computation processing unit architectures realized the routine production of MD simulations at the  $\mu\text{m}$  length scale and  $\mu\text{s}$  time scale. Nevertheless, expanding the boundaries to mesoscopic systems often requires simplified models.

The use of CG Hamiltonians can provide such a solution. Along this objective, the CG Martini force field has been successfully applied to simulations of a wide range of biomolecular processes.<sup>20, 51-53</sup> Although it was initially designed to model lipid systems, continuous efforts

expanded its applicability to protein-membrane systems, carbohydrates, polymers, and nanoparticles, while preserving its invaluable transferability. Recently, Martini 2.2P reparameterization led to 2.3P,<sup>54</sup> an improvement toward studying peripheral protein membrane binding by capturing cation- $\pi$  interactions. These achievements highlight the readiness of the developmental group to overcome the force field's inherent pitfalls as a response to the strong interest of the community. For instance, a known caveat of the Martini model is the overestimation of the protein dimerization free energies compared to corresponding experimental values by several kcal/mol both in aqueous solution and in the membrane.<sup>21, 22</sup> When dimerization free energies in simulation models are significantly overestimated, they may result in instantaneous, nonselective, and irreversible binding, thereby leading to the formation of unrealistic protein oligomers or overaggregation. This overestimation in the free energy of dimerization further complicates the interpretation of simulation data for dimerization interfaces: if the binding between two proteins is disproportionately strong and fast, then the dimerization interface predicted by CG simulations may be inaccurate. Also, aggregates resulting from excessive dimerization free energies may give rise to unrealistically pronounced confinement effects for lipids and proteins diffusing in the membrane plane, thereby impairing the understanding of membrane lateral dynamics and organization. The Martini 3 open beta version, which is currently under development, aims at dealing with this issue. Indeed, the reparameterization of the LJ potential in a bead-size-dependent manner as well as the careful consideration of the corresponding bond lengths could result in the appropriate bead solvation free energy, thus leading to a reduction of artificially high dimerization energies.

In this study, using Martini 3 in comparison to the Martini 2.2P force field in various aqueous and transmembrane systems, we observed noticeable differences on the estimation of the

dimerization free energy profiles resulting from these two different force field versions. Martini 3 reparameterization was performed with an eye to adjust overestimated dimerization barriers using Martini 2.2P. However, we observe that the estimated energies now underestimate the experimentally-calculated free energy of dimerization by a considerable amount in all the cases that we studied with the exception of the EphA1 dimer (Table 2). The improved solvation model achieved in Martini 3, however, allows the identification of more discreet local minima on the dimerization profile.

The diversity of the systems chosen in this study in terms of the dimerization driving forces enables a decomposition of the energetic terms that are more influential in each force field. The transmembrane protein dimerization of the two systems studied here exhibits a good overall consistency with both force fields. The binding interface of rhodopsin was estimated in a 2012 study<sup>34</sup> with Martini 2.2P, in remarkable agreement with the experimental data. In this study, we were able to obtain the same interface as in previous study and the experimental cryoEM structure (H1/H8-H1/H8) with both Martini 2.2P and Martini 3. The global minimum for the free energy of rhodopsin association estimated here (point Y, Figure 1c) with Martini 2.2P is different from and approximately 13 kcal/mol lower compared to the study of Ref.<sup>34</sup>, possibly due to the difference in the force field version (Martini 2.1 vs 2.2P). The minimum corresponding to the cryoEM structure is identified as the third lowest minimum (point X, Figure 1c). Other identified minima correspond to other low-energy dimerization interfaces observed in Ref.<sup>34</sup>. Multiple local minima were also obtained with Martini 3. Martini 3 captures the experimentally evaluated interface as a local minimum, and again the global minimum and other local minima correspond to conformations described in Ref.<sup>34</sup> as low-energy dimerization interfaces, which facilitate higher-order architectures of the rhodopsin assembly (Figure 1d).<sup>55</sup>

The free energy of dimerization estimated with Martini 3 for the experimental structure (point X, -7 kcal/mol) is, however, sufficiently higher, demonstrating the attention given to the excessive aggregation issues of the former force field version. Nonetheless, in the absence of the experimentally determined value, the ability of the force field in determining the rhodopsin dimerization free energy remains undetermined for this particular case. The dimer interface of the EphA1 protein as estimated from both force field versions as a global minimum was in close proximity to the corresponding NMR structure. The free energy profile computed with Martini 3 displayed improved results compared to Martini 2.2P, and estimates a higher free energy of binding (-4.9 kcal/mol, Table 3) closer to the experimental one (-3.68 kcal/mol, Table 2). Given the above, the reparametrization of Martini advances the studies of transmembrane protein interactions.

For insulin dimerization, Martini 3 was able to capture several microstates close to the experimentally determined global minimum. These states had been previously characterized in an atomistic-resolution study and correspond to intermediate states during the dimerization procedure. Martini 3 with dihedral corrections successfully estimates the global minimum of the dimerization free energy profile to a corresponding structure close to the experimentally determined one in all cases, however, it underestimates the free energy of association compared to the experimental value (-4.1 vs -7.2 kcal/mol). Martini 2.2P also estimated the experimentally-determined structure of the insulin dimer as a minimum in the PMF albeit not as the lowest energy structure of the four identified minima. The free energy of H-Ras/Raf[RBD] dimerization studied with Martini 2.2P was overestimated at -39.5 kcal/mol compared to the experimental value of -9.6 kcal/mol; also the experimental structure was identified as the third minimum in ranking. The free energy profile using Martini 3 displays a minimum in a distinct area compared



to the experimental binding interface with an estimated energy, which is smaller than the experimental value (-7.1 kcal/mol, RMSD = 5.5 Å), but much closer compared to the Martini 2.2P value. For the barnase/barstar system, Martini 2.2P overestimates the free energy of dimerization compared to the experimental value (-32.3 kcal/mol vs -19.2 kcal/mol) but correctly identifies the experimental structure as the global minimum at an RMSD = 2.7 Å (point A, Figure 3c). Martini 3 for the barnase/barstar again underestimates the association free energy (-5.6 kcal/mol) predicting a global minimum with significant divergence from the experimental crystal structure (RMSD = 11.5 Å). The dimer interfaces of the H-Ras/Raf[RBD] and barnase/barstar are governed mainly by strong electrostatic interactions (salt bridges) and especially in the case of H-Ras/Raf[RBD] by specific hydrogen bonds.<sup>50, 56</sup> The success of Martini 3 in the case of insulin dimer, which is dictated mainly by the proper formation of hydrogen bonds, suggests that erroneous electrostatic interactions could cause the deficiency of Martini 3 in obtaining the correct global minimum in these cases. The absence of the polarizability scheme in Martini 3 could result in this discrepancy because of the absence of explicit screening of electrostatics.

From previous published works as well as from the released notes of the Martini 3 beta release, the developers aim at a totally reparameterized force field to compensate for bead-size-dependent Lennard-Jones cross interactions. The parameterized tiny-type beads are used in the aromatic rings in order to reproduce realistic stacking distances. These tiny beads have their effective size ( $\sigma$ ) of the Lennard-jones parameters, describing their non-bonded interactions, reduced from the typical 0.47 to 0.32 nm. Indeed, in our example of insulin dimerization (where the Phe24, Phe25 and Tyr26 bear the tiny type of beads on their side chains), the Martini 3 predicts discrete minima that correspond to different packings of the Phe side chains. However,

the predicted dimerization energies are higher than the experimental energy and the energies calculated from previous computational studies. Using Martini 3, different microstates can be observed during the dissociation of the insulin dimer (multiple minima with RMSD  $< 6 \text{ \AA}$ ) (Figure 2c). Interestingly, all of these minima have higher energy compared to experiment. Hence, even though the size of the beads allows for adequate solvation via the solute-solvent interactions, the solute-solute interactions of the tiny beads seem to be weak. Moreover, a new water model, represented by a single bead, was also introduced in Martini 3 with the aim to improve the desolvation of the overstacking of the new bead types. The reduction of the Lennard-Jones  $\epsilon$  parameter for the cross interactions of these tiny beads could result in the higher energy observed. The successful identification of the global minimum of the insulin dimer in the case of Martini 3 indicates that the hydrogen bonding is well represented in this new version, since this type of directional interactions would have been crucial for obtaining such small RMSD value ( $< 2 \text{ \AA}$ ). On the contrary, even in the absence of the side-chain dihedral corrections, where the aromatic side chains would have obtained the energetically optimal packing, Martini 3 underestimates their free energy of association suggesting an arguable point at the parameterization procedure. Regarding the transmembrane systems, the minima correspond to structures that agree with the experimentally-observed dimer interfaces as well as with other interfaces previously observed in computational studies in the case of rhodopsin.

Overall, Martini 2.2P systematically overestimates barriers in distinct areas, which could lead to larger complexes, when multiple monomers are present. However, the estimated free energies of binding that correspond to minima with low RMSD approximate in an efficient agreement the experimental binding free energy. Hence, the overbinding seems to result from artificially considering structures in distinct areas (high RMSD) rather than the overestimation of

the binding energy at the crystallographic interface (low RMSD). Regarding the use of Martini 3, we expect that the fictitious supramolecular protein aggregation, which often results from Martini 2.2P, will be less often encountered in Martini 3, thus obtaining more realistic protein clusters. Nevertheless, the computed free energies of protein-protein binding seem to be systematically underestimated, at least in the aqueous environment. One potential area of improvement would be the addition of a polarizable water model in Martini 3. In Martini 2.1 four water molecules were represented by one neutral particle that had as an effect the absence of any explicit screening of electrostatic interactions, which was implicitly introduced with the use of a uniform dielectric constant. Hence, it was not able to respond to electrostatic fields and charge reorganization. To account for polarization effects, a polarizable water model consisting of one central neutral and two oppositely charged particles each of them bonded with the neutral one in an angle  $\theta$  was adopted in Martini 2.2P. The 2.2P version adopts the use of this polarizable water model in combination with charged dummy particles on polar amino acids, which shows increased performance in local electrostatics screening. Martini 3, however, includes the single-bead water model, which provides a three-fold enhanced computational efficiency at the expense of accuracy in some applications. In our case studies, the dimerization of the barnase-barstar system, which is electrostatically driven, should be mostly affected. Indeed, as we observe, the free energy of association predicted from Martini 2.2P is in better agreement with the experimental value compared to the Martini 3, which underestimates the experimental value by 4-fold. Hence, it would be expected that the introduction of a polarizable water model in Martini 3 could enhance the performance of the force field in protein dimerization studies and, in particular, of protein-associated membrane systems as the polarizability of water has proven to significantly affect free energies of association to membranes.<sup>57</sup>

An important outcome of this study is the use of converged enhanced sampling simulations with a refined coarse-grained force field, and the appropriate defined collective variables can provide a robust approach with reasonable throughput and accuracy for the calculation of the protein dimerization process. However, one should examine all energy minima present in the free energy landscape of protein dimerization as the ranking of energy minima may not always be in the correct order.

In the current study, we have used Martini 3.0.beta.3.2 as the only publicly available version of Martini 3. To the best of our knowledge, there is only one study<sup>29</sup> that refers to the more recent version called Martini 3.0.beta.4.17, which however is not openly available. In the aforementioned work, the authors compare their results with that obtained from their previous work<sup>58</sup> using the 3.0.beta.3.2 version. Interestingly, they concluded that the main force field issue in their case (protein overcompactness resulting from poor protein solvation) was persistent in the newer v.3.0.beta.4.17 version. This finding highlights the complexity of a coarse-grained force field parameterization. Towards the advancement of coarse-grained force fields to more accurate energetic estimations, the driving force of dimerization should also be taken into consideration. For example, in the case of the barnase-barstar dimer the interface is stabilized by the charge complementarity of the two monomers at the binding interface, while the insulin and H-Ras/Raf[RBD] dimer interface is driven by the hydrogen bond association of an intermolecular antiparallel  $\beta$ -sheet. Thus, it is envisioned that an accurate description of these bonds would require the implementation of directional interactions. Explicit dipole contributions from the protein backbone could advance the description of this type of interfaces. This is a direction that could be taken in future reparameterization efforts for the MARTINI and other CG force fields.

## 4. METHODS

### Coarse-grained Protein Models

The initial dimer conformations were obtained from their respective experimental structures from the PDB: 1BRS<sup>59</sup> for the barnase-barstar, 4INS<sup>48</sup> for the insulin dimer, 4G0N<sup>50</sup> for the H-Ras/Raf[RBD], 2K1L<sup>37</sup> for the EphA1, and 6OFJ<sup>36</sup> for rhodopsin. The polarizable Martini force field version 2.2P and the beta version Martini 3 were employed.<sup>32, 60</sup> The martinize python script<sup>61</sup> was used for generating the coarse-grained structures and their corresponding topologies for all systems. The protein complexes were solvated in water boxes with the corresponding water models (PW for Martini 2.2P and WN for Martini 3) for each force field, and ions were added to neutralize the systems (Table S1).

Martini 3 is used in the case of the insulin dimer in two different schemes. In the first scheme, we use a Martini 3 version, where dihedral potentials are added to the force field for two consecutive C $\alpha$ s and their first sidechain bead ([first side chain bead-C $\alpha$ <sub>i</sub> (bead)-C $\alpha$ <sub>i+1</sub> (bead)-first side chain bead]), where *i* runs over all residues. The force field terms of these potentials are fitted to the dihedrals of the atomistic crystallographic structure and added to Martini 3 as side chain dihedral corrections. In the second scheme, the Martini 3 force field without these side chain dihedral corrections is used. We evaluate the effect of the addition of these dihedral potentials to the Martini 3 force field accuracy in studying protein-protein interactions.

The transmembrane systems were built with the insane tool<sup>62</sup> adding the lipids and the solvent beads (Table S1) as well as incorporating the protein inside the membrane.

The atomistic systems were prepared initially with the psfgen tool, using the CHARMM force field representations, in order to include missing atoms in the structure. For barnase-barstar, the system was prepared as in Plattner et al.<sup>11</sup> starting from crystal structure with PDB ID 1BRS.<sup>59</sup>

The histidine protonation states (such as His102 of the barnase/barstar system) and other protonation states were kept as mentioned in the literature; if no information could be retrieved for the protonation state, these were assigned from the psfgen module of VMD. Terminal residues were kept neutral in the case they were different from the protein terminal residues.

### **Simulation Settings**

The simulation parameters follow those proposed by de Jong et al.<sup>63</sup> The Verlet neighbor search algorithm<sup>64</sup> was used to update the neighbor list with a straight cutoff of 11 Å. The temperature of the system was kept constant at 300 K with the stochastic velocity rescaling thermostat,<sup>65</sup> while pressure was maintained isotropically at 1 bar using the Parrinello-Rahman barostat<sup>66</sup> with a time constant of 12 ps and a compressibility of  $3 \times 10^{-4} \text{ bar}^{-1}$  representing an NPT ensemble. A cut-off of 11 Å was used for Lennard Jones and Coulombic interactions. For the latter, a reaction-field potential was applied.<sup>67</sup> In all simulations, the elastic network with default options, i.e. a force constant of  $1.2 \text{ kcal/mol } \text{Å}^2$  and an upper cut-off of 9 Å, was employed for the protein. In all cases, the elastic bonds network scheme was applied to each monomer. A time step of 10 fs was used. Martini 3 force field parameters were obtained from the open-beta release version 3.0.b.3.2.<sup>60</sup> Side-chain dihedral corrections as implemented in Martini 3<sup>60</sup> were added to each protein model based on its crystallographic structure unless stated otherwise. The dielectric constant for Martini 3 was set to 15, which is the default value for non-polarizable Martini versions. The same simulation parameters were used for the simulations with Martini 2.2P, except for the dielectric constant, which was set in this case to 2.5, as required by this force field's water model. Initially, the systems were minimized using the steepest descent algorithm for 1000 steps to remove steric clashes. The unbiased simulations were run for 15  $\mu\text{s}$ . Table 1

summarizes the different simulations that were performed and their corresponding timescales. For the transmembrane systems the semi-isotropic pressure coupling on the x-y plane from the z-axis was used.

### **Parallel Tempering Metadynamics in the Well-Tempered Ensemble (PT-MetaD-WTE) and Potential of Mean Force Calculations**

PT-MetaD-WTE simulations were performed using Plumed 2.5<sup>68</sup> across two CVs, namely a) the center-of-mass distance between the C $\alpha$  atoms of each monomer ( $D_{\text{COM}}$ ) and b) the root mean square deviation (RMSD) of the C $\alpha$  atoms from the corresponding crystal structures. A total of 11 replicas were run for each system in the temperature range between 300K and 320K with a temperature step of 2 K, and random exchanges between them were attempted every 100 steps. The average exchange rate between the neighboring replicas in each of the simulations is reported in Table S2. The biasing factor and the initial height of the hills set as referred in the Table 1. The width of the Gaussian hills were set to 0.1 Å for both the RMSD and  $D_{\text{COM}}$ , deposited every 1 ps.

The PT-MetaD-WTE simulations were run until convergence was reached. Convergence was monitored by estimating the free energy profile in regular intervals (Figures S1, S3, S5, S6b, S9, and S10), the diffusion of the CVs as a function of the simulation time (Figures S2, S11, S12) as well as the convergence of the associated error as calculated from the block analysis<sup>69</sup> (Figures S13-S15). The error in the estimation of the free energies (Table 2) was calculated as the average of the errors calculated for the blocks 900-1000. The Gromacs 2019.2<sup>70</sup> release package was used for the MD simulations as well as for the analysis. The metadynamics bias potential was written every 1000 steps and the Potential of Mean Force (PMF) was calculated by

post processing the output file using the `sum_hills` utility. The total simulation time for each system is shown in Table 1.

### Free Energy of Association Calculation

Supporting Information Figures S1, S3, S5, S6, S10 show the 1D PMF as a function of center of mass distance and RMSD. These PMFs were computed based on normalized distance distributions ( $p(r)$ ) between the center of geometry of the two proteins, including volumetric (also called entropic) correction  $2 \times \ln(r)$ .<sup>71</sup> This correction is commonly used to compensate the higher probability to find the dissociate state, as this is dependent of the box volume (or area in the simulations performed in bilayers).

$$PMF(r) = -RT \ln(p(r)) + (n_c - 1) \times \ln(r) \quad (1)$$

where  $n_c$  is the dimensionality of the system.

The distance  $r$  was shifted to take into account the average radial volume of the proteins, which was estimated based on the minimum distance allowed for the monomers to interact. For the insulin dimer, this distance is 1.75 nm; for barnase-barstar the distance is 2.25 nm; for H-Ras/Raf[RBD] it is 2.4 nm; and for the transmembrane systems EphA1 and rhodopsin the distances are 0.65 nm and 3 nm, respectively. The entire PMF was shifted to zero in the unbound state (plateau region of the 1D PMFs) before the PMF integration. The binding free energies (displayed in Table 2) were estimated by integrating the PMF over the distance  $r$  as<sup>71, 72</sup>

$$K_{bind} = \int_0^{r_c} 4\pi r^2 e^{-\frac{PMF(r)}{RT}} dr \quad (2)$$

$$\Delta G_{bind}^0 = -RT \ln(K_{bind} C^0) \quad (3)$$

with the standard concentration  $C_0$  equal to  $(1/1.25) \text{ nm}^{-3}$  for the insulin dimer with a cubic simulation box edge = 11 nm and two monomers in the box,  $(1/1.51) \text{ nm}^{-3}$  for barnase/barstar



with a box edge = 13 nm and two monomers in the box,  $(1/1.51) \text{ nm}^{-3}$  for H-Ras/Raf[RBD] dimer with a box edge = 13 nm and two monomers in the box,  $(1/0.98) \text{ nm}^{-3}$  for rhodopsin dimer with a cubic simulation box edge = 15 nm and two monomers in the box, and  $(1/1.249) \text{ nm}^{-3}$  for EphA1 with a cubic simulation box edge = 11 nm and two monomers in the box. The chosen cutoff ( $r_c$ ) was set to the fully dissociated state, i.e., 4 nm for EphA1, 6 nm for rhodopsin, and 5 nm for all other systems. For the transmembrane proteins, the 2D version of the equation above, integrating using a radially symmetric PMF in the x-y plane, was used as the proteins are restrained in the bilayer.

$$K_{bind} = \int_0^{r_c} 2\pi r^2 e^{-\frac{PMF(r)}{RT}} dr$$

The free energy of association was also calculated directly from the minima of the PMF 1D profiles, and the two values are compared in Table 2.

The PMFs were calculated for the replica at 300 K because this temperature coincides with the temperature used for the experiments with which we compare the values obtained here. The replicas are used only in order to enhance the sampling of the conformational space by exploring structures at higher temperatures and later accepting those with a probability to lower temperature replicas.

Supporting information Figures S3, S5, S6, S10 show the 1D PMF as a function of RMSD. These plots were calculated using the RMSD with respect to each crystallographic structure and the free energy included in the free energy surface files (produced by sum\_hills utility of PLUMED). From this plot, several minima are identified corresponding to the free energy of the dimer structures. In order to identify the exact location of these minima, we selected the RMSD

values corresponding to the minima free energy. Local minima are acquired in increments of 1 Å (see minima.py).

### **MD Analysis-Visualization**

Clustering was performed for the last 14 μs of the unbiased MD using the gmx cluster tool for a RMSD of 2.5 Å of the Cα atoms. Backmapping to atomistic resolution was achieved using the backward.py<sup>61</sup> script. VMD<sup>73</sup> was used for visualization of the trajectories.

### **Acknowledgements**

This work was supported by computational time granted from the Greek Research & Technology Network (GRNET) in the National HPC facility—ARIS—under Project IDs pr007019/kras. The research project was supported by the Hellenic Foundation for Research and Innovation (H.F.R.I.) under the “1st Call for H.F.R.I. Research Projects to support Faculty members and Researchers and the procurement of high-cost research equipment grant” (Project Number: 1780), awarded to ZC. The authors acknowledge the use of Fenix Infrastructure resources, which are partially funded from the European Union's Horizon 2020 Research and Innovation programme through the ICEI project under grant agreement No. 800858. The authors thank Dr. Vittorio Limongelli and Alexis Chatzigoulas for helpful discussions, and Dr. Mario Valle, Dr. Giuseppe Lo Re and Aspasia Vozi for help with uploading the data.

## Supporting Information

The Supporting Information is available free of charge via the Internet at <https://pubs.acs.org/doi/10.1021/acs.jctc.0c00507>.

Figures S1-S15 show one-dimensional free energy plot as a function of center-of-mass distance in different simulation times and as a function of RMSD collective variables, diffusion of the RMSD and distance reaction coordinates, comparison of Martini 3 derived structures with the crystallographic structure of the insulin dimer, probability densities of the RMSD for unbiased MD simulations using Martini 3, estimated error in the calculation of the free energies; Discussion on results obtained with the side-chain dihedral corrections in Martini 3; Table S1 contains the number of beads of each species for each simulated system; Table S2 contains the average exchange rate (%) between neighboring replicas for each system ([PDF](#)).

All data (input files, output files, trajectories, analysis scripts, and results) are available in the link: <https://hdl.handle.net/21.17101/cscs-lts-b04b2d27-1f74-4e02-b25d-847936c2bc4b?noredirect>. The archival space was granted by the second joint PRACE-ICEI call in the Fenix research infrastructure at CSCS, Switzerland.

## References

1. Bonetta, L., Protein-Protein Interactions: Interactome under Construction. *Nature* **2010**, *468*, 851-854.
2. Zhou, M.; Li, Q.; Wang, R., Current Experimental Methods for Characterizing Protein-Protein Interactions. *ChemMedChem* **2016**, *11*, 738-756.
3. Spreitzer, E.; Usluer, S.; Madl, T., Probing Surface Accessibility in the Dynamic Meshwork of Protein-Protein Interactions. *J Mol Biol* **2020**.
4. Sokolovski, M.; Cveticanin, J.; Hayoun, D.; Korobko, I.; Sharon, M.; Horovitz, A., Measuring Inter-Protein Pairwise Interaction Energies from a Single Native Mass Spectrum by Double-Mutant Cycle Analysis. *Nat Commun* **2017**, *8*, 212.
5. Keskin, O.; Tuncbag, N.; Gursoy, A., Predicting Protein-Protein Interactions from the Molecular to the Proteome Level. *Chem Rev* **2016**, *116*, 4884-4909.
6. 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.
7. Schreiber, G.; Haran, G.; Zhou, H. X., Fundamental Aspects of Protein-Protein Association Kinetics. *Chem Rev* **2009**, *109*, 839-860.

8. Pan, A. C.; Jacobson, D.; Yatsenko, K.; Sritharan, D.; Weinreich, T. M.; Shaw, D. E., Atomic-Level Characterization of Protein-Protein Association. *Proc Natl Acad Sci USA* **2019**, *116*, 4244-4249.
9. Barducci, A.; Bonomi, M.; Prakash, M. K.; Parrinello, M., Free-Energy Landscape of Protein Oligomerization from Atomistic Simulations. *Proc Natl Acad Sci USA* **2013**, *110*, E4708-4713.
10. Johnston, J. M.; Wang, H.; Provasi, D.; Filizola, M., Assessing the Relative Stability of Dimer Interfaces in G Protein-Coupled Receptors. *PLoS Comput Biol* **2012**, *8*, e1002649.
11. Plattner, N.; Doerr, S.; De Fabritiis, G.; Noé, F., Complete Protein-Protein Association Kinetics in Atomic Detail Revealed by Molecular Dynamics Simulations and Markov Modelling. *Nat Chem* **2017**, *9*, 1005-1011.
12. Webb, G. A., *Nuclear Magnetic Resonance*. Royal Society of Chemistry: Cambridge, **2004**; Vol. 33, p 586.
13. Singh, N.; Li, W., Recent Advances in Coarse-Grained Models for Biomolecules and Their Applications. *Int J Mol Sci* **2019**, *20*.
14. Bruininks, B. M. H.; Souza, P. C. T.; Marrink, S. J., A Practical View of the Martini Force Field. *Methods in molecular biology (Clifton, N.J.)* **2019**, *2022*, 105-127.
15. Corradi, V.; Sejdiu, B. I.; Mesa-Galloso, H.; Abdizadeh, H.; Noskov, S. Y.; Marrink, S. J.; Tieleman, D. P., Emerging Diversity in Lipid-Protein Interactions. *Chem Rev* **2019**, *119*, 5775-5848.
16. Marrink, S. J.; Corradi, V.; Souza, P. C. T.; Ingólfsson, H. I.; Tieleman, D. P.; Sansom, M. S. P., Computational Modeling of Realistic Cell Membranes. *Chem Rev* **2019**, *119*, 6184-6226.
17. Angelikopoulos, P.; Sarkisov, L.; Cournia, Z.; Gkeka, P., Self-Assembly of Anionic, Ligand-Coated Nanoparticles in Lipid Membranes. *Nanoscale* **2017**, *9*, 1040-1048.
18. Wang, Y.; Gkeka, P.; Fuchs, J. E.; Liedl, K. R.; Cournia, Z., Dppc-Cholesterol Phase Diagram Using Coarse-Grained Molecular Dynamics Simulations. *Biochim Biophys Acta* **2016**, *1858*, 2846-2857.
19. Podewitz, M.; Wang, Y.; Gkeka, P.; von Grafenstein, S.; Liedl, K. R.; Cournia, Z., Phase Diagram of a Stratum Corneum Lipid Mixture. *J Phys Chem B* **2018**, *122*, 10505-10521.
20. Marrink, S. J.; Tieleman, D. P., Perspective on the Martini Model. *Chem Soc Rev* **2013**, *42*, 6801-6822.
21. Stark, A. C.; Andrews, C. T.; Elcock, A. H., Toward Optimized Potential Functions for Protein-Protein Interactions in Aqueous Solutions: Osmotic Second Virial Coefficient Calculations Using the Martini Coarse-Grained Force Field. *J Chem Theory Comput* **2013**, *9*.
22. Javanainen, M.; Martinez-Seara, H.; Vattulainen, I., Excessive Aggregation of Membrane Proteins in the Martini Model. *PLoS ONE* **2017**, *12*, e0187936.
23. Lu, S.; Zhang, J., Small Molecule Allosteric Modulators of G-Protein-Coupled Receptors: Drug-Target Interactions. *J. Med. Chem.* **2018**.
24. Alessandri, R.; Souza, P. C. T.; Thallmair, S.; Melo, M. N.; de Vries, A. H.; Marrink, S. J., Pitfalls of the Martini Model. *J Chem Theory Comput* **2019**, *15*, 5448-5460.
25. Souza, P. C. T.; Thallmair, S.; Marrink, S. J.; Mera-Adasme, R., An Allosteric Pathway in Copper, Zinc Superoxide Dismutase Unravels the Molecular Mechanism of the G93a Amyotrophic Lateral Sclerosis-Linked Mutation. *J Phys Chem Lett* **2019**, *10*, 7740-7744.

26. Faustino, I.; Abdizadeh, H.; Souza, P. C. T.; Jeucken, A.; Stanek, W. K.; Guskov, A.; Slotboom, D. J.; Marrink, S. J., Membrane Mediated Toppling Mechanism of the Folate Energy Coupling Factor Transporter. *Nat Commun* **2020**, *11*, 1763.
27. Rao, S.; Bates, G. T.; Matthews, C. R.; Newport, T. D.; Vickery, O. N.; Stansfeld, P. J., Characterizing Membrane Association and Periplasmic Transfer of Bacterial Lipoproteins through Molecular Dynamics Simulations. *Structure* **2020**, *28*, 475-487.e473.
28. Boags, A.; Samsudin, F.; Khalid, S., Details of Hydrophobic Entanglement between Small Molecules and Braun's Lipoprotein within the Cavity of the Bacterial Chaperone Lola. *Scientific Reports* **2019**, *9*, 3717.
29. Larsen, A. H.; Wang, Y.; Bottaro, S.; Grudinin, S.; Arleth, L.; Lindorff-Larsen, K., Combining Molecular Dynamics Simulations with Small-Angle X-Ray and Neutron Scattering Data to Study Multi-Domain Proteins in Solution. *PLoS Comput Biol* **2020**, *16*, e1007870.
30. Vazquez-Salazar, L. I.; Selle, M.; de Vries, A. H.; Marrink, S. J.; Souza, P. C. T., Martini Coarse-Grained Models of Imidazolium-Based Ionic Liquids: From Nanostructural Organization to Liquid-Liquid Extraction. *Green Chem.* **2020**.
31. Srinivasan, S.; Zoni, V.; Vanni, S., Estimating the Accuracy of the Martini Model Towards the Investigation of Peripheral Protein-Membrane Interactions. *Faraday Discuss.* **2020**.
32. de Jong, D. H.; Singh, G.; Bennett, W. F. D.; Arnarez, C.; Wassenaar, T. A.; Schäfer, L. V.; Periole, X.; Tieleman, D. P.; Marrink, S. J., Improved Parameters for the Martini Coarse-Grained Protein Force Field. *J Chem Theory Comput* **2013**, *9*, 687-697.
33. Yesylevskyy, S. O.; Schäfer, L. V.; Sengupta, D.; Marrink, S. J., Polarizable Water Model for the Coarse-Grained Martini Force Field. *PLoS Comput Biol* **2010**, *6*, e1000810.
34. Periole, X.; Knepp, A. M.; Sakmar, T. P.; Marrink, S. J.; Huber, T., Structural Determinants of the Supramolecular Organization of G Protein-Coupled Receptors in Bilayers. *J Am Chem Soc* **2012**, *134*, 10959-10965.
35. Salom, D.; Lodowski, D. T.; Stenkamp, R. E.; Le Trong, I.; Golczak, M.; Jastrzebska, B.; Harris, T.; Ballesteros, J. A.; Palczewski, K., Crystal Structure of a Photoactivated Deprotonated Intermediate of Rhodopsin. *Proc Natl Acad Sci USA* **2006**, *103*, 16123-16128.
36. Zhao, D. Y.; Pöge, M.; Morizumi, T.; Gulati, S.; Van Eps, N.; Zhang, J.; Miszta, P.; Filipek, S.; Mahamid, J.; Plietzko, J. M.; Baumeister, W.; Ernst, O. P.; Palczewski, K., Cryo-Em Structure of the Native Rhodopsin Dimer in Nanodiscs. *J Biol Chem* **2019**, *294*, 14215-14230.
37. 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.
38. Artemenko, E. O.; Egorova, N. S.; Arseniev, A. S.; Feofanov, A. V., Transmembrane Domain of EphA1 Receptor Forms Dimers in Membrane-Like Environment. *Biochim Biophys Acta* **2008**, *1778*, 2361-2367.
39. Chavent, M.; Chetwynd, A. P.; Stansfeld, P. J.; Sansom, M. S. P., Dimerization of the EphA1 Receptor Tyrosine Kinase Transmembrane Domain: Insights into the Mechanism of Receptor Activation. *Biochemistry* **2014**, *53*, 6641-6652.
40. Strazza, S.; Hunter, R.; Walker, E.; Darnall, D. W., The Thermodynamics of Bovine and Porcine Insulin and Proinsulin Association Determined by Concentration Difference Spectroscopy. *Arch Biochem Biophys* **1985**, *238*, 30-42.

41. Rudolph, M. G.; Linnemann, T.; Grunewald, P.; Wittinghofer, A.; Vetter, I. R.; Herrmann, C., Thermodynamics of Ras/Effector and Cdc42/Effector Interactions Probed by Isothermal Titration Calorimetry. *J Biol Chem* **2001**, *276*, 23914-23921.
42. Schreiber, G.; Fersht, A. R., Interaction of Barnase with Its Polypeptide Inhibitor Barstar Studied by Protein Engineering. *Biochemistry* **1993**, *32*, 5145-5150.
43. Zoete, V.; Meuwly, M.; Karplus, M., Study of the Insulin Dimerization: Binding Free Energy Calculations and Per-Residue Free Energy Decomposition. *Proteins* **2005**, *61*, 79-93.
44. Banerjee, P.; Mondal, S.; Bagchi, B., Insulin Dimer Dissociation in Aqueous Solution: A Computational Study of Free Energy Landscape and Evolving Microscopic Structure Along the Reaction Pathway. *J Chem Phys* **2018**, *149*, 114902.
45. Ganim, Z.; Jones, K. C.; Tokmakoff, A., Insulin Dimer Dissociation and Unfolding Revealed by Amide I Two-Dimensional Infrared Spectroscopy. *Phys Chem Chem Phys* **2010**, *12*, 3579-3588.
46. Kim, T.; Rhee, A.; Yip, C. M., Force-Induced Insulin Dimer Dissociation: A Molecular Dynamics Study. *J Am Chem Soc* **2006**, *128*, 5330-5331.
47. Hjorth, C. F.; Norrman, M.; Wahlund, P.-O.; Benie, A. J.; Petersen, B. O.; Jessen, C. M.; Pedersen, T. Å.; Vestergaard, K.; Steensgaard, D. B.; Pedersen, J. S.; Naver, H.; Hubálek, F.; Poulsen, C.; Otzen, D., Structure, Aggregation, and Activity of a Covalent Insulin Dimer Formed During Storage of Neutral Formulation of Human Insulin. *J Pharm Sci* **2016**, *105*, 1376-1386.
48. Baker, E. N.; Blundell, T. L.; Cutfield, J. F.; Cutfield, S. M.; Dodson, E. J.; Dodson, G. G.; Hodgkin, D. M.; Hubbard, R. E.; Isaacs, N. W.; Reynolds, C. D., The Structure of 2zn Pig Insulin Crystals at 1.5 Å Resolution. *Philos. Trans. R. Soc. Lond. B. Biol. Sci* **1988**, *319*, 369-456.
49. Bourne, H. R.; Sanders, D. A.; McCormick, F., The Gtpase Superfamily: A Conserved Switch for Diverse Cell Functions. *Nature* **1990**, *348*, 125-132.
50. Fetics, S. K.; Guterres, H.; Kearney, B. M.; Buhman, G.; Ma, B.; Nussinov, R.; Mattos, C., Allosteric Effects of the Oncogenic Rasq61l Mutant on Raf-Rbd. *Structure* **2015**, *23*, 505-516.
51. Castillo, N.; Monticelli, L.; Barnoud, J.; Tieleman, D. P., Free Energy of Walp23 Dimer Association in Dmpc, Dppc, and Dopc Bilayers. *Chem Phys Lipids* **2013**, *169*, 95-105.
52. Monticelli, L.; Tieleman, D. P.; Fuchs, P. F. J., Interpretation of 2H-NMR Experiments on the Orientation of the Transmembrane Helix Walp23 by Computer Simulations. *Biophys J* **2010**, *99*, 1455-1464.
53. 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.
54. Khan, H. M.; Souza, P. C. T.; Thallmair, S.; Barnoud, J.; de Vries, A. H.; Marrink, S. J.; Reuter, N., Capturing Choline-Aromatics Cation- $\Pi$  Interactions in the Martini Force Field. *J Chem Theory Comput* **2020**.
55. Gunkel, M.; Schöneberg, J.; Alkhalidi, W.; Irsen, S.; Noé, F.; Kaupp, U. B.; Al-Amoudi, A., Higher-Order Architecture of Rhodopsin in Intact Photoreceptors and Its Implication for Phototransduction Kinetics. *Structure* **2015**, *23*, 628-638.
56. Dong, F.; Vijayakumar, M.; Zhou, H.-X., Comparison of Calculation and Experiment Implicates Significant Electrostatic Contributions to the Binding Stability of Barnase and Barstar. *Biophys J* **2003**, *85*, 49-60.

57. Salassi, S.; Simonelli, F.; Bochicchio, D.; Ferrando, R.; Rossi, G., Au Nanoparticles in Lipid Bilayers: A Comparison between Atomistic and Coarse-Grained Models. *The Journal of Physical Chemistry C* **2017**, *121*, 10927-10935.
58. Bottaro, S.; Bengtsen, T.; Lindorff-Larsen, K. Integrating Molecular Simulation and Experimental Data: A Bayesian/Maximum Entropy Reweighting Approach. In *Structural Bioinformatics: Methods and Protocols*, Gáspári, Z., Ed.; Springer US: New York, NY, **2020**, pp 219-240.
59. Buckle, A. M.; Schreiber, G.; Fersht, A. R., Protein-Protein Recognition: Crystal Structural Analysis of a Barnase-Barstar Complex at 2.0-Å Resolution. *Biochemistry* **1994**, *33*, 8878-8889.
60. Lin, W.-C.; Iversen, L.; Tu, H.-L.; Rhodes, C.; Christensen, S. M.; Iwig, J. S.; Hansen, S. D.; Huang, W. Y.; Groves, J. T., H-Ras Forms Dimers on Membrane Surfaces Via a Protein-Protein Interface. *Proceedings of the National Academy of Sciences* **2014**, *111*, 2996-3001.
61. Muratcioglu, S.; Chavan, T. S.; Freed, B. C.; Jang, H.; Khavrutskii, L.; Freed, R. N.; Dyba, M. A.; Stefanisko, K.; Tarasov, S. G.; Gursoy, A., Gtp-Dependent K-Ras Dimerization. *Structure* **2015**, *23*, 1325-1335.
62. Wassenaar, T. A.; Ingólfsson, 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.
63. 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.
64. Verlet, L., Computer "Experiments" on Classical Fluids. I. Thermodynamical Properties of Lennard-Jones Molecules. *Phys. Rev.* **1967**, *159*, 98-103.
65. Bussi, G.; Donadio, D.; Parrinello, M., Canonical Sampling through Velocity Rescaling. *J Chem Phys* **2007**, *126*, 014101.
66. Parrinello, M., Polymorphic Transitions in Single Crystals: A New Molecular Dynamics Method. *J Appl Phys* **1981**, *52*, 7182.
67. Tironi, I. G.; Sperb, R.; Smith, P. E.; van Gunsteren, W. F., A Generalized Reaction Field Method for Molecular Dynamics Simulations. *J Chem Phys* **1995**, *102*, 5451-5459.
68. 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.
69. Flyvbjerg, H.; Petersen, H. G., Error Estimates on Averages of Correlated Data. *J Chem Phys* **1989**, *91*, 461-466.
70. Berendsen, H. J. C.; van der Spoel, D.; van Drunen, R., GROMACS: A Message-Passing Parallel Molecular Dynamics Implementation. *Comput Phys Commun* **1995**, *91*, 43-56.
71. Souza, P. C. T.; Thallmair, S.; Conflitti, P.; Ramírez-Palacios, C.; Alessandri, R.; Raniolo, S.; Limongelli, V.; Marrink, S. J., Protein-Ligand Binding with the Coarse-Grained Martini Model. *Nat Commun* **2020**, *11*, 3714.
72. Limongelli, V., Ligand Binding Free Energy and Kinetics Calculation in 2020. *WIREs Computational Molecular Science* **2020**, *10*, e1455.
73. Humphrey, W.; Dalke, A.; Schulten, K., Vmd: Visual Molecular Dynamics. *J Mol Graph* **1996**, *14*, 33-38, 27.