Modified Potential Functions Result in Enhanced Predictions of a Protein Complex by All-Atom MD Simulations, Confirming a Step-wise Association Process for Native PPIs

Zhen-lu Li[1] and Matthias Buck[1,2]

[1]Department of Physiology and Biophysics, Case Western Reserve University, School of Medicine, 10900 Euclid Avenue, Cleveland, Ohio 44106, U. S. A.

[2]Department of Pharmacology; Department of Neurosciences, Case Western Reserve University, School of Medicine, 10900 Euclid Avenue, Cleveland, Ohio 44106, U. S. A.

E-mail: matthias.buck@case.edu

ABSTRACT

Native protein-protein interactions (PPIs) are the cornerstone for understanding the structure, dynamics and mechanisms of function of protein complexes. In this study, we investigate the association of the SAM domains of the EphA2 receptor and SHIP2 enzyme by performing a combined total of 48 μs all-atom molecular dynamics (MD) simulations. While the native SAM heterodimer is only predicted at a low rate of 6.7% with the original CHARMM36 force field, the yield is increased to 16.7% and to 18.3% by scaling the vdW solute-solvent interactions (better fitting the solvation free energy of amino acid side chain analogues) and by an increase of vdW radius of guanidinium interactions, and thus a dramatic reduction of electrostatic interaction between Arg and Glu/Asn in CHARMM36m, respectively. These modifications effectively improve the overly sticky association of proteins, such as ubiquitin, using the original potential function. By analyzing the 25 native SAM complexes formed in the simulations, we find that their formation involves a pre-orientation guided by electrostatic interaction, consistent with an electrostatic steering mechanism. The complex could then transform to the native protein interaction surfaces directly from a well pre-orientated position (Δinterface-RMSD < 5Å). In other cases, modest (< 90°) orientational and/or translational adjustments are needed (5 Å <Δi-RMSD <10 Å) to the native complex. Although the tendency for non-native complexes to dissociate has nearly doubled with the modified potential functions, a re-association to the correct complex structure is still rare. Instead a most non-native complexes are undergoing configurational changes/surface searching, which do not lead to native structures on a timescale of 250 ns. These observations provide a rich picture on mechanisms of protein-protein complex formation, and suggest that computational predictions of native complex protein-protein interactions could be improved further.

Keywords: protein-protein complex formation; potential function; electrostatic steering; encounter complex; association mechanism

INTRODUCTION

Biological function that arises from the formation (and/or dissociation) of protein complexes critically depends on the strength and kinetics of the underlying protein-protein interactions (PPIs), where specific/native interactions typically compete with less specific/non-native contacts. Our detailed knowledge of protein-protein interactions and understanding of complex formation is far from complete [1-5]. Especially, an appreciation of the protein association processes involved is still lacking for the

great majority of systems at the molecular level. This is in part due to the limited temporal, if not spatial resolution of experimental methods. Despite early successes in Brownian and coarse grained molecular dynamics simulations of (mostly tightly bound) complexes [2,6-10], there have, as yet, been very few simulations of protein association processes using unrestrained all-atom simulations [11-16]. In many cases it remains difficult to computationally predict the correct protein complexes structures ab initio. Only relatively recently have computational protein docking programs started to incorporate mainchain (or even sidechain) flexibility [17-19]; also the role of local, as well as global dynamics in protein complexes and in complex formation is now becoming well appreciated [20-23]. Indeed, a community wide challenge, the critical assessment of predicted interactions (CAPRI), established in 2001, has yielded steady improvements over the years; currently most "easy to moderately difficult" protein complex target structures can be predicted reliably to be among the top or top 5 (or 10) of calculated solutions [24], although often sequence/evolution-based as well as experimental information are used as restraints.

Classical all-atom molecular dynamics simulations are purely based on empirically derived physical potential functions. In principle, these simulations can provide information on the unrestrained and dynamic processes of protein association/dissociation as well as on the physical-chemical details of the interactions, which are not easily accessible by other computational methods or by current experiments. For example, recently, we reported all-atom simulations on the dissociation process of a small protein heterodimer, the EphA2-SHIP2 SAM: SAM domain complex [25]. In that study we showed that a simple (e.g. a single) dissociation pathway was absent and that instead a step-wise process was involved, allowing for multiple pathways [25]. A number of inferences could be made from that study, including the suggestion that the reverse process, that of protein association, would involve transient encounter complexes and may proceed via intermediary bound states. Another process associated with protein complex formation, electrostatic steering which directs the formation of encounter complexes to be close to the native protein configuration, has been reported and discussed over several decades [26-28].

Insufficient sampling is a major obstacle in the accurate prediction of protein complexes but so are inaccuracies in the potential function. Once two proteins are trapped into a local minimum/non-native complex, they find it hard to dissociate or to convert to a native complex, as the relaxation time of dissociation or orientation transition could be far beyond the simulation time scale. This could be the case with our system as well: here, using CHARMM36, the SAM domains associated on a reasonable timescale, but the initial associated states were often "trapped", in that they showed few interconversions and fewer dissociation events, than would be expected from initial non-native encounter complexes [9, 29, 30]. Consequently, the yield rates of native protein complexes were typically low within the simulation time-limit. However, extending the simulations to several μs did not resolve the issue (data not shown, and as also seen the lack of transitions between certain configurational states [22]). In fact, the low yield of correct structures, approx.7-8% is actually similar to those seen in the early Brownian Dynamics simulations [9,10]. Because of this "kinetic trapping issue", the molecular dynamics community has recently appreciated that several of the current potential functions appear to lack accuracy, especially for simulations involving protein association [14] and for the behavior of intrinsically denatured proteins (IDPs) [31]. In case of the former, many proteins -even those having only non-specific non-covalent self-association (such as ubiquitin)- showed a strong and unphysical aggregation [14,32]. Similarly, IDPs became too compact in the simulations [33, 34]. Both effects arise

from an underestimation of protein solvation and both are in disagreement with experimental observations.

To address these problems, researchers have tried to improve parameters in the potential functions in different ways. In one study, Piana et al. developed a new water model, which could better reproduce intrinsically disordered proteins [35], and this has recently been refined to also work on folded proteins [34]. A most recently released version of the CHARMM potential function, CHARMM36m, improves the prediction of α-helices and intrinsic disorder in polypeptides [36]. Best et al. have explored a scaling of protein-water interactions with the Amber potential function [37], which showed an improved prediction for IDPs and non-specific protein associations. A study of Feig and colleagues found that the protein translational and especially rotational diffusion slowed down following an adjustment of the potential function by globally scaling protein-water interactions [14]. Several other studies have rescaled the protein-protein interactions or protein-water interactions, or carried out a mixed rescaling with different water models (such as TIP4P) in order to get a better performance in the prediction of native protein configurations or protein-protein association [38-42]. At this time, there are still only a few studies of unrestrained protein-protein association involving binding partners with moderate binding affinity using classical molecular dynamics simulations. Furthermore, it was not clear to what extent the potential, such as the recently revised version of CHAMM36, CHARMM36m, may accurately predict the native protein-protein interactions.

Availability of computer resources to extensively simulate the association events motivated us to carry out unrestrained all-atom simulations to study protein complex formation of one system. We performed a set of 60 simulations (each of 250 ns) on the association between EphA2 SAM (E-SAM) and SHIP2 SAM domain (S-SAM). Previously, we found that the SAM heterodimer protein complex is an intrinsically dynamic structure, using restraints from solution NMR experiments [43]. This view was reinforced by extensive classical all-atom MD simulations which showed that the protein complex could convert between several configurational states on a ~ 100 ns timescale in the absence of restraints [22]. The complex has a moderate binding affinity with a dissociation constant (K_d) value at 2.2 μM, a value typical for the interaction of cell signaling proteins. As mentioned above, in a recent computational study, we investigated the dissociation pathways of the complex resulting from swaps of domain bridging charged sidechains which weakened and perturbed the energy landscape of the complex [25]. SAM domains are particularly suitable for computational studies of protein-protein association, considering their small size (around 70 residues) and well folded structure.

When using the original CHARMM36 potential function in the present study, we found that the protein association is rather sticky and the simulations have a rather low yield rate of the native complex at 6.7%. By scaling the solvent-solute vdW interactions to better fit the solvation free energy of amino acid side chain analogues, the yield of the native complex is increased to 16.7%. A reduction of interactions of the guanidinium group (i.e. the Arg sidechain) makes in CHAMRM36m, also results in a weakened protein association and consequentially a better yield ratio of the native complex at 18.3%. These improved potential functions also enhance the prediction of non-specific self-association between Ubiquitin. Finally, we examined the protein association pathways of the 25 cases where native complexes of E-SAM: S-SAM are formed and again find a multi-state process is at play.

METHODS

The simulations of protein associations between E-SAM and S-SAM were performed with the all-atom CHARMM potential function [44-46], 60 simulations each, for three sets of parameters, CHARMM36, CHARMM36m and CHARMM36s; the latter denoting vdW scaled solute-water interactions. The E-SAM and S-SAM domains of the NMR derived complex [pdb.id 2kso, ref. 43] were separated to a distance of 40 Å between their center of mass. Both SAM domains were then rotated randomly with respect to their center of mass in each of the 60 simulations to produce different initial orientations for each simulation with CHARMM36. In simulations with CHARMM36s and CHARMM36m, we used the same set-ups of initial orientations between two SAMs as in simulations with CHARMM36, thus there are no bias between three sets of parameters in terms of starting simulation configurations. The nearest distance between any atoms of the two SAMs are about 6-18 Å, mostly (in 48 of 60 simulations) larger than 10 Å. The whole system (Ubiquitin or E-SAM: S-SAM) was solvated by TIP3P water [47] in a simulation box of 90×90×90 $Å^3$, setting up periodic boundary conditions. Sodium and Chloride ions were added to a near-physiological concentration of 150 mM and to neutralize the system. Sidechain groups were charged as expected at pH 7.0 (Histidine with hydrogen on delta carbon-HSD/HID, was used). In total, the simulation box contained about 71,000 atoms. The system set-up followed previous protocols (e.g. initial heating and equilibration for 40 ps and 1 ns, respectively [e.g. 22]). Each simulation was 250 ns -300 ns long. We also performed simulations of self-association of Ubiquitin. Two Ubiquitins [pdb id. 1ubq] were separated by 40 Å at the beginning of simulation. Following an equilibrium for 20 ns, we performed 2 μs production simulations on the Anton2 supercomputer for each of the potential functions.

The standard free energy perturbation method was used to calculate the solvation free energy of amino acid side chain analogues [48-51]:

$$\Delta G = -k_B T \sum_{k=1}^{N} ln < \exp\left[-\frac{\mathrm{H}(\mathbf{r},\mathbf{p};\lambda_{k+1})-\mathrm{H}(\mathbf{r},\mathbf{p};\lambda_k)}{k_B T}\right] >_k ,$$

where λ is a scale factor (and a small perturbation) that gradually tune the potential energy function of a system. Specifically in our simulation, the potential energy function between the side chain analogue and water was first tuned gradually from λ = 1 to λ = 0 by a decrement δλ = 0.02 (forward simulation) in a 2 ns simulation; and is gradually turned back on over from λ = 0 to λ = 1 in a followed simulation period of 2.0 ns (backward simulation). 50 intermediate λ-states or "windows" (N=50) are produced with an interval of δλ = 0.02. In order to avoid charge singularities at λ = 0 or 1, a soft-core potential is used smoothly switch off the interactions. Specifically, outgoing atoms see their electrostatic interactions with the environment decoupled during λ = 0 to 0.5, while van der Waals interactions are gradually decoupled from λ = 0 to 1. Therefore, the electrostatic interaction is fully switched off before the van der Waals interaction is tuned to 0. Each window is simulated for 40 ps, with a 4 ps lengthy equilibrium and a 36 ps simulation for data collection. Overall, then for each calculation of solvation free energy for a specific amino acid analogue, the system is first equilibrated for 2.0 ns. A subsequent 4.0 ns of the simulation (forward and backward simulation) was used to perform the free energy perturbation. Three repeats were performed to estimate the statistic errors.

In all simulations, the van der Waals (vdW) potential was truncated at 12 Å (default for CHARMM36/-m) and smoothly shifted to zero between 10 and 12 Å to reduce cutoff noise. The Particle-Mesh Ewald (PME) method was used for calculating the long distance electrostatic interactions [52]. The SHAKE

algorithm was applied for all covalent bonds to hydrogen. The time step was set as 2 fs. The systems were coupled to a thermostat at 300 K and a barostat at 1 bar with the Langevin scheme [53]. All these systems were simulated using the NAMD/2.12 package [54]. The simulations of Ubiquitin were further transferred to Anton supercomputer [55] to simulate the self-association at a longer time scale of 2 μs. Analysis (interface-RMSD, residues-residues contact, surface electrostatics, orientational angle, pair interactions, and solvent accessible surface) were done with VMD or in NAMD with in-lab scripts (for descriptions see figure legends/main text). The first time of contact between the two SAM domains was noted when any atoms of the two structures came within 4 Å from one another. The same cutoff was used to construct contact maps. In calculation of interface-RMSD (i-RMSD), the Cα position of interface residues 949-962 of E-SAM and residues 1215-1239 of S-SAM were used for an otherwise standard RMSD calculation, with respect to the average Cα location in the NMR ensemble [43]. A protein complex with an i-RMSD < 4.0 Å with respect to native complex is considered as native-like, given that this value encompasses the different orientation states of the two SAM domains relative to one another from the original solution NMR analysis [43]. Incidentally, 4 Å is also the cut-off chosen for i-RMSD to indicate acceptable protein complex structures in the Critical Assessment of Predicted Interactions (CAPRI) experiment [24]. The contrast similarities of residue-residue contact maps were made between current simulations of protein association and prior simulations starting from a native complex of E-SAM: S-SAM. Native contacts were considered as interactions between: L951: H1219/N1220/G1221, P952: N1220, G953: N1220/G1221/W1222, H954: G1221, K956: W1222/D1235, R957: D1224/F1227, Y960: D1230. The similarities are calculated by comparing the prevalence of these interactions in simulations with different potential functions to the prevalence of these interactions in simulations starting from a native complex.

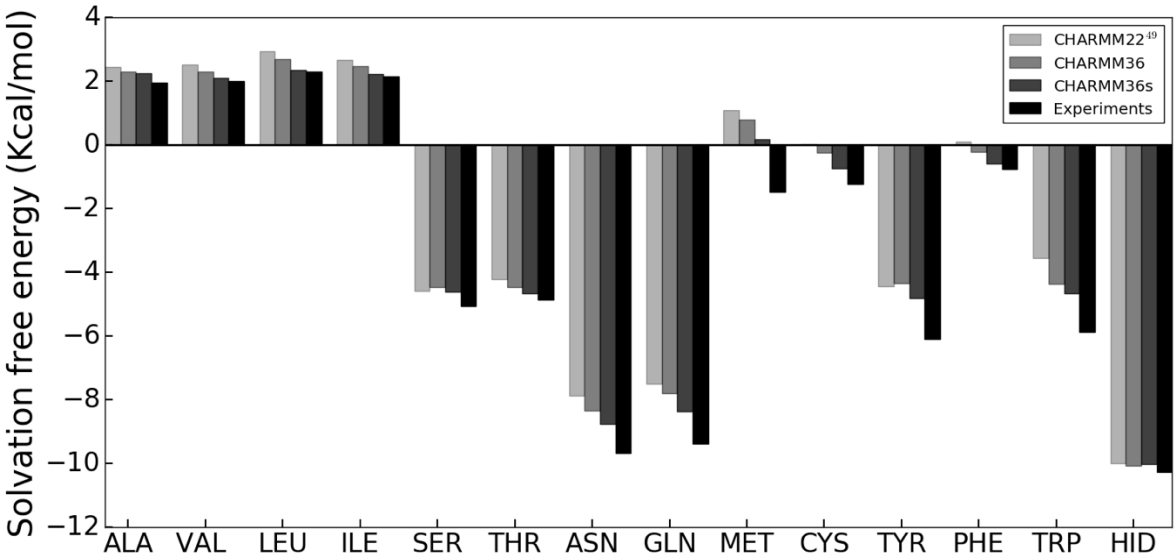

Figure 1: Solvation Free Energy of Amino Acid Side Chain Analogues. Comparisons are made with experimental data (ref. 56) and the molecular dynamics simulation results with CHARMM22 (by Shirts et al., ref. 49), CHARMM36, and CHARMM36s. HID indicates Histidine with H atom bonded to N^{δ}.

RESULTS

Optimization of Solvation Free Energy of Amino Acid Side Chains by Scaling Protein Atom-Water Interactions

Side chains of amino acids are dominant groups that mediate the protein-protein interactions. However, in many of the prevalent potential functions (e.g. CHARMM and Amber), the solvation free energy of amino acid side chain analogue is underestimated relative to the experimental value. The differences vary from -0.2 kcal/mol to as large as -2.5 kcal/mol. The inaccuracy is particularly large for amino acids that typically constitute "hot spots" at protein-protein interface (PPI) such as hydrophobic residue ILE and aromatic residues TYR and TPR [Fig. 1 and Table 1]. The underestimated solvation promotes protein-protein association. A scaling of side chain: water interaction is a straightforward and empirical way to increase the solvation of amino acid side chain. We made modification on the CHARMM36 potential function parameter by scaling the solute-solvent vdW interaction potential energy. Firstly, the interactions between -CH2, –CH3 group and water is increased by a small scale factor of 1.03 (i.e. an increase of 3%). These changes help to better fit the solvation of hydrophobic amino acids of ALA, VAL, LEU, ILE. To further increase the solvation of polar residue, a scale factor of 1.10 was applied to polar S, N and C at the terminal of SER, THR, MET, CYS. The solvation free energy was not calculated for charged residues where solvation free energy calculations are made difficult by the frequent presence of counter ions [48-51]. However, due to stereo-similarity at the side chain level to their ASN/GLN polar counterparts, these scalings were also applied to basic and acidic residues LYS, ARG, ASP and GLU. At last, a scale factor of 1.05 was applied to the benzene group in PHE, TYR and TRP. It improves the solvation of benzene; in the meantime, it increases the solvation of the aromatic amino acids {see Table 1}. All the changes applied to individual amino acids are shown in Figure S1.

Table 1: Solvation Free Energy of Amino Acid Side Chain Analogue (Kcal/mol).

Amino Acid Side Chain Analogue	Experiment (ref. 56)	CHARMM36	CAHRMM36s	CHARMM22 (ref. 49)
ALA (methane)	1.94	2.29+/-0.07	2.24+/-0.03	2.44
VAL (propane)	1.99	2.28+/-0.04	2.08+/-0.10	2.52
LEU (isobutene)	2.28	2.68+/-0.05	2.33+/-0.06	2.94
ILE (butane)	2.15	2.46+/-0.09	2.21+/-0.12	2.67
SER (methanol)	-5.06	-4.48+/-0.11	-4.63+/-0.08	-4.59
THR (ethanol)	-4.88	-4.47+/-0.08	-4.68+/-0.09	-4.22
ASN (methanethiol)	-9.68	-8.38+/-0.09	-8.77+/-0.12	-7.89
GLN (propionamide)	-9.38	-7.82+/-0.07	-8.39+/-0.01	-7.51
MET (methyle ethylSulfide)	-1.48	0.78+/-0.07	0.18+/-0.12	1.08
CYS (methanethiol)	-1.24	-0.25+/-0.03	-0.74+/-0.02	0.02
Benzene	-0.87	-0.34+/-0.05	-0.72+/-0.11	NA
TYR (p-cresol)	-6.11	-4.35+/-0.21	-4.83+/-0.19	-4.46
PHE (toluene)	-0.76	-0.22+/-0.08	-0.60+/-0.20	0.09
TRP (methylindole)	-5.88	-4.38+/-0.08	-4.66+/-0.14	-3.57
HID (methylimidazole)	-10.27	-10.08+/-0.05	-10.04+/-0.06	-10.00

As listed in Table1 and Fig.1, all of the amino acids achieve better solvation with vdW scaling of the solute-water interactions [56]. On average, the solvation (excluding ALA and and protonated His, HSD) is increased by 0.36 kcal/mol for all residues. In a pioneering study by Best et al and Feig et al, similar scale on solute-water has been applied [14, 37]. However, in these two studies, a much higher vdW scaling factor of 1.10 times was applied to all atoms. The solvation free energy calculation showed that a 1.10 scaling affords too much solvation for the amino acids acids, already apparent in the study of Best et al [37]. In addition, modifications on the main chain atoms likely made the protein fold unstable (data not shown). Therefore, in this study, the scale was only applied to some of the amino acid side chain atoms; and overall, a much lower scaling factor (mostly 1.03, and few atoms with 1.05 and 1.10)

was used. Next, we tested the influence of these changes on a model system for protein-protein association. Last year, a modified version-CHARMM36m was released. A major change was made in the CMAP potential; and with respect to protein-protein association, to the interaction between guanidinium (-NH3 group of ARG) and of the carboxylate (-COO group of ASP/GLU) was reduced by modifying vdw radii from 0.3550 nm to 0.3637 nm [36]. Thus, in the present study, we tested three sets of simulation parameters/potential functions, denoted ---CHARMM36, CHARMM36s, and CHARMM36m.

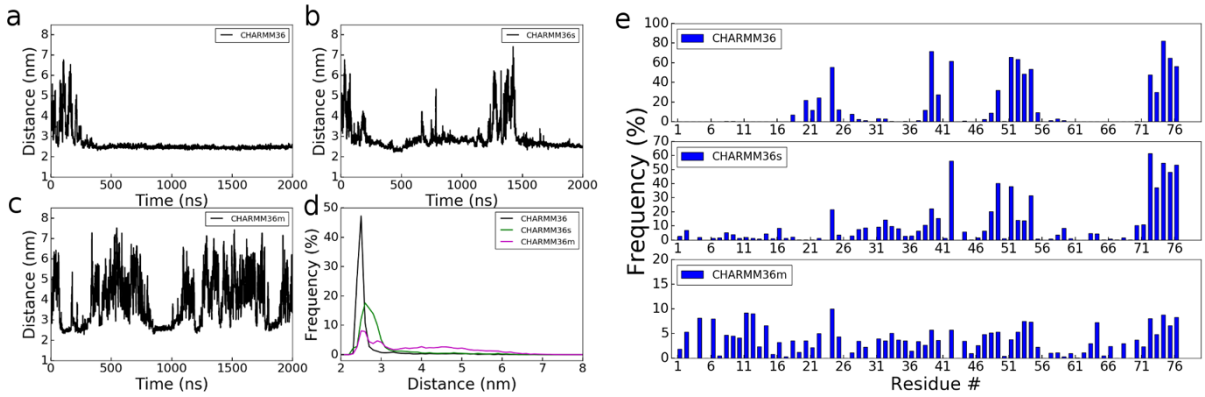

Figure 2: *Self-association of Ubiquitin over time course of a 2,000 ns simulation with different potential functions. (a-c) Time evolution of distance between center of mass of ubiquitin. (d) Distribution of distance between center of mass of the two ubiquitins (e) Frequency of individual residues at ubiquitin-ubiquitin interface.*

Simulations of weak Non-Specific Interactions between Ubiquitin

Based on the NMR results, Ubiquitin has a non-covalent self-association with a very weak binding affinity, Kd at 4.4 - 4.9 mM [57], which is likely to manifest in frequent dissociation and association events which should be captured even in a short simulation. However, the study by Abriata et al. showed that in molecular dynamics simulation Ubiquitin has a rather strong self-association and aggregate easily and unphysically [32]. Similarly, in our simulation, with the original CHARMM36 potential function ubiquitin inevitably associates into a dimer which does not dissociate over the remainder of the 2 µs simulation (Fig. 2a). With improved solvation of the amino acid side chains in CHARMM36s, "overly sticky" protein-protein interactions are partially released (Fig. 2b). The protein complex undergoes a few dissociation-(re)association events, but eventually forms a relatively stable complex from 1.5 to 2 µs. With CHARMM36m, the protein complex undergoes rather fast and frequent dissociation-(re)association (Fig. 2c). Fig. 2d plots the distribution of distance between center of mass of proteins. The distance between the center of masses of two Ubiquitin is within 3 nm for 90.5%, 77.6% and 34.2% of time in the simulations trajectories, using CHARMM36, CHARMM36s and CHARMM36m respectively. (By comparison, the radius of gyration of ubiquitin is 1.2 nm.) Based on the experimental K_d value above, the concentration of Ubiquitin (4.8 mM) in the simulation, and supposing that the kinetics of association and dissociation are fast, the fraction of Ubiquitin dimer should be 49.3% to 54.2%. Clearly, both potential function parameter changes improve the predictions of monomer-dimer equilibrium of Ubiquitin, but CHARMM36s still overestimates the bound population whereas CHARMM36m may slightly underestimate it. The NMR study indicated that residues 4-12, 42-51, and 62-71 are the major contacts at the protein-protein interface (due to the high flexibility, res. 72 to 76 were not analyzed [57]). In the simulations with CHARMM36, the prevalence of PPIs at res. 4-12 or

62-71 is not observed. Simulations with CHAMM36s and CHARMM36m both show some interactions at res. 4-12 and 62-71, but CHAMMR36s shows only a very low frequency for these interactions. Surprisingly, compared to CHARMM36/-s, CHARMM36m shows a rather wide spread/non-specific interaction across the entire protein.

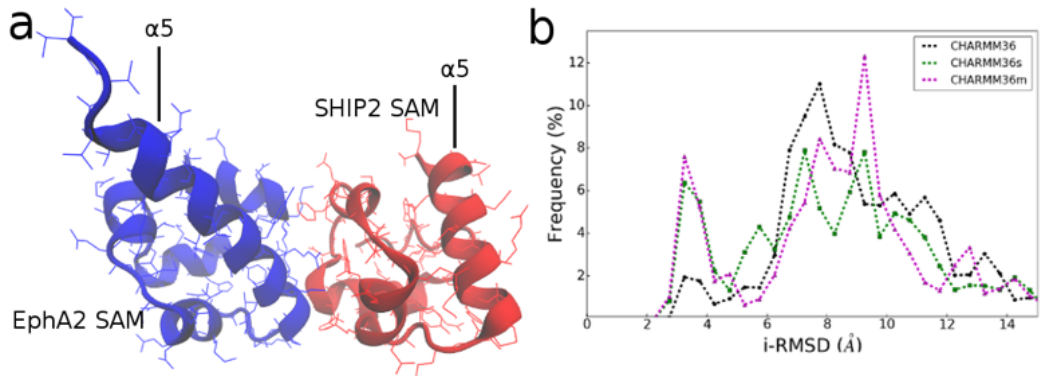

Figure 3: Enhanced prediction of native complex of E-SAM: S-SAM. (a) Heterodimer of E-SAM: S-SAM solved via NMR. E-SAM is in blue color and S-SAM is in red color throughout the paper. (b)Distribution of i-RMSD for different simulation sets (i-RMSD for structures in the last 50 ns of the 60 simulations).

There is only one correction in CHARMM36m that is directly related to protein-protein interactions: the vdW minimum potential between atom type of NC2 (ARG) and OC (GLU/ASP) is shifted from 0.3550 nm to 0.3637 nm. The influence of this correction is dramatic. While the change is small for the vdW potential energy, the change in inter-atom distance is large in terms of electrostatics, considering that HC (connected to NC2) and OC have partial charges of 0.46 and -0.76 e respectively. The longer the equilibrium distance between HC and OC, the more reduced are the electrostatic interactions between the ARG and GLU/ASP sidechains. A rough estimation (without considering the vicinity of water or possible counterions) indicates that the interactions between –COO and $–C(NH2)_2$ group is increased by 3.24 kcal/mol from -107.48 kcal/mol with CHARMM36 to -104.24 kcal/mol with CHARMM36m (Figure. S2). In the self-association of ubiquitin, the association is largely maintained by guandinium-carboxylate interactions as well as interactions between LYS and ASP/GLU. Therefore, the reduction of guanidinium-carboxylate interaction in CHARMM36m largely attenuates its self-association. However, considering the fraction of Ubiquitin dimer is less than that expected in simulations with CHARMM36m, and it results in a rather even distribution of residues at PPI, the reduction of guanidinium-carboxylate interactions may overly weaken the protein-protein interactions. However, this point still needs to be cautiously examined in the other systems and further examined via free energy calculation of protein-protein binding.

Enhanced Prediction of E-SAM: S-SAM Association with Optimized Force Field Parameter

Due to our previous extensive experimental and computational studies of EphA2-SAM: SHIP2 SAM complex, we chose this complex to examine a protein-protein association process. The complex has a moderate binding affinity with a Kd of 2 µM. Using the original potential function, CHAMRM36, we found that even long simulations of 2-12 µs, carried out on Anton could not dislodge several of the trapped states that exist shortly after association (data not shown). Raising the simulation temperature (from 300K to 323K or 350K) or increasing the salt concentration (from 150 mM to 1.5 M NaCl) did not overcome this problem either. In order to increase simulation sampling, we performed 60 simulations

each of 250-300 ns for each set of simulation parameters, CHARMM36, /-s and /-m. Fig. 3 plots the distribution of the interface RMSD (i-RMSD) defined by the root mean square distance to native contacts (see Methods). Clearly, the simulations with modified potential function parameters have a better performance in sampling native-like structures than simulations with the original CHARMM36 potential function. Specifically, with original CHAMRM36, only 4 of 60 simulations yields a native complex. However, with the modified potential functions in CHARMM36s, the simulations yield 10 native complexes; with CHAMRM36m, 11 native complexes were observed. So the yield of native structures was increased from 6.7% to 16.7%, and 18.3% respectively. A significant difference between the CHARMM36s and /-m potentials is that the former has several structures in the range of 5-6 Å i-RMSD whereas the latter has very few, suggesting that the free energy minimum of the native complex is well separated by having energy barriers that restrict access to surrounding minima. This observation is likely to increase the stability of the native-like complexes; for example, in our study of mutant complexes we found that the dissociation of mutant complexes is greatly facilitated by their ability to shift to weaker binding interfaces, from which the dissociation event occurred [25]. Consequently, an energy barrier between near-native and native structures may not allow an easy transition from one to the other, and it is likely to be preferably to have a smooth energy landscape, funneling complexes to the native state. Thus, the method of improving protein solvation may have a superior feature in this regard.

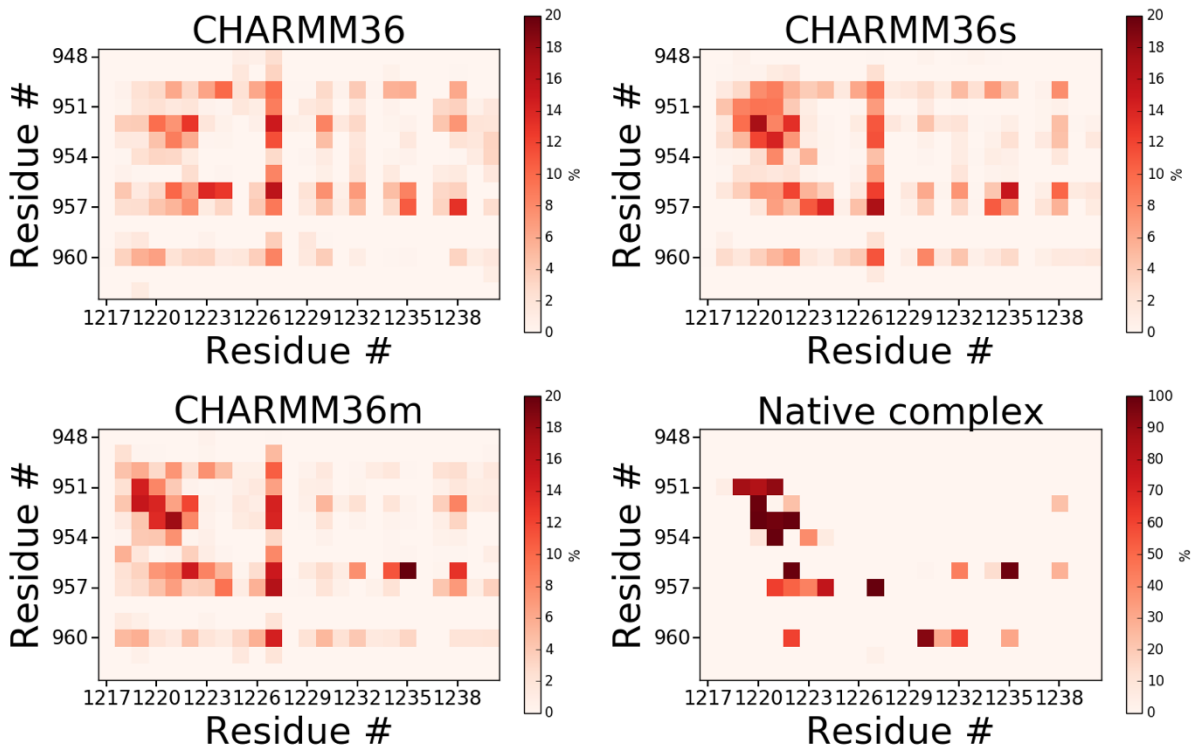

Figure 4: Residue-residue contact maps for different simulation sets (subset of map showing regions of native-like interactions only). The last 50 ns of the trajectories were used. Contact maps (bottom, right) using a weighted ensemble of native structures state 1 and state 2 [22].

Fig. 4 compares the difference in residue-residue contact patterns (for the last 50 ns of the simulations), between simulations with original and modified parameters. Counting major interacting spots in native complex, the contact similarity between the native configuration and the latter part of the association simulations with CHARMM36, CHARMM36s and CHARMM36m are 5.5%, 12.5 % and 12.5%

respectively (see Methods for calculation details). As non-specific/transient contacts are averaged, the pattern of native contacts are enhanced in simulation with modified parameters with reference to the residue-residue contact map obtained from our previous simulations starting from the native complex [22]. Furthermore, remarkably, the major region of native contacts, E-SAM res. 952-957: S-SAM res. 1220-1224, 1227 and 1235 is particularly enhanced in simulations with the modified potential function, since even in cases where a native-like complex is not formed, one of the two proteins typically uses the correct native-like/high affinity interface for making contacts. The enhanced sampling with improved potential function indicates that the inaccuracy of CHARMM36 potential function accounts at least in part for the low yield rate of native structure in original simulations.

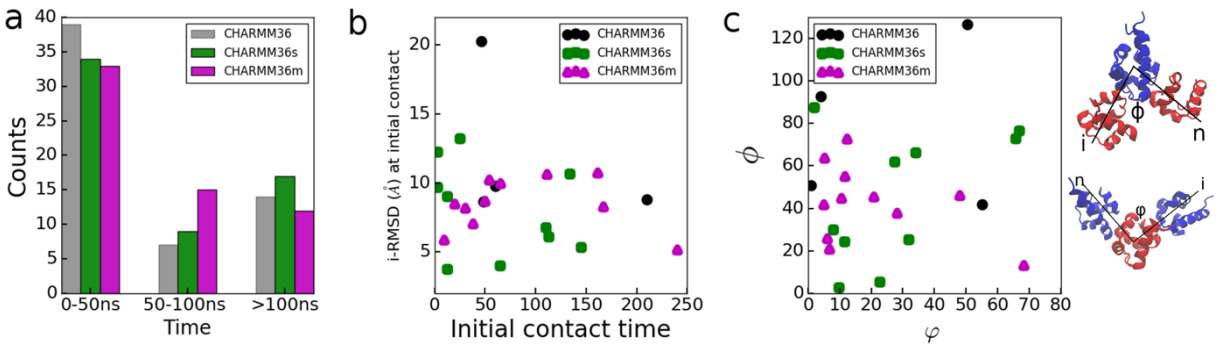

Figure 5: Difference between simulations with original and modified potential function. (a) Initial contact time for 60 independent simulations. (b) Changes of i-RMSD and (c) orientation change between initial contact and final structures in simulations that get to a native complex. ϕ denotes the orientation changes of S-SAM relative to E-SAM (E-SAM is superimposed), comparing the structures at the initial contact time and the time that a native-like complex is formed and φ denotes the orientation change of E-SAM relative to S-SAM (S-SAM is superimposed). Center of mass of E-SAM or S-SAM were used to measure the angles. See Table S1 for results of individual simulations.

Comparing Simulations with Original Potential Function and with Modified Parameters

Given that the potential function modifications overall weaken the protein-protein interactions, we may anticipate a delay in the time to first substantial contact. Indeed, simulations with CHARMM36s and /-m provide a good number of trajectories where contacts are made later than in simulations run with the original potential function (Fig. 5a). However, the simulations with CHARMM36s and /-m also have significantly more dissociation events, 12 and 15 respectively, than simulations with CHARMM36 (8 dissociation events). In each simulation set, there is one native-like complex that is formed after a protein-dissociation event, suggesting that by itself a dissociation does not create conditions for a favorable subsequent association process. For simulations that lead to a native complex, we plot the i-RMSD at the initial contact time (Fig. 5b). It shows that at the initial contact time (between 20-250 ns), 2, 5 and 5 of the complexes have already positioned themselves into a configuration, (i.e. with an Δi-RMSD $\leq$ 5.0Å, in comparison to the 4Å criteria of the formation of a native complex), that is in vicinity of the native complex in the CHARMM36, -/s and /-m simulations respectively. Such a better pre-orientation in the simulations with the modified potential may be observed because during the approach of the two proteins, they are less easily trapped in non-native configurations, but also they have more time to pre-orientate. On the other hand, once the proteins are associated, with the modified potential function, the proteins could also more easily undergo orientation changes to form a native complex. In Fig. 5c, we compare the orientation changes ϕ (or φ) of one protein relative to the other

protein from initial contact time to the formation of the native-like complex. Angle ϕ describes the surface movement of S-SAM relative to E-SAM, while φ is related to the self-rotation of S-SAM (See definition in Fig. 5c). For the CHARMM36, /-s and /-m simulations 0, 5 and 6 of the successful simulations need to undergo a modest orientation change < 45° for each angle. However, 2, 5, 5 simulations, respectively, undergo orientation changes for both angels <90° with at least one angel having a > 45° change (see also Table S1 in Supporting Information). Taken together, the modification on the potential function partially eliminates overly sticky between E-SAM: S-SAM, by allowing a better pre-orientation and easier surface movement, at least over short reorientations of one domain relative to the other. Both processes provide mechanisms to release complexes from trapped states, leading to an enhanced prediction of native-like structures.

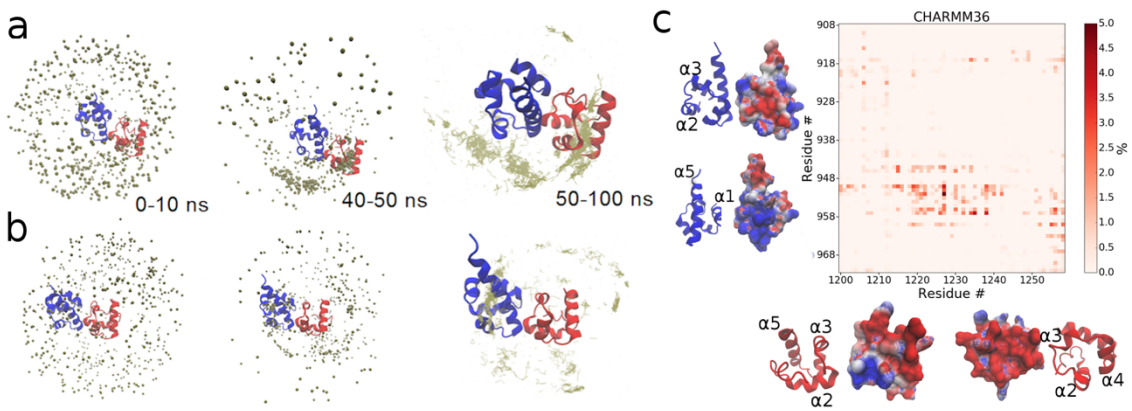

Figure 6: Pre-orientation during the early stage of protein association. (a-b) The early stages of E-SAM and S-SAM association for simulation with the original potential function. Pre-orientation of S-SAM (E-SAM) relative to E-SAM (S-SAM). Proteins are superimposed on E-SAM (a) or S-SAM (b), respectively. The center of mass of S-SAM (or E-SAM) is denoted by small beads for 0-10 ns and 40-50 ns and by lines for 50-100 ns in grey color. (c) E-SAM: S-SAM residue-residue contact map during the early stage of protein-protein association in simulations with original potential function. Simulation trajectories of 0-50 ns are used to construct the contact maps, averaged over 60 trajectories. The color scale (=contact occupancy; max near ~4.8%) is from grey to red as indicated. The contact occupancy is still low as not all proteins are extensively interacting with the same residues. Electrostatic surfaces are shown alongside the respective E- and S-SAM residue axes, with SAM orientations, beside them as mainchain cartoons. The electrostatic surface was calculated with the APBS module of VMD and is shown at ± 4 kT from negative (red), neutral (white) to positive electrostatic potential (blue). Results from the other two potential functions are shown as as Figure S3 in SI.

Protein Association Pattern and Pathway between E-SAM and S-SAM

Electrostatic pre-orientation during the early stage of protein association: The initial contact between E-SAM and S-SAM are made within 50 ns in more than half of the simulations, irrespective of parameter modifications (Fig. 5a). The preferential distribution of the centers of mass of S-SAM relative to E-SAM that results from such pre-orientation and translation events is shown in Fig. 6a for simulations with CHARMM36. The S-SAM move towards to E-SAM, and by 50 ns mostly concentrate on one region made up of helices α4 and α5 (and less so α1) of E-SAM (the coordinate frames are superimposed on E-SAM in all molecular snapshots shown). Intriguingly, while initial contacts are confined to residues ~945-960 on the E-SAM side, no such dramatic confinement exists on the side of

S-SAM (Fig. 6b). Fig. 6c shows the E-SAM: S-SAM residue contact map, averaged over the first 50 ns of the simulations. Consisted with Fig. 6a-b, the surface made by helix α2 and α3 (residues 922-938) of E-SAM (visible as a blue ribbon on left) is rarely involved in the interactions with S-SAM.

The pre-orientation of S-SAM relative to E-SAM is expected to be directed by electrostatic interactions. As shown in Fig. 6c, the primary binding region of S-SAM has many negative charges and thus a negative surface electrostatic potential (esp. far right). However, the region of E-SAM which contains helices α2 and α3 is also negatively charged. Therefore, the interaction between these regions is unfavorable, reflected by the absence of center of mass localization of S-SAM around this region of E-SAM. Instead, the E-SAM region containing helix α4 and α5 is positively charged (electrostatic surface depicted on top left). The two proteins associate with each other through this preferred interface which shows a strong complementarity between the positive and negative electrostatic potential. This supports the viewpoint that an electrostatic pre-orientation (or electrostatic steering) of the domains is involved to direct the early stages of protein association. Negative charged sidechains are more extensively distributed on the S-SAM surface, even though negative charge is more concentrated on one side of the protein. This may explain why there is no dramatic confinement of E-SAM relative to the S-SAM surface. These patterns of preorientation, over the first 50 ns, only change slightly upon modification of the potential function; results for CHARMM36m and /-s are shown in the SI (Fig. S3). This result is not too surprising as the modifications involve only short range vdW contacts directly (and in the case of /-s only vdW contacts to water). Thus, the effect of the potential function change is more pronounced at shorter interaction distances, i.e. once the proteins have associated. Another important point is, however, that the distributions of center of mass of the SAM domains around each other is very similar over the first 10 ns and is extensive, showing a wide sampling of starting orientations of the 60 simulations in each condition.

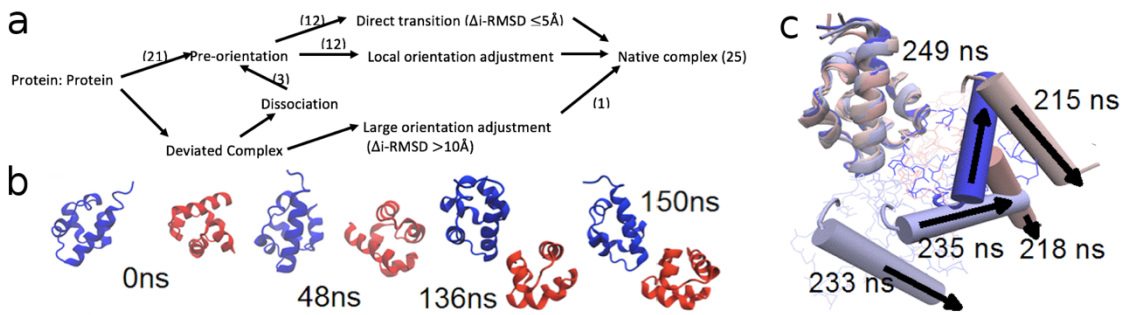

Figure 7: Protein association process between E-SAM and S-SAM. (a) Summary of pathways of all 25 protein-protein associations leading to native complex. (b) Snapshot for a pre-orientation process of the unbound SAM domains (shown for Simulation #33 with CHARMM36). (c) A dynamic transition towards to the native structure occurs in the bound state (shown for Simulation #17 with CHARMM36). Helix 5 of S-SAM are shown in cartoon representation in right panel.

Pathways towards to a native complex: Fig. 7a summaries the formation process of native complex in the simulations with all three potential function. In most of the 25 simulations, the S-SAM is typically already positioned in vicinity of interfaces made of α4, α5 and α1 of E-SAM at the initial contact. For 12 simulations, the change of i-RMSD between the structure at the initial contact time and final native structure is less than or equals 5Å, as noted above. Only relatively small changes are needed to form a

native complex from the pre-orientated positions. The angle change is typically less than 45º. In the other simulations, after initial contact, moderate orientation changes are needed (sliding or rotations). The changes of orientation angle are mostly below 90º (12 simulations) with only two larger than 90º. Taken together, the formation of complex is predominantly through a pre-orientation that is in vicinity of the native complex. If the initial association is too distant from the native structure (for example, S-SAM binding to the opposite side of E-SAM), in principle, the proteins may also directly move around one another by rotational translation/sliding along the binding partner's protein surface to get to the native structure. This relative movement of the two proteins involves a considerable amount of breaking and the establishment of intermediary/new residue-residue interactions. Therefore, if this needs to occur over longer distances/periods of time, it may be less favorable for the formation of a native-like complex. So a trapped sub-stable complex may better convert to a native-like configuration by going through a dissociation-(re)association process, rather than by an extensive movement along the other domains surface to which it is already bound. However, even the complex has a pre-orientation that are close to native complex, with i-RMSD 5-10 Å at the initial contact time, the complex could still become trapped in a non-native structure over the remainder of the simulation time.

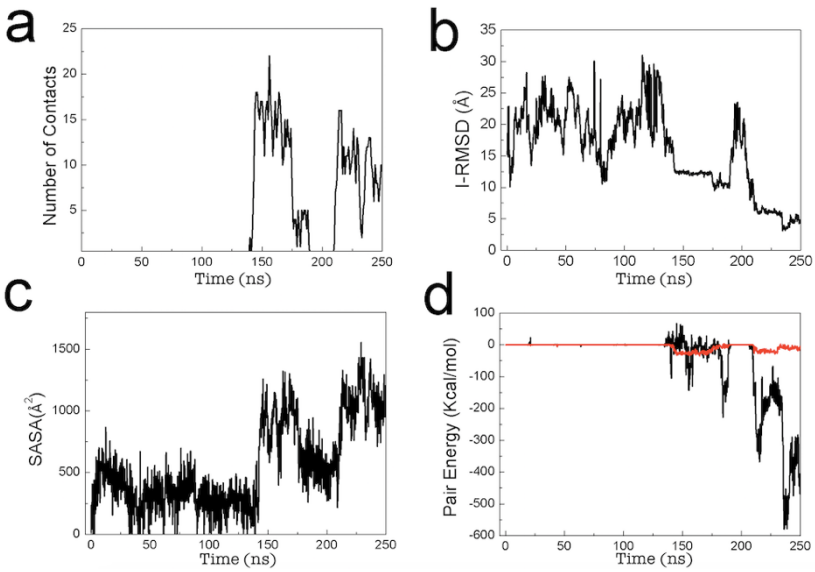

Figure 8: A protein association process (from simulation Simulation #17 with CHARMM36). (a) Number of contacts between residues of the E-SAM and S-SAM domain (b) i-RMSD and (c) buried surface accessible surface area (SASA) of the E-SAM and S-SAM. (d) Protein-protein Electrostatic (black) and vdW interactions (red) between E-SAM and S-SAM.

Fig. 7b and Fig. 7c shows two examples of protein-protein association process with CHARMM36 (see Fig. S4 in SI for simulations with CHARMM36s and –m). In Fig. 7b (shown for simulation #33), a native complex is the result of a good pre-orientation which occurs relatively late at 136 ns. In Fig. 7c (shown for simulation #17), after the initial contact, the S-SAM is largely rotated to transform to the native complex. After first making protein-protein contacts at 215 ns, S-SAM slides along the surface of E-SAM (215-233 ns) and eventually adjusts itself to a native structure through a major rotation (233-249 ns). Some residues at the SAM domain interfaces (in this transition, E-SAM G953/R957 and S-SAM H1219) could be critical for the transitions, acting as pivots, consistent with the concept of "anchoring residue" introduced by Camaco and Colleagues [58]. Before the association at 215 ns, indeed there is a dissociation event. Fig. 8 shows the variation of the number of residue-residue contacts,

interface RMSD (i-RMSD; see Methods), buried surface area, and electrostatic interaction as the proteins form a native-like complex in simulation #17. As the proteins get closer, more residue-residue contacts are established, and more surface area is buried. But at the beginning, the complex still only represents a non-native complex. At a time around 190 ns, the protein dissociated from the non-native structure and established contacts again at 215 ns. After structure adjustment in the next 35 ns, a native-like structure with low i-RMSD value was formed, and in the meantime, a more buried surface accessible surface area (SASA) and lower potential energy for pair interactions are achieved. Protein-protein electrostatic interactions drive the orientation of the proteins to form the complex, but the actual complex formation involves a sizeable desolvation of the protein-protein interface region as a major driving force, reversing the process we noted upon SAM domain dissociation [25].

DISCUSSION

This study shows that the formation of native protein-protein interactions can be enhanced by apparently subtle changes to the potential. By scaling solute-water vdW interactions (in CHARMM36s) to bring the solvation free energies of amino acid sidechains closer to experimental values, we also found that the same increased protein solvation is beneficial for the formation of native PPIs. In CHARMM36m, by reducing the guanidinium-carboxylate interactions between ARG and ASP/GLU, largely the electrostatic attraction between two proteins is reduced and this allows for a greater number of dissociations. The quantum chemistry calculations in potential function parameterization may not fully appreciate the roles of water and thus result in an underestimated solvation of amino acids. It should be noted that improvements in the water model used, did not completely improve the solvation of amino acid sidechains. Except for TIP3P_MOD (which is rarely used and is subjected to a large change of the vdW radius and strength of water), water models TIP3P, TIP4P and SPC all showed underestimated solvation of amino acids [50]. A scaling of solute-water interactions is an empirical way to eliminate the bias for solute-solute and water-water interactions. The scale factor is very small and it only affects the interactions of amino acids side chain with water. The increased solvation may be also beneficial for the prediction of the behavior of intrinsically disordered proteins (IDPs) where solvation is of key importance, or for the protein-protein association of the other systems where the protein-protein interactions are largely mediated by hydrophobic or aromatic residues. CHARMM36m is clearly beneficial for the predictions of native complex. But our work suggests that the reduction of guanidinium-carboxylate interaction may go further than necessary. The recent benchmark study by D. E. Research of multiple popular potential functions indicated that some proteins show reduced stability with CHARMM36m [36]. Therefore, while it largely improves the protein-protein prediction, further studies are still needed to assess the stability of a wide variety of individual proteins and accuracy of CHAMRM36m in protein association in several systems.

The still relatively modest yield of native structures is possibly due to inaccuracies that remain in other aspects of the current potential function parameters, but it is also likely that much longer sampling on the order of tens, if not hundreds of microseconds is required for finding native-like protein complexes [e.g. 16, 59]. Ideally, in protein-protein complex formation, the native structure corresponds to a deep but narrow free energy well, while the non-native sub-stable structures are more wildly spread and have broader minima that are (again ideally) connected to the native well. For example, we notice that the distribution of i-RMSD over the last 50 ns of CHARMMM36m trajectories (Fig. 4) have a minimum between approx. 4.5-6 Å, suggesting that complexes in this range are at the crest of a funnel that quickly

moves them either towards or away from the native-like complex. Beyond this region the energy landscape is likely to be an even more complicated terrain. This complexity, it is suggested, makes the process of finding and maintaining the native structure especially hard in simulations that start from separated proteins and only use physical parameters. Indeed, a number of laboratories are pursuing the use of knowledge-based potentials or combinations of knowledge-based and physical potentials [60-62]. Enhanced sampling methods, such as temperature- or Hamiltonian-based replica exchange or simulated annealing molecular dynamics have been used in sampling configurational space of individual proteins, intrinsically disordered proteins and peptides [16, 59], but are just beginning to be used in the prediction of protein-protein association [e.g. 63,64].

The formation of native complexes largely occurs through a pre-orientation process that could position the S-SAM in vicinity of the PPI of native complex. Are the association and dissociation processes of the SAM-SAM complex the reverse of one another? We could show that the protein-protein association indeed involves less sliding or reorientation if the encounter complex is already close to a native complex. Therefore, in these cases, the association is not the reverse as the dissociation process we observed previously, where (albeit with affinity reduced mutants) transitions to a more weakly binding surface appeared to be obligatory before the proteins could separate [25]. Generally, and despite modification of the potential function, the mechanism of binding and finding the native state complex is overall similar to the reverse of the dissociation trajectories we had with complex destabilizing mutants [25]. Similar to the viewing an association in reverse, coming out of a deep energy minimum for dissociation seems best accomplished by a surface sliding or pivoting movement to a weaker binding interface first. We should caution, however, that our conclusion may not be general as it is based on the study of one model system only; a SAM-SAM complex, which has highly complementary electrostatics and yet –as a cell signaling/adaptor domain interaction- has a much weaker binding affinity than classical protein complexes (e.g. Barnase-Barstar which interact very tightly with a deep, funnel-like energy landscape [65]). The greater number of dissociation-(re-)association events could in principle also help the formation of a better pre-orientated complex but this is more rarely observed. It is intriguing that the effects of parameter change on the ease of dissociation and ease of configurational changes within the complex are similar, while a great number of simulations succeed in going to the native state without a dissociation step. This suggests that the rate determining interactions rely on similar close range contacts. These observations are in agreement with the findings of Feig, Sugita and colleagues [66] who, in all-atom simulations of the cytoplasm of a bacterial cell discovered that metabolites spend a considerable time rolling on protein surfaces, rather than diffusing between proteins.

Since the early pioneering work of R. Wade and colleagues, it is known that the formation of encounter states can be fast and is well modeled by Brownian Dynamics which employ rigid particles with an emphasis on long and medium range interactions [9,10]. On one hand, a close correspondence between theoretical and experimental parameters, such as salt dependence and electrostatics modifying mutants is possible (most experiments measure the formation of the encounter state, rather than of the native state). On the other hand, only a very small fraction of encounter complexes has a sufficient number of native-like contacts, to give small i-RMSD values and convert to high quality complex structures. In many circumstances, the second step, that of a configurational transition of the already bound state, is thought to be rate limiting for the formation of the native complex. Recently, spin-labeling and other types of NMR experiments have opened the door to a residue specific characterization of the initial

bound encounter, if not transient, intermediate non-native bound states of protein complexes [67,68]. Such experiments have started on the SAM: SAM complex in our laboratory.

CONCLUSION and PERSPECTIVE

In summary, we investigated the protein-protein association with an original potential function and two slightly modified potential functions. Using the original CHARMM36 potential function, the native-like structures of the protein complex are obtained in 6.7% of the all-atom unbiased molecular dynamics simulations of the EphA2 (E-)SAM: SHIP2 (S-)SAM heterodimer within 250 ns. By improving the solvation free energy of amino acids side chains in a potential function we denote CHARMM36s, the prediction of native complex is increased to 16.7%. Due to a decrease of electrostatic attraction between ARG and GLU/ASP, the recently published CHAMMR36m potential function largely reduces the protein-protein interactions between Ubiquitin homodimers, and also of the E-SAM: S-SAM complex, yielding native complexes in 18.3% of simulations. In all trajectories, electrostatic interactions between charged residues are a key force in directing protein-protein association. In order to form a native complex, one protein needs to pre-orientate to a position that is in vicinity of the native protein-protein interface of the other. From such a pre-orientated positon, the initial complex could in some cases directly transform to the native complex with small changes (Δi-RMSD≤5Å); Alternatively, in other cases, the formation of a native complex involves moderate changes (5Å <Δi-RMSD <= 10Å; a rotation of one of both domains, mostly by less than 90º). Overall, the study provides a rich picture on mechanisms of protein-protein complex formation in a small model system. In future, computational modeling by itself should be able to predict protein-protein interactions with high accuracy. This will help to direct the site mutagensis experiments, and accelerate the release of structure-level protein interactomes. The reliable prediction of protein complexes would also greatly reduce experimental expense, especially considering that only a minor fraction of structures is currently available of the large protein interactome (predicted to be of at least 130,000 protein complexes for humans [69]). The current study suggests that improvement in the potential function by itself can significantly increase the prediction of protein-protein association. The CHARMM36m and, alternatively, a CHAMM36s potential function with scaled solute-solvent interactions, partially remove current potential function inaccuracies. With increased computing power, physical protein association trajectories may go to 100s of microseconds if not several miliseconds and in combination with further optimized potential functions yield a higher rate of native complex formation.

ADDED NOTE: After submission of an earlier version of this manuscript on bioRxiv, two other groups submitted all-atom simulations on protein association, either using enhanced sampling and a manipulation of the potential function or both. Specifically, in a study by D. E. SHAW research, the researchers developed an enhanced sampling method that could accurately predicate native complexes in lengthy MD simulations. The simulations use a strategy that by constantly scaling down the protein-protein interactions (the vdW and electrostatic potential function parameters; the extent of parameter modification is protein dependent) to help the protein escape from non-native complexes. [https://www.biorxiv.org/content/10.1101/303370v1]. A study of the Chong lab. [Lillian Chong, Univ. of Pittsburgh.] uses a weighted ensemble method to guide the protein association along a pathway ultimately guided by a reduction of i-RMSD. https://www.biorxiv.org/content/10.1101/453985v1.

ACKNOWLEDGEMENTS

This work was supported by NIGMS grant R01GM112491 to the Buck lab. This work used the Extreme Science and Engineering Discovery Environment (XSEDE) Stampede at Texas Advanced Computing Center (TACC); Ohio Supercomputer Center (OSC); as well as local computing resource in the core facility for Advanced Research Computing at Case Western Reserve University. Anton Computer time was provided by the Pittsburgh Supercomputing Center (PSC) through Grant R01GM116961 from the National Institutes of Health.

AUTHOR CONTRIBUTION

ZLL and MB conceived and designed the study; ZLL performed the simulation; ZLL and MB analyzed the data; ZLL and MB wrote the manuscript.

REFERENCES

1. Janin, J. 1997. Kinetics of protein-protein recognition. Proteins: Struct. Fnct. Genen. 28:153-161.
2. Schreiber, G., and G. Haran, and H. Zhou. 2009. Fundamental aspects of protein-protein association kinetics. Chem. Rev. 109:839-860.
3. Khare S. D., S. Fleishman. 2013. Emerging themes in the computational design of novel enzymes and protein–protein interfaces. FEBS Lett. 587:1147-1154.
4. Moal, I. H., R. Moretti, D. Baker, and J. Fernández-Recio. 2013. Scoring functions for protein-protein interactions. Curr. Opin. Struct. Biol. 6:862-867.
5. Vakser, I. A. Protein-protein docking: from interaction to interactome. Biophys. J. 2014. 107:1785-1793
6. Northrup, S. H., and H. P. Erickson. 1992. Kinetics of protein-protein association explained by Brownian dynamics computer simulation. Proc. Natl. Acad. Sci. U.S.A 89: 3338-3342.
7. Camacho C. J., and S. Vadja. 2001. Protein docking along smooth association pathways. Proc. Natl. Acad. Sci. U.S.A 98:10636-10641.
8. Gabdoulline R. R., and R. C. Wade. 2001. Protein-protein association: investigation of factors influencing association rates by Brownian dynamics. J. Mol. Biol. 306:1139-1155.
9. Harel M, Spaar A, and Schreiber G. 2009. Fruitful and futile encounters along the association reaction between proteins. Biophys J. 96:4237-48.
10. Elcock AH, Gabdoulline RR, Wade RC, and McCammon JA. 1999. Computer simulation of protein-protein association kinetics: acetylcholinesterase-fasciculin. J Mol Biol. 291:149-62
11. Ahmad, M., W. Gu, and V. Helms. 2008. Mechanism of fast peptide recognition by SH3 domains. Angew. Chem. Int. Ed. Engl. 47:7626-7630.
12. Chong, S. H., and S. Ham. 2012. Impact of chemical heterogeneity on protein self-assembly in water. Proc. Natl. Acad. Sci. USA. 109:7636-7641.
13. Blöchliger, N., M. Xu, and A. Caflisch. 2015. Peptide Binding to a PDZ Domain by Electrostatic Steering via Nonnative Salt Bridges. Biophys. J. 108:2362-2370.
14. Nawrocki G., P. Wang, I. Yu, Y. Sugita, and F. Michael. 2017. Slow-Down in Diffusion in Crowded Protein Solutions Correlates with Transient Cluster Formation. J. Phys. Chem. B, 121:11072–11084.
15. Schmidt AG, Xu H, Khan AR, O'Donnell T, Khurana S, King LR, Manischewitz J, Golding H, Suphaphiphat P, Carfi A, Settembre EC, Dormitzer PR, Kepler TB, Zhang R, Moody MA, Haynes BF, Liao HX, Shaw DE, and Harrison SC. (2013) Preconfiguration of the antigen-binding site

during affinity maturation of a broadly neutralizing influenza virus antibody. Proc Natl Acad Sci U S A. 110:264-269.
16. Paul F, Wehmeyer C, Abualrous ET, Wu H, Crabtree MD, Schöneberg J, Clarke J, Freund C, Weikl TR, Noé F. 2017 Protein-peptide association kinetics beyond the seconds timescale from atomistic simulations. Nat Commun. 8:1095.
17. Khare, S. D., and S. Fleishman. 2013. Emerging themes in the computational design of novel enzymes and protein–protein interfaces. FEBS Lett. 587:1147-1154.
18. Bonvin, A. M. 2006. Flexible protein-protein docking. Curr. Opin. Struct. Biol. 16:194-200.
19. Zacharias, M. 2010. Accounting for conformational changes during protein-protein docking. Curr. Opin. Struct. Biol. 20:180-186.
20. Kimura, S. R., R. C. Brower, S. Vajda, and C. J. Camacho. 2001. Dynamical view of the positions of key side chains in protein-protein recognition. Biophys J. 80:635-642.
21. Pabon, N. A., and C. J. Camacho. 2017. Probing protein flexibility reveals a mechanism for selective promiscuity. Elife. 6: e22889
22. Zhang, L., and M. Buck. 2013. Molecular Simulations of a Dynamic Protein Complex: Role of Salt-Bridges and Polar Interactions in Configurational Transitions. Biophys.J. 105:2412-2417.
23. Zhang, L., and M. Buck. 2017. Molecular Dynamics Simulations Reveal Isoform Specific Contact Dynamics between the Plexin Rho GTPase Binding Domain (RBD) and Small Rho GTPases Rac1 and Rnd1. J Phys Chem B. 121:1485-1498
24. Lensink MF, Velankar S, Wodak SJ. 2017. Modeling protein-protein and protein-peptide complexes: CAPRI 6th edition. Proteins. 85:359-377.
25. Zhang, L., S. Borthakur, and M. Buck. 2016. Dissociation of a Dynamic Protein Complex Studied by All-Atom Molecular Simulations. Biophys. J. 110:877-886.
26. Kozack R. E., M. J. d'Mello, S. Subramaniam. 1995. Computer modeling of electrostatic steering and orientational effects in antibody-antigen association. Biophys. J. 68:807-814.
27. Ubbink M. 2009. The courtship of proteins: understanding the encounter complex. FEBS Lett. 583:1060-1066
28. Tsai, M. Y., W. Zheng, D. Balamurugan, N. P. Schafer, B. L. Kim, M. S. Cheung, and P. G. Wolynes. 2016. Electrostatics, structure prediction, and the energy landscapes for protein folding and binding. Protein Sci. 25:255-269
29. Kozakov, D., K. Li, D. R. Hall, D. Beglov, J. Zheng, P. Vakili, O. Schueler-Furman, I. C. Paschalidis, and G. M. Clore. 2014. Encounter complexes and dimensionality reduction in protein–protein association. eLife 3:e01370
30. Zhou HX, and Bates PA. 2013. Modeling protein association mechanisms and kinetics. Curr Opin Struct Biol. 23:887-93.
31. Petrov, D., and B. Zagrovic. 2014. Are Current Atomistic Force Fields Accurate Enough to Study Proteins in Crowded Environments? PLoS Comput. Biol. 10:e1003638.
32. Abriata, L. A., and D. P. Matteo. 2015. Assessing the potential of atomistic molecular dynamics simulations to probe reversible protein-protein recognition and binding. Sci. Rep. 5:10549.
33. Rauscher, S., V. Gapsys, M. J. Gajda, M. Zweckstetter, B. L. de Groot, and H. Grubmuller. 2015. Structural Ensembles of Intrinsically Disodered Proteins Depend Strongly on Force Field: A Comparision to Experiment. J. Chem. Theory Comput. 11:5513-5524.
34. Robustell,i P., S. Piana, and D. E. Shaw. 2018. Developing a molecular dynamics force field for

both folded and disordered protein states. Proc. Natl. Acad. Sci. U S A. DOI: 10.1073/pnas.1800690115.
35. Piana, S., A. G. Donchev, P. Robustelli, and D. E. Shaw. 2015. Water Dispersion Interactions Strongly Influence Simulated Structural Properties of Disordered Protein States. J. Phys. Chem. B, 119:5113-5123.
36. Huang. J., S. Rauscher, G. Nawrocki, T. Ran, M. Feig, B. L. de Groot, H. Grubmüller, and A. D. MacKerell Jr. 2017. CHARMM36m: an improved force field for folded and intrinsically disordered proteins. Nat. Methods. 14:71-73.
37. Best, R. B., W. Zheng, and J. Mittal. 2014. Balanced Protein–Water Interactions Improve Properties of Disordered Proteins and Non-Specific Protein Association. J. Chem. Theory Comput. 10:5113–5124.
38. Nerenberg, P. S., and T. Head-Gordon. 2011. Optimizing Protein-Solvent Force Fields to Reproduce Intrinsic Conformational Preferences of Model Peptides. J. Chem. Theory Comput. 7:1220–1230.
39. Nerenberg, P. S., B. Jo, C. So, A. Tripathy, and T. Head-Gordon. 2012. Optimizing Solute–Water van der Waals Interactions To Reproduce Solvation Free Energies. J. Phys. Chem. B 116:4524–4534.
40. Stark, A. C., C. T. Andrews, and A. H. Elcock. 2013. 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. 9:4176-4185.
41. Lindorff-Larsen, K., P. Maragakis, S. Piana, M. P. Eastwood, R. O. Dror, and D. E. Shaw. 2013. Systematic Validation of Protein Force Fields against Experimental Data. PLOS One 8:10.
42. Debiec K. T., M. J. Whitley, L. M. I. Koharudin, L. T. Chong, and A. M. Gronenborn. Intergrating NMR, SAXS, and atomistic simulations: structure and dynamics of a two-domain protein. Biophys. J. 114: 839-855.
43. Lee, H. J., P. K. Hota, P. Chugha, H. Miao, L. Zhang, S. J. Kim, R. S. Alviani, L. Stetzig, B. Wang, and M. Buck. 2012. NMR structure of a heterodimeric SAM: SAM complex: Characterization and manipulation of EphA2 binding reveal new cellular functions of SHIP2. Structure 20:41-55.
44. Mackerell, A. D., et al. 1998. All-atom empirical potential for molecular modeling and dynamics studies of proteins. J. Phys. Chem. B 102: 3586-3616.
45. Huang, J., and A. D. MacKerell. 2013. CHARMM36 all-atom additive protein force field: validation based on comparison to NMR data. J. Comput. Chem. 34:2135-2145.
46. Buck, M., S. Bouguet-Bonnet, R. W. Pastor, A. D. MacKerell. 2016. Importance of the CMAP correction to the CHARMM22 protein force field: dynamics of hen lysozyme. Biophys. J. 90: L36-L39.
47. Jorgensen W. L., J. Chandrasekhar, J. D. Madura, R. W. Impey, and M. L. Klein. 1983. Comparison of simple potential functions for simulating liquid water. J. Chem. Phys. 79:926–926.
48. MacCallum JL, Tieleman DP. 2003. Calculation of the water-cyclohexane transfer free energies of neutral amino acid side-chain analogs using the OPLS all-atom force field. J. Comput. Chem. 24:1930-1935.
49. Shirts MR, Pitera JW, Swope WC, Pande VS. 2003. Extremely precise free energy calculations of amino acid side chain analogs: Comparison of common molecular mechanics force fields for proteins. J. Chem. Phys. 119: 5740.
50. Shirts MR, Pande VS. 2005. Solvation free energies of amino acid side chain analogs for common molecular mechanics water models. J Chem Phys.122:134508.

51. Chang J, Lenhoff AM, and Sandler SI. 2007. Solvation Free Energy of Amino Acids and Side-Chain Analogues. J. Phys. Chem. B, 111, 2098–2106.
52. Essmann U., L. Perera, M. L. Berkowitz, T. Darden, H. Lee, and L. G. Pedersen. 1995. A smooth particle mesh Ewald method. J. Chem. Phys. 103:8577-8593.
53. Ryckaert J. P., G. Ciccotti, and H. J. C. Berendsen. 1977. Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. J. Comput. Phys. 23:327–341.
54. Phillips, J. C., R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, E. Villa, C. Chipot, R. D. Skeel, L. Kale, and K. Schulten. Scalable molecular dynamics with NAMD. 2005. J. Comput. Chem. 26:1781-1802.
55. Shaw DE, Grossman JP, Bank JA, Batson B, Butts JA, Chao JC, Deneroff MM, Dror RO, Even A, Fenton CH, Forte A, Gagliardo J, Gill G, Greskamp B, Ho CR, Ierardi DJ, Iserovich L, Kuskin JS, Larson RH, Layman T, et al. 2014. Anton 2: Raising the bar for performance and programmability in a special-purpose molecular dynamics supercomputer. SC14: International Conference for High Performance Computing, Networking, Storage and Analysis, vol.00, no., pp.41-53. doi:10.1109/ SC.2014.9.
56. Wolfenden, R.; Andersson, L.; Cullis, P. M.; Southgate, C. C. B. 1981. Biochemistry, 20, 849-855.
57. Liu Z, Zhang WP, Xing Q, Ren X, Liu M, Tang C. 2012. Noncovalent dimerization of ubiquitin. Angew Chem Int Ed Engl. 51:469-472.
58. Keskin, O., N. Tuncbag, and A. Gursoy. 2016. Predicting Protein-Protein Interactions from the Molecular to the Proteome Level. Chem. Rev. 116:4884-4909.
59. London, N., B. Rave, O. Schueler-Furman. 2013. Peptide docking and structure-based characterization of peptide binding: from knowledge to know-how. Curr. Opin. Struct. Biol. 23:894-902
60. Lee, H., M. Baek, C. R. Lee, S. Park, and C. Seok. 2017. Template-based modeling and ab initio refinement of protein oligomer structures using GALAXY in CAPRI round 30. Proteins 85:399-407.
61. Rajamani, D., S. Thiel, S. Vajda, and C. J. Camacho. 2004. Anchor residues in protein-protein interactions. Proc. Natl. Acad. Sci. U S A. 101:11287-11292.
62. DeGrave AJ, Ha JH, Loh SN, Chong LT. 2018. Large enhancement of response times of a protein conformational switch by computational design. Nat Commun. 9:1013.
63. Pan AC, Jacobson J, Yatsenko K, Sritharan D, Weinrich TM, and Shaw DE. 2018 Atomic-level characterization of protein-protein association. Preprint posted online at bioRxiv on Apr. 17th., http://dx.doi.org/10.1101/303370.
64. Huang W, Peng Y, Kiselar J, Zhao X, Albaqami A, Mendez D, Chen Y, Chakravarthy S, Gupta S, Ralston C, Kao HY, Chance MR, Yang S. 2018. Multidomain architecture of estrogen receptor reveals interfacial cross-talk between its DNA-binding and ligand-binding domains. Nat Commun. 9:3520.
65. Wang L, Siu SW, Gu W, and Helms V. 2010 Downhill binding energy surface of the barnase-barstar complex. Biopolymers. 93:977-85.
66. Yu, I., T. Mori, T. Ando, R. Harada, J. Jung, Y. Sugita, and M. Feig. Biomolecular interactions modulate macromolecular structure and dynamics in atomistic model of a bacterial cytoplasm. 2016. eLife. 5: e19274.

67. Andrałojć W, Hiruma Y, Liu WM, Ravera E, Nojiri M, Parigi G, Luchinat C, Ubbink M. 2017. Identification of productive and futile encounters in an electron transfer protein complex. Proc Natl Acad Sci U S A. 114:E1840-E1847.
68. An SY, Kim EH, Suh JY. 2018. Facilitated Protein Association via Engineered Target Search Pathways Visualized by Paramagnetic NMR Spectroscopy. Structure. 26:887-893.e2.
69. Kavitha Venkatesan, Jean-François Rual, Marc Vidal. An empirical framework for binary interactome mapping. Nature Methods 2009: e6, 83–90.