Protein Association Pathways and Dynamics with Improved Potential Functions in all atom MD Simulations

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

Characterizing the features that are important for protein-protein interactions is the cornerstone for understanding the structure, dynamics and function of protein complexes. In this study, we investigate the heterodimer association of SAM domains of the EphA2 receptor and SHIP2 enzyme by performing a combined total of 60 µs of all atom molecular dynamics simulations. In the native complex, the two SAM domains interact using charged surfaces which are highly complementary. However, in simulations of 250 ns, many of the initial protein complexes are trapped into non-native configurations. A few SAM domains associate to form heterodimers through pre-orientation towards positions that are close to those in the native-complex. In this case, only small adjustments are needed, but in other trajectories, following protein association, large configurational movements (sliding and pivoting) of one SAM domain on the protein surface of the other are seen. As part of this mechanism, dissociation-(re-)association events are observed as well, helping the formation of native-like complexes. We tested a set of potential functions, by slightly weakening the interaction between guanidinium and carboxylate (applied in the most recently released CHARMM36m), or by slightly increasing the solvation of protein polar sidechain groups (5% scaling of the vdW interaction energy) in the CHARMM36 potential function. In both cases the prediction of native-like SAM complexes is significantly enhanced from a 10% yield of native complexes with the original potential function to 18%-25% with the modified parameters. Native-like protein complexes are formed by more easily allowing conformational transitions as well as dissociation events. These observations point to a way for the improvement of computational predictions of protein-protein interactions and complexes in general, where – analogously to protein folding- trapped intermediate states are considered as detrimental kinetically to the protein association process.

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

Running title: Not getting stuck in protein association

INTRODUCTION

Biological function that arises from the formation (and/or dissociation) of protein complexes critically depends on their population dynamics, which are, in turn, determined by the strength and kinetics of the underlying protein-protein interactions. Despite the significance of these features for molecular

recognition, detailed knowledge of protein-protein interactions and understanding of complex formation is far from complete [1-5]. Especially, an appreciation of protein association processes is still lacking for the great majority of systems at the molecular level. This is in part due to the limited temporal-spatial resolution of experimental methods. Despite early successes in Brownian and coarse grained molecular dynamics simulations, matching kinetic trends of protein complex formation [2,6-8], to our knowledge, there have as yet been very few simulations of ab initio protein association processes using all atom simulations [9-12]. Furthermore, until relatively recently most protein docking computer protocols could not incorporate sidechain (or even mainchain) flexibility [13-15], but the role of local, as well as global dynamics in protein complexes and complex formation is now becoming well appreciated [16-19]. All atom molecular dynamics simulations can provide information on the dynamic process of protein association, that is not accessible by current experiments, including on the physical-chemical details of the interactions. For example, recently, we reported all-atom simulations on the dissociation process of a small protein heterodimer, the EphA2-SHIP2 SAM: SAM domain complex [20]. 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 [20]. 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 proceed via intermediary bound states. States and processes associated with protein complex formation, such as electrostatic steering have been reported and discussed in the literature over decades [21-23]. Availability of computer resources to extensively simulate the association events, motivated us to carry out unrestrained all-atom simulations to study protein complex formation between the SAM domains of EphA2 and SHIP2.

The protein domain Sterile alpha motif (SAM) is found in more than 200 human proteins [24]. As an example, the SAM domain is an important component of the intracellular region of the transmembrane Ephrin receptor [25]. This largest tyrosine kinase family in the human genome is involved in axon guidance and cell migration and also plays a pivotal role in several cancers [26]. Moreover, in some cases, the SAM domain is used to form self-associated structures to mediate protein-protein interactions. For example, a heterodimer formed by the EphA2 SAM domain (denoted E-SAM henceforth) and the SAM domain of the enzyme SHIP2 (denoted S-SAM) has been solved by us, using restraints from solution NMR experiments and was found to be an intrinsically dynamic structure [27]. This view was reinforced by extensive all atom MD simulations which showed that the protein complex could convert between multiple configurational states on a ~ 100 ns timescale [18]. 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 perturbed the energy landscape of the complex [20]. The SAM domain is particularly suitable for computational studies of protein-protein association, considering its small size (around 70 residues) and well folded structure with 5 alpha helices (Fig. 1a). The distribution of charged residues is remarkable (a region of the surface of the EphA2 SAM is entirely positively, another region of the SHIP2 SAM domain is entirely negatively charged; Fig. 2). Furthermore, upon association the protein undergoes negligible internal conformational, but substantial internal dynamics changes [20]. 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.

In this study, we performed multiple and extensive simulations of EphA2 SAM: SHIP2 SAM (E-SAM: S-SAM) association, initially with the original CHARMM36 potential function. However, while the

SAM domains associated on a reasonable timescale, the initial associated states were often "trapped", in that they showed few interconversions and fewer dissociation events, than would be expected from encounter complexes [28]. Indeed, native protein complexes are very hard to be captured in the unrestrained simulations. The yield rates of native protein complexes are typically low in free self-assembly simulations of the two SAM domains, it appears, because the proteins are often trapped in states that are far from the native structure. If the energy barriers to the correct configuration of the complex are are high, the protein domains appear kinetically trapped within the simulation time limit. We also observed such trapped states in a few of the native-like simulations in our original report, which showed that configurational transitions involve intermediary salt bridges in this system [18]. In recent years, the molecular dynamics community has appreciated that the current popular potential functions appear to lack accuracy in cases of simulations involving protein association and in the behavior of intrinsically denatured proteins (IDPs) [29]. In the former, many proteins -even those having only non-specific non-covalent self-association (such as ubiquitin)- showed strong and unphysical aggregation [12,30]. IDPs also became too compact in the simulations [31, 32]. Both effects likely arise from an underestimation of protein solvation and both are in disagreement with experimental observations. Supposing the interface between two soluble proteins is overall hydrophilic, the underestimation of protein solvation likely over-enhances protein-protein association. To address these problems, researchers have tried to improve potential function parameters in different ways. In one study Piana et al. developed a new water model, which could better reproduce intrinsically disordered proteins [33], but has recently been refined to also work on folded proteins [32]. A most recently released version of the CHARMM potential function, CHARMM36m, improves the prediction of α-helices and intrinsic disorder in polypeptides [34]. Best et al. have explored a scaling of protein-water interactions with the Amber potential function [35], which showed an improved prediction in intrinsic disordered protein and non-specific protein association. 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 [12]. Several other studies have rescaled the protein-protein interactions or protein-water interactions, or mixed rescaling with different water models (such as TIP4P) in order to get a better performance in prediction of protein configurations or protein-protein association [36-40].

Here, we investigated the SAM: SAM associations with all atom simulations. We examined the performance of the simulations in predicting the native protein-protein complex structure, in a most straightforward way by simply counting the yield of native-like structures after an extensive amount of sampling (60 simulations for each potential function tried).We compared the simulation results with the original parameters, with those using the original CHARMM36 potential function and with the most recently released CHARMM36m function, and with simulations using a modified potential function by scaling certain solute-solvent vdW interactions. The simulations with modified potential function parameter showed an an improved prediction in native contacts. In the early stage of protein association, an electrostatic patch directs the two proteins to interact by interfaces with a high level of electrostatic potential complementarity. But there are several competing charged patches on the SAM domain surfaces, so only a few of those interfaces make contacts to adequately correspond to the native protein complex structure. In cases where either or both SAM pre-orientate with respect to the other, assuming positions and orientations that are close to those of the native-complex, only small adjustments are needed after association to form the native structure, however, in other trajectories, large configurational

movements (sliding and pivoting) are involved. Dissociation-(re-)association events are also observed in a few of the trajectories and are predicted to be beneficial in forming a native-like complex. Implications of these findings for protein association and all-atom potential functions are discussed.

METHODS

The E-SAM and S-SAM domains of the NMR derived complex (pdb.id 2kso, see Fig. 1a) were separated to a distance of 4 nm between their center of mass. Then both proteins are rotated randomly with respect to their center of mass to produce different initial orientations for each simulation, but the same initial structures for each set. Following a very brief equilibration (see below), the initial deviation between native-like interfaces, i-RMSD, was essentially the same at 8-24 Å, with an average at 16 Å (not shown). The whole system was solvated by TIP3P water in a simulation box of $9\times9\times9$ nm^3, setting up periodic boundary conditions. Sodium and Chloride ions were added to a near-physiological concentration of 150 mM. Sidechain groups were charged as expected at pH 7.0 (HSD was used for neutral histidine). In total, the simulation box has 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. 18]). 60 independent simulations were performed. For the first 10 simulations, a harmonic potential with force constant 100 Kcal/mol/$Å^2$ was applied if the two proteins were greater than 5 nm apart. For the other 50 simulations, there are no restraints---two SAM are totally free in the simulation box, and upon leaving the box they re-enter on the opposite side as per period boundary imaging conventions.

Simulations were performed with the original CHARMM36 potential function including CMAP correction [41-43]. We also tested the CHARMM36m, which was recently released with major changes in the CMAP potential, and a slight reduction of guanidinium and carboxylate salt bridge interaction by modifying vdW radii [34]. We made a slightly different modification on the original CHARMM36 potential function parameter [42] by scaling the solute-solvent vdW interaction potential energy to different extents (by 3%, 5% and by 10%). The principle of our approach was to make a minimal modification to the current potential function parameters. For soluble proteins, the polar/charged residues are mostly exposed to bulk water while the nonpolar residues are largely buried during the hydrophobic collapse. Therefore, the polar/charged residues (more specifically, their sidechain groups) are involved in protein association. In addition, modifications on the mainchain atoms made the protein fold unstable (data not shown). So here, instead scaling protein-water for all protein atoms [12, 35], only the side chain atoms, specifically the terminal group atoms, were subjected to changes. Changes consisted of a scaling of the atoms interaction with water with scale factor λ. These atoms affected are indicated in Table S1 and the scaled term is described in the Figure legend. Scale factor λ was set as 1.03, 1.05 and in a final set of simulations to 1.10. In this way, the interaction of the sidechain of the polar and charge residues with the water was more favorable and the solvation of SAM domains is increased. The early part of the simulations with scale factor of 1.10 (10 simulations of 50-100 ns) showed that, the proteins have a difficult time to make contacts and form complexes, suggesting that λ at 1.10 affords too much solvation and does not improve the performance of the protein association. So a scaling with such a value and beyond was not considered further. In summary, we have four kinds of different set-ups: simulations with original CHARMM36 potential function (λ=1.00), simulations with CHARMM36 with a scale factor λ at1.03 or at 1.05, and simulation with CHARMM36m potential function. For brevity, these set-ups are denoted as S_{100}, S_{103}, S_{105} and S_{100m}.

In order to enhance sampling, 60 independent simulations were performed for each kind of set-up. Each simulation was 250 ns long. In all simulations, the van der Waals (vdW) potential was truncated at 1.2 nm (default for CHARMM36/-m). The Particle-Mesh Ewald (PME) method was used for calculating the long distance electrostatic interactions. 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. All these systems were simulated using the NAMD/2.12 package [44]. Analysis (interface-RMSD, residues-residues contact, surface electrostatics, orientational angle, pair interactions, and solvent accessible surface) were done with VMD and NAMD with in-lab scripts (for descriptions see figure legends/main text). 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 considered. A protein complex with an i-RMSD < 4 Å with respect to native complex is considered as native-like.

RESULTS

The early stage of E-SAM: S-SAM association:

Prior to the formation of a stable complex, the early stage of protein association is a rather dynamic process, in that the two proteins typically adjust the relative orientation with respect to each other. However, the initial contact between E-SAM and S-SAM are made within 50 ns in about half of the simulations (Figure 1b). Prior to initial contact formation, pre-orientation events of S-SAM relative to E-SAM were observed (Fig. 1c). 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. 1d).

Fig. 2 shows the E-SAM: S-SAM residue contact map, averaged over the first 50 ns of the SAM: SAM association process seen in 60 independent simulations. Clearly, several residues are significantly involved in < 4Å contracts, while other residues are rarely involved in the inter-protein interactions. The surface made by helix α-2 and α-3 (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. 2, 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 α-2 and α-3 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 α-4 and α-5 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 are no dramatic confinement of E-SAM on the S-SAM surface.

E-SAM: S-SAM association pattern, dynamics and pathway with the original CHARMM36 potential function:

Within 250 ns of the simulation start, the great majority of the trajectories enter into relatively stable association states. However, most of the simulations (54/60) yielded configurations in the final E-SAM: S-SAM complexes, that are far from the native structures, either displaced by translation and/or by rotation.

Native-like structures are obtained in only 6 of 60 simulations based on the deviation of the final structure relative to the native complex (interface RMSD, see Method). In three of the simulations that form a native-like complex, the process of pre-orientation – that precedes formation of substantial contacts- is used to direct the two proteins toward a position that is close to the native structure. They can then directly transform into the native-like structure with only small structural readjustment (Fig. 3a; shown for simulation S_{100}-7 where pre-orientation continues to 16.3 ns, see others in Fig. S1). In the other three simulations, the formation of native-like complex involves additional movement after the initial association (e.g. S_{100}-17 in Fig. 3b). After protein-protein contact 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, a concept of "anchoring residue" introduced by Camaco and Colleagues [45]. It should be recognized that although there are large adjustments after the initial association to yield a native-like structure, the adjustments first proceeded in a direction away from the target structure, as the initial association position was actually already close to that of the native structure, especially S-SAM was already bound to regions of α4, α5, α1 of E-SAM. Dissociation events are found in 9 of 60 simulations, but only one yields a native complex after dissociation and (re-)association (also S_{100}-17 shown in Fig. 3c).

As an example, for simulation S_{100}-17, Fig. 4a-d show the variation of the number of residue-residue contact, i-RMSD, buried surface area, and electrostatic interaction as the proteins form a native-like complex. 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 time around 190 ns, the protein dissociated from the non-native structure and established contacts again at 215 ns. As mentioned above, after structure adjustment in the next 35 ns, a native-like structure with low i-RMSD value structure was formed, and in the meantime, lower buried SASA and lower pair interactions are achieved. Protein-protein electrostatic interactions drive the orientation of the proteins to form the complex, but the actual complex formation involves the desolvation of the protein-protein interface region as a major driving force, reversing the process we noted upon SAM domain dissociation [20].

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. However, this process is seen in very few of the simulations with the original potential function. 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 dissociation-(re)association process, rather than by extensive movement along the other domains surface to which it is already bound.

Enhanced prediction of E-SAM: S-SAM association with several different modifications to the potential function:

Despite the role of long range electrostatic interactions in establishing a pre-oriented, but not yet bound complex, we find that the dissociation and re-association as well as the configurational transitions once bound are relatively rare with the standard CHARMM36 potential function. Raising the simulation temperature (from 300K to 323K or 350K) or increasing the salt concentration (from 150 mM to 1.5 M) did not overcome this problem, even if simulations are run to the length of several μs (data not shown). Thus, long range electrostatics, while important for the initial encounter complexes, are not able to distinguish the major interfaces from competing interfaces. The "culprit", inferred from several papers along similar lines in the literature (see introduction) appears to be polar-solvent interactions that could be too weak; thus we also ran 60 simulations with the vdW interaction energy between certain protein polar groups and water scaled up by a factor of 1.03, 1.05. Also we tested the performance of the newly released CHARMM36m, where the salt bridge interactions between protein sidechain guanidinium and carboxylate groups were reduced.

With the modified potential functions, the association process yields more native-like protein complex structures. Fig. 5 plots the distribution of i-RMSD. Clearly the simulations with modified force parameters have a better performance than simulations with original CHARMM36 potential function. Specifically, while only 6 simulations get to the native-complex in S_{100}, 13, 15 and 11 native structures are obtained in simulations sets of S_{103}, S_{105}, and S_{100m}. So the yield of native structures was increased from 10% to 21%, 25%, and 18%, respectively. Fig. 6 compares the difference in residue-residue contact patterns (for the last 50 ns of the simulations), between simulations with original and modified parameters. The pattern and population of native contacts are largely enhanced in simulation with modified parameters with reference to the residue-residue contact map obtained from our previous simulations starting from the native complex (State 1 and State 2) [18]. Clearly, the major region of native contacts, E-SAM res. 952-957: S-SAM res. 1220-1227 is enhanced in simulations with the modified potential function, surprisingly even in cases where a native-like complex is not formed, one of the two proteins uses the correct high affinity interface for making contacts.

Difference between simulations with original potential function and with modified parameters

As in the simulations with original potential function, in the early stage of protein association with modified potential function, electrostatic interaction driven pre-orientations similarly direct the initial protein association. Protein-protein association was established within 50 ns for about half of the complexes (Fig. S2a). The S-SAM positions itself to a region close to helices α4 and α5 of E-SAM (Fig. S2b), but the confinement of E-SAM on the surface of S-SAM is less dramatic as mentioned above (Fig. S2c). Overall , based on the residue contact maps for the first 50 ns of the simulations (Fig. S3), there is little difference in the quality of the pre-orientations of S-SAM towards to α4 and α5 of E-SAM, thus the enhanced success of the simulations with enhanced parameters must originate elsewhere.

In Fig. 7, the pair interactions between E-SAM and S-SAM domains in the final 50 ns simulations complex are plotted as a function of i-RMSD, averaged for each set of simulations. Overall, the native-like structures (low i-RMSD) correspond to lower (i.e. stronger) pair interaction energies. The farther the complex deviates from the native structure, the higher (i.e. the less favorable) is the pair interaction energy (vdW and electrostatic interaction). Of course, the direct residue pair interactions between E-SAM and S-SAM only contribute to a part of the total free energy. But it is still one of the most significant contributions and it roughly represents the low-energy property of the native structure. The E-SAM: S-SAM interaction interface is overall hydrophilic with residues pairs across the protein-protein interface formed between charged residues and polar residues. As noted previously, there is only a small hydrophobic patch surrounding Trp1222 [20]. As we increase the protein solvation in our simulations, the side chains of polar residues will interact with the solvent more favorably, so we expect pair interactions between E-SAM and S-SAM to become less favorable. This is revealed in Fig. 7 as an overall trend, comparing the simulations of S_{103} and S_{105} to simulations with original potential function. However, the correction of guanidinium and carboxylate salt bridge interaction (typically established between arginines [but not NH3+ or NH2 of Lysine and Asn/Gln] with glutamic acid and aspartic acid) in CHARMM36m is also good for sampling native-like structures, without a major change in SAM-domain pair interaction energies.

Comparing the length of time spent from protein initial contact to the formation of a native-like complex (Fig. 8a), the simulations with modified the potential function overall take a slightly longer time to form a native-like complex. This implies that the simulations with the modified potential function more easily allows transitions form less favorable states to native-like state, and/or that these kinds of transitions may take a longer time. In order to characterize the dynamic change of the protein complex from the initial contact to the formation of a native-like complex, we calculated the angle change of S-SAM (or E-SAM) relative to E-SAM (or S-SAM), ϕ (or φ). 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. 4b). As shown in Fig. 8c, in some of the simulations, both the angle changes are relative small. This means that at the time of initial contact, the complex is already reasonably close to the native-like structure. With the potential function modification, there are more cases where both angles have larger changes (> 40°); so in these cases, the enhanced yield of native-like complex is due to a better transition on the surface of the protein once associated. This is likely because weaken protein-protein interactions make the proteins less easy to be trapped into sub-stable states. Fig. 8d-e shows two large configuration change events. In simulation of S_{105}-33, the S-SAM slides from a position top of E-SAM to the bottom E-SAM, then it shifts to native-like structure (Fig. 8d). In simulation of S_{105}-55, S-SAM makes a large orientation change from a distal position left of E-SAM towards to the bottom of E-SAM, then shifts to a region close to α4 and α5 of E-SAM; from this position, S-SAM rotates ~180° with respect to itself to eventually reach the native-like state (Fig. 8e). These transitions indicate that the modification of the potential function, allows large surface movement of one relative domain relative to the other. In addition, there are significantly more dissociation events in simulations with modified parameters (9, 15, 16 and 16 simulations have dissociation events for simulation set S_{100}, S_{103}, S_{105} and S_{100m}, respectively). For these four different kinds of set-ups respectively, 1, 3, 3 and 1 simulations yield a native-like complex through protein association-(re)association process (Fig. 8f).

Taken together, by a 5% scaling-up the protein-solvent interactions or (less so) by correction the guanidinium and carboxylate van der Waals radii, more extensive movements of one protein on the surface of the other are made possible and the parameter change also allows more dissociation events. Both processes provide mechanisms to release complexes from trapped states, leading to an enhanced prediction of native-like structures.

DISCUSSIONS

The relative weakening in protein-protein interactions, evident by the reduced pair-energy (y-axis) of plots in Fig. 7, helps the E-SAM: S-SAM complex escape from some of the non-native structures it may be trapped in. However, since only the protein-solvent interaction was scaled up, the effect must be indirect. Specifically, we previously inferred a sizeable contribution of protein-solvent interactions at/near the interface in the bound state/process of dissociation [20]. Such interactions were strengthened in the potential function modification, including waters that bridge protein-protein interactions, to the detriment of pure protein-protein interactions. This scenario of indirect effects illustrates the well-known difficulty associated with making modifications or even scaling of certain contributions to the potential function. Specifically, for the problem of finding the native-like complex configurations, our scaling is expected to raise the energy minimum of complexes, possibly making them less stable, due to a smaller energy gap, to higher energy alternative configurations. It is intriguing that the effect on dissociation and ease of configurational changes within the complex is similar due to the modifications in CHARMM36m. This suggests that the rate determining interactions rely on similar close range contacts. These observations are similar to the findings of Feig, Sugita and colleagues [46] 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.

We may also comment on whether the association and dissociation of the SAM-SAM complex are the reserve of one another. Despite the potential function modification, the mechanism of binding and finding the native state complex is similar to the reverse of the dissociation trajectories we had with the swap mutants [20]. However, in some cases, it may be not similar; as several association events started from a very favorable pre-orientation of the SAM domains that is close to the native-like complex. By contrast, dissociation from the native-like complex faces the task of coming out of a deep energy minimum, which seems best accomplished by a surface sliding or pivoting movement to a weaker binding interface first. We should also caution that our conclusions are based on the study of one model system only; a SAM-SAM complex, which has highly complementary electrostatics. Nevertheless, a preference for on surface configurational changes have implications for protein-protein and protein-ligand docking algorithms.

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 50ns of the trajectories (Fig. 5) have a minimum between approx.5-6A, 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 [47-49]. Enhanced sampling method, such as temperature- or Hamitonian-based replica exchange molecular dynamics have been used in sampling configurational space of intrinsically disordered proteins, but have rarely been used in the prediction of protein-protein association. 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. For example, protein structure refinement could in some systems be accomplished recently with lengthy but partly restrained simulations [50]. While this may be true, the lesson suggested by the current study is that tinkering with certain aspects of the potential function can ameliorate one problem of "trapped encounter complexes and intermediate states" (by enhancing protein-protein dissociations) while possibly having undesirable effects on another aspects of the protein-protein complex energy landscape (decreasing the difference in the energy between of states). The situation is similar in that respect to the process of protein folding, where populated/long lived intermediate states are now seen as a detriment to the kinetics of the folding process and in some cases, leading to side reactions, such as the formation of aggregated proteins [51]. Nevertheless, we also suggest that certain steps of the association process, specifically the initial pre-orientation of the domains, could be largely independent of the latter steps.

CONCLUSION

In summary, we investigated the association of the SAM domains of the EphA2 receptor and SHIP2 enzyme by performing close to two hundred 250 ns all atom molecular dynamics simulations. Using the CHARMM36 potential function the NMR-like structures of the protein complex are obtained in approx. 10% of the real-space unbiased molecular dynamics simulations, while many other simulations are trapped in the alternate/non-native states. The patterns of initial protein contacts are found to be directed in many cases by long range electrostatic interactions. In the formation of a native-like complex with the original potential function, the protein forms the native structure from a pre-oriented position that is close to the native structure, but in some other cases, large configurational movements (sliding and pivoting) of one SAM domain on the protein surface of the other are seen. Dissociation-(re-)association events also help the formation of native-like complexes. The study provides a rich picture on mechanisms of protein-protein complex formation of a small model system, which has remarkable electrostatic complementarity between protein surfaces. Importantly, we found that CHARMM36m and a 5% protein-solvent interaction scaled potential function performs significantly better (18% and 25%, respectively) in yielding the native-like protein complex compared to CHARMM36. With a modified potential function parameter for a slightly increased protein solvation, the overestimation of initial protein-protein contacts and their stability is found to be reduced, and the overall prediction of native contacts is improved. A more systematic re-parameterization of the potential function may be warranted to further improve the prediction of protein association, but the results presented here point towards possible alternative strategies for the case of protein complex structure prediction, by either substantially improving sampling within an already associated state or by inclusion of knowledge-based potentials.

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), 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. Ahmad, M., W. Gu, and V. Helms. 2008. Mechanism of fast peptide recognition by SH3 domains. Angew. Chem. Int. Ed. Engl. 47:7626-7630.
10. 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.
11. 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.
12. 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.
13. 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.
14. Bonvin, A. M. 2006. Flexible protein-protein docking. Curr. Opin. Struct. Biol. 16:194-200.
15. Zacharias, M. 2010. Accounting for conformational changes during protein-protein docking. Curr. Opin. Struct. Biol. 20:180-186.
16. 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.
17. Pabon, N. A., and C. J. Camacho. 2017. Probing protein flexibility reveals a mechanism for selective promiscuity. Elife. 6: e22889

18. 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.
19. 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
20. 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.
21. 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.
22. Ubbink M. 2009. The courtship of proteins: understanding the encounter complex. FEBS Lett. 583:1060-1066
23. 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
24. Qiao, F., and J. U. Bowie. 2005. The many faces of SAM. Sci. STKE. 286: re7.
25. Shi, X., V. Hapiak, J. Zheng, J. Muller-Greven, D. Bowman, R. Lingerak, M. Buck, B. C. Wang, and A. W. Smith. 2017. A role of the SAM domain in EphA2 receptor activation. Sci. Rep. 7:45084.
26. Pasquale, E. B. 2010. Eph receptors and ephrins in cancer: bidirectional signaling and beyond. Nat. Rev. Cancer. 10:165-180.
27. 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.
28. 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
29. Petrov, D., and B. Zagrovic. 2014. Are Current Atomistic Force Fields Accurate Enough to Study Proteins in Crowded Environments? PLoS Comput. Biol. 10:e1003638.
30. 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.
31. 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.
32. 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.
33. 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.
34. 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.
35. 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.
36. 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.
37. 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.
38. 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.
39. 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.
40. 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.
41. 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.
42. 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.
43. 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.
44. 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.
45. 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.
46. 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.
47. Keskin, O., N. Tuncbag, and A. Gursoy. 2016. Predicting Protein-Protein Interactions from the Molecular to the Proteome Level. Chem. Rev. 116:4884-4909.
48. 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
49. 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.
50. Heo L., and Feig M., 2018. PREFMD: a web server for protein structure refinement via molecular dynamics simulations. Bioinformatics. 34:1063-1065
51. Dobson CM. 2003. Protein folding and misfolding. Nature. 426:884-890.

Figures:

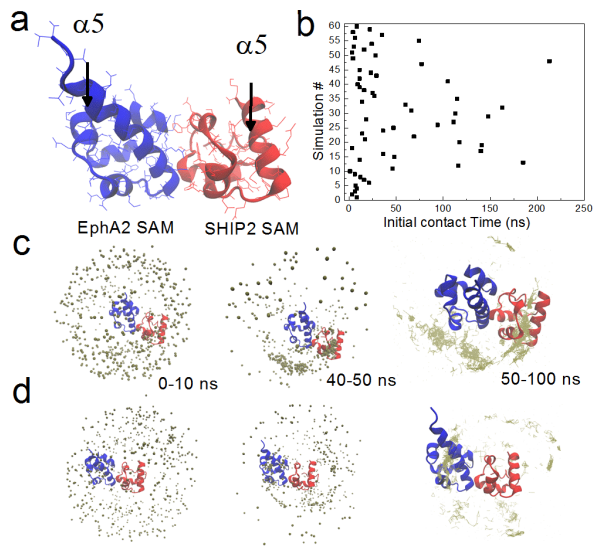

Figure 1: Native complex of EphA2 SAM: SHIP2 SAM heterodimer [18] (a) and the early stages of E-SAM and S-SAM association for simulation with the original potential function (b-d). (b) Initial contact time for 60 independent simulations. (c-d) Pre-orientation of S-SAM (S-SAM) relative to E-SAM (S-SAM). Proteins are superimposed on E-SAM (c) or S-SAM (d). In (c) and (d) 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.

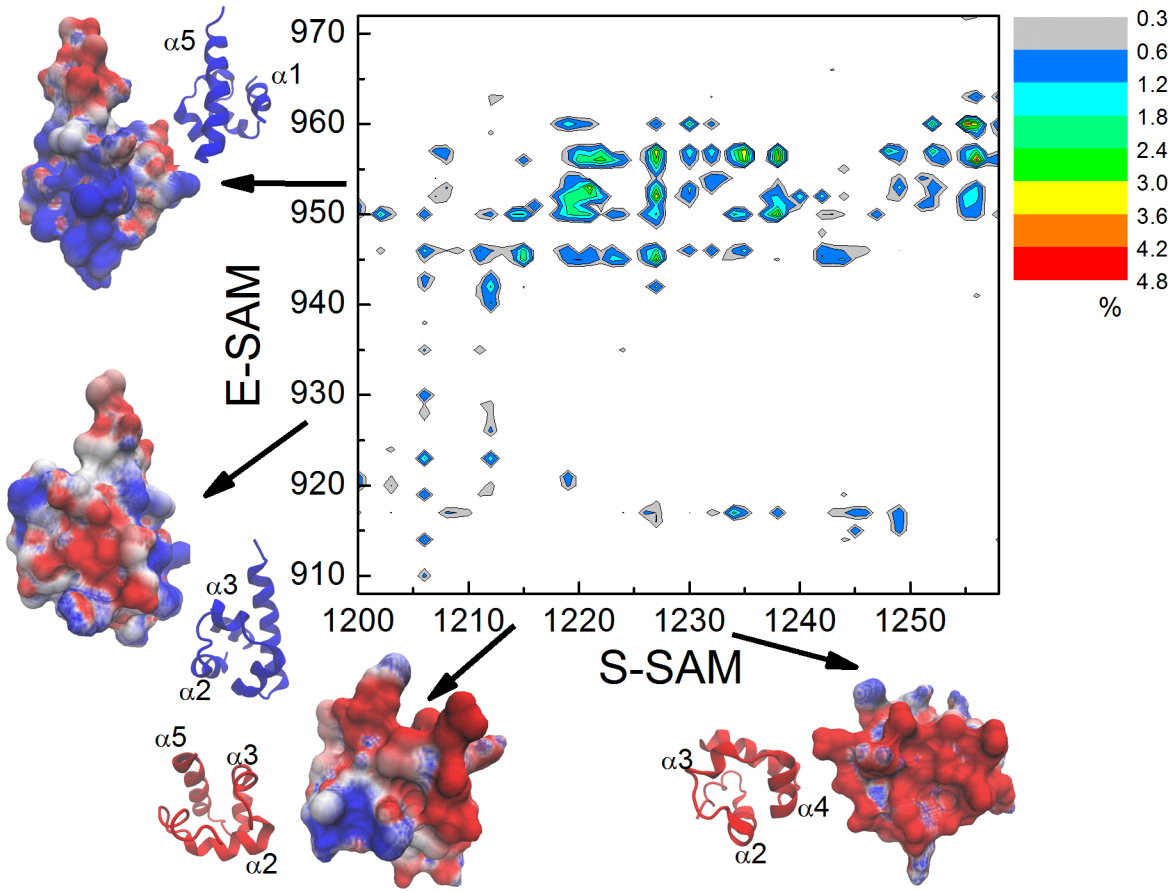

Figure 2: E-SAM: S-SAM residue-residue contact map during the early stage of protein-protein association in simulations with original potential function (S_{100}). Simulation trajectories of 0-50 ns are used to construct the contact maps, averaged over 60 trajectories. The color scale (=contact occupancy; max near ~4.8%) is from grey to red as indicated. The contact occupancy is still low as not all proteins are extensively interacting. Electrostatic maps are shown alongside the respective E- and S-SAM residue axes, with SAM orientations, beside them as mainchain cartoons. Red/bue indicate negative/positive electrostatic potential, respectively.

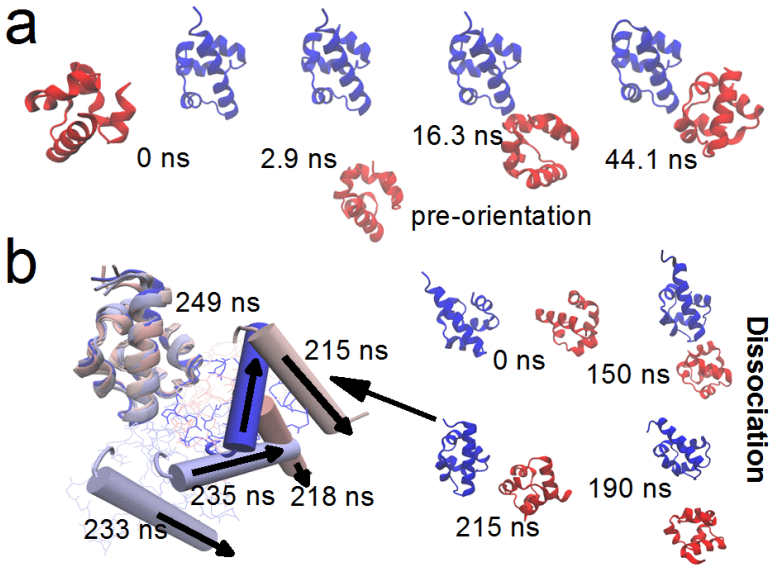

Figure 3: Protein association process between E-SAM and S-SAM. (a) Snapshot for a pre-orientation process (shown for S_{100}-7). (b) A dissociation-(re)association event followed by a dynamic transition towards to the native structure (shown for S_{100}-17).

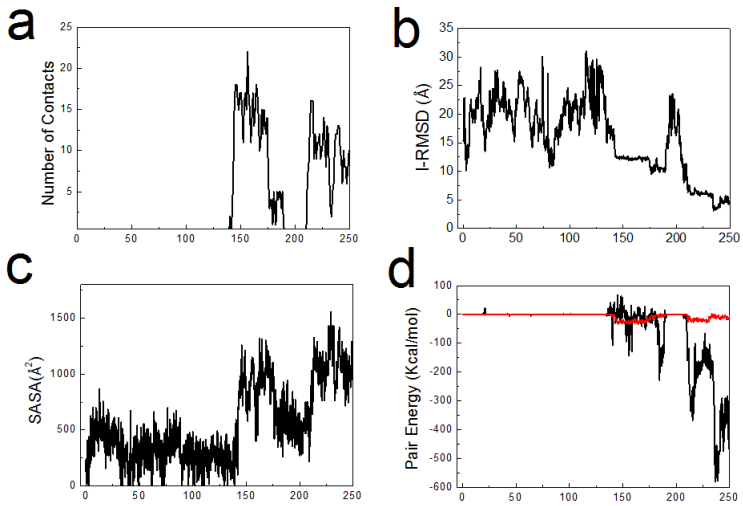

Figure 4: A protein association process (from simulation S_{100}-17). (a) Number of contacts between residues of E-SAM and S-SAM domain. A contact event is registered if a residue is within 0.5 nm of a residue on the other protein. (b) i-RMSD and (c) buried surface accessible surface area (SASA) of the E-SAM and S-SAM. (d) Protein-protein Electrostatic (black) and vdW interactions (red) between E-SAM and S-SAM.

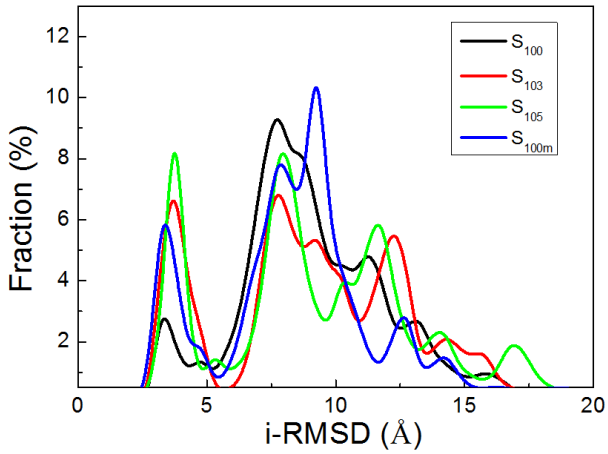

Figure 5: Distribution of i-RMSD for different simulation sets (i-RMSD for structures in the last 50 ns of the 60 simulations).

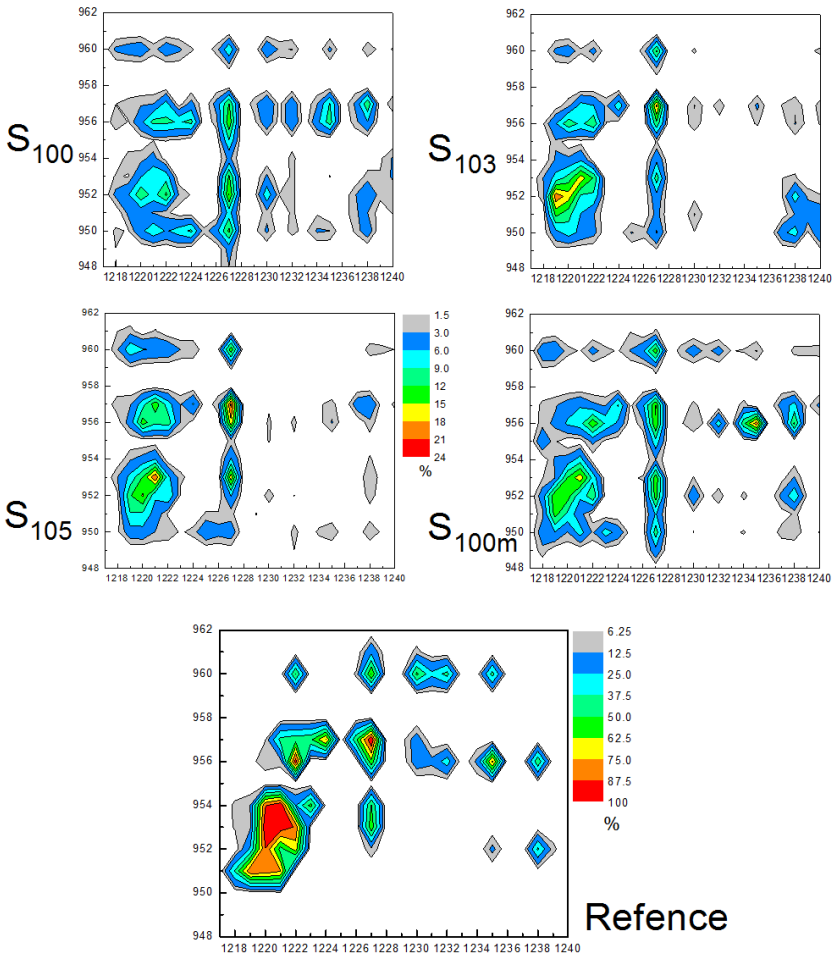

Figure 6: Residue-residue contact maps for different simulation sets (subset of map showing regions of native-like interactions only). The last 50 ns of the trajectories were used. Reference contact maps (bottom) of simulations starting from native structures state 1 and state 2 [18].

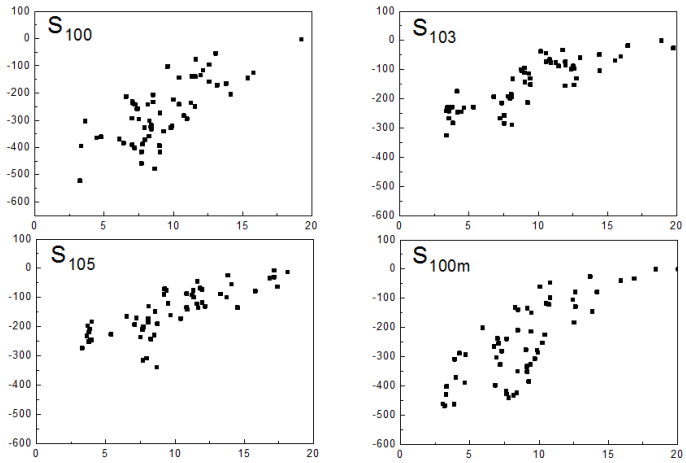

Figure 7: Pair interaction energy between E-SAM and S-SAM (in kcal/mol) as a function of i-RMSD (in Å) for the last 50 ns of the simulations.

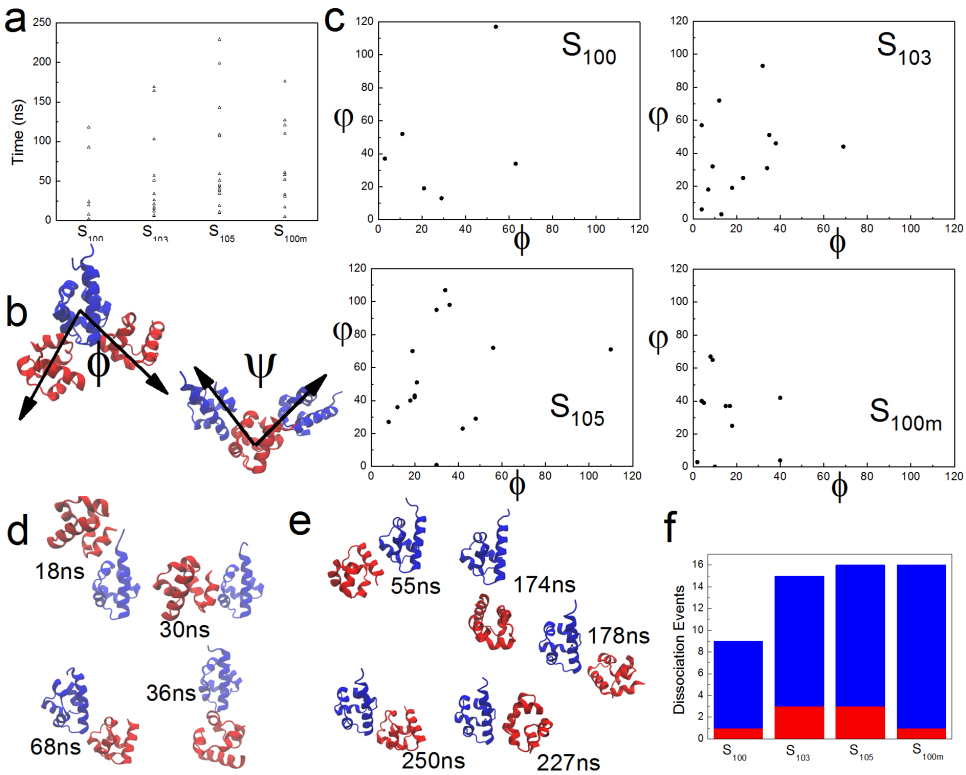

Figure 8: Difference between simulations with original and modified potential function. (a) Length of time from initial contact to the formation of a native-like complex. (b) Definition of angle ϕ and φ. ϕ denotes the orientation changes of S-SAM relative to E-SAM (E-SAM is superimposed, comparing the 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. (c) Change in ϕ and φ in the "successful" simulations from initial association to the native-state of each set. (d) A major surface movement of S-SAM relative to E-SAM (shown for S_{105}-33). (e) A major surface movement of S-SAM relative to E-SAM (55-178ns) and a large rotation of S-SAM (178-250 ns). (f) Number of dissociation events in different simulation sets. Red indicates the of number cases where a dissociation event leads to a native-like protein complex upon rebinding.