Pattern, Pathways and Dynamics of EphA2 SAM and SHIP2 SAM Heterodimer Association Revealed by All-Atom Molecular Dynamics Simulation

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 multi-microsecond all atom molecular dynamics simulations. In the native complex, the SAM domains interact using charged surfaces which are highly complementary. However, in simulations of 100-200 ns, most of the initial protein complexes are trapped into non-native configurations. However, a few SAM domains associate to form heterodimers from orientations that are close to those in the native-complex. In this case, only minor adjustments are needed, but in other trajectories, 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. Importantly, by slightly increasing the solvation of protein polar sidechain groups (scaling of the vdW interaction energy in the CHARMM36 potential function), the prediction of native-like SAM complexes is enhanced by more easily allowing configurational transitions and dissociation events. These observations likely point to a way for the improvement of computational predictions of protein-protein interactions and complexes in general.

Keywords: protein-protein association; protein complex formation; molecular dynamics; potential function parameters

INTRODUCTION

Biological function that arises from the formation (and/or dissociation) of protein complexes critically depends on their dynamics, which are, in turn, determined by the strength and kinetics of the underlying protein-protein interactions for molecular recognition. Despite the significance of these features, a detailed knowledge of protein-protein interactions and understanding of complex formation is far from complete [1-5]. Especially, an appreciation of protein association processes are still lacking for the great majority of systems at the molecular level, 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 [9-11]. Furthermore, until relatively recently most protein docking

computer protocols could not incorporate sidechain (or mainchain) flexibility [12-14], but the role of local, as well as global dynamics in protein complexes and complex formation is now becoming well appreciated [15-18]. 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 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 [19]. 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. A number of inferences could be made from this study, including the suggestion that the reverse process, that of protein association, would proceed via transient encounter and intermediary states. States and processes associated with protein complex formation, such as electrostatic steering have been reported and discussed in the literature over decades [20-22]. Availability of computer resources to extensively simulate the association events, motivated us to carry out unrestrained all-atom simulations of the same small heterodimer EphA2-SHIP2 SAM: SAM domain to study the process of protein-protein complex formation.

The protein domain Sterile alpha motif (SAM) is found in more than 200 human proteins [23]. As an example, the SAM domain is an important component of the intracellular region of the transmembrane Ephrin receptor [24]. 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 [25]. 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 [26]. 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 [17]. As mentioned above, in a recent computational study, we also investigated the dissociation pathways of the complex resulting from swaps of domain bridging charged sidechains which had perturbed the energy landscape of the complex [19]. 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 (one surface of the EphA2 SAM is entirely positively, one 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 [19]. The complex has a moderate binding affinity with 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 standard CHARMM36 potential function. However, we quickly realized that while the SAM domains associated on a reasonable timescale, the initial associated states were often "stuck", in that they showed few interconversions and fewer dissociation events, than would be expected from encounter complexes [27]. 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 the behavior of intrinsically

2

denatured proteins (IDPs) [28-30]. In the former, proteins such as ubiquitin showed strong aggregation and the IDPs became too compact. 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 model the intrinsic disorder of IDPs [31]. A most recently released model of CHARMM36m improved the prediction of α helices and intrinsic disorder in polypeptides, buck acknowledged that more re-parameterization had to be done [32]. Several other studies have rescaled the protein-protein interactions or protein-water interactions, in order to get a better performance in prediction of protein configurations or protein-protein association [33-37].

Here, by scaling certain solute-solvent vdW interactions and by carrying out dozens of simulations, we compared the simulation results, using the original CHARMM potential function and the modified potential function, where the latter shows an increased dynamics in the protein association process and an improved prediction in native contacts. In the early stage of protein association, an electrostatic patch directs the two proteins to interact through interfaces with a high level of electrostatic potential complementarity. But there are several competing charged patches, so only approx. half of those interfaces make contacts to exactly correspond to the native protein complex structure. In cases where either or both SAM domains pre-orientate to positions that are close to those of the native-complex, only minor adjustments are needed after association to form the native structure. However, in other trajectories, large configurational movements (sliding and pivoting) are observed. Dissociation-(re-)association events are also seen in a few (especially in the parameter modified) trajectories 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. The whole system was solvated by TIP3P water in a simulation box of $9 \times 9 \times 9$ 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). The system set-up followed previous protocols (e.g. initial heating and equilibration for 40 ps and 1 ns, respectively [e.g. 17]). During the simulations a step potential was applied to prevent the escape of E-SAM or S-SAM from the primary simulation box. Similarly, a step potential was used to restrain the relative distance between E-SAM and S-SAM. Specifically, no force was applied if the distance between the center of E-SAM and S-SAM was within 5 nm, and a harmonic potential with force constant 100 Kcal/mol/Å^2 was applied if the two proteins were greater than 5 nm apart. Twenty independent simulations were performed. In the first 10 simulations, S-SAM was randomly rotated about its center of mass to produce different initial simulation structures. In the last 10 simulations, both proteins are rotated randomly with respect to their center of mass.

Simulations were performed with the CHARMM36 potential function including CMAP correction [38-40]. We made a modification on the CHARMM36 potential function parameter [39] by scaling the

3

solute-solvent vdW interaction to test whether this enhances the prediction of protein-protein association. 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, 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 1.1. 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.

In order to enhance sampling, 20 simulations were performed for each kind of set-up with λ at 1.00, 1.03 or 1.05 (denoted as ST, MA and MB for brevity) and each simulation was at least 100 ns long. 10 simulations were performed with λ at 1.1 (denoted as MC, see Table S2 for a summary in the Supporting Information). In all simulations, the van der Waals (vdW) potential was cut at 1.2 nm. 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 [41]. Analysis (interface-RMSD, residues-residues contact, surface electrostatics, orientation angle, pair interactions, and solvent accessible surface) were done with VMD and NAMD with in-lab scripts. In calculation of interface-RMSD, interface residues 912-921, 949-962 of E-SAM and residues 1215-1239 of S-SAM were considered.
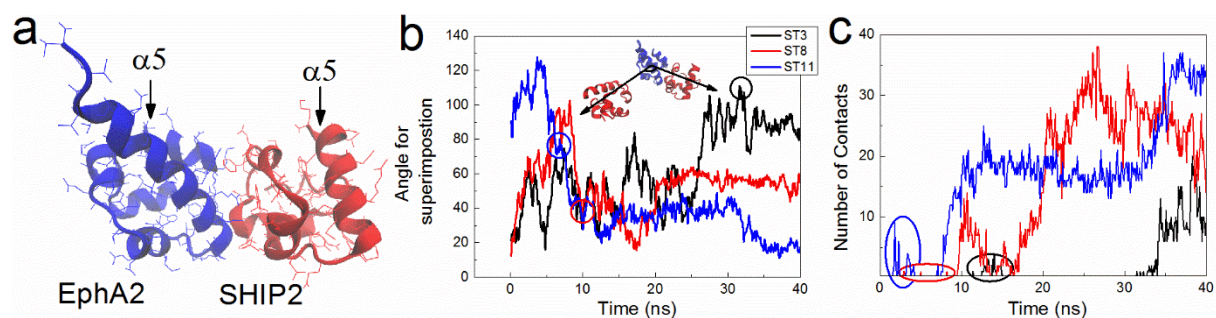

Figure 1: Native complex of EphA2 SAM: SHIP2 SAM heterodimer [17] (a) and the early stage of E-SAM and S-SAM association (b-c). (b) Adjustment of relative orientation between E-SAM and S-SAM protein along with the simulation times. Proteins are superimposed on E-SAM. The relative orientation is defined as the angle of S-SAM (center of mass) with respect to its position in the reference complex (pdb id. 2kso). A contact event is registered if a residue is within 0.5 nm of a residue on the other protein. The circles indicate the initial time that two proteins associate (defined as the simulation time that two proteins make contacts that lead to a full contact). (c) Number of contacts between residues of E-SAM and S-SAM domain. The ellipsoids mark the tentative contacts. By contrast an initial tentative contact is defined as an interaction that is not immediately followed by a full protein-protein association. See Fig. S1 and S2 for results for all ST, MA and MB simulations.

RESULTS AND DISCUSSION

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. During the early stages of protein association, the two proteins constantly adjust the relative orientation with respect to each other. A SAM domain revolves significantly both around its own center of mass and by -what appears as a translation on the surface of a sphere- around the center of the other SAM domain. For example, Fig. 1b shows the relative angle between E-SAM and its native complex reference as a function of simulation time for a few representative simulations (ST3, ST8 and ST11; see also Fig. S1 for full results). Orientational adjustments around the time (before and after) of the initial association event for these three simulations are shown in Fig. S3 with a few snapshots. The adjustment of the relative orientation happens before the first contact of two proteins is established, but also occurs through another process: from initial protein contact to the formation of a relatively stable configuration, where further residue contacts have been established (Fig. 1c, see also Fig. S2). Besides, in a few of the simulations, early contact events are observed that do not result in persistent protein association, (Fig. 1c). If the early contacts are unfavorable for the formation of a stronger association, the two proteins will separate until a more favorable orientation /contacts can be established.

Fig. 2 shows the E-SAM: S-SAM residues contact map, averaged over the early stage of the SAM: SAM association process seen in 20 independent simulations. Clearly, several residues are significantly involved in < 5Å contracts, while other residues are rarely involved in the inter-protein interactions. For example, residues 922-938 of E-SAM are rarely involved in interactions. This is further revealed in the inset diagrams which show a projection of the position of the center of mass of S-SAM (or E-SAM) around E-SAM (or S-SAM) when the coordinate frames are superimposed on the latter structure. While the surface of S-SAM is well sampled by E-SAM (right), 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 surface electrostatic potential pattern of the two SAM domains is expected to be a major determinant of the pattern of binding. As shown in Fig. 2, the primary binding region of S-SAM has many negative charges and thus a large region of 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 expected to be unfavorable, reflected by the absence of center of mass localization of S-SAM around this region of E-SAM (lack of orange density in left inset). Instead, the E-SAM region containing helix α-4 and α-5 is positively charged (electrostatic surface depicted on top left). 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, even though negatively charge is largely concentrated on one side of the protein (see below). Still, overall, the two proteins associate with each other through interfaces which show a reasonably complementarity between the positive and negative electrostatic potential. This supports the viewpoint that an electrostatic pre-orientation of the domains is involved to direct the early stages of protein association.

Table I: *Comparison of complexes and contacts in the different simulation sets... Protein complex with i-RMSD < 5 Å with respect to native complex is considered as native-like structure. *Indicates non-ordinary association process involving large configurational change or dissociation-(re-)association process. *** % number of frames where the interaction is present. *** weighted average, the contact frequency in the three simulation sets are scaled by the native averages.*

|	ST	MA	MB
Native-like (final)	4 (1*)	5 (3*)	5 (4*)
Native-like (visited)	4	7	8
Dissociation Events	2	4	8
Native contacts (time average of state1 and 2 in [17])	Contact frequency in simulation sets		
Y960-D1235 42.4%**	1.6%	6.0%	3.2%
Y960-T1232 42.0%	4.0%	8.2%	0.2%
Y960-D1230 52.5%	7.3%	15.3%	4.2%
Y960-F1227 63.2%	16.1%	11.2%	8.5%
Y960-W1222 50.7%	2.4%	9.4%	4.9%
R957-F1227 100.0%	14.8%	23.7%	33.6%
R957-E1226 47.8%	9.6%	3.8%	2.2%
R957-D1224 76.2%	10.9%	15.3%	17.0%
R957-D1223 47.1%	14.1%	9.7%	15.8%
R957-W1222 45.5%	3.8%	5.8%	17.0%
R957-G1221 41.5%	13.4%	13.3%	32.5%
K956-E1238 47.3%	11.3%	9.4%	5.1%
K956-D1235 78.6%	15.8%	17.5%	8.9%
K956-F1227 59.0%	19.6%	16.8%	11.5%
K956-W1222 94.1%	19.7%	20.4%	12.7%
H954-F1227 50.6%	7.9%	2.2%	11.1%
H954-D1223 54.4%	0.1%	2.5%	5.1%
H954-G1221 98.1%	9.1%	20.7%	14.2%
H954-N1220 86.3%	5.4%	16.9%	5.5%
G953-F1227 58.2%	14.7%	13.9%	14.5%
G953-W1222 85.6%	5.0%	13.1%	5.2%
G953-G1221 96.3%	18.8%	28.3%	31.4%
G953-N1220 97.7%	17.7%	27.4%	16.0%
P952-W1222 68.2%	5.3%	19.0%	12.1%
P952-G1221 44.9%	11.9%	21.3%	21.0%
P952-N1220 87.1%	20.9%	32.3%	22.5%
L951-G1221 76.1%	3.2%	12.15%	6.2%
L951-N1220 84.2%	12.2%	21.2%	16.4%
L951-H1219 85.3%	13.8%	21.2%	14.0%
Weighted** Average	**7.7%**	**11.4%**	**9.3%**
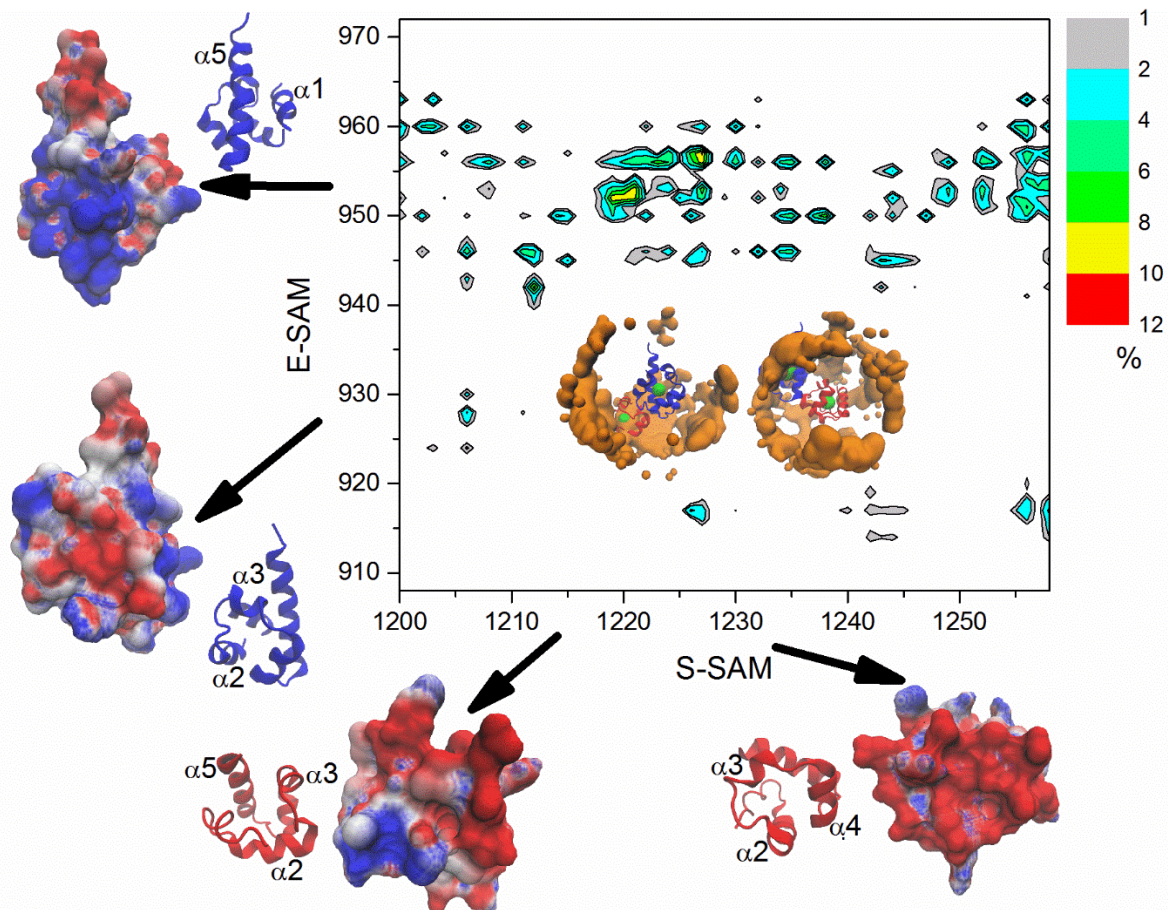
Figure 2: E-SAM: S-SAM residue-residue contact map during the early stage of protein-protein association in simulations with standard force field (ST). Simulation trajectories up to 10 ns before as well as after the first association event are used to construct the contact maps, averaged over 20 trajectories. The color scale (=contact occupancy; max near ~12%) is from grey to red as indicated. The inset diagrams (near bottom half of map) show the extent of sampling around each protein (the center of mass) with respect to the other protein (aligned on E-SAM, left and S-SAM, right; shown in same orientation as complexes with electrostatic surface; the location of the center of mass of the NMR derived complexes are shown as green sphere. Electrostatic maps are shown alongside the respective E- and S-SAM residue axes, with SAM orientations, beside them as mainchain cartoons.

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

Within 50 ns of the simulation start, the great majority of the trajectories enter into relatively stable association states, as shown by a plateau in the number of contacts (Fig. S2). Thus, the association is strengthened compared to the initial association pattern during the early stage. This might suggest a preferred set of interfaces in ST simulations, but, as noted above, many of the simulations yielded configurations in the final parts of the trajectories, that are far from the native E-SAM: S-SAM structures, with interfaces either displaced by translation and/or by rotation. Remarkably, there are no obvious differences between the first 10 and second 10 simulations, showing that the association process does not rely on the initial orientations of the two proteins (which is more randomized in the latter 10 simulations).
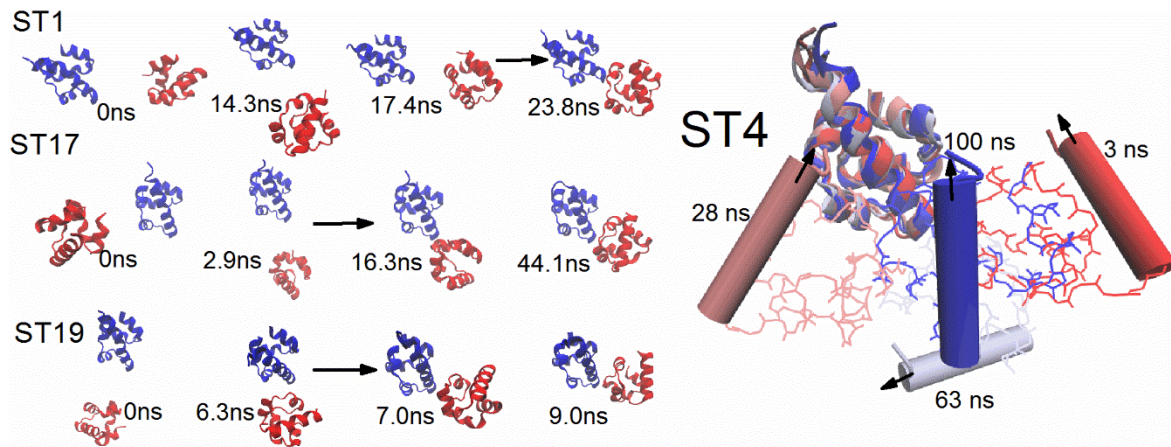
7

Figure 3: Protein association process between E-SAM and S-SAM. Left: The pre-orientation (denoted by arrows) before protein association in simulation ST1, 17, 19. Right: A dynamic transition towards to the native structure in simulation ST4.
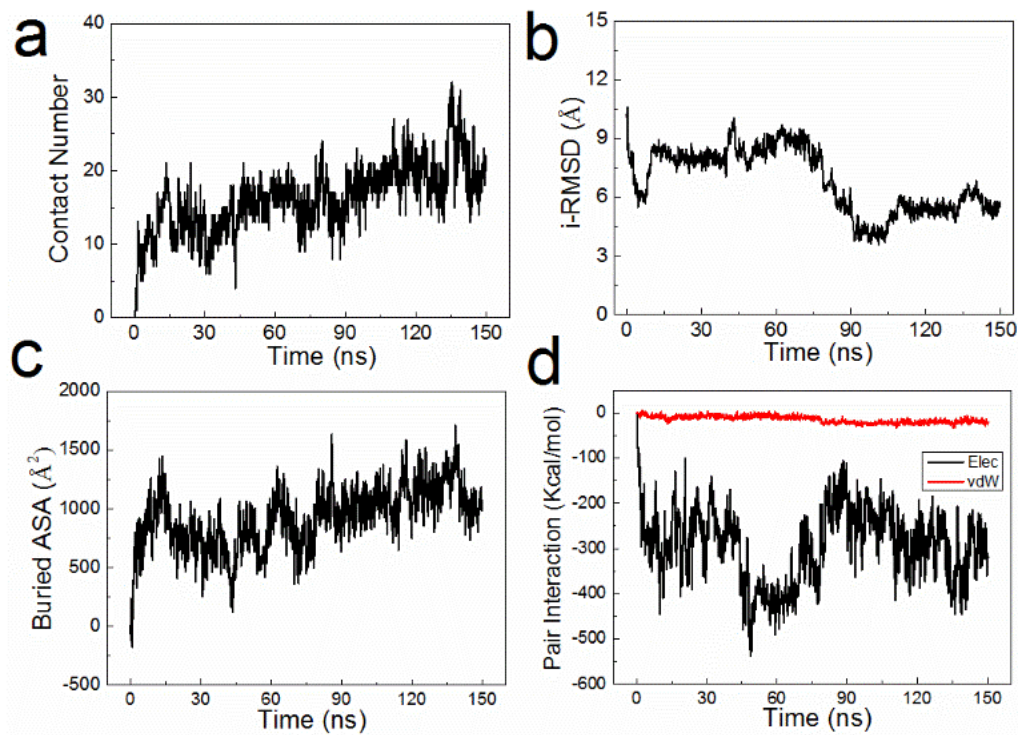

Figure 4: A protein association process (from simulation ST4). (a) Number of residue-residue contacts. (b) i-RMSD and (c) buried ASA of the E-SAM and S-SAM (see Fig. S10 for results for all simulation sets). (d) Electrostatic and vdW interactions between E-SAM and S-SAM.
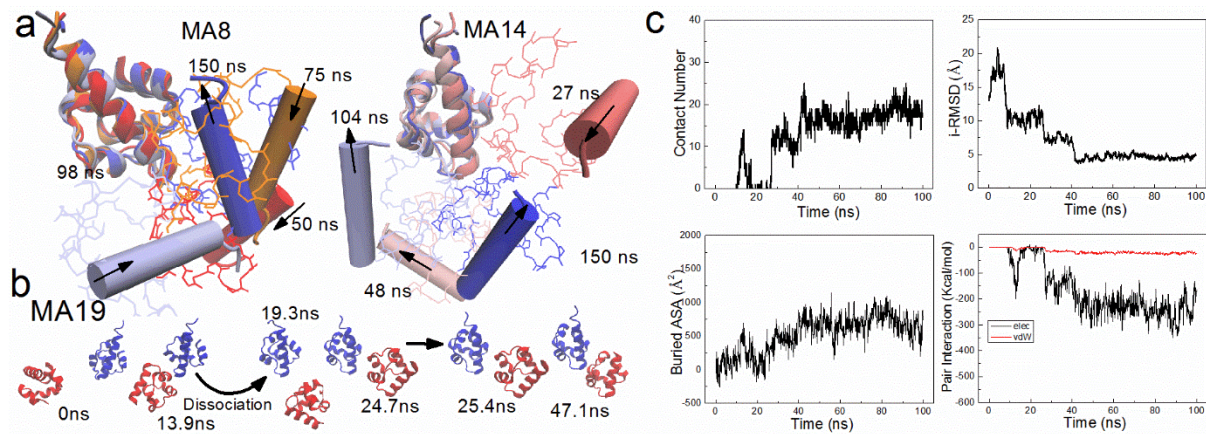
Figure 5: (a) Dynamic transition towards to the native-like structures in simulations MA8 and MA14. (b) Dissociation-(re)association events in MA19. c) Number of contacts, I-RMSD, solvent accessible surface area buried and pairwise protein-protein potential energy for the MA19 association trajectory.

Native-like structures are obtained in 4 of 20 simulations, based on the deviation of the final structure relative to the native complex (interface RMSD, see Method and Fig. S4). Fig. 3 shows the E-SAM: S-SAM association process that yield a native-like structure in simulations ST1, ST4, ST17 and ST19. In simulations ST1, ST17 and ST19, a process of pre-orientation − that precedes formation of substantial contacts- is used to direct the two proteins toward to a position that is close to the native structure. From there, they directly transform into the native-like structure with only minor structural readjustment. Differently, in ST4, the formation of native-like complex involves additional movement after the initial association (Fig. 3 right). The S-SAM domain slides along the surface of E-SAM (3-20 ns) and eventually adjusts itself to a native structure (at 100+ ns). It should be realized that although there are large adjustments after initial association to yield a native-like structure in ST4, the initial association position was actually already reasonably close to that of the native structure.

Fig. 4 plots the variation of buried surface area, electrostatic and van der Waals (vdW) interaction for the ST4 trajectory, as the proteins form a native-like complex (low i-RMSD). As the proteins get closer, more residue-residue contacts are established, and more surface area is buried. 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 [19].

E-SAM: S-SAM association with a modified 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. 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) appear to be polar-solvent interactions that could be too weak; thus we

9

also ran 20 simulations with the vdW polar group – water interaction energy scaled up by a factor of 1.03, 1.05 and 10 simulations with a scaling factor of 1.10 (see Methods). In these latter simulations, 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 calculations.

The patterns of contacts during the latter part of the association/in the bound states are shown in Fig. S5 and S6 considering, respectively, the first contact events (similar to Fig.2) and the last 20 ns of simulations with the modified parameters. The initial contact the contact pattern is changed already to become more native-like while the regions of surface covered are more widely distributed, suggesting more extensive sampling of orientation states. Native-like structures are visited in these simulations (7 and 8 times in the MA and MB sets, respectively), although eventually native-like structures are obtained towards the end of the simulations, in only 5 of 20 trajectories in each system (MA, MB). However, in the simulations with modified potential function, the native-like structures may arise as a result of the increased frequency of dissociation processes and from dynamic transition processes in bound states, rather than from changes to the pre-orientation process. An important point is that scaling the vdW polar group-solvent interaction energy will not affect the longer range electrostatics, and indeed no difference in the extent of native-like pre-orientation prior to binding is observed. Specifically, in simulations MA7, MA11, MB5, a process of pre-orientation – that precedes formation of substantial contacts- is used to direct the two proteins toward to a position that is close to the native structure (Fig. S7-8). In simulations MA8, MA14, MB2, MB4 and MB9, a large relative movement between E-SAM and S-SAM is observed (snapshots in Fig. 5 for MA8, MA14 and Fig. S8 for MB2, MB4, MB9). For example, in MA8, a large rotational movement is seen of the S-SAM domain relative to E-SAM before the NMR-like structure is observed. The orientation of α-helix 5 (and the S-SAM domain) is fully rotated by approx. -270° (or $+90^{o}$) from 30 ns to 150 ns in this process. It should be noted that such rotations are unlike those observed between the native-like states in a previous simulation study [17] with the unmodified potential function. There, orientations such as that seen at 30 ns were "stuck" (as shown in supplementary material of that paper), strongly suggesting here that modification of the interaction potential with water (see below) has enabled such transitions. Interestingly, some residues at the SAM domain interfaces could be critical for the transitions, acting as pivots [see also 19], a concept of "anchoring residues" was introduced by Camaco and colleagues [42]. For example, residues pairs between E-SAM R957 and S-SAM W1222/F1227 are found to act as a pivot.

Dissociation-(re)association processes that yield a native-like structure are observed in simulations MA19, MB9, and MB14 (Fig. 5b-c and Fig. S8). If the initial association is too distant from the native structure, in principle, the proteins may also directly move around one another by rotational translation/sliding along the partner protein surface to get to the native structure. However, this process is seen in none of simulations, although we do observe the large movement of one protein around the other, while maintaining extensive surface to surface contacts, in a few simulations (such as MB1, Fig. S9), without yielding a native-like structure. This relative movement of the two proteins involves frequent breaking and the establishment of residue-residue interactions, as discussed in ref. 17, without a change in the overall number of contacts. Scaling-up protein-solvent interactions, likely, by competition makes the energy barrier for directly converting a sub-stable state to the native

10

structure also somewhat lower, but appears not as effective than lowering the barrier for protein dissociation. Thus, if the two proteins bind to form a non-native complex structure, our observations suggest a straightforward mechanism in that the two proteins would dissociate and rebind together towards a position that is possibly closer to the native structure.
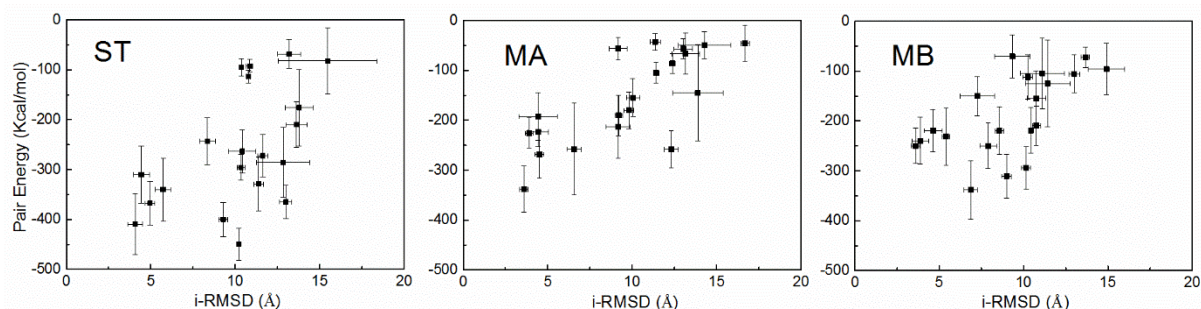

Figure 6: Pair residue interaction energy between E-SAM and S-SAM as a function of i-RMSD. Simulations with λ at 1.00 (a), 1.03 (b) and 1.05 (c).

Improved prediction of SAM heterodimers with an optimized protein-water interaction parameter

In Fig. 6, the pair interactions between E-SAM and S-SAM domains in the final 20 ns simulations (or 20 ns ahead of dissociation if it happens) 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 further 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 several 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 [19]. As we increase the protein solvation in our simulations, the side chain 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. 6 as an overall trend, comparing the simulations with the modified parameter to simulations with standard parameter potential function. As the pair interaction between E-SAM and S-SAM increases in the simulations with the modified potential functions, more dissociation or dissociation-(re-)association events are observed (Table I).

Table I and Fig. 7 compare the difference in performance and contact patters, respectively between simulations with standard and modified parameters. There is not a big difference in obtaining the native-like structure when the final structures are compared, but indeed, as noted above, the native-like structure are more visited in the simulations with λ at 1.03, 1.05 (sets MA, MB) compared to those with λ at 1.00 (set ST, all in Table I). Similarly, comparing pattern of contacts in the final stage (20 ns) of the E-SAM: S-SAM complexes in Fig. 7, the pattern and population of native contacts are enhanced in simulation with λ at 1.03, 1.05 with reference to the residue-residue contact map obtained from our previous simulations starting from the native complex (Cluster 1 and Cluster 2) [17]. 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.
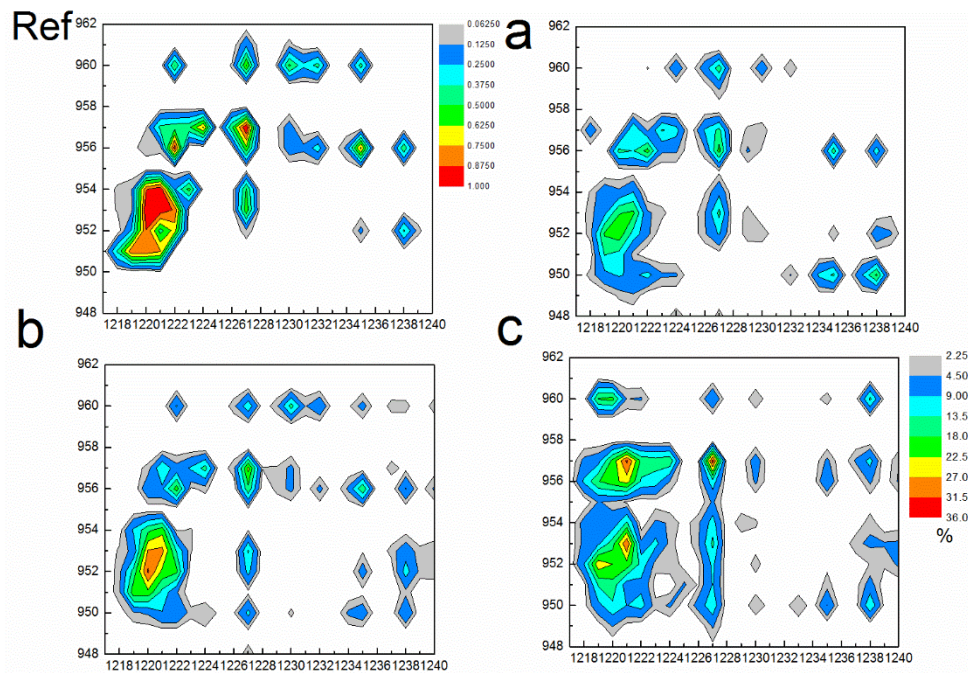

Figure 7: (a-c) Residue-residue contact maps for ST, MA, MB simulations. The last 20 ns of the trajectories were used. Reference contact maps (top left) of simulations starting from natives structures cluster #1 and #2 [17]. Fig. 7a-c is zoomed-in from Fig. S6, where the full association patterns are shown.

The relative weakening in protein-protein interactions, evident by the reduced pair-energy (y-axis) of plots in Fig. 6, helps the E-SAM: S-SAM complexes escape from some of the bound non-native structures, in this way, it may help direct the protein toward to the native structure. 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 [19]. Such interactions would be strengthened in the potential function modification, including waters that bridge protein-protein interactions, to the likely detriment of pure protein-protein interactions. This scenario of indirect effects illustrates the well-known difficulty associated with making modifications or even just 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, likely making them less stable, due to a smaller energy gap to higher energy alternative configurations.

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. In reality the energy landscape is an even more complicated terrain. This complexity, it is suggested, makes the process of finding and maintaining the native structure challenging in simulations that start from separated proteins and only use physical parameters. Indeed, a number of laboratories are pursuing the use of knowledge-based

12

potentials or combinations of knowledge-based and physical potentials [43-45]. 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 relatively modest yield of native structures is likely due to inaccuracies that remain in other aspects of the current potential function parameters, but it is also possible that much longer sampling on the order of tens, if not hundreds of microseconds is required for finding native-like protein complexes. 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 - here of "struck encounter complexes"- (by enhancing protein-protein dissociations) while having possibly undesirable effects on another aspects of the protein-protein complex energy landscape (here decreasing the difference in the energy between of states). 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 dozens of all atom molecular dynamics simulations. The NMR-like structures of the protein complex are obtained in a few 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 by long range electrostatic interactions. In the formation of a native-like complex with the original potential function, the protein mostly forms the native structure from a position that is close to the native structure. In the simulation sets with a modified potential function the balance between direct association leading to native-like states and processes involving substantial adjustments (protein sliding and pivoting of one protein on the surface of the other protein) is shifted to the latter, but this also includes trajectories where the domains go through a dissociation-rebinding process to reach a position that is close to the native structure.

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, with a modified potential function parameter for a slightly increased protein solvation, the overestimation of initial protein-protein contacts and their stability is reduced, and the overall prediction of native contacts is improved. At the same time the interaction energy of the native states is reduced however, which may explain why the improvement in finding the native-complexes is not as dramatic as may be expected. A more systematic re-parametrization of the potential function is warranted to further improve the prediction of protein association, but the results presented here point towards a possible alternative strategies, by either substantially improving sampling or by inclusion of knowledge-based potentials.

13

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. 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.

13. Bonvin, A. M. 2006. Flexible protein-protein docking. Curr. Opin. Struct. Biol. 16:194-200.

14. Zacharias, M. 2010. Accounting for conformational changes during protein-protein docking. Curr. Opin. Struct. Biol. 20:180-186.

15. 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.

16. Pabon, N. A., and C. J. Camacho. 2017. Probing protein flexibility reveals a mechanism for selective promiscuity. Elife. 6: e22889

17. 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.

18. 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

19. 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.

20. 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.

21. Ubbink M. 2009. The courtship of proteins: understanding the encounter complex. FEBS Lett. 583:1060-1066

22. 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

23. Qiao, F., and J. U. Bowie. 2005. The many faces of SAM. Sci. STKE. 286: re7.

24. 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.

25. Pasquale, E. B. 2010. Eph receptors and ephrins in cancer: bidirectional signaling and beyond. Nat. Rev. Cancer. 10:165-180.

26. 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.

27. 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

28. Petrov, D., and B. Zagrovic. 2014. Are Current Atomistic Force Fields Accurate Enough to Study Proteins in Crowded Environments? PLoS Comput. Biol. 10:e1003638.

29. 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.

30. 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.

31. 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.

32. 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.

33. 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.

34. 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.

35. 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.

36. 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.

37. 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.

38. 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.

39. 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.

40. 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.

41. 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.

42. 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.

43. Keskin, O., N. Tuncbag, and A. Gursoy. 2016. Predicting Protein-Protein Interactions from the Molecular to the Proteome Level. Chem. Rev. 116:4884-4909.

44. 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

45. 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.