JCTC Journal of Chemical Theory and Computation

pubs.acs.org/JCTC Article

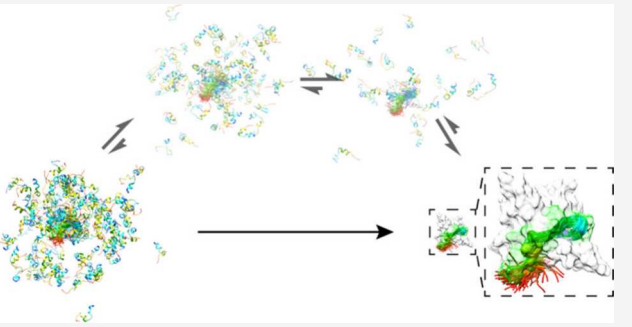
Deciphering the Mystery in p300 Taz2−p53 TAD2 Recognition

Tongtong Li, Stefano Motta, and Yi He*

Cite This: https://doi.org/10.1021/acs.jctc.4c00541 Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: Intrinsically disordered proteins (IDPs) engage in various fundamental biological activities, and their behavior is of particular importance for a better understanding of the verbose but well-organized signal transduction in cells. IDPs exhibit uniquely paradoxical features with low affinity but simultaneously high specificity in recognizing their binding targets. The transcription factor p53 plays a crucial role in cancer suppression, carrying out some of its biological functions using its disordered regions, such as N-terminal transactivation domain 2 (TAD2). Exploration of the binding and unbinding processes between proteins is challenging, and the inherently disordered properties of these regions further complicate the issue. Computer simulations are a powerful tool to complement the experiments to fill gaps to explore the binding/unbinding processes between proteins. Here, we investigated the binding mechanism between p300 Taz2 and p53 TAD2 through extensive molecular dynamics (MD) simulations using the physics-based UNited RESidue (UNRES) force field with additional Gō-like potentials. Distance restraints extracted from the NMR-resolved structures were imposed on intermolecular residue pairs to accelerate binding simulations, in which Taz2 was immobilized in a native-like conformation and disordered TAD2 was fully free. Starting from six structures with TAD2 placed at different positions around Taz2, we observed a metastable intermediate state in which the middle helical segment of TAD2 is anchored in the binding pocket, highlighting the significance of the TAD2 helix in directing protein recognition. Physics-based binding simulations show that successful binding is achieved after a series of stages, including (1) protein collisions to initiate the formation of encounter complexes, (2) partial attachment of TAD2, and finally (3) full attachment of TAD2 to the correct binding pocket of Taz2. Furthermore, machine-learning-based PathDetect-SOM was used to identify two binding pathways, the encounter complexes, and the intermediate states.



INTRODUCTION

Intrinsically disordered proteins (IDPs) or intrinsically disordered regions (IDRs) in proteins have received widespread attention from researchers due to their conformational ensembles and biological functions.[1−4] The typical structural feature of IDPs, namely conformational heterogeneity, endows them with the potential for one-to-many binding[5] by arranging themselves into distinct geometries suitable for different binding sites. For example, most transactivation domains (TADs) in transcription factors, maintaining gene stability by mediating the cell cycle and the gene transcription, have been reported to adopt heterogeneous structures under physiological conditions.[6] TADs, such as p53[7] TAD2, exhibit specific local structures when in complex with their binding partners.[8,9] In addition, post-transcriptional modification (e.g., phosphorylation) of specific amino acids is a common way to modulate IDP/IDR functions.[10] Furthermore, prevalent evidence demonstrated that protein mutations at hotspots result in loss of function.[11,12] This suggests that the physical properties of a few key amino acid residues affect the recognition kinetics and deserve to be explored in detail to reveal the binding mechanism. However, advancing our understanding of the recognition mechanism, by which IDPs accurately bind to their partners, is a significant but challenging task. When it comes to the IDP binding mechanism, two models, namely induced fit (IF)[13] and conformational selection (CS),[14−16] have been proposed to describe numerous protein binding scenarios. The induced fit model holds that binding to target proteins can induce IDP structural rearrangements to better fit the binding pocket.[13] In simulations of binding between N_{TAIL} α-MoRE and XD, α-MoRE with less helical content was observed in the transition states, and the high structural flexibility gave rise to a more productive binding than α-MoRE with a higher helical content. In addition, prefolded α-MoRE tends to unfold before binding to XD.[17] Conversely, based on the conformational selection model, proteins prefer to selectively recruit preorganized IDPs that resemble those in the complex, and

Received: April 23, 2024
Revised: July 21, 2024
Accepted: August 2, 2024

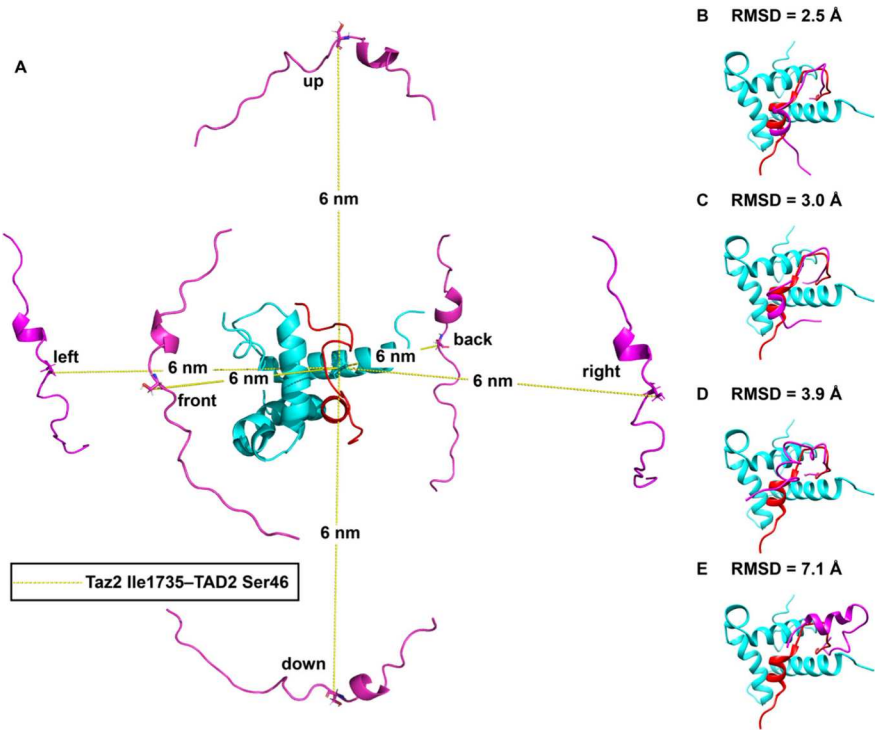


Figure 1. Six starting structures in UNRES binding simulations are shown, with Taz2 Ile1735 and TAD2 Ser46 separated by 6 nm (A), and some snapshots are extracted with complex RMSD values annotated in panels B–E. In the native Taz2–TAD2 complex (PDB ID 2MZD[8]), Taz2 is colored in cyan and TAD2 is colored in red. In the starting structures in panel (A) and sampled structures in panels B–E, TAD2 is colored in magenta with Taz2 aligned with the native complex.

then the IDP ensemble undergoes a structural shift toward the preorganized IDP structure.[18] The flux[19] through IF/CS for a specific protein complex depends on many factors, such as the population of preorganized structures, protein concentration, etc. As more work in this field is established, we will gain a better understanding of protein recognition mechanisms, which facilitate rational drug design to precisely modulate the specified signaling pathways.

Two proposed mechanisms for protein–protein recognition involving IDPs, conformational selection and induced fit, differ in the sequence of folding and binding.[14−16,19−22] To get an in-depth understanding of the folding and binding involving IDPs, over past years, theories and computational approaches have been developed to explore IDP coupled binding and folding processes, at both atomistic[23,24] and coarse-grained simulations of IDPs,[25−28] as well as the development and application of Gō-like models[29,30] to probe IDP interactions. It is known that both p300 and p53 play essential biological functions related to cancers, and the intrinsically disordered nature of p53 TAD2 makes it difficult to uncover the mechanism behind it. Our prior work[31] explored the disassociation process of the p300 Taz2 with p53 TAD2 using steered MD simulations. Although the unbinding mechanism does provide valuable insights into the interplay of two proteins, it is more straightforward to study protein recognition directly through binding simulations. Large-scale conformational changes occur during IDP folding and binding over a long period of time, making it very challenging, if not nearly impossible, to reveal IDP binding mechanisms with all-atom force fields. In contrast, coarse-grained force fields[25−28] show obvious advantages in studying long-timescale dynamics and large proteins and their complexes. In addition to using coarse-grained force fields to speed up simulations, previous work has performed simulations using native contact-based Gō-like potential energy functions[29,30,32] to study protein folding/binding.[33−35] Unbiased simulations[36,37] can reconstruct reliable binding processes due to no intermolecular interaction restraints, allowing IDPs to freely approach their binding partners and search for optimal binding pockets. Non-native contacts[38,39] were reported to play a role in the recognition process, and simulations biased with native contacts may overestimate their contribution and ultimately lead to artifacts. However, we need to remember that unbiased simulations require more computational resources to search for accurate binding sites than their biased counterparts, and the Gō-like model efficiently leverages prior knowledge of protein complexes to study their binding mechanisms.

Over the past years, we have been developing/testing a strategy to explore protein structure/dynamics and structural transitions during protein association/dissociation processes at the atom/residue level.[31,35,40−48] In this work, we performed MD simulations using the physics-based UNited RESidue (UNRES)[40,47,49−52] force field with additional Gō-like potentials to investigate the binding mechanism of p300 Taz2 and p53 TAD2. Although UNRES has been used to probe the structure and dynamics of many biological systems previously,[47,53,54] the recent introduction of the UNRES online server[55] provides an easy-to-use solution for the community to explore protein systems of interest with ease.[26,56−60] Taking advantage of the physical intermolecular distance restraints in simulations, conformational sampling can leverage existing experimental knowledge to efficiently generate bound complexes and reveal possible binding pathways. By aligning the encounter ensembles in the binding trajectories onto the native complex, our results indicate that regions surrounding the Taz2 binding pocket, especially the flexible loops, are the

https://doi.org/10.1021/acs.jctc.4c00541
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

primary regions responsible for the initial collisions. In most cases, the TAD2 helix was already formed prior to collision with Taz2. Following the formation of the encounter complex, the intermediate species were found before the transition to the final protein complex, suggesting that structures containing partial intermolecular interactions lie in local energy minimal basins. Furthermore, machine-learning-based PathDetect-SOM[61] analysis revealed the binding pathways appear to go through the structural ensembles of the encounter complex and the intermediate state described above.

METHODS

UNRES MD Simulation Setup. We employed the physics-based coarse-grained UNited RESidue (UNRES) force field (NEWCT-9P)[62] to perform binding simulations from six starting structures, where TAD2 was placed at six different positions relative to Taz2 (referred to as up, down, left, right, front, and back afterward) in a water box with a side length of 15.4 nm and a distance between Taz2 Ile1735 and TAD2 Ser46 of 6 nm, as shown in Figure 1A. The conformation of TAD2 in the starting structure is obtained from our previous MD simulation[46] results of TAD2. Through cluster analysis, the representative structure of the primary cluster (58% of the 60,000 frames, the other cluster accounts for 42%) is used in this work. Considering the diversity of conformations adopted by TAD2, the selected initial structure may have little impact on the study of the Taz2−TAD2 binding mechanism. One-hundred and twenty independent simulations were carried out for each starting structure using a Langevin thermostat to control the simulation temperature at 300 K. Periodic boundary conditions (PBCs)[63] were applied in three dimensions with a cutoff of 25 Å for long-range interactions. Snapshots were taken every 10,000 steps with dt = 0.1, resulting in a total of 2600 snapshots per simulation. In the simulations, we applied the Gō-like model by extracting the residue pairs within the Taz2 domain and between Taz2 and TAD2 from the NMR-resolved structure[8] if the distance between Cα atoms of a pair of residues is less than 8.9 Å. To impose intermolecular chemical cross-link restraints between Cα atoms, we applied a flat-bottom quartic function to define the restraints, where the well depth and well-wall breadth have the same value of 1. The well depth represents the confidence of the corresponding distance in the protein complex. More details can be found in a previously published paper from Lubecka and Liwo.[64] The restrained distances between residue pairs were within ±1 Å of the distance in the native complex. There is no preference as to the nature of the restraint residues; as long as they meet the restraint criteria, they can be hydrophobic, hydrophilic, or charged residues. In total, we generated 450 restraints, including 296 pairs of Taz2 intramolecular restraints and 154 pairs of intermolecular restraints (see Table S1 for a list of cross-link restraints imposed in our simulations). These distance restraints allowed Taz2 to possess a native-like structure and the simulation to be biased with the intermolecular cross-link restraints to produce bound complexes. However, there are no restraints within TAD2, ensuring the free dynamics of TAD2 during the simulation. In this way, we can study the binding and folding process of TAD2 in the proximity to Taz2.

Root-Mean-Square Deviation (RMSD) Calculation and the Selected Trajectories. The RMSD of the p300 Taz2−p53 TAD2 complex was calculated relative to the first model of NMR-resolved structures.[8] The RMSD of p53 TAD2 was calculated in the same way. For p53 TAD2, RMSD values were analyzed for different regions, including the $TAD2_{47-55}$ helix and the entire $TAD2_{35-59}$ domain. Based on the complex RMSD values, 190 out of the 720 replicas were identified as having an average RMSD value (in the last 100 frames, frames 2500−2599) of less than or equal to 2.5 Å. This complex RMSD cutoff is used as a criterion for complex binding, as structures with RMSD values of 3 and 2.5 Å (Figures 1B and 1C) show good binding to Taz2, whereas those with RMSD values of 3.9 and 7 Å (Figures 1D and 1E) show offset or loose binding. Taking into account the formation of the $TAD2_{47-55}$ helix, 182 of the 190 replicas (Table S2) were further selected using a $TAD2_{47-55}$ helix RMSD (average of the last 100 frames) cutoff of 1.4 Å. The eight excluded replicas with a $TAD2_{47-55}$ helix RMSD greater than 2.9 Å indicate the final unfolding of TAD2. Based on the RMSD results, these 182 trajectories with successful binding and TAD2 folding were selected to carry out further analyses. To monitor the progress of the binding simulations, the native contact fraction was computed for each frame by comparison with the native contacts formed in the experimental structure with a distance cutoff of 5.5 Å between any heavy atoms of two residues.

Contact Formation Order and Cluster Analysis. To track the formation of intermolecular contacts of the Taz2−TAD2 complex and contacts within TAD2, the frame number at which a pair of residues made contact for the first time was recorded and ranked to generate the contact formation order in a single trajectory. A pair of residues was considered to be in contact with each other if the distance between the heavy atoms is less than 5.5 Å. The mean and standard deviation of the contact formation order on the 182 trajectories were then calculated. By analyzing the contact formation sequence, initial contacts between Taz2 and TAD2, corresponding to small formation orders, can be found. The relative formation order between intermolecular contacts and TAD2 internal contacts roughly provides information about the folding and binding sequences of TAD2. Hierarchical cluster analysis in the principal component (PC) space was conducted in Bio3D[65] along the first two largest principal components to generate five representative structures from a given structural ensemble.

PathDetect-SOM. Self-organizing maps (SOMs) are a type of unsupervised artificial neural networks which offers a clear visual representation of data on a two-dimensional map.[66,67] They have gained wide recognition for analyzing various data types, including conformations obtained from molecular dynamics (MD) simulations.[68−70]

To explore conformational details at atomic resolution, we back-mapped simulations to all atoms using PULCHRA,[71] and we calculated intermolecular distances between the Cβ atoms of TAD2 and Taz2 for each frame resulting in 2250 input features. The same distances were used in a previous publication.[31] The SOM was trained over 5000 cycles. During each cycle, input vectors representing individual conformations were randomly presented to the map and assigned to the neuron with the closest weights, known as the best matching unit (BMU). During training, the weights of the BMU and its neighboring neurons are adjusted to align with the input vector. The extent of this adjustment decreased with distance from the BMU and progressed during training. This process was iterated, preserving the topological relationships between neurons and ensuring that similar original input data remained close to the map. In a subsequent step, we grouped neurons into a small yet representative number of clusters using

https://doi.org/10.1021/acs.jctc.4c00541
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

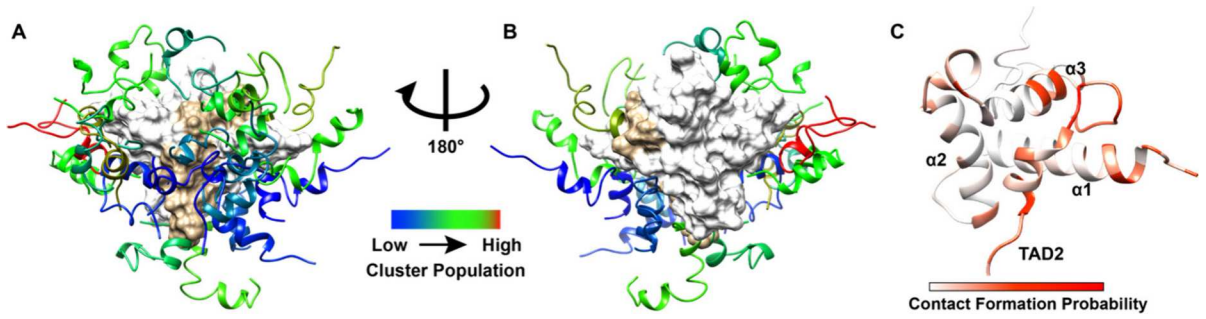


Figure 2. Taz2−TAD2 encounter complexes originating from effective collisions. In panels A and B, the native Taz2−TAD2 complexes (PDB ID 2MZD[8]) are shown as a surface view, with Taz2 in white and TAD2 in tan. The encounter complex ensemble is grouped into 20 clusters and visualized by overlaying Taz2 onto the native complex with TAD2 colored according to the corresponding cluster population. In panel C, residues forming contacts in the encounter complexes are colored in red.

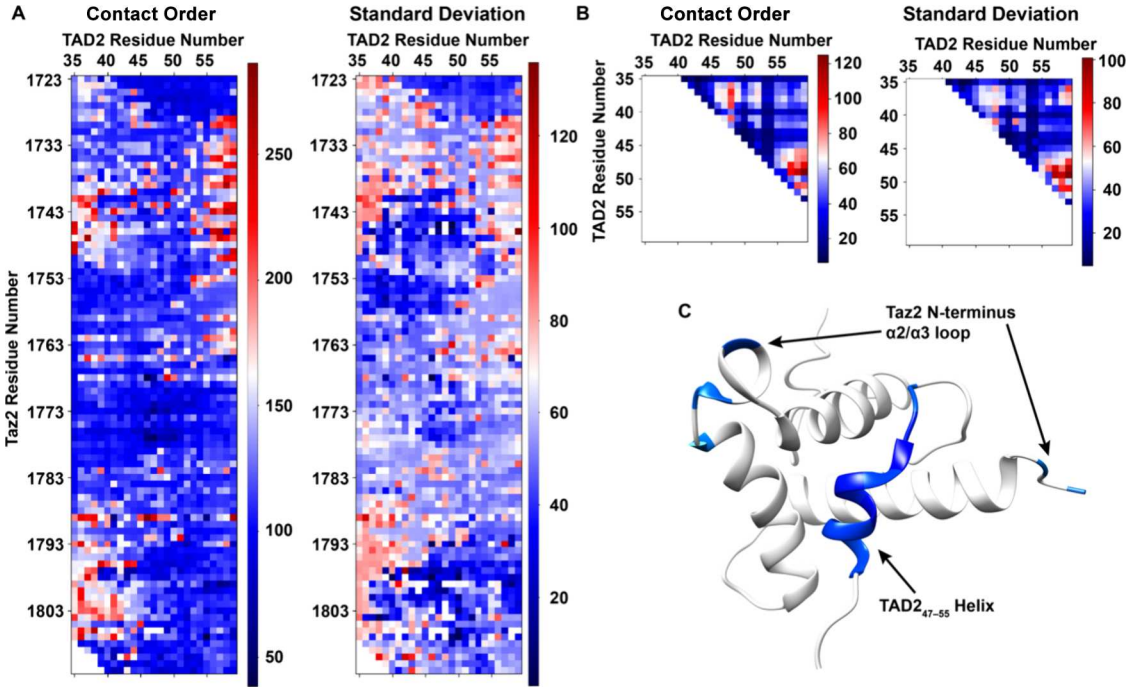


Figure 3. Contact formation order of the Taz2−TAD2 complex and the internal TAD2. Means and standard deviations (SD) of the formation orders for the intermolecular and intramolecular contact are shown in panels A and B. Residues involved in initial contacts (the first 100 contacts formed) are highlighted in panel C, with darker blue implying a higher frequency of initial contacts being formed.

agglomerative hierarchical clustering based on the Euclidean distance and complete linkage. We determined the optimal number of clusters (in this case, 12) using Silhouette profiles (Figure S1).

For this study, we employed a 10 × 10 sheet-shaped SOM without periodicity across the boundaries, using a hexagonal lattice shape. We set a distance value cap of 1.5 nm to eliminate uninformative differences between frames in unbound conformations. The trajectories of the molecular dynamics simulations are mapped onto the SOM based on the BMU annotations, allowing us to construct the path covered by each simulation on the map. Using this time-dependent annotation, we estimated an approximate transition matrix between each pair of neurons. Only transitions with at least three observations were retained. This matrix was then converted into a row-stochastic matrix, and we constructed a graph with nodes representing neurons. The edges of this graph were assigned values corresponding to the negative logarithms of the transition probabilities between the respective neurons. Due to the pruning of low-sampled transitions, some neurons may remain isolated from others and are not represented in the graphs.

Finally, we visualized the helical content of TAD2 by assigning a color code to each neuron based on the average value of the property for the frames belonging to that neuron. All analyses were carried out using the PathDetect-SOM tool[61,72] and the igraph package[73] for graph construction. All trajectories were truncated for SOM analysis when the following RMSD criteria against the native structure were satisfied: complex RMSD < 2.5 Å, TAD2 RMSD < 2.5 Å, and TAD2 helix RMSD < 1.4 Å. Following the satisfaction of the RMSD criteria, 10 additional frames were retained to ensure that the process of complete binding was captured for each replica.

RESULTS

The association mechanism was revealed by monitoring the progression of complex binding and TAD2 helix folding in

https://doi.org/10.1021/acs.jctc.4c00541
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

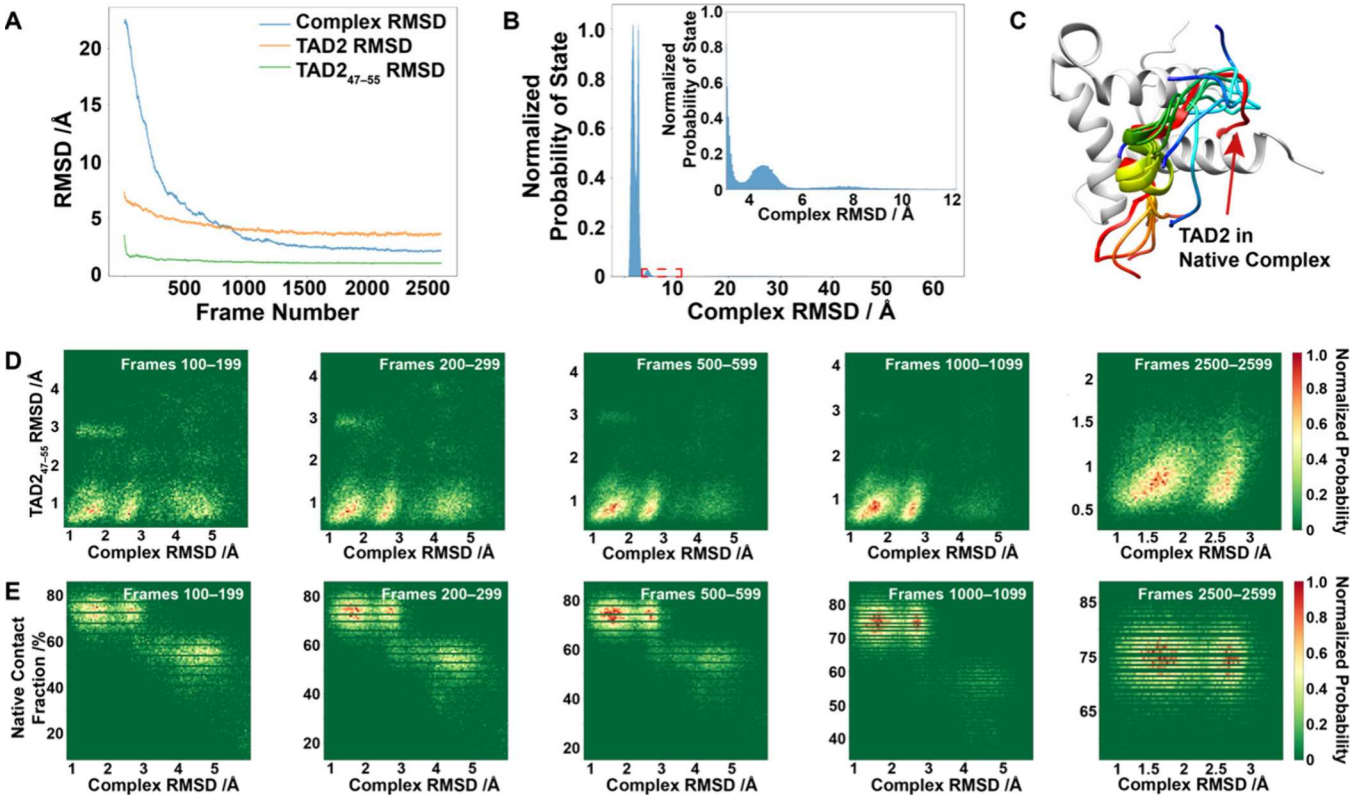


Figure 4. Root-mean-square deviation (RMSD) analysis of the Taz2−TAD2 complex, the TAD2 helix (residue 47−55) RMSD, and the native contact fraction between Taz2 and TAD2. In panel A, the average RMSD of the complex, the entire TAD2, and the TAD2 helix over all 182 replicas decreases with simulations. The complex RMSD distribution is shown in panel B. Inset: RMSD distribution, highlighted in the red dashed box. By aligning Taz2, the intermediate states (conformations in panel B with the complex RMSD in the range of 3.5−5.5 Å) are overlaid with the native complex (C), where native TAD2 is shown in red and edged ribbon view and TAD2 in the intermediate states is colored in rainbow colors (blue−red representing the N−C terminus) and shown in a rounded ribbon view. 2D-RMSD distributions (D) were calculated for different binding stages by extracting frames 100−199, 200−299, 500−599, 1000−1099, and 2500−2599. The native contact fraction between Taz2 and TAD2 relative to the complex RMSD (E) is plotted at the same binding stages as in panel D.

UNRES binding simulations between Taz2 and TAD2. By analyzing 182 replicas, we extracted snapshots of encounter complexes and intermediate species from trajectories and identified binding pathways from PathDetect-SOM analysis.

Encounter Complexes and Initial Intermolecular Contacts. Interpretation of the binding mechanism relies on the detection of key species during the association process and on the order of their occurrence. To form an encounter complex, the first question is whether TAD2 randomly collides with Taz2. By monitoring the minimum distance between Taz2 and TAD2 (Figure S2), two distinct collision scenarios were detected using a minimum intermolecular distance cutoff of 2 Å as the threshold for the formation of encounter complexes. In some cases, more than one collision was detected, including ineffective and finally productive collisions, while in other cases a single collision was sufficient to prompt subsequent structural transition. Considering encounter complexes specifically originating from effective collisions, we extracted an encounter structure for each trajectory and grouped them into 20 clusters. When we align Taz2 in clusters with the native complex,[8] TAD2 shows a strong tendency to collide with the side of the binding site, as shown in Figure 2A. Few effective collisions occurred on the opposite side of the binding site (Figure 2B). In Figure 2A, a majority of the initial collisions occurred around the binding site of Taz2 (TAD2 in green), especially the $\alpha 2/\alpha 3$ loop (the loop region between α-helix 2 and α-helix 3 of Taz2, TAD2 in red). To delve into it in-depth, we mapped the residues contributing to intermolecular contacts in the encounter complexes onto a ribbon view of the native complex, as shown in Figure 2C. Loop regions on Taz2 are more prone to collide with TAD2, such as the N-terminus on the right and the $\alpha 2/\alpha 3$ loop on the upper left. Therefore, this loop region may play a role in the initial stage of protein binding. Key residues in the encounter complexes starting from different structures were mapped onto the native complex, and we can see modest differences between these residues in Figure S3.

As the starting point for binding events, the encounter complex will lead to the formation of more contacts and the positioning of TAD2 to the correct binding pocket. We exploited the order of contact formation to analyze the initial contacts. Identifying initial intermolecular contacts can help us understand the structural transition from encounter complexes to native-like complexes. The initially formed contacts (blue areas on the left panels of Figures 3A and 3B) with low standard deviations (blue areas on the right panel of Figures 3A and 3B) are of great significance. If we compare the blue areas in Figure 3A (initial contacts) with the native contacts found in the NMR-resolved Taz2−TAD2 complex,[8] we note that the initial contacts are often non-native contacts. For intramolecular contacts in TAD2, the initial contacts involve residues within the TAD2 helix (residues 47−55), suggesting that the helix can form earlier than other segments within TAD2. As shown in Figure S4, we counted the number of

https://doi.org/10.1021/acs.jctc.4c00541
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

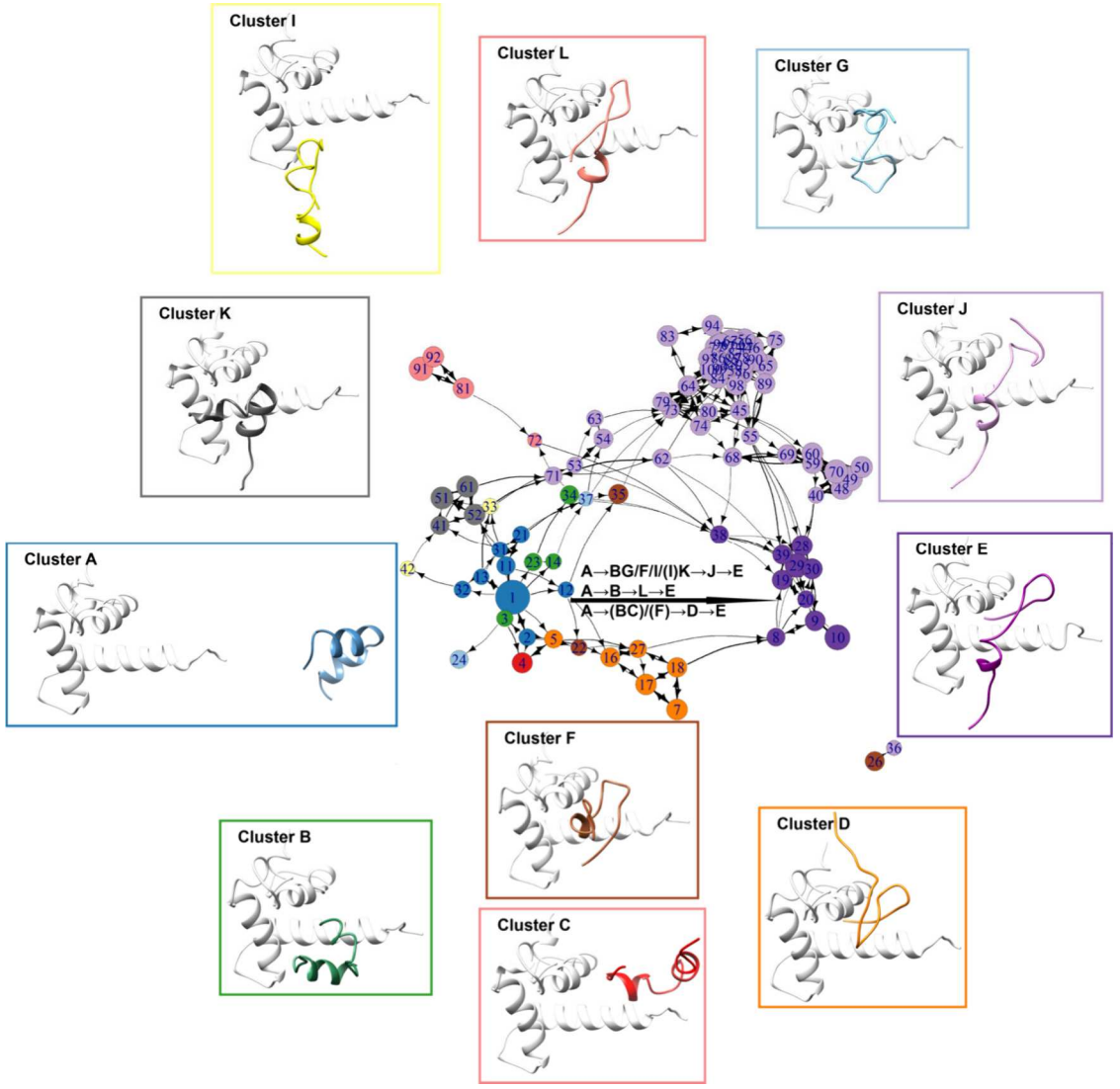


Figure 5. Transition network diagram that represents the PathDetect-SOM transition matrix between neurons. Arrows represent structural switches between two neurons. Representative structures for each cluster are shown with borders and TAD2 colored according to the color of the clusters in Figure S8A. Clusters A and E represent the unbound and bound states, respectively. Clusters B−C and F−I appear to be the encounter complex mentioned in previous sections, while clusters J−L appear to be the intermediate state carrying a dynamic TAD2 N-terminus, and cluster D appears to be a less populated intermediate state with the TAD2 N-terminus attached to the Taz2 binding pocket instead. SOM neuron numbers can be found in Figure S8A.

occurrences of Taz2 and TAD2 residues in different contact formation order ranges. We can see that the N-terminus (residues 1723−1728) and the $\alpha 2/\alpha 3$ loop (residues 1770−1773, 1776) of Taz2 can make contacts with TAD2 at an early stage (the blue and orange bars with formation order in the range of 1−100). To visualize the residues involved in the initial contacts, we mapped these residues to the native complex and colored them according to the frequency with which they formed initial contacts (the first 100 contacts formed), as shown in Figure 3C. Regions surrounding the binding site are responsible for the initial contacts, confirming our findings in the encounter complex shown in Figures 2A and 2B. For TAD2, the helical region (residues 47−55) first contacts Taz2 (Figure S4). However, when we extract intermolecular contacts with a standard deviation of less than 20 and count the corresponding number of occurrences (Figure S5), there are fewer residues that make initial contacts in the same formation order in all trajectories, indicating that there is no common formation sequence and the structural transition to the native-like complex is different in all replicas.

Intermediate States. Chemicals during reactions sometimes get stuck and accumulate in intermediate states. In the same way, this also happens during the folding/binding process. To roughly monitor the binding and folding progression and capture transient intermediate species, we calculated the RMSD of the Taz2−TAD2 complex, TAD2 (residues 35−59), and the TAD2 helix (residues 47−55). As shown in Figure 4A, the RMSD values of the TAD2 helix and TAD2 decrease at the beginning of simulations, while the complex RMSD takes longer to reach low values. We also made a histogram plot of the complex RMSD (Figure 4B), with the two main peaks representing the bound complex and the greater RMSD value indicating that the complex is in the bound state but with the TAD2 terminus deviating from the native state. Additionally, in the enlarged RMSD profile, we noticed two small peaks in the range of 3.5−5.5 and 6−10 Å.

https://doi.org/10.1021/acs.jctc.4c00541
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

Ensembles with complex RMSDs of 6−10 Å are encounter complexes, and ensembles with RMSDs of 3.5−5.5 Å are intermediate species. To explore the structural space of the complex along simulations, we took snapshots at different stages (frames 100−199, 200−299, 500−599, 1000−1099, and 2500−2599), and plotted the TAD2 helix RMSD versus the complex RMSD, as shown in Figure 4D. We also performed cluster analysis on frames 100−199, 200−299, 500−599, 1000−1099, and 2500−2599, and the population of bound complexes gradually increased as the simulation progressed (Figure S6). In the 2D-RMSD plots for frames 2500−2599, two major energy basins, corresponding to the two peaks of bound complexes in Figure 4B (RMSD less than 3.5 Å), suggest that bound structures have been formed for all replicas. For the remaining plots in Figure 4D, we truncated the complex RMSD to below 6 Å, including the bound structures and intermediate species (the entire RMSD plots are shown in Figure S7). We can clearly observe a local energy basin with the complex RMSD in the range of 3.5−5.5 Å at frames 100−199 and 200−299, and a reduction in its population at frames 500−599 and 1000−1099 suggests a structural shift from the intermediate states to bound complexes. This intermediate species has a low RMSD value for the TAD2 helix, indicating the presence of a folded helix in TAD2. We extracted the intermediate species from all replicas and performed cluster analyses to group them into five clusters. We aligned these clusters to the native complex (Figure 4C) and found that for most replicas, the TAD2 helix was formed and anchored in the binding pocket, with slight fluctuations in the TAD2 N-terminus. The presence of the intermediate states indicates that the TAD2 helix readily folds and binds to Taz2, giving rise to a local energy minimum before the formation of a fully bound complex. Additionally, a less obvious energy basin was observed in the upper left of the 2D-RMSD plots (TAD2 helix RMSD around 3 Å and complex RMSD around 1.5−2.5 Å) at frames 100−199, 200−299, 500−599, and 1000−1099. This minor intermediate species represents a bound complex with an attached TAD2 N-terminus and unfolded TAD2 helix. In addition, the native contact fraction between Taz2 and TAD2 were used to reveal the progression of protein binding. Notably, the native contact fraction results confirm our observation of the intermediate state. When the complex RMSD is in the range of 3.5−5.5 Å, the native contact fraction is in the range of 40−60% (Figure 4E), suggesting the presence of a metastable species before the formation of the native-like structure.

Binding Pathways Identified in PathDetect-SOM. The analysis in the previous section identified two different intermediate species, implying there are multiple possible binding pathways, so PathDetect-SOM analysis was used to precisely get an overview of the types of pathways sampled in the different replicas. A SOM map containing 10 × 10 neurons was obtained after iterative training, which were further grouped into 12 macrostates, with cluster A and cluster E representing unbound and bound structures, respectively. We can capture the detailed structural transition network from the diagram representing the neuron transition matrix (Figure 5), and the neuron transition matrix was also projected onto the SOM map (Figure S8B). The transition network shows that multiple pathways through different clusters can link cluster A to cluster E. Based on the SOM result, there are two main types of pathways: pathway 1, which passes from the neurons on top of the graph (clusters G, I, and J−L), first binds at the small helix of TAD2, and then arranges of the extreme portion of the TAD2; and pathway 2, which passes from the neurons at the bottom of the graph (clusters C and D), first binds at the TAD2 N-terminus, and then arranges in the secondary structure of TAD2. For example, as shown in the upper part of the transition network diagram in Figure 5, from cluster A, simulations may evolve through cluster K, then cluster J, and finally transform into cluster E. In this pathway, representative structures of clusters K and J possess structures resembling the intermediate species described above, in which the TAD2 helix is anchored in the binding pocket but the N-terminus still fluctuates near the binding pocket. Sometimes, cluster A first transforms into cluster I and then enters into cluster K (clusters A−I−K−J−E), with cluster I like the encounter complex mentioned above. TAD2 then correctly resides in the Taz2 binding pocket, as more local contacts are formed. We also tracked a few similar pathways passing clusters, such as clusters A−BG/F/I−J−E (where the structural transition spans clusters BG, F, or I between clusters A and J) and clusters A−B−L−E. These pathways cover clusters J or L that have a structure similar to the intermediate state and can be considered as the first type of pathway (pathway 1), suggesting that many replicas undergo the binding of the TAD2 helix prior to the TAD2 N-terminus. Alternatively, we also observed a second type of pathway (pathway 2, lower part of the transition network diagram in Figure 5), which subsequently passes through clusters A−(BC)/(F)−D−A. Cluster D has the TAD2 N-terminus attached and the helix unfolded and wiggling, corresponding to the local energy basin in Figure 4D (the complex RMSD and TAD2 helix RMSD of approximately 2 and 3 Å, respectively). Compared to the specific structural ensembles identified in previous sections, clusters B−C and F−I resemble encounter complexes, clusters J−L resemble the intermediate species, and cluster D resembles the minor intermediate species distinct from clusters J−L. The number of replicas that pass through cluster D is 53 (about 30%) while 110 (about 60%) are those that pass through cluster J. Then, a small percentage (17 replicas, about 10%) does not pass through either cluster but follows alternative routes (for example, through cluster L or through neuron 37). The two remaining trajectories passed both cluster D and cluster J. In addition, the TAD2 helical fraction was monitored and mapped onto the SOM, as shown in Figures S8C and S8D. Pathway 1 covering intermediates J−L exhibits a helical preference of 8% (up to 19% across 25 residues) before conversion to the bound structure; in contrast, pathway 2 covering intermediate D has a lower helix content of about 5%. The helical fraction in the unbound state suggests that on one hand, p53 TAD2 prefers to adopt a helical structure, and on the other hand, it also indicates an overestimation of helical stability in the UNRES force field.

Briefly, pathways 1 and 2 detected in the SOM neuron transition network reveal a three-stage process illustrating a possible binding mechanism, in which TAD2 first collides with Taz2 to form an encounter complex, then TAD2 helix binds to the binding pockets to form an intermediate state in pathway 1 (whereas in pathway 2, the TAD2 N-terminus binds first), and finally the remaining regions of TAD2 search for the correct binding site to complete the entire binding process. Nevertheless, these pathways do not rule out other possible pathways, for example, pathways with rapid protein binding do not cover many clusters. The distance restraints imposed during simulations may also have an influence on the sampled

https://doi.org/10.1021/acs.jctc.4c00541
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

conformations and then on the captured structural transition pathways.

DISCUSSION

Binding Mechanism of Taz2 and TAD2. Protein recognition, especially when a protein is disordered, is highly complex, because the process involves binding and IDP folding. Its identification sheds light on protein engineering and rational drug design targeting specific protein−protein interactions. In this work, we applied computational tools to study the binding mechanism of Taz2 and TAD2 making use of the Gō-like model to boost the simulations. Although we applied restraints during simulations, the presence of non-native intermolecular contacts observed in the contact formation order analysis (Figure 3) demonstrates that the chemical cross-link restraints did not conceal contributions of non-native interactions. Hence, the Gō-like model can still be utilized to reveal the mechanism. An adequate interpretation of the recognition mechanism requires characterization of the important species. First, as shown in Figure 2, as the starting point of the association process, the encounter complex is a collection of structures in which TAD2 is widely present on the side of the Taz2 binding site rather than the other side. The collision probability is extremely high in the vicinity of the flexible regions of Taz2 (the N-terminus and the $\alpha 2/\alpha 3$ loop), indicating the importance of these regions in protein binding. TAD2 then slides across the Taz2 surface in search of the binding pocket. As shown in Figure 2, we recorded and ranked the contact formation order for each replica. Initial contacts identified in all 182 replicas (blue and orange bars in Figure S4) involve Taz2 residues, such as those in the terminal regions and the $\alpha 2/\alpha 3$ loop (residues 1770−1773, 1776), and TAD2 residues, such as those in the helical region. If we consider initial contacts with a standard deviation of less than 20 (orange bars in Figure S5), we find that the early formed contacts involved the TAD2 N-terminus rather than the helix, possibly due to helical residues filtered through the standard deviation sieve. This implies that the formation order of TAD2 helix residues is diverse in different replicas, which does not conflict with the initial contacts observed in Figure S4. In addition, based on analyses of the complex RMSD and native contact fraction in Figure 4, an intermediate state was found, corresponding to the structure in which the folded TAD2 helix binds to Taz2 while simultaneously bearing a swinging N-terminus. A minor intermediate state attached through the TAD2 N-terminus was also discovered. The results reported here do not contradict previous observations on other systems, which suggest that fluctuations in loop regions may have little to no effect on the initial encounter between the IDP and its binding partners.[74,75] The specific region identified as the encounter site consists of three positively charged residues (Lys1772, Arg1773, and Lys1774) that attract the acidic residues of TAD2. Several factors potentially contribute to this observation. Flexible loops that cover a larger spatial range may increase the probability of encounters when complexes involve IDPs. Additionally, the fly casting effect can be largely offset by a slower diffusion. In our simulations, no explicit solvent was used, resulting in faster diffusion compared with regular MD simulations. Furthermore, coarse-grained (CG) models have smoother free energy landscapes, leading to faster diffusion. These combined factors of electrostatic interactions and loop flexibility may synergistically lead to the observation that intermolecular interactions predominantly occur in the flexible Taz2 loop region. However, the contribution of fast diffusion and the smoother free energy landscape of a coarse-grained model may amplify the role of fluctuations in the loop region.

The PathDetect-SOM results reveal multiple pathways from the unbound to bound structures, suggesting that the association process is complicated. However, most pathways can be divided into two types. In pathway 1, the folded TAD2 helix anchors to the binding site first, followed by further binding of the TAD2 N-terminus. This partially bound complex was identified as the intermediate species in the association process, highlighting a significant contribution of the TAD2 helix to the Taz2−TAD2 recognition. In contrast, pathway 2 through cluster D represents a different binding pathway in which the TAD2 N-terminus binds to the Taz2 binding pocket before the helix. In general, the binding of Taz2 and TAD2 undergoes three stages including protein collision, TAD2 partial binding, and complete binding. The presence of a few replicas with fast binding suggests that the binding and helical folding of TAD2 are rapid (providing few insights into the recognition mechanism). Furthermore, NMR studies indicated that the binding rate between Taz2 and TAD2 is comparable to the helix folding rate at the high protein concentration used in the kinetic study and should be slightly slower at *in vivo* concentration.[76] This is consistent with the presence of the TAD2 helix in pathway 1 prior to complete binding. In summary, the identification of key species and typical pathways in the association process paints a detailed picture for us to better understand the recognition mechanism.

Previous experimental works provided valuable information for the interactions between TAD2 and Taz2, such as the structural characterization of the p300 Taz2−p53 TAD2 complex.[8] Furthermore, the binding and folding rates in the p53 TAD2 and CBP (paralog of p300) Taz2 system calculated using NMR spectroscopy showed that the binding rate was fast and even comparable to the helical folding rate at the concentrations in the experiments.[76] This is consistent with our observations that TAD2 is prefolded before approaching Taz2. Apart from these findings, few experimental works have focused on the binding mechanism of our protein system. However, previous studies on protein binding of a similar NCBD and disordered ACTR system showed that the ACTR helical propensity in the preformed structure was significant for molecular recognition based on results from NMR spectroscopy and fluorescence-monitored stopped-flow kinetic techniques.[77] For more concise knowledge of the binding mechanism and kinetic process of disordered proteins, please read these reviews: refs 20 and 78−80.

Computational Strategies for Protein Binding Mechanisms. The computational approaches used in this work can be applied for the analysis of other protein systems, providing a detailed description at the atom/residue level of protein activities. By capturing key species and structural conversions in the association processes, we gain insight into how proteins recognize each other. With specially tailored restraints, TAD2 is able to capture Taz2 with high flexibility, adjusting its conformation accordingly during simulations. This aspect holds significant relevance for investigating the behavior of disordered proteins during protein recognition. This is the first time that PathDetect-SOM analysis has been used in protein association studies. Through the key species in the simulations and the structural transition diagram identified by the SOM, we can briefly grasp the mechanism behind the extensive replicas.

https://doi.org/10.1021/acs.jctc.4c00541
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

However, while computational approaches offer valuable insights into protein association processes, experimental validation is indispensable for confirming the accuracy of these findings and providing a comprehensive understanding of the underlying mechanisms. Protein binding processes can be studied using methods such as NMR,[76] stopped-flow and time-resolved assays,[81,82] Φ value analysis,[83] among other methods.[84] For example, to confirm the role of the flexible regions of Taz2 in the encounter complex, the binding rates of native Taz2 and its mutant counterpart can be monitored, in which Taz2 can be engineered by fixing the geometry of the Taz2 $\alpha 2/\alpha 3$ loop. To verify the order of contact formation, NMR titration methods can be used to track changes in the chemical shifts of specific amino acid residues. We deduce that the initial contacts identified in our work would undergo chemical shift perturbations before later-formed contacts. Furthermore, surface plasmon resonance has been used to study the low-populated binding intermediate of a protein−RNA system by designing mutations that interfere with critical interactions in the intermediate species.[84] A combination of experimental and computational techniques is required to elucidate puzzling research from multiple aspects.

CONCLUSIONS

In summary, we modeled the association process between Taz2 and TAD2 using coarse-grained UNRES simulations with additional Gō-like potentials and studied the binding mechanism by analyzing the structural transition features over time and identifying pathways for the structural transition between neurons. By imposing Gō-like restraints, we can significantly speed up binding simulations, making it possible to study a long-timescale biological activity with large spatial conformational changes. Based on the 2D-RMSD plots in Figure S7, the TAD2 helix is dynamic but prefers to adopt a helical structure in the isolated state. When approaching Taz2, TAD2 tends to collide with the side of the Taz2 binding site. As TAD2 slides across the Taz2 surface in search of the binding pocket, multiple non-native interactions occurred transiently to stabilize the corresponding structural ensembles. Prior to conversion to the final product, an intermediate species was detected in which TAD2 carrying a dynamic N-terminus was bound to the binding pocket via the TAD2 helix, which was confirmed in pathway 1 based on the PathDetect-SOM analysis.

ASSOCIATED CONTENT

Data Availability Statement

The data underlying this article are available in OneDrive at https://tinyurl.com/4dtknt2p.

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jctc.4c00541.

Chemical cross-link restraints imposed during UNRES simulations and additional figures on binding mechanisms, such as contact formation order, intermediate species, and SOM analysis (PDF)

AUTHOR INFORMATION

Corresponding Author

Yi He − *Department of Chemistry & Chemical Biology, The University of New Mexico, Albuquerque, New Mexico 87131, United States; Translational Informatics Division, Department of Internal Medicine, The University of New Mexico, Albuquerque, New Mexico 87131, United States;* orcid.org/0000-0002-6884-5312; Email: yihe@unm.edu

Authors

Tongtong Li − *Department of Chemistry & Chemical Biology, The University of New Mexico, Albuquerque, New Mexico 87131, United States;* orcid.org/0009-0001-9724-7418

Stefano Motta − *Department of Earth and Environmental Sciences, University of Milano-Bicocca, Milan 20126, Italy;* orcid.org/0000-0002-0812-6834

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jctc.4c00541

Author Contributions

The manuscript was written through the contributions of all authors. All authors have given approval to the final version of the manuscript.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

This work was supported by the National Science Foundation (Grant 2237369). This work was also supported by the Substance Use Disorders Grand Challenge Pilot Research Award, the startup fund from the University of New Mexico, and the University of New Mexico Office of the Vice President for Research WeR1 Faculty Success Program.

REFERENCES

(1) Dyson, H. J.; Wright, P. E. Intrinsically Unstructured Proteins and Their Functions. *Nat. Rev. Mol. Cell Biol.* **2005**, *6* (3), 197−208.

(2) Handa, T.; Kundu, D.; Dubey, V. K. Perspectives on Evolutionary and Functional Importance of Intrinsically Disordered Proteins. *Int. J. Biol. Macromol.* **2023**, *224*, 243−255.

(3) Van Der Lee, R.; Buljan, M.; Lang, B.; Weatheritt, R. J.; Daughdrill, G. W.; Dunker, A. K.; Fuxreiter, M.; Gough, J.; Gsponer, J.; Jones, D. T.; Kim, P. M.; Kriwacki, R. W.; Oldfield, C. J.; Pappu, R. V.; Tompa, P.; Uversky, V. N.; Wright, P. E.; Babu, M. M. Classification of Intrinsically Disordered Regions and Proteins. *Chem. Rev.* **2014**, *114* (13), 6589−6631.

(4) Chen, J.; Kriwacki, R. W. Intrinsically Disordered Proteins: Structure, Function and Therapeutics. *J. Mol. Biol.* **2018**, *430* (16), 2275−2277.

(5) Hsu, W.-L.; Oldfield, C. J.; Xue, B.; Meng, J.; Huang, F.; Romero, P.; Uversky, V. N.; Dunker, A. K. Exploring the Binding Diversity of Intrinsically Disordered Proteins Involved in One-to-Many Binding. *Protein Sci.* **2013**, *22* (3), 258−273.

(6) Liu, J.; Perumal, N. B.; Oldfield, C. J.; Su, E. W.; Uversky, V. N.; Dunker, A. K. Intrinsic Disorder in Transcription Factors. *Biochemistry* **2006**, *45* (22), 6873−6888.

(7) Hafner, A.; Bulyk, M. L.; Jambhekar, A.; Lahav, G. The Multiple Mechanisms That Regulate P53 Activity and Cell Fate. *Nat. Rev. Mol. Cell Biol.* **2019**, *20* (4), 199−210.

(8) Miller Jenkins, L. M.; Feng, H.; Durell, S. R.; Tagad, H. D.; Mazur, S. J.; Tropea, J. E.; Bai, Y.; Appella, E. Characterization of the P300 Taz2-P53 TAD2 Complex and Comparison with the P300 Taz2-P53 TAD1 Complex. *Biochemistry* **2015**, *54* (11), 2001−2010.

(9) Okuda, M.; Nishimura, Y. Extended String Binding Mode of the Phosphorylated Transactivation Domain of Tumor Suppressor P53. *J. Am. Chem. Soc.* **2014**, *136* (40), 14143−14152.

(10) Bah, A.; Forman-Kay, J. D. Modulation of Intrinsically Disordered Protein Function by Post-Translational Modifications. *J. Biol. Chem.* **2016**, *291* (13), 6696−6705.

(11) Schöneberg, T.; Liebscher, I. Mutations in G Protein-Coupled Receptors: Mechanisms, Pathophysiology and Potential Therapeutic Approaches. *Pharmacol. Rev.* **2021**, *73* (1), 89−119.
(12) Uversky, V. N.; Davé, V.; Iakoucheva, L. M.; Malaney, P.; Metallo, S. J.; Pathak, R. R.; Joerger, A. C. Pathological Unfoldomics of Uncontrolled Chaos: Intrinsically Disordered Proteins and Human Diseases. *Chem. Rev.* **2014**, *114* (13), 6844−6879.
(13) Koshland, D. E., Jr; Némethy, G.; Filmer, D. Comparison of Experimental Binding Data and Theoretical Models in Proteins Containing Subunits. *Biochemistry* **1966**, *5* (1), 365−385.
(14) TSAI, C.-J.; KUMAR, S.; MA, B.; NUSSINOV, R. Folding Funnels, Binding Funnels, and Protein Function. *Protein Sci.* **1999**, *8* (6), 1181−1190.
(15) Ma, B.; Kumar, S.; Tsai, C.-J.; Nussinov, R. Folding Funnels and Binding Mechanisms. *Protein Eng. Des. Sel.* **1999**, *12* (9), 713−720.
(16) Tsai, C.-J.; Ma, B.; Sham, Y. Y.; Kumar, S.; Nussinov, R. Structured Disorder and Conformational Selection. *Proteins Struct. Funct. Bioinforma.* **2001**, *44* (4), 418−427.
(17) Robustelli, P.; Piana, S.; Shaw, D. E. Mechanism of Coupled Folding-upon-Binding of an Intrinsically Disordered Protein. *J. Am. Chem. Soc.* **2020**, *142* (25), 11092−11101.
(18) Monod, J.; Wyman, J.; Changeux, J. P. On the Structure of Allosteric Transitions: A Plausible Model. *J. Mol. Biol.* **1965**, *12*, 88−118.
(19) Hammes, G. G.; Chang, Y.-C.; Oas, T. G. Conformational Selection or Induced Fit: A Flux Description of Reaction Mechanism. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106* (33), 13737−13741.
(20) Csermely, P.; Palotai, R.; Nussinov, R. Induced Fit, Conformational Selection and Independent Dynamic Segments: An Extended View of Binding Events. *Trends Biochem. Sci.* **2010**, *35* (10), 539−546.
(21) Ma, B.; Nussinov, R. Enzyme Dynamics Point to Stepwise Conformational Selection in Catalysis. *Curr. Opin. Chem. Biol.* **2010**, *14* (5), 652−659.
(22) Espinoza-Fonseca, L. M. Reconciling Binding Mechanisms of Intrinsically Disordered Proteins. *Biochem. Biophys. Res. Commun.* **2009**, *382* (3), 479−482.
(23) Ponder, J. W.; Case, D. A. Force Fields for Protein Simulations. In *Advances in Protein Chemistry*; Academic Press, 2003; Vol. *66*, pp 27−85. DOI: 10.1016/S0065-3233(03)66002-X.
(24) Huang, J.; MacKerell, A. D. Force Field Development and Simulations of Intrinsically Disordered Proteins. *Curr. Opin. Struct. Biol.* **2018**, *48*, 40−48.
(25) Lee, K. H.; Chen, J. Multiscale Enhanced Sampling of Intrinsically Disordered Protein Conformations. *J. Comput. Chem.* **2016**, *37* (6), 550−557.
(26) Li, T.; Hendrix, E.; He, Y. Simple and Effective Conformational Sampling Strategy for Intrinsically Disordered Proteins Using the UNRES Web Server. *J. Phys. Chem. B* **2023**, *127* (10), 2177−2186.
(27) Terakawa, T.; Takada, S. Multiscale Ensemble Modeling of Intrinsically Disordered Proteins: P53 N-Terminal Domain. *Biophys. J.* **2011**, *101* (6), 1450−1458.
(28) Gong, X.; Zhang, Y.; Chen, J. Advanced Sampling Methods for Multiscale Simulation of Disordered Proteins and Dynamic Interactions. *Biomolecules.* **2021**, *11*, 1416.
(29) Munshi, S.; Rajendran, D.; Ramesh, S.; Subramanian, S.; Bhattacharjee, K.; Kumar, M. R.; Naganathan, A. N. Controlling Structure and Dimensions of a Disordered Protein via Mutations. *Biochemistry* **2020**, *59* (2), 171−174.
(30) Liu, X.; Chen, J.; Chen, J. Residual Structure Accelerates Binding of Intrinsically Disordered ACTR by Promoting Efficient Folding upon Encounter. *J. Mol. Biol.* **2019**, *431* (2), 422−432.
(31) Li, T.; Motta, S.; Stevens, A. O.; Song, S.; Hendrix, E.; Pandini, A.; He, Y. Recognizing the Binding Pattern and Dissociation Pathways of the P300 Taz2-P53 TAD2 Complex. *JACS Au* **2022**, *2* (8), 1935−1945.
(32) Karanicolas, J.; Brooks, C. L. Improved Gō-like Models Demonstrate the Robustness of Protein Folding Mechanisms Towards Non-Native Interactions. *J. Mol. Biol.* **2003**, *334* (2), 309−325.
(33) Naganathan, A. N.; Orozco, M. The Native Ensemble and Folding of a Protein Molten-Globule: Functional Consequence of Downhill Folding. *J. Am. Chem. Soc.* **2011**, *133* (31), 12154−12161.
(34) Levy, Y.; Cho, S. S.; Onuchic, J. N.; Wolynes, P. G. A Survey of Flexible Protein Binding Mechanisms and Their Transition States Using Native Topology Based Energy Landscapes. *J. Mol. Biol.* **2005**, *346* (4), 1121−1145.
(35) Hendrix, E.; Motta, S.; Gahl, R. F.; He, Y. Insight into the Initial Stages of the Folding Process in Onconase Revealed by UNRES. *J. Phys. Chem. B* **2022**, *126* (40), 7934−7942.
(36) Kurcinski, M.; Kolinski, A.; Kmiecik, S. Mechanism of Folding and Binding of an Intrinsically Disordered Protein as Revealed by Ab Initio Simulations. *J. Chem. Theory Comput.* **2014**, *10* (6), 2224−2231.
(37) Chong, S.-H.; Im, H.; Ham, S. Explicit Characterization of the Free Energy Landscape of PKID-KIX Coupled Folding and Binding. *ACS Cent. Sci.* **2019**, *5* (8), 1342−1351.
(38) Huang, Y.; Liu, Z. Nonnative Interactions in Coupled Folding and Binding Processes of Intrinsically Disordered Proteins. *PLoS One* **2010**, *5* (11), No. e15375.
(39) Chen, T.; Song, J.; Chan, H. S. Theoretical Perspectives on Nonnative Interactions and Intrinsic Disorder in Protein Folding and Binding. *Curr. Opin. Struct. Biol.* **2015**, *30*, 32−42.
(40) He, Y.; Xiao, Y.; Liwo, A.; Scheraga, H. A. Exploring the Parameter Space of the Coarse-Grained UNRES Force Field by Random Search: Selecting a Transferable Medium-Resolution Force Field. *J. Comput. Chem.* **2009**, *30* (13), 2127−2135.
(41) Stevens, A. O.; Luo, S.; He, Y. Three Binding Conformations of BIO124 in the Pocket of the PICK1 PDZ Domain. *Cells* **2022**, *11* (15), 2451.
(42) Gil Pineda, L. I.; Milko, L. N.; He, Y. Performance of CHARMM36m with Modified Water Model in Simulating Intrinsically Disordered Proteins: A Case Study. *Biophys. Reports* **2020**, *6* (2−3), 80−87.
(43) He, Y.; Liwo, A.; Weinstein, H.; Scheraga, H. A. PDZ Binding to the BAR Domain of PICK1 Is Elucidated by Coarse-Grained Molecular Dynamics. *J. Mol. Biol.* **2011**, *405* (1), 298−314.
(44) Stevens, A. O.; Kazan, I. C.; Ozkan, B.; He, Y. Investigating the Allosteric Response of the PICK1 PDZ Domain to Different Ligands with All-atom Simulations. *Protein Sci.* **2022**, *31* (12), No. e4474.
(45) Song, S.; Li, T.; Stevens, A. O.; Raad, T.; He, Y. Investigating the Mechanical Properties and Flexibility of N-BAR Domains in PICK1 by Molecular Dynamics Simulations. *Curr. Protein Pept. Sci.* **2023**, *24* (10), 865−877.
(46) Li, T.; Stevens, A. O.; Gil Pineda, L. I.; Song, S.; Ameyaw Baah, C. A.; He, Y. Changes in Structure and Flexibility of P53 TAD2 upon Binding to P300 Taz2. *J. Theor. Comput. Chem.* **2020**, *19* (04), 2040007.
(47) He, Y.; Mozolewska, M. A.; Krupa, P.; Sieradzan, A. K.; Wirecki, T. K.; Liwo, A.; Kachlishvili, K.; Rackovsky, S.; Jagieła, D.; Ślusarz, R.; Czaplewski, C. R.; Ołdziej, S.; Scheraga, H. A. Lessons from Application of the UNRES Force Field to Predictions of Structures of CASP10 Targets. *Proc. Natl. Acad. Sci. U. S. A.* **2013**, *110* (37), 14936−14941.
(48) He, Y.; Chen, C.; Xiao, Y. United-Residue (UNRES) Langevin Dynamics Simulations of Trpzip2 Folding. *J. Comput. Biol.* **2009**, *16* (12), 1719−1730.
(49) Liwo, A.; Ołdziej, S.; Pincus, M. R.; Wawak, R. J.; Rackovsky, S.; Scheraga, H. A. A United-Residue Force Field for off-Lattice Protein-Structure Simulations. I. Functional Forms and Parameters of Long-Range Side-Chain Interaction Potentials from Protein Crystal Data. *J. Comput. Chem.* **1997**, *18* (7), 849−873.
(50) Liwo, A.; Pincus, M. R.; Wawak, R. J.; Rackovsky, S.; Ołdziej, S.; Scheraga, H. A. A United-Residue Force Field for off-Lattice Protein-Structure Simulations. II. Parameterization of Short-Range Interactions and Determination of Weights of Energy Terms by Z-Score Optimization. *J. Comput. Chem.* **1997**, *18* (7), 874−887.

https://doi.org/10.1021/acs.jctc.4c00541
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

(51) Liwo, A.; Kaźmierkiewicz, R.; Czaplewski, C.; Groth, M.; Ołdziej, S.; Wawak, R. J.; Rackovsky, S.; Pincus, M. R.; Scheraga, H. A. United-Residue Force Field for off-Lattice Protein-Structure Simulations: III. Origin of Backbone Hydrogen-Bonding Cooperativity in United-Residue Potentials. *J. Comput. Chem.* **1998**, *19* (3), 259−276.
(52) Liwo, A.; He, Y.; Scheraga, H. A. Coarse-Grained Force Field: General Folding Theory. *Phys. Chem. Chem. Phys.* **2011**, *13* (38), 16890−16901.
(53) Gołaś, E.; Maisuradze, G. G.; Senet, P.; Ołdziej, S.; Czaplewski, C.; Scheraga, H. A.; Liwo, A. Simulation of the Opening and Closing of Hsp70 Chaperones by Coarse-Grained Molecular Dynamics. *J. Chem. Theory Comput.* **2012**, *8* (5), 1750−1764.
(54) Zhou, R.; Maisuradze, G. G.; Suñol, D.; Todorovski, T.; Macias, M. J.; Xiao, Y.; Scheraga, H. A.; Czaplewski, C.; Liwo, A. Folding Kinetics of WW Domains with the United Residue Force Field for Bridging Microscopic Motions and Experimental Measurements. *Proc. Natl. Acad. Sci. U. S. A.* **2014**, *111* (51), 18243−18248.
(55) Czaplewski, C.; Karczyńska, A.; Sieradzan, A. K.; Liwo, A. UNRES Server for Physics-Based Coarse-Grained Simulations and Prediction of Protein Structure, Dynamics and Thermodynamics. *Nucleic Acids Res.* **2018**, *46* (W1), W304−W309.
(56) Al-Shuhaib, M. B. S.; Hashim, H. O.; Al-Shuhaib, J. M. B. Epicatechin Is a Promising Novel Inhibitor of SARS-CoV-2 Entry by Disrupting Interactions between Angiotensin-Converting Enzyme Type 2 and the Viral Receptor Binding Domain: A Computational/Simulation Study. *Comput. Biol. Med.* **2022**, *141*, 105155.
(57) Żyła, A.; Martel, A.; Jurczak, P.; Moliński, A.; Szymańska, A.; Kozak, M. Human Cystatin C Induces the Disaggregation Process of Selected Amyloid Beta Peptides: A Structural and Kinetic View. *Sci. Rep.* **2023**, *13* (1), 20833.
(58) Barrón-García, O. Y.; Gaytán-Martínez, M.; Ramírez-Jiménez, A. K.; Luzardo-Ocampo, I.; Velazquez, G.; Morales-Sánchez, E. Physicochemical Characterization and Polyphenol Oxidase Inactivation of Ataulfo Mango Pulp Pasteurized by Conventional and Ohmic Heating Processes. *LWT* **2021**, *143*, 111113.
(59) Wesołowski, P. A.; Wales, D. J.; Pracht, P. Multilevel Framework for Analysis of Protein Folding Involving Disulfide Bond Formation. *J. Phys. Chem. B* **2024**, *128* (13), 3145−3156.
(60) Graul, M.; Karska, N.; Wąchalska, M.; Krupa, P.; Ślusarz, M. J.; Lubocki, M.; Bieńkowska-Szewczyk, K.; Rodziewicz-Motowidło, S.; Sieradzan, A. K.; Lipińska, A. D. The N-Terminal Proline Hinge Motif Controls the Structure of Bovine Herpesvirus 1-Encoded Inhibitor of the Transporter Associated with Antigen Processing Required for Its Immunomodulatory Function. *J. Mol. Biol.* **2023**, *435* (5), 167964.
(61) Motta, S.; Callea, L.; Bonati, L.; Pandini, A. PathDetect-SOM: A Neural Network Approach for the Identification of Pathways in Ligand Binding Simulations. *J. Chem. Theory Comput.* **2022**, *18* (3), 1957−1968.
(62) Liwo, A.; Sieradzan, A. K.; Lipska, A. G.; Czaplewski, C.; Joung, I.; Zmudzińska, W.; Hałabis, A.; Ołdziej, S. A General Method for the Derivation of the Functional Forms of the Effective Energy Terms in Coarse-Grained Energy Functions of Polymers. III. Determination of Scale-Consistent Backbone-Local and Correlation Potentials in the UNRES Force Field and Force-F. *J. Chem. Phys.* **2019**, *150* (15), 155104.
(63) Sieradzan, A. K. Introduction of Periodic Boundary Conditions into UNRES Force Field. *J. Comput. Chem.* **2015**, *36* (12), 940−946.
(64) Lubecka, E. A.; Liwo, A. Introduction of a Bounded Penalty Function in Contact-Assisted Simulations of Protein Structures to Omit False Restraints. *J. Comput. Chem.* **2019**, *40* (25), 2164−2178.
(65) Grant, B. J.; Skjærven, L.; Yao, X. Q. The Bio3D Packages for Structural Bioinformatics. *Protein Sci.* **2021**, *30* (1), 20−30.
(66) Kaski, S.; Kangas, J.; Kohonen, T. Bibliography of Self-Organizing Map (SOM) Papers: 1981−1997. *Neural Comput. Surv.* **1998**, *1* (3&4), 1−176.
(67) Miljković, D. Brief Review of Self-Organizing Maps. In *2017 40th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)*; IEEE, 2017; pp 1061−1066.
(68) Pandini, A.; Fraccalvieri, D.; Bonati, L. Artificial Neural Networks for Efficient Clustering of Conformational Ensembles and Their Potential for Medicinal Chemistry. *Curr. Top. Med. Chem.* **2013**, *13* (5), 642−651.
(69) Bouvier, G.; Desdouits, N.; Ferber, M.; Blondel, A.; Nilges, M. An Automatic Tool to Analyze and Cluster Macromolecular Conformations Based on Self-Organizing Maps. *Bioinformatics* **2015**, *31* (9), 1490−1492.
(70) Motta, S.; Siani, P.; Donadoni, E.; Frigerio, G.; Bonati, L.; Di Valentin, C. Metadynamics Simulations for the Investigation of Drug Loading on Functionalized Inorganic Nanoparticles. *Nanoscale* **2023**, *15* (17), 7909−7919.
(71) Rotkiewicz, P.; Skolnick, J. Fast Procedure for Reconstruction of Full-Atom Protein Models from Reduced Representations. *J. Comput. Chem.* **2008**, *29* (9), 1460−1465.
(72) Motta, S.; Pandini, A.; Fornili, A.; Bonati, L. Reconstruction of ARNT PAS-B Unfolding Pathways by Steered Molecular Dynamics and Artificial Neural Networks. *J. Chem. Theory Comput.* **2021**, *17* (4), 2080−2089.
(73) Csardi, G.; Nepusz, T. The Igraph Software Package for Complex Network Research. *InterJournal, Complex Syst.* **2006**, *1695* (5), 1−9.
(74) Ganguly, D.; Zhang, W.; Chen, J. Electrostatically Accelerated Encounter and Folding for Facile Recognition of Intrinsically Disordered Proteins. *PLoS Comput. Biol.* **2013**, *9* (11), No. e1003363.
(75) Zhang, W.; Ganguly, D.; Chen, J. Residual Structures, Conformational Fluctuations, and Electrostatic Interactions in the Synergistic Folding of Two Intrinsically Disordered Proteins. *PLoS Comput. Biol.* **2012**, *8* (1), No. e1002353.
(76) Arai, M.; Ferreon, J. C.; Wright, P. E. Quantitative Analysis of Multisite Protein-Ligand Interactions by NMR: Binding of Intrinsically Disordered P53 Transactivation Subdomains with the TAZ2 Domain of CBP. *J. Am. Chem. Soc.* **2012**, *134* (8), 3792−3803.
(77) Iešmantavičius, V.; Dogan, J.; Jemth, P.; Teilum, K.; Kjaergaard, M. Helical Propensity in an Intrinsically Disordered Protein Accelerates Ligand Binding. *Angew. Chemie Int. Ed.* **2014**, *53* (6), 1548−1551.
(78) Zhou, H.-X.; Pang, X.; Lu, C. Rate Constants and Mechanisms of Intrinsically Disordered Proteins Binding to Structured Targets. *Phys. Chem. Chem. Phys.* **2012**, *14* (30), 10466−10476.
(79) Mollica, L.; Bessa, L. M.; Hanoulle, X.; Jensen, M. R.; Blackledge, M.; Schneider, R. Binding Mechanisms of Intrinsically Disordered Proteins: Theory, Simulation, and Experiment. *Front. Mol. Biosci.* **2016**, *3* (SEP), 1−18.
(80) Yang, J.; Gao, M.; Xiong, J.; Su, Z.; Huang, Y. Features of Molecular Recognition of Intrinsically Disordered Proteins via Coupled Folding and Binding. *Protein Sci.* **2019**, *28* (11), 1952−1965.
(81) Walker, M.; Zhang, X.-Z.; Jiang, W.; Trinick, J.; White, H. D. Observation of Transient Disorder during Myosin Subfragment-1 Binding to Actin by Stopped-Flow Fluorescence and Millisecond Time Resolution Electron Cryomicroscopy: Evidence That the Start of the Crossbridge Power Stroke in Muscle Has Variable Geometry. *Proc. Natl. Acad. Sci. U. S. A.* **1999**, *96* (2), 465−470.
(82) Lento, C.; Wilson, D. J. Subsecond Time-Resolved Mass Spectrometry in Dynamic Structural Biology. *Chem. Rev.* **2022**, *122* (8), 7624−7646.
(83) Dogan, J.; Mu, X.; Engstrom, A.; Jemth, P. The Transition State Structure for Coupled Binding and Folding of Disordered Protein Domains. *Sci. Rep.* **2013**, *3*, 3−8.
(84) Borkar, A. N.; Bardaro, M. F.; Camilloni, C.; Aprile, F. A.; Varani, G.; Vendruscolo, M. Structure of a Low-Population Binding Intermediate in Protein-RNA Recognition. *Proc. Natl. Acad. Sci. U. S. A.* **2016**, *113* (26), 7171−7176.

https://doi.org/10.1021/acs.jctc.4c00541
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX