Deciphering the Mystery in p300 Taz2-p53 TAD2 Recognition

Tongtong Li[a], Stefano Motta[b], Yi He[a,c]*

[a]Department of Chemistry & Chemical Biology, The University of New Mexico, Albuquerque, New Mexico 87131, United States

[b]Department of Earth and Environmental Sciences, University of Milano-Bicocca, Milan 20126, Italy

[c]Translational Informatics Division, Department of Internal Medicine, The University of New Mexico, Albuquerque, New Mexico 87131, United States

*Email: yihe@unm.edu

KEYWORDS: binding mechanism, intrinsically disordered TAD2 of p53, p300 Taz2, coarse-grained MD simulations, the UNited RESidue (UNRES) force field, Gō-like model

ABSTRACT: Intrinsically disordered proteins 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 the 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 in such exploration processes of binding/unbinding processes between proteins. Here, we investigated the binding mechanism between p300 Taz2 and p53 TAD2 through extensive Gō-like molecular dynamics (MD) simulations using the physics-based UNited RESidue (UNRES) force field. 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 varied structures and biological functions.[1–3] The typical structural feature of IDPs, namely conformational heterogeneity, endows them with the potential for one-to-many binding[4] 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.[5] TADs, such as p53[6] TAD2, exhibit specific local structures when in complex with their binding partners.[7,8] In addition, post-transcriptional modification (e.g., phosphorylation) of specific amino acids is a common way to modulate IDP/IDR functions.[9] Furthermore, prevalent evidence demonstrated that protein mutations at hotspots result in loss of function.[10,11] 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 the 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)[12] and conformational selection (CS)[13], 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.[12] 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 higher helical content. In addition, prefolded α-MoRE tends to unfold before binding to XD.[14] Conversely, based on the conformational selection model, proteins prefer to selectively recruit preorganized IDPs that resemble those in the complex, and then the IDP ensemble

undergoes a structural shift toward the preorganized IDP structure.[13] The flux[15] 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 facilitates rational drug design to precisely modulate specified signaling pathways.

Computational modeling is a promising strategy to explore protein structure/dynamics and structural transitions during protein association/dissociation processes at the atom/residue level as demonstrated in our previous work.[16–26] It is known that both p300 and p53 play essential biological functions related to cancers, the intrinsically disordered nature of p53 TAD2 makes it difficult to uncover the mechanism behind it. Our prior work[23] 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[27] 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[28] to study protein folding/binding[17,29,30]. Unbiased simulations[31,32] 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[33,34] 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.

In this work, we performed Gō-like MD simulations using the physics-based UNited RESidue (UNRES)[17,19,24–26,35–38] force field to investigate the binding mechanism of p300 Taz2 and p53 TAD2. Taking advantage of the native physical intermolecular interactions present 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 primary regions responsible for 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 analysis revealed the binding pathways appear to go through the structural ensembles of encounter complex and the intermediate state described above.

Methods

UNRES MD Simulation Setup

We performed binding simulations using the physics-based coarse-grained UNited RESidue (UNRES) force field [17,19,24–26,35–38] from six starting structures, where TAD2 with a random structure was placed at six different positions relative to Taz2 (referred to as up, down, left, right, front, and back afterward) and the distance between Taz2 Ile1735 and TAD2 Ser46 was 6 nm, as shown in Figure 1A. 120 independent simulations were carried out for each starting structure. 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 interactions within the Taz2 domain and intermolecular interactions between Taz2 and TAD2 from the NMR-resolved structure[7] if the distance between Cα atoms of a pair of residues is less than 8.9 Å. The restrained distances between residue pairs were within ±1 Å of the distance in the native complex. These distance restraints allowed Taz2 to possess a native-like structure and the simulation to be biased with the intermolecular interactions to produce bound complexes. However, there are no restraints within TAD2, ensuring free dynamics of TAD2 during the simulation. In this way, we can study the binding and folding process of TAD2 in 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[7]. 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 (Figure 1B,C) show good binding to Taz2, whereas those with RMSD values 3.9 and 7 Å (Figure 1D,E) show offset or loose binding. Taking into account the formation of the $TAD2_{47–55}$ helix, 182 of the 190 replicas 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 $TAD2_{47–55}$ 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.

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 contacts 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 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 sequence of TAD2. Hierarchical cluster analysis in principal component (PC) space was conducted in Bio3D[39] along the first two largest principal components to generate five representative structures from a given structural ensemble.

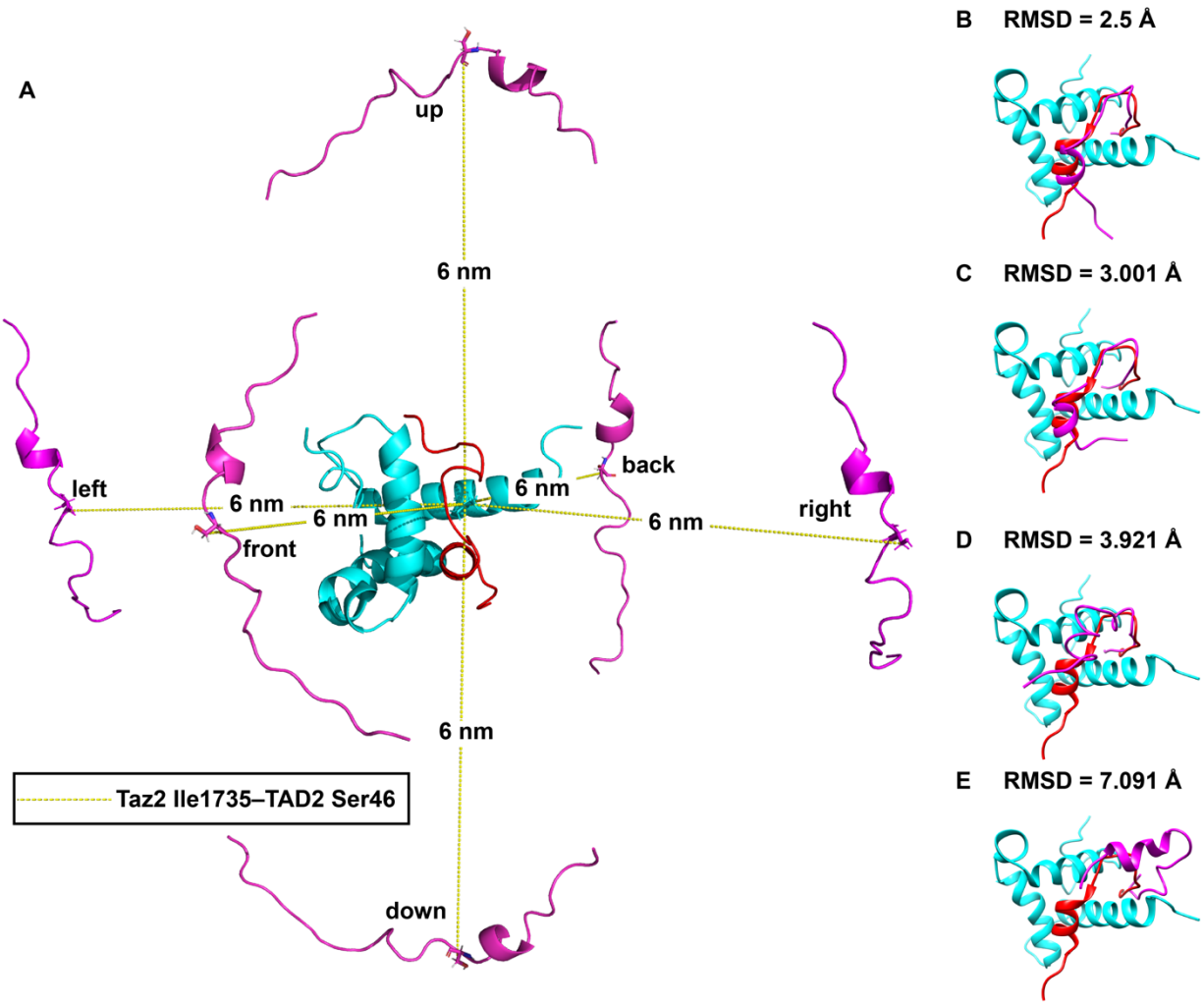

Figure 1. Six starting structures in UNRES binding simulations are shown, with Taz2 Ile1735 and TAD2 Ser46 separated by 6 nm **(A)**, and some snapshots are extracted with complex RMSD values annotated **(B–E)**. In the native Taz2-TAD2 complex (PDB ID, 2MZD[7]), Taz2 is colored in cyan and TAD2 is colored 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.

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[40,41]. They have gained wide recognition for analyzing various data types, including conformations obtained from molecular dynamics (MD) simulations[42–44].

To explore conformational details at atomic resolution, we back-mapped simulations to all atoms using PULCHRA[45], 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[23]. 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 on the map. In a subsequent step, we grouped neurons into a small yet representative number of clusters using agglomerative hierarchical clustering based on 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 10x10 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 3 observation 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 logarithm of the transition probabilities between the respective neurons.

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 [46,47] and the igraph package[48] 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 UNRES binding simulations between Taz2 and TAD2. By analyzing 182 replicas, we extracted snapshots of encounter complexes and the intermediate species from trajectories and identified binding pathways from PathDetect-SOM analysis.

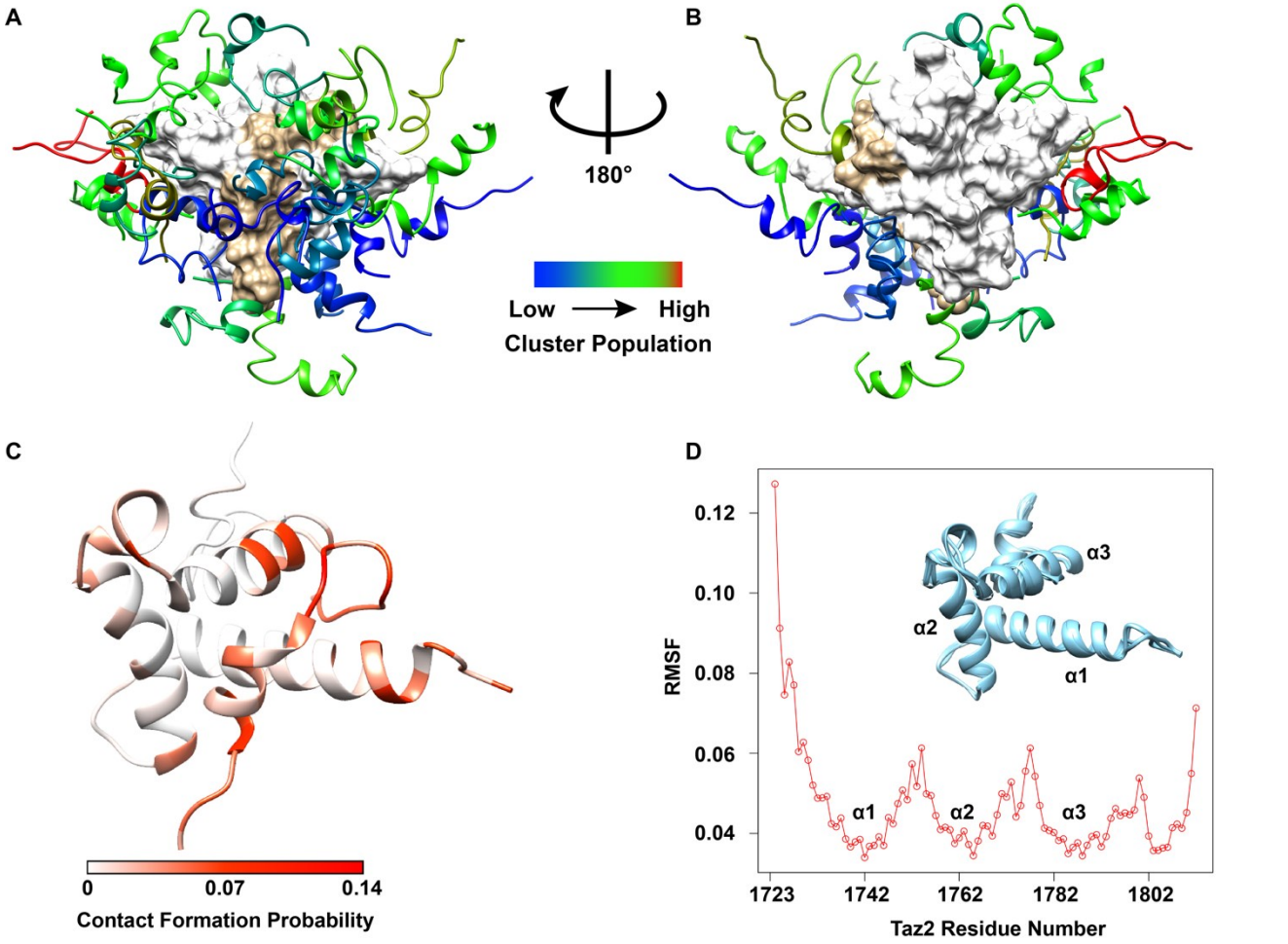

Figure 2. Taz2-TAD2 encounter complexes originating from effective collisions. In panels **(A–B)**, the native Taz2-TAD2 complexes (PDB ID, 2MZD[7]) are shown as a surface view, with Taz2 in white and TAD2 in tan. The encounter complex ensemble is grouped into 20 clusters and visualized by overlaying 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. The root-mean-square fluctuation (RMSF) of Taz2 during simulations is shown in panel **(D)**. Inset: the superimposed Taz2 in the encounter complexes.

Encounter Complexes and Initial Intermolecular Contacts

Interpretation of the binding mechanism relies on the detection of key species during the association process and 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 collision, while in other cases a single collision was sufficient to prompt subsequent structural transition. Considering specifically encounter complexes 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[7], 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 α2/α3 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 α2/α3 loop on the upper left. Such high collision frequencies on these residues are expected to result from high mobility and correspondingly large capture radius, according to the fly-casting mechanism[49]. 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. Root-mean-square fluctuation (RMSF) analysis was performed to confirm the flexible characteristics of the terminal and loop regions in Taz2, as shown in Figure 2D. The RMSF plot supports our claim of high flexibility in regions with frequent collisions.

As the starting point for binding events, the encounter complex will lead to forming more contacts and positioning TAD2 to the correct binding pocket. We exploited the order of contact formation to analyze initial contacts. Identifying initial intermolecular contacts can help understand the structural transition from encounter complexes to native-like complexes. The initially formed contacts (blue areas on the left panels of Figure 3A,B) with low standard deviations (blue areas on the right panel of Figure 3A,B) 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[7], 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 occurrences of Taz2 and TAD2 residues in different contact formation order ranges. We can see that the N-terminus (residues 1723–1728) and the α2/α3 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 Figure 2A,B. 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.

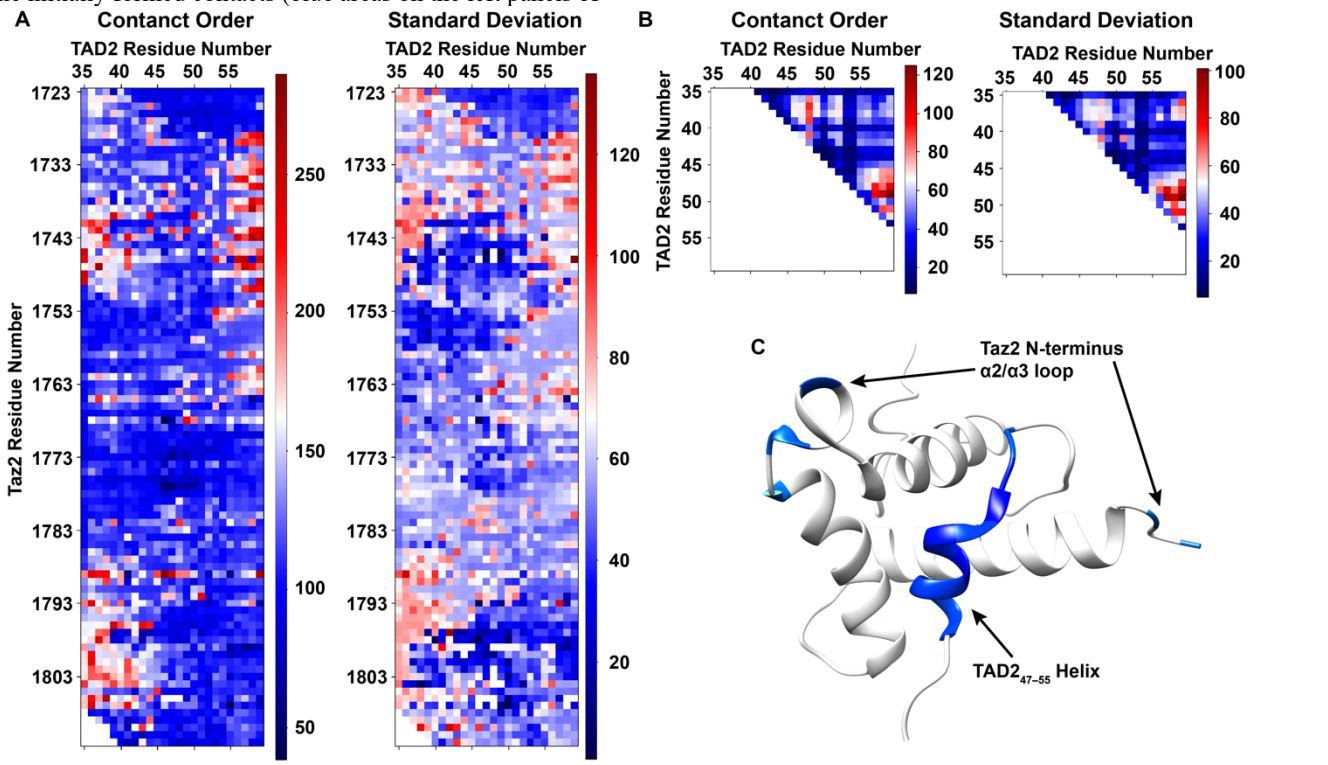

Figure 3. Contact formation order of the Taz2-TAD2 complex and internal TAD2. Means and standard deviations (SD) of the formation orders for the intermolecular and intramolecular contact are shown in panels **(A–B)**. Residues involved in initial contacts (the first 100 contacts formed) are highlighted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

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 Å and 6–10 Å. 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

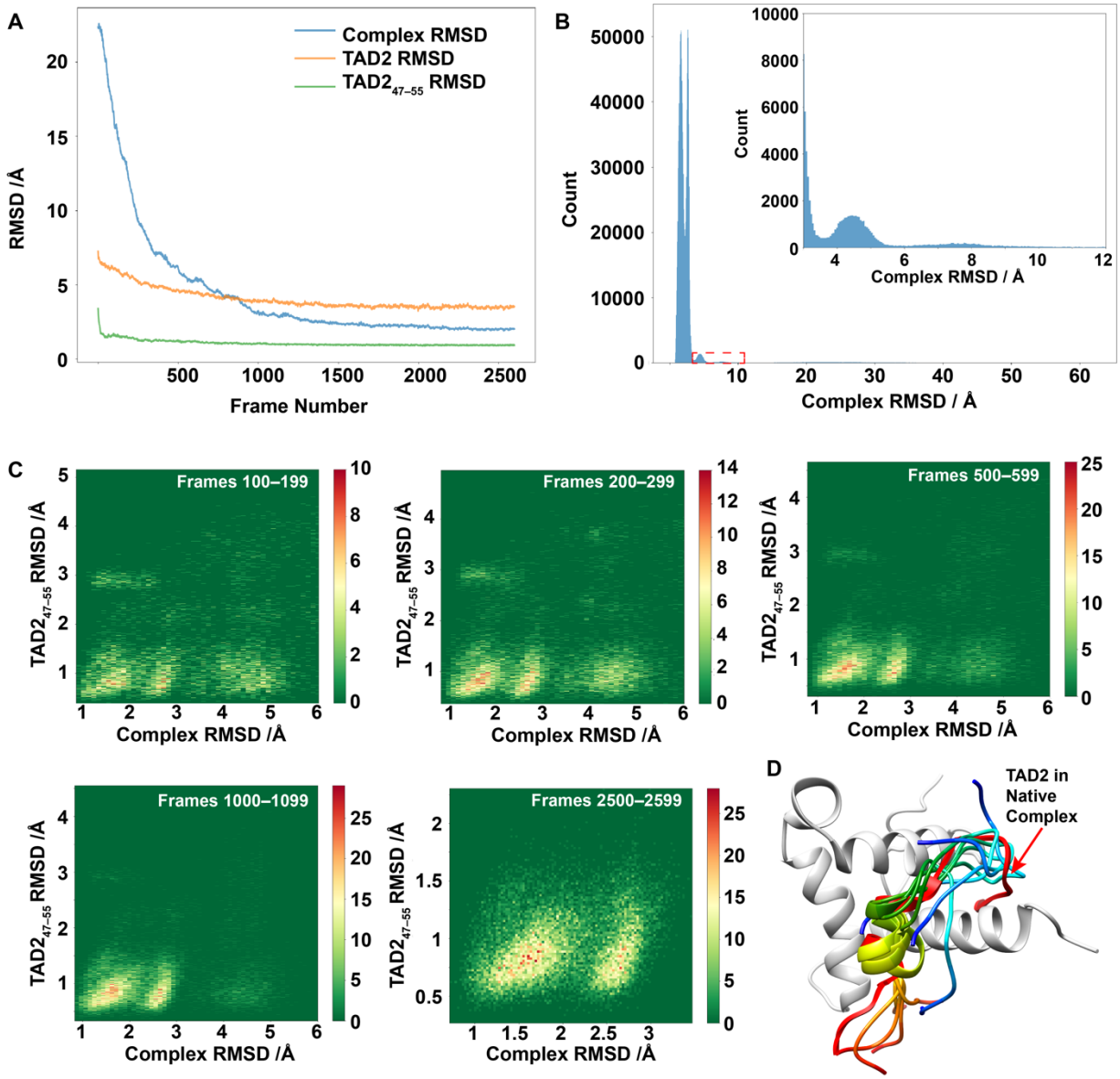

Figure 4. Root-mean-square deviation (RMSD) distribution of the Taz2-TAD2 complex and the TAD2 helix (residue 47–55). In panel **(A)**, the average RMSD of the complex, the entire TAD2, and the TAD2 helix over all 182 replicas decreases with simulations. The complex RMSD distribution is shown in panel **(B)**. Inset: RMSD distribution that is highlighted in the red dashed box. 2D-RMSD distributions **(C)** were calculated for different binding stages by extracting frames 100–199, 200–299, 500–599, 1000–1099, and 2500–2599. By aligning Taz2, the intermediate states (complex RMSD shown in panels **(B–C)** in the range of 3.5–5.5 Å) are overlaid with the native complex **(D)**, 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.

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 4C. 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 4C, 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 4D) 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 forming 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.

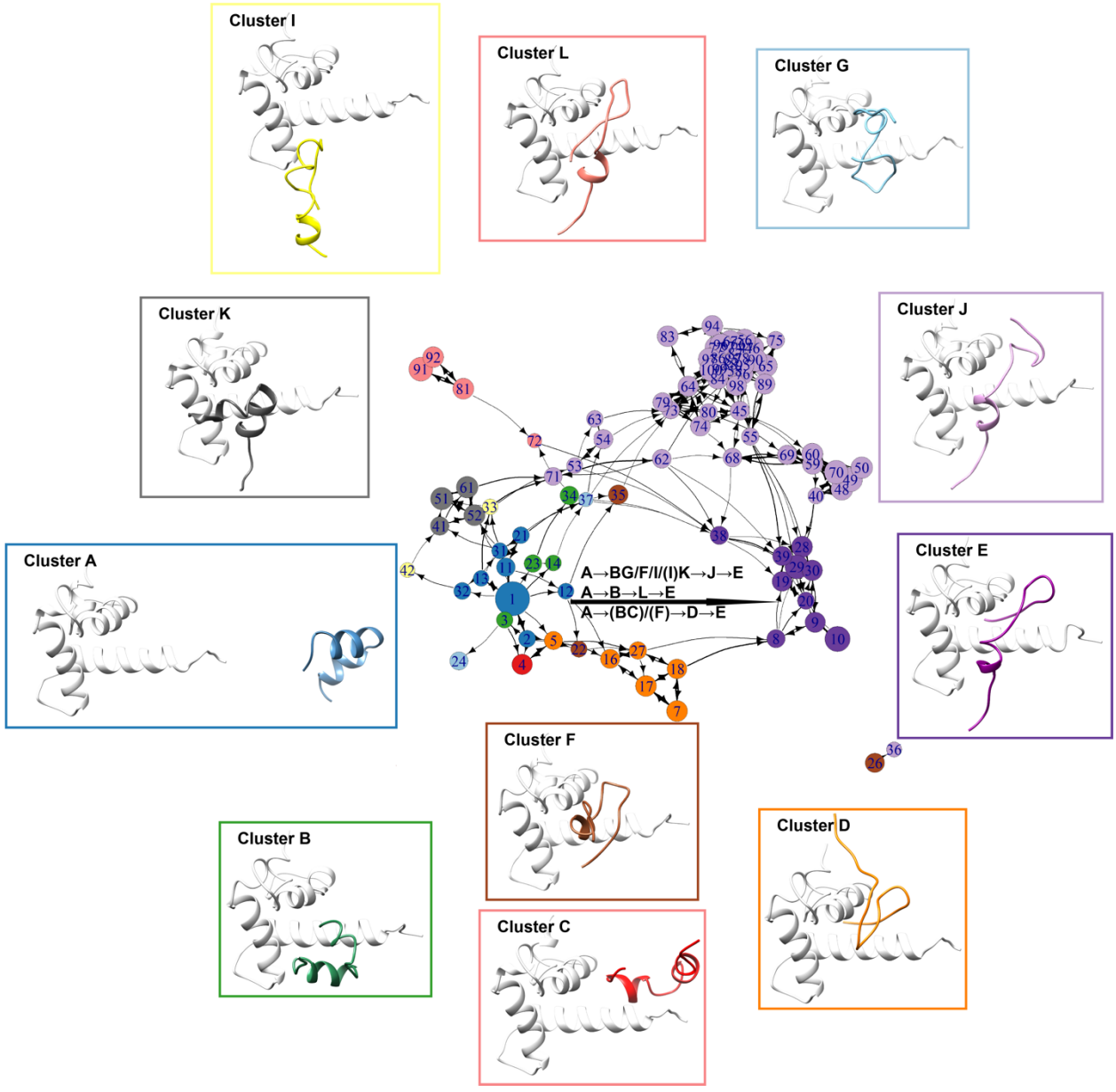

Figure 5. The graph representing the PathDetect-SOM transition matrix between neurons. Arrows represent structural switches between two neurons. Representative structures for each cluster are shown with borders and TAD2 colored according to the color of clusters in Figure S8A. Clusters A and E represent 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.

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 10x10 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, J–L), first binds at the small helix of TAD2 and then arranges of the extreme portion of the TAD2. Pathway 2, which passes from the neurons at the bottom of the graph (clusters C, D), first binds at the TAD2 N-terminus and then arranges of 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 similar structure 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 that a second type of pathway (pathway 2, lower part of the transition network diagram in Figure 5) 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 4C (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). In addition, the TAD2 helical fraction was monitored and mapped onto the SOM, as shown in Figure S8C,D. 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 the one hand, p53 TAD2 prefers to adopt a helical structure and on the other hand, 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 searches for the correct binding site to complete the entire binding process. Nevertheless, these pathways don't rule out other possible pathways, for example, pathways with rapid protein binding don't cover many clusters.

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 and the implication of the Gō-like model facilitates 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 native contact-based restraints didn't 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 the 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 α2/α3 loop). According to the fly-casting mechanism, the dynamic TAD2 and the flexible regions of Taz2 have large capture radii, accelerating mutual capture and protein collisions. 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 α2/α3 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 doesn't conflict with the initial contacts observed in Figure S4. In addition, based on the complex RMSD analysis 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 merely through the TAD2 N-terminus was also discovered.

The PathDetect-SOM results reveal multiple pathways from the unbound to the bound structures, suggesting that the association process is complicated. But 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 is 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.[50] This is consistent with the presence of the TAD2 helix in pathway 1 prior to the 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.

Computational Strategies for Protein Binding Mechanisms

The computational approaches used in this work can be applied to analyze other protein systems of interest and provide an atom/residue-level description of protein activities. By capturing key species and structural conversions in the association processes, we can visualize how proteins recognize each other. With specially tailored restraints, TAD2 is able to capture Taz2 with high flexibility and adjust its conformation accordingly during simulations, which is significant for studying the behavior of disordered proteins during protein recognition. To the best of our knowledge, this is the first time that PathDetect-SOM analysis has been used in protein association studies involving disordered proteins. Through the key species in simulations and the structural transition diagram identified by the SOM, we can briefly grasp the mechanism behind the extensive replicas.

Experimental techniques play a critical role in determining protein binding mechanisms and validating computational results. Protein binding processes can be studied using methods such as NMR[50], stopped-flow and time-resolved assays[51,52], Φ value analysis[53], among other methods[54]. For example, to confirm the role of the flexible regions of Taz2 in the encounter complex, the binding rates of the native Taz2 and its mutant counterpart can be monitored, in which Taz2 can be engineered by fixing the geometry of the Taz2 α2/α3 loop. To verify the order of contact formation, NMR titration methods can be used to track changes in 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.[54] 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 Gō-like coarse-grained UNRES simulations and studied the binding mechanism by analyzing the structural transition features over time and identifying pathways for structural transition between neurons. By imposing Gō-like restraints of intermolecular interactions, 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 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

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website.
The data underlying this article are available in OneDrive at https://tinyurl.com/4dtknt2p.

AUTHOR INFORMATION

Corresponding Author

Yi He - Department of Chemistry & Chemical Biology, The University of New Mexico, Albuquerque, New Mexico 87131, United States; https://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; https://orcid.org/0009-0001-9724-7418
Stefano Motta - Department of Earth and Environmental Sciences, University of Milano-Bicocca, Milan 20126, Italy; https://orcid.org/0000-0002-0812-6834

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.

ACKNOWLEDGMENT

This work was supported by the National Science Foundation (Grant No. 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 6, 197–208 (2005).
2. Handa, T., Kundu, D. & Dubey, V. K. Perspectives on evolutionary and functional importance of intrinsically disordered proteins. Int J Biol Macromol 224, 243–255 (2023).
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 114, 6589–6631 (2014).
4. 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 Science 22, 258–273 (2013).
5. Liu, J., Perumal, N. B., Oldfield, C. J., Su, E. W., Uversky, V. N. & Dunker, A. K. Intrinsic Disorder in Transcription Factors. Biochemistry 45, 6873–6888 (2006).
6. Hafner, A., Bulyk, M. L., Jambhekar, A. & Lahav, G. The multiple mechanisms that regulate p53 activity and cell fate. Nat Rev Mol Cell Biol 20, 199–210 (2019).
7. LM, M. J. L. M., Feng, H., Durell, S. R., Tagad, H. D. & Mazur, S. J. Characterization of the p300 Taz2-p53 TAD2 complex and comparison with the p300 Taz2-p53 TAD1 complex. Biochemistry 54, 2001–2010 (2015).
8. Okuda, M. & Nishimura, Y. Extended string binding mode of the phosphorylated transactivation domain of tumor suppressor p53. J Am Chem Soc 136, 14143–14152 (2014).
9. Bah, A. & Forman-Kay, J. D. Modulation of intrinsically disordered protein function by post-translational modifications. Journal of Biological Chemistry 291, 6696–6705 (2016).
10. Schöneberg, T. & Liebscher, I. Mutations in G Protein–Coupled Receptors: Mechanisms, Pathophysiology and Potential Therapeutic Approaches. Pharmacol Rev 73, 89 LP – 119 (2021).
11. 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 114, 6844–6879 (2014).
12. Koshland Jr, D. E., Némethy, G. & Filmer, D. Comparison of experimental binding data and theoretical models in proteins containing subunits. Biochemistry 5, 365–385 (1966).
13. Monod, J., Wyman, J. & Changeux, J. P. On the structure of allosteric transitions: a plausible model. J. Mol. Biol 12, 88–118 (1965).
14. Robustelli, P., Piana, S., Shaw, D. E. & Shaw, D. E. Mechanism of Coupled Folding-upon-Binding of an Intrinsically Disordered Protein. J Am Chem Soc 142, 11092–11101 (2020).
15. 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 106, 13737–13741 (2009).
16. Stevens, A. O., Luo, S. & He, Y. Three Binding Conformations of BIO124 in the Pocket of the PICK1 PDZ Domain. Cells 2022, Vol. 11, Page 2451 11, 2451 (2022).
17. 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 126, 7934–7942 (2022).
18. 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 Rep (2020). doi:10.1007/s41048-020-00107-w
19. 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 405, 298–314 (2011).
20. 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 Science 31, (2022).
21. Song, S., Li, T., O Stevens, A., Raad, T. & He, Y. Investigating the mechanical properties and flexibility of N-BAR domains in PICK1 by molecular dynamics simulations. Curr Protein Pept Sci 24, (2023).
22. 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 2040007, 2040007 (2020).
23. 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 2, (2022).
24. 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 110, 14936–14941 (2013).
25. He, Y., Chen, C. & Xiao, Y. United-residue (UNRES) Langevin dynamics simulations of trpzip2 folding. J Comput Biol 16, 1719–1730 (2009).
26. 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 30, 2127–2135 (2009).
27. Kmiecik, S., Gront, D., Kolinski, M., Wieteska, L., Dawid, A. E. & Kolinski, A. Coarse-Grained Protein Models and Their Applications. Chem Rev 116, 7898–7936 (2016).
28. Karanicolas, J. & Brooks, C. L. Improved Gō-like Models Demonstrate the Robustness of Protein Folding Mechanisms Towards Non-native Interactions. J Mol Biol 334, 309–325 (2003).
29. Naganathan, A. N. & Orozco, M. The Native Ensemble and Folding of a Protein Molten-Globule: Functional Consequence of Downhill Folding. J Am Chem Soc 133, 12154–12161 (2011).
30. 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 346, 1121–1145 (2005).
31. 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 10, 2224–2231 (2014).
32. Chong, S.-H., Im, H. & Ham, S. Explicit Characterization of the Free Energy Landscape of pKID–

KIX Coupled Folding and Binding. ACS Cent Sci 5, 1342–1351 (2019).

33. Huang, Y. & Liu, Z. Nonnative Interactions in Coupled Folding and Binding Processes of Intrinsically Disordered Proteins. PLoS One 5, e15375 (2010).

34. Chen, T., Song, J. & Chan, H. S. Theoretical perspectives on nonnative interactions and intrinsic disorder in protein folding and binding. Curr Opin Struct Biol 30, 32–42 (2015).

35. 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 127, 2177–2186 (2023).

36. Krupa, P., Mozolewska, M. A., Wiśniewska, M., Yin, Y., He, Y., Sieradzan, A. K., Ganzynkowicz, R., Lipska, A. G., Karczyńska, A., Slusarz, M., Slusarz, R., Giełdoń, A., Czaplewski, C., Jagieła, D., Zaborowski, B., Scheraga, H. A. & Liwo, A. Performance of protein-structure predictions with the physics-based UNRES force field in CASP11. Bioinformatics 32, 3270–3278 (2016).

37. Liwo, A., He, Y. & Scheraga, H. A. Coarse-grained force field: general folding theory. Phys Chem Chem Phys 13, 16890–16901 (2011).

38. Liwo, A., Baranowski, M., Czaplewski, C., Gołaś, E., He, Y., Jagieła, D., Krupa, P., Maciejczyk, M., Makowski, M., Mozolewska, M. A., Niadzvedtski, A., Ołdziej, S., Scheraga, H. A., Sieradzan, A. K., Ślusarz, R., Wirecki, T., Yin, Y. & Zaborowski, B. A unified coarse-grained model of biological macromolecules based on mean-field multipole-multipole interactions. J Mol Model 20, (2014).

39. Grant, B. J., Skjærven, L. & Yao, X. Q. The Bio3D packages for structural bioinformatics. Protein Science 30, 20–30 (2021).

40. Kaski, S., Kangas, J. & Kohonen, T. Bibliography of self-organizing map (SOM) papers: 1981–1997. Neural computing surveys 1, 1–176 (1998).

41. Miljković, D. Brief review of self-organizing maps. in 2017 40th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO) 1061–1066 (IEEE, 2017).

42. 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 13, 642–651 (2013).

43. 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 31, 1490–1492 (2015).

44. 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 15, 7909–7919 (2023).

45. Rotkiewicz, P. & Skolnick, J. Fast procedure for reconstruction of full-atom protein models from reduced representations. J Comput Chem 29, 1460–1465 (2008).

46. 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 17, 2080–2089 (2021).

47. 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 18, 1957–1968 (2022).

48. Csardi, G. & Nepusz, T. The igraph software package for complex network research. InterJournal, complex systems 1695, 1–9 (2006).

49. Shoemaker, B. A., Portman, J. J. & Wolynes, P. G. Speeding molecular recognition by using the folding funnel: The fly-casting mechanism. Proc Natl Acad Sci U S A 97, 8868–8873 (2000).

50. 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 134, 3792–3803 (2012).

51. 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. Proceedings of the National Academy of Sciences 96, 465–470 (1999).

52. Lento, C. & Wilson, D. J. Subsecond Time-Resolved Mass Spectrometry in Dynamic Structural Biology. Chem Rev 122, 7624–7646 (2022).

53. Dogan, J., Mu, X., Engstrom, A. & Jemth, P. The transition state structure for coupled binding and folding of disordered protein domains. Sci Rep 3, 3–8 (2013).

54. 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. Proceedings of the National Academy of Sciences 113, 7171–7176 (2016).

SYNOPSIS TOC