JACS Au
AN OPEN ACCESS JOURNAL OF THE AMERICAN CHEMICAL SOCIETY

pubs.acs.org/jacsau

Article

Investigating the Conformational Dynamics of a Y-Family DNA Polymerase during Its Folding and Binding to DNA and a Nucleotide

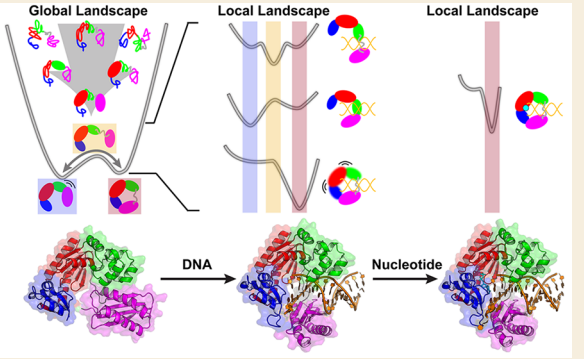
Xiakun Chu, Zucai Suo, and Jin Wang*

Cite This: *JACS Au* 2022, 2, 341−356

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: During DNA polymerization, the Y-family DNA polymerases are capable of bypassing various DNA damage, which can stall the replication fork progression. It has been well acknowledged that the structures of the Y-family DNA polymerases have been naturally evolved to undertake this vital task. However, the mechanisms of how these proteins utilize their unique structural and conformational dynamical features to perform the translesion DNA synthesis are less understood. Here, we developed structure-based models to study the precatalytic DNA polymerization process, including DNA and nucleotide binding to DPO4, a paradigmatic Y-family polymerase from *Sulfolobus solfataricus*. We studied the interplay between the folding and the conformational dynamics of DPO4 and found that DPO4 undergoes first unraveling (unfolding) and then folding for accomplishing the functional "open-to-closed" conformational transition. DNA binding dynamically modulates the conformational equilibrium in DPO4 during the stepwise binding through different types of interactions, leading to different conformational distributions of DPO4 at different DNA binding stages. We observed that nucleotide binding induces modulation of a few contacts surrounding the active site of the DPO4−DNA complex associated with a high free energy barrier. Our simulation results resonate with the experimental evidence that the conformational change at the active site led by nucleotide is the rate-limiting step of nucleotide incorporation. In combination with localized frustration analyses, we underlined the importance of DPO4 conformational dynamics and fluctuations in facilitating DNA and nucleotide binding. Our findings offer mechanistic insights into the processes of DPO4 conformational dynamics associated with the substrate binding and contribute to the understanding of the "structure−dynamics−function" relationship in the Y-family DNA polymerases.



KEYWORDS: *Conformational dynamics, Localized frustration, Energy landscape, DNA polymerization, Protein−DNA recognition, Nucleotide binding*

DNA replication, an essential process occurring in all living organisms, is finely tuned by the DNA polymerases. During *in vivo* DNA polymerization, these protein machines may frequently encounter lesions in the DNA template, which can potentially block the normal progression of replication forks. To resolve this critical issue, the Y-family DNA polymerases can perform the translesion synthesis and bypass the DNA lesions.[1,2] Meanwhile, the Y-family DNA polymerases catalyze DNA synthesis with low catalytic efficiency, low processivity, and low fidelity with both undamaged and damaged DNA,[2] compared to the DNA polymerases in the A- and B-families.[3] Structural analysis revealed that the Y-family DNA polymerases have conserved architectures,[4−8] which are different from those of the high-fidelity replicative DNA polymerases. In light of the "structure−function" paradigm, it has been acknowledged that the function of the Y-family DNA polymerase is characterized by its unique structure.[9,10] Nevertheless, a clear picture of how the Y-family DNA polymerases regulate translesion DNA synthesis through the structure and associated conformational dynamics is still not present.

As a paradigmatic Y-family DNA polymerase, DNA polymerase IV (DPO4) from *Sulfolobus solfataricus* has a conserved polymerase core composed of a finger (F), a palm (P), and a thumb (T) domain, as well as a C-terminal little finger (LF) domain tethered to the T domain through a flexible linker.[4] Prior to the catalytic process, there are two essential substrate binding processes, including DPO4 binding to DNA and subsequently recruiting a nucleotide to the active site in the DPO4−DNA complex. Crystal structures of DPO4 in the apo state, DPO4−DNA binary, and DPO4−DNA−

Received: August 20, 2021
Published: December 16, 2021

https://doi.org/10.1021/jacsau.1c00368
JACS Au 2022, 2, 341−356

nucleotide ternary complexes revealed a global conformational change in DPO4 occurring during DNA binding through the relocation of the LF domain relative to the polymerase core and the slightly changed global conformation of DPO4 retaining between the binary and the ternary forms.[4,11] The large-scale "open-to-closed" DPO4 conformational transition induced by DNA has been found to result in a dynamical protein−DNA recognition process that may contribute to the low-fidelity DNA synthesis.[12] Besides, a slight local structural adaption in the F domain of DPO4 was identified to stabilize the bound incoming nucleotide.[11] However, the dynamical and full picture of the structural rearrangements of DPO4 from the apo state, then to the DNA binding binary complex, and finally to the nucleotide binding ternary complex remains largely elusive.

Akin to the other Y-family DNA polymerases, DPO4 is a typical multidomain protein. DPO4 has been observed to undergo stepwise unfolding with the intermediate observed in both the experiment[13] and the simulations.[14−16] The unfolding intermediate, which shows an extended linker and unstable LF domain interfaces with well-folded individual domains in DPO4, was further hypothesized to benefit the formation of multiple DPO4 conformations during its binding to DNA or proliferating cell nuclear antigen (PCNA).[13] This fact indicates a positive role of the DPO4 unfolding in facilitating the functional binding processes. Recent theoretical work found that the weakly formed domain interfaces in DPO4 are the key to realizing the high efficiency of folding and DNA binding, simultaneously.[16] Currently, the interplay between the global (un)folding and domain spatial rearrangements in DPO4, in particular how the (un)folding affects the functional conformational dynamics in DPO4, is still in need of a quantitative investigation.

DPO4−DNA binding was characterized to be a complex process that shows multistep characteristics associated with dynamically arranging the DPO4 conformational distribution.[17−22] This picture of DPO4 binding to DNA with conformational fluctuations may help the intricate regulation of DPO4 binding to the replication forks during the translesion synthesis through coordinating the movements of the LF domain, which can contribute to the polymerase switching between DPO4 and a replicative DNA polymerase.[23] Currently, it is still not clear how the conformational transition in DPO4 and the DPO4−DNA interactions evolve during DNA binding. After DNA binding, an incoming nucleotide binds to the DPO4−DNA binary complex to form the precatalytic ternary complex. The conformational rearrangements of the active site in the DPO4−DNA complex induced by nucleotide binding have been considered to be the rate-limiting step of the whole enzymatic process.[17,24,25] The results based on the stopped-flow Förster resonance energy transfer (FRET) study suggested that the F domain motions should account for the slow conformational rearrangement in DPO4 during nucleotide incorporation, but the process is less understood at the single-molecular level and the underlying mechanism remains unclear.[19]

Here, we addressed the DPO4 conformational dynamics at the precatalytic steps, i.e., an initial DNA binding followed by a nucleotide binding to DPO4. Technically, we developed structure-based models (SBMs) to study the folding, conformational transition, and substrate binding of DPO4. Motivated by the recent experimental evidence that DPO4 is largely in the apo structure and able to adopt the DNA-binding structure with a minor population,[11,26] we extended the single-basin SBM, which was used in our previous studies,[14−16] to the double-well one that incorporates the structural information from the apo and ternary DPO4 forms. The model not only generated the experimentally consistent simulation results[13,26] but also enabled us to simultaneously study the global folding and local conformational dynamics of DPO4. We uncovered that the "open-to-closed" functional conformational transition in DPO4 occurs at the bottom of the folding energy landscape and pre-exists in the absence of DNA. The binding of DPO4 to DNA undergoes multiple steps associated with the different conformational distributions of DPO4 that are determined by different interactions. Furthermore, we found that there is a high free energy barrier during nucleotide binding. Careful examinations show subtle destabilization in the interactions surrounding the active site of the DPO4−DNA complex during nucleotide binding and give a hint on the origin of the binding free energy barrier. By performing the localized frustration analyses, we found that the DPO4 conformational dynamics induced by substrate binding are closely related to the highly frustrated interactions present in the native structures. Our theoretical work provides mechanistic insights into the rate-limiting, prechemistry step of the DPO4 catalyzed reaction and helps the understanding of translesion DNA synthesis by the Y-family polymerases.

RESULTS

Global Folding and Local Conformational Dynamics of DPO4

We built a double-well two-bead SBM to study DPO4 folding and conformational dynamics. Each residue, except glycine, was modeled as two beads, representing the backbone and side chain, respectively. In our previous study,[16] we found that the one-bead homogeneous SBM may overweight the contribution of the interdomain interactions in the total energy in the native structure of DPO4. We further suggested that weakening the strengths of interdomain interactions in the SBM can optimize the folding and DNA binding of DPO4. Here, we found that the two-bead homogeneous SBM can naturally lead to a decreased proportion of interdomain interactions in the total energy for stabilizing the native structures with respect to the one-bead homogeneous SBM, possibly due to the fact that an improved native contact map was used with the presence of the side chain in the two-bead coarse-grained model (see Materials and Methods). In addition, considering the highly charged property of DPO4 as a DNA binding protein, we further included the electrostatic interactions described by the Debye−Hückel model and placed the charges onto the side chain of the designated residues (one positive charge for arginine and lysine, one negative charge for aspartic and glutamic acid). The model takes into account the effects of salt concentration through the Debye screening length. Unless otherwise specified, we used the salt concentration of 0.05 M throughout the simulations in accordance with previous experiments for DPO4 folding and substrate binding.[13,24,25,28] The double-well model is realized by a mixture of native contact maps of DPO4 in the apo DPO4 structure ($DPO4_A$)[11] and ternary DPO4−DNA−nucleotide structure ($DPO4_T$)[4] and aims to produce two basins at the energy landscape in order to describe the "open-to-closed" conformational transition of DPO4 (see Supporting Information). Crystallographic structural analysis revealed that the major differences

https://doi.org/10.1021/jacsau.1c00368
JACS Au 2022, 2, 341−356

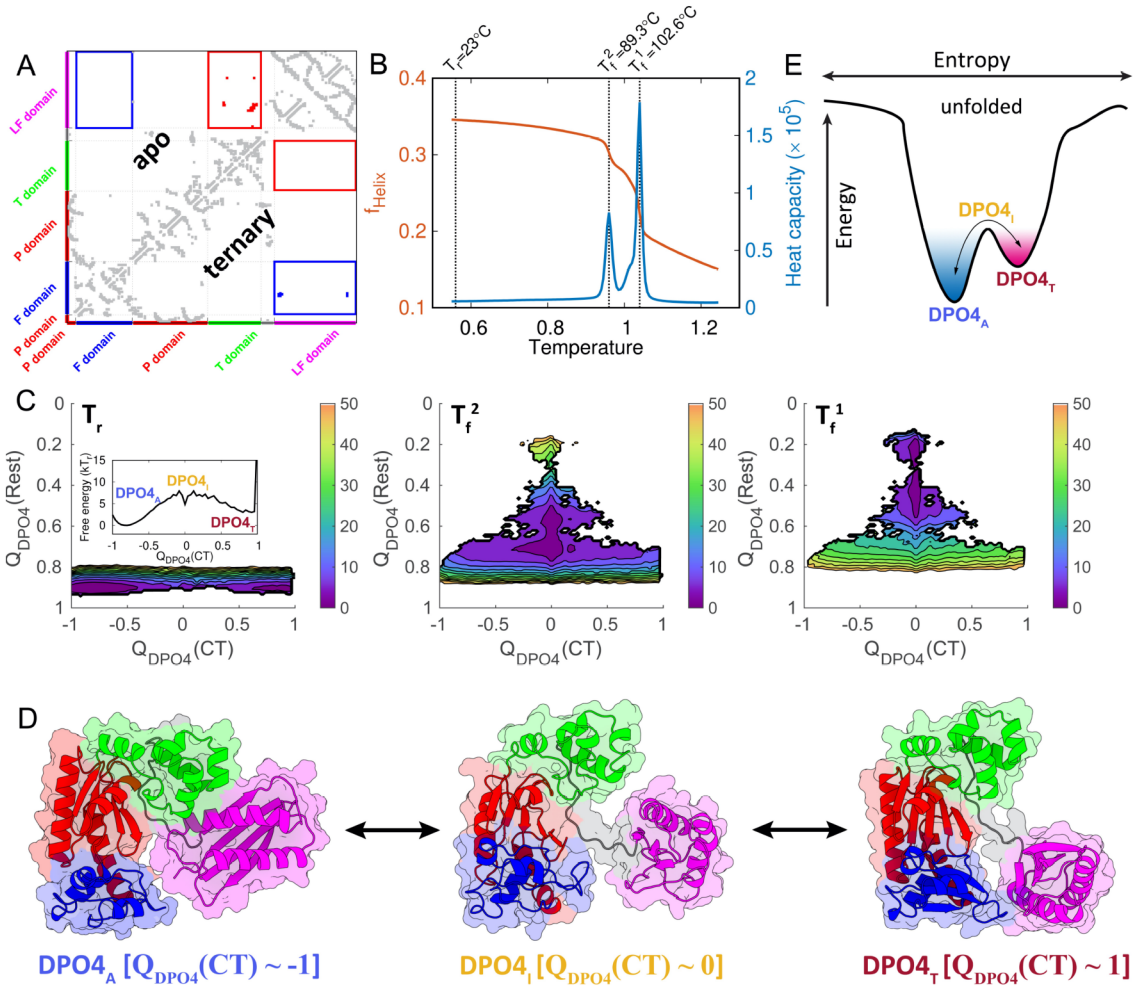


Figure 1. Folding and conformational dynamics of DPO4. (A) The contact map of DPO4 at the apo structure (PDB: 2RDI[11]) (top left) and the ternary DPO4−DNA−nucleotide structure (PDB: 1JX4[4]) (bottom right). The domains in DPO4 are finger domain (F domain, residues 11−70, blue), palm domain (P domain, residues 1−10 and 71−166, red), thumb domain (T domain, residues 167−229, green) and little finger domain (LF domain, residues 245−341, magenta). The flexible linker (residues 230−244) that tethers the T and LF domains is colored gray. The red and blue rectangles indicate the major change of the contacts in DPO4 between the apo and the ternary structures, corresponding to the contacts formed at the T−LF and F−LF interfaces, respectively. (B) Proportion of helical formation in DPO4 and heat capacity curve along with the temperature. Due to the lack of Ramachandran angles in our coarse-grained model, we defined the formation of a helical segment as the one that has at least three continuous dihedrals within the range of −35°−145°.[27] The experimental temperatures are approximately mapped to the simulation temperatures with the knowledge of folding temperatures and further using a linear relation.[13] (C) Free energy landscapes of DPO4 at the room temperature T_r (left), the first folding transition temperature T_f^1 (right), and the second folding transition temperature T_f^2 (middle). The free energy profiles are projected onto $Q_{DPO4}(Rest)$ and $Q_{DPO4}(CT)$. $Q_{DPO4}(CT) = Q_{DPO4}(F-LF) - Q_{DPO4}(T-LF)$, where $Q_{DPO4}(T-LF)$ is the fraction of the interdomain native contacts between the T and the LF domains in the apo structure and $Q_{DPO4}(F-LF)$ is the fraction of the interdomain native contacts between the F and the LF domains in the ternary structure. $Q_{DPO4}(Rest)$ is the fraction of the native contacts in DPO4, excluding the ones at the T−LF and F−LF domain interfaces. The free energy is in the unit of kT. T is the corresponding temperature where the free energy was calculated. (D) Structural illustrations of DPO4 at the apo ($DPO4_A$), intermediate ($DPO4_I$), and ternary states ($DPO4_T$). The domains in DPO4 have the same color schemes as the ones at the axes in (A). (E) Scheme illustrating the energy landscape of DPO4 folding and conformational dynamics.

of DPO4 between the apo and the ternary structures are attributed to the spatial position and rotation of the LF domain that forms interactions with the T domain in the $DPO4_A$ and the F domain in the $DPO4_T$, respectively (Figure 1A). Meanwhile, the other segments of DPO4, including the individual domain structures and domain−domain interfaces, remain largely the same. Therefore, the conformational transition of DPO4 between the $DPO4_A$ and the $DPO4_T$ corresponds to the rearrangements of the interfacial LF domain contacts.

We performed Replica-Exchange Molecular Dynamics (REMD) simulations to explore DPO4's folding and conformational dynamics.[29] With the Weighted Histogram Analysis Method (WHAM),[30] we investigated the thermodynamics of DPO4 folding, including the heat capacity curve and the melting curve (Figure 1B). We observed an apparent two-step DPO4 folding process exhibiting two folding temperatures (the low melting temperature T_f^2 = 0.96 and the high melting temperature T_f^1 = 1.04 in Figure 1B; the temperature is in reduced units), consistent with the observations in the experiments, which identified two melting temperatures of

https://doi.org/10.1021/jacsau.1c00368
JACS Au 2022, 2, 341−356

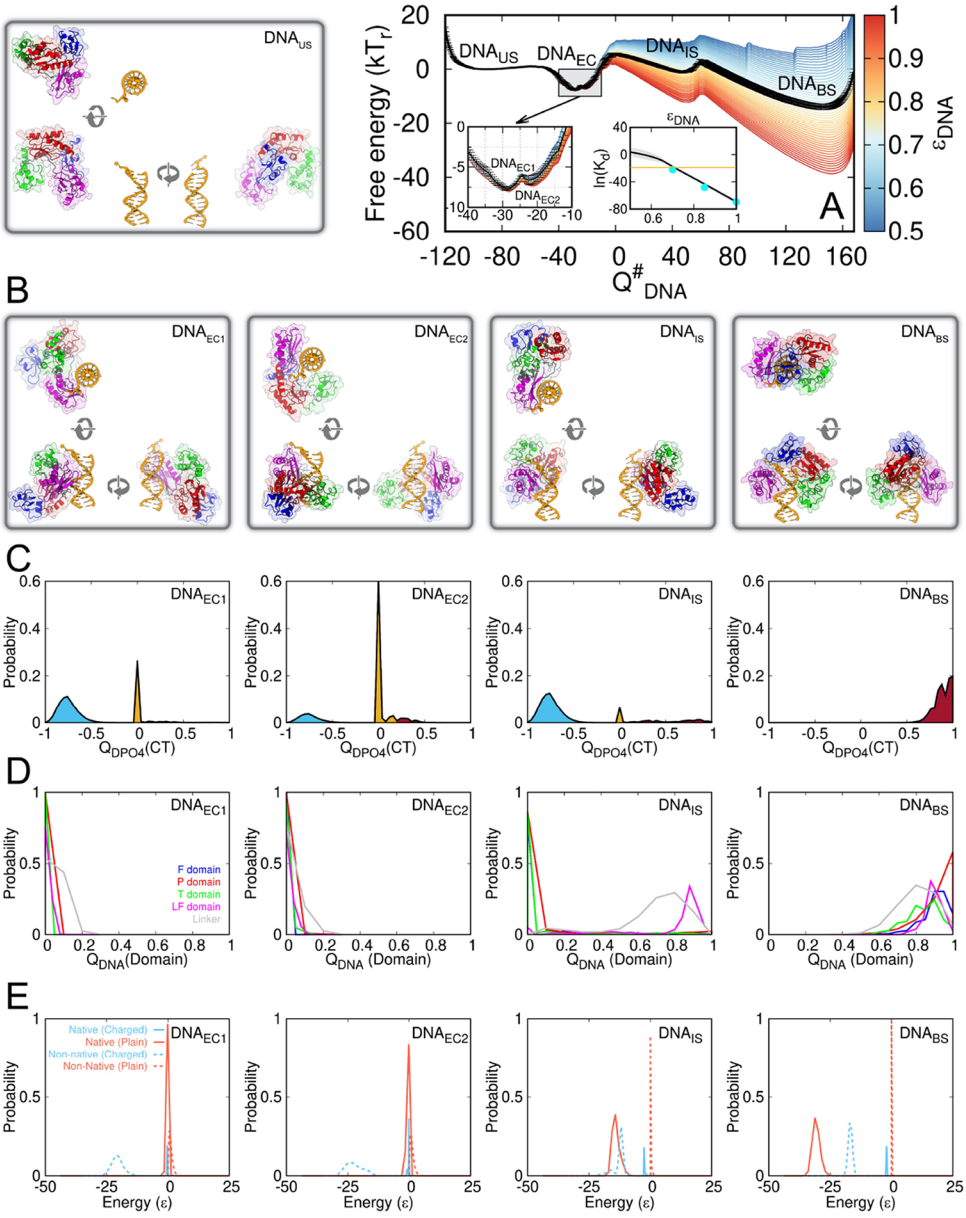


Figure 2. DPO4−DNA binding. (A) Free energy landscapes of DPO4 binding to DNA projected onto the binding reaction coordinate $Q^{\#}_{DNA}$ at different ϵ_{DNA}, where $Q^{\#}_{DNA} = N_{DNA}Q_{DNA} - dRMS_{DNA}$, Q_{DNA} is the fraction of DPO4−DNA interchain native contacts, N_{DNA} is the number of DPO4−DNA interchain native contacts, and $dRMS_{DNA}$ is the difference of the distance of native contact pairs between DPO4 and DNA with deviation from 0 indicating deviation from the native structure.[35,36] $dRMS_{DNA}$ is in the unit of Å. ϵ_{DNA} is the strength of the native contacts between DPO4 and DNA. The free energy landscapes show 4 minima, which are denoted as "Unbinding State (DNA_{US})", "Encounter Complex (DNA_{EC})", "Intermediate State (DNA_{IS})", and "Bound State (DNA_{BS})". Inset plots are the zoom-in free energy landscapes at the region of the DNA_{EC} state (left) and binding affinity (K_d) along with ϵ_{DNA} (right). The free energy landscapes and K_d at different ϵ_{DNA} were calculated from reweighting the thermodynamics at ϵ_{DNA} = 1.0 (see Supporting Information). The black line in the free energy plot (ϵ_{DNA} = 0.70), which matches with the experimental K_d (3−10 nM),[24,28] was obtained from the direct umbrella sampling simulations. In the zoom-in free energy landscape plot, the DNA_{EC} state is further divided into the DNA_{EC1} and DNA_{EC2} states, separated by a minor free energy barrier. In the K_d plot, the gray shadow region corresponds to the standard error of the mean value (black line) and the yellow line indicates the experimental affinity. The cyan points in the K_d plot are the results from the direct umbrella sampling simulations. (B) Typical DPO4−DNA structures in the DNA_{US}, DNA_{EC1}, DNA_{EC2}, DNA_{IS}, and DNA_{BS} states extracted from the simulations. The structure is shown in three different views for each binding state. (C) Conformational dynamics of DPO4 at each binding state shown by the probability distribution along with $Q_{DPO4}(CT)$. (D) Probability distribution of the fraction of native contacts formed by the individual domains and the linker in DPO4 with DNA. (E) Probability distribution of the interaction energy between DPO4 and DNA.

DPO4 at 89.3 and 102.6 °C.[13] It is worth noting that our previous simulations with single-basin one-bead SBMs generated only one peak on the heat capacity curve and resulted in sigmoidal-like melting curves.[14−16] The results suggest that the presence of the side chain bead and electrostatics in the SBM is critical to recapture the global folding behaviors of DPO4.[31] Due to the simplified interactions and the coarse-grained nature in the SBM, the

https://doi.org/10.1021/jacsau.1c00368
JACS Au 2022, 2, 341−356

simulation temperatures cannot directly correspond to the experimental ones. In this regard, we assumed a linear temperature dependence on the energy and provided an approximate connection to bridge the simulation temperatures and the experimental ones with the knowledge of the folding temperatures (see Supporting Information).

We quantified the free energy landscapes of DPO4 onto the fraction of native contacts of folding ($Q_{DPO4}(Rest)$) and conformational transition between the $DPO4_A$ and $DPO4_T$ ($Q_{DPO4}(CT)$) at room temperature T_r and the two folding temperatures T_f^2 and T_f^1 (Figure 1C). $Q_{DPO4}(CT)$ is the subtraction of the fraction of the interdomain native contacts between the T and the LF domains in the apo structure ($Q_{DPO4}(T - LF)$) and the fraction of interdomain native contacts between the F and LF domains in the ternary structure ($Q_{DPO4}(F - LF)$), so DPO4 in the apo and ternary structure has $Q_{DPO4}(CT)$ equal to −1 and 1, respectively. In order to see whether our model can successfully capture the structures of the $DPO4_A$ and $DPO4_T$, we also quantified the free energy landscapes of DPO4 projected onto the root-mean-square deviation (RMSD) toward the apo ($RMSD_A$) and ternary structure ($RMSD_T$) of DPO4 (Figure S7).

At room temperature, we can see that when $RMSD_A$ and $RMSD_T$ are small close to 0, $Q_{DPO4}(CT)$ values approach −1 and 1, respectively. This suggests formations of the $DPO4_A$ and $DPO4_T$ structures at room temperature, and $Q_{DPO4}(CT)$ is capable of describing the transitions between the $DPO4_A$ and the $DPO4_T$ forms. The conformational dynamics of DPO4 at room temperature is limited and entirely attributed to the transition between the $DPO4_A$ and $DPO4_T$ (Figure 1C). Besides, there is an intermediate state of DPO4 ($DPO4_I$) formed during the transition between the $DPO4_A$ and the $DPO4_T$ (Figure 1D). The intermediate state $DPO4_I$, at which the LF domain in DPO4 shows no interactions with either the F or T domain and the other regions of DPO4 are well folded, is an inevitable on-pathway intermediate state. In other words, DPO4 at the $DPO4_I$ exhibits an extended flexible linker and serves as the bridge to connect the structurally distinct $DPO4_A$ and $DPO4_T$. The observation of $DPO4_I$ here is consistent with the melting experiment,[13] resonating with the fact that the flexible linker is the key to realize the DPO4 substrate binding through the conformational dynamics of DPO4.[12] With increasing the temperature to the low melting temperature T_f^2, the $DPO4_I$ state becomes more populated than the $DPO4_A$ and $DPO4_T$ states, indicating that the $DPO4_I$ state is entropically favored. Structural analysis on the free energy minimum at $Q_{DPO4}(Rest) \sim 0.7$ shows that the LF domain interfaces are entirely broken while other regions in DPO4 remain folded (Figure S8). Moreover, a new free energy minimum emerges on the landscape at $Q_{DPO4}(CT) \sim 0$ and $Q_{DPO4}(Rest) \sim 0.6$, signifying an intermediate state for DPO4 (un)folding. We found that the LF domain at the intermediate state is unfolded associated with a fully folded polymerase core (Figure S8). Continuously increasing the temperature to the high dominant melting temperature T_f^1 results in two new minima on the free energy landscape, corresponding to an additional folding intermediate ($Q_{DPO4}(CT) \sim 0$ and $Q_{DPO4}(Rest) \sim 0.4$) and the unfolded state ($Q_{DPO4}(CT) \sim 0$ and $Q_{DPO4}(Rest) \sim 0.2$), respectively. The structural analysis on the intermediate state at T_f^1 shows that DPO4 has the folded F and P domains and the formed F−P domain interface (Figure S8). The multiple intermediate unfolding states of DPO4 were observed in our previous studies as a result of "divided-and-conquer" domain-wise folding.[14,16]

Free energy profiles show that $Q_{DPO4}(CT)$ can reach values higher than 0.5 or lower than −0.5 only when $Q_{DPO4}(Rest)$ is higher than 0.7. This indicates that the interfacial LF domain interactions, which are responsible for the functional conformational dynamics of DPO4, can only be formed when the other regions of DPO4 have accomplished folding. It suggests the vulnerable structural characteristics of the LF domain interfaces in DPO4 are responsible for the functional purpose. We also found that modulating the model parameters related to the strengths of the LF domain interacting with the T and F domains has minor effects on changing the global folding temperatures (Figure S9). This indicates the functional conformational dynamics of DPO4 has a minimal impact on the global folding process. From the energy landscape perspective (Figure 1E), the apo and ternary states of DPO4 represent two energy basins that are globally located at the bottom of the energy landscapes. In other words, the structures and interactions within DPO4 that underpin its function can only be formed at the late stages of folding, and DPO4 transforms from an inactive ($DPO4_A$) state to an active ($DPO4_T$) state through the local unfolding, which has been largely regarded as an effect of the frustration on the energy landscape in favor of the protein conformational dynamics.[32−34]

Conformation-, Interaction- and Salt-Dependent Multistep DPO4−DNA Binding

On the basis of the double-well SBM of DPO4, we further studied DPO4−DNA binding (in the absence of nucleotide). The DNA binding model includes the interactions of the DPO4−DNA native contacts derived from the ternary crystal structure of DPO4−DNA−nucleotide[4] and electrostatic interchain interactions between DPO4 and DNA.[37,38] To achieve sufficient sampling, we performed umbrella sampling simulations for the DPO4−DNA binding process. We implemented the biasing potentials along with the binding reaction coordinate $Q_{DNA}^{\#}$. $Q_{DNA}^{\#} = N_{DNA}Q_{DNA} - dRMS_{DNA}$, containing both the information on the fraction of the interchain native contacts Q_{DNA} and the Euclidean distance of the interchain native contacts to the bound structure $dRMS_{DNA}$ (N_{DNA} is the number of interchain native contacts). A high (low) value of $Q_{DNA}^{\#}$ corresponds to a high (low) degree of native similarity to the DNA binding in the native structure. It has been recognized that Q_{DNA} and $dRMS_{DNA}$ are good at describing the process after the native contacts start to establish and the unbound states with no native contacts formed, respectively.[39] Thus, $Q_{DNA}^{\#}$ can provide a more precise description at both unbound states and binding states, compared to our previous studies.[16,22] We further calibrated our model to the binding affinity by modulating the strengths of the DPO4−DNA interchain native contacts. From the quantified binding free energy landscape (Figure 2A), we identified four free energy minima that separate the DNA-binding process into three stages: from the completely dissociative unbound state (DNA_{US}) to the initially anchoring encounter complex (DNA_{EC}), then to the partially bound intermediate state (DNA_{IS}), and finally to the fully bound state (DNA_{BS}). The multistep DNA binding picture obtained here is consistent with experimental observations.[17−21] Further analysis shows the DNA_{EC} is made up of two metastable states on the free energy landscape (DNA_{EC1} and DNA_{EC2}).

https://doi.org/10.1021/jacsau.1c00368
JACS Au 2022, 2, 341−356

We note that the DNA_{EC} state was not able to be detected as a metastable state in our previous study,[16] where the one-bead SBM without calibrations to the experiments was used along with $dRMS_{DNA}$ as the reaction coordinate. Here, the careful determination on the DNA_{EC} enabled us to characterize the conformational distribution of DPO4 and further dissect the conformational dynamics mechanism of DPO4 during DNA binding.

We found that the conformational dynamics of DPO4 is modulated by DNA during binding (Figure 2B,C and Figure S14). When DPO4 is in the DNA_{EC} stage, it exhibits remarkable population in the $DPO4_I$ form ($Q_{DPO4}(CT) \sim 0.0$), in particular in the DNA_{EC2}. Bearing in mind that DPO4 is largely in the $DPO4_A$ form when isolated, DPO4 has significantly shifted its conformational equilibrium toward the $DPO4_I$ form in forming the DNA_{EC}. Native contact and interaction energy analyses revealed a negligible amount of native contacts (interactions) between DPO4 and DNA formed in both the DNA_{EC1} and the DNA_{EC2}, where the interchain interactions are purely non-native electrostatic (Figure 2D,E). These features suggest that the formation of the DNA_{EC} state is nonspecific and driven by the non-native electrostatic interactions. The formations of the DNA_{EC1} and DNA_{EC2} from the DNA_{US} can contribute to the "facilitated diffusion" in the protein−DNA recognition process by reducing the dimensionality of the searching space from 3D to 1D.[40−43] The non-native electrostatic interactions play the key role in forming these two states, where DPO4 is populated in the $DPO4_I$ form associated with the extended and flexible linker. Since the linker region is positively charged, we further removed the positive charges in the linker region and performed additional DPO4−DNA binding simulations to examine the effects of the charges in the linker on the DNA binding process. We found that overall the free energy landscape after DPO4 initializing DNA binding ($Q^{\#}_{DNA} > -60$) was elevated from the original one (Figure S15). This implies that the charged interactions between the linker in DPO4 and DNA can increase the stability of the DPO4−DNA binding states. A significant decrease was observed in the barrier height of the transition from the DNA_{EC} to the DNA_{US} state after removing the positive charges in the linker region. The results indicate that the flexible and extended, positively charged linker in DPO4 prevents the dissociation between DPO4 and DNA, thus in favor of the "facilitated diffusion".

In the DNA_{IS}, the LF domain and the linker region in DPO4 accomplish DNA binding, and DPO4 is largely in the $DPO4_A$ form. It indicates that the transition from the nonspecific DNA_{EC} to the partially specific DNA_{IS} involves the modulation of DPO4 conformational dynamics from the $DPO4_I$ to the $DPO4_A$ form coupled with DNA binding. In the last stage, the DPO4−DNA binding and DPO4 conformational transition to the $DPO4_T$ form were found to be strongly coupled. Overall, the stepwise DPO4−DNA binding with the nonmonotonic adaptation of DPO4 conformational dynamics underlines the complexity of the DPO4−DNA binding process.

To investigate the effects of the specific and nonspecific interactions on DPO4−DNA binding, we changed the strengths of the DPO4−DNA native contact interactions and salt concentrations, which result in different strengths of electrostatic interactions in the system. We found an apparent decrease of barrier height for the transition of $DNA_{EC} \rightarrow DNA_{IS}$ with increasing the strength of the DPO4−DNA native contact interactions (Figure 3A). This indicates that a strong

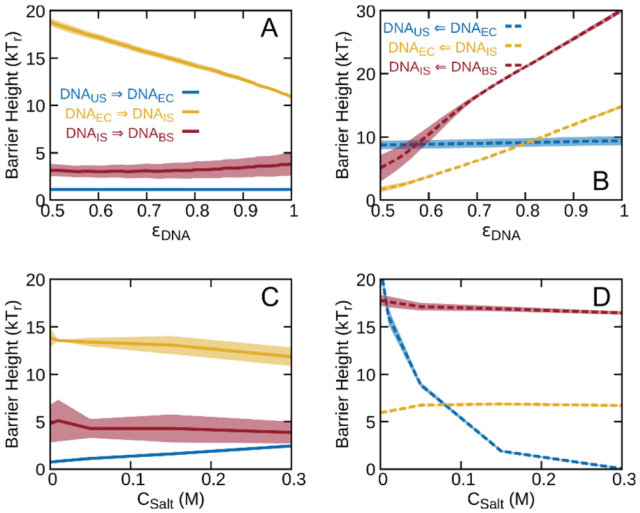


Figure 3. Barrier heights of DPO4−DNA binding and unbinding. The barrier heights at different stages along with ϵ_{DNA} for (A) binding and (B) unbinding processes. The barrier heights at different stages along with C_{Salt} for (C) binding and (D) unbinding processes. Shadow regions represent the standard errors at the corresponding mean values.

specific DPO4−DNA interaction can help the formation of the DNA_{IS} state. However, strengthening the DPO4−DNA native contacts has minor effects in accelerating the transition from the DNA_{IS} to the DNA_{BS}. Our previous work has demonstrated that the flexible domain interface in DPO4 plays a significant role in inducing the DNA_{IS} toward the DNA_{BS}.[16] These results together suggest that the last stage of DPO4−DNA binding is controlled by the intrinsic conformational dynamics of DPO4, rather than the interactions between DPO4 and DNA. Further calculation of the conformational distribution in DPO4 along with DNA binding shows that DPO4 has a notable population in the $DPO4_T$ form at the transition state (barrier region, $Q^{\#}_{DNA} \sim 60$) between the DNA_{IS} and DNA_{BS} (Figure S16). This feature signifies a "conformational selection"[44−46] for the last stage of the DPO4−DNA binding process, in line with the theoretical inference that the slow and large-scale conformational dynamics of the proteins favor the "conformational selection" mechanism.[47−49] For the unbinding process, we observed constantly accelerating effects led by decreasing the strength of the DPO4−DNA native contact interactions on the transitions from the DNA_{BS} to DNA_{IS} and then to the DNA_{EC} (Figure 3B). This is an intuitive finding as both DNA_{BS} and DNA_{EC} are stabilized by the specific DPO4−DNA interactions (Figure 2E).

Decreasing the salt concentrations (increasing the strength of electrostatic interactions) decreases the barrier height for DPO4 capturing DNA (Figure 3C), likely because of the "fly-casting" effects enhanced by the strengthening of the nonspecific DPO4−DNA electrostatic interactions at low salt concentrations.[50] Meanwhile, the barrier heights for forming the specific DPO4−DNA complex during the following two stages are slightly decreased by weakening the electrostatic interactions. It is possibly due to the fact that, during DNA binding, the magnitude of electrostatic interactions, which are largely non-native, exhibits only a slight decrease (Figure 2E). Similarly, it is also expected that the salt concentration plays a minor role in modulating the unbinding process from the

https://doi.org/10.1021/jacsau.1c00368
JACS Au 2022, 2, 341−356

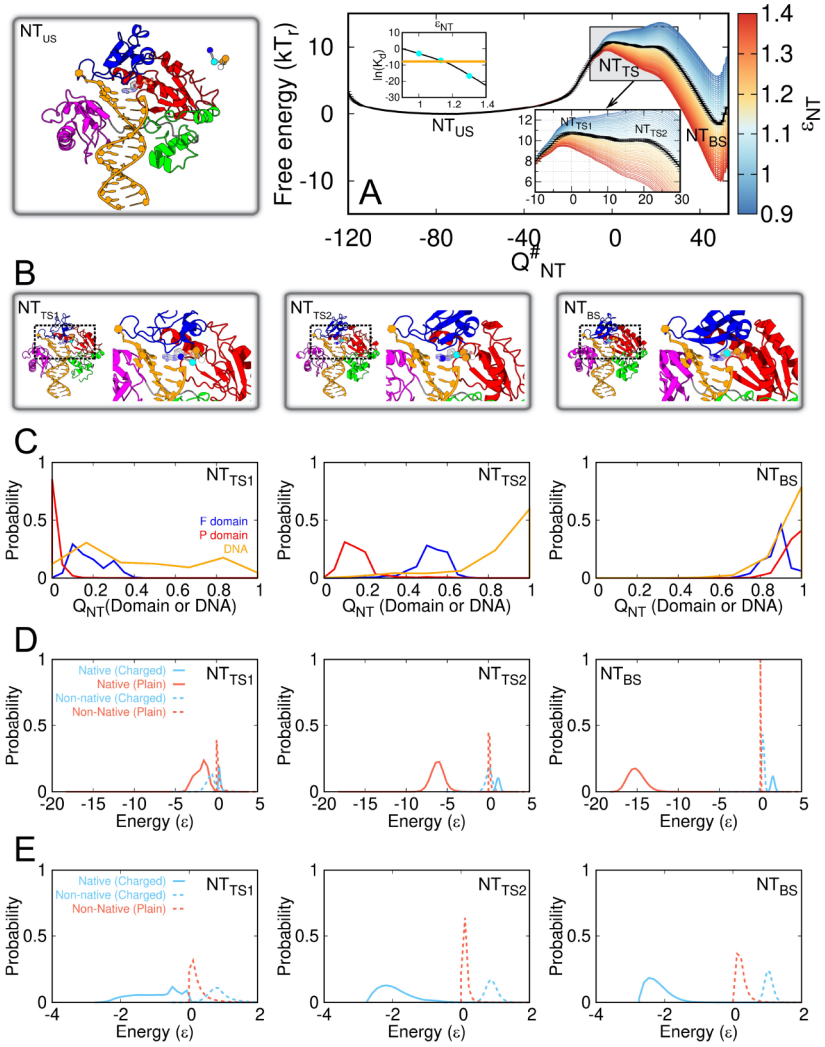


Figure 4. Nucleotide binding to the DPO4−DNA complex. (A) Free energy landscapes of nucleotide binding to the DPO4−DNA complex projected onto the binding reaction coordinate $Q^{\#}_{NT}$ at different ϵ_{NT}, where $Q^{\#}_{NT} = N_{NT}Q_{NT} - dRMS_{NT}$, Q_{NT} is the fraction of nucleotide interchain native contacts, N_{NT} is the number of nucleotide interchain native contacts, and $dRMS_{NT}$ is the difference of the distance of native contact pairs between the nucleotide and the DPO4−DNA complex with deviation from 0 indicating deviation from the native structure. $dRMS_{NT}$ is in the unit of Å. ϵ_{NT} is the strength of the nucleotide interchain native contacts. The free energy landscapes show two minima and one transition state, which are denoted as "Unbinding State (NT_{US})", "Transition State (NT_{TS})", and "Bound State (NT_{BS})". Inset plots are the zoom-in free energy landscapes at the region of the NT_{TS} (bottom) and binding affinity (K_d) along with ϵ_{NT} (top). The free energy landscapes and K_d at different ϵ_{NT} values were calculated from reweighting the thermodynamics at ϵ_{NT} = 1.00. The black line in the free energy plot (ϵ_{NT} = 1.13), which matches with the experimental K_d (200−800 μM),[24,25] was obtained from the direct umbrella sampling simulations. In the zoom-in free energy landscape plot, the NT_{TS} state is further divided into the NT_{TS1} and NT_{TS2} states. In the K_d plot, the gray shadow region corresponds to the standard error of the mean value (black line), and the yellow line indicates the experimental affinity. The cyan points in the K_d plot are the results from the direct umbrella sampling simulations. (B) Typical structures of the ternary DPO4−DNA−nucleotide system in the NT_{US}, NT_{TS1}, NT_{TS2}, and NT_{BS} states extracted from the simulations. The structure is shown in global view and zoom-in view for the nucleotide binding site at each binding state except the NT_{US} state. (C) Probability distribution of the fraction of native contacts formed by the individual domains and the linker in DPO4 and DNA with the nucleotide. The probability distribution of the interaction energy (D) between nucleotide and DPO4 and (E) between nucleotide and DNA.

DNA_{BS} toward the DNA_{EC} (Figure 3D). However, the electrostatic interactions were found to significantly impact the transition of $DNA_{EC} \rightarrow DNA_{US}$, which becomes the rate-limiting step for the unbinding process at the very low salt concentration.

Based on the effects of the interactions on the DPO4−DNA binding and unbinding, we conclude that different stages of the DPO4−DNA binding process are controlled by different types of interactions. For binding, the nonspecific electrostatic interactions provide the driving forces to initialize DPO4−DNA binding; then, the native contacts promote the formation of the partially bound complex, and finally the conformational dynamics in DPO4 drag the transition from the DNA_{IS} to the DNA_{BS} through the "conformational selection" mechanism. For unbinding, the native contact interactions between DPO4 and DNA play significant roles in the stages of transitions from the bound complex to the encounter complex; finally, the dissociation rate of DPO4−DNA is determined by the nonspecific electrostatic interactions. Our simulation results show the complex DPO4−DNA binding and unbinding

https://doi.org/10.1021/jacsau.1c00368
JACS Au 2022, 2, 341−356

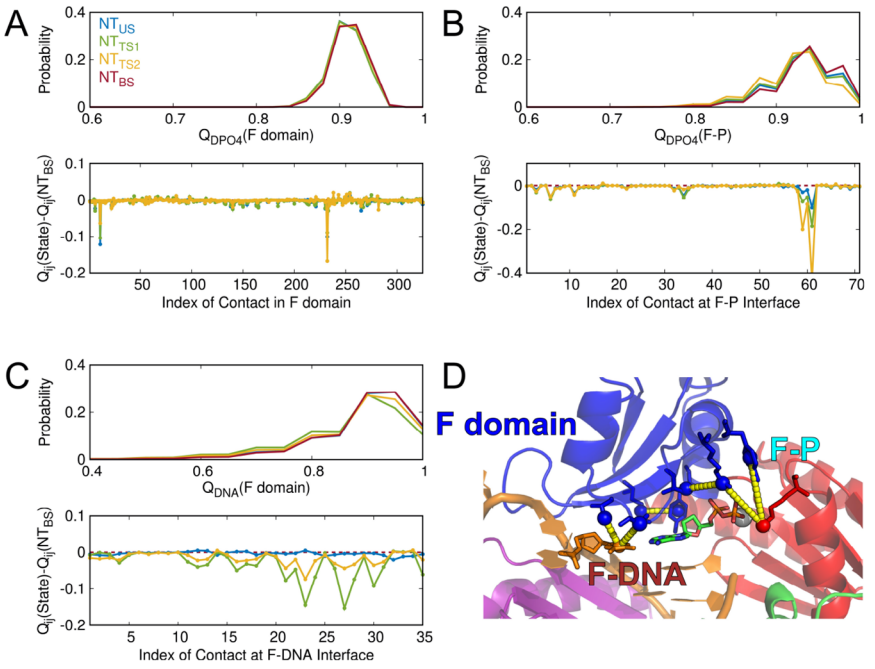


Figure 5. Formations of native contacts (A) within the F domain in DPO4, (B) at the F−P domain interface in DPO4, and (C) between the F domain in DPO4 and DNA, formed at the NT_{US}, NT_{TS1}, NT_{TS2}, and NT_{BS} states during the binding of nucleotide to the DPO4−DNA complex. In each part, the top panel shows the probability distribution of Q and the bottom panel shows the changes in the individual native contact (Q_{ij}, the probability of the individual native contact formed between bead i and bead j in the SBM) from the binding state to the bound state (NT_{BS}). (D) Illustration of the contacts that have large discrepancies during nucleotide binding. These contacts were identified with $Q_{ij}(State) - Q_{ij}(NT_{BS}) < -0.1$.

processes that are strongly dependent on the intrachain and interchain interactions as well as the ionic environments.

Two-State Nucleotide Binding to the DPO4−DNA Complex

In the last section, we studied the process of nucleotide binding to the DPO4−DNA complex, which completes the precatalytic steps for nucleotide incorporation. Our model, based on the DPO4−DNA binding model, further includes the native contacts and electrostatic interactions between the incoming nucleotide and the DPO4−DNA complex.[4] Here, we applied the umbrella sampling simulation strategy as we did in studying the DPO4−DNA binding with a focus on nucleotide binding. From the quantified free energy landscapes (Figure 4A), we observed a two-state binding process with a free energy barrier around 10 kT, which indicates a highly cooperative process. The free energy landscape changes upon different strengths of the nucleotide binding native contact interactions (ϵ_{NT}), leading to a switch of the location associated with the highest free energy barrier. When ϵ_{NT} is small, close to 0.9, the highest free energy barrier is located at $Q_{NT}^{\#} \sim 21$; when ϵ_{NT} is big, close to 1.4, the highest free energy barrier is located at $Q_{NT}^{\#} \sim 0$. This indicates an essential change of the rate-limiting step in nucleotide binding due to the change of the interactions associated with the nucleotide.

The contact analysis shows that the nucleotide initializes the binding process by forming the preliminary contacts with the F domain in DPO4 and DNA at the first transition state NT_{TS1} (Figure 4B,C and Figure S21). Upon proceeding to the second transition state NT_{TS2}, the nucleotide continues to stabilize the interactions with DPO4 and almost accomplishes forming contacts with the DNA at the final NT_{BS}. This indicates that the nucleotide at the NT_{TS2} has arrived at the correct spatial position on the DNA and the transition of $NT_{TS2} \rightarrow NT_{BS}$ mainly corresponds to the stabilization of the native contacts between the nucleotide and DPO4. Further analysis on interaction energy shows that the driving forces for nucleotide binding at the early stage in forming NT_{TS1} are both the native contacts and the non-native electrostatic interactions (Figure 4D,E), different from DPO4 binding to DNA. The transition from the NT_{TS2} to the NT_{BS} is promoted by the native contacts between the nucleotide and DPO4, so the increase of the nucleotide native contact strength can decrease the barrier height at the NT_{TS2} more than that at the NT_{TS1} (Figure S22), resulting in the switching of the transition state region.

In order to see how nucleotide binding influences the DPO4−DNA complex, we performed further analyses on the conformational dynamics of the DPO4−DNA complex. Overall, the DPO4−DNA complex exhibits very similar structures and interactions during nucleotide binding (Figure S24). However, careful examinations revealed that there are mild changes primarily associated with the F domain in DPO4 during nucleotide binding. To further assess the origins of the changes in the probability distribution of Q, we calculated the probability of the individual native contact formed at these 4 nucleotide binding states and made the comparison to it at the NT_{BS} state (Figure 5). We found that most of the native contacts remain similar to those at the NT_{BS} state during nucleotide binding. However, there are notable changes in a few native contacts within the F and P domains and at the F−P and F−DNA interfaces. In this regard, the slight changes in $Q_{DPO4}(F\ domain)$, $Q_{DPO4}(F-P)$, and $Q_{DNA}(F\ domain)$ during nucleotide binding are contributed by the destabilization of a small number of native contacts. The findings indicate that a few native contacts within the F domain and at the F−P and F−DNA interfaces are distorted by the nucleotide during its

https://doi.org/10.1021/jacsau.1c00368
JACS Au 2022, 2, 341−356

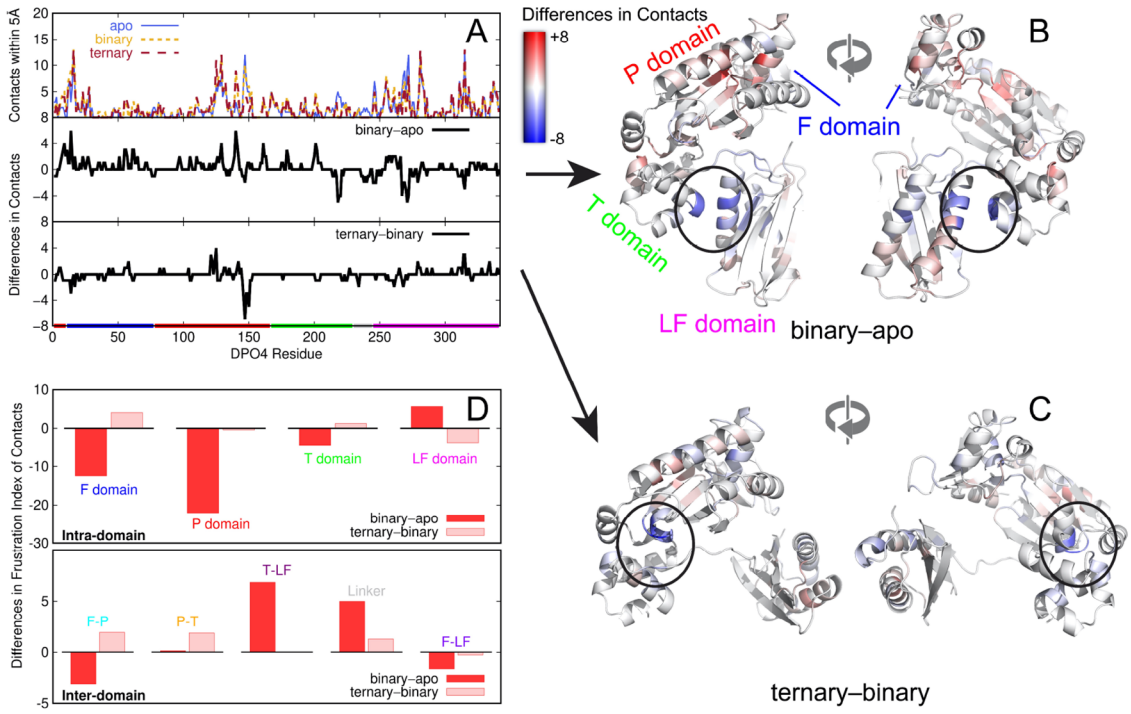


Figure 6. Localized frustration in DPO4. (A) Number of highly frustrated interactions in the vicinity of each residue in apo (PDB: 2RDI[11]), binary (PDB: 2RDJ[11]), and ternary (PDB: 1JX4[4]) states (top). The differences between the apo and the binary states; the binary and ternary states are respectively shown at the middle and bottom in (A). The *x*-axis is colored according to the domain index in DPO4, same as that in Figure 1. (B) and (C) are DPO4 structures colored according to the differences in contacts shown in (A). (D) Differences of the frustration indexes of contacts in DPO4 between the apo and binary state; the binary and ternary states are calculated based on the intradomain (top) and interdomain (bottom) interactions.

binding. We further characterized these contacts and mapped them onto the structure (Figure 5D). We found that all these contacts are located at or proximate to the nucleotide binding site; thus, the partial breaking of these contacts at the binding transition states can open the binding site in order to accommodate the incoming nucleotide. Our results suggest that the opening of the active site in the DPO4–DNA complex may facilitate nucleotide binding, similar to what was observed previously in a protein kinase with opening its active site for the ATP recruitment.[51,52]

The metal ions are indispensable when the DNA polymerase incorporates the nucleotide into the DNA molecule through the phosphoryl transfer reaction.[53] Since it is still challenging to accurately describe and model the ion interactions in a classical molecular dynamics simulation,[54,55] for simplicity, we coupled the ion and nucleotide binding by establishing the bonded interactions between the ion and nucleotide in our model. To see the effects of the ion on nucleotide binding, we removed the ion and its related interactions and performed the simulations again. We still observed a high free energy barrier for nucleotide binding with two transition states when the ion was absent (Figure S25). Further analyses on the native contacts and interactions between the nucleotide and DPO4–DNA complex showed similar results with and without the ion. This suggests that the binding pathways should not be substantially altered by the ion. However, we found a significant decrease in the stability of the bound state when the ion is absent. The result indicates that the interactions from the ion can help to form a stable ternary DPO4–DNA–nucleotide complex.

Localized Frustration in DPO4 at the apo, DNA Binary, and DNA–Nucleotide Ternary States

Naturally foldable proteins are deemed to obey the "principle of minimal frustration",[56] which efficiently guides the folding on the funneled energy landscapes. In reality, proteins often endure a limited fraction of interresidue interactions that conflict with others. Although these interactions generally weaken the stability of folded structures, they can promote specific conformational movements, which may be related to the functional purposes. Our results have indicated that the local conformational dynamics, rather than the global unfolding, has important effects on the substrate binding processes. To study the local functional conformational dynamics in DPO4 by taking into account the energetic frustrations, we quantified and compared the frustrations in DPO4 at the native apo, DNA binary, and DNA–nucleotide ternary states based on the method introduced by Ferreiro et al.[32]

In all three states of DPO4, we see that the interactions in DPO4 are dominated in the minimally frustrated way (Figure S26), indicating that DPO4 possesses a globally funneled folding energy landscape. The highly frustrated contacts are generally located on the surfaces of individual domains in DPO4, similar to the observations in the single domain proteins.[32,57] Interestingly, there are a notable amount of highly frustrated contacts formed between the T and the LF domains in the apo state of DPO4 (Figure S26A). This signifies a frustrated T–LF domain interface that is prone to unravel or crack in favor of the "open-to-closed" state transitions of DPO4. Our results are in line with the previous

https://doi.org/10.1021/jacsau.1c00368
JACS Au 2022, 2, 341–356

findings that the highly frustrated interactions in proteins are often enriched at the regions responsible for the large-scale conformational changes.[58]

To see how DNA and nucleotide affect the frustrations in DPO4, we compared the differences of the highly frustrated contacts formed in the vicinity of each residue in DPO4 between the apo and the binary states, as well as the binary and ternary states (Figure 6A). We found that the presence of DNA overall increases the degree of frustration for the residues in the F and P domains (Figure 6A, middle). This indicates that the residues in the F and P domains of DPO4 are more mobile in the binary state than they are in the apo state. Meanwhile, DNA binding decreases the degree of frustration at several residues located at the interface between the T and the LF domains (Figure 6B). This indicates that the highly frustrated contacts that favor the "open-to-closed" state transitions in DPO4 are diminished after DNA binding. Nucleotide binding has a much weaker effect on modulating the frustration in DPO4 than DNA binding does (Figure 6A, bottom, and Figure 6C). However, we note that a short segment (residues 145−152) in the P domain possesses less highly frustrated contacts when the nucleotide is present. We found that this region is located at the surface of the P domain, interacting with the T domain. Thus, our results indicate that the mobility of the P−T domain interface is weakened by nucleotide binding.

We further studied the effects of substrate binding on the highly frustrated contacts formed by the intra- and interdomain interactions through measuring the changes of the frustration index of contacts upon substrate binding. The frustration index measures how favorable a particular contact is relative to the set of all possible contacts in that location normalized by the variance of that distribution.[32] Thus, a low (high) value of the frustration index corresponds to a strongly (weakly) frustrated contact. We see that DNA binding destabilizes all of the three individual domains in the conserved polymerase core (the F, P, and T domains) through decreasing the frustration index of the corresponding intradomain contacts (Figure 6D, top). In addition, the F−P domain interface is considered to be more flexible, with a lower frustration index in the binary state than in the apo state. Meanwhile, the frustrated contacts formed by the T−LF domain interface and linker in the DPO4 apo state are significantly minimized by DNA binding, indicating that the conformational dynamics related to the T−LF domain interface and linker in DPO4 vanish in the binary state. Further binding of nucleotide to the binary state stabilizes the F domain and domain interfaces of F−P and P−T through increasing the frustration index of the contacts (Figure 6D, bottom). Together, our results suggest that the F domain and the F−P domain interface in DPO4 are more unstable at the binary state than at the apo and ternary states, and the P−T domain interface is more stable at the ternary state than at the binary state. The frustration analysis echoes our SBM simulation results that the F domain and the domain interfaces of F−P and P−T have high propensities to enable the specific conformational motions during the nucleotide binding process.

DISCUSSION AND CONCLUSIONS

Crystal structures revealed that DPO4 adopts distinct conformations with and without substrate binding.[4,11] With the double-well SBM, we studied the conformational transition of DPO4 between the inactive ($DPO4_A$) and the active ($DPO4_T$) form. During the transition, DPO4 forms an inevitable $DPO4_I$ state, which shows an extended linker connecting the T domain and the LF domain. A moderate increase of the temperature from room temperature leads to an increase of the population of the $DPO4_I$ state, where the individual domains and domain interfaces except the ones involving the LF domain remain folded (Figure 1C and Figure S6). The $DPO4_I$ state was found to be dominant when DPO4 forms the nonspecific encounter complex with the DNA (Figure 2C). This indicates that an elevated temperature leads to the increased population of the $DPO4_I$ state and thus may promote the DPO4−DNA binding process. On the other hand, the crystallographic DPO4−PCNA structure revealed that the LF domain in DPO4 is adapted from it in the $DPO4_A$ form to anchor PCNA for forming the complex.[12] This also implies that the $DPO4_I$ state may promote the DPO4−PCNA binding through breaking the interface of the LF domain in the $DPO4_A$ form. Since the $DPO4_I$ is an entropy-driven state, we proposed a positive role of the temperature in facilitating the binding of DPO4 to the substrates/proteins by inducing the formation of the $DPO4_I$ state. It is worth noting that, in our recent studies with the single-basin one-bead model,[15,16] the $DPO4_I$ state was not detected during DPO4 unfolding. With considering the local conformational transition in DPO4 and improving the coarse-grained level of the model to the two-bead double-well SBM, here we characterized the $DPO4_I$ as the intermediate state for both the "open-to-closed" transition and unfolding of DPO4. An unfolding intermediate state with similar structural characteristics of the $DPO4_I$ was previously observed in the melting experiments.[13] In this regard, the results generated by the current model are in good agreement with the experiments, suggesting that DPO4 undergoes partial unfolding to accomplish the functional conformational transition. The finding enriches the current understanding of conformational flexibility and frustration in the multidomain protein native structures for promoting the functional structure arrangements.[59,60]

Our simulations show that the conformational transition of DPO4 occurs through adapting the interfacial domain interactions involved by the LF domain while the other regions in DPO4 remain structurally unaltered. The results indicate that the domain interfaces of the LF domain in DPO4, which are responsible for the functional conformational dynamics, are more fragile than the others, which are responsible for maintaining the DPO4's folded structure. This has led to a globally funneled energy landscape of DPO4 with two small basins at the bottom of the funnel, corresponding to the inactive and active DPO4 conformational states (Figure 7). The transition between the $DPO4_A$ state and the $DPO4_T$ state has to go through the entropy-driven $DPO4_I$ state located at the upper layer of the energy landscape compared to that of these two states. All three states of DPO4 are located at the bottom of the funnel-like energy landscape, so the functional conformational dynamics of DPO4 is restricted to an efficient local structural rearrangement of the domain interfaces rather than a slow global unfolding.[34,61] This leads to first unraveling and then folding for the dynamical scenario of the conformational change.

The DPO4−DNA encounter complex is stabilized by the non-native electrostatic interactions with DPO4 largely in the $DPO4_I$ form (Figure 7). Given the fact that the linker in DPO4 is positively charged, we performed additional DPO4−DNA binding simulations with the linker in DPO4 free of positive charges. Despite the notable destabilization in the binding

https://doi.org/10.1021/jacsau.1c00368
JACS Au 2022, 2, 341−356

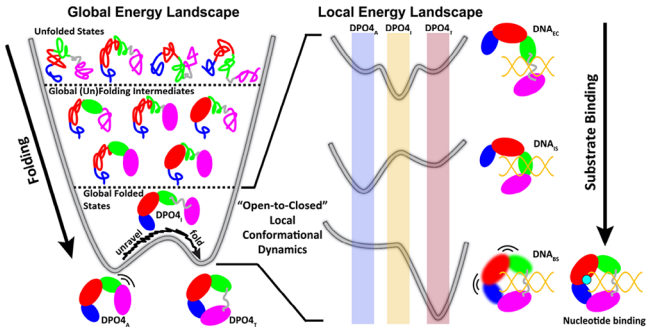


Figure 7. Scheme illustrating DPO4 folding, conformational dynamics, and substrate binding from the energy landscape perspective. For folding, the global energy landscape of DPO4 is funnel-like with two basins at the bottom. For substrate binding, the local energy landscapes responsible for the functional "open-to-closed" conformational dynamics of DPO4 are illustrated. DPO4 is in cartoon plot with each domain and the linker region colored by the same scheme used in Figure 1.

states led by removing the positive charges in the linker of DPO4, the binding barrier heights remain almost the same regardless of the presence of the positive charges in the linker. A significant decrease of the barrier height was observed in the transition from the DNA_{EC} to the DNA_{IS} state. Thus, the extended, positively charged linker of DPO4 can prevent the dissociation of the DPO4–DNA encounter complex, thus facilitating the binding process by restricting the searching in a 1D manner. From the DNA_{EC} to the DNA_{IS}, DPO4 undergoes a short-range translocation on DNA by forming the native contacts with DNA, primarily through the LF domain and linker region (Figure 7). Further analysis on the structural distribution of DPO4 during DNA binding shows that DPO4 has significantly decreased the population in the $DPO4_I$ form and increased the population in the $DPO4_A$ form upon forming the DNA_{IS} state (Figure S16). It is worth noting that the most populated forms of DPO4 in the DNA_{US}, DNA_{EC}, and DNA_{IS} are the $DPO4_A$, $DPO4_I$, and $DPO4_A$, respectively. The observation leads to the "backtracking" of DPO4 conformation during DNA binding.[62,63] In addition, we found that the transformation of the nonspecific DNA_{EC} to the specific DNA_{IS} is the rate-limiting step for DPO4–DNA binding and can be accelerated by strengthening the DPO4–DNA native contact interactions. This underlines the importance of the specific interactions in guiding and promoting DPO4–DNA binding. In cells, the coordination of the DNA polymerase is usually undertaken by the sliding clamps (PCNA and bacterial β-clamp).[64−66] Structural and biochemical studies revealed that DPO4 binds to PCNA with multiple conformations.[12] As revealed by our recent study,[67] the specific conformational adaption of DPO4 coupled with PCNA binding may be advantageous to regulate the activity and the accessibility of DPO4 at the replication site. Here, we found that the translocation of DPO4 to the replication site on DNA is slow because of the energetically frustrating protein–DNA landscape led by the nonspecific electrostatic interaction during the DNA searching process.[43,68,69] In this regard, we suggest that this process can be accelerated by PCNA *in vivo*, which provides the guiding interactions to position DPO4 to the spatial proximity of the DNA replication site. Although our previous SBM simulations observed the similar multistep DPO4–DNA binding process,[16,22] the conformational distribution of the DPO4 in the dissociative state and the DNA-binding interactions in the models were not calibrated to the experiments. Furthermore, the binding reaction coordinate was not optimally chosen, so that the precise characterization of the DPO4–DNA binding process was not possible. Here with the current well-calibrated model, we determined the binding mechanisms of the complex DPO4–DNA binding process, including the "backtracking" in DPO4 upon forming the DNA_{IS} state and "conformational selection" of DPO4 during the last transition of the DNA_{IS} to DNA_{BS} state.

Therefore, we found that the double-well two-bead well-calibrated SBM developed here goes beyond our previous model in studying DPO4 folding[14,15,22] and DNA binding[16,22] because of the following three aspects. First, we upgraded the one-bead model to the two-bead one. We demonstrated that the two-bead model can naturally reduce the contribution of the interdomain interactions in the total energy from the one-bead model. The weak interdomain interactions in DPO4 are requested by the efficient folding and DNA binding.[16] In addition, it has been recognized that the presence of the side chain in the two-bead model can have better placement of the charges than the one-bead model,[37] considering the fact that the electrostatic interactions are important for both the DPO4 folding and the DNA binding processes.[13,17] Second, the simulations of DPO4 with the double-well SBM led to the observation of a metastable $DPO4_I$ state, which was not able to be characterized by the single-basin SBM developed for DPO4 folding to the apo structure.[15,16] The $DPO4_I$ state was further identified as the intermediate state for both folding and conformational transition, enlightening the understanding of the interplay between DPO4 folding and conformational dynamics. Third, the DPO4–DNA binding model was calibrated to the experiments, and simulations were performed with a carefully determined reaction coordinate. This has enabled us to dissect the underlying mechanisms of conformational dynamics in DPO4 during its multistep binding to DNA.

Nucleotide binding goes through a typical two-state process associated with a high energy barrier. There are only minor structural changes in the DPO4–DNA complex after nucleotide binding, leading to the similar energy landscapes of DPO4 with and without nucleotide binding (Figure 7). However, there are notable changes in a few contacts in the DPO4–DNA complex during the nucleotide binding process. The nucleotide can destabilize several interactions surrounding the active site within the F domain, at the F–P domain interface and the interface between the F domain and DNA at the binding transition states. Protein structure opening for recruiting a substrate via partial protein unfolding, particularly at the binding site, was previously found in other protein systems.[51,52] This again underlines the importance of frustrations in protein structure for functional purposes. Interestingly, we found that the changes in interactions led by nucleotide binding are mainly associated with the F domain in DPO4. The flexibility inside the F domain and the fluctuating interactions at the interfaces of the F domain were previously characterized to have a potential contribution in catalyzing the translesion synthesis across various DNA lesions,[16,70,71] the *in vivo* role of DPO4 as a Y-family polymerase.[1,2] Here, we suggest a positive role of the small and intrinsically fluctuating F domain in facilitating nucleotide binding.

Our frustration analyses show that the T–LF domain interface is highly frustrated in the apo DPO4 state, and DNA binding increases the degree of frustration in the F domain and

https://doi.org/10.1021/jacsau.1c00368
JACS Au 2022, 2, 341–356

the domain interfaces of F−P and P−T. The frustrated regions and interactions have a high propensity to promote the specific conformational changes during substrate binding. Therefore, the results through calculating the energetic frustration at the native state[32] resonated with our SBM simulation findings and further provided a different way to dissect the roles of local DPO4 conformational dynamics in its functional substrate binding processes. It has been suggested that DPO4 can readily accept the damaged or mismatched base pairs during low-fidelity DNA polymerization due to the small energetic cost of adapting the DPO4 conformation to accommodate the base pair at the active site.[72] The notable enhancement of frustration in the polymerase core at the DPO4−DNA binary state observed in our study can induce the conformational flexibility in the DPO4−DNA complex, in particular at the active site, thus in favor of the recruitment of the incoming nucleotide. We further performed similar frustration analyses on a high-fidelity DNA polymerase, the DNA polymerase I large fragment from a thermostable strain of *Bacillus stearothermophilus* (Bacillus fragment, BF) (Figure S27).[73−75] We found that DNA binding induces the stabilization of the F domain in BF by decreasing the degree of the localized frustration (Figure S28). Meanwhile, the frustration index of the contacts within the P domain and at the F−P domain interface have only subtle changes upon DNA binding. The observations are very different from those of DPO4, where the F and P domains, as well as the F−P domain interface, are destabilized by DNA binding with decreasing the frustration index of the associated contacts. It has been well-known that the F domain is critical for modulating the fidelity of the DNA polymerase as it forms the contacts with the replicating base pair.[76] In this regard, we suggest that the stable F domain in the BF−DNA binary complex contributes to the sterically tight active site. This further promotes establishing the contacts between the F domain and the replicating base pair, responsible for the fidelity-checking mechanisms of nucleotide incorporation. The distinct results from the frustration analyses on DPO4 and BF indicate the potential connections of the localized frustrations to the polymerase fidelity. Therefore, our study provides a plausible explanation on the origin of the low-fidelity DNA polymerization by DPO4 from the conformational frustration and dynamical perspective.

In this study, we developed the SBMs to study the DPO4's global folding, local conformational transition, DNA binding, and nucleotide binding. We provided a full picture of conformational dynamics in DPO4 during its precatalytic substrate binding processes and characterized its relation and impacts on the substrate binding. Together with the localized frustration analyses, we emphasized the importance of the conformational dynamics and structural fluctuations of DPO4 in promoting the conformational transition from the inactive to active state, which forms the bound DPO4−DNA complex and facilitates nucleotide binding. Our findings provided mechanistic insights into the DPO4 conformational dynamics upon substrate binding. We anticipate that the results from the DPO4 study can be used to understand the conformational dynamics of other Y-family DNA polymerases, as they have the conserved structural architecture[4−8] with the flexible charged linker, which promotes the intermediate state formation.[13]

MATERIALS AND METHODS

A coarse-grained SBM was developed for studying the DPO4 conformational dynamics and its binding to DNA and a nucleotide. SBM is inspired by the energy landscape theory,[56,77] which assumes a "minimally frustrated" funnel-like energy landscape with biasing to the native state of folding and binding. Thus, SBM only considers the interactions in the protein native structure, so the relevant protein folding and binding processes can be accelerated. SBMs have been widely applied in studying various protein dynamics, including the protein folding,[78,79] the protein−DNA recognition,[42,43,80] the intrinsically disorder proteins' binding−folding,[27,38,81,82] and protein aggregation.[83] The results obtained from these simplified models were found to be consistent with experiments in many aspects,[78,84,85] confirming the validity of the SBMs.

For DPO4, we adapted the SBM, which often exhibits one basin representing the native state, to the double-well SBM, which has two basins corresponding to the apo DPO4 state and ternary DPO4−DNA−nucleotide state. Each residue in DPO4 is represented by two beads (except glycine), with one bead placing at the C_α position and the other placing at the centroid of the side chain. One unit charge was assigned to lysine and arginine (positive) and glutamic and aspartic acid (negative), respectively. The SBM potential for DPO4 used in our study is expressed as follows:

$$V_{\mathrm{SBM}}^{\mathrm{DPO4}} = V_{\mathrm{Local}}^{\mathrm{DPO4}} + V_{\mathrm{Native}}^{\mathrm{DPO4}} + V_{\mathrm{Non\text{-}native}}^{\mathrm{DPO4}} + V_{\mathrm{Electrostatic}}^{\mathrm{DPO4}}$$

$V_{\mathrm{Local}}^{\mathrm{DPO4}}$ describes the local interactions, including the bond stretching, angle bending, dihedral rotation, and chirality maintenance. Each term of $V_{\mathrm{Local}}^{\mathrm{DPO4}}$ (except bond stretching) has two potential minima with the positions adapted from the DPO4 apo and ternary crystal structures; $V_{\mathrm{Native}}^{\mathrm{DPO4}}$ is the nonlocal native biasing potential, based on a mixture contact map from the DPO4 apo and ternary crystal structures; $V_{\mathrm{Non\text{-}native}}^{\mathrm{DPO4}}$ represents the volume-excluding potential; and $V_{\mathrm{Electrostatic}}^{\mathrm{DPO4}}$ describes the electrostatic interactions through the Debye−Hückel model.

In our previous study,[16] we found that the application of the default homogeneous strength of the intra- and interdomain native contacts in the SBM does not result in the efficient folding and DNA binding processes for DPO4 from the kinetic aspects. From the evolutionary perspective, proteins are deemed to be evolved to optimize folding and function.[59] Slightly decreasing the interdomain native contacts in the homogeneous SBM can accelerate DPO4 folding and achieve efficient DPO4−DNA binding. The findings appear to be reasonable considering the fact that there are a large number of hydrophobic residues within the domains of DPO4,[4] so the intradomain interactions have been naturally strengthened in stabilizing the native structure of DPO4. Thus, it is important to take into account the heterogeneity of the interactions and weaken the interdomain interactions in the SBM. However, there is no experimental data serving as quantitative guidance to determine the strength of the interdomain interaction; here, we used our previous study as a reference.[16]

We used the single-basin one-bead SBM to study DPO4 folding and DNA binding previously.[16] We found that the optimal strength of the interdomain native contacts should be rescaled to 0.7−0.8 in order to achieve efficient DPO4 folding and DNA (un)binding. This results in a 10.95%−12.51% proportion of interdomain energetic contribution to the total, regarded as the optimal values. Here, we calculated the proportions of the energetic contribution of the interdomain interactions to the total energy of the apo structure and ternary structure with the default parameters of the double-well two-bead SBM. We found that the percentages are 12.91% and 12.40%, respectively. These two values are close to the range suggested by our previous study using the single-basin one-bead SBM. In other words, our current model with default parameters on the intra- and interdomain interaction strengths naturally generates an optimal energetic contribution of the interdomain interactions to the total energy for the efficient DPO4 folding and DNA binding. Therefore, we used the default intra- and interdomain interaction strength in the current model.

Further calibration on the strengths of the native contacts from the apo and ternary structure in building the mixed contact map was performed. This was realized by modulating the strengths and

https://doi.org/10.1021/jacsau.1c00368
JACS Au 2022, 2, 341−356

generating the probability distribution of DPO4 at the $DPO4_A$, $DPO4_I$, and $DPO4_T$ state. In principle, strengthening the contacts derived from the apo (ternary) structure should increase the probability of the $DPO4_A$ ($DPO4_T$) state (Figure S5). We determined the strengths of the native contacts based on the following two experimental observations. First, the crystal structure of DPO4 indicates that DPO4 should be mainly in the apo structure at room temperature.[11] Second, increasing temperature leads to an increase of the population of DPO4 in the ternary structure in solution.[26] In practice, we applied the thermodynamic reweighing method to the data generated at the default SBM simulations to obtain the thermodynamic results at the other designated parameters.[16,39,86]

For DNA, we used the short DNA segment (primer/template 13/17-mer DNA substrate) present in the ternary crystal structure.[4] Each nucleotide was reduced into three beads, representing the sugar, base, and phosphate groups, respectively. The phosphate pseudobead was modeled to carry one negative charge. In the simulations, the short DNA segment was used and set to be rigid. It is due to the following two facts. First, the binding of DPO4 to DNA is coordinated by PCNA *in vivo*.[65,66] During DPO4−DNA binding, PCNA binds with DPO4 and relocates DPO4 toward the vicinity of the DNA replication sites, so DPO4 does not have to perform the 1D diffusion on a long DNA molecule. Instead, a combination of the short-range 3D diffusion and local-range 1D diffusion appears to be appropriate to describe DPO4−DNA binding. Second, DNA has a high stiffness with a persistence length of ∼50 nm (∼150 bp).[87] The effects of the conformation and flexibility of the DNA molecule should be negligible on DPO4 binding, considering that the short DNA segment was used. We note that further improvement of the DNA model can be made by taking into account the DNA conformational flexibility while still using the SBMs for the proteins.[88−90]

The potential of the DPO4−DNA system is expressed as follows:

$$V_{\mathrm{SBM}}^{\mathrm{DPO4-DNA}} = V_{\mathrm{SBM}}^{\mathrm{DPO4}} + V_{\mathrm{Non\text{-}local}}^{\mathrm{DPO4-DNA}}$$

where $V_{\mathrm{Non\text{-}local}}^{\mathrm{DPO4-DNA}}$ is made up of the native, non-native, and electrostatic interaction potentials of interchain DPO4−DNA. The strength of the interchain native contacts between DPO4 and DNA was calibrated in accordance with the experimental affinity.[24,28]

For nucleotide, we determined the native contacts in the ternary structure.[4] The nucleotide was coarse-grained into five beads, representing the base, sugar, two phosphate groups, and one calcium ion. The potential of the DPO4−DNA−nucleotide system is expressed as follows:

$$V_{\mathrm{SBM}}^{\mathrm{DPO4-DNA-NT}} = V_{\mathrm{SBM}}^{\mathrm{DPO4-DNA}} + V_{\mathrm{SBM}}^{\mathrm{NT}} + V_{\mathrm{Non\text{-}local}}^{\mathrm{DPO4,DNA-NT}}$$

where $V_{\mathrm{SBM}}^{\mathrm{NT}}$ is biasing to the native structure of nucleotide in the crystal structure with a typical SBM expression and $V_{\mathrm{Non-local}}^{\mathrm{DPO4,DNA-NT}}$ describes the nonlocal interactions of nucleotide with DPO4 and DNA. The strength of the interchain native contacts between the nucleotide and the DPO4−DNA complex was calibrated in accordance with the experimental affinity.[24,25]

Simulations were performed by Gromacs software (version 4.5.7).[91] Reduced units were used throughout the simulations, except the length is in the units of nm or Å. For DPO4 folding, we performed two sets of REMD simulations starting from DPO4 structures in the apo and ternary forms, respectively. For DNA and nucleotide binding, we performed umbrella sampling simulations along with the corresponding binding reaction coordinates $Q^{\#}$, which is expressed as $Q^{\#} = NQ - dRMS$, where Q is the fraction of interchain native contacts for substrate binding (DNA or nucleotide), N is the number of the interchain native contacts, and $dRMS$ is the difference of the distance of native contact pairs. For the SBMs, Q was deemed as a good reaction coordinate for describing the protein folding[92] and adding the biased potentials during the umbrella sampling simulations.[93,94] However, for protein binding, Q was found to be incapable of discriminating among different unbound conformations,[39] which all have interchain Q values equal to 0. Discriminating the unbound states is critical to determine the binding and unbinding pathways. Previously, we used $dRMS$, which measures the degree of the dissociation relative to the bound structure. Although $dRMS$ has been proved effective for studying protein binding when applying the umbrella sampling simulations,[95,96] we previously found that $dRMS$ does not well capture the conformational differences after the ligand anchors the target protein.[36] In this regard, we applied the biased potentials on the reaction coordinate $Q^{\#}$, which contains the information from Q and $dRMS$. When the substrate is unbound from the DPO4, $Q \sim 0$ and the change of $Q^{\#}$ strongly depends on the change of $dRMS$, which is competent to discriminate the unbound states; when the substrate approaches the binding site, $dRMS$ becomes small, close to 0, so $Q^{\#}$ mainly relies on Q, which has been proven to be an optimal reaction coordinate for the SBMs. In this regard, $Q^{\#}$ can provide a comprehensive description of both the unbound states and the states after the substrate initializes the interactions with DPO4, thus resulting in the characterizations of the (un)binding pathways.

The umbrella sampling simulations were conducted with the aid of the PLUMED plugin (version 2.5.0).[97] Three sets of umbrella sampling simulations with different initial structures at one binding contact strength or salt concentration were performed. The multiple trajectories in one set of the simulation were analyzed by the Weighted Histogram Analysis Method (WHAM).[30] The trajectories were further analyzed by the reweighting method, which used the principle of statistical mechanics to obtain the thermodynamic results at other parameters in the SBMs.[16,39,86] The details of the models and simulations can be found in the Supporting Information.

Frustration analyses were carried out by the frustratometer server.[98] The server used the associative memory, water mediated, structure, and energy model (AWSEM), in which a coarse-grained representation of residue with interaction parameters optimized from landscape theory is used.[99] The latest version of AWSEM, which considers the electrostatic interactions,[100] is included in the frustratometer server and was used in this study. We used the crystal structures of DPO4 at the apo (PDB: 2RDI[11]), binary (PDB: 2RDJ[11]), and ternary states (PDB: 1JX4[4]) to perform the frustration analyses. The details of the method can be found here.[98]

The necessary files for setting up Gromacs (version 4.5.7 with PLUMED version 2.5.0) simulations and analysis programs/scripts are publicly available at https://osf.io/sj86k/.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/jacsau.1c00368.

Additional materials and methods, including the figures of model setup, simulation methodology details, frustration calculation, and additional results (PDF)

■ AUTHOR INFORMATION

Corresponding Author

Jin Wang − *Department of Chemistry and Department of Physics and Astronomy, State University of New York at Stony Brook, Stony Brook, New York 11794, United States;* orcid.org/0000-0002-2841-4913; Email: jin.wang.1@stonybrook.edu

Authors

Xiakun Chu − *Department of Chemistry, State University of New York at Stony Brook, Stony Brook, New York 11794, United States;* orcid.org/0000-0003-3166-7070

Zucai Suo − *Department of Biomedical Sciences, College of Medicine, Florida State University, Tallahassee, Florida 32306, United States;* orcid.org/0000-0003-3871-3420

Complete contact information is available at: https://pubs.acs.org/10.1021/jacsau.1c00368

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

We acknowledge the support from the National Institute of General Medical Sciences of the National Institutes of Health under Award Number R01GM124177. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. The authors would like to thank Stony Brook Research Computing and Cyberinfrastructure and the Institute for Advanced Computational Science at Stony Brook University for access to the high-performance SeaWulf computing system, which was made possible by a $1.4M National Science Foundation grant (#1531492).

■ REFERENCES

(1) Ohmori, H.; et al. The Y-Family of DNA Polymerases. *Mol. Cell* **2001**, *8*, 7−8.
(2) Sale, J. E.; Lehmann, A. R.; Woodgate, R. Y-family DNA polymerases and their role in tolerance of cellular DNA damage. *Nat. Rev. Mol. Cell Biol.* **2012**, *13*, 141.
(3) McCulloch, S. D.; Kunkel, T. A. The fidelity of DNA synthesis by eukaryotic replicative and translesion synthesis polymerases. *Cell Res.* **2008**, *18*, 148−161.
(4) Ling, H.; Boudsocq, F.; Woodgate, R.; Yang, W. Crystal structure of a Y-family DNA polymerase in action: a mechanism for error-prone and lesion-bypass replication. *Cell* **2001**, *107*, 91−102.
(5) Silvian, L. F.; Toth, E. A.; Pham, P.; Goodman, M. F.; Ellenberger, T. Crystal structure of a DinB family error-prone DNA polymerase from Sulfolobus solfataricus. *Nat. Struct. Biol.* **2001**, *8*, 984−989.
(6) Trincao, J.; Johnson, R. E.; Escalante, C. R.; Prakash, S.; Prakash, L.; Aggarwal, A. K. Structure of the catalytic core of S. cerevisiae DNA polymerase η: implications for translesion DNA synthesis. *Mol. Cell* **2001**, *8*, 417−426.
(7) Uljon, S. N.; Johnson, R. E.; Edwards, T. A.; Prakash, S.; Prakash, L.; Aggarwal, A. K. Crystal structure of the catalytic core of human DNA polymerase kappa. *Structure* **2004**, *12*, 1395−1404.
(8) Biertümpfel, C.; Zhao, Y.; Kondo, Y.; Ramón-Maiques, S.; Gregory, M.; Lee, J. Y.; Masutani, C.; Lehmann, A. R.; Hanaoka, F.; Yang, W. Structure and mechanism of human DNA polymerase η. *Nature* **2010**, *465*, 1044−1048.
(9) Fleck, O.; Schär, P. Translesion DNA synthesis: little fingers teach tolerance. *Curr. Biol.* **2004**, *14*, R389−R391.
(10) Goodman, M. F. Error-prone repair DNA polymerases in prokaryotes and eukaryotes. *Annu. Rev. Biochem.* **2002**, *71*, 17−50.
(11) Wong, J. H.; Fiala, K. A.; Suo, Z.; Ling, H. Snapshots of a Y-family DNA polymerase in replication: substrate-induced conformational transitions and implications for fidelity of Dpo4. *J. Mol. Biol.* **2008**, *379*, 317−330.
(12) Xing, G.; Kirouac, K.; Shin, Y. J.; Bell, S. D.; Ling, H. Structural insight into recruitment of translesion DNA polymerase Dpo4 to sliding clamp PCNA. *Mol. Microbiol.* **2009**, *71*, 678−691.
(13) Sherrer, S. M.; Maxwell, B. A.; Pack, L. R.; Fiala, K. A.; Fowler, J. D.; Zhang, J.; Suo, Z. Identification of an unfolding intermediate for a DNA lesion bypass polymerase. *Chem. Res. Toxicol.* **2012**, *25*, 1531−1540.
(14) Wang, Y.; Chu, X.; Suo, Z.; Wang, E.; Wang, J. Multidomain protein solves the folding problem by multifunnel combined landscape: Theoretical investigation of a Y-family DNA polymerase. *J. Am. Chem. Soc.* **2012**, *134*, 13755−13764.
(15) Chu, X.; Suo, Z.; Wang, J. Confinement and Crowding Effects on Folding of a Multidomain Y-Family DNA Polymerase. *J. Chem. Theory Comput.* **2020**, *16*, 1319−1332.
(16) Chu, X.; Suo, Z.; Wang, J. Investigating the trade-off between folding and function in a multidomain Y-family DNA polymerase. *eLife* **2020**, *9*, e60434.
(17) Xu, C.; Maxwell, B. A.; Brown, J. A.; Zhang, L.; Suo, Z. Global conformational dynamics of a Y-family DNA polymerase during catalysis. *PLoS Biol.* **2009**, *7*, e1000225.
(18) Maxwell, B. A.; Xu, C.; Suo, Z. Conformational dynamics of a Y-family DNA polymerase during substrate binding and catalysis as revealed by interdomain Forster resonance energy transfer. *Biochemistry* **2014**, *53*, 1768−1778.
(19) Raper, A. T.; Suo, Z. Investigation of intradomain motions of a Y-family DNA polymerase during substrate binding and catalysis. *Biochemistry* **2016**, *55*, 5832−5844.
(20) Brenlla, A.; Markiewicz, R. P.; Rueda, D.; Romano, L. J. Nucleotide selection by the Y-family DNA polymerase Dpo4 involves template translocation and misalignment. *Nucleic Acids Res.* **2014**, *42*, 2555−2563.
(21) Raper, A. T.; Reed, A. J.; Suo, Z. Kinetic mechanism of DNA polymerases: contributions of conformational dynamics and a third divalent metal ion. *Chem. Rev.* **2018**, *118*, 6000−6025.
(22) Chu, X.; Liu, F.; Maxwell, B. A.; Wang, Y.; Suo, Z.; Wang, H.; Han, W.; Wang, J. Dynamic Conformational Change Regulates the Protein-DNA Recognition: An Investigation on Binding of a Y-Family Polymerase to Its Target DNA. *PLoS Comput. Biol.* **2014**, *10*, e1003804.
(23) Maxwell, B. A.; Suo, Z. Kinetic basis for the differing response to an oxidative lesion by a replicative and a lesion bypass DNA polymerase from Sulfolobus solfataricus. *Biochemistry* **2012**, *51*, 3485−3496.
(24) Fiala, K. A.; Suo, Z. Mechanism of DNA polymerization catalyzed by Sulfolobus solfataricus P2 DNA polymerase IV. *Biochemistry* **2004**, *43*, 2116−2125.
(25) Fiala, K. A.; Sherrer, S. M.; Brown, J. A.; Suo, Z. Mechanistic consequences of temperature on DNA polymerization catalyzed by a Y-family DNA polymerase. *Nucleic Acids Res.* **2008**, *36*, 1990−2001.
(26) Lee, E.; Fowler, J. D.; Suo, Z.; Wu, Z. Backbone assignment of the binary complex of the full length Sulfolobus solfataricus DNA polymerase IV and DNA. *Biomol. NMR Assignments* **2017**, *11*, 39−43.
(27) de Sancho, D.; Best, R. B. Modulation of an IDP binding mechanism and rates by helix propensity and non-native interactions: association of HIF1α with CBP. *Mol. BioSyst.* **2012**, *8*, 256−267.
(28) Sherrer, S. M.; Brown, J. A.; Pack, L. R.; Jasti, V. P.; Fowler, J. D.; Basu, A. K.; Suo, Z. Mechanistic studies of the bypass of a bulky single-base lesion catalyzed by a Y-family DNA polymerase. *J. Biol. Chem.* **2009**, *284*, 6379−6388.
(29) Sugita, Y.; Okamoto, Y. Replica-exchange molecular dynamics method for protein folding. *Chem. Phys. Lett.* **1999**, *314*, 141−151.
(30) Kumar, S.; Rosenberg, J. M.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A. The weighted histogram analysis method for free-energy calculations on biomolecules. I. The method. *J. Comput. Chem.* **1992**, *13*, 1011−1021.
(31) Nie, Q.-M.; Sun, L.-Z.; Li, H.-B.; Chu, X.; Wang, J. Effects of electrostatic interactions on global folding and local conformational dynamics of a multidomain Y-family DNA polymerase. *Phys. Chem. Chem. Phys.* **2021**, *23*, 20841−20847.
(32) Ferreiro, D. U.; Hegler, J. A.; Komives, E. A.; Wolynes, P. G. Localizing frustration in native proteins and protein assemblies. *Proc. Natl. Acad. Sci. U. S. A.* **2007**, *104*, 19819−19824.
(33) Ferreiro, D. U.; Komives, E. A.; Wolynes, P. G. Frustration in biomolecules. *Q. Rev. Biophys.* **2014**, *47*, 285−363.
(34) Whitford, P. C.; Onuchic, J. N. What protein folding teaches us about biological function and molecular machines. *Curr. Opin. Struct. Biol.* **2015**, *30*, 57−62.
(35) Zagrovic, B.; Snow, C. D.; Khaliq, S.; Shirts, M. R.; Pande, V. S. Native-like mean structure in the unfolded ensemble of small proteins. *J. Mol. Biol.* **2002**, *323*, 153−164.
(36) Chu, X.; Wang, J. Position-, disorder-, and salt-dependent diffusion in binding-coupled-folding of intrinsically disordered proteins. *Phys. Chem. Chem. Phys.* **2019**, *21*, 5634−5645.

https://doi.org/10.1021/jacsau.1c00368
JACS Au 2022, 2, 341−356

(37) Azia, A.; Levy, Y. Nonnative electrostatic interactions can modulate protein folding: molecular dynamics with a grain of salt. *J. Mol. Biol.* **2009**, *393*, 527−542.
(38) Chu, X.; Wang, Y.; Gan, L.; Bai, Y.; Han, W.; Wang, E.; Wang, J. Importance of electrostatic interactions in the association of intrinsically disordered histone chaperone Chz1 and histone H2A.Z-H2B. *PLoS Comput. Biol.* **2012**, *8*, e1002608.
(39) Cao, H.; Huang, Y.; Liu, Z. Interplay between binding affinity and kinetics in protein-protein interactions. *Proteins: Struct., Funct., Genet.* **2016**, *84*, 920−933.
(40) von Hippel, P. H.; Berg, O. Facilitated target location in biological systems. *J. Biol. Chem.* **1989**, *264*, 675−678.
(41) Halford, S. E.; Marko, J. F. How do site-specific DNA-binding proteins find their targets? *Nucleic acids research* **2004**, *32*, 3040−3052.
(42) Givaty, O.; Levy, Y. Protein sliding along DNA: dynamics and structural characterization. *J. Mol. Biol.* **2009**, *385*, 1087−1097.
(43) Chu, X.; Muñoz, V. Roles of conformational disorder and downhill folding in modulating protein-DNA recognition. *Phys. Chem. Chem. Phys.* **2017**, *19*, 28527−28539.
(44) Kumar, S.; Ma, B.; Tsai, C.-J.; Sinha, N.; Nussinov, R. Folding and binding cascades: dynamic landscapes and population shifts. *Protein science* **2000**, *9*, 10−19.
(45) Tsai, C.-J.; Ma, B.; Sham, Y. Y.; Kumar, S.; Nussinov, R. Structured disorder and conformational selection. *Proteins: Struct., Funct., Genet.* **2001**, *44*, 418−427.
(46) Bosshard, H. R. Molecular recognition by induced fit: how fit is the concept? *Physiology* **2001**, *16*, 171−173.
(47) Zhou, H.-X. From induced fit to conformational selection: a continuum of binding mechanism controlled by the timescale of conformational transitions. *Biophys. J.* **2010**, *98*, L15−L17.
(48) Okazaki, K.-i.; Takada, S. Dynamic energy landscape view of coupled binding and protein conformational change: induced-fit versus population-shift mechanisms. *Proc. Natl. Acad. Sci. U. S. A.* **2008**, *105*, 11182−11187.
(49) 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*, 13737−13741.
(50) 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.* **2000**, *97*, 8868−8873.
(51) Hyeon, C.; Jennings, P. A.; Adams, J. A.; Onuchic, J. N. Ligand-induced global transitions in the catalytic domain of protein kinase A. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106*, 3023−3028.
(52) Chu, W.-T.; Chu, X.; Wang, J. Binding mechanism and dynamic conformational change of C subunit of PKA with different pathways. *Proc. Natl. Acad. Sci. U. S. A.* **2017**, *114*, E7959−E7968.
(53) Steitz, T.; Smerdon, S.; Jager, J.; Joyce, C. A unified polymerase mechanism for nonhomologous DNA and RNA polymerases-Comment/reply. *Science* **1994**, *266*, 2022.
(54) Riahi, S.; Roux, B.; Rowley, C. N. QM/MM molecular dynamics simulations of the hydration of Mg (II) and Zn (II) ions. *Can. J. Chem.* **2013**, *91*, 552−558.
(55) Lev, B.; Roux, B.; Noskov, S. Y. Relative free energies for hydration of monovalent ions from QM and QM/MM simulations. *J. Chem. Theory Comput.* **2013**, *9*, 4165−4175.
(56) Bryngelson, J. D.; Wolynes, P. G. Spin glasses and the statistical mechanics of protein folding. *Proc. Natl. Acad. Sci. U. S. A.* **1987**, *84*, 7524−7528.
(57) Chen, M.; Chen, X.; Schafer, N. P.; Clementi, C.; Komives, E. A.; Ferreiro, D. U.; Wolynes, P. G. Surveying biomolecular frustration at atomic resolution. *Nat. Commun.* **2020**, *11*, 5944.
(58) Ferreiro, D. U.; Hegler, J. A.; Komives, E. A.; Wolynes, P. G. On the role of frustration in the energy landscapes of allosteric proteins. *Proc. Natl. Acad. Sci. U. S. A.* **2011**, *108*, 3499−3503.
(59) Bigman, L. S.; Levy, Y. Proteins: molecules defined by their trade-offs. *Curr. Opin. Struct. Biol.* **2020**, *60*, 50−56.
(60) Levy, Y. Protein assembly and building blocks: beyond the limits of the LEGO brick metaphor. *Biochemistry* **2017**, *56*, 5040−5048.
(61) Whitford, P. C.; Sanbonmatsu, K. Y.; Onuchic, J. N. Biomolecular dynamics: order-disorder transitions and energy landscapes. *Rep. Prog. Phys.* **2012**, *75*, 076601.
(62) Gosavi, S.; Chavez, L. L.; Jennings, P. A.; Onuchic, J. N. Topological frustration and the folding of interleukin-1β. *J. Mol. Biol.* **2006**, *357*, 986−996.
(63) Gosavi, S.; Whitford, P. C.; Jennings, P. A.; Onuchic, J. N. Extracting function from a β-trefoil folding motif. *Proc. Natl. Acad. Sci. U. S. A.* **2008**, *105*, 10384−10389.
(64) Kuriyan, J.; O'Donnell, M. Sliding clamps of DNA polymerases. *J. Mol. Biol.* **1993**, *234*, 915−925.
(65) Friedberg, E. C.; Lehmann, A. R.; Fuchs, R. P. Trading places: how do DNA polymerases switch during translesion DNA synthesis? *Mol. Cell* **2005**, *18*, 499−505.
(66) Jansen, J. G.; Fousteri, M. I.; de Wind, N. Send in the clamps: control of DNA translesion synthesis in eukaryotes. *Mol. Cell* **2007**, *28*, 522−529.
(67) Chu, W.-T.; Suo, Z.; Wang, J. Binding-induced conformational changes involved in sliding clamp PCNA and DNA polymerase DPO4. *Iscience* **2020**, *23*, 101117.
(68) Marcovitz, A.; Levy, Y. Frustration in protein-DNA binding influences conformational switching and target search kinetics. *Proc. Natl. Acad. Sci. U. S. A.* **2011**, *108*, 17957−17962.
(69) Leven, I.; Levy, Y. Quantifying the two-state facilitated diffusion model of protein-DNA interactions. *Nucleic Acids Res.* **2019**, *47*, 5530−5538.
(70) Bauer, J.; Xing, G.; Yagi, H.; Sayer, J. M.; Jerina, D. M.; Ling, H. A structural gap in Dpo4 supports mutagenic bypass of a major benzo [a] pyrene dG adduct in DNA through template misalignment. *Proc. Natl. Acad. Sci. U. S. A.* **2007**, *104*, 14905−14910.
(71) Vyas, R.; Efthimiopoulos, G.; Tokarsky, E. J.; Malik, C. K.; Basu, A. K.; Suo, Z. Mechanistic basis for the bypass of a bulky DNA adduct catalyzed by a Y-family DNA polymerase. *J. Am. Chem. Soc.* **2015**, *137*, 12131−12142.
(72) Mizukami, S.; Kim, T. W.; Helquist, S. A.; Kool, E. T. Varying DNA base-pair size in subangstrom increments: evidence for a loose, not large, active site in low-fidelity Dpo4 polymerase. *Biochemistry* **2006**, *45*, 2772−2778.
(73) Kiefer, J. R.; Mao, C.; Hansen, C. J.; Basehore, S. L.; Hogrefe, H. H.; Braman, J. C.; Beese, L. S. Crystal structure of a thermostable Bacillus DNA polymerase I large fragment at 2.1 Å resolution. *Structure* **1997**, *5*, 95−108.
(74) Kiefer, J. R.; Mao, C.; Braman, J. C.; Beese, L. S. Visualizing DNA replication in a catalytically active Bacillus DNA polymerase crystal. *Nature* **1998**, *391*, 304−307.
(75) Johnson, S. J.; Taylor, J. S.; Beese, L. S. Processive DNA synthesis observed in a polymerase crystal suggests a mechanism for the prevention of frameshift mutations. *Proc. Natl. Acad. Sci. U. S. A.* **2003**, *100*, 3895−3900.
(76) Bębenek, A.; Ziuzia-Graczyk, I. Fidelity of DNA replication-a matter of proofreading. *Curr. Genet.* **2018**, *64*, 985−996.
(77) Bryngelson, J. D.; Onuchic, J. N.; Socci, N. D.; Wolynes, P. G. Funnels, pathways, and the energy landscape of protein folding: a synthesis. *Proteins: Struct., Funct., Genet.* **1995**, *21*, 167−195.
(78) Clementi, C.; Nymeyer, H.; Onuchic, J. N. Topological and energetic factors: what determines the structural details of the transition state ensemble and "en-route" intermediates for protein folding? An investigation for small globular proteins. *J. Mol. Biol.* **2000**, *298*, 937−953.
(79) Whitford, P. C.; Noel, J. K.; Gosavi, S.; Schug, A.; Sanbonmatsu, K. Y.; Onuchic, J. N. An all-atom structure-based potential for proteins: Bridging minimal models with all-atom empirical forcefields. *Proteins: Struct., Funct., Genet.* **2009**, *75*, 430−441.

https://doi.org/10.1021/jacsau.1c00368
JACS Au 2022, 2, 341−356

(80) Levy, Y.; Onuchic, J. N.; Wolynes, P. G. Fly-casting in protein-DNA binding: frustration between protein folding and electrostatics facilitates target recognition. *J. Am. Chem. Soc.* **2007**, *129*, 738−739.
(81) Wang, J.; Wang, Y.; Chu, X.; Hagen, S. J.; Han, W.; Wang, E. Multi-scaled explorations of binding-induced folding of intrinsically disordered protein inhibitor IA3 to its target enzyme. *PLoS Comput. Biol.* **2011**, *7*, e1001118.
(82) Chu, X.; Gan, L.; Wang, E.; Wang, J. Quantifying the topography of the intrinsic energy landscape of flexible biomolecular recognition. *Proc. Natl. Acad. Sci. U. S. A.* **2013**, *110*, E2342−E2351.
(83) Yang, S.; Levine, H.; Onuchic, J. N. Protein oligomerization through domain swapping: role of inter-molecular interactions and protein concentration. *J. Mol. Biol.* **2005**, *352*, 202−211.
(84) Levy, Y.; Wolynes, P. G.; Onuchic, J. N. Protein topology determines binding mechanism. *Proc. Natl. Acad. Sci. U. S. A.* **2004**, *101*, 511−516.
(85) 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*, 1121−1145.
(86) Li, M.; Cao, H.; Lai, L.; Liu, Z. Disordered linkers in multidomain allosteric proteins: Entropic effect to favor the open state or enhanced local concentration to favor the closed state? *Protein Sci.* **2018**, *27*, 1600−1610.
(87) Hagerman, P. J. Flexibility of DNA. *Annu. Rev. Biophys. Biophys. Chem.* **1988**, *17*, 265−286.
(88) Sambriski, E.; Schwartz, D.; De Pablo, J. A mesoscale model of DNA and its renaturation. *Biophys. J.* **2009**, *96*, 1675−1690.
(89) Terakawa, T.; Kenzaki, H.; Takada, S. p53 searches on DNA by rotation-uncoupled sliding at C-terminal tails and restricted hopping of core domains. *J. Am. Chem. Soc.* **2012**, *134*, 14555−14562.
(90) Daitchman, D.; Greenblatt, H. M.; Levy, Y. Diffusion of ring-shaped proteins along DNA: case study of sliding clamps. *Nucleic Acids Res.* **2018**, *46*, 5935−5949.
(91) Hess, B.; Kutzner, C.; van der Spoel, D.; Lindahl, E. GROMACS 4: algorithms for highly efficient, load-balanced, and scalable molecular simulation. *J. Chem. Theory Comput.* **2008**, *4*, 435−447.
(92) Cho, S. S.; Levy, Y.; Wolynes, P. G. P versus Q: Structural reaction coordinates capture protein folding on smooth landscapes. *Proc. Natl. Acad. Sci. U. S. A.* **2006**, *103*, 586−591.
(93) Badasyan, A.; Liu, Z.; Chan, H. S. Probing possible downhill folding: Native contact topology likely places a significant constraint on the folding cooperativity of proteins with 40 residues. *J. Mol. Biol.* **2008**, *384*, 512−530.
(94) Giri Rao, V. V. H.; Gosavi, S. On the folding of a structurally complex protein to its metastable active state. *Proc. Natl. Acad. Sci. U. S. A.* **2018**, *115*, 1998−2003.
(95) Kim, Y. C.; Hummer, G. Coarse-grained models for simulations of multiprotein complexes: application to ubiquitin binding. *J. Mol. Biol.* **2008**, *375*, 1416−1433.
(96) Domanski, J.; Sansom, M. S.; Stansfeld, P. J.; Best, R. B. Balancing force field protein-lipid interactions to capture transmembrane helix-helix association. *J. Chem. Theory Comput.* **2018**, *14*, 1706−1715.
(97) Tribello, G. A.; Bonomi, M.; Branduardi, D.; Camilloni, C.; Bussi, G. PLUMED 2: New feathers for an old bird. *Comput. Phys. Commun.* **2014**, *185*, 604−613.
(98) Parra, R. G.; Schafer, N. P.; Radusky, L. G.; Tsai, M.-Y.; Guzovsky, A. B.; Wolynes, P. G.; Ferreiro, D. U. Protein Frustratometer 2: a tool to localize energetic frustration in protein molecules, now with electrostatics. *Nucleic Acids Res.* **2016**, *44*, W356−W360.
(99) Davtyan, A.; Schafer, N. P.; Zheng, W.; Clementi, C.; Wolynes, P. G.; Papoian, G. A. AWSEM-MD: protein structure prediction using coarse-grained physical potentials and bioinformatically based local structure biasing. *J. Phys. Chem. B* **2012**, *116*, 8494−8503.
(100) Tsai, M.-Y.; Zheng, W.; Balamurugan, D.; Schafer, N. P.; Kim, B. L.; Cheung, M. S.; Wolynes, P. G. Electrostatics, structure prediction, and the energy landscapes for protein folding and binding. *Protein Sci.* **2016**, *25*, 255−269.

https://doi.org/10.1021/jacsau.1c00368
JACS Au 2022, 2, 341−356