Free energy simulations reveal molecular mechanism for functional switch of a DNA helicase

Wen Ma[1,2,3,4], **Kevin D Whitley**[1,3], **Yann R Chemla**[1,3,4*], **Zaida Luthey-Schulten**[1,2,3,4,5*], **Klaus Schulten**[1,2,3,4†]

***For correspondence:**
ychemla@illinois.edu (YRC);
zan@illinois.edu (ZLS)

†In memoriam.

[1]Center for the Physics of Living Cells, University of Illinois at Urbana-Champaign; [2]Beckman Institute for Advanced Science and Technology, University of Illinois at Urbana-Champaign; [3]Center for Biophysics and Computational Biology, University of Illinois at Urbana-Champaign; [4]Department of Physics, University of Illinois at Urbana-Champaign; [5]Department of Chemistry, University of Illinois at Urbana-Champaign

Abstract

Helicases play key roles in genome maintenance, yet it remains elusive how these enzymes change conformations and how transitions between different conformational states regulate nucleic acid reshaping. Here we developed a computational technique combining structural bioinformatics approaches and atomic-level free energy simulations to characterize how the *E. coli* DNA repair enzyme UvrD changes its conformation at the fork junction to switch its function from unwinding to rezipping DNA. The lowest free energy path shows that UvrD opens the interface between two domains, allowing the bound ssDNA to escape. The simulation results predict a key metastable "tilted" state during ssDNA strand switching. By simulating FRET distributions with fluorophores attached to UvrD, we show that the new state is supported quantitatively by single-molecule measurements. The present study deciphers key elements for the "hyper-helicase" behavior of a mutant, and provides an effective framework to characterize directly structure-function relationships in molecular machines.

Introduction

Helicases are ubiquitous motor proteins that move along nucleic acids and separate duplex DNA or RNA into its component strands. This role is critical for various aspects of DNA and RNA metabolism; defects in helicase function in humans can lead to genomic instability and a predisposition to cancer (***van Brabant et al., 2000***; ***Brosh, 2013***). Characterizing the atomistic mechanism for helicase function, though challenging, is crucial to link protein structure with their function and help engineering helicases with novel activities (***Arslan et al., 2015***).

DNA helicases can unwind double-stranded DNA (dsDNA) into single-stranded DNA (ssDNA), which are later copied during DNA replication or modified in DNA repair processes (***Wu and Spies, 2013***; ***Lohman et al., 2008***). They are classified into six superfamilies (SF), among which SF1 and SF2 helicases are the largest superfamilies and share many similar conserved motifs. The minimal functional units for SF1 and SF2 helicases are monomers that contain two RecA-like motor domains for ATP hydrolysis (***Singleton et al., 2007***). SF1 helicases can unwind dsDNA by translocating on a ssDNA strand as shown in Figure 1a. Such translocation happens in a stepwise manner,

during which the chemical energy from ATP hydrolysis is used to break the bonds in dsDNA via conformational changes of the motor domains (**Yang, 2010**; **Patel and Donmez, 2006**). An exemplary *E. coli* helicase, UvrD, belonging to SF1, has many cellular roles such as methyl-directed mismatch repair (**Iyer et al., 2006**; **Spies and Fishel, 2015**) and nucleotide excision repair (**Sancar, 1996**) by unwinding duplex DNA. UvrD can also prevent deleterious recombination by removing RecA filaments from ssDNA (**Lestini and Michel, 2007**). Along with its homologous proteins PcrA and Rep, UvrD has been demonstrated in experiments to translocate on ssDNA progressively 3′ to 5′ (**Matson, 1986**; **Mechanic et al., 1999**; **Dillingham et al., 2000**; **Myong et al., 2005**; **Fischer et al., 2004**). Structures of UvrD-like SF1 helicase solved so far share a four-subdomain tertiary arrangement (1A/2A/1B/2B) (**Singleton et al., 2007**), including two RecA-like domains (1A/2A) which contain the ATP binding site and are proposed to function as the translocase (**Dillingham et al., 2001**; **Lee and Yang, 2006**), and a flexible domain (2B) which is believed to play a regulatory role in helicase activity (**Lohman et al., 2008**; **Dillingham, 2011**). In particular, the 2B domain is known to adopt different conformations (**Velankar et al., 1999**; **Brendza et al., 2005**; **Jia et al., 2011**; **Nguyen et al., 2017**) and has been proposed to act as a "molecular switch" controlling UvrD unwinding (**Comstock et al., 2015**).

Combining optical tweezers and single-molecule FRET, Comstock et al. (**Comstock et al., 2015**) demonstrated that UvrD can switch its activity from DNA unwinding to rezipping (measured by optical tweezers) by dramatically changing its conformation between two states (detected by FRET). The transition from unwinding to rezipping activities was proposed to occur through switching ssDNA strands, accompanied by rotation of the 2B domain (see Figure 1b). In this model, the GIG motif on 2B serves as an anchor point on dsDNA above the fork junction, such that rotation of 2B can position the 1A/2A translocase domains on either ssDNA strand, leading to 3′ to 5′ UvrD translocation either toward (unwinding) or away from (rezipping) the DNA fork. Two crystal structures seem to support this strand-switching model (see Figure 1c): one structure of UvrD (pdb code: 2IS2) (**Lee and Yang, 2006**) bound to a dsDNA-ssDNA junction is expected to be the "unwinding" state (defined here as the "closed" state) because its 1A/2A domains would translocate UvrD into the DNA fork; the other structure (pdb code: 3LFU) (**Jia et al., 2011**) solved without DNA is expected to represent the "rezipping" state (defined here as the apo state) because the 1A/2A domains presumably would be bound to the opposing strand, translocating UvrD away from the DNA fork. The structural differences between closed and apo states mainly involve a simple rotation of the 2B domain (Figure 1c).

However, in order for the ssDNA strand-switching to happen, the rezipping state must adopt a conformation with a gap between the 1B and 2B domain that is large enough for the bound ssDNA to escape, whereas in both the closed and apo structures the four domains 1B-1A-2A-2B form a closed ring topologically. As we show here, contrary to the common assumption that the apo structure is a functional state of UvrD, the FRET signal simulated using real fluorophores attached to the apo-state structure does not match the experimentally observed signal of the rezipping state, nor the unwinding state. Furthermore, it has been reported that cross-linking the 2B and 1B domains of the SF1 helicase Rep can change it into a superhelicase (**Arslan et al., 2015**), capable of unwinding thousands of base pairs processively. What are the key regulatory factors for the functional switch and is it possible to design mutants with different activities?

To characterize the conformational states of UvrD at the fork junction and the transitions between those states, we use MD simulations, which are well-suited to study atomic-level mechanisms in conjunction with crystallography, single-molecule and biochemical techniques (**Russel et al., 2009**; **Zhao et al., 2010**; **Arkhipov et al., 2013**; **Cheng et al., 2017**; **Latorraca et al., 2017**). However, due to the very long time-scale of conformational changes, brute-force simulations are challenging in the case of large molecular motors such as UvrD. Here, we employed a novel computational approach which integrates advanced sampling simulations with bioinformatics tools that survey structural information from homologs. We were able to identify modes of motions for function switching from principal component analysis of a "trajectory" derived from the alignment of various

surveyed crystal structures. Using the first two principal components as reaction coordinates, the subsequent all-atom Hamiltonian replica exchange simulations (totaling 12μs) predict a metastable "tilted" conformation, which has significantly lower free energy than the apo state. The lowest free energy path is determined to describe the transition between the closed state to the "tilted" state. After the closed-to-tilted transition takes place, 2B and 1B domains are separated with enough distance from each other to enable strand-switching to happen. We demonstrate that ssDNA can be disengaged from the ssDNA binding domains of UvrD in the tilted state. Furthermore, the tilted UvrD structure is shown to be able to form stable interactions with the opposing strand after ssDNA strand switching has occurred. We also highlight the role of the GIG motif in assisting 2B domain diffusion along dsDNA during strand-switching. These findings suggest principles underlying mechanisms of related molecular machines beyond what we have known from existing structures.

The properties obtained from the transition pathway are consistent with the single-molecule data (***Comstock et al., 2015***) as well as mutagenesis studies (***Meiners et al., 2014***). Firstly, we carried out equilibrium simulations of UvrD site-specifically labeled with FRET dye pair AlexaF555/AlexaF647 for both the closed state and the tilted state. The calculated average FRET efficiencies for the two states are in good agreement with those for the unwinding and rezipping states measured in single-molecule experiments, respectively. These simulations also allow us to obtain key fluorophore conformations in the tilted state to explain the shape of the experimental FRET distribution. Secondly, we illustrate the molecular basis for hyper-helicase activity of a UvrD double mutant (D403A/D404A) for the first time. Finally, a physical model integrating the simulation results and the measured equilibrium constant from optical tweezers experiments is provided to explain the helicase function-switching mechanism.

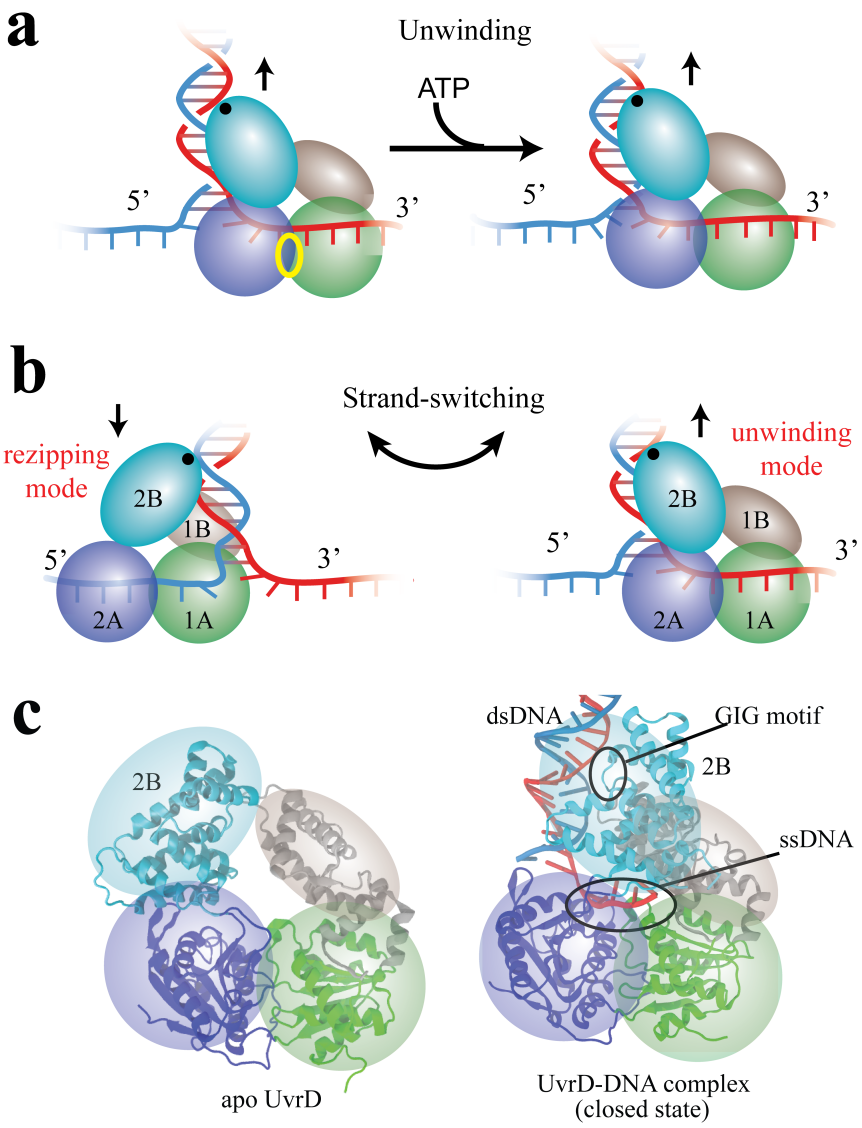


Figure 1. a. Schematic illustration of unwinding by a DNA helicase (UvrD). The helicase uses chemical energy from ATP (hydrolyzed at the site labeled with a yellow ellipse between two motor domains) to unwind dsDNA. b. A proposed model of UvrD functional switching at the fork junction. 1A, 2A, 1B and 2B domains are labeled in green, blue, gray and cyan, respectively. The UvrD conformation on the left represents the rezipping state whereas the conformation on the right represents the unwinding state. GIG motif (residues 414 to 422), drawn as a black dot, is important for UvrD interacting with dsDNA. c. Crystal structures for apo UvrD (3LFU) and UvrD-DNA complex (2IS2, the closed state).

Results

Structural bioinformatics analysis of conformational ensembles of UvrD-like proteins

Our goal is to characterize UvrD conformational changes that switch its function. Recently free energy simulation methods have been successfully applied to study transitions between two functional conformational states of complex molecular machines (***Moradi and Tajkhorshid, 2013***; ***Ma and Schulten, 2015***; ***Czub et al., 2017***). However for UvrD all the known structures bound to the DNA fork junction belong to the closed (unwinding) state (Figure 1c). It is unclear whether the apo state of UvrD could bind to the dsDNA-ssDNA junction. By aligning the apo state to the closed state, we found geometrical clashes between the fork junction and the apo state (Fig.1-figure supplement 1b). We thus forced UvrD at the fork junction to rotate from the closed state to the apo state using targeted molecular dynamics (***Schlitter et al., 1994***) (see Fig.1-figure supplement 1c for details). However, such an operation experienced large resistance (DNA was free to move), and the protein returned back to the vicinity of the closed state after the external force was released. We thus need to find new conformations that can represent the rezipping state.

In order to reach the rezipping state while bound to the fork junction, UvrD must reach some hidden metastable states, which can be far away from the 2B-domain-rotation pathway around the dsDNA axis. To identify such states, we developed an approach based on surveying the pdb database (details in Methods). We used protein-protein BLAST (basic local alignment search tool) to search the swissport database with the UvrD sequence as the query sequence. Then we downloaded the pdb files of these homologs with 40% or more sequence identity. A subsequent principal component analysis (PCA) was carried out to find out the most significant degrees of structural variations among UvrD and its homologs. The coordinates of the homolog structures were then projected onto the first two principal components (PC1 and PC2) (Figure 2). Three distinguishable clusters are shown in Figure 2: one represents the canonical closed conformation, one represents the canonical apo state, and another one represents an interesting structure (from the replication initiator protein) in which the 2B domain is tilted from the dsDNA axis. All the structures belonging to the apo state are without nucleic acids bound. The structure in the "tilted" cluster only has ssDNA bound, and thus very likely it is not a functional state of UvrD because of the absence of dsDNA interactions. To characterize the functional state for rezipping, we need to carry out all-atom free energy simulations (the next subsection).

We next calculated the so-called involvement coefficients (***Lei et al., 2009***) (ICs), which are often used to show the contribution of individual modes to the overall structural displacement. For the displacement between the closed structure and the tilted structure, the ICs of the first two PCs are very high (see Fig.2-figure supplement 1a), indicating that the first two PCs are sufficient to describe the protein conformational changes based on the available UvrD homolog structures. Directions of motions along the first two PCs are shown on the closed structure (Fig.2-figure supplement 1b). We noted that PC1 is in a similar direction as the rotational movement between the closed and apo states. PC2 represents a tilting motion orthogonal to the rotation. Since the closed-to-apo rotation of the 2B domain cannot bring UvrD to the rezipping state due to steric clashes, we suspect that PC2 might make a very important contribution to UvrD conformational switching when bound to the junction.

Free energy landscape of UvrD conformational ensembles when bound to the fork junction

Based on the information revealed by the PCA analysis, we would like to find the UvrD conformation responsible for the rezipping state when bound to the dsDNA-ssDNA junction. For this purpose, extensive enhanced sampling simulations (12 μs in total) were carried out to characterize the free energy landscape of UvrD conformations and detect any interesting metastable states in it. See Methods for the setup and simulation details.

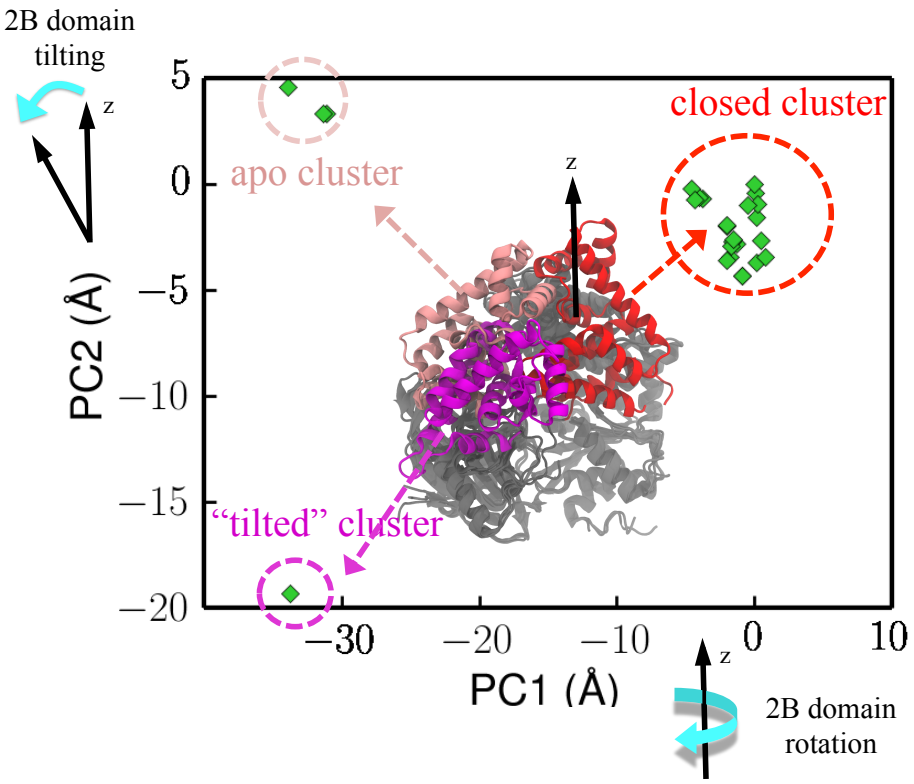


Figure 2. Projection of crystal structures onto the first two PCs from PCA. The structures were obtained through a pdb survey. The structures in the middle of the panel show three structural clusters of the 2B domain, labeled red (closed cluster), magenta (tilted cluster) and pink (apo cluster), respectively. The remaining three domains (1A/2A/2B), which are labeled in gray, only have very small structural variation among the homologs. PC1 represents a rotation motion of the 2B domain around the z-axis, whereas PC2 represents a tilting motion away from the z-axis (see Fig.2-figure supplement 1b).

We first characterized the 2D potential of mean force (PMF) using the first two PCs as coordinates (middle panel of Figure 3). We identified two conformations located in the two local minima of the 2D PMF map, respectively (right and left panels of Figure 3). These two conformations are defined as "closed" and "tilted" states. The tilted state has features that have not been found in any of the existing crystal structures, as we show in the following sections. The PDB file for the newly found tilted state is provided as Supplementary File 1.

The closed and tilted conformations served as the initial and final states for a transition path finding protocol, which was employed to find the lowest free energy path between them (see Methods). The most probable transition happens in two phases, during which the 2B domain undergoes coupled rotational and tilting motions. In the first phase (closed→IM), 2B carries out a large-scale tilting motion along PC2, overcoming a 4.4 kcal/mol barrier before reaching an intermediate state IM. In the second phase (IM→TS→tilted), 2B performs mostly a rotational motion along PC1, overcoming a 1 kcal/mol barrier ($G_{TS} - G_{IM}$) at the global transition state (TS) before reaching the tilted state. Thus the rate-limiting step is the first phase, which involves mostly a tilting motion. Fig.3-figure supplement 1 provides the PMF values and intermediate conformations along the lowest free energy path. Movie S1 shows the conformational changes of UvrD during the transition.

One can notice that the region the apo structure represents has a high energy value, which is more than 8 kcal/mol higher than the initial state. This demonstrates that the apo state, which is connected to the closed state by 2B domain rotation, is very unfavorable at the dsDNA-ssDNA junction.

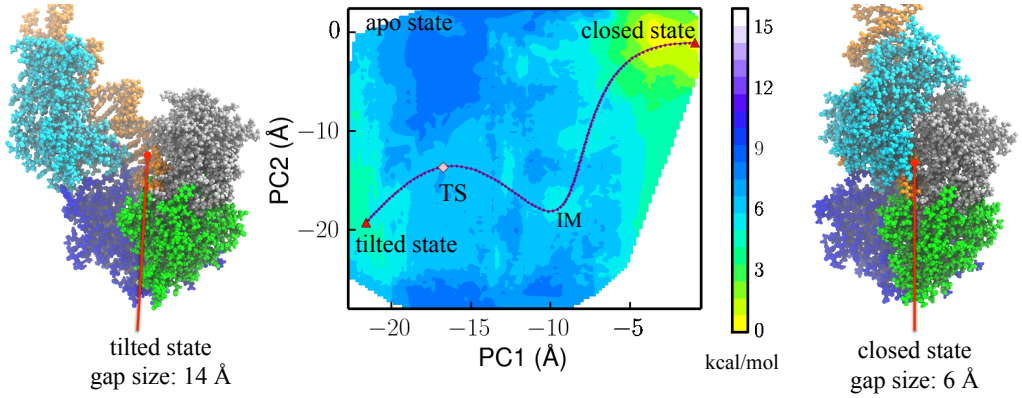


Figure 3. Middle panel: free energy profile generated using the projections along the first 2 PCs. The transition state (TS) and an intermediate state (IM) are located between the closed-to-tilted transition (the predicted tilted state PDB is provided as Supplementary File 1). Right and left panels: snapshots for the closed and tilted states are shown along with the gap size, which is defined by the minimal C_α distance between 2B and 1B domains.

We took the representative protein structure in the final and initial states and measured the gap size, which is defined by the closest C_α atom distance between the 2B and 1B domain. The extended ssDNA has a diameter around 10 Å (***Landy et al., 2013***). The initial closed state has a very small gap size of 6 Å, through which the ssDNA cannot pass. The final tilted state has a gap size of 14 Å, which is open enough for ssDNA to pass through.

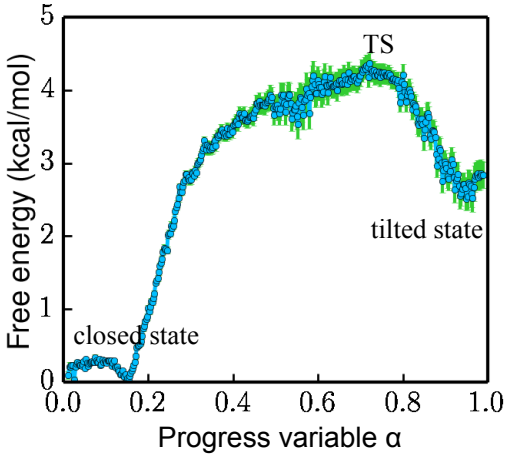


Figure 4. Free energy projected along the progress variable α, of which the value ranges from 0 to 1.0 (the closed state being 0 and the tilted state being 1.0). The standard error is calculated by a bootstrapping error analysis procedure.

The overall free energy landscape projected along a progress variable α is plotted in Figure 4. α is proportional to the projection on PC1 and is scaled from 0 to 1.0 between the closed state and the tilted state. The free energy for the metastable tilted state is about 2.5 kcal/mol higher than that of the closed state. The system has to overcome a 4.2 kcal/mol energy barrier at the transition state (TS) to reach the tilted state.

Validation of the predicted tilted state

We first tested if the ssDNA can escape from the tilted structure. To accelerate the process, we used targeted molecular dynamics by adding a harmonic potential to the coordination number between UvrD and ssDNA. The targeted coordination number was forced to change from an initial value of 18 to 0 in 30 ns. As shown in Movie S2, the ssDNA is seen disengaged from the ssDNA binding

domains of UvrD. The final interaction energy between ssDNA and the 1A/2A/1B domains of UvrD gradually drops to zeros (see Fig.5-figure supplement 4). Further below, we also show that this tilted structure can bind stably to the opposing strand to complete the strand-switching process (Figure 8d).

To quantitatively validate our simulation results against experimental data, we compared the FRET efficiency distributions predicted for the closed and tilted states computationally to those of the functional states measured experimentally. We first obtained the 1D FRET efficiency distributions for the unwinding and rezipping state based on the raw single-molecule data (***Comstock et al., 2015***) (see Methods for details). The distributions, shown in Figure 5a, have peak positions at 0.66 and 0.29 for unwinding and rezipping, respectively. By explicitly simulating UvrD in the two states with fluorophore labels (AlexaFluor555/AlexaFluor647) as in the single-molecule experiments, we also determined FRET efficiencies for the closed and tilted states (Figure 5b). The simulations accumulated 500 ns for each state, and we considered the orientation factor of the fluorophores in determining the FRET efficiency (Methods). The predicted FRET efficiency peak for the closed state is at 0.72, whereas the peak for the tilted state is around 0.31. The close agreement between experimental and simulated FRET distributions reaffirms that the tilted state should be the protein conformation responsible for rezipping. As a control, we simulated the apo-state structure with the fluorophore labels for 500 ns as well. The apo-state FRET distribution, which peaks at 0.16, is quite different from the rezipping-state distribution (Fig.5-figure supplement 1), suggesting that the apo structure is not the conformation for UvrD rezipping at the junction.

We further examined the representative fluorophore pair conformations at the local maxima (FRET Efficiency=0.3 and 0.6) of the tilted state FRET distribution (green curve in Figure 5b). It appears that the fluorophores have different conformations at the two different FRET values (Fig.5-figure supplement 3), due to the conformational dynamics of the dyes with the long linkers. The "shoulder" of the tilted-state FRET distribution curve at 0.6 efficiency is caused by a metastable conformation of AlexaFluor555 with different pair-distance and orientation comparing to the conformation at 0.3 efficiency.

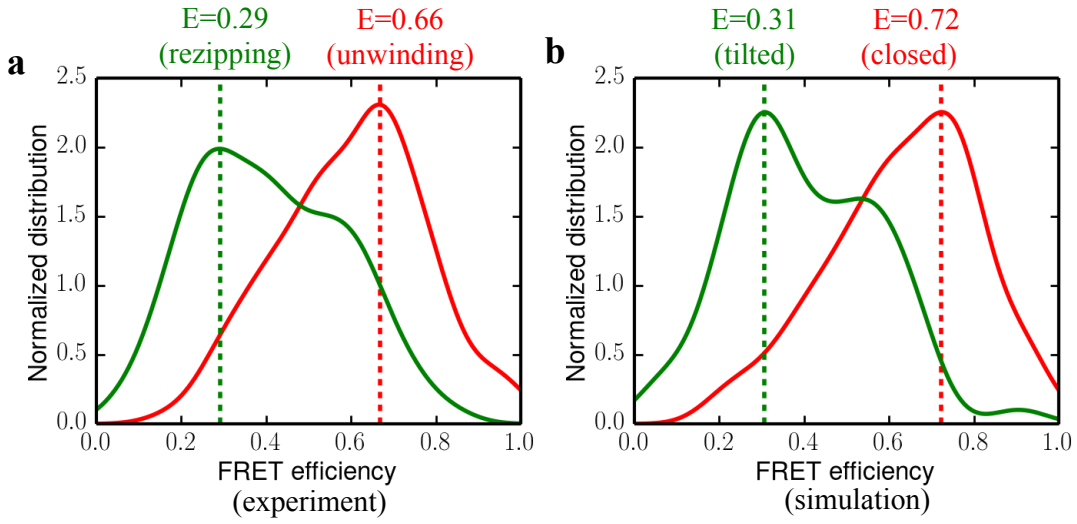


Figure 5. Comparing experimental FRET efficiency distributions to the distributions obtained from simulations. a. Experimental distributions for the unwinding and rezipping states. The dotted lines show the peak positions for the two states. b. Simulated FRET efficiency distribution for the closed and tilted states.

UvrD diffusion along dsDNA

In the UvrD functional switching model, the 2B domain of UvrD has to maintain contact with dsDNA; otherwise the protein might disassociate from the fork junction during the ssDNA strand exchanging.

It is known that the GIG motif (motif IVc) of UvrD plays a key role in interacting with dsDNA (***Myong et al., 2005***), and T422 (a representative residue of GIG) is important for UvrD activity (***Lee and Yang, 2006***). We thus monitored the changes in the interaction between GIG and dsDNA. Figure 6 shows a free energy landscape plotted against the DNA base ID in contact with GIG and the distance between them. For each simulation frame, we calculated the distances between every DNA residue's O2P atom and the OG1 atom of T422. Then the minimal distance and the corresponding DNA base ID were used as the two coordinates. Note the two strands of dsDNA share the same base ID here: for residue x in strand A (indexing according to pdb), the complementary residue in strand B has the same ID x. In the present case, frames with base ID 18 only involve strand A - T422 interaction; whereas frames with base ID 14 only involve strand B - T422 interaction.

In the closed state, residue 18 of strand A contacts the GIG motif, whereas in the tilted state, residue 14 of strand B contacts the GIG motif. Thus, there is a diffusional motion along the dsDNA during the conformational change (see Figure 6). In such a way, UvrD is able to switch the binding dsDNA strand and finds an energetically favorable configuration for the ssDNA strand-switching that will happen in the next step. The diffusion happens in a way that the DNA and T422 are disengaged first, and T422 then re-engages with another DNA residue along the double strand. The base ID in contact with T422 during the transition path from the closed state to the tilted state is shown in Fig.6-figure supplement 1. One can see that the 2B diffusion happens late during the transition. Although UvrD diffuses along dsDNA during the transition, there is no base pair unwound during the closed-to-tilted transition.

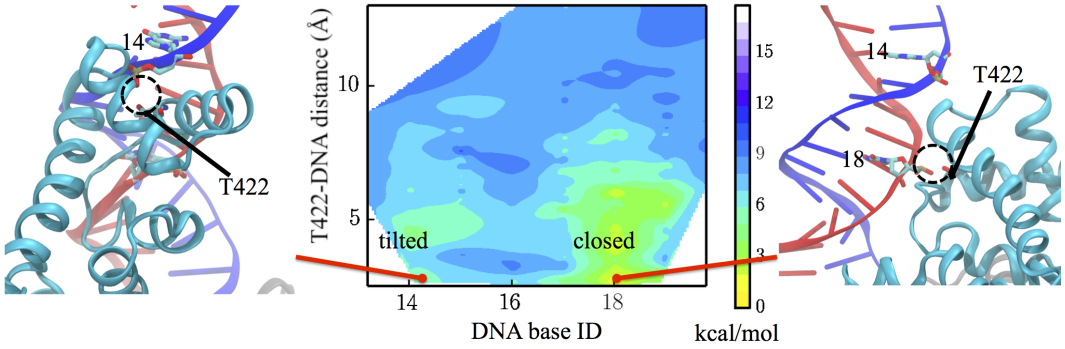


Figure 6. Interaction changes between the GIG motif and dsDNA. Here we use the DNA base ID to represent the closest DNA residue in contact with T422 (part of the GIG motif) on strand A (red) or its complementary residue on strand B (blue). T422 engages with the backbone phosphate of residue 18 of strand A in the closed state whereas it engages with the phosphate of residue 14 of strand B in the tilted state.

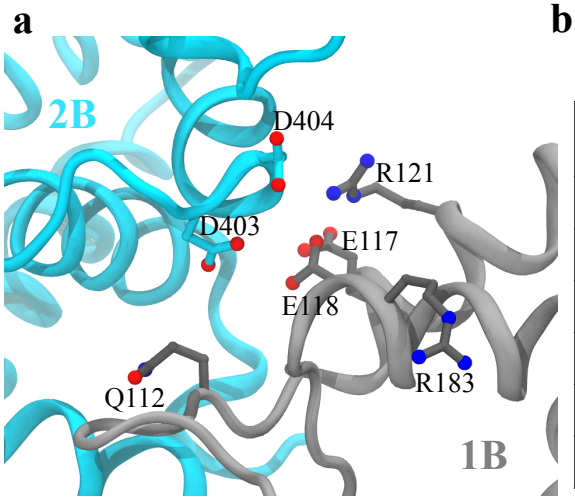
Molecular mechanism for the UvrD303 mutant

Our simulations provide a molecular explanation for the hyper-activity reported for a mutant (UvrD303) that involves two important aspartic acid residues at the 2B-1B interface. Previous experimental work (***Meiners et al., 2014***) discovered that UvrD303 with substitution of two residues, 403 and 404 (both from Asp to Ala), in the 2B domain exhibits a "hyper-helicase" unwinding activity in vitro. The authors suggested that such mutations will reduce the 1B-2B domain interactive contacts and thus yield an intermediate conformation instead of a closed conformation. Such an intermediate state they argued would result in the hyper-activity. However, this explanation is not consistent with the single-molecule measurements (***Comstock et al., 2015***) showing that the closed conformation is responsible for unwinding activity.

To reconcile the conflict, we estimated $\Delta\Delta G_{\text{bind}}$ for the binding free energy between the 1B and 2B domains upon mutating D403 and D404 into alanine, based on our enhanced sampling trajectory. Here $\Delta\Delta G_{\text{bind}} = \Delta G^{\text{mutant}}_{\text{bind}} - \Delta G^{\text{WT}}_{\text{bind}}$, where $\Delta G^{\text{mutant}}_{\text{bind}}$ is the binding free energy for the mutant and $\Delta G^{\text{WT}}_{\text{bind}}$ is that for the wild type. $\Delta\Delta G_{\text{bind}}$ calculated for the closed state is around -2.85 kcal/mol, showing

a stabilization effect of the double alanine mutant. On the other hand, $\Delta\Delta G_{\text{bind}}$ calculated for the tilted state is around 0. This indicates that UvrD303 actually favors the closed conformation and thus will lead to better unwinding activity. The so-called MM/PBSA method (molecular mechanics Poisson-Bolzmann surface area) (***Kollman et al., 2000***; ***Homeyer and Gohlke, 2012***) was used for calculating ΔG_{bind}.

Figure 7a shows the configuration of D403/D404 and key residues on 1B that contribute most significantly to the binding energy change upon the mutation in the closed state. The first five residues on 1B with the largest contribution to $\Delta\Delta G_{\text{bind}}$ are listed in Figure 7b (for the tilted state, all the individual residue contributions to $\Delta\Delta G_{\text{bind}}$ become zero). We noted that there are not many positively charged residues on 1B that are very close to D403/D404. The maximum number of hydrogen bonds formed between D403/D404 and the 1B domain is around two pairs during the simulations. Considering that there are also negatively charged residues of 1B (E118/E117) near D403/D404, mutating the two aspartic acid residues into alanine will not decrease but rather increase the interaction strength between 1B and 2B. We also found that there are significant numbers of nonpolar residues located around residues 403 and 404 (L186, A184, L114, I113, L122). Thus, mutating the two charged residues into hydrophobic residues instead increases the interaction strength between the nonpolar groups and the two alanine residues. Overall, the stabilization of the closed state of UvrD303 leads to consistent unwinding of UvrD helicase, reconciling the biochemical measurement (***Meiners et al., 2014***) with the single-molecule experiment (***Comstock et al., 2015***).



Contribution to $\Delta\Delta G_{\text{bind}}$ from different residues of 1B

Residues of 1B	$\Delta\Delta G_{\text{bind}}$ (kcal/mol)
All	-2.85
R121	0.73
R183	0.66
E117	-0.72
E118	-1.01
Q112	-1.38

Figure 7. a. The configuration of key residues involved in the interaction between D403/D404 (belonging to 2B) and the 1B domain. 2B is shown in cyan whereas 1B is shown in gray. b. A table showing the contribution to $\Delta\Delta G_{\text{bind}}$ from key residues of 1B upon the mutation. Only residues with $|\Delta\Delta G_{\text{bind, x}}| > 0.6$ kcal/mol are shown, where x is the residue index. Positive values indicate destabilization effects of the mutation; negative values indicate stabilization effects.

Discussion

We have characterized the conformational dynamics and a key metastable state of UvrD at a fork junction with a hybrid computational approach. The transition pathway as well as the free energy landscape for UvrD functional switching at the fork junction was obtained, and we found that the opening of the 2B domain involves a major tilting motion followed by a major rotational motion. Diffusion of 2B along the dsDNA happens in the late stage of the transition, during which the GIG motif switches its contact from one strand of dsDNA to the other strand. The transition leads to a gap opening between 2B and 1B, which enables the ssDNA to escape presumably allowing the motor domains to strand-switch.

A physical model for UvrD functional switching

A schematic model can be established based on the simulation results (Figure 8a, b). The corresponding molecular models are shown in Figure 8d. The UvrD functional switching happens in a two-step manner. A first step is the opening of the 2B domain, followed by a second step of the switching of the bound ssDNA strand, in which the original ssDNA disengages from the 1A/2A/1B domain binding site and the other strand fills in.

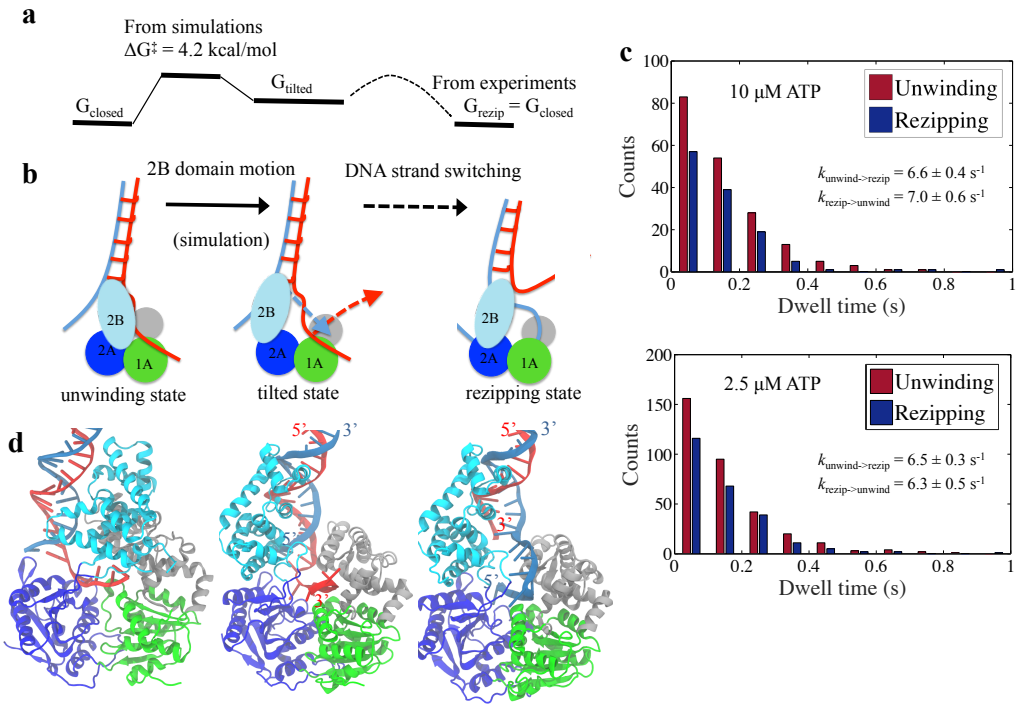


Figure 8. a. Illustration for the whole free energy landscape of ssDNA strand switching enabled by UvrD conformational transition. b. Schematic representation showing the 2-step process of how UvrD switches the ssDNA strand along which the motor domain walks. c. Dwell time distributions for the unwinding and rezipping states at 10 μM and 2.5 μM ATP concentration based on the measured traces from optical tweezers (Methods). d. The structural models for the unwinding, tilted and rezipping states are shown from left to right. The structural model for the rezipping state was obtained from the tilted state after the ssDNA strand switching as illustrated in Methods. Strands A and B of the dsDNA are shown in red and blue, respectively.

To obtain the free energy difference between the unwinding and rezipping states, we performed a dwell time analysis based on past single-molecule measurements (***Comstock et al., 2015***). The dwell times of the unwinding and rezipping states of UvrD monomers are plotted in a histogram and the calculated averaged rates for both transitions are almost equal ($k_{\text{unwind}\to\text{rezip}}$ = 6.6 s^{-1} and $k_{\text{rezip}\to\text{unwind}}$ = 7.0 s^{-1}) at 10 μM ATP (see Figure 8c). Thus the equilibrium constant is around 1 and the unwinding and the re-zipping conformations should have similar free energy. This is consistent

with the picture that the tilted state is a little bit less favorable than the initial state but as soon as the ssDNA releases and the other ssDNA strand binds to the UvrD, the system returns to a lower free energy (the rezipping state) (Figure 8a). For the mutant UvrD303, the free energy for the closed state drops around 3 kcal/mol, whereas G_{tilted} and G_{rezip} remain the same. The relative stabilization of the closed state leads to more persistent unwinding.

The strand switching is mostly driven by Brownian motion and does not require energy from ATP hydrolysis. To address the possible effect of ATP on strand switching, we (1) analyzed additional data from the optical tweezers experiments and compared the switching rate at two different ATP concentrations and (2) also analyzed the x-ray structures of the closed state with and without ATP. The dwell time distributions at 10 μM ATP and 2.5 μM ATP concentration are plotted in Figure 8c. The equilibrium constants of switching measured for the two concentrations are very similar (both around 1), which suggests that strand switching is likely an ATP-independent process. Furthermore, although our simulated system is based on an ATP-free UvrD crystal structure (2IS2), our computational approach covered the structural information from ATP (or its analogs) bound structures (Fig.8-Figure supplement 1). One can see that the 2B motion between the ATP-substrate bound and empty UvrD in the closed state is small relative to the large closed-tilted conformational change. Therefore, it is not very likely that ATP binding has a noticeable impact on the closed-to-tilted transition.

DNA-UvrD conformation at the rezipping state

To explore the structure of the rezipping state further, we built a rezipping structure starting from the tilted conformation after ssDNA strand switching has occurred (see Figure 8d and Methods). After a 100 ns equilibration simulation, the modeled system was stable and had a rmsd around 3Å from the tilted state (Fig.8-Figure supplement 2). The newly obtained rezipping structure satisfies the following considerations: (1) The protein conformation is very similar to ~~the~~ the tilted conformation. (2) The interaction configuration between 2B and dsDNA remains the same between the tilted state and the rezipping state. Note that during our simulations of the closed to tilted transition, the 2B domain changed its contact from one strand of dsDNA to the other (4 bp shift, from the red strand A to the blue strand B in Figure 6). (3) The ATPase domains 1A-2A are in the correct orientation along the ssDNA (3' to 5'), pointing away from the junction. Such a conformation enables UvrD to translocate along ssDNA, allowing the duplex to rezip behind it. In examining the rezipping structure, we found a small loop forming between the dsDNA junction and the ssDNA-1A binding site. A similar feature was proposed by a translocation model of PcrA helicase in ***Park et al.*** (***2010***), which suggested that PcrA can extrude a ssDNA loop while it attaches to dsDNA and translocates the 5' ssDNA tail in an open conformation.

The apo state seen in the crystal structures without DNA bound is likely not a functional state of UvrD at the fork junction. First, the simulated apo-structure FRET distribution is quite different from the rezipping-state FRET distribution. Upon completion of the ssDNA strand-switch, we expect the conformation of UvrD to stay close to the tilted structure. The FRET signal from the single-molecule experiment shows a clear two-state distribution, and our FRET distribution for the tilted state from simulations agrees very well with the experiment. Second, we aligned the apo state structure to the tilted state in Fig.8-Figure supplement 3, and there are serious clashes between the apo structure and the dsDNA. Thirdly, the apo state is highly unfavorable at the fork junction according to our simulations.

Functional insights for UvrD and its homologs

Our simulations, backed by the single-molecule measurements, provide functional insights for UvrD in several important biological processes. For example, frequent strand switching of UvrD due to 2B conformational transition results in unwinding over short distances (***Comstock et al., 2015***), which is consistent with the small number of basepairs unwound during nucleotide excision repair (***Kisker et al., 2013***). On the other hand, UvrD303 is associated with a recombination-deficient

phenotype (***Centore et al., 2009***), possibly due to lacking such a structural transition as the closed state is over-stabilized. It has been reported that UvrD can dismantle RecA filaments from the ssDNA at a stalled replication fork (***Veaute et al., 2005***; ***Lestini and Michel, 2007***). As RecA has a central role in homologous recombination (***Cox, 2007***), a population shift towards the closed state could enhance UvrD's ability to disrupting RecA-ssDNA filaments and impair recombinational repair.

The tilted state and related motions found here can possibly help connect structural information with function for other SF1 helicases. A highly homologous helicase, PcrA, is known to efficiently strip RecA filaments off ssDNA in an "open" conformation (***Park et al., 2010***). The low-FRET "open" conformation of PcrA could be similar to the tilted conformation revealed in this study. In this case, PcrA is anchored to the dsDNA and translocates the 5' ssDNA strand in the direction towards ~~from~~ the junction. A different mode of PcrA is binding to the 3' ssDNA and the dsDNA while unwinding the duplex in the closed form (***Velankar et al., 1999***; ***Niedziela-Majka et al., 2007***). Another UvrD homolog RecB, by mostly tilting its 2B domain from the putative closed state, forms interactions with other subunits in the RecBCD complex (***Singleton et al., 2004***; ***Wilkinson et al., 2016***), which has a key role in initiating recombinational repair (***Spies and Kowalczykowski, 2005***).

It may be possible to engineer UvrD-like helicases with tunable unwinding activities. Experiments have shown that cross-linking Rep and PcrA in the closed form resulted in superhelicase activity (***Arslan et al., 2015***). We demonstrated that mutating the 2 aspartic acid residues into alanine on 2B domain stabilizes the UvrD closed conformation. The ~~contribution~~ analysis of the binding free energy change upon the mutation (Figure 7b) provides potential target residues to guide future experimental designs. For example, mutating some negatively charged residues on 1B might also result in hyper-helicase behavior. Our findings for the conformational dynamics of UvrD and the related computational strategy establish a foundation for future studies to reveal principles employed by other related helicase systems.

Methods and Materials

Structural bioinformatics analysis of UvrD homologs

Our computational study is based on analyzing the structural ensemble of UvrD homologs. There are two representative structures for UvrD: one being the so-called "closed" state (e.g. 2IS2); the other one being the apo state (e.g. 3LFU). As stated in Results, the apo state is likely not a functional structure of UvrD at the DNA fork junction. To explore the conformational space of UvrD as much as possible, we performed a structural survey for possible UvrD homologue conformations using bioinformatics sequence and structure alignment tools (***Altschul et al., 1997***; ***Cock et al., 2009***; ***Bakan and Bahar, 2009***; ***Bakan et al., 2014***). The initial sequence alignments were obtained using NCBI blastp search (***Altschul et al., 1997***) of Protein Data Bank database sequences, with UvrD as the query sequence. Twenty-six structures were selected from the surveyed structures with sequence identity better than 40% and query sequence coverage larger than 60%. The structure alignment was generated by ProDy (***Bakan and Bahar, 2009***; ***Bakan et al., 2014***) from the pairwise sequence alignments by Biopython (***Cock et al., 2009***). The resulting 26 structures can be interpreted in terms of a "trajectory" with the coordinates $\mathbf{r}(k) = (r_1(k), r_2(k), ..., r_{3N}(k))^{\intercal}$, of which each frame k ($k = 1, 2, ..., 26$) contains $3N$ coordinates (from N C_α atoms) of the homologous structures that were mapped onto the original UvrD chain.

We then performed principal component analysis (PCA) (***García, 1992***; ***Bakan and Bahar, 2009***; ***Raveh et al., 2016***) with ProDy to determine a number of modes for reducing the phase space of UvrD motion. Only the C_α coordinates of the 2B domain were used for the PCA calculations, after aligning the 1A/2A/1B domains of all the 26 structures to those of the closed state structure (2IS2). The covariance matrix σ for PCA is determined via $\sigma = \langle (\mathbf{r}(k) - \langle \mathbf{r}(k) \rangle)(\mathbf{r}(k) - \langle \mathbf{r}(k) \rangle)^{\intercal} \rangle$, where the angular brackets $\langle\rangle$ denote the average over k (all the frames). The eigenvectors $\mathbf{v_i}$ (principal components or PCs) of the σ matrix are determined by $\lambda_i \mathbf{v_i} = \sigma \mathbf{v_i}$. These PCs, which are ranked by their corresponding eigenvalues, represent different directions of conformational motion away from the original closed state.

The homologous structures were then projected onto the first two PCs with the largest eigenvalues. As stated in Results, a "tilted" structure based on pdb 1UAA was found as an outstanding cluster among the homologous structures. To see the contributions of different PCs to the displacement between the closed structure and tilted structure, we further calculated the involvement coefficiency η_i (***Ma and Karplus, 1997***; ***Lei et al., 2009***) of the ith PC. η_i is defined as $|\mathbf{v_i} \cdot \Delta\mathbf{R}|$, where $\Delta\mathbf{R}$ is the unit vector describing the displacement from the closed structure to the tilted structure. Only the first two PCs contribute significantly to the overall motion (Fig.2-Figure supplement 1a). PC1 and PC2 are used later as coordinates to compute the free energy landscape.

MD simulation setup

Our simulations were initiated from the closed state (pdb 2IS2) of UvrD (see Figure 1c). The protein-DNA system was solvated in a 100 Å × 100 Å × 130 Å water box with 55 mM NaCl (the system had ∼140K atoms in total). A 2×10^4-step energy minimization was carried out and the system was then heated to 310 K in 30 ps, employing harmonic constraints with 1 kcal/(mol Å^2) spring constant to the C_α atoms. Keeping the spring constant, a 1 ns equilibration in the NPT ensemble (1 atm at 310 K) was performed with a Langevin thermostat for temperature coupling. This was followed by a 1 ns NVT-ensemble simulation, during which the spring constant was gradually decreased to zero. The system was then equilibrated for 60 ns, and the resulting configuration is referred to as the closed state. All MD simulations in our study were performed using NAMD 2.10 (***Phillips et al., 2005***) with the CHARMM36 force field (***Best et al., 2012***; ***Hart et al., 2012***).

Free energy simulation protocol

To determine the free energy profile along a reaction coordinate, we employed the Hamiltonian replica-exchange (HREX) method (***Park et al., 2012***; ***Jiang et al., 2012***, ***2014***). HREX uses a series

of replicas ($j = 1, 2, ..., M$) of the system, which are simulated concurrently with slightly different Hamiltonians and are exchanged frequently among themselves based on the Metropolis exchange criterion (***Sugita et al., 2000***). HREX can be very powerful in reconstructing rugged free energy landscapes by exchanging external biasing potentials, which, with different biasing parameters, are added to the replicas to enhance the sampling throughout the reaction coordinate (RC). The biasing potential (or the window potential) for each replica j usually assumes the form of $U_m(\xi_j) = k_m(\xi_j - p_m)^2/2$, where ξ_j is the current value of the reaction coordinate for replica j, m ($m = 1, 2, ...,$ M) is the index for the biasing potentials (windows), p_m is the preassigned parameter for the center of the harmonic potential, and k_m is the spring constant. The centers of the biasing potentials (p_m) are selected as an ordered list of values ($p_1 < p_2 < ... < p_M$) all over the RC to cover the reaction of interest fully. Exchanges between two neighboring replicas (replicas with neighboring p_m values) are attempted periodically during the simulations. Without the replica-exchange strategy, this protocol reduces to the conventional umbrella sampling, which often suffers from the inefficient sampling of degrees of freedom orthogonal to the reaction coordinate (***Jiang et al., 2012***).

The present study chooses the projection on the first PC ($\mathbf{v_1}$) as the reaction coordinate ξ and includes M = 120 biasing windows between the closed state and the tilted state. The initial configurations for the M windows were generated through a 5 ns targeted MD simulation (***Schlitter et al., 1994***), by driving UvrD from the closed state to the tilted state. The distribution of the obtained initial snapshots was roughly uniform along the first PC. An exchange between two neighboring replicas was attempted every 10 ps and the spring constant of the harmonic potential was set to 100 kcal/(mol Å^2). The production run of each replica lasted 100 ns, and the total simulation time added up to 12 μs (100 ns $\times$ 120). Eventually the weighted histogram analysis method (WHAM) (***Kumar et al., 1992***) was applied to obtain the unbiased 1D and 2D free energy landscapes in Figure 4 and 3. We performed the Monte Carlo bootstrap error analysis (***Stine, 1989***; ***Hub et al., 2010***) to estimate the uncertainty along the reaction coordinate. The basic idea of bootstrapping is to obtain several estimates (we obtained 10 trials) for the free energy based on randomly generated subpopulations from the histogram in each window. Our simulations with HREX benefitted from a scalable multiple copy algorithm (***Jiang et al., 2014***) which enables simulating hundreds of replicas simultaneously on a petascale supercomputer.

As stated in Results, the tilted state structure was identified as one of the most important metastable states. Based on the free energy landscape using the projections on the first two PCs, the lowest free energy path describing the most probable reaction mechanism was localized between closed state and the tilted state using the optimization algorithm in ***Ensing et al.*** (***2005***). The path was then smoothed and 120 images were chosen uniformly along the 2D pathway applying the curve-fitting protocol in ***Ma and Schulten*** (***2015***).

FRET efficiency calculation based on simulations with dye molecules

To check if the simulated closed and tilted states generate the FRET signals of the respective unwinding and rezipping states measured by the single-molecule experiments, we carried out equilibrium simulations with the actual dye molecules for both states. AlexaFluor555 and AlexaFluor647 maleimides (Molecular Probes, Eugene, OR) were modeled according to ***Vrljic et al.*** (***2010***) and ***Gust et al.*** (***2014***) (see Fig.5-Figure supplement 2). Then the two dyes were, respectively, attached to UvrD residues 473 and 100, which were mutated to cysteine from alanine. Force field parameters for the dyes linked to a cysteine residue were obtained from the CHARMM General Force Field (CGenFF) (***Vanommeslaeghe et al., 2010***) using the ParamChem server. The total charges were set to 0 and -3 for the two dyes respectively (***Gust et al., 2014***). Partial charges on the atoms were further refined by the Force Field Toolkit (ffTK) (***Mayne et al., 2013***) in VMD (***Humphrey et al., 1996***). Parameters for bonds, angles and dihedrals from CGenFF with high penalty scores were validated or refined by ffTK.

To sample dye dynamics efficiently, we launched 50 independent standard MD simulations with random initial velocity seeds for the closed, tilted and apo states. Every single simulation lasted

10 ns and a total 500 ns simulation time was accumulated for each state.

The FRET efficiency was determined by $E = R_0^6/(R^6 + R_0^6)$, where R is the distance between the donor and acceptor, and R_0 is the Föster radius (or the 50% energy transfer distance). R_0 is given by the relationship (***Wu and Brand, 1994***) $R_0 = (8.79 \times 10^{-5}\, n^{-4} \phi_D J \kappa^2)^{1/6}$, where n is the index of refraction, ϕ_D is the donor quantum yield, J is the spectral overlap integral, and κ^2 is the orientation factor. R_0 is determined to be 51 Å when κ^2 equals 2/3, assuming that the dyes randomize their orientations by rapid diffusion prior to energy transfer. Such an assumption can be problematic, and in the present study the orientation factor is calculated using $\kappa^2 = (\cos\theta_T - 3\cos\theta_D \cos\theta_A)^2$, where θ_T is the angle between the donor and acceptor transition dipole moments and θ_D and θ_A are the angles between these two dipoles and the vector connecting the donor and acceptor (***Corry and Jayatilaka, 2008***). The transition dipole moments for AlexaFluor555/647 or very similar dyes have been determined in ***Corry and Jayatilaka*** (***2008***) and ***Graen*** (***2009***). The simulated FRET data were integrated to 4 ns per point to obtain its probability distribution using the density kernel estimation method (***Parzen, 1962***).

Analysis of single-molecule data

To validate our simulation results, analysis based on the raw data from single-molecule experiments (***Comstock et al., 2015***) was carried out. ***Comstock et al.*** (***2015***) combined optical tweezers (to detect UvrD unwinding activity) and single-molecule FRET (to detect UvrD conformation) measuring both simultaneously. Example raw time traces of UvrD activity and conformation are shown in Fig. 3 and Fig. S5 in (***Comstock et al., 2015***) (at 10 μM ATP concentration). Time traces from the optical tweezers were sampled at 267 Hz. Time traces for donor and acceptor intensities were integrated to 30-60 ms per data point. The time-dependent FRET efficiency $E(t)$ was calculated by $E(t) = 1/(1+\gamma(I_D(t)/I_A(t)))$ (***Ha et al., 1999***; ***Choi et al., 2010***), where $I_D(t)$ and $I_A(t)$ are the measured donor and acceptor intensities, and γ is a correction factor accounting for the different detection efficiencies for the two dyes, and can be measured from photobleaching events. $\gamma = \Delta I_A/\Delta I_D$ is determined to be 0.78 from 20 acceptor photobleaching events, where ΔI_A and ΔI_D are the acceptor and donor intensity changes upon acceptor photobleaching, respectively.

To measure the FRET efficiency distribution for the unwinding and rezipping states individually, we needed to assign each raw data point to the two states separately. Since the helicase velocity and FRET efficiency were measured concurrently, we used helicase velocity to define whether each data point in the traces belonged to unwinding or rezipping states (see Fig. S5 in ***Comstock et al.*** (***2015***)). Time intervals were determined during which the helicase was either unwinding, rezipping, or paused (positive velocity indicates UvrD is in the unwinding state; negative velocity indicates rezipping; absolute unwinding velocity smaller than 20 bp/s indicates a pause). Paused states were not considered in the analysis. FRET efficiencies over each time interval were collected for the unwinding and rezipping states from 141 time intervals (13 molecules in total). We then used the density kernel estimation method to obtain the experimental FRET distribution (***Parzen, 1962***; ***Comstock et al., 2015***)). A density kernel plot is a summation of small Gaussians centered at each FRET data point. We used a standard deviation of 0.06 for the Gaussians.

We also analyzed the dwell times for both the unwinding state (high FRET) and the rezipping state (low FRET) of UvrD monomers. For this purpose, the duration of each time interval defined above was measured using the traces from optical tweezers measurements. We chose to select intervals and calculate the dwell time using the tweezers signal because it has a higher time resolution than the FRET signal (about one order of magnitude higher). The dwell time distribution was obtained by histogramming the collected duration values for the unwinding and rezipping state separately. In order to assess the effect of ATP concentration on UvrD functional switching, we analyzed optical tweezers data of UvrD activity at two different ATP concentrations. Figure 8c plots the distributions of dwell times at 10 μM and 2.5 μM ATP concentration. The rates of the transitions were estimated by the inverse of the averaged dwell times.

Modeling of the rezipping-state structure

To construct a structure of the rezipping state (after ssDNA strand switching) starting from the tilted conformation (Figure 8d), we consider the following constraints: (1) the 2B domain maintains its contact with the dsDNA while the ssDNA binding domains (1A and 2A) disassociate from one ssDNA strand and bind to the other ssDNA strand. Otherwise, the entire protein would dissociate from DNA. (2) The interaction configuration between ssDNA and the motor domains (1A-2A) must remain the same after strand switching. The motor domains move from 3′ to 5′ on the ssDNA in both the unwinding and rezipping modes. With these considerations, we created a structural model in which we repositioned the UvrD-bound ssDNA segment from the 3′ ssDNA tail (strand A) to the 5′ ssDNA tail (strand B) of the junction in the tilted state. This was achieved by attaching the 5′ terminus of strand B to the 3′ terminus of strand A, and by cutting the ssDNA (strand A) at the junction position. We then equilibrated the modeled system in a water box with 55 mM NaCl for 100 ns.

Acknowledgments

The present contribution is dedicated to Klaus Schulten (1947-2016), whose visionary developments in high performance simulation tools for molecular dynamics permit characterizing biomolecular motor action at biologically relevant timescales. This work has been supported by grants from the NIH (9P41GM104601 to Z.L.-S and K.S.; R01 GM120353 to Y.R.C.) and from the NSF PHY-1430124 (Center for the Physics of Living Cells to Y.R.C., Z.L.-S. and K.S.). The authors gladly acknowledge supercomputer time provided by the Texas Advanced Computing Center via Extreme Science and Engineering Discovery Environment grant NSF-MCA93S028 and the Blue Waters sustained-petascale computing project, which is supported by NSF (OCI-0725070 and ACI-1238993) and the State of Illinois. We also thank former Chemla lab member Dr. Matthew J. Comstock for scientific discussions.

Additional files

Movie S1: A movie showing the changes in the molecular structure along the optimal transition path from the closed state to the tilted state.

Movie S2: A movie showing how the ssDNA disengages from its binding domains of UvrD in the tilted state.

Supplementary File 1: The PDB file for the predicted tilted-state structure of the UvrD-DNA complex.

Supplementary File 2: CHARMM36 force field parameters for the fluorophore molecules used in the simulations (AlexaFluor555 and AlexaFluor647).

References

Altschul SF, Madden TL, Schaffer AA, Zhang JH, Zhang Z, Miller W, Lipman D. Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. Nucleic Acids Res. 1997; 25:3389–3402.

Arkhipov A, Shan Y, Das R, Endres NF, Eastwood MP, Wemmer DE, Kuriyan J, Shaw DE. Architecture and Membrane Interactions of the EGF Receptor. Cell. 2013; 152(3):557–569.

Arslan S, Khafizov R, Thomas CD, Chemla YR, Ha T. Engineering of a superhelicase through conformational control. Science. 2015; 348(6232):344–347.

Bakan A, Bahar I. The intrinsic dynamics of enzymes plays a dominant role in determining the structural changes induced upon inhibitor binding. Proc Natl Acad Sci USA. 2009; 106(34):14349–14354.

Bakan A, Dutta A, Mao W, Liu Y, Chennubhotla C, Lezon TR, Bahar I. Evol and ProDy for bridging protein sequence evolution and structural dynamics. Bioinformatics. 2014; 30(18):2681.

Best RB, Zhu X, Shim J, Lopes PEM, Mittal J, Feig M, MacKerell AD. Optimization of the Additive CHARMM All-Atom Protein Force Field Targeting Improved Sampling of the Backbone ϕ, ψ and Side-Chain $\chi 1$ and $\chi 2$ Dihedral Angles. J Chem Theor Comp. 2012; 8(9):3257–3273. PMID: 23341755.

Brendza KM, Cheng W, Fischer CJ, Chesnik MA, Niedziela-Majka A, Lohman TM. Autoinhibition of *Escherichia coli* Rep monomer helicase activity by its 2B subdomain. Proc Natl Acad Sci USA. 2005; 102:10076–10081.

Brosh RM. DNA helicases involved in DNA repair and their roles in cancer. Nat Rev Cancer. 2013 August; 13(8):542–558.

Centore RC, Leeson MC, Sandler SJ. UvrD303, a Hyperhelicase Mutant That Antagonizes RecA-Dependent SOS Expression by a Mechanism That Depends on Its C Terminus. Journal of Bacteriology. 2009; 191(5):1429–1438.

Cheng MH, Torres-Salazar D, Gonzalez-Suarez AD, Amara SG, Bahar I. Substrate transport and anion permeation proceed through distinct pathways in glutamate transporters. eLife. 2017 jun; 6:e25850.

Choi UB, Stop P, Vrljic M, Chu S, Brunger A, Weninger KR. Single-molecule FRET-derived model of synaptotagmin 1-SNARE fusion complex. Nat Struct Mol Biol. 2010; 17:318–331.

Cock PJA, Antao T, Chang JT, Chapman BA, Cox CJ, Dalke A, Friedberg I, Hamelryck T, Kauff F, Wilczynski B, de Hoon MJL. Biopython: freely available Python tools for computational molecular biology and bioinformatics. Bioinformatics. 2009; 25(11):1422.

Comstock MJ, Whitley KD, Jia H, Sokoloski J, Lohman TM, Ha T, Chemla YR. Direct observation of structure-function relationship in a nucleic acid-processing enzyme. Science. 2015; 348(6232):352–354.

Corry B, Jayatilaka D. Simulation of Structure, Orientation, and Energy Transfer between AlexaFluor Molecules Attached to MscL. Biophys J. 2008; 95(6):2711 – 2721.

Cox MM. Motoring along with the bacterial RecA protein. Nat Rev Mol Cell Biol. 2007; 8(2):127–138.

Czub J, Wieczór M, Prokopowicz B, Grubmüller H. Mechanochemical Energy Transduction during the Main Rotary Step in the Synthesis Cycle of F1-ATPase. J Am Chem Soc. 2017; 139(11):4025–4034. PMID: 28253614.

Dillingham MS. Superfamily I helicases as modular components of DNA-processing machines. Biochem Soc Trans. 2011; 39(2):413–423.

Dillingham MS, Soultanas P, Wiley P, Webb MR, Wigley DB. Defining the roles of individual residues in the single-stranded DNA binding site of PcrA helicase. Proc Natl Acad Sci USA. 2001; 98(15):8381–8387.

Dillingham MS, Wigley DB, Webb MR. Demonstration of unidirectional single-stranded DNA translocation by PcrA helicase: measurement of step size and translocation speed. Biochemistry. 2000; 39:205–212.

Ensing B, Laio A, Parrinello M, Klein ML. A Recipe for the Computation of the Free Energy Barrier and the Lowest Free Energy Path of Concerted Reactions. J Phys Chem B. 2005; 109(14):6676–6687. PMID: 16851750.

Fischer CJ, Maluf NK, Lohman TM. Mechanism of ATP-dependent translocation of E. coli UvrD monomers along single-stranded DNA. J Mol Biol. 2004; 344:1287–1309.

García AE. Large-amplitude nonlinear motions in proteins. Phys Rev Lett. 1992 Apr; 68:2696–2699.

Graen T. Dye'ing for FRET: The HTH motif in smFRET-MD simulation. Bachelor thesis, Georg-August-Universität Göttingen; 2009.

Gust A, Zander A, Gietl A, Holzmeister P, Schulz S, Lalkens B, Tinnefeld P, Grohmann D. A Starting Point for Fluorescence-Based Single-Molecule Measurements in Biomolecular Research. Molecules. 2014; 19(10):15824–15865.

Ha T, Ting AY, Liang J, Caldwell WB, Deniz AA, Chemla DS, Schultz PG, Weiss S. Single-molecule fluorescence spectroscopy of enzyme conformational dynamics and cleavage mechanism. Proc Natl Acad Sci USA. 1999; 96(3):893–898.

Hart K, Foloppe N, Baker CM, Denning EJ, Nilsson L, MacKerell AD. Optimization of the CHARMM Additive Force Field for DNA: Improved Treatment of the BI/BII Conformational Equilibrium. J Chem Theor Comp. 2012; 8(1):348–362. PMID: 22368531.

Homeyer N, Gohlke H. Free Energy Calculations by the Molecular Mechanics Poisson?Boltzmann Surface Area Method. Molecular Informatics. 2012; 31(2):114–122.

Hub JS, de Groot BL, van der Spoel D. g_wham-A Free Weighted Histogram Analysis Implementation Including Robust Error and Autocorrelation Estimates. J Chem Theor Comp. 2010; 6:3713–3720. doi: 10.1021/ct100494z.

Humphrey W, Dalke A, Schulten K. VMD – Visual Molecular Dynamics. J Mol Graphics. 1996; 14:33–38.

Iyer RR, Pluciennik A, Burdett V, Modrich PL. DNA Mismatch Repair: Functions and Mechanisms. Chem Rev. 2006; 106(2):302–323.

Jia H, Korolev S, Niedziela-Majka A, Maluf NK, Gauss GH, Myong S, Ha T, Waksman G, Lohman TM. Rotations of the 2B Sub-domain of E. coli UvrD Helicase/Translocase Coupled to Nucleotide and DNA Binding. J Mol Biol. 2011; 411(3):633 – 648.

Jiang W, Luo Y, Maragliano L, Roux B. Calculation of Free Energy Landscape in Multi-Dimensions with Hamiltonian-Exchange Umbrella Sampling on Petascale Supercomputer. J Chem Theor Comp. 2012; 8(11):4672–4680.

Jiang W, Phillips J, Huang L, Fajer M, Meng Y, Gumbart J, Luo Y, Schulten K, Roux B. Generalized Scalable Multiple Copy Algorithms for Molecular Dynamics Simulations in NAMD. Comput Phys Commun. 2014; 185:908–916.

Kisker C, Kuper J, Van Houten B. Prokaryotic Nucleotide Excision Repair. Cold Spring Harbor Perspectives in Biology. 2013; 5(3).

Kollman PA, Massova I, Reyes C, Kuhn B, Huo S, Chong L, Lee M, Lee T, Duan Y, Wang W, Donini O, Cieplak P, Srinivasan J, Case DA, Cheatham TE. Calculating Structures and Free Energies of Complex Molecules:? Combining Molecular Mechanics and Continuum Models. Acc Chem Res. 2000; 33(12):889–897. PMID: 11123888.

Kumar S, Bouzida D, Swendsen RH, Kollman PA, Rosenberg JM. The Weighted Histogram Analysis Method for Free-Energy Calculations on Biomolecules. I. The Method. J Comp Chem. 1992; 13:1011–1021.

Landy J, Lee Y, Jho Y. Limiting law excess sum rule for polyelectrolytes. Phys Rev E. 2013 Nov; 88:052315.

Latorraca NR, Fastman NM, Venkatakrishnan AJ, Frommer WB, Dror RO, Feng L. Mechanism of Substrate Translocation in an Alternating Access Transporter. Cell. 2017 2017/07/07; 169(1):96–107.e12.

Lee JY, Yang W. UvrD helicase unwinds DNA one base pair at a time by a two-part power stroke. Cell. 2006; 187:1349–1360.

Lei M, Velos J, Gardino A, Kivenson A, Karplus M, Kern D. Segmented Transition Pathway of the Signaling Protein Nitrogen Regulatory Protein C. J Mol Biol. 2009; 392(3):823 – 836.

Lestini R, Michel B. UvrD controls the access of recombination proteins to blocked replication forks. EMBO J. 2007; 26(16):3804–3814.

Lohman TM, Tomko EJ, Wu CG. Non-hexameric DNA helicases and translocases: mechanisms and regulation. Nat Rev Mol Cell Biol. 2008; 9:391–401.

Ma J, Karplus M. Ligand-induced conformational changes in ras p21: a normal mode and energy minimization analysis. J Mol Biol. 1997; 274(1):114 – 131.

Ma W, Schulten K. Mechanism of substrate translocation by a ring-shaped ATPase motor at millisecond resolution. J Am Chem Soc. 2015; 137:3031–3040.

Matson SW. Escherichia coli helicase II (urvD gene product) translocates unidirectionally in a 3' to 5' direction. J Biol Chem. 1986; 261(22):10169–10175.

Mayne CG, Saam J, Schulten K, Tajkhorshid E, Gumbart JC. Rapid parameterization of small molecules using the Force Field Toolkit. J Comp Chem. 2013; 34:2757–2770.

Mechanic LE, Hall MC, Matson SW. Escherichia coli DNA Helicase II Is Active as a Monomer. J Biol Chem. 1999; 274(18):12488–12498.

Meiners MJ, Tahmaseb K, Matson SW. The UvrD303 Hyper-helicase Exhibits Increased Processivity. J Biol Chem. 2014; 289(24):17100–17110.

Moradi M, Tajkhorshid E. Mechanistic picture for conformational transition of a membrane transporter at atomic resolution. Proc Natl Acad Sci USA. 2013; 110(47):18916–18921.

Myong S, Rasnik I, Joo C, Lohman TM, Ha T. Repetitive shuttling of a motor protein on DNA. Nature. 2005; 322:1321–1325.

Nguyen B, Ordabayev Y, Sokoloski JE, Weiland E, Lohman TM. Large domain movements upon UvrD dimerization and helicase activation. Proc Natl Acad Sci USA. 2017; 114(46):12178–12183.

Niedziela-Majka A, Chesnik MA, Tomko EJ, Lohman TM. Bacillus stearothermophilus PcrA monomer is a single-stranded DNA translocase but not a processive helicase in vitro. J Biol Chem. 2007; 282:27076–27085.

Park J, Myong S, Niedziela-Majka A, Lee KS, Yu J, Lohman TM, Ha T. PcrA Helicase Dismantles RecA Filaments by Reeling in DNA in Uniform Steps. Cell. 2010; 142(4):544–555.

Park S, Kim T, Im W. Transmembrane Helix Assembly by Window Exchange Umbrella Sampling. Phys Rev Lett. 2012 Mar; 108:108102.

Parzen E. On Estimation of a Probability Density Function and Mode. Ann Math Statist. 1962 09; 33(3):1065–1076.

Patel SS, Donmez I. Mechanisms of Helicases. J Biol Chem. 2006; 281(27):18265–18268.

Phillips JC, Braun R, Wang W, Gumbart J, Tajkhorshid E, Villa E, Chipot C, Skeel RD, Kale L, Schulten K. Scalable Molecular Dynamics with NAMD. J Comp Chem. 2005; 26:1781–1802.

Raveh B, Karp JM, Sparks S, Dutta K, Rout MP, Sali A, Cowburn D. Slide-and-exchange mechanism for rapid and selective transport through the nuclear pore complex. Proc Natl Acad Sci USA. 2016; 113(18):E2489–E2497.

Russel D, Lasker K, Phillips J, Schneidman-Duhovny D, Velázquez-Muriel JA, Sali A. The structural dynamics of macromolecular processes. Curr Opin Cell Biol. 2009; 21(1):97 – 108. Cell structure and dynamics.

Sancar A. DNA excision repair. Annu Rev Biochem. 1996; 65:43–81.

Schlitter J, Engels M, Krüger P. Targeted Molecular Dynamics: A New Approach for Searching Pathways of Conformational Transitions. J Mol Graph. 1994; 12:84–89.

Singleton MR, Dillingham MS, Wigley DB. Structure and mechanism of helicases and nucleic acid translocases. Annu Rev Biochem. 2007; 76:23–50.

Singleton MR, Dillingham MS, Gaudier M, Kowalczykowski SC, Wigley DB. Crystal structures of RecBCD enzyme reveals a machine for processing DNA breaks. Nature. 2004; 432:187–193.

Spies M, Fishel R. Mismatch Repair during Homologous and Homeologous Recombination. Cold Spring Harbor Perspectives in Biology. 2015; 7(3).

Spies M, Kowalczykowski SC. Homologous Recombination by the RecBCD and RecF Pathways. American Society of Microbiology; 2005.

Stine R. An Introduction to Bootstrap Methods. Sociological Methods & Research. 1989; 18(2-3):243–291.

Sugita Y, Kitao A, Okamoto Y. Multidimensional replica-exchange method for free-energy calculations. The Journal of Chemical Physics. 2000; 113(15):6042–6051.

690 **van Brabant AJ**, Stan R, Ellis NA. DNA helicases, genomic instability, and human genetic disease. Annu Rev
691 Genomics Hum Genet. 2000; 1:409–459.

692 **Vanommeslaeghe K**, Hatcher E, Acharya C, Kundu S, Zhong S, Shim J, Darian E, Guvench O, Lopes P, Vorobyov I,
693 MacKerell Jr AD. CHARMM General Force Field: A Force Field for Drug-Like Molecules Compatible with the
694 CHARMM All-Atom Additive Biological Force Fields. J Comp Chem. 2010; 31(4):671–690.

695 **Veaute X**, Delmas S, Selva M, Jeusset J, Cam EL, Matic I, Fabre F, Petit MA. UvrD helicase, unlike Rep helicase,
696 dismantles RecA nucleoprotein filaments in Escherichia coli. EMBO J. 2005; 24:180–189.

697 **Velankar SS**, Soultanas P, Dillingham MS, Subramanya HS, Wigley DB. Crystal structures of complexes of PcrA
698 DNA helicase with a DNA substrate indicate an inchworm mechanism. Cell. 1999; 97:75–84.

699 **Vrljic M**, Strop P, Ernst JA, Sutton RB, Chu S, Brunger AT. Molecular mechanism of the synaptotagmin-SNARE
700 interaction in Ca2+-triggered vesicle fusion. Nat Struct Mol Biol. 2010 03; 17(3):325–331.

701 **Wilkinson M**, Chaban Y, Wigley DB. Mechanism for nuclease regulation in RecBCD. eLife. 2016; 5:e18227.

702 **Wu CG**, Spies M. Overview: What Are Helicases? New York, NY: Springer New York; 2013.

703 **Wu PG**, Brand L. Resonance Energy Transfer: Methods and Applications. Analytical Biochemistry. 1994; 218(1):1
704 – 13.

705 **Yang W**. Lessons Learned from UvrD Helicase: Mechanism for Directional Movement. Annual Review of
706 Biophysics. 2010; 39(1):367–385.

707 **Zhao Y**, Terry D, Shi L, Weinstein H, Blanchard SC, Javitch JA. Single-molecule dynamics of gating in a neurotrans-
708 mitter transporter homologue. Nature. 2010; 465(7295):188–193.

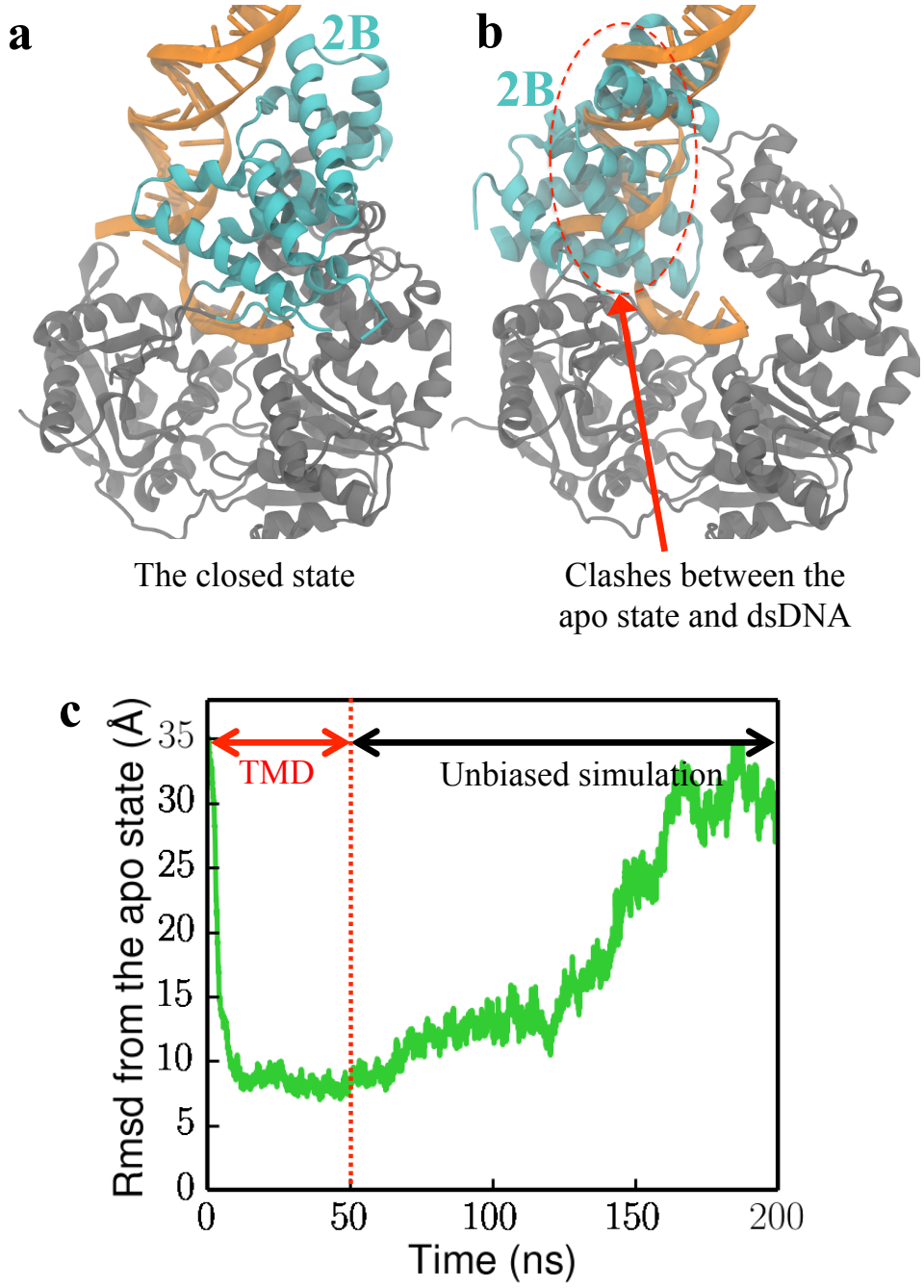


Figure 1–Figure supplement 1. UvrD is found unlikely to adopt the apo conformation when bound to a fork junction. a. The closed-state structure of UvrD (2IS2). The 2B domain is shown in cyan whereas the remaining 1A/2A/1B domains are shown in gray. The fork junction is shown in orange. b. The apo-state structure (3LFU) is aligned to the closed-state structure. Major clashes are detected between 2B of the apo state and the fork junction. c. In a forced rotation simulation of the 2B domain at the DNA junction (DNA is free to move), the change of RMSD from the apo state (calculated using C_α atoms of the 2B domain) is shown. The plot shows a two-stages simulation: starting from RMSD = 35 Å, the first stage is a targeted MD simulation driving UvrD to the apo state from the closed state in 50 ns; the second stage is an unbiased simulation which lasts 150 ns. Eventually UvrD returns to the closed state, indicating that the apo state is very unfavorable.

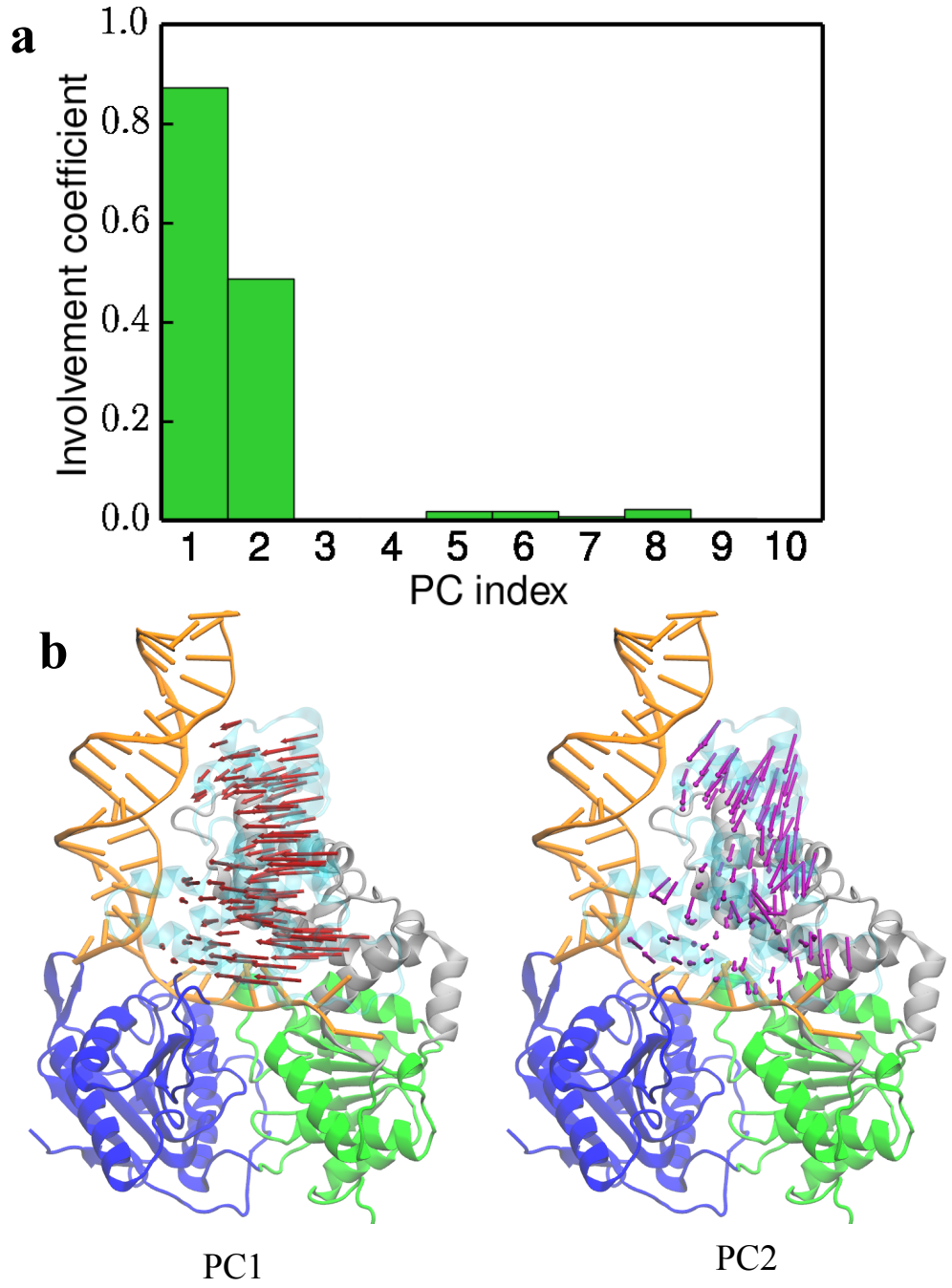


Figure 2–Figure supplement 1. a. Involvement co-efficiency of the first 10 PCs for the displacement between the closed cluster and the tilted cluster. b. Directions of motions along the first two PCs shown on the closed state structure.

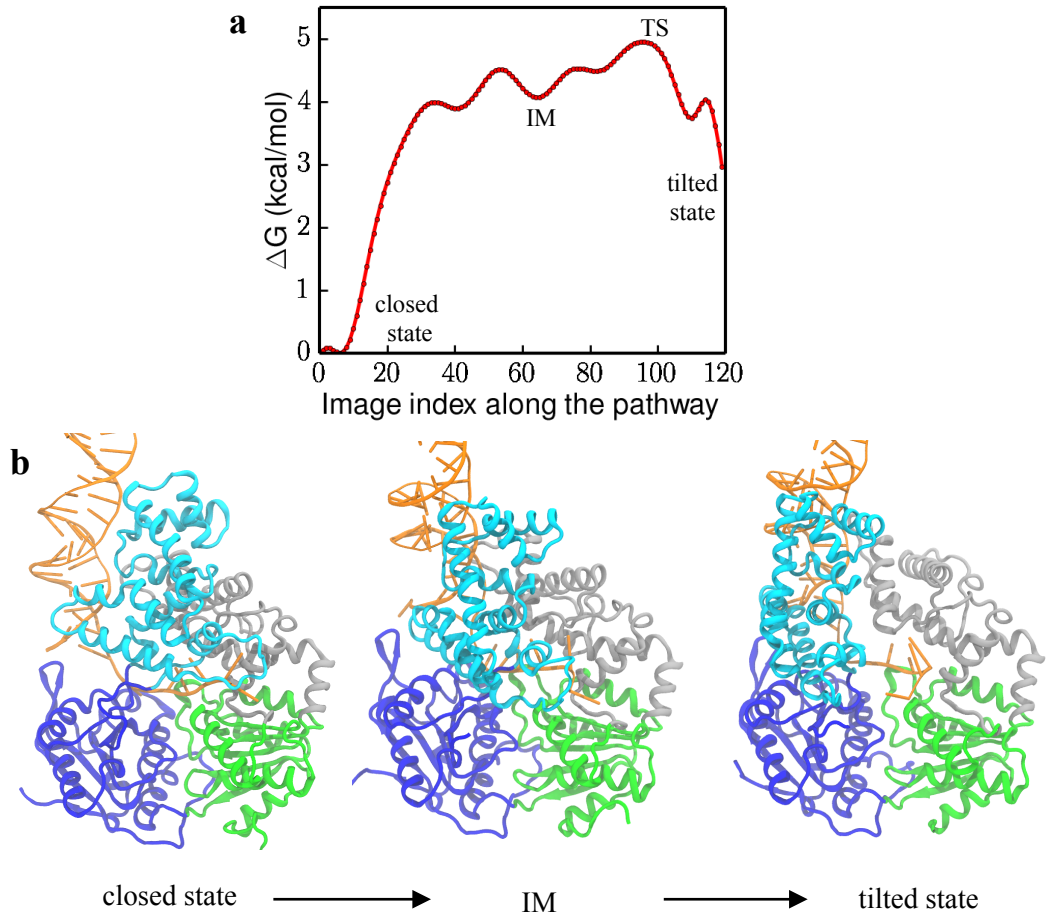


Figure 3–Figure supplement 1. a. 1D free energy profile along the lowest free energy path (LFEP) (Figure 3). The x-axis is the index of the 120 data points (images) along the LFEP. The positions of the transition state TS and the intermediate state IM are labeled in the figure. b. Conformations of UvrD-DNA complex during the closed to tilted transition.

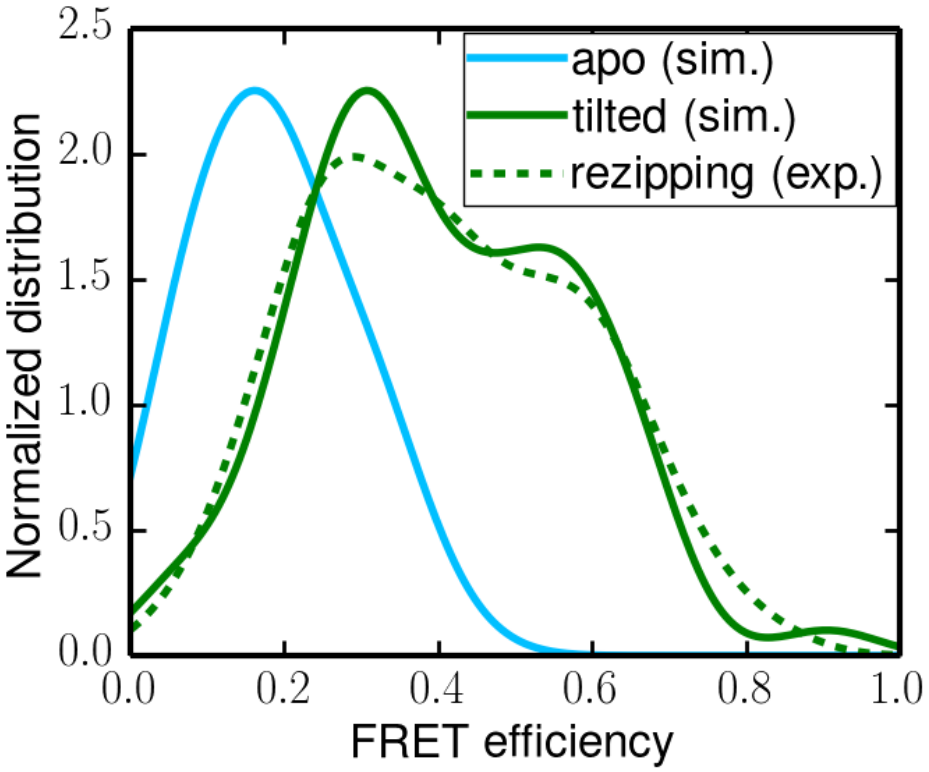


Figure 5–Figure supplement 1. Simulated apo-state FRET distribution (solid cyan curve) and its comparison with the distributions for the rezipping state (dotted green curve, from experiments) and the tilted state (solid green curve, from simulations).

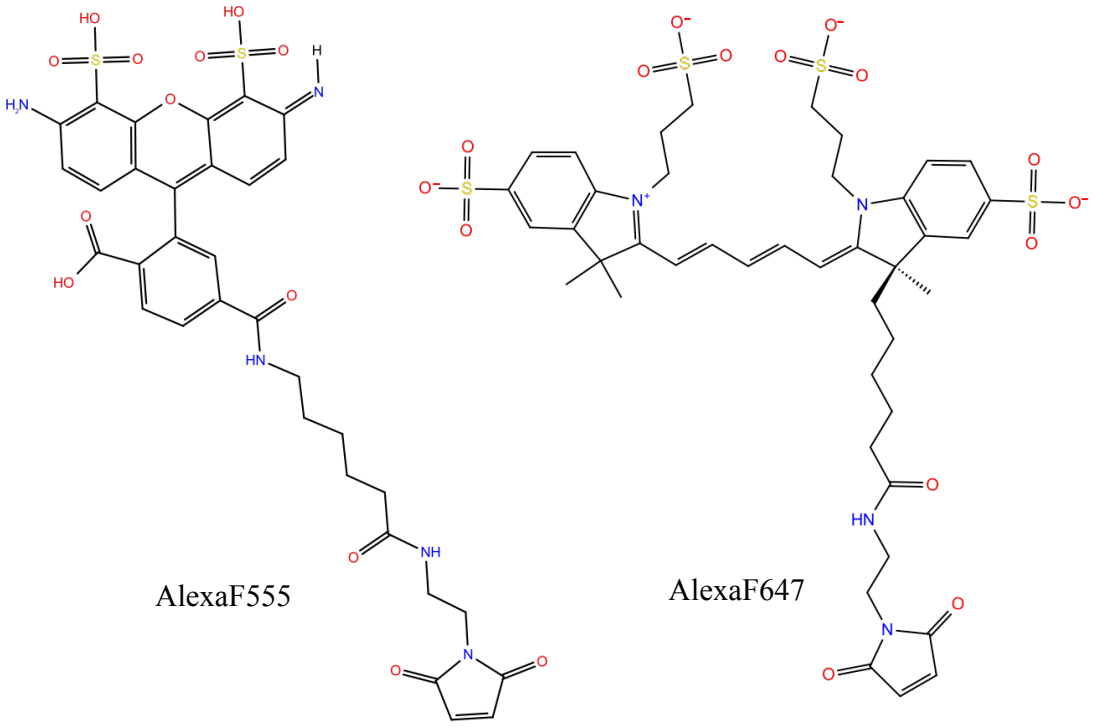


Figure 5–Figure supplement 2. Molecular structures of the dyes (AlexaF555/AlexaF647) used in the simulations.

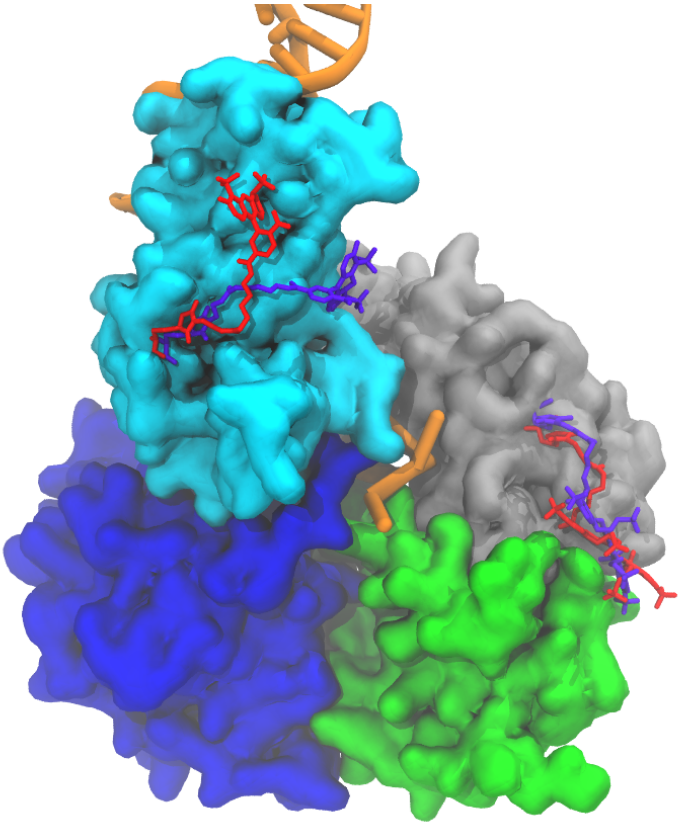


Figure 5–Figure supplement 3. Averaged fluorophore pair conformations in the tilted state at different FRET efficiency values. The averaged conformation of the pair at 0.3 FRET efficiency is shown in red. The averaged conformation of the pair at 0.6 FRET efficiency is shown in purple.

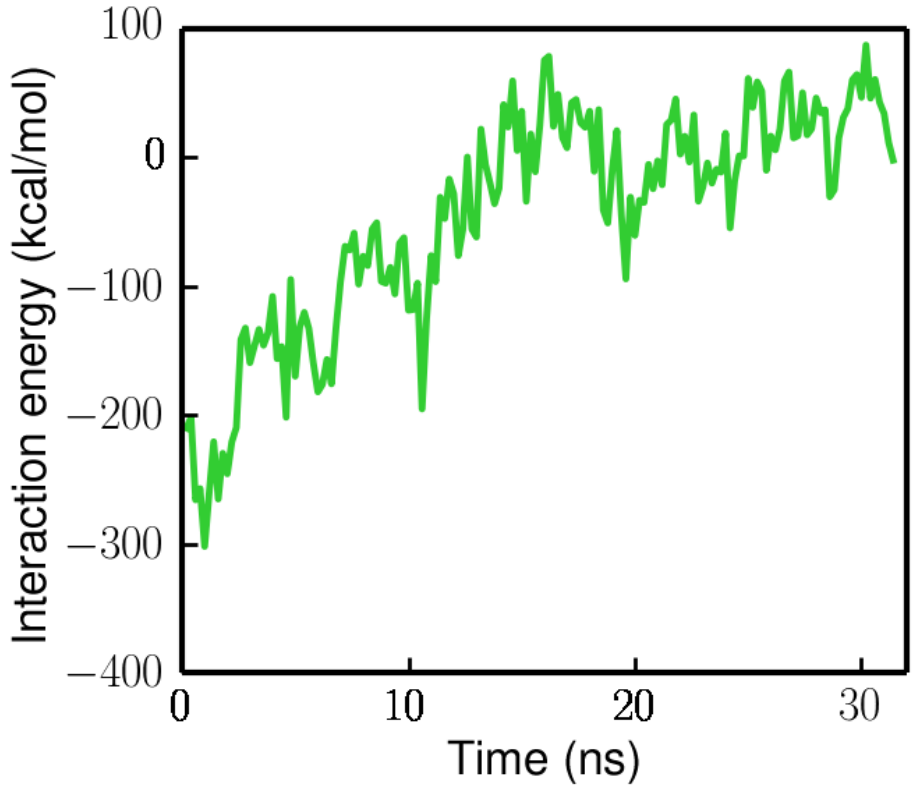


Figure 5–Figure supplement 4. Interaction energy (electrostatic + Van der Waals) between the ssDNA and the ssDNA-binding domains (1A/2A/1B) during the ssDNA disengagement simulation (the protein is kept in the tilted state).

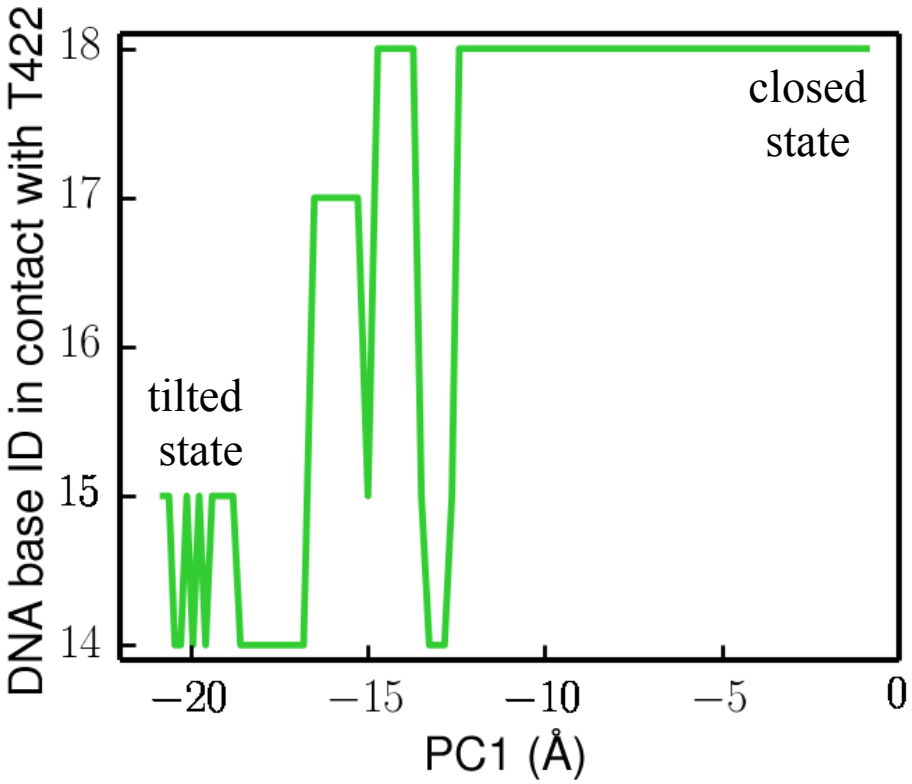


Figure 6–Figure supplement 1. Evolution of the DNA base ID in contact with the GIG motif during the transition from the closed state to the tilted state (along the lowest free energy path).

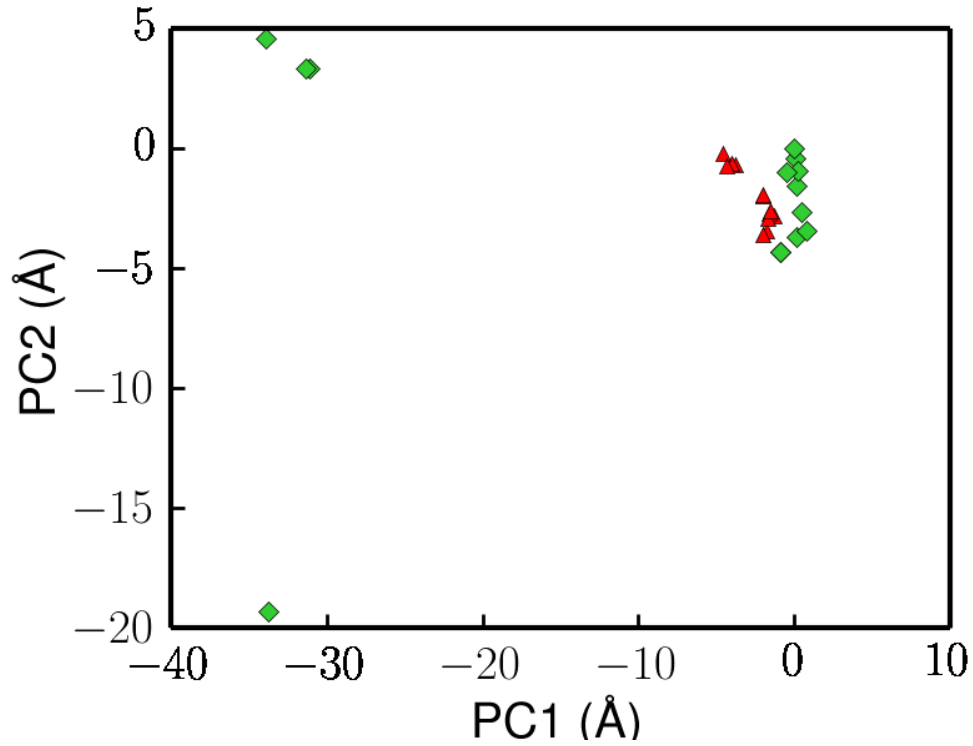


Figure 8–Figure supplement 1. Projection of crystal structures onto the first two principal components from PCA (same data points used in Figure 2). Here, the UvrD structures with ATP or non-hydrolysable analog molecules bound are colored in red.

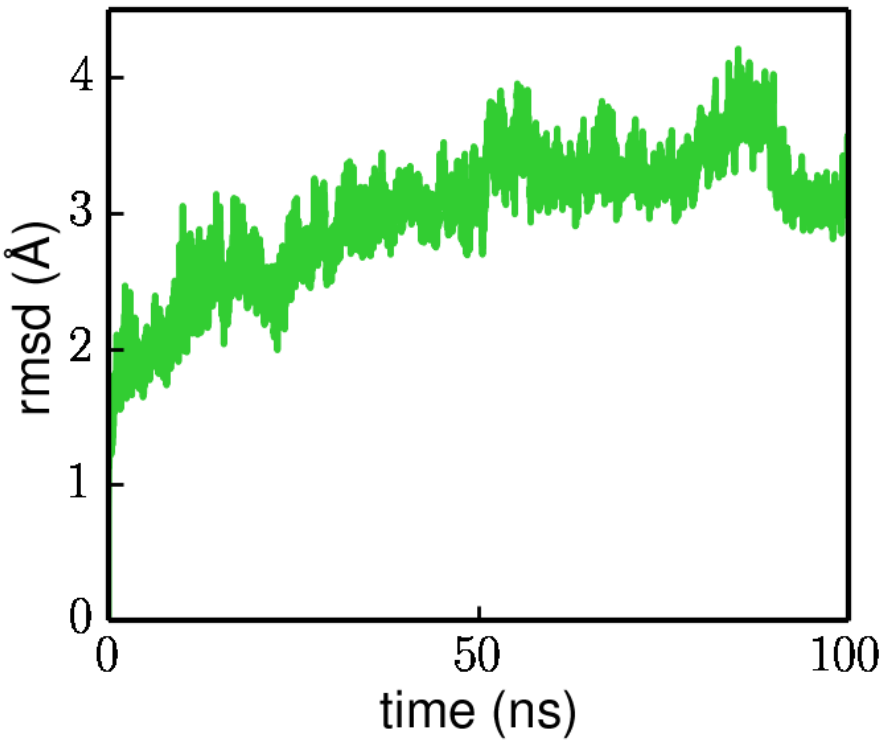


Figure 8–Figure supplement 2. Rezipping-state RMSD (calculated using the protein non-hydrogen atoms) from the tilted state during the 100 ns equilibration simulation.

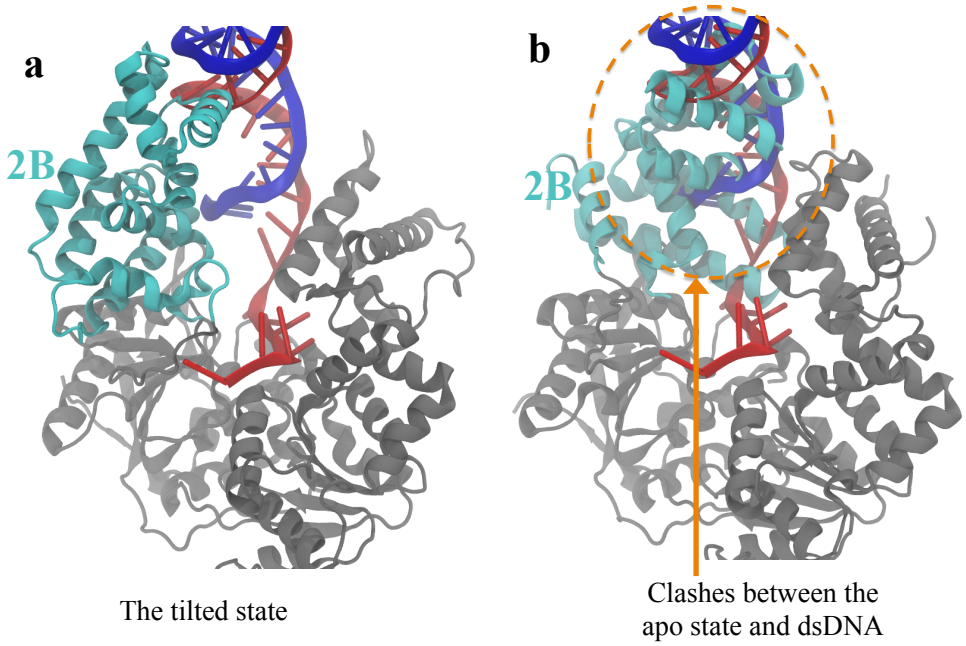


Figure 8–Figure supplement 3. a. The tilted conformation obtained from the simulations. The 2B domain is shown in cyan whereas 1A/2A/1B are shown in gray. Strands A and B of the dsDNA are shown in red and blue, respectively. b. After aligning the apo structure to the tilted structure, geometric clashes between the apo 2B domain and the dsDNA of the tilted state are detected (shown in the orange dotted circle).