



Computationally exploring the mechanism of bacteriophage T7 gp4 helicase translocating along ssDNA

Shikai Jin (金石开)^{a,b}, Carlos Bueno^b, Wei Lu^{b,c}, Qian Wang^d, Mingchen Chen^e, Xun Chen^{b,f}, Peter G. Wolynes^{a,b,c,f,1}, and Yang Gao^{a,1}

Edited by Gerhard Hummer, Max Planck Institute for Biophysics, Frankfurt am Main, Germany; received February 8, 2022; accepted July 5, 2022 by Editorial Board Member Kiyoshi Mizuuchi

Bacteriophage T7 gp4 helicase has served as a model system for understanding mechanisms of hexameric replicative helicase translocation. The mechanistic basis of how nucleoside 5'-triphosphate hydrolysis and translocation of gp4 helicase are coupled is not fully resolved. Here, we used a thermodynamically benchmarked coarse-grained protein force field, Associative memory, Water mediated, Structure and Energy Model (AWSEM), with the single-stranded DNA (ssDNA) force field 3SPN.2C to investigate gp4 translocation. We found that the adenosine 5'-triphosphate (ATP) at the subunit interface stabilizes the subunit-subunit interaction and inhibits subunit translocation. Hydrolysis of ATP to adenosine 5'-diphosphate enables the translocation of one subunit, and new ATP binding at the new subunit interface finalizes the subunit translocation. The LoopD2 and the N-terminal primase domain provide transient protein-protein and protein-DNA interactions that facilitate the large-scale subunit movement. The simulations of gp4 helicase both validate our coarse-grained protein-ssDNA force field and elucidate the molecular basis of replicative helicase translocation.

helicase | gp4 | motor proteins | coarse-grained model

Helicases are nucleotide triphosphatase (NTPase)-coupled motors that travel along DNA or RNA (1). Helicases play important roles in many physiological processes including genomic DNA replication. Replicative helicases run at the forefront of the replication fork and separate the double-stranded (ds) parental DNA into two single-stranded (ss) daughter strands, which then serve as templates for DNA synthesis (2, 3). Moreover, helicases are organization hubs for DNA replication by physically interacting with DNA polymerases, primases, ssDNA binding proteins, and adaptor proteins. During their operations, replicative helicases encircle one of the daughter strand ssDNA along which they translocate and sterically exclude the other strand to drive strand separation (2, 3). According to their conserved sequence motifs, helicases can be classified into six superfamilies (SF), with SF1 and SF2 monomeric and SF3 to SF6 hexameric (1). Replicative helicases are hexameric and belong to SF3, SF4, and SF6 families. Helicases in bacteria, bacteriophage, and mitochondria belong to the SF4 family along with RecA-like ATPase domains and display 5'-3' polarity, while archaeal and eukaryotic SF6 helicases and viral SF3 helicase have AAA+ ATPase domains and display 3'-5' polarity in their translocation. The structures and mechanisms of hexameric helicase translocation have been extensively studied (2, 3). The homo- or heterohexamers assemble into ring or lockwasher shapes with coiled ssDNA within the central channel. One or two DNA binding loops from the six subunits form a staircase that holds the DNA backbone (4–10). Each subunit in SF3 E1 and SF5 Rho helicases binds one nucleotide, while each subunit from the SF4 and SF6 helicases holds two nucleotides. The DNA binding loops take on distinct conformations in SF3 and SF5 helicases to form a staircase along the DNA backbone. In contrast, the DNA binding loops are rigid in SF4 and SF6 helicases. NTPase sites are located at each subunit interface. Biochemical and single-molecule studies have suggested that NTPs are hydrolyzed sequentially within the helicase hexamer and only one NTPase site fires at a time (11–13). Consistent with that idea, gradual conformational changes of the NTPase sites along the hexameric ring are observed in several helicase-DNA structures, suggesting ordered sequential hydrolysis (4, 5, 7, 8). Taken together, a sequential hand-over-hand mechanism has been proposed for hexameric helicases. An NTPase cycle will drive the DNA binding loop or the subunit at one end of DNA to migrate to the other end so as to form new protein-DNA contacts. Sequential movement of the six subunits enables processive translocation along ssDNA. Nevertheless, how the NTPase cycle is coupled to translocation is unknown, and how a subunit or DNA binding loop migrates a long distance to reach the distal DNA end is unclear.

Molecular dynamics (MD) simulations can give insights about dynamic molecular processes that are challenging to obtain using purely experimental methods. Because of

Significance

Hexameric helicase is a central component in DNA replication and many replicative helicases have been proposed as drug targets for disease treatment. Knowledge of both the structures of replicative helicases and their mechanisms of translocation is essential for a better understanding of DNA replication and replication stress response, as well as for therapeutic targeting of helicases. The very large size of the helicase and the large-scale conformational changes that occur during helicase translocation have hindered investigations of helicase structure and dynamics. Simulations with our newly developed coarse-grained protein-single-stranded DNA force field recapitulated the large-scale translocation of the gp4 subunit with limited computational cost and revealed numerous mechanistic details about gp4 helicase translocation.

Author contributions: S.J., P.G.W., and Y.G. designed research; S.J. performed research; S.J., C.B., W.L., and P.G.W. contributed new reagents/analytic tools; S.J., Q.W., M.C., X.C., P.G.W., and Y.G. analyzed data; and S.J., P.G.W., and Y.G. wrote the paper.

The authors declare no competing interest.

This article is a PNAS Direct Submission. G.H. is a guest editor invited by the Editorial Board.

Copyright © 2022 the Author(s). Published by PNAS. This open access article is distributed under Creative Commons Attribution-NonCommercial-NoDerivatives License 4.0 (CC BY-NC-ND).

¹To whom correspondence may be addressed. Email: pw8@rice.edu or yg60@rice.edu.

This article contains supporting information online at <http://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2202239119/-DCSupplemental>.

Published August 1, 2022.

the large size of the helicase–DNA complex and the lack of proper force fields for protein–DNA complexes, there have been only a handful of attempts to simulate the helicase translocation process. Coarse-grained simulations have been carried out for SF3 E1 helicase, hepatitis C virus helicase, and the multimeric ATPase chaperonin GroEL (14–16). In another study on LTag helicase, Langevin dynamics simulation has been applied to investigate the protein–DNA interaction in SF3 simian virus 40 helicase (17). However, the coarse-grained DNA models employed in these studies lack the physical benchmark of the ssDNA model and the protein–DNA interactions. Recently, an all-atom simulation on SF5 Rho has revealed how the ATPase cycle is coupled to the transitions of the DNA binding loops (18). So far, there have been no simulation analyses on any SF4 and SF6 helicase family members, which are the major replicative helicases for all three domains of life. Moreover, the DNA conformations and the DNA–protein interactions in the SF4 and SF6 helicases are distinct from those for the SF3 and SF5 helicases. Translocation of SF4 and SF6 helicases has been proposed to involve large-scale conformational changes of an entire subunit, which are absent for the SF3 and SF5 helicases (4, 7, 8).

The replicative system from bacteriophage T7 provides a model system for studying DNA replication. T7 gp4 encodes a dual functional protein with primase on its N-terminal domain (NTD) and SF4 helicase on its C-terminal domain. The gp4 helicase exists as heptamers and hexamers in the absence of DNA, with the hexameric form being responsible for DNA unwinding and the heptameric form being possibly responsible for DNA loading (19). In vivo, the gp4 helicase can physically interact with gp5 DNA polymerase and gp2.5 ssDNA binding protein (20). At a replication fork, a single gp4 hexamer and multiple gp5 molecules work cooperatively to catalyze parental DNA unwinding and both leading and lagging strand synthesis (21–23), similar to what happens for other replication systems (24, 25). Recent structures of T7 gp4 with an ssDNA substrate show that the gp4 helicase domain forms a lockwasher-shaped hexamer and interacts with A-form-like ssDNA. The two subunits at the two ends of the hexamer are separated by over 20 Å. The terminal subunit of the lockwasher existed in three distinct conformations, at the 5′-end of DNA, at the 3′-end of DNA, or in the middle, which suggests a subunit translocation pathway. Moreover, the structure suggests that the ATPase site at the 5′-end DNA hydrolyzes ATP first, consistent with the sequential model that has been proposed based on biochemical and single-molecular studies (11, 12).

In this report we construct a hybrid coarse-grained force field for protein–ssDNA complexes by combining the OpenAWSEM (Associative memory, Water-mediated, Structure and Energy Model) model for protein and a modified Open3SPN2 model of the nucleic acid components (26). Simulations of gp4 helicase translocation with our force field reveal that ATP hydrolysis is the key determinant that enables subunit translocation. Moreover, our simulation results capture several intermediate states and identify transient protein–DNA and protein–protein interactions that facilitate the long-distance subunit translocation. In summary, the transferable force field developed here is able to simulate motor translocation with large-scale movement.

Results

The Benchmark of the Weights in the ssDNA Model. The 3SPN.2C force field was developed by the de Pablo group that aimed at reproducing the thermodynamic properties of dsDNA (27). The 3SPN.2C force field has been parameterized based on the physicochemical properties such as the free energy of nucleic acid hybridization, the intrastrand base stacking energy, the DNA persistence length, and the width of minor and major grooves (28). In order to better capture the thermodynamics properties of ssDNA, we modified the Open3SPN2 force field (26). The original base-pairing and cross-stacking terms are adapted in the current force field to fit the potential hairpin interactions between two distant regions of one ssDNA. The cutoffs for both terms were set as 1.2 nm. The remaining energy terms include the bonded terms (bond, angle, and di-hedral) and the nonbonded terms (base-stacking, exclusion, and electrostatics). Since the original weights have already been tuned to fit the DNA backbone properties, we set a series of weights (0.5, 0.8, 1, 1.2, and 1.5) for base-stacking, base-pairing, and cross-stacking terms to achieve the best agreement with experimental persistence length values. The same exclusion and electrostatics terms are kept from the previous OpenAWSEM with Open3SPN2 force field to treat the protein–DNA interaction.

To determine parameters for energy terms, we have benchmarked the persistence length and the melting temperature of ssDNA (Table 1) to make sure they fall within the theoretical estimate(28). We benchmarked the ssDNA persistence length following the same way as in the 3SPN.2C model (27). The persistence length is calculated as $l_p = \langle R_e^2 \rangle / 2L$, where l_p is the persistence length, R_e is the end-to-end distance, and L is the contour length. For benchmarking, we have used the same

Table 1. The thermodynamic data of the benchmark of the ssDNA part of the force field

		Persistence length		
Sequence and length		Ionic strength, mM	Experimental values, nm (32)	Simulation values, nm
Base stacking weight				
0.5	Poly(A) 144 bp	150	2–4	3.45
0.8	Poly(A) 144 bp	150	2–4	3.40
1.0	Poly(A) 144 bp	150	2–4	3.37
1.2	Poly(A) 144 bp	150	2–4	3.32
1.5	Poly(A) 144 bp	150	2–4	3.33
Melting temperature				
Sequence	Length	Ionic strength, mM	Experimental values, K (34)	Simulation values, K
CAGTACAGTATTTTTTACTGTACTG	25	119	336.6	330–335
AGTAGTAATCACACCTTTT–GGTGTGATTACTACT	35	119	343.3	340–345

ssDNA charge spacing (4.3 Å) per base that was used in the benchmark of 3SPN2 for the contour length L . We studied a total of five weight combinations (28). A 144-base-pair (bp) Poly(A) sequence was used to avoid any hairpin formation and the average persistence length for the best weight combination (1.2) is 3.32 nm. The values of the ssDNA persistence length cover a range in the literature due to the use of different experimental techniques, sequences, and thermodynamic conditions (29, 30). The tuned value of 3.32 nm is still in reasonable agreement with the experimental values that were provided in table 2 of the previous 3SPN2.C paper for ssDNA (27, 31, 32). The melting temperature is defined as the temperature where the free energies of the hybridized DNA and dehybridized DNA become equal (33). To sample the melting of ssDNA, we built the DNA hairpin with the sequence given in Table 1 as the initial structures. The simulations with umbrella sampling used the distance between two intermediate points of the chains as the collective variable at 300 K. In the case of the 25-bp sequence, the intermediate points are the centers of mass of the 5th and the 25th nucleotide. In the other case of the 35-bp sequence, the intermediate points were chosen as the center of mass of the 5th and the 30th nucleotides. The free energy profiles were extended every 5 K from 300 K to 350 K and the melting temperature was determined where the free energy in the two basins was equal (*SI Appendix, Fig. S1*). These melting temperatures agree with the experimental values (34). In addition, the presently optimized force field well maintains ssDNA local structures such as ssDNA loop during test runs.

Constructing the gp4-DNA Complex for Coarse-Grained Simulation. The structures of the gp4 helicase-DNA complex with the transition subunit F at two ends of the lockwasher (Protein Data Bank [PDB] IDs: 6n7n and 6n7t) are used as the initial models (*SI Appendix, Fig. S2*) (8). It has been proposed that during the translocation the mobile subunit F will separate from the E-F interface at the DNA 5'-end and translocate to the DNA 3'-end to form new interactions with subunit A. Complete models for each chain in gp4 helicase structures were built using Modeler software (35). Then, a short equilibration simulation with the CHARMM27 force field (36) was applied to align the short N-terminal helix in the Modeler model to the cryogenic electron microscopy (cryo-EM) structure (Fig. 1A). The ssDNA was extended to 50 bp from an ideal B-form DNA generated by Open3SPN2. The ATP molecule is described in coarse-grained fashion based on the same topology from Open3SPN2 (Fig. 1B). The root-mean-square deviation (RMSD) values between the final model and the deposited PDB structures were around 1.1 Å for all chains. The top and the side views of the final model for the initial structures are shown in Fig. 1 C and D.

In addition, we analyzed the all-atom frustration patterns of gp4 dimers with apo, ADP bound, and ATP bound at the subunit dimer interfaces (37). Naturally occurring proteins generally have a smooth, funnel-like energy landscape with minimal kinetic traps, promoting robust and rapid folding into a single native structure (38). Frustration in the proteins occurs when a molecule is unable to simultaneously achieve a minimum energy for each interaction individually. Quantifying the degree

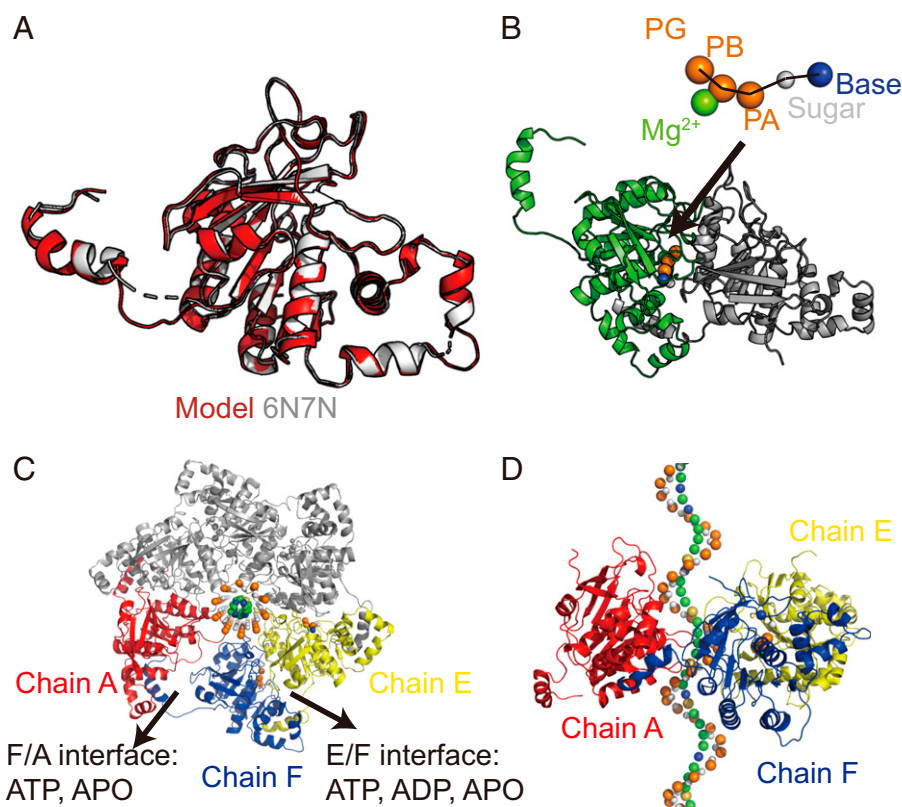


Fig. 1. The structures used for simulation. (A) Model of a gp4 helicase subunit (shown as chain F of 6n7n). White color shows the cryo-EM structure. Red color shows our final model after the energy minimization by NAMD. (B) The coarse-grained ATP molecule is shown at the interface of two chains. A zoom-in view of the ATP molecule is shown in the upper right corner. (C) The top view of the final homohexamer model built from 6n7n. The chains A, E, and F are colored in red, yellow, and blue, respectively. All other chains are colored in gray. (D) The side view of the final homohexamer model built from 6n7n. The chains B, C, and D are hidden for better visualization.

of local frustration helps us understand the stability of any given residue interaction. Minimally frustrated interactions generally are stable interactions, while the highly frustrated interactions are typically unstable (39). Frustration analysis reveals that the subunit–subunit interface is unstable in the apo states due to unfavorable interactions among positively charged side chains K520, R522, and R504 (*SI Appendix, Table S1*). The addition of the ADP to the subunit interface helps stabilize the interface by increasing the number of minimally frustrated interactions compared to apo form (*SI Appendix, Fig. S3*). Compared with ADP, ATP contributes two more minimally frustrated interactions and reduces five highly frustrated interactions. The negatively charged triphosphate groups may balance the positively charged active site (23).

ATP Hydrolysis Assists the Initialization of Translocation. To probe the mechanism of chemomechanical coupling, a total of six sets of simulations were designed to delineate the roles of ATP binding, hydrolysis, and release during a subunit translocation. We designed ATP, ADP, and apo forms in the chain E/F interface and ATP or apo forms in the chain F/A interface (Fig. 1C). An interpolated 50-frame trajectory from the two initial states before and after the translocation was generated. Umbrella sampling was used to capture the large-scale

conformational changes for the free energy landscape analysis. Each structure was individually equilibrated and then treated as the input for umbrella sampling.

The one-dimensional (1D) free energy profiles of EFAPO_AFATP, EFADP_AFATP, and EFATP_AFATP simulations are shown in Fig. 2A. We noticed there are several intermediate basins located at the Q diff value around 0.55, 0.45 to 0.5, and 0.3 to 0.4. Furthermore, we have extended the 1D free energy profile to two dimensions with an additional axis using the largest principal component of the motions in a separate very long EFADP_AFATP trajectory (details in *Materials and Methods*). We divided the 1D free energy profile into three segments based on the corresponding two-dimensional (2D) free energy profile, including local basins I and EX1, IM1 and IM2, and IM3 and F, respectively. These local basins are generally arranged in the position from the lower-right corner to the upper-left corner. The sampling of I, IM2, IM3, and F is consistent among the different sets of simulations, as exemplified by comparing representative IM2 structures from different simulations (*SI Appendix, Fig. S4 and Table S2*). In contrast, the local basins before the translocation starts as IM1 with given suffixes in the different simulations. The translocation step always happens between the IM1 and IM2 basins.

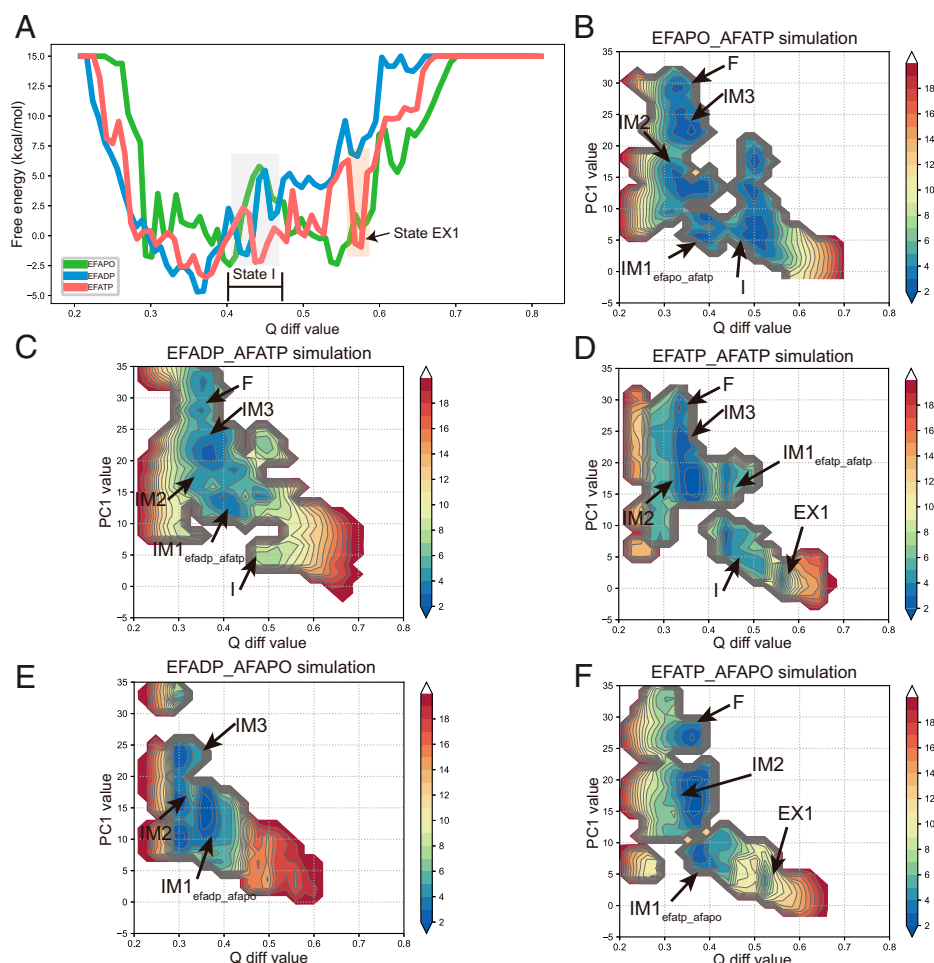


Fig. 2. The 1D and 2D free energy profiles of helicase translocation simulations. (A) The 1D free energy profile of the three different states of the ATP forms in the chain E/F interface. (B–F) The 2D free energy profile of EFAPO_AFATP, EFADP_AFATP, EFATP_AFATP, EFADP_AFAPO, and EFATP_AFAPO simulation. The x axis is the value of Q diff, for which the value 1 indicates the structure is the same as the initial state and 0 indicates the structure is the same as the final state. The y axis uses the largest PC generated from PC analysis of a very long EFADP_AFATP simulation. The color bars on the trajectories of the free energy profiles are in units of kilocalories per mole. The state I indicates “initial”, IM indicates “intermediate”, and “F” indicates final. The state I has Q diff values ranging from 0.45 to 0.55. The states IM2, IM3, and F are located at similar Q diff values between 0.3 and 0.4 but they have different PC1 values at 15 to 18, 22 to 26, and 28 to 32, respectively.

We noticed that adding an ATP molecule to the chain E/F interface created a local basin with an energy barrier around 7 kcal/mol, indicating that the ATP molecule prevents the contact breaking of chain E/F. Comparing the 2D free energy profiles from these simulations (Fig. 2 *B–D*), we found a local basin termed “EX1” that only exists in the EFATP simulations. A representative structure of the EX1 basin (*SI Appendix, Fig. S5*) suggested there are close contacts between chain F and chain E, indicating that the presence of ATP is detrimental to subunit translocation. In contrast, it is easier to bypass this initial energy trap when only an ADP molecule is bound but an additional local basin appears at a Q diff value around 0.43. In the EFAPD simulations, the free energy is smooth after crossing the large energy barrier between state I and IM1_{efapo_afatp}. It is likely that the ADP molecule is released before the translocation starts. The state IM1_{efapo_afatp} is the closest to the intermediate state captured in cryo-EM structure 6n7s among all states in EFAPD_AFATP simulation with average Q value of 0.533 (*SI Appendix, Fig. S6*), where subunit F is apo and released from chain E/F interface but has not moved to subunit A yet. The dTTP binding state of 6n7s coincides with our hypothesis that the ATP/ADP is released before the large-scale translocation.

An ATP Molecule Helps Stabilize the New Interface. To probe the role of the new ATP molecule binding at the FA interface, we compared the simulations of AFATP and AFAPD forms during the translocation. We found a high-energy barrier located between the basin IM2 and the basin F when there is no ATP on the FA interface. With an ATP molecule, the transition from IM2 to the final states becomes much easier (Fig. 2 *C* and *D*). We found similar patterns in the 2D free energy profiles computed using a second axis as the ATP binding angle between the arginine finger of chain F and Walker A motif of chain A (*Materials and Methods*). A large binding angle rotation was observed between basin IM2 and basin F in the EFATP_AFATP simulation (Fig. 3*A*). We then checked the electrostatic potential surfaces for selected structures within the IM2 and IM3 basins. The surface electrostatic density indicates there is a positively charged cavity in the binding surface of IM2 basin structure. An ATP molecule binding at that cavity relieves the repulsion due to the positive charges (Fig. 3 *B* and *C*).

Conformational Changes along the Translocation. During the translocation, the general internal structures of each domain are maintained in the simulation, consistent with those observed in cryo-EM (*SI Appendix, Fig. S7*). Here we used the representative structures from EFATP_AFATP trajectory for the analysis of the translocation pathway. The angles between the interacting subunits E–F and F–A were calculated for evaluating the rotational motion during the translocation progress, which can reflect the radial and tangential motions between the two subunits. From the morphing movie based on the representative structures of the intermediate states, we can propose a translocation pathway (Fig. 4). The overall translocation starts from the state I, with chain F undergoing a clockwise rotation of almost 15° relative to chain E. In the presence of an ATP molecule, chain E/F will be held at a “closed” conformation (EX1 basin) with extensive chain E/F interactions bridged by the ATP. ATP hydrolysis triggers the relative rotation of chain E/F toward opposite directions and opens the E/F interface. From IM1_{efatp_afatp} to IM2, subunit F translocates to the new DNA end and starts to contact chain A. As shown in *SI Appendix, Fig. S8*, we also noticed there are transient contacts between

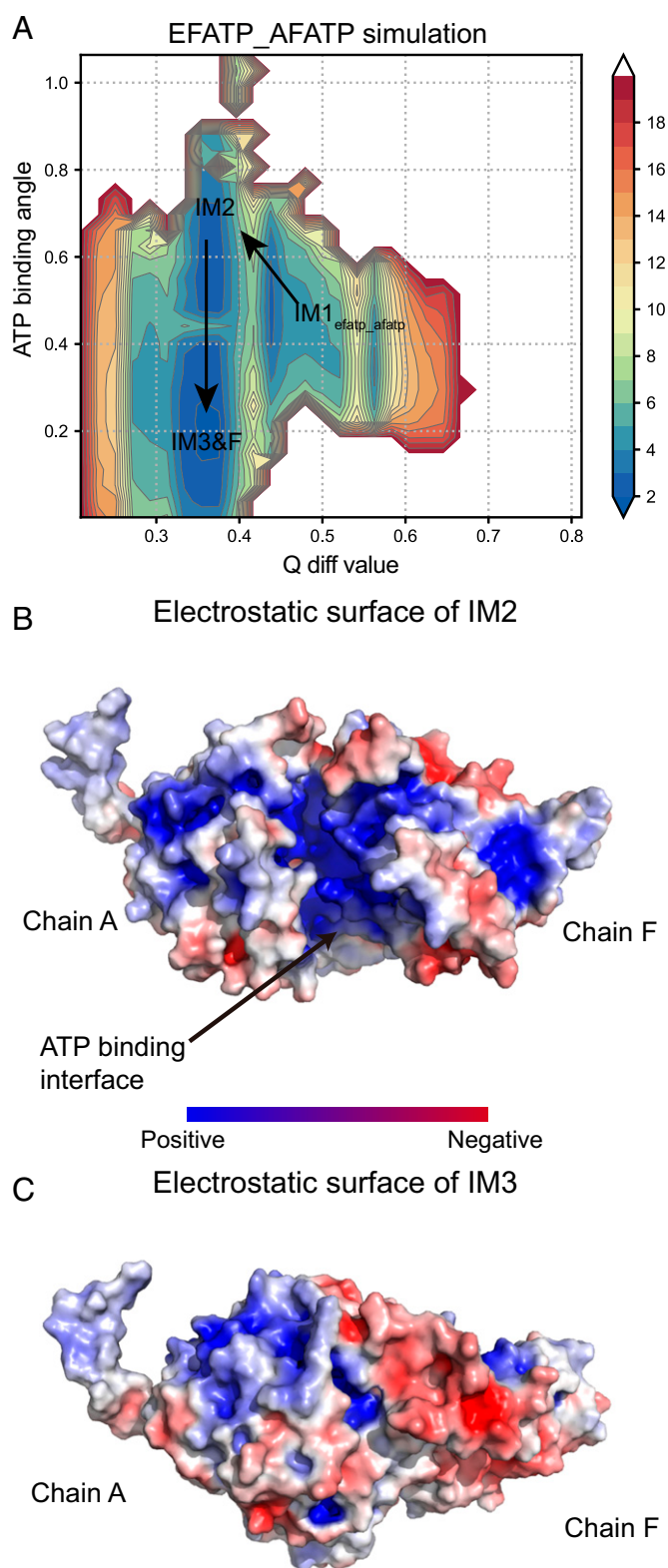


Fig. 3. The electrostatics analysis of chain F/A interface. (A) The 2D free energy profile of Q diff and ATP binding angle for EFATP_AFATP simulation. We found there is a local basin when we put an ATP molecule on the F/A interface during the transition from IM1_{efatp_afatp} to F. To reach the state F, the simulation goes into state IM2 with a repulsive positive charge cavity in ATP binding pocket. (B and C) The electrostatic potential surfaces for representative structures from IM2 (B) and IM3 (C) are shown. The colorbar shows the relative charges of the surface.

chain E and chain F during IM1_{efatp_afatp} to IM2 transition. From IM2 to IM3, the conformational changes involve tangential rotation and an anticlockwise rotation of around 10° of

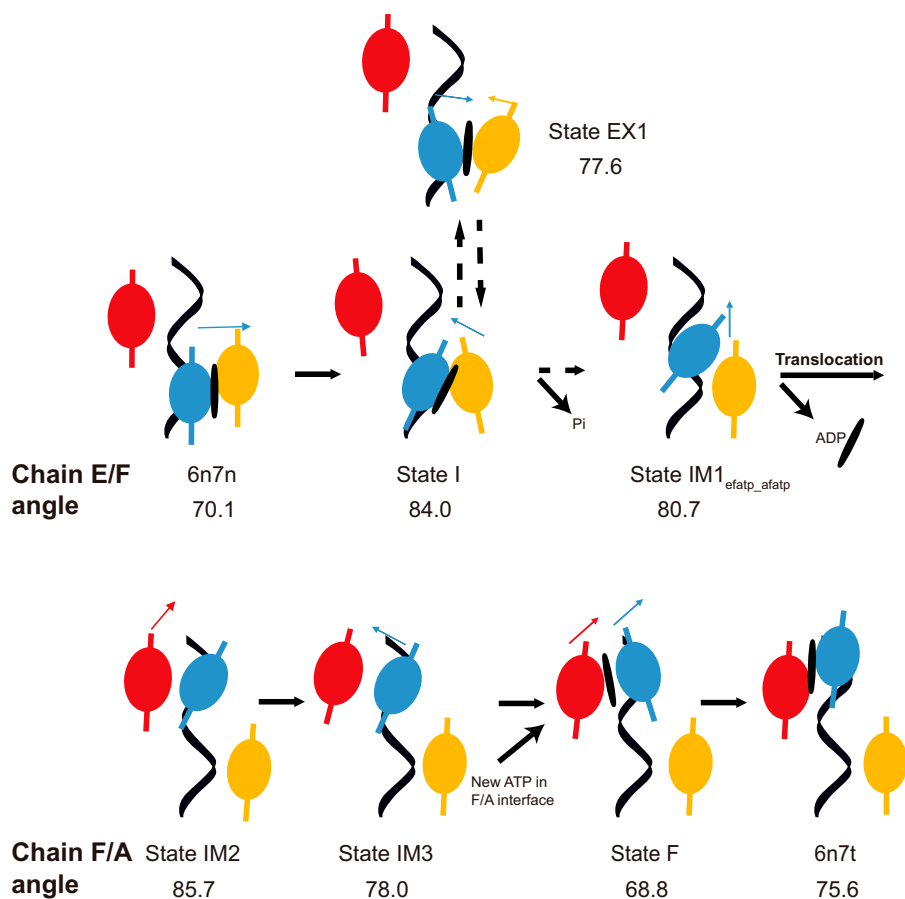


Fig. 4. The overall translocation steps of EFATP_AFATP simulation. The MD simulation reveals several intermediate states. The red, blue, and yellow colors are for chains A, E, F, respectively, in both the schematic figures and the structures. An oval and a stick are used to represent the overall orientation of each chain. We propose that the ATP hydrolyzes between state I and state IM1 while ADP is released before translocation. The dashed line shows there is an extra local state EX1 when an ATP molecule resides in the E/F interface. The arrows indicate the direction of the proposed motion of each local state. We also point out the chain E/F relative angle and the chain F/A relative angle in the corresponding states.

chain F relative to the DNA backbone. Chain F cannot form a stable interface with chain A without a new ATP molecule in state IM3. The final states are achieved when a new ATP molecule binds to the new interface.

We also evaluated the protein–DNA contacts for the intermediate states. We collected all sampled conformations within each of the local states and plotted the contacts between chain F and the DNA chain that exist in more than half of the structures. There are a total of three protein loops that are within 9.5-Å distances to interact with the DNA backbone (*SI Appendix, Fig. S9*). Among them, LoopD1 (residues 487 to 490) and LoopD2 (residues 467 to 473) have been observed interacting with DNA in the gp4–DNA structure. The Arg487 on LoopD1 and Lys467, Asn468, Lys471, and Lys473 on LoopD2 possibly contact the DNA backbone. The third loop (residues 432 to 435) is missing in the cryo-EM structures. However, there are no positively charged residues in the third loop, and the distance is not close enough for the loop to form a direct interaction with the DNA backbone.

We compared the average contacts between chain F and DNA right before and after the translocation (state IM1 and IM2) and found a significant role for LoopD2. The LoopD2 formed the consecutive contacts with DNA during the IM1_{efatp_afatp} to IM2 transition, while the LoopD1–DNA interaction is only found in state I. A 2D free energy profile was made based on the Q diff value and the nearest distance between the nucleotide and the center of mass of LoopD2 (Fig. 5). Several representative structures were picked from states IM1_{efatp_afatp} and IM2 and the transition state in Fig. 5. The several positively charged side chains on

LoopD2 can simultaneously interact with more than one segment of DNA. It is likely that LoopD2 plays a role of anchor to support the large-scale translocation during the subunit translocation.

The Primase Domain Increases the Helicase Activity of gp4. Most replicative helicases contain an NTD that is capable of binding DNA. However, the role of the NTD in helicase translocation is unclear. The gp4 NTD encodes a topoisomerase-primase (TOPRIM) subdomain that binds ssDNA weakly (40). Biochemical assays have shown that deletion of the entire NTD primase domain reduces gp4 helicase activity (41). There is a domain swap between the helicase and primase domain, i.e., the primase domain of chain E is on top of the helicase domain of chain F. The linker between the helicase domain and the primase domain forms extensive interactions with the neighboring helicase domain and holds the hexamer together. Yet, direct primase–helicase interactions were not observed in previous structures.

To explore the role of the NTD in helicase translocation, we built a model of hexameric gp4 primase–helicase complex (residues 67 to 549). In Fig. 6A, we used the same metric to evaluate the 2D free energy profile of the complex umbrella sampling simulation. The addition of NTD decreases the energy barriers between local basins, especially the one from state I to state IM1_{efapo_afatp} in the helicase simulation. In Fig. 6B and C, we show the structures of chain E/F and a zoom-in view of the binding interface in representative structures from states I and IM1_{efapo_afatp}. The NTD domain of subunit E contacts the

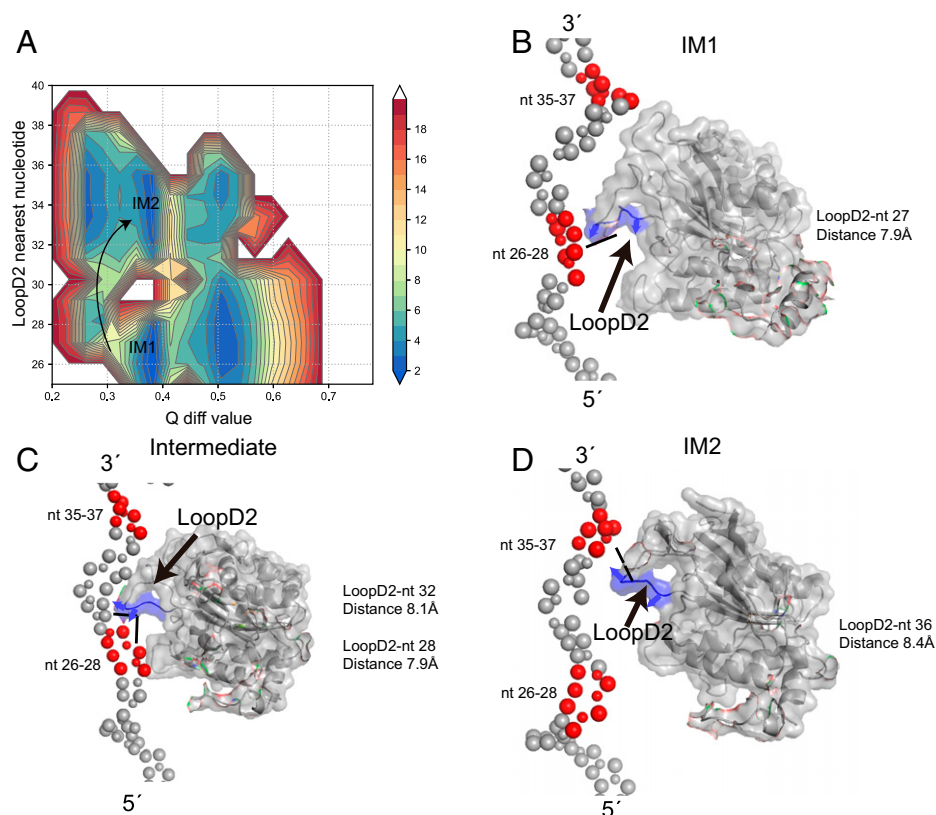


Fig. 5. The role of LoopD2 in the translocation. (A) Two-dimensional free energy profile for the Q diff value and the location of the nearest nucleotide in the DNA to the LoopD2. (B–D) The representative structures from IM1 (B), intermediate (C), and IM2 (D) show the contact between LoopD2 and DNA backbone contact. The nearest nucleotides with the lowest energy in A (26–28, 35–37) are colored in red. The LoopD2 is colored in blue and the distance between the nearest nucleotide with the center of mass of LoopD2 is connected by a dashed line. The value of the distance is commented on the right. A blue arrow indicates the side chain direction of the charged residues in LoopD2.

neighbor chain F helicase domain with a β -sheet (residues 184 to 189) and a short helix (residues 199 to 206) besides interactions through the linker. In addition, two loops (residues 113 to 117 and residues 134 to 137) and a helix (residues 213 to 220) in NTD from subunit F form contacts with DNA during the translocation (Fig. 6D). Residue K137 also contributes to the RNA synthesis activity of the primase domain (42). The primase–DNA contact in chain F and the transient primase–helicase interactions between neighbor chains may facilitate the helicase subunit translocation.

Discussion

We previously introduced OpenAWSEM and Open3SPN2 as coarse-grained models for protein (AWSEM) and DNA (3SPN.2) MD simulations within the OpenMM framework (26). Using graphics processing units, a 30-fold speedup has been achieved in protein and protein–DNA simulations over the existing LAMMPS-based implementations running on a single central processing unit core. In this work, we further optimized the coarse-grained model to better fit the ssDNA based on the previous architecture of OpenAWSEM with Open3SPN2. Our model faithfully reproduces the physical properties of ssDNA. The force field uses excluded volume and electrostatics terms to treat the protein–DNA interactions. This new development provides a transferable model with top-down parameterization, which can be easily applied to other giant protein–nucleic acid systems. The OpenMM framework provides a full-power Python application programming interface that makes modifications or new terms implementations much easier than other platforms. Moreover, all source codes in this paper are publicly available and a tutorial has been provided to benefit the general community to use the present model.

A key question in the helicase field is how the chemical energy from ATP binding and hydrolysis is coupled to protein conformational changes (17). Previous studies suggested a single electrostatic charge in the ATPase active site can control the global conformational changes and stabilize the subunit interface in a helicase homolog (43). In the present study, we investigated gp4 helicase translocation along ssDNA. Our results suggest that the negatively charged ATP is the key to stabilize the subunit–subunit interface. Only in the presence of ATP is the interface minimally frustrated. The tightly bound ATP prevents subunit dissociation and retains the subunit at a local conformational basin “EX1.” Hydrolysis of ATP to ADP significantly lowers the energy barrier for subunit dissociation. On the other hand, new ATP binding helps establish the new interface at the other end of DNA. The observed kinetic scheme is consistent with experimental measurements and previous all-atom simulation of Rho helicases (18).

The ssDNA in SF4 and SF6 helicases all take on an A-like form. During the translocation, the mobile helicase subunit travels more than 20 Å and over 12 nucleotides in distance to interact with the downstream DNA. During this process, the subunit–subunit contacts via the ATPase site are lost, and the subunit is only held together by a long and flexible linker. The simulation demonstrates that the subunit will not perform a free three-dimensional search during translocation, but instead transient DNA–protein and protein–protein interactions guide the long-distance translocation. Cryo-EM has identified two DNA interacting loops, LoopD1 and LoopD2. The LoopD2 is highly positively charged. Although only Lys467 and Asn468 have been shown to directly interact with DNA in the static cryo-EM structure, mutating Lys471 and Lys473 reduces DNA unwinding (44). Our simulations show that LoopD2 contacts

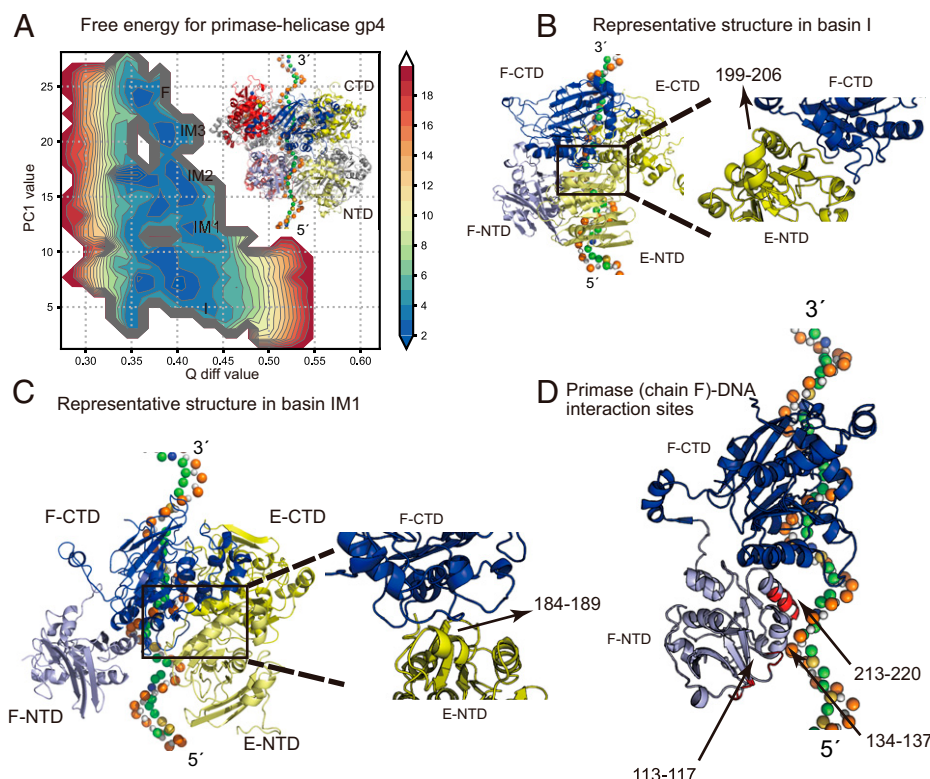


Fig. 6. Analysis of the primase-helicase interaction. (A) The 2D free energy profile of the primase-helicase complex. The energy barriers are decreased when compared with the EFAPQ_AFATP simulation. The representative structure of the complete model is shown in the upper right corner. The color schemes are the same as the helicase-only structures. The primase domain in each chain is depicted as a lighter color than the corresponding helicase part. (B and C) Representative structures from state I (B) and IM1 (C) with a zoom-in view for the helicase-primase interface. The NTD-CTD interaction residues are labeled. (D) A representative structure of the primase-helicase complex shows the key interaction sites found by MD simulation. The interaction sites are labeled.

different segments of DNA through the translocation with its many positively charged residues, which guides the long-distance subunit translocation. Moreover, the NTD can form transient interactions with the helicase subunit and ssDNA to facilitate its transition.

In summary, our development provides a transferable model with top-down parameterization, which can be easily applied to other giant protein-nucleic acid systems. This model allows the analysis of a nonequilibrium process driven by ATP hydrolysis using an equilibrium coarse-grained force field. With the emergence of many cryo-EM structures of giant protein-nucleic acid complexes, we believe our system can greatly aid the mechanistic understanding of essential physiological processes in molecular biology.

Materials and Methods

Here, we give the details of our force field. First, the introduction of the force field is described. Then, we introduce the umbrella sampling technique and how we pick out the collective variables for umbrella sampling. Moreover, we detail how to analyze the contact and calculate the free energy profile.

A Coarse-Grained Force Field for a Protein-ssDNA-ATP System. Because they employ reduced representations, coarse-grained force fields must be benchmarked to correctly represent real biomolecule systems. While coarse-grained force fields that include both protein and DNA exist, they only focus on special cases and rely on corresponding atomistic MD simulations for their parameterization (45, 46). Here we introduced a transferrable coarse-grained force field, and no further benchmarks are required for studying the specific protein-DNA systems.

The coarse-grained protein folding force field known as AWSEM is the newest version of a series of models that have been optimized based on the principles of the energy landscape theory of protein folding which provide a quantitative machine-learning strategy (47). Three explicit atoms, CA, CB, and O, and three

virtual sites, C, N, and H (except for proline and glycine), are modeled to represent one residue in AWSEM simulation (26). The latest updated version of AWSEM, AWSEM-Suite, has participated in protein structure prediction competition CASP13 and won the top three server predictions in two cases (48). The protein model AWSEM has been successfully applied to several different systems and is available to the public as an online server (49).

The coarse-grained DNA model we employed includes many detailed aspects of DNA architecture such as the specific hydrogen bonding and base-stacking potentials from the canonical B-DNA structure (27). We have used the 3SPN.2C force field to investigate the dynamics of the nuclear factor κ B heterodimer binding problem (50). The detailed benchmark process of this ssDNA force field is included in *Results*.

The representations of the ATP molecule and the ADP molecule are transplanted from the Open3SPN2 force field. However, the backbone and angle terms are parameterized based on the average distances and angles of multiple crystal structures that contain ATP molecules. The extra coarse-grained PB and PG atoms are positioned at the same place as PB and PG atoms in the cryo-EM structure with -1 and -2 charges, respectively. The ATP-protein interactions include two parts: exclusion and electrostatics. While the exclusion term uses the same value as in Open3SPN2, the dielectric constant of electrostatics term has been changed. Typically, the macromolecule is considered to be a low dielectric medium while the water phase is modeled as a homogeneous medium with a dielectric constant of 80. However, for the coarse-grained protein model the dielectric constant is too low for protein-ATP electrostatics, and $\epsilon = 5 \sim 10$ fits better for the current model (51, 52). Here we choose the value $\epsilon = 8$. In order to evaluate the role of the ATP molecule during the translocation in its binding pocket, we added a distance bias term to the corresponding residues 504 and 535 in the full chain. The magnesium ion was modeled as a single bead with a radius of 2.35 for its van der Waals radius and mass 24.305 with two positive charges.

These terms are combined as detailed in the following equation:

$$V_{\text{total}} = V_{\text{AWSEM}} + V_{\text{ss-3SPN2}} + V_{\text{protein-DNA}} + V_{\text{ligand}} + V_{\text{protein-ligand}}$$

To build the initial models of the helicase domain (residues 264 to 549), Modeler version 9.23 was used to fix 15 missing residues in the cryo-EM structures

(35). The structures were loaded in VMD, the corresponding topology files were generated, and then padding with water molecules and sodium and chlorine ions was added. Then 10,000 steps on equilibration simulation with CHARMM27 force field was carried at NAMD.

Free Energy Calculations Based on Umbrella Sampling Technique.

Umbrella sampling is an enhanced sampling technique that could force the exploration of regions of state space that would otherwise have insufficient sampling. It uses a series of independent windows along a selected collective variable, which serves as a continuous parameter to describe the system from a higher-dimensional space to model a conformational transition (53). The umbrella sampling along with an order parameter Q diff between the two translocation states from cryo-EM structures was used to project the free energy landscapes onto a single dimension. The harmonic biasing potential used for constant temperature umbrella sampling simulations for 8 million steps was scaled to 1,000 kcal/mol. The biasing center values were chosen to be equally spaced from 0 to 1 with an increment of 0.02. The weighted histogram analysis method (WHAM) is used to reconstruct the unbiased free energy landscapes from the umbrella sampling data (54). An additional dimension related to the specific conformation changes of interest was added for plotting the 2D free energy profile. To generate a series of initial structures for umbrella sampling, we first used the morph command in PyMOL to generate a total of 50 interpolation states from the two endpoint structures. Then, the Q diff value was added as an external bias with strength of 2,000 kcal/mol for 50,000 steps. After that, we ran another 50,000 steps without additional constraints to remove steric clashes and the final structures were used as the input for umbrella sampling.

Parameters Used for Umbrella Sampling and Free Energy Calculation.

To compute the relative free energy of the helicase translocation along the DNA, we used Q diff to sample structures both near the limits and intermediate between the two topologies. The Q value was introduced in 2001 for evaluating the structure similarity like RMSD (55). Q diff is a variant of Q value that has been used in our previous work of the aggregation free energy landscape (56):

$$Q \text{ diff} = \frac{q - q_2}{q_1 - q_2}$$

$$q(r_{ij}) = \frac{1}{(N-2)(N-3)} \sum_{j=i+2} [e^{-(r_{ij}-r_{ij}^{N1})^2/2\sigma_q^2} - e^{-(r_{ij}-r_{ij}^{N2})^2/2\sigma_q^2}].$$

In the above equation, $\sigma_{ij} = |j - i|^{0.15}$, while $q_1 = q(r_{ij}^{N1})$ and $q_2 = q(r_{ij}^{N2})$. $N1$ and $N2$ indicate distances evaluated in the starting and final states of the translocation (PDB IDs 6n7n and 6n7t) (48). The harmonic biasing potential used for constant temperature umbrella sampling simulations along Q diff is shown in the following equation:

$$V_{Q-bias} = \frac{1}{2} k_{Q-bias} (Q_{diff} - Q_0)^2.$$

Principal component (PC) analysis is a powerful method to probe large-scale conformational changes of proteins (57). Here we used all the atoms of chain A, E, and F from a very long (8 million steps for each window, 50 windows) EFADP_AFATP umbrella sampling simulation to compute the top five PCs. All of the calculations were carried out with the ProDy package (58). The largest PC contributes to a total of 40% of the dynamics, and the relative motion of each residue is computed in *SI Appendix, Fig. S10*. The overall motion corresponding to the largest PC is shown in *Movie S1*. The largest PC was picked out to describe the transition in 2D free energy profiles.

The ATP binding angle is selected as the angle of the center of mass of the coarse-grained ATP molecule, the Walker A motif (residue Ser319 to Thr320) and

the arginine finger (Arg522) of the neighboring chain. Both Walker A motif and arginine finger have been reported as the key binding sites of ATP. The nearest nucleotide toward the LoopD2 is selected as the shortest distance between the atom of S in each nucleotide and the center of mass of LoopD2.

Methods for Representative Structure Selection and Electrostatics Surface Analysis.

The representative structures of IM2 and IM3 were picked out from the clustering analysis of all the structures that fall in the IM2 and IM3 basins in EFATP_AFATP simulation. This clustering uses the pairwise Q value of chain A, chain E, and chain F as the similarity metric (48). We generated a hierarchically clustered heat map using the clustermap module with default parameters in Seaborn package and picked out the center one of the largest cluster (upper-left corner) as the representative structure. The final clustermaps of the state IM2 and IM3 are shown in *SI Appendix, Fig. S11*. We used the module APBS from PyMOL to perform the electrostatics surface analysis (59).

Metrics for Evaluating the Motions in the Translocation. We used the chain E/F com angle and chain F/A com angle to evaluate the subunit translocation. For the chain E/F com angle, it is the angle of the center of mass of chain E and chain F with the center of mass of the chain E N-terminal helix (residues 264 to 279) as the centering point. For the chain F/A com angle, it is the angle of the center of mass of chain F and chain A, with chain F N-terminal helix. The threshold of 9.5 Å was used to judge whether there are contacts between protein and DNA as well as two protein domains.

All-Atom Frustration Calculation. All-atom frustration analysis can be used in protein structure refinement, drug design, and binding pocket analysis (37, 39). The all-atom frustration analysis using Rosetta software package is detailed in another survey paper (37). A review by Ferreira et al. describes the overall property and definition of frustration can be consulted by readers who are interested in the analysis (38).

Data Availability. The source code for the force field and data from the simulation and analysis used in this paper have been deposited at https://github.com/CryoSky/helicase_gp4. The OpenAWSEM force field is available at <https://github.com/npschafer/openawsem>. The documentations and references of the AWSEM force field can be found at awsem-md.org. The Open3SPN2 force field is available for download from the GitHub repository at <https://github.com/cabb99/open3spn2>. All other study data are included in the article and/or supporting information.

ACKNOWLEDGMENTS. This work was supported by the Center for Theoretical Biological Physics, sponsored by NSF Grant PHY-2019745. Additionally, we recognize the D.R. Bullard Welch Chair at Rice University, Grant C-0016 (to P.G.W.). This work is also supported by Grant R01 GM44557 from the National Institute of General Medical Sciences (to P.G.W.), by the Cancer Prevention and Research Institute of Texas Recruitment of First-time Tenure-track Faculty Members Fund (Grant RR190046), and by National Institute of General Medical Sciences Grant 1R35GM142722 (to Y.G.). We thank Dr. Min-Yeh Tsai and Dr. Xiakun Chu for helpful discussions during the development of the force field and simulation.

Author affiliations: ^aDepartment of Biosciences, Rice University, Houston, TX 77005; ^bCenter for Theoretical Biological Physics, Rice University, Houston, TX 77005; ^cDepartment of Physics, Rice University, Houston, TX 77005; ^dDepartment of Physics, University of Science and Technology of China, Hefei 230026, China; ^eDepartment of Research and Development, neoX Biotech, Beijing 100206, China; and ^fDepartment of Chemistry, Rice University, Houston, TX 77005

1. M. R. Singleton, M. S. Dillingham, D. B. Wigley, Structure and mechanism of helicases and nucleic acid translocases. *Annu. Rev. Biochem.* **76**, 23–50 (2007).
2. M. E. O'Donnell, H. Li, The ring-shaped hexameric helicases that function at DNA replication forks. *Nat. Struct. Mol. Biol.* **25**, 122–130 (2018).
3. Y. Gao, W. Yang, Different mechanisms for translocation by monomeric and hexameric helicases. *Curr. Opin. Struct. Biol.* **61**, 25–32 (2020).
4. O. Itsathitphaisarn, R. A. Wing, W. K. Eliason, J. Wang, T. A. Steitz, The hexameric helicase DnaB adopts a nonplanar conformation during translocation. *Cell* **151**, 267–277 (2012).
5. N. D. Thomsen, J. M. Berger, Running in reverse: The structural basis for translocation polarity in hexameric helicases. *Cell* **139**, 523–534 (2009).
6. S. Bailey, W. K. Eliason, T. A. Steitz, Structure of hexameric DnaB helicase and its complex with a domain of DnaG primase. *Science* **318**, 459–463 (2007).
7. M. Meagher, L. B. Epling, E. J. Enemark, DNA translocation mechanism of the MCM complex and implications for replication initiation. *Nat. Commun.* **10**, 3117 (2019).
8. W. Yang, M. M. Seidman, W. D. Rupp, Y. Gao, Replisome structure suggests mechanism for continuous fork progression and post-replication repair. *DNA Repair (Amst.)* **81**, 102658 (2019).
9. P. Eickhoff et al., Molecular basis for ATP-hydrolysis-driven DNA translocation by the CMG helicase of the eukaryotic replisome. *Cell Rep.* **28**, 2673–2688.e8 (2019).
10. R. Georgescu et al., Structure of eukaryotic CMG helicase at a replication fork and implications to replisome architecture and origin initiation. *Proc. Natl. Acad. Sci. U.S.A.* **114**, E697–E706 (2017).
11. D. J. Crampton, S. Mukherjee, C. C. Richardson, DNA-induced switch from independent to sequential dTTP hydrolysis in the bacteriophage T7 DNA helicase. *Mol. Cell* **21**, 165–174 (2006).
12. B. Sun et al., ATP-induced helicase slippage reveals highly coordinated subunits. *Nature* **478**, 132–135 (2011).

13. M. Schlierf, G. Wang, X. S. Chen, T. Ha, Hexameric helicase G40P unwinds DNA in single base pair steps. *eLife* **8**, e42001 (2019).
14. C. Hyeon, G. H. Lorimer, D. Thirumalai, Dynamics of allosteric transitions in GroEL. *Proc. Natl. Acad. Sci. U.S.A.* **103**, 18939–18944 (2006).
15. K. Yoshimoto, K. Arora, C. L. Brooks III, Hexameric helicase deconstructed: Interplay of conformational changes and substrate coupling. *Biophys. J.* **98**, 1449–1457 (2010).
16. H. Flechsig, Computational biology approach to uncover hepatitis C virus helicase operation. *World J. Gastroenterol.* **20**, 3401–3409 (2014).
17. H. Liu, Y. Shi, X. S. Chen, A. Warshel, Simulating the electrostatic guidance of the vectorial translocations in hexameric helicases and translocases. *Proc. Natl. Acad. Sci. U.S.A.* **106**, 7449–7454 (2009).
18. W. Ma, K. Schulten, Mechanism of substrate translocation by a ring-shaped ATPase motor at millisecond resolution. *J. Am. Chem. Soc.* **137**, 3031–3040 (2015).
19. E. A. Toth, Y. Li, M. R. Sawaya, Y. Cheng, T. Ellenberger, The crystal structure of the bifunctional primase-helicase of bacteriophage T7. *Mol. Cell* **12**, 1113–1123 (2003).
20. S. J. Lee, C. C. Richardson, Choreography of bacteriophage T7 DNA replication. *Curr. Opin. Chem. Biol.* **15**, 580–586 (2011).
21. H. J. Geertsema, A. W. Kulczyk, C. C. Richardson, A. M. van Oijen, Single-molecule studies of polymerase dynamics and stoichiometry at the bacteriophage T7 replication machinery. *Proc. Natl. Acad. Sci. U.S.A.* **111**, 4073–4078 (2014).
22. K. E. Duderstadt *et al.*, Simultaneous real-time imaging of leading and lagging strand synthesis reveals the coordination dynamics of single replisomes. *Mol. Cell* **64**, 1035–1047 (2016).
23. Y. Gao *et al.*, Structures and operating principles of the replisome. *Science* **363**, eaav7003 (2019).
24. P. McInerney, A. Johnson, F. Katz, M. O'Donnell, Characterization of a triple DNA polymerase replisome. *Mol. Cell* **27**, 527–538 (2007).
25. J. S. Lewis *et al.*, Tunability of DNA polymerase stability during eukaryotic DNA replication. *Mol. Cell* **77**, 17–25.e5 (2020).
26. W. Lu *et al.*, OpenAWSEM with Open3SPN2: A fast, flexible, and accessible framework for large-scale coarse-grained biomolecular simulations. *PLoS Comput. Biol.* **17**, e1008308 (2021).
27. G. S. Freeman, D. M. Hinckley, J. P. Lequieu, J. K. Whitmer, J. J. de Pablo, Coarse-grained modeling of DNA curvature. *J. Chem. Phys.* **141**, 165103 (2014).
28. D. M. Hinckley, G. S. Freeman, J. K. Whitmer, J. J. de Pablo, An experimentally-informed coarse-grained 3-Site-Per-Nucleotide model of DNA: Structure, thermodynamics, and dynamics of hybridization. *J. Chem. Phys.* **139**, 144903 (2013).
29. H. Chen *et al.*, Ionic strength-dependent persistence lengths of single-stranded RNA and DNA. *Proc. Natl. Acad. Sci. U.S.A.* **109**, 799–804 (2012).
30. M. C. Murphy, I. Rasnik, W. Cheng, T. M. Lohman, T. Ha, Probing single-stranded DNA conformational flexibility using fluorescence spectroscopy. *Biophys. J.* **86**, 2530–2537 (2004).
31. B. Tinland, A. Pluen, J. Sturm, G. Weill, Persistence length of single-stranded DNA. *Macromolecules* **30**, 5763–5765 (1997).
32. C. G. Baumann, S. B. Smith, V. A. Bloomfield, C. Bustamante, Ionic effects on the elasticity of single DNA molecules. *Proc. Natl. Acad. Sci. U.S.A.* **94**, 6185–6190 (1997).
33. W. K. Olson *et al.*, "DNA simulation benchmarks as revealed by X-ray structures" in *Computational Studies of RNA and DNA*, J. Sponer, F. Lancaš, Eds. (Springer, 2006), pp. 235–257.
34. R. Owczarzy *et al.*, Effects of sodium ions on DNA duplex oligomers: Improved predictions of melting temperatures. *Biochemistry* **43**, 3537–3554 (2004).
35. B. Webb, A. Sali, "Comparative protein structure modeling using MODELLER" in *Current Protocols in Bioinformatics* (John Wiley & Sons, Inc., 2016), pp. 5.6.1–5.6.37.
36. A. D. Mackerell *et al.*, All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J. Phys. Chem. B* **102**, 3586–3616 (1998).
37. M. Chen *et al.*, Surveying biomolecular frustration at atomic resolution. *Nat. Commun.* **11**, 5944 (2020).
38. D. U. Ferreira, E. A. Komives, P. G. Wolynes, Frustration in biomolecules. *Q. Rev. Biophys.* **47**, 285–363 (2014).
39. M. Chen *et al.*, Protein structure refinement guided by atomic packing frustration analysis. *J. Phys. Chem. B* **124**, 10889–10898 (2020).
40. B. Zhu, S. J. Lee, C. C. Richardson, Direct role for the RNA polymerase domain of T7 primase in primer delivery. *Proc. Natl. Acad. Sci. U.S.A.* **107**, 9099–9104 (2010).
41. L. E. Bird, K. Håkansson, H. Pan, D. B. Wigley, Characterization and crystallization of the helicase domain of bacteriophage T7 gene 4 protein. *Nucleic Acids Res.* **25**, 2620–2626 (1997).
42. M. Kato, T. Ito, G. Wagner, C. C. Richardson, T. Ellenberger, Modular architecture of the bacteriophage T7 primase couples RNA primer synthesis to DNA synthesis. *Mol. Cell* **11**, 1349–1360 (2003).
43. D. M. Kim, H. Zheng, Y. J. Huang, G. T. Montellione, J. F. Hunt, ATPase active-site electrostatic interactions control the global conformation of the 100 kDa SecA translocase. *J. Am. Chem. Soc.* **135**, 2999–3010 (2013).
44. A. K. Satapathy *et al.*, Residues in the central beta-hairpin of the DNA helicase of bacteriophage T7 are important in DNA unwinding. *Proc. Natl. Acad. Sci. U.S.A.* **107**, 6782–6787 (2010).
45. J. J. Uusitalo, H. I. Ingólfsson, S. J. Marrink, I. Faustino, Martini coarse-grained force field: Extension to RNA. *Biophys. J.* **113**, 246–256 (2017).
46. A. K. Sieradzian *et al.*, A new protein nucleic-acid coarse-grained force field based on the UNRES and NARES-2P force fields. *J. Comput. Chem.* **39**, 2360–2370 (2018).
47. N. P. Schafer, B. L. Kim, W. Zheng, P. G. Wolynes, Learning to fold proteins using energy landscape theory. *Isr. J. Chem.* **54**, 1311–1337 (2014).
48. S. Jin *et al.*, Protein structure prediction in CASP13 using AWSEM-suite. *J. Chem. Theory Comput.* **16**, 3977–3988 (2020).
49. S. Jin *et al.*, AWSEM-Suite: A protein structure prediction server based on template-guided, coevolutionary-enhanced optimized folding landscapes. *Nucleic Acids Res.* **48**, W25–W30 (2020).
50. D. A. Potoyan, C. Bueno, W. Zheng, E. A. Komives, P. G. Wolynes, Resolving the NFκB heterodimer binding paradox: Strain and frustration guide the binding of dimeric transcription factors. *J. Am. Chem. Soc.* **139**, 18558–18566 (2017).
51. A. Warshel, A. Papayan, Electrostatic effects in macromolecules: Fundamental concepts and practical modeling. *Curr. Opin. Struct. Biol.* **8**, 211–217 (1998).
52. L. Li, C. Li, Z. Zhang, E. Alexov, On the dielectric "constant" of proteins: Smooth dielectric function for macromolecular modeling and its implementation in DelPhi. *J. Chem. Theory Comput.* **9**, 2126–2136 (2013).
53. Q. Liao, Enhanced sampling and free energy calculations for protein simulations. *Prog. Mol. Biol. Transl. Sci.* **170**, 177–213 (2020).
54. E. Gallicchio, M. Andrec, A. K. Felts, R. M. Levy, Temperature weighted histogram analysis method, replica exchange, and transition paths. *J. Phys. Chem. B* **109**, 6722–6731 (2005).
55. M. P. Eastwood, C. Hardin, Z. Luthey-Schulten, P. G. Wolynes, Evaluating protein structure-prediction schemes using energy landscape theory. *IBM J. Res. Dev.* **45**, 475–497 (2001).
56. W. Zheng, M. Y. Tsai, M. Chen, P. G. Wolynes, Exploring the aggregation free energy landscape of the amyloid-β protein (1–40). *Proc. Natl. Acad. Sci. U.S.A.* **113**, 11835–11840 (2016).
57. X. Lin *et al.*, Forging tools for refining predicted protein structures. *Proc. Natl. Acad. Sci. U.S.A.* **116**, 9400–9409 (2019).
58. A. Bakan, L. M. Meireles, I. Bahar, ProDy: Protein dynamics inferred from theory and experiments. *Bioinformatics* **27**, 1575–1577 (2011).
59. E. Jurrus *et al.*, Improvements to the APBS biomolecular solvation software suite. *Protein Sci.* **27**, 112–128 (2018).