Structural and dynamic basis of DNA capture and translocation by mitochondrial Twinkle helicase

Zhao Li 1, Parminder Kaur 2,3,*, Chen-Yu Lo 1, Neil Chopra 1, Jamie Smith 1, Hong Wang 2,3,4 and Yang Gao 1,*

1BioSciences Department, Rice University, Houston, TX 77005, USA, 2Physics Department, North Carolina State University, Raleigh, NC 27695, USA, 3Center for Human Health and the Environment, North Carolina State University, Raleigh, NC 27695, USA and 4Toxicology Program, North Carolina State University, Raleigh, NC 27695, USA

Received August 06, 2022; Revised October 21, 2022; Editorial Decision October 25, 2022; Accepted October 26, 2022

ABSTRACT

Twinkle is a mitochondrial replicative helicase which can self-load onto and unwind mitochondrial DNA. Nearly 60 mutations on Twinkle have been linked to human mitochondrial diseases. Using cryo-electron microscopy (cryo-EM) and high-speed atomic force microscopy (HS-AFM), we obtained the atomic-resolution structure of a vertebrate Twinkle homolog with DNA and captured in real-time how Twinkle is self-loaded onto DNA. Our data highlight the important role of the non-catalytic N-terminal domain of Twinkle. The N-terminal domain directly contacts the C-terminal helicase domain, and the contact interface is a hotspot for disease-related mutations. Mutations at the interface destabilize Twinkle hexamer and reduce helicase activity. With HS-AFM, we observed that a highly dynamic Twinkle domain, which is likely to be the N-terminal domain, can protrude ~5 nm to transiently capture nearby DNA and initialize Twinkle loading onto DNA. Moreover, structural analysis and subunit doping experiments suggest that Twinkle hydrolyzes ATP stochastically, which is distinct from related helicases from bacteriophages.

INTRODUCTION

A mitochondrion, the powerhouse of a cell, harbors its own genome, which is replicated and maintained differently from its nuclear counterpart. The circular double-stranded mitochondrial genome contains two separate origins (1,2). The replication of the two DNA strands is asynchronous and does not involve Okazaki fragment synthesis (3,4). The mitochondrial replisome shares a distant homology with the replisomes from bacteria and bacteriophages (2,5). Its core components include DNA polymerase γ. Twinkle helicase and mitochondrial single-stranded DNA-binding protein (6). DNA synthesis on both strands is initialized with RNA transcripts produced by the mitochondrial RNA polymerases (7,8). Mutations and deletions in the mitochondrial genome are correlated with numerous neuromuscular diseases and premature aging (9).

Twinkle is a mitochondrial replicative helicase (10,11). The expression level of Twinkle is correlated with the copy number of mitochondrial DNA (12), and the homozygous deletion of Twinkle is embryonically lethal (13). Nearly 60 point mutations on Twinkle have been linked to various human mitochondrial diseases (11,14). Patients with Twinkle mutations exhibit reduced mitochondrial DNA copy numbers or partial deletions of the mitochondrial DNA (9). Moreover, Twinkle also participates in mitochondrial DNA repair (15,16) and degradation (17). Consistent with its diverse functions, Twinkle interacts with a number of proteins involved in mitochondrial DNA replication and repair (6,16–18).

Helicases are chemo-mechanical motors that couple ATP hydrolysis to directional translocation on DNA or RNA (19,20). Based on sequence conservations, helicases are divided into six superfamilies (SFs), with SF1 and SF2 helicases being monomeric and SF3–SF6 helicases being hexameric (20). Replicative helicases are all hexameric and belong to SF3, 4 and 6. The six helicase subunits form a ring or lock-washer-shaped hexamer, binding ATP at each of the subunit interfaces and DNA in its central channel. DNA-binding loops from different subunits stagger like a staircase to interact with the DNA backbone. While hexameric helicases may display different directionality, step sizes and other mechanistic details, a sequential hand-over-hand translocation mechanism has been proposed for most of them (21–27). Following ATP hydrolysis and product release at one end of the hexamer, the DNA-binding loops or the entire subunit move from one end of the DNA to the other end to form a new ATP-binding site. Sequential hydrolysis along the hexamer leads to unidirectional and progressive translocation on DNA. A similar hand-over-hand translocation mechanism has been proposed for many hexa-
americ protein translocases (28). Furthermore, the hexameric rings of replicative helicases pose topological restrictions for their loading onto the DNA substrates. To overcome this challenge, bacterial (SF4), archaean (SF6) and eukaryotic nuclear (SF6) helicases utilize specialized loaders to crack their rings open (29). In contrast, Twinkle (SF4) and bacteriophage T7 gp4 (SF4) helicases are capable of self-loading onto DNA (30,31), but their mechanisms are poorly understood.

Twinkle is a member of the SF4 family helicases and shares low homology with bacterial DnaB and bacteriophage T7 gp4 (10,32). The N-terminal domain (NTD) of Twinkle is derived from the TOPPRIM family primase (33). However, key residues in the Twinkle primase active site from vertebrates are mutated, and no primase activity has been detected (11). The helicase domain is located on the C-terminal half of Twinkle. Similar to gp4 and DnaB (21,22), Twinkle forms a homo-hexamer and migrates in the 5′ to 3′ direction on the single-stranded DNA (ssDNA) (32,34). Interestingly, the reported Twinkle helicase activity is significantly lower than that of DnaB or gp4 (35). In addition, Twinkle can bind to and migrate on double-stranded DNA (dsDNA) and Holliday junction DNA (35,36). To date, structures of Twinkle in the absence of DNA have been reported (34,37–39). However, how Twinkle functions mechanistically on DNA and how disease-related mutations affect Twinkle function are still largely unclear. Here we report cryo-electron microscopy (cryo-EM), biochemical and high-speed atomic force microscopy (HS-AFM) characterization of a vertebrate Twinkle homolog. Our data elucidate unique structural and dynamic features of Twinkle helicase, and support a novel molecular model explaining Twinkle self-loading, DNA unwinding and human disease-related mutations on Twinkle.

MATERIALS AND METHODS

Sequence analysis

Twinkle homolog sequences from Arabidopsis thaliana, bacteriophage T7, Dictyostelium discoideum, Drosophila melanogaster, Gallus gallus and Homo sapiens (Hs) were used as baits for the BLAST search (https://www.uniprot.org/blast/). The target database was set to be the UniProtKB reference proteomes plus Swiss-Prot, with the E-value set to be 10. Redundant sequences were eliminated, and sequences with <400 residues or >800 residues were removed. The remaining sequences were aligned with the MUSCLE sequence alignment tool (40). The sequence identity matrix was plotted with Matlab.

Plasmid construction

The Lates calcarifer (Lc) Twinkle (Entrez ID 108894827, synthesized by GeneUniversal Inc.) was cloned into a modified pET28a vector with an N-terminal histidine tag and a PreScission protease cleavage site. Mutations of LcTwinkle were generated using methods described in the QuikChange mutagenesis kit (Agilent). Sequences of the LcTwinkle constructs were confirmed by sequencing the entire reading frames of each construct. The active site mutation of Homo sapiens (Hs) (uniport Q8CIW5) and Danio rerio (Dr) (uniport A0A0R41CC1) Twinkle were all synthesized by GeneUniversal Inc. and cloned into the same pET28a vector.

Protein expression and purification

The LcTwinkle plasmid was transformed into Escherichia coli BL21 (DE3) (Novagen). Isopropyl-β-D-thiogalactopyranoside (1 mM) was added to the cell culture at an optical density of 0.8 for protein induction. The E. coli cells were further incubated at 16°C for 16 h under shaking at 150 rpm for protein expression. The cells were harvested and disrupted by sonication in a Lysis buffer containing 50 mM Tris–HCl, pH 8.0, 1 M NaCl and 5% glycerol. The soluble fraction was loaded onto a 5 ml HisTrap column (GE Healthcare), which was pre-equilibrated with the Lysis buffer plus 20 mM imidazole. The column was subsequently washed with 300 ml of Lysis buffer plus 50 mM imidazole, and the protein was eluted with 5 ml of Lysis buffer plus 300 mM imidazole. The eluted protein was treated with 10 U of PreScission protease (Sigma) at 4°C for 2 h. Afterward, the protein was diluted into a MonoQ buffer containing 50 mM Tris pH (8.0), 0.1 mM ATP, 1 mM MgCl₂, 150 mM KCl, 3 mM dithiothreitol (DTT) and 5% glycerol, and loaded onto a MonoQ column (GE Healthcare) pre-equilibrated with the MonoQ buffer. LcTwinkle was eluted by a gradient KCl concentration. The purified LcTwinkle was aliquoted and flash-frozen at −80°C for further study. Mutant LcTwinkle proteins were expressed and purified according to the same protocol. Hs, Mm and Dr Twinkle were expressed and purified similarly to LcTwinkle. After elution from HisTrap columns, the proteins were mixed with ssDNA (5′-TGTTCTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT
DNA unwinding assay

The helicase reactions were carried out in a buffer containing 50 mM Tris–HCl (pH 8.0), 50 mM KCl, 5% glycerol, 0.1 mg/ml bovine serum albumin (BSA), 5 mM MgCl$_2$ and 4 mM ATP. Each reaction contained 25 nM FAM-labeled fork DNA substrate, which was annealed with 5′-FAM-CCTAGCTCAGGTTCAGTACTCGAACTCTACATACTATACATGAATATCATAACTAATAA-3′ and 5′-TTATTAGTTATGATTACATGTATAGTTATCATCTCAAGCTCATG-3′. The reactions were all performed at 37°C for 30 min. The reactions were stopped with a quench buffer containing 20 mM EDTA, 1% SDS and 0.2 mg/ml proteinase K. The substrate and product were separated on a 12% native polyacrylamide gel, and the images were analyzed using a Sapphire Biomolecular Imager-RGB IS1025 (Azure BioSystems). All results were based on at least three independent tests.

DNA binding assay

The DNA binding assay was performed in 100 μl of reaction buffer containing 25 mM Tris–HCl (pH 8.0), 150 mM KCl, 0.5 mM ATP, 1 mM MgCl$_2$, 20 mM Cy5-labeled ssDNA (5′-Cy5-GGATTATTTACATTGGCAGA TTCACC-3′) and varying concentrations of WT or mutant LcTwinkle. The reaction mix was loaded onto 96-well plates and the fluorescence anisotropy was measured with 645 nm excitation and 670 nm emission using an Infinite M1000 Pro fluorescence Spectrophotometer, Agilent, Santa Clara, CA, USA. The total protein concentration was 0.2 mg/ml for each of the proteins. The protein sample was scanned at room temperature using a fluorometer (Cary Eclipse Fluorescence Spectrophotometer, Agilent, Santa Clara, CA, USA). During the scan, a 430 nm filter was used as the excitation wavelength and the emission between 430 and 570 nm was collected. For subunit doping experiments, WT and mutant Twinkles were buffer exchanged to 500 KCl, 50 mM Tris–HCl (pH 8.0) and 3 mM DTT; 1 mM ATP and 2 mM MgCl$_2$ were added when indicated. The protein concentration was 0.2 mg/ml for each of the proteins. The protein sample was scanned at room temperature using a fluorometer (Cary Eclipse Fluorescence Spectrophotometer, Agilent, Santa Clara, CA, USA). During the scan, a 430 nm filter was used as the excitation wavelength and the emission between 430 and 570 nm was collected. For subunit doping experiments, WT and mutant Twinkles were buffer exchanged to 500 KCl, 50 mM Tris–HCl (pH 8.0) and 3 mM DTT. The total protein concentration after mixing was 2 μM. All mixed proteins were incubated at room temperature for 5 min before the activity assay. The ATPase, helicase and DNA binding activities were determined as described above. Mixed LcTwinkle at 0.2 μM was used for the helicase assay and at 0.05 μM for the ATPase assay.

Cryo-EM sample preparation

E325Q LcTwinkle was mixed with ssDNA (5′-TGGTCTTTTTTTTTTTTTTTTTTTTTTTTT-TT3′) at a 1:1.5 molar ratio in a buffer containing 50 mM Tris (pH 8.0), 150 mM KCl, 3 mM DTT, 1 mM ATP and 2 mM MgCl$_2$. Following incubation on ice for 10 min, the LcTwinkle–DNA complex was loaded onto a Superose6 increase column (10/300 GL, GE Healthcare) equilibrated in the same buffer. The eluted LcTwinkle–DNA complex was concentrated to 1 μM (~0.4 mg/ml), as determined by the Bradford assay. A 3 μl aliquot of LcTwinkle–DNA sample was deposited onto a freshly glow-discharged Quantifoil R1.2/1.3 300 mesh grid and blotted using a Vitrobot (FEI) with the standard Vitrobot filter paper, Ø55/20 mm (Ted Pella). The blotting time was set to 4 s, the blotting force was set to 4 and the blotting was done under 100% humidity at 20°C. The grids were flash-frozen in liquid ethane and stored in liquid nitrogen.

Cryo-EM data collection and processing

A total of 6490 micrographs of LcTwinkle–DNA complexes were collected on a Titan Krios electron microscope operated at 300 kV (cryo-EM core facility at the University of Texas McGovern Medical School) using the super-resolution mode with a nominal magnification of 130 K (calibrated pixel size of 1.07 Å on the sample level, corresponding to 0.535 Å in super-resolution mode). Movies were recorded with a K2 Summit camera, with the dose rate at the detector set to 7 e⁻/Å²/s. The total exposure time for each video was 7 s, which was fractionated into 35 frames of subimages. The defocus values ranged between 0.6 and 3 μm. MotionCor2 (42) was used for drift correction and electron dose weighting for all movies. The defocus values were estimated on non-dose-weighted micrographs with Gctf (43). A total of 1 976 802 particles were picked from 6334 manually screened micrographs with Gautomatch (developed by Kai Zhang, http://www.mrc-lmb.cam.ac.uk/kzhang/Gautomatch/). After 2D classification with RELION 3.0 (41), 939 599 particles were selected. An ab initio hexamer LcTwinkle model was generated in cryoSPARC (44). The particles from 2D classification were

Subunit doping

Cyan fluorescent protein (CFP)–Twinkle and yellow fluorescent protein (YFP)–Twinkle were purified similarly to WT Twinkle. For the fluorescence resonance energy transfer (FRET) experiments, the individual CFP–Twinkle and YFP–Twinkle proteins or the mixture of two were prepared in a buffer containing 500 KCl, 50 mM Tris–HCl (pH 8.0) and 3 mM DTT; 1 mM ATP and 2 mM MgCl$_2$ were added when indicated. The protein concentration was 0.2 mg/ml for each of the proteins. The protein sample was scanned at room temperature using a fluorometer (Cary Eclipse Fluorescence Spectrophotometer, Agilent, Santa Clara, CA, USA). During the scan, a 430 nm filter was used as the excitation wavelength and the emission between 430 and 570 nm was collected. For subunit doping experiments, WT and mutant Twinkles were buffer exchanged to 500 KCl, 50 mM Tris–HCl (pH 8.0) and 3 mM DTT. The total protein concentration after mixing was 2 μM. All mixed proteins were incubated at room temperature for 5 min before the activity assay. The ATPase, helicase and DNA binding activities were determined as described above. Mixed LcTwinkle at 0.2 μM was used for the helicase assay and at 0.05 μM for the ATPase assay.

Negative staining EM

WT and mutant Twinkle were diluted to 0.02 mg/ml in a buffer containing 50 mM Tris (pH 8.0), 150 mM KCl, 3 mM DTT, 1 mM ATP and 2 mM MgCl$_2$. The ssDNA substrate (5′-TGGTCTTTTTTTTTTTTTTTTTTTTTTTTT-3′) was added at a 1:2 molar ratio relative to the LcTwinkle hexamers. A 3 μl aliquot of LcTwinkle–DNA sample was deposited onto a freshly glow-discharged CF400-CU carbon grid (EMS Inc.). After absorbing the extra protein solution with filter paper, 3 μl of 2% uranyl acetate were added to the grid. Following a 20 s incubation, the extra uranyl acetate was removed with filter paper. The grids were imaged on a JEOL2100 microscope operated at 200 kV voltage and ×400 000 magnifications. The contrast transfer function (CTF) estimation and 2D classification were performed with RELION 3.0 (41).
classified into six 3D classes with the RELION 3D classification tool (41). Particles from two of the six classes (one in hexamer form and the other in heptamer form) were selected for subsequent processing. Further 3D classification was performed with cryoSPARC (44) to clean the datasets and separate different conformations of LcTwinkle. For the LcTwinkle–DNA complex, the particles from cryoSPARC were analyzed with Ctf Refine and polishing in RELION (41). Final refinements were done with the non-uniform refinement in cryoSPARC (44).

Model building and refinement
A homolog model of LcTwinkle was generated with the I-TASSER server (https://zhanglab.ccmb.med.umich.edu/I-TASSER/) and used for the initial rigid-body search. ssDNA coordinates adapted from the gp4 DNA structure (PDB ID: 6N7V) were used as the starting model for DNA. Each of the protein chains and the DNA were manually docked into the cryo-EM density maps in Chimera (45). The models were first manually adjusted in COOT (46), with real-space refinement and secondary structure and geometry restraints. Due to the low resolution of the NTDs, polyalanine models were used during the refinement, except for the residues at the NTD–CTD interfaces with good side chain density. For the lower resolution LcTwinkle–DNA2, LcTwinkleα, and LcTwinkleβ structures, the monomers from the refined LcTwinkle–DNA model were docked into the cryo-EM density with Chimera (45) and the connections were adjusted in COOT (46). Statistics of all cryo-EM data collection and structure refinement are shown in Supplementary Table S1.

High-speed atomic force microscopy imaging in liquids
LcTwinkle (35 nM hexamer) was incubated with the linear DNA substrate containing a 37 nt ssDNA gap positioned at 23% from one DNA end (48) in Twinkle Reaction Buffer (20 mM HEPES pH 7.6, 150 mM NaCl, 7.5 mM MgCl2) containing ATP (4 mM) for 1 min. The sample was diluted in Twinkle Reaction Buffer and deposited onto a freshly prepared 1-(3-aminopropyl)silatrane (APS)-treated mica surface. The sample was incubated with mica (SPI) for 2 min. The APS–mica surface containing the sample was washed with Twinkle Reaction Buffer and scanned in Twinkle Reaction Buffer using a Cypher VRS AFM (Asylum Research). BlueDrive Photothermal Excitation was used to drive a BioloVer fast (AC10DS) cantilever with the resonance frequency (f) at ~1500 kHz and the spring constant (k) at ~0.1 N/m. The images were scanned at 0.8–2 frames/s. The images were analyzed using commercially available Asylum MFP3D software.

RESULTS
The overall structure of Twinkle helicase
Previously, poor protein solubility hampered structural studies of Twinkle (34,49). Although a high-resolution structure of HsTwinkle was reported recently (39), it is with a disease-related mutant W315L and the DNA and ATP are absent. We attempted to identify a Twinkle homolog similar to the HsTwinkle and amenable to high-resolution structural determination. Sequence alignment suggested that vertebrate Twinkle proteins share high identities and may have similar structures and functional properties (Supplementary Figure S1). We picked Hs, Mm, Dr and LcTwinkle homologs to test their stability and protein–DNA complex formation. LcTwinkle shares 65% identity and 90% similarity with HsTwinkle, with most of the essential residues for Twinkle activities and disease-related mutations conserved (Supplementary Figure S1B). Two variants of LcTwinkle sequences were predicted by different databases, one longer variant and one shorter variant lacking the N-terminal mitochondrial targeting sequence (MTS) and the zinc-binding domain (ZBD) (Figure 1A; Supplementary Figure S1C). The core regions, including the helicase and primase-like domains, are identical in the two variants. The shorter LcTwinkle was chosen for cloning. Hs, Mm, Dr and LcTwinkle homologs with active site mutations corresponding to gp4 E343Q (50) were preliminarily purified with an affinity column. Considering the similarities between gp4 and Twinkle, the same ssDNA substrate as in gp4 structure determination (5′-TGGTCTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT-3′) (21) was used for Twinkle complex formation, although the primase recognition motif TGGTC turned out to be unnecessary at the end. Notably, among all Twinkle proteins tested, only LcTwinkle migrated as a single sharp peak in the gel filtration column in the presence of ATP and ssDNA (Figure 1B; Supplementary Figure S2).

WT and E325Q LcTwinkle were thoroughly purified for further characterization. The purified WT LcTwinkle displayed robust ATP-dependent helicase activity (Figure 1C). E325Q mutation eliminated the helicase-unwinding activity of LcTwinkle (Figure 1C). Although WT LcTwinkle can bind ssDNA in the presence of non-hydrolyzable ATP analogs (Supplementary Figure S2D), the resulting ssDNA complex was unstable. It migrated as a much higher molecular weight peak in gel filtration (Supplementary Figure S2E). Consistently, the WT LcTwinkle–ssDNA–ATPγS complexes appeared as aggregates on cryo-EM grids (Supplementary Figure S2F). Thus, the purified E325Q LcTwinkle–DNA–ATP complex from gel filtration (Figure 1B) was used for cryo-EM analysis. We determined the LcTwinkle–DNA complex structure with cryo-EM at 3.5 Å resolution (Figure 1D; Supplementary Figure S3; Supplementary Table S1). The local resolutions of the C-terminal helicase domains (CTDs) were near 3 Å (Supplementary Figure S3E). Atomic structures of all six CTDs can be built (Figure 1D). The N-terminal primase-like domains (NTDs) were at 5–8 Å resolution (Supplementary Figure S3E). Atomic structures of all six CTDs can be built (Figure 1D). The N-terminal primase-like domains (NTDs) were at 5–8 Å resolution (Supplementary Figure S3E). The α-helices in the NTDs can be traced, and bulky side chains near the NTD–CTD interfaces are visible (Supplementary Figure S4A). The homology model of the LcTwinkle NTD can be confidently docked into the cryo-EM densities in four out of the six subunits (Figure 1D). The CTDs of LcTwinkle are similar to the apo structures of HsTwinkle and AlphaFold (51) predicted structures, with root mean square deviations (RMSDs) ~1 Å (Supplemen-
Figure 1. Overall structure of LcTwinkle. (A) Domain structures of Hs and LcTwinkle. The disease-related Twinkle mutations are marked as red bars. MTS, mitochondrial targeting sequence; Zn, zinc-binding domain; Pri, primase; Hel, helicase; C-tail, C-terminal tail. (B) E325Q LcTwinkle–DNA complex formation revealed by size exclusion chromatography. (C) Helicase activity of WT and E325Q LcTwinkle. Twinkle hexamers at 25–400 nM were used for the assays. (D, E) Two views of the LcTwinkle–DNA complex. The subunits A to F are color-coded from blue to red, following their order on DNA. DNA is colored orange. The NTDs from subunits E and F are disordered and their positions are marked by semi-transparent triangles. The disordered part of the N–C loop in subunit F is indicated by a dashed line. (F) Structure of LcTwinkle–DNA. The mobile subunit F and the structural gaps between F and its neighboring subunits are highlighted. (G, H) Structures of the apo LcTwinkle hexamer (G) and heptamer (H). The structural gaps in the oligomers are marked.

LcTwinkle forms a homo-hexamer wrapping DNA in its central channel (Figure 1D). The six subunits are labeled A–F following their order on DNA in the 3′ to 5′ direction. The NTDs are attached to the side of the CTDs. The domain structure of LcTwinkle is consistent with previously published low-resolution structural models of Twinkle (34,37,38), but different from the apo structure of HsTwinkle and predicted Twinkle structures (39), where the NTD interacts with the CTD from the same subunit (Supplementary Figure S5A). This arrangement of the NTDs is also distinct from all the other replicative helicases, in which the NTDs are all located on top of the helicase rings at the 5′ side of the DNA (Supplementary Figure S5D–I) (21–24,26). Compared with gp4, the LcTwinkle NTD swings nearly 90° to the side (Supplementary Figure S5J). Furthermore, the pseudo-primase active site is flipped to face downward,
making interactions with the unwound DNA impossible (Supplementary Figure S5J). An extended linker connects the NTD and the CTD (N–C linker) in a domain-swapped manner (Figure 1D, E). The N–C linker from one subunit interacts with the CTD of the second subunit on its 5′ side of DNA. The CTDs form a non-planar lock-washer-shaped hexamer (Figure 1D), similar to gp4 and DnaB (21, 22). The subunits at the two ends of the lock-washer are bridged by the N–C linker, making the hexamer a complete circle (Figure 1D), ATPase sites are located at each of the subunit interfaces (Figure 2A), and five ATP molecules can be refined (Supplementary Figure S4E). A 12 nt ssDNA binds to the central channel of LcTwinkle, with an average of 2 nt per subunit (Figure 2A). The DNA conformation in Twinkle is similar to that in gp4, DnaB, CMG and MCM helicases, but distinct from the DNA in E1 helicase (Supplementary Figure S5K).

Careful 3D classification yielded three additional structures of LcTwinkle with distinct conformations (Figure 1F–H; Supplementary Figure S3A). One structure obtained at 7.5 Å resolution contains DNA but exhibits altered CTDs (LcTwinkle–DNA2). In LcTwinkle–DNA2, the terminal helicase subunit (subunit F) has departed from subunit E and travelled half-way to the 3′ end of the DNA (Figure 1F; Supplementary Figure S6A). This structure may represent an intermediate state during Twinkle translocation. Although the LcTwinkle sample was prepared in the presence of DNA, two apo LcTwinkle structures were identified, with one in the hexamer form (LcTwinkle6) and the other in the presence of DNA, two apo LcTwinkle structures were identified, with one in the hexamer form (LcTwinkle6) and the other existing as a heptamer (LcTwinkle7) (Figure 1G, H). These structures are similar to what was previously observed with apo Twinkle (34, 37–39). LcTwinkle6 stays in a similar non-planar lock-washer configuration as the LcTwinkle–DNA complex (Figure 1G; Supplementary Figure S6B), different from the planar rings formed by gp4 or DnaB in the absence of DNA (52–54). LcTwinkle7 at 8.5 Å resolution contains an enlarged central channel of ~40 Å (Figure 1H), similar to the HsTwinkle heptamer (29). Interestingly, a structural gap is also present in LcTwinkle7. However, the NTDs are completely disordered in both apo structures, suggesting that the NTDs are highly dynamic without DNA (Figure 1G, H). It has been reported that in the absence of ssDNA, the NTD conformation is controlled by ATP binding (38). The NTDs are ordered without ATP but become disordered in the presence of ATP (38, 39), consistent with our structures in the presence of ATP.

ATP- and DNA-binding interfaces of Twinkle

The LcTwinkle helicase contains five intact ATP-binding sites, each formed by residues from two adjacent subunits (Figure 2A). One Mg2+ is associated with each ATP molecule (Supplementary Figure S4E). The Walker A motif from one subunit wraps the triphosphate group of the ATP with its backbone amide groups (Figure 2B). The K301 side chain on the Walker A motif directly interacts with the γ-phosphate of the ATP. The T302 helps coordinate the Mg2+ ion. The D396 on the Walker B motif is ~4 Å from the Mg2+ and may coordinate the Mg2+ through water-mediated interaction. The mutated Q325 interacts with the γ-phosphate group of the ATP and Mg2+. Two side chains, F501 and R481, sandwich the adenine base of the ATP (Figure 2B, Supplementary Figure S7A). Similar to HsTwinkle (35), LcTwinkle is promiscuous and can utilize nucleotides with either an adenine or thymine base (Supplementary Figure S7B). In contrast, gp4 prefers dTTP over ATP (55). The R481 on LcTwinkle comes from a secondary structural element different from that in gp4 and the F501 is the equivalent of the gp4 Y535 (Supplementary Figure S7A). Mutations of F501A and R481A reduce the Kd of ATP by ~10-fold (Supplementary Table S2). Surprisingly, F501Y reduces ATP binding as well and making LcTwinkle favor the adenine over the thymine base (Supplementary Figure S7B). The K487 and R489 residues (the arginine finger) from the adjacent subunits contact the γ-phosphate of the ATP. In addition, the H438 side chain and the Q461 backbone are also in proximity to the γ-phosphate and may stabilize the water for nucleophilic attack during ATP hydrolysis (Figure 2B). Mutations of R489A and H438A significantly reduce the ATPase activity and eliminate DNA binding and helicase activities of LcTwinkle, confirming their critical roles in ATP sensing and translocation (Figure 2C, D; Supplementary Table S2).

DNA binding in LcTwinkle is mediated by two adjacent loops (LoopD1 and LoopD2 in Figure 2A and B). The G455 and the A457 backbone amide groups and the S456 side chain from LoopD1, and the R440 and the K441 from LoopD2 interact with the phosphate backbones of DNA (Figure 2B). In addition, the Y412 side chain contacts the O4′ atom on the sugar ring (Figure 2B). The DNA-binding residues in each subunit span a distance of 4 nt, with residues on LoopD1 and the R440 on LoopD2 interacting with two nucleotides, and the K441 side chain grabbing two additional nucleotides on the 3′ side of DNA (Figure 2B). DNA bound to subunit F lacks the interaction with K441 due to the absence of a 5′-end neighbor. Two similar DNA-binding loops were found in gp4 helicase (Supplementary Figure S7C) (21). However, the corresponding position of F454 is an arginine in gp4, which is essential for gp4 DNA binding (56). LoopD2 in gp4 contains additional positively charged residues and packs on the 3′ side of the coiled ssDNA (Supplementary Figure S7C) (57). Besides directly binding DNA, the positively charged LoopD2 was proposed to facilitate the large-scale domain movement during translocation and prevent DNA from slipping out of gp4 (Supplementary Figure S7D) (27, 57). In contrast, LoopD2 in LcTwinkle is shorter and negatively charged. The DNA is exposed when observed from the C-terminal side of LcTwinkle (Supplementary Figure S7E). Overall, our structure suggested that LcTwinkle has reduced DNA interactions compared with that of gp4. Direct mutagenesis confirmed the structural model of Twinkle–DNA interaction. Mutations R440A and S456A eliminate DNA binding and helicase activities (Figure 2C, D). K441A LcTwinkle had low solubility and could not be purified.

Each LcTwinkle helicase subunit is similar within the LcTwinkle hexamer, with an average RMSD of ~0.2 Å (Supplementary Figure S6C). The DNA-binding loops are superimposable, except that K441 from subunit A at the DNA 3′ end takes a different conformation (Figure 2E). Moreover, the ATP-binding sites at the domain interfaces
Figure 2. ATP- and DNA-binding interfaces of LcTwinkle. (A) ATP- and DNA-binding sites in the LcTwinkle–DNA complex. The CTDs are shown as transparent surfaces and structural elements involved in ATP and DNA binding are shown in the cartoon. (B) Zoom-in view of the ATP- and DNA-binding sites between subunits C and D. Key residues are shown as sticks. WA, WB, D1 and D2 stand for Walker A, Walker B, LoopD1 and LoopD2 motifs, respectively. The position of the predicted water molecule for the nucleophilic attack is highlighted by a dotted circle. (C) Helicase activity of the WT and the active site mutants of LcTwinkle. (D) Fluorescence anisotropy DNA binding of the WT and the active site mutants of LcTwinkle. (E) DNA-binding loops from different subunits of the LcTwinkle–DNA complex take the same configuration. (F) Structural alignments of ATP-binding sites across the hexamer suggest that they are superimposable. The five ATP-binding sites were extracted from the LcTwinkle–DNA complex, and subunits at the 5′ side of the DNA in each dimer are aligned. The 5′-side subunit is shown as gray surfaces, and the active site residues on the 3′ side subunits are drawn as cartoons and sticks. (G, H) Subunit doping experiments of ATPase (G) and helicase (H) activities with WT and E325Q LcTwinkle. Mixed LcTwinkle at 0.2 μM was used for the helicase assay and 0.05 μM for the ATPase assay.

are almost identical (Figure 2F; Supplementary Figure S6D). This is distinct from most other hexameric helicases, where gradual conformational changes of ATP-binding sites along the helicase hexamer were thought to correspond to sequential ATP hydrolysis and unidirectional translocation (21,23–26). When the planar apo gp4 is used as a reference, one subunit in an LcTwinkle dimer rotates 14° relative to its neighboring subunit (Supplementary Figure S6E). As a comparison, the gp4 subunits rotate 17–23° from the planar conformation (Supplementary Figure S6E). The almost identical configurations of the ATP-binding sites in the LcTwinkle hexamer suggested that all subunits have a similar chance of hydrolyzing ATP. Moreover, the similar hexameric structures with and without DNA (Supplementary Figure S6B) indicated that DNA binding would not significantly stimulate Twinkle ATPase activity. Notably, gp4 dTTPase activity is stimulated 40- to 100-fold by DNA, whereas Twinkle ATPase activity is only increased by <2-fold upon DNA binding (Supplementary Table S2) (35,50).

To further investigate the mechanism of LcTwinkle ATP hydrolysis and translocation, we performed subunit doping experiments (50,58). When mutated enzymes with defective catalysis were titrated to the WT enzymes, the change in the activity level of the hetero-oligomer correlates with the subunit cooperativity. If the nucleotide hydrolysis is stochastic within the oligomer, the activity will decrease linearly; on the other hand, when the nucleotide hydrolysis is highly correlated, the activity will change exponentially. Previous subunit doping experiments indicated that the dTTPase activity is highly correlated in gp4 helicase (50), whereas ATP hydrolysis and translocation in archaeal MCM helicase is only moderately correlated (58). A prerequisite for subunit doping experiments is the efficient formation of the hetero-oligomeric helicase complex. To confirm
that different types of LcTwinkle can form proper hetero-oligomers, we constructed and purified LcTwinkle with either N-terminal CFP or YFP (Supplementary Figure S8A, B). Mixing of the two types of LcTwinkle produced a FRET signal at a wavelength of ~536 nm, corresponding to YFP emission, while the CFP peak at ~470 nm was reduced (Supplementary Figure S8C). Moreover, the FRET signal emerged within 2 min upon mixing (Supplementary Figure S8D), suggesting efficient and fast hetero-oligomer formation. Similar to the gp4 subunit doping experiments (30), we picked E325Q LcTwinkle, which eliminates ATP hydrolysis but does not affect the ATP or the DNA binding (Figure 2C). When E325Q LcTwinkle is titrated to the WT LcTwinkle, both the ATPase activity and the helicase activity drop linearly (Figure 2G, H). The linear decrease suggested a mechanism of stochastic ATP hydrolysis within the LcTwinkle hexamer. Similar linear decreases for the ATPase and helicase activities were observed with H438A and R489A, despite their different roles in ATP hydrolysis and translocation (Supplementary Figure S8). In addition, we tested subunit doping with the DNA binding mutant R440A (Supplementary Figure S8I, J). The DNA binding affinity is reduced but much slower than the predicted linear decrease when R440A LcTwinkle was titrated. The helicase activity decreases linearly with the increasing amount of the R440A LcTwinkle, similar to those of ATPase site mutants.

**Disease-related Twinkle mutations**

Nearly 60 mutations on Twinkle have been implicated in human diseases (Figure 1A; Supplementary Figure S1B). Disease-related mutations on HsTwinkle can be mapped onto the LcTwinkle structure, with most of them conserved (Figure 3; Supplementary Figure S1B). The N–C linker and the NTD form extensive interactions with the CTD of its 3′-side subunit (Figure 3A). The domain-swapped interface is a hotspot for disease-related mutations (Figure 3A). The NTD directly contacts the CTD with an α-helix formed by residues 195–204 (Figure 3A; Supplementary Figure S9A). Although the overall resolution is low for the NTDs, the helix is well ordered, and densities for the W195 (HsTwinkle W315) and the K199 (HsTwinkle K319) side chains are visible (Supplementary Figures S4A and S9A). W195 (HsTwinkle W315) is in proximity to the Y386 (HsTwinkle Y506) side chain, and K199 (HsTwinkle K319) interacts with the backbone oxygen of L387 (HsTwinkle V507) (Supplementary Figure S9A). Residues near the interface, including W195 (HsTwinkle W315), S198 (HsTwinkle A318), K199 (HsTwinkle K319), R214 (HsTwinkle R334) and P215 (HsTwinkle P335) on the NTD, and L387 (HsTwinkle V507) and Y388 (HsTwinkle Y508) on the CTD, have been found in human patients (Figure 3A). Although Y386 (HsTwinkle Y506) mutation has not been correlated with human disease, Y386A reduces the helicase activity similarly to the disease-related mutations K199E (HsTwinkle K319), W195L (HsTwinkle W315) and Y388C (HsTwinkle Y508) (Supplementary Figure S9B). The N–C interface may stabilize Twinkle hexamer during translocation (38). WT LcTwinkle forms hexamers and heptamers on the negative staining EM grid (Supplementary Figure S9C, E). In contrast, Y386A (HsTwinkle Y506), K199E (HsTwinkle K319), W195L (HsTwinkle W315) and Y388C (HsTwinkle Y508) LcTwinkle samples contain significantly reduced oligomers, with the abundant oligomer form of these mutants being heptamers or larger oligomers (Supplementary Figure S9C–J), similar to what was observed for HsTwinkle (38,39). In contrast, the NTDs are located at different places relative to the CTD, and the N–C interface is not preserved in the apo HsTwinkle structure (Supplementary Figure S9K), possibly due to the W315L mutation (corresponding to LcTwinkle W195L) on the N–C interface (39).

The N–C linker near the NTD is sandwiched by the NTD and the CTD (Figure 3A). Six mutations on the NTD–linker junction are related to human diseases. The C-terminal end of the N–C linker forms a helix and stacks in a hydrophobic groove on the CTD (Figure 3A). L261 (HsTwinkle L381), F250 (HsTwinkle F370) and I247 (HsTwinkle I367) interact with the CTD through hydrophobic interactions, while R254 (HsTwinkle R374) forms salt links with D352 (HsTwinkle D472) (Figure 3A). Mutations of L261 (HsTwinkle L381), R254 (HsTwinkle R374), F250 (HsTwinkle F370), S249 (HsTwinkle S369) and I247 (HsTwinkle I367) on the N–C linker, and R343 (HsTwinkle R463), A340 (HsTwinkle A460), A355 (HsTwinkle A475), L336 (HsTwinkle L456), T337 (HsTwinkle T457), Q338 (HsTwinkle Q458), W354 (HsTwinkle W474), F358 (HsTwinkle F478), L361 (HsTwinkle L481) and F365 (HsTwinkle F485) on the CTD cause human diseases. The linker in the HsTwinkle structure interacts with the CTD through the same interface, although the linker helix is lifted ~1 Å and the interface is reduced (Supplementary Figure S9J). The N–C linker may play a similar role to that of the NTDs in stabilizing Twinkle hexamer, as suggested by previous biochemical and structural data (38).

Besides mutations at the domain-swapped interface, 11 mutations are located on the CTD (Figure 3B). Residues W321 (HsTwinkle W441), N318 (HsTwinkle N438), S306 (HsTwinkle S426) and F304 (HsTwinkle F424) are at the core of the CTD and may be important for CTD folding and stability. R423 (HsTwinkle R543), T291 (HsTwinkle T411), R446 (HsTwinkle K566), N279 (N399) and R280 (HsTwinkle R400) are near the subunit–subunit interface on the CTD. These residues may contribute to the CTD subunit–subunit interactions and ATP hydrolysis. In addition, seven mutations are mapped to the surface of the NTD, and most of them are charged residues (Figure 3C). Of note, A243 (human disease W363L), R228 (human disease G348R) and R446 (human disease K566R) are different from the corresponding human residues. The differences in A243 (HsTwinkle W363) and R228 (HsTwinkle G348) are possibly due to changes in their surrounding residues (Supplementary Figure S9L, M). R446 (HsTwinkle K566) is part of LoopD2 but does not directly contact DNA (Supplementary Figure S7C). How human K566R mutation would affect Twinkle function is unclear.

**HS-AFM imaging of LcTwinkle DNA binding dynamics**

Our recent AFM imaging in liquids showed that human Twinkle subunits could self-assemble into hexamers and higher order complexes, and switch between open and closed configurations (59). However, due to the low tempo-
Figure 3. Disease-related mutations on LcTwinkle. (A) Disease-related mutations on the N–C interface and the N–C linker. Two adjacent subunits C (green) and D (yellow) are selected to illustrate the positions of the disease-related mutations. The mutations on subunit C are highlighted in blue and the mutations on subunit D are highlighted in orange. The corresponding human disease-related mutations are indicated in parentheses. The locations of each panel relative to the overall structure are indicated on the small diagrams in the top right corner in (A). (B) Disease-related mutations on the CTD. (C) Disease-related mutations on the NTD.

Figure 3. Disease-related mutations on LcTwinkle. (A) Disease-related mutations on the N–C interface and the N–C linker. Two adjacent subunits C (green) and D (yellow) are selected to illustrate the positions of the disease-related mutations. The mutations on subunit C are highlighted in blue and the mutations on subunit D are highlighted in orange. The corresponding human disease-related mutations are indicated in parentheses. The locations of each panel relative to the overall structure are indicated on the small diagrams in the top right corner in (A). (B) Disease-related mutations on the CTD. (C) Disease-related mutations on the NTD.

Figure 3. Disease-related mutations on LcTwinkle. (A) Disease-related mutations on the N–C interface and the N–C linker. Two adjacent subunits C (green) and D (yellow) are selected to illustrate the positions of the disease-related mutations. The mutations on subunit C are highlighted in blue and the mutations on subunit D are highlighted in orange. The corresponding human disease-related mutations are indicated in parentheses. The locations of each panel relative to the overall structure are indicated on the small diagrams in the top right corner in (A). (B) Disease-related mutations on the CTD. (C) Disease-related mutations on the NTD.

...eral resolution of the conventional AFM imaging, we could not elucidate the real-time conformational changes of Twin-...
**DISCUSSION**

Most replicative helicases contain NTDs in addition to the C-terminal helicase domains (Supplementary Figure S5) (21–24,26). The gp4 NTD encodes a primase domain, whereas the NTDs in bacterial, archaeal and eukaryotic replicative helicases physically interact with the primases (52,61). The NTDs are always on the 5’ side of the DNA relative to the helicase domains to assist primer synthesis (Supplementary Figure S5). Like gp4, Twinkle helicases from plants and lower eukaryotes can catalyze primer synthesis (62,63). However, the primase active site residues are mutated in vertebrate Twinkle (11). Furthermore, the NTD rotates 90° and flips 180° away from the unwound DNA (Supplementary Figure S5J). The unique NTD arrangement is consistent with the lack of Okazaki fragment synthesis in animal mitochondria (3,4). Moreover, the NTD conformations are likely to be controlled by both DNA and ATP. The NTDs are on top of the CTD ring in apo HsTwinkle structures (39). In our LcTwinkle structures prepared with a high concentration of ATP, the NTDs are disordered in the absence of DNA (Figure 1G) but attach to the side of the CTD ring in the presence of DNA (Figure 1D).

Although they do not have enzymatic activities, the Twinkle NTDs probably play important roles in Twinkle oligomerization, additional DNA binding and protein–protein interactions, similar to the NTDs in gp4 and DnaB (52,64). Our LcTwinkle structure shows that the NTD and the N–C linker directly contact the CTD in its neighboring subunit in the DNA-bound state (Figures 1D and 3A). The interaction stabilizes Twinkle hexamers around DNA. Mutations at the interface alter oligomerization of Twinkle and reduce helicase unwinding (Supplementary Figure S9).
(38). The domain-swapped interface is a hotspot for disease-related mutations (Figure 3A). In addition, the NTD may facilitate Twinkle DNA capture and loading (as discussed below). Moreover, strand annealing and exchange activities have been reported for human Twinkle (35), and these activities require more than one DNA-binding interface. The six NTDs may capture multiple ssDNA molecules to assist their pairing. In addition, Twinkle is reported to interact with multiple factors in DNA replication and repair (6,16–18). Any proteins operating on the lagging strand DNA may contact Twinkle through the NTD. Seven disease-related mutations are mapped onto the NTD. They are far away from the CTD or the NTD–CTD interface, and thus not likely to be directly involved in Twinkle unwinding. These mutations may affect Twinkle protein or DNA interactions.

Loading of replicative helicases onto the genomic DNA is a prerequisite and often a critical regulatory step in DNA replication (29). In the nucleus of bacteria, archaea and eukaryotes, an origin recognition complex recognizes the replication origin and recruits the helicase. In addition, a helicase loader helps open the helicase ring for loading (29). Neither an origin recognition complex nor a helicase loader has been identified in mitochondria. Twinkle itself must search and capture the DNA substrate in 3D. Remarkably, HS-AFM imaging in liquids uncovered a novel proximal DNA-induced conformational change of Twinkle. Using HS-AFM, we observed that when LcTwinkle is close to DNA (LcTwinkle–DNA distance <20 nm), a domain could move freely in solution and protrude ~5 nm in length, away from the major portion of LcTwinkle oligomers to search for and capture nearby DNA (Figure 4). We propose that the mobile domain is the NTD. The NTDs do not interact with each other (Figure 1D), while the CTDs can oligomerize in the presence of ATP, which was included in our AFM imaging buffer. Our apo LcTwinkle and previous apo HsTwinkle structures (39) confirmed the dynamic nature of the NTDs (Figure 1G, H). The N–C linker spans >30 residues and can account for the long-distance protrusion when it becomes fully extended. Therefore, we termed the domain protrusion observed in HS-AFM as the ‘N-protrusion’ (Figure 4; Supplementary Videos S2–S6). The N-protrusion is probably guided by electrostatic interactions between the NTDs and DNA. There are several positively charged patches on the NTD that may potentially interact with the negatively charged DNA electrostatically over long distances (Figure 4F). It is worth noting that the Debye screening length around dsDNA is ~1 nm at the ionic strength used in HS-AFM imaging (65). However, the Debye screening length for DNA on the mica surface is unknown. In addition, since both proteins and DNA were mobile during AFM imaging, the precise distance between LcTwinkle and DNA that activates N-protrusion could be significantly shorter than what we measure based on individual AFM images. The length of the N-protrusion could also be augmented by mobile DNA attached to the NTD.

The extraordinary conformational change of the N-protrusion suggests a plausible mechanism for Twinkle DNA loading in the absence of a helicase loader (Figure 5). Individual NTDs from an apo Twinkle molecule either sit on its own CTD without domain swap (39) or extend out to search for nearby DNA (Figure 4; Supplementary Videos S2–S6). After DNA capture, the protruded NTD helps bring the DNA into its central channel for loading (Figure 4D; Supplementary Video S6). In addition, the Twinkle hexameric ring connected by the N–C linkers must open to allow DNA entrance. Apo LcTwinkle is likely to be prone to ring opening due to the lack of domain-swapped N–C interactions. HS-AFM revealed that LcTwinkle could switch between open and closed ring conformations in the absence of DNA (Figure 4A, panels I and II; Supplementary Video S2). Our HS-AFM data also confirmed that Twinkle could capture DNA at its central channel through ring opening (Figure 4D; Supplementary Video S6) without subunit association or dissociation. Taken together, the cryo-EM and HS-AFM data suggest two separate DNA-binding regions on Twinkle, the NTD for initial transient DNA capture and the helicase central channel for DNA unwinding. Our model of Twinkle loading (Figure 5) is also consistent with a model proposed for gp4 helicase, which involves the initial binding of the DNA to the NTD outside of the helicase ring, a conformational transition followed by the migration of the DNA into the central channel, and ring closure (31). Although not directly observed, we cannot exclude alternative possibilities of LcTwinkle loading by assembling of monomers or dissociation of a subunit from a heptamer.

The ATP binding and hydrolysis empower helicase translocation on DNA. While several models of ATP hydrolysis have been proposed in hexameric ATPases, the sequential ATP hydrolysis model is widely accepted (66). In contrast, our structural and biochemical analyses suggest that the ATP hydrolysis is stochastic in LcTwinkle. The random ATP hydrolysis was also reported for the hexameric ClpX peptide translocase, where the loss of catalytic residues in one or more ATP-binding sites only reduces but does not eliminate translocation (67). Similarly, ATP-binding sites in the hetero-hexameric CMG helicase are not all required for its translocation (26). Helicase translocation is associated with frequent futile cycles, slipping and backtracking (55). Possibly, ATP hydrolysis near the 5′ end of DNA will lead to translocation of one to two subunits, whereas ATP hydrolysis in the middle or close to the 3′ end may result in futile ATPase cycles (Figure 5). The en bloc movement of multiple subunits also explains how inactive subunits can be tolerated in Twinkle. CMG and ClpX. Twinkle can move on DNA unidirectionally in the 5′ to 3′ direction (32). Our LcTwinkle–DNA structure suggested that subunit F on the 5′ end of DNA has reduced DNA binding, as one of the key DNA-binding residues K441 from a neighboring subunit is missing at the 5′ end of DNA. It is possible that the subtle difference in DNA binding renders subunit F most mobile. On the other hand, the directional movement could be affected by the NTDs. SF4 G40P helicase has a minimal NTD and can translocate in both directions on ssDNA (68). In addition, a comparison of DNA-binding loops in gp4 and LcTwinkle also suggested that DNA may easily slide out from the Twinkle DNA-binding channel (Supplementary Figure S7E). Thus, our structural and biochemical analyses suggest that Twinkle is inefficient in unwinding DNA due to the stochastic ATP hydrolysis and reduced DNA binding. Indeed, Twinkle ATP hydrolysis is 20- to 100-fold slower.
Figure 5. Model of LcTwinkle loading and translocation. Twinkle in the absence of DNA is prone to ring opening (State I). When DNA is nearby, the positively charged NTD captures the negatively charged DNA over long distances and the CTD hexamer opens further for DNA loading (State II). After loading, Twinkle forms a lock-washer-shaped hexamer with the NTDs attaching to the side of the CTDs and the DNA binding to the CTD central channel (State III). ATP hydrolysis is stochastic in Twinkle. When ATP is hydrolyzed at the subunit interfaces near the DNA 5’ end (State IV and State VI), one or two subunits dissociate from their neighbor and travel to the other end of the DNA (State V and State VII). However, when ATP is hydrolyzed in the middle or close to the DNA 3’ end (State VIII), the translocation is unfavored, leading to a futile cycle of ATP hydrolysis.

than that of gp4 or DnaB (35). It is possible that Twinkle only stays in a partially active form and additional binding partners or post-translational modifications fully activate Twinkle.

In summary, our biochemical, biophysical and structural data illustrate unique structural and dynamic features of the mitochondrial replicative helicase Twinkle. Our data highlight the important role of the enzymatically inactive NTD in Twinkle loading and unwinding, and the NTD–CTD interface represents a hotspot for human disease-related mutations.

DATA AVAILABILITY
All original data and materials are available upon request. The 3D cryo-EM density maps for LcTwinkle complexes have been deposited in the EM Database under the accession codes EMD-27842, EMD-27843, EMD-27844 and EMD-27845, and the coordinates for the structure have been deposited in the Protein Data Bank under accession code PDB 8E2L.

SUPPLEMENTARY DATA
Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS
We thank Wenhua Guo from the Rice University Shared Equipment Authority, Isaac Forester from the cryo-EM core facility at Baylor College of Medicine (supported by CPRIT RP190062), and Venkata Mallimapalli and Guizhen Fan from the cryo-EM core facility at the University of Texas McGovern Medical School (supported by CPRIT RP190062) for their help in cryo-EM sample preparation, screening and data collection. We thank Ryan Fuerer and Keith Jones at Asylum Research for access to Cypher VRS AFM, Yuri Lyubchenko at the University of Nebraska for sharing the APS synthesis protocol, and Dorothy Erie at UNC-Chapel Hill for sharing APS. We thank Dr. Wei Yang from the NIH for her support in the initial cloning of the LcTwinkle gene.

FUNDING
The work is supported by the Cancer Prevention & Research Institute of Texas (CPRIT) [RR190046]; the National Institutes of Health [R35GM142722 to Y.G.]; the National Institutes of Health [RO1GM123246 to H.W.]; and the P30 ES025128 Pilot Project Grant to H.W. and P.K. through the Center for Human Health and the Environment at NCSU.
**Conflict of interest statement.** The authors declare no conflicts of interest.

**REFERENCES**

1. Robberson, D.L., Kasamatsu, H. and Vinograd, J. (1972) Replication of mitochondrial DNA. Circular replicative intermediates in mouse 1 cells. *Proc. Natl Acad. Sci. USA*, 69, 737–741.

2. Gustafsson, C.M., Falkenberg, M. and Larsson, N.G. (2016) Maintenance and expression of mammalian mitochondrial DNA. *Annu. Rev. Biochem.*, 85, 133–160.

3. Yasukawa, T., Reyes, A., Cluett, T.J., Yang, M.Y., Bowmaker, M., Jacobs, H.T. and Holt, I.J. (2006) Replication of vertebrate mitochondrial DNA entails transient ribonucleotide incorporation throughout the lagging strand. *EMBO J.*, 25, 5358–5371.

4. Miralles Fuste, J., Shi, Y., Wanrooij, S., Zhu, X., Jemt, E., Persson, O., Sabouri, N., Gustafsson, C.M. and Falkenberg, M. (2014) In vivo occupancy of mitochondrial single-stranded DNA binding protein supports the strand displacement mode of DNA replication. *PLoS Genet.*, 10, e1004852.

5. Falkenberg, M. (2018) Mitochondrial DNA replication in mammalian cells: overview of the pathway. *Essays Biochem.*, 52, 286–297.

6. Korhonen, J.A., Pham, X.H., Pellegrini, M. and Falkenberg, M. (2004) Reconstitution of a minimal mtDNA replisome in vitro. *EMBO J.*, 23, 2423–2429.

7. Wanrooij, S., Fuste, J.M., Farge, G., Shi, Y., Gustafsson, C.M. and Falkenberg, M. (2008) Human mitochondrial RNA polymerase primes lagging-strand DNA synthesis in vitro. *Proc. Natl Acad. Sci. USA*, 105, 11122–11127.

8. Fuste, J.M., Wanrooij, S., Jemt, E., Granycome, C.E., Cluett, T.J., Shi, Y., Atanassova, N., Holt, I.I., Gustafsson, C.M. and Falkenberg, M. (2010) Mitochondrial RNA polymerase is needed for activation of the origin of light-strand DNA replication. *Mol. Cell.*, 37, 67–78.

9. DeBalsi, K.L., Hoff, K.E. and Copeland, W.C. (2004) Role of the mitochondrial DNA replication machinery in mitochondrial DNA mutagenesis, aging and age-related diseases. *Aging Res. Rev.*, 33, 89–104.

10. Spellbrink, J.N., Li, F.Y., Tiranti, V., Nikali, K., Yuan, Q.P., Tariq, M., Wanrooij, S., Garrido, N., Comi, G., Morandi, L. et al. (2001) Human mitochondrial DNA deletions associated with mutations in the gene encoding twinkle, a phage T7 gene 4-like protein localized in mitochondria. *Nat. Genet.*, 28, 223–231.

11. Peter, B. and Falkenberg, M. (2020) TWINKLE and other human mitochondrial DNA helicases: structure, function and disease. *Genes (Basel)*, 11, 408.

12. Tynnismaa, H., Sembongi, H., Bokori-Brown, M., Granycome, C., Ashley, N., Poulton, J., Jalanko, A., Spellbrink, J.N., Holt, I.I. and Suomalainen, A. (2004) Twinkle helicase is essential for mtDNA maintenance and regulates mtDNA copy number. *Hum. Mol. Genet.*, 13, 3219–3227.

13. Milenkovic, D., Satic, M., Kuhl, J., Ruzzenente, B., Freyer, C., Jemt, E., Park, C.B., Falkenberg, M. and Larsson, N.G. (2013) TWINKLE is an essential mitochondrial helicase required for synthesis of nascent D-loop strands and complete mtDNA replication. *Hum. Mol. Genet.*, 22, 1983–1993.

14. Baloh, R.H., Salavaggione, E., Milbrandt, J. and Pestrunk, A. (2007) Familial parkinsonism and ophthalmoplegia from a mutation in the mitochondrial DNA helicase gene. *Arch. Neurol.*, 64, 998–1000.

15. Pohjoismaki, J.L., Williams, S.L., Boettger, T., Goftart, S., Kim, J., Suomalainen, A., Moraes, C.T. and Braun, T. (2013) Overexpression of Twinkle-helicase protects cardiomyocytes from genotoxic stress caused by reactive oxygen species. *Proc. Natl Acad. Sci. USA*, 110, 19408–19413.

16. Sykora, P., Kanno, S., Akbari, M., Kulikowicz, T., Baptiste, B.A., Leandro, G.S., Lu, H., Tian, J., May, A., Becker, K.A. et al. (2017) DNA polymerase beta participates in mitochondrial DNA repair. *Mol. Cell Biol.*, 37, e00257-17.

17. Phee, V., Blei, D., Trombly, G., Corsi, S., Szukzo, M.J., Rebele-Guimard, P., Gammage, P.A., Kudin, A.P., Becker, C., Altmuller, J. et al. (2018) Linear mitochondrial DNA is rapidly degraded by components of the replication machinery. *Nat. Commun.*, 9, 1727.
39. Riccio,A.A., Bouvette,J., Perera,L., Longley,M.J., Krahn,J.M., Williams,J.G., Dutcher,R., Borgia,M.J. and Copeland,W.C. (2022) Structural insight and characterization of human twinkle helicase in mitochondrial disease. *Proc. Natl Acad. Sci. USA*, 119, e2207459119.

40. Edgar,R.C. (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.*, 32, 1792–1797.

41. Zivanov,J., Nakane,T., Forsberg,B.O., Kimanius,D., Hagen,W.J., Lindahl,E. and Scheres,S.H. (2018) New tools for automated high-resolution cryo-EM structure determination in RELION-3. *Elife*, 7, e42166.

42. Zheng,S.Q., Palovcak,E., Armache,J.P., Verba,K.A., Cheng,Y. and Agard,D.A. (2017) MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nat. Methods*, 14, 331–332.

43. Zhang,K. (2016) Gctf: real-time CTF determination and correction. *J. Struct. Biol.*, 193, 1–12.

44. Punjani,A., Rubinstein,J.L., Fleet,D.J. and Brubaker,M.A. (2017) cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. *Nat. Methods*, 14, 290–296.

45. Peetersen,E.F., Goddard,T.D., Huang,C.C., Couch,G.S., Greenblatt,D.M., Meng,E.C. and Ferrin,T.E. (2004) UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.*, 25, 1605–1612.

46. Emsley,P., Lohkamp,B., Scott,W.G. and Cowtan,K. (2010) Features and development of COOT. *Acta Crystallogr. D Biol. Crystallogr.*, 66, 480–501.

47. Afonine,P.V., Poon,B.K., Read,R.J., Sobolev,O.V., Terwilliger,T.C., Urzhumtsev,A. and Adams,P.D. (2018) Real-space refinement in PHENIX for cryo-EM and crystallography. *Acta Crystallogr. D Struct. Biol.*, 74, 531–544.

48. Countryman,P., Fan,Y., Gorthi,A., Pan,H., Strickland,J., Kaur,P., Wang,X., Lin,J., Lei,X., White,C. et al. (2018) Cohesin SA2 is a sequence-independent DNA-binding protein that recognizes DNA replication and repair intermediates. *J. Biol. Chem.*, 293, 1054–1069.

49. Ziebarth,T.D., Farr,C.L. and Kaguni,L.S. (2007) Modular architecture of the hexameric human mitochondrial DNA helicase. *J. Mol. Biol.*, 367, 1382–1391.

50. Crampton,D.J., Mukherjee,S. and Richardson,C.C. (2006) DNA-induced switch from independent to sequential dTTP hydrolysis in the bacteriophage T7 DNA helicase. *Mol. Cell*, 21, 165–174.

51. Jumper,J., Evans,R., Pritzl,A., Green,T., Figurnov,M., Ronneberger,O., Tunyasuvunakool,K., Bates,R., Zidek,A., Potapenko,A. et al. (2021) Highly accurate protein structure prediction with alphafold. *Nature*, 596, 583–589.

52. Bailey,S., Eliason,W.K. and Steitz,T.A. (2007) Structure of hexameric DnaB helicase and its complex with a domain of DnaG primase. *Science*, 318, 459–463.

53. Singleton,M.R., Sawaya,M.R., Ellenberger,T. and Wigley,D.B. (2000) Crystal structure of T7 gene 4 ring helicase indicates a mechanism for sequential hydrolysis of nucleotides. *Cell*, 101, 589–600.

54. Toth,E.A., Li,Y., Sawaya,M.R., Cheng,Y. and Ellenberger,T. (2003) The crystal structure of the bifunctional primase-helicase of bacteriophage T7. *Mol. Cell*, 12, 1113–1123.

55. Sun,B., Johnson,D.S., Patel,G., Smith,B.Y., Pandey,M., Patel,S.S. and Wang,M.D. (2011) ATP-induced helicase slippage reveals highly coordinated subunits. *Nature*, 478, 132–135.

56. Washington,M.T., Rosenberg,A.H., Griffin,K., Studier,F.W. and Patel,S.S. (1996) Biochemical analysis of mutant T7 primase/helicase proteins defective in DNA binding, nucleotide hydrolysis, and the coupling of hydrolysis with DNA unwinding. *J. Biol. Chem.*, 271, 26825–26834.

57. Satapathy,A.K., Kochaniak,A.B., Mukherjee,S., Crampton,D.J., van Oijen,A. and Richardson,C.C. (2010) Residues in the central beta-hairpin of the DNA helicase of bacteriophage T7 are important in DNA unwinding. *Proc. Natl Acad. Sci. USA*, 107, 6782–6787.

58. Moreau,M.J., McGeoch,A.T., Lowe,A.R., Izhaki,L.S. and Bell,S.D. (2007) ATPase site architecture and helicase mechanism of an archaeal MCM. *Mol. Cell*, 28, 304–314.

59. Kaur,P., Longley,M.J., Pan,H., Wang,W., Countryman,P., Wang,H. and Copeland,W.C. (2020) Single-molecule level structural dynamics of DNA unwinding by human mitochondrial twinkle helicase. *J. Biol. Chem.*, 295, 5564–5576.

60. Shlyakhtenko,L.S., Gall,A.A. and Lyubchenko,Y.L. (2013) Mica functionalization for imaging of DNA and protein–DNA complexes with atomic force microscopy. *Methods Mol. Biol.*, 931, 295–312.

61. Sun,J., Shi,Y., Georgec,R.E., Yuan,Z., Chait,B.T., Li,H. and O’Donnell,M.E. (2015) The architecture of a eukaryotic replisome. *Nat. Struct. Mol. Biol.*, 22, 976–982.

62. Harman,A. and Barth,C. (2018) The Dictyostelium discoideum homologue of twinkle, twm1, is a mitochondrial DNA helicase, an active primase and promotes mitochondrial DNA replication. * BMC Mol. Biol.*, 19, 12.

63. Diray-Arce,J., Liu,B., Cupp,J.D., Hunt,T. and Nielsen,B.L. (2013) The arabidopsis At1g30680 gene encodes a homologue to the phage T7 gp4 protein that has both DNA primase and DNA helicase activities. *BMC Plant Biol.*, 13, 36.

64. Yang,W., Seidman,M.M., Rupp,W.D. and Gao,Y. (2019) Replisome structure suggests mechanism for continuous fork progression and post-replication repair. *DNA Repair (Anast. J.)*, 81, 102658.

65. Sorgenfrei,S., Chiu,C.Y., Johnston,M., Nuckolls,C. and Shepard,K.L. (2011) Debye screening in single-molecule carbon nanotube field-effect sensors. *Nano Lett.*, 11, 3739–3743.

66. Lyubimov,A.Y., Strycharska,M. and Berger,J.M. (2011) The nuts and bolts of ring-translocase structure and mechanism. *Curr. Opin. Struct. Biol.*, 21, 240–248.

67. Martin,A., Baker,T.A. and Sauer,R.T. (2005) Rebuilt AAA+ motors reveal operating principles for ATP-fuelled machines. *Nature*, 437, 1115–1120.

68. Mesa,P., Alonso,J.C. and Ayora,S. (2006) *Bacillus subtilis* bacteriophage SPPI G40P helicase lacking the N-terminal domain unwinds DNA bidirectionally. *J. Mol. Biol.*, 357, 1077–1088.