The interaction between transport-segment DNA and topoisomerase IA—crystal structure of MtbTOP1 in complex with both G- and T-segments

Shomita Ferdous[1,2,†], Tumpa Dasgupta[1,2,†], Thirunavukkarasu Annamalai[1,2], Kemin Tan[3,*] and Yuk-Ching Tse-Dinh [iD][1,2,*]

[1]Department of Chemistry and Biochemistry, Florida International University, Miami, FL 33199, USA, [2]Biomolecular Sciences Institute, Florida International University, 11200 SW 8th St, Miami, FL 33199, USA and [3]Structural Biology Center, X-ray Science Division, Advanced Photon Source, Argonne National Laboratory, 9700 S. Cass Avenue, Lemont, IL 60439, USA

ABSTRACT

Each catalytic cycle of type IA topoisomerases has been proposed to comprise multistep reactions. The capture of the transport-segment DNA (T-segment) into the central cavity of the N-terminal toroidal structure is an important action, which is preceded by transient gate-segment (G-segment) cleavage and succeeded by G-segment religation for the relaxation of negatively supercoiled DNA and decatenation of DNA. The T-segment passage in and out of the central cavity requires significant domain–domain rearrangements, including the movement of D3 relative to D1 and D4 for the opening and closing of the gate towards the central cavity. Here we report a direct observation of the interaction of a duplex DNA in the central cavity of a type IA topoisomerase and its associated domain–domain conformational changes in a crystal structure of a *Mycobacterium tuberculosis* topoisomerase I complex that also has a bound G-segment. The duplex DNA within the central cavity illustrates the non-sequence-specific interplay between the T-segment DNA and the enzyme. The rich structural information revealed from the novel topoisomerase–DNA complex, in combination with targeted mutagenesis studies, provides new insights into the mechanism of the topoisomerase IA catalytic cycle.

GRAPHICAL ABSTRACT
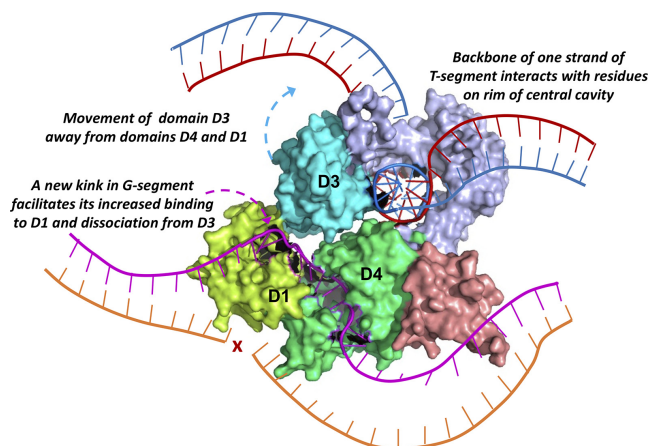
INTRODUCTION

Type IA topoisomerases are present in all kingdoms of life for resolving DNA topological barriers encountered in essential genomic transactions that require the passage of an intact DNA, either single-stranded (ssDNA) or double-stranded (dsDNA) through the break of an ssDNA (1–4). They, in conjunction with partner proteins such as helicases and RNA polymerase, play diverse roles in genome functions including replication, transcription and recombination (4–8). Type IA topoisomerases are also the only topoisomerase family with RNA topoisomerase activity that may have been present in the LUCA (last universal common ancestor) (9). Among type IA topoisomerases, bacterial topoisomerase I relaxes negatively supercoiled DNA efficiently to alleviate the global and local level of negative DNA

supercoiling (10–12). In organisms such as *Mycobacterium tuberculosis* with no topoisomerase III or IV present, topoisomerase I may also be responsible for the decatenation of DNA intermediates formed in replication and recombination (13).

The generally accepted catalytic mechanism of type IA topoisomerases was first proposed based on the structure of the *Escherichia coli* topoisomerase I N-terminal domains (D1–D4) that form a torus-like structure (14). In the first step of a catalytic cycle, type IA topoisomerases require binding of an ssDNA segment (gate segment or G-segment) primarily by D4, followed by a nucleophilic attack by the active site tyrosine (from D3) to form a covalent intermediate associated with a break of the phosphodiester linkage on the G-segment to create a DNA gate (15,16). Associated with the opening of the DNA gate, the movement of D3 away from D1 and D4 (opening of the toroid hole) is required to create an entry from the D3 side of the torus to its central cavity, allowing for the passage of the second DNA segment (T-segment) in and out of the central cavity (14,17) (Figure 1). Therefore, the sequential steps of the type IA topoisomerase catalytic cycle require a well-concerted opening/closing of the DNA gate and toroid hole. Although the details of the mechanism remain largely speculative, this gate opening model in general has been supported by computational analysis (18) and biochemical studies that include single-molecule experiments (19–24). For example, the gate openings of *E. coli* topoisomerase I and topoisomerase III were directly detected in single molecule assays based on transient extension of a topoisomerase-bound DNA strand (22). The ∼6.6 nm DNA transient change in extension was obtained by molecular simulations and was believed to result from the opening of the toroid hole. However, in all previously determined crystal structures of type IA topoisomerases, D3 associates closely with D1 and D4 to keep the toroidal structure in a closed state, and transport segment (T-segment) DNA has never been observed to be present in the interior of the central cavity. The interaction between T-segment DNA and the interior of the central cavity remains to be characterized.

We report here a crystal structure of *M. tuberculosis* topoisomerase I domains D1–D5 (MtbTOP1-704t) in complex with a bound G-segment and a duplex DNA that is captured inside the central cavity of the toroidal structure formed by D1–D4. The presence of the DNA duplex inside the cavity is associated with a new conformation for the G-segment and as well as interdomain movements of MtbTOP1-704t in comparison with other previously determined MtbTOP1-704t structures (25,26). The movement of D3 away from D1 and D4 leaves the 3′ region of the G-segment more closely attached to D1, forcing a new (second) kink in the G-segment (26,27) (Figure 1F). Moreover, the interaction pattern between the phosphate backbone of the captured T-segment DNA and the positively charged and polar residues on the cavity rim provides new insights into how these toroid hole residues may form a binding surface to facilitate the passage of the T-segment during the catalytic cycle of type IA topoisomerases.

MATERIALS AND METHODS

Protein expression and purification

MtbTOP1-704t (residues 2–704) was expressed with a previously described clone derived from pET-His6-Mocr TEV-LIC cloning vector (2O-T) with the Mocr (monomeric mutant of OCR from bacteriophage T7) tag (25) in *E. coli* T7 Express Crystal strain (New England BioLabs). Cells were grown in LB (Miller) broth at 37°C to the exponential phase ($OD_{600} = 0.4$) and left at room temperature for 20 min before recombinant protein expression was induced by adding 1 mM isopropyl-β-D-thiogalactopyranoside (IPTG). After overnight induction at 22°C, the cells were pelleted and then subjected to 3× freeze–thaw lysis in a buffer containing 20 mM sodium phosphate pH 7.4, 0.5 M NaCl, 20 mM imidazole and 1 mg/ml lysozyme. The soluble lysate with the His6-Mocr-tagged protein was collected after centrifugation at 35 000 rpm for 1 h and then incubated with Ni Sepharose 6 Fast Flow resin (GE Healthcare) for 1 h at 4°C. The affinity resin packed into a column was washed extensively with buffer of 20 mM sodium phosphate pH 7.4, 0.5 M NaCl and 20 mM imidazole before elution of the His6-Mocr-tagged protein with a buffer containing 20 mM sodium phosphate, pH 7.4, 0.5 M NaCl and 500 mM imidazole. The eluted His6-Mocr-tagged protein was incubated with tobacco etch virus (TEV) protease (1 mg of TEV protease for every 40 mg of fusion tagged protein) at 20°C for 6 h, followed by 4°C overnight in the presence of 50 mM Tris, pH 8, 0.5 mM EDTA and 1 mM dithiothreitol (DTT). The TEV-digested sample was then passed through the Ni Sepharose again to remove the fusion tags. The protein from the flow-through was further purified with S300 size exclusion chromatography (20 mM Tris–HCl, pH 8, and 300 mM KCl). The pure fractions were then concentrated for crystallization or dialyzed into the storage buffer (100 mM potassium phosphate pH 7.4, 1 mM EDTA and 50% glycerol).

Site-directed mutagenesis of MtbTOP1 was conducted with the Q5 high fidelity DNA polymerase with the 2O-T (Addgene plasmid # 29710) clone of MtbTOP1 (28) as template according to the protocol provided by New England BioLabs. The primers used for each mutant MtbTOP1 are listed in Supplementary Table S1. The entire MtbTOP1-coding region of mutagenized plasmid isolated from *E. coli* strain NEB 5-alpha was sequenced to confirm that only the desired mutation is present in the mutant clone. Expression and purification of wild-type and mutant MtbTOP1 from *E. coli* C41(DE3) strain (from Lucigen) as host were carried out with the same procedures used for MtbTOP1-704t described above.

Crystallization

MtbTOP1-704t was concentrated to ∼1.1 mM (∼85 mg/ml). The protein, MTS2-12 oligonucleotide (5′-CTTCCGCTTGAC-3′, custom synthesized by Sigma-Aldrich, ∼2 mM) and $MgCl_2$ (from 50 mM stock solution) were mixed at a molar ratio of 0.55:1.25. The mixture was then incubated on ice for ∼2 h. With a Mosquito nanoliter liquid handler (TTP LabTech), crystallization condition screening was set up using the sitting drop vapor diffusion
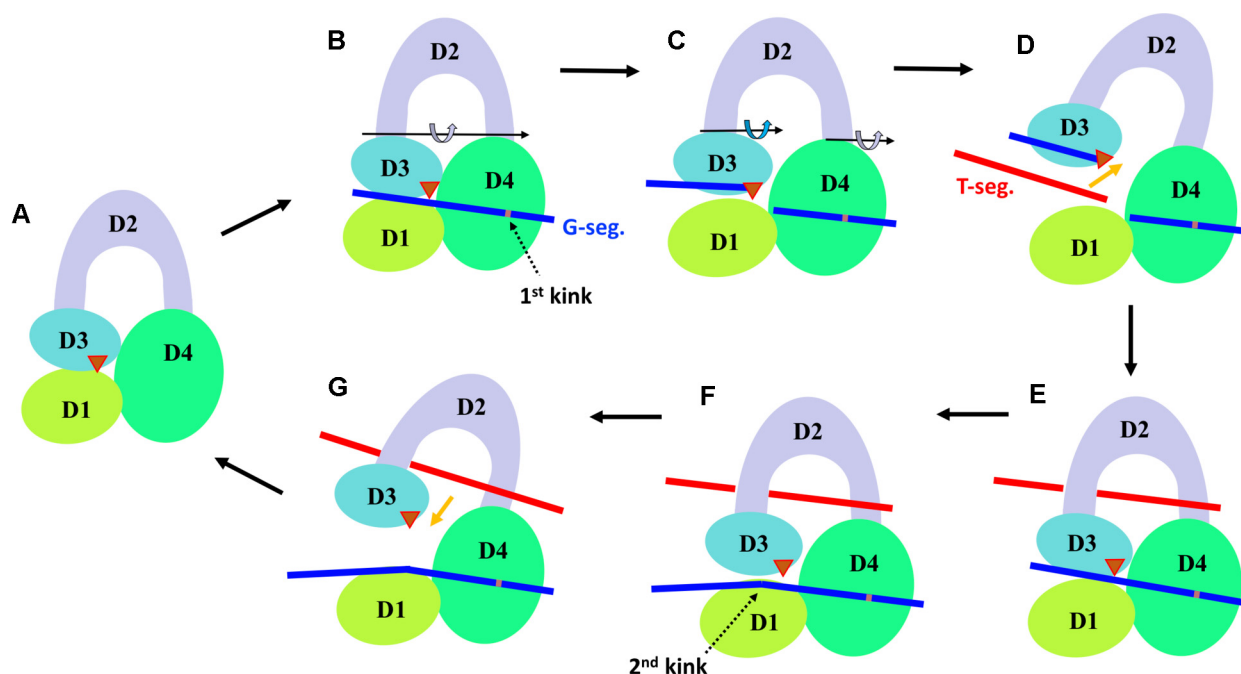
Figure 1. A schematic diagram of one catalytic cycle of type IA topoisomerases. Only the four N-terminal domains (D1–D4) that form a torus-like assembly are considered here for their required movements during catalysis. The red triangle represents the active site tyrosine residue in D3. Steps C and F are the specific focus in this study. The conformational change upon G-segment binding (25) in Step A is not illustrated. When D3 is closely associated with D1 and D4 as exemplified in Step B, the swing of D2 is restricted by two hinge joints, one between D2 and D3 and the other one between D2 and D4. When D3 disassociates from D1 and D4 as shown in Step C, these two hinge joints can then operate individually. The rotational axes of the hinge joints displayed in Step B and to a larger degree in Step C are largely diagrammatic. The T-segment could be either an ssDNA (for relaxation of supercoiled DNA) or a dsDNA (for DNA decatenation).

method with four 96-condition screens: Index, PEG/ION and SaltRx from Hampton Research; and NeXtal PEGs II from Molecular Dimensions, at 16°C. Each crystallization droplet containing 0.4 μl of protein–DNA mixture, and 0.4 μl of precipitating agent was equilibrated against a reservoir containing 140 μl of precipitating solution. About 10 days after the setting, typical needle-like crystals appeared in crystallization drops under several conditions. The structures of these crystals were determined to be the same as the MtbTOP1-704t–ssDNA complex as reported previously (PDB codes: 6CQI and 6CQ2) (26). About 8 months later, a single rectangular-shaped crystal appeared under a new condition comprising 0.1 M sodium acetate trihydrate, pH 7.0 and 12% (w/v) polyethylene glycol (PEG) 3350. This crystal was harvested and treated with a cryoprotectant solution (25% glycerol in its mother liquor) and then flash-frozen in liquid nitrogen before data collection.

X-ray diffraction and structure determination

One set of X-ray diffraction data was collected at 100 K from the cryocooled crystal at the 19-ID beamline of the Structural Biology Center at the Advanced Photon Source at Argonne National Laboratory (29). The HKL3000 program suite was used for data processing, including intensity integration, scaling and merging (Table 1) (30). The dataset was initially processed to a resolution limit of 2.92 Å. The anisotropy U and B factors of the dataset indicated no anisotropy (data not shown). The earlier reported

MtbTOP1-704t/MTS2-11 complex structure (PDB code: 6CQI) was used as a search template (Table 1) in the structural determination of the crystal with the molecular replacement method (31). By comparing the initial model and electron density maps, the D1, D4 and D5 domains of MtbTOP1-704t fit electron density well. The D3 domain that hosts catalytic residue Y342 does not fit into the electron density, indicating a conformational change of the domain compared with the MtbTOP1-704t/MTS2-11 template structure. Also, the electron density for the top of the arch-like D2 domain is mostly not well defined. The G-segment oligo that is mainly buried inside the DNA-binding groove in D4 is also well defined in electron density maps except for its 3′ region that is located in the interface between the D1 and D3 domains. After model rebuilding (32), particularly of the D3 domain and a part of the D2 domain and subsequent refinement, densities for DNA phosphate backbones that are present in the central cavity of the toroidal structure of the N-terminal domains became apparent. However, densities for bases were much less well defined for their positions in modeling when compared with the modeling of the DNA backbone. X-ray diffraction data were then reprocessed to a resolution limit cut-off at 2.78 Å to include more weak reflections (33). With more than 3300 unique reflections (an increase of ∼15.7%) being added to the dataset, the quality of overall electron density maps becomes significantly improved. It helped to recognize a homo base-paired parallel-stranded (ps) duplex with its 3′–3′ end portion running through the central cavity of the enzyme. The 5′–5′ end portion of the ps duplex partici-

Table 1. Data collection and refinement statistics

Data collection	MtbTOP1-704t/3MTS2-12
Space group	$P2_12_12$
Unit cell dimensions	
a, b, c (Å)	92.50, 98.63, 104.1
α, β, γ (°)	90, 90, 90
Protein molecular weight Da (residue)	77 468.3 (703)[a]
DNA molecular weight Da (residue)	10 954.02 (36)[b]
Mol or Complex (AU)	1
Wavelength (Å)	0.9792
Resolution (Å)	2.78–46.25
No. of unique reflections	24 482 (1195)[c]
Completeness (%)	99.9 (100)[c]
Redundancy	6.1 (5.8)[c]
R_{merge}	0.092 (1.372)[c]
R_{pim}	0.040 (0.614)[c]
$CC_{1/2}$	0.982 (0.490)[c]
$I/\sigma(I)$	23.5 (1.1)[c]
Wilson B-factors (Å^2)	82.8
Phasing[d]	
Resolution (Å)	3.00–46.25
Correlation coefficient	0.552
Refinement	
Resolution (Å)	2.78–46.25
No. of reflections (work/test)	13 197/1198
R_{work}/R_{free}	24.9/26.9
No. of atoms	
Protein/DNA	4930/698
Water/others	0/62
B-factors (Å^2)	
Protein/DNA	94.3/104
Water/others	NA/112
Root mean square deviation	
Bond length (Å)	0.003
Bond angle (°)	0.557
Ramachandran plot (%)[e]	
Preferred regions	95.02
Allowed regions	4.98
Outliers	0.00
All-atom clashscore[e]	6.37
Rotamer outliers (%)[e]	0.83
Cβ deviations[e]	0.00
PDB code	8CZQ

[a]Not including three N-terminal vector-derived residues, SNA; [b]including three MTS2-12 (5′-CTTCCGCTTGAC-3′) oligos; [c](last resolution bin, 2.78–2.83 Å); [d]molecular replacement; [e]MolProbity statistics (35).

pates in the formation of a rare tetraplex intercalated motif (i-motif) with its corresponding part of a symmetry-related MtbTOP1-704t/3MTS2-12 complex.

In the final model, two loops from the arch-like D2 domain are missing due to lack of electron density. One of them (from D226 to P236) connects two ceiling strands (β1 and β2) of the central cavity (25). The other one is from T474 to R488, including the long loop located in the front of the cavity (25). No magnesium ion is identified in the structure. Moreover, it is noticeable that a few positive peaks in difference electron density maps were observed in an interstitial space between the MtbTOP1-704t/oligo complexes in the crystal. It is likely that these peaks were from short fragments of ssDNA with disordered bases. In the final model, the majority of these peaks were interpreted as phosphate groups with a few associated electron densities left unaccounted for. The final model was refined using the pro-

gram Phenix.refine (Table 1) (34). Structural validation was performed using the program MolProbity (35).

E. coli AS17 complementation assay

The mutant 2O-T MtbTOP1 plasmids were transformed into the *E. coli* AS17 [F$^-$ *topA17*(*am*) pLL1(*Tetsup D43,74*)] strain which does not grow at 42°C without complementation because of a temperature-sensitive *topA* mutation (36,37). As positive and negative control, the wild-type 2O-T MtbTOP1 clone and 2O-T empty vector were also each transformed into AS17. Transformants were first isolated at 30°C. Single colonies of the AS17 transformants were grown at 30°C overnight in LB (Miller) broth with 100 μg/ml carbenicillin. The overnight cultures were diluted with LB so that OD_{600} equals 0.1. Ten-fold serial dilutions were made in LB, and 5 μl of each dilution was spotted on LB agar plates with 100 μg/ml carbenicillin. The plates in duplicates were incubated at 30°C and 42°C for 36 h and 18 h, respectively.

Relaxation activity assay

The 20 μl relaxation activity assay reaction contained 10 mM Tris–HCl, pH 8.0, 50 mM NaCl, 0.1 mg/ml gelatin, 2 mM $MgCl_2$ and 0.3 μg of negatively supercoiled pBAD/thio plasmid DNA (5.5 nM) purified by cesium chloride gradient centrifugation. Serial dilutions of the wild-type and mutant MtbTOP1 enzymes were added to each reaction and incubated at 37°C for 30 min. The reactions were stopped by adding 4 μl of stop solution (50 mM EDTA, 50% glycerol and 0.5% v/v bromophenol blue). The relaxation reaction products were analyzed by electrophoresis in a 1% agarose gel with TAE (40 mM Tris-acetate, pH 8.0, 2 mM EDTA) buffer at 25 V (1 V/cm) for 18 h. The gel was stained for 1 h with 1 μg/ml ethidium bromide solution, de-stained with deionized water for 15 min and photographed using UV light with the Alpha Imager Mini.

Differential scanning fluorimetry (DSF) assay

Melting curves for wild-type and mutant MtbTOP1 were obtained with the DSF assay. Thermal denaturation of the recombinant proteins was carried out in a 96-well microplate in the presence of SYPRO Orange dye using the Bio-Rad CFX96TM Real-Time polymerase chain reaction (PCR) instrument (38). Each 10 μl assay solution contains 3 μM enzyme in 25 mM KCl, 50 mM potassium phosphate, pH 7.6, 1 mM DTT, 10% glycerol, 5× SYPRO Orange dye (diluted from 5000× solution supplied by Thermo Fisher). Heating from 25°C to 95°C was programmed with a 30 s hold time, 0.5°C increment per cycle followed by reading of the fluorescence signal using the channel with an LED–filter photodiode combination designated for single-color fluorescence resonance energy transfer (FRET) with an λex of 490 nm and λem of 570 nm, close to the maximum λex and λem for SYPRO Orange fluorescence (480 and 568 nm) (39). The negative first derivative of the fluorescence signal in each melting curve was plotted versus temperature.
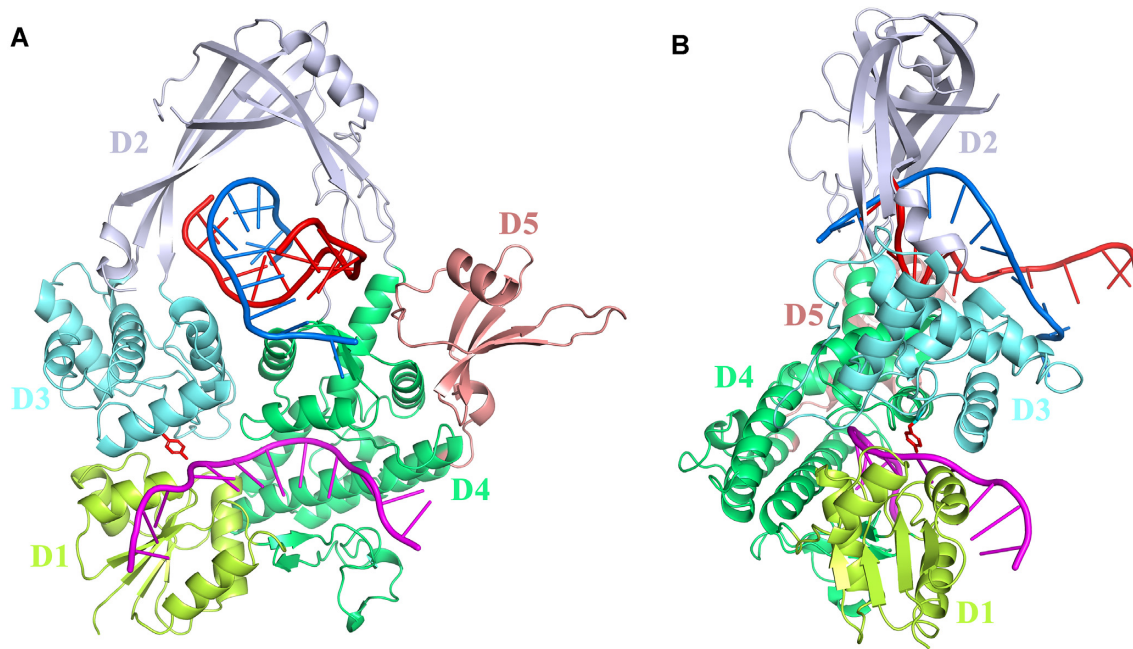
Figure 2. A ribbon diagram of the MtbTOP1-704t/3MTS2-12 structure. (**A** and **B**) A front view and a side view of the MtbTOP1-704t/3MTS2-12 complex, respectively. The N-terminal domains (D1–D4) and the first C-terminal domain (D5) of MtbTOP1-704t are drawn in lime, light blue, cyan, green and dark salmon pink, respectively. The key catalytic residue Y342 from D3 is drawn in stick format to highlight its position. The G-segment ssDNA is colored in magenta. The two strands of the duplex DNA that is captured inside the central cavity are colored in red (Strand A) and blue (Strand A′), respectively.

RESULTS

A new co-crystal of MtbTOP1-704T with a 12-mer oligonucleotide

During initial co-crystallization screening of MtbTOP1-704t with oligonucleotides of different sequences and lengths, one crystal of unique morphology appeared nearly 8 months after the others (26) from a different crystallization condition. The structure of the crystal was subsequently determined to a resolution limit of 2.78 Å by using the molecular replacement method. The structure reveals a complex of MtbTOP1-704t with three 12-mer oligonucleotides (MTS2-12, 5′-CTTCCGCTTGAC-3′). For the convenience of the following description and discussion, the complex will be referred as MtbTOP1-704t/3MTS2-12 hereafter.

Overall structure of MtbTOP1-704T/3MTS2-12

The construct MtbTOP1-704t (A2–T704) consists of four N-terminal domains (D1–D4) and the first C-terminal domain (D5) out of the full-length MtbTOP1 (934 residues). The four N-terminal domains form a toroidal structure with a central cavity, as observed in apo and G-segment-bound MtbTOP1-704t structures (Figure 2) (25,26), but with significant conformational changes as elaborated later. The association of the C-terminal D5 with the last N-terminal domain (D4) is mostly the same as described earlier (25). An intact MTS2-12 oligonucleotide (G-segment in magenta, Figure 2) binds to MtbTOP1-704t with the polarity (5′ to 3′) from the well-defined binding groove within the D4 domain to its extended binding site between D1 and D3. The G-segment, as detailed later, has an additional kink in its 3′

region in the new MtbTOP1-704t/3MTS2-12 complex (Figure 3) that has not been seen in previous crystal structures of type IA topoisomerases with a bound G-segment. The second kink makes the 3′ end of the G-segment more closely attached to D1. The G-segment overall is less twisted in comparison with its counterparts observed in other G-segment-bound MtbTOP1-704t structures (PDB codes: 6CQI and 6CQ2) (26).

Remarkably, there is a duplex DNA (A and A′ strands in red and blue, respectively, Figure 2) that is captured in the central cavity in the new structure. Based on the proposed catalytic mechanism of relaxation of negatively supercoiled DNA by type IA topoisomerases, the capture of T-segment DNA within the central cavity is an important step that should occur after G-segment transient cleavage and before G-segment religation (2,10) (Figure 1). DsDNA could also be captured as T-segment DNA when DNA catenation/decatenation is catalyzed by type IA topoisomerases. However, the passage of T-segment DNA in and out of the central cavity and the interaction of T-segment DNA with the central cavity have been largely uncharacterized for type IA topoisomerases. We propose that this MtbTOP1-704t/3MTS2-12 complex represents how a type IA topoisomerase interacts with a T-segment DNA within its central cavity.

Conformational change upon capture of T-segment DNA in the central cavity

The unique torus-like assembly formed by the N-terminal domains of type IA topoisomerases provides the structural flexibility for the enzyme to adopt different conformations in different catalytic steps. From the apo form of
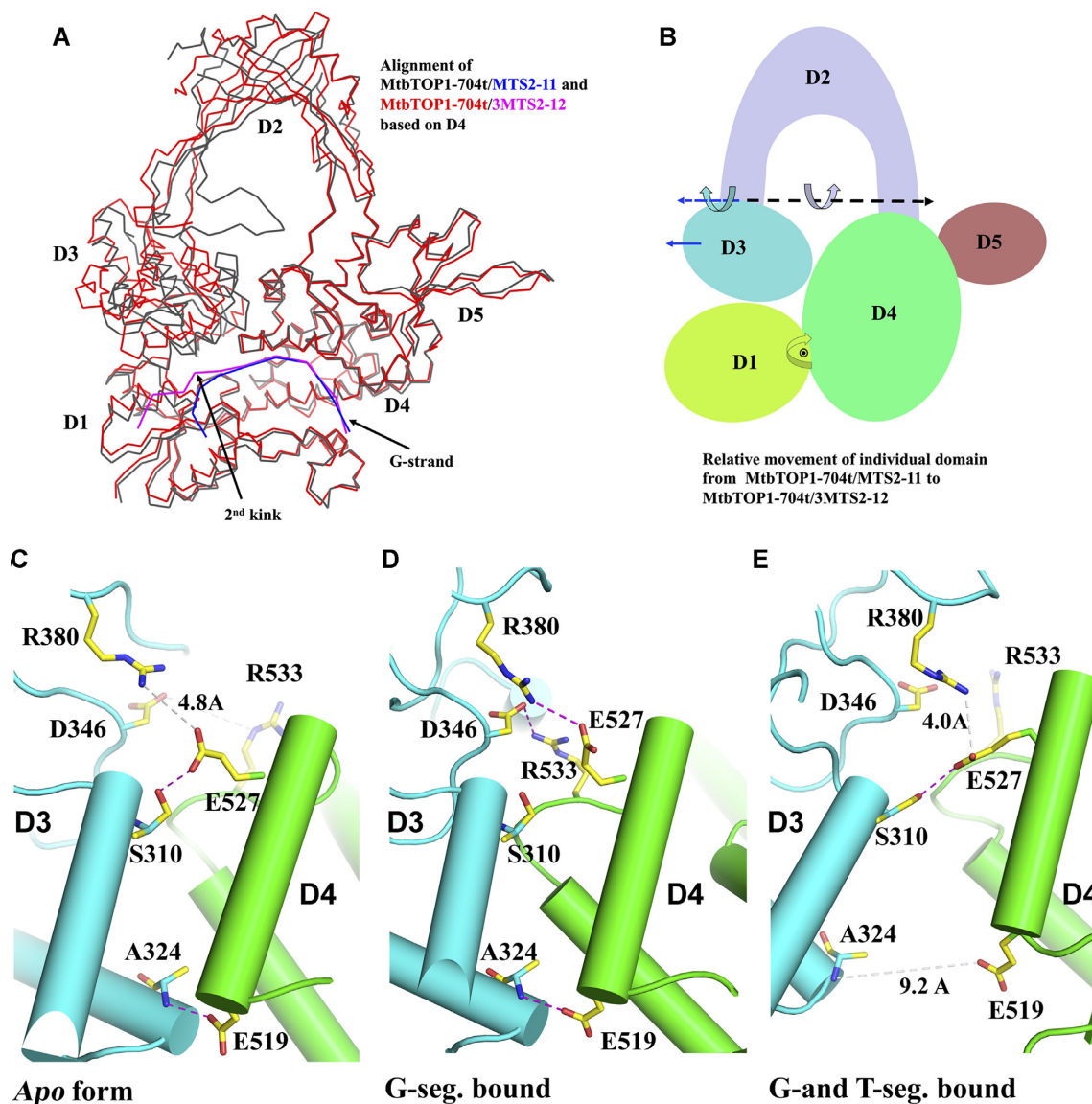
Figure 3. Conformational change when a duplex DNA is captured inside the central cavity. (**A**) Overlay of only the G-segment-bound MtbTOP1-704t/MTS2-11 structure (PDB: 6CQI) onto the MtbTOP1-704t/3MTS2-12 structure based on the alignment of their D4 domains. The duplex DNA in MtbTOP1-704t/3MTS2-12 is not shown for clarity. The color schemes for the two structures including G-segments are shown on the top right of the panel. (**B**) Relative movement of individual N-terminal domains from MtbTOP1-704t/MTS2-11 to MtbTOP1-704t/3MTS2-12, induced by the capture of duplex DNA inside the central cavity. Major translation and rotation are represented by a solid arrow and dashed arrow rotation axis, respectively. All translation directions and rotation axis orientations are approximate. (**C–E**) The specific interactions across D3 and D4 in the apo form (PDB 5D5H), G-segment-bound (PDB 6CQI) and G- and T-segments-bound MtbTOP1-704t (reported here) structures, respectively. The magenta dashed line represents either a hydrogen bond or a salt bridge (bond distances ≤3.5 Å). The gray dashed line with a label simply represents an atom–atom distance.

MtbTOP1-704t to its G-segment-bound form, we have observed previously the intradomain conformational changes for individual N-terminal domains as well as relative interdomain conformational changes (25,26). The conformational changes from the apo form to the G-segment-bound form bring the catalytic residues from different domains into one spatial proximity to form an active catalytic site for the initiation of DNA gate opening, as in Step B in Figure 1. The new MtbTOP1-704t/3MTS-12 complex could represent a snapshot of how the torus-like N-terminal domains can undergo additional conformational changes with a T-segment DNA captured in the central cavity after G-

segment religation and before T-segment release, as shown for Step F in Figure 1.

As a quick assessment of the conformational changes of N-terminal domains resulting from T-segment DNA binding within the central cavity, a simple structural alignment of MtbTOP1-704t (protein) only from the MtbTOP1-704t/MTS2-12 complex and MtbTOP1-704t/MTS2-11 (PDB entry: 6CDI) results in a root mean square deviation (RMSD) value of 2.21 Å (40), indicating that the presence of the additional DNA segment in the central cavity of MtbTOP1-704t is associated with a significant conformational change, especially an interdomain rearrangement

Table 2. Interdomain conformational change in MtbTOP1-704t/3MTS2-12 relative to single G-segment-bound MtbTOP1-704t/MTS2-11

Conformational change[a]	D1	D2	D3	D4	D5	D3 (when D2 is aligned)
Translation (Å)	1.37	3.78	4.45	0.06	1.19	5.21
Rotation (°)	6.4	12.2	22.6	0	4.6	22.5

[a]After D4 domains from MtbTOP1-704t/3MTS2-12 and MtbTOP1-704t/MTS2-11 are aligned, the translation and rotation of each individual domain in MtbTOP1-704t/3MTS2-12 relative to its corresponding domain in MtbTOP1-704t/MTS2-11 are calculated based on their transformation matrix obtained from their structure alignment by using the program Secondary-structure Matching (SSM) (40). The last column shows D3 relative movement when D2 domains from the two structures are aligned.

from the protein structure with only the G-segment bound. In order to appreciate the relative movement of each individual domain with respect to each other, only the D4 domain is used for an alignment of the two structures (Figure 3A) to show how other domains move when DNA is captured in the central cavity (Table 2).

D1 slightly tilts upward to have a closer contact with the 3′ end of the G-segment, as described below. D3, where the catalytic residue Y342 is located, contributes to the major conformational changes upon the presence of DNA in the central cavity (Figure 2; Supplementary Movie S1). Different residues in the D3 domain shift differently due to a combination of a translation and a rotation of the domain (Figure 3B). Some residues can shift >10 Å. For example, P412 of D3 moves ~12.8 Å. The catalytic residue Y342 (Figure 2) shifts ~3.3 Å, significantly moving away from other critical active site residues as well as the G-segment scissile phosphate at the cleavage site. As expected, the movement of D3 reduces the association between D3 and D4 (Figure 3C–E). The buried interface areas between these two domains decreased from ~1022 Å^2 to 590 Å^2. Also due to the partial separation of D3 and D4, two specific interactions, i.e. one hydrogen bond (between the main chain N atom of D3 A324 and the Oε1 atom of D4 E519) and one salt bridge (between D3 D346 and D4 R533), are lost. D4 E527 switches its interaction back to D3 S310 (a hydrogen bond) from D3 R380 (a salt bridge) (Figure 3).

The arch-like D2 also moves notably, as shown in Figure 3. The base of the arch expands ~3 Å from 27 Å to 30 Å, measured by the Cα–Cα distance between two internal residues of the toroidal hole, L468 and A515 that are located at the bottom of each side of the arch, respectively. The central loop, the loop that positions in front of the central cavity (25), is missing in the MtbTOP1-704t/3MTS2-12 model, indicating its disorder in the structure. The loop in the structure apparently moves away from the front of the central cavity to allow the passage of the T-segment DNA into the central cavity. The disordered structure of this loop may be related to its dynamic features, as discussed later.

The 3′ region of the G-segment becomes more tightly bound to domain D1

In only the G-segment-bound MtbTOP1-704t/MTS2-11 structure, the bases of the ssDNA from positions –4 to at

least +3 are stacked in a conformation similar to one B-DNA strand (26). A backbone kink between the –4 and –5 nucleotides is associated with the preference for a cytosine at the –4 position relative to the scissile phosphate at the cleavage site for all bacterial topoisomerase I (15–16,41). Similar to what was observed in the crystal structure of the *E. coli* topoisomerase I covalent complex (16), the cytosine base in the –4 position fits into a sterically restrictive cavity formed by conserved residues including Y176 of MtbTOP1 (corresponding to Y177 of *E. coli* topoisomerase I) which wedges between the –4 cytosine base and the next base at the –5 position to cause a kink in the backbone. However, the G-segment in this newly obtained MtbTOP1-704t/3MTS2-12 structure has a different conformation at the 3′ region compared with the previously determined structure with the G-segment only, forming a second kink between the +2 and +3 nucleotides. The base on the +3 nucleotide (G) rotates relatively ~180°, and the bases that follow on the +4 and +5 nucleotides resume the stacking like a B-DNA strand. The second kink makes the G-segment in the MtbTOP1-704t/3MTS2-12 structure appear less twisted (Figure 4). Interestingly, the newly observed kink is associated with a concerted increase in interaction between the 3′ region of the G-segment and the D1 domain, which slightly tilts upward towards the G-segment as described above. The increased interactions (Figure 4B) include a salt bridge between R30 and the phosphate group of G (+3) nucleotide, a bidentate hydrogen bond between E43 and A (+4) nucleotide, and a hydrogen bond between E95 and C (+5) nucleotide. The addition of these interactions not seen in previous G-segment-bound only MtbTop1 structure (Supplementary Figure S1) greatly increases the interface area between the G-segment and D1 domain, from ~349 Å^2 to 615 Å^2. Meanwhile the interface area between the G-segment and D3 domain is slightly reduced from 176 Å^2 to 166 Å^2, with a net loss of a salt bridge between R344 and the phosphate group of T (+1) nucleotide. The increased interactions between the 3′ region of the G-segment and D1 could provide the energy for the formation of the second kink on the G-segment, which will help D3 move away from the G-segment and D1 as well. The G-segment conformation shown in Figure 4 represents a possible catalytic intermediate after a T-segment DNA is captured inside the central cavity and the G-segment is religated. This shift in protein–DNA interaction may be required for the opening of the toroid hole following G-segment religation, as illustrated for Step F in Figure 1.

Parallel-stranded DNA homoduplex and its interaction with MtbTOP1-704T

Since the 12-mer MTS2-12 oligonucleotide (5′-CTTCCGCTTGAC-3′) is the only DNA molecule used in the co-crystallization with MtbTOP1-704t, the DNA duplex that is trapped in the central cavity has to be formed from two MTS2-12 oligonucleotides. The DNA duplex is an unusual ps duplex (Figure 5) (42–45). While this ps duplex does not resemble the regular B-DNA duplex in the DNA relaxation reaction, the strand of the ps duplex that interacts with the MtbTOP1 toroid hole residues as discussed below can be a model for the T-segment of a
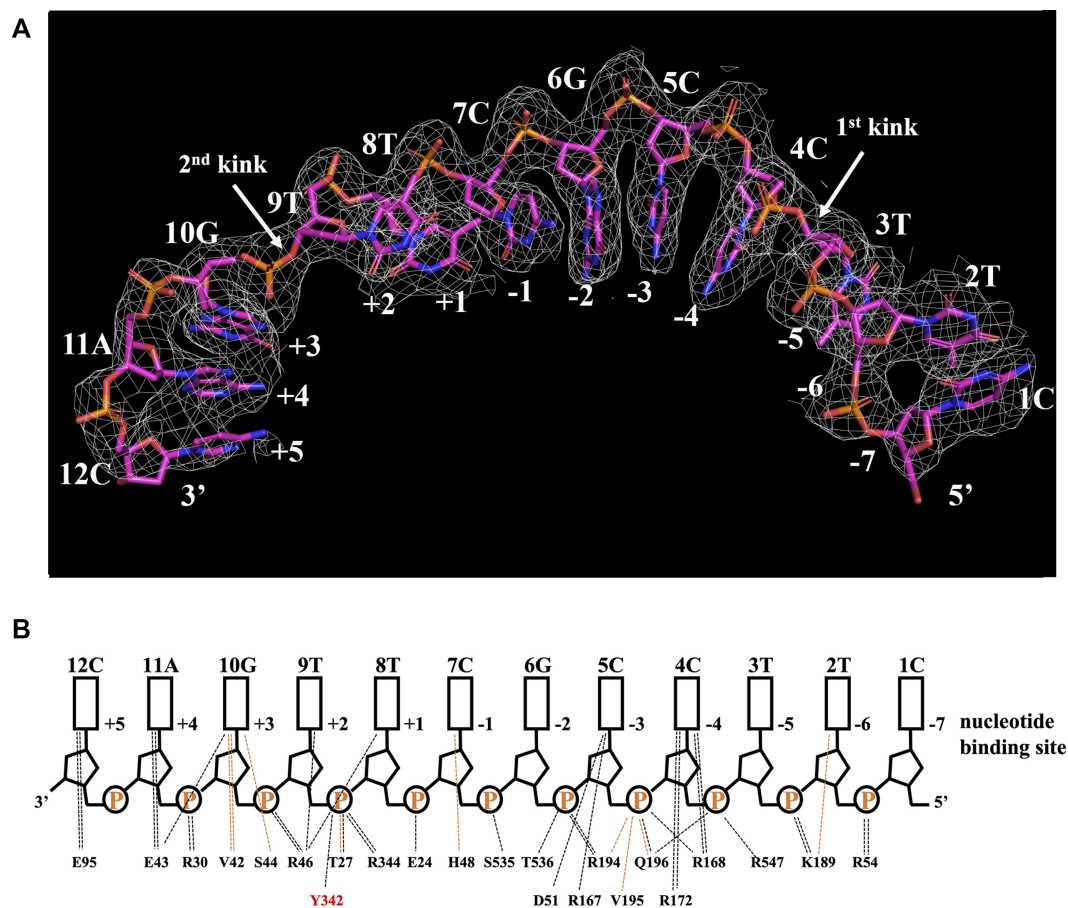
Figure 4. Conformation of the G-segment and its interaction pattern. (**A**) The conformation of the G-segment drawn in stick format. Compared with other G-segment-bound topoisomerase IA structures, the G-segment in MtbTOP1-704t/3MTS2-12 forms a second kink in its 3′ region, resulting in its tightened binding to D1. The gray mesh represents the σA-weighted $2F_o–F_c$ electron density map with a contour level of 1σ. (**B**) The interaction pattern of G-segment DNA with its surrounding protein residues. The key catalytic residue Y342 is highlighted to show the drift of its interaction from the scissile phosphodiester bond between the –1 and +1 nucleotides. Black and orange dashed lines represent the interactions contributed by side chain and main chain atoms of protein residues, respectively.

regular DNA duplex that enters into the toroid hole during the DNA relaxation reaction.

Based on electron density maps, one oligonucleotide (referred to as the A strand hereafter) has all 12 nucleotides visible. The other strand (referred to as the A′ strand hereafter) has a disordered 5′-end nucleotide (1C). The bases of 2T nucleotides from the two strands are not paired. Therefore, there are 10 non-canonical homo base pairs within the ps duplex. The ps homo duplex can be divided into two portions. From 6G to the 3′–3′ end 12C, the seven base pairs include a GA cross-strand stacking at steps 10G and 11A. The cross-strand stacking at GA dinucleotide steps is a signature of a ps duplex (45) and could be a stabilizing factor for the formation of a ps duplex. Another portion of the duplex that includes three 5′–5′ region nucleotides (3T4C5C) is not involved in the interaction of the duplex with the central cavity of MtbTOP1-704t. Instead, this portion is involved in the formation of an unusual, intercalated base pair stacking with the same part of a duplex from a neighboring symmetry-related MtbTOP1-704t/3MTS2-12 complex in the crystal (Supplementary Figure S2). The interaction of two duplexes forms a quadruple helix or rather a

tetraplex DNA with an i-motif (45–48), Figure 5. Although the i-motif observed in this structure is very interesting in terms of its unique sequence and as an example of the diversity of non-canonical DNA structures, the formation of the i-motif or the 2-fold dimeric MtbTOP1-704t/3MTS2-12 assembly is likely to be a molecular packing artifact of the crystal. There is currently no evidence that type IA topoisomerases are involved in the interaction with the i-motif. Therefore, we will not focus on the structure and the potential function of the i-motif observed here in this report, and only focus on the portion of the DNA that interacts with the MtbTOP1-704t cavity.

Both the ps DNA duplex and antiparallel-stranded (aps) B-DNA duplex are right-handed helices with similar diameters (~20 Å). One major difference between the ps duplex and aps duplex is that the ps duplex has two grooves of similar width and depth, while the aps duplex is well known to have a major groove and a minor groove (42,43). Despite the differences in the structures of the ps and aps DNA duplex, there are notable features in the way in which the ps duplex interacts with MtbTOP1-704t that may be used for modeling the interaction with B-DNA or ssDNA within the
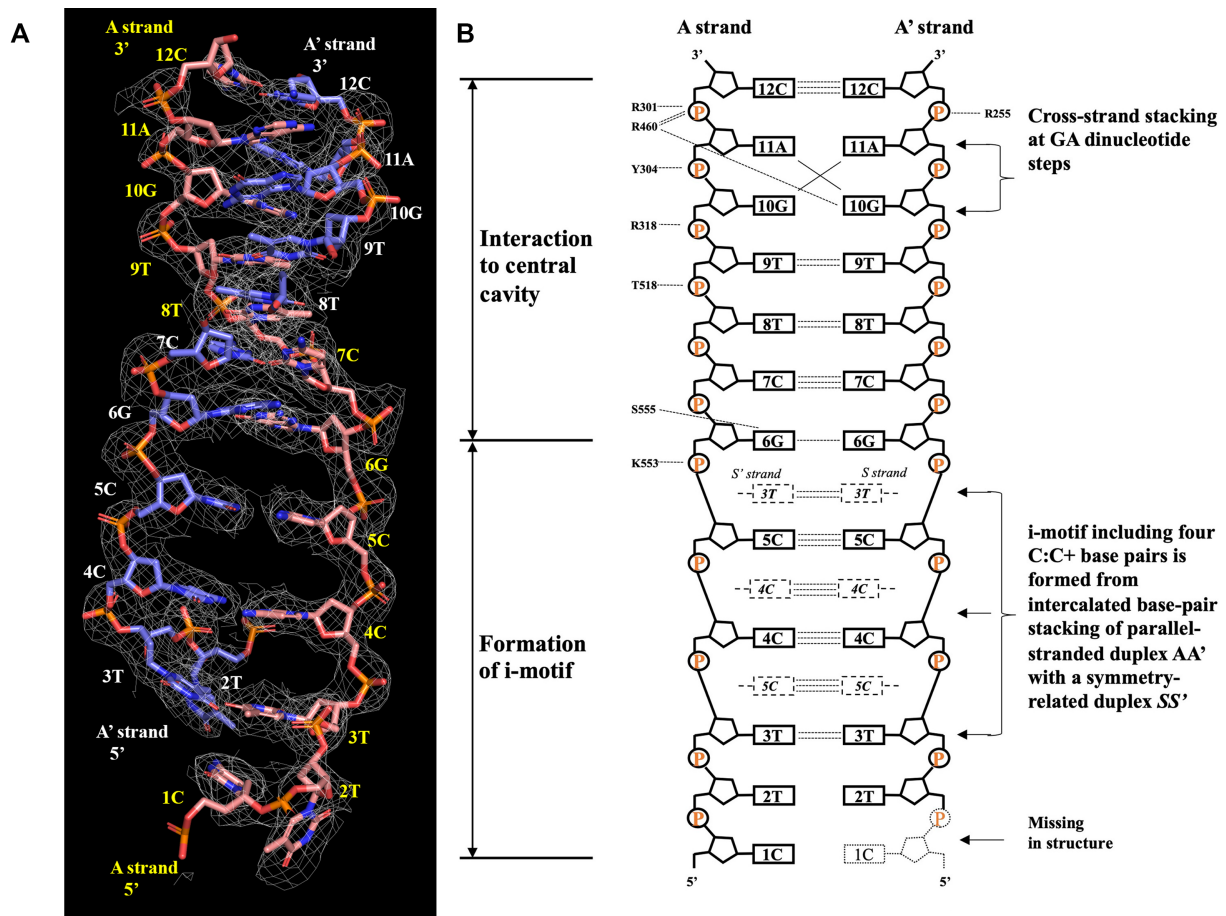
Figure 5. Formation of the ps DNA homoduplex and i-motif. (**A**) A stick representation of the ps DNA homoduplex formed from a pair of MTS2-12 oligonucleotides. The carbon atoms of bases and sugars of two strands, A and A′ are colored in dark orange and blue, respectively. The gray mesh surrounding the duplex represents the σA-weighted $2F_o$–F_c electron density map contoured at the 1σ level. (**B**) A schematic diagram of the interaction pattern (including base stacking) between the two ps homoduplex-forming oligonucleotides and with residues in MtbTOP1. The 5′–5′ portion of the duplex forms an i-motif with its counterpart from a symmetry-related complex. The 3′–3′ portion of the duplex is responsible for the interactions with the central cavity. Black dashed lines represent the interactions between protein atoms and ps DNA.

cavity. Firstly, all interactions of the ps duplex with MtbTOP1-704t are mostly contributed by the A strand (Figure 5). The other strand, the A′ strand, is only positioned by the A strand through their non-canonical base pairings. The position of the ps duplex is asymmetric within the central cavity, closer to the side where D3 is connected to D2, and the opening site of the toroid hole between D3 and D1 is nearby. Secondly, the A strand interacts with MtbTOP1-704t mostly through its outside sugar–phosphate backbone. These interactions are independent of the dsDNA sequence for either the ps duplex or the aps B-DNA duplex. Additionally, although a ps duplex cannot be wholly superimposed with a B-form DNA, the backbone of one strand of the ps duplex (A-strand, 5′-CTTGAC-3′, within the cavity) can be aligned with the backbone of one strand of a typical B-form DNA (PDB code:1D8G) with an RMSD value of 1.16 Å based on 54 atoms on the backbone (Supplementary Figure S3). Considering the similarities of the helix diameter and sugar–phosphate backbone twist of one strand of ps and aps duplexes, we propose that an aps B-DNA duplex that is trapped inside the central cavity will interact with the internal rim of the central cavity in a simi-

lar way. Therefore, the polarity of the ps duplex will be unlikely to play a role in such an interaction. In the same line of thought, both strands of the aps B-DNA duplex would have equal potential to interact with the central cavity when an aps B-DNA duplex is captured. The interaction with mostly the A strand only and the sequence-independent interaction of the ps duplex may indicate that a captured ssDNA could enter and interact with the central cavity in a manner similar to the A strand of the ps duplex.

The key T-segment binding residues

In order to characterize the interaction between the T-segment DNA and the central cavity, we have targeted selected residues within the rim of the cavity that form bonds with the captured DNA. They form a mixture of hydrogen bonds and salt bridges with the A strand backbone (Figure 5). The residues involved in the interactions with the A strand of the ps duplex are R301, Y304, R318, R460, T518, K553 and S555 (Figure 6A). The residue R460 also forms a hydrogen bond with the base of the 10G nucleotide of the A′ strand, which is the very nucleotide that is involved in
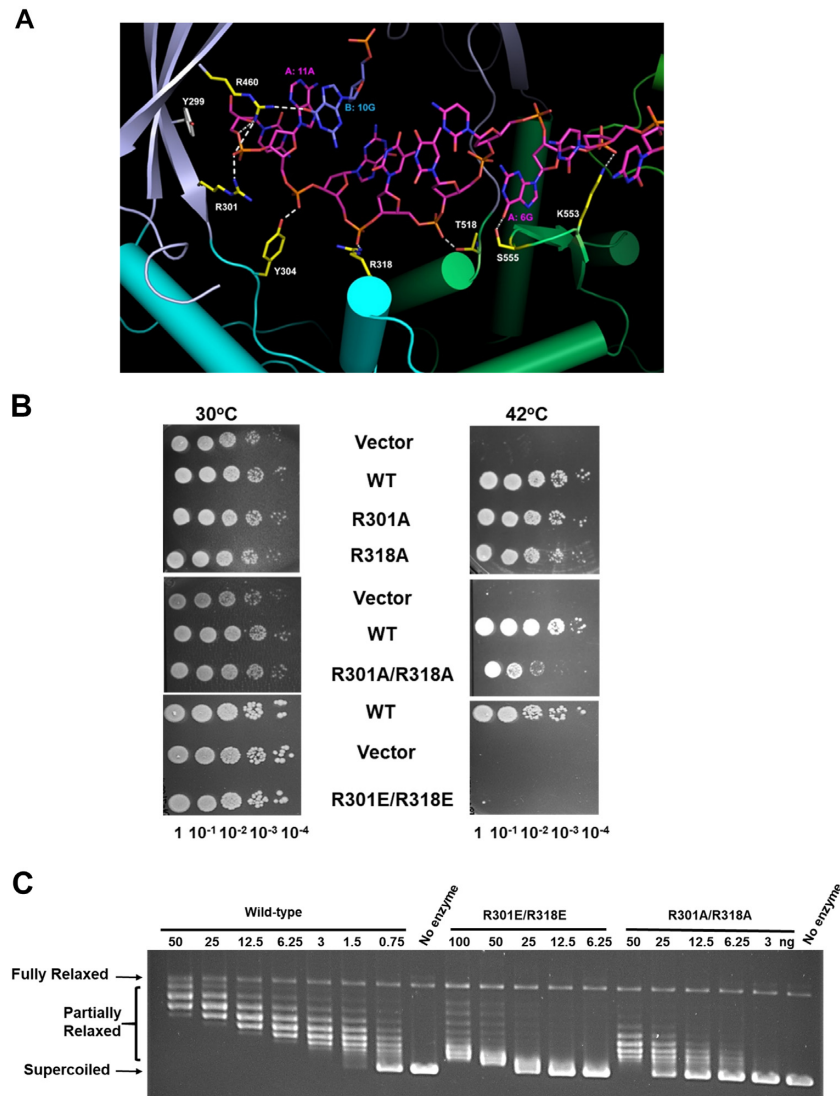
Figure 6. Role of the central cavity residues in interactions with the T-segment and catalysis. (**A**) Interactions between the T-segment and central cavity residues as observed in the crystal structure. Since the A strand contributes most of the interactions for the ps DNA, only the A strand is shown in the figure for clarity. One nucleotide from the A′ strand, 10G that participates in cross-strand stacking, is shown for its contribution of a hydrogen bond to R460 from the rim of the central cavity. (**B**) *In vivo* complementation of temperature-sensitive *topA* in *E. coli* AS17 by R301A, R318A, R301A/R318A or R301E/R318E mutant MtbTOP1. Ten-fold serial dilutions of the cultures of the AS17 transformants were spotted on LB plates with carbenicillin and incubated at 30°C or the non-permissive temperature of 42°C. Vector, 2O-T cloning vector; WT, wild-type MtbTOP1 clone. (**C**) Assay of the relaxation activity of MtbTOP1 with R301 and R318 mutations. Serial dilutions of wild-type or mutant MtbTOP1 with R301E/R318E or R301A/R318A substitutions were incubated with supercoiled plasmid DNA for 30 min at 37°C.

cross-strand stacking. These residues are located on only one side of the internal rim of the arch-like D2 and this side of the rim is close to the opening site of the toroid hole between D1 and D3. This interaction site could potentially facilitate the movement of the T-segment in and out of the central cavity. Alignment of bacterial topoisomerase I sequences shows that residues at these positions are not highly conserved (Supplementary Figure S4). However, we do often see residues with either positively charged or hydrogen bond-forming side chains at these positions. We postulate that a surface that can form salt bridges and hydrogen bonds with the backbone of the A strand of the T-segment is common in the internal rim of the bacterial topoisomerase I cavity close to the opening site of the toroid hole. The

T-segment most probably can slide within the cavity with multiple possible patterns of T-segment interactions.

The effect of single alanine substitutions at R301, Y304, R318, R460, T518, K553 and S555 was evaluated by transforming the mutant MtbTOP1 clones in the 2O-T vector into *E. coli* AS17. It has been shown in previous studies that leaky expression of MtbTOP1 from the T7 promoter can support the growth of AS17 at the non-permissive temperature of 42°C (41). The loss of interaction from these single alanine substitutions did not affect the *in vivo* complementation of the temperature-sensitive *topA* mutation in AS17 (Figure 6B; Supplementary Figure S5). However, the combination of the R301A and R318A mutations led to an ~100-fold decrease in the *in vivo* complementation of AS17

(Figure 6B) and an 8-fold decrease in relaxation activity (Figure 6C). Conversion of the positively charged Arg side chains at these positions to negatively charged Glu abolished the *in vivo* complementation (Figure 6B) and resulted in an ~67-fold reduction of relaxation activity (Figure 6C). The effect from the site-directed substitutions at these two basic residues provides support that the T-segment interactions observed in the crystal structure are relevant for the catalytic activity of MtbTOP1.

Domain–domain interface residues for toroid hole opening/closing

The movement of D3 away from D1 and D4 effectively removes the restriction of the movement of D3 relative to D2 and unlocks significant restriction on the hinge joint between D2 and D4. The movement of D2 relative to D4 is probably the primary factor in opening the toroid hole between D3 and D2 to allow the passage of the T-segment (Figures 1 and 3). Therefore, we propose that the initial disassociation of D3 away from D1 and D4 is the first step of gate opening. When D3 movement is unrestricted by D1 and D4, the gate opening could become wider due to the movements of D3 relative to D2 and of D2 relative to D4, respectively, as shown in Step C in Figure 1. This sequence of events is to be reversed during gate closing for G-segment religation. After G-segment religation, D3 will again dissociate from D1 and D4 for the release of the T-segment in Step G shown in Figure 1. A previous molecular simulation on type IA topoisomerase gate dynamics has suggested that D3 first breaks contacts with D4 before breaking contacts with D1 (22). Here we targeted some of the residues within the interface between D3 and D4 (Figure 3C–E) for their potential roles in the opening/closing of the toroid hole.

E519 is the first residue of the α6 helix of D4 and is located at the boundary between D2 and D4. This glutamate residue is strictly conserved in a region of high similarity for bacterial topoisomerase I (Supplementary Figure S4). In the apo form and the G-segment-bound form structures, the E519 side chain forms a hydrogen bond with the main chain amide group A324, which is the first residue of the α2 helix of D3 (Figure 3C). The negatively charged side chain of E519 apparently also forms an electrostatic interaction with the positively charged N-terminus of the α2 helix (49,50) (Figure 3C). These interactions are conserved in apo and G-segment-bound MtbTOP1-704t structures (PDB 5D5H and 6CQI). While the interaction of E519 with the D3 domain disappears when D3 moves away from D4 as observed in the new open structure of MtbTOP1-704t/3MTS2-12, E519 interactions with D3 may be critical for the enzyme to return to a closed conformation during the catalytic cycle. Site-directed mutagenesis of E519 was carried out to investigate its function. We were able to obtain a clone of MtbTOP1-E519A under the control of the T7 promoter in the 2O-T plasmid in *E. coli* strain NEB5alpha. However, no viable transformants could be obtained in strains that have the T7 RNA polymerase, including T7 Express Crystal (used for expression of wild-type MtbTOP1), as well as C41(DE3) from Lucigen for toxic protein expression (51). A single transformant was obtained in BL21AI

designed to minimize toxic protein expression from the T7 promoter (52) and only seven transformants were obtained for C41(DE3)pLysS, a frequency that is >1000-fold lower than the wild-type MtbTOP1 clone. The E519A plasmids in C41(DE3)pLysS have deletions that removed the T7 promoter and part of the mutant MtbTOP1-E519A coding sequence, even though the agar plates used for transformation contained 2% glucose to suppress recombinant protein expression. Induction with arabinose and IPTG did not produce any MtbTOP1-E519 protein in the BL21AI transformant, suggesting the loss of the inducible T7 RNA polymerase in the transformant. We propose that the carboxyl side chain of E519 is critical for D3 and D4 to reassociate following gate opening. Ala substitution at this position may result in a highly toxic mutant topoisomerase that would accumulate the covalent complex in an open conformation, but we do not have definitive biochemical evidence to prove that the E519A mutant is hyper-cleavage.

E527 is another strictly conserved glutamate residue (Supplementary Figure S4) in D4 of MtbTOP1 that appears to contribute to the association of D4 with D3. The relaxation activity of the E527A mutant MtbTOP1 shows that the loss of this glutamate side chain results in a 2- to 4-fold reduction of relaxation activity at 37°C (Figure 7A). The negative first derivative curves from the thermal shift assay using differential scanning fluorometry (Figure 7B) showed that wild-type MtbTOP1 has two melt peaks corresponding to protein melting temperatures of 44°C and 54°C. We propose that dissociation of the MtbTOP1 domain D3 away from domains D4 and D1 occurs at the first melting temperature, corresponding to the loss of its quaternary structure, which is the toroidal assembly in the type IA topoisomerase. The second step of protein denaturation of such a multiple domain topoisomerase will be the unfolding of its tertiary structures or domains. Since the MtbTOP1 polypeptide chain travels back and forth within the four N-terminal domains D1–D4, the denaturation of these domains is expected to entail cooperative unfolding, which is likely to be represented by the second melt peak. While we cannot exclude other possibilities for the assignments of these two transitions, the E527A mutation results in two decreased protein melting temperatures of 40°C and 47°C, respectively, consistent with the contribution of this Glu residue in both the association of domain D4 with D3, and the overall polypeptide stability of MtbTOP1. Additional evidence in future experiments is needed to confirm the interpretation of the two DSF peaks as proposed.

R380 is unique to MtbTOP1, and not a conserved residue among topoisomerase I enzymes from bacteria (Supplementary Figure S4). MtbTOP1 with the R380A mutation has nearly the same relaxation activity (Figure 7A) as wild-type MtbTOP1. The first protein melting temperature decreased from 44°C to 42°C, while the second protein melting temperature remained the same when compared with wild-type MtbTOP1 (Figure 7B). This suggests that the non-conserved R380 has a more modest contribution to MtbTOP1 protein folding and stability when compared with E527. In the G-segment-bound structure of MtbTOP1-704t/MTS2-12 (Figure 3D, PDB 6CQI), the side chain of E527 seems to be attracted away from S310 by R380 to form a salt bridge with the positively charged
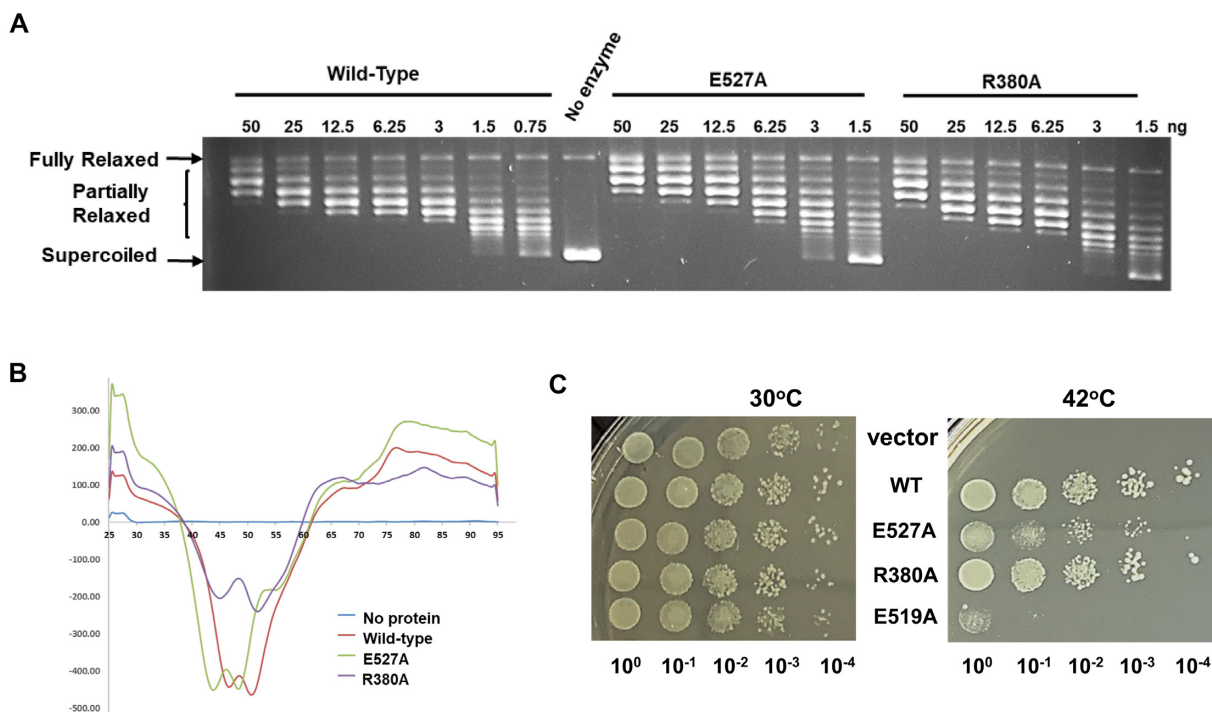
Figure 7. Effect of Ala substitutions targeting D3–D4 interdomain interactions. (**A**) Relaxation activity of MtbTOP1 with E527A or R380A substitution. Serial dilutions of wild-type or mutant MtbTOP1 enzyme were incubated with supercoiled plasmid DNA for 30 min at 37°C. (**B**) Comparison of the melting profiles of wild-type, E527A and R380A MtbTOP1 by differential scanning fluorometry. (**C**) Complementation of the temperature-sensitive *topA* mutation in *E. coli* AS17 with MtbTOP1 with mutations at the D3–D4 interface. Ten-fold serial dilutions of the cultures of the AS17 transformants were spotted on LB plates with carbenicillin and incubated at 30°C or the non-permissive temperature of 42°C. Vector, 2O-T cloning vector; WT, wild-type MtbTOP1 transformant.

residue. When R380 is absent, S310 can probably restore the interaction between E527 and D3.

These site-directed mutations targeting D3–D4 interactions in MtbTOP1 were further tested for their effect on *in vivo* complementation of the temperature-sensitive *topA* in *E. coli* strain AS17. The results of the complementation assay (Figure 7C; Supplementary Figure S5) showed that the R380A and S310A mutants are similar to wild-type MtbTOP1 in growth complementation. The E527A mutant transformant showed slower growth than the wild-type transformants at 42°C, and the number of viable colonies appeared to decrease by ∼10-fold. In contrast, the viability of AS17 complemented by MtbTOP1-E519A at 42°C was 10^3-fold less than that of AS17 complemented by wild-type MtbTOP1. This is consistent with wild-type-like relaxation activity for the R380A mutant, reduced relaxation activity and/or protein stability for the E527A mutant and lack of *in vivo* relaxation activity from the E519A mutant MtbTOP1.

DISCUSSION

We have determined a crystal structure of MtbTOP1-704t in complex with a bound G-segment and a ps duplex that is captured inside the central cavity of the torus-like N-terminal domains. The observation of the interaction of the captured DNA with the central cavity and its associated protein conformational changes provide new insights into the mechanism of capturing T-segment DNA by type IA

topoisomerases during its catalytic cycle. The interaction pattern that we can see in this structure is expected to be a snapshot of the dynamic interaction between the T-segment and the cavity rim. Nevertheless, this pattern of employing multiple hydrogen bonds and salt bridges to bind mainly the phosphate backbone of a single strand of the T-segment can be utilized by diverse type IA topoisomerases for relaxation of supercoiled DNA as well as catenation/decatenation of DNA or RNA. This mechanism of T-segment interaction does not require strict conservation of the residues at the cavity rim and would be largely independent of the nucleotide sequence of the T-segment.

Capturing T-segment DNA strand(s) into the central cavity involves transient cleavage of the G-segment and also the opening of the toroid hole between D1 and D3. The opening and closing of the toroid hole have been long proposed to be critical features in the catalytic cycle of type IA topoisomerases. The observation of the opening of the toroid hole has remained elusive until a study on bacterial type IA topoisomerases that applied a single molecule assay to directly measure the opening of the DNA gate by employing magnetic tweezers (22). The reported 6.6 nM opening of the DNA gate should be sufficient for capturing T-segment DNA strand(s) into the central cavity of the N-terminal toroidal structure. Recently, similar gate opening (∼8.5 nM) as well as ssDNA or dsDNA trapping inside the central cavity have also been measured for the human TopoIIIa–RMI1–RMI2 (TRR) complex using optical tweezers and fluorescence microscopy (24). However, the
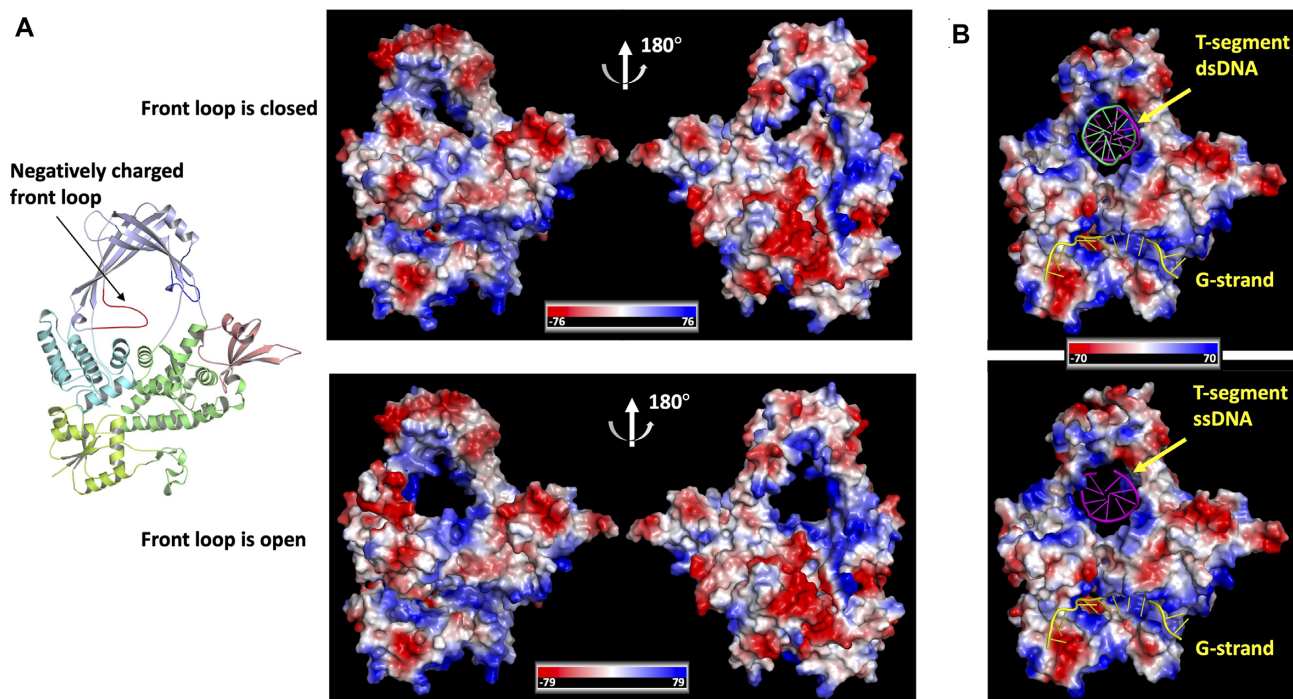
Figure 8. The potential regulatory role of the loop in front of the central cavity and T-segment interaction models. (**A**) Left: a ribbon diagram of MtbTOP1-704t in its closed conformation (PDB code: 5D5H). The negatively charged front loop is highlighted in red. Right: electrostatic potential surface representations of MtbTOP1-704t when the front loop is closed before and after T-segment binding (top panel) and when the front loop is open during T-segment binding (bottom panel), respectively. The surface of the molecule is colored according to amino acid residue charge, ranging from blue (positive) to red (negative). (**B**) Models of a B-form aps duplex DNA (top panel) and ssDNA (bottom panel) captured inside the central cavity. The aps duplex model was built based on a DNA backbone alignment (Supplementary Figure S3) of one strand of a B-form aps duplex DNA onto the 3′ region of the A strand of the ps DNA duplex in the structure of MtbTOP1-704t/3MTS2-12. The model of ssDNA inside the cavity is derived from the duplex DNA model (top panel) by keeping one strand that corresponds to the A strand in the structure of MtbTOP1-704t/3MTS2-12.

detailed molecular mechanism of toroid hole opening and T-segment DNA capture remains unclear.

According to the proposed type IA topoisomerase mechanism, the arch-like D2 should be mobile in comparison with the other N-terminal domains. However, from all known MtbTOP1 structures, D2 itself, even its base, has limited intradomain conformational change. In the new structure described here, the base of D2 expands only ∼3 Å (∼11% increase) upon dsDNA binding within the central cavity. Therefore, the opening of the central cavity for the access of the T-segment may be mainly determined by interdomain rearrangement instead of conformational change of individual domains. There are two hinge joints between D2 and D3 and between D2 and D4, respectively, which are expected to play primary roles in the opening of the central cavity (Figure 1). Since D3 is tethered to D1 and D4 in a closed conformation, the motion of D2 is hereby largely limited to the back-and-forth rotation around an axis that goes between the two hinge joints. After cleavage of the G-segment, however, it becomes possible for D3 with a covalently bound 5′ G-segment fragment to move away from D4 and D1. When D3 detaches from D4 and D1, it adds extra dimensions of rotation to the hinge joint between D2 and D4, and makes it possible for the central cavity to open widely on the D3 side (Figure 1D). The detachment of D3 from D4 and D1 also frees the restriction of the swing of D3 with respective to D2, which was observed in an earlier structural study of D2 and D3 domains of EcTOP1 (17). In this line of thinking, we propose that the detachment of D3 from D4 and D1 is a triggering factor for the opening of the central cavity. The two steps in the catalytic cycle that require the opening of the central cavity are represented as Steps C and F in Figure 1. Besides the steps shown in the figure, there are likely to be other distinctive transient steps, which could be potentially associated with multiple conformations observed in single molecule experiments (22–23,53).

In order to capture T-segment DNA inside the central cavity, besides the opening of the DNA gate and toroid hole, the negatively charged loop in front of the cavity should also move aside (Figure 8A). In the closed conformation of MtbTOP1, the loop partially covers the central cavity. Additionally, the negative charges it carries can neutralize some of the positive charges on the rim of the central cavity. When the loop moves aside, the electrostatic potential of the central cavity rim becomes wholly positive (Figure 8A). This creates a favorable environment to attract negatively charged DNA molecules into the cavity. We propose that the front loop may play a regulatory role during the catalytic cycle of type IA topoisomerases. Nevertheless, due to the loop's high mobility, we are not able to locate its position in the MtbTOP1-704t/3MTS2-12 structure. Its exact roles in the catalytic mechanism need to be further investigated.

In the structure presented here, a ps duplex uses mainly the sugar–phosphate backbone of one of its two strands to interact non-specifically with the internal rim of the central cavity. We propose that a B-form aps duplex would interact with the central cavity in a similar manner to that shown in the top panel of Figure 8B and described in the previously proposed model of the dsDNA decatenation mechanism for type IA topoisomerases (14). Both catenation and decatenation involve the passage of a dsDNA through a transient break on the G-segment to enter the central cavity. After the religation of the G-segment, the dsDNA will have to be released from the central cavity.

Type IA topoisomerases are found in all kingdoms of life to provide the ability to disentangle intermediates formed in replication and recombination through their decatenation activity (1,4–5,54–55). Topoisomerase III in archaea has been shown to function as a decatenase (56). Topoisomerase III is also the more efficient decatenase in organisms such as *E. coli* that have both topoisomerase I and topoisomerase III (57,58) so that topoisomerase III can supplement the role of topoisomerase IV in unlinking daughter chromosomes during DNA replication (59,60). However, *M. tuberculosis* is an example of a bacterium that does not have topoisomerase III or topoisomerase IV. A role for topoisomerase I in chromosome segregation has been proposed and supported by the mapping of *Mycobacterium smegmatis* topoisomerase I binding and cleavage sites in the Ter region (13). The catenation/decatenation activity of MtbTOP1 may be important for the enzyme's physiological function.

The capture of a ssDNA (T-segment) by the central cavity is also a critical step for the relaxation of negatively supercoiled DNA and catenation/decatenation of ssDNA. As described in the Results, the ps duplex interacts with the central cavity asymmetrically. Not only is the duplex positioned closer to the D2–D3 hinge joint than to the D2–D4 hinge joint, but also only one strand (A strand) is responsible for most of these interactions between the duplex and the central cavity. We propose that the DNA–protein interactions observed for A strand binding to the central cavity may illustrate how a single-stranded T-segment DNA runs through the central cavity after it is captured (Figure 8B, bottom panel).

A crystal is formed from the orderly packing of molecules. The likelihood of seeing MtbTOP1 with its central cavity wide open in a crystal structure is slim. It is conceivable when the T-segment is captured inside the central cavity that the interactions between the positively charged rim of the central cavity and negatively charged backbone of T-segment DNA could assist in the closing of the central cavity until the G-segment is religated. After the ligation, D3 will free the 3′ end of the G-segment to D1 and then swing away from D1 and D4 to initiate another opening of the central cavity for release of trapped T-segment DNA (Figure 1, Step G). However, the details of the T-segment releasing mechanism remain unknown.

We propose that the structure of MtbTOP1-704t/3MTS2-12 reported here represents a topoisomerase IA–T-segment complex that corresponds to a snapshot of one moment after G-segment religation (Figure 1, Step F). The observed structure demonstrates that the central cavity can accommodate a DNA molecule as postulated for the catalytic mechanism. Based on the strong negative selection in *E. coli* to avoid the expression of MtbTOP1 protein with the E519A mutation, we hypothesize that the electrostatic and hydrogen bond interactions involving the negatively charged side chain of highly conserved E519 in D4 and the N-terminus of the α2 helix in D3 are critical for the closing of the central cavity. The interaction between E527 in D3 and residues in D4 can affect the opening of the central cavity and the folding stability of the protein, but appears to be less critical than E519 for the enzyme to complete the catalytic cycle

Additional aspects of the proposed topoisomerase IA catalytic mechanism remain to be fully elucidated. After D3 separates from D4 and D1, conformational changes arising out of relative domain–domain rotations around D2–D3 and D2–D4 hinge joints are needed to make the central toroid cavity fully accessible for capturing the T-segment DNA. Further investigation is required on the triggering mechanism of these domain–domain rearrangements as well as the conformational change of the negatively charged loop in front of the central cavity.

DATA AVAILABILITY

Atomic coordinates and structure factors for the reported crystal structure have been deposited in the Protein Data Bank under accession number 8CZQ.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

REFERENCES

1. Wang,J.C. (2002) Cellular roles of DNA topoisomerases: a molecular perspective. *Nat. Rev. Mol. Cell Biol.*, **3**, 430–440.
2. Dasgupta,T., Ferdous,S. and Tse-Dinh,Y.C. (2020) Mechanism of type IA topoisomerases. *Molecules*, **25**, 4769.
3. Garnier,F., Debat,H. and Nadal,M. (2018) Type IA DNA topoisomerases: a universal core and multiple activities. *Methods Mol. Biol.*, **1703**, 1–20.

4. Bizard,A.H. and Hickson,I.D. (2020) The many lives of type IA topoisomerases. *J. Biol. Chem.*, **295**, 7138–7153.

5. Bocquet,N., Bizard,A.H., Abdulrahman,W., Larsen,N.B., Faty,M., Cavadini,S., Bunker,R.D., Kowalczykowski,S.C., Cejka,P., Hickson,I.D. *et al.* (2014) Structural and mechanistic insight into Holliday-junction dissolution by topoisomerase IIIalpha and RMI1. *Nat. Struct. Mol. Biol.*, **21**, 261–268.

6. Goto-Ito,S., Yamagata,A., Takahashi,T.S., Sato,Y. and Fukai,S. (2017) Structural basis of the interaction between topoisomerase IIIbeta and the TDRD3 auxiliary factor. *Sci. Rep.*, **7**, 42123.

7. Cheng,B., Zhu,C.X., Ji,C., Ahumada,A. and Tse-Dinh,Y.C. (2003) Direct interaction between *Escherichia coli* RNA polymerase and the zinc ribbon domains of DNA topoisomerase I. *J. Biol. Chem.*, **278**, 30705–30710.

8. Banda,S., Cao,N. and Tse-Dinh,Y.C. (2017) Distinct mechanism evolved for mycobacterial RNA polymerase and topoisomerase I protein–protein interaction. *J. Mol. Biol.*, **429**, 2931–2942.

9. Forterre,P. and Gadelle,D. (2009) Phylogenomics of DNA topoisomerases: their origin and putative roles in the emergence of modern organisms. *Nucleic Acids Res.*, **37**, 679–692.

10. Terekhova,K., Gunn,K.H., Marko,J.F. and Mondragon,A. (2012) Bacterial topoisomerase I and topoisomerase III relax supercoiled DNA via distinct pathways. *Nucleic Acids Res.*, **40**, 10432–10440.

11. Brochu,J., Breton,E.V. and Drolet,M. (2020) Supercoiling, R-loops, replication and the functions of bacterial type 1A topoisomerases. *Genes (Basel)*, **11**, 249.

12. Vos,S.M., Tretter,E.M., Schmidt,B.H. and Berger,J.M. (2011) All tangled up: how cells direct, manage and exploit topoisomerase function. *Nat. Rev. Mol. Cell Biol.*, **12**, 827–841.

13. Rani,P. and Nagaraja,V. (2019) Genome-wide mapping of topoisomerase I activity sites reveal its role in chromosome segregation. *Nucleic Acids Res.*, **47**, 1416–1427.

14. Lima,C.D., Wang,J.C. and Mondragon,A. (1994) Three-dimensional structure of the 67K N-terminal fragment of *E. coli* DNA topoisomerase I. *Nature*, **367**, 138–146.

15. Tse,Y.C., Kirkegaard,K. and Wang,J.C. (1980) Covalent bonds between protein and DNA. Formation of phosphotyrosine linkage between certain DNA topoisomerases and DNA. *J. Biol. Chem.*, **255**, 5560–5565.

16. Zhang,Z., Cheng,B. and Tse-Dinh,Y.C. (2011) Crystal structure of a covalent intermediate in DNA cleavage and rejoining by *Escherichia coli* DNA topoisomerase I. *Proc. Natl Acad. Sci. USA*, **108**, 6939–6944.

17. Feinberg,H., Lima,C.D. and Mondragón,A. (1999) Conformational changes in *E. coli* DNA topoisomerase I. *Nat. Struct. Biol.*, **6**, 918–922.

18. Xiong,B., Burk,D.L., Shen,J., Luo,X., Liu,H., Shen,J. and Berghuis,A.M. (2008) The type IA topoisomerase catalytic cycle: a normal mode analysis and molecular dynamics simulation. *Proteins*, **71**, 1984–1994.

19. Li,Z., Mondragón,A. and DiGate,R.J. (2001) The mechanism of type IA topoisomerase-mediated DNA topological transformations. *Mol. Cell*, **7**, 301–307.

20. Dekker,N.H., Rybenkov,V.V., Duguet,M., Crisona,N.J., Cozzarelli,N.R., Bensimon,D. and Croquette,V. (2002) The mechanism of type IA topoisomerases. *Proc. Natl Acad. Sci. USA*, **99**, 12126–12131.

21. Leelaram,M.N., Bhat,A.G., Godbole,A.A., Bhat,R.S., Manjunath,R. and Nagaraja,V. (2013) Type IA topoisomerase inhibition by clamp closure. *FASEB J.*, **27**, 3030–3038.

22. Mills,M., Tse-Dinh,Y.C. and Neuman,K.C. (2018) Direct observation of topoisomerase IA gate dynamics. *Nat. Struct. Mol. Biol.*, **25**, 1111–1118.

23. Spakman,D., Bakx,J.A.M., Biebricher,A.S., Peterman,E.J.G., Wuite,G.J.L. and King,G.A. (2021) Unravelling the mechanisms of type 1A topoisomerases using single-molecule approaches. *Nucleic Acids Res.*, **49**, 5470–5492.

24. Bakx,J.A.M., Biebricher,A.S., King,G.A., Christodoulis,P., Sarlós,K., Bizard,A.H., Hickson,I.D., Wuite,G.J.L. and Peterman,E.J.G. (2022) Duplex DNA and BLM regulate gate opening by the human TopoIIIα–RMI1–RMI2 complex. *Nat. Commun.*, **13**, 584.

25. Tan,K., Cao,N., Cheng,B., Joachimiak,A. and Tse-Dinh,Y.C. (2016) Insights from the structure of *Mycobacterium tuberculosis* topoisomerase I with a novel protein fold. *J. Mol. Biol.*, **428**, 182–193.

26. Cao,N., Tan,K., Annamalai,T., Joachimiak,A. and Tse-Dinh,Y.C. (2018) Investigating mycobacterial topoisomerase I mechanism from the analysis of metal and DNA substrate interactions at the active site. *Nucleic Acids Res.*, **46**, 7296–7308.

27. Changela,A., DiGate,R.J. and Mondragon,A. (2001) Crystal structure of a complex of a type IA DNA topoisomerase with a single-stranded DNA molecule. *Nature*, **411**, 1077–1081.

28. Cao,N., Tan,K., Zuo,X., Annamalai,T. and Tse-Dinh,Y.C. (2020) Mechanistic insights from structure of *Mycobacterium smegmatis* topoisomerase I with ssDNA bound to both N- and C-terminal domains. *Nucleic Acids Res.*, **48**, 4448–4462.

29. Rosenbaum,G., Alkire,R.W., Evans,G., Rotella,F.J., Lazarski,K., Zhang,R.G., Ginell,S.L., Duke,N., Naday,I., Lazarz,J. *et al.* (2006) The Structural Biology Center 19ID undulator beamline: facility specifications and protein crystallographic results. *J. Synchrotron Radiat.*, **13**, 30–45.

30. Minor,W., Cymborowski,M., Otwinowski,Z. and Chruszcz,M. (2006) HKL-3000: the integration of data reduction and structure solution—from diffraction images to an initial model in minutes. *Acta Crystallogr. D Biol. Crystallogr.*, **62**, 859–866.

31. Vagin,A. and Teplyakov,A. (2010) Molecular replacement with MOLREP. *Acta Crystallogr. D Biol. Crystallogr.*, **66**, 22–25.

32. Emsley,P., Lohkamp,B., Scott,W.G. and Cowtan,K. (2010) Features and development of Coot. *Acta Crystallogr. D Biol. Crystallogr.*, **66**, 486–501.

33. Luo,Z., Rajashankar,K. and Dauter,Z. (2014) Weak data do not make a free lunch, only a cheap meal. *Acta Crystallogr. D Biol. Crystallogr.*, **70**, 253–260.

34. Afonine,P.V., Grosse-Kunstleve,R.W., Echols,N., Headd,J.J., Moriarty,N.W., Mustyakimov,M., Terwilliger,T.C., Urzhumtsev,A., Zwart,P.H. and Adams,P.D. (2012) Towards automated crystallographic structure refinement with Phenix.refine. *Acta Crystallogr. D Biol. Crystallogr.*, **68**, 352–367.

35. Chen,V.B., Arendall,W.B., Headd,J.J., Keedy,D.A., Immormino,R.M., Kapral,G.J., Murray,L.W., Richardson,J.S. and Richardson,D.C. (2010) MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D Biol. Crystallogr.*, **66**, 12–21.

36. Wang,Y., Lynch,A.S., Chen,S.-J. and Wang,J.C. (2002) On the molecular basis of the thermal sensitivity of an *Escherichia coli* topA mutant. *J. Biol. Chem.*, **277**, 1203–1209.

37. Narula,G., Annamalai,T., Aedo,S., Cheng,B., Sorokin,E., Wong,A. and Tse-Dinh,Y.C. (2011) The strictly conserved Arg-321 residue in the active site of *Escherichia coli* topoisomerase I plays a critical role in DNA rejoining. *J. Biol. Chem.*, **286**, 18673–18680.

38. Huynh,K. and Partch,C.L. (2015) Analysis of protein stability and ligand interactions by thermal shift assay. *Curr. Protoc. Protein Sci.*, **79**, 28.29.1–28.29.14.

39. Seabrook,S.A. and Newman,J. (2013) High-throughput thermal scanning for protein stability: making a good technique more robust. *ACS Comb. Sci*, **15**, 387–392.

40. Krissinel,E. and Henrick,K. (2004) Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. *Acta Crystallogr. D Biol. Crystallogr.*, **60**, 2256–2268.

41. Annamalai,T., Dani,N., Cheng,B. and Tse-Dinh,Y.C. (2009) Analysis of DNA relaxation and cleavage activities of recombinant *Mycobacterium tuberculosis* DNA topoisomerase I from a new expression and purification protocol. *BMC Biochem.*, **10**, 18.

42. Pattabiraman,N. (1986) Can the double helix be parallel? *Biopolymers*, **25**, 1603–1606.

43. Yang,X.L., Sugiyama,H., Ikeda,S., Saito,I. and Wang,A.H. (1998) Structural studies of a stable parallel-stranded DNA duplex incorporating isoguanine:cytosine and isocytosine:guanine basepairs by nuclear magnetic resonance spectroscopy. *Biophys. J.*, **75**, 1163–1171.

44. Szabat,M. and Kierzek,R. (2017) Parallel-stranded DNA and RNA duplexes—structural features and potential applications. *FEBS J.*, **284**, 3986–3998.

45. Tripathi,S., Zhang,D. and Paukstelis,P.J. (2015) An intercalation-locked parallel-stranded DNA tetraplex. *Nucleic Acids Res.*, **43**, 1937–1944.

46. Phan,A.T. and Leroy,J.L. (2000) Intramolecular i-Motif structures of telomeric DNA. *J. Biomol. Struct. Dyn.*, **17**, 245–251.

47. Garabedian,A., Butcher,D., Lippens,J.L., Miksovska,J., Chapagain,P.P., Fabris,D., Ridgeway,M.E., Park,M.A. and Fernandez-Lima,F. (2016) Structures of the kinetically trapped i-motif DNA intermediates. *Phys. Chem. Chem. Phys.*, **18**, 26691–26702.

48. Abou Assi,H., Garavís,M., González,C. and Damha,M.J. (2018) i-Motif DNA: structural features and significance to cell biology. *Nucleic Acids Res.*, **46**, 8038–8056.

49. Hol,W.G. (1985) Effects of the alpha-helix dipole upon the functioning and structure of proteins and peptides. *Adv. Biophys.*, **19**, 133–165.

50. Sitkoff,D., Lockhart,D.J., Sharp,K.A. and Honig,B. (1994) Calculation of electrostatic effects at the amino terminus of an alpha helix. *Biophys. J.*, **67**, 2251–2260.

51. Schlegel,S., Genevaux,P. and de Gier,J.-W. (2015) De-convoluting the genetic adaptations of *E. coli* C41(DE3) in real time reveals how alleviating protein production stress improves yields. *Cell Reports*, **10**, 1758–1766.

52. Studier,F.W. (1991) Use of bacteriophage T7 lysozyme to improve an inducible T7 expression system. *J. Mol. Biol.*, **219**, 37–44.

53. Gunn,K.H., Marko,J.F. and Mondragon,A. (2017) An orthogonal single-molecule experiment reveals multiple-attempt dynamics of type IA topoisomerases. *Nat. Struct. Mol. Biol.*, **24**, 484–490.

54. Tang,S., Wu,M.K.Y., Zhang,R. and Hunter,N. (2015) Pervasive and essential roles of the Top3-Rmi1 decatenase orchestrate recombination and facilitate chromosome segregation in meiosis. *Mol. Cell*, **57**, 607–621.

55. Zhu,Q., Pongpech,P. and DiGate,R.J. (2001) Type I topoisomerase activity is required for proper chromosomal segregation in Escherichia coli. *Proc. Natl Acad. Sci. USA*, **98**, 9766–9771.

56. Bizard,A.H., Yang,X., Debat,H., Fogg,J.M., Zechiedrich,L., Strick,T.R., Garnier,F. and Nadal,M. (2018) TopA, the *Sulfolobus solfataricus* topoisomerase III, is a decatenase. *Nucleic Acids Res.*, **46**, 861–872.

57. Li,Z., Mondragon,A., Hiasa,H., Marians,K.J. and DiGate,R.J. (2000) Identification of a unique domain essential for *Escherichia coli* DNA topoisomerase III-catalysed decatenation of replication intermediates. *Mol. Microbiol.*, **35**, 888–895.

58. Terekhova,K., Marko,J.F. and Mondragon,A. (2014) Single-molecule analysis uncovers the difference between the kinetics of DNA decatenation by bacterial topoisomerases I and III. *Nucleic Acids Res.*, **42**, 11657–11667.

59. Lee,C.M., Wang,G., Pertsinidis,A. and Marians,K.J. (2019) Topoisomerase III acts at the replication fork to remove precatenanes. *J. Bacteriol.*, **201**, e00563-18.

60. Nurse,P., Levine,C., Hassing,H. and Marians,K.J. (2003) Topoisomerase III can serve as the cellular decatenase in *Escherichia coli*. *J. Biol. Chem.*, **278**, 8653–8660.