1-2016

Insights from the Structure of Mycobacterium tuberculosis Topoisomerase I with a novel protein fold

Kemin Tan
Argonne National Laboratory

Nan Cao
Department of Chemistry and Biochemistry, Florida International University, ncao@fiu.edu

Bokun Cheng
New York Medical College

Andrezej Joachimiak
Argonne National Laboratory

Yuk-Ching Tse-Dinh
Biomolecular Sciences Institution, Florida International University, ytsedinh@fiu.edu

Follow this and additional works at: https://digitalcommons.fiu.edu/biomolecular_fac

Recommended Citation

Insights from the Structure of *Mycobacterium tuberculosis* Topoisomerase I with a novel protein fold

Kemin Tan,[1*] Nan Cao,[2] Bokun Cheng,[3] Andrzej Joachimiak,[1] Yuk-Ching Tse-Dinh[2,4*]

[1] Structural Biology Center, Biosciences Division, Argonne National Laboratory, Argonne, IL 60439, USA

[2] Department of Chemistry and Biochemistry, Florida International University, Miami, FL 33199, USA.

[3] Department of Biochemistry and Molecular Biology, New York Medical College, Valhalla, NY 10595, USA.

[4] Biomolecular Sciences Institute, Florida International University, Miami, FL 33199, USA.

*Corresponding authors

E-mail: ktan@anl.gov (KT), ytsedinh@fiu.edu (YT)

Tel: 1-630-252-3948 (KT), 1-305-348-4956 (YT)

Abstract

The DNA Topoisomerase I enzyme of *Mycobacterium tuberculosis* (MtTOP1) is essential for the viability of the organism and survival in a murine model. This topoisomerase is being pursued as a novel target for the discovery of new therapeutic agents for the treatment of drug resistant tuberculosis (TB). In this study, we succeeded in obtaining a structure of MtTOP1 by first predicting that the C-terminal region of MtTOP1 contains four repeated domains that do not involve the Zn-binding tetra-cysteine motifs seen in the C-terminal domains of *E. coli* topoisomerase I. A construct (a.a. A2-T704), MtTOP1-704t, that includes the N-terminal domains (D1-D4) and the first predicted C-terminal domain (D5) of MtTOP1 was expressed and found to retain DNA cleavage-religation activity and catalyze ssDNA catenation. MtTOP1-704t was crystallized, and a structure of 2.52 Å resolution limit was obtained. The structure of the MtTOP1 N-terminal domains has features that have not been observed in the other previously available bacterial topoisomerase I crystal structures. The first C-terminal domain D5 forms a novel protein fold of a four-stranded antiparallel β-sheet stabilized by a crossing-over α-helix. Since there is only one type IA topoisomerase present in Mycobacteriaceae and related Actinobacteria, this subfamily of type IA topoisomerase may be required for multiple functions in DNA replication, transcription, recombination, and repair. The unique structural features observed for MtTOP1 may allow these topoisomerase I enzymes to carry out physiological functions associated with topoisomerase III enzyme in other bacteria.

Key words

TB, TopA, protein fold, protein domains, crystallization

Abbreviations

MtTOP1, *M. tuberculosis* topoisomerase I

TB, tuberculosis

Introduction

Topoisomerases are needed in all organisms to maintain the optimal level of DNA supercoiling for physiological functions, and to resolve topological barriers encountered during processes of replication, transcription, recombination, and repair.[1, 2] In most organisms, multiple topoisomerases are present, and each cellular topoisomerase has specific roles associated with these physiological processes.[3] The genome of the pathogen *Mycobacterium tuberculosis* encodes only one type I topoisomerase that catalyzes topological transitions by cutting and rejoining a single-strand of DNA and one type II topoisomerase that carries out catalysis by cutting and rejoining a double-strand of DNA. The type II topoisomerase (DNA gyrase) of *M. tuberculosis* is the target of moxifloxacin, a fluoroquinolone antibacterial agent being utilized in a new tuberculosis (TB) drug combination regimen and found to be beneficial for reducing TB treatment duration.[4] The only type I topoisomerase in *M. tuberculosis* (MtTOP1, Rv3646c) is a type IA topoisomerase that has been found to be essential for growth, and its depletion was shown to decrease survival following treatment with novobiocin and isoniazid.[5] Each year, around 9 million people worldwide develop TB caused by *Mycobacterium tuberculosis*, accounting for 1.3 million deaths. MtTOP1 has been actively investigated as a target for novel TB drugs[6-8] that are very much needed for the treatment of drug-resistant TB. The structure of MtTOP1 will be essential for structure-based rational drug discovery efforts.

Our early trials to obtain diffraction quality crystals of the full length MtTOP1 protein were unsuccessful. Constructs of the MtTOP1 N-terminal region corresponding to the N-terminal domains D1-D4 found in *E. coli* topoisomerase I (EcTOP1) crystal

structures did not express well as soluble proteins. In a recently determined structure of full length EcTOP1 in complex with ssDNA[9] (PDB ID: 4RUL), the C-terminal region of EcTOP1 comprises three 4-Cys zinc ribbon domains and two zinc ribbon-like domains, and a modular ssDNA-binding mode was observed. This suggests that the C-terminal region of MtTOP1 may also contain multiple domains with repeated and conserved sequence motifs even though the sequences of EcTOP1 and MtTOP1 C-terminal regions are very different with essentially no homologies. We analyzed the structural aspects of the C-terminal region of MtTOP1. The recognition of individual domains in this region helped us redesign constructs for functional studies and enabled us to obtain diffraction quality crystals for structural characterization of MtTOP1. We report here the crystal structure of MtTOP1-704t with the N-terminal domains D1-D4 and C-terminal domain D5 (a.a. A2-T704) and its functional characterizations.

Results

Analysis of C-terminal domains of MtTOP1

MtTOP1 does not have easily recognizable sequence motifs (such as the repeated 4-Cys Zn-binding motifs present in EcTOP1) that can help divide the MtTOP1 C-terminal region into domains. Nevertheless, MtTOP1 does bear four GxxGPY sequence motifs[10] in its C-terminal region. Based on the locations of the four GxxGPY repeats, we divided the region into four repeat domains (Fig. 1). A secondary structure prediction indicated that each domain comprises three or four β-strands and one C-terminal α-helix.[11] The predicted folding of the MtTOP1 C-terminal domains is different from the zinc ribbon domains or zinc ribbon-like domains found in EcTOP1 and

Thermotoga maritima topoisomerase I.[12] After the fourth or last C-terminal domain there is an additional long highly positively charged tail (~ 26 amino acids) in MtTOP1. Besides the basic C-terminal tail, there are two more stretches of basic amino acids.[13] Based on the prediction, one is a part of the insertion on the loop between the β3 and β4 strands of D6, and the other is a part of the linker between the D7 and D8 domains (Fig. 1). Deletions of the stretches of basic amino acids in the C-terminal region of MtTOP1 have been shown to affect DNA binding and DNA strand passage.[13] The deletions would unlikely make an impact on the folding of the individual C-terminal domains but may affect the relative orientations of the C-terminal domains and the inter-domain flexibility.

MtTOP1-704t is one of the constructs that have been made based on the C-terminal domain prediction above. It spans from amino acid A2 to T704 and includes the N-terminal domains and the first predicted C-terminal domain D5 (Fig.1 and Graphical Abstract). The functional and structural characterization of MtTOP1-704t is the focus of this study.

Biochemical activities of MtTOP1-704t

MtTOP1-704t can be expressed readily as a soluble recombinant protein in *E. coli* host strain T7 Express Crystal. The protein was purified to >99% homogeneity and assayed for relaxation of negatively supercoiled plasmid DNA. The results in Fig. 2 showed that while relaxation activity can be detected with 6.2 ng (60 fmole) of the full length MtTOP1, relaxation activity was not observed with 100 ng (1.3 pmole) of MtTOP1-704t.

Although relaxation activity could not be detected for MtTOP1-704t, the protein was found to have DNA cleavage activity similar to that of full length MtTOP1 (Fig. 3A). The same preferred cleavage site with a cytosine base that is four nucleotides upstream of the cleavage site was recognized for formation of the cleavage product.[14] Intramolecular religation was promoted by adding Mg(II) to shift the cleavage-religation equilibrium toward religation along with 1 M NaCl to dissociate the non-covalent enzyme-DNA complex. The results in Fig. 3B showed that MtTOP1-704t is efficient in DNA religation, so the C-terminal sequences not found in MtTOP1-704t do not affect the cleavage and rejoining of ssDNA by MtTOP1.

There are two types of IA topoisomerases in most species of bacteria that have been studied. *E. coli* topoisomerase III (EcTOP3) is much more efficient than EcTOP1 in DNA catenation/decatenation, while EcTOP1 can remove negative supercoils much more rapidly.[15] Since *M. tuberculosis* only has one type of IA topoisomerase enzyme, MtTOP1 is likely to be required for the resolution of single-stranded DNA recombination and replication intermediates. The catenation activity of *Mycobacterium smegmatis* topoisomerase I has been demonstrated previously with M13 single-stranded DNA.[16] Since MtTOP1-704t was found to have robust DNA cleavage-religation capabilities, the truncated enzyme was assayed for catenation of single-stranded M13mp18 DNA circles to determine if it can catalyze passage of a single-stranded DNA through the break created by enzyme cleavage on another single-strand. The results (Fig. 4) showed that MtTOP1-T704 has comparable catenation activity to that of full length MtTOP1.

Overall structure of MtTOP1-704t

The crystal structure of MtTOP1-704t was determined with molecular replacement[17] and alternative cycles of model building and refinement (Table 1). In the asymmetric unit, there is one MtTOP1-704t as shown in Fig. 5. The 15 N-terminal residues (A2 to S16) are not visible in the electron density maps of the structure. This indicates the flexible nature of the N-terminal leading peptide, and the flexibility is consistent with its glycine-rich (five Gly residues) composition. Interestingly, the first 15 residues (A2 to S16) of the peptide are "extra" compared to EcTOP1, Fig. 6. Additionally, the leading peptide is positively charged (including one Arg and two Lys residues, Supplementary Fig. 1). The functionality of this leading peptide is currently unknown.

The model of the crystal structure of MtTOP1-704t starts from G17, which corresponds to the starting residue G2 of the EcTOP1 structure (Fig. 6). There is no indication for a stable quaternary assembly based on a PISA (Protein, Interface, Structures, and Assemblies) analysis,[18] which is consistent with the observation that MtTOP1 is a monomer in solution. The four N-terminal domains (D1-D4, G17-A638) have a similar toroid assembly as EcTOP1 (G2-N590) while the first C-terminal domain D5 shows sequential and structural novelties among known protein folds as discussed later.

To evaluate the similarity of the N-terminal assembly of MtTOP1 to that of EcTOP1, the N-terminal domains of these two molecules were aligned,[19] resulting in a root mean square deviation (rmsd) of 1.92 Å and sequence identity of 44.9% (Fig.7A). Pair-wise alignments of individual domains resulted in rmsd values of 1.01 Å, 1.46 Å, 1.35 Å, and 1.39 Å for D1, D2, D3, and D4, respectively. Their sequence identities are

56.0%, 26.9%, 49.6%, and 49.2%, respectively. Apparently, D1 is the most conserved while D2 is the least conserved between MtTOP1 and EcTOP1.

Similarly, when the N-terminal region of MtTOP1 is superimposed on EcTOP3 (Fig. 7B), the calculated rmsd value was 3.66 Å with a sequence identity of 21.5%. Relative domain-domain movement of these enzymes in their crystals could impact the superposition involving multiple domains. Pair-wise individual domain alignments of MtTOP1 and EcTOP3 yielded rmsd values of 1.76 Å, 1.93 Å, 2.11 Å, and 1.85 Å for D1, D2, D3 and D4, respectively.   Their sequence identities are 28.3%, 19.5%, 31.3%, and 25.2%, respectively. From these comparisons it is clear that MtTOP1 is closer to EcTOP1 than EcTOP3. However, MtTOP1 has some local structural features similar to EcTOP3, to be elaborated on later. There is no metal binding observed at the active site of the Apo form of MtTOP1.

Unique features of individual N-terminal domains of MtTOP1

The D1 or so called TOPRIM (TOpoisomerase-PRIMse) domain is the most conserved domain in the N-terminal region among these enzymes based on the alignments above. In MtTOP1 it includes residues G17 to D157 except the region from R50 to S86, which sticks out from D1 toward D4 (colored grey in Fig. 5). In the EcTOP1 structure, a similar protruding region is defined as a part of D4 because of its close association with the main body of D4 and its partial role in forming a ssDNA-binding groove on D4.[20, 21] For comparison, we define this region in MtTOP1 as a part of D4 as well, although it is described as a part of D1 in the EcTOP3 structure.[22] In MtTOP1, this protruding region has no sequence homology to the corresponding region in EcTOP1

and is 15 amino acids shorter (Figs. 5 and 6). Though this protruding region from R50

to S86 does not have a prominent α-helix seen in EcTOP1 (Fig. 7A), it is well

structured, comprising five β turns and a short β hairpin. Interestingly, in EcTOP3, this

portion is also a tight loop, similarly consisting of multiple β turns[22] (Fig. 7B).   The

significance of the structural variation of the protrusion remains unknown.

The hinge domain D2 comprises two sequentially separated regions: F213 to

R302 (β1-β2-α1-β3) and D434 to R516 (β4-β5-α2-β6) (Figs. 5 and 6). Though the

sequence identities of the domain between MtTOP1 and EcTOP1 or EcTOP3 are as

low as 26.9% and 19.5%, respectively, the hinge domain of MtTOP1 essentially has a

similar structure to that of the two *E.coli* topoisomerases except for two rather prominent

insertions in MtTOP1. The first insertion (T252 to I270, colored magenta in Figs. 5A

and 7) creates a large loop between β2 and α1 on the right side behind the toroid hole.

The magenta loop is rich in charged residues and seemingly creates a positively

charged edge on the top of the toroid hole (Fig. 5). A smaller insertion (loop214-255)

was found in EcTOP3, which forms a helical motif instead (Fig 7B).[22] The second

insertion (G480 to A485, colored red in Figs. 5A and 7) is part of a long loop between α2

and β6 in the bottom front of the toroid hole. This loop is negatively charged. The two

charged loops (colored in magenta and red) positioned on the two sides of the toroid

hole of MtTOP1 might play an important coordinating role in the proposed route of the

passing strand during catalysis of DNA relaxation by attracting or repelling DNA. The

magenta colored insertion T252 to I270 is unique to MtTOP1 and EcTOP3 and may

play roles that differentiate MtTOP1 and EcTOP3 from EcTOP1.

D3 is an all α-helical domain (P303 to A433), including the catalytic residue Y342 on a loop between the α2 and α3 helices. The catalytic site of the Apo form of MtTOP1 centered at Y342 shows no significant conformational variation compared to the catalytic sites of the Apo forms of either EcTOP1[23] or EcTOP3,[22] indicating the same catalytic mechanism for these type IA topoisomerases. The catalytic site also includes the DxDREG TOPRIM sequence motif from D1, which is quite conserved (Fig. 6). The only difference is the variation in residue x, which is Gly112 in MtTOP1, a Leu in EcTOP1, and a Pro in EcTOP3. Additionally, between the α3 and α4 helices of D3, there is a long loop of 35 residues, including the sequence motif Q386EAHEAIRP that is highly conserved in type IA topoisomerases, with the invariant histidine proposed to play a role in catalysis[24]. In all of the EcTOP1 structures determined to date, a part of the loop upstream of the conserved sequence is partially disordered. In MtTOP1, in contrast, this loop is well structured without any break and seemingly correlates to the formation of the α4 helix (highlighted in blue, Fig. 5), which degenerates to a β-turn in EcTOP1 (Fig. 7A).

In MtTOP1, a feature of D4 may be related to one of its physiological functions; the last helix in D4 (α8) has an insertion of 14 residues (G610 to G623, colored purple in Figs. 5 and 7) in the middle of the helix, which is not found in EcTOP1 and EcTOP3. The purple insertion is right below the decatenation loop of EcTOP3,[22] which is also from D4 (Fig. 7B). The actual functionality of this unusual insertion in the α8 helix needs to be explored. The α8 helix is then followed by a linker to the C-terminal domains. The linker is very different in sequence and structure among MtTOP1, EcTOP1, and EcTOP3, indicating the connection between the N- and C-terminal domains varies

significantly among these topoisomerases. However, the DNA-binding groove is mostly conserved (Fig. 5).   Two negatively charged sulfate groups likely from the crystallization buffer were identified inside the positively charged DNA-binding groove (not shown). They may mimic two phosphate groups of DNA in their interactions within the groove.

Novel MtTOP1 C-terminal domain

D5 is the first C-terminal domain, from R638 to P705. This small domain of 67 residues is composed of a four-stranded antiparallel β-sheet with a C-terminal helix packed against the β-sheet on one side (Fig. 8). Hydrophobic interaction dominates the interaction between the β-sheet and the helix. An additional salt bridge (R668-E699) and hydrogen bonds are also observed. The helix apparently plays an important role in stabilizing the β-sheet and the small domain. The other side of the β-sheet is open and hosts several residues that may be involved in DNA binding. The conserved tyrosine Y665, the third residue of the β3-strand, could be a key residue in the interaction of D5 with DNA by forming π–stacking(s) with the base(s) of DNA (Fig. 5). The other residues Y656 and R658 on the β2-strand, R681 on the β4-strand and K661 on the sharp β2-β3 turn (Fig. 5) can potentially form π–stacking, π–cation, or other forms of interaction with DNA.

A Dali structural homology search[25] of MtTOP1 results in no significant hits (Z score ≤ 4.3, rmsd ≥ 2.6, and sequence identity ≤ 14%) even though a protein fold containing a helix packed against one face of an anti-parallel β-sheet is common. This indicates the structural novelty of the MtTOP1 C-terminal domain.

Classification of topoisomerase C-terminal domains.

From Fig. 1, it is clear that the four MtTOP1 C-terminal domains exemplified by D5 represent one subgroup (Topo_C_Rpt) of bacterial topoisomerase C-terminal repeats. In this subgroup a C-terminal domain has a four-stranded antiparallel β-sheet compounded or stabilized by a crossing-over α-helix on one side of the β-sheet. Besides MtTOP1, Topo_C_Rpt can also be found in other *Mycobacterium* species as well as related species of *Actinobacteria* (Supplementary Fig. 1). The second subgroup of the C-terminal domain of topoisomerases, namely Topo_C_ZnRpt, is represented by the 4-Cys zinc ribbon domains of EcTOP1, in which one four-stranded antiparallel β-sheet is stabilized by a Zn-binding site on the top of each domain[9] (PDB ID: 4RUL). Variations of these two subgroups or even new subgroups are expected with further structural characterization of topoisomerases. For example, the only C-terminal domain of *Thermotoga maritima* topoisomerase I can be regarded as a member of Topo_C_ZnRpt but without Zn binding.

Discussion

Trapping of the bacterial topoisomerase I covalent complex on cleaved chromosomal DNA has been shown to result in rapid bacterial cell death, validating this class of enzyme as a useful target for novel bactericidal antibiotics[26, 27]. The structure of MtTOP1-704t reported here provides details of the active site and its vicinity that would be useful for the discovery of poison-type inhibitors for trapping of the topoisomerase covalent intermediate. Topoisomerase poisons are typically interfacial inhibitors

interacting with both DNA and the topoisomerase.[28] Interactions between potential

poison-type inhibitors and MtTOP1 protein residues near the active site are needed to

confer high binding affinity and selectivity. For example, while m-AMSA has been

shown to inhibit *M. tuberculosis* topoisomerase I and growth of *M. tuberculosis,*[29] this

compound has much greater selectivity for eukaryotic type II topoisomerases. A binding

pocket for functional groups in m-AMSA present near the active site of human type II

topoisomerases accounts for the specific interactions with m-AMSA and its

effectiveness as an anticancer drug. The construct of MtTOP1-704t could be utilized in

the future for co-crystallization of an enzyme-DNA-inhibitor ternary complex to further

enable the discovery of new TB therapies.

Catalytic inhibitors of MtTOP1 could also be useful for treating TB by inhibiting

the essential function of the enzyme. MtTOP1-704t has no detectable relaxation

activity. Some or all other C-terminal parts of MtTOP1 including D6, D7, and D8 or

even the flexible tail may be required for relaxation of DNA supercoiling. MtTOP1-704t

can carry out DNA cleavage-religation and catalyze the catenation of ssDNA with

efficiencies comparable to those of full-length MtTOP1. All *Mycobacterium* species and

most members of the *Actinobacteria* phylum have only one type IA topoisomerase

encoded in their genomes. These include the topoisomerase I with a domain

organization similar to MtTOP1 found in the pathogens *Corynebacterium diphtheria* and

Actinomyces odontolyticus (Supplementary Fig. 1) that are likely to also be essential for

viability. Unique features observed in the N-terminal domains of MtTOP1, such as the

charged loops on opposite sides of the toroid hole and the insertion at the position near

where the decatenation loop of EcTOP3 can be found, may also have evolved for

MtTOP1 to carry out all the catalytic functions that may be associated with bacterial type IA topoisomerases. Further studies are required to determine the biochemical and physiological significance of these unique structural features.

The novel structure of the MtTOP1 C-terminal domain D5 reported here provides new insights into the folding of diverse bacterial topoisomerase I proteins, particularly their C-terminal regions. The organization of repeated bacterial topoisomerase I C-terminal domains may have arisen during evolution through gene duplication. Gene segments that encode a specific protein fold that can interact with ssDNA have been combined in the evolutionary process with the segments encoding the N-terminal domains that have the DNA cleavage-religation active site at the entrance of a toroid hole. Different repeats of the Topo_C_Rpt or Topo_C_ZnRpt domains or their likenesses are utilized in each bacterial species for the DNA-protein or protein-protein interactions that are required for the catalytic activity and physiological functions of this ubiquitous enzyme. Interaction between the C-terminal domains of EcTOP1 and RNA polymerase is important for the efficient removal of transcription-driven negative supercoiling during transcription elongation.[30] The C-terminal domains of MtTOP1 may interact similarly with *M. tuberculosis* RNA polymerase for relieving the torsional stress generated during transcription.

The flexibility of the linkers between the individual C-terminal domains may account for the difficulty encountered in the crystallization of full-length bacterial topoisomerase I proteins. This could be addressed by introducing a DNA binding substrate or protein interaction partner to limit the degree of interdomain movement, facilitating the crystallization of full length proteins. Successful determination of full-

length *Mycobacterium* topoisomerase I crystal structures will provide additional new insights into the mechanism and function of this essential enzyme.

Materials and Methods

Cloning, expression, and purification

A clone for expression of MtTOP1-704t was constructed using the Gibson Assembly Master Mix from New England BioLabs, placing the coding sequence of residues 2-704 of MtTOP1 under the control of the T7 promoter in pET-His$_6$-Mocr TEV-LIC cloning vector (2O-T) that was a gift from Scott Gradia (Addgene plasmid # 29710). The MtTOP1-704t protein was expressed in the *E. coli* T7 Express Crystal strain (New England BioLabs). Cells were grown in Luria Bertani broth at 37°C to the exponential phase (OD$_{600}$ = 0.4) and left at room temperature for 20 min. Recombinant protein expression was induced by adding 1 mM IPTG. After overnight induction at 22°C, the cells were collected and subjected to freeze-thaw lysis in buffer (20 mM sodium phosphate pH 7.4, 0.5 M NaCl, 20 mM imidazole, 1 mg/ml lysozyme). The recombinant protein in the soluble lysate was allowed to bind to Ni Sepharose 6 Fast Flow (GE Healthcare) and packed into a column. After extensive washing with buffer (20 mM sodium phosphate pH 7.4, 0.5 M NaCl, 20 mM imidazole), the topoisomerase protein was eluted with buffer (20 mM sodium phosphate, pH 7.4, 0.5 M NaCl, 500 mM imidazole). Following treatment with TEV protease, the protein was passed through the Ni Sepharose again to remove the His$_6$-Mocr tag. Additional purification was achieved with single-stranded DNA cellulose column chromatography (Sigma) using an increasing concentration gradient of KCl to elute the protein. The final purification step

utilized an S300 size exclusion chromatography column as described previously.[31] The

purified protein was concentrated to about 85 mg/ml for crystallization.

Crystallization

The screening for crystallization conditions was set up with a Mosquito nanoliter liquid

handler (TTP LabTech) using the sitting drop vapor diffusion technique in 96-well

CrystalQuick plates (Greiner). For each condition, 0.2 µl of protein and 0.2 µl of

crystallization formulation were mixed, and the mixture was equilibrated against 140 µl

of the crystallization solution in each reservoir well. The crystallization screens used

were MCSG-1–4 at 16°C. Diffraction quality crystals appeared under the following

condition: 0.2 M Ammonium Sulfate, 0.1 M Tris:HCl pH 8.5, and 25% (w/v) PEG3350.

Prior to data collection, the crystals were treated with cryoprotectant (20% glycerol in

crystallization buffer) and cryocooled directly in liquid nitrogen.

X-ray diffraction and structure determination

Sets of single-wavelength diffraction data were collected at 12.66 keV at 100 K

from cryocooled MtTOP1-704t crystals. All data were obtained at the 19-ID beamline of

the Structural Biology Center at the Advanced Photon Source at Argonne National

Laboratory using the program SBCcollect.[32] The intensities of each data set were

integrated, scaled, and merged with the HKL3000 program suite[33] (Table 1). The best

data set measured has a resolution limit of 2.52 Å. The structures of the MtTOP1-704t

crystal were determined using the molecular replacement method.[17] The best search

template used was the structures of the N-terminal domains (D1-D4) of full-length

EcTOP1 (PDB ID: <u>4RUL</u>). Subsequent model rebuilding for the N-terminal domains and building the first C-terminal domain (D5) based on difference maps was performed using the program COOT[34]. After several alternative cycles of model building and refinement, the final model was refined using the program Phenix.refine[35] (Table 1). Structural validation was performed using the program MolProbity.[36]

Relaxation activity assay

The relaxation reaction was carried out in 20 µl of 10 mM Tris-HCl (pH 8.0) 50 mM NaCl, 0.1 mg/mL gelatin, and 5 mM $MgCl_2$ with 250 ng of supercoiled pBAD/Thio plasmid DNA purified by CsCl gradient centrifugation. Following incubation at 37°C for 30 min, the reactions were terminated and analyzed by agarose gel electrophoresis as previously described.[37]

Assay of oligonucleotide DNA cleavage and religation

The 32 base oligonucleotide substrate 5'-CAGTGAGCGAGCTTCCGCT↓TGACATCCCAATA-3' with a preferred cleavage site for MtTOP1 indicated by the arrow has been previously described.[13] It was labeled at the 5'-end with γ-^{32}P-ATP and T4 polynucleotide Kinase. Cleavage activity was assayed by incubating the labeled oligonucleotide (0.5 pmole) with MtTOP1 and MtTOP1-704t in 5 µl of 10 mM Tris-HCl pH 8.0 at 37°C for 30 min. Intramolecular religation was initiated by addition of $MgCl_2$ followed by further incubation for 15 min. The reactions were terminated by the addition of an equal volume of sequencing gel loading buffer. The reaction substrate and the 19 base-long cleavage product were separated by

electrophoresis in a 15% sequencing gel followed by Phosphor-Imager analysis of the gel.

Catenation activity assay

Catenation of M13 ssDNA circles was assayed with a protocol similar to previously described conditions[16]. The reaction volume of 20 μl contained 20 mM Tris-HCl (pH 8.0), 20 mM KCl, 2 mM $MgCl_2$, 0.05 mg/ml BSA, 10 mM spermidine, 30% glycerol, and 0.5 μg of M13 ssDNA (from Bayou Biolabs). After incubation with the indicated amount of topoisomerase protein for 1 h at 52°C, the reactions were terminated with the addition of 1% SDS, 5 mM EDTA, and 12 μg of proteinase K, and left at 37°C for 1 h before electrophoresis in 0.8% agarose gel.

Accession Number

The atomic coordinates and structure factors of *M. tuberculosis* topoisomerase I have been deposited in the Protein Data Bank with accession code 5D5H.

References

1.   Vos S. M., Tretter E. M., Schmidt B. H. & Berger J. M. (2011) All tangled up: how cells direct, manage and exploit topoisomerase function. Nat. Rev. Mol. Cell Biol. 12, 827-841. 10.1038/nrm3228; 10.1038/nrm3228.

2.   Chen S. H., Chan N. L. & Hsieh T. S. (2013) New mechanistic and functional insights into DNA topoisomerases. Annu. Rev. Biochem. 82, 139-170. 10.1146/annurev-biochem-061809-100002; 10.1146/annurev-biochem-061809-100002.

3.   Forterre P., Gadelle D. (2009) Phylogenomics of DNA topoisomerases: their origin and putative roles in the emergence of modern organisms. Nucleic Acids Res. 37, 679-692. 10.1093/nar/gkp032.

4.   Dawson R., Diacon A. H., Everitt D., van Niekerk C., Donald P. R., Burger D. A., Schall R., Spigelman M., Conradie A., Eisenach K., Venter A., Ive P., Page-Shipp L., Variava E., Reither K., Ntinginya N. E., Pym A., von Groote-Bidlingmaier F. & Mendel C. M. (2015) Efficiency and safety of the combination of moxifloxacin, pretomanid (PA-824), and pyrazinamide during the first 8 weeks of antituberculosis treatment: a phase 2b, open-label, partly randomised trial in patients with drug-susceptible or drug-resistant pulmonary tuberculosis. Lancet 385, 1738-1747. S0140-6736(14)62002-X [pii].

5.   Ahmed W., Menon S., Godbole A. A., Karthik P. V. & Nagaraja V. (2014) Conditional silencing of topoisomerase I gene of *Mycobacterium tuberculosis* validates its essentiality for cell survival. FEMS Microbiol. Lett. 353, 116-123. 10.1111/1574-6968.12412 [doi].

6.   Sridevi J. P., Suryadevara P., Janupally R., Sridhar J., Soni V., Anantaraju H. S., Yogeeswari P. & Sriram D. (2015) Identification of potential *Mycobacterium tuberculosis*

topoisomerase I inhibitors: a study against active, dormant and resistant tuberculosis. Eur. J. Pharm. Sci. 72, 81-92. 10.1016/j.ejps.2015.02.017 [doi].

7.   Godbole A. A., Ahmed W., Bhat R. S., Bradley E. K., Ekins S. & Nagaraja V. (2015) Targeting *Mycobacterium tuberculosis* topoisomerase I by small-molecule inhibitors. Antimicrob. Agents Chemother. 59, 1549-1557. 10.1128/AAC.04516-14 [doi].

8.   Ravishankar S., Ambady A., Awasthy D., Mudugal N. V., Menasinakai S., Jatheendranath S., Guptha S., Sharma S., Balakrishnan G., Nandishaiah R., Ramachandran V., Eyermann C. J., Reck F., Rudrapatna S., Sambandamurthy V. K. & Sharma U. K. (2015) Genetic and chemical validation identifies *Mycobacterium tuberculosis* topoisomerase I as an attractive anti-tubercular target. Tuberculosis (Edinb) S1472-9792(15)20742-1 [pii].

9.   Tan K., Zhou Q, Cheng B., Zhang Z., A, Joachimiak A. & tse-Dinh, Y.C.(2015) Structural basis for suppression of hypernegative DNA supercoiling by *E. coli* topoisomerase I. Nucleic Acids Res., Oct 20. pii: gkv1073. [Epub ahead of print]

10.   Schuster-Bockler B., Schultz J. & Rahmann S. (2004) HMM Logos for visualization of protein families. BMC Bioinformatics 5, 7. 10.1186/1471-2105-5-7 [doi].

11.   Drozdetskiy A., Cole C., Procter J. & Barton G. J. (2015) JPred4: a protein secondary structure prediction server. Nucleic Acids Res. 43, W389-94. 10.1093/nar/gkv332 [doi].

12.   Hansen G., Harrenga A., Wieland B., Schomburg D. & Reinemer P. (2006) Crystal structure of full length topoisomerase I from *Thermotoga maritima*. J. Mol. Biol. 358, 1328-1340. S0022-2836(06)00332-9 [pii].

13.	Ahmed W., Bhat A. G., Leelaram M. N., Menon S. & Nagaraja V. (2013) Carboxyl terminal domain basic amino acids of mycobacterial topoisomerase I bind DNA to promote strand passage. Nucleic Acids Res. 41, 7462-7471. 10.1093/nar/gkt506 [doi].

14.	 Annamalai T., Dani N., Cheng B. & 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. 10.1186/1471-2091-10-18.

15.	Terekhova K., Marko J. F. & Mondragon A. (2013) Studies of bacterial topoisomerases I and III at the single-molecule level. Biochem. Soc. Trans. 41, 571-575. 10.1042/BST20120297 [doi].

16.	Bhaduri T., Bagui T. K., Sikder D. & Nagaraja V. (1998) DNA topoisomerase I from *Mycobacterium smegmatis*. An enzyme with distinct features. J. Biol. Chem. 273, 13925-13932. .

17.	Vagin A., Teplyakov A. (2010) Molecular replacement with MOLREP. Acta Crystallogr. D Biol. Crystallogr. 66, 22-25. 10.1107/S0907444909042589 [doi].

18.	Krissinel E., Henrick K. (2007) Inference of macromolecular assemblies from crystalline state. J. Mol. Biol. 372, 774-797. S0022-2836(07)00642-0 [pii].

19.	 Krissinel E., 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.

20.	Perry K., Mondragon A. (2003) Structure of a complex between *E. coli* DNA topoisomerase I and single-stranded DNA. Structure 11, 1349-1358. .

21.   Cheng B., Feng J., Mulay V., Gadgil S. & Tse-Dinh Y. C. (2004) Site-directed mutagenesis of residues involved in G Strand DNA binding by *Escherichia coli* DNA topoisomerase I. J. Biol. Chem. 279, 39207-39213. 10.1074/jbc.M405891200.

22.   Mondragon A., DiGate R. (1999) The structure of *Escherichia coli* DNA topoisomerase III. Structure 7, 1373-1383. st7b02 [pii].

23.   Lima C.D., Wang J.C. & MondragonA. (1994) Three-dimensional structure of the 67K N-terminal fragment of *E. coli* DNA topoisomerase I. Nature, 367, 138-146.

24.   Perry K., Mondragon A. (2002) Biochemical characterization of an invariant histidine involved in *Escherichia coli* DNA topoisomerase I catalysis. J. Biol. Chem. 277, 13237-13245. 10.1074/jbc.M112019200.

25.   Holm L., Rosenstrom P. (2010) Dali server: conservation mapping in 3D. Nucleic Acids Res. 38, W545-9. 10.1093/nar/gkq366 [doi].

26.   Tse-Dinh Y. C. (2015) Targeting bacterial topoisomerase I to meet the challenge of finding new antibiotics. Future Med. Chem. 7, 459-471. 10.4155/fmc.14.157 [doi].

27.   Tse-Dinh Y. C. (2009) Bacterial topoisomerase I as a target for discovery of antibacterial compounds. Nucleic Acids Res. 37, 731-737. 10.1093/nar/gkn936.

28.   Pommier Y. (2013) Drugging topoisomerases: lessons and challenges. ACS Chem. Biol. 8, 82-95. 10.1021/cb300648v; 10.1021/cb300648v.

29.   Godbole A. A., Ahmed W., Bhat R. S., Bradley E. K., Ekins S. & Nagaraja V. (2014) Inhibition of *Mycobacterium tuberculosis* topoisomerase I by m-AMSA, a eukaryotic type II topoisomerase poison. Biochem. Biophys. Res. Commun. 446, 916-920. 10.1016/j.bbrc.2014.03.029 [doi].

30. Cheng B., Zhu C. X., Ji C., Ahumada A. & 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. 10.1074/jbc.M303403200 [doi]; M303403200 [pii].

31. Zhang Z., Cheng B. & 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. U. S. A. 108, 6939-6944. 10.1073/pnas.1100300108.

32. Rosenbaum G., Alkire R. W., Evans G., Rotella F. J., Lazarski K., Zhang R. G., Ginell S. L., Duke N., Naday I., Lazarz J., Molitsky M. J., Keefe L., Gonczy J., Rock L., Sanishvili R., Walsh M. A., Westbrook E. & Joachimiak A. (2006) The Structural Biology Center 19ID undulator beamline: facility specifications and protein crystallographic results. J. Synchrotron Radiat. 13, 30-45. S0909049505036721 [pii].

33. Minor W., Cymborowski M., Otwinowski Z. & 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. 10.1107/S0907444906019949.

34. Emsley P., Lohkamp B., Scott W. G. & Cowtan K. (2010) Features and development of Coot. Acta Crystallogr. D Biol. Crystallogr. 66, 486-501. 10.1107/S0907444910007493.

35. 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. & Adams P. D. (2012) Towards automated crystallographic structure refinement with phenix.refine. Acta Crystallogr. D Biol. Crystallogr. 68, 352-367. 10.1107/S0907444912001308 [doi].

36.   Davis I. W., Leaver-Fay A., Chen V. B., Block J. N., Kapral G. J., Wang X.,

Murray L. W., Arendall W. B.,3rd, Snoeyink J., Richardson J. S. & Richardson D. C.

(2007) MolProbity: all-atom contacts and structure validation for proteins and nucleic

acids. Nucleic Acids Res. 35, W375-83. gkm216 [pii].

37.   Zhu C. X., Tse-Dinh Y. C. (2000) The acidic triad conserved in type IA DNA

topoisomerases is required for binding of Mg(II) and subsequent conformational

change. J. Biol. Chem. 275, 5318-5322.

Figure Legends

Fig. 1. Division of the C-terminal region of *M. tuberculosis* topoisomerase I into four domains and a flexible tail. The conserved residues among the four domains are highlighted in red, including the GxxGPY sequence motif. Besides the conserved tyrosine residues within the GxxGPY motifs, other possible DNA binding residues are highlighted in magenta (for possible π stacking to the bases of nucleotides) or blue (for possible π–cation or salt bridge or hydrogen bond to DNA).   The secondary structures of D5 are indicated above the appropriate sequence of D5. The highly positively charged tail after D8 is listed below the alignment, and it is presumably structurally flexible.

Fig. 2. Assay of relaxation activity of full length MtTOP1 and MtTOP1-704t. Supercoiled pBAD/Thio plasmid DNA was incubated with the indicated amount of MtTOP1 or MtTOP1-704t at 37°C for 30 min before separation of the supercoiled DNA from relaxed topoisomers by agarose gel electrophoresis.

Fig. 3. DNA cleavage and religation by MtTOP1-704t (A) Comparison of the formation of DNA cleavage product by MtTOP1 and MtTOP1-704t. O: Oligonucleotide substrate; P: Major cleavage product. (B) Intramolecular DNA religation by MtTOP1 and MtTOP1-704t. Lane 1: Oligonucleotide substrate with no enzyme; Lanes 2 and 5: DNA cleavage reaction. (C) DNA religation (R) with no Mg^{2+} present; Lanes 3 and 6:

DNA religation (R) with 5 mM MgCl$_2$ and 1 M NaCl added; Lanes 4 and 7: DNA religation (R) with 10 mM MgCl$_2$ and 1 M NaCl added.

Fig. 4. Catenation of single-stranded M13mp18 DNA by MtTOP1 and MtTOP1-704t. Single-stranded M13mp18 DNA was incubated with the indicated amount of MtTOP1 or MtTOP1-704t before proteinase K treatment. C: circular ssDNA; L: linear ssDNA; CC: catenated circular ssDNA.

Fig. 5. Structure of MtTOP1-704t. (A) A ribbon diagram of the structure of MtTOP1-704t. The main bodies of the N-terminal D1, D2, D3, and D4 domains and the C-terminal D5 domain are colored in salmon, yellow, cyan, green, and pink, respectively. Numbering of the secondary structures are based on each individual domain. The protrusion from D1 that is regarded as a part of D4, as described in the EcTOP1 structure, is colored in grey. In D2, the two loops that include unique insertions are colored in magenta (the loop between β2 and α1) and red (the loop between α2 and β3). In D3, the α4 helix that degenerated to a β-turn in the EcTOP1 structures is highlighted in blue. The catalytic residue Y342 is drawn in stick format. In D4, the unusual insertion in the middle of the α8 helix is highlighted in purple. (B) and (C) The molecular surface representations of MtTOP1-704t. The orientation of the molecule in (B) is the same as that in (A). The orientation of the molecule in (C) has a 180° rotation around the vertical axis with respect to the view in (B). (D) and (E) The electrostatic potential surface representations of MtTOP1-704t. The surface of MtTOP1-704t is colored according to amino acid residue charge, ranging from blue (positive) to red

(negative). The orientation of the molecule in (D) is similar to that in (B), but with a -15°

rotation around the horizontal axis with respect to the view in (B) for a better view of the

DNA binding groove.   The orientation of the molecule in (E) is the same as that in (C).

Figures 5, 7, and 8 are prepared with the program PyMOL (http://www.PyMOL.org).

Fig. 6. N-terminal domain arrangement of MtTOP1 and multiple sequence

alignment.   (A) The N-terminal domain arrangement of MtTOP1. Domains D1, D2, D3,

and D4 are colored in salmon, yellow, cyan, and green, respectively. The protruded

region from D1 in MtTOP1 and EcTOP1, is regarded as a part of D4. In EcTOP3, it is

described as a part of D1.[22] The asterisk indicates the location of the catalytic tyrosine

residue. (B) Structure-based Multiple Sequence Alignment with EcTOP1 and EcTOP3.

The residues that are identical between MtTOP1 and EcTOP1 are colored in red. The

residues of EcTOP3 that are conserved in both MtTOP1 and EcTOP1 are also colored

in red. The residues in regions of similar secondary structures but different

conformations between MtTOP1 and EcTOP1 are colored in green. The residues in

regions with neither sequential nor structural alignments are colored in blue. The

disordered residues in the structures of EcTOP1 are shown in lower case. The

secondary structures of MtTOP1, α-helices as coils and β-strands as arrows, are drawn

above the appropriate sequence of MtTOP1. The numbering of these secondary

structures is based on each domain for descriptive convenience. The same color

schemes for domains used in (A) are used for highlighting their corresponding

sequences of MtTOP1 and EcTOP1 in the structural alignment.

Fig. 7. N-terminal region structural alignments of MtTOP1 with EcTOP1 and EcTOP3. (A) Ribbon diagrams of the superposition of MtTOP1 with EcTOP1 in a front view and in a side view. The color scheme for the MtTOP1 domains and featured structural parts is the same as in Fig. 5. EcTOP1 is colored in grey. (B) Ribbon diagrams of the superposition of MtTOP1 with EcTOP3 in a front view and in a side view. EcTOP3 is colored in grey, and its decatenation loop is labeled.

Fig. 8. **Structure of the first MtTOP1 C-terminal domain (D5).** (A) A ribbon diagram of D5. The residues (G660, G663, and Y665) that are part of the highly conserved sequence motif GxxGPY in each MtTOP1 C-terminal domain are highlighted in magenta. Several residues (including Y665) that are potentially involved in DNA binding are drawn in stick format. (B) A stereo view of the β-sheet of D5 and several exposed residues on the β-sheet mentioned in (A). The electron density drawn in grey mesh is calculated from a weighted 2Fo-Fc map and contoured at the 1σ level.

β1   β2   β3

D5 642 NSIKLFDDTHGRPIYVRVGKNGPYLERLVAGDTGE----------PTPQRANLSDSITPDELTLQVAEELFATP 705

D6 706 GLDPETGHEIVAREGRFGPYVTEILPEPAADAAAAQGVKKRQKAAGPKPRTGSLLRSMDLQTVTLEDALRLLSLP 787

D7 788 RVVGVDPASGEEITAQNGRYGPYLKR------------------GNDSRSLVTEDQIFTITLDEALKIYAEP 841

D8 842 KRRGRQSASAPPLRELGTDPASGKPMVIKDGRFGPYVTD--------------GETNASLRKGDDVASITDERAAELLADR 908

β4   α1

Positively Charged Tail   909 RARGPAKRPARKAARKPAKKAAKRD 934

Figure 1

Figure 2

A
DNA Cleavage

MtTOP1   MtTOP1-704t

0 .03 .06 .12 .25 .5 1.0 0 .03 .06 .12 .25 .5 1 pmole

← O

← P

B

No
Enz   MtTOP1   MtTOP1-704t

        C   R   R   C   R   R
        0   5   10   0   5   10   mM Mg

← O

← P

1   2   3   4   5   6   7
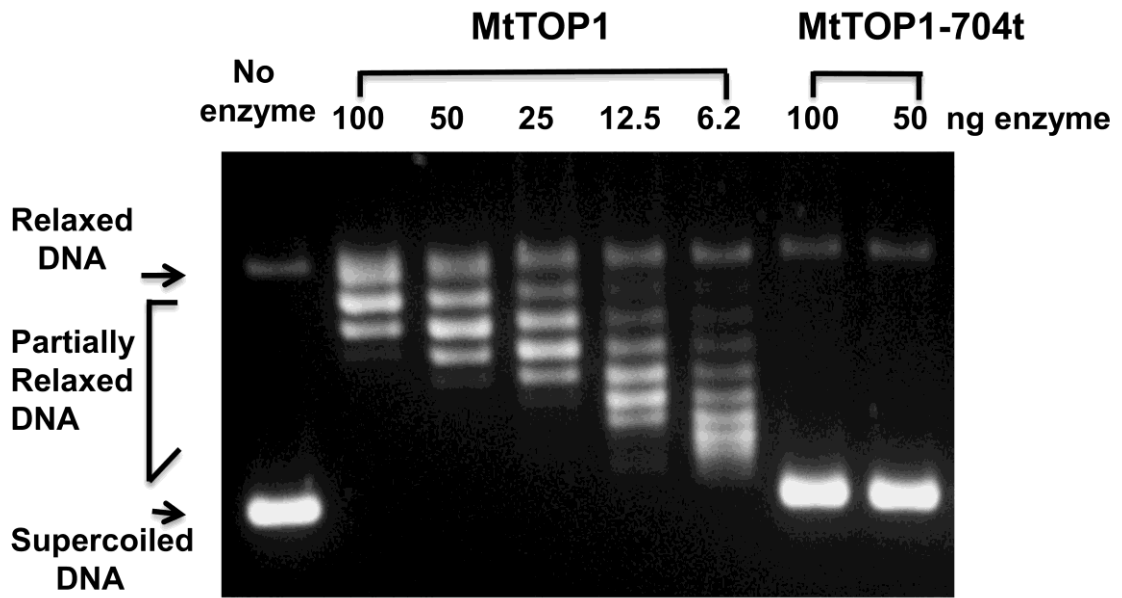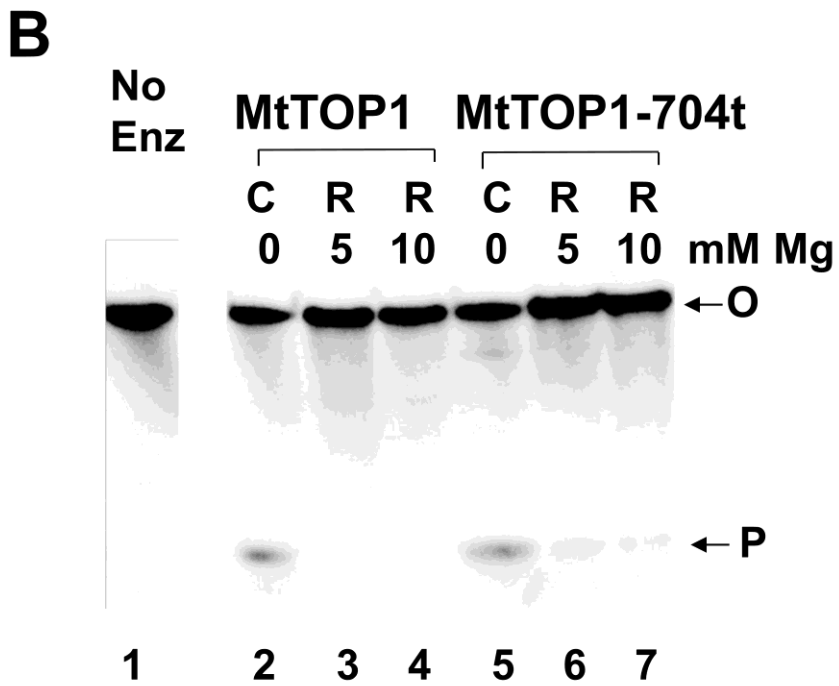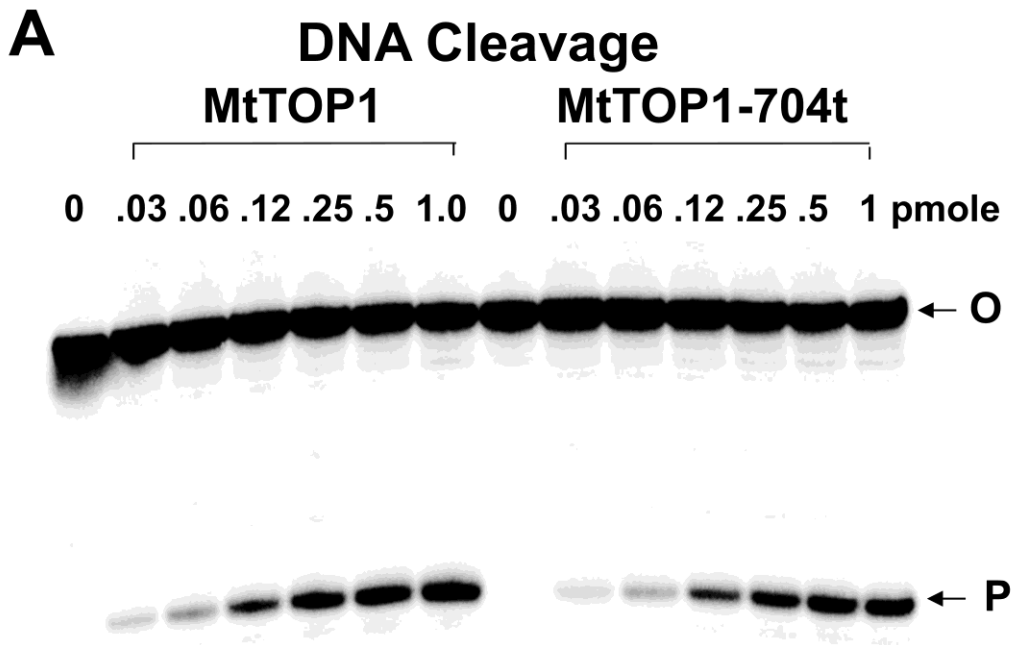
Figure 3

Figure 4
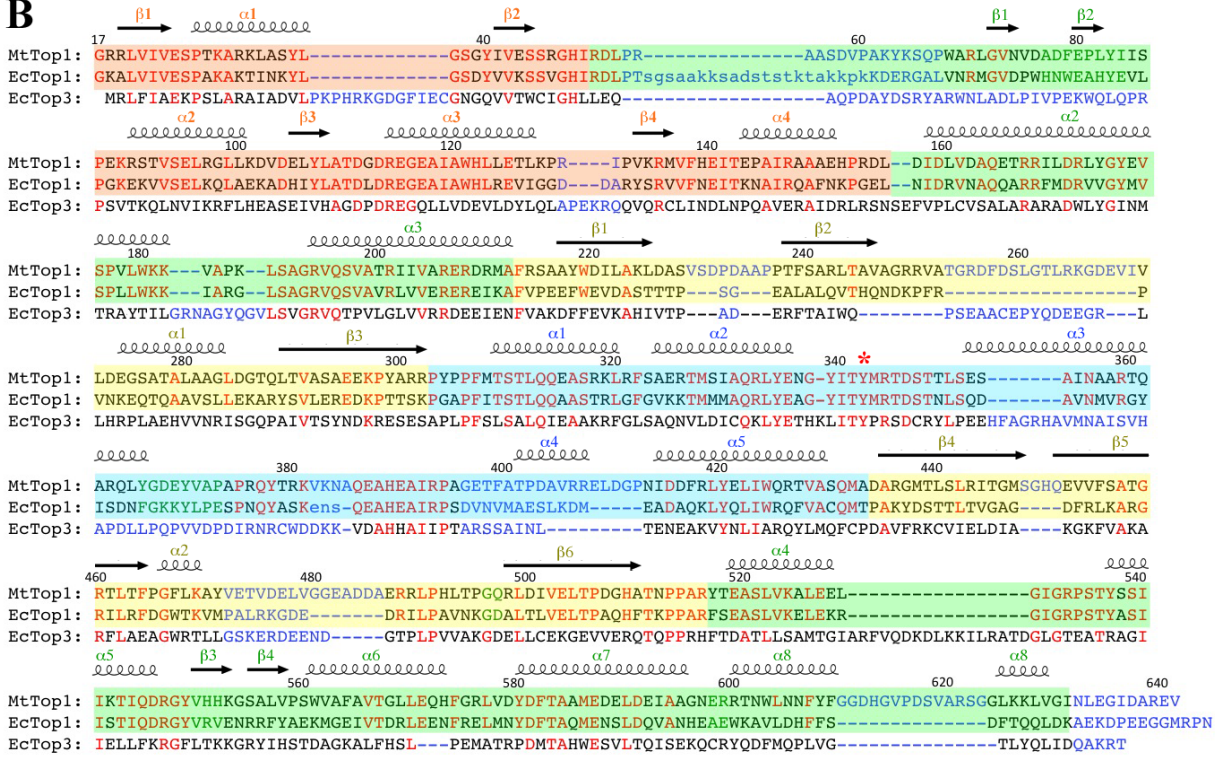
Figure 5

Figure 6

protruding helix
of EcTOP1

decatenation
loop of EcTOP3
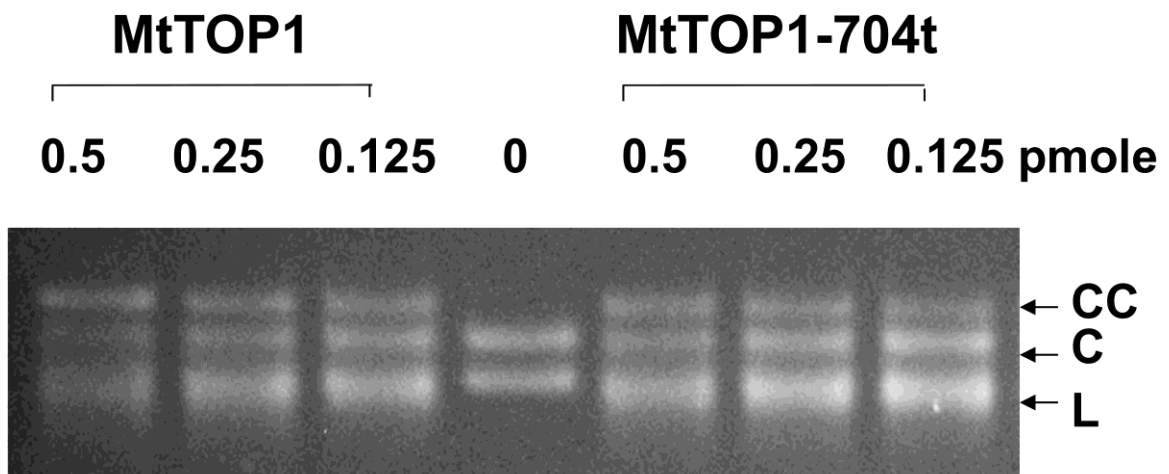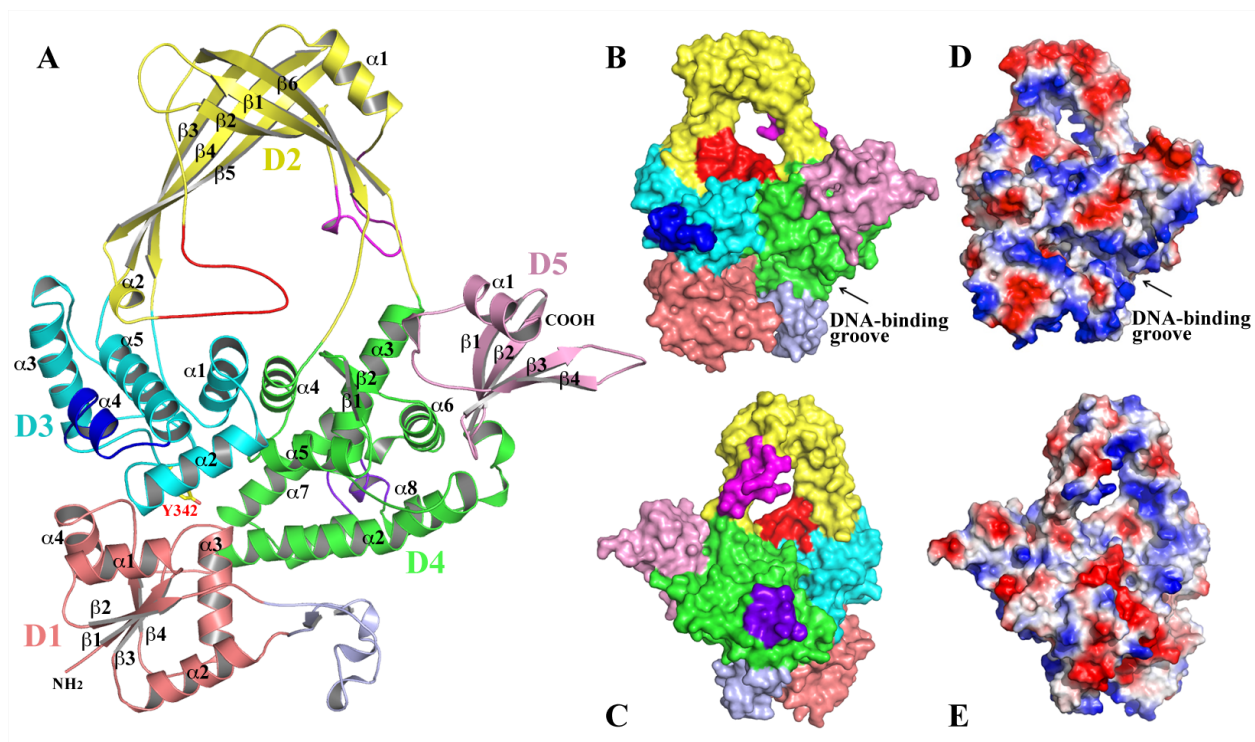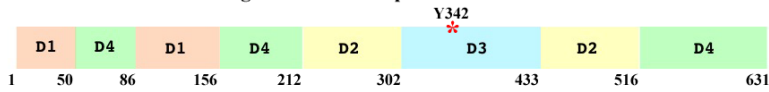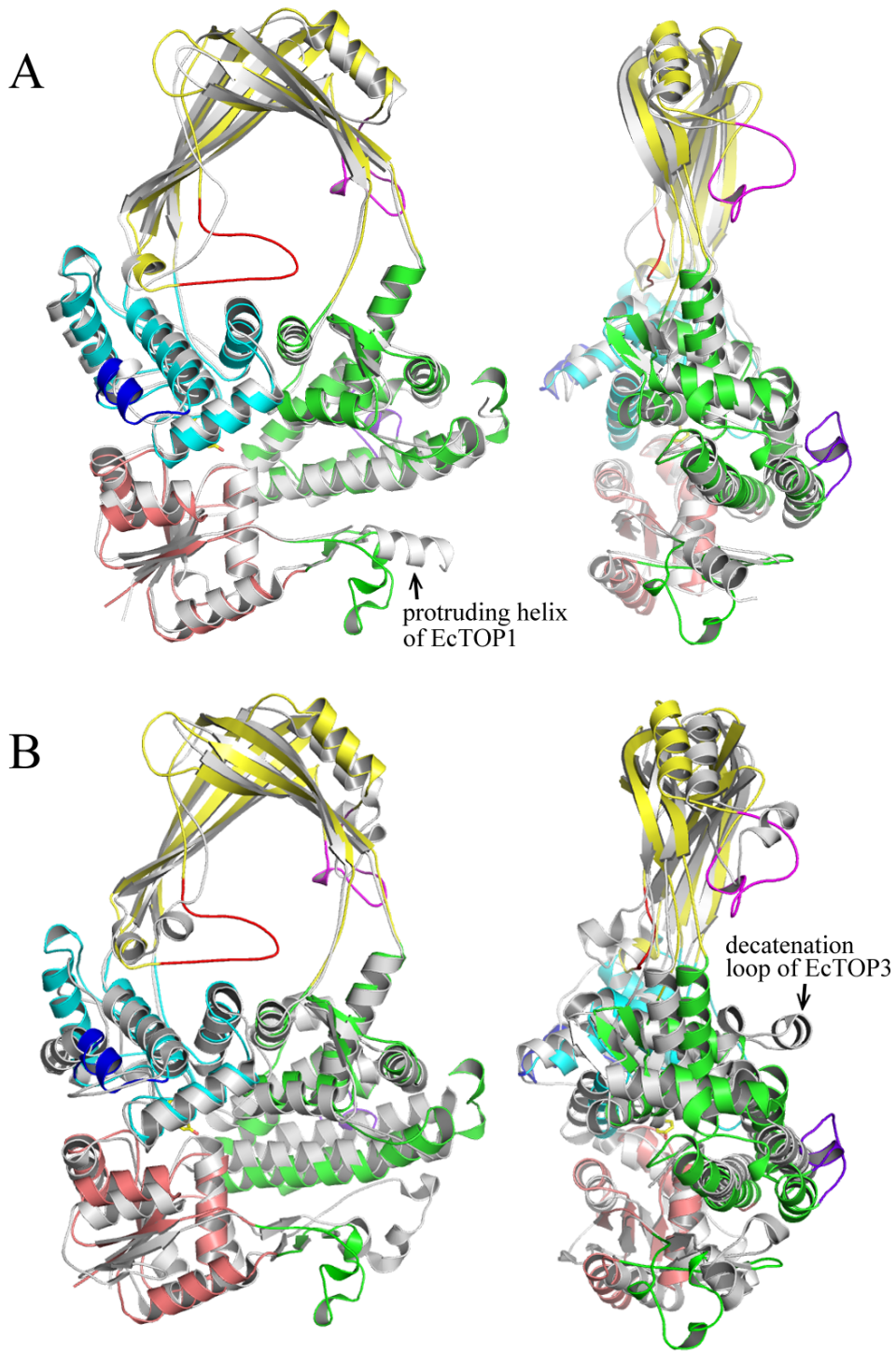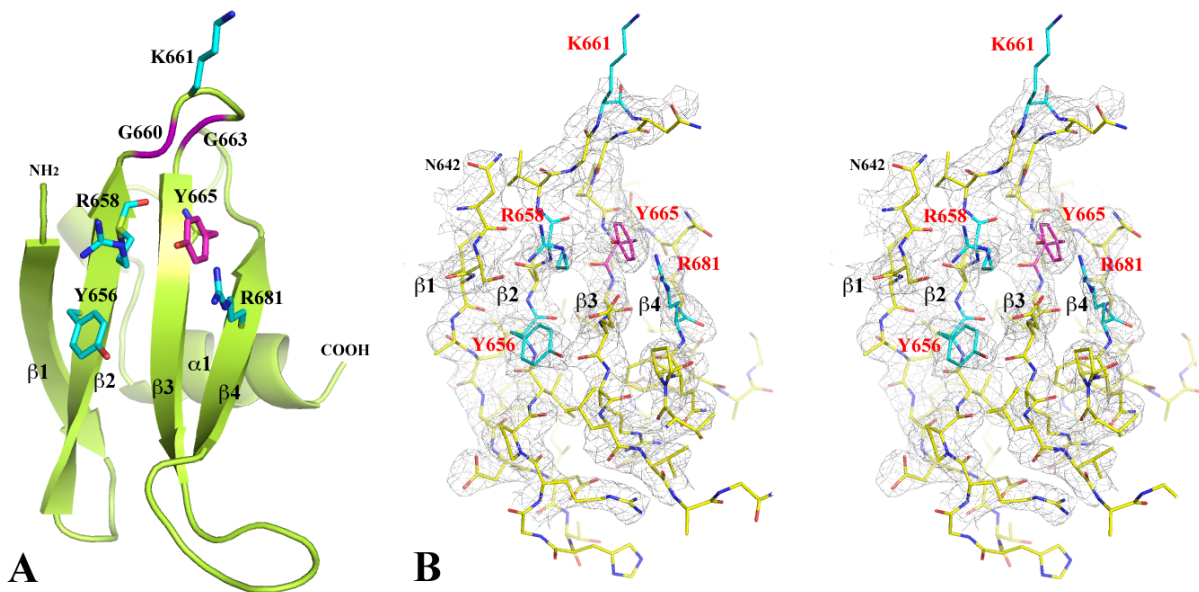
Figure 7

Figure 8

Table 1: Data collection and refinement statistics

Data collection	MtTOP1
Space group	$P2_12_12_1$
Unit Cell dimensions	
a, b, c (Å)	64.00,93.15, 140.2
α, β, γ (°)	90, 90, 90
MW Da (residue)	77,468.3 (703) [1]
Mol (AU)	1
Wavelength (Å)	0.9792
Resolution (Å)	42.0-2.52
Number of Unique reflections	28,858
Completeness (%)	99.2 (97.6) [2]
Redundancy	5.6 (5.6) [2]
R_{merge}	0.098 (0.696) [2]
$I/\sigma(I)$	14.0 (2.0) [2]
Solvent content (%)	54.3
Phasing	
Resolution (Å)	42.0-3.00
Correlation coefficient [3]	0.416
Refinement	
Resolution	42.0-2.52
No. reflections (work/test)	26,111/1,380
R_{work}/R_{free}	0.174/0.241
No.of atoms	
Protein	5334
Water/Others	72/47
B-factors (Å^2)	
Protein	55.0
Water/Others	40.7/77.47
R.m.s deviation	
Bond length (Å)	0.008
Bond angle (°)	1.083
PDB code	5D5H

[1] Not including three N-terminal vector-derived residues, SNA; [2] (Last resolution bin, 2.52-2.56

Å); [3] Molecular replacement
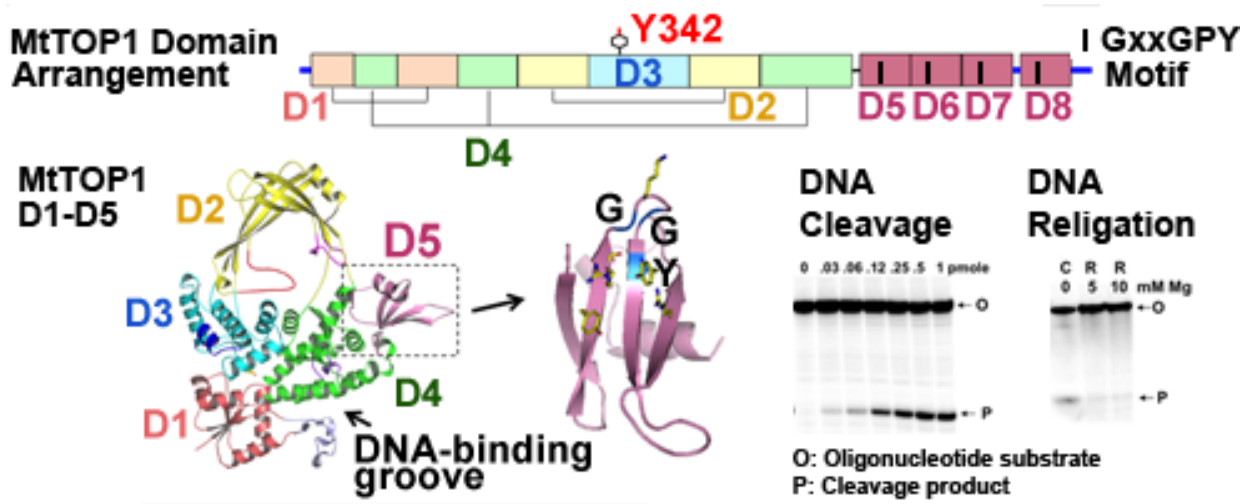
Graphical abstract

Highlights

- A long-awaited crystal structure of *M. tuberculosis* Topoisomerase I (MtTOP1)

- Recognition of individual MtTOP1 C-terminal domains and their organization

- Rational design of truncated MtTOP1 enables production of diffraction quality crystals

- Charged loops in the toroid hole are unique features of MtTOP1 N-terminal domains

- Novel protein fold of MtTOP1 C-terminal domains defines a subset of topoisomerases