11-22-2016

Identification of proximal sites for unwound DNA substrate in Escherichia coli topoisomerase I with oxidative crosslinking.

Bokun Cheng
New York Medical College

Qingxuan Zhou
Biomolecular Sciences Institute and Department of Chemistry and Biochemistry, Florida International University, qinzhou@fiu.edu

Liwei Weng
Johns Hopkins University

John D. Leszyk
University of Massachusetts Medical School

Marc M. Greenberg
Johns Hopkins University

See next page for additional authors

Recommended Citation
Cheng, B., Zhou, Q., Weng, L., Leszyk, J. D., Greenberg, M. M. and Tse-Dinh, Y.-C. (2017), Identification of proximal sites for unwound DNA substrate in Escherichia coli topoisomerase I with oxidative crosslinking. FEBS Lett, 591: 28–38. doi:10.1002/1873-3468.12517

Received Date : 01-Oct-2016

Revised Date : 21-Nov-2016

Accepted Date : 22-Nov-2016

Article type : Research Letter

Identification of proximal sites for unwound DNA substrate in *E. coli* topoisomerase I with oxidative crosslinking

Bokun Cheng[1], Qingxuan Zhou[2,3], Liwei Weng[4], John D. Leszyk[5], Marc M. Greenberg[4], Yuk-Ching Tse-Dinh[2,3*]

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

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

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

[4] Department of Chemistry, Johns Hopkins University, Baltimore, Maryland 21218.

[5] Department of Biochemistry and Molecular Pharmacology and Proteomics and Mass Spectrometry Facility, University of Massachusetts Medical School, Worcester, MA 01605,

[*]Corresponding author:

Yuk-Ching Tse-Dinh

Tel: 1-305-348, 4956; Fax: 1-305-348-3772

email:ytsedinh@fiu.edu

Abstract

Topoisomerases catalyze changes in DNA topology by directing the movement of DNA strands through consecutive cleavage-rejoining reactions of the DNA backbone. We describe the use of a phenylselenyl-modified thymidine incorporated into a specific position of a partially unwound DNA substrate in crosslinking studies of *Escherichia coli* topoisomerase I to gain new insights into its catalytic mechanism. Crosslinking of the phenylselenyl-modified thymidine to the topoisomerase protein was achieved by the addition of a mild oxidant. Following nuclease and trypsin digestion, lysine residues on topoisomerase I crosslinked to the modified thymidine were identified by mass spectrometry. The crosslinked sites may correspond to proximal sites for the unwound DNA strand as it interacts with enzyme in the different stages of the catalytic cycle.

Keywords:
topoisomerase, DNA-protein interactions, crosslinking

Introduction

Topoisomerase functions are critical for vital genomic processes including replication, transcription, recombination and repair [1-3]. Topoisomerases catalyze the resolution of topological barriers that arise during these processes. The enzymes catalyze the movement of DNA strands via the cleavage and rejoining of DNA phosphodiester backbone. The opposing activities of bacterial topoisomerase I and DNA gyrase play important roles in the homeostatic regulation of DNA supercoiling [4-9]. *Escherichia coli* topoisomerase I (EcTopA) encoded by the *topA* gene interacts with two single stranded segments in underwound duplex DNA to catalyze negative supercoiling removal [10]. The process involves transport of a single strand segment of DNA (T-DNA) through the "gate" created by the topoisomerase cleavage of the opposite strand of DNA (G-DNA) (Scheme 1) [11-13]. EcTopA recognition of underwound duplex DNA is important for the relief of transcription-driven negative supercoiling during transcription elongation [14-17]. The action of EcTopA prevents hypernegative supercoiling of DNA and accumulation of R-loops. The interactions between the enzyme and single-stranded regions of the underwound duplex DNA and the resulting movement of T-DNA strand are

critical for the overall catalytic mechanism [18,19]. A crystal structure of a single strand segment of DNA covalently linked to Tyr319 at the active site formed by the N-terminal domains of EcTopA provides a view of the break formed on the G-strand of DNA as a catalytic intermediate [20,21]. A more recent crystal structure of full length EcTopA showed the interaction of the T-strand segment of single-stranded DNA interacting with the C-terminal domains of EcTopA [10,22]. The proposed catalytic mechanism for EcTopA requires movement of the T-strand DNA into the proximity of the active site opening in order for passage of the T-strand through the DNA break on the G-strand (Step 3 of Scheme 1). The C-terminal domains of EcTopA contain flexible linkers [10] that could facilitate significant conformational change of the complex with both G-strand and T-strand single-stranded DNA segments bound to EcTopA. However, there is no structural or biochemical data that support the proposed movement of the T-strand into the proximity of the G-strand DNA opening at the active site. Mapping of the sites on topoisomerase I that are proximal to a specific nucleotide present on the T-strand of the partially unwound duplex DNA may provide such evidence.

DNA containing photoactivatable nucleotides have been used previously for crosslinking to proximal residues in interacting proteins [23-31]. Diazirine, halide or thio-substituted nucleotides and phenyl-azide modified phosphates have been used to probe DNA interactions by proteins involved in replication, transcription and recombination [23,28,29,32-34]. Despite the existence of multiple molecular tools, identifying specific DNA-protein interaction sites using these approaches is challenging. Often a laser [23,35] or custom built UV apparatus is required to obtain a high yield of the crosslinked complex [27]. In addition, the large number of negative charges on the cross-linked DNA phosphodiester backbone make it difficult to identify the crosslinked peptides by mass spectrometry [25].

DNA-protein crosslinks are produced from a phenyl selenide derivative of thymidine (PhSeT) without the requirement of a laser or custom built UV apparatus (Figure 1A). The reactive entity can be generated by reaction with mild oxidants such as sodium periodate, or photochemically using routinely available photon sources [36,37]. Another useful feature of PhSeT is that it can be introduced into DNA via standard solid phase synthesis methods or via the respective nucleotide triphosphate, which is accepted as a substrate by DNA polymerase. The phenyl selenide has been used to probe interactions between histone tails and nucleosomal DNA

[36]. The electrophilic species produced reacts with the side chains of cysteine, histidine, and lysine amino acids. The interaction sites were inferred by analysis of crosslinking products, their yields, and the rate constants for their formation from histone variants with mutations at amino acid residues that are potential targets for crosslinking due to the technical challenge for direct identification of the cross-linked DNA-peptide adducts [36]. We report here the incorporation of PhSeT at a specific position on the T-strand of a partial duplex oligonucleotide substrate for crosslinking to *E. coli* DNA topoisomerase I (EcTopA) and direct identification of the crosslinking sites by LC/MS/MS as a new approach to probe the catalytic mechanism of this type IA topoisomerase.

Materials and Methods

Oligonucleotide synthesis

The oligonucleotide Substrate-Y with two segments of single-stranded DNA linked to a double stranded stem as model for partially unwound DNA substrate for EcTopA is shown in Figure 1. EcTopA cleaves the substrate at a unique site 4 nucleotides downstream from a dC indicated by the arrow on the G-strand (bottom strand) to form the covalent intermediate. The bottom strand of Substrate-Y containing the PhSeT (**X**) at the 5'-end of the T-strand of DNA was synthesized using previously reported methods [38], followed by hybridization to the top strand (T-strand) to form Substrate-Y. Oxidative crosslinking with Substrate-Y alone showed very low degree of inter-strand crosslinking (Figure S1) as X is positioned in the single-stranded region of the substrate.

Assay of cleavage of DNA substrate by EcTopA

The G-strand or T-strand oligonucleotide of Substrate-Y was labeled with ^{32}P at the 5'-end with γ-^{32}P-ATP and T4 polynucleotide kinase before hybridization with the unlabeled opposite strand. Substrate-Y formed by hybridization was then incubated with EcTopA in 10 µl of 20 mM sodium phosphate, pH 7.2, 1 mM EDTA with 10 or 100 mM NaCl at room temperature for 30 min. The reactions were stopped by addition of an equal volume of sequencing gel loading buffer and electrophoresed in a 15% denaturing sequencing gel, followed by Phosphorimager analysis.

Crosslinking Reaction

To demonstrate initially that Substrate-Y labeled with ^{32}P at the 5'-end of the T-strand can be crosslinked to EcTopA protein via the PhSeT, the crosslinking reaction was carried out under conditions similar to those used in a previous study [36] in buffer (20 mM sodium phosphate, pH 7.2, 1 mM EDTA, 10 or 100 mM NaCl). The crosslinking reaction was initiated by the addition of 5 mM $NaIO_4$ and allowed to proceed in the dark at room temperature for up to 2 hours before quenching the reaction by the addition of 50 mM Na_2SO_3. After 5 minutes, equal volume of loading buffer for SDS PAGE was added, and the reaction products were analyzed by electrophoresis in 10% SDS gel, followed by Phosphorimager analysis of the dried gel.

Processing of topoisomerase-DNA crosslinked complex for mass spectrometry analysis

Following the crosslinking reaction between 60 pmol each of unlabeled Substrate-Y and EcTopA, the EcTopA-DNA adduct was enriched by applying the reaction mixture to a DE52 cartridge (with diethylaminoethyl resin, from GE Healthcare) equilibrated with the crosslinking buffer and washed with buffer containing 100 mM NaCl. The fractions enriched in EcTopA-DNA adduct were eluted with buffer containing 250 mM NaCl. The eluted fractions were concentrated and exchanged by ultrafiltration with Amicon Ultra (30,000 kDa Molecular weight cut off) into buffer of 20 mM sodium phosphate, pH 7.4, 1 mM EDTA, 100 mM NaCl, 0.5 mM $MgCl_2$. Following digestion with 10,000 U of Micrococcal nuclease (New England BioLabs) for 1 h at 37°C, the digest was loaded onto a SDS 5-10% gradient polyacrylamide gel, followed by Coomassie blue staining. The Coomassie blue stained gel band containing the mixture of the nuclease digested EcTopA adduct and unreacted EcTopA protein co-eluted in the DE52 fractions was excised and submitted to the Proteomics and Mass Spectrometry Facility at the University of Massachusetts Medical School for tryptic digest and LC/MS/MS analysis.

In Gel Digestion

Gel slices were cut into 1x1 mm pieces and placed in 1.5 ml Eppendorf tubes with 1 ml of water for 30 minutes. The water was removed and 50 µl of 250 mM ammonium bicarbonate was added. For reduction 5 ul of a 45 mM solution of 1, 4 dithiothreitol (DTT) was added and the samples were incubated at 50°C for 30 minutes. The samples were cooled to room

temperature and then 5 µl of a 100 mM iodoacetamide solution (to effect alkylation) was added and allowed to react for 30 minutes. The gel slices were washed 2 X with 1 ml water aliquots. The water was removed and 1ml of 50:50 (50 mM ammonium bicarbonate: acetonitrile) was placed in each tube and samples were incubated at room temperature for 1 hour. The solution was then removed and 200 µl of acetonitrile was added to each tube at which point the gels slices turned opaque white. The acetonitrile was removed and gel slices were further dried in a Speed Vac. Gel slices were rehydrated in 50 µl of 2 ng/µl trypsin (Sigma) in 0.01% ProteaseMAX Surfactant (Promega), 50 mM ammonium bicarbonate. Additional bicarbonate buffer was added to ensure complete submersion of the gel slices. Samples were incubated at 37^{o}C for 21 hours. The supernatant of each sample was then removed and placed in a separate 1.5 ml Eppendorf tube. Gel slices were further dehydrated with 100 µl of 80:20 (Acetonitrile: 1% formic acid). The extract was combined with the supernatants of each sample. The samples were then dried down in a Speed Vac. Samples were dissolved in 25 µl of 5% Acetonitrile in 0.1% trifluoroacetic acid prior to injection for LC/MS/MS analysis.

LC/MS/MS on Q Exactive

A 3.0 µl aliquot was directly injected onto a custom packed 2 cm x 100 µm C18 Magic 5 µm particle trap column. Peptides were then eluted and sprayed from a custom packed emitter (75 µm x 25 cm C18 Magic 3 µm particle) with a linear gradient from 95% solvent A (0.1% formic acid in water) to 35% solvent B (0.1% formic acid in Acetonitrile) in 40 minutes at a flow rate of 300 nanoliters per minute on a Waters Nano Acquity UPLC system. Data dependent acquisitions were performed on a Q Exactive mass spectrometer (Thermo Scientific) according to an experiment where full MS scans from 300-1750 m/z were acquired at a resolution of 70,000 followed by 12 MS/MS scans acquired under HCD fragmentation at a resolution of 35,000 with an isolation width of 1.2 Da. Raw data files were processed with Proteome Discoverer (version 1.4) prior to searching with Mascot Server (version 2.5) against the NCBInr database. Search parameters utilized were fully tryptic with 2 missed cleavages, parent mass tolerances of 10 ppm and fragment mass tolerances of 0.05 Da. A fixed modification of carbamidomethyl cysteine and variable modifications of acetyl (protein N-term), pyroglutamic for N-term glutamine, oxidation of methionine, and crosslinking of cysteine and lysine with XPhospho, XGPhospho, XGTPhospho (with X being the PhSeT) on

cysteine and lysine were considered. Search results were loaded into the Scaffold Viewer (Proteome Software, Inc.) for peptide validation and quantitation. Data was later analyzed again with XGPhospho on histidines being added as a potential modification.

Results

Cleavage of Substrate-Y by EcTopA at a single position on the G-strand of Substrate-Y

EcTopA has a binding pocket in domain D4 for a cytosine base located at a position on the single-stranded G-strand DNA segment that is four nucleotides upstream of the phosphodiester bond being cleaved by the active site tyrosine, accounting for the sequence selectivity observed for its DNA cleavage sites [21,39]. This cleavage target selectivity is shared by all bacterial topoisomerase I and reverse gyrase enzymes [40]. EcTopA cleaves both single-stranded T-strand (Figure 2, lanes 2,3) or G-strand (lanes 8,9) labeled with ^{32}P at the 5'-end in the absence of the complementary strand to produce faster migrating products. When labeled G-strand is hybridized to T-strand to form Substrate-Y, the preferred cleavage site (arrow, Figure 1) remains in the single-stranded region of G-strand to be recognized by EcTopA, resulting in the same cleavage product (lanes 11,12). The fraction of cleavage product quantitated by densitometry analysis for Substrate-Y in lanes 11 and 12 (11% and 6% respectively) is similar to that observed previously for a 39-base long oligonucleotide substrate used to study EcTopA residues involved in G-strand binding [41]. In previous studies, oxidative crosslinking reactions with PhSeT had included 100 mM NaCl [42]. The cleavage efficiency Substrate-Y by EcTopA was reduced at 100 mM NaCl (lane 12) compared to reactions with 10 mM NaCl (lane 11). When the T-strand is hybridized to the G-strand in Substrate-Y a cytosine base is not present in the single-stranded region to align a phosphodiester bond in the active site for cleavage by EcTopA [39]. Hence, little or no cleavage product was observed in lanes 5 and 6. The absence of cleavage product in lanes 5 and 6 for the labeled T-strand when hybridized to unlabeled G-strand provides support for the formation of the partially double-stranded Substrate-Y structure as postulated in Figure 1.

Oxidative crosslinking of Substrate-Y to EcTopA

When the PhSeT incorporated at the 5'-end of ^{32}P-labeled T-strand of the Substrate-Y is activated by $NaIO_4$ following binding to EcTopA, the crosslinked complex with EcTopA is

observed as a high molecular weight ^{32}P-labeled species following electrophoresis of the reaction mixture in a 10% SDS gel and analysis of the dried gel by Phosphorimager. The unreacted labeled DNA substrate co-migrates with the dye front. The yield of the crosslinked complex increased when the reaction time was extended from 5 min to 2 h (Figure 3A). The yield of crosslinked complex also increased when a higher ratio of Substrate-Y to EcTopA was used (Figure 3B). As much as 41% of input Substrate-Y could be crosslinked with EcTopA (Lane 7, Figure 3B). The appearance of the crosslinked complex as a doublet could be due to crosslinking of a second molecule of Substrate-Y to EcTopA with a molecule of Substrate-Y already bound. The yield of crosslinked complex was about the same with either 10 or 100 mM NaCl. However, since cleavage efficiency was higher in 10 mM NaCl, subsequent crosslinking reactions were carried out in reaction buffer with 10 mM NaCl.

Processing of oligonucleotide in cross-linked complex by Micrococcal nuclease digestion

The crosslinking reaction was scaled up with unlabeled substrate-Y at 1:1 EcTopA:Substrate-Y ratio. The cross-linked complex was observed as a band above the position of EcTopA in a SDS PAGE gel using Coomassie blue staining (Figure 4A, lane 2). The crosslinked complex formed between EcTopA and Substrate-Y was enriched over unreacted EcTopA using a DE52 cartridge. The desired product(s) eluted with high salt (lanes 2, 4 in Figure 4B) and were extensively digested with Micrococcal nuclease to remove most of the DNA from the complex so that mass spectrometry analysis could be carried out for identification of the crosslinked EcTopA residue. The DNA component in the substrate-Y:EcTopA adduct was effectively trimmed by the nuclease, as evidenced by the co-migration of the nuclease treated EcTopA adduct with the unreacted EcTopA in the SDS gel, following 2 h of nuclease treatment (lanes 3, 5 Figure 4B). Additional higher molecular weight complexes were observed under these reaction conditions. The band in lane 5 of Figure 4B (nuclease digestion product of lane 4 species) was selected for further analysis as lane 4 had the higher ratio of crosslinked SubstrateY-EcTopA complex:unreacted EcTopA protein when compared to lane 2.

Identification of lysines in EcTopA crossslinked to PhSeT

Tryptic digest and LC/MS/MS analysis of Micrococcal nuclease treated Substrate Y-EcTopA complex mixed with unreacted EcTopA identified peptides crosslinked to XGphospho (X being PhSeT) with the added molecular mass of 649 (Table 1). The extensive digestion with

Micrococcal nuclease processed the covalently bound DNA to a dinucleotide unit. Analysis of the spectral data from the fragments (Figure S2) are consistent with the dinucleotide crosslinked to lysines at EcTopA positions 19, 28, 279, 303, and 679 found in domains D1 (K19, K28), D2 (K279), D3 (K303) and D6 (K679) as shown in Figure 5A. Although the spectral data could not rule out that in the last peptide H681 may be the crosslinking site instead of K679, no other potential crosslinking site to histidines was observed. Crosslinking adducts from reactions of cysteines also were not detected in the spectral data. Consideration of the crosslinked lysines reveals that K279, located in N-terminal domain D2 of EcTopA, is proximal to a nucleotide binding site identified previously (PDB 1CY1) in the central hole of the complex between the EcTopA 67 kDa N-terminal fragment and trinucleotides pTTT [16] (Figure 5B). K679 is shown in a close-up view (Figure 5C) of the C-terminal domains D5 and D6 in the structure of full length EcTopA (PDB 4RUL) in complex with a single-stranded oligonucleotide [10]. Other crosslinked lysines, K28 in N-terminal domain D1 and K303 in N-terminal domain D3, are positioned on opposite sides of the active site in the structure of the EcTopA 67KDa in covalent complex (PDB 3PX7) with a cleaved single-stranded oligonucleotide (Figure 5D), in which the 5'-end of the cleaved G-strand DNA is linked by a phosphotyrosine to the active site nucleophile Y319 in N-terminal domain D3 [20].

Discussion

Substrate Y was designed to model the partially unwound duplex DNA in a negatively supercoiled DNA substrate of EcTopA. Cleavage by EcTopA requires the presence of the single-stranded DNA region [19] and the preferred cleavage sites have a cytosine 4 nucleotides upstream of the cleaved phosphodiester linkage [21]. Consequently, only one of the oligonucleotide strands in Substrate-Y can be utilized as the G-strand for DNA cleavage (Figure 1). The goal of this study was to provide insight concerning the interactions between EcTopA and partially unwound duplex DNA beyond what can be determined from structural studies on complexes formed between EcTopA and small single-stranded oligonucleotides. The reactive thymidine analogue, PhSeT, was placed nine nucleotides from the junction of the single-strand and double strand regions, at the end of the single-stranded region of the T-strand of Substrate-Y to probe potential interaction sites on EcTopA in solution. Crosslinking reactions of PhSeT

following oxidative activation likely involve reaction with nucleophilic lysine, histidine, or cysteine residues [36]. The molecule does not react with other residues, including tyrosine, serine, threonine, glutamine, or asparagine. Crosslinking to cysteine residues was not observed in this study. Out of the 14 cysteines in EcTopA, twelve of them are part of tetracysteine motifs engaged in coordination to Zn(II) for structural organization of the C-terminal domains [10,43]. Therefore, it is unlikely that these cysteines are available as crosslinking partners. There are 70 lysines and 12 histidines in EcTopA. The MS data unambiguously identified four lysines (K19, K28, K279, K303) as crosslinking sites for the reactive PhSeT incorporated into Substrate-Y. DNA linkage on the fifth peptide is likely to be at K679, although H681 could not be ruled out as a crosslinking site based on the spectral data. Alignment of different bacterial topoisomerase I amino acid sequences (Figure S3) shows that K19 is highly conserved among both gram-positive and gram-negative bacteria. Furthermore, only positively charged lysine or arginine residues are present at position 303. Lysine, arginine, or asparagine is found at position 279. K28, K679 and H681 are replaced by hydrophobic amino acids in some topoisomerases.

In the proposed catalytic mechanism of EcTopA (Scheme 1), the enzyme-DNA complex undergoes multiple transitions and conformational changes, likely involving different sets of protein-DNA interactions [44]. Solvent exposed lysine residues such as K19, K28, and K679 could be involved in initial interactions between the enzyme and single-stranded regions of DNA (step 1 in Scheme 1). We propose that some of the crosslinked lysines might also be involved in interactions with the designed partial duplex substrate that correspond to interactions in the later steps in the catalytic cycle. When the T-strand oligonucleotide with PhSeT was used as substrate for oxidative crosslinking to EcTopA in the single-stranded form, the total yield of crosslinked complex was much higher (Figure S4), but subsequent Micrococcal nuclease treatment and tryptic digest did not produce any peptide adduct in sufficient yield for identification by LC/MS/MS analysis. This is most likely because crosslinking of the single-stranded T-stand oligonucleotide occurred with a much larger number of sites. It is expected that a single-stranded oligonucleotide without a preferred cleavage site can slide after initial binding and occupy a very large number of different sites in the enzyme. In contrast, binding of the T-strand oligonucleotide as part of the partial duplex Substrate-Y with a single preferred cleavage site on the G-strand would be constrained to a more limited collection of complexes consisting of distinct protein-DNA interactions. This would restrict the possible crosslinking sites for the

reactive PhSeT present on the T-strand of the Substrate-Y to a smaller number of EcTopA residues. The selectivity of the oxidative crosslinking sites for studying other protein-DNA interactions is expected to be greater when there are more constraints on the interactions based on either base recognition or DNA structure requirement.

From the crystal structure of full length EcTopA with DNA bound to the C-terminal (PDB 4RUL), it was proposed that the T-strand of partial duplex DNA substrate is held by hydrophobic stacking interactions between the DNA bases and aromatic amino acids in the C-terminal domains of EcTopA [10]. This is consistent with an earlier solution structure of domains D8 and D9 determined by NMR [45]. Direct interactions between the bound DNA and lysines in the C-terminal domains were not observed in this crystal structure. However, if the single-stranded oligonucleotide segment present in the structure shown in Figure 5C extends further as part of a longer partial duplex DNA substrate, it might reach the vicinity of K679.

A number of nucleotide binding sites have been identified in crystal structures with mononucleotide or trinucleotides bound to the 67 kDa N-terminal domain of EcTopA [46]. One of the nucleotide binding sites proximal to K279 (Figure 5B) is found in the central hole and may represent an interaction between the passing T-strand of DNA and residues in the central hole after the passing T-strand has entered the hole (step 3 in Scheme 1).

The passing T-strand of DNA has to move through the break in the cleaved G-strand and must move to the vicinity of the active site at the junction of domains D1 and D3. It is expected that interdomain protein-protein interactions near the active site nucleophile Y319 must be disrupted to increase the distance between the 5'-phosphostyrosyl and 3'-hydroxyl ends of the cleaved G-strand to create an opening for the entrance to the central hole so that the T-strand of single-stranded DNA can pass through the break. K28 and K303 are located on opposite sides of the active site tyrosine (Figure 5A, 5D) and may interact with T-strand of single-stranded DNA as it moves towards the opening to the central hole between Steps 3 and 4 in Scheme 1. The 5'-phosphotyrosine linkage primarily holds the region of G-strand DNA downstream of the cleavage site (Figure 5D). K28 and K303 may also interact with this segment of the G-strand DNA immediately downstream from the scissile phosphate. The interaction between EcTopA and single-stranded G-strand DNA upstream of the cleavage site is well understood from biochemical and structural studies [20,39,47]. The structures of the non-covalent complex

(1MW8) and covalent complex (3PX7) show that the region of G-strand DNA upstream of the cleavage site is bound in a DNA binding groove [47] that contains conserved residues (in region from R168 to Q198). This region of the protein interacts with the DNA sugar or nucleobase components, including a cytosine 4 bases upstream of the cleavage site that accounts for the G-strand cleavage site selectivity [39].

In this study, we developed a method that identifies exact sites of crosslinking between protein lysine residues and a specific nucleotide on the DNA substrate using mass spectrometry. The method does not require prior purification of the individual peptides containing the crosslinked adduct. In addition, this crosslinking approach does not require any special UV irradiation equipment, but instead utilizes a commonly available reagent ($NaIO_4$) to initiate crosslinking. LC/MS/MS analysis of the crosslinked DNA was carried out after digestion to dinucleotides with a single nuclease [48]. We identified several sites on the *E. coli* topoisomerase I structure that could represent interaction sites proximal to the passing T-strand of single-stranded DNA in the different stages of the catalytic cycle. The roles of lysines on the solvent accessible surface of bacterial topoisomerase I have not been considered previously. The positively charged lysines may serve as sources for the initial recognition of complementary charged, single-stranded nucleic acid present as unwound DNA in negatively supercoiled DNA. Relaxation of the negatively supercoiled DNA would require movement of the T-strand to the vicinity of the G-strand opening. The interaction of the T-strand with nucleotide binding site found within the central hole was demonstrated directly by the identification of K279 as a crosslinking site in this study. The new result provides support for the proposed movement of the T-strand through the opening on the G-strand into the central hole as a required step of the catalytic cycle. The described crosslinking approach, followed by Micrococcal nuclease digestion of the crosslinked protein-DNA adduct can be adopted to investigate other protein-DNA interactions in solution.

Author Contributions

YT and MMG conceptually designed the overall research project. JDL contributed to the LC/MS/MS experimental design. BC, QZ, LW performed experiments. All authors contributed to the analysis of the data and writing of the manuscript.

Funding

This work was supported by National Institute of Health grants R01 GM054226 to YT, R01 GM054996 to MMG.

References

[1] 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, 827-41.

[2] Wang, J.C. (2002). Cellular roles of DNA topoisomerases: a molecular perspective. *Nat. Rev. Mol. Cell Biol.* 3, 430-40.

[3] Chen, S.H., Chan, N.L., and Hsieh, T.S. (2013). New mechanistic and functional insights into DNA topoisomerases. *Annu. Rev. Biochem.* 82, 139-70.

[4] Ferrandiz, M.J., Martin-Galiano, A.J., Arnanz, C., Camacho-Soguero, I., Tirado-Velez, J.M., and de la Campa, A.G. (2016). An increase in negative supercoiling in bacteria reveals topology-reacting gene clusters and a homeostatic response mediated by the DNA topoisomerase I gene. *Nucleic Acids Res.* 44, 1541-52.

[5] Blot, N., Mavathur, R., Geertz, M., Travers, A., and Muskhelishvili, G. (2006). Homeostatic regulation of supercoiling sensitivity coordinates transcription of the bacterial genome. *EMBO Rep.* 7, 710-5.

[6] Ahmed, W., Menon, S., D N B Karthik, P.V., and Nagaraja, V. (2016). Autoregulation of topoisomerase I expression by supercoiling sensitive transcription. *Nucleic Acids Res.* 44, 1549-52.

[7] Menzel, R. and Gellert, M. (1983). Regulation of the genes for *E. coli* DNA gyrase: homeostatic control of DNA supercoiling. *Cell* 34, 105-13.

[8] Goldstein, E. and Drlica, K. (1984). Regulation of bacterial DNA supercoiling: plasmid linking numbers vary with growth temperature. *Proc. Natl. Acad. Sci. U. S. A.* 81, 4046-50.

[9] Tse-Dinh, Y.C. (1985). Regulation of the *Escherichia coli* DNA topoisomerase I gene by DNA supercoiling. *Nucleic Acids Res.* 13, 4751-63.

[10] Tan, K., Zhou, Q., Cheng, B., Zhang, Z., Joachimiak, A., and Tse-Dinh, Y.C. (2015). Structural basis for suppression of hypernegative DNA supercoiling by *E. coli* topoisomerase I. *Nucleic Acids Res.* 43, 11031-46.

[11] 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, 138-46.

[12] Terekhova, K., Marko, J.F., and Mondragon, A. (2013). Studies of bacterial topoisomerases I and III at the single-molecule level. *Biochem. Soc. Trans.* 41, 571-5.

[13] 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, 10432-40.

[14] Drolet, M., Broccoli, S., Rallu, F., Hraiky, C., Fortin, C., Masse, E., and Baaklini, I. (2003). The problem of hypernegative supercoiling and R-loop formation in transcription. *Front. Biosci.* 8, d210-21.

[15] Liu, L.F. and Wang, J.C. (1987). Supercoiling of the DNA template during transcription. *Proc. Natl. Acad. Sci. U. S. A.* 84, 7024-7.

[16] Masse, E., Phoenix, P., and Drolet, M. (1997). DNA topoisomerases regulate R-loop formation during transcription of the *rrnB* operon in *Escherichia coli. J. Biol. Chem.* 272, 12816-23.

[17] Drolet, M. (2006). Growth inhibition mediated by excess negative supercoiling: the interplay between transcription elongation, R-loop formation and DNA topology. *Mol. Microbiol.* 59, 723-30.

[18] Schoeffler, A.J. and Berger, J.M. (2008). DNA topoisomerases: harnessing and constraining energy to govern chromosome topology. *Q. Rev. Biophys.* 41, 41-101.

[19] Depew, R.E., Liu, L.F., and Wang, J.C. (1978). Interaction between DNA and *Escherichia coli* protein omega. Formation of a complex between single-stranded DNA and omega protein. *J. Biol. Chem.* 253, 511-8.

[20] 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. U. S. A.* 108, 6939-44.

[21] 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, 5560-5.

[22] Beran-Steed, R.K. and Tse-Dinh, Y.C. (1989). The carboxyl terminal domain of *Escherichia coli* DNA topoisomerase I confers higher affinity to DNA. *Proteins* 6, 249-58.

[23] Naryshkin, N., Druzhinin, S., Revyakin, A., Kim, Y., Mekler, V., and Ebright, R.H. (2009). Static and kinetic site-specific protein-DNA photocrosslinking: analysis of bacterial transcription initiation complexes. *Methods Mol. Biol.* 543, 403-37.

[24] Forget, D. and Coulombe, B. (2003). Site-specific protein-DNA photocross-linking of purified complexes: topology of the RNA polymerase II transcription initiation complex. *Methods Enzymol.* 370, 701-12.

[25] Steen, H. and Jensen, O.N. (2002). Analysis of protein-nucleic acid interactions by photochemical cross-linking and mass spectrometry. *Mass Spectrom. Rev.* 21, 163-82.

[26] Geyer, H., Geyer, R., and Pingoud, V. (2004). A novel strategy for the identification of protein-DNA contacts by photocrosslinking and mass spectrometry. *Nucleic Acids Res.* 32, e132.

[27] Rosenfeld, K.K., Ziv, T., Goldin, S., Glaser, F., and Manor, H. (2011). Mapping of DNA binding sites in the Tetrahymena telomerase holoenzyme proteins by UV cross-linking and mass spectrometry. *J. Mol. Biol.* 410, 77-92.

[28] Kovach, M.J., Tirumalai, R., and Landy, A. (2002). Site-specific photo-cross-linking between lambda integrase and its DNA recombination target. *J. Biol. Chem.* 277, 14530-8.

[29] Winnacker, M., Welzmiller, V., Strasser, R., and Carell, T. (2010). Development of a DNA photoaffinity probe for the analysis of 8-OxodG-binding proteins in a human proteome. *Chembiochem* 11, 1345-9.

[30] Winnacker, M., Breeger, S., Strasser, R., and Carell, T. (2009). Novel diazirine-containing DNA photoaffinity probes for the investigation of DNA-protein-interactions. *Chembiochem* 10, 109-18.

[31] Shigdel, U.K., Zhang, J., and He, C. (2008). Diazirine-based DNA photo-cross-linking probes for the study of protein-DNA interactions. *Angew. Chem. Int. Ed Engl.* 47, 90-3.

[32] Savkina, M., Temiakov, D., McAllister, W.T., and Anikin, M. (2010). Multiple functions of yeast mitochondrial transcription factor Mtf1p during initiation. *J. Biol. Chem.* 285, 3957-64.

[33] Pestryakov, P.E., Khlimankov, D.Y., Bochkareva, E., Bochkarev, A., and Lavrik, O.I. (2004). Human replication protein A (RPA) binds a primer-template junction in the absence of its major ssDNA-binding domains. *Nucleic Acids Res.* 32, 1894-903.

[34] Paratkar, S. and Patel, S.S. (2010). Mitochondrial transcription factor Mtf1 traps the unwound non-template strand to facilitate open complex formation. *J. Biol. Chem.* 285, 3949-56.

[35] Pingoud, V., Geyer, H., Geyer, R., Kubareva, E., Bujnicki, J.M., and Pingoud, A. (2005). Identification of base-specific contacts in protein-DNA complexes by photocrosslinking and mass spectrometry: a case study using the restriction endonuclease SsoII. *Mol. Biosyst* 1, 135-41.

[36] Weng, L., Zhou, C., and Greenberg, M.M. (2015). Probing interactions between lysine residues in histone tails and nucleosomal DNA via product and kinetic analysis. *ACS Chem. Biol.* 10, 622-30.

[37] Hong, I.S., Ding, H., and Greenberg, M.M. (2006). Radiosensitization by a modified nucleotide that produces DNA interstrand cross-links under hypoxic conditions. *J. Am. Chem. Soc.* 128, 2230-1.

[38] Hong, I.S. and Greenberg, M.M. (2005). Efficient DNA interstrand cross-link formation from a nucleotide radical. *J. Am. Chem. Soc.* 127, 3692-3.

[39] Narula, G. and Tse-Dinh, Y.C. (2012). Residues of *E. coli* topoisomerase I conserved for interaction with a specific cytosine base to facilitate DNA cleavage. *Nucleic Acids Res.* 40, 9233-43.

[40] 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.

[41] Cheng, B., Feng, J., Mulay, V., Gadgil, S., and 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, 39207-13.

[42] Ding, H., Majumdar, A., Tolman, J.R., and Greenberg, M.M. (2008). Multinuclear NMR and kinetic analysis of DNA interstrand cross-link formation. *J. Am. Chem. Soc.* 130, 17981-7.

[43] Tse-Dinh, Y.C. and Beran-Steed, R.K. (1988). Escherichia coli DNA topoisomerase I is a zinc metalloprotein with three repetitive zinc-binding domains. *J. Biol. Chem.* 263, 15857-9.

[44] Feinberg, H., Lima, C.D., and Mondragon, A. (1999). Conformational changes in *E. coli* DNA topoisomerase I. *Nat. Struct. Biol.* 6, 918-22.

[45] Yu, L., Zhu, C.X., Tse-Dinh, Y.C., and Fesik, S.W. (1995). Solution structure of the C-terminal single-stranded DNA-binding domain of *Escherichia coli* topoisomerase I. *Biochemistry* 34, 7622-8.

[46] Feinberg, H., Changela, A., and Mondragon, A. (1999). Protein-nucleotide interactions in *E. coli* DNA topoisomerase I. *Nat. Struct. Biol.* 6, 961-8.

[47] Perry, K., Mondragon, A. (2003). Structure of a complex between *E. coli* DNA topoisomerase I and single-stranded DNA. Structure 11, 1349-58.

[48] Sulkowski, E., Odlyzko, A.M., and Laskowski, M.S. (1970). Preparation of 3'-dinucleotides from homoribopolymers by digestion with micrococcal nuclease. *Anal. Biochem.* 38, 393-400.

[49] Keller, A., Nesvizhskii, A.I., Kolker, E., and Aebersold, R. (2002). Empirical statistical model to estimate the accuracy of peptide identifications made by MS/MS and database search. *Anal. Chem.* 74, 5383-92.

FIGURE LEGENDS

Scheme 1. Proposed mechanism of relaxation of negatively supercoiled DNA by bacterial topoisomerase I. Step 1: Single-stranded DNA regions in negatively supercoiled DNA bind to N-terminal domains as G-strand and C-terminal domains as T-strand. Step 2: G-strand is cleaved to form covalent intermediate with phosophotyrosine linkage to active site tyrosine (yellow star). Step 3: Enzyme undergoes conformational change to allow entry of T-strand into central hole. Step 4: The cleaved G-strand is religated. Step 5: Enzyme produces DNA product with increased winding of one helical turn to reduce negative DNA supercoiling by one linking number.

Figure 1. Incorporation of PhSeT in partial duplex Substrate-Y for oxidative crosslinking to EcTopA. (A) Oxidative crosslinking of PhSeT to lysine, histidine and cysteine residues in proteins. (B) Partial duplex Substrate-Y formed by hybridization of G-strand and T-strand oligonucleotides. The arrow in the G-strand sequence corresponds to the site of cleavage by EcTopA to form the covalent enzyme-DNA intermediate.

Figure 2. Cleavage of the G-strand in Substrate-Y by EcTopA. The 5'-^{32}P labeled T-strand or G-strand oligonucleotides (5 pmol) were used as substrate for cleavage by EcTopA in either single-stranded form (lanes 1-3, 7-9) or following hybridization to the unlabeled complementary strand to form Substrate-Y (lanes 4-6, 10-12). Lanes 1, 4, 7, 10 are control reactions with no enzyme added. EcTopA (1 pmol) was added to the other reactions. Lanes 3, 6, 9, 12 reactions had 100 mM NaCl added. The other lanes had 10 mM NaCl.

Figure 3. Oxidative crosslinking of Substrate-Y to EcTopA analyzed by SDS PAGE and Phosphorimaging. (A) Time course of formation of crosslinked EcTopA complex from crosslinking reaction with 2 pmol of Substrate-Y labeled with ^{32}P at the 5'-end of the T-strand, and 40 pmol of EcTopA in reaction buffer containing 100 mM NaCl. (B) Increase in cross-linked complex with increasing concentration of Substrate-Y. Lane 1: control reaction of 0.25 pmol of labeled Substrate-Y with no enzyme added; Lanes 2-9 had 40 pmol of EcTopA. NaCl concentration: 10 mM in lanes 1-5, and 100 mM in lanes 6-9. Lanes 2, 6: 0.25 pmol Substrate-Y; lanes 3, 7: 0.5 pmol Substrate-Y; lanes 4, 8: 1 pmol Substrate-Y; lanes 5, 9: 2 pmol Substrate-Y.

Figure 4. Scaled up reaction for oxidative crosslinking of substrate-Y to EcTopA and the processing of the crosslinked complex by Micrococcal nuclease digestion. The Coomassie blue stained SDS gels show: (A) Crosslinked complex formed between Substrate-Y and EcTopA in reaction with 60 pmol each of EcTopA and Substrate-Y migrating above unreacted EcTopA (lane 2). (B) Enrichment of crosslinked complex by DE52 chromatography, followed by digestion with Micrococcal nuclease. Each fraction eluted from DE52 cartridge (lanes 2, 4) was digested with Micrococcal nuclease (lanes 3, 5). Lane 1: Molecular weight standards.

Figure 5. Positions of lysines identified as crosslinked residues in the structures of EcTopA. (A) Position of lysine residues in different domains of apo structure of full length EcTopA generated from PDB 4RUL. (B) A close up view of of EcTopA N-terminal fragment in complex with trinucleotides pTTT (PDB 1CY1). (C) A close up view of the full length EcTopA with single-stranded oligonucleotide bound to the C-terminal domains (PDB 4RUL). (D) A close up view of EcTopA N-terminal fragment forming a covalent complex with cleaved single-stranded G-strand oligonucleotide (PDB 3PX7). The images were generated using PyMOL.

Table 1. Tryptic peptides in EcTopA found to have modification of XGphospho from increase of molecular mass by 649. The lysines predicted to form the crosslinking adduct are underlined. H681 in the last peptide is an alternative crosslinking site if the XGphospho linkage to histidine is included as a potential modification in search criteria. The two cysteines in the last peptide have carbamidomethyl (+57) modification.

Residues	Sequence	Peptide Identification Probability	Actual Minus Calculated Peptide Mass (PPM)
16-28	TIN<u>K</u>YLGSDYVVK	88%	-1.3
20-35	YLGSDYVV<u>K</u>SSVGHIR	91%	-0.37
272-296	EDKPTTS<u>K</u>PGAPFITSTLQQAASTR	87%	-0.80
303-310	<u>K</u>TMMMAQR	85%	-1.8
679-700	<u>K</u>LHVCGNNPTCDGYEIEEGEFR	99%	-1.4

Scaffold (version Scaffold_4.4.3, Proteome Software Inc., Portland, OR) was used to validate MS/MS based peptide and protein identifications. Peptide identifications were accepted if they could be established at greater than 80.0% probability by the Peptide Prophet algorithm [49] with Scaffold delta-mass correction.

Scheme 1

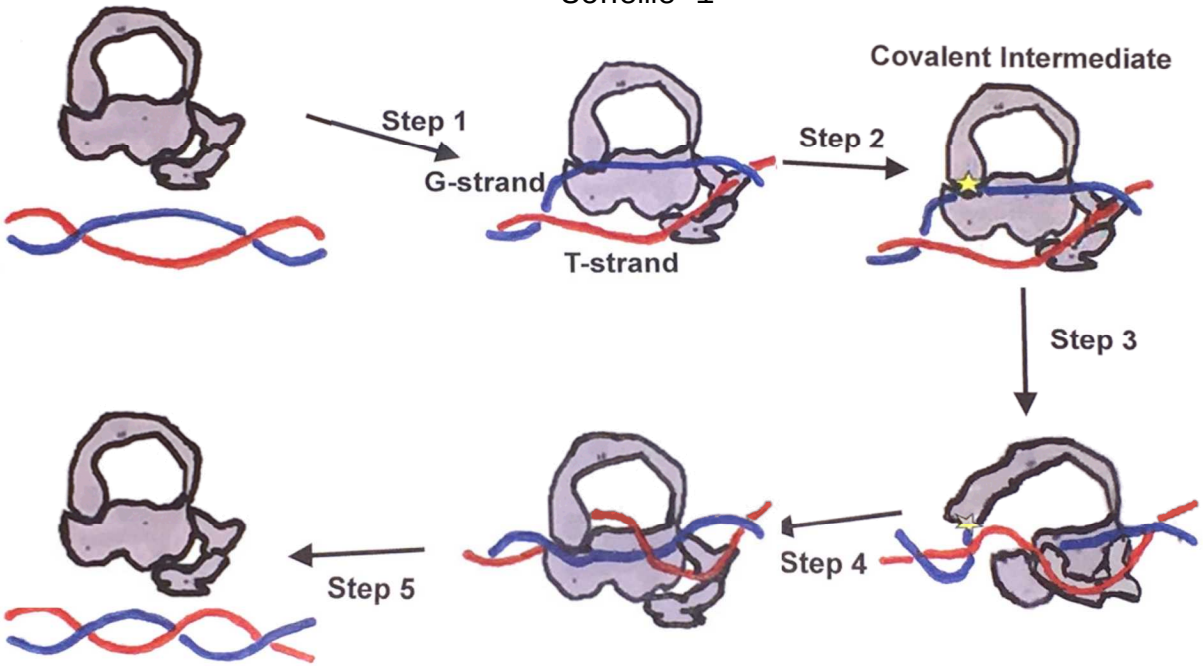
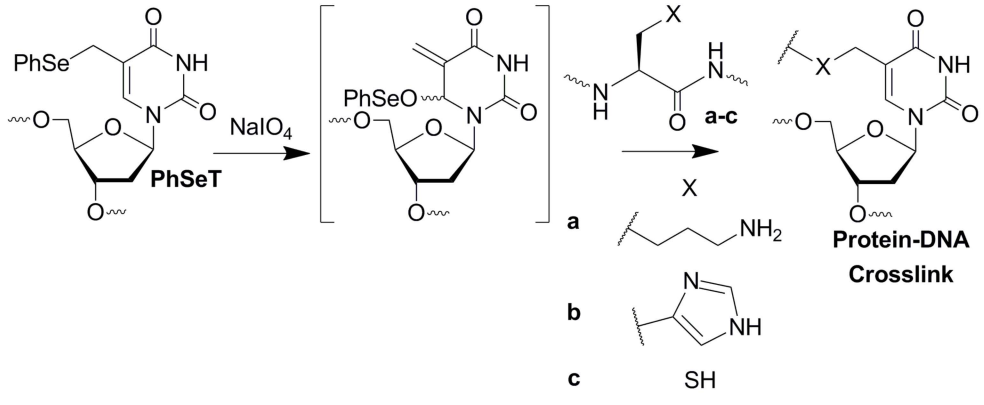

Step 1
G-strand
T-strand
Step 2
Covalent Intermediate
Step 3
Step 4
Step 5

Figure 1

A



B

5′-GCTAACCTGAAAGATTATGCGAT↓TTGGG-3′ G-strand
3′-CGATTGGACTTTCTAATAC
GTGTTATG**X**-5′ T-strand

X= PhSeT (phenylselenyl-modified thymidine)

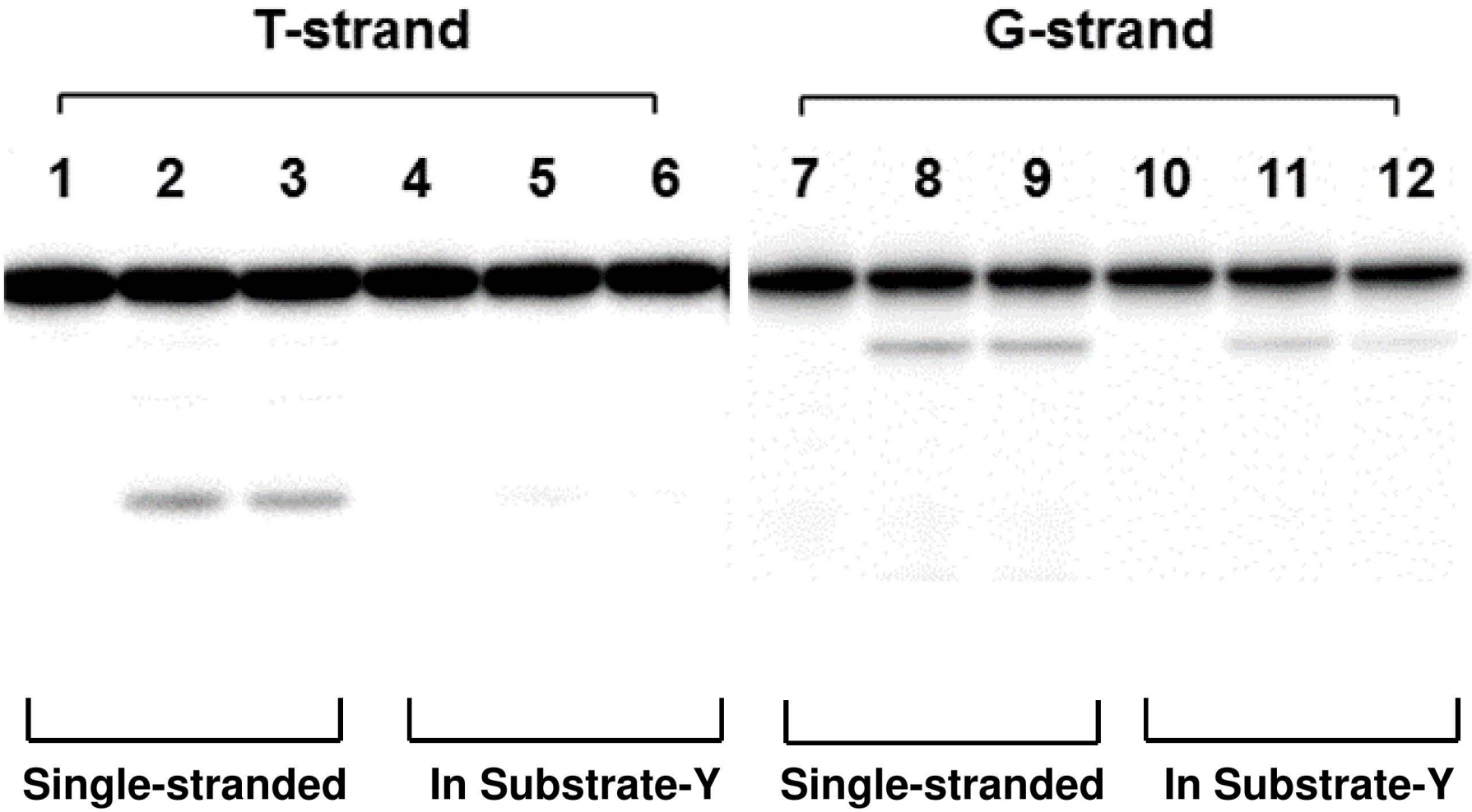

T-strand
G-strand
1
2
3
4
5
6
7
8
9
10
11
12
Single-stranded
In Substrate-Y
Single-stranded
In Substrate-Y

Figure 2

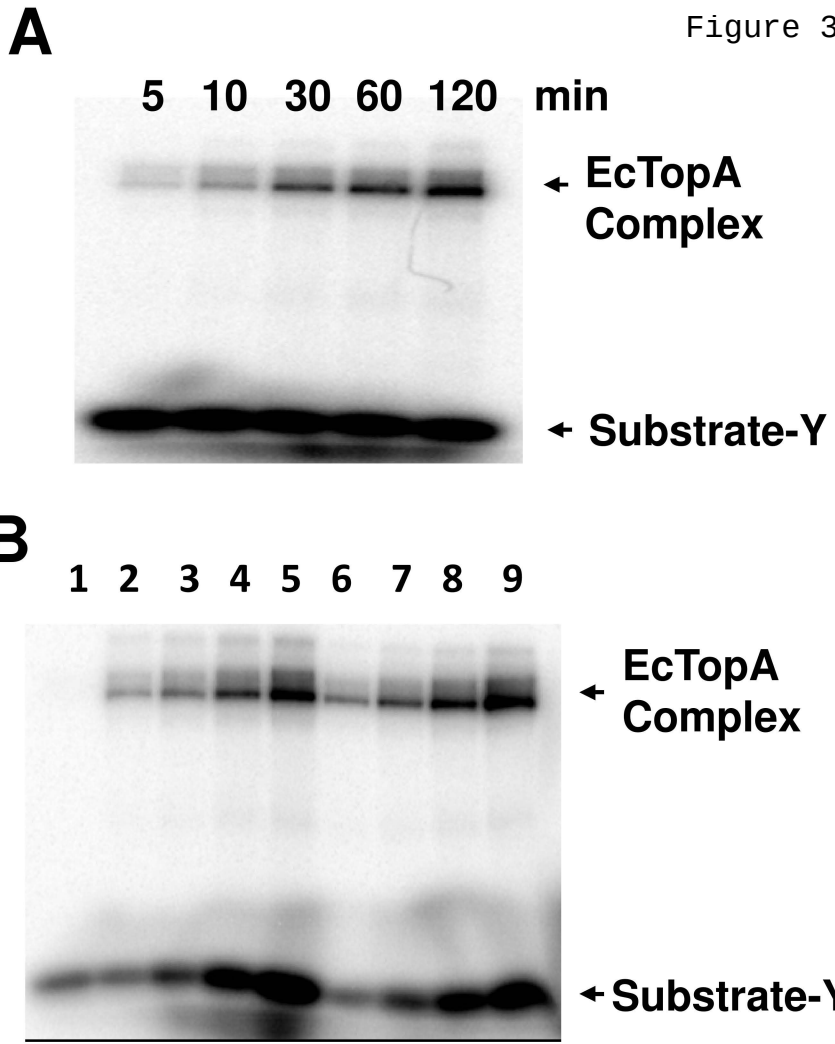
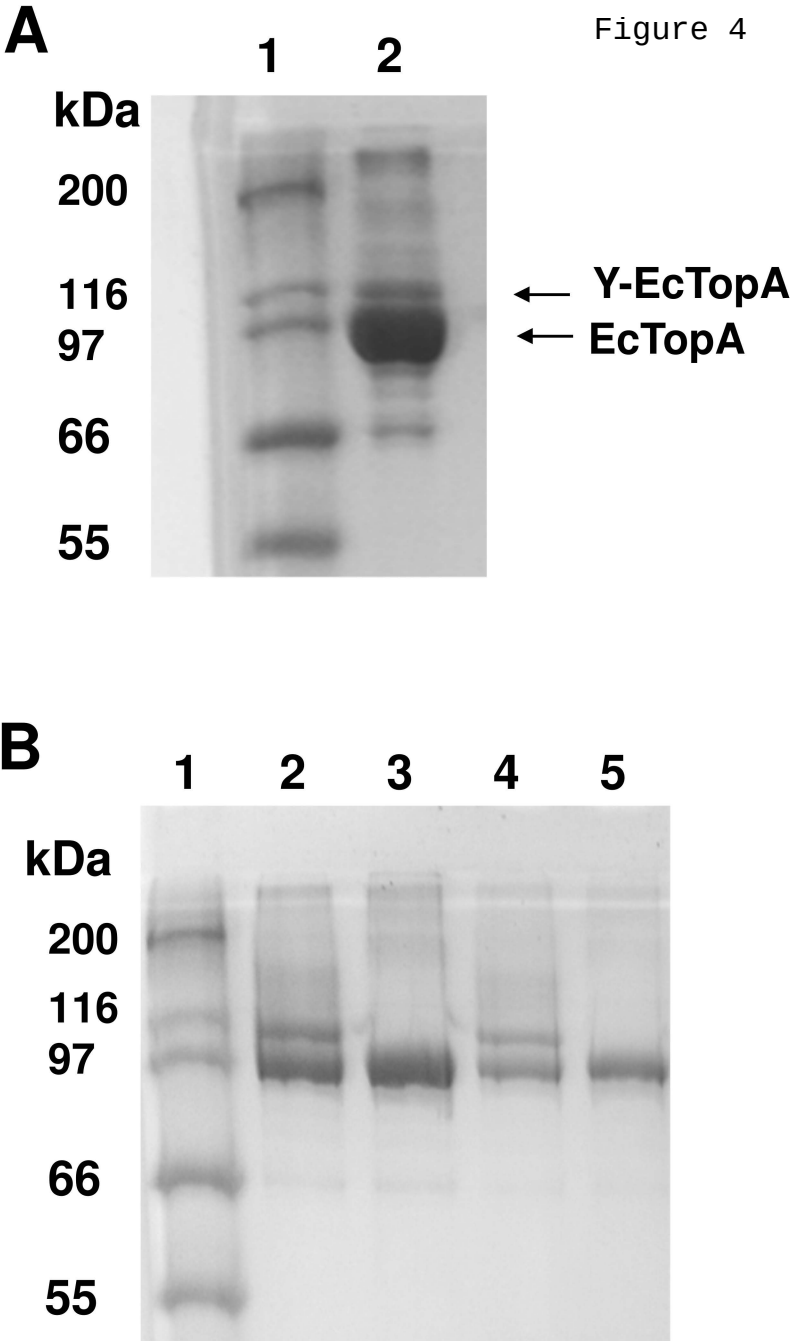

A
5 10 30 60 120 min
EcTopA Complex
Substrate-Y
B
1 2 3 4 5 6 7 8 9
EcTopA Complex
Substrate-Y

Figure 4


A
1
2
kDa
200
116
97
66
55
Y-EcTopA
EcTopA
B
1
2
3
4
5
kDa
200
116
97
66
55

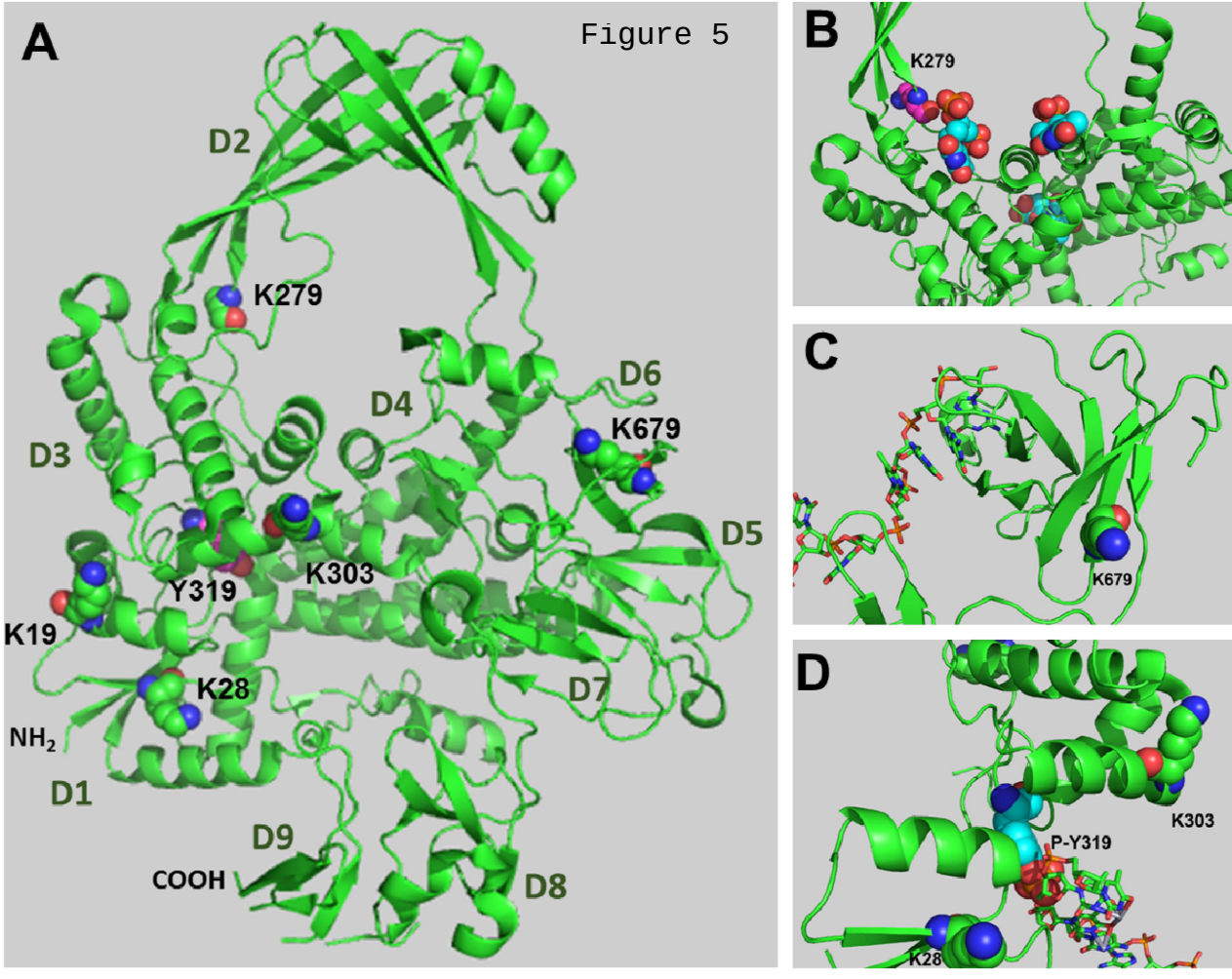


Figure 5
A
D2
K279
D6
D4
K679
D3
D5
K303
Y319
K19
K28
D7
NH2
D1
D9
COOH
D8
B
K279
C
K679
D
K303
P-Y319
K28