The structure of Escherichia coli DNA topoisomerase III

ARTICLE *in* STRUCTURE · DECEMBER 1999

Impact Factor: 5.62 · DOI: 10.1016/S0969-2126(00)80027-1 · Source: PubMed

The structure of *Escherichia coli* DNA topoisomerase III

Alfonso Mondragón[1]* and Russell DiGate[2]

Background: DNA topoisomerases are enzymes that change the topology of DNA. Type IA topoisomerases transiently cleave one DNA strand in order to pass another strand or strands through the break. In this manner, they can relax negatively supercoiled DNA and catenate and decatenate DNA molecules. Structural information on *Escherichia coli* DNA topoisomerase III is important for understanding the mechanism of this type of enzyme and for studying the mechanistic differences among different members of the same subfamily.

Results: The structure of the intact and fully active *E. coli* DNA topoisomerase III has been solved to 3.0 Å resolution. The structure shows the characteristic fold of the type IA topoisomerases that is formed by four domains, creating a toroidal protein. There is remarkable structural similarity to the 67 kDa N-terminal fragment of *E. coli* DNA topoisomerase I, although the relative arrangement of the four domains is significantly different. A major difference is the presence of a 17 amino acid insertion in topoisomerase III that protrudes from the side of the central hole and could be involved in the catenation and decatenation reactions. The active site is formed by highly conserved amino acids, but the structural information and existing biochemical and mutagenesis data are still insufficient to assign specific roles to most of them. The presence of a groove in one side of the protein is suggestive of a single-stranded DNA (ssDNA)-binding region.

Conclusions: The structure of *E. coli* DNA topoisomerase III resembles the structure of *E. coli* DNA topoisomerase I except for the presence of a positively charged loop that may be involved in catenation and decatenation. A groove on the side of the protein leads to the active site and is likely to be involved in DNA binding. The structure helps to establish the overall mechanism for the type IA subfamily of topoisomerases with greater confidence and expands the structural basis for understanding these proteins.

Addresses: [1]Department of Biochemistry, Molecular Biology and Cell Biology, Northwestern University, 2153 Sheridan Road, Evanston, IL 60208-3500, USA and [2]Department of Pharmaceutical Science, School of Pharmacy and Medical Biotechnology Center, University of Maryland Biotechnology Institute, Baltimore, MD 21201, USA.

*Corresponding author.
E-mail: a-mondragon@nwu.edu

Key words: conformational changes, crystal structure, DNA topology, enzyme mechanism, topoisomerases

Received: 8 June 1999
Revisions requested: 15 July 1999
Revisions received: 29 July 1999
Accepted: 9 August 1999

Published: 29 October 1999

Structure November 1999, 7:1373–1383

0969-2126/99/$ – see front matter

Introduction

DNA topoisomerases are enzymes that change the topological state of DNA. They are involved in several crucial cellular processes such as replication, transcription, chromosome segregation, and chromosome condensation (for a review see [1]). All topoisomerases work by transiently breaking one or two strands of the DNA phosphodiester backbone to allow passage of single-stranded DNA (ssDNA) or double-stranded DNA (dsDNA) through the break. By this simple mechanism, topoisomerases can change the superhelicity of DNA, or catenate or decatenate DNA molecules.

Topoisomerases are classified into two major families: type I enzymes break only one DNA strand, whereas type II enzymes catalyze the concerted cleavage of a DNA duplex. In both cases the transient cleavage of the phosphodiester backbone involves the formation of one or two phosphotyrosine bonds. Type I enzymes are further classified into two subfamilies on the basis of their biochemical properties: type IA enzymes form a transient covalent bond to the 5′ phosphoryl of the broken strand, whereas type IB enzymes form a 3′-end covalent bond. In addition to their common biochemical mechanism, members of the same subfamily share significant sequence similarities and are very likely to be structurally related. Moreover, the structure determination of two type IB enzymes [2–4] confirmed the anticipated structural dissimilarity between the type IA and IB subfamilies.

The initial identification of *Saccharomyces cerevisiae* DNA topoisomerase III [5] as a type IA enzyme together with the identification of a ‘reverse gyrase’ in archaebacteria [6] and the identification of a human type IA molecule [7] clearly suggests that the type IA subfamily enzymes are widespread in nature and not confined to prokaryotes as originally thought. Type IB enzymes have only been positively identified in eukaryotes and possibly in archaebacteria [8]. Until recently, all known type II enzymes seemed to be related and to form a single unified family spanning all kingdoms, although the identification in *Sulfulobus shibatae* of a type II topoisomerase suggests that at least two type II subfamilies may also exist [9].

Four topoisomerases have been identified in *E. coli*: topoisomerases I–IV. Topoisomerases I and III belong to the type IA subfamily and share sequence and mechanistic similarities whereas topoisomerases II and IV are related type II enzymes. *In vitro*, topoisomerases I and III catalyze similar reactions but with different preferences: topoisomerase III is a more efficient decatenating enzyme, whereas topoisomerase I is more efficient at relaxing supercoiled DNA [10]. For these reasons it has been suggested that topoisomerase III is likely to be involved in the unlinking of nascent daughter chromosomes [10], a reaction not catalyzed by topoisomerase I [11]. The molecular organization of these two enzymes is related. There is significant sequence similarity in a region spanning approximately the first 600 amino acids of the proteins (1–581 in topoisomerase I, 1–609 in topoisomerase III) and no discernible sequence similarity in the rest of the protein. In both cases, the C termini of the proteins are involved in interactions with DNA [12,13]. It has been shown that the 44 amino acids at the C terminus of topoisomerase III are involved in binding to ssDNA but they can be removed without total loss of enzymatic activity [13]. The first 609 amino acids of topoisomerase III confer full enzymatic activity, although the truncated enzyme is less processive [13]. In contrast, *E. coli* DNA topoisomerase I has a much larger DNA-binding domain at the C terminus, which is needed for full activity [14]. This difference is probably reflected in the size of the DNA region interacting with the protein: topoisomerase I protects a region of approximately 40 bases whereas topoisomerase III only protects a 14-base region [15]. Interestingly, the ssDNA cleavage pattern generated by topoisomerase III does not change between the full-length and the truncated enzyme [13], suggesting that the weak sequence preference exhibited does not reside in the last 44 amino acids. Strong sequence similarity around the active site suggests that the DNA cleavage and re-ligation reactions proceed in the same way in both proteins. A major difference is that *in vitro* topoisomerase III cleaves RNA [16] as well as DNA, and it also has been shown to decatenate RNA molecules [17], making it the only topoisomerase known to have an RNA-decatenating activity. It is not known whether this RNA-decatenating activity is significant *in vivo*.

The structure of a 67 kDa N-terminal fragment of *E. coli* DNA topoisomerase I (residues 1–596) [18] shows that the protein has a toroidal shape formed by four domains. The protein has a large hole in the center of approximately 25 Å diameter, which can easily accommodate DNA. The active site is at the interface between two domains. It has been suggested [18] that the enzyme breaks one DNA strand in synergy with a large conformational change to open a gate in the broken DNA strand. The opening allows passage of the other strand(s) into the central hole in the protein. After passage the protein closes again, allowing re-ligation of the broken strand. Opening of the protein to release the captured strand(s) completes the reaction. No external energy source, such as ATP, is required for the reaction. This mechanism explains all of the reactions catalyzed by this type of protein, including relaxation, catenation and decatenation.

A similar mechanism involving the formation of DNA gates has been suggested for the type II enzymes [19–21]. The structure of a large fragment of yeast topoisomerase II [20] shows that this protein also has a large hole formed by a protein dimer, with each monomer containing one active site. The proposed mechanism of action also involves the opening of the hole to allow the DNA to enter the protein. In the case of topoisomerase II, however, two different openings are involved, one to allow entrance and one to allow exit of the DNA [20,21]. ATP hydrolysis is required to return the protein to its original conformation, but not for the cleavage or re-ligation reactions [22]. It is clear that although the type IA and type II enzymes are very different, there are some remarkable parallelisms in their mode of action.

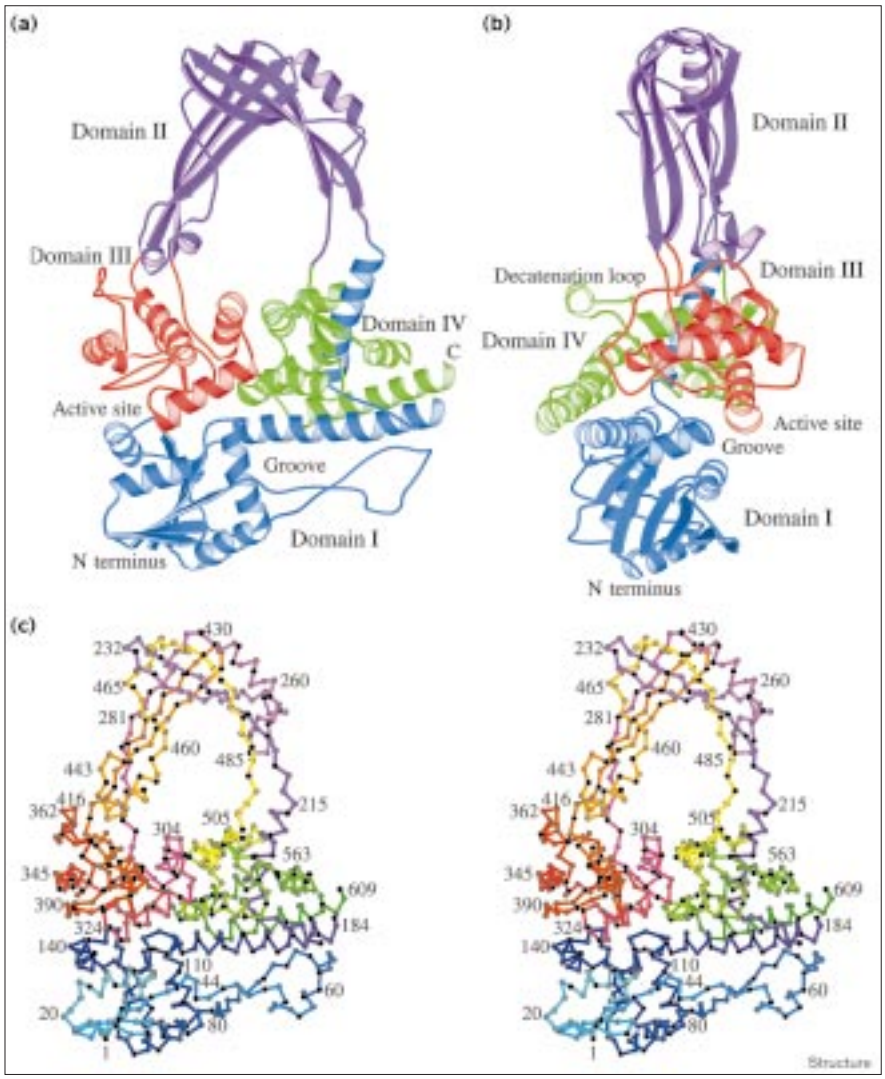
Here we present the first structure of an intact and fully active type IA DNA topoisomerase, *E. coli* DNA topoisomerase III. By comparing this structure with that of the 67 kDa N-terminal fragment of *E. coli* DNA topoisomerase I, another member of the type IA subfamily, we are able to address important mechanistic questions that were raised by the structure of topoisomerase I.

Results

Overall structure of *E. coli* DNA topoisomerase III

All members of the type IA subfamily share significant sequence similarity [23] and it is therefore not surprising that the structure of topoisomerase III has the now familiar fold of the type IA enzymes. It consists of four domains (I–IV) that form a torus with a hole of approximately 25 Å in the center (Figure 1). In the crystal structure of topoisomerase III most of the amino acids are visible, except in two adjacent regions: a loop at the surface of the protein (185–190) and the last 44 amino acids (610–653). Domain I (residues 1–216) is an α/β domain with a six-stranded mixed β sheet at its core. Domain II (residues 218–287 and 418–488) forms over half of the torus and achieves its curvature through several β strands. Domain III (residues 288–417) is mainly helical and is located next to domains I and IV. The last domain, IV (residues 489–609), forms part of the main body of the protein and makes contacts with domains I and III. The C terminus of the protein is located in domain IV. The last ordered residue, Thr609, is at the end of a long α helix distal to domain I. As mentioned above, the last 44 amino acids form an ssDNA-binding domain. This domain starts at residue 610 and is unlikely to approach the active site, as the distance from the last ordered residue to the active site is too long to be spanned easily by such a small domain.

Figure 1



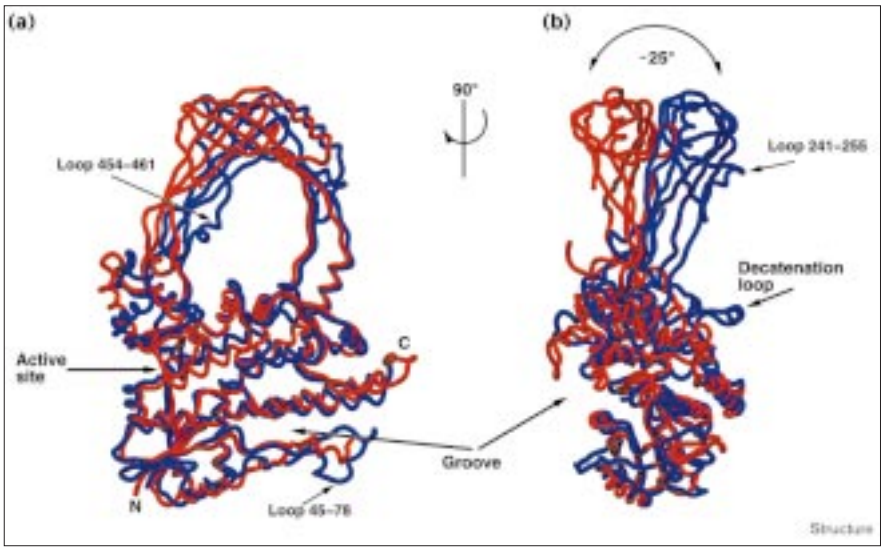
Overall architecture of *E. coli* DNA topoisomerase III. (a,b) Ribbon representation showing the four domains that constitute the protein. (a) Frontal view; (b) side view. The domains are colored differently to distinguish them (domain I, blue; domain II, purple; domain III, red; and domain IV, green). The two views emphasize the unusual shape of the protein, which is due not only to the presence of the central hole but also to the relatively narrow width of the protein. The active site is located at the interface of domains I and III and is inaccessible in this conformation. The active-site tyrosine is located in domain III. The proposed mechanism of action calls for domain III and domain II to move in order to create an opening or gate. The putative binding groove for ssDNA and the decatenation loop are marked. (c) Stereoview of the Cα trace of the protein in the same orientation as in (a). The protein has been colored using a color ramp to emphasize the path followed by the protein. In particular, it shows that domain II is formed by residues from two different regions in the sequence. For reference, some residue numbers are shown. The drawing was made with the program MolScript [38].

Comparison with *E. coli* DNA topoisomerase I

A comparison of the topoisomerase III structure with the structure of the 67 kDa N-terminal domain of *E. coli* DNA topoisomerase I reveals significant differences (Figure 2). Figure 3 shows a structure-based sequence alignment that highlights the different domains and the secondary structure for both proteins. Table 1 summarizes the results of the superposition of the different domains, and the sequence similarity for each domain. The differences between each structural unit are not as large as when the entire structure is compared, indicating that a relative rearrangement of the domains has occurred. The nature of the reorientation of the domains can be best understood by looking at the superposition of domains I and IV (Figure 2). This shows that domain II has rotated approximately 25° relative to the rest of the protein. The movement is hinged at the base of domains II and III and leaves the active-site region largely unaffected. The observed movement could be due to crystal-packing interactions, or it may reflect a real difference between the two proteins. In either case, it reflects the flexibility of this domain with respect to the rest of the protein, supporting the notion that domains II and III could move away from the rest of the molecule to create an open conformation.

The comparison also reveals that domain I in topoisomerase III has an extra two β strands on the central β sheet, but that the α/β structure is very similar. A long loop that emanates from domain I and intertwines with domain IV is very different in both structures (residues 34–86 in topoisomerase I; 45–78 in topoisomerase III) and forms a region where different members of the same subfamily have insertions of different lengths and sequence [23]. In the structure of topoisomerase I this loop is largely

Figure 2



Comparison of the structures of *E. coli* DNA topoisomerases I (red) and III (blue). (a) Front view; (b) side view. For the comparison the mainchain atoms in domains I and IV were superposed. This accentuates the different orientation of domain II. The active site remains relatively unchanged as most of the movement occurs in the top domain (domain II), which seems to tilt sideways with respect to the main body of the protein. The superposition also emphasizes the fact that although both proteins are very similar overall, there are also considerable local differences due to insertions and deletions. In the crystal structure of topoisomerase III there are more ordered regions than in the crystal structure of topoisomerase I and hence more residues are visible. The presence of an extra ordered loop (454–461) on the side of the central hole makes the hole seem smaller in topoisomerase III, but it is likely that the hole is very similar in both structures. Another loop (241–255) in domain II also contributes to the central hole shape. Despite the absence of some residues lining the groove of topoisomerase I, the two proteins have a very similar groove in the main body. In topoisomerase III the groove contains many residues in the long 45–78 loop that forms part of domain I. The drawing was made with the program SETOR [39].

disordered, so a detailed comparison is impossible. At the end of domain I there is a long α helix that connects it to domain II. The orientation of this helix is slightly different, and suggests a mechanism by which interactions near domain I or IV could translate into movements of domain II. Domain II is also very similar in both proteins, except for one insertion in a short loop (residues 241–255). Domain III contains the active-site tyrosine and other highly conserved amino acids; it seems to be structurally the most different as there is a small deletion in topoisomerase III that causes a large rearrangement of an α helix. Nevertheless, the common regions in this domain are clearly very similar, particularly the ones containing residues that form the active site. The last domain, IV, has a large insertion (residues 502–519) in a loop connecting two helices and located at the base of the central hole (Figure 1). The loop, named the decatenation loop, is positively charged, with an arginine and three lysines facing the central hole. At the C terminus the 67 kDa fragment of topoisomerase I extends for a few more residues than the structure of topoisomerase III; suggesting that in both proteins the C terminus wraps around the protein and points towards a groove formed by domains I and IV.

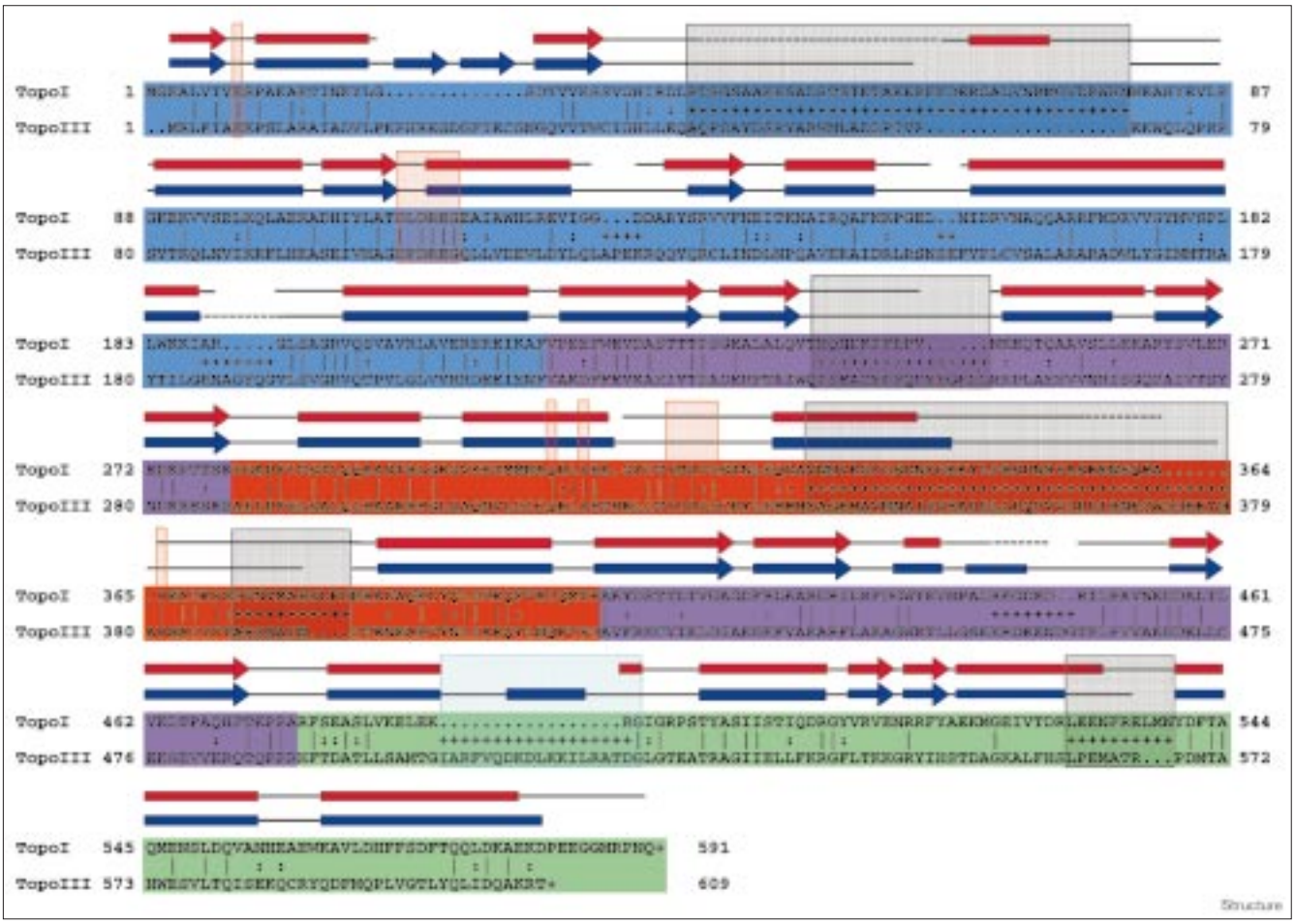
The structural comparison of the two molecules allows the alignment of the primary sequences in a slightly different manner. Figure 3 shows the primary sequence alignment of topoisomerases I and III. The percentage identity and similarity for the four domains is shown in Table 1. It is impossible to align the structures in some regions, as the structure is either disordered or too dissimilar. Domain III is the one with the highest sequence identity and smallest root mean square deviation (rmsd) for the structurally common regions, but it is also the domain that has the largest number of insertions and deletions and hence the lowest percentage of structurally similar regions. This domain is structurally the most different, but in the common regions it is the most similar. The alignment shows that for the first 609 amino acids of topoisomerase III the percent identity with topoisomerase I is around 26% and that important regions for activity, such as the active site, are highly conserved. This structure-based alignment should also help to realign the sequences of the growing family of type IA topoisomerases.

Active site and putative DNA-binding regions

The active site is located at the interface between domains I and III and is inaccessible in this conformation of the protein. As domain III is covalently linked to domain II it is clear that it cannot be separated from the rest of the protein without a large conformational change, such as the one suggested for *E. coli* DNA topoisomerase I [18]. Thus, the present structure represents one conformation of a very dynamic molecule. The conformational change is not only needed to expose the active site but also to provide a possible entrance to the hole at the center of the protein. The interface between the two domains is small. The buried surface area is 1390 Å^2 (1370 Å^2 in domain III and 1420 Å^2 in domains I and IV). There are also few interactions across the interface.

Domain IV meets domains I and III at a position adjacent to the active site (Figure 1), but no residues from domain

Figure 3



Sequence alignment of *E. coli* DNA topoisomerases I and III. The primary sequences were aligned on the basis of structural similarities, disregarding any sequence similarities. Nevertheless, the resulting alignment is similar to the one obtained by aligning the primary sequences of several type IA topoisomerases [23]. The four domains are identified using the same color scheme as in Figure 1 (domain I, blue; domain II, purple; domain III, red; and domain IV, green). The secondary structural elements for both proteins are shown above the sequence (topoisomerase I, red bars; topoisomerase III, blue bars; α helix, solid bars; β strand, arrowed bars), the dotted lines represent regions whose structure is not known, and gaps represent sequence insertions. Red boxes surround residues that form the active site. Black boxes surround residues that are structurally different and cannot be aligned. The cyan box surrounds the decatenation loop. The hinge region corresponds to the border between domain II and domains I and IV (blue/purple border and purple/green border). The sequence is shown using single-letter amino acid notation. Identical residues are indicated by a vertical bar, similar residues with a colon (:). Regions that could not be aligned owing to large structural differences or disordered regions are marked with +. Insertions in the sequences are indicated by a period (·). The C-terminal residues that are not present in the structures are not shown.

IV seem to form part of the active site itself. At this intersection the protein has a narrow cross-section. A clear groove can be seen that extends from domain IV to the active site of the protein (Figures 2,4). The groove is between 10 and 20 Å wide and 25 and 30 Å long and can be several angstroms deep in some areas. It provides a clear path of entrance to the active site.

The central hole has 17 lysines and arginines facing the interior, giving this region an overall positive charge (Figure 4). The size and charge of the hole are such that dsDNA or ssDNA could be accommodated easily inside the hole, even in the closed conformation. Trapping experiments have shown that both dsDNA and ssDNA can be trapped inside the protein during the catalytic cycle (RD, unpublished results). In topoisomerase III there is an extra ordered region on one side of the hole that makes the hole seem smaller, but overall the holes in both proteins are similar (see Figure 2). A significant difference is the presence of two loops in topoisomerase III, one at the interface of domains III and IV (502–519) and the other in domain II (241–255). They serve to create two extra ledges on the side of the hole, giving it a slightly different appearance (Figure 4). The decatenation loop at

Table 1

Comparison between *E. coli* DNA topoisomerases I and III.

	Domain I	Domain II	Domain III	Domain IV	Overall
Rmsd for C_α atoms* (Å)	1.7	1.8	0.9	2.0	3.7
Sequence identity† (%)	25.6	24.6	34.1	22.9	26.3
Sequence similarity‡ (%)	34.5	32.2	43.9	33.3	35.3
Structurally similar residues§ (%)	77.8	83.1	63.1	79.3	76.2

*Only those residues that are clearly structurally similar were included in the comparison (see Figure 3). †Percentage of residues in topoisomerase III that are structurally similar and identical to the corresponding residues in topoisomerase I. ‡Percentage of residues in topoisomerase III that are structurally similar and identical or similar to the corresponding residues in topoisomerase I. §Percentage of residues in topoisomerase III that are structurally similar to the corresponding residues in topoisomerase I and hence were included in the comparisons.

the intersection of domains III and IV is only present in some members of this family and contains several arginines and lysines that face the central hole. Deletion of this loop abolishes decatenation but not relaxation activity (RD, unpublished results). This loop must play a central role in decatenation activity, although the exact mechanism is unclear. One possibility is that by changing the charge and shape of the central hole the protein selects for DNA molecules with different topologies.

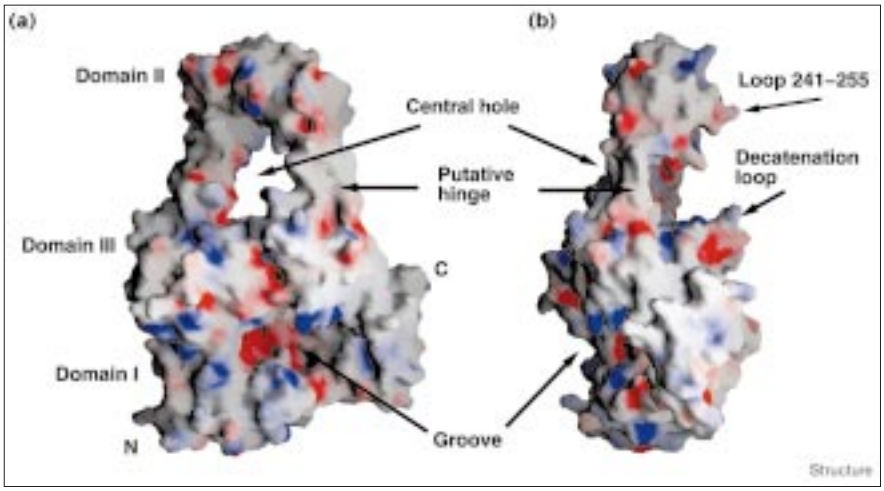
Discussion

Active site and cleavage mechanism

In the conformation of the protein in the crystal the active site is buried between domains I and III. Comparison with the structure of topoisomerase I as well as with the sequences of other members of the same subfamily reveals that this region is the most highly conserved. Figure 5 shows the active site of topoisomerase III. It includes Glu7, Asp103, Asp105 and Glu107 in domain I, and Glu317, Tyr320, Glu321, Tyr328, Arg330, Asp332 and His381 in domain III. Of these ten residues, five are completely conserved and the other five are highly conserved [23]. Several of these residues are not in the immediate vicinity of the active-site tyrosine and form a secondary shell. They may play a role in protein–DNA interactions during the reaction.

The equivalent residues in topoisomerase I are Glu9, Asp111, Asp113, Glu115, Glu309, Tyr312, Glu313, Tyr319, Arg321, Asp323 and His365. Several of these residues have been mutated to alanine [24,25], but mutation of only some residues — Glu9, Asp111, Glu115, Asp113, Tyr319 and Arg321 — results in loss or reduction of relaxation activity. Arg321 can be mutated to a lysine without loss of activity [24], suggesting that the important feature is the positive charge of the amino acid. Glu9 can be replaced by a glutamine, leading to loss of relaxation activity but not DNA cleavage activity; this may indicate that this amino acid interacts directly with DNA [24].

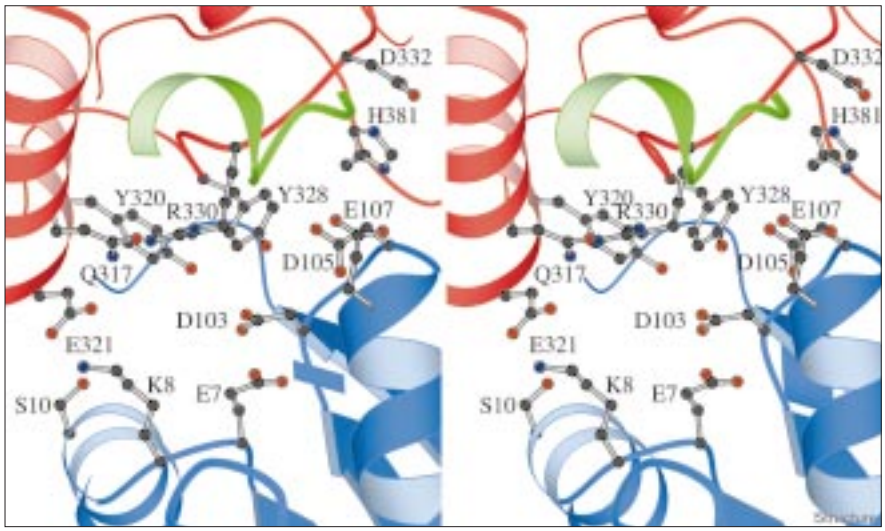
Figure 4



Surface representation of *E. coli* DNA topoisomerase III. (a) Front view; (b) side view. The molecular surface is colored according to charge. Red areas correspond to negative charge, blue to positive charge. The potential levels correspond to $-15\ k_bT/e$ for blue and $+15\ k_bT/e$ for red. The surface representation emphasizes the presence of the putative groove or binding site for ssDNA, the central positively charged hole and the decatenation loop. The central hole has two ledges formed by the decatenation loop and loop 241–255, giving it a different appearance from the hole in topoisomerase I. The decatenation loop is positively charged in the region that faces the central hole. The size of the DNA-binding groove is such that ssDNA could be accommodated inside easily. The presence of positively charged loops at the rims of the groove suggests that ssDNA could bind with the bases facing the protein and the phosphodiester backbone away from it. This binding mode has been observed in RPA, an ssDNA-binding protein [29]. The putative hinge region is marked. This is the region that might serve as a hinge point to allow opening of the protein. The surface representation was drawn with the program GRASP [40].

Figure 5

Stereo representation of the active site of *E. coli* DNA topoisomerase III. The mainchain of the protein is colored using the same color scheme as in Figure 1. This region contains some of the most highly conserved residues among all type IA topoisomerases. The active-site tyrosine, Tyr328, is surrounded by several acidic residues in domain I. In the open conformation, these acidic residues may serve to bind magnesium and through it ssDNA. The drawing was made with the program MolScript [38].



Mechanistically, it is clear that domain III has to separate from domain I in order to expose the active site and to accept the strand to be cleaved, and also to open a gate for entrance (or exit) of the other strand(s) into the central cavity. Once the domains separate, the acidic residues in domain I could be involved in magnesium binding. These amino acids are highly conserved and are spatially arranged in such a manner that they could easily adopt a geometry appropriate for this. As has been pointed out before [18], the spatial arrangement of these acidic residues is similar to the one observed in the exonuclease site of Klenow fragment [26] and suggests that they could be involved in magnesium binding.

In order to form the phosphotyrosine bond involved in the transient cleavage reaction a proton has to be abstracted from the OH of the tyrosine. It has been suggested that type IB enzymes work by general acid–base catalysis [27] and this could be true of type IA enzymes as well. The pH profiles of both topoisomerase I and III (K Perry and AM, unpublished results, and RD, unpublished results) show bell-shaped curves for the relaxation activity against pH. The relaxation activity disappears at around pH 6.5, consistent with an ionizable histidine. The structures of topoisomerase I and III reveal the presence of a strictly conserved histidine (residues 365/381) in the vicinity of the active-site tyrosine that could be a plausible candidate, but the mutagenesis results seem to rule out this histidine as a general base as the protein is still active [24]. According to the mutagenesis studies, the only alternative candidate is a glutamate (residues 7/9), but in both structures the glutamate is too far from the tyrosine to be able to interact directly without a large conformational change. It is still impossible to assign specific roles to any residues in the vicinity of the active-site tyrosine, and further studies are required to elucidate the atomic mechanism of the cleavage and re-ligation reactions.

Mechanistic implications

As mentioned above, topoisomerase III does not require the last 44 amino acids for activity or for the recognition and cleavage of ssDNA, even though these amino acids are involved in ssDNA binding [13]. This suggests that the protein has more than one DNA-binding region. One possible DNA-binding region is the positively charged hole at the center of the protein. It is likely that this hole serves to accommodate DNA, either single- or double-stranded, during the strand passage steps and not to bind or recognize DNA. Its role is likely to be more passive, to act simply as a transient recipient of DNA during the topological transformations. Type IA topoisomerases require ssDNA for activity, and recognition probably involves the main body of the protein. In the structure of topoisomerase III there is a large groove that extends from the active-site region to the other end of the main body of the protein (Figure 4). A similar groove or depression is present in the structure of topoisomerase I, but disordered amino acids in this region make it less apparent. In the structures of *E. coli* DNA topoisomerase I in complex with nucleotides [28] it was observed that one binding site is near the active site at one end of the groove, further suggesting that ssDNA could bind in this region. The groove is only slightly charged at the point nearest the active site, where two charged amino acids are present. The wall of the groove is formed by a long α helix emanating from the α/β domain and connecting it to the beginning of domain II. The nature of the amino acids facing the groove suggests that ssDNA could enter the groove with the bases facing the groove and the phosphate backbone facing away. A small positively charged ledge above the groove

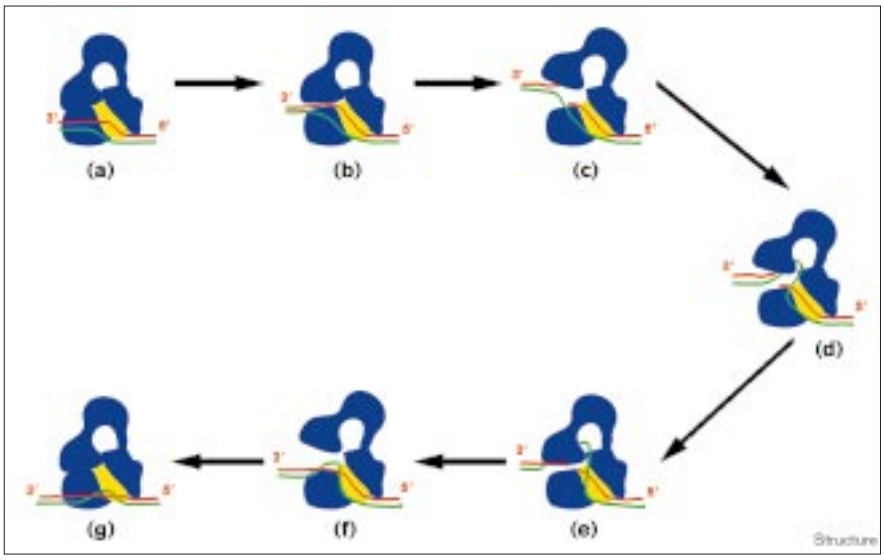
could help stabilize the complex. This orientation of the DNA with the bases buried in the protein could provide a mechanism for ssDNA recognition and also a way of maintaining a single-stranded region during the reaction. The position of the groove also suggests that upon binding to ssDNA a conformational change could be triggered by interactions with the helix connecting domains I and II. The proposed binding mode is similar to the one observed in the structure of Replication Protein A (RPA) ssDNA-binding domains with DNA, where ssDNA binds with the bases facing the protein and the phosphate backbone facing the opposite way [29]. The size of the binding groove in both RPA and topoisomerase III is similar.

Topoisomerase III protects at least two bases 3′ of the cleavage site and at least twelve bases 5′ of the cleavage site [15]. This suggests that ssDNA could enter the groove and extend towards the active site with the 3′ end pointing towards the active site and several bases entering the groove. This orientation and positioning of the DNA is also consistent with the expected interactions with the protein. After cleavage, the strand 3′ to the cleavage site stays noncovalently associated with the protein. Protein–DNA interactions in the groove would ensure that the DNA does not leave the protein, but remains noncovalently attached during the strand-passage steps.

The structure presented here for topoisomerase III represents a fully active topoisomerase. It is thus possible to propose a mechanism of action for the protein without having to invoke a role for the disordered C-terminal region. The mechanism should also apply to other members of the same subfamily, as it is clear that they are all structurally and mechanistically related. The essence of the mechanism of action of type IA topoisomerases was captured by the structure of *E. coli* DNA topoisomerase I [18] and is illustrated in Figure 6. It involves the opening of a gate in the DNA to allow passage of another strand(s) through it. The proposed mechanism involves the following steps: first, recognition and entrance into the active site of an ssDNA region (Figure 6a,b); second, cleavage of the ssDNA concomitant with a conformational change that separates domains I and III (Figure 6c); third, strand passage through an ssDNA gate and into the central hole of the protein (Figure 6d); fourth, re-ligation of the broken strand by bringing together the ends of the broken strand via a reversal of the conformational change (Figure 6d,e); and finally, expulsion of the unbroken strand(s) from the central hole (Figure 6f,g). The structure of topoisomerase III is consistent with this proposed mechanism and further validates it.

Topoisomerase III is a much better catenating and decatenating enzyme than topoisomerase I. Although

Figure 6



Proposed mechanism of action for type IA DNA topoisomerases, depicting possible steps in the reaction. The protein is shown in blue with a yellow patch corresponding to the putative binding groove for ssDNA. Two DNA strands are shown, one in red and the other in green. They could represent, for example, the two strands of a closed-circular duplex or two ssDNA circles. The putative orientation of the red strand, the strand to be cleaved, is indicated. The proposed mechanism applies to any reaction catalyzed by this family of enzymes, such as DNA relaxation, catenation or decatenation. The proposed mechanism involves the following steps. (a) Recognition of an ssDNA region by the protein. (b) Binding of the ssDNA region to the groove in the protein. The binding is followed, or occurs simultaneously, with the opening of the protein to allow entrance of one strand into the active site. (c) Cleavage of the DNA. The 5′ end of the broken strand remains covalently attached to one region of the protein (domain III) through the active-site tyrosine while the 3′ end of the broken strand remains noncovalently attached to a different region of the protein (domains I and IV). The noncovalent binding is probably aided by interactions along the groove and also by the highly conserved acidic residues in domain I. (d) Passage of the unbroken strand through the opening, or gate, formed by the broken strand. After passage through the opening the unbroken strand enters the central hole of the protein. (e) Once the unbroken strand is inside, the reaction is followed by a reversal of the protein conformational change to bring the two ends of the broken strand together and allow re-ligation. The re-ligation reaction is probably the reversal of the cleavage reaction. (f) Once the broken strand is re-ligated the protein opens up again to expel the strand trapped inside. (g) Final product of the reaction. One strand (green) has passed through the break formed in the other (red). The two DNA strands have now changed their topology. If they belong to a duplex, the linking number has been increased by 1. If they are separate single-stranded circles, the final product is catenation or decatenation.

there are considerable differences in the two structures, the major difference between the two proteins that could affect the reactions is the presence of the decatenation loop. Its removal leads to loss of decatenation activity but not of relaxation activity (RD, unpublished results). The role of this loop is unclear. One possibility is that it helps select for catenated molecules. It may also provide a second DNA-binding region on the opposite side of the groove and in this way allow interaction with more than one DNA molecule. The positively charged character of the loop (Figure 4) supports the notion that it could be a DNA-binding region.

The last 44 amino acids are disordered in the structure, although it is probable that they become ordered upon DNA binding. This DNA-binding domain is likely to be located distal to the active site and closer to the putative hinge region and the groove. If this is the case, the protein–DNA interactions would extend all along the main body of the protein. For this to happen the DNA has to be oriented with its long axis perpendicular to the central hole of the protein. This face of the protein is clearly involved in extensive contacts with DNA. The other side of the protein could also interact with DNA if the DNA were to wrap around the protein and approach other regions, such as the decatenation loop.

The opening of the protein probably occurs in two steps. The first one involves a smaller separation of the domains to allow ssDNA to enter the active site. This opening may be induced by the entrance of ssDNA into the groove and allows DNA to be positioned and aligned properly in the active site. Once cleavage occurs, with one end of the broken strand covalently attached to domain III and the other noncovalently attached to domain I, a second and larger conformational change occurs. This is important in order to ensure that no futile cleavage/re-ligation occurs and to open a gate in the DNA to allow strand passage. The proposed orientation of the DNA in the groove, with the 3′ end closer to the active site, suggests that the noncovalently attached strand remains in the groove during the topological transformations. The exact nature of the conformational change is unknown, but it is likely to involve a movement of domains II and III away from domains I and IV and be hinged at the two strands connecting these domains (Figures 2,4). Domains II and III are also likely to change relative conformation, moving in a sideways rotation as was observed in the structure of domains II and III of *E. coli* DNA topoisomerase I [30]. As the domains move away, domain III pulls the intact strand(s) towards the central hole. Strand passage occurs as a direct consequence of this movement of the domains. Once the strand(s) are inside the main body of the protein, the conformational change is reversed and the two ends of the broken strand come together. Re-ligation is most likely to occur by the reverse mechanism of cleavage. Expulsion of the DNA from the protein is the last step of the reaction and completes the cycle.

Biological implications

E. coli DNA topoisomerase III is a type IA topoisomerase that can relax negatively supercoiled DNA as well as catenate and decatenate DNA very efficiently. Comparison of the structures of *E. coli* DNA topoisomerases I and III shows that the two proteins are very similar in overall fold, but the differences indicate important elements that may be responsible for their activities. The two domains that form part of the central hole of the protein (domains II and III) are positioned differently in the two proteins, supporting the idea that these domains can move in order to create an opening in the protein and expose the active site. The structure also reveals the presence of a deep groove on the side of the protein that leads to the active site and that may be responsible for ssDNA binding and recognition. The groove was not recognized in the structure of *E. coli* DNA topoisomerase I, but the comparison suggests that it may be present in both proteins. Some differences in the structures may be due to the differences in mechanism and substrate specificity. In particular, *E. coli* DNA topoisomerase III has a 17 amino acid loop insertion that extends from the base of the central hole, which may be involved in catenation and decatenation of DNA. Taken together, all of these observations help to refine the proposed mechanism of action and to identify regions of the protein that may be involved in DNA binding and in the cleavage and re-ligation reactions, although the precise role of many amino acids in the active-site region remains unclear.

The mechanism of action described here explains all of the reactions catalyzed by type IA enzymes. It has interesting parallels with the mechanism of type II enzymes, which also pass strands by a mechanism involving a hole in the protein and large conformational rearrangements. There are still many details of the reaction and of the interactions with DNA that remain to be elucidated, but the overall mechanism for this type of enzyme is now more firmly established with the structure of a second member of this subfamily.

Materials and methods

Crystallization

Intact *E. coli* DNA topoisomerase III was purified as described elsewhere [11]. Crystals were grown by microdialysis against 1.65 M Na, K phosphate (pH 6.3) at 14°C. Typical crystals are hexagonal pyramids with a base of 50–100 μm and a length of 300–500 μm. They belong to space group $P6_122$ with cell constants of 236 Å × 236 Å × 108 Å, hexagonal. There is one molecule in the asymmetric unit, giving a solvent content of over 80%.

Data collection and heavy-atom derivatives

The crystals are small, diffract weakly, and are very radiation sensitive. Hence all characterization and data collection were performed at

Table 2

Data collection and phasing statistics.

	Data set			
	Native I	Native II	$PtCl_4$	EMTS
Data collection				
Detector	CCD/CHESS	MAR IP/SSRL	MAR IP/SSRL	MAR IP/SSRL
Resolution (Å)	3.0	3.5	3.5	3.5
Number of crystals	3	2	2	1
Unique/total reflections	33,563/136,735	22,147/91,488	17,588/36,512	18,874/35,252
Completeness (%)	95.9	97.5	80	84.5
R_{sym}*† (%)	8.0 (26.0)	8.4 (19.8)	10.5 (29.9)	9.6 (26.9)
MFID‡	–	–	19.1	16.2
MIR analysis (50–3.5 Å)				
Number of sites	–	–	5	1
R_{cullis}§				
Acentric			0.87	0.98
Centric			0.92	0.99
Phasing power¶				
Acentric			0.74	0.24
Centric			0.53	0.17

Mean figure of merit = 0.207. Solvent modification (50–3.0 Å): average phase change for acentric reflections 75.3°; 62% of centric phases did not change. *$R_{sym} = \sum |I - \langle I \rangle| / \sum I$, where I = observed intensity, and $\langle I \rangle$ = average intensity obtained from multiple measurements. †Numbers in parentheses correspond to the highest resolution shell. ‡MFID (mean fractional isomorphous difference) = $\sum ||Fph| - |Fp|| / \sum |Fp|$, where $|Fp|$ = protein structure-factor amplitude and $|Fph|$ = heavy-atom derivative structure-factor amplitude. §$R_{cullis} = \sum ||Fh(obs)| - |Fh(calc)|| / \sum |Fh(obs)|$ where $|Fh(obs)|$ = observed heavy-atom structure-factor amplitudes, and $|Fh(calc)|$ = calculated heavy-atom structure-factor amplitude. ¶Phasing power = root mean square ($|Fh| / E$), where $|Fh|$ = heavy atom structure-factor amplitude and E = residual lack of closure error.

synchrotron radiation laboratories (Photon Factory, SSRL and CHESS). The crystals can be cryo-cooled by quickly transferring them into a solution of 25% glycerol, 2.0 M Na, K phosphate (pH 6.3). Cryo-cooling does not fully prevent radiation damage, and the crystals diffract to high resolution for only a few minutes. Typically, each crystal was translated twice to limit radiation damage. Data were integrated with the program DENZO [31], and all subsequent scaling and processing were carried out using programs from the CCP4 [32] suite, unless otherwise noted. Heavy-atom searches were performed by collecting a small region of reciprocal space. Two derivatives were identified: one by soaking for 12 h in 10 mM K_2PtCl_4 and another by soaking for 1 h in 1 mM ethylmercurythiosalicylate. The position of one of the platinum sites was identified using the program RSPS and used to locate all the other sites. Table 2 shows data collection and phasing statistics.

Table 3

Refinement statistics.

Resolution (Å)	10–3.0
Number of reflections:	
Total	32,569
Working set	30,926
Test set	1,643
R factor* (%)	
Working set	24.7
Test set	29.7
Rms bond (Å)	0.013
Rms angle (°)	2.6
Average B factor (Å^2)	
Mainchain	49.3
Sidechain	50.8

*R factor = $\sum ||Fp| - |Fc|| / \sum |Fp|$, where $|Fp|$ = protein structure-factor amplitude and $|Fc|$ = calculated structure-factor amplitude.

Phasing, model building, and refinement

Phases were calculated with the program MLPHARE using the native data and the two derivatives data sets from SSRL. The map calculated from these phases was untraceable, with no distinguishable features. Solvent modification using the program SOLOMON [33] was used to improve the map. The starting envelope was calculated to 6 Å and the resolution was extended slowly to 3.5 Å. The resulting map was greatly improved, and the fold of the protein was clearly recognizable. The resolution was extended using calculated phases for the 3.5–3.0 Å shell and the CHESS native data set. A partial polyalanine model of the structure was built using the bones option of the program O [34]. Several cycles of refinement with X-PLOR [35] and phase combination with SIGMAA helped to extend the model and add all sidechains. At all stages of the refinement the free R factor was monitored to test the validity of the procedure [36]. Once most of the model was built it was further refined using the simulated-annealing protocol [37] in X-PLOR. The refinement was continued using the program REFMAC, with strict stereochemical constraints on the secondary structure elements. Towards the end of the refinement a bulk-solvent correction was applied in REFMAC using the protocol in X-PLOR. Because of the resolution of the data, no water molecules were added and only two temperature factors were assigned per residue, one for the mainchain atoms and one for the sidechain atoms. The final R factor between 10–3.0 Å is 24.5% and the free R factor is 29%. There are 4800 atoms in 603 residues with good stereochemistry. Over 90% of all residues are in allowed regions of the Ramachandran plot and none is in disallowed regions. Table 3 shows the relevant refinement statistics.

Accession numbers

The coordinates for *E. coli* topoisomerase III have been submitted to the PDB under accession code 1D6M.

Acknowledgements

We acknowledge the contributions of A Sharma to the early stages of this project. We thank V Grum, C Lima, A Patera and A Sharma for help with data collection as well as the staff at the Photon Factory, SSRL, and

CHESS. We thank JP Abrahams for an unreleased version of SOLOMON and L Frappier and A Bochkarev for the coordinates of RPA. We thank L Godley, M Gwynn, T Jardetzky, T O'Halloran, A Tackle, J Widom and members of the Mondragón laboratory for comments and suggestions. This work was supported by the NIH (GM51350 to AM and GM48445 to RD).

References

1. Wang, J.C. (1996). DNA Topoisomerases. *Annu. Rev. Biochem.* 65, 635-692.
2. Cheng, C., Kussie, P., Pavletich, N. & Shuman, S. (1998). Conservation of structure and mechanism between eukaryotic topoisomerase I and site-specific recombinases. *Cell* 92, 841-850.
3. Redinbo, M.R., Stewart, L., Kuhn, P., Champoux, J.J. & Hol, W.G. (1998). Crystal structures of human topoisomerase I in covalent and noncovalent complexes with DNA. *Science* 279, 1504-1513.
4. Stewart, L., Redinbo, M.R., Qiu, X., Hol, W.G. & Champoux, J.J. (1998). A model for the mechanism of human topoisomerase I. *Science* 279, 1534-1541.
5. Wallis, J.W., Chrebet, G., Brodsky, G., Rolfe, M. & Rothstein, R. (1989). A hyperrecombination mutation in *S cerevisiae* identifies a novel eukaryotic topoisomerase. *Cell* 58, 409-419.
6. Kikuchi, A., Asai, K., Confalonieri, F., Elie, C., Nadal, M., Bouthier de La Tour, C., Forterre, P. & Duguet, M. (1984). Reverse gyrase: a helicase-like domain and a type I topoisomerase I in the same polypeptide. *Nature* 309, 677-681.
7. Hanai, R., Caron, P. R. & Wang, J. C. (1996). Human TOP3: A single-copy gene encoding DNA topoisomerase III. *Proc. Natl Acad. Sci. USA* 93, 3653-3657.
8. Slesarev, A. I., Stetter, K. O., Lake, J. A., Gellert, M., Krah, R. & Kozyavkin, S. A. (1993). DNA topoisomerase V is a relative of eukaryotic topoisomerase I from a hyperthermophilic prokaryote. *Nature* 364, 735-737.
9. Bergerat, A., de Massy, B., Gadelle, D., Varoutas, P.-C., Nicolas, A. & Forterre, P. (1997). An atypical topoisomerase II from archaea with implications for meiotic recombination. *Nature* 386, 414-417.
10. DiGate, R.J. & Marians, K.J. (1988). Identification of a potent decatenating enzyme from *Escherichia coli*. *J. Biol. Chem.* 263, 13366-13373.
11. Hiasa, H., DiGate, R.J. & Marians, K.J. (1994). Decatenating activity of Escherichia coli DNA gyrase and topoisomerases I and III during oriC and pBR322 DNA replication *in vitro*. *J. Biol. Chem.* 269, 2093-2099.
12. Beran-Steed, R.K. & Tse-Dinh, Y.C. (1989). The carboxy terminal domain of *Escherichia coli* DNA topoisomerase I confers higher affinity to DNA. *Proteins* 6, 249-258.
13. Zhang, H.L. & DiGate, R.J. (1994). The carboxyl-terminal residues of *Escherichia coli* topoisomerase III are involved in substrate binding. *J. Biol. Chem.* 269, 9052-9059.
14. Zumstein, L. & Wang, J. C. (1986). Probing the structural domains and function *in vivo* of *Escherichia coli* DNA topoisomerase I by mutagenesis. *J. Mol. Biol.* 191, 333-340.
15. Zhang, H.L., Malpure, S. & DiGate, R.J. (1995). *Escherichia coli* DNA topoisomerase III is a site-specific DNA binding protein that binds asymmetrically to its cleavage site. *J. Biol. Chem.* 270, 23700-23705.
16. DiGate, R.J. & Marians, K.J. (1992). *Escherichia coli* topoisomerase III-catalyzed cleavage of RNA. *J. Biol. Chem.* 267, 20532-20535.
17. Wang, H., Di Gate, R.J. & Seeman, N.C. (1996). An RNA topoisomerase. *Proc. Natl Acad. Sci. USA* 93, 9477-9482.
18. Lima, C.D., Wang, J.C. & Mondragon, A. (1994). Three-dimensional structure of the 67 K N-terminal fragment of *E. coli* DNA topoisomerase I. *Nature* 367, 138-146.
19. Roca, J. & Wang, J.C. (1994). DNA transport by at type II DNA topoisomerase: evidence in favor of a two gate mechanism. *Cell* 77, 609-616.
20. Berger, J.M., Gamblin, S.J., Harrison, S.C. & Wang, J.C. (1996). Structure and mechanism of DNA topoisomerase II. *Nature* 379, 225-232.
21. Roca, J., Berger, J.M., Harrison, S.C. & Wang, J.C. (1996). DNA transport by a type II topoisomerase: Direct evidence for a two-gate mechanism. *Proc. Natl Acad. Sci. USA* 93, 4057-4062.
22. Lindsley, J.E. & Wang, J.C. (1993). On the coupling between ATP usage and DNA transport by yeast DNA topoisomerase II. *J. Biol. Chem.* 268, 8096-8104.
23. Caron, P.R. & Wang, J.C. (1994): Alignment of primary sequences of DNA topoisomerases. In *Advances in Pharmacology.* (Lui, L.F., ed.), pp 271-297, Boca Raton, Academic Press.
24. Chen, S.J. & Wang, J.C. (1998). Identification of active site residues in Escherichia coli DNA topoisomerase I. *J. Biol. Chem.* 273, 6050-6056.
25. Zhu, C.X., Roche, C.J., Papanicolaou, N., DiPietrantonio, A. & Tse-Dinh, Y.C. (1998). Site-directed mutagenesis of conserved aspartates, glutamates and arginines in the active site region of *Escherichia coli* DNA topoisomerase I. *J. Biol. Chem.* 273, 8783-8789.
26. Beese, L.S. & Steitz, T.A. (1991). Structural basis for the 3′5′ exonuclease activity of *Escherichia coli* DNA polymerase I: a two metal ion mechanism. *EMBO J.* 10, 25-33.
27. Stivers, J.T., Shuman, S. & Mildvan, A.S. (1994). *Vaccinia* DNA topoisomerase I: single-turnover and steady-state kinetic analysis of the DNA strand cleavage and ligation reactions. *Biochemistry* 33, 327-339.
28. Feinberg, H., Changela, A. & Mondragon, A. (1999). Protein-nucleotide interactions in *E. coli* DNA topoisomerase I. *Nat. Struct. Biol.* 6, 961-968.
29. Bochkarev, A., Pfuetzner, R.A., Edward, A.M. & Frappier, L. (1996). Structure of the single-stranded-DNA-binding domain of replication protein A bound to DNA. *Nature* 385, 176-181.
30. Feinberg, H., Lima, C.D. & Mondragon, A. (1999). Conformational changes in *E. coli* DNA topoisomerase I. *Nat. Struct. Biol.* 6, 918-922.
31. Otwinowski, Z. (1993). Oscillation data reduction program. In *Data Collection and Processing*. (Sawyer, L., Isaacs, N. & Bailey, S., eds), pp. 56-62. Daresbury Laboratory, Warrington, UK.
32. Collaborative Computational Project 4 (1994). The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D* 50, 760-763.
33. Abrahams, J.P. & Leslie, A.G.W. (1996). Methods used in the structure determination of bovine mitochondrial F1 ATPase. *Acta Crystallogr. D* 52, 30-42.
34. Jones, T.A., Zou, J.Y., Cowan, S.W. & Kjeldgaard, M. (1991). Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr. A* 47, 110-119.
35. Brünger, A.T. (1992). *X-PLOR, Version 3.1, A System for X-ray Crystallography and NMR*. Yale University Press, New Haven, CT.
36. Brünger, A.T. (1992). The free R value: a novel statistical quantity for assessing the accuracy of crystal structure. *Nature* 355, 472-474.
37. Brünger, A.T., Kuriyan, J. & Karplus, M. (1987). Crystallographic R factor refinement by molecular dynamics. *Science* 235, 458-460.
38. Kraulis, P.J. (1991). MolScript: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallogr.* 24, 946-950.
39. Evans, S.V. (1993). SETOR: Hardware lighted three-dimensional solid model representations of macromolecules. *J. Mol. Graphics* 11, 134-138.
40. Nicholls, A., Sharp, K.A. & Honig, B.H. (1991). Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons. *Proteins* 11, 281-286.