Structure and function of an archaeal topoisomerase VI subunit with homology to the meiotic recombination factor Spo11

Matthew D. Nichols1, Kristen DeAngelis, James L. Keck and James M. Berger2

Department of Molecular and Cell Biology, University of California at Berkeley, 229 Stanley Hall-3206, Berkeley, CA 94720 and
1Department of Biology, Amherst College, Amherst, MA 01002, USA
2Corresponding author
E-mail: jmberger@uclink4.berkeley.edu

In all organisms, type II DNA topoisomerases are essential for untangling chromosomal DNA. We have determined the structure of the DNA-binding core of the Methanococcus jannaschii DNA topoisomerase VI A subunit at 2.0 Å resolution. The overall structure of this subunit is unique, demonstrating that archaeal type II enzymes are distinct from other type II topoisomerases. However, the core structure contains a pair of domains that are also found in type IA and classic type II topoisomerases. Together, these regions may form the basis of a DNA cleavage mechanism shared among these enzymes. The core A subunit is a dimer that contains a deep groove that spans both protomers. The dimer architecture suggests that DNA is bound in the groove, across the A subunit interface, and that the two monomers separate during DNA transport. The A subunit of topoisomerase VI is homologous to the meiotic recombination factor, Spo11, and this structure can serve as a template for probing Spo11 function in eukaryotes.

Keywords: DNA-binding protein/Spo11/topoisomerase

Introduction

Type II DNA topoisomerases are ubiquitous proteins that catalyze the ATP-dependent transport of one DNA duplex through another. During the DNA transport reaction, these enzymes generate transient double strand DNA breaks (DSBs) through tyrosine-mediated covalent attachments to both DNA strands at the 5' positions (for review, see Wang, 1996). The type II topoisomerase reaction is essential for cell viability, as it removes the tangles and knots that arise between stretches of DNA during replication and recombination. Over 20 years of research have been devoted to the study of the biochemical mechanism of these proteins, as well as of their important pharmacological value as chemotherapeutic and antimicrobial targets.

An archaeal type II topoisomerase activity capable of catalyzing ATP-driven relaxation and decatenation of duplex DNA circles was first discovered in Sulfolobus shibatae (Bergerat et al., 1994). The protein is an $A_B_2$ heterotetramer; such an oligomeric composition is also observed in prokaryotic type II topoisomerases. However, initial sequence analysis of the S. shibatae type II topoisomerase genes revealed that the N-terminal third of the B subunit has only sparse homology to the ATP-binding region of mesophilic type II topoisomerases, whereas no sequence homology was apparent to classic type II enzymes in the rest of the protein (Bergerat et al., 1997; Keeney et al., 1997). More recently, a motif present in the A subunit has been proposed to correspond to a fold within the DNA-binding and cleavage core of classic type II topoisomerases (Aravind et al., 1998). The archaeal type II protein is now termed DNA topoisomerase VI (Bergerat et al., 1997), and related enzymes have been found in a number of other archaeal organisms.

Coincident with the discovery of topoisomerase VI in S. shibatae, a new DNA cleavage activity was characterized in the yeast Saccharomyces cerevisiae. This activity generates the DSBs that initiate homologous recombination during meiosis (reviewed in Haber, 1997). Following DSB formation in wild-type cells, additional proteins exonucleolytically resect the 5' DNA termini of the DSBs to generate 3' single-stranded DNA tails. In certain mutant yeast strains, the breaks are not resected, and the primary DNA cleavage species instead accumulates with protein covalently attached to the 5' strand termini on either side of the break, apparently via a phosphotyrosine linkage (de Massy, 1995; Keeney and Kleckner, 1995; Liu et al., 1995). This covalent protein–DNA complex has been purified and the protein identified as Spo11p (Keeney et al., 1997), indicating that Spo11 is the catalytic subunit of the meiotic DNA-cleaving activity. Mutation of an invariant tyrosine in Spo11 to phenylalanine abolishes DSB formation (Bergerat et al., 1997), suggesting that it is likely to be the active site residue. Because Spo11 and the A subunit of the archaeal type II enzyme topoisomerase VI (topo VI-A) are homologous, the topo VI-A subunit is also likely to bind and cleave DNA via a 5' phosphotyrosyl linkage (Bergerat et al., 1997; Keeney et al., 1997).

The observation that the sequence of the topo VI-A is smaller than and not homologous to the DNA-binding region of classic type II topoisomerases raises a number of questions regarding the structure and function of the archaeal protein. For example, is topo VI-A a minimized, diverged version of the topoisomerase II DNA-binding region or does it possess a novel architecture? Which features of the type II topoisomerase mechanism are in common between topoisomerase VI and classic type II topoisomerases? How is one DNA segment transported through another by topoisomerase VI? Does the transported DNA enter and leave from the same side (‘one-gate’) or, as has been proposed for classic type II topoisomerases, from different sides (‘two-gate’) of the enzyme (Roca and Wang, 1994)? Finally, how is a type II DNA cleavage activity used for meiotic recombination?

To understand better the mechanisms of topoisomerase VI and Spo11, as well as their relationship to classic...
type II topoisomerases, we determined the structure of a core fragment of topo VI-A from the archaeon *Methanococcus jannaschii*. This region comprises all the sequence elements conserved between archaeal topo VI-A proteins and eukaryotic Spo11 homologs. The overall shape and fold of the protein is strikingly different from that of the classic type II topoisomerases, showing that the archaeal topoisomerases constitute a new class of type II proteins. However, a DNA-binding and a metal-binding motif were found to be common not only between topo VI-A and classic type II topoisomerases, but with type IA topoisomerases as well. A number of these structural similarities can be extended to certain metal-dependent 5′ to 3′ nucleases such as the T4 RNase H and FEN-1 proteins (Mueser et al., 1996; Hosfield et al., 1998), supporting the proposal that type IA and type II topoisomerases employ a metal-assisted mechanism for DNA cleavage. Finally, the structure of the topo VI-A protein suggests how DNA might be bound by the protein, how a ‘two-gate’ mechanism might be used by topoisomerase VI, and why this type of DSB-forming module might have been adapted for meiotic recombination.

Results

Activity of a proteolytically defined core of topo VI-A

The full-length (42 kDa) *M. jannaschii* topo VI-A subunit was cloned, expressed and purified from *Escherichia coli*. Limited proteolysis studies of the protein revealed a 34 kDa species that was resistant to degradation. Microsequencing identified the N-terminus of this fragment as Thr69; mass spectrometry gave a molecular weight consistent with an untruncated C-terminus (not shown). The region spanning residues 69–369 of the topo VI-A gene subsequently was cloned and expressed, and will be referred to as topo VI-A'. The N-terminal region removed by proteolysis is not conserved among topo VI-A/Spo11 proteins; indeed, the homologous protein from the archaeal organism *Pyrococcus horikoshii* lacks the N-terminal region altogether.

The topo VI-A' protein is a stable dimer as assayed by both gel filtration and analytical ultracentrifugation (not shown), and binds short DNA duplexes as evidenced by gel mobility shift experiments (Figure 1A). A mutation in an acidic residue (Glu197) that is invariant among all topo VI-A/Spo11 family members and lies in a proposed DNA-binding groove (discussed below) diminishes the affinity of the protein for DNA by at least 10- to 20-fold, confirming that this DNA-binding activity is specific. This mutation has no apparent effect on the overall structure of the protein for DNA by at least 10- to 20-fold, and binds short DNA duplexes as evidenced by gel filtration and analytical ultracentrifugation (not shown).

Structure determination of topo VI-A'

Topo VI-A' crystallized in the space group P2₁, with one dimer per asymmetric unit. The structure was solved by multiple isomorphous replacement and refined at 2.0 Å resolution to a free R-value of 24.2% and a working R-factor of 18.8% (Table I). PROCHECK (Laskowski et al., 1993) analysis shows that the molecule has good overall geometry, with no residues falling in either generously allowed or disallowed regions of Ramachandran space. Traceable electron density was observed for all amino acids from 71 to 369, excepting residues 266–273. The topo VI-A' monomer has overall dimensions of 65×48×30 Å and is composed of two distinct domains (residues 69–140 and residues 147–369) that represent the N-terminal third and C-terminal two-thirds of the molecule, respectively (Figure 2A). The two domains are linked by a short tether spanning residues 141–146. A search of the structural database by DALI (Holm and Sander, 1993) revealed no close matches to any known protein. However, two regions of topo VI-A' were found to be similar to small domains or folds observed in other proteins.

The N-terminal domain of topo VI-A' (residues 71–140) contains five α-helices and two β-strands, and displays a fold related to the DNA-binding domain of the *E.coli* catabolite activator protein (CAP) (Schultz et al., 1991). Helices α1, α2 and α4 make up the three-helix bundle characteristic of the CAP domain, and are backed by a
Table I. Multiple isomorphous replacement and refinement statistics

<table>
<thead>
<tr>
<th>Data collection</th>
<th>Native</th>
<th>K₂PtCl₄</th>
<th>Ter-Pt</th>
<th>SmCl₃</th>
<th>Se-met</th>
</tr>
</thead>
<tbody>
<tr>
<td>Resolution (Å)</td>
<td>20–2.0</td>
<td>20–2.1</td>
<td>20–2.1</td>
<td>20–2.7</td>
<td>20–2.6</td>
</tr>
<tr>
<td>Rₚsym (last shell) (%)</td>
<td>3.6 (13.1)</td>
<td>4.5 (12.2)</td>
<td>4.0 (9.2)</td>
<td>8.2 (18.2)</td>
<td>6.8 (15.1)</td>
</tr>
<tr>
<td>Completeness (last shell)</td>
<td>99.7 (99.3)</td>
<td>97.4 (95.0)</td>
<td>99.9 (99.6)</td>
<td>99.1 (98.3)</td>
<td></td>
</tr>
<tr>
<td>Rₚiso b (last shell) (%)</td>
<td>15.6 (18.7)</td>
<td>11.2 (13.3)</td>
<td>10.8 (15.3)</td>
<td>10.9 (15.2)</td>
<td></td>
</tr>
<tr>
<td>No. of sites</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>7</td>
</tr>
<tr>
<td>Phasing power a (last shell)</td>
<td>1.4 (1.2)</td>
<td>0.7 (0.5)</td>
<td>0.7 (0.4)</td>
<td>1.4 (1.2)</td>
<td></td>
</tr>
<tr>
<td>Phasing power anomalous</td>
<td>1.4 (1.1)</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td></td>
</tr>
<tr>
<td>Cullis R d (last shell)</td>
<td>0.82 (0.82)</td>
<td>0.92 (0.91)</td>
<td>0.90 (0.92)</td>
<td>0.80 (0.82)</td>
<td></td>
</tr>
<tr>
<td>Cullis R anomalous</td>
<td>0.84 (0.89)</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td></td>
</tr>
<tr>
<td>Figure of merit (20–2.8Å)</td>
<td>Acentric</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Overall (last shell)</td>
<td>0.53 (0.38)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Refinement

| Resolution (Å)              | 20–2.0 |
| No. of reflections          | 43 761 |
| working                     | 40 108 |
| free (% total)              | 3653 (8%) |
| Rₚwork, Rfree c (%)         | 19.7 (24.2) |
| Sigma cut-off               | 0 |

Structure and stereochemistry

| No. of atoms               | 4979 |
| protein                    | 4648 |
| water                      | 326  |
| ions                       | 5    |
| R.m.s.d. bond lengths (Å)  | 0.006|
| R.m.s.d. bond angles (°)   | 1.4  |

a Rₚsym = Σ|Iᵢ – 〈I〉|/ΣIᵢ, where Iᵢ is the intensity measurement for reflection j and 〈I〉 is the mean intensity for multiply recorded reflections.

b Rₚiso = Σ|Fᵦ – |Fᵦ|/Σ|Fᵦ|, where Fᵦ and Fᵦ are the derivative and native structure factors, respectively.

Phasing power = |Fᵦ|/|Fᵦ|, where (Fᵦ) is the root-mean-square heavy atom structure factor and E is the residual lack of closure error.

c Cullis R = Σ|Fᵦ| ± |Fᵦ| – |Fᵦ| |Fᵦ| |Fᵦ|, where Fᵦ is the calculated heavy atom structure factor.

d Rₚwork, free = Σ|Fᵦ| – |Fᵦ|/|Fᵦ|, where the working and free R-factors are calculated using the working and free reflection sets, respectively. The free reflections (8% of the total) were held aside throughout refinement.

Fig. 2. Structure of the topo VI-A’ monomer. (A) Stereo ribbon diagram showing the overall tertiary structure of the topo VI-A’ monomer. The regions are color-coded as follows: gold, N-terminal domain; green, metal-coordinating Rossmann-like fold of the C-terminal domain; blue, other regions of the C-terminal domain; magenta, linker region. (B) Representative electron density map using observed amplitudes and experimental (flattened) phases showing the region around the Mg²⁺ ion (green sphere). Coordinating acidic groups are shown as ball-and-stick, while water molecules are small red spheres. Invariant acidic groups which coordinate the ion are labeled. Contours are at 1.0σ (orange) and 2.5σ (cyan). (A) was generated by MOLSCRIPT (Kraulis, 1991), (B) by MOLSCRIPT and RASTER-3D (Kraulis, 1991; Merritt and Murphy, 1994).

small, two-stranded β-sheet composed of β1 and β2. Helices α2 and α4 correspond to the two adjacent helices of the CAP helix–turn–helix (HTH) motif, with α4 being analogous to the HTH ‘recognition helix’ that inserts into the major groove of the DNA (Harrison, 1991; Pabo and Sauer, 1992). A short 3₁₀ helix, α₃, resides in the turn of
the HTH motif. The N-terminal domain of topo VI-A’ lacks a third strand normally inserted between β1 and β2 in other CAP-like domains; instead, β2 pairs directly with β1, and is followed immediately by the linker region.

The C-terminal domain of topo VI-A’ can be divided into three smaller subdomains. The first subdomain includes residues 147–190 and is a five-stranded, antiparallel β-sandwich (β3–β7). The second subdomain, consisting of residues 191–284 and 350–362, folds into a four-stranded, parallel β-sheet (β8–β11) sandwiched between two pairs of α-helices (α6, α8, α9 and α15). β3 of the first subdomain packs edgewise against β9 of the second subdomain, forming a continuous β-sheet. The second subdomain belongs to the Rossmann fold family of structures (Branden and Tooze, 1991), found in a number of proteins known to bind nucleotides, phosphorylated amino acids or other small organic molecules. The final subdomain of the C-terminal domain (residues 285–349) is a small α/β hairpin (α10–α14 and β12–β13) that is inserted between β11 and α15 of the second subdomain.

A single, strong electron density peak (>3.5σ in experimental electron density maps) was found nestled within the loops of the Rossmann-like fold, coordinated by side chains and water molecules in an octahedral configuration. This type of liganding geometry, coupled with the close (2.1–2.2 Å) atomic interaction distances and the presence of MgCl₂ in the crystallization conditions, suggested that Mg²⁺ is bound by this subdomain. The ion is coordinated by Glu197 from the β8/α6 junction, and by Asp249 and Asp251, which reside in the β10–α9 loop (Figure 2B); these residues are invariant among all topo VI-A/Spoo1 homologs. Oe2 of Glu197 and Oβ2 of Asp249 directly ligand the Mg²⁺ ion, while both oxygens of Asp251 are hydrogen-bonded to two coordinating water molecules. Two additional water molecules are also observed, completing the octahedral coordination geometry.

Core structural motifs in the topo VI-A’ monomer are shared between type IA and IIA topoisomerases

Following the nomenclature that has been used to distinguish different families of type I topoisomerases as either type IA or type IB, we will refer to classic and archaeal type II proteins as type IIA and type IIB topoisomerases, respectively (Wang, 1996; Liu and Wang, 1999). Other than a small pattern of conserved acidic groups (Aravind et al., 1998), there are no substantial stretches of sequence similarity between the A subunit of type IIB topoisomerases and either type IA or IIA proteins. Not surprisingly, there is also no similarity between their overall tertiary or quaternary structures. However, the N-terminal CAP-like domain and the C-terminal Rossmann fold structure are also observed in both type IA and type IIA topoisomerases (Berger et al., 1998).

The N-terminal CAP-like domain of topo VI-A’ superposes on the S.cerevisiae topoisomerase II (Berger et al., 1996) CAP-like domain with an r.m.s.d. of 2.8 Å for 54 Cα positions (Figure 3A). Similar folds are found in domains III and IV of E.coli DNA topoisomerase I (Lima et al., 1994; Murzin, 1994; Berger et al., 1998). Furthermore, as in yeast topoisomerase II and domain III of E.coli DNA topoisomerase I, the CAP-like region of topo VI-A’ also contains the proposed active site tyrosine (Tyr103) (Berger et al., 1997; Keeney et al., 1997). Evidence that Tyr103 is the active site tyrosine derives from its sequence invariance in all known members of the topo VI-A/Spoo1 family, as well as from mutational studies of Spo11, demonstrating that its corresponding residue, Tyr135, is essential for DSB formation in yeast (Berger et al., 1997). Superposition of the yeast topoisomerase II CAP-like domain and the topo VI-A’ N-terminal domain places the active site tyrosine of these proteins 14 Å from each other (Figure 3A). The active site tyrosine of the archaeal protein is on a different secondary structural element than in the yeast and bacterial enzymes (helix α2 of topo VI-A’ versus the β2–β3 turn of topoisomerases I and II), yet in all cases arginines (Arg99 in topo VI-A’ versus Arg781 in topoisomerase II and Arg321 in topoisomerase I) lie nearby. These arginine residues are also invariant within each topoisomerase family. Moreover, in the case of topoisomerases I and II, the conserved arginine appears to play a role in the topoisomerase mechanism, as alanine mutations at this position impair the ability of these enzymes to relax negatively supercoiled DNA (Caron and Wang, 1994; Chen and Wang, 1998; Liu and Wang, 1998; Zhu et al., 1998). Given the similarity of this region in all topoisomerases that cleave at the 5’ end of DNA, we will refer to this domain as a 5Y-CAP domain.

The central C-terminal subdomain of topo VI-A’, the Rossmann-like fold, superposes with the central region of domain I of E.coli DNA topoisomerase I with an r.m.s.d. of 2.3 Å over 96 Cα positions (Figure 3B and C). The three invariant acidic groups in this region of topo VI-A’ (Glu197, Asp249 and Asp251) align even more precisely (r.m.s.d. of 1.2 Å) with an analogous set of highly conserved residues in topoisomerase I (Glu9, Asp111 and Asp113) (Figure 3B). This same fold, including the invariant acidic groups, furthermore is found in the B’ core region of topoisomerase II (Berger et al., 1998). This Rossmann-like fold has been proposed to coordinate divalent metals in type IA and IIA topoisomerases, specifically Mg²⁺ (Lima et al., 1994; Berger et al., 1998), but no direct structural evidence for this claim currently exists. This region in the topo VI-A’ structure does appear to coordinate a Mg²⁺ ion through the conserved E, DxD trio (Figure 2B).

The Rossmann-like fold in type IA and II topoisomerases consists of a four-stranded, parallel β-sheet of β2β1β3β4 topology sandwiched by α-helices; the conserved acidic groups nestle in the loops of this fold, at the β1–α1 and β3–α3 junctions (Berger et al., 1998). Additional sequence evidence recently has emerged suggesting that this fold, complete with acidic residues, is present in a number of proteins beyond type IA and type II proteins, including DnaG-type primases, OLD family nucleases and RecR proteins (Aravind et al., 1998). This fold has been termed a ‘toprim’ motif. We have examined the structures of a number of other metal-assisted phosphotransfer and phosphodiesterase enzymes outside of these protein families, and found that certain 5’ to 3’ nucleases such as T4 RNase H and FEN-1 (Mueser et al., 1996; Hsiouf et al., 1998), as well as response-regulator proteins such as CheY and NarL (Stock et al., 1993; Baikalov et al., 1996), also exhibit this motif (Figure 3D and E). In these cases, metal ions are often bound.
Fig. 3. Comparison of type IA, IIA and IIB topoisomerase structural motifs. (A) Superposition of CAP-like regions from yeast topoisomerase II (gray) (Berger et al., 1996) and topo VI-A (gold) represented as a ribbon. The respective active site tyrosines and neighboring invariant arginines are shown as ball and stick. (B) Superposition of the Rossmann folds from domain I of E.coli DNA topoisomerase I (Lima et al., 1994) and the central C-terminal subdomain of topo VI-A'. The backbone is drawn as a ribbon and colored gray (topoisomerase I) or green (topo VI-A'). Conserved acidic groups are shown as ball and stick, while the bound Mg$^{2+}$ ion in the topo VI-A' structure is shown as a magenta sphere. (C) Close up of the metal-binding pocket in (A), showing the superposition of the acidic groups. Side chains are colored and labeled in gray or green for topoisomerase I or topo VI-A', respectively. (D) Topology diagram of the core metal-binding fold. Both folds share the same β2β1β3β4 strand order. Helix α4 (dashed, light-gray) is not present in T4 RNase H, and two insertions present in this enzyme between β1 and α1, and β2 and α2 are indicated as triangles labeled IR1 and IR2. The E.coli DNA topoisomerase I domain insertion between β1 and α1 and the yeast topoisomerase II β4–α4 insertion are also indicated as IT1 and IT2, respectively. Additional conserved acidic groups present in certain 5′ to 3′ nucleases (see text) and in type IA and IIA topoisomerases are indicated by open circles and labeled ‘E’ or ‘T’, respectively. (E) Comparison of the core metal-binding fold from T4 RNase H (Mueser et al., 1996) (left) and topo VI-A' (right). Secondary structural elements are colored green (β-strand), cyan (α-helix) and orange (coil). Acidic residues and metals in similar positions between topo VI-A' and T4 RNase H are indicated as ball-and-stick and colored magenta; an additional acidic residue and metal seen in T4 RNase H are colored gray. (A–C and E) generated by Ribbons (Carson, 1991).

by the acidic groups within the loops of the fold, where they have been directly implicated in assisting catalysis of phosphotransfer or hydrolysis. Furthermore, the localization of an HTH DNA-binding element and a metal-binding motif within the same polypeptide chain is not unique to type IA, IIA and IIB topoisomerases: T4 RNase H, FEN-1 and NarL proteins all contain nucleic acid interaction domains that lie adjacent to the metal-binding region (Baikalov et al., 1996; Mueser et al., 1996; Hosfield et al., 1998).

The topo VI-A' dimer
Each asymmetric unit of the crystal contains a 'U'-shaped dimer of topo VI-A' (Figure 4A and B). The dimer has overall dimensions of 65×48×60 Å and contains a channel ~18 Å wide, 18 Å deep and 50 Å long that spans both monomers. The primary dimer contact between the A subunits is formed by a pseudo-continuous β-sheet, constructed from the first subdomain of the C-terminal region of each monomer. This interface contains a small, packed hydrophobic core (residues 163, 165, 170–171, 173, 175, 178 and 184) that is flanked by numerous polar and van der Waals interactions (residues 146–147, 156, 169, 172, 174, 176, 179–180 and 199–200). The primary contact buries ~600 Å$^2$ of surface area per protomer, and generates the central floor of the channel (Figure 4C). A secondary dimer interface is formed by dimer-related N- and C-terminal domains (residues 85, 89, 94–97, 99–100, 102–104 and 106–107 with 198, 200, 203–204, 207–208, 297, 299, 340, 342–343, 345–346 and 349). This interaction builds part of the channel walls, burying ~500 Å$^2$ per protomer between the two regions on each side. The contacts in this secondary region are primarily ionic, and numerous waters are found buried within the interface. Finally, there are a small number of contacts between the linker region and the C-terminal domain of the dimer-related protomer (residues 142–144 with 147, 199, 200, 222 and 342). The overall dimer arrangement of topo VI-A' places Tyr103 from the N-terminal domain of one protomer 9.6 Å from the bound Mg$^{2+}$ ion in the toprim
motif of the other protomer (Figure 4B). This tyrosine is $>30$ Å from the metal-binding pocket of its own protomer, which is located on the opposite wall of the channel.

A comparison of the topo VI-A’ dimer with the known structures of the DNA-binding and cleavage cores of *E.coli* DNA topoisomerase I and yeast topoisomerase II shows that the overall structures are very different (Figure 5). DNA topoisomerase I is a monomeric protein with a large ($>20$ Å) hole in its interior. Topoisomerase II is a dimeric enzyme that contains an even larger interior hole ($>45$ Å). Outside of the sparse sequence patterns manifest in the toprim region (Caron and Wang, 1993; Aravind *et al*., 1998), these proteins show no obvious sequence similarity, either with each other or with topo VI-A. Furthermore, the order and placement of the structurally similar 5Y-CAP and toprim domains within the primary amino acid sequence differs between topoisomerase families (Figure 5B). Despite these differences, however, the architectures of *E.coli* DNA topoisomerase I and *M.jannaschii* topo VI-A’ both position their active site tyrosines close to the conserved acidic pockets of their toprim folds (Figure 5C); there is evidence that this 5Y-CAP/toprim arrangement may occur in type IIA topoisomerases as well (Yoshida *et al*., 1991; Fass *et al*., 1999).

**Discussion**

The topo VI-A’ structure contains key catalytic motifs seen in other topoisomerases that form phosphotyrosyl linkages to the 5’ ends of DNA, yet the overall structure of the protein is unique to the type IIB class of topoisomerases. A close inspection of the similarities and differences between the topo VI-A’ structure and type IA and IIA topoisomerases reveals how to reconcile these issues. In particular, these analyses suggest how topoisomerase VI...
binds, cleaves and transports DNA. These studies also provide a unified framework for considering DNA cleavage by type IA and II topoisomerases, and help answer certain questions regarding the function of the meiotic recombination factor, Spo11.

**DNA binding by topoisomerase VI**

A number of features of the topo VI-A' dimer suggest how the protein may interact with DNA. First, the deep channel formed by the two protomers is the most electrostatically positive region of the dimer, making it a plausible
A conserved DNA cleavage mechanism for type IA, IIA and IIB topoisomerases

In addition to a role in DNA binding, Mg\(^{2+}\) may also assist the topoisomerase VI DNA cleavage and religation reactions. All type IA, IIA and IIB topoisomerases link covalently to the 5’ end of DNA and require metal for full activity. However, for type IA and IIA topoisomerases, the precise role of metal is still debated (Goto et al., 1984; Osheroff, 1987; Domanico and Tse-Dinh, 1991; Chen and Wang, 1998; Zhu et al., 1998; Liu and Wang, 1999). The presence of Mg\(^{2+}\) ions bound within the toprim fold of topo VI-A’ provides the first direct structural evidence that the conserved acidic groups of the toprim fold act as a metal-liganding center and that these residues are therefore not likely to participate directly in general acid–base catalysis. Whether metal is directly involved in the chemistry of phosphotransfer remains to be determined. Nevertheless, the observation that the toprim-type metal-binding domain is found in other nuclease and phosphotransferase families hints that a mode of direct, metal-assisted phosphotransfer may exist in the type IA and II topoisomerases.

Should the toprim fold assist DNA phosphotransfer directly, it would appear to act in trans for the topoisomerase VI cleavage reaction, such that the metal-binding site of one protomer provides part of the catalytic environment to the tyrosine of its dimer mate (Figure 4C). This arrangement has been proposed for type IIA enzymes, in which each toprim domain is thought to associate with the dimer-related 5Y-CAP domain, after the regions are

surface for DNA binding. Second, the groove is of the appropriate size (~18 Å diameter) to accommodate DNA. Third, both the recognition helix and proposed active site tyrosine of the 5Y-CAP region, as well as the Mg\(^{2+}\)-binding pocket of the toprim fold, face the groove and together generate much of the channel wall. Finally, mutation of one of the invariant acidic residues that coordinates Mg\(^{2+}\) (Glu197) dramatically diminishes DNA binding, yet does not affect the overall protein structure (Figure 1). Together, these features support a model in which DNA binds within the central channel of the dimer (Figure 6), and where the metal-binding site is important for this interaction.
brought together by a series of intramolecular hinges (Berger et al., 1998; Liu and Wang, 1999). Type IA proteins likewise can be considered to undergo an analogous ‘pseudo-trans’ reaction. This is because despite the localization of the 5Y-CAP and toprim components on a single polypeptide chain, the tertiary structure of these proteins is such that they are brought together over a distance by a long, flexible extension arm (Lima et al., 1994). The 5Y-CAP and toprim domains are embedded within very different tertiary and quaternary structural contexts between type IA, IIA and IIB proteins (Figure 5), yet their conservation and similar quaternary relationship to each other suggest that a conserved mechanism for phosphotransfer to the tyrosine is utilized by these enzymes.

DNA transport by topoisomerase VI

DNA binding and cleavage are only part of the topoisomerase reaction, since decatenation also requires that one DNA segment be passed through the transient break in another. In type IA and IIA topoisomerases, an ‘enzyme bridge’ model has been proposed for this reaction: the proteins first cleave, then open and span the broken halves of a DNA segment while passing a second segment through the break (Tse et al., 1980; Brown and Cozzarelli, 1981). The broken DNA is anchored to the bridge by both covalent and non-covalent protein–DNA interactions, while the bridge itself is formed by the protein architecture. By contrast, type IB topoisomerases, which form phosphotyrosyl linkages to the 3' end of DNA, require no metal ion for catalysis, and contain neither the 5Y-CAP nor toprim structural elements. Type IB enzymes are related to the integrase family of proteins and their catalytic mechanism does not appear to involve DNA transport (Cheng et al., 1998; Redinbo et al., 1998; Sherrat and Wigley, 1998). For its part, DNA topoisomerase VI has been shown to be capable of passing duplex DNA segments through one another (Bergerat et al., 1994; Buhler et al., 1998). Thus, topoisomerase VI would be expected to function as a enzyme bridge.

The structure of the topo VI-A’ dimer suggests how topoisomerase VI may act through an enzyme-bridging mechanism. Upon DNA binding, if DNA resides in the topo VI-A’ channel as proposed, it would rest directly on the dimer interface. This feature implies that after DNA cleavage and during DNA separation, the dimer interface must be opened simultaneously, until a gap sufficiently wide to allow passage of a second DNA duplex is attained (~20–25 Å). This notion of domain separation has precedents in both type IA and IIA topoisomerases, where protein–protein interfaces are thought to be formed and broken numerous times during the catalytic cycle in order to open the cleaved DNA (Wigley et al., 1991; Lima et al., 1994; Berger et al., 1996). The fact that the topo VI-A’ protein appears to exist as a stable dimer and contains a sizeable primary dimer interface (~600 Å²) is not an impediment to this opening model; for example, DNA topoisomerase II exists as a stable dimer in solution (Kₐ = 10⁻¹¹ M) (Tennyson and Lindsley, 1997), yet the primary dimer interface is ~800 Å² in area and is also thought to separate during catalysis (Roca and Wang, 1994; Roca et al., 1996). Following the opening of the protein–DNA complex, our proposed DNA-binding configuration predicts that as the transport DNA passes through the cleaved DNA, it would move from one side of the A subunit to the other, crossing the dimer interface. This is consistent with a ‘two-gate’ reaction mechanism, in which the passed DNA segment enters and exits from different sides of the enzyme; such a model has been similarly proposed for DNA topoisomerase II (Roca and Wang, 1994; Wang, 1996). In order to prevent dissociation of the enzyme and subsequent formation of free DSBs upon DNA opening, a second factor must be present to hold the assembly together (Figure 6C). It seems plausible that the B subunit of topoisomerase VI provides this role, ‘bridging’ the DNA gap during DNA transport. The ATP-binding regions of the B subunits may coordinate the timing of these events and initiate the conformational changes that are required for the transport reaction (Buhler et al., 1998).

If topoisomerase VI does function as an enzyme bridge, there are several aspects of its reaction mechanism that may differ from either type IA or IIA topoisomerases. For example, the DNA-binding and cleavage cores of both type IA and IIA proteins contain large intramolecular holes that are thought to serve as temporary storage areas for the passed DNA segments during the transport cycle (Lima et al., 1994; Berger et al., 1996; Cabral et al., 1997) (Figure 5B). The analogous binding and cleavage core in topoisomerase VI, the A subunit, lacks such a storage hole. Furthermore, in both type IA and IIA enzymes, the two 5Y-CAP domains can be oriented such that the first helices of the HTH motifs directly abut one another (Lima et al., 1994; Berget et al., 1996; Cabral et al., 1997). For topo VI-A’, this orientation is unlikely, as the tether linking the N-terminal, 5Y-CAP domain to the C-terminal domain is too short to permit such a conformation, and DNA binding to the channel would block direct contact between the topo VI-A' CAP-like domains. Together, these architectural features of the topoisomerase VI A subunit suggest that the specific topoisomerase VI DNA opening and transport reactions are distinct from the mechanisms used by either type IA or IIA topoisomerases.

Spo11

Eukaryotes have been known for some time to use type IIA topoisomerases, rather than topoisomerase VI, for duplex DNA transport reactions and untangling their chromosomes (Wang, 1996). It was therefore surprising to find that homologs of topo VI-A exist in eukaryotes (Bergerat et al., 1997; Keeney et al., 1997; McKim and Hayashi-Hagihara, 1998; Dernburg et al., 1998). The yeast protein, Spo11, does not appear to be a topoisomerase, but is instead part of the meiotic recombination complex, forming the DSBs necessary to initiate recombination (Keeney and Kleckner, 1995; Keeney et al., 1997). Given the sequence similarity between topo VI-A and Spo11, it is likely that the general structures and DNA-binding and cleavage mechanisms of the two proteins are comparable (Figure 7). Consistent with this notion, mutations in the acidic pocket of the Spo11 toprim region that would be expected to perturb Mg²⁺ binding do, in fact, impair or abolish the ability of this protein to generate DSBs in vivo (S. Keeney, personal communication).

During meiotic recombination, Spo11 needs at least
Numbering is for every tenth residue of the sequence. The N-terminal ends of both the topoisomerase VI A (α-helices) are placed over the primary sequence, labeled and colored elements observed in topo VI-A and Spo11 family members; red, invariant within Sequences are colored based on conservation: green, invariant among both topo VI-A and Spo11 family members; red, invariant among different from other topoisomerases, suggesting that topoisomerase VI is distinct from other topoisomerases in how it recognizes and transports DNA. However, key functional motifs within topo VI-A' are shared with type IA and IIA topoisomerases, suggesting that in addition to their common ability to bridge broken DNA ends during DNA transport, all topoisomerases that cleave at the 5' end of DNA may also use a common mechanism for cleaving and religating DNA. Future work is needed to understand the precise interaction of members of the topo VI-A/Spo11 family with DNA or with their respective protein complements. Such studies will in turn help clarify how similar collections of catalytic motifs that perform identical reactions can be exploited in different contexts to diversify a biochemical reaction and provide different biological consequences.

Materials and methods

**Protein purification**

The topo6A gene was amplified by PCR from TIGR/ATCC Microbial Genome Special Collection construct AMBUQ03 and cloned into pET24b (Novagen). Protein was overexpressed in E.coli BL21(DE3) cells containing a plasmid coding for the overexpression of several rare rRNA codons (Kim et al., 1998). Cells were induced with 1 mM isopropyl-β-D-thiogalactopyranoside (IPTG) at A600 = 0.3 and grown for 5 h at room temperature. Cells were then harvested by centrifugation, resuspended in buffer A [50 mM Tris–HCl pH 7.5; 10% glycerol; 1 mM EDTA; 1 mM EGTA; 1 mM pepstatin A; 1 mM leupeptin; 1 mM dithiothreitol (DTT); 1 mM phenylmethylsulfonyl fluoride (PMSF)] + 50 mM KCl, and frozen dropwise into liquid nitrogen. Purification of topo VI-A was performed as follows: cells were thawed, sonicated and centrifuged. The clarified lysate was placed at 60°C for 10 min, on ice for 15 min, then recentrifuged. The supernatant was adsorbed to a 30 ml phosphocellulose column (Whatmann) and eluted using a 50–500 mM KCl gradient in buffer A. Peak fractions were determined by SDS–PAGE, pooled and concentrated to 2 ml by ultrafiltration (Amicon, Centriprep-30). The concentrated protein was then passed over a Sephacyrl S300 gel-filtration column (Pharmacia) equilibrated in buffer A + 300 mM KCl. Peak fractions were pooled and concentrated to 15–20 mg/ml [as measured by UV absorbance (Edelhoch, 1967)]. Purity was judged to be >98% by Coomassie-stained SDS–PAGE. Limited proteolytic mapping using trypsin and chymotrypsin identified a 34 kDa, stable, protease-resistant fragment; subsequent protein sequencing identified the N-terminal end of this species as residue Thr67 (trypsin) or Thr69 (chymotrypsin).

The coding sequence including residues 69–369 of M.jannaschii topo VI-A subunit was prepared as described (V an Duyne et al., 1998). Cells were induced with 1 mM isopropyl-β-D-thiogalactopyranoside (IPTG) at A600 = 0.3 and grown for 5 h at room temperature. Cells were then harvested by centrifugation, resuspended in buffer A + 300 mM KCl. Peak fractions were pooled and concentrated to 15–20 mg/ml [as measured by UV absorbance (Edelhoch, 1967)]. Purity was judged to be >98% by Coomassie-stained SDS–PAGE. Limited proteolytic mapping using trypsin and chymotrypsin identified a 34 kDa, stable, protease-resistant fragment; subsequent protein sequencing identified the N-terminal end of this species as residue Thr67 (trypsin) or Thr69 (chymotrypsin).

For crystallization trials, purified topo VI-A protein was dialyzed overnight against 10 mM Tris–HCl, pH 8.5; 200 mM NaCl, and the protein concentration adjusted to 10 mg/ml. Topo VI-A was crystallized by hanging drop vapor diffusion. The initial well solution consisted of 20–25% 2-methyl-2,4-pentanediol (MPD), 13–16% PEG3000; 400–425 mM NaCl; 160–200 mM MgCl2; 50 mM sodium cacodylate–HCl, pH 6.1. Protein and well solutions were mixed 1:1. Crystals (trapezoidal prisms) grew within 12–36 h at 20°C. Crystals were looped directly from the drops and flash-cooled in a gaseous nitrogen stream at –160°C. Derivatives other than the selenomethionine protein were prepared by soaking crystals overnight in well solution with heavy metal at either 1 mM [K2PtCl4, dichloro(2,2'-6',2'-terpyridine)platinum(II) (Ter-Pt)] or 2 mM (SnCl2). All data sets except the Ter-Pt were collected on a RaxisIIe detector using a Rigaku RU-200 rotating anode at a wavelength of 1.54 Å. The Ter-Pt set was collected using a RaxisIV on a Rigaku RU-300. Topo VI-A crystallizes in the space group P21, with unit cell dimensions a = 66.51 Å, b = 59.04 Å, c = 87.26 Å, α = γ = 90.0° and β = 94.05°, and one dimer per asymmetric unit.

For crystallization trials, purified topo VI-A’ protein was dialyzed overnight against 10 mM Tris–HCl, pH 8.5; 200 mM NaCl, and the protein concentration adjusted to 10 mg/ml. Topo VI-A’ was crystallized by hanging drop vapor diffusion. The initial well solution consisted of 20–25% 2-methyl-2,4-pentanediol (MPD), 13–16% PEG3000; 400–425 mM NaCl; 160–200 mM MgCl2; 50 mM sodium cacodylate–HCl, pH 6.1. Protein and well solutions were mixed 1:1. Crystals (trapezoidal prisms) grew within 12–36 h at 20°C. Crystals were looped directly from the drops and flash-cooled in a gaseous nitrogen stream at –160°C. Derivatives other than the selenomethionine protein were prepared by soaking crystals overnight in well solution with heavy metal at either 1 mM [K2PtCl4, dichloro(2,2'-6',2'-terpyridine)platinum(II) (Ter-Pt)] or 2 mM (SnCl2). All data sets except the Ter-Pt were collected on a RaxisIIe detector using a Rigaku RU-200 rotating anode at a wavelength of 1.54 Å. The Ter-Pt set was collected using a RaxisIV on a Rigaku RU-300. Topo VI-A crystallizes in the space group P21, with unit cell dimensions a = 66.51 Å, b = 59.04 Å, c = 87.26 Å, α = γ = 90.0° and β = 94.05°, and one dimer per asymmetric unit.

Data for all crystals were indexed and reduced with DENZO/SCALEPACK (Owinowski and Minor, 1997). The CCP4 set of programs...
was used for truncating and scaling native and derivative data sets (Collaborative Computational Project Number 4, 1994). Phase calculation and density modification were carried out with the program REVEAL (Lamzin and Wilson, 1993), respectively (Abrahams and Leslie, 1996; Fortelle and Bricogne, 1997). Model building was done with O (Jones and Kjeldgaard, 1997). All but 12 residues of the topo VI-A sequence (amino acids 69–70 and 266–273) were traceable into the original density-modified maps; the final structure comprises amino acids 71–265 and 274–369 of the native topo VI-A sequence. Model refinement was carried out using a Refmac procedure (Lamzin and Wilson, 1993; Murshudov et al., 1997); non-crystallographic symmetry (NCS) and phase restraints were used until the free R-factor reached 26.1%. The molecule was rebuilt using 2Fo–Fc maps, and subsequently refined with Refmac using NCS restraints until the free R-factor converged at 24.2% with a working R-factor of 19.7%. Geometry was monitored throughout using PROCHECK (Laskowski et al., 1993), with all residues of the final structure falling within either the most-favored regions (93.4%) or the additionally allowed regions (6.6%). For modeling studies, a 24 bp B-DNA duplex was placed in the groove of topo VI-A, such that the pseudo-dyad of the DNA was coincident with the 2-fold axis of the protein dimer. No direct steric clashes were observed, except for a few side chains, which can be readily moved away using preferred rototmers.

**Gel mobility shift and circular dichroism experiments**

To assist purification of mutant and wild-type proteins, and to minimize artifacts that might arise from different protein preparations, we altered the protein purification procedure as follows: the wild-type topo VI-A′-coding sequence was subcloned from pET24b into pET28b (Novagen) to create a protein fused-in-frame with a hexa-histidine N-terminal tag. The Ghu197Ala topo VI-A′:cepET24b construct using the Quik-change mutagenesis kit (Stratagene), verified by restriction digest mapping and DNA sequencing, and also subcloned into pET28b. Cells containing the his-tagged constructs were grown, induced and harvested as described for the untagged mutant and wild-type tagged proteins were lysed by sonication and concentrated to 3–5 mg/ml. Aliquots of the purified proteins were with nickel–agarose resin, but this time the flowthrough was collected 10% glycerol, and cleaving with thrombin (1 U/mg of protein, 25°C, 1 hour), histidine tags were removed by diluting the proteins to 1.0 mg/ml with 25 mM Tris–HCl, pH 8.0; 250 mM glycine; 5 mM MgCl2. Gels were run at 200 V for 2 h at 20°C, dried on filter paper (Whatmann) and exposed onto phosphoimaging plates (Bio-Rad). The secondary structure contents of both wild-type and mutant proteins were measured using circular dichroism on an A VIV 62DS spectrometer with a Peltier temperature-controlled sample holder in a 0.1 cm path length cuvette at 4°C. The samples contained 2 mg/ml protein (55 µM) in 20 mM Tris–HCl, pH 8.0, 100 mM NaCl.

**Acknowledgements**

The authors would like to thank Deborah Fass, Scott Keeney, Rachel Fennell-Fezzio and Dan Minor for critical reading of the manuscript and helpful discussions. J.L.K. is a Fellow of the Jane Coffin Childs Memorial Fund for Medical Research. J.M.B. is grateful to the Charlotte Geyer Foundation, which funded the work described in this paper. We would also like to thank J. Leipe, D. Aravind, L. Aravind and E. Koonin for helpful discussions. J.L.K. is a Fellow of the Jane Coffin Childs Memorial Fund for Medical Research. J.M.B. is grateful to the Charlotte Geyer Foundation, which funded the work described in this paper. We would also like to thank J. Leipe, D. Aravind, L. Aravind and E. Koonin for helpful discussions.

References


Kraulis,P. (1991) MOLSCRIPT: a program to produce both detailed photorealistic molecular graphics.


Received July 19, 1999; revised and accepted September 13, 1999.