Novel Insights into Catalytic Mechanism from a Crystal Structure of Human Topoisomerase I in Complex with DNA†

Matthew R. Redinbo,*,‡,§ James J. Champoux,|| and Wim G. J. Hol*,‡,⊥

Department of Biological Structure and Biomolecular Structure Center, Department of Microbiology, and Department of Biochemistry and Howard Hughes Medical Institute, University of Washington School of Medicine, Seattle, Washington 98195

Received November 23, 1999; Revised Manuscript Received March 13, 2000

ABSTRACT: Human topoisomerase I helps to control the level of DNA supercoiling in cells and is vital for numerous DNA metabolic events, including replication, transcription, and recombination. The 2.6 Å crystal structure of human topoisomerase I in noncovalent complex with a DNA duplex containing a cytosine at the −1 position of the scissile strand rather than the favored thymine is reported. The hydrogen bond between the O2 position of this −1 base and the ϵ-amino of the conserved Lys-532 residue, the only base-specific contact observed previously in the human topoisomerase I−DNA interaction, is maintained in this complex. Several unique features of this structure, however, have implications for the DNA-binding and active-site mechanisms of the enzyme. First, the ends of the DNA duplex were observed to shift by up to 5.4 Å perpendicular to the DNA helical axis relative to structures reported previously, suggesting a novel degree of plasticity in the interaction between human topoisomerase I and its DNA substrate. Second, 12 additional residues at the NH_2 terminus of the protein (Trp-203−Gly-214) could be built in this structure, and they were found to pack against the putative hinge region implicated in the clamping of the enzyme around duplex DNA. Third, a water molecule was observed adjacent to the scissile phosphate and the active-site residues; the potential specific base character of this solvent molecule in the active-site mechanism of the enzyme is discussed. Fourth, the scissile phosphate group was found to be rotated by 75°, bringing Lys-532 into hydrogen-bonding distance of one of the nonbridging phosphate oxygens. This orientation of the scissile phosphate group implicates Lys-532 as a fifth active-site residue, and also mimics the orientation observed for the 3′-phosphotyrosine linkage in the covalent human topoisomerase I−DNA complex structure. The implications of these structural features for the mechanism of the enzyme are discussed, including the potential requirement for a rotation of the scissile phosphate group during DNA strand cleavage and covalent attachment.

DNA topoisomerases are a ubiquitous class of enzymes that solve the DNA topological problems arising from numerous nuclear processes, including DNA replication, transcription, recombination, and chromatin remodeling (*1*). All topoisomerases utilize a catalytic tyrosine residue to cleave one strand of DNA and form a transient phosphotyrosine bond with one end of the nicked strand. Topoisomerases are classified as type I and type II on the basis of their distinct differences in sequence and function. Type II enzymes are dimeric, break both strands of a duplex DNA (using catalytic tyrosine residues from each monomer), and pass an intact DNA duplex through this transient double-stranded break in an ATP-dependent manner. The type I topoisomerases, in contrast, are monomeric, require no additional energy cofactor, and manage DNA superhelical tension in cells by cleaving one strand of duplex DNA. Type I enzymes are further classified as type IA and type IB, which share no sequence similarity (*2*). While type IA topoisomerases were initially characterized in prokaryotes, homologues have now been identified in several eukaryotes, including humans, and termed topoisomerases III (*1*). Type IA topoisomerases are able to relax only negatively supercoiled DNA, require magnesium and a single-stranded stretch of DNA for function, and become transiently covalently attached to the 5′-end of the nicked single strand. Topoisomerases IB, however, are able to relax both positively and negatively supercoiled DNA with equal efficiency, do not require a single-stranded region of DNA or metal ions for function, and form a transient 3′-phosphotyrosine bond with the nicked DNA strand. Type IB enzymes have only been identified in eukaryotic cells and vaccinia virus to date. With the structural characterizations of fragments of *Es-*

† Supported by a Burroughs Wellcome Fund Career Award in the Biomedical Sciences (M.R.R.), by Grant GM49156 from the NIH (J.J.C.), by a major equipment grant from the Murdock Charitable Trust to the Biomolecular Structure Center at the University of Washington, and by Grant NCI CA65656 from the NIH (W.G.J.H.).

* To whom correspondence should be addressed. M.R.R.: Department of Chemistry, CB#3290, University of North Carolina, Chapel Hill, NC 27599-3290; e-mail, redinbo@unc.edu; phone, (919) 843-8910. W.G.J.H.: Biomolecular Structure Center, Box 357742, University of Washington, Seattle, WA 98195-7742; e-mail, hol@gouda.bmsc.washington.edu; phone, (206) 685-7044.

‡ Department of Biological Structure and Biomolecular Structure Center, University of Washington School of Medicine.

§ Present address: Department of Chemistry, CB#3290, University of North Carolina, Chapel Hill, NC 27599-3290.

|| Department of Microbiology, University of Washington School of Medicine.

⊥ Department of Biochemistry and Howard Hughes Medical Institute, University of Washington School of Medicine.

10.1021/bi992690t CCC: $19.00
Published on Web 05/19/2000

cherichia coli topoisomerase I (a type IA enzyme; *3*, *4*), yeast topoisomerase I (*5*), and vaccinia topoisomerase (*6*), and the more recently determined structures of human topoisomerase I in complex with DNA (*7*−*9*) and of a catalytic fragment of vaccinia topoisomerase (*10*), it is clear that the type IA and IB topoisomerases share no structural similarity (*11*).

Human topoisomerase I is an ∼91 kDa type IB enzyme that was shown by limited proteolysis to be composed of four major domains: a highly charged 24 kDa NH_2-terminal domain, a positively charged 56 kDa core domain, a 7 kDa linker domain, and a 6 kDa COOH-terminal domain which contains the catalytic Tyr-723 (*12*). The NH_2-terminal domain is thought to be only partially structured in the expressed enzyme and has been shown to be dispensable for catalytic activity in vitro (*13*). This region of the enzyme, however, appears to physically associate with numerous key nuclear proteins, including p53 (*14*, *15*), TATA-binding protein and TATA-associated factors of TFIID (*16*−*18*), SV-40 large T-antigen (*19*−*21*), and nucleolin (*22*), and may play a role in the targeting of topoisomerase I to active replication and transcription complexes within the nucleus. Recently determined crystal structures of human topoisomerase I in complex with DNA have revealed a bi-lobed enzyme that completely wraps around its DNA substrate (*7*−*9*). The enzyme makes numerous protein−DNA contacts, and the vast majority of these contacts (37 of 41) are protein to DNA phosphate interactions. Only one base-specific interaction was observed, and that was between the ϵ-amino group of the conserved Lys-532 and O2 of the thymine base in the scissile strand (the strand cleaved by the enzyme) upstream from the cleavage site (i.e., −1 position). Joining the catalytic Tyr-723 around the scissile phosphate group were two arginines (Arg-488 and Arg-590) and a potentially positively charged histidine (His-632); all of these residues are conserved in the topoisomerases IB with known sequences. An active-site mechanism was proposed (*8*) in which Tyr-723 attacks the phosphorus atom of the scissile phosphate group; the positively charged character of Arg-488, Arg-590, and His-632 was proposed to stabilize the transition state of the scissile phosphate, with His-632 perhaps also functioning as a general acid to protonate the leaving 5′-hydroxyl group. No amino acid side chain was found adjacent to the active site that could serve as a general base in the activation of the essential tyrosine hydroxyl. A "controlled rotation" mechanism was also proposed for the relaxation of superhelical tension by the enzyme (*8*).

Been et al. (*23*) analyzed DNA sites effectively cleaved by eukaryotic topoisomerases I, and identified a preferred consensus sequence of (A/T)(G/C)(T/A)**T**∧ for positions −4 (A/T) to −1 (**T**) in the scissile strand adjacent to the cleavage site (∧). The strongest observed preference was for a thymine base (**T** in the above sequence) adjacent to the cleavage site. Anderson et al. (*24*) also described a high-affinity eukaryotic topoisomerase I binding site from the rDNA of *Tetrahymena*. This sequence contained nucleotides ACTT∧ at positions −4 to −1 in the scissile strand adjacent to the cleavage site, which is in agreement with the preferences identified by Been et al. (*23*). The single base-specific interaction identified between human topoisomerase I and its DNA substrate occurs between O2 of the highly preferred thymine base at the −1 position and Lys-532 (*7*). Cytosine, which was found to be the next most preferred nucleotide base at the −1 position, also contains an O2 pyrimidine oxygen. To examine further the nature and potential variability of human topoisomerase I−DNA interactions, we determined the crystal structure to 2.6 Å resolution of a 70 kDa form of the enzyme in noncovalent complex with a 22 bp DNA oligonucleotide containing a cytosine base at the −1 position of the scissile strand. The results from this structural analysis reveal a novel orientation of the active-site region of the protein−DNA complex and provide new insights into the active-site chemistry of the enzyme.

MATERIALS AND METHODS

A 70 kDa NH_2-terminally truncated form of human topoisomerase I (comprising residues 175−765) containing a catalytic Tyr-723 to Phe mutation was expressed using a baculovirus/insect cell system and purified as described previously (*13*). DNA oligonucleotides were synthesized and purified by reverse phase chromatography by DNA Express (Fort Collins, CO). The following 22 bp duplex DNA oligonucleotide was used in this structure determination:

```
 -10        -5       -1 +1       +5            +12
5′-A A A A A G A C T C^ A G A A A A A T T T T T-3′ scissile
3′-T T T T T C T G A G  T C T T T T T A A A A A-5′ intact
```

The ∧ indicates the position of the cleavage site in the scissile strand; **C** and **G** indicate the changes at the −1 position of the scissile and intact strands, respectively, relative to the high-affinity binding site containing a thymine at the −1 position of the scissile strand (*23*, *24*) used in previous structural studies of topoisomerase I in complex with DNA. Additional oligonucleotides containing G and A at their scissile −1 positions were also synthesized (along with their complementary strands), but these duplex oligonucleotides failed to yield crystals.

Crystals in space group $P2_1$ containing both DNA and protein were grown and cryoprotected as described previously (*8*, *9*). Briefly, crystals were grown over the course of 1 week in sitting drops at 22 °C by mixing 6 μL of crystallant [5 mM Tris-HCl (pH 6.0), 20 mM MES-HCl (pH 6.8), 27% PEG-400, 145 mM $MgCl_2$, and 30 mM dithiothreitol], 2 μL of water, 2 μL of duplex oligonucleotide (0.1 mM in 6 mM NaCl), and 4 μL of protein [5 mg/mL in 10 mM Tris-HCl (pH 7.5), 1 mM EDTA, and 5 mM dithiothreitol]. Crystal size was improved by macroseeding every 4 days into freshly prepared drops of the same composition, which produced diffraction-quality crystals after a minimum of four transfers. Crystals were cryoprotected by soaking in a solution of 20% ethylene glycol and 80% crystallant for 5 min prior to cooling in either liquid nitrogen or a nitrogen gas stream at 100 K. Data to 2.6 Å resolution were collected from a single crystal at 100 K at the Advanced Photon Source, beamline 19-ID, using a CCD detector and incident radiation with a wavelength of 1.07 Å. These data were processed and reduced using the HKL-2000 package (Table 1; *25*).

The structure of this protein−DNA complex was determined by molecular replacement (*26*) using the crystal structure of a 70 kDa form of human topoisomerase I in noncovalent complex with DNA (*8*, *9*) as a search model. Rigid-body, positional, and individual *B*-factor refinement was carried out with X-PLOR (*27*), including a bulk solvent

Table 1: Crystallographic Statistics

resolution (Å; highest shell)	15−2.6 (2.72−2.60)
space group	$P2_1$
unit cell dimensions	$a = 56.98$ Å, $b = 124.92$ Å, $c = 72.29$ Å, $\beta = 93.84°$
no. of total reflections	105936
no. of unique reflections	29834
mean redundancy	3.6
R_{sym} (highest shell)	0.068 (0.288)
completeness (%) (highest shell)	96.8 (87.0)
mean I/σ (highest shell)	21.2 (4.2)
R_{cryst} (highest shell)	0.214 (0.320)
R_{free} (highest shell)	0.281 (0.366)
rmsd	
bond lengths (Å)	0.015
bond angles (deg)	1.74
dihedrals (deg)	22.1
impropers (deg)	1.58
no. of protein atoms	3966 (residues 203−633 and 714−765)
mean *B*, protein atoms (Å^2)	53.6
no. of DNA atoms	896
mean *B*, DNA atoms (Å^2)	51.5
no. of solvent atoms	353
mean *B*, solvent atoms (Å^2)	66.2

model and an overall anisotropic *B*-factor correction. Manual adjustments to the model were performed in conjunction with refinement using the program O (*28*). R_{free} was used as a cross-validation tool at all times during refinement (*29*); as such, 7% of the data were set aside for this purpose prior to any structural refinement. The Wilson *B*-factor for the data was 56.2 Å^2. Novel regions of the structure (particularly the deoxyribonucleotides at the −1 position, residues 203−214, and the ends of the DNA oligonucleotides) were positioned into σ_A-weighted difference maps (*30*). The linker domain region of the molecule (Pro-636−Lys-712) exhibited individual *B*-factors in excess of 80 Å^2 (and frequently greater than 120 Å^2) and exhibited no interpretable electron density in $2F_{obs} - F_{calc}$ maps; thus, the linker domain was considered disordered and is not present in the completed structure. At later stages of refinement, 353 solvent sites were added to the model. The final model exhibited good geometry (Table 1) and no Ramachandran outliers as assessed using the program PROCHECK (*31*).

Coordinate superpositions and the calculations of root-mean-square deviations were performed using X-PLOR (*27*). Figures were created using MOLSCRIPT (*32*), Bobscript (*33*), and Raster3D (*34, 35*). Coordinates have been deposited at the RCSB Protein Data Bank (entry 1EJ9).

RESULTS

Overall Structure of the Complex. To examine further the nature of the interaction between human topoisomerase I and its DNA substrate, we pursued the structural characterizations of a 70 kDa form of the enzyme in noncovalent complexes with DNA oligonucleotides containing either guanine, adenine, or cytosine at the −1 position in the scissile strand rather than the favored thymine. It had been shown previously that thymine is preferred approximately 10-fold over the other three nucleotide bases in this scissile strand position (*23*). In addition, several crystal structures of human topoisomerase I in complex with DNA revealed that the side chain of Lys-532 forms a hydrogen bond with pyrimidine O2 on the thymine base at this position, and that this is the only base-specific interaction observed between human topoisomerase I and DNA (*7*−*9*). Crystals of human topoisomerase I DNA complexes were obtained only in the presence of the duplex oligonucleotide containing a cytosine at the −1 position in the scissile strand, which is the only other base with an oxygen in the same position as in thymine. Crystallization trials in the presence of DNA duplexes containing either guanine or adenine at this position yielded only amorphous precipitates.

The 2.6 Å crystal structure of the noncovalent complex between human topoisomerase I and a 22 bp duplex DNA oligonucleotide with cytosine at the −1 position of the scissile strand was refined using X-PLOR with a crystallographic *R*-factor of 0.214 and a cross-validating R_{free} of 0.281 (see Table 1; *27*, *29*). This represents the highest-resolution and best refined structure of a 70 kDa form of human topoisomerase I in noncovalent complex with DNA to date, exceeding the 2.8 Å structure with a final *R*-factor and an R_{free} of 0.224 and 0.312, respectively, reported previously (*8, 9*).

The structure of the complex reported here (which we will call the "minus 1C" structure) is quite similar overall to the structures of human topoisomerase I−DNA complexes determined previously[1] (Figure 1; *7*−*9*). The 22 bp duplex DNA oligonucleotide still adopts an essentially B-form conformation, and the enzyme contacts the DNA over the central 10 base pairs (from position −5 to +4; see Materials and Methods for base numbering) surrounding the cleavage site. The overall fold of the protein is conserved, with the core and COOH-terminal domains lining up remarkably well with the form 1[1] structure, as shown in Figure 1. In addition, similar to a subset of several form 1 complexes described previously (*9*), the linker domain region (Pro-636−Lys-712) of the minus 1C structure was found to be disordered and is not present in the final model.

The root-mean-square deviations (rmsds) between the minus 1C structure and the other structures of human topoisomerase I in complex with DNA reported previously (*7, 8*) are between 1.2 and 1.5 Å when all equivalent nonsolvent atoms are superimposed (Table 2). The protein regions of the complexes align even more closely together, exhibiting rmsds of only 1.0−1.3 Å over all equivalent protein atoms and only 0.7−1.2 Å over equivalent Cα positions. When the core subdomain III regions of the minus 1C and a minus 1T structure are superimposed as in Figure 1, the "cap" regions (core subdomains I and II as shown in blue and yellow, respectively, for the minus 1C structure) of each molecule deviate little in structure; they are related by a rigid-body rotation of only 1.6° about an axis close to vertical when the complex is oriented as in Figure 1. The DNA oligonucleotide regions of the molecule, however, deviate more significantly in structure. The rmsds over all equivalent DNA atoms are 2.1−2.3 Å, and over phosphorus atoms are 2.1−2.4 Å (Table 2). As can be seen in Figure 1, the ends of the DNA oligonucleotide have shifted relative

[1] Previous structures all contained a minus 1T oligonucleotide, and included a 70 kDa form of the enzyme in a noncovalent complex with DNA (form 1; *8, 9*), and two reconstituted forms of the enzyme (*7*), one in a noncovalent complex (form 2) and one in a covalent complex (form 6) with DNA. Reconstituted human topoisomerase I is the noncovalent complex between the 58 kDa core and the 6.3 kDa COOH-terminal domains.

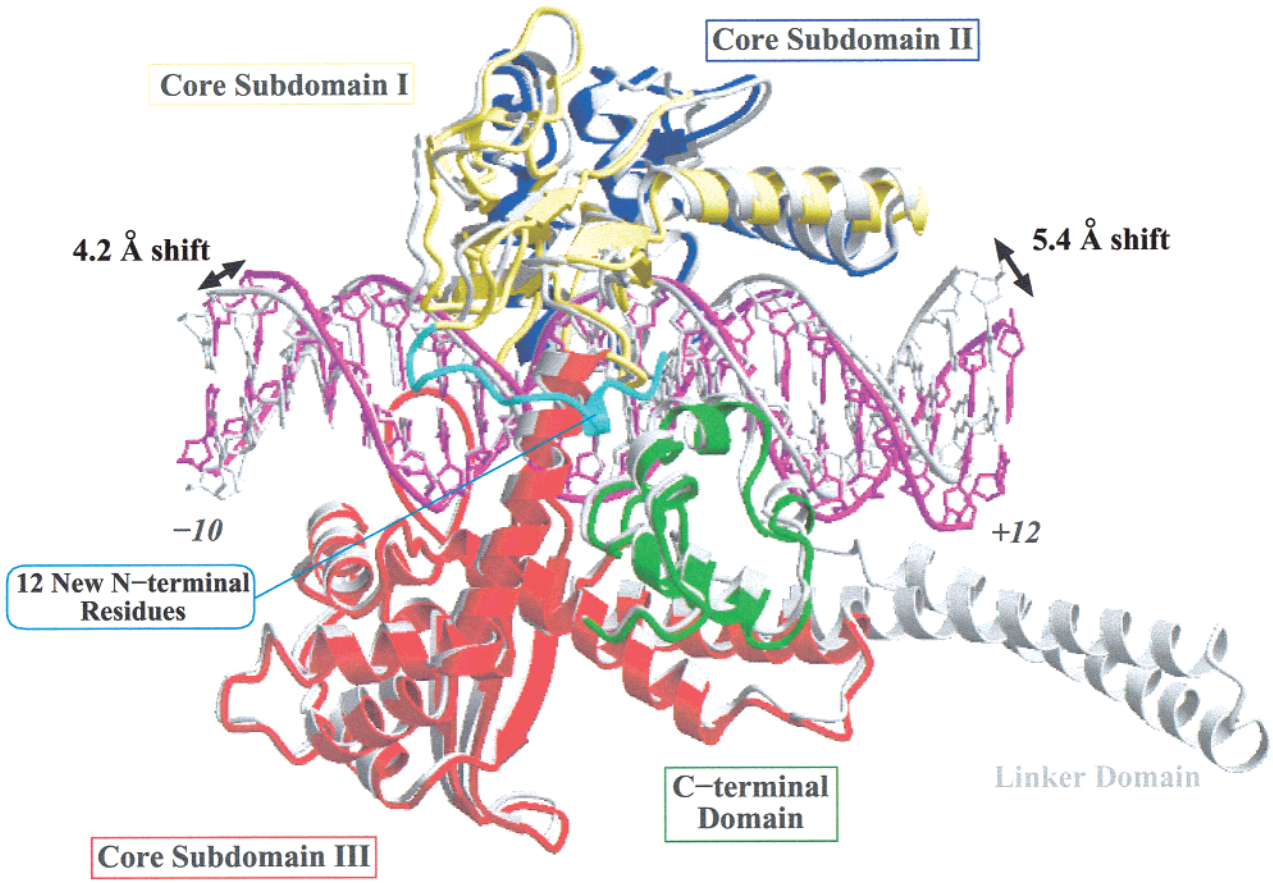

FIGURE 1: Overview of the minus 1C structure reported here, superimposed on the structure of a 70 kDa form of human topoisomerase I in a noncovalent complex with DNA reported previously (minus 1T form 1, rendered in gray; *8*, *9*). The minus 1C structure is rendered with core subdomains I−III in yellow, blue, and red, respectively, the COOH-terminal domain in green, and the 22 bp DNA oligonucleotide in magenta. The 12 NH_2-terminal residues novel to this structure are shown in cyan. The structures were superimposed using the 200 Cα positions of core subdomain III. Note that the linker domain was disordered in the minus 1C structure and is therefore not present in the model.

to that in the form 1 DNA duplex. The observed shifts were perpendicular to the DNA helical axis and were up to 4.2 Å in magnitude upstream of the cleavage site (at the −10 phosphate), and up to 5.4 Å downstream of the cleavage site (at the +12 phosphate). These are the largest shifts in the DNA position between any human topoisomerase I−DNA complex structures observed to date; the maximum shift observed previously was 1.8 Å between two minus 1T structures (*9*).

Because it was unclear whether this alteration in the DNA region of the structure was due to a rigid-body shift or an induced bend in the DNA duplex, we superimposed the DNA oligonucleotides of each structure using only their 42 equivalent phosphorus positions. The results of this superposition are shown in Figure 2. After a rigid-body rotation of 4.5° about an axis centered between the −1 and −2 base pairs (see Materials and Methods), the minus 1T DNA duplex superimposes onto that of the minus 1C structure with an rmsd of 2.4 Å (Table 2 and Figure 2, bottom). The helical axes of each duplex [the central lines in each DNA duplex in Figure 2, calculated using NUPARM (*36*)] appear to align well after superposition, as the distance between the upstream and downstream regions of the oligonucleotide decreased from 4.5 and 2.7 Å, respectively, to 1.8 and 0.9 Å. The phosphate groups at the ends of the DNA oligonucleotide, however, align less well, indicating that a certain degree of flexibility in the DNA is allowed in the regions outside that which is contacted by the protein. Therefore, it appears that the DNA duplex in the minus 1C structure is shifted relative to the minus 1T structure largely by a rigid-body rotation, but that some level of flexibility in DNA structure is accommodated by the protein−DNA complex away from the DNA-binding site.

Additional Residues at the NH_2 Terminus. Compared to the minus 1T structures, 12 additional residues (Trp-203−Gly-214) at the NH_2 terminus of the molecule could be fitted into σ_A-weighted difference maps and were refined in the minus 1C structure (Figure 3). This extends the NH_2 terminus of the structurally known region of human topoisomerase I from Ile-215 to Trp-203. Although the protein constructs used in all the structure characterizations of human topoisomerase I begin at Lys-175, interpretable density at the NH_2 terminus of the molecule had never been extended back beyond Ile-215 (*7*−*9*). It is not clear what is unique about the current structure that allowed us to extend the model by these 12 residues at the NH_2 terminus. Trp-203 of this newly visible region makes a 3.3 Å parallel aromatic stacking interaction with His-346 of core subdomain I; each of these residues is conserved in the known vertebrate topoisomerase I sequences. It has been shown previously that histidine residues pack preferentially against tryptophan side chains in proteins using such a parallel stacking arrangement (*37*). This stretch

Table 2: Root-Mean-Square Deviations between the Noncovalent Minus 1C Structure and Other Human Topoisomerase I−DNA Complex Structures[a]

	Topo70 noncovalent minus 1T (form 1)	reconstituted noncovalent minus 1T (form 2)	reconstituted covalent minus 1T (form 6)
all nonsolvent atoms			
rmsd (Å)	1.49	1.26	1.38
no. of atoms	4680	4675	4583
all protein atoms			
rmsd (Å)	1.31	1.01	1.13
no. of atoms	3784	3779	3691
Cα positions			
rmsd (Å)	1.20	0.68	0.84
no. of atoms	471	471	458
all DNA atoms			
rmsd (Å)	2.12	2.16	2.34
no. of atoms	896	896	892
P positions			
rmsd (Å)	2.32	2.16	2.43
no. of atoms	42	42	41

[a] Form 1 is the noncovalent complex between an NH_2-terminally truncated 70 kDa form of human topoisomerase I with an inactivating Tyr-723-Phe mutation and a 22 bp duplex DNA oligonucleotide (Topo70; *8*, *9*). Forms 2 and 6 both contain a reconstituted form of the enzyme [58 kDa core and 6.3 kDa COOH-terminal domains in a 1:1 noncovalent complex (*41*)] in complex with a similar 22 bp DNA oligonucleotide. The form 2 structure is a noncovalent complex with DNA (with an inactivating Tyr-723-Phe mutation), while in the form 6 structure, the active-site Tyr-723 is intact and the enzyme is trapped in its covalent complex with DNA via a 3′-phosphotyrosine linkage (*7*). Forms 1, 2, and 6 all have a thymine at the −1 position of the scissile strand (minus 1T).

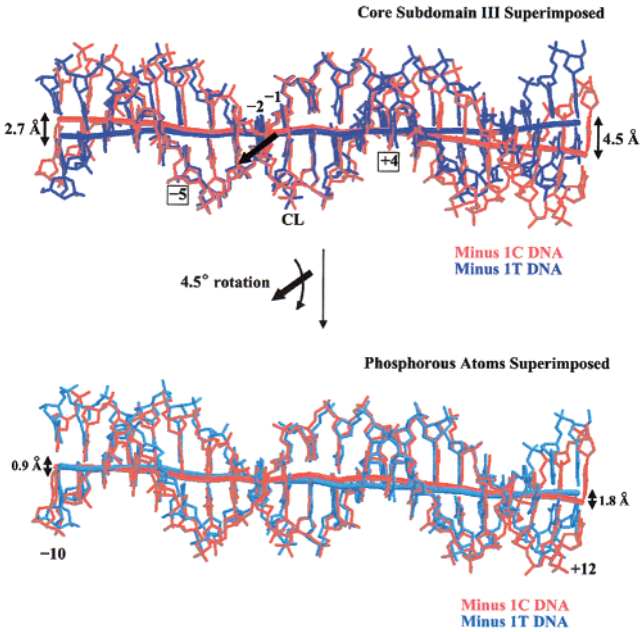

FIGURE 2: Rigid-body shift in the position of the DNA oligonucleotides between the minus 1C and minus 1T form 1 structures. The DNA oligonucleotides are first shown in the same orientation as in Figure 1, aligned by the superposition of the Cα positions of the core subdomain III regions of the enzyme, with the minus 1C DNA in red and the minus 1T DNA in blue. The minus 1T oligonucleotide (light blue) was subsequently superimposed on the minus 1C (red) oligonucleotide using the positions of the 42 equivalent phosphorus atoms in the oligonucleotides. This superposition resulted in a 4.5° rigid-body rotation of the minus 1T oligonucleotide about an axis normal to the page and centered between the −1 and the −2 base pairs. The DNA helical axes of each duplex [calculated using NUPARM (*36*)] are rendered as solid lines. The site of DNA strand cleavage is indicated (CL).

of 12 amino acids also interacts with the upper portion of α-helix 8 of core subdomain III, and the side chain of Trp-205 is 3.7 Å from Gly-437 (Figure 3). A glycine residue is conserved in this position in 13 of 14 cellular topoisomerases I with known sequences. It is thought that the upper region of α8 functions as part of the "hinge" by which the enzyme opens to clamp around its DNA substrate (*7*). The proper positioning of this stretch of 12 amino acids at the NH_2 terminus of the molecule may be critical for the opening and closing mechanism of human topoisomerase I as discussed below.

Lys-532 to Pyrimidine O2 Interaction. The interaction between Lys-532 and O2 of the pyrimidine base at the −1 position in the scissile strand is intact in the current minus 1C structure, similar to structures of human topoisomerase I in complex with DNA determined previously (Figure 4A; *7*−*9*). In all the minus 1C and minus 1T structures, the length of the observed hydrogen bond between the ϵ-amino group of Lys-532 and the O2 atom of the pyrimidine base in the −1 position of the scissile strand is between 2.8 and 3.0 Å (Table 3). Thus, the replacement of the favored thymine base in this position with cytosine did not abrogate this hydrogen-bonding interaction. In addition, it remains the only base-specific interaction between human topoisomerase I and its DNA substrate observed in any structure. As can be seen in Figure 4B, the interaction between Lys-532 and the cytosine O2 atom in the minus 1C structure is analogous to the Lys-532−thymine O2 interaction seen in the minus 1T structure. The existence of the N2 group in the −1 guanine base in the intact strand (which base pairs with the scissile −1 C base) does not appear to greatly distort this region of the structure (Figure 4A,B). Lys-532, however, is shifted slightly down and away from the N2 group of the −1 guanine base relative to the minus 1T structure, and this shift may be caused by the presence of the N2 group in the minor groove.

Water Adjacent to the Active Site. A water molecule was observed adjacent to the active-site residues in the minus 1C structure. The solvent site appeared as a 3.8σ peak in a σ_A-weighted (see Materials and Methods) difference density map, and at the completion of structural refinement had a thermal displacement parameter (*B*-factor) of 62.8 Å^2 (the mean *B*-factor for all solvent sites was 66.2 Å^2; see Table 1). It makes a short, 2.5 Å hydrogen-bonding interaction with the guanidinium group of Arg-590 (Figure 5). In addition, if a hydroxyl group ("OH" in Figure 5) is added to the Phe-723 side chain to create a "pseudo-tyrosine", this hydroxyl oxygen would be 3.0 Å from the phosphorus atom of the scissile phosphate group and 2.3 Å from this adjacent water molecule. Since no amino acid side chain is observed to be positioned properly to function as a general base to activate the nucleophilic tyrosine, it was suggested that the proton might be abstracted by a hypothetical water molecule as the reaction proceeds (*8*). The solvent site observed in the current structure appears to be appropriately placed to function as this putative catalytic water.

Rotation of the Scissile Phosphate Group. One of the most striking features detected in the minus 1C structure was the observed rotation of the scissile phosphate group by 75°

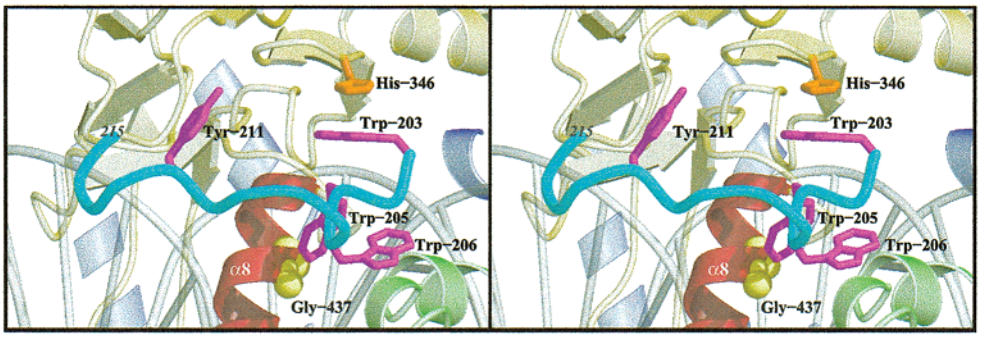

FIGURE 3: Stereoview of the 12 new residues built at the NH_2 terminus of the minus 1C structure. The orientation of the molecule in this figure is similar to that in Figure 1. The amino acid backbone from Trp-203 to Gly-214 is rendered in cyan, and the aromatic residues in this stretch of the molecule are shown in magenta. Gly-437, a putative hinging residue near the top of α8 in core subdomain III (red), is rendered in yellow, and His-346 of core subdomain I (yellow), which stacks in a planar fashion on Trp-203, is rendered in orange.

about a pseudobond between the O5′, P, and O3′ atoms and the subsequent reorientation of the nonbridging oxygen atoms. σ_A-weighted difference density maps calculated in the initial stages of structural refinement clearly indicated that a shift in the position of the nonbridging oxygens of the scissile phosphate was in order (Figure 6A). When the rebuilding of this area and structural refinement of the minus 1C structure were complete, the ϵ-amino group of Lys-532 was observed to be within hydrogen-bonding distance (3.2 Å) of O1P (Figure 5). In addition, the interactions between the active-site residues and the nonbridging oxygens of the scissile phosphate were observed to have changed significantly. As shown in Table 4 and Figures 5 and 6B, the two active-site arginine residues, Arg-488 and Arg-590, make 3.0 and 2.8 Å interactions, respectively, with the O1P atom, and His-632 makes a 2.9 Å interaction with the O2P atom in the minus 1T form 1 structure (similar interactions are observed in the form 2 structure; *8*). While the His-632−O2P interaction is still conserved in the minus 1C structure (at 2.6 Å), O1P is now observed to interact with Lys-532 and Arg-488 (at 3.2 and 2.6 Å, respectively), and Arg-590 appears to be intermediate (3.3 and 3.6 Å) between both nonbridging oxygen atoms. Thus, the proximity between Lys-532 and the scissile phosphate oxygen implicates Lys-532 as an active-site residue. The involvement of a lysine residue in the catalytic mechanism of the type IB topoisomerases had been suggested on the basis of mutational analysis of vaccinia topoisomerase and XerD, a structurally and functionally related tyrosine recombinase from *Salmonella typhimurium* (*38*, *42*). In addition, the orientation of the scissile phosphate group in the minus 1C noncovalent structure is similar to that observed in the covalent minus 1T complex of human topoisomerase I and DNA reported previously (*7*). As shown in Figure 6C, the O1P and O2P atoms are similarly oriented in these structures, as opposed to the orientation observed for this group in the noncovalent minus 1T form 1 and form 2 structures (Figure 6B). These results suggest that a greater degree of structural flexibility exists at the active site of the enzyme than was initially appreciated, and suggest alterations to the proposed active-site mechanism for the enzyme as discussed below.

DISCUSSION

Several novel features of the 2.6 Å crystal structure reported here of human topoisomerase I in noncovalent complex with a DNA duplex containing a cytosine in the −1 position of the scissile strand have implications for the proposed DNA-binding and active-site mechanisms of the enzyme. Overall, the structure is quite similar to the human topoisomerase I−DNA complex structures described previously, with rmsds between 1.3 and 1.5 Å between all equivalent atoms. The linker domain of the minus 1C structure, however, was found to be disordered (Figure 1), similar to that of a subset of form 1 structures described previously (*9*). This observation lends further support to the notion that the linker domain of human topoisomerase I is highly flexible. This flexibility has been proposed to play a critical role in the proposed "controlled rotation" mechanism for the relaxation of superhelical tension by the enzyme (*8*, *9*).

The DNA oligonucleotide region of the minus 1C structure was also observed to be shifted relative to form 1 structures described previously (*9*). These shifts (up to 5.4 Å in magnitude) were the largest yet observed for the DNA oligonucleotide of a topoisomerase I−DNA complex structure (Figures 1 and 2). Although they appear to be the result of a rigid-body shift in the position of the DNA duplex, they nonetheless suggest that the human topoisomerase I−DNA interaction can accommodate a greater degree of flexibility than has been previously appreciated. Because the DNA oligonucleotides are confined within the crystalline lattice, stacking head-to-tail to form a pseudo-continuous DNA helix in all the crystal forms of human topoisomerase I−DNA complexes described thus far, it has been difficult to assess the full degree to which the enzyme might distort its DNA substrate. These results give us the first indication that this type of rigid-body shift is accommodated within the topoisomerase I−DNA interaction.

The 12 additional residues added to the NH_2 terminus of the structure (Trp-203−Gly-214) pack against the human topoisomerase I protein in a region involving small portions of core subdomain I, the COOH-terminal domain, and α-helix 8 of core subdomain III (Figures 1 and 3). The upper portion of α8 has been proposed to function as part of the hinging mechanism that allows human topoisomerase I to open up and clamp around its substrate DNA duplex (*7*); this region of the enzyme is protease sensitive in the absence of DNA and becomes relatively protease resistant when DNA is added (*12*). Trp-205 and Trp-206 are part of a group of large aromatic residues that cluster around the top of α8 and the adjacent region of the COOH-terminal domain; these

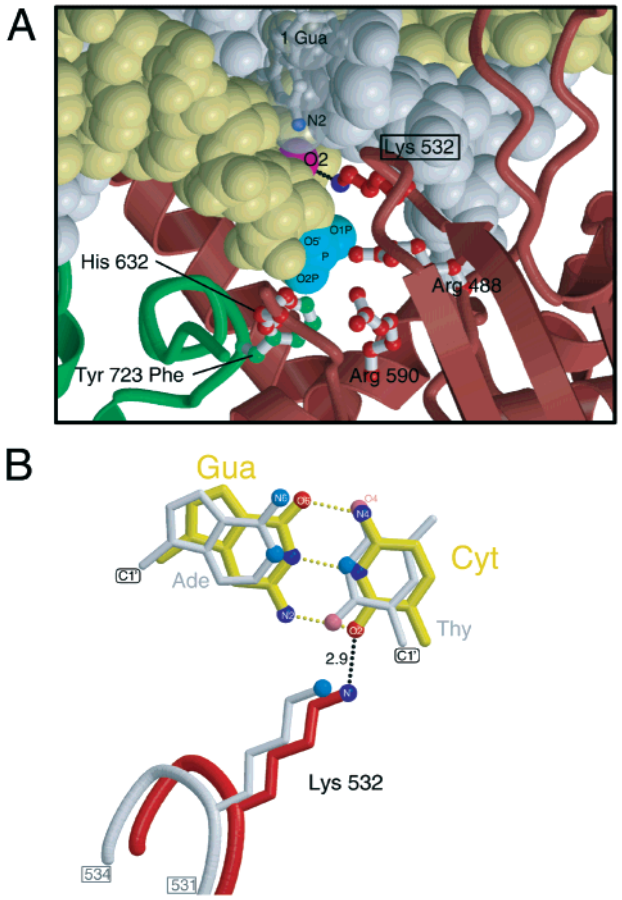

FIGURE 4: (A) Positions of the active-site residues and Lys-532 in the minus 1C structure. The molecule in this figure is rotated approximately 180° about the vertical axis relative to Figure 1. The scissile and intact DNA strands are rendered in yellow and light gray, respectively, and the scissile phosphate is shown in cyan. Lys-532 (red) makes a 2.9 Å hydrogen bond with O2 (magenta) of the cytosine base in the scissile −1 position; the base-pairing guanine at this position is rendered partially transparent so that the position of its N2 (blue) can be seen relative to Lys-532. The remaining active-site residues, Arg-488, Arg-590, and His-632 (seen essentially end-on), are rendered in white and red, or in white and green (for Tyr-723-Phe). (B) Closeup view of the interactions between Lys-532 and the pyrimidine O2 atoms in the minus 1C and minus 1T form 1 structures. The C·G base pair at the −1 position in the minus 1C structure is shown in yellow and Lys-532 in red, and the equivalent regions of the minus 1T form 1 structure (including the T·A base pair at the −1 position) are in gray. The positions of the C1′ atoms linking the bases to the deoxyribose rings of the sugar−phosphate backbone are also indicated. The minus 1T form 1 structure was superimposed on the minus 1C structure using all equivalent Cα positions.

Table 3: Lys-532 to Pyrimidine O2 Hydrogen-Bonding Distances in Human Topoisomerase I−DNA Complex Structures

	distance (Å)[a]
minus 1C	2.82
form 1[b]	3.03
form 2[b]	2.94
form 6[b]	2.86

[a] The length of the hydrogen bond between the ϵ-amino group of Lys-532 and the O2 atom of the −1 scissile pyrimidine base. [b] See the footnote of Table 2 and the text for definitions.

aromatic residues include Trp-441 and Trp-754 in addition to Trp-205 and -206. Each of these residues is conserved in the higher eukaryotic topoisomerases I with known se-

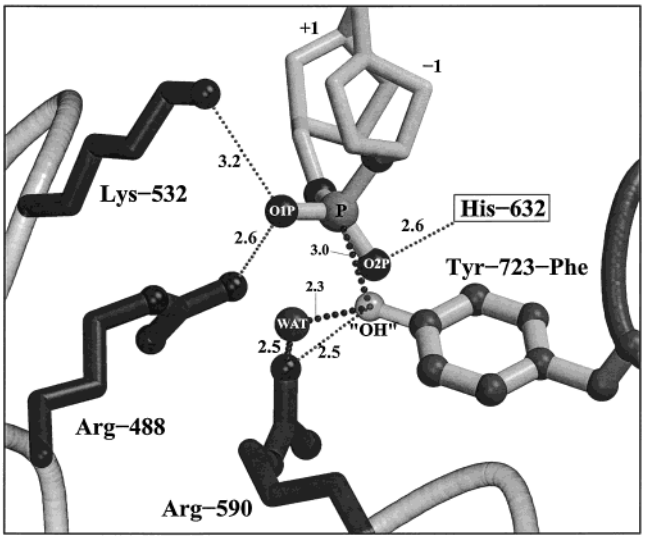

FIGURE 5: Putative catalytic water (WAT) and catalytic side chains in the active site of the minus 1C structure. Lys-532 and Arg-488 make 3.2 and 2.6 Å hydrogen bonds, respectively, with the nonbridging O1P of the scissile phosphate, and His-632 (shown schematically) lies 2.6 Å from the other nonbridging oxygen, O2P. The potential catalytic water, proposed to function as a specific base in the strand cleavage mechanism of the enzyme, makes a 2.5 Å hydrogen bond with Arg-590. The tyrosyl hydroxyl ("OH"), if it were to exist on the Tyr-723-Phe residue, would be 3.0 Å from the phosphorus atom, 2.3 Å from the putative catalytic water, and 2.5 Å from the guanidinium group of Arg-590.

quences. It has been shown that the replacement of the residue equivalent to Gly-437 in yeast topoisomerase I (Gly-369) with aspartic acid leads to a diminished level of DNA binding by the enzyme (*39*). In addition, it has been suggested that the main-chain flexibility of the conserved glycine residue at position 437 in human topoisomerase I (and position 369 in the yeast enzyme) may be critical for the opening and closing mechanism of the protein (*39*). Replacement of Gly-437 with a non-glycine residue may reduce the DNA binding affinity of the enzyme simply by restricting the $\phi-\psi$ flexibility of this critical residue. (Gly-437 is found in human topoisomerase I in an α-helical region of the protein, with ϕ and ψ angles of −60° and −24°, respectively.) In addition, the proper positioning of the Trp-203−Gly-214 region, including the formation of the conserved hydrophobic cluster that includes residues Trp-205, Trp-206, Trp-441, and Trp-754, may also be a critical final step in the proper closing of the enzyme around its DNA substrate. This step might be disrupted by the introduction of a charged amino acid in place of Gly-437, which would require the repositioning of several of these hydrophobic residues. These results suggest that the conserved hydrophobic cluster made up in part by the 12-amino acid stretch observed for the first time in this structure might play an important role in the opening and closing mechanism of the enzyme.

A water molecule was positioned adjacent to the active-site residues, including Tyr-723-Phe, in the minus 1C structure (Figure 5). Because no protein residue was observed in an appropriate position to function as a general base during catalysis, we previously proposed that the proton on the attacking nucleophile was lost to a water molecule as the covalent intermediate was formed (*8*). At the time of the publication of this mechanism, no ordered water molecule

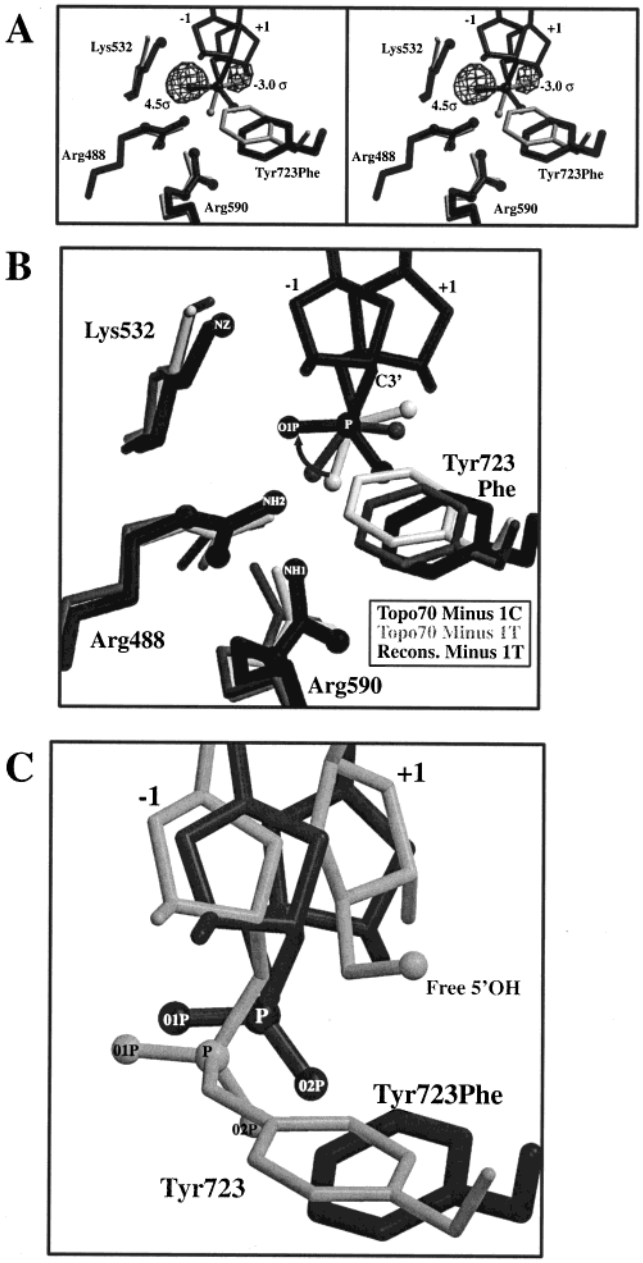

FIGURE 6: (A) Stereoview of the shift in the positions of the nonbridging oxygens of the minus 1C scissile phosphate. A σ_A-weighted difference map contoured at 4.5σ (light gray) and −3.0σ (dark gray) calculated prior to any rebuilding of the model in this area indicates that a positional shift of these atoms is appropriate. The minus 1C structure is in dark gray, while the form 1 structure (light gray) is shown as it was originally positioned after molecular replacement and rigid-body refinement. (B) Superposition of active-site residues of the Topo70 noncovalent minus 1C (black, thicker bonds), Topo70 noncovalent minus 1T complex (Topo70 Minus 1T; white), and reconstituted noncovalent minus 1T complex (Recons. Minus 1T; medium gray) structures. The shift in position of the nonbridging oxygens of the minus 1C scissile phosphate (indicated by the arrow) is evident relative to the other two structures. In addition, the position of the Tyr-723-Phe residue has also shifted. The structures were superimposed using the Cα positions of the active-site residues, Tyr-723-Phe, Arg-590, Arg-488, Lys-532, and His-632. (C) Closeup of the active site of the noncovalent minus 1C complex (black) and the reconstituted covalent minus 1T complex (form 6; gray) superimposed using the Cα positions of the active-site residues, as in panel B. Note that the nonbridging oxygens of the scissile phosphates of each structure (whether in a phosphodiester linkage or a 3′-phosphotyrosine linkage) are oriented in the similar fashion.

Table 4: Distances (in Å) between the Nonbridging Oxygens of the Scissile Phosphate and the Non-Tyrosine Active-Site Residues in the Minus 1C and Form 1 Structures

	noncovalent minus 1C		noncovalent form 1 (minus 1T)	
	O1P	O2P	O1P	O2P
Lys-532 (NZ)	3.17[a]	5.48	5.61	6.00
Arg-488 (NH2)	2.56[a]	3.30	3.04[a]	4.99
Arg-590 (NH1)	3.60	3.30	2.82[a]	5.28
His-632 (NE2)	5.00	2.56[a]	3.72	2.92[a]

[a] Interactions within hydrogen-bonding distance.

had been observed in any of the human topoisomerase I−DNA complex structures. However, in the subsequent refinement of the minus 1T form 2 structure (PDB entry 1A35; *7*), a water molecule was positioned like that observed in the minus 1C structure described here. Both solvent sites form a close (2.5 Å in the minus 1C structure) hydrogen-bonding interaction with the guanidinium group of Arg-590 and are positioned adjacent to Tyr-723-Phe (2.3 Å from the OH of a pseudo-tyrosine in the minus 1C structure; Figure 5). As we still see no evidence for an amino acid side chain adjacent to the active site that may function as a general base, we propose that a water molecule positioned like that observed in the minus 1C and form 2 structures may serve as a specific base to activate the catalytic Tyr-723. In addition, the guanidinium group of Arg-590, which lies 2.5 Å from the hydroxyl group of the pseudo-tyrosine of the Tyr-732-Phe residue, may cause a significant downward shift of the pK_a of this tyrosine hydroxyl (*40*). This would facilitate the formation of a deprotonated, phenolate form of Tyr-723 at physiological pH, further readying this essential catalytic residue for in-line nucleophilic attack on the scissile phosphate group.

The hydrogen-bonding interaction between pyrimidine O2 of the −1 scissile cytosine and the ϵ-amino group of Lys-532, the only protein−DNA base-specific interaction observed between human topoisomerase I and its DNA substrate, is intact in the minus 1C structure (Figure 4A). In fact, this interaction is very similar to that observed in the previously determined human topoisomerase I−DNA complex structures (Table 3 and Figure 4B; *7*−*9*). Thus, the presence of the N2 group on the base-pairing guanine base at the −1 position of the intact strand did not eliminate the interaction between Lys-532 and the −1 cytosine. Crystals were not obtained, however, in the presence of a non-pyrimidine base at the scissile −1 position. That fact, taken together with the observations of Been et al. (*23*) and Anderson et al. (*24*) that thymine is primarily (with cytosine secondarily) favored by topoisomerase I in that position, suggests that the observed interaction between the pyrimidine O2 atom and Lys-532 plays a critical role in the association of the enzyme with DNA.

The nonbridging oxygens of the scissile DNA phosphate group were observed to shift in position by a 75° rotation about a pseudobond between the O5′, P, and O3′ atoms in the minus 1C structure relative to their positions in the previous noncovalent minus 1T structures (Figure 6A). This shift brings the Lys-532 ϵ-amino group into hydrogen-bonding distance of one of the nonbridging oxygens of the scissile phosphate (Figure 5), and also changes the pattern of contact between the other active-site residues and the

scissile phosphate group (Figure 6B and Table 4). Whereas in the previous structures of noncovalent human topoisomerase I−DNA complexes Arg-488 and Arg-590 contacted one nonbridging oxygen (O1P) and His-632 contacted the other (O2P), now in the minus 1C noncovalent complex Lys-532 and Arg-488 contact O1P and Arg-590 is observed to be intermediate (3.3 and 3.6 Å) between the two nonbridging oxygens. In addition, His-632 makes an even closer contact (2.6 vs 2.9 Å) with O2P in the minus 1C structure than previously observed (Table 4). His-632 may perform two roles in the active-site cleavage mechanism of the enzyme (*8*). First, its positive charge character may assist in the stabilization of the pentavalently coordinated transition state of the reaction. Second, as the cleavage mechanism progresses, it may further function to protonate the leaving 5′-oxygen to create the free 5′-hydroxyl. In this formulation, as catalysis proceeds the positive charge of His-632 is less available for stabilization of the transition state, since this residue becomes involved in the protonation of the leaving 5′-hydroxyl group. Lys-532 may then function as an active-site residue by utilizing its positive charge character to stabilize this transition state at the later stages.

The orientation of the scissile phosphate group in the minus 1C structure is unique relative to the noncovalent minus 1T structures examined to date (Figure 6B), but is similar to that of the 3′-phosphotyrosine linkage observed in the structure of the covalent human topoisomerase I−DNA complex (Figure 6C; *7*). Perhaps a critical step in the cleavage mechanism of the enzyme is the rotation of the scissile phosphate, in part so that Lys-532, functioning as an active-site residue, can be brought into proximity to stabilize the transition state. The reason cytosine is less favored than thymidine in the scissile −1 position may be because it facilitates a premature rotation of the scissile phosphate group. Such a premature rotation may be needed with a C•G base pair in the −1 position to stabilize Lys-532 in the presence of the guanine N2 group in the minor groove. In summary, we propose that the active site of human topoisomerase I is made up of Tyr-723, Arg-590, Arg-488, His-632, Lys-532, and a catalytic water molecule. In addition, the catalytic pathway of the enzyme may involve a rotation of the scissile phosphate and a reorientation of the nonbridging oxygens relative to the active site residues.

ACKNOWLEDGMENT

We thank Stewart Turley, Dave Chudzik, Arnthor AEvarsson, Rongbao Li, Jungwoo Choe, Paulene Quigley, Stephen Suresh, and Bianca Hovey for considerable help in data collection. In addition, thanks are due to Heidrun Interthal, Zheng Yang, Colleen Kelleher, Jamie Winshell, and Brad Bernstein for helpful discussions and to Francis Athappilly, Christophe Verlinde, and Ethan Merritt for technical assistance.

REFERENCES

1. Wang, J. C. (1996) *Annu. Rev. Biochem. 65*, 635−692.
2. Champoux, J. J. (1998) *Prog. Nucleic Acid Res. Mol. Biol. 60*, 111−132.
3. Lima, C. D., Wang, J. C., and Mondragon, A. (1994) *Nature 367*, 138−146.
4. Yu, L., Zhu, C.-X., Tse-Dinh, Y.-C., and Fesik, S. W. (1995) *Biochemistry 34*, 7622−7628.
5. Leu, N., Sharma, A., and Mondragon, A. (1995) *Structure 3*, 1315−1322.
6. Sharma, A., Hanai, R., and Mondragon, A. (1994) *Structure 2*, 767−777.
7. Redinbo, M. R., Stewart, L., Kuhn, P., Champoux, J. J., and Hol, W. G. J. (1998) *Science 279*, 1504−1513.
8. Stewart, L., Redinbo, M. R., Qiu, X., Hol, W. G. J., and Champoux, J. J. (1998) *Science 279*, 1534−1541.
9. Redinbo, M. R., Stewart, L., Champoux, J. J., and Hol, W. G. J. (1999) *J. Mol. Biol. 292*, 685−696.
10. Cheng, C., Kussie, P., Paveltich, N., and Shuman, S. (1998) *Cell 92*, 841−850.
11. Redinbo, M. R., Champoux, J. J., and Hol, W. G. J. (1999) *Curr. Opin. Struct. Biol. 9*, 29−36.
12. Stewart, L., Ireton, G. C., and Champoux, J. J. (1996) *J. Biol. Chem. 271*, 7602−7608.
13. Stewart, L., Ireton, G. C., Parker, L. H., Madden, K. R., and Champoux, J. J. (1996) *J. Biol. Chem. 271*, 7593−7601.
14. Gobert, C., Bracco, L., Rossi, F., Olivier, M., Tazi, J., Lavelle, F., Larsen, A. K., and Riou, J.-F. (1996) *Biochemistry 35*, 5778−5786.
15. Albor, A., Kaku, S., and Kulesz-Martin, M. (1998) *Cancer Res. 58*, 2091−2094.
16. Merino, A., Madden, K. R., Lane, W. S., Champoux, J. J., and Reinberg, D. (1993) *Nature 365*, 227−232.
17. Kretzschmar, M., Meisterernst, M., and Roeder, R. G. (1993) *Proc. Natl. Acad. Sci. U.S.A. 90*, 11508−11512.
18. Shykind, B. M., Kim, J., Stewart, L., Champoux, J. J., and Sharp, P. A. (1997) *Genes Dev. 11*, 397−407.
19. Simmons, D. T., Melendy, T., Usher, D., and Stillman, B. (1996) *Virology 222*, 365−374.
20. Pommier, Y., Kohlhagen, G., Wu, C., and Simmons, D. T. (1998) *Biochemistry 37*, 3818−3823.
21. Haluska, P., Saleem, A., Edwards, T. K., and Rubin, E. H. (1998) *Nucleic Acids Res. 26*, 1841−1847.
22. Bharti, A. K., Olson, M. O. J., Kufe, D. W., and Rubin, E. H. (1996) *J. Biol. Chem. 271*, 1993−1997.
23. Been, M. D., Burgess, R. R., and Champoux, J. J. (1984) *Nucleic Acids Res. 12*, 3097−3114.
24. Andersen, A. H., Gocke, E., Bonven, B. J., Neilsen, O. F., and Westergaard, O. (1985) *Nucleic Acids Res. 13*, 1543−1557.
25. Otwinowski, Z., and Minor, W., Eds. (1993) *Data Collection and Processing*, Daresbury Laboratory, Warrington, U.K.
26. Navaza, J. (1994) *Acta Crystallogr. A50*, 157−163.
27. Brünger, A. T. (1993) *X-PLOR*, version 3.1, Yale University Press, New Haven, CT.
28. Jones, T. A., Zou, J.-Y., Cowan, S. W., and Kjeldgaard, M. (1991) *Acta Crystallogr. A47*, 110−119.
29. Brünger, A. T. (1993) *Acta Crystallogr. D49*, 24−36.
30. Read, R. A. (1986) *Acta Crystallogr. A42*, 140−149.
31. Laskowski, R. A., MacArthur, M. W., Moss, D. S., and Thornton, J. M. (1993) *J. Appl. Crystallogr. 26*, 283−291.
32. Kraulis, P. J. (1991) *J. Appl. Crystallogr. 24*, 946−950.
33. Esnouf, R. M. (1999) *Acta Crystallogr. D55*, 938−940.
34. Bacon, D. J., and Anderson, W. F (1988) *J. Mol. Graphics 6*, 219−220.
35. Merritt, E. A., and Murphy, M. E. P. (1994) *Acta Crystallogr. D50*, 869−873.
36. Bhattacharyya, D., and Bansal, M. (1989) *J. Biomol. Struct. Dyn. 6*, 635−653.
37. Samanta, U., Pal, D., and Chakrabarti, P. (1999) *Acta Crystallogr. D55*, 1421−1427.
38. Cao, Y., and Hayes, F. (1999) *J. Mol. Biol. 289*, 517−527.
39. Hann, C. L., Carlberg, A. L., and Bjornsti, M.-A. (1998) *J. Biol. Chem. 273*, 31519−31527.
40. Fersht, A. (1999) *Structure and Mechanism in Protein Science*, W. H. Freeman and Company, New York.
41. Stewart, L., Ireton, G. C., and Champoux, J. J. (1997) *J. Mol. Biol. 269*, 355−372.
42. Wittschieben, J., and Shuman, S. (1997) *Nucleic Acids Res. 25*, 3001−3008.

BI992690T