A Model for the Mechanism of Human Topoisomerase I

Lance Stewart,*† Matthew R. Redinbo,* Xiayang Qiu,‡
Wim G. J. Hol, James J. Champoux§

The three-dimensional structure of a 70-kilodalton amino terminally truncated form of human topoisomerase I in complex with a 22–base pair duplex oligonucleotide, determined to a resolution of 2.8 angstroms, reveals all of the structural elements of the enzyme that contact DNA. The linker region that connects the central core of the enzyme to the carboxyl-terminal domain assumes a coiled-coil configuration and protrudes away from the remainder of the enzyme. The positively charged DNA-proximal surface of the linker makes only a few contacts with the DNA downstream of the cleavage site. In combination with the crystal structures of the reconstituted human topoisomerase I before and after DNA cleavage, this information suggests which amino acid residues are involved in catalyzing phosphodiester bond breakage and religation. The structures also lead to the proposal that the topoisomerization step occurs by a mechanism termed "controlled rotation."

Topoisomerases are ubiquitous and essential enzymes that solve the topological problems that accompany DNA replication, transcription, chromatin assembly, recombination, and chromosome segregation by introducing transient breaks into the helix (*1–3*). Strand cleavage by all topoisomerases involves nucleophilic attack by a catalytic tyrosine residue on the scissile phosphodiester bond that culminates in the formation of a covalent bond between the enzyme and one end of the broken strand. The accompanying report (*4*) describes the three-dimensional structure of a form of human topoisomerase I (topo I) reconstituted from two fragments of the protein (core and COOH-terminal domains), complexed either noncovalently or covalently with a 22–base pair (bp) DNA oligonucleotide. Here we present the crystal structure of an NH_2 terminally truncated 70-kD form of the human enzyme composed of residues Lys^{175} to Phe^{765} (topo70) in a noncovalent complex with double-stranded DNA. This structure includes the linker region (residues Pro^{636} to Lys^{712}) that is missing in the reconstituted enzyme. The spatial organization of the amino acid residues in the vicinity of the scissile phosphate in the DNA, in conjunction with information from mutagenesis studies and the pattern of conserved amino acid residues, suggests a chemical mechanism for the nicking-closing reaction. The presence of two positively charged surfaces downstream of the cleavage site that make minimal contacts with the DNA helix would appear to be more consistent with a modified form of the "free rotation" model for DNA relaxation, called "controlled rotation," than with the "strand passage" model favored for *Escherichia coli* topo I (*1, 2, 5*).

Crystals of an inactive mutant form of topo70 [in which Tyr^{723} is mutated to Phe (Y723F)] in complex with a 22-bp oligonucleotide were obtained after expression of topo70 in insect cells and after screening more than 50 different oligonucleotide variations of a high-affinity topo I–binding site from the ribosomal DNA of *Tetrahymena thermophila* (*6*). Repeated cycles of macroseeding were required to obtain crystals of sufficient size for data collection, and both cryocooling and synchrotron radiation were necessary to obtain data beyond 3 Å resolution (Table 1). Although crystals could be grown reproducibly, they consistently exhibited both a high degree of nonisomorphism and a large mosaic spread. The *b* axis of the monoclinic crystals varied by 12 Å, and the β angle showed a range of 10°. This variation made the identification of useful isomorphous heavy-atom derivatives extremely difficult, despite our having collected ~120 data sets after examining more than 700 crystals with the x-ray beam.

Molecular replacement approaches (*7*), using a variety of available search models, allowed us to position a 22-bp straight B-form DNA into the monoclinic cell along the *c* axis (*8*). However, the phases calculated from this model, alone or in combination with weak phases from a crystal with three iodinated bases in the DNA, were not sufficient to solve the structure of the topo70-DNA complex. We could only achieve the structure solution by molecular replacement using the structure of the reconstituted human topo I (*4*). However, despite numerous rounds of structural refinement (*9*), the phases derived from the reconstituted model were not sufficient to allow the linker domain to be traced. Ultimately, by combining these phases with weak multiwavelength anomalous dispersion (MAD) phasing information from a crystal containing a hexabrominated oligonucleotide–topo70 complex, the structure of the linker domain was eventually elucidated (Table 1).

The 77-residue linker domain appears to consist mainly of two long α helices connected by a short turn formed by residues Met^{675} to Ala^{678}, creating an antiparallel coiled-coil (Fig. 1, A and B). The tip of the coiled-coil has low temperature factors probably as a result of crystal contacts of this region with a neighboring topo70 molecule in the crystal. Residues Asn^{711} to Leu^{716}, which connect the linker to the COOH-terminal domain, display quite high temperature factors, and residues Pro^{636} to Phe^{640}, which bridge the core and linker domains, are disordered. The two long helices (α18 and α19) of the linker domain are well defined in density and interact with each other through a number of classical hydrophobic leucine-leucine contacts (Fig. 1) and also through contacts involving the aliphatic portions of long, charged side chains as exemplified by the interaction between residues Lys^{654} and Leu^{698} (Fig. 1B). There is clear evidence of a heptad repeat in the two ~30-residue-long helices making up the coiled-coil (Fig. 1C).

A DALI search (*10*) revealed a very high degree of structural similarity of the linker domain with the two helices of the natural repressor of primer (Rop) protein from *E. coli* (*11*) and the GreA transcript cleavage factor from *E. coli*. The highest degree of similarity was found with a modified form of Rop that has a two–amino acid insertion (Rop<2aa>) in the turn region that more effectively registers the heptad repeat of the hydrophobic residues (Fig. 1, A and C). A total of 60 residues of

L. Stewart, M. R. Redinbo, X. Qiu, Biomolecular Structure Center and Department of Biological Structure, School of Medicine, University of Washington, Seattle, WA 98195–7742, USA.
W. G. J. Hol, Departments of Biological Structure and Biochemistry, Biomolecular Structure Center, and Howard Hughes Medical Institute, School of Medicine, University of Washington, Seattle, WA 98195–7742, USA.
J. J. Champoux, Department of Microbiology, School of Medicine, University of Washington, Seattle, WA 98195–7242, USA.

*These authors contributed equally to this work.
†Present address: Emerald BioStructures, Incorporated, 7865 Northeast Day Road West, Bainbridge Island, WA 98110, USA. E-mail: emerald_biostructures@rocketmail.com
‡Present address: Department of Structural Biology, SmithKline Beecham Pharmaceuticals, King of Prussia, PA 19406, USA.
§To whom correspondence should be addressed: E-mail: champoux@u.washington.edu

Rop<2aa> and the linker can be superimposed with a root-mean square (rms) deviation for Cα atoms of 0.8 Å, yet the amino acid sequence identity is only 3.3% (Fig. 1C). Bringing the two linker helices together results in the formation of three interhelix salt bridges and nine hydrophobic pairwise interactions and buries 1030 Å^2 of surface area (Fig. 1C). The solvent-exposed surfaces of the linker helices (α18 and α19) are highly charged and generally polar in nature. The DNA proximal (top) side of the coiled-coil has a large net positive charge, with nine lysine and arginine residues and only two aspartic acid side chains. In contrast, the bottom surface of the linker region is only slightly positively charged with seven lysine and arginine residues counterbalanced by six aspartic and glutamic acid residues (Fig. 1C). It seems likely that the asymmetrical organization of charges along the surface of the linker plays a functional role in the mechanism of human topo I.

The linker domain helices protrude away from the core and COOH-terminal domains, with contacts occurring only between the last helix (α17) of core subdomain III and the COOH-terminal end of the second linker helix (α19) (Fig. 1B). Residues Pro^{613}, Leu^{617}, and the aliphatic portion of the Arg^{624} side chain in core helix 17 engage in hydrophobic interactions with Val^{703}, Ile^{714}, and Leu^{716} of linker helix 19. In addition, there are two interdomain salt bridges, Arg^{624}-Glu^{710} and Arg^{621}-Asp^{707}, between helices 17 and 19 (Fig. 1B). No other interactions are observed between the linker domain helices and the remainder of the topo70 molecule. As such, the linker domain might be free to shift with respect to the core by moving hydrophobic surfaces along each other, a phenomenon that could explain the extreme degree of non-

Table 1. Crystallographic structure determination of residues 641 to 716 of 70-kD human topo I. MAD data were collected from a crystal containing six 5′-bromo-3′ deoxyuridine nucleotides on the DNA oligonucleotide, at positions -9, +1, +3, +4, +5, and +6 on the intact strand. The mean figures of merit [$\langle|\Sigma P(\alpha)e^{i\alpha}/\Sigma P(\alpha)|\rangle$, where α is the phase and $P(\alpha)$ is the phase probability distribution] of the MAD phasing information alone before phase combination were 0.32 and 0.20 for acentric (acen.) and centric (cen.) data, respectively (iso, isomorphous; anom, anomalous). Data used for refinement were collected from a crystal containing an unsubstituted DNA oligonucleotide. For the crystallization procedure human topo70 Y723F mutant was purified from a baculovirus-insect cell (SF9) expression system as described (*19*). Duplex oligonucleotides were prepared as described (*4*). Topo70-DNA cocrystals were grown by vapor diffusion at 22°C from sitting drops that were prepared by mixing 6 μl of crystallant (5 mM tris-HCl, pH 6.0, 20 mM MES-HCl, pH 6.8, 27% v/v 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 increased by multiple rounds of macroseeding into freshly prepared crystallization drops once every 3 days. For structure determination and refinement, crystals were cryoprotected by soaking them in a 4:1 mixture of crystallant plus ethylene glycol for 4 min at 22°C and were flash-frozen in liquid nitrogen or a nitrogen gas stream cooled to 100 K. Data were collected at 100 K at Stanford Synchrotron Radiation Laboratory, Cornell High Energy Synchrotron Source, and Brookhaven National Laboratory. Crystals were of space group $P2_1$ with a = 56.6 Å, $b \cong 123$ Å, c = 71.8 Å, and $\beta \cong 97°$, but they exhibited marked crystal-to-crystal nonisomorphism. The monoclinic b lengths and β angles varied from 118 to 130 Å and 92° to 102°, respectively, among the ~120 data sets collected, which led to high R_{iso} values (10 to 25%) between native data sets. Extensive alterations in cryoprotecting and flash-freezing protocols, including freezing in liquid propane, did not alleviate this problem. The structure was solved by molecular replacement with AMORE (*7*) by using the structure of the noncovalent complex of reconstituted human topo I bound to a 22-bp DNA duplex oligonucleotide as the search model (*4*). However, interpretable electron density for the 77-residue linker domain (~10% of the asymmetric unit by mass) was not evident in σ_A-weighted (*20*) partial model maps or in σ_A-weighted phase-combined maps that used single isomorphous replacement phases from a crystal with three 5-iodo-deoxyuridine substitutions on the DNA oligonucleotide. A four-wavelength MAD experiment was conducted at Stanford Synchrotron Radiation Laboratory, beamline 1-5, with a single crystal containing six 5′-bromo-3′ deoxyuridine substitutions on the DNA oligonucleotide (a = 56.9 Å, b = 120.3 Å, c = 71.5 Å, β = 100.7°). Heavy-atom refinement and MAD phasing were conducted with SHARP (*21*) with the high-energy reference data set (0.8996 Å) as the "pseudo" native. The σ_A-weighted phase combination between model phases and MAD phases to 3.2 Å resolution (Res.) gave clear, interpretable density for 72 of the 77 residues of the linker domain (Fig. 1, A and C), which were built with the program O (*22*). The model was then positioned into a 2.8 Å resolution native data set (a = 56.5 Å, b = 118.4 Å, c = 71.5 Å, β = 101.2°), and the remainder of the linker domain (with the exception of residues 636 to 640) was traced and side chains were built into σ_A-weighted $2|F_{obs}| - |F_{calc}|$ and $|F_{obs}| - |F_{calc}|$ maps. The model was refined by X-PLOR (*9*) with simulated annealing (*23*) and iterative model adjustments with O. The final complex model contains residues 215 to 635 and 641 to 765 of human topo I and a 22-bp DNA duplex oligonucleotide, with good geometry and no Ramachandran outliers (*24*). Figures were created with MOLSCRIPT (*25*), Raster3D (*26, 27*), and GRASP (*28*).

MAD phasing information

Wavelength	Res. (Å) (last shell)	Reflections unique/total	Complete (%) (last shell)	R_{sym}* (%) (last shell)	R_{anom}† (%)	f'_{obs}/f''_{obs}‡ (electrons)	*PhP*§		R_{cullis}‖	
							Acen. (iso/anom)	Cen.	Acen. (iso/anom)	Cen.
0.8996 Å	30–3.2 (3.3–3.2)	15,385/51,036	97.9 (78.0)	6.8 (34.4)	4.6	−3.0/3.5	—/0.74	—	—/0.91	—
0.9196 Å	30–3.2 (3.3–3.2)	15,443/86,569	98.0 (96.7)	6.8 (37.2)	3.8	−6.8/4.4	0.92/1.10	0.79	0.49/0.89	0.57
0.9203 Å	30–3.2 (3.3–3.2)	15,387/54,437	98.0 (95.1)	6.1 (35.4)	3.9	−8.1/1.9	0.19/0.49	0.23	0.65/0.94	0.68
0.9342 Å	30–3.2 (3.3–3.2)	15,522/52,913	98.0 (83.1)	6.3 (38.6)	4.2	−3.7/0.4	0.28/0.32	0.25	0.63/0.95	0.67

Structural refinement

Res. (Å) (last shell)	Reflections unique/total	Redundancy	Complete (%) (last shell)	R_{sym}* (%) (last shell)	No. protein and DNA atoms	R¶ (%)	R_{free}# (%)	Rmsd** bonds (Å)	Rmsd** angles (°)
8–2.8 (2.9–2.8)	21,217/63,314	3.0	94.7 (86.8)	5.5 (30.5)	5369	24.2	33.7	0.012	1.70

*$R_{sym} = \Sigma|I - \langle I\rangle|/\Sigma I$, where I is the observed intensity and $\langle I\rangle$ the average intensity of multiple symmetry-related observations of that reflection. †$R_{anom} = \Sigma||I^+_{PH} - I^-_{PH}||/\Sigma(|I_{PH}|)$, where I^+_{PH} and I^-_{PH} are the intensities of the Bijvoet pairs and $\langle I_{PH}\rangle$ is the average of these Bijvoet intensities. ‡Mean scattering factors (f) observed for heavy-atom substitutions. §*PhP* (phasing power) = (rms$|F_h|$)/(rms E), where $|F_h|$ is the heavy-atom structure factor amplitude and E the residual lack of closure error. ‖$R_{cullis} = \Sigma|E|/\Sigma||F_{PH}| - |F_P||$. ¶$R = \Sigma||F^{obs}_P| - |F^{calc}_P||/\Sigma|F^{obs}_P|$. #$R_{free} = \Sigma||F^{obs}_P| - |F^{calc}_P||/\Sigma|F^{obs}_P|$ where $|F^{obs}_P|$ is from a test set not used in structural refinement (7% of the observed data). **Rmsd values are rms deviations from ideal bond lengths and bond angles.

isomorphism seen in the crystals.

Because the NH_2-terminal domain of human topo I has no defined spatial organization (*4*), the three-dimensional structure of the topo70-DNA complex provides a view of all the well-defined structural domains of human topo I that interact with DNA. As has been seen in all six crystal forms of human topo I–DNA complexes obtained thus far, the blunt-ended oligonucleotide duplexes are stacked head-to-tail in the crystals (*4*, *8*). The core and COOH-terminal domains of topo70 form a clamp around the DNA in a manner very similar to that seen in the reconstituted enzyme (*4*). A view of the electrostatic surface potential of topo70 demonstrates that the DNA-proximal surface of the linker domain is highly positively charged (Fig. 2A), as is the central hole of the enzyme, which matches very well the negative charges of the DNA sugar-phosphate backbone. Comparing the −4 to +6 base pair regions of the DNA in the topo70-DNA complex with the reconstituted topo I–DNA complexes (*4*) yields an rms deviation of 0.6 Å. The remaining flanking base pairs display an rms deviation of 1.3 Å. In addition, the core and COOH-terminal modules of topo70 are also very similar in structure to those of the reconstituted enzymes, with an overall rms deviation between Cα atoms of 1.2 Å. Significant deviations only occur near the active site where residues Gly^{717} to Asn^{722} adopt a short helical conformation (α20) in the topo70-DNA complex (Figs. 1B and 2B).

In our crystals the long axis of the linker domain points ~30° away from the DNA helical axis (Fig. 2, A and B). The tip of the linker is 35 Å removed from the nearest phosphate group in the pseudocontinuous DNA helix. Moreover, the highly positively charged top surface of the coiled-coil,

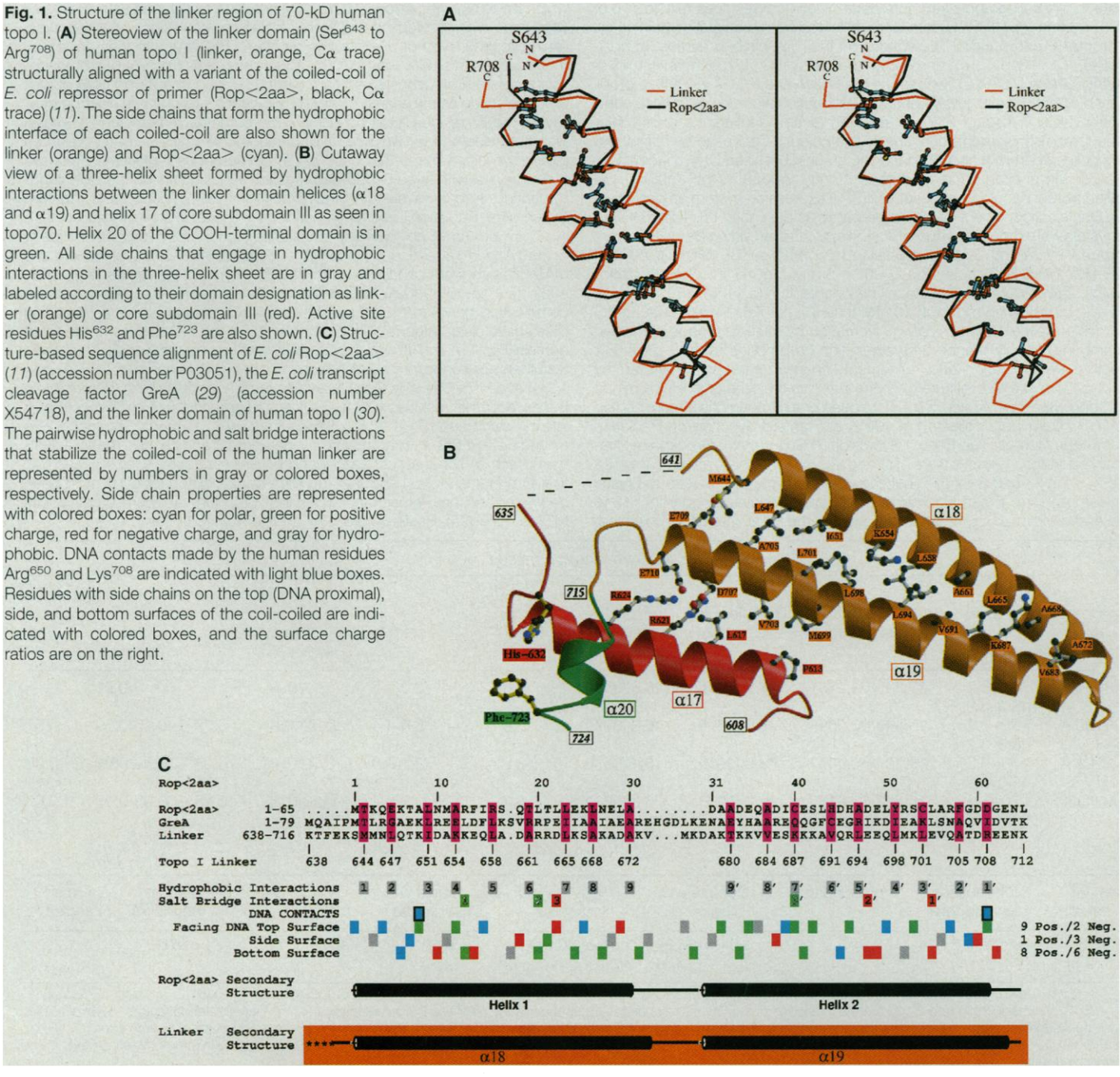

Fig. 1. Structure of the linker region of 70-kD human topo I. (**A**) Stereoview of the linker domain (Ser^{643} to Arg^{708}) of human topo I (linker, orange, Cα trace) structurally aligned with a variant of the coiled-coil of *E. coli* repressor of primer (Rop<2aa>, black, Cα trace) (*11*). The side chains that form the hydrophobic interface of each coiled-coil are also shown for the linker (orange) and Rop<2aa> (cyan). (**B**) Cutaway view of a three-helix sheet formed by hydrophobic interactions between the linker domain helices (α18 and α19) and helix 17 of core subdomain III as seen in topo70. Helix 20 of the COOH-terminal domain is in green. All side chains that engage in hydrophobic interactions in the three-helix sheet are in gray and labeled according to their domain designation as linker (orange) or core subdomain III (red). Active site residues His^{632} and Phe^{723} are also shown. (**C**) Structure-based sequence alignment of *E. coli* Rop<2aa> (*11*) (accession number P03051), the *E. coli* transcript cleavage factor GreA (*29*) (accession number X54718), and the linker domain of human topo I (*30*). The pairwise hydrophobic and salt bridge interactions that stabilize the coiled-coil of the human linker are represented by numbers in gray or colored boxes, respectively. Side chain properties are represented with colored boxes: cyan for polar, green for positive charge, red for negative charge, and gray for hydrophobic. DNA contacts made by the human residues Arg^{650} and Lys^{708} are indicated with light blue boxes. Residues with side chains on the top (DNA proximal), side, and bottom surfaces of the coil-coiled are indicated with colored boxes, and the surface charge ratios are on the right.

which faces the DNA, only makes two contacts with the duplex involving residues Lys[650] and Arg[708]. These residues, which are located at the core-proximal end of the linker and are conserved as lysine or arginine in all cellular type I topoisomerases (8), respectively make interactions with the phosphate between the +7 and +8 nucleotides on the noncleaved strand and the phosphate between the +9 and +10 nucleotides on the cleaved strand (Fig. 2C). In contrast, the remaining seven positively charged residues and seven polar residues on the top side of the coiled-coil make no interactions with phosphates or any other oligonucleotide atom.

This phenomenon of positively charged residues on a pair of helices facing but not

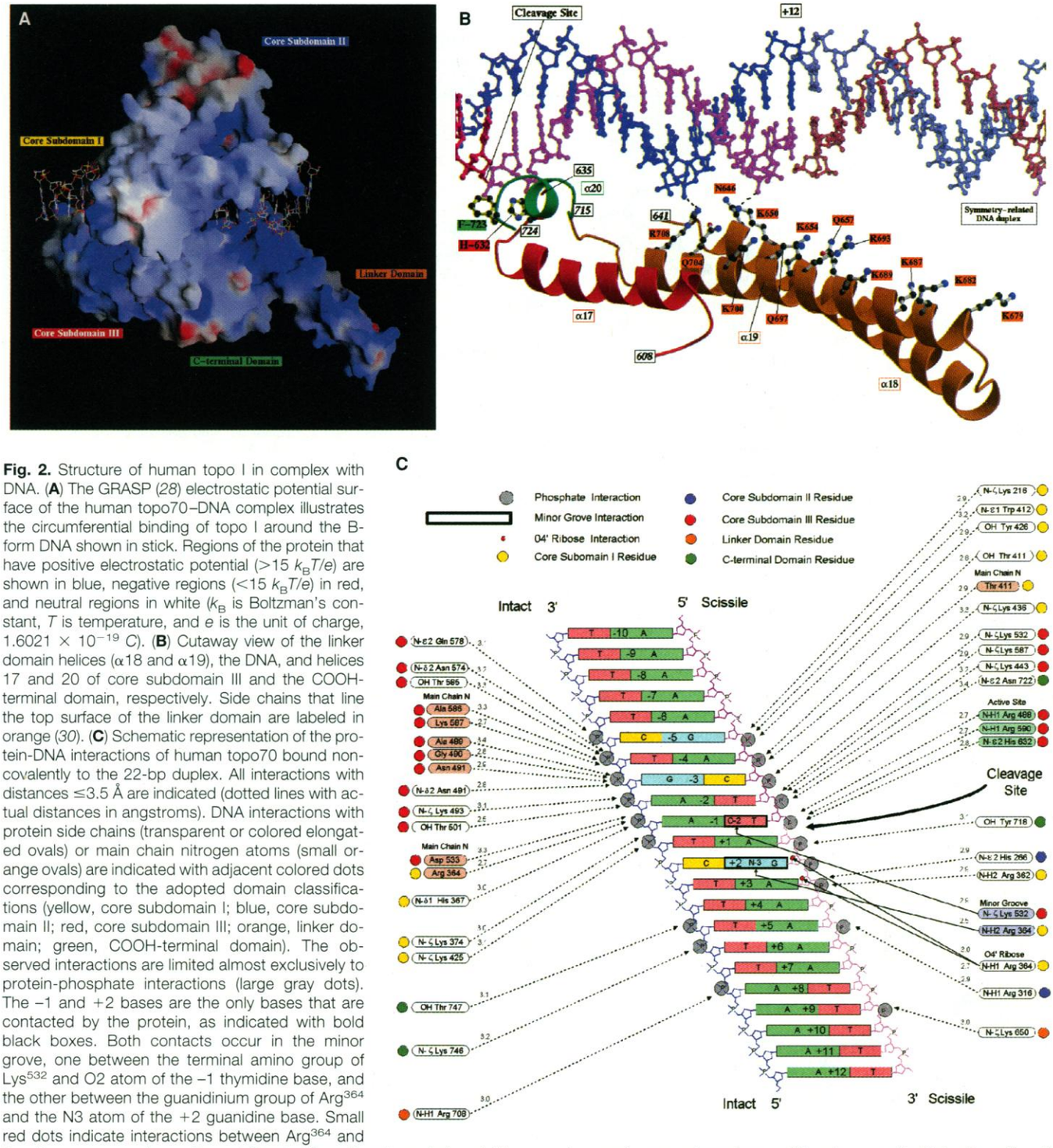

Fig. 2. Structure of human topo I in complex with DNA. (**A**) The GRASP (*28*) electrostatic potential surface of the human topo70–DNA complex illustrates the circumferential binding of topo I around the B-form DNA shown in stick. Regions of the protein that have positive electrostatic potential (>15 k_BT/e) are shown in blue, negative regions (<15 k_BT/e) in red, and neutral regions in white (k_B is Boltzman's constant, T is temperature, and e is the unit of charge, 1.6021×10^{-19} C). (**B**) Cutaway view of the linker domain helices (α18 and α19), the DNA, and helices 17 and 20 of core subdomain III and the COOH-terminal domain, respectively. Side chains that line the top surface of the linker domain are labeled in orange (*30*). (**C**) Schematic representation of the protein-DNA interactions of human topo70 bound noncovalently to the 22-bp duplex. All interactions with distances ≤3.5 Å are indicated (dotted lines with actual distances in angstroms). DNA interactions with protein side chains (transparent or colored elongated ovals) or main chain nitrogen atoms (small orange ovals) are indicated with adjacent colored dots corresponding to the adopted domain classifications (yellow, core subdomain I; blue, core subdomain II; red, core subdomain III; orange, linker domain; green, COOH-terminal domain). The observed interactions are limited almost exclusively to protein-phosphate interactions (large gray dots). The −1 and +2 bases are the only bases that are contacted by the protein, as indicated with bold black boxes. Both contacts occur in the minor grove, one between the terminal amino group of Lys[532] and O2 atom of the −1 thymidine base, and the other between the guanidinium group of Arg[364] and the N3 atom of the +2 guanidine base. Small red dots indicate interactions between Arg[364] and the ribose oxygens in positions +2 and +3 on the cleaved strand. The phosphate-deoxyribose backbone of the intact strand is shown in blue, and that of the scissile strand is shown in magenta and pink to indicate regions upstream or downstream of the cleavage site. Note, the active site Tyr[723] has been mutated to Phe in our crystal structure and does not contact the DNA.

contacting the phosphate-deoxyribose chain is repeated elsewhere in the complexes of human topo I with DNA. The "nose cone" helices 5 and 6 of core subdomains I and II also have DNA-proximal segments (residues Thr^{303} to Tyr^{338}) with six positively charged residues pointing in the direction of the duplex phosphates (*4*). However, only Arg^{316} makes direct contact with the phosphate between the +5 and +6 nucleotides on the cleaved strand (Fig. 2C). The other five positive charges do not contact the oligonucleotide duplex in any of the three crystal structures we have solved so far [this report and (*4*)]. Apparently, human topo I contains two positively charged surfaces, facing but not contacting the DNA, with each surface residing in intriguing helical portions of the enzyme. It is most likely that these helices play an important role in the mechanism of topoisomerization.

The three-dimensional structures of reconstituted human topo I in covalent and noncovalent complex with DNA (*4*) and the topo70 structure with its linker domain described here provide close-up views of the active site region before and after DNA cleavage (Fig. 3, A to C). In combination with a structure-based align-

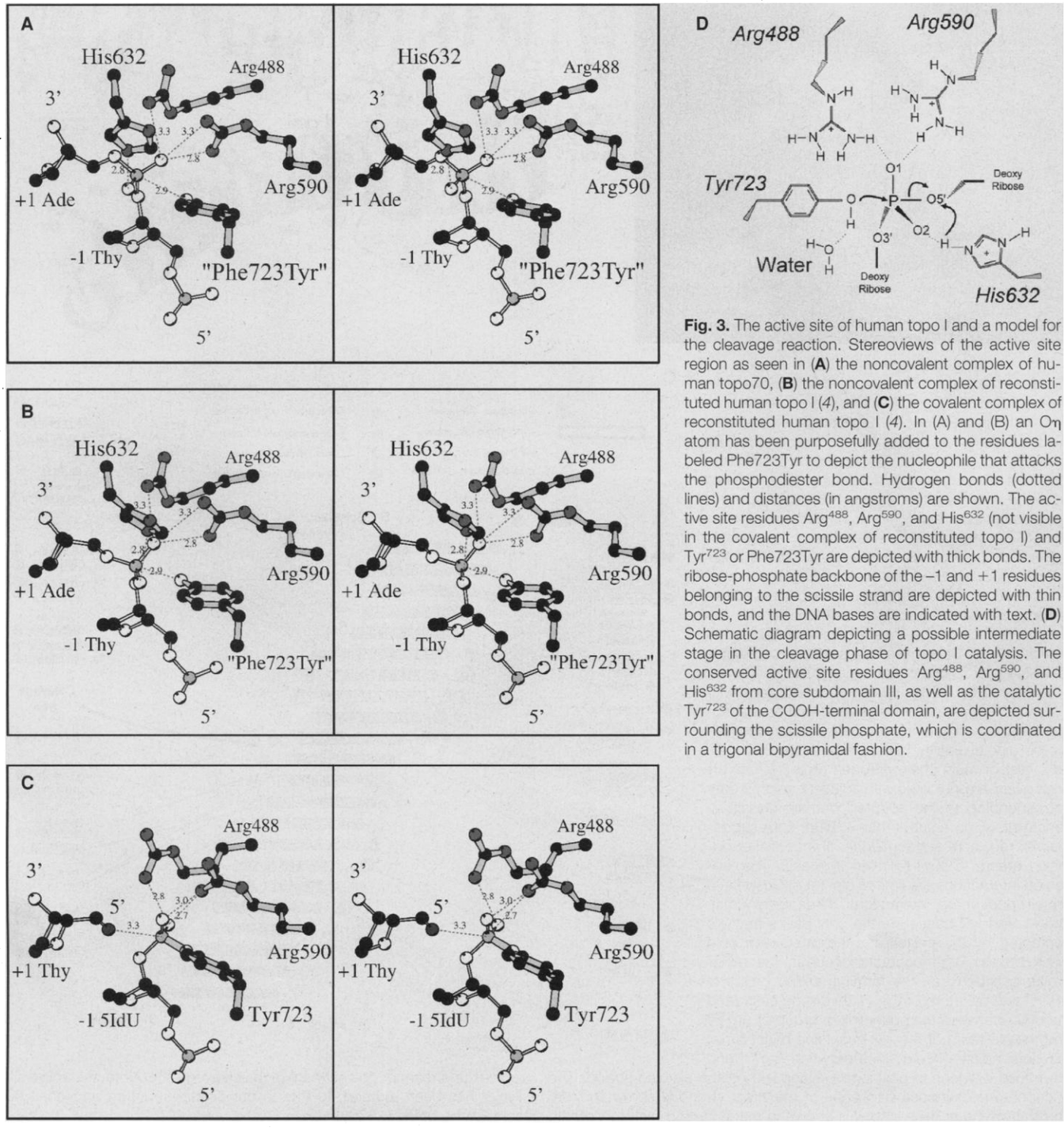

Fig. 3. The active site of human topo I and a model for the cleavage reaction. Stereoviews of the active site region as seen in (**A**) the noncovalent complex of human topo70, (**B**) the noncovalent complex of reconstituted human topo I (*4*), and (**C**) the covalent complex of reconstituted human topo I (*4*). In (A) and (B) an $O\eta$ atom has been purposefully added to the residues labeled Phe723Tyr to depict the nucleophile that attacks the phosphodiester bond. Hydrogen bonds (dotted lines) and distances (in angstroms) are shown. The active site residues Arg^{488}, Arg^{590}, and His^{632} (not visible in the covalent complex of reconstituted topo I) and Tyr^{723} or Phe723Tyr are depicted with thick bonds. The ribose-phosphate backbone of the −1 and +1 residues belonging to the scissile strand are depicted with thin bonds, and the DNA bases are indicated with text. (**D**) Schematic diagram depicting a possible intermediate stage in the cleavage phase of topo I catalysis. The conserved active site residues Arg^{488}, Arg^{590}, and His^{632} from core subdomain III, as well as the catalytic Tyr^{723} of the COOH-terminal domain, are depicted surrounding the scissile phosphate, which is coordinated in a trigonal bipyramidal fashion.

ment of all eukaryotic topoisomerase I sequences (*8*), the results of several mutagenesis studies (*12*), and a structural comparison of topo I with bacteriophage integrases (*13*), this information suggests a mechanism for phosphodiester bond breakage and religation by topo I. Replacing the observed Phe^{723} with a tyrosine residue in the Y723F mutant form of the topo70 structure (Fig. 3A), or of the reconstituted protein structure (Fig. 3B), brings the Oη oxygen to within 2.9 Å of the phosphorus atom in both crystal forms. The tyrosine hydroxyl is positioned perfectly for nucleophilic attack and subsequent covalent attachment to the 3′ end of the broken strand because it is colinear with the O5′-P scissile bond. Other residues near the phosphodiester bond that is cleaved are as follows: Arg^{488}, interacting with one nonbridging oxygen O1 (pro-*R* atom); Arg^{590}, also near O1; and His^{632} next to the other nonbridging oxygen, O2. These residues are absolutely conserved among all cellular and viral topo I enzymes (*8*), and mutagenesis studies of human, yeast, and vaccinia topo I have indicated that each of these residues plays an essential role in the nicking-closing reaction (*12*).

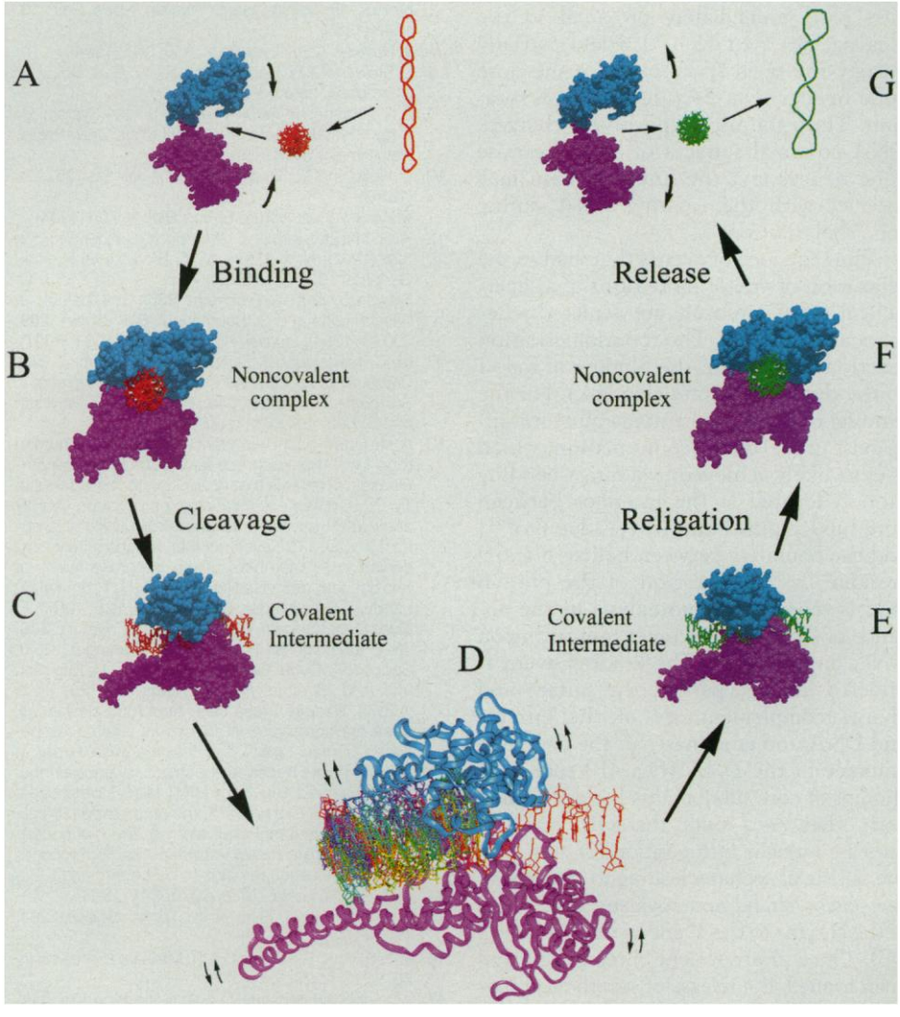

Fig. 4. The controlled rotation mechanism of human topo I. The highly negatively superhelical substrate DNA (red with right handed writhe) becomes partially relaxed through steps (**A**) through (**G**) and is converted to the less supercoiled state depicted in green. Human topo I (topo70) is rendered as a bilobed structure with core subdomains I and II forming the "cap" lobe (cyan), and core subdomain III plus the COOH-terminal and linker domains forming the "catalytic" lobe (magenta). The structure shown in (**D**) is expanded by a factor of 2 and shows the Cα trace of the protein with the rotating DNA depicted as a series of different-colored rotation states that show the DNA segment at 30° intervals. Because the shape of the enzyme is complementary to but not always in direct contact with the surface of the substrate DNA, small rocking movements (small arrows) may be allowed during the events of controlled rotation.

For the catalytic mechanism of the cellular topo I enzymes, it seems reasonable to propose that the absolutely conserved Arg^{488} and Arg^{590} residues in human topo I contribute to the stabilization of the pentavalent coordination state through hydrogen bonding to one of the nonbridging oxygen atoms of the scissile phosphate (Fig. 3D). The other nonbridging oxygen would be stabilized in the pentavalent intermediate through hydrogen bonding to the Nε2 atom of His^{632}, which is also positioned near the 5′-oxygen of the scissile bond, and might in addition function as a general acid by donating a proton to the 5′-leaving group during the cleavage reaction (Fig. 3, A, B, and D).

In the crystal structures there is no side chain atom within a 4 Å radius of the hydroxyl of Tyr^{723} that could act as a general base. The closest such amino acid is His^{632}, which is unlikely to be involved as a general base because the Nε2 atom of His^{632} is 4.3 Å away from what would be the hydroxyl oxygen of Tyr^{723} in the crystal structure of the reconstituted noncovalent complex (*4*) and 5.5 Å away in the topo70 noncovalent complex reported here. One possible explanation is that the presence of phenylalanine instead of tyrosine at position 723 in these two structural forms shifts His^{632} away from its normal position where it does indeed act as a general base to abstract the proton from the attacking hydroxyl. Alternatively, it may be that the proton is transferred to water as catalysis proceeds (Fig. 3D). Further work will be required to establish the role of a base in the cleavage reaction of human topo I.

In the structure of the covalent complex (*4*), the 5′-SH group of the +1 nucleotide is not in line for nucleophilic attack on the phosphorus atom of the tyrosine-phosphate linkage (Fig. 3C). However, if one invokes a rotation of ~180° about the C4′-C5′ bond of the terminal deoxyribose group, a 5′-OH in place of the 5′-SH would assume an ideal position for nucleophilic attack on the phosphorus atom. Our crystal structure does not reveal why the 5′-SH is inactive in religation, but it is unlikely to be due to the sulfur being intrinsically a weak nucleophile because the cysteine sulfhydryl of some protein tyrosine phosphatases has been shown to be an effective nucleophile in the hydrolysis of a phosphotyrosine bond (*14*, *15*). It seems more likely that the failure lies in the difference in the geometry of the 5′ carbon-sulfhydryl group as compared with the 5′ carbon-hydroxyl group of the natural substrate for religa-

tion. The His^{632} residue could serve as a general base to increase the nucleophilicity of the attacking 5′-OH; however, this role for His^{632} must remain provisional because this residue is not visible in the structure of the covalent complex (*4*).

Despite considerable debate, the mechanism of DNA relaxation after formation of the covalent complex and before religation remains elusive (*1*, *2*). The free rotation and strand passage models represent the two extremes of a continuum in the conceptual framework for how topo I might effect changes in linking number. The free rotation model supposes that the 5′ terminus of the broken strand is released from the active site and is allowed to rotate freely about the complementary unbroken strand. According to this model, multiple winding events could occur for each cleavage-religation cycle, which is in agreement with the observations of Stivers *et al.* on the vaccinia viral topo I (*16*). In contrast, the strand passage model proposes that the unbroken strand is passed through an enzyme-bridged gate that is formed by covalent attachment to the 3′ end and by noncovalent binding to the 5′ end of the broken strand. According to this model, only a single winding event would occur for each cleavage-religation cycle.

The structures presented here and in the accompanying article (*4*) do not reveal any protein-DNA contacts that indicate a "tight-fisted" grip on any portion of the DNA downstream of the cleavage site. Therefore, the strand passage model in its extreme form is not likely to be the mode by which topo I promotes topoisomerization (*17*). Likewise, a model in which the helix downstream of the break site is completely free to rotate around the phosphodiester bond in the intact strand can be ruled out on the basis of the physical constraints imposed by the structure itself. Thus, we conclude that relaxation occurs by an intermediate mechanism called controlled rotation, in which ionic interactions between the DNA and both the nose cone helices and the linker domain regulate the winding process.

The considerations that led to the controlled rotation hypothesis are as follows. Initial modeling studies of topoisomerization demonstrated that severe protein-DNA clashes would occur if the DNA downstream of the cleavage site were allowed to rotate about any one of the five rotatable bonds spanning the sugar phosphate bond network between the −1 and +1 bases of the intact strand. Such clashes would occur even if core subdomains I and II were allowed to fully dissociate from the DNA substrate. This observation eliminates all models that involve the simple rotation about a single bond in the intact strand. We next examined the predicted movements of the downstream DNA, assuming that the flexibility in the DNA could be mimicked by a pseudobond anchored on the 3′ bridging oxygen atom of the phosphate between the −1 and +1 nucleotides in the intact strand. If we modeled the DNA rotation about such a pseudobond that was oriented along the helical axis and tilted ~10° downward away from the nose cone helices, the downstream region of the DNA helix would come into close proximity with the distal portion of the linker domain, and after ~180° of rotation, just pass the nose cone helices (Fig. 4D). In addition, the base pairs immediately proximal to the cleavage site (+1 to +3) would partially occupy the open space between the nose cone helices (Fig. 2A) during DNA rotation. Thus, the highly positively charged, DNA-proximal surfaces of both the nose cone helices and the linker domain may interact with the rotating DNA during topoisomerization.

The sequence of events that lead to the relaxation of one or more turns of a superhelical DNA molecule are depicted schematically in Fig. 4. The topoisomerization reaction begins with the binding of topo I to the duplex substrate (Fig. 4A). For the binding to occur, the enzyme must initially exist in an "open" conformation, which is most likely achieved by a hinge-bending motion located at the interface between core subdomains I and III (residue Pro^{431}) and the boundary between helices 8 and 9 (residue Lys^{452}), a region of the protein that is sensitive to proteolysis in the absence of DNA and becomes resistant upon DNA binding (*18*). The binding event is directed in large part by the surface and charge complementarity of the enzyme and DNA and culminates in the complete embrace of the DNA (Fig. 4B) such that the lips of core subdomains I and III touch each other. As a result, the active site residues are brought into position for attack on the scissile phosphate, leading to cleavage of the scissile strand and covalent attachment of the enzyme to the 3′ end of the DNA (Fig. 4C). Once the covalent intermediate has been formed, the release of superhelical tension can occur through one or more cycles of controlled rotation (Fig. 4D). To illustrate this process, we magnified the protein by a factor of 2 and rendered it as a Cα trace, and the rotating DNA downstream of the break is depicted as a rainbow of conformers at 30° intervals (Fig. 4D). Any rotation at this stage is presumably driven by torsional strain within the DNA. Subsequently, the covalent intermediate (Fig. 4E) is religated with concomitant release of the Tyr^{723} from the end of the DNA (Fig. 4F). Finally, a DNA molecule with reduced superhelicity is released (Fig. 4G), allowing the enzyme to undergo another cycle of DNA binding and relaxation.

REFERENCES AND NOTES

1. J. J. Champoux, *Prog. Nucleic Acid Res. Mol. Biol.*, in press.
2. J. J. Champoux, in *DNA Topology and its Biological Effects*, J. C. Wang and N. R. Cozzarelli, Eds. (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1990), pp. 217–242.
3. J. C. Wang, *Annu. Rev. Biochem.* **65**, 635 (1996).
4. M. R. Redinbo, L. Stewart, P. Khun, J. J. Champoux, W. G. J. Hol, *Science* **279**, 1534 (1998).
5. C. D. Lima, J. C. Wang, A. Mondragon, *Nature* **367**, 138 (1994).
6. A. H. Andersen, E. Gocke, B. J. Bonven, O. F. Nielsen, O. Westergaard, *Nucleic Acids Res.* **13**, 1543 (1985).
7. J. Navaza, *Acta Crystallogr.* **A50**, 157 (1994).
8. L. Stewart, M. R. Redinbo, X. Qiu, W. G. J. Hol, J. J. Champoux, data not shown.
9. A. T. Brünger, *X-PLOR (Version 3.1): A System for X-ray Crystallography and NMR* (Yale Univ. Press, New Haven, CT, 1993).
10. L. Holm and C. Sander, *Nucleic Acids Res.* **25**, 231 (1997).
11. M. Vlassi *et al.*, *Nature Struct. Biol.* **1**, 706 (1994).
12. K. R. Madden and J. J. Champoux, *Cancer Res.* **52**, 525 (1992); N. A. Levin, M. A. Bjornsti, G. R. Fink, *Genetics* **133**, 799 (1993); A. D. Jensen and J. Q. Svejstrup, *Eur. J. Biochem.* **236**, 389 (1996); J. Wittschieben and S. Shuman, *J. Biol. Chem.* **269**, 29978 (1994); *Nucleic Acids Res.* **25**, 3001 (1997); M. D. Megonigal, J. Fertala, M. A. Bjornsti, *J. Biol. Chem.* **272**, 12801 (1997); C. Cheng, L. K. Wang, J. Sekiguchi, S. Shuman, *ibid.*, p. 8263; B. O. Petersen and S. Shuman, *ibid.*, p. 3891.
13. A structure-based sequence alignment of human topo I with the integrase family of tyrosine recombinases (*8*) reveals that in addition to the active site Tyr^{723}, the two essential arginine residues, Arg^{488} and Arg^{590}, have structural equivalents in the lambda, HP1, and Cre systems that are absolutely conserved in all members of the integrase family of tyrosine recombinases [H. J. Kwon, R. Tirumalai, A. Landy, T. Ellenberger, *Science* **276**, 126 (1997); F. Guo, D. N. Gopaul, G. D. VanDuyne, *Nature* **389**, 40 (1997); A. B. Hickman, S. Waninger, J. J. Scocca, F. Dyda, *Cell* **89**, 227 (1997); D. Esposito and J. J. Scocca, *Nucleic Acids Res.* **25**, 3605 (1997)]. Despite these clear structural similarities, there are some distinct functional differences between human topo I and the integrase family of recombinases because the structural equivalent of the conserved His^{280} of HP1, His^{308} of lambda, and His^{289} of Cre is Lys^{587} in human topo I. Moreover, the functional equivalent of the highly conserved and catalytically important His^{632} of topo I in the recombinases remains to be determined.
14. J. A. Stuckey *et al.*, *Nature* **370**, 571 (1994).
15. D. Barford, A. J. Flint, N. K. Tonks, *Science* **263**, 1397 (1994).
16. J. T. Stivers, T. K. Harris, A. S. Mildvan, *Biochemistry* **36**, 5212 (1997).
17. Although the capacity of eukaryotic topo I to catenate circular DNAs containing a single-strand break was originally interpreted as strong evidence in support of the strand passage hypothesis [P. O. Brown and N. R. Cozzarelli, *Science* **206**, 1081 (1979)], an alternative pathway involving a linear intermediate has also been proposed to account for the catenation reaction [W. K. McCoubrey Jr. and J. J. Champoux, *J. Biol. Chem.* **261**, 5130 (1986)].
18. L. Stewart, G. C. Ireton, J. J. Champoux, *J. Biol. Chem.* **271**, 7602 (1996).
19. L. Stewart, G. C. Ireton, L. H. Parker, K. R. Madden, J. J. Champoux, *ibid.*, p. 7593.
20. R. J. Read, *Acta Crystallogr.* **A42**, 140 (1986).
21. E. de La Fortelle and G. Bricogne, *Methods Enzymol.* **276**, 472 (1997).

22. T. A. Jones, J.-Y. Zou, S. W. Cowan, *Acta Crystallogr.* **A47**, 110 (1991).
23. A. T. Brünger, *J. Mol. Biol.* **203**, 803 (1988).
24. J. M. Moore, W. J. Chazin, R. Powls, P. E. Wright, *Biochemistry* **27**, 7806 (1988).
25. P. J. Kraulis, *J. Appl. Crystallogr.* **24**, 946 (1991).
26. D. J. Bacon and W. F. Anderson, *J. Mol. Graphics* **6**, 219 (1988).
27. E. A. Merritt and M. E. P. Murphy, *Acta Crystallogr.* **D50**, 869 (1994).
28. A. Nicholls, K. A. Sharp, B. Honig, *Proteins Struct. Funct. Genet.* **11**, 282 (1991).
29. C. E. Stebbins *et al.*, *Nature* **373**, 636 (1995).
30. Single-letter abbreviations for the amino acid residues are as follows: A, Ala; C, Cys; D, Asp; E, Glu; F, Phe; G, Gly; H, His; I, Ile; K, Lys; L, Leu; M, Met; N, Asn; P, Pro; Q, Gln; R, Arg; S, Ser; T, Thr; V, Val; W, Trp; and Y, Tyr.
31. We thank G. C. Ireton, E. Pohl, S. Turley, B. Bernstein, C. Verlinde, F. Athappilly, F. van den Akker, D. Chudzik, J. Yeh, and R. Li for their technical assistance and many engaging conversations; and Y. Pocker for helpful comments concerning the catalytic mechanism; H. Bellamy and P. Kuhn at the Stanford Synchrotron Radiation Laboratory for help with MAD and isomorphous data collection; and the staffs of Brookhaven National Laboratory, the Cornell High Energy Synchrotron Source, and the European Synchrotron Radiation Facility for their support with data collection. This work was supported by grant GM49156 from NIH (to J.J.C.), by American Cancer Society grant PF-3905 (to L.S.), by an NIH fellowship grant GM16713 to X.Q., and by a major equipment grant from the Murdock Charitable Trust to the Biomolecular Structure Center and grant CA65656 from NIH (to W.G.J.H).

1 December 1997; accepted 28 January 1998

Natural Ligand of Mouse CD1d1: Cellular Glycosylphosphatidylinositol

Sebastian Joyce,* Amina S. Woods, Jonathan W. Yewdell, Jack R. Bennink, A. Dharshan De Silva, Alina Boesteanu, Steven P. Balk, Robert J. Cotter, Randy R. Brutkiewicz

Mouse CD1d1, a member of the CD1 family of evolutionarily conserved major histocompatibility antigen–like molecules, controls the differentiation and function of a T lymphocyte subset, NK1$^+$ natural T cells, proposed to regulate immune responses. The CD1d1 crystal structure revealed a large hydrophobic binding site occupied by a ligand of unknown chemical nature. Mass spectrometry and metabolic radiolabeling were used to identify cellular glycosylphosphatidylinositol as a major natural ligand of CD1d1. CD1d1 bound glycosylphosphatidylinositol through its phosphatidylinositol aspect with high affinity. Glycosylphosphatidylinositol or another glycolipid could be a candidate natural ligand for CD1d1-restricted T cells.

The *CD1* region encodes a family of evolutionarily conserved proteins that closely resemble the classical antigen presenting major histocompatibility complex (MHC) molecules (*1*, *2*). CD1d controls the function of NK1$^+$ natural T (NKT) cells (*3*, *4*), an unusual subset of T lymphocytes that express receptors for natural killer cells (NKRP-1C) and T cells [αβ or γδ T cell receptor (TCR)] (*5*). NKT cells are thought to play an immunoregulatory role in responses to foreign and self antigens (*6*, *7*). Maturation of NKT cells depends on the interaction of the TCR with CD1d1 (*8*–*10*). It is unclear whether this interaction requires the display of specific ligands in the CD1d1 groove akin to those presented by MHC molecules. Because the TCR repertoire of NKT cells is highly restricted (*11*–*13*), they probably interact with ligand-free CD1d1 or with CD1d1 displaying a highly conserved ligand. The three-dimensional structure of CD1d1 revealed a large hydrophobic binding site occupied by ligands whose chemical nature could not be determined with the diffraction data (*2*).

Natural ligands of CD1d1 were isolated from purified CD1d1 molecules expressed by TAP (transporters associated with antigen processing)-deficient human T2 cells infected with a recombinant vaccinia virus expressing the *CD1d1* gene (*3*, *14*). Low molecular weight ligands associated with CD1d1 and HLA class I molecules (the latter serving as controls) were isolated and fractionated by reversed-phase high-performance liquid chromatography (RP-HPLC) as described (*15*, *16*). Several peaks were observed at 210 nm (detects peptide and double bonds) that were associated only with CD1d1-derived material (Fig. 1A). Very little, if any, absorbance was detected at 254 or 280 nm (detects aromatic amino acids) (*17*). The major peak present in CD1d1-derived material (Fig. 1A, arrow) corresponds to a minor peak in the fractions from HLA-associated ligands. Each of the six fractions constituting the major peak (about 82 and 85 min) were refractionated by RP-HPLC with a mixture of C_{18} and cation-exchange (1:1) matrix (*18*). This yielded single peaks that behaved in an indistinguishable manner (*17*). Similar refractionation of the HLA class I–associated material eluting between about 82 and 85 min did not yield any signal (*17*), probably because the chemical constitution of this material prevented its recovery.

Thus, the major peak eluting between about 82 and 85 min represents a CD1d1 ligand. To identify this material, we subjected it to matrix-assisted laser-desorption/ionization mass spectrometric (MALDI-MS) analysis (*19*). The positive ion mass spectra of each of the six refractionated fractions contained peaks at m/z 598, 614, 636, 1211, and 1227 (Fig. 1B), conclusively showing their identity to one another. MALDI-MS analysis after digestion of these samples with either carboxypeptidase or aminopeptidase indicated that the ligands associated with CD1d1 were not peptidic (*17*).

The CD1d1-associated natural ligand was incubated with ammonium sulfate (*19*) to enhance the ion signal in the mass spectrometer (*20*). This resulted in the loss of signals at m/z 1211 and 1227 and the gain of signal at m/z 887 (Fig. 1C). The peak at m/z 887 was consistent with the mass of a protonated molecular ion (MH^+) of a phospholipid (*21*). The negative ion mass spectrum (*19*) of the sample treated with ammonium sulfate revealed a $(MH)^-$ ion of the same species at m/z 885.3 and a fragment ion at m/z 705.6 (Fig. 1D) resulting from the characteristic, and nearly diagnostic, loss of an inositol head group (*22*).

From these three mass spectra, one solution to the identity of the peak at m/z 1211 is that it is the MH^+ ion for a phosphatidylinositol-diglycoside containing stearic and arachidonic acids, glucosamine, and mannose, with a calculated mass of 1211.4 daltons. Because it contains a negatively charged phosphate group, it is unstable and hence of relatively low intensity in the positive ion mass spectrum (Fig. 1B). Under high laser power it decomposes, cleaving the C^2-C^3 bond of glycerol to form a stable fragment ion at m/z 614 that lacks a phosphate. The m/z 614 ion was the most abundant ion in the mass spectrum and was used to determine the identities of the acyl

S. Joyce, A. D. De Silva, A. Boesteanu, Department of Microbiology and Immunology, The Pennsylvania State University College of Medicine, Milton S. Hershey Medical Center, Hershey, PA 17033–0850, USA.
A. S. Woods and R. J. Cotter, Departments of Oncology and Pharmacology and Molecular Sciences, The Johns Hopkins University School of Medicine, Baltimore, MD 21205–2185, USA.
J. W. Yewdell, J. R. Bennink, R. R. Brutkiewicz, Laboratory of Viral Diseases, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, MD 20892–0440, USA.
S. P. Balk, Hematology-Oncology Division, Beth Israel Deaconess Medical Center, Harvard Medical School, Boston, MA 02215, USA.

*To whom correspondence should be addressed. E-mail: sjoyce@bcmic.hmc.psu.edu