Structure, Vol. 10, 237–248, February, 2002, ©2002 Elsevier Science Ltd. All rights reserved. PII S0969-2126(02)00707-4

The Crystal Structure of Human Tyrosyl-DNA Phosphodiesterase, Tdp1

Douglas R. Davies,[1] Heidrun Interthal,[2] James J. Champoux,[2] and Wim G.J. Hol[1,3,4]

[1]Department of Biochemistry
Box 357742
University of Washington
Seattle, Washington 98195

[2]Department of Microbiology
Box 357242
School of Medicine
University of Washington
Seattle, Washington 98195

[3]Howard Hughes Medical Institute
University of Washington
Seattle, Washington 98195

Summary

Tyrosyl-DNA phosphodiesterase (Tdp1) catalyzes the hydrolysis of a phosphodiester bond between a tyrosine residue and a DNA 3′ phosphate. The enzyme appears to be responsible for repairing the unique protein-DNA linkage that occurs when eukaryotic topoisomerase I becomes stalled on the DNA in the cell. The 1.69 Å crystal structure reveals that human Tdp1 is a monomer composed of two similar domains that are related by a pseudo-2-fold axis of symmetry. Each domain contributes conserved histidine, lysine, and asparagine residues to form a single active site. The structure of Tdp1 confirms that the protein has many similarities to the members of the phospholipase D (PLD) superfamily and indicates a similar catalytic mechanism. The structure also suggests how the unusual protein-DNA substrate binds and provides insights about the nature of the substrate in vivo.

Introduction

Tyrosyl-DNA phosphodiesterase (Tdp1) catalyzes the hydrolysis of a phosphodiester bond between a tyrosine residue and a DNA 3′ phosphate [1]. The only known example of such a linkage in eukaryotic cells occurs in the enzyme-DNA covalent complex formed when a type IB DNA topoisomerase cleaves DNA. Therefore, it has been suggested that Tdp1 is responsible for cleaving such complexes when they become stalled on the DNA in the cell [1]. Although there is little evidence for the formation of such complexes in vivo, there is abundant in vitro evidence showing that a variety of conditions can prevent or slow religation by topoisomerase I and therefore cause the trapping of the enzyme on the DNA. These conditions include DNA damage leading to the generation of nicks, gaps, or abasic sites, the incorporation of nucleotide analogs, such as Ara-C, or treatment with drugs, such as camptothecin, that poison the topoisomerase reaction (reviewed in [2–4]). The cytocidal effect of camptothecin appears to require DNA synthesis [5–8], suggesting that it is the collision of a replication fork with the stalled topoisomerase I that leads to cell death [9, 10]. Since Tdp1 theoretically counteracts the effects of camptothecin and its derivatives, drugs that inhibit the phosphodiesterase would be predicted to act synergistically with this class of topoisomerase poisons and could be valuable in a combined drug therapy regimen.

It has recently been shown, based on sequence comparisons and on a mutational analysis of active site residues in human Tdp1, that the enzyme belongs to the phospholipase D (PLD) superfamily [11]. This conclusion is confirmed by the demonstration of a phosphoenzyme covalent intermediate in the reaction [11]. Other members of the PLD superfamily, besides phospholipases D, are bacterial phosphatidyl serine, cardiolipin synthetases, a bacterial nuclease (Nuc) from *Salmonella typhimurium*, a bacterial toxin, and a number of poxvirus envelope proteins [12–14]. The N-terminal regions of these proteins are poorly conserved, and, indeed, the first 148 amino acids of human Tdp1 can be removed without an effect on catalysis in vitro [11]. With the exception of the poxvirus proteins, which may lack enzymatic activity, all of the enzymes in the superfamily catalyze a phosphoryl transfer reaction where the acceptor is either an alcohol or water. The similar reaction chemistry for the members of the family appears to derive from a pair of motifs (referred to as HKD motifs), each of which contains a highly conserved histidine, lysine, and aspartate residue in the sequence $HxK(x)_4D$ [14]. Since the Tdp1 motifs contain the conserved histidine and lysine residues but lack the aspartate residues, this enzyme represents a distinct class within the PLD superfamily [11]. It has been suggested, based on an analysis of reaction intermediates and products, that catalysis proceeds with the formation of a phosphoenzyme intermediate that is subsequently cleaved to form the final product [15–17]. In two cases, a histidine in one of the HKD motifs has been directly implicated as the nucleophile in the reaction leading to the covalent intermediate [18, 19]. Crystal structures of Nuc and a phospholipase D reveal that the conserved active site histidine and lysine residues are clustered and bound to phosphate or tungstate ligands in a way that is consistent with the proposed chemistry [20, 21].

The exact substrate for Tdp1 in the repair of topoisomerase I-DNA covalent complexes in vivo is uncertain, but a complex containing the native form of human topoisomerase I is cleaved poorly, if at all, by human Tdp1 (H. Interthal and J. Champoux, unpublished data). This result is understandable in view of the structure of the topoisomerase I-DNA covalent complex, where the tyrosyl-DNA phosphodiester bond is completely buried within the complex [22]. Interestingly, a bacteriophage Int protein-DNA complex (which also contains a 3′ DNA

[4]Correspondence: hol@gouda.bmsc.washington.edu

Key words: camptothecin; DNA repair; PLD superfamily; topoisomerase I; tyrosyl-DNA phosphodiesterase

Table 1. Data Collection and Refinement Statistics

	Peak	Inflection	Remote
Data Collection			
λ (Å)	0.9792	0.9794	0.9640
Space group	$P2_12_12$	$P2_12_12$	$P2_12_12$
Unit cell dimensions, a, b, c (Å)	52.03, 56.31, 186.45	52.11, 56.36, 186.60	52.17, 56.39, 186.71
Resolution (Å)	1.69	1.69	1.69
Mosaicity (°)	0.238	0.281	0.329
Unique reflections	59,625	59,162	57,856
Completeness (highest shell; %)	96.0 (76.4)	95.2 (72.9)	93.1 (66.2)
Average I/σ[a]	13.42	16.06	18.80
Redundancy	4.48	3.58	3.33
R_{sym} (%)[b]	6.3 (19.2)	4.8 (23.4)	5.1 (42.4)
Figure of merit	0.540		
Refinement			
Resolution range (Å)	30.0–1.69		
Reflections (free)	66,989 (3,349)		
$R_{crystal}/R_{free}$ (%)[c]	19.8/22.9		
Rms deviations from ideality			
Bond lengths (Å)	0.018		
Bond angles (°)	1.813		
Chirality	0.137		

[a] I/σ is the mean reflection intensity divided by the estimated error.
[b] $R_{sym} = (\Sigma|I_{hkl} - <I>|)/(\Sigma I_{hkl})$, where the average intensity $<I>$ is taken over all symmetry equivalent measurements and I_{hkl} is the measured intensity for any given reflection.
[c] $R_{crystal} = ||F_o| - |F_c||/|F_o|$, where F_o and F_c are the observed and calculated structure factor amplitudes, respectively. R_{free} is equivalent to $R_{crystal}$, but calculated for 5% of the reflections chosen at random and omitted from the refinement process.

phosphotyrosine linkage) is poorly cleaved by *Saccharomyces cerevisiae* Tdp1; however, prior heating of the Int-DNA complex to 65°C greatly increased cleavage by the enzyme [1]. These observations suggest that the initial topoisomerase I-DNA covalent complex must be processed in some way to make the tyrosine-DNA phosphodiester bond accessible to Tdp1 [1].

Here, we present the crystal structure of a fully active N-terminal truncation of human Tdp1 protein that is missing the first 148 amino acids. The crystal structure firmly establishes the relationship with members of the PLD superfamily and provides important insights into the chemistry of catalysis. Moreover, our structure suggests how this unusual enzyme may interact with its protein-DNA substrate.

Results and Discussion

Description of Human Tdp1 Δ1−148 Structure

The crystal structure of human Tdp1 was solved by multiwavelength anomalous dispersion methods from selenomethionine-labeled protein. The 1.69 Å structure has a crystallographic R factor of 19.8% and an R_{free} of 22.9%. Data collection, processing, and structural refinement statistics are listed in Table 1, and a representative view of the electron density map is shown in Figure 1. The protein construct was an N-terminal truncation of human Tdp1 missing the first 148 amino acid residues that retained full catalytic activity [11]. Using the numbering system of the full-length protein, the three-dimensional model extends from residue Gly149 to the C terminus of human Tdp1 (Ser608), with three disordered loops between His382 and Ser391, Gly424 and Val435, and Lys555 and Ala568. The construct also contains a 25-residue N-terminal tag that includes a $(His)_6$ affinity sequence, a linker, and a thrombin cleavage site. Although this affinity tag is intact in the crystallized protein, only four residues at the C-terminal end of the tag are visible in the electron density map and are designated as Glu145′–Pro148′. The structure includes a *cis*-prolyl peptide bond between Leu576 and Pro577. Evaluation of the model with PROCHECK [23] shows that 89.5% of the residues lie in the most favored regions of the Ramachandran plot, 9.7% in the additionally allowed regions, 0.8% in the generously allowed regions, and none in the disallowed regions.

The Tdp1 structure (Figures 2A and 2B) contains 11 α helices and 17 β strands as identified by DSSP [24]. For simplicity, the designations of DSSP were modified so that α turns that included only a single hydrogen bond were not classified as numbered α-helical segments. The tertiary structure can be described as an α-β-α-β-α sandwich composed of two α-β-α domains that are related by a pseudo-2-fold axis of symmetry. The N-terminal domain extends from Gly149 to Thr350 (Figure 2, blue), and the C-terminal domain runs from Asn351 to Ser608 (Figure 2, yellow). Each domain has in common a seven-stranded mixed parallel-antiparallel β sheet and two conserved α helices (α1 and α2 of the N-terminal domain correspond to α6 and α7 of the C-terminal domain, respectively). The directionality and position of the β strands are conserved between the two seven-stranded sheets. Each domain contains one of the conserved HKD sequence motifs. A superposition of the two domains (Figure 2C) reveals an rms difference for 111 conserved α carbons between the two domains of 5.20 Å, with sequence identity of 10.0%. Beyond this nearly symmetrical scaffold, however, several differ-

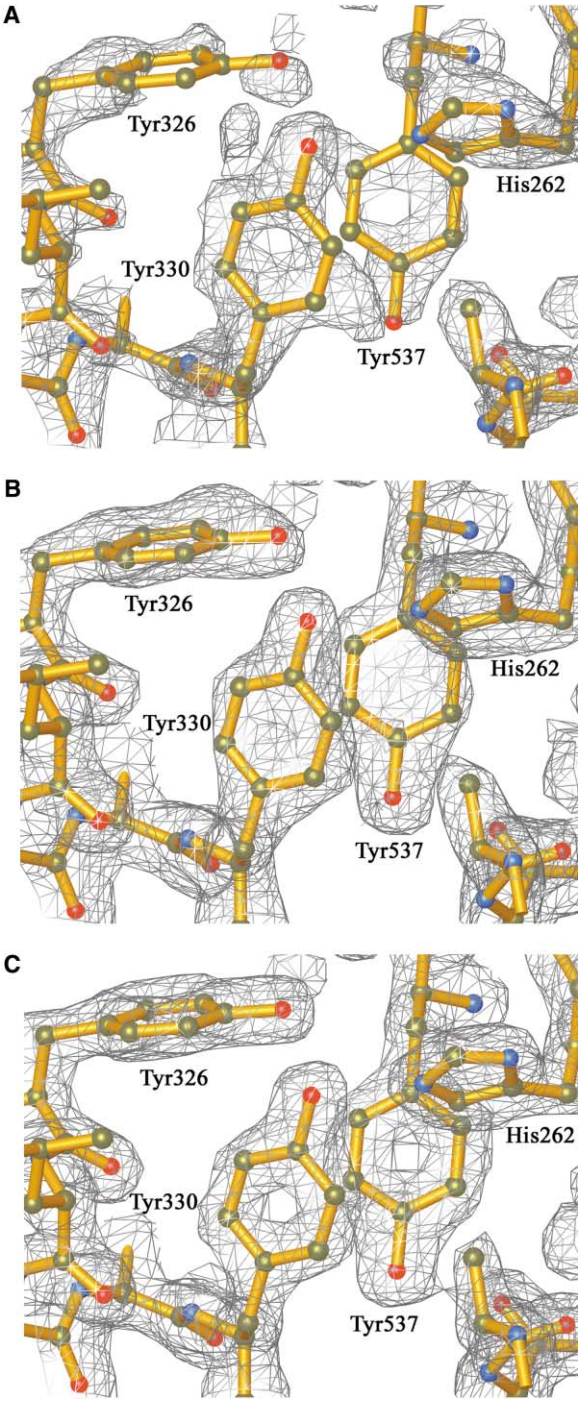

Figure 1. Electron Density Associated with a Cluster of Tyrosine Residues in the C-terminal Domain of Tdp1

(A) The initial electron density map calculated by SOLVE [32].
(B) The solvent-flattened electron density output from Resolve [33].
(C) The final σ_A-weighted $2F_o - F_c$ map calculated by Refmac5 [27].
In all cases, the maps are contoured at 1.5 σ. Figures 1, 2, 6, 7, and 8 were generated with the program Ribbons [37].

ences distinguish the two domains of Tdp1. The N-terminal domain has three α-helical segments that are not found in the C-terminal domain (α3, α4, and α5). A coil consisting of eight N-terminal residues of Tdp1 Δ1–148 (including four residues of the affinity tag linker) protrudes from one side of the globular protein, in a conformation that is stabilized by crystal-packing interactions with a neighboring Tdp1 monomer. The C-terminal domain is 60 residues larger than the N-terminal domain. Four α helices that are not present in the N-terminal domain (α8, α9, α10, and α11), an additional β sheet consisting of three short antiparallel strands (β9, β15, and β16), and an extended coil at the C terminus contribute to the larger size. The variations in secondary structure elements and loops that surround the symmetrical core of Tdp1 create a highly asymmetric molecular surface.

The active site region containing the pairs of highly conserved histidine and lysine residues of the HKD motifs of the PLD superfamily is centered at the pseudo-2-fold axis of symmetry between the two domains, at a point where the convex sides of the large β sheets from each domain come in closest contact with one another. The active site residues lie near the middle of an elongated cleft that runs along one entire surface of the protein. Loops β4-β5 and β12-β13 contribute the key histidine and lysine residues of the HKD motifs that have been shown by mutagenesis studies to be important for catalysis: His263, Lys265, His493, and Lys495 [11] (Figure 2B). Four additional residues that are conserved in all Tdp1 orthologs, Asn283, Gln294, Asn516, and Glu538, are also located in the active site. Asn283 and Gln294 are present on loop β6-β7, Asn516 is found on loop β14-α11, and Glu538 is in strand β17. The possible role of some of these active site residues in catalysis will be discussed below.

The constellation of residues in the active site retains the symmetry of the overall tertiary structure, with each domain contributing a histidine, lysine, and asparagine residue (Figure 3). In each domain, the Nζ of an active site lysine is within hydrogen bonding distance of the backbone carbonyl oxygen of the histidine residue from the same domain. The conserved asparagine residues Asn283 and Asn516 are symmetrically oriented toward the center of the active site, and each interacts with a water molecule in the active site via hydrogen bonds to Nδ2. The symmetry of the active site is broken by the nature of two residues that interact with the active site histidines. The Nδ1 atom of His263 on the N-terminal domain is within hydrogen bonding distance of a carboxyl oxygen of Glu538 from the C-terminal domain, which is in turn within hydrogen bonding distance of Oγ of Ser514. The backbone carbonyl oxygen of Ser514 is also within hydrogen bonding distance of Nδ1 of His263, completing a hydrogen bonding circuit between the three residues. A similar hydrogen bonding pattern is found on the other domain related by the pseudo-2-fold axis of symmetry, except that His493 interacts with glutamine (Gln294) and threonine (Thr281) residues instead of with glutamic acid and serine. Another asymmetry in the active site can be found in the arrangement of two water molecules (designated Wat1 and Wat2) that are well ordered in the electron density map. Whereas Wat1 is within hydrogen bonding distance of Nδ2 of Asn283, Nϵ2 of His263, and Nϵ2 of His493, Wat2 is positioned within hydrogen bonding distance of Nζ of Lys495, Nδ2 of Asn516, Nϵ2 of His263, and Nϵ2 of

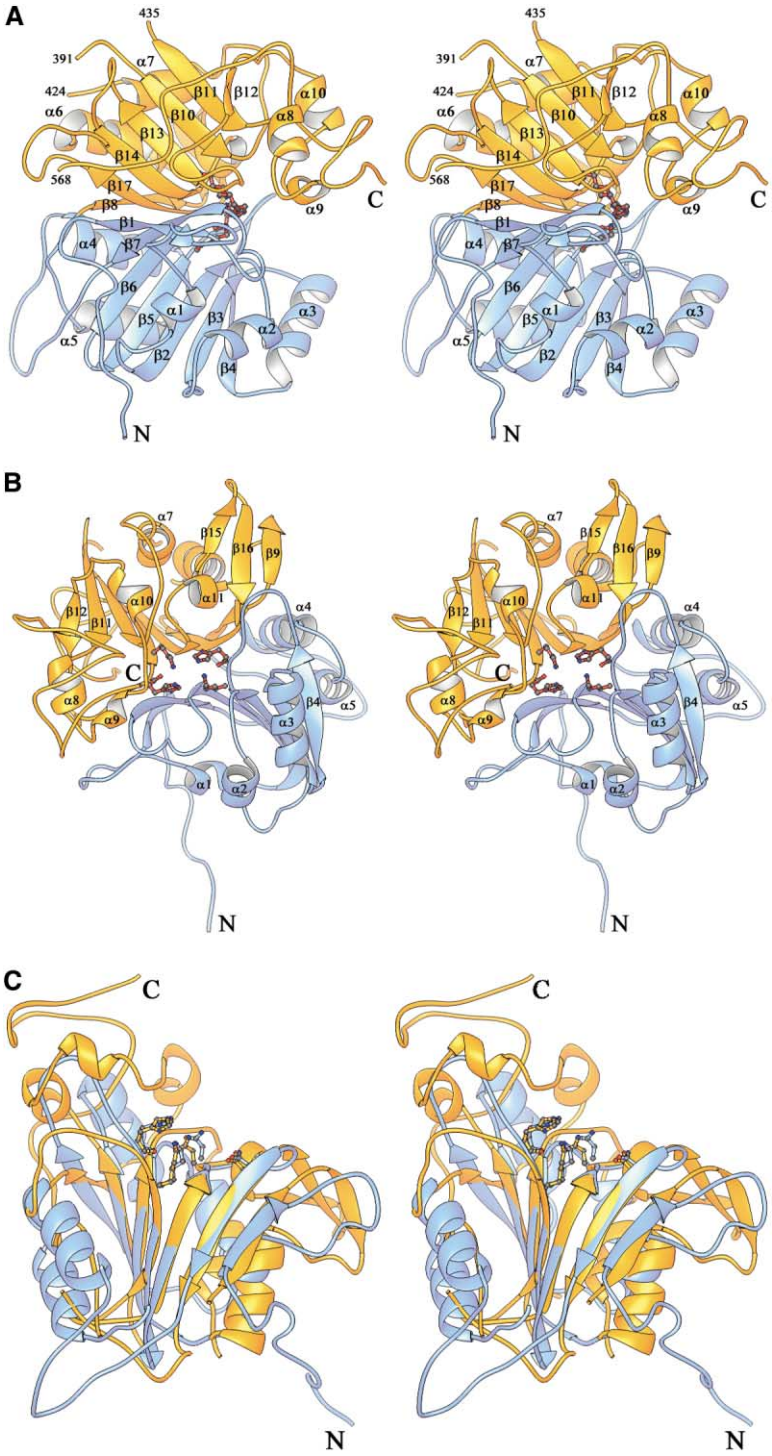

Figure 2. Ribbon Representations of the Structure of Human Tdp1

(A) Stereoview of Tdp1 colored to show the two-domain structure of the protein. The N-terminal domain of Tdp1 (residues 149–350) is colored blue, and the C-terminal domain (residues 351–608) is colored yellow. The view is perpendicular to the pseudo-2-fold axis of symmetry between the two domains. The active site residues His263, Lys265, His493, and Lys495 are depicted as red ball and stick structures.

(B) Same as (A), rotated to view the molecule parallel to the pseudo-2-fold axis of symmetry.

(C) Superposition of the two domains of Tdp1. Coloration is the same as in (A) and (B), except that the active site residues have the same coloration as their respective domains.

His493. Thus, the overall arrangement of the active site is pseudosymmetrical, with the two key histidine residues existing in distinctly different chemical environments.

The active site lies at the bottom and near the center of an approximately 40 Å-long cleft that runs along the entire "top" surface of Tdp1, nearly perpendicular to the domain-domain boundary (Figures 2B and 4). The shape of this substrate binding cleft is well defined, as all three of the disordered loops in the Tdp1 structure are located on the opposite side of the molecule. As may be expected for an enzyme that hydrolyzes protein-DNA covalent complexes, the substrate binding cleft is asymmetrical in terms of both shape and charge distribution (Figure 4). On one side of the active site, the cleft is relatively narrow and lined with predominantly positively charged amino acid residues. On the other side of the active site, the cleft has a greater negative charge distribution and flares out into a more open, bowl-shaped

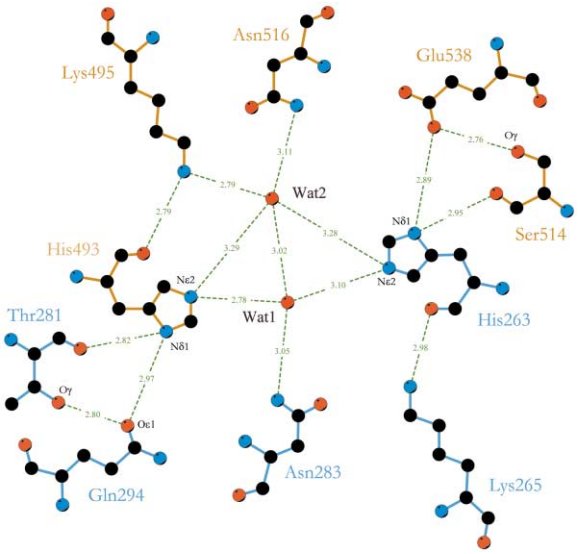

Figure 3. Schematic View of the Hydrogen Bonding Pattern Observed in the Active Site of Tdp1

Bonds of residues from the N-terminal domain of Tdp1 are colored blue, and residues from the C-terminal domain are colored yellow. This view was created with the program LIGPLOT [38].

basin. The shape and surface charge of the substrate binding cleft suggest that the narrow, positively charged side may act as a DNA binding region, while the more open side with the more negative charge distribution is better suited to binding the protein moiety of the substrate. The dimensions of the putative DNA binding side of the substrate binding cleft may provide clues to the general mode of DNA binding by the enzyme: the cleft is narrow (approximately 8–10 Å wide) and deep (10–15 Å), with the deepest and narrowest portions directly above the active site. This portion of the cleft could accommodate a single strand of DNA but is too narrow to accommodate a DNA duplex without a conformational change. In addition to having a predominantly positively charged cleft, the surface of the Tdp1 protein on the putative DNA binding side of the cleft also has a predominately positive charge (Figure 4) and may be suited for nonspecific binding of DNA.

The putative protein binding side of the cleft is much larger and roughly bowl shaped, flaring from a width of 8 Å directly above the active site to a width of roughly 20 Å. The walls of this cleft extend approximately 15 Å above the deepest part of the cleft. Although this side of the substrate binding cleft is too small to accommodate an intact molecule of topoisomerase I, it could easily accommodate a molecule larger than a single tyrosine residue or a short peptide (see below).

Catalytic Mechanism of Tdp1

In the two-step catalytic mechanism for the enzymes in the PLD superfamily outlined by Stuckey and Dixon [20], an active site histidine residue from one domain acts as a nucleophile that attacks the substrate phosphorous atom and cleaves the phosphodiester linkage, resulting in the formation of a phosphoenzyme intermediate. The active site histidine residue from the other domain functions as a general acid, donating a proton to the leaving group. The formation of a covalent intermediate has been shown for several members of the PLD superfamily, including human Tdp1 [11, 18, 19, 25]. In the cases of Nuc and the bacterial toxin, the intermediate has been shown to involve a linkage to one of the active site histidines. In the second step of the proposed catalytic reaction, the other active site histidine may act as a general base to activate a water molecule that subsequently hydrolyzes the phosphoenzyme intermediate.

In the structure of Tdp1, the Nδ1 atom of the catalytic histidine from the N-terminal domain, His263, is within hydrogen bonding distance of one of the side chain oxygens of the conserved acidic residue, Glu538, from the C-terminal domain. This arrangement is roughly maintained on the other side of the active site, except that the active site histidine residue from the C-terminal domain, His493, is hydrogen bonded to the Oε1 atom of a glutamine residue, Gln294. These interactions would be expected to stabilize the Nδ1-H tautomer of both active site histidines, leaving the substrate proximal Nε2 atoms of both histidines as possible candidates for the nucleophile in the first step of the reaction. However, His263 is in close contact with a negatively charged residue, whereas His493 is in a more neutral environment. This asymmetry suggests that His263, rather than His493, is the nucleophile, since the negatively charged side chain of Glu538 would be expected to stabilize the developing positive charge on the imidazole ring of His263 during the formation of the phosphohistidine intermediate. In addition, the negative glutamate side chain could enhance the nucleophilicity of His263 Nε2 for attack on the phosphorus atom of the substrate. That His263 is the nucleophile in the first step of the reaction is supported by the finding that the H263A mutant has virtually no catalytic activity and does not form a covalent intermediate [11].

Mutation of the conserved lysine residues of PLD superfamily members impairs their catalytic activity [11, 18, 19, 20, 25, 26], and it has been proposed that these residues are involved in binding to the phosphate moiety of substrate and neutralizing its negative charge [20, 21]. It is also likely that the conserved lysine residues may play a key role in transition state stabilization. The conserved asparagine residues interact with well-ordered water molecules in the active site and may also be involved in substrate binding. In conclusion, it seems that Tdp1 essentially follows the previously proposed catalytic mechanism for PLD superfamily members; however, Tdp1 is unique among the PLD superfamily members with known structure, in that it possesses a chemically asymmetrical active site, which suggests that His263 acts as the nucleophile in the first step of the reaction.

Comparison to Other Enzymes in the PLD Superfamily

As mentioned earlier, comparison of active site composition and conservation of sequence motifs have shown Tdp1 to belong to the PLD superfamily [11]. The three-dimensional crystal structures of two PLD family mem-

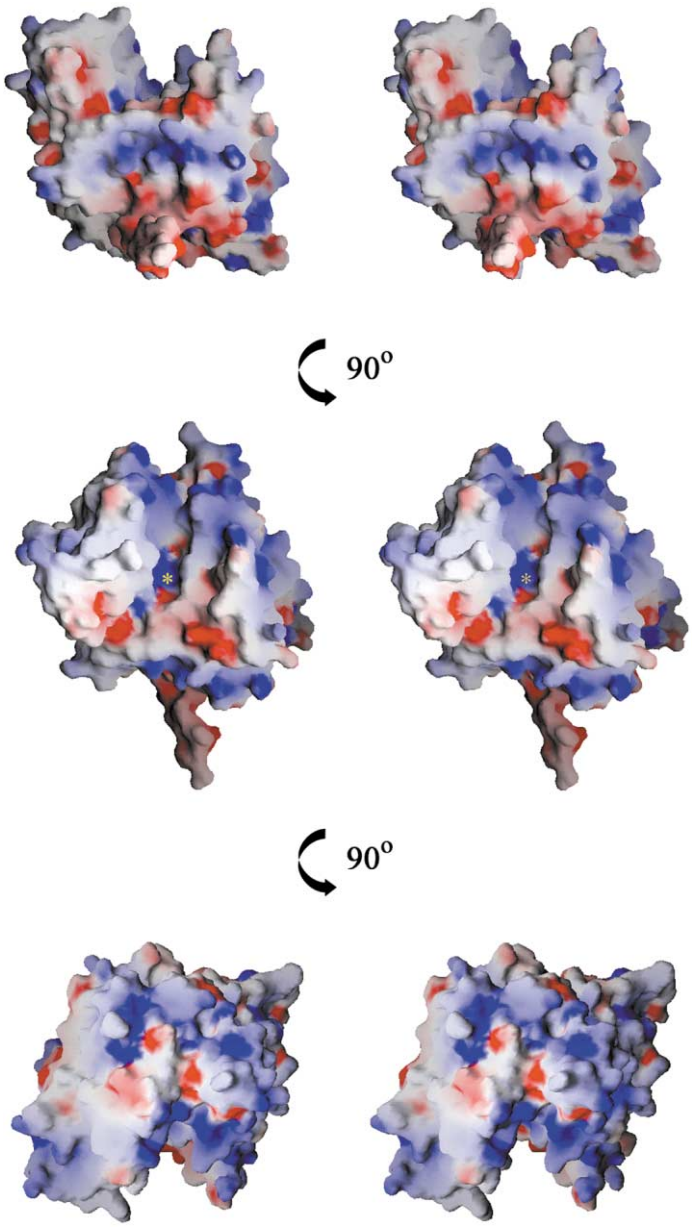

Figure 4. Three Stereoviews of the Electrostatic Potential Surface of Tdp1

Each view was generated with the program GRASP [39]. Electrostatic surfaces are colored between −10 kT (red) and +10 kT (blue). Each view is rotated 90° away from the orientation in the adjoining section of the figure.

(Top) View along the substrate binding cleft from the putative protein binding side. The cleft is located at the top of the molecule in this orientation.

(Middle) View parallel to the pseudo-2-fold axis of symmetry between the two domains of Tdp1 (this orientation is identical to that of the ribbon diagram in Figure 2B). The location of the active site is marked with a yellow asterisk.

(Bottom) View along the substrate binding cleft from the putative DNA binding side. The cleft is located at the bottom of the molecule in this orientation.

bers, a bacterial nuclease from *Salmonella typhimurium* (Nuc) and a PLD from *Streptomyces* sp. PMF have previously been solved [20, 21]. Despite very low sequence identity outside of the conserved active site residues, the overall structures of *Streptomyces* PLD and Nuc have many similarities with Tdp1 (Figures 5 and 6). Like Tdp1, *Streptomyces* PLD is a two-domain structure, with each domain consisting of a β sheet surrounded by α helices, while Nuc is a homodimer, with each subunit corresponding to one domain of Tdp1 or *Streptomyces* PLD. The major β sheet in each domain of Tdp1 consists of seven strands each, whereas the sheets in Nuc consist of eight strands, and *Streptomyces* PLD contains one sheet of eight strands and one sheet of nine strands. In each case, however, the first seven β strands have the same topology, 1765234, with strands 2, 3, and 4 parallel and the rest antiparallel. The topology and approximate positions of five α helices are also conserved between the three structures.

As noted above, human Tdp1 and its orthologs form a distinct class within the PLD superfamily, because they lack the conserved aspartate residue of the HKD sequence motif. The crystal structures of Nuc and *Streptomyces* PLD show that the conserved aspartate residue does not participate in catalysis, since it is located some 19 Å away from center of the active site of the enzyme. In Nuc, this residue lies on a turn between β5 and β6 and is positioned to form four hydrogen bonds with a tight turn between α1 and β2. The corresponding aspartate residues on both domains of *Streptomyces* PLD are involved in similar interactions and are thought to be important for stabilization of the tertiary structure of the molecule [21]. In Tdp1, the amino acid positions corresponding to the conserved aspartate in *Streptomy-*

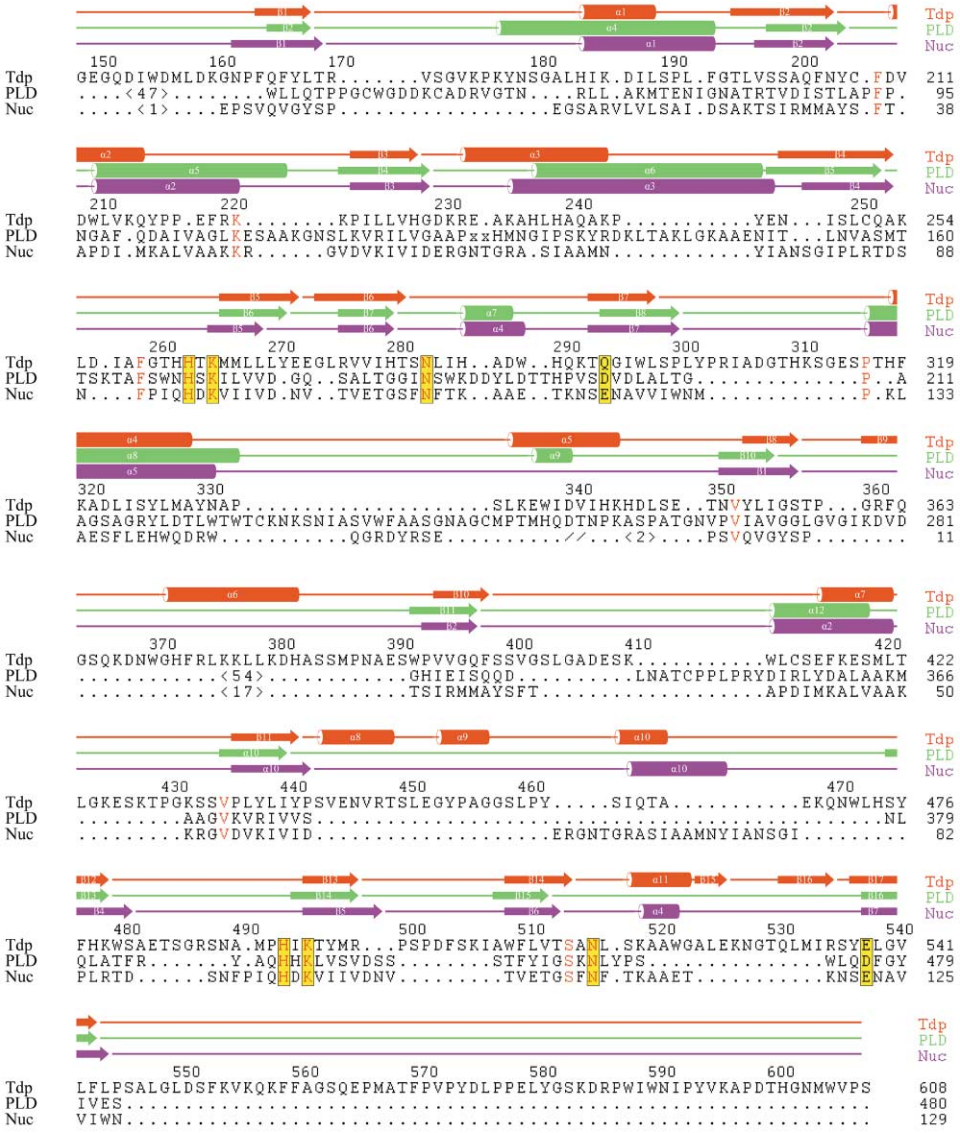

Figure 5. Structure-Based Sequence Alignment and Secondary Structure Comparison for Human Tdp1, Phospholipase D, and Bacterial Nuclease

Alignment is based on regions of three-dimensional overlap between the crystal structures of Tdp1 (Tdp, red), phospholipase (PLD, green; RCSB accession number 1F0I) and bacterial nuclease (Nuc, purple; RCSB accession number 1BYS). Large stretches of PLD and Nuc sequence that do not overlap with Tdp1 have been omitted, and a double slash (//) marks the boundary between the two monomers of Nuc. Areas of sequence identity are marked with red text, and the conserved His, Lys, Asn, and acidic residues of the catalytic active sites of these enzymes are boxed and highlighted in gold.

ces PLD and Nuc are buried in the hydrophobic core of the molecule, because the β strands in the sheets are longer than those in *Streptomyces* PLD and Nuc. In the N-terminal domain of Tdp1, there is a glutamic acid residue that resides on the loop between β5 and β6, but it only has two potential hydrogen bonding interactions with the loop between α1 and β2. Overall, it appears that the tight, aspartate-stabilized turns found in *Streptomyces* PLD and Nuc are not necessary for the structural stability of Tdp1.

Structural overlaps between Tdp1 and *Streptomyces* PLD and Nuc were created with the program LSQKAB [27]. Between Tdp1 and the PLD, 173 Cα atoms overlap, with an rms difference of 2.57 Å. When Tdp1 is superimposed on a dimer of Nuc, 156 Cα atoms overlap, with an overall rms difference of 3.13 Å. The overlapping residues of Tdp1 and *Streptomyces* PLD share only 12.1% sequence identity, whereas the overlapping residues of Tdp1 and Nuc share 17.3% sequence identity.

Although the three enzymes share very low sequence identity overall, the active site residues superimpose remarkably well. Figure 7 shows the three-dimensional alignment of the active site residues of Tdp1, *Streptomyces* PLD, and Nuc. It is striking that the overall side chain conformations are very similar, even though Tdp1 was solved in the absence of substrate or substrate

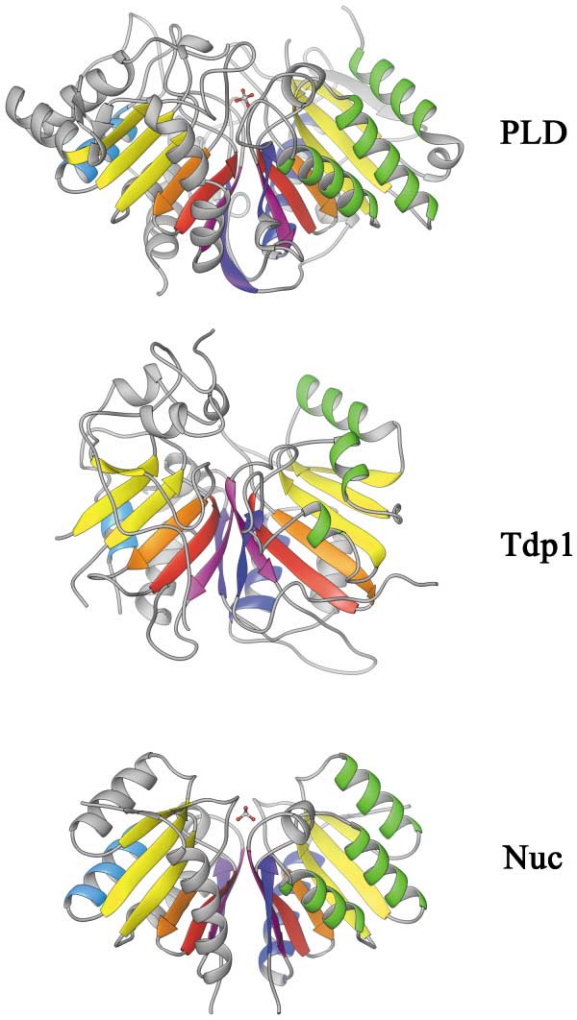

Figure 6. Three-Dimensional Structural Comparison of Tdp1, PLD, and Nuc

Ribbon representations of PLD (top), Tdp1 (middle), and bacterial nuclease (Nuc, bottom), are all shown in the same orientation, viewed perpendicular to the 2-fold (or pseudo-2-fold) axis of symmetry between the two domains of the protein. The orientation of Tdp1 in this figure is the same as that in Figure 2A but rotated 90° counter clockwise in the plane of the paper. Coloration of the models highlights regions of structurally conserved secondary structure elements. For Tdp1, these conserved elements are: β1 and β8 (dark blue); β2, β3, β4, β10, β11, and β12 (yellow); α1, α2, and α3 (green); α4 (blue); β5 and β13 (orange); β6 and β14 (red); β7 and β17 (purple); α7 (cyan). Loops and regions of secondary structure that are not conserved among all three structures are colored gray. Ball and stick representations in the PLD and Nuc structures indicate the positions of phosphate and tungstate ions, respectively, and mark the location of the active sites.

fragments, whereas the *Streptomyces* PLD and Nuc structures contain phosphate and tungstate ions, respectively. Although PLD, Nuc, and Tdp1 act on distinctly different substrates, the substrates for all three enzymes have in common a phosphodiester moiety. Each enzyme is expected to bind phosphate at the center of the active site, in presumably similar orientations. Interestingly, the oxygens of the two water molecules in the Tdp1 active site occupy positions close to those occupied by two oxygen atoms of the tungstate and phosphate ligands in Nuc and *Streptomyces* PLD. The major difference in side chain conformations between the three structures is that the catalytic histidine on the N-terminal domain of *Streptomyces* PLD is rotated approximately 90° about χ_1 relative to the conformations found in Nuc and Tdp1, such that Nδ1, not Nε2, is within hydrogen bonding distance of the phosphate. Interestingly, Leiros et al. [21] reported that in a tungstate-bound structure of *Streptomyces* PLD, the histidine on the N-terminal domain adopts a conformation similar to that found in Nuc, with the Nε2 atom of the residue oriented toward the tungstate. The tungstate ion in this unpublished *Streptomyces* PLD structure was reported to be closer to the histidine on the N-terminal domain than to the histidine on the C-terminal domain. The authors cited this finding as evidence suggesting that the N-terminal histidine residue may be the more probable nucleophile in the catalytic mechanism. Such a finding is consistent with the suggestion, based on our structure above, that the N-terminal domain of Tdp1 contains the nucleophilic histidine. Interestingly, experimental evidence from *Streptomyces antibioticus* has indicated that the active nucleophile resides on the C-terminal domain for that particular enzyme [25].

Structural Clues to the Nature of the Tdp1 Substrate

Tdp1 has been implicated as a specific repair enzyme for topoisomerase I-DNA lesions, but it is not likely to act on the intact protein-DNA complex. The covalent human topoisomerase I-DNA complex consists of a 91 kDa protein linked to the DNA via a 3′ tyrosyl phosphodiester bond buried deep within the complex [22]. It seems apparent, a priori, that the substrate of Tdp1 cannot become accessible to the enzyme without some alteration of the topoisomerase-DNA complex. Yang et al. [1] proposed that denaturation and/or proteolysis of the topoisomerase, through some unknown mechanism, might be a necessary precursor to hydrolysis of the 3′ phosphodiester bond. Recent evidence has indicated that repair of lesions by Tdp1 occurs only after double-strand break formation resulting from collision of a replication fork with a topoisomerase-DNA complex [28]. Thus, both the DNA and the protein moieties of the Tdp1 substrate appear to undergo some form of degradation and processing before Tdp1 can separate them. The exact nature of the Tdp1 substrate is still unknown, but the structure of the Tdp1 protein helps provide some important clues that limit the range of possibilities.

Recent studies have shown that the cleavage efficiency of *S. cerevisiae* Tdp1 is dependent on the structure of the attached DNA [28]. These studies showed that, whereas yeast Tdp1 did not efficiently cleave substrates with 3′ tyrosyl phosphodiester bonds in the middle of a nicked duplex, very little preference was shown between substrates that were single stranded, blunt ended and double stranded, or double stranded with a 5′ tail. These results are somewhat difficult to rationalize in light of the structure of the human Tdp1 substrate binding cleft. As described above, the putative DNA binding cleft in the human Tdp1 structure is only wide

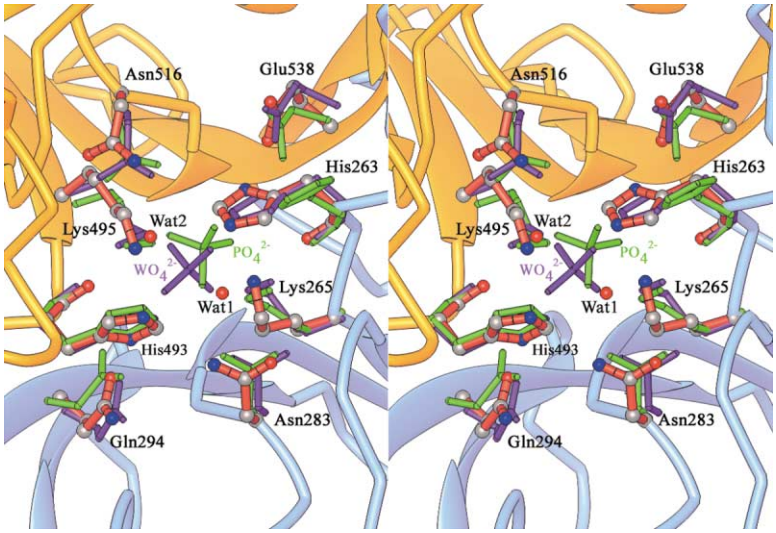

Figure 7. Comparison of the Active Sites of Tdp1, PLD, and Nuc

Tdp1 is depicted as a ribbon structure with the same coloration as in Figure 2. The active site residues of Tdp1 are shown as ball and stick structures with red-colored bonds. The active site residues from PLD are shown as green stick structures, and the active site residues from Nuc are shown as violet stick structures. Similarly, the phosphate group in the active site of PLD is shown in green, and the tungstate ion present in the structure of Nuc is shown in violet. Two water molecules found in the active site of Tdp1 are depicted as red spheres.

enough to accommodate a single strand of DNA. Two possible mechanisms can be envisioned for cleavage of a substrate involving a double-stranded DNA moiety. One possibility is that only single-stranded DNA can bind near the active site of DNA, and that melting of the first few base pairs of the duplex must occur prior to binding of the phosphotyrosine in the active site of Tdp1. If unpairing of the DNA occurs prior to or during binding, the positive electrostatic potential on the face of the putative DNA binding side of Tdp1 may help by binding the DNA of the complementary strand. Another possibility is that a conformational change takes place in Tdp1 that widens the DNA binding cleft to accommodate double-stranded DNA.

The exact nature of the protein moiety of the Tdp1 substrate is another open question. Since full-length topoisomerase I-DNA complexes are apparently not cleaved by human Tdp1 (H. Interthal and J. Champoux, unpublished data), many in vitro studies have employed minimal protein moieties, such as a single tyrosine residue in the case of *S. cerevisiae* Tdp1 [1, 28, 29] or a relatively short peptide in the case of human Tdp1 [11]. However, the broad area of the putative protein binding cleft suggests that a much larger protein moiety could bind to Tdp1. Inspired by the shape of the putative protein binding region of the Tdp1 substrate binding cleft, modeling studies were carried out in which the 6.3 kDa C-terminal domain of topoisomerase I (which contains the catalytic tyrosine residue) was manually docked into the putative protein binding cleft of Tdp1 (Figure 8). For the creation of this model, the C-terminal 46 residues from the crystal structure of human topoisomerase I in covalent complex with DNA were used (Protein Data Bank accession number 1A31) [22]. The phosphate moiety of the phosphotyrosine residue was positioned in the active site of Tdp1 so that the phosphate overlapped with the tungstate binding site of the analogous Nuc structure. The phosphate group was held static while the rest of the C-terminal domain was rotated to find an optimal fit in the binding cleft of Tdp1. The resultant model displays a good complementary fit between the topoisomerase fragment and Tdp1, with no clashes between the two proteins. In support of this model, a substrate consisting of single-stranded DNA covalently bound to the 6.3 kDa C-terminal domain of human topoisomerase I has been shown to be a suitable substrate for human Tdp1 (H. Interthal and J. Champoux, unpublished data). These experiments suggest that the protein moiety of the natural Tdp1 substrate could be as large as the C-terminal domain of human topoisomerase I. These structural considerations will certainly be informative in future attempts to define the Tdp1 substrate in vivo.

Biological Implications

When eukaryotic topoisomerase I becomes stalled on the DNA in the cell, the resulting lesion consists of a phosphodiester bond between a tyrosine residue and a DNA 3′ phosphate. This unique linkage can occur as a result of various forms of DNA damage or from treatment with anticancer drugs, such as camptothecin, and tyrosyl-DNA phosphodiesterase (Tdp1) has been implicated in the repair of such lesions in vivo [1, 28]. Since the activity of Tdp1 can theoretically counteract the effects of camptothecin, drugs that inhibit Tdp1 could potentially act synergistically with topoisomerase poisons in combined drug therapy regimens. Elucidation of the structure of Tdp1 is an important step in any such inhibitor design effort.

The overall tertiary structure of human Tdp1 is strikingly similar to the known structures of two phospholipase D (PLD) superfamily members, a *Streptomyces* PLD [21] and a bacterial nuclease (Nuc) [20]. The protein is an α-β-α-β-α sandwich composed of two α-β-α domains that are related by a pseudo-2-fold axis of symmetry. Each domain contributes conserved histidine, lysine, and asparagine residues to form a single active site. Unlike the active sites of PLD and Nuc, the two active site histidines of Tdp1 exist in distinct chemical environments, with the histidine residue on the N-terminal domain being the probable nucleophile for the cleavage step of the reaction.

The scissile phosphodiester bond in a topoisomerase

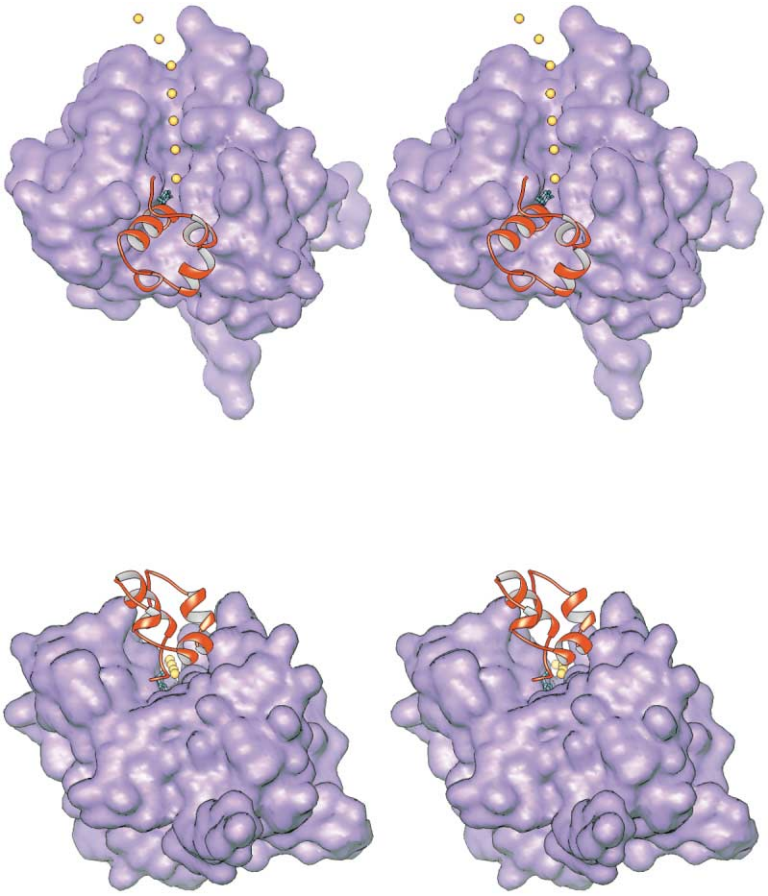

Figure 8. Two Stereoviews of a Hypothetical Model of the Human Topoisomerase I C-terminal Domain Docked into the Putative Peptide Binding Cleft of Tdp1

The molecular surface of Tdp1 is depicted in purple. The red ribbon structure represents the C-terminal domain of human topoisomerase I from residue 720 to 765, and the phosphotyrosine residue (Tyr723) is depicted as a cyan ball and stick structure. The yellow spheres on the other side of the substrate binding cleft represent the possible location of a single strand of DNA that would be covalently attached to the phosphotyrosine. The top stereo pair shows a view of the model roughly parallel to the pseudo-2-fold axis of symmetry of Tdp1, and the orientation is essentially the same as that in Figure 2b and the middle of Figure 4. The bottom stereo pair shows the model from the putative protein binding side of the cleft, and the orientation is the same as that in the top of Figure 4.

I-DNA linkage is buried deep within the complex. Several lines of evidence suggest that both the protein and DNA moieties of the complex undergo some sort of processing and degradation before Tdp1 can separate them [1, 28]. The structure of Tdp1 provides several clues to the nature of the substrate in vivo. Tdp1 has an asymmetrical, elongated substrate binding cleft that appears well suited to binding a protein moiety on one side of the cleft and the DNA on the other. The structure suggests that a protein moiety at least as large as the 6.3 kDa C-terminal domain of human topoisomerase I may be accommodated by the protein binding side of the cleft but that only a single strand of DNA could be accommodated on the DNA binding side without a conformational rearrangement.

Experimental Procedures

Crystallization, Selenomethionine Derivatives, Freezing Conditions, and Data Collection

Human Tdp1 Δ1−148 was expressed and purified as previously described [11]. Purified protein was dialyzed into a buffer containing 250 mM NaCl, 15 mM Tris (pH 8.2), 1 mM EDTA, and 2 mM Tris (2-carboxyethyl) phosphine (TCEP) and concentrated to 5–6 mg/mL. Tdp1 was crystallized by vapor diffusion in sitting drops by combining equal volumes of protein solution and a crystallant containing 11%–15% PEG 8000, 100 mM CHES (pH 9.6), and 2–15 mM spermine. Two crystal forms were produced under these conditions, a monoclinic form ($P2_1$; a = 49.0 Å, b = 190.3 Å, c = 54.5 Å, and β = 100.3°) and an orthorhombic form ($P2_12_12$; a = 186.5 Å, b = 54.2 Å, and c = 50.0 Å), although the monoclinic form appeared primarily under higher concentrations of PEG and spermine. Dozens of crystals appeared in each drop in less than 24 hr and usually reached full size in less than 4 days. Typical orthorhombic crystals reached sizes of 0.2 × 0.2 × 0.3 mm, and monoclinic crystals were usually 0.01 × 0.15 × 0.4 mm in size. Both crystal forms diffracted to 2.8 Å or better in house and between 1.5 Å and 2.0 Å with synchrotron radiation. To prepare monoclinic crystals for flash cooling, they were transferred to a cryoprotectant containing 25% glycerol, 15% PEG 8000, 300 mM NaCl, 10 mM spermine, 100 mM CHES (pH 9.6), 1 mM EDTA, and 2 mM TCEP. The same cryoprotectant was used for the orthorhombic crystal form; however, the orthorhombic crystals were much more sensitive to mechanical shock and cracked if transferred from crystallization drops to cryoprotectant solutions. Hence, orthorhombic crystals were instead prepared for flash cooling by adding relatively large volumes of cryoprotectant directly to crystallization drops. Next, orthorhombic crystals were immediately removed from drops and frozen by plunging directly into liquid nitrogen.

Selenomethionine derivatives were prepared by expressing Tdp1 Δ1−148 in a nonauxotrophic *E. coli* strain in M9 minimal media supplemented with selenomethionine and other amino acids (100 mg each of Lys, Thr, and Phe and 50 mg each of Val, Leu, Ile, and SeMet per liter of culture) in order to suppress de novo methionine biosynthesis as described by Van Duyne, et al. [30]. Purification and crystallization of selenomethionine-labeled protein was carried out exactly as it was for unlabeled protein, except that 5 mM TCEP was present in all buffers to prevent oxidation of selenomethionine residues. Electrospray ionization mass spectroscopy showed that the level of selenomethionine substitution was greater than 99%. The SeMet data set was collected at the Advanced Light Source beamline 5.0.2 with an ADSC Quantum 4 CCD detector. The data were integrated, scanned, and reduced with the HKL software [31].

Phase Determination, Model Building, and Structural Refinement

The structure of Tdp1 was determined using SOLVE version 1.18 [32] and three wavelengths of a MAD data set collected to 1.69 Å resolution. SOLVE found seven heavy-atom sites out of a possible 11 methionine residues present in the protein. Inspection of the final model shows that 3 of the 11 methionine residues present in the protein lie on disordered loops or in the disordered portion of the His-tag and linker. Density modification of the original solution was carried out by RESOLVE version 1.05 [33]. These initial phases were input into wARP/ARP [34] for automated model building and phase improvement. wARP/ARP was able to trace the backbone and place side chains for approximately 85% of the model of the protein. Additional model building was carried out using XtalView [35]. Further refinement was carried out using Refmac 5 [27, 36] and alternating cycles of manual rebuilding in XtalView.

Acknowledgments

We would like to thank Dr. Howard Nash for kindly providing the human Tdp1 cDNA. We also gratefully acknowledge the help of Dr. Keith Henderson at beamline 5.0.2 of the Advanced Light Source for assistance with synchrotron data collection. We also acknowledge the staff of Structural Biology Center at the Advanced Photon Source for assistance with early characterization of the crystals. D.R.D. acknowledges Stewart Turley and Drs. Abhinav Kumar and Francis Athappilly for assistance with data collection and computing. W.G.J.H. acknowledges the Murdock Charitable Trust for a major equipment grant to the Biomolecular Structure Center at the University of Washington. This work was supported by grants from the National Institutes of Health to J.J.C. (GM 49156) and to W.G.J.H. (CA 65656).

Received: September 20, 2001
Revised: December 11, 2001
Accepted: December 13, 2001

References

1. Yang, S.W., Burgin, A.B., Jr., Huizenga, B.N., Robertson, C.A., Yao, K.C., and Nash, H.A. (1996). A eukaryotic enzyme that can disjoin dead-end covalent complexes between DNA and type I topoisomerases. Proc. Natl. Acad. Sci. USA *93*, 11534–11539.
2. Strumberg, D., Pilon, A.A., Smith, M., Hickey, R., Malkas, L., and Pommier, Y. (2000). Conversion of topoisomerase I cleavage complexes on the leading strand of ribosomal DNA into 5′-phosphorylated DNA double-strand breaks by replication runoff. Mol. Cell. Biol. *20*, 3977–3987.
3. Pourquier, P., and Pommier, Y. (2001). Topoisomerase I-mediated DNA damage. Adv. Cancer Res. *80*, 189–216.
4. Pommier, Y. (1998). Diversity of DNA topoisomerases I and inhibitors. Biochimie *80*, 255–270.
5. Holm, C., Covey, J.M., Kerrigan, D., and Pommier, Y. (1989). Differential requirement of DNA replication for the cytotoxicity of DNA topoisomerase I and II inhibitors in Chinese hamster DC3F cells. Cancer Res. *49*, 6365–6368.
6. Hsiang, Y.H., Lihou, M.G., and Liu, L.F. (1989). Arrest of replication forks by drug-stabilized topoisomerase I-DNA cleavable complexes as a mechanism of cell killing by camptothecin. Cancer Res. *49*, 5077–5082.
7. D'Arpa, P., Beardmore, C., and Liu, L.F. (1990). Involvement of nucleic acid synthesis in cell killing mechanisms of topoisomerase poisons. Cancer Res. *50*, 6919–6924.
8. Ryan, A.J., Squires, S., Strutt, H.L., Evans, A., and Johnson, R.T. (1994). Different fates of camptothecin-induced replication fork-associated double-strand DNA breaks in mammalian cells. Carcinogenesis *15*, 823–828.
9. Tsao, Y.P., Russo, A., Nyamuswa, G., Silber, R., and Liu, L.F. (1993). Interaction between replication forks and topoisomerase I-DNA cleavable conplexes: studies in a cell-free SV40 DNA replication system. Cancer Res. *53*, 5908–5914.
10. Nitiss, J.L., and Wang, J.C. (1996). Mechanisms of cell killing by drugs that trap covalent complexes between DNA topoisomerases and DNA. Mol. Pharmacol. *50*, 1095–1102.
11. Interthal, H., Pouliot, J.J., and Champoux, J.J. (2001). The tyrosyl-DNA phosphodiesterase Tdp1 is a member of the phospholipase D superfamily. Proc. Natl. Acad. Sci. USA *98*, 12009–12014.
12. Koonin, E.V. (1996). A duplicated catalytic motif in a new superfamily of phosphohydrolases and phospholipid synthases that includes poxvirus envelope proteins. Trends Biochem. Sci. *21*, 242–243.
13. Morris, A.J., Engebrecht, J., and Frohman, M.A. (1996). Structure and regulation of phospholipase D. Trends Pharmacol. Sci. *17*, 182–185.
14. Ponting, C.P., and Kerr, I.D. (1996). A novel family of phospholipase D homologues that includes phospholipid synthases and putative endonucleases: identification of duplicated repeats and potential active site residues. Protein Sci. *5*, 914–922.
15. Stanacev, N.Z., and Stuhne-Sekalec, L. (1970). On the mechanism of enzymatic phosphatidylation. Biosynthesis of cardiolipin catalyzed by phospholipase D. Biochim. Biophys. Acta *210*, 350–352.
16. Bruzik, K., and Tsai, M.D. (1984). Phospholipids chiral at phosphorus. Synthesis of chiral phosphatidylcholine and stereochemistry of phospholipase D. Biochemistry *23*, 1656–1661.
17. Raetz, C.R., Carman, G.M., Dowhan, W., Jiang, R.T., Waszkuc, W., Loffredo, W., and Tsai, M.D. (1987). Phospholipids chiral at phosphorus. Steric course of the reactions catalyzed by phosphatidylserine synthase from Escherichia coli and yeast. Biochemistry *26*, 4022–4027.
18. Gottlin, E.B., Rudolph, A.E., Zhao, Y., Matthews, H.R., and Dixon, J.E. (1998). Catalytic mechanism of the phospholipase D superfamily proceeds via a covalent phosphohistidine intermediate. Proc. Natl. Acad. Sci. USA *95*, 9202–9207.
19. Rudolph, A.E., Stuckey, J.A., Zhao, Y., Matthews, H.R., Patton, W.A., Moss, J., and Dixon, J.E. (1999). Expression, characterization, and mutagenesis of the *Yersinia pestis* murine toxin, a phospholipase D superfamily member. J. Biol. Chem. *274*, 11824–11831.
20. Stuckey, J.A., and Dixon, J.E. (1999). Crystal structure of a phospholipase D family member. Nat. Struct. Biol. *6*, 278–284.
21. Leiros, I., Secundo, F., Zambonelli, C., Servi, S., and Hough, E. (2000). The first crystal structure of a phospholipase D. Struct. Fold. Des. *8*, 655–667.
22. Redinbo, M.R., Stewart, L., Kuhn, P., Champoux, J.J., and Hol, W.G.J. (1998). Crystal structures of human topoisomerase I in covalent and noncovalent complexes with DNA. Science *279*, 1504–1513.
23. Laskowski, R.A., MacArthur, M.W., Moss, D.S., and Thornton, J.M. (1993). PROCHECK: a program to check the stereochemical quality of protein structures. J. Appl. Crystallogr. *26*, 283–291.
24. Kabsch, W., and Sander, C. (1983). Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. Biopolymers *22*, 2577–2637.
25. Iwasaki, Y., Horiike, S., Matsushima, K., and Yamane, T. (1999). Location of the catalytic nucleophile of phospholipase D of *Streptomyces antibioticus* in the C-terminal half domain. Eur. J. Biochem. *264*, 577–581.
26. Sung, T.C., Roper, R.L., Zhang, Y., Rudge, S.A., Temel, R., Hammond, S.M., Morris, A.J., Moss, B., Engebrecht, J., and Frohman, M.A. (1997). Mutagenesis of phospholipase D defines a superfamily including a trans-Golgi viral protein required for poxvirus pathogenicity. EMBO J. *16*, 4519–4530.
27. (CCP4) Collaborative Computational Project 4. (1994). The CCP4 suite: programs for protein crystallography. Acta Cryst. D Biol. Crystallogr. *50*, 760–763.
28. Pouliot, J.J., Robertson, C.A., and Nash, H.A. (2001). Pathways for repair of topoisomerase I covalent complexes in *Saccharomyces cerevisiae.* Genes Cells *6*, 677–687.
29. Pouliot, J.J., Yao, K.C., Robertson, C.A., and Nash, H.A. (1999). Yeast gene for a Tyr-DNA phosphodiesterase that repairs topoisomerase I complexes. Science *286*, 552–555.
30. Van Duyne, G.D., Standaert, R.F., Karplus, P.A., Schreiber, S.L., and Clardy, J. (1993). Atomic structures of the human immu-

nophilin FKBP-12 complexes with FK506 and rapamycin. J. Mol. Biol. *229*, 105–124.
31. Otwinowski, Z., and Minor, W. (1997). Processing of X-ray diffraction data collected in oscillation mode. Methods Enzymol. *276*, 307–326.
32. Terwilliger, T.C., and Berendzen, J. (1999). Automated MAD and MIR structure solution. Acta Crystallogr. *D55*, 849–861.
33. Terwilliger, T.C. (2000). Maximum-likelihood density modification. Acta Crystallogr. *D56*, 965–972.
34. Perrakis, A., Morris, R., and Lamzin, V.S. (1999). Automated protein model building combined with iterative structure refinement. Nat. Struct. Biol. *6*, 458–463.
35. McRee, D.E. (1999). XtalView/Xfit—a versatile program for manipulating atomic coordinates and electron density. J. Struct. Biol. *125*, 156–165.
36. Murshudov, G.N., Vagin, A.A., Lebedev, A., Wilson, K.S., and Dodson, E.J. (1999). Efficient anisotropic refinement of macromolecular structures using FFT. Acta Crystallogr. *D55*, 247–255.
37. Carson, M. (1997). Ribbons. Methods Enzymol. *277*, 493–505.
38. Wallace, A.C., Laskowski, R.A., and Thornton, J.M. (1995). LIGPLOT: a program to generate schematic diagrams of protein-ligand interactions. Protein Eng. *8*, 127–134.
39. Nicholls, A., Sharp, K.A., and Honig, B. (1991). Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons. Proteins *11*, 281–296.

Accession Numbers

Atomic coordinates and structure factors have been deposited in the Protein Data Bank (with the accession code 1jy1) at Rutgers University.