The tyrosyl-DNA phosphodiesterase Tdp1 is a member of the phospholipase D superfamily

Heidrun Interthal*, Jeffrey J. Pouliot†, and James J. Champoux*‡

*Department of Microbiology, Box 357242, School of Medicine, University of Washington, Seattle, WA 98195-7242; and †Laboratory of Molecular Biology, National Institute of Mental Health, Building 36, Room 1B08, Bethesda, MD 20892-4034

Communicated by Howard A. Nash, National Institutes of Health, Bethesda, MD, August 15, 2001 (received for review May 14, 2001)

The phospholipase D (PLD) superfamily is a diverse group of proteins that includes enzymes involved in phospholipid metabolism, a bacterial toxin, poxvirus envelope proteins, and bacterial nucleases. Based on sequence comparisons, we show here that the tyrosyl-DNA phosphodiesterase (Tdp1) that has been implicated in the repair of topoisomerase I covalent complexes with DNA contains two unusual HKD signature motifs that place the enzyme in a distinct class within the PLD superfamily. Mutagenesis studies with the human enzyme in which the invariant histidines and lysines of the HKD motifs are changed confirm that these highly conserved residues are essential for Tdp1 activity. Furthermore, we show that, like other members of the family for which it has been examined, the reaction involves the formation of an intermediate in which the cleaved substrate is covalently linked to the enzyme. These results reveal that the hydrolytic reaction catalyzed by Tdp1 occurs by the phosphoryl transfer chemistry that is common to all members of the PLD superfamily.

The members of the phospholipase D (PLD) superfamily comprise a highly diverse group of proteins that include plant, mammalian and bacterial PLDs, bacterial phosphatidylserine and cardiolipin synthases, a bacterial toxin, several poxvirus envelope proteins, and some bacterial nucleases (1–3). Sequence alignments reveal that with the exception of two nucleases (1) the proteins arose as a result of a gene duplication event with each half of the protein containing four repeated motifs. Motifs 3 and 4 contain the highly conserved $HxK(x)_4D(x)_6GSxN$ sequence, termed the HKD motif (4), which has been implicated in the catalytic mechanism (see below). The crystal structure of the *Salmonella typhimurium* Nuc protein (4), one of the bacterial nucleases, shows that the active form of the enzyme is a dimer with the HKD motifs contributed by each subunit organized in roughly the same spatial arrangement as the two HKD motifs found in the crystal structure of the monomeric PLD from *Streptomyces* sp. strain PMF (5).

For those members of the superfamily known to have catalytic activity, the enzymatic reactions all involve phosphoryl transfer from a donor to an acceptor that is either an alcohol or water. In the cases of the phospholipid synthases, a phosphatidyl moiety is transferred either from a cytidine 5′-diphosphate–diacylglycerol donor to serine or from phosphatidylglycerol to another phosphatidylglycerol to synthesize phosphatidylserine or cardiolipin, respectively. The PLDs either hydrolyze the phosphodiester bond in the phospholipid to produce phosphatidic acid and a free head group (often choline) or catalyze the exchange of one head group for another (transphosphatidylation). The nucleases appear to simply catalyze the hydrolysis of DNA phosphodiester bonds.

The similar chemistry underlying these reactions suggests that the enzymes in the superfamily share a similar catalytic mechanism. Early evidence showing retention of configuration at the substrate phosphorous in the reactions catalyzed by cabbage PLD and *Escherichia coli* phosphatidylserine synthase suggested a "ping pong"-type mechanism with the formation of a covalent phosphoenzyme intermediate (6–8). The direct demonstration of such a covalent intermediate for *S. typhimurium* Nuc and *Yersinia pestis* murine toxin substantiates this view and reveals that the active site nucleophile in the reaction is one of the highly conserved histidines of the HKD motif (9, 10). In support of the importance of the histidines and lysines of both HKD motifs, the known crystal structures reveal that all four residues are clustered together in the active center of the enzymes (4, 5).

When a type IB topoisomerase cleaves DNA, the enzyme becomes covalently attached to the DNA through a phosphodiester bond between a tyrosine residue in the enzyme and the 3′ end of the cleaved strand (11). Subsequent religation of the DNA restores the DNA phosphodiester bond and releases the enzyme from the DNA. Yang *et al.* (12) described a phosphodiesterase activity in a variety of eukaryotic cell types that cleaves the bond between tyrosine and a DNA 3′ phosphate. They speculated that the activity might be important *in vivo* for the repair of topoisomerase I-DNA covalent complexes that occur in the cell when the topoisomerase fails to religate the DNA or that accumulate after blocking religation with the anticancer drug camptothecin (12–15). Subsequently the *Saccharomyces cerevisiae* gene for the tyrosyl-DNA phosphodiesterase (Tdp1) was cloned and characterized, and orthologs were identified in other eukaryotic but not prokaryotic species (16). In certain repair-deficient backgrounds, yeast null mutants for the phosphodiesterase were found to be hypersensitive to conditions that result in the accumulation of topoisomerase I-DNA complexes (16, 17). These results in combination with substrate specificity studies suggest that Tdp1 functions as part of a double-strand break repair pathway to resolve the structures formed when a replication fork collides with a topoisomerase I stalled on DNA (17).

Here we show by sequence alignments that Tdp1 defines a distinct class of proteins within the PLD superfamily. This observation in combination with a mutagenesis analysis involving the critical residues in the HKD motif elucidate the reaction mechanism of human Tdp1.

Materials and Methods

Cloning and Site-Directed Mutagenesis. The 5′ terminus of the 1,824-bp human *TDP1* coding sequence was identified by rapid amplification of cDNA ends of a CLONTECH Marathon human cDNA library. The full-length coding sequence was PCR-amplified directly from the library and cloned into the *Bam*HI sites of vector pET15b (Novagen), which adds a His tag and a thrombin cleavage site (MGSSHHHHHHSSGLVPRGSHMLEDP) to the N terminus of Tdp1. The *TDP1* coding sequence in the resulting plasmid pHN1894S differs in the following positions from the published sequence of the predicted human gene FLJ11090 (National Center for Biotechnology Information accession no. NM_018319). Nucleotides 282 T to C and 318 G to T do not lead to amino acid changes. Nucleotides 964 G to A (D322N), 983 T to C (M328T), 1165 C to G (P389A), and 1644

Abbreviations: PLD, phospholipase D; Tdp1, tyrosyl-DNA phosphodiesterase.

‡To whom reprint requests should be addressed. E-mail: champoux@u.washington.edu.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

BIOCHEMISTRY

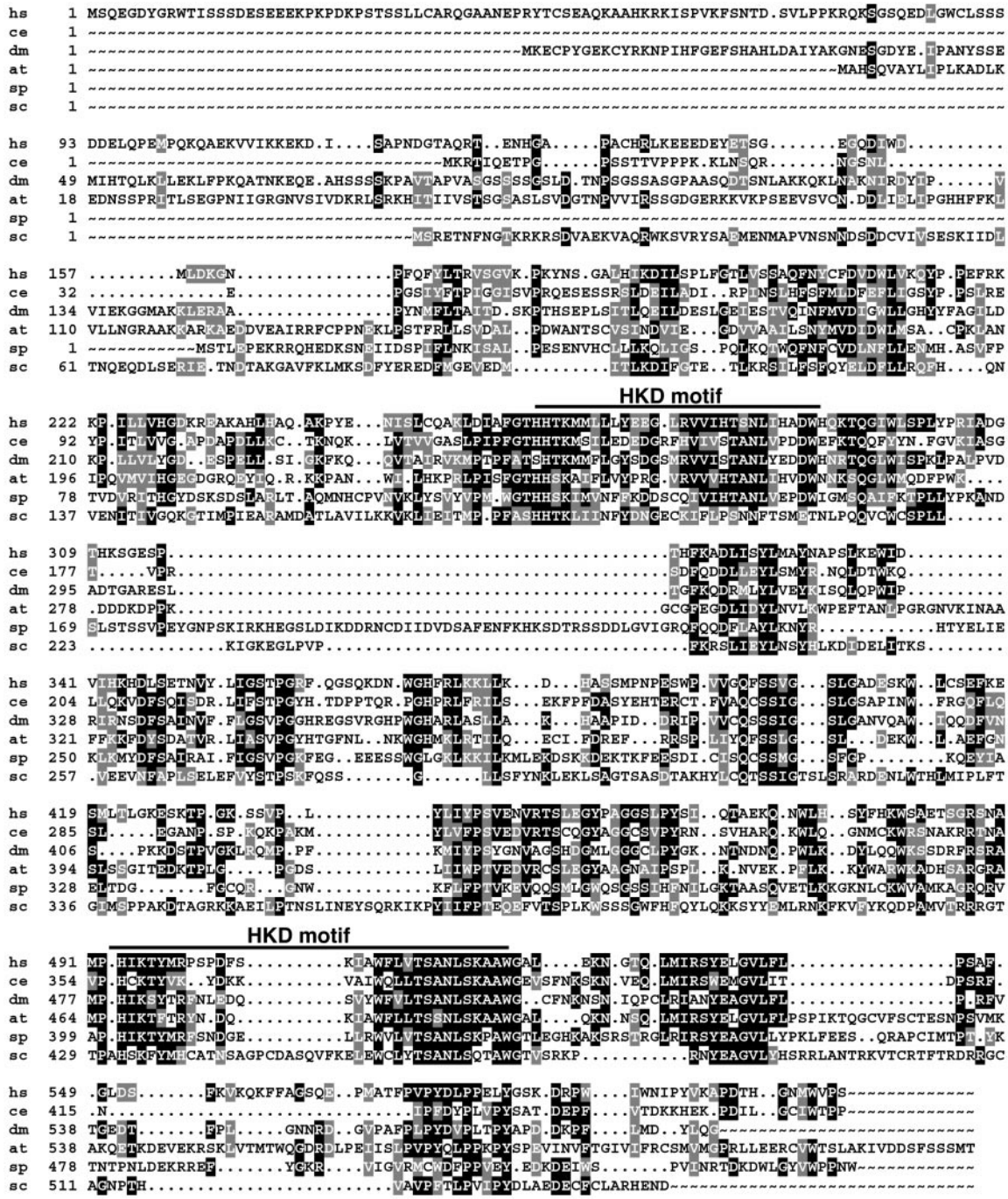


Fig. 1. Sequence alignment of Tdp1 orthologs. The program PILEUP of the GCG package was used with some manual modification to create the alignment. BOXSHADE was used to shade residues black that are identical in ≥ 50% of the sequences and similar residues were shaded in gray. The HKD motifs are indicated by black bars above the alignment. Abbreviations and accession numbers are as follows: hs (*Homo sapiens* Tdp1 predicted protein FLJ11090, National Center for Biotechnology Information RefSeq accession no. NP_060789), ce (*C. elegans*, GenPept: AAC68960), dm (*D. melanogaster* glaikit protein, GenPept: CAB86488), at (*A. thaliana*, GenPept: CAB89327), sp (*S. pombe*, hypothetical protein SPCP31B10.05, PIR: T41695), sc (*S. cerevisiae* Tdp1, gene product of ORF YBR223c, GenPept: CAA85186). The *C. elegans* and *A. thaliana* protein sequences were predicted based on genomic sequences.

T to A (F548L) lead to the indicated changes in the amino acid sequence. Plasmid pHN1894S(ΔTDP1) was generated by removing the *Bam*HI fragment of pHN1894S that contains the entire *TDP1* coding sequence. Mutant alleles of human *TDP1* in plasmid pHN1894S, including the deletion mutant Δ1–148, were generated by using the QuickChange site-directed mutagenesis kit from Stratagene. Detailed information concerning the cloning of *TDP1* and the sequences of the oligonucleotides used for mutagenesis are available as supporting information, which is published on the PNAS web site, www.pnas.org.

Expression and Purification of Wild-Type and Mutant Human Tdp1 Proteins. N-terminally His-tagged wild-type Tdp1 and the mutant proteins Δ1–148, H263A, H493A, H493N, K495S, and K265S were expressed in *E. coli* BL21 (DE3) cells by using the pET expression system (Novagen). In addition, a mock purification

using plasmid pHN1894S(ΔTDP1) was carried out for use as a negative control. Plasmids were freshly transformed into BL21 cells, and the cells were grown in LB medium at pH 7.5 with 1% glucose and 50 μg/ml ampicillin at 30°C. Two hours after induction with 1 mM isopropyl β-D-thiogalactoside, cells were harvested, and the dry pellets were stored at −80°C.

For protein purification HiTrap Chelating columns (Amersham Pharmacia) and the His-bind buffer kit from Novagen were used. Cell pellets were thawed on ice, resuspended in binding buffer (0.5 M NaCl/5 mM imidazole/20 mM Tris·HCl, pH 7.9, with protease inhibitors), and broken by sonication. After centrifugation, the supernatant was squeezed through a 22-gauge needle, filtered through a 0.45-μm filter, and passed through a $NiSO_4$-charged chelating column. The column was first washed with binding buffer and then with wash buffer with protease inhibitors (0.5 M NaCl, 60 mM imidazole, 20 mM Tris·HCl, pH 7.9). Elution was carried out with elution buffer (0.5 M NaCl/1 M imidazole/20 mM Tris·HCl, pH 7.9/protease inhibitors), and the proteins were stored in 50 mM KCl, 50 mM Tris·HCl (pH 7.5), 1 mM EDTA, 2 mM DTT, and 50% glycerol at −20°C. The N-terminal truncation mutant Δ1–148 was chromatographed on a Whatman Cellulose Phosphate P11 column before binding to the chelating column.

All enzyme preparations, except the H263A and K265S preparations, were estimated to be more than 90% pure. Although H263A and K265S are only ≈50% pure, owing to a lower recovery of protein, mixing controls demonstrated that the reduced activity of H263A was not the result of contaminating inhibitors. Enzyme concentrations were estimated by using a Bradford assay and by the intensity of Coomassie-stained bands after SDS/PAGE.

Substrate Preparation. The 14-mer oligonucleotide CL14N (5′-GAAAAAAGAGTT↓AG-3′) was ^{32}P end-labeled with T4 polynucleotide kinase in the presence of [γ-^{32}P]ATP and annealed to the complementary 25-mer CP25N (5′-TAAAAATTTTTCTAACTCTTTTTTC-3′), which was cold-phosphorylated at its 5′ end. Reaction mixtures containing 0.125 μM oligonucleotide suicide substrate CL14N/CP25N and 0.75 μM topo70 [a 70-kDa form of human topoisomerase I that is missing the nonessential N terminus (18)] in 150 mM KCl, 10 mM Tris·HCl (pH 7.5), 1 mM EDTA, and 1 mM DTT were incubated at room temperature (23°C) for 60 min to allow suicide cleavage and covalent binding of topo70 to the 3′ end of the resulting 12-mer oligonucleotide (12-topo, see Fig. 3*A*). The arrow in the CL14N sequence above indicates the preferred cleavage site of human topoisomerase I. Subsequently, topo70 was digested with 1 μg/μl of trypsin at 37°C for 2 h to generate the Tdp1 substrate 12-pep, which consists of a 12-mer oligonucleotide with a small trypsin-resistant peptide covalently bound to it via the active site tyrosine. Trypsin was inactivated by the addition of PMSF to 0.7 μg/μl.

Tdp1 Activity Assay. For the Tdp1 activity assay, the proteins were serially diluted 5-fold in reaction buffer starting at a concentration of ≈14 μM. Reactions (15 μl) contained 150 mM KCl, 10 mM Tris·HCl (pH 7.5), 1 mM EDTA, 1 mM DTT, 6.7 nM 12-pep substrate, and 1 μl of the diluted enzyme. After incubation for 15 min at room temperature, reactions were stopped by the addition of 15 μl of formamide loading dye (96% formamide, 20 mM EDTA, 0.03% xylene cyanol, 0.03% bromophenol blue). Six microliters of each reaction was analyzed on a 15% sequencing gel. Image retrieval and quantitation was carried out by using a PhosphorImager in conjunction with IMAGEQUANT software (Molecular Dynamics).

Covalent Intermediate Trapping Experiment. The time-course experiments were started by the addition of the indicated Tdp1 proteins (final concentration ≈100 nM) or mock purified protein to reaction mixtures containing 6.7 nM 12-pep substrate, 50 mM Tris·HCl (pH 7.5), 1 mM DTT, and 1 mM EDTA. All reactions were carried out on ice to slow the reaction. At the indicated time points, 45 μl of the reaction mixes was stopped by addition to 12 μl of 5× SDS loading buffer (300 mM Tris·HCl, pH 6.8/10% SDS/50% glycerol/25% β-mercaptoethanol/0.005% bromophenol blue]. Fifteen-microliter aliquots were boiled for 3 min and analyzed by 10% SDS/PAGE to visualize the labeled proteins. Another 15-μl aliquot of these same samples was mixed with 2 μl of 0.1 M $MgCl_2$ and 85 μl of cold ethanol to precipitate the DNA and any DNA-protein complexes. The pellets were resuspended in 20 μl of 80% formamide loading dye. Samples were boiled for 3 min, and 5 μl of each was analyzed on a 15% sequencing gel. SDS/PAGE and sequencing gels were dried and analyzed as described above.

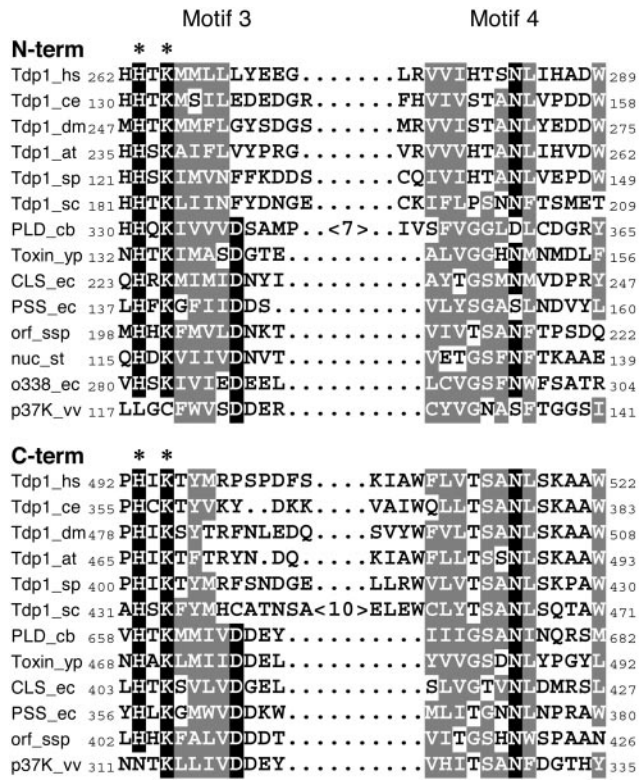


Fig. 2. Alignment of the HKD motifs found in Tdp1 orthologs with the HKD motifs of a representative member from each PLD superfamily class. The program CLUSTALW with additional manual modifications was used to create the alignment. Proteins that possess only one HKD motif are grouped with the N-terminal motifs of the proteins containing two HKD motifs. Highlighted in gray are key residues that fall into one of the three categories established by Ponting and Kerr (1) for the members of the PLD superfamily: hydrophobic residues (A, C, I, L, V, M, F, Y, W), small residues (G, A, S), and acidic and amides of acidic residues (D, E, N, Q). Black boxes indicate the most conserved amino acids in the motif. * marks positions that were mutated in human Tdp1. Numbers flanking the sequences indicate the positions of the first and last amino acids shown in that protein sequence. Numbers in brackets represent the numbers of amino acids not shown in the alignment. Accession codes and sequences represented are the same as in Fig. 1 for the Tdp1 orthologs. Representatives of the other PLD superfamily classes are [roman numerals indicate the class according to Ponting and Kerr (1)]: PLD_cb (I, castor bean PLD, PIR: A54850), Toxin_yp (II, *Y. pestis* murine toxin, GenPept: CAA63386), CLS_ec (III, *E. coli* cardiolipin synthase, SWISS-PROT: P31071), PSS_ec (IV, *E. coli* phosphatidylserine synthase, SWISS-PROT: P23830), orf_ssp (VI, *Synchocystis sp.* product of comE ORF1, GenPept: BAA10416), nuc_st (VII, *S. typhimurium* Nuc protein precursor, PIR: S41475), o338_ec (VIII, *E. coli* hypothetical 38-kDa protein in FEC1-FIMB intergenic region, SWISS-PROT: P39369), p37K_vv (V, *Vaccinia* virus p37K major envelope protein, SWISS-PROT: P20638).

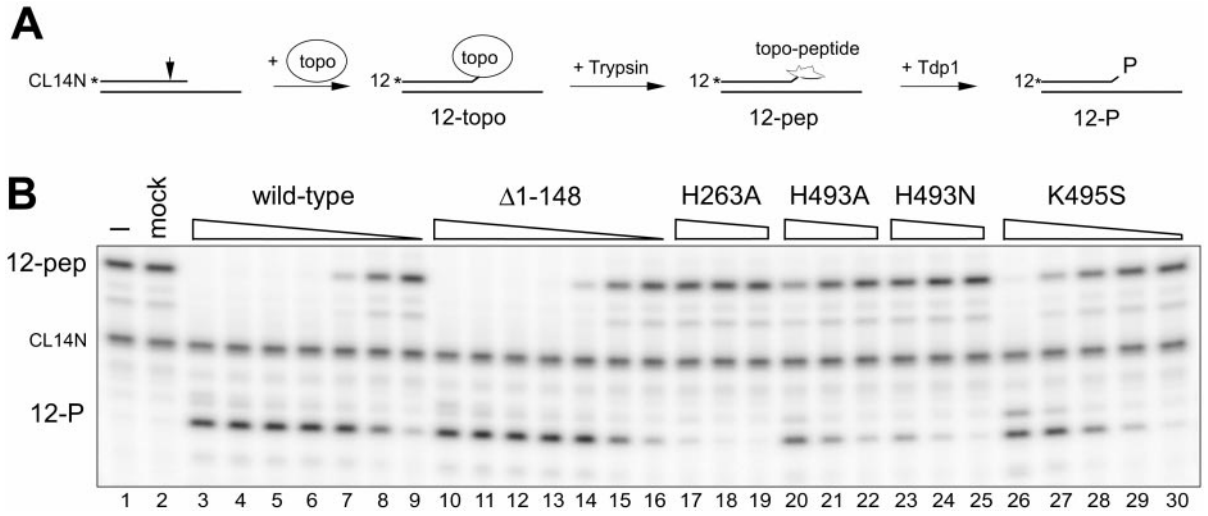


Fig. 3. Tdp1 activity assay. (*A*) Schematic representation of substrate preparation and Tdp1 activity assay. The 5′ end-labeled topoisomerase I oligonucleotide suicide substrate CL14N/CP25N is cleaved by topoisomerase I (site indicated with arrow), and the 12-mer cleavage product remains covalently bound to the topoisomerase (12-topo). Subsequent digestion with trypsin leaves a small peptide covalently bound to the DNA via the active site tyrosine. This is the substrate (12-pep) for Tdp1, which cleaves specifically between the peptide and the DNA to produce a 12-mer oligonucleotide with a 3′ phosphate group (12-P). * indicates the ^{32}P 5′ end-labeled scissile strand (CL14N). (*B*) Sequencing gel analysis of reaction products. Serial 5-fold dilutions of the various proteins (as indicated by the triangles above the lanes) were incubated for 15 min with identical amounts of substrate 12-pep (lane 1). The band labeled CL14N present in every lane is unreacted 14-mer suicide substrate. The product of the Tdp1 hydrolysis is 12-P. The identities of the bands were verified by using oligonucleotides with known structures (data not shown).

Results and Discussion

Tdp1 and Its Homologs Are Members of the PLD Superfamily. Pouliot *et al.* (16) identified the *S. cerevisiae* gene *TDP1* that codes for the phosphodiesterase Tdp1 that specifically removes covalently bound tyrosine from the 3′ end of DNA. Recent database searches revealed a full-length human cDNA (National Center for Biotechnology Information accession no. NM_018319) encoding a 608-aa Tdp1 ortholog with the predicted protein designated FLJ11090. The gene coding for this 68.5-kDa protein is located on chromosome 14. Furthermore, predicted Tdp1 orthologs are present in a number of other eukaryotic organisms. An alignment of the single orthologous protein sequences from *Caenorhabditis elegans*, *Drosophila melanogaster*, *Arabidopsis thaliana*, *Schizosaccharomyces pombe*, and *S. cerevisiae* is shown in Fig. 1.

The homology between the Tdp1 orthologs from different species extends from approximately amino acid 150 of the human protein to the C terminus. The N-terminal regions of the Tdp1 orthologs are poorly conserved and vary substantially in length. Among the regions that are best conserved are amino acids 262–289 and 492–522 of the human protein. A sequence alignment of these conserved regions with motifs 3 and 4 of representative members from each class within the PLD superfamily defined by Ponting and Kerr (1) reveals a high degree of similarity (Fig. 2). Importantly, motifs 3 and 4 encompass the highly conserved HKD sequence previously identified as a hallmark of the PLD superfamily (1–3). Highlighted in Fig. 2 are only the positions that were previously identified by Ponting and Kerr (1) as key elements of the HKD motif. Notably, the Tdp1 homologs lack the otherwise invariantly conserved aspartate highlighted in black in Fig. 2, although most do contain an aspartate 1–3 amino acids beyond this point. Interestingly, in the two crystal structures of PLD superfamily members that have been solved (4, 5), the conserved histidines, lysines, and asparagines from both HKD motifs (see Fig. 2, black shading) surround the active site pocket. However, the conserved aspartate is positioned a long way from the active site and appears to be important for protein folding or stability as mutant *Y. pestis* murine toxin and *S. typhymurium* endonuclease Nuc with amino acid changes at this site could not be expressed or purified from *E. coli* (9, 10).

Tdp1 proteins obviously possess a unique HKD motif that differs from other PLD superfamily members because the conserved aspartate residue at the usual position is absent. Furthermore, the remainder of the human Tdp1 protein sequence cannot reasonably be aligned with any of the other PLD family members listed in Fig. 2. Ponting and Kerr (1) subdivided the PLD superfamily into classes, where members within a class can be accurately aligned throughout their length, but the sequence similarities between classes are confined to the HKD motifs. Based on this definition, our findings indicate that the Tdp1 orthologs represent a distinct class within the PLD superfamily.

Mutations of the Most Conserved Amino Acids in the HKD Motif Drastically Reduce Tdp1 Activity. To confirm that human Tdp1 is a member of the PLD superfamily, several mutations were introduced in the coding region for the most conserved amino acids in the HKD motifs. N-terminally His-tagged human Tdp1 and the mutant proteins Δ1–148, H263A, H493A, H493N, K495S, and K265S (see Fig. 2) were expressed in *E. coli*, purified by using Ni-chelating chromatography, and tested for activity. In addition, a mock purification using a strain containing a plasmid deleted for the *TDP1* coding sequence [pHN1894S(ΔTDP1)] was carried out to use as a negative control.

Previously, the substrate used for assaying the *S. cerevisiae* Tdp1 was an oligonucleotide containing a 3′ phosphotyrosine. Preliminary experiments showed that human Tdp1 could cleave this substrate confirming that it was an active ortholog of the yeast enzyme (J.J.P., unpublished work). Instead of the tyrosine-oligonucleotide substrate used previously (12, 16), the substrate used here for all of the activity assays consisted of a partially double-stranded DNA oligonucleotide linked to a small trypsin-resistant topoisomerase I peptide via a 3′ phosphodiester bond to a tyrosine residue (12-pep, Fig. 3*A*; see *Materials and Methods*).

To compare the activities of the various mutant proteins with

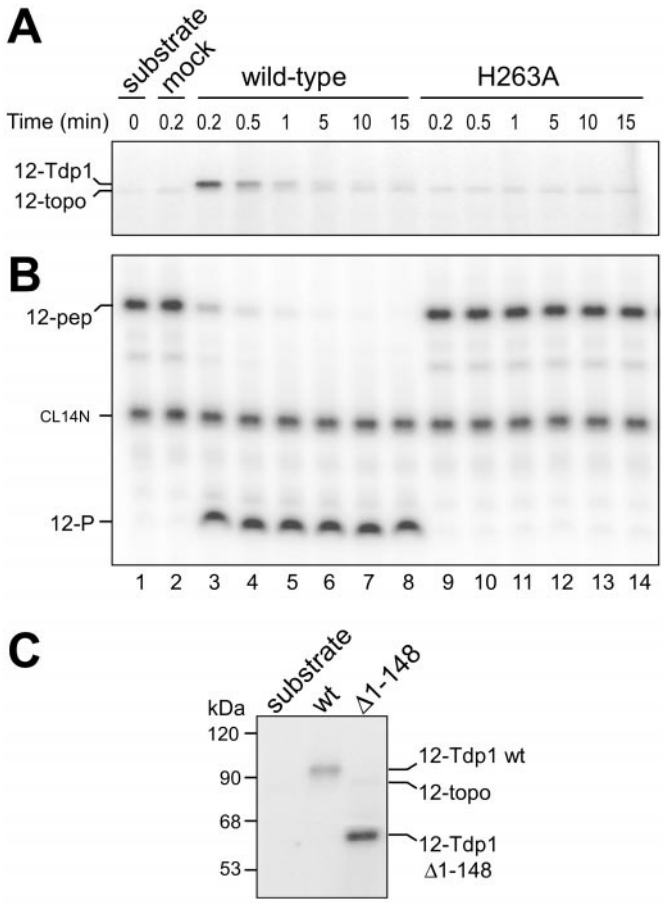


Fig. 4. Identification of a covalent intermediate in the Tdp1 reaction. (*A*) SDS/PAGE analysis of reactions stopped at the indicated times with SDS. The covalent reaction intermediate is labeled 12-Tdp1. The faint band labeled 12-topo consists of residual undigested topo I covalently bound to the 12-mer (12-topo, Fig. 3*A*). (*B*) Sequencing gel analysis of the same time-course experiment samples as in *A*. Bands are labeled as in Fig. 3*B*. (*C*) SDS/PAGE analysis of reaction intermediates trapped with SDS for wild-type (wt) and Δ1–148 Tdp1. Both proteins are shifted to a slower mobility when they are covalently bound to the substrate.

that of the wild-type Tdp1, serial dilutions of the proteins were incubated with identical amounts of the labeled oligonucleotide-peptide substrate (12-pep, Fig. 3*A*), and the products were analyzed by sequencing gel electrophoresis. The activity of the truncated Δ1–148 mutant was virtually indistinguishable from the wild-type activity (Fig. 3*B*, compare lanes 10–16 with 3–9). Thus, the poorly conserved and highly hydrophilic 148 N-terminal amino acids of Tdp1 (see Fig. 1) are not required for activity *in vitro*. This result shows that the phosphodiesterase activity is indeed confined to the conserved regions of the orthologous yeast and human Tdp1 proteins. It is extremely unlikely that the activity observed here is caused by a contaminating *E. coli* protein because *E. coli* does not possess a tyrosyl-DNA phosphodiesterase (12, 16) and no activity was detected in the mock-purified preparation when tested at the highest concentration (Fig. 3*B*, lane 2). Despite the apparent structural similarities with PLD, attempts to detect PLD activity in Tdp1 using phosphatidyl choline as the substrate were unsuccessful (J.J.C. and H.I., unpublished work).

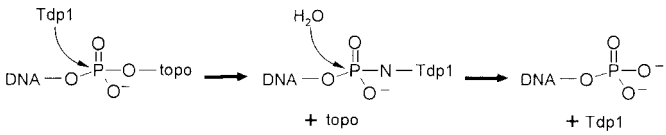
All amino acid changes in the conserved HKD motifs dramatically reduced the phosphodiesterase activity. The H263A mutant protein was more than 4 orders of magnitude less active than the wild-type protein (Fig. 3*B*, compare lanes 17 and 9) with an apparent activity very near the threshold of detection. The H493A mutant protein was about 3,000 times less active (Fig. 3*B*, lanes 20 and 8), whereas substitution with asparagine in H493N caused a 15,000-fold reduction in activity (Fig. 3*B*, lanes 23 and 9). Thus, the histidine residues in both HKD motifs of human Tdp1 are critical for activity. Mutation of K495 in the second HKD motif to serine led to an ≈125-fold reduction in activity (Fig. 3*B*, lanes 27 and 7). Although the effect of mutating this lysine residue was not as severe as mutating the conserved histidines, the results are consistent with those obtained for other members of the family, and thus implicate this conserved lysine in substrate binding and probably catalysis (4, 5, 9, 10, 19–21). In a subsequent experiment we observed that a K265S mutation reduced the activity to a nearly undetectable level (data not shown) consistent with an essential role for K265 in catalysis.

Mutations in the conserved residues, especially the histidines and lysines, in the HKD motifs of other well-studied PLD family members result without exception in an impressive loss of activity (9, 10, 19–21). Furthermore, the conserved histidines have been directly implicated in catalysis by experiments with *S. typhimurium* Nuc and *Y. pestis* murine toxin where it was demonstrated that a phosphohistidine intermediate was formed (9, 10). In addition, the crystal structures of PLD family members show close interactions between the histidines, lysines, and asparagines from both HKD motifs with the phosphate or tungstate atoms in the active site pocket (4, 5). The effects of the mutations in the HKD motifs of the human Tdp1 enzyme strongly support the hypothesis that the catalytic mechanism of this enzyme is similar to that of other PLD family members.

The Reaction Mechanism of Tdp1 Involves a Covalent Intermediate. A critical test of the proposition that Tdp1 is functionally similar to other members of the PLD superfamily is to demonstrate that a covalent enzyme-DNA intermediate forms during catalysis. To trap and detect such a potential Tdp1 reaction intermediate, a time-course experiment with the wild-type enzyme was carried out at 0°C to slow down the reaction. The H263A mutant and mock-purified proteins were used as negative controls for the experiment. After rapid denaturation with SDS and boiling, analysis by SDS/PAGE demonstrated that ^{32}P-labeled DNA was bound covalently to Tdp1 (Fig. 4*A*, lanes 3–5, 12-Tdp1). The observed protein band is indeed Tdp1 with bound labeled DNA because neither the mock-purified protein nor the inactive H263A mutant protein produced this band (Fig. 4*A*, lanes 2 and 9–14). The faint bands labeled 12-topo that are present in all lanes of Fig. 4*A* at equal intensities consist of topo70-DNA covalent complex that had not been degraded by trypsin during the preparation of the 12-pep substrate. Because topo70 and His-tagged Tdp1 are similar in size (592 vs. 633 aa), the topo70-DNA complex (12-topo) provides a convenient marker for the mobility of a protein containing a covalently attached 12-mer oligonucleotide. The fact that the labeled Tdp1 comigrated with 12-topo provides further evidence that we have trapped the reaction intermediate containing bound DNA. Furthermore, the N-terminal truncation mutant Δ1–148 can be similarly labeled and also is shifted to a slower mobility when it is covalently bound to DNA (Fig. 4*C*).

A true reaction intermediate is expected to appear during the course of the reaction and disappear when the substrate has been consumed. Fig. 4*B* shows a sequencing gel analysis with the samples from the same time course experiment that is shown in Fig. 4*A*. Importantly, the decay of the labeled wild-type Tdp1 reaction intermediate with time (Fig. 4*A*, lanes 3–6) correlates with the observed decrease of the unreacted substrate band in the sequencing gel (Fig. 4*B*, 12-pep). Consistent with the intermediate representing less than 1% of the unreacted substrate at any given time, a much longer exposure was required to observe the intermediate (Fig. 4*A*) than was needed to follow the

disappearance of substrate (Fig. 4*B*). These data reveal that the Tdp1 reaction mechanism includes a covalent reaction intermediate in accordance with the following suggested reaction scheme:



We conclude that the trapped intermediate contains covalently bound DNA linked most likely to a histidine in one of the HKD motifs as previously identified in Nuc and the murine toxin (9, 10). Accordingly, in the first step of the reaction, nucleophilic attack by this histidine releases the topoisomerase-derived peptide and forms an intermediate with the enzyme covalently bound to the 3′ phosphate end of the DNA. The proximity of the conserved lysines to the nonbridging oxygens of the phosphate in the crystal structures (4, 5) suggests that they are involved in stabilizing the transition state during catalysis. The conserved histidine residue in the other HKD motif could function as a general acid in the initial cleavage reaction to protonate the oxygen atom of the leaving tyrosine and subsequently act as a general base to activate the water for hydrolysis of the phosphohistidine intermediate. The observation that human Tdp1 catalyzes phosphodiester bond cleavage with the formation of a covalent enzyme-DNA intermediate is consistent with the idea that Tdp1 is a member of the PLD family of enzymes.

Perspectives. The sequence features found in the HKD motifs that are common to all of the protein classes in the PLD superfamily, including the Tdp1 protein, presumably derive from a common ancestral gene. Based on the known crystal structures (4, 5), it appears that this conservation of sequence reflects the underlying phosphoryl transfer chemistry that is characteristic of these enzymes. However, the enzymes within the family act on substrates with very diverse structures, ranging from phospholipids to DNA. It seems likely that the lack of substantial sequence similarity outside of the HKD motifs when comparing family members from different classes reflects, at least in part, the different structural requirements for binding different substrates.

The addition of Tdp1 to the PLD superfamily expands the family to include a specialized DNA repair function. Failure of topoisomerase I to religate DNA leaves the topoisomerase covalently attached to the 3′ end of the DNA at the site of the break. The importance of a repair pathway for such damage is underscored by the observation that prolonged treatment with the anticancer drug camptothecin, which causes the accumulation of topoisomerase I-DNA covalent complexes, is lethal for the cell (15, 22). Evaluating the contribution of Tdp1 to the repair of such lesions in human cells will require further studies. Such studies will be aided by the isolation of specific inhibitors of the enzyme, which should be expedited by our identification of its membership in the PLD superfamily.

We thank Dae Sung (Tony) Han for excellent technical assistance and Greg Ireton for preliminary studies. We thank Achim Schnaufer, Sharon Schultz, and Jamie Winshell for critical reading of the manuscript and all members of the Champoux lab for stimulating discussions. This work was supported by Grant GM49156 from the National Institutes of Health.

1. Ponting, C. P. & Kerr, I. D. (1996) *Protein Sci.* **5,** 914–922.
2. Morris, A. J., Engebrecht, J. & Frohman, M. A. (1996) *Trends Pharmacol. Sci.* **17,** 182–185.
3. Koonin, E. V. (1996) *Trends Biochem. Sci.* **21,** 242–243.
4. Stuckey, J. A. & Dixon, J. E. (1999) *Nat. Struct. Biol.* **6,** 278–284.
5. Leiros, I., Secundo, F., Zambonelli, C., Servi, S. & Hough, E. (2000) *Struct. Fold. Des.* **8,** 655–667.
6. Stanacev, N. Z. & Stuhne-Sekalec, L. (1970) *Biochim. Biophys. Acta* **210,** 350–352.
7. Raetz, C. R., Carman, G. M., Dowhan, W., Jiang, R. T., Waszkuc, W., Loffredo, W. & Tsai, M. D. (1987) *Biochemistry* **26,** 4022–4027.
8. Bruzik, K. & Tsai, M. D. (1984) *Biochemistry* **23,** 1656–1661.
9. Gottlin, E. B., Rudolph, A. E., Zhao, Y., Matthews, H. R. & Dixon, J. E. (1998) *Proc. Natl. Acad. Sci. USA* **95,** 9202–9207.
10. Rudolph, A. E., Stuckey, J. A., Zhao, Y., Matthews, H. R., Patton, W. A., Moss, J. & Dixon, J. E. (1999) *J. Biol. Chem.* **274,** 11824–11831.
11. Wang, J. C. (1996) *Annu. Rev. Biochem.* **65,** 635–692.
12. Yang, S. W., Burgin, A. B., Jr., Huizenga, B. N., Robertson, C. A., Yao, K. C. & Nash, H. A. (1996) *Proc. Natl. Acad. Sci. USA* **93,** 11534–11539.
13. Sastry, S. & Ross, B. M. (1998) *J. Biol. Chem.* **273,** 9942–9950.
14. Liehr, J. G., Giovanella, B. C. & Verschraegen, C. F., eds. (2000) *Ann. N.Y. Acad. Sci.* **922,** 1–360.
15. Hsiang, Y. H., Hertzberg, R., Hecht, S. & Liu, L. F. (1985) *J. Biol. Chem.* **260,** 14873–14878.
16. Pouliot, J. J., Yao, K. C., Robertson, C. A. & Nash, H. A. (1999) *Science* **286,** 552–555.
17. Pouliot, J. J., Robertson, C. A. & Nash, H. A. (2001) *Genes Cells* **6,** 677–687.
18. Stewart, L., Ireton, G. C., Parker, L. H., Madden, K. R. & Champoux, J. J. (1996) *J. Biol. Chem.* **271,** 7593–7601.
19. Sung, T. C., Roper, R. L., Zhang, Y., Rudge, S. A., Temel, R., Hammond, S. M., Morris, A. J., Moss, B., Engebrecht, J. & Frohman, M. A. (1997) *EMBO J.* **16,** 4519–4530.
20. Xie, Z., Ho, W. T. & Exton, J. H. (2000) *J. Biol. Chem.* **275,** 24962–24969.
21. Iwasaki, Y., Horiike, S., Matsushima, K. & Yamane, T. (1999) *Eur. J. Biochem.* **264,** 577–581.
22. Pourquier, P. & Pommier, Y. (2001) *Adv. Cancer Res.* **80,** 189–216.