T4 DNA ligase structure reveals a prototypical ATP-dependent ligase with a unique mode of sliding clamp interaction

Ke Shi[1], Thomas E. Bohl[1], Jeonghyun Park[1], Andrew Zasada[1], Shray Malik[1], Surajit Banerjee[2], Vincent Tran[1], Na Li[1], Zhiqi Yin[1], Fredy Kurniawan[1], Kayo Orellana[1] and Hideki Aihara [1,*]

[1]Department of Biochemistry, Molecular Biology, and Biophysics, University of Minnesota, 6–155 Jackson Hall, 321 Church Street S.E. Minneapolis, MN 55455, USA and [2]Northeastern Collaborative Access Team, Cornell University, Advanced Photon Source, Lemont, Illinois, 60439, USA

ABSTRACT

DNA ligases play essential roles in DNA replication and repair. Bacteriophage T4 DNA ligase is the first ATP-dependent ligase enzyme to be discovered and is widely used in molecular biology, but its structure remained unknown. Our crystal structure of T4 DNA ligase bound to DNA shows a compact α-helical DNA-binding domain (DBD), nucleotidyl-transferase (NTase) domain, and OB-fold domain, which together fully encircle DNA. The DBD of T4 DNA ligase exhibits remarkable structural homology to the core DNA-binding helices of the larger DBDs from eukaryotic and archaeal DNA ligases, but it lacks additional structural components required for protein interactions. T4 DNA ligase instead has a flexible loop insertion within the NTase domain, which binds tightly to the T4 sliding clamp gp45 in a novel α-helical PIP-box conformation. Thus, T4 DNA ligase represents a prototype of the larger eukaryotic and archaeal DNA ligases, with a uniquely evolved mode of protein interaction that may be important for efficient DNA replication.

INTRODUCTION

DNA ligases seal breaks in the phosphodiester backbone of DNA (1–3). They are essential enzymes for life and are found in all kingdoms of life as well as in many viruses (4–6). In DNA replication, DNA ligase joins the Okazaki fragments produced during the lagging strand synthesis to generate a continuous DNA strand. In DNA repair and recombination, DNA ligase seals strand breaks as the ultimate step, after processing of DNA lesions or exchange of DNA strands. DNA ligases from eukaryotes, archaea, and some bacteria that play an essential role in replication physically associate with the DNA polymerase processivity factor known as the sliding clamp (PCNA and its homologs) (7–10). The ligase-sliding clamp interaction helps recruit DNA ligase to nicks between newly synthesized Okazaki fragments or the sites of DNA lesions, and thereby facilitates completion of DNA replication and repair (11–14). However, an interaction between virally encoded ligase and a sliding clamp has not been reported.

All classic DNA/RNA ligases join a 5′-phosphorylated and a 3′-hydroxyl (OH) end to generate a phosphodiester linkage, employing either ATP or NAD+ as the energy source for the formation of the new chemical bond (15). These enzymes all share a conserved nucleotidyl-transferase (NTase) domain that harbors the adenine nucleotide-binding pocket. In case of DNA ligases, the NTase domain is coupled to a C-terminal OB-fold domain. The OB domain in ATP-dependent ligases is essential for the transfer of AMP moiety to an active site lysine residue in the NTase domain during the first step of the DNA ligation reaction (16), whereas the structurally distinct domain Ia plays an analogous role in NAD$^+$-dependent ligases (17,18). The OB domain is also essential in positioning the nicked DNA substrate during all following steps of the ligation reaction, which entails transfer of AMP to the DNA 5′-phosphate and an attack of the 3′-hydroxyl group on the phosphate to seal the nick (19). It is believed that relative conformation of the NTase and OB domains changes significantly between the lysine-adenylation step prior to DNA-binding and in the ligase–DNA complex. The NTase and OB domains represent the minimal ATP-dependent DNA ligase, as found for the bacteriophage T7 ligase (20). These two domains are also conserved in the eukaryotic mRNA-capping enzymes that catalyze the transfer of guanine nucleotide to the 5′-phosphorylated terminus of RNA (21).

---

*To whom correspondence should be addressed. Tel: +1 612 624 1491; Email: aihar001@umn.edu

In addition to NTase and OB domains that constitute the catalytic core of DNA ligase, the ATP-dependent ligases from eukaryotes and archaea have a third DNA-binding domain (DBD) in their N-terminus (10,19,22–24). The DBDs consist of 250–300 residues and are structurally a bundle of 13–15 α-helices (Supplementary Figure S1). DBD provides most of the DNA-binding affinity in case of human ligase 1 (hLig1) and it allows the three-domain (DBD-NTase-OB) ligases to completely wrap around the nicked double-stranded DNA substrate (19). While the viral DNA ligases structurally characterized to date lack the N-terminal DBD and are smaller than cellular ligases, the ATP-dependent DNA ligase from Chlorella virus uses a unique β-hairpin insertion within the OB domain (latch) to similarly wrap around the DNA substrate (25). The complete encircling of DNA is a common feature also found in NAD+-dependent DNA ligases; the *Escherichia coli* NAD+-dependent DNA ligase uses a set of HhH, Ia, and zinc-finger domains structurally distinct from the DBD of ATP-dependent ligases to wrap around the bound DNA substrate (26).

DNA ligase from bacteriophage T4 was the first ATP-dependent ligase enzyme as well as one of the first polynucleotide ligases to be identified (27–29). Since its discovery a half century ago, T4 DNA ligase has been widely used in various applications ranging from molecular cloning, library construction and high-throughput DNA sequencing (30–32), to the detection of small RNAs (33). The amino acid sequence of T4 DNA ligase suggests that it has the NTase-OB domains catalytic core as found in all DNA ligases, but it has a ∼130 amino acids extension on its N-terminus. This is in contrast to the other structurally characterized viral DNA ligases; as mentioned above, bacteriophage T7 DNA ligase lacks an additional structural domain outside the NTase-OB core (20), and the Chlorella virus ligase has an extended β-hairpin insertion within the OB domain (25). The N-terminal extension of T4 DNA ligase does not show significant sequence homology to, and is much smaller than the DBD of eukaryotic and archaeal DNA ligases. However, deletion of this N-terminal region severely compromises the DNA ligation activity of T4 DNA ligase in vitro (34) suggesting that this N-terminal region facilitates DNA-binding or nick-sensing. Specific mode of substrate DNA-binding enabled by the N-terminal domain may confer T4 DNA ligase the useful characteristics for *in vitro* applications, such as the robust activity toward blunt-end double-stranded DNA substrates.

Bacteriophage T4 has served as a model system to study basic mechanisms of DNA replication (35–43), recombination, and repair (44). We reasoned that structural information of T4 DNA ligase could help improve comprehensive understanding of these processes. In addition, given the popularity of T4 DNA ligase in various applications, knowledge of atomic details of its DNA-interactions could enable design of engineered DNA ligases fit for specific applications. Here we report crystal structure of T4 DNA ligase in complex with DNA, identification of a sliding clamp (gp45)-interaction motif, and crystallographic analysis of the ligase-gp45 interaction. The results highlight both conserved and unique structural features of T4 DNA ligase and suggest its evolutionary relationship to the eukaryotic DNA ligases.

MATERIALS AND METHODS

Protein production

The coding sequences for T4 DNA ligase (gp30) and gp45 were amplified by polymerase chain reaction from bacteriophage T4 (ATCC). Full-length wild-type T4 DNA ligase, T4 DNA ligase NTase-OB domain fragment (136–487), and human Lig1 (233–919) were expressed with an N-terminal His$_6$-tag (MGSSHHHHHHSSGLVPRGSH) using pET28a, and T4 DNA ligase DBD fragment (1–129), internally truncated ligase T4LΔ(222–247)GS, and the full-length gp45 were expressed with a C-terminal His$_6$-tag (LEHHHHHH) using the pET24a vector. T4 clamp-loader was prepared by co-expressing gp44 and gp62 as reported previously (45). All proteins were expressed in *E. coli* strain BL21(DE3) and purified using nickel-affinity and size-exclusion chromatography. His$_6$-tags were not removed.

Crystallization

T4 DNA ligase–DNA complex was assembled by mixing purified T4 DNA ligase with an equimolar amount of a nicked double-stranded DNA substrate (GCTGATGCGTddC/pGTCGGACTGA/TCAGTCCG ACGACGCATCAGC) at an approximate final protein concentration of 12 mg ml^{-1} in 50 mM Tris–HCl pH 7.5, 10 mM MgCl$_2$, 2 mM ATP, 10 mM dithiothreitol (DTT), 0.1 M NaCl, 5% glycerol and incubating the mixture at room temperature for 15 min. The complex was crystallized by sitting-drop vapor diffusion at 20°C, in a drop formed by mixing 0.1 μl of the complex with 0.1 μl of the reservoir solution containing 0.1 M Bis–Tris pH 5.5, 0.1 M ammonium acetate, and 17% polyethylene glycol (PEG) 10 000 (JCSG+ suite, H6). Crystals were transferred to the reservoir solution supplemented with 25% glycerol and flashed cooled in liquid nitrogen. Selenomethionine (SeMet)-labeled T4 DNA ligase DBD fragment (1–129) was crystallized by hanging-drop vapor diffusion by mixing 1 μl of purified protein at ∼17 mg ml^{-1} in 10 mM Tris–HCl pH 7.4, 0.2 M NaCl, 10 mM β-mercaptoethanol with 1 μl of a reservoir solution containing 30% PEG 4000 and 0.1 M Tris–HCl pH 8.5, and the crystals were harvested similarly to above. Gp45-peptide complex crystals were obtained by hanging-drop vapor diffusion by mixing 1 μl of purified gp45 at ∼30 mg ml^{-1} in 10 mM Tris–HCl (pH 7.4), 0.2 M NaCl, 10 mM β-mercaptoethanol and 2 mM T4 ligase peptide (KKEPEGLDFLFDA) with 1 μl of a reservoir solution containing 6% PEG 10 000. The crystals were subsequently further soaked with 1 mM of the same peptide in the reservoir solution supplemented with 20% ethylene glycol for 5 min before being flash cooled for data collection.

Structure determination

X-ray diffraction data were collected at the Advanced Photon Source 24-ID-C and 24-ID-E (NE-CAT) beamlines. Full-length T4 DNA ligase–DNA complex crystal diffraction data were integrated and scaled with HKL2000 (46). T4 DNA ligase DBD and gp45-peptide complex diffraction

data were integrated and scaled with XDS (47). The structure of the SeMet-derivatized T4 ligase DBD was solved by SAD phasing using selenium anomalous signals. Five selenium atoms were located using SHELXD (48) and phasing was done using PHASER (49). Density modification was performed with PARROT (50) to improve the initial phases. BUCCANEER (51) found 100% of the amino acids by autobuilding. The model was refined using REF-MAC (52). Structure solution of the full-length T4 DNA ligase–DNA complex was achieved by molecular replacement in PHASER, using the Chlorella virus DNA ligase–DNA complex (PDB ID: 2Q2T) and the T4 DNA ligase DBD structure above as search models. Two copies of the complexes were found in an asymmetric unit. Iterative refinement and model building were done using PHENIX (53) and COOT (54). T4 gp45-peptide complex structure was solved by molecular replacement using PHASER and the T4 gp45 structure (PDB ID: 1CZD) as the search model. One copy of the gp45 trimer was found in an asymmetric unit. Subsequent refinement using PHENIX generated a map that clearly showed the bound ligase peptide. The peptide was built using COOT and the model was further refined using PHENIX. The data collection and final refinement statistics are show in Table 1. Figures of molecular structures were generated using PyMOL (www.pymol.org) and electrostatic potentials were calculated using PDB2PQR (55) and APBS (56). DNA structural parameters were calculated using CURVES+ (57) and 3DNA (58) and plotted with Veusz licensed under the GPL.

Isothermal titration calorimetry (ITC)

ITC was performed to determine the dissociation constant (K_d) for binding of the T4 ligase loop peptide to gp45. Purified gp45 was dialyzed against 0.2 M NaCl, 10 mM Tris–HCl pH 7.4, and 10 mM β-mercaptoethanol at 4°C in a 3.5 kDa cutoff membrane. An HPLC-purified loop peptide KKEPEGLDFLFDA (>95% pure, Selleck Chemicals) was dialyzed against the same buffer in a 2 kDa cutoff Slide-A-lyzer cassette (Thermo Fisher). Peptide and gp45 were adjusted to the desired concentrations with used dialysis buffer as measured by A_{257} (ε = 220 M^{-1} cm^{-1}) and A_{280} (ε = 19 940 M^{-1} cm^{-1}), respectively. ITC was performed on a Nano ITC Low Volume isothermal titration calorimeter (TA Instruments). ITC was run at 25°C with 300 rpm stirring. The instrument was equilibrated for 30 min to reach a stable baseline. After a small injection of 0.49 μl to remove buffer that had diffused into the needle, 2.49 μl of loop peptide (525 μM) was injected into the experimental cell with an initial 300 μl gp45 (51 μM) every 5 min for 20 injections. The reference cell was always filled with water. Four replicates were performed as described above, and a blank experiment was performed with loop peptide injected into used dialysis buffer. A fifth experiment was performed with 100 μM gp45 in the experimental cell. The data were analyzed in the NanoAnalyze software (TA Instruments).

DNA ligation assays

For testing blunt-end ligation, a 15 bp substrate phosphorylated on both termini (pGTCGGACTGAT TCGG/pCCGAATCAGTCCGAC) at 50 μM was incubated with indicated concentrations of the proteins in 50 mM Tris–HCl pH 7.5, 10 mM MgCl$_2$, 1 mM ATP, 10 mM DTT in a total volume of 20 μl. After 10-minute incubation at room temperature, the reaction was stopped by the addition of EDTA to 0.1 M final concentration. The reaction products were separated by native PAGE on a 6% acrylamide-TBE gel and visualized by ethidium bromide staining. The DNA substrate used for testing ligation of multiple nicks on a partially double-stranded circular DNA was prepared by annealing five oligonucleotides (5′ 6-FAM-AGTGCCAAGCTTGCATGCCTGCAGG TCGACTCT/pAGAGGATCCCCGGGTACCGAGC /pTCGAATTCGTAATCATGGTCAT/pAGCTGTTT CCTGTGTGAAATTG/pTTATCCGCTCACAATTCC ACAC) with M13mp18 single-stranded DNA (Bayou Biolabs). The DNA substrate at 100 nM was first incubated with indicated amounts of gp45, gp44/62, or BSA for 10 min at room temperature in 50 mM Tris–HCl pH 7.5, 10 mM MgCl$_2$, 1 mM ATP, 10 mM DTT, and 150 mM KCl. Subsequently T4 DNA ligase was added and the mixture was further incubated for 1 hr at room temperature. Reaction was stopped by the addition of formamide (87.5% v/v final) and heating to 98°C. 10 μl of the sample was separated on a 15% acrylamide–TBE–urea gel and the gel was scanned using a Typhoon FLA9500 imager. The band intensities were measured using ImageQuant TL software. We confirmed that our gp45 and gp44/62 preparations have no detectable DNA ligase activity (data not shown).

Size-exclusion chromatography

Proteins were separated on the Superdex200 10/300 size-exclusion column equilibrated and operating with a running buffer containing 20 mM Tris–HCl pH 7.4, 0.2 M NaCl, and 10 mM β-mercaptoethanol at the flow rate of 0.5 ml min^{-1}. 500 μl of gp45 (∼3 mg ml^{-1}), wild-type or the internally truncated T4 DNA ligase (∼2 mg ml^{-1}), or their mixtures at indicated molar ratios were injected. The proteins were detected by UV absorption at 280 nm. Molecular mass standards to calibrate the column were bovine γ-globulin (158 kDa), ovalbumin (44 kDa), horse myoglobin (17 kDa), vitamin B$_{12}$ (1.35 kDa).

Fluorescence anisotropy

Wild-type T4 ligase, T4 ligase 1–129, T4 ligase 136–487, and human Lig1 were dialyzed against 0.1 M NaCl, 20 mM Tris–HCl pH 7.4, 0.01% Triton X-100 (Acros), and 5 mM β-mercaptoethanol at 4°C in 3.5 kDa cutoff dialysis tubes (Thermo) overnight. The protein concentrations after dialysis were determined by UV absorption at 280 nm. Two-fold dilution series of the ligases were prepared in dialysis buffer with 100 μg/ml bovine serum albumin (Sigma A7030) and 15 nM of a 5′ fluorescein-labeled DNA probe. Intact and nicked dsDNA probes were prepared by annealing the 5′ fluorescein-labeled strand (5′ 6-FAM-GGTAGCCACAGCCAGTCAGCCGATTGCGGGAC-3′) with either 5′-GTCCCGCAATCGGCTGACTGGCT GTGGCTACC-3′ or 5′-GTCCCGCAATCGGCTG-3′ and PO$_4$-ACTGGCTGTGGCTACC-3′, respectively.

Table 1. Data collection and refinement statistics

	Se – DBD	T4 DNA ligase/DNA	gp45/T4 ligase–peptide
PDB ID	5WFY	6DT1	6DRT
Data Collection			
Wavelength (Å)	0.979	0.979	0.979
Resolution range (Å)	51.03–1.40 (1.45–1.40)	47.22–2.75 (2.85–2.75)	58.76–2.11 (2.19–2.11)
Space group	$P2_1$	$P1$	$P2_12_12_1$
Unit cell			
a, b, c (Å)	28.81, 40.80, 51.07	63.53, 67.43, 114.64	63.76, 90.95, 151.40
α, β, γ (°)	90, 92.35, 90	88.78, 80.94, 62.97	90, 90, 90
Total reflections	178 856 (16 893)	84 628 (8465)	254 087 (25 965)
Unique reflections	23 250 (2281)	42 899 (4259)	50 446 (4893)
Multiplicity	7.7 (7.4)	2.0 (2.0)	5.0 (5.2)
Completeness	0.99 (0.98)	0.98 (0.98)	0.99 (1.00)
$I/\sigma(I)$	16.3 (1.2)	7.1 (1.4)	16.9 (1.7)
R-merge (%)	15.11 (156.3)	9.06 (54.65)	6.60 (100.9)
R-meas (%)	16.19 (168.2)	12.82 (77.3)	7.36 (112.0)
$CC_{1/2}$	0.997 (0.386)	0.983 (0.598)	0.999 (0.634)
Refinement			
Reflections used	23 228 (2281)	42 747 (4254)	50 244 (4893)
Reflections for R-free	1190 (119)	2154 (205)	2548 (250)
R-work (%)	14.13 (28.20)	16.98 (27.79)	19.88 (30.50)
R-free (%)	18.15 (33.88)	21.90 (32.25)	24.03 (33.01)
Number of non-H atoms	1197	9525	5848
Macromolecules	1063	9089	5514
Ligands	6	167	28
Protein residues	132	923	717
r.m.s.d.			
Bond length (Å)	0.007	0.009	0.001
Bond angles (°)	0.87	0.81	0.38
Ramachandran plot			
Favored (%)	98.46	97.27	96.6
Allowed (%)	1.5	2.73	3.4
Outliers (%)	0	0	0
Average B-factor (Å^2)	18.50	76.03	58.47
Macromolecules (Å^2)	17.16	76.82	58.54
Ligands (Å^2)	22.43	75.60	70.62
Solvent (Å^2)	29.45	49.48	55.95

Statistics for the highest-resolution shell are shown in parentheses.

Samples were prepared in black, flat bottom, low volume 384-well plates (Corning 3821BC) at 20 μl except for the samples with nicked dsDNA and human Lig1, which were prepared at 15 μl. Fluorescence polarization was measured with a Tecan Spark 10M with the excitation and emission wavelengths set to 459 and 517 nm. The Z-position for the measurement was calculated from a representative well, and the gain and instrument G-factor were optimized for the run. The maximum concentration of protein used was 100 μM except for T4 ligase 1–129 where it was 400 μM. Each dilution series was prepared in triplicate. The average polarization of the DNA probes in the absence of ligase was subtracted from the data, and the data were fit by the ligand-depletion single-site binding model in GraphPad Prism, which also calculated the 95% confidence intervals for the dissociation constants.

RESULTS

Structure determination of T4 DNA ligase

We obtained crystals of the full-length T4 DNA ligase in complex with a 21 base-pair (bp) nicked duplex DNA, in the presence of ATP and magnesium ion. The DNA substrate was assembled from a 11-mer upstream, a 10-mer down-stream, and a 21-mer template strand, and had the same sequence as that used in the crystallographic study of hLig1 (19) on either side of the nick. We used a 5′-phosphorylated downstream strand to allow enzymatic transfer of AMP to its 5′-terminus, and an upstream strand terminated with a 2′,3′-dideoxyribonucleotide to block the final step of the nick sealing reaction. As poor crystal reproducibility precluded experimental phasing of the T4 ligase–DNA complex structure, we have separately crystallized and determined the structure of the N-terminal domain of T4 DNA ligase (residues 1 – 129) at 1.40 Å resolution by the selenomethionine single-wavelength anomalous diffraction (SAD) phasing (Table 1, Supplementary Figure S2B). The structure of the T4 ligase–DNA complex was in turn determined by molecular replacement phasing using the structures of the T4 DNA ligase N-terminal domain and the Chlorella virus ligase–DNA complex (25) as search models and was refined to 2.75 Å resolution (Table 1, Supplementary Figure S2A). In the T4 DNA ligase–DNA crystals, there are two ligase–DNA complexes in the asymmetric unit with essentially same structures (backbone r.m.s.d. for protein = 0.4 Å for 459 amino acids). As one of the two ligase–DNA complexes showed clearer electron density with a lower average B-factor in the refined coordinates than the other (average B values of 41.9 versus 101.6 Å^2), we will

focus on description of the better ordered complex (chains A, B, C, D).

Overall structure of T4 DNA ligase–DNA complex

The refined atomic model of the T4 DNA ligase–DNA complex (Figure 1A–C) includes the entire 21 bp DNA and the whole ligase polypeptide from residue 1 through 487, with the exception of a disordered loop spanning residues 222–247 (discussed in more detail below). Electron density map showed that an adenylate moiety is attached to the 5′-terminus of the downstream strand at the nick via a pyrophosphate linkage (Figure 4D), but not to the active site residue Lys159, indicating that the structure represents a snapshot of DNA ligation reaction prior to the final step of nick sealing. T4 DNA ligase consists of three structural domains, an α-helical N-terminal DNA-binding domain (DBD; residues 1–129), the NTase (residues 133–367), and the OB-fold (residues 370–487) domains. The three domains connected through short linkers together wrap around the bound DNA substrate, with each of the domains positioned primarily over the DNA minor groove and making extensive backbone contacts. The complete encircling of DNA by a C-shaped protein clamp is reminiscent of the modes of DNA-binding observed for mammalian DNA ligases (19,22), and the *E. coli* NAD+-dependent DNA ligase (26). As observed for these ligases, a salt bridge between Lys384 from the OB domain and Asp112 from DBD completes the encircling of DNA by T4 DNA ligase (Figure 1B, Supplementary Figure S2A). The loop harboring Lys384 corresponds to (albeit shorter than) the β-hairpin segment (latch) in Chlorella virus ligase (25).

The extensive DNA contacts made by T4 DNA ligase involve a total of 55 amino acid residues from the three structural domains and span a 18 bp region centered on the nick (Supplementary Figure S3), burying a total of 2503 Å^2 surface areas of the protein and DNA. These interactions hold the nicked double-stranded DNA in a distorted conformation that deviates from the canonical B-form; the DNA is slightly underwound and bent (16°) toward the N-terminal DBD, and shows a significantly widened minor groove adjacent to the nick bound by NTase and OB domains. The nucleotides take the 3′-endo sugar pucker for a span of several base-pairs around the nick (nucleotides G9–G12 and A30–A33, Supplementary Figure S3), corresponding to the A-form RNA-like conformation upstream of the nick. The DNA helical axis has a large offset (>5 Å) across the nick, which is stabilized by an inter-strand aromatic stacking between two guanine bases (G12 and G32) from complementary strands across the nick (Figure 1D) and van der Waals interaction between the 3′-terminal base upstream of the nick (C11) and the deoxyribose moiety of the 5′-terminal nucleotide downstream of the nick (G12). Overall, the DNA substrate bound to T4 DNA ligase shows a similar conformation to those observed for hLig1 (19) and the Chlorella virus ligase (25), except that the DNA bound to T4 DNA ligase shows a slightly smaller degree of bending (Supplementary Figure S4). The binding affinity (K_d) of T4 DNA ligase for nicked and unnicked DNA, respectively, was determined using fluorescence anisotropy to be 6.3 (95% confidence interval: 4.8–8.2) and 6.6 (5.6–7.8) μM, which are comparable to the K_d of hLig1 measured under the same conditions (7.7 [4.2–14] and 11 [6.2–20] μM) (Supplementary Figure S5). However, unlike the case of hLig1 where DBD accounts for bulk of the affinity (19), T4 DNA ligase NTase-OB domain core binds DNA with higher affinities (23 [18–29] and 17 [13–21] μM) than the isolated DBD (138 [97–201] and 119 [98–144] μM), indicating that relative contributions of the different domains vary between the enzymes.

N-terminal DNA-binding domain (DBD)

The N-terminal DBD of T4 DNA ligase is a bundle of seven α-helices, and it engages two regions of the DNA minor groove separated by one turn of the helix (Figure 2B, Figure 3). Notably, T4 DNA ligase DBD shows significant structural similarity to the core DNA-binding components of the DBDs of larger mammalian and archaeal DNA ligases (Figure 2, Supplementary Figure S1), despite not sharing appreciable sequence homology. Although DBD of hLig1 is a larger domain with 14 α-helices and contains approximately twice as many amino acids as the T4 DNA ligase DBD (19), the α-helices in the core of hLig1 DBD that contact DNA are positioned similarly to the helices of T4 DNA ligase DBD (Figure 2C) and the two DBDs can be superimposed with an r.m.s.d. of 2.8 Å over 124 Cα atoms. Accordingly, the mode of DNA-binding of T4 DNA ligase DBD closely mimics that of hLlig1 (19) and hLig3 DBD (22). As pointed out for the DBD of hLig1 and for the structurally distinct HhH domain of *E. coli* LigA (26,59), the structure of T4 DNA ligase DBD exhibits an internal pseudo dyad that mirrors the two-fold symmetry of the double-stranded DNA it is bound to. The DNA molecules bound to T4 DNA ligase and hLig1 show a better fit when the two complexes are superimposed based on the DBDs rather than the catalytic (NTase) domains of the enzymes (Supplementary Figure S4), consistent with the role of DBD as a major DNA-binding platform for ligation.

Most amino acid residues of the T4 DNA ligase DBD involved in DNA backbone phosphate contacts are from the N-terminal ends of α-helices and the short loops connecting the antiparallel helices. These include Ser14, Thr15 and Lys16 from α2, Gly116, Ser118, Val119, Ser120, Ile121 and Lys124 from α7, Arg79 from the loop connecting α4 to α5, and Thr82 and Asn84 from α5 (Figure 3, Supplementary Figure S3B). In addition, an extended loop between α3 and α4, which includes Gln44, Tyr45, Tyr46, Lys48, and Lys49, runs parallel to the DNA backbone and makes extensive phosphate contacts with the nicked DNA strand upstream of the nick. Of note, we observed a sausage-shaped electron density with one of its ends anchored in a deep pocket adjacent to this loop. The density was interpreted as part of a PEG molecule present in the crystallization condition. The PEG chain tracks along the minor groove of DNA and bridges between DBD and NTase domain, apparently stabilizing the ligase–DNA complex (Supplementary Figure S6).

In contrast to the high structural similarity of DNA-binding elements between the DBDs of T4 DNA ligase and hLig1, the smaller T4 ligase DBD lacks the extra structural
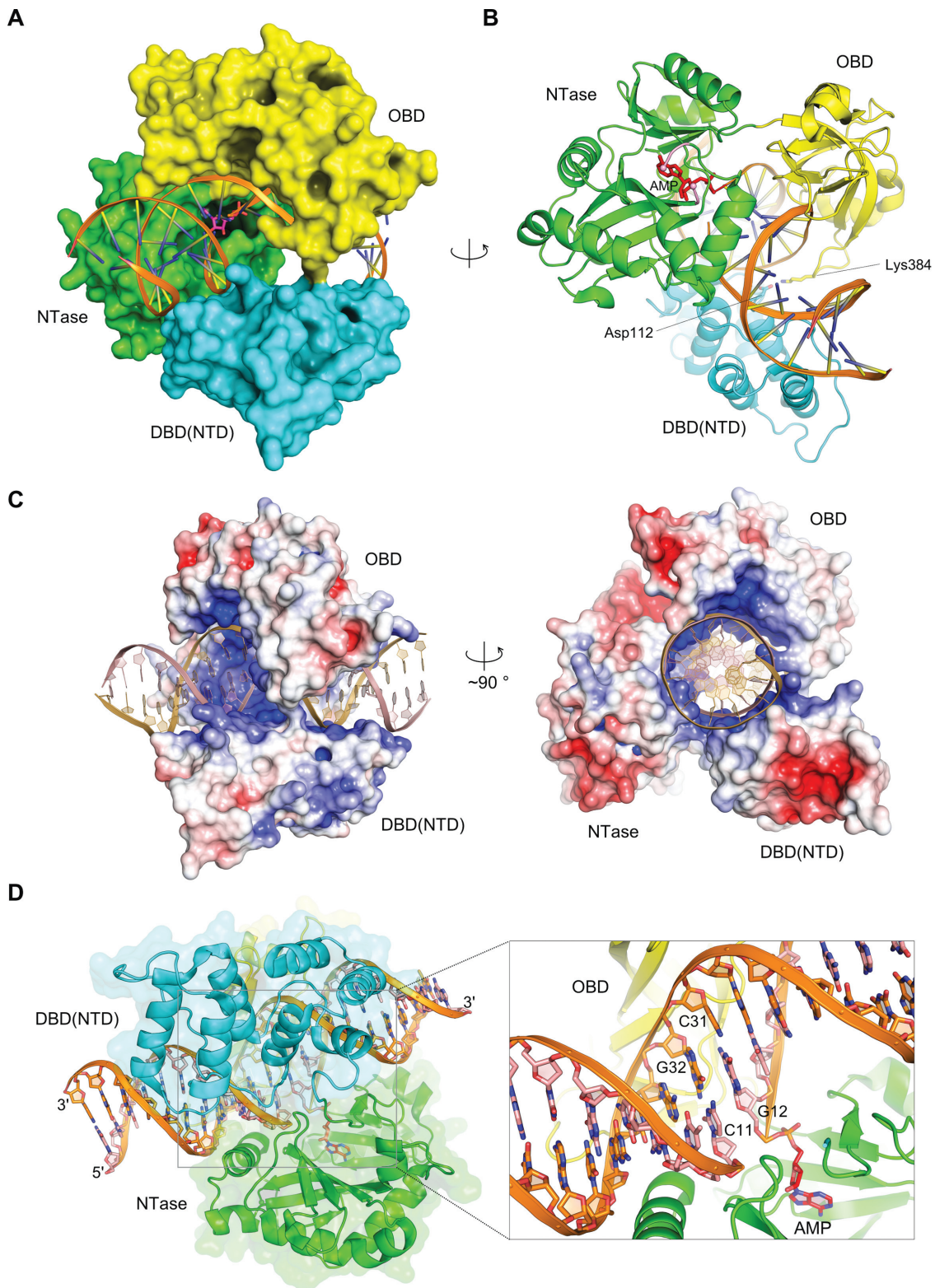
Figure 1. Overall structure of the T4 ligase–DNA complex. (**A**, **B**) Molecular surface (**A**) and cartoon representation (**B**) of the T4 ligase–DNA complex, with the three structural domains of ligase color-coded. A salt bridge between OB-fold domain residue Lys384 and DBD residues Asp112 completes the enveloping of DNA by the ligase. (**C**) Protein surface colored according to electrostatic potential (red, $-5kT\ e^{-1}$, to blue, $+5kT\ e^{-1}$). (**D**) Inter-strand base stacking (G12 and G32) across the nick stabilizing a large shift in the DNA helical axis.
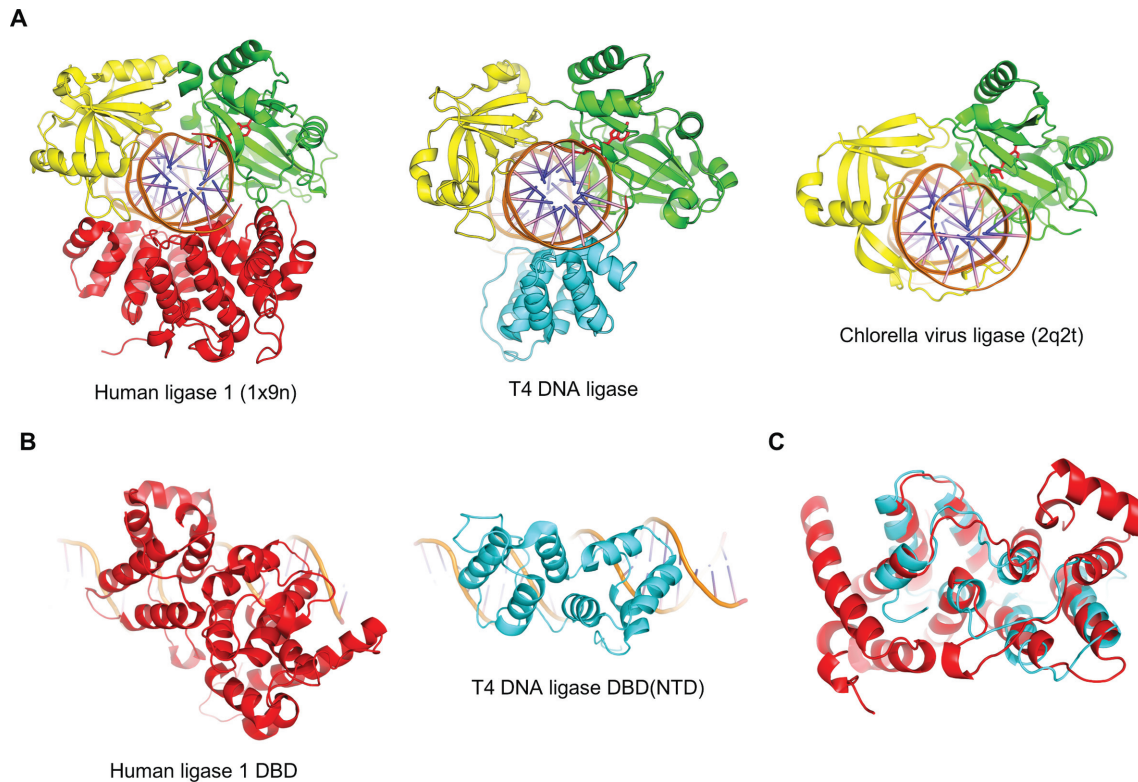
Figure 2. Structural comparison between ATP-dependent DNA ligases. (**A**) Human Ligase 1 (hLig1, DBD in red) (19), T4 DNA ligase (DBD in cyan), and Chlorella virus ligase–DNA complex (25) structures shown in similar orientations, from left to right. (**B, C**) Side-by-side comparison (B) and superposition (C) of the DBDs from hLig1 and T4 DNA ligase. Color scheme follows that in (A).
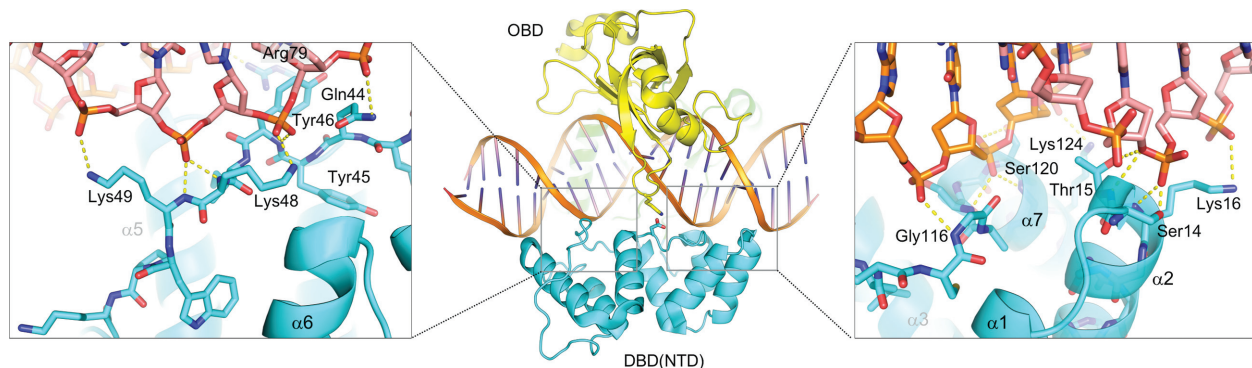

Figure 3. DBD-DNA contacts. T4 DNA ligase DBD–DNA interactions, with details shown for two regions in magnified views. The yellow dotted lines highlight hydrogen bonding or salt bridge interactions.

elements found on either side of the core DNA-binding helices in the DBDs of eukaryotic and archaeal DNA ligases (Figure 2B, C, Supplementary Figure S1). This results in less extensive contacts between DBD and the NTase domain for T4 DNA ligase. Importantly, the peripheral structural elements of the DBDs of mammalian and archaeal ligases have been shown to mediate interaction with the DNA polymerase processivity factor PCNA (10,60), or other repair factors (23). The structure of T4 DNA ligase DBD does not change significantly upon interaction with DNA, as shown by a close superposition of the DNA-bound and free crystal structures (r.m.s.d. of 0.55 Å for all atoms).

DNA ligase catalytic core (AMP interactions)

The NTase and OB-fold domains constitute the catalytic core of DNA ligase. Among the structurally characterized polynucleotide ligases, the NTase-OB domain core of T4 DNA ligase is most similar to that of Chlorella virus ligase (r.m.s.d. = 1.8 Å for 161 Cα atoms), consistent with their close phylogenetic relationship (amino acid sequence alignment and the conserved motifs are shown in Supplementary Figure S7). The NTase domain fold comprises two lobes, each consisting of a layer of β-sheet flanked by α-helices, and the cleft formed between the two lobes serves as the adenine nucleobase-binding site. The adenine base
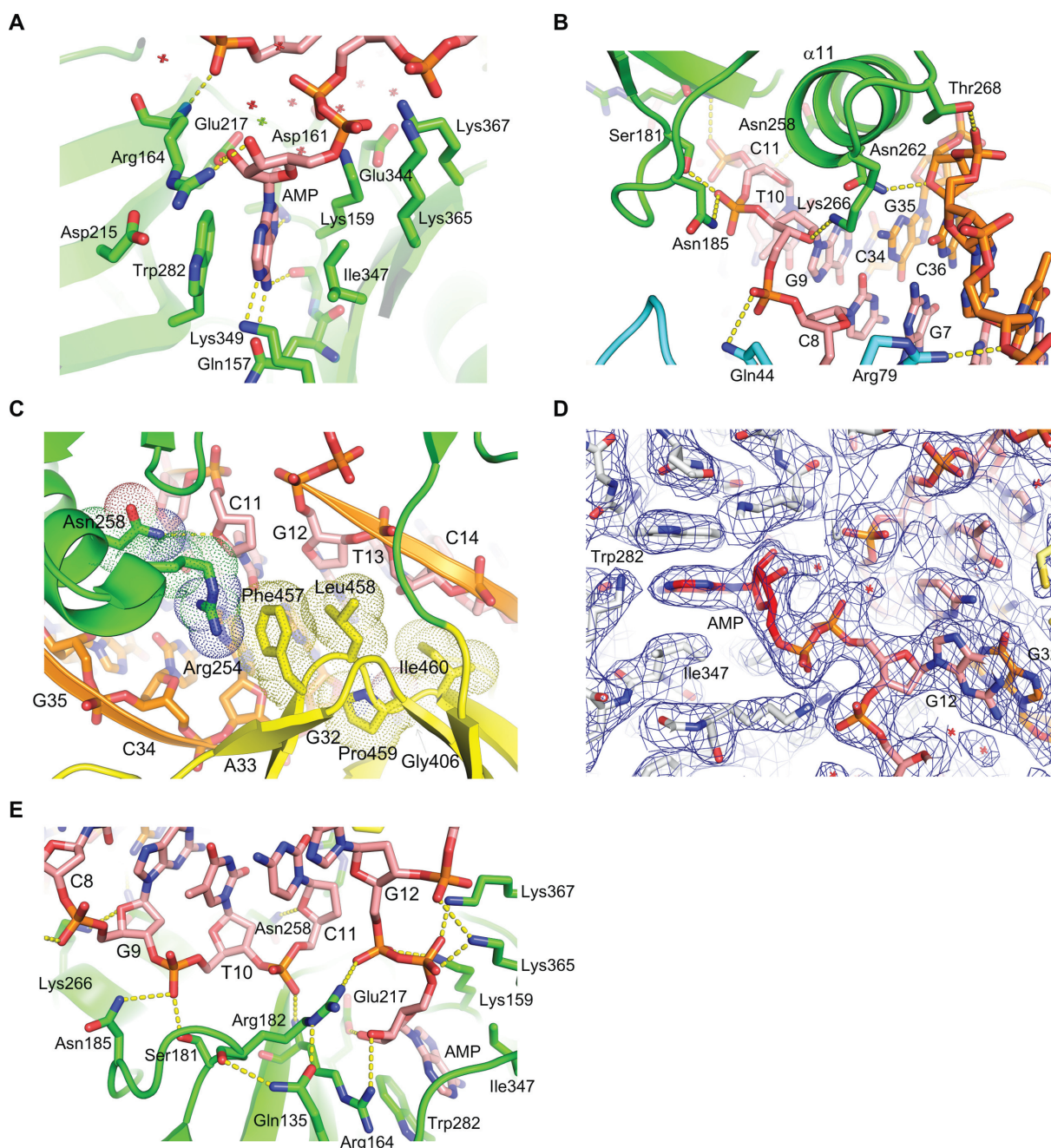
Figure 4. T4 ligase–DNA interactions around the nick. (**A**) AMP interactions with the NTase domain residues. Red and green crosshairs represent water molecules and a putative magnesium ion, respectively. (**B**) Helix α11 from the NTase domain making DNA minor groove contacts. Viewed from the C-terminal end of α11. (**C**) A continuous minor groove interactions made by the N-terminal end of α11 from NTase domain and hydrophobic residues from OBD. (**D**) 2mFo-DFc electron density map contoured at 1.0σ shown for the region around the active site. (**E**) DNA backbone contacts by the NTase domain residues across and flanking the nick. The yellow dotted lines highlight hydrogen bonds or salt bridges.

is tightly stacked between Trp282 (from motif **IIIa**) and Ile347 (from motif **IV**), which respectively play analogous roles to the corresponding Phe and Met sidechains from hLig1 and Chlorella virus ligase (Figure 4A). Specificity for the adenine-nucleotide is conferred by hydrogen-bonds made by Gln157 (corresponds to Glu566 in hLig1, Thr25 in Chlorella virus ligase) and the backbone carbonyl group of Leu158 to Ade N6, Lys349 (Lys in hLig1, Arg in Chlorella virus) to Ade N1, and the amide group of Ala160 to Ade

N6 (Figure 4A). The 2′-hydroxyl group of AMP in *syn* conformation is hydrogen-bonded to Glu217 (from motif **III**); a conserved interaction that explains the strong preference for rATP (over 2′-deoxy ATP) during the ligation reaction. The 3′-hydroxyl group of AMP interacts with the side chain of Arg164 (motif **I**), which is positioned in place by a network of hydrogen-bonds involving Asp215 (motif **III**) and the main chain carbonyl of Met136. The main chain amide group of Arg164 interacts with the phosphate group of the

3′-terminal nucleotide, allowing this residue to span across the nick (Figure 4A, E).

The AMP phosphate group in the active site of T4 DNA ligase is within hydrogen-bonding distance from three lysine residues; Lys159 (motif **I**), Lys365, and Lys367 (motif **V**). Lys159 is the catalytic residue that is adenylated in the ligase-AMP intermediate, while the other two lysines may form an oxyanion hole to stabilize the negatively charged intermediates of the adenylyl transfer reactions. The adjacent DNA 5′-phosphate interacts with Arg182, which in turn is stabilized by hydrogen-bonding to Gln135 (Figure 4E). These interactions together align the leaving AMP moiety for the nucleophilic attack during the nick-sealing reaction. Glu344 (motif **IV**) interacts closely with Lys159 to form an ion pair. Glu344, Glu217 and intervening Asp161 (motif **I**) are likely to coordinate metal ions essential for activating the catalytic Lys159 residue (61) and stabilizing the negatively charged intermediate of the nick sealing reaction. We observed electron density for several water molecules or putative magnesium ions surrounded by these residues (Figure 4A). Of these, one that interacts with Glu217 and is positioned close to the 5′ phosphate and the apposed 3′ DNA end at the nick was modeled as magnesium.

DNA contacts by NTase and OB domains

The OB domain is a barrel of 7 β-strands, capped on either end by two short α-helices. The β-barrel with an oblong shape fits in the minor groove of DNA, making extensive contacts with backbones of both strands as well as inserting hydrophobic residues into the groove (Figure 5). The DNA-binding surface is extended by two symmetrically positioned loops that project out from opposite ends of the barrel and track along the DNA backbones, which include Ser447, Asp448, and Arg450 interacting with the template strand upstream of the nick (Figure 5A). Like DBD, the OB domain shows a ∼2-fold symmetrical overall fold, mirroring the dyad of the bound DNA molecule. The OB domain and α-helix 11 (residues 254–266) from the NTase domain form a continuous surface that inserts into the widened DNA minor groove on both sides of the nick (Figure 5B). On the upstream of the nick, the NTase domain residues N262, K266, and N258 from α-helix 11 form hydrogen bonds with the deoxyribose O4′ atoms of nucleotides from both DNA strands (Figure 4B). In particular, Asn258 interacts with the 3′-terminal nucleotide (C11) by the nick, possibly helping to position it for the nick-sealing reaction (Figure 4C). Arg254 at the N-terminal end of this helix stacks on Phe457 from OB domain, where both these residues make van der Waals contacts with the deoxyribose and base moieties of the 3′-terminal nucleotide C11, respectively. This cation-π stacking is a unique mode of interaction to bridge between the DNA-binding surfaces of the NTase and OB domains in a DNA ligase. Arg254 is positioned in the DNA major groove similarly to Arg200 of *E. coli* LigA (26), where it also makes a water-mediated hydrogen bond to the N3 atom of A33 on the template strand.

On the downstream side of the nick, the OB domain residue Leu458 makes a van der Waals contact with the deoxyribose moiety of the 5′-terminal nucleotide G12 (Figure 4C). All eukaryotic and archaeal DNA ligases structurally characterized to date have phenylalanine at this position while *E. coli* ligase A has valine, thus varying in size of the corresponding hydrophobic side chain. Ile460 makes a van der Waals contact with the deoxyribose moiety of the second nucleotide (T13) from the 5′-terminus, extending the hydrophobic patch including Phe457 and Leu458. On the template strand across from the nick, the OB domain residues Gly406 and Pro459 pack against the deoxyribose moieties in the C3′-endo conformation of C31 and G32, respectively (Figure 4C, Supplementary Figure S3B). In addition to these interactions in the minor groove, Ser407 donates a hydrogen bond to the O2 position of the T13 base or to the O4′ atom of the third nucleotide (C14) (Figure 5C, Supplementary Figure S3B). The former interaction is the only direct base contact made by T4 DNA ligase, as noted for the corresponding Thr249 of Chlorella virus ligase that accepts a hydrogen bond from the N3 atom of an adenine base. The Ser side chain at this position likely allows interactions with any base, consistent with sequence non-specific nature of DNA ligation reaction. The OB domain interaction with the template strand backbone spans up to Lys384, which connects with the DBD residue Asp112 (Figure 5A). Asp112 is also engaged in a water-mediated hydrogen-bond network involving Tyr39 and a DNA phosphate, allowing for uninterrupted backbone contacts by the two domains.

Gp45-binding loop

T4 DNA ligase has a stretch of 26 amino acids (222–247) within the NTase domain, for which electron density is very weak presumably due to high conformational flexibility. The two ends of this stretch are positioned close to each other (His221 and Ala248 Cα atoms are 5.8 Å apart), rendering this region looped out (Figure 5B). An internally truncated construct T4LΔ(222–247)GS, which has a di-amino acid Gly-Ser substituted for residues 222–247, shows comparable or slightly better *in vitro* DNA ligation activity than the wild-type protein (Supplementary Figures S8, S10). Thus, the residues in this loop are dispensable for the enzymatic activity of T4 DNA ligase. Moreover, the crystal structure of T4LΔ(222–247)GS bound to DNA shows essentially an identical structure to that of the wild-type protein (data not shown), confirming that the looped-out residues are not integral part the structure of T4 DNA ligase. Of note, T4LΔ(222–247)GS-DNA complex crystals were more readily reproducible than the original T4 DNA ligase–DNA complex crystals, suggesting that the severe crystal reproducibility problem for the full-length protein was at least in part caused by this flexible loop within the NTase domain.

The flexible loop comprising residues 222–247 is centered on a stretch of hydrophobic amino acids ([231]**EPEGLDFLFD**[236]), which shows significant sequence similarity to the very C-terminal region of the T4 DNA polymerase gp43 ([889]**EKASLDFLFG**[898], Table 2). The C-terminal region of gp43 is known to mediate the interaction with its processivity factor gp45 sliding clamp (62), a trimeric protein ring that tethers DNA polymerase to DNA analogously to the eukaryotic and archaeal PCNA (35). Therefore, we hypothesized that the flexible loop of T4 DNA ligase binds gp45. Mixing purified gp45 trimer
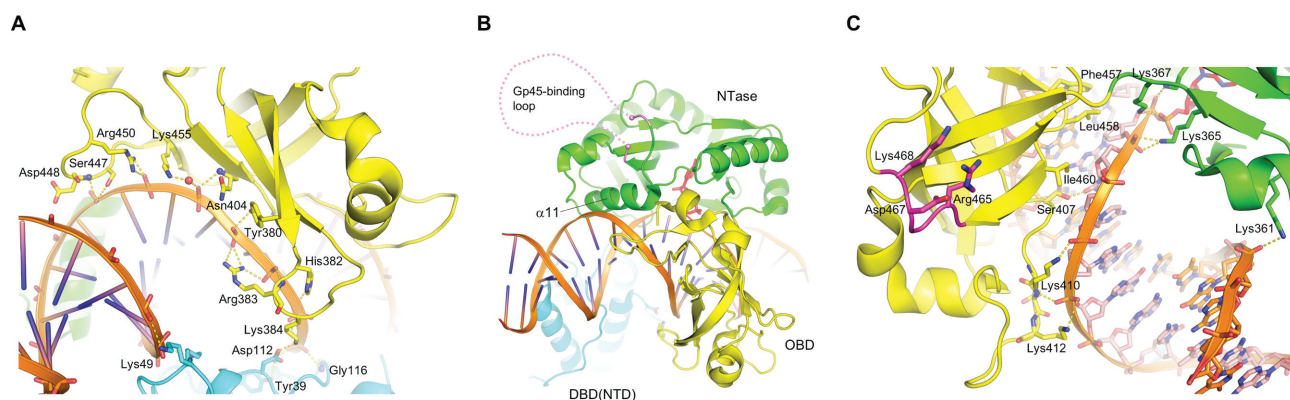
Figure 5. OBD-DNA contacts. (**A**) Backbone contacts by OB-fold domain with the template (unnicked) DNA strand. A water molecule mediating protein-DNA interaction is shown by a small red sphere. (**B**) An overview of the engagement of the DNA minor groove by NTase-OBD ligase core. The disordered gp45-binding loop (residues 222–247) is shown by a pink dotted line, with the last ordered residues on either side of the loop highlighted by small pink spheres. (**C**) Contacts made by the OB-fold domain with the nicked DNA strand and base-pairs in the minor groove. Motif **VI** residues are shown in magenta.

with the wild-type T4 DNA ligase in the absence of DNA generated a species larger than either protein alone in size-exclusion chromatography (SEC), demonstrating that these two proteins indeed form a stable complex (Figure 6A). In contrast, T4LΔ(222–247)GS did not show the co-migration with gp45 in SEC. Thus, the flexible loop is essential for the ligase–clamp interaction. We have further found using isothermal titration calorimetry (ITC) that a synthetic peptide (**KKEPEGLDFLF**DA) corresponding to the residues 229 to 237 of T4 DNA ligase binds to gp45 with a K_d of 10.7 ± 3.2 µM, showing that the central part of the loop including the hydrophobic residues is sufficient for the interaction. The titration curve in the ITC experiment suggested that each subunit within the gp45 trimer can independently bind the ligase-derived peptide with no apparent cooperativity (Figure 6B).

Ligase - Sliding clamp interaction

To gain further insights into the interaction between T4 DNA ligase and the gp45 sliding clamp, we determined a crystal structure of the T4 gp45 trimer in complex with the ligase-derived peptide at 2.1 Å resolution (Table 1, Supplementary Figure S2C). The structure shows that each of the three subunits within the trimeric gp45 ring is bound by a ligase peptide, confirming the binding stoichiometry observed in the ITC experiment. The conformation of the gp45 trimer in this complex shows little change from that in its free state (35). As observed in the crystal structure of a C-terminal peptide from bacteriophage RB69 DNA polymerase bound to the gp45 clamp (63) and those of the eukaryotic PIP-box peptides bound to PCNA (64), the T4 DNA ligase peptide binds to a hydrophobic pocket on the β-sheet surface of gp45 facing outside of the protein ring. The ligase peptide forms an amphipathic α-helix and fits into the pocket with good shape complementarity (Figure 7A, C). Within the pocket, Leu231, Phe233, Leu234 and Phe235 side chains make hydrophobic contacts with the gp45 surface. The peptide conformation for the residues interacting with gp45 is essentially the same for all three

molecules in the crystal (Figure 7B). The canonical α-helical conformation with regular hydrogen bonds observed for the T4 DNA ligase peptide is unique and distinct from the highly conserved PIP-box conformation containing an anchoring 3_{10} helix, which was observed for eukaryotic and archaeal PCNA-interacting proteins (64–67) as well as the bacteriophage RB69 polymerase bound on their cognate sliding clamp surfaces (Figure 7D) (63).

Whereas the small T4 DNA ligase-derived peptide can occupy all of the three equivalent binding sites on a gp45 trimer as shown by ITC and x-ray crystallography, only single molecule of the full-length T4 DNA ligase can bind each trimeric gp45 ring. In the SEC analyses, addition of an excess of ligase beyond the molar ratio of 1:3 between the ligase and gp45 monomers did not lead to formation of a larger complex but led to accumulation of free ligase (Supplementary Figure S9). This could be due to the binding of additional ligase molecules being occluded by the first ligase molecule bound to the gp45 ring, suggesting a relatively tight spacing between the two proteins. The observation is analogous to the mutually exclusive binding between Lig1 and Pol δ, or Lig1 and FEN1, to a PCNA ring (3). Eukaryotic and archaeal DNA ligases have been shown to interact with PCNA or the 9-1-1 DNA-repair clamp, and in some cases, the physical interaction stimulates nick-sealing activity of the ligase (10,68). It has been hypothesized that such stimulation results from a templating effect of the protein clamp to facilitate closure of the multi-domain ligase around a DNA substrate or alternatively other allosteric mechanisms (10). T4 gp45 did not stimulate the blunt-end ligation activity of T4 DNA ligase on a short linear DNA substrate (Supplementary Figure S8), suggesting that the clamp interaction does not directly modulate the ligase's catalytic activity. However, we observed modest stimulation of T4 DNA ligase sealing multiple nicks on a circular DNA by the presence of both gp45 and the T4 clamp-loader gp44/62 (Supplementary Figure S10). These results suggest that the T4 ligase–clamp interaction may help localize T4 DNA ligase to its substrates and thereby facilitate ligation reactions.
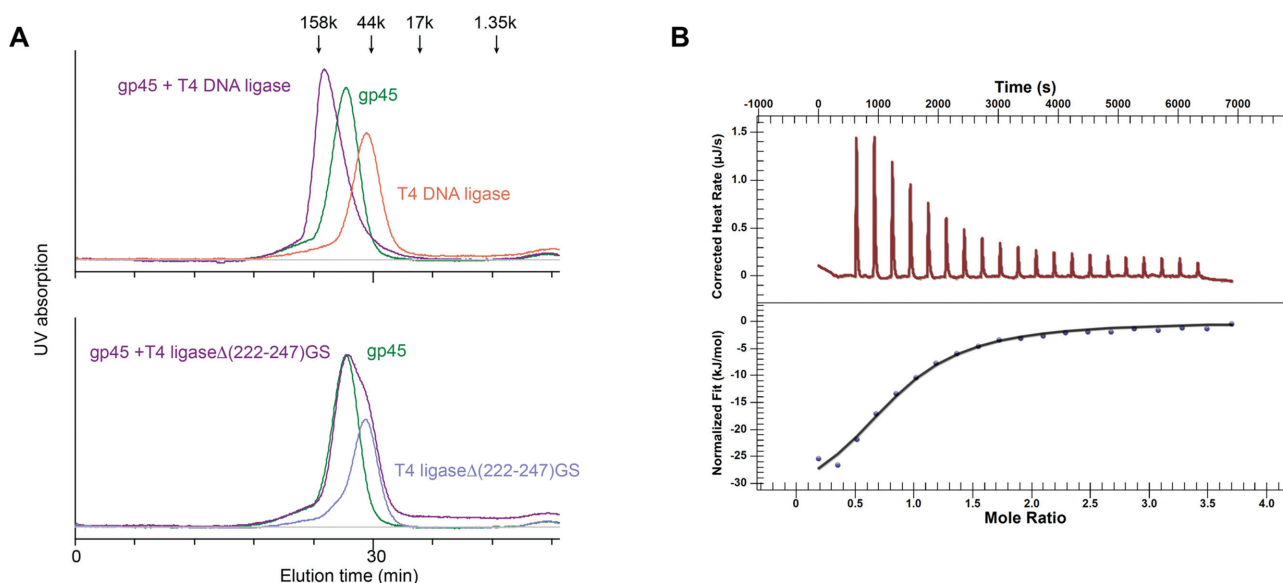
Figure 6. T4 DNA ligase-gp45 interaction. (**A**) Size-exclusion chromatography profiles showing complex formation between the wild-type T4 DNA ligase and gp45 (upper panel), but not between the internally truncated ligase lacking residues 222–247 and gp45 (lower panel). (**B**) A representative ITC experiment wherein the T4 ligase loop peptide (525 μM) was titrated into gp45 (51 μM). The data were fit to an independent binding model implying a lack of cooperativity between the binding sites of the gp45 trimer. Five replicates gave an average K_d of 10.69 μM with a standard deviation of 3.16 μM.

Table 2. Sliding clamp-binding motifs

| T4 DNA ligase (gp30)	^{227}E P E G **L** D **F** **L** **F** D^{236}
T4 DNA polymerase (gp43)	^{889}E K A S **L** D **F** **L** **F** G^{898}
T4 RNA polymerase-associated protein (gp33)	^{104}K T N T **L** D **F** **L** **L** ^{112}
T4 RNA polymerase σ-factor (gp55)	^{173}D S P S **L** D **F** **L** **Y** E^{182}
T4 flap endonuclease (rnh)	^{1}M D **L** E **M** **M** **L** D ^{8}
RB69 DNA polymerase (gp43)	^{893}K K A S **L** F D **M** **F** D^{902}
P21^{CIP1} Cyclin-dependent kinase inhibitor-1	^{143}R **Q** T S **M** T D **F** **Y** H^{152}
Human flap endonuclease FEN-1	^{336}T **Q** G R **L** D D **F** **F** K^{345}
Human DNA Ligase 1 (hLig1)	^{1}M **Q** R S **I** M S **F** **F** H^{10}
Human polymerase δ (p66)	^{455}R **Q** V S **I** T G **F** **F** Q^{464}

Alignment of the T4 gp45-binding motifs with that from the related RB69 phage and eukaryotic PIP-boxes.

DISCUSSION

The structure of T4 DNA ligase–DNA complex shows a conserved general architecture of ligase–DNA complexes, in which the multiple ligase domains together wrap around a nicked DNA substrate. The complete envelopment of DNA by T4 DNA ligase is facilitated by the N-terminal α-helical DBD that closely resembles the core DNA-binding components of the much larger DBDs of mammalian and archaebacterial ATP-dependent DNA ligases. This is in contrast to the mode of DNA envelopment by another viral DNA ligase from Chlorella virus, which employs a unique β-hairpin insertion within the OB domain. Interestingly, the NTase-OB core of T4 DNA ligase is phylogenetically more closely related to, and accordingly shows higher structural similarity to, that of Chlorella virus ligase than any of the other ATP-dependent ligases characterized to date. Thus, these smaller viral DNA ligases are minimal prototypical DNA ligases with alternative modes of DNA encircling,

and T4 DNA ligase may represent a progenitor of larger cellular ATP-dependent DNA ligases containing the α-helical DBD (Supplementary Figure S1). The larger DBDs of cellular DNA ligases have extra structural elements compared to the minimal DBD of T4 DNA ligase, which likely have evolved to facilitate interactions with other protein factors including PCNA. In addition, the eukaryotic DNA ligases have acquired additional domains that mediate their specialized functions in distinct biological contexts, such as the BRCT and ZnF domains (3).

We have shown that T4 DNA ligase forms a stable complex with T4 sliding clamp gp45, the processivity factor for DNA polymerase. The interaction is analogous to that between the eukaryotic and archaeal replicative DNA ligases with PCNA, which is thought to facilitate joining of Okazaki fragments during the discontinuous lagging strand synthesis. However, the interaction between T4 DNA ligase and gp45 is mediated by a sequence motif in the middle of a loop inserted within the NTase domain, rather than
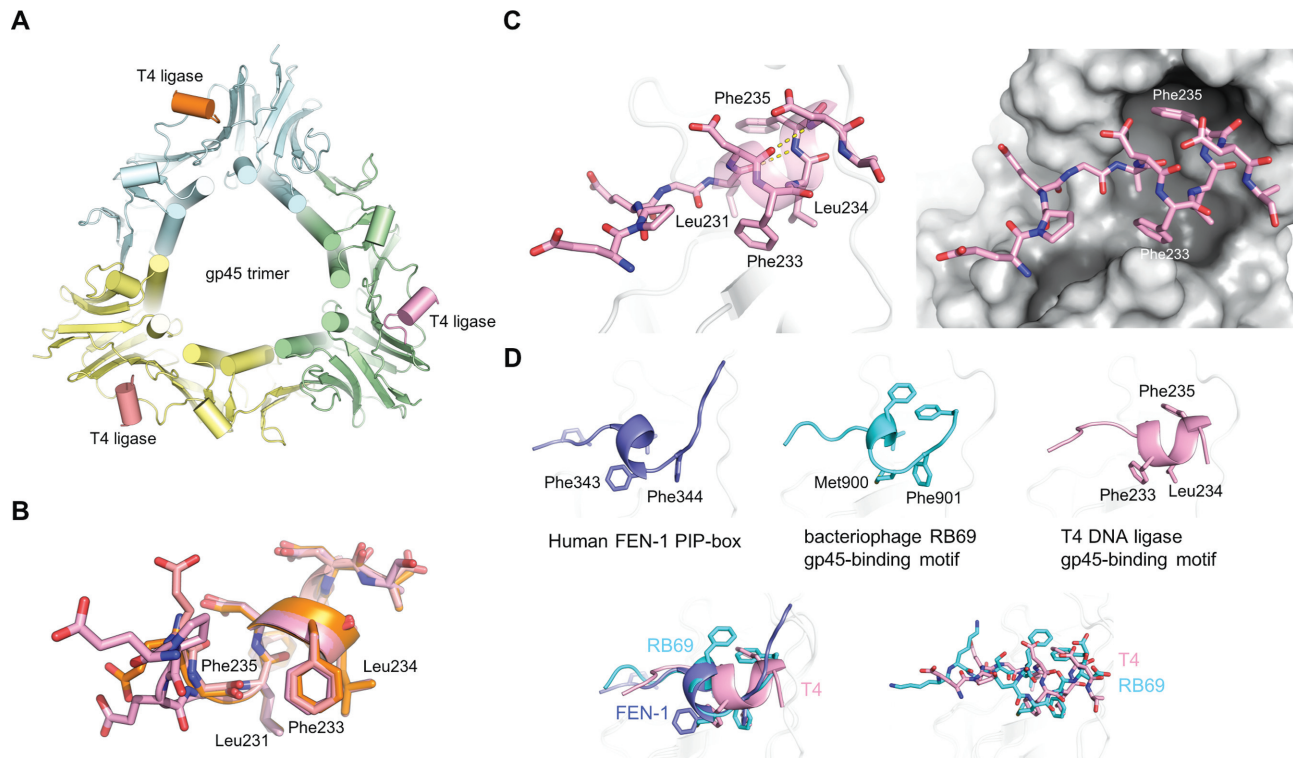
Figure 7. Structure of the gp45-binding loop of T4 DNA ligase. (**A**) A gp45 trimer bound by three copies of the T4 ligase peptide. (**B**) Superposition of the three ligase-derived peptides showing their similar α-helical conformation and positioning of hydrophobic side chains. (**C**) Close-up views showing the α-helical conformation of the T4 ligase peptide with canonical hydrogen-bonds highlighted (left), and fitting within a hydrophobic pocket on the gp45 surface (right). (**D**) Human FEN1 PIP-box motif bound to PCNA (slate) (64), phage RB69 polymerase bound to gp45 (cyan) (63), and T4 ligase bound to gp45 (pink). hFEN-1 and RB69 have similar conformations with a 3_{10} helix, whereas T4 ligase forms an α-helix with three consecutive residues fit in the pocket on gp45. The superpositions in the lower panels highlight the distinct α-helical conformation of the T4 ligase peptide.

the structural elements within DBD or a PIP-box motif outside the three essential ligase domains as observed in the mammalian/archaeal DNA ligases (10,11,60). A corresponding insertion is not found in the NTase domain of any DNA/RNA ligase or related enzymes, suggesting that T4 DNA ligase acquired this distinct mode of sliding clamp interaction through independent evolution. Like the cellular genome replication, the T4 phage replication requires coordinated syntheses of leading and lagging DNA strands at the replication fork (37), and T4 DNA ligase (gp30) plays an essential role in the phage DNA synthesis (38–43). The DNA ligase-gp45 interaction may therefore be key for efficient viral genome replication. In addition to the DNA polymerase and ligase, T4 RNaseH (rnh), which removes RNA primers from the 5′ end of Okazaki fragments, has also been proposed to interact with gp45 (36). This would allow the activities of three essential enzymes required for the lagging strand synthesis to be coordinated through interactions with gp45, similar to the interactions of Pol δ, Lig1, and FEN-1 with PCNA important in eukaryotic DNA replication.

The T4 gp45-interaction motif (**LDFLF**) adopts a novel α-helical conformation distinct from the canonical conformation including a 3_{10} helix observed for the sliding clamp-binding motifs (PIP-box) from eukaryotes, archaea, and bacteriophages (65–67), including the DNA polymerase C-terminal peptide from a related bacteriophage RB69

(63). The α-helical conformation is also distinct from the extended conformation of β clamp-interacting peptides in bacteria (9,69,70). The three consecutive hydrophobic residues (Phe233, Leu234, Phe235) of T4 DNA ligase cluster on one side of an amphipathic α-helix (Supplementary Figure S2C) and insert into the hydrophobic pocket on the gp45 surface, which is probably a shared feature among the T4 gp45-interacting proteins including the DNA polymerase gp43 and an RNA polymerase-associated protein gp33 based on their amino acid sequences (Table 2). However, despite the distinct backbone conformations, the gp45-interacting motifs from RB69 and T4 have spatially similar positioning of hydrophobic sidechains pointing toward the gp45 surface, possibly explaining why the gp43 polymerase from these two bacteriophages can partially substitute for each other in phage DNA replication (71).

The physical association with a sliding clamp would confer processivity to DNA ligase, which is inherently a nonprocessive enzyme. For every ligation reaction, DNA ligase needs to release DNA and take a distinct conformation incompatible with DNA-binding, where the 'motif VI' within the OB domain is presented to the active site to catalyze a lysine adenylation reaction (16,24). As observed for other DNA ligases (19,22,25,26), the motif VI residues of T4 DNA ligase including Arg465, Asp467, and Lys468, are positioned opposite the DNA-binding surface of the OB domain, facing outward in the ligase–DNA complex
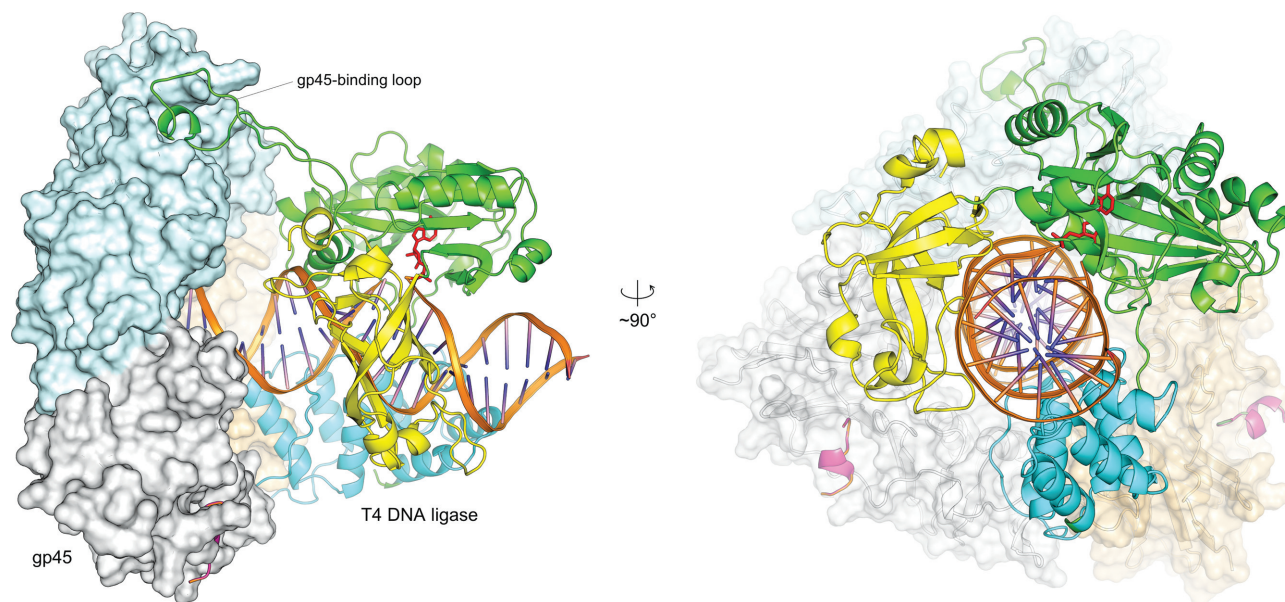
Figure 8. A hypothetical model of the T4 ligase-gp45-DNA ternary complex. Molecular surface of a gp45 ring and ribbon model of T4 DNA ligase are shown. The color scheme for T4 DNA ligase follows that in Figure 1. The peptides occupying two remaining binding sites on the gp45 trimer are shown in pink. The modeling exercise suggests that the DBD/NTD might potentially interact with an inter-subunit cleft on the gp45 trimer.

(Figure 5C). With the NTase domain flexibly tethered to a gp45 ring that is threaded on a double-stranded DNA, T4 DNA ligase would be able to go through multiple cycles of nick-sealing reactions involving large movement of the OB domain, without falling off from DNA. Modeling suggests that T4 DNA ligase could engage a nick in a double stranded DNA molecule while being tethered to the adjacent gp45 ring loaded on the upstream side of the nick (Figure 8). *Sulfolobus solfataricus* DNA ligase has been shown to interact with PCNA through a flexible loop within DBD, which allows the ligase to take a remarkably extended conformation while bound to PCNA (10). The gp45-bound T4 DNA ligase may similarly sample a range of conformations for efficient DNA ligation or coordination of activities between the multiple enzymes that associate with gp45.

Despite the high overall structural similarity to other DNA ligases in the NTase-OB ligase catalytic core, T4 DNA ligase shows distinct features in DNA interactions, including unique sets of hydrogen-bonding and van der Waals contacts made in the minor groove. The deoxyribose moiety of the 5′-terminal nucleotide adjacent to the nick is loosely contacted by Leu458 of T4 DNA ligase, which corresponds to Val383 of *E. coli* LigA (26), Phe286 of Chlorella virus ligase (25), and Phe872 of hLig1 (19) (Supplementary Figure S11). The aromatic side chain at this position is conserved across all eukaryotic and archaeal DNA ligases, and Phe872 of hLig1, Phe717 of hLig3 (22) and Phe578 of hLig4 (72) pack more tightly against the deoxyribose ring than Leu458 of T4 DNA ligase. By contrast, the smaller valine side chain in *E. coli* LigA leaves a larger gap between the van der Waals radii at this interface. Interestingly, despite the universally conserved A-form-like DNA conformation including the C3′-endo sugar pucker for nucleotides around the nick in various ligase–DNA complexes (19,22,25,26), these DNA ligases show varied activities on ribonucleotide-containing

DNA substrates. The eukaryotic DNA ligases do not efficiently ligate nicked substrates containing even a single 5′-terminal ribonucleotide adjacent to the nick (73). T4 DNA ligase shows better ligation activity on the same substrate, although a fraction of the adenylated intermediate fail to complete the nick-sealing reaction. *E. coli* LigA ligates this substrate as well as it does a regular DNA substrate (73). Thus, the varied activities seem to inversely correlate with the extent of the interaction between the deoxyribose moiety and the abutting hydrophobic side chain, which suggests that a steric clash of the 2′-OH group is the basis for the selectivity against 5′-terminal ribonucleotide. The different tolerance on 5′-terminal ribonucleotides by eukaryotic and prokaryotic DNA ligases may reflect difference in the repair mechanisms of mis-incorporated ribonucleotides in respective organisms (74).

In summary, the structure of the T4 DNA ligase–DNA complex highlights interesting evolutionary relationships to eukaryotic DNA ligases, including the common DNA-binding mechanism conserved through divergent evolution, and distinct modes of sliding clamp interactions likely independently acquired through convergent evolution. The structure may also provide bases for rationally designing engineered ligases fit for applications, such as point mutations or insertion/fusion of foreign DNA-binding elements to augment affinity toward specific type of DNA substrates (75), or replacement of the gp45-interaction motif with peptide elements capable of binding to heterologous proteins for efficient recruitment of T4 DNA ligase to the sites of DNA strand breaks.

DATA AVAILABILITY

The atomic coordinates and structure factors have been deposited in the Protein Data Bank under the accession codes

5WFY (SeMet DBD), 6DT1 (T4 DNA ligase–DNA complex), and 6DRT (gp45 complexed with ligase-derived peptide).

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

FUNDING

REFERENCES

1. Lehman,I.R. (1974) DNA ligase: structure, mechanism, and function. *Science*, **186**, 790–797.
2. Shuman,S. (2009) DNA ligases: progress and prospects. *J. Biol. Chem.*, **284**, 17365–17369.
3. Tomkinson,A.E., Vijayakumar,S., Pascal,J.M. and Ellenberger,T. (2006) DNA ligases: structure, reaction mechanism, and function. *Chem. Rev.*, **106**, 687–699.
4. Lindahl,T. and Barnes,D.E. (1992) Mammalian DNA ligases. *Annu. Rev.Biochem.*, **61**, 251–281.
5. Wilkinson,A., Day,J. and Bowater,R. (2001) Bacterial DNA ligases. *Mol. Microbiol.*, **40**, 1241–1248.
6. Doherty,A.J. and Suh,S.W. (2000) Structural and mechanistic conservation in DNA ligases. *Nucleic Acids Res.*, **28**, 4051–4058.
7. Kiyonari,S., Takayama,K., Nishida,H. and Ishino,Y. (2006) Identification of a novel binding motif in Pyrococcus furiosus DNA ligase for the functional interaction with proliferating cell nuclear antigen. *J. Biol. Chem.*, **281**, 28023–28032.
8. Levin,D.S., Bai,W., Yao,N., O'Donnell,M. and Tomkinson,A.E. (1997) An interaction between DNA ligase I and proliferating cell nuclear antigen: implications for Okazaki fragment synthesis and joining. *Proc. Natl. Acad. Sci. U.S.A.*, **94**, 12863–12868.
9. Pandey,P., Tarique,K.F., Mazumder,M., Rehman,S.A., Kumari,N. and Gourinath,S. (2016) Structural insight into beta-Clamp and its interaction with DNA Ligase in Helicobacter pylori. *Sci. Rep.*, **6**, 31181.
10. Pascal,J.M., Tsodikov,O.V., Hura,G.L., Song,W., Cotner,E.A., Classen,S., Tomkinson,A.E., Tainer,J.A. and Ellenberger,T. (2006) A flexible interface between DNA ligase and PCNA supports conformational switching and efficient ligation of DNA. *Mol. Cell*, **24**, 279–291.
11. Levin,D.S., McKenna,A.E., Motycka,T.A., Matsumoto,Y. and Tomkinson,A.E. (2000) Interaction between PCNA and DNA ligase I is critical for joining of Okazaki fragments and long-patch base-excision repair. *Curr. Biol.*, **10**, 919–922.
12. Tom,S., Henricksen,L.A., Park,M.S. and Bambara,R.A. (2001) DNA ligase I and proliferating cell nuclear antigen form a functional complex. *J. Biol. Chem.*, **276**, 24817–24825.
13. Montecucco,A., Rossi,R., Levin,D.S., Gary,R., Park,M.S., Motycka,T.A., Ciarrocchi,G., Villa,A., Biamonti,G. and Tomkinson,A.E. (1998) DNA ligase I is recruited to sites of DNA replication by an interaction with proliferating cell nuclear antigen: identification of a common targeting mechanism for the assembly of replication factories. *EMBO J.*, **17**, 3786–3795.
14. Dionne,I., Nookala,R.K., Jackson,S.P., Doherty,A.J. and Bell,S.D. (2003) A heterotrimeric PCNA in the hyperthermophilic archaeon Sulfolobus solfataricus. *Mol. Cell*, **11**, 275–282.
15. Pascal,J.M. (2008) DNA and RNA ligases: structural variations and shared mechanisms. *Curr. Opin. Struct. Biol.*, **18**, 96–105.
16. Sriskanda,V. and Shuman,S. (1998) Mutational analysis of Chlorella virus DNA ligase: catalytic roles of domain I and motif VI. *Nucleic Acids Res.*, **26**, 4618–4625.
17. Gajiwala,K.S. and Pinko,C. (2004) Structural rearrangement accompanying NAD+ synthesis within a bacterial DNA ligase crystal. *Structure*, **12**, 1449–1459.
18. Sriskanda,V. and Shuman,S. (2002) Conserved residues in domain Ia are required for the reaction of Escherichia coli DNA ligase with NAD+. *J. Biol. Chem.*, **277**, 9695–9700.
19. Pascal,J.M., O'Brien,P.J., Tomkinson,A.E. and Ellenberger,T. (2004) Human DNA ligase I completely encircles and partially unwinds nicked DNA. *Nature*, **432**, 473–478.
20. Subramanya,H.S., Doherty,A.J., Ashford,S.R. and Wigley,D.B. (1996) Crystal structure of an ATP-dependent DNA ligase from bacteriophage T7. *Cell*, **85**, 607–615.
21. Shuman,S. and Lima,C.D. (2004) The polynucleotide ligase and RNA capping enzyme superfamily of covalent nucleotidyltransferases. *Curr. Opin. Struct. Biol.*, **14**, 757–764.
22. Cotner-Gohara,E., Kim,I.K., Hammel,M., Tainer,J.A., Tomkinson,A.E. and Ellenberger,T. (2010) Human DNA ligase III recognizes DNA ends by dynamic switching between two DNA-bound states. *Biochemistry*, **49**, 6165–6176.
23. Ochi,T., Gu,X. and Blundell,T.L. (2013) Structure of the catalytic region of DNA ligase IV in complex with an Artemis fragment sheds light on double-strand break repair. *Structure*, **21**, 672–679.
24. Nishida,H., Kiyonari,S., Ishino,Y. and Morikawa,K. (2006) The closed structure of an archaeal DNA ligase from Pyrococcus furiosus. *J. Mol. Biol.*, **360**, 956–967.
25. Nair,P.A., Nandakumar,J., Smith,P., Odell,M., Lima,C.D. and Shuman,S. (2007) Structural basis for nick recognition by a minimal pluripotent DNA ligase. *Nat. Struct. Mol. Biol.*, **14**, 770–778.
26. Nandakumar,J., Nair,P.A. and Shuman,S. (2007) Last stop on the road to repair: structure of E. coli DNA ligase bound to nicked DNA-adenylate. *Mol. Cell*, **26**, 257–271.
27. Becker,A., Lyn,G., Gefter,M. and Hurwitz,J. (1967) The enzymatic repair of DNA, II. Characterization of phage-induced sealase. *Proc. Natl. Acad. Sci. U.S.A.*, **58**, 1996–2003.
28. Cozzarelli,N.R., Melechen,N.E., Jovin,T.M. and Kornberg,A. (1967) Polynucleotide cellulose as a substrate for a polynucleotide ligase induced by phage T4. *Biochem. Biophys. Res. Commun.*, **28**, 578–586.
29. Weiss,B. and Richardson,C.C. (1967) Enzymatic breakage and joining of deoxyribonucleic acid, I. Repair of single-strand breaks in DNA by an enzyme system from Escherichia coli infected with T4 bacteriophage. *Proc. Natl. Acad. Sci. U.S.A.*, **57**, 1021–1028.
30. Reuter,J.A., Spacek,D.V. and Snyder,M.P. (2015) High-throughput sequencing technologies. *Mol. Cell*, **58**, 586–597.
31. Schmidt,D., Wilson,M.D., Spyrou,C., Brown,G.D., Hadfield,J. and Odom,D.T. (2009) ChIP-seq: using high-throughput sequencing to discover protein-DNA interactions. *Methods*, **48**, 240–248.
32. Son,M.S. and Taylor,R.K. (2011) Preparing DNA libraries for multiplexed paired-end deep sequencing for Illumina GA sequencers. *Curr. Protoc. Microbiol.*, doi:10.1002/9780471729259.mc01e04s20.
33. Maroney,P.A., Chamnongpol,S., Souret,F. and Nilsen,T.W. (2008) Direct detection of small RNAs using splinted ligation. *Nat. Protoc.*, **3**, 279–287.

34. Rossi,R., Montecucco,A., Ciarrocchi,G. and Biamonti,G. (1997) Functional characterization of the T4 DNA ligase: a new insight into the mechanism of action. *Nucleic Acids Res.*, **25**, 2106–2113.

35. Moarefi,I., Jeruzalmi,D., Turner,J., O'Donnell,M. and Kuriyan,J. (2000) Crystal structure of the DNA polymerase processivity factor of T4 bacteriophage. *J. Mol. Biol.*, **296**, 1215–1223.

36. Mueser,T.C., Hinerman,J.M., Devos,J.M., Boyer,R.A. and Williams,K.J. (2010) Structural analysis of bacteriophage T4 DNA replication: a review in the Virology Journal series on bacteriophage T4 and its relatives. *Virol. J.*, **7**, 359.

37. Noble,E., Spiering,M.M. and Benkovic,S.J. (2015) Coordinated DNA replication by the bacteriophage T4 replisome. *Viruses*, **7**, 3186–3200.

38. Bolle,A., Epstein,R.H., Salser,W. and Geiduschek,E.P. (1968) Transcription during bacteriophage T4 development: requirements for late messenger synthesis. *J. Mol. Biol.*, **33**, 339–362.

39. Hosoda,J. (1967) A mutant of bacteriophage T4 defective in alpha-glucosyl transferase. *Biochem. Biophys. Res. Commun.*, **27**, 294–298.

40. Hosoda,J. and Mathews,E. (1968) DNA replication in vivo by a temperature-sensitive polynucleotide ligase mutant of T4. *Proc. Natl. Acad. Sci. U.S.A.*, **61**, 997–1004.

41. Masamune,Y. and Richardson,C.C. (1968) Enzymatic breakage and joining of deoxyribonucleic acid. IV. DNA synthesis in E. coli infected with ligase-negative mutants of phage T4. *Proc. Natl. Acad. Sci. U.S.A.*, **61**, 1328–1335.

42. Sugimoto,K., Okazaki,T. and Okazaki,R. (1968) Mechanism of DNA chain growth, II. Accumulation of newly synthesized short chains in E. coli infected with ligase-defective T4 phages. *Proc. Natl. Acad. Sci. U.S.A.*, **60**, 1356–1362.

43. Warner,H.R. and Hobbs,M.D. (1967) Incorporation of uracil-14C into nucleic acids in Escherichia coli infected with bacteriophage T4 and T4 amber mutants. *Virology*, **33**, 376–384.

44. Liu,J. and Morrical,S.W. (2010) Assembly and dynamics of the bacteriophage T4 homologous recombination machinery. *Virol. J.*, **7**, 357.

45. Kelch,B.A., Makino,D.L., O'Donnell,M. and Kuriyan,J. (2011) How a DNA polymerase clamp loader opens a sliding clamp. *Science*, **334**, 1675–1680.

46. Otwinowski,Z. and Minor,W. (1997) Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.*, **276**, 20.

47. Kabsch,W. (2010) XDS. *Acta Crystallogr. D, Biol. Crystallogr.*, **66**, 125–132.

48. Sheldrick,G.M. (2008) A short history of SHELX. *Acta Crystallogr. A, Found. Crystallogr.*, **64**, 112–122.

49. McCoy,A.J., Grosse-Kunstleve,R.W., Adams,P.D., Winn,M.D., Storoni,L.C. and Read,R.J. (2007) Phaser crystallographic software. *J. Appl. Crystallogr.*, **40**, 658–674.

50. Zhang,K.Y., Cowtan,K. and Main,P. (1997) Combining constraints for electron-density modification. *Methods Enzymol.*, **277**, 53–64.

51. Cowtan,K. (2006) The Buccaneer software for automated model building. 1. Tracing protein chains. *Acta Crystallogr. D, Biol. Crystallogr.*, **62**, 1002–1011.

52. Murshudov,G.N., Vagin,A.A. and Dodson,E.J. (1997) Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. D, Biol. Crystallogr.*, **53**, 240–255.

53. Adams,P.D., Afonine,P.V., Bunkoczi,G., Chen,V.B., Davis,I.W., Echols,N., Headd,J.J., Hung,L.W., Kapral,G.J., Grosse-Kunstleve,R.W. *et al.* (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D, Biol. Crystallogr.*, **66**, 213–221.

54. Emsley,P., Lohkamp,B., Scott,W.G. and Cowtan,K. (2010) Features and development of Coot. *Acta Crystallogr. D, Biol. Crystallogr.*, **66**, 486–501.

55. Dolinsky,T.J., Nielsen,J.E., McCammon,J.A. and Baker,N.A. (2004) PDB2PQR: an automated pipeline for the setup of Poisson-Boltzmann electrostatics calculations. *Nucleic Acids Res.*, **32**, W665–W667.

56. Baker,N.A., Sept,D., Joseph,S., Holst,M.J. and McCammon,J.A. (2001) Electrostatics of nanosystems: application to microtubules and the ribosome. *Proc. Natl. Acad. Sci. U.S.A.*, **98**, 10037–10041.

57. Lavery,R., Moakher,M., Maddocks,J.H., Petkeviciute,D. and Zakrzewska,K. (2009) Conformational analysis of nucleic acids revisited: Curves+. *Nucleic Acids Res.*, **37**, 5917–5929.

58. Lu,X.J. and Olson,W.K. (2008) 3DNA: a versatile, integrated software system for the analysis, rebuilding and visualization of three-dimensional nucleic-acid structures. *Nat. Protoc.*, **3**, 1213–1227.

59. Lee,J.Y., Chang,C., Song,H.K., Moon,J., Yang,J.K., Kim,H.K., Kwon,S.T. and Suh,S.W. (2000) Crystal structure of NAD(+)-dependent DNA ligase: modular architecture and functional implications. *EMBO J.*, **19**, 1119–1129.

60. Song,W., Pascal,J.M., Ellenberger,T. and Tomkinson,A.E. (2009) The DNA binding domain of human DNA ligase I interacts with both nicked DNA and the DNA sliding clamps, PCNA and hRad9-hRad1-hHus1. *DNA Repair (Amst.)*, **8**, 912–919.

61. Unciuleac,M.C., Goldgur,Y. and Shuman,S. (2017) Two-metal versus one-metal mechanisms of lysine adenylylation by ATP-dependent and NAD+-dependent polynucleotide ligases. *Proc. Natl. Acad. Sci. U.S.A.*, **114**, 2592–2597.

62. Berdis,A.J., Soumillion,P. and Benkovic,S.J. (1996) The carboxyl terminus of the bacteriophage T4 DNA polymerase is required for holoenzyme complex formation. *Proc. Natl. Acad. Sci. U.S.A.*, **93**, 12822–12827.

63. Shamoo,Y. and Steitz,T.A. (1999) Building a replisome from interacting pieces: sliding clamp complexed to a peptide from DNA polymerase and a polymerase editing complex. *Cell*, **99**, 155–166.

64. Bruning,J.B. and Shamoo,Y. (2004) Structural and thermodynamic analysis of human PCNA with peptides derived from DNA polymerase-delta p66 subunit and flap endonuclease-1. *Structure*, **12**, 2209–2219.

65. De Biasio,A. and Blanco,F.J. (2013) Proliferating cell nuclear antigen structure and interactions: too many partners for one dancer? *Adv. Protein Chem. Struct. Biol.*, **91**, 1–36.

66. Park,S.Y., Jeong,M.S., Han,C.W., Yu,H.S. and Jang,S.B. (2016) Structural and functional insight into proliferating cell nuclear antigen. *J. Microbiol. Biotechnol.*, **26**, 637–647.

67. Winter,J.A. and Bunting,K.A. (2012) Rings in the extreme: PCNA interactions and adaptations in the archaea. *Archaea*, **2012**, 951010.

68. Wang,W., Lindsey-Boltz,L.A., Sancar,A. and Bambara,R.A. (2006) Mechanism of stimulation of human DNA ligase I by the Rad9-rad1-Hus1 checkpoint complex. *J. Biol. Chem.*, **281**, 20865–20872.

69. Bunting,K.A., Roe,S.M. and Pearl,L.H. (2003) Structural basis for recruitment of translesion DNA polymerase Pol IV/DinB to the beta-clamp. *EMBO J.*, **22**, 5883–5892.

70. Patoli,A.A., Winter,J.A. and Bunting,K.A. (2013) The UmuC subunit of the E. coli DNA polymerase V shows a unique interaction with the beta-clamp processivity factor. *BMC Struct. Biol.*, **13**, 12.

71. Yeh,L.S., Hsu,T. and Karam,J.D. (1998) Divergence of a DNA replication gene cluster in the T4-related bacteriophage RB69. *J. Bacteriol.*, **180**, 2005–2013.

72. Kaminski,A.M., Tumbale,P.P., Schellenberg,M.J., Williams,R.S., Williams,J.G., Kunkel,T.A., Pedersen,L.C. and Bebenek,K. (2018) Structures of DNA-bound human ligase IV catalytic core reveal insights into substrate binding and catalysis. *Nat. Commun.*, **9**, 2642.

73. Tumbale,P., Williams,J.S., Schellenberg,M.J., Kunkel,T.A. and Williams,R.S. (2014) Aprataxin resolves adenylated RNA-DNA junctions to maintain genome integrity. *Nature*, **506**, 111–115.

74. Potenski,C.J. and Klein,H.L. (2014) How the misincorporation of ribonucleotides into genomic DNA can be both harmful and helpful to cells. *Nucleic Acids Res.*, **42**, 10226–10234.

75. Wilson,R.H., Morton,S.K., Deiderick,H., Gerth,M.L., Paul,H.A., Gerber,I., Patel,A., Ellington,A.D., Hunicke-Smith,S.P. and Patrick,W.M. (2013) Engineered DNA ligases with improved activities in vitro. *Protein Eng. Des. Sel.*, **26**, 471–478.