Structures of LIG1 uncover a lack of sugar discrimination against a ribonucleotide at the 3'-end of nick DNA

MELIKE CAGLAYAN (caglayanm@ufl.edu)
UNIVERSITY OF FLORIDA  https://orcid.org/0000-0003-1107-1042

Qun Tang
UNIVERSITY OF FLORIDA

Mitchell Gulkis
UNIVERSITY OF FLORIDA  https://orcid.org/0000-0001-5831-8802

Article

Keywords: Ribonucleotide, rNTP, DNA polymerase, DNA ligase I, DNA ligase IIIα/XRCC1 complex, mutagenic ligation, DNA repair, DNA replication, genome stability

Posted Date: September 27th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1998177/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License
Abstract

Ribonucleotides can be incorporated by DNA polymerases and the subsequent joining of 3'-OH and 5'-P ends in the phosphodiester backbone at the nick by DNA ligase during DNA replication and repair is critical for maintaining genome stability. Although it has been extensively studied for DNA polymerases across families, the sugar discrimination mechanism of a human DNA ligase at atomic resolution is entirely missing. Here, for the first time, we determine X-ray structure of DNA ligase I (LIG1) in complex with nick DNA containing rG:C at the 3'-end and capture the ligase at the final phosphodiester bond formation step of the ligation reaction involving an adenylate (AMP) release. Moreover, we show mutagenic end joining of the nick DNA substrate with preinserted 3'-rG:C by LIG1 in vitro. Our findings reveal an important source of ribonucleotides embedded in genomic DNA, which could come from the failure of LIG1 to discriminate against a ribonucleotide at the 3'-end during nick sealing step of DNA replication and repair.

Introduction

The concentrations of ribonucleotide triphosphates (rNTPs) are much more abundant than those of deoxyribonucleotide triphosphates (dNTPs) within a cell, and therefore, their misincorporation into genomic DNA by DNA polymerases occur at higher frequencies than mismatch nucleotide insertions (1–4). It has been estimated that > 10,000 or > 1,000,000 ribonucleotides can be incorporated into yeast and mouse genomes, respectively, making genomic ribonucleotides the most prevalent source of cellular DNA damage (5–13). Ribonucleotides embedded into genomic DNA can confound structural and chemical integrity of duplex DNA, increase its susceptibility to endogenous or exogenous damage, and lead to several types of genome instability such as replication blockage, mutagenesis, aberrant recombination, protein-DNA crosslinks, double-strand breaks (DSBs), and chromosome alterations if not repaired by ribonucleotide excision repair (RER) (14–20).

In human cells, ribonucleotide incorporation can occur during almost all cellular DNA transactions such as DNA replication by replicative polymerases (pols α, δ, and ε), non-homologous end joining (NHEJ) of DSB repair by pol μ and base excision repair (BER) by pols β and λ, translesion synthesis (TLS) by pol ι, and mitochondrial repair by pol γ (6–11, 21–26). Due to the high cellular rNTP/dNTP imbalance, discrimination against ribonucleotide incorporation by these DNA polymerases is essential for the maintenance of genomic integrity (1–4). It has been extensively shown by structure/function studies that most DNA polymerases use a side chain named the “steric gate” (i.e., B-, and Y- family pols) or a protein backbone segment in minor groove nucleotide binding pocket (i.e, X-family pols) to govern ribonucleotide exclusion (6–26). Replicative DNA polymerases effectively remove mismatches through their 3'-5' exonuclease domains, which is found to be relatively weaker at proofreading ribonucleotides during DNA synthesis (27). The kinetic basis of inefficient rNTP incorporation by DNA polymerases is mainly governed by weaker binding and slower rate of incorporation compared to dNTPs (21).
DNA polymerases and DNA ligases have fundamental roles in maintaining genome stability during the downstream steps of DNA replication and repair (35). Indeed, the efficiency of DNA ligation by DNA ligase at the final step relies on the accuracy of DNA synthesis that requires a Watson-Crick base pair between 5'-phosphate (P) and 3'-hydroxyl (OH) termini of the nick after DNA polymerase incorporates a correct dNTP (36). In our previous studies, we reported the importance of the coordination between DNA polymerase (pol) β and DNA ligase I (LIG1) or DNA ligase IIIα/XRCC1 (LIGIIIα/XRCC1) complex at the downstream steps of BER pathway to maintain accurate repair (37–47). Our biochemical studies have revealed that, in certain situations, BER responses can lead to mutagenic outcomes and the DNA ligation step of the BER pathway can be compromised in case of damaged (8-oxodGTP) or mismatched nucleotide incorporation by polβ (35–52). Furthermore, at atomic resolution, we recently reported that the LIG1 discriminates against the mutagenic mismatches distinctly depending on the architecture of 3'-terminus at the nick DNA. Our structures of LIG1/nick DNA complexes with G:T and A:C mismatches revealed the ligase strategies that favor or deter the ligation of base substitution errors at the final step of BER pathway (53).

Although the distinct features of the sugar selection mechanism have been extensively reported for DNA polymerases across families by X-ray crystallography studies (7–34), the mechanism of ribonucleotide discrimination by human DNA ligases at atomic resolution is entirely missing. Given the fact of high cellular rNTP/dNTP ratio, the mutagenic ligation of ribonucleotides by a DNA ligase at the final step following their incorporation by a DNA polymerase at a non-negligible rate during DNA replication or repair could constitute a major threat to genome stability. In the present study, we aimed to elucidate the mechanism by which human LIG1 facilitates or hinders sealing ribonucleotide at the 3'-end of nick DNA. Our LIG1/nick DNA structure uncovers a lack of a proficient sugar discrimination and demonstrates that the ligase active site can accommodate 3'-riboG opposite template base C. For the first-time, we showed a phosphodiester bond formation between 3'-OH and 5'-P ends of DNA termini at the nick, which refers to the final step of the ligation reaction involving an adenylate (AMP) release. Moreover, we demonstrated the mutagenic ligation of nick DNA substrate containing 3'-preinserted rG:C by LIG1 and LIGIIIα/XRCC1 complex in the ligation assays in vitro. Overall, our findings provide a novel insight into the characterization of ribonucleotide selectivity on the downstream events of DNA repair and replication as well as a lack of sugar discrimination by LIG1, demonstrating its distinctive ability to seal nick DNA with 3'-rG:C at atomic resolution.

**Results**

**Structures of LIG1/rG:C demonstrate a failure to discriminate against a ribonucleotide at the 3'-end of nick DNA**

To elucidate the sugar discrimination mechanism of LIG1 at atomic resolution, we determined the structure of LIG1 in complex with nick DNA containing rG:C at the 3'-end (Table 1). Our LIG1/3'-rG:C structure shows density for the 2'-hydroxyl (2'-OH) group on the ribose ring and demonstrates the formation of a phosphodiester bond between adjacent 3'-OH and 5'-P ends of DNA termini at the nick.
(Figure 1). This LIG1 structure represents the first structure of human DNA ligase that engages with a ribonucleotide-containing nick DNA during the final step of the ligation reaction where 3'-OH terminus of the nick attacks the 5'-P terminus downstream of the nick to form a phosphodiester bond that is coupled to a release of adenylate (AMP).

In our previously solved LIG1/A:T structure, we reported that the active site can accommodate a cognate A:T base pair where DNA-AMP intermediate is formed (53). The comparison of the 2Fc-Fo maps at 3'-OH and 5'-P ends of the nick DNA demonstrated continuous and discontinuous maps in the LIG1 structures of rG:C and A:T, respectively (Figure 2A-B). This means that both DNA ends are positioned for nick sealing in the LIG1/A:T structure, which refers to the step 2 of the ligation reaction, and a phosphodiester bond between 3'-OH and 5'-P ends is formed in the LIG1/3'-rG:C structure. The Fo-Fc map at 3σ (Figure 1B) and the 2Fo-Fc map at 1σ (Figure 2B) demonstrate that the AMP is incomplete in the LIG1/rG:C and shows less density than the LIG1/A:T structure. This observation indicates that AMP is still linked to the 5'-P end to form DNA-AMP intermediate in the LIG1/A:T structure, while it's already released in the LIG1/3'-rG:C structure, which refers to the step 3 of the ligation reaction. The overlay of both LIG1 structures shows a superimposed Cα root mean square deviation of 0.398 Å (Figure 2C) and no significant conformational difference in the domain organization of the LIG1 catalytic core (Supplementary Figure 1).

Our previously solved LIG1/mismatch structures demonstrated that the A:C is positioned at the initial step (step 1) where the ligase active site stays adenylated (LIG1-AMP) at its active site lysine residue (K568), and that the LIG1/G:T mismatch is captured during step 2 when an adenylate is transferred to the 5'-end of the nick (53), while the LIG1/rG:C structure shows a phosphodiester bond formation coupled to AMP release at step 3 of the ligation reaction (Supplementary Figure 2). We also compared the map of LIG1/rG:C (EE/AA) with the previously solved structure of LIG1/C:G (wild-type) (54,55). The overlay of both LIG1 structures demonstrated the 2Fc-Fo discontinuous map at 3'-OH and 5'-P ends of the nick for LIG1/C:G and a continuous map for LIG1/rG:C (Supplementary Figure 3). We observed a Watson-Crick conformation for rG:C and sugar pucker analysis of LIG1/3'-rG:C structure showed that the ribose adopts a C3'-endo conformation, which is similar to that of the LIG1/A:T structure (Supplementary Figure 4).

Mutagenic ligation of nick DNA containing a ribonucleotide at 3'-end by LIG1

To investigate the sugar discrimination mechanism of LIG1 against a ribonucleotide at the 3'-end of the nick \textit{in vitro}, we tested the ligation efficiency of wild-type and EE/AA mutant of LIG1 using the nick DNA substrate with 3'-preinserted rG opposite template base C that could be formed during DNA replication or repair when DNA polymerase incorporates a rNTP (Supplementary Scheme 1). In the control reactions, we used the nick DNA substrate with 3'-dG:C.

Our results demonstrated the mutagenic ligation of nick DNA with 3'-rG:C by wild-type (Figure 3A, lines 9-14) and EE/AA mutant (Figure 3C, lines 9-14) of LIG1. This was comparable with the ligation products of nick DNA containing a cognate 3'-dG:C (Figure 3A and 3C, lines 2-7). We observed a similar amount of ligation products for both DNA substrates (3'-rG:C \textit{versus} 3'-dG:C) with wild-type (Figure 3B) and the low
fidelity mutant of the ligase (Figure 3D). Similarly, our results demonstrated no significant difference between the mutagenic ligation efficiencies of both LIG1 proteins (Supplementary Figures 5-6).

In our structure studies, we use the EE/AA mutant of LIG1 carrying Glu(E)346 and Glu(E)592 mutations at the Mg²⁺-dependent high-fidelity site to crystallize the ligase in the conditions lacking Mg²⁺ (53,55). In the reaction conditions that mimic the LIG1 crystals, we also observed efficient nick sealing of the DNA substrate with 3’-rG:C in the absence of Mg²⁺ for longer time points of incubation (Supplementary Figure 7).

In our previous studies (38,42,45,53), we reported that LIG1 can ligate nick DNA containing a Watson-Crick like mismatch (G:T) and oxidative damage (8-oxoG:A). We next compared the mutagenic ligation efficiency of LIG1 for the nick DNA substrates containing a mismatched, damaged, and ribonucleotide-containing ends (Supplementary Figure 8). Our results showed that the ligation of 3’-8-oxoG:A and 3’-G:T yields nick sealing products along with the accumulation of DNA-AMP intermediates (Supplementary Figure 8A, lines 2-7 and 9-14, respectively), while there was only mutagenic ligation products in the presence of the nick DNA with 3’-rG:C (Supplementary Figure 8A, lines 16-21). These results demonstrate more efficient nick sealing of a ribonucleotide than a mismatched or damaged base at the 3’-end, which could explain why the ligase active site engages with those DNA ends at different steps of the ligation reaction (Supplementary Figure 8B).

Role of LIG1 active site residues for sugar discrimination against 3’-rG:C at the nick DNA

We previously reported that the LIG1 active site shows distinct DNA conformations depending on the identity of the 3’-terminus (cognate versus mismatched) (53). The overlay of LIG1 structures in complex with nick DNA containing 3’-rG:C and 3’-A:T showed no conformational change at the positions of the LIG1 active site residues F635, R871, and F872, which reside in the close proximity to the 2’-OH of the ribose (Figure 4A). Furthermore, we did not observe a shift in the position of F872 at the ligase active site engaging with rG:C as we reported previously in the LIG1/A:C structure (Figure 4B).

To further investigate this observation, we tested the importance of the ligase active site residue F872 for the end joining efficiency of the nick DNA substrate with 3’-rG:C by LIG1 in the ligation assays (Supplementary Scheme 1). For this purpose, we used the LIG1 mutants carrying a F872A mutation in the wild-type (F872A) and low fidelity mutant (EE/AA) background. Our results demonstrated that LIG1 F872A (Figure 5A, lines 2-7) and LIG1 EE/AA/F872A (Figure 5A, lines 8-13) can also efficiently seal nick DNA with 3’-rG:C. There were slightly less ligation products at the initial time points of the reaction (Figure 5B and Supplementary Figure 9) in comparison with the wild-type and EE/AA mutant of LIG1 (Figure 3). These findings demonstrate that the active site residue F872 that is located upstream of the nick and positioned close to the deoxyribose moiety of the nucleotide at the 5’-end has no role in sugar discrimination by LIG1.

Discussion
Ribonucleotides (rNTPs) are found at a concentration several orders of magnitude higher than dNTPs in most cells (1–3). It has been shown that the eukaryotic replicative polymerases insert around 10,000 rNTPs per round of DNA replication and repair polymerases are also capable of inserting rNTPs during DNA synthesis with some being better at discriminating against rNTP insertion than others (6, 14). Therefore, rNMPs are the most common type of DNA lesion found in the genomic DNA with critical effects on overall genome stability (27). Although DNA polymerases have evolved selectivity against rNTPs favoring dNTP insertion and have an intrinsic exonuclease proofreading function to remove rNTPs, this is not sufficient to prevent large amounts of rNTPs remaining in the DNA backbone (6–26). The RER pathway, involving RNase H2 function, can efficiently remove ribonucleotides from RNA:DNA hybrids by cleaving at the 5'-end of the ribonucleotide, which is followed by strand displacement synthesis by pol δ and flap removal by Flap Endonuclease 1 (FEN1) and/or Dna2, and finally ligation step by LIG1 (56–58).

In addition to its role in the RER pathway, LIG1 is responsible for the ligation of over 10 million Okazaki fragments during each round of DNA replication and nick sealing of the repair products at the final step of all DNA repair pathways which process DNA damage that could be formed due to a variety of endogenous and exogenous sources (59–61). The nick repair products are indeed potentially toxic intermediates, and this may lead to DNA strand breaks that could trigger harmful nuclease activities, cytotoxicity, double strand breaks in replicating DNA, and ultimately genomic instability (61–63). Therefore, DNA ligases could be challenged by various genomic insults such as oxidative damage, non-canonical mismatches, and ribonucleotides embedded into genomic DNA due to their incorporations by DNA polymerases on the downstream steps of DNA repair and replication (35, 36). DNA ligases have evolved a tolerance for ribonucleotides at 3'-end while heavily discriminating against a ribonucleotide at 5'-end (54, 59–63). On the other hand, the mechanism of sugar discrimination during nick sealing reaction by DNA ligase at atomic resolution remains unknown. ATP-dependent DNA ligases catalyze the ligation reaction including three sequential steps (64–69): (i) enzyme adenylation at the active site lysine residue (i.e., K568) and formation of the ligase-AMP intermediate, (ii) transfer of AMP group to 5'-P end of nick DNA resulting in the formation of DNA-AMP intermediate, and finally (iii) phosphodiester bond formation between adjacent 3'-OH and 5'-P DNA termini at a nick.

Our previous and present LIG1 structures revealed the features of DNA substrate and LIG1 interaction that dictate accurate versus mutagenic outcomes (53). Our structures contribute to the understanding of the mechanism by which LIG1 discriminates against non-canonical DNA intermediates during nick sealing at each individual step of the ligation reaction as well as how the ligase active site engages with nick DNA distinctly depending on the architecture of cognate (step 2), mismatched (step 1 or 2), or ribonucleotide (step 3)-containing ends (Supplementary Table 1). Overall LIG1 structures in complex with nick DNA substrates have provided considerable mechanistic details about steps 1 and 2 of the ligation reaction while relatively less information has been reported for step 3 (53–55, 70, 71). In addition to our present study of the LIG1/rG:C structure showing the human ligase active site at step 3 in the crystal conditions lacking Mg²⁺ using EDTA pre-treated LIG1 protein, another step 3 structure has been shown in ATP-dependent ligase from Prochlorococcus marinus (Pmar-Lig) (72). In this structure, Pmar-Lig (step 2)
crystals were immersed in cryoprotectant buffer containing MgCl₂ and then frozen immediately to get the step 3 Pmar-Lig/DNA structure, which is poised for phosphodiester bond formation with partially occupied AMP in the active site. Since it is well known that Mg²⁺ is required for step 3 of the ligation reaction, it is at first puzzling why our LIG1/rG:C without Mg²⁺ shows step 3 (73). However, this is likely because of the 3'-endo sugar pucker that ribonucleotides adopt. In B-form DNA, the sugar moieties usually adopt the 2'-endo sugar pucker conformation, while in A-form DNA or RNA duplexes the sugar moieties usually adopt the 3'-endo sugar pucker conformation (74). The ligase active site forces the DNA upstream of the nick to adopt an A-form conformation which causes the sugar at the 3'-end to assume the 3'-endo conformation, both of which are necessary for ligation (54). However, Mg²⁺ plays an important role in this step. In previous studies, it has been shown that the catalytic Mg²⁺ can activate the 3'-OH for nucleophilic attack on the 5'-P (73). However, this catalytic Mg²⁺ plays a dual role where the interaction between Mg²⁺ and the 3'-OH also stabilizes the 3'-endo conformation of the 3'-end. When we compare the Mg²⁺-containing structure of LIG1 wild-type/3'-ddC:G (dideoxycytidine), which lacks a 3'-OH group, as well as the structure of Pmar-Lig/3'-dC:G with non-active Mn²⁺, the divalent metal atom is in almost the same position, which demonstrates that the position of the divalent metal is conserved for the ligases. If we imagine the position of the 3'-OH in the LIG1/3'-ddC:G structure, it is also in the same position as Pmar-Lig/3'-dC:G; additionally, the divalent metal interacts with both 3'-OH and 5'-P (Supplementary Fig. 10). Since rG at the 3'-end adopts a stable 3'-endo conformation due to its identity as a ribonucleotide, it does not require the rigidity enforced by the catalytic Mg²⁺ to properly align the 3'-OH for catalysis. We assume that this is the reason LIG1/rG:C shows step 3 of ligation reaction even without Mg²⁺. Indeed, this is further supported biochemically by the ligation of nick DNA substrate with 3'-rG:C in the absence of Mg²⁺ (Supplementary Fig. 7).

To understand the mechanism of sugar discrimination by human DNA ligases more comprehensively, further structural and biochemical studies are required with other repair ligases such as DNA ligase IIIα and DSB repair ligase IV (75–81). We previously reported that the BER DNA ligases, LIG1 and DNA ligase IIIα/XRCC1 complex, are compromised by subtle changes in all 12 possible mismatched base pairs at the 3'-end of the nick repair intermediate (42). In the present study, our ligation assays with DNA ligase IIIα/XRCC1 complex also demonstrated mutagenic nick sealing of DNA substrate with 3'-rG:C as efficient as a cognate 3'-dG:C ligation (Supplementary Fig. 11), and we did not observe any significant difference between the mutagenic ligation efficiencies of LIG1 and DNA ligase IIIα/XRCC1 complex (Supplementary Fig. 6). Although these findings indicate a lack of sugar discrimination against a ribonucleotide by LIG1 for 3'-rG:C nick DNA only, further structure/function characterizations of LIG1 against the nick DNA intermediates containing all possible ribonucleotides mismatches at either 3'- or 5'-end at a nick are necessary to provide a comprehensive insight into how LIG1 bestows substrate specificity during DNA replication and repair, which is important to understand the disease mechanisms and to develop potential therapeutic targets and strategies (82–84).

**Methods**
Preparation of nick DNA substrates

DNA substrates with a 6-carboxyfluorescein (FAM) label were prepared as described previously (35-53). Nick DNA substrates containing 3'-preinserted dG:C, dG:T, 8oxodG:A, and rG:C were used in the ligation assays including DNA ligase I (wild-type, and F872A, EE/AA, and EE/AA/F872A mutants) or DNA ligase IIIα/XRCC1 complex (Supplementary Table 2). Nick DNA substrates for LIG1 X-ray crystallography were prepared by annealing upstream, downstream, and template primers (Supplementary Table 3).

Protein purifications

His-tag (pET-24b) human C-terminal (Δ261) wild-type and E346A/E592A (EE/AA) mutant of LIG1 proteins were purified as described previously (35-53). Briefly, they were overexpressed in Rosetta (DE3) E. coli cells in Terrific Broth (TB) media with kanamycin (50 μg/ml) and chloramphenicol (34 μg/ml) at 37 °C. The cells were induced with 0.5 mM isopropyl β-D-thiogalactoside (IPTG) when the OD600 was reached to 1.0, and the overexpression was continued for overnight at 28 °C. Cells were lysed in the lysis buffer [50 mM Tris-HCl (pH 7.0), 500 mM NaCl, 20 mM imidazole, 10% glycerol, 1 mM PMSF, an EDTA-free protease inhibitor cocktail tablet] by sonication at 4 °C and the cell lysate was pelleted at 31,000 x g for 1 h at 4 °C. Proteins were purified by a HisTrap HP column with an increasing imidazole concentration (20-300 mM) after being equilibrated in the binding buffer [50 mM Tris-HCl (pH 7.0), 500 mM NaCl, 20 mM imidazole, 10% glycerol] at 4 °C. The collected fractions were subsequently loaded onto a HiTrap Heparin column that was equilibrated with the binding buffer [50 mM Tris-HCl (pH 7.0), 50 mM NaCl, 1.0 mM EDTA, and 10% glycerol], and then eluted with a linear gradient of NaCl up to 1 M. LIG1 proteins were further purified by Superdex 200 10/300 columns in the buffer [20 mM Tris-HCl (pH 7.0), 200 mM NaCl, and 1 mM DTT]. The plasmid DNA constructs encoding human C-terminal (Δ261) LIG1 genes carrying a single F872A and triple EE/AA/F872A mutations were generated by side-directed mutagenesis using the plasmid DNA of LIG1 wild-type and E346A/E592A (EE/AA) mutant, respectively, as previously described (35-53). His-tag LIG1 mutant proteins F872A and EE/AA/F872A were overexpressed and purified as described above. Human his-tag DNA ligase IIIα/XRCC1 complex (pDuet1) was purified as previously described (35-53). Briefly, the complex protein was overexpressed in BL21(DE3) E. coli cells in LB media at 37 °C and the cell lysis was prepared as described above. The supernatant was loaded onto a HisTrap HP column (GE Health Sciences) and purified with an increasing imidazole gradient (20-300 mM) elution at 4 °C. The collected fractions were subsequently loaded onto a HiTrap Heparin column that was equilibrated with the binding buffer [25 mM Tris-HCl (pH 7.5), 50 mM NaCl, and 10% glycerol], and then eluted with a linear gradient of NaCl up to 1 M. The protein complex was further purified by Superdex 200 Increase 10/300 chromatography (GE Healthcare) in the buffer [25 mM Tris-HCl (pH 7.0), 200 mM NaCl, glycerol 5%, and 1 mM DTT]. All proteins purified in this study were dialyzed against storage buffer including 25 mM Tris-HCl (pH 7.0), 200 mM NaCl, concentrated, frozen in liquid nitrogen, and stored at -80 °C.

Crystallization and structure determination
LIG1 C-terminal (∆261) EE/AA mutant that was pre-treated with EDTA before using in the crystallization experiments in the absence of Mg$^{2+}$ as described previously (53). All LIG1-nick DNA complex crystals were grown at 20 °C using the hanging drop method. LIG1 (at 27 mg ml$^{-1}$)/DNA complex solution was prepared in 20 mM Tris-HCl (pH 7.0), 200 mM NaCl, 1 mM DTT, 1 mM EDTA and 1 mM ATP at 1:4:1 DNA:protein molar ratio and then mixed with 1 μl reservoir solution containing 100 mM MES (pH 6.1), 100 mM lithium acetate, and 14% (w/v) polyethylene glycol PEG3350. All crystals grew in one day. They were then washed in the reservoir solution with 20% glycerol and flash cooled in liquid nitrogen for data collection. The crystals were maintained at 100° K during X-ray diffraction data collection using the beamline 7B2 at Cornell High Energy Synchrotron Source (CHESS). The diffraction images were indexed and integrated using HKL2000. All structures were solved by the molecular replacement method using PHASER with PDB entry 7SUM as a search model (53,85,86). Iterative rounds of model building were performed in COOT and the final models were refined with PHENIX or REFMAC5 (86-89). 3DNA program was used for sugar pucker analysis (90). All structural images were drawn using PyMOL (The PyMOL Molecular Graphics System, V0.99, Schrödinger, LLC). Detailed crystallographic statistics are provided in Table 1.

**DNA ligation assays**

The nick sealing efficiency of DNA ligase (DNA ligase I or DNA ligase IIIα/XRCC1 complex) was investigated to determine the sugar discrimination mechanism against 3'-preinserted rG opposite template base C in the ligation assays (Supplementary Scheme 1) as described previously (35-53). Briefly, the reaction mixture containing 50 mM Tris-HCl (pH 7.5), 100 mM KCl, 10 mM MgCl$_2$, 1 mM ATP, 1 mM DTT, 100 μg ml$^{-1}$ BSA, 10% glycerol, and 500 nM nick DNA substrate was initiated by the addition of 100 nM DNA ligase, incubated at 37 °C, and stopped at the time points indicated in the figure legends. The reaction products were then quenched with an equal amount of gel loading buffer [95% formamide, 20 mM ethylenediaminetetraacetic acid, 0.02% bromophenol blue and 0.02% xylene cyanol]. After incubation at 95 °C for 3 min, the reaction products were separated by electrophoresis on an 18% denaturing polyacrylamide gel. The gels were scanned with a Typhoon Phosphor Imager (Amersham Typhoon RGB), and the data were analyzed using ImageQuant software. The ligation assays were performed similarly with nick DNA substrates containing 3’-dG:C, 3’-8oxodG:A, or 3’-dG:T. In addition, nick sealing of DNA with 3’-rG:C by EE/AA mutant of LIG1 in the absence of Mg$^{2+}$ was performed as described above in the reaction buffer lacking 10 mM MgCl$_2$.

**Data availability**

Atomic coordinates and structure factors for the reported crystal structure of LIG1 in complex with nick DNA/3’-riboG:C has been deposited in the RCSB Protein Data Bank under accession number 8DVT. All relevant data are available from the authors upon reasonable request.

**References**

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


43. Çağlayan, M. Pol β gap filling, DNA ligation and substrate-product channeling during base excision repair opposite oxidized 5-methylcytosine modifications. DNA Repair 95,102945 (2020).


Declarations

Funding

This work was supported by a grant 1R35GM147111-01 from the National Institute of General Medical Sciences (NIGMS).

Acknowledgements

This work is based upon research conducted at the Center for High Energy X-ray Sciences (CHEXS), which is supported by the National Science Foundation under award DMR-1829070, and the Macromolecular Diffraction at CHESS (MacCHESS) facility, which is supported by award 1-P30-GM124166-01A1 from the National Institute of General Medical Sciences, National Institutes of Health, and by New York State’s Empire State Development Corporation (NYSTAR). The authors thank Dr. Robert McKenna and Jacob Combs (University of Florida) for beamline 7B2 at CHESS, and an assistance with crystal shipment and data collection.

Author contributions
Conceptualization M.Ç., methodology and investigation T.Q., M.G., M.Ç.; writing-original draft T.Q., M.G., M.Ç.; writing-reviewing and editing T.Q., M.G., M.Ç.; funding acquisition M.Ç.

## Tables

**Table 1. X-ray data collection and refinement statistics of LIG1/nick DNA duplex with a ribonucleotide (rG:C) at the 3'-end.**

<table>
<thead>
<tr>
<th></th>
<th>LIG1&lt;sup&gt;EE/AA&lt;/sup&gt; Nick DNA with 3'-rG:C</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>PDB entry ID</strong></td>
<td></td>
</tr>
<tr>
<td><strong>Data collection</strong></td>
<td></td>
</tr>
<tr>
<td>Space group</td>
<td>P2₁2₁2₁</td>
</tr>
<tr>
<td>Cell dimensions</td>
<td></td>
</tr>
<tr>
<td>a, b, c (Å)</td>
<td>65.065, 115.633, 126.338</td>
</tr>
<tr>
<td>a, b, g (°)</td>
<td>90</td>
</tr>
<tr>
<td>Resolution (Å)</td>
<td>20-3.0 (3.05-3.0)</td>
</tr>
<tr>
<td>(R_{sym})</td>
<td>0.181 (1.451)</td>
</tr>
<tr>
<td>(I / \sigma (I))</td>
<td>17.7 (1.7)</td>
</tr>
<tr>
<td>(CC_{1/2})</td>
<td>1.000 (0.655)</td>
</tr>
<tr>
<td>(CC^*)</td>
<td>1.000 (0.890)</td>
</tr>
<tr>
<td>Completeness (%)</td>
<td>96.3 (100)</td>
</tr>
<tr>
<td>Redundancy</td>
<td>12.5 (13.6)</td>
</tr>
<tr>
<td><strong>Refinement</strong></td>
<td></td>
</tr>
<tr>
<td>Resolution (Å)</td>
<td>20-3.0</td>
</tr>
<tr>
<td>No. reflections</td>
<td>17321</td>
</tr>
<tr>
<td>(R_{work}/R_{free})</td>
<td>0.172/0.206</td>
</tr>
<tr>
<td>Non-H atoms:</td>
<td></td>
</tr>
<tr>
<td>Protein/DNA</td>
<td>5665</td>
</tr>
<tr>
<td>AMP</td>
<td>21</td>
</tr>
<tr>
<td>Average B-factors (Å²):</td>
<td>60.1</td>
</tr>
<tr>
<td>R.M.S.D</td>
<td></td>
</tr>
<tr>
<td>Bond lengths (Å)</td>
<td>0.009</td>
</tr>
<tr>
<td>Bond angles (°)</td>
<td>1.174</td>
</tr>
</tbody>
</table>

**Figures**
Figure 1

**Structure of LIG1 bound to nick DNA duplex containing 3’-rG:C.** A, X-ray crystal structure of LIG1/nick DNA duplex containing a ribonucleotide (rG) at 3’-terminus of the nick DNA. Simulated annealing omit map (Fo-Fc) of LIG1/rG:C (cyan) is displayed for 2’-OH of 3’-terminus contoured at 3σ. Schematic view of DNA including DNA:RNA hybrid used in the LIG1/nick DNA duplex crystallization. B, Simulated annealing omit map (Fo-Fc) of LIG1/rG:C is displayed for 3’-OH terminus and AMP contoured at 3σ. DNA and AMP are shown as sticks and the map is depicted in green.
LIG1/rG:C structure demonstrates a phosphodiester bond formation and AMP release during the final ligation step at atomic resolution. **A-B**, The 2Fo - Fc density map (1σ) is displayed for 3'- and 5'-terminus of the nick in the LIG1/rG:C (cyan) and LIG1/A:T (magenta) structures. AMP is incomplete and shows lower density in the LIG1/rG:C (step 3) than the LIG1/A:T (step 2). DNA and AMP are shown as sticks and the map is depicted in grey. **C**, Overlay of LIG1 structures bound to the nick DNA duplexes containing rG:C (cyan) and A:T (PDB: 7SUM, magenta) ends shows the difference at the positions of 3' and 5' terminus of the nick.
Figure 3

Mutagenic ligation of the nick DNA with 3’-ribonucleotide by LIG1. A-C, Lanes 1 and 8 are the negative enzyme controls of the nick DNA substrates with 3’-dG:C and 3’-rG:C, respectively. Lanes 2-7 and 9-14 are the ligation products in the presence of 3’-dG:C and 3’-rG:C, respectively, by wild-type (A) and EE/AA mutant (C) of LIG1, and correspond to time points of 0.5, 1, 3, 5, 8, and 10 min. B-D, Graphs show time-dependent change in the amount of ligation products. The data represent the average from three independent experiments ± SD.
**Figure 4**

Comparison of LIG1 active site residues between the structures of LIG1/nick DNA complexes with rG:C versus cognate and mismatched ends. Overlay of LIG1 structures bound to the nick DNA duplexes containing rG:C (cyan), A:T (PDB: 7SUM, magenta), and A:C (PDB: 7SX5, grey) ends shows the differences in the positions of 3'- and 5'-terminus at the nick DNA and no difference in the LIG1 active site F635, R871, and F872 residues around 2'-OH of the ribose.

**Figure 5**

- **A**
  - Ligation product
  - Ligation substrate
  - Time (min)
  - F872A
  - EE/AA/F872A

- **B**
  - Ligation Product (%)
  - Time (min)
  - F872A
  - EE/AA/F872A
Role of LIG1 active side residue F872 for mutagenic ligation of the nick DNA with 3'-rG:C. A, Lane 1 is the negative enzyme control of the nick DNA substrate with 3'-rG:C. Lanes 2-7 and 8-13 are the ligation reaction products by LIG1 F872 and EE/AA/F872A mutants, respectively, and correspond to time points of 0.5, 1, 3, 5, 8, and 10 min. B, Graph shows time-dependent change in the amount of ligation products. The data represent the average from three independent experiments ± SD.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [SupplementaryData.pdf](#)