Structures of LIG1 uncover the mechanism of sugar discrimination against a ribonucleotide at 3'- and 5'-end of the nick DNA

Qun Tang, Mitchell Gulkis, and Melike Çağlayan*

Department of Biochemistry and Molecular Biology, University of Florida, Gainesville, FL 32610, USA

*To whom correspondence should be addressed. Tel.: +1 352-294-8383; E-mail: caglayanm@ufl.edu

Human DNA ligase I (LIG1) is the main replicative ligase that seals Okazaki fragments and finalizes DNA repair pathways by joining canonical 3'-OH and 5'-P ends of the nick DNA in a three-step ligation reaction. Ribonucleotides can be misincorporated by DNA polymerases resulting in a nick with 3'-ribonucleotide while RNase H2 mediated cleavage leaves a nick harboring 5'-ribonucleotide during ribonucleotide excision repair. However, how LIG1 surveils DNA ends with a "wrong" sugar at atomic resolution is unknown. Here, we determine X-ray structures of LIG1/nick DNA complexes with 3'- or 5'-single ribonucleotide during different stages of the ligation reaction. Our LIG1/5'-rG:C structure reveals a global conformational change, which discriminates against 5'-RNA/DNA junctions at the initial step when the ligase-AMP intermediate is formed. Furthermore, we capture LIG1/3'-RNA-DNA heteroduplexes that are tolerated at the active site where AMP is transferred to nick DNA (step 2) and final phosphodiester bond formation occurs (step 3). Finally, we demonstrate the mutagenic and defective ligation of the nick DNA with 3'- and 5'-ribonucleotide, respectively, *in vitro*. Together, these results uncover how LIG1 encounters ribonucleotides embedded into genome during nuclear replication and the last step of DNA repair pathways to maintain genome integrity.

Mammalian DNA ligases contain a conserved three domain core architecture consisting of the oligonucleotide binding (OBD), the adenylation (AdD), and the DNA binding (DBD) domains, which seal 3'-hydoxyl (OH) and 5'-phosphate (P) termini of the nick DNA in a conserved ligation reaction involving the three consecutive chemical steps[1,2] (Extended Data Scheme 1): the formation of the DNA ligase-adenylate intermediate (LIG-AMP) in step 1, subsequent transfer of AMP moiety to the 5'-P end of the nick (DNA-AMP) in step 2, and final phosphodiester bond formation coupled to AMP release in step 3[3,4]. Despite these similarities, human DNA ligases exhibit different ligation fidelity which relies on the accuracy of DNA synthesis that requires a Watson-Crick base pair when DNA polymerase incorporates a correct nucleotide during DNA replication and repair[5,6].

The concentration of ribonucleotide triphosphates (rNTPs) is 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, making genomic ribonucleotides the most prevalent source of cellular DNA damage[7-10]. In human cells, ribonucleotide incorporation can occur during almost all cellular DNA transactions[11-17], and discrimination against ribonucleotide incorporation by replicative and repair DNA polymerases is essential for the maintenance of genomic integrity[18-27]. 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, and chromosome alterations[28-31].

Among the four known human DNA ligases, LIG1 plays a critical role in Okazaki fragment maturation during DNA replication and finalizes the last ligation step of most of DNA repair pathways[1-4]. In the first solved structure of LIG1, it has been shown in the ligation assays *in*

50 *vitro* that the ligase displays a great discrimination against the DNA substrate when the 5'-
51 phosphorylated strand is completely RNA, while it can ligate the RNA strand that is located at
52 the upstream of the nick[32]. Yet, the mechanism of sugar discrimination by LIG1 at atomic
53 resolution is entirely missing. Recently solved LIG1 structures demonstrated that the ligase
54 employs a Mg^{2+}-reinforced nick DNA-binding mode to ensure high fidelity ligation, and this
55 accuracy is a critical determinant of faithful replication of the nuclear genome[33-35].
56 Furthermore, our structures revealed that LIG1 active site discriminates against mutagenic
57 mismatches distinctly depending on the architecture of the 3'-terminus at the nick DNA[36]. The
58 ligation efficiency for the ribonucleotides at positions near the ends of nick DNA has been also
59 reported for DNA ligases from *Chlorella virus, Thermus thermophilus*, and *Saccharomyces*
60 *cerevisiae*[37-41].
61 Ribonucleotide Excision Repair (RER) is the primary mechanism through which
62 ribonucleotides embedded into genomic DNA can be repaired[42,43]. The RER pathway is
63 initiated by RNase H2-mediated incision of the DNA backbone at the 5'-end of the
64 ribonucleotide, followed by strand displacement synthesis by pol δ, then flap removal by Flap
65 Endonuclease 1 and/or Dna2, and finally nick sealing by LIG1[44-46]. Previous studies also
66 demonstrated that LIG1 attempts to ligate the RNA-DNA lesion following RNase H2 cleavage,
67 which results in the formation of abortive ligation products with a 5'-adenylated-ribonucleotide
68 that can be resolved by Aprataxin (APTX)[47,48]. Despite of the fact that LIG1 involves in the
69 RER process, how it surveils the RNA-DNA junctions harboring a 5'-ribonucleotide at atomic
70 resolution remains unknown.
71 In the present study, we aimed to elucidate the mechanism by which human LIG1 discriminates
72 against a single ribonucleotide at the 3'- and 5'-ends of nick DNA. Our LIG1 structures uncover
73 a lack of proficient sugar discrimination and mutagenic ligation of nick DNA with a
74 ribonucleotide at the 3'-end during different steps of ligation reaction (Extended Data Scheme

75 1). We also demonstrate a conformational change of LIG1 in the presence of 5'-RNA-DNA
76 junction, which deters nick sealing during initial step. Overall, our findings provide a novel
77 insight into the characterization of ribonucleotide selectivity by LIG1 on the downstream
78 events of DNA replication and repair, demonstrating its ability to surveil ribonucleotide-
79 containing ends of nick DNA distinctly depending on the position of an incorrect sugar.

80 **LIG1 fails to discriminate against a ribonucleotide at 3'-end of the nick DNA**

81 We solved the structures of LIG1/nick DNA complexes containing rA:T, rC:G, and rG:C at
82 the 3'-end (Fig. 1 and Extended Data Table 1) for $LIG1^{WT}$ and $LIG1^{EE/AA}$ that harbors E346A
83 and E592A mutations resulting in the ablation of the high-fidelity site, which has been utilized
84 in previous structures with non-canonical substrates[33-36]. The root mean square deviation
85 (RMSD) of the LIG1 structures for 3'-rA:T, 3'-rC:G, and 3'-rG:C was no higher than 0.7 Å for
86 all main chain atoms, which demonstrates that all LIG1 structures show similar global
87 conformations (Extended Data Table 2). We observed the ligase active site in step 2 of the
88 ligation reaction when the AMP moiety is transferred to the 5'-end of the nick DNA, forming
89 a DNA-AMP intermediate, in the structures of $LIG1^{WT}$/3'-rA:T and $LIG1^{EE/AA}$ for 3'-rA:T, 3'-
90 rC:G, and 3'-rG:C. Similarly, for the LIG1/nick DNA complexes containing canonical ends,
91 $LIG1^{WT}$/3'-dA:T and $LIG1^{EE/AA}$/3'-dG:C structures showed step 2 (Fig. 1a-c,e,f,h). In addition,
92 we captured $LIG1^{EE/AA}$/3'-rA:T and $LIG1^{EE/AA}$/3'-rG:C during the final step of the ligation
93 reaction where the 3'-OH terminus attacks the 5'-P terminus downstream of the nick to form a
94 phosphodiester bond (Fig. 1d,g). In the LIG1/nick DNA structures captured in step 2, the Fo-
95 Fc maps at 3σ of the AMP moiety show a density for almost all the atoms in the AMP (Fig.
96 1a-c,e,f,h). In the step 3 structures, the Fo-Fc maps at 3σ of the AMP moiety are incomplete,
97 suggesting that the AMP is released (Fig. 1d,g). The only difference in the crystallization
98 conditions between step 2 (one day) and step 3 (three days) structures of LIG1 is the amount
99 of time during which the ligase/nick DNA crystals were allowed to grow before freezing.

100 The superimposition of LIG1 structures showing step 2 ($LIG1^{WT}$/3'-dA:T, $LIG1^{WT}$/3'-rA:T,
101 $LIG1^{EE/AA}$/3'-rA:T) *versus* step 3 ($LIG1^{EE/AA}$/3'-rA:T) of the ligation reaction demonstrates that
102 the C3' atom of the ribose on the 3'-end moves closer to the 5'-P at the nick in the ligase
103 structures exhibiting a final phosphodiester bond formation (Fig. 1i). We observed a similar
104 difference in the overlay of G:C structures showing step 2 ($LIG1^{EE/AA}$/3'-dG:C, $LIG1^{EE/AA}$/3'-
105 rG:C) *versus* step 3 ($LIG1^{EE/AA}$/3'-rG:C) (Fig. 1j). The distance between C3' atom of the ribose
106 and 5'-P was ~ 1.0 Å shorter (~ 3.5-4.0 Å and 2.6 Å which correspond to steps 2 and 3,
107 respectively) in the step 3 structures of LIG1 (Extended Data Fig. 1). This is consistent with
108 the formation of a covalent bond between the 3'- and 5'-ends of the nick. Furthermore, the
109 comparison of the 2Fo-Fc maps at the 3'-OH and 5'-P ends of the nick demonstrated continuous
110 and discontinuous maps in the LIG1 structures captured in the steps 2 and 3, respectively
111 (Extended Data Fig. 1). Finally, we observed a Watson-Crick base pair at the 3'-end for all
112 LIG1/nick DNA structures containing 3'-ribonucleotides and sugar pucker analyses
113 demonstrate that the ribose adopts a C3'-endo conformation (Extended Data Figs. 2 and 3).

114 **LIG1 discriminates against RNA/DNA junctions harboring a ribonucleotide at the 5'-end**

115 We captured the LIG1/nick DNA complex containing a 5'-ribonucleotide during the initial step
116 1 of the ligation reaction where an AMP moiety remains covalently bond to the active side
117 chain of K568 (Fig. 2a). The superimposition of overall step 2 and 3 structures of the present
118 study demonstrated that they are similar with the global conformation (Fig. 2b). When we
119 overlayed the three domains (DBD, AdD, OBD) of $LIG1^{EE/AA}$ for 5'-rG:C and 5'-dG:C
120 structures individually, we found that the RMSD value was 1 Å, demonstrating no major
121 conformational change within the individual domains (Fig. 2c,d and Extended Data Table 2).
122 In previously solved structures of LIG1/nick DNA duplexes containing canonical, damaged
123 (8-oxoG:A), and mismatched ends[32-36], no global ligase conformational change has been
124 reported and all the RMSD values have been found to be less than 1 Å. However, when we

125 overlayed the entire catalytic core of LIG1$^{EE/AA}$ for 5'-rG:C and 5'-dG:C, as well as our other
126 LIG1 structures of 3'-ribonucleotides, the RMSD value was found to be higher than 2 Å,
127 indicating a large global conformational change (Fig. 2c and Extended Data Table 2).
128 To determine the cause of the global conformational change, we aligned the DBD and AdD
129 domains of LIG1$^{EE/AA}$ individually for 5'-dG:C and 5'-rG:C structures. The overlay of DBD
130 demonstrates a conformational change in the AdD domain at the loop 1 (541-549 aa), and the
131 overlay of AdD domain reveals a conformational change in the OBD domain at the loop 2
132 (749-759 aa) (Fig. 2e,f). These subtle changes observed in the loop 1 and loop 2 result in the
133 large global conformational change because of the misalignment of entire domains relative to
134 each other. A similar structural rearrangement has been previously reported for LIG4
135 undergoing a conformational change from an "open" LIG-AMP state to "closed" DNA-AMP
136 conformation[49].
137 Additionally, we aligned the AdD domain of LIG1$^{EE/AA}$ for 5'-dG:C and 5'-rG:C structures and
138 demonstrated that loop 3 (733-742 aa), which contains Lys(R)738 that shifts more than 10 Å
139 closer to the 5'-P in the 5'-rG:C structure compared to the 5'-dG:C structure and forms a putative
140 hydrogen bond with the 5'-P (Fig. 2g). We posit that this additional interaction deters AMP
141 transfer from the ligase active site (K568) to the 5'-end of the nick (step 2) by increasing the
142 energetic barrier to the adenylate transfer.
143 Lastly, the overlay of LIG1$^{EE/AA}$/5'-rG:C (step 1) with all LIG1^{WT} or LIG1$^{EE/AA}$ structures for
144 3'-RNA-DNA complexes and canonical DNA (steps 2 and step 3) as well as LIG1$^{EE/AA}$ with a
145 3'-dA:C mismatch[36] (PDB: 7SX5, step 1), revealed a difference in the DNA conformation (Fig.
146 3a). The RMSD using main chain atoms of LIG1$^{EE/AA}$ for 5'-rG:C with 5'-dG:C and 3'-dA:C
147 structures was found to be ~ 2.7Å (Extended Data Table 2). The main difference is the rotation
148 of the 5'-P downstream of the nick DNA which widens the major groove downstream of the
149 nick DNA and results in a narrowing of the minor groove (Fig. 3b,c). The minor groove is the

150 interaction site of the OBD domain and narrowing the minor groove changes the conformation
151 of OBD between the LIG1$^{EE/AA}$/5'-dG:C and 5'-rG:C structures (Fig. 3d).

152 **Ligation efficiency of nick DNA containing a ribonucleotide at 3'- or 5'-end by LIG1**

153 We finally tested the ligation efficiency of LIG1 using the nick DNA substrates with
154 preinserted 3'-ribonucleotides and 5'-rG:C (Extended Data Scheme 2). Our results
155 demonstrated the mutagenic ligation of the nick DNA substrates containing 3'-rA:T, 3'-rC:G,
156 and 3'-rG:C (Fig. 4a). As expected, we obtained efficient nick sealing of canonical 3'-dA:T, 3'-
157 dC:G, and 3'-dG:C ends (Extended Data Fig. 4) demonstrating no major difference in the
158 amount of ligation products (Fig. 4b). On the other hand, our results demonstrated a lack of
159 nick sealing by LIG1 in the presence of 5'-rG:C for the time points (10-60 sec) where we
160 obtained efficient ligation of 3'-ribonucleotides (Fig. 4c, lanes 4-7) showing ~100-fold
161 difference in the ligation products between 3'-ribonucleotides and 5'-rG:C (Fig. 4d). Consistent
162 with previous report[47], we observed ligase failure products with 5'-AMP-RNA for longer time
163 points of the reaction (Fig. 4c, lanes 8-11 and Extended Data Fig. 5).

164 **Discussion**

165 LIG1 is responsible for the ligation of over 10 million Okazaki fragments during each round
166 of DNA replication and nick sealing of the repair products at the final step of all DNA repair
167 pathways[1-4]. Previously solved structures of LIG1/nick DNA duplexes with canonical,
168 damaged, and mismatched ends have provided considerable details about steps 1 and 2 of the
169 ligation reaction while relatively less information has been reported for final step[32-36]. Our
170 structures uncover the features of 3'-RNA-DNA substrate and LIG1 interaction resulting in
171 mutagenic outcomes at atomic resolution, which was entirely missing. This lesion mimics a
172 DNA repair intermediate with 3'-ribonucleotides embedded in the genomic DNA because of
173 their misincorporation by replicative and repair DNA polymerases in nucleus or
174 mitochondria[11-27]. Using EDTA pre-treated LIG1$^{EE/AA}$ protein, we captured the ligase active

175 site while engaging with 3'-ribonucleotides (rA, rG, rC) during steps 2 and 3 in the crystal
176 conditions lacking Mg^{2+}. The lack of Mg^{2+} is confirmed in $LIG1^{WT}$ structures for 3'-dA:T and
177 3'-rA:T, which contain the Mg^{HiFi} site, but show no density for Mg^{2+} (Extended Data Fig. 6).
178 Previously, in the structure of ATP-dependent ligase from *Prochlorococcus marinus* (Pmar-
179 Lig), the phosphodiester bond formation referring to step 3 of ligation reaction has been
180 reported[50]. In B-form DNA, the sugar moieties usually adopt the 2'-endo sugar pucker
181 conformation, while in A-form DNA or RNA duplexes the sugar moieties usually adopt the 3'-
182 endo sugar pucker conformation[51]. The ligase active site forces the DNA upstream of the nick
183 to adopt an A-form conformation which causes the sugar at the 3'-end to assume the 3'-endo
184 conformation. Mg^{2+} can activate the 3'-OH for nucleophilic attack on the 5'-P and the 3'-end of
185 nick DNA can also stabilize the 3'-endo conformation. We suggest that the $LIG1^{EE/AA}$
186 structures for 3'-rG:C and 3'-rA:T that we captured in step 3 of the ligation reaction without
187 Mg^{2+} in the crystal conditions is due to the 3'-endo conformation of sugar pucker that
188 ribonucleotides naturally adopt.
189 Our study also contributes to the understanding of the mechanism by which LIG1 discriminates
190 against a wrong sugar at the 5'-end of nick DNA. This 5'-RNA-DNA junction mimics the lesion
191 could be inserted into DNA during RNase H2-dependent excision repair and trigger RNA-
192 DNA damage response as previously reported[45-48]. In the LIG1/5'-RNA-DNA structure, we
193 found a different closed state due to the conformational change of the DNA caused the
194 ribonucleotide at the 5'-end. The OBD domain of LIG1 normally binds in the minor groove
195 downstream of the nick DNA; however, in the LIG1/5'-rG:C structure, we observed a
196 narrowing of the minor groove which initializes a major global conformational change in LIG1.
197 Although it has been extensively reported that most DNA polymerases exhibit a "steric gate"
198 (*i.e.*, B-, and Y- family pols) or a protein backbone segment in minor groove nucleotide binding
199 pocket (*i.e*, X-family pols) which governs ribonucleotide exclusion[16-27], our study represents

200 the first solved structure of RNA-DNA nick complexes for a human DNA ligase. Overall, our
201 structures of LIG1 containing a single ribonucleotide at the 3'- or 5'-end of nick DNA contribute
202 to understanding the molecular mechanism which governs ribonucleotide selectivity by a
203 human ligase, LIG1 which functions in nuclear replication and repair. Future studies with other
204 human DNA ligases will be aimed at untangling the impact of improper sugar on the ligation
205 reaction to gain a deeper understanding of how fidelity of nick sealing is ensured to maintain
206 genome integrity.

207 **Methods**

208 **Protein purifications**

209 Recombinant proteins with 6x-his tag for the C-terminal (Δ261) wild-type and E346A/E592A
210 (EE/AA) mutant of human DNA ligase I (LIG1) were overexpressed in Rosetta (DE3) *E. coli*
211 cells in Terrific Broth (TB) media with kanamycin (50 $\mu g ml^{-1}$) and chloramphenicol
212 (34 $\mu g ml^{-1}$) at 37 °C[52-58]. The cells were induced with 0.5 mM isopropyl β-D-thiogalactoside
213 (IPTG) when the OD_{600} was reached to 1.0, and the overexpression was continued for overnight
214 at 28 °C. Cells were lysed in the lysis buffer containing 50 mM Tris-HCl (pH 7.0), 500 mM
215 NaCl, 20 mM imidazole, 10% glycerol, 1 mM PMSF, an EDTA-free protease inhibitor cocktail
216 tablet by sonication at 4 °C. The cell lysate was pelleted at 31,000 x g for 1 h at 4 °C. Proteins
217 were purified by a HisTrap HP column with an increasing imidazole concentration (20-300
218 mM) after being equilibrated in the binding buffer containing 50 mM Tris-HCl (pH 7.0),
219 500 mM NaCl, 20 mM imidazole, 10% glycerol at 4 °C. The collected fractions were
220 subsequently loaded onto a HiTrap Heparin column that was equilibrated with the binding
221 buffer containing 50 mM Tris-HCl (pH 7.0), 50 mM NaCl, 1.0 mM EDTA, and 10% glycerol,
222 and then eluted with a linear gradient of NaCl up to 1 M. LIG1 proteins were further purified
223 by Superdex 200 10/300 column in the buffer containing 20 mM Tris-HCl (pH 7.0), 200 mM

224 NaCl, and 1 mM DTT. All proteins concentrated, frozen in liquid nitrogen, and stored at -80
225 °C.

226 **Crystallization and structure determination**

227 Nick DNA substrates for LIG1 X-ray crystallography were prepared by annealing upstream,
228 downstream, and template primers (Extended Data Table 4). LIG1 C-terminal (Δ261) wild-
229 type ($LIG1^{WT}$) and EE/AA mutant ($LIG1^{EE/AA}$) proteins were pre-treated with EDTA before
230 using in the crystallization experiments in the absence of Mg^{2+} as we reported previously[36]. All
231 LIG1-nick DNA complex crystals were grown at 20 °C using the hanging drop method. LIG1
232 (at 27 $mgml^{-1}$)/DNA complex solution was prepared in 20 mM Tris-HCl (pH 7.0), 200 mM
233 NaCl, 1 mM DTT, 1 mM EDTA and 1 mM ATP at 1:4:1 DNA:protein molar ratio and then
234 mixed with 1 µl reservoir solution. The crystal conditions are described in Extended Data Table
235 5. Crystals were harvested and submerged in cryoprotectant solution containing reservoir
236 solution mixed with glycerol to a final concentration of 20% glycerol before being flash cooled
237 in liquid nitrogen for data collection (HKL Research, Inc). Crystals were kept at 100 °K during
238 X-ray diffraction data collection using the beamlines APS-22-ID and CHESS-7B2. All
239 structures were solved by the molecular replacement method using PHASER with PDB entry
240 7SUM as a search model[59]. Iterative rounds of model building were performed in COOT and
241 the final models were refined with PHENIX or REFMAC5[60-62]. 3DNA program was used for
242 sugar pucker analysis[63]. All structural images were drawn using PyMOL (The PyMOL
243 Molecular Graphics System, V0.99, Schrödinger, LLC). Detailed crystallographic statistics are
244 provided in Extended Data Table 1.

245 **DNA ligation assays**

246 Nick DNA substrates with a preinserted ribonucleotide at 3'- or 5'-end were prepared to
247 evaluate the ligation efficiency of LIG1 in the presence of 3'-rA:T, 3'-rC:G, 3'-rG:C or 5'-rG:C
248 (Extended Data Scheme 2). In addition, we tested the nick DNA substrates with cognate ends

249 3'-dA:T, 3'-dC:G, 3'-dG:C and 5'-dG:C (Extended Data Table 6). The reaction containing 50
250 mM Tris-HCl (pH: 7.5), 100 mM KCl, 10 mM $MgCl_2$, 1 mM ATP, 1 mM DTT, 100 μgml^{-1}
251 BSA, 10% glycerol, and nick DNA substrate (500 nM) was initiated by the addition of LIG1
252 (100 nM). The reaction samples were incubated at 37 °C, stopped by quenching with an equal
253 amount of the buffer containing 95% formamide, 20 mM ethylenediaminetetraacetic acid,
254 0.02% bromophenol blue and 0.02% xylene cyanol, and collected at the time points indicated
255 in the figure legends. The reaction products were then separated by electrophoresis on an 18%
256 denaturing polyacrylamide gel. The gels were scanned with a Typhoon Phosphor Imager
257 (Amersham Typhoon RGB), and the data were analyzed using ImageQuant software[52-58].

258 **Data availability**

259 Atomic coordinates and structure factors for the reported crystal structure of LIG1 in complex
260 with nick DNA complexes has been deposited in the RCSB Protein Data Bank under accession
261 numbers 8FYK, 8FZX, 8FZL, 8G1X, 8FZS, 8G0J, 8G1O, 8G11, 8G37. All relevant data are
262 available from the authors upon reasonable request.

263 **Funding**

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

266 **Acknowledgements**

267 This work is based upon research conducted at the Center for High Energy X-ray Sciences
268 (CHEXS), which is supported by the National Science Foundation under award DMR-
269 1829070, and the Macromolecular Diffraction at CHESS (MacCHESS) facility, which is
270 supported by award 1-P30-GM124166-01A1 from the National Institute of General Medical
271 Sciences, National Institutes of Health, and by New York State's Empire State Development
272 Corporation (NYSTAR). This research used resources of the Advanced Photon Source (APS),
273 U.S. Department of Energy (DOE) Office of Science user facility operated for the DOE Office

of Science by Argonne National Laboratory under Contract No. DE-AC02-06CH11357. The authors thank to Dr. Craig Vander Kooi (University of Florida) for his exceptional support with the APS beam time 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.

References

1. Ellenberger, T. & Tomkinson, A. E. Eukaryotic DNA ligases: Structural and functional insights. *Annu. Rev. Biochem*. **77**, 313-338 (2008).
2. Timson, D. J., Singleton, M. R. & Wigley, D. B. DNA ligases in the repair and replication of DNA. *Mutat. Res*. **460**, 301-318 (2000).
3. Tomkinson, A. E., Vijayakumar, S., Pascal, J. M. & Ellenberger, T. DNA ligases: structure, reaction mechanism, and function. *Chem. Rev*. **106**, 687-699 (2006).
4. Doherty, A. J. & Suh, S. W. Structural and mechanistic conservation in DNA ligases. *Nucleic Acids Res*. **28**, 4051-4058 (2000).
5. Çağlayan, M. Interplay between DNA polymerases and DNA ligases: Influence on substrate channeling and the fidelity of DNA ligation. *J. Mol. Biol*. **431**, 2068-2081 (2019).
6. Çağlayan, M. The ligation of pol β mismatch insertion products governs the formation of promutagenic base excision DNA repair intermediates. *Nucleic Acids Res*. **48**, 3708-3721 (2020).
7. Klein, H. L. Genome instabilities arising from ribonucleotides in DNA. *DNA Repair* **56**, 26-32 (2017).

298 8. Potenski, C. J. & Klein, H. L. How the misincorporation of ribonucleotides into
299 genomic DNA can be both harmful and helpful to cells. *Nucleic Acids Res*. **42**, 10226-
300 10234 (2014).
301 9. Nick McElhinny, S. A. et al. Genome instability due to ribonucleotide incorporation
302 into DNA. *Nat. Chem. Biol*. **6**, 774-781 (2010).
303 10. Yao, N. Y., Schroeder, J. W., Yurieva, O., Simmons, L. A. & O'Donnell, M. E. Cost
304 of rNTP/dNTP pool imbalance at the replication fork. *Proc. Natl. Acad. Sci*. **110**,
305 12942-12947 (2013).
306 11. Clausen, A. R., Zhang, S., Burgers, P. N., Lee, M. Y. & Kunkel, T. A. Ribonucleotide
307 incorporation, proofreading and bypass by human DNA polymerase delta. *DNA Repair*
308 **12**, 121-127 (2013).
309 12. Nick McElhinny, S. A. et al. Abundant ribonucleotide incorporation into DNA by yeast
310 replicative polymerases. *Proc. Natl. Acad. Sci. U. S. A*. **107**, 4949-4954 (2010).
311 13. Clausen, A. R. et al. Structure-function analysis of ribonucleotide bypass by B family
312 DNA replicases. *Proc. Natl. Acad. Sci. U.S.A.,* **110**, 16802-16807 (2013).
313 14. Clausen, A. R., Zhang, S., Burgers, P. N., Lee, M. Y. & Kunkel, T. A. Ribonucleotide
314 incorporation, proofreading and bypass by human DNA polymerase delta. *DNA Repair*
315 **12**, 121-127 (2013).
316 15. Gosavi, R. A., Moon, A. F., Kunkel, T. A., Pedersen, L. C. & Bebenek, K. The catalytic
317 cycle for ribonucleotide incorporation by human DNA pol lambda. *Nucleic Acids Res.*
318 **40**, 7518-7527 (2012).
319 16. Brown, J. A. et al. A novel mechanism of sugar selection utilized by a human X-family
320 DNA polymerase. *J. Mol. Biol*. **395**, 282-290 (2010).
321 17. Moon, A. F. et al. Structural accommodation of ribonucleotide incorporation by the
322 DNA repair enzyme polymerase μ. *Nucleic Acids Res*. **45**, 9138-9148 (2017).

323 18. Cavanaugh, N. A., Beard, W. A. & Wilson, S. H. DNA polymerase β ribonucleotide
324 discrimination. *J. Biol. Chem*. **285**, 24457-24465 (2010).
325 19. Kasiviswanathan, R. & Copeland, W. C. Ribonucleotide discrimination and reverse
326 transcription by the human mitochondrial DNA polymerase. *J. Biol. Chem*. **286**, 31490-
327 31500 (2011).
328 20. Joyce, C. M. Choosing the right sugar: how polymerases select a nucleotide substrate.
329 *Proc. Natl. Acad. Sci. U. S. A.* **94**, 1619-1622 (1997).
330 21. Brown, J. A. & Suo, Z. Unlocking the sugar steric gate of DNA polymerases.
331 *Biochemistry* **50**, 1135-1142 (2011).
332 22. Donigan, K. A., McLenigan, M. P., Yang, W., Goodman, M. F. & Woodgate, R. The
333 steric gate of DNA polymerase iota regulates ribonucleotide incorporation and
334 deoxyribonucleotide fidelity. *J. Biol. Chem*. **289**, 9136-9145 (2014).
335 23. DeLucia, A. M., Grindley, N. D. & Joyce, C. M. An error-prone family Y DNA
336 polymerase (DinB homolog from Sulfolobus solfataricus) uses a 'steric gate' residue
337 for dis- crimination against ribonucleotides. *Nucleic Acids Res*. **31**, 4129-4137 (2003).
338 24. Brown, J. A. & Suo, Z. Unlocking the sugar 'steric gate' of DNA polymerases.
339 *Biochemistry* **50**, 1135-1142 (2011).
340 25. Ruiz, J. F. et al. Lack of sugar discrimination by human pol μ requires a single glycine
341 residue. *Nucleic Acids Res*. **31**, 4441-4449 (2014).
342 26. Kasiviswanathan, R. & Copeland, W. C. Ribonucleotide discrimination and reverse
343 transcription by the human mitochondrial DNA polymerase. *J. Biol. Chem*. **286**, 31490-
344 31500 (2011).
345 27. Vaisman, A. & Woodgate, R. Ribonucleotide discrimination by translesion synthesis
346 DNA polymerases. *Crit. Rev. Biochem. Mol. Biol*. **53**, 382-402 (2018).
347 28. Caldecott, K. W. Ribose-an internal threat to DNA. *Science* **343**, 260-261 (2014).

348 29. Hovatter, K. R. & Martinson, H. G. Ribonucleotide-induced helical alteration in DNA
349 prevents nucleosome formation. *Proc. Natl. Acad. Sci. U. S. A*. **84**, 1162-1166 (1987).
350 30. Conover, H. N., Lujan, S. A., Chapman, M. J., Cornelio, D. A., Sharif, R., Williams, J.
351 S., Clark, A. B., Camilo, F., Kunkel, T. A. & Argueso, J. L. Stimulation of
352 chromosomal re- arrangements by ribonucleotides. *Genetics* **201**, 951-961 (2015).
353 31. Huang, S. N., Williams, J. S., Arana, M. E., Kunkel, T. A. & Pommier, Y.
354 Topoisomerase I-mediated cleavage at unrepaired ribonucleotides generates DNA
355 double-strand breaks. *EMBO J*. **36**, 361-373 (2016).
356 32. Pascal, J. M. et al. Human DNA ligase I completely encircles and partially unwinds
357 nicked DNA. *Nature* **432**, 473-478 (2004).
358 33. Tumbale, P. P. et al. Two-tiered enforcement of high-fidelity DNA ligation. *Nat.*
359 *Commun.* **10**, 5431 (2019).
360 34. Williams, J. S. et al. High-fidelity DNA ligation enforces accurate Okazaki fragment
361 maturation during DNA replication. *Nat. Commun*. **12**, 482 (2021).
362 35. Jurkiw, T. J. *et al.* LIG1 syndrome mutations remodel a cooperative network of ligand
363 binding interactions to compromise ligation efficiency. *Nucleic Acids Res*. **49**, 1619-
364 1630 (2021).
365 36. Tang, Q., Gulkis, M., McKenna, R. & Çağlayan, M. Structures of LIG1 that engage
366 with mutagenic mismatches inserted by polβ in base excision repair. *Nat. Commun*. **13**,
367 3860 (2022).
368 37. Luo, J. & Barany, F. Identification of essential residues in *Thermus thermophilus* DNA
369 ligase. *Nucleic Acids Res*. **24**, 3079-3085 (1996).
370 38. Tomkinson, A. E., Tappe, N. J. & Friedberg, E. C. DNA ligase I from *Saccharomyces*
371 *cerevisiae*: physical and biochemical characterization of the CDC9 gene product.
372 *Biochemistry* **31**, 11762-11771 (1992).

373 39. Sriskanda, V. & Shuman, S. Specificity and fidelity of strand joining by *Chlorella virus*
374 DNA ligase. *Nucleic Acids Res*. **26**, 3536-3541 (1998).
375 40. Shuman, S. Vaccinia virus DNA ligase: specificity, fidelity, and inhibition.
376 *Biochemistry* **34**, 16138-16147 (1995).
377 41. Bhagwat, A. S., Sanderson, R. J. & Lindahl, T. Delayed DNA joining at 3'mismatches
378 by human DNA ligases. *Nucleic Acids Res*. **27**, 4028-4033 (1999).
379 42. Sparks, J. L. *et al.* RNase H2-initiated ribonucleotide excision repair. *Mol. Cell* **47**, 980-
380 986 (2012).
381 43. Cerritelli, S. M. & Crouch, R. J. Ribonuclease H: the enzymes in eukaryotes. *FEBS J.*
382 **276**, 1494-1505 (2009).
383 44. Hiller, B., Achleitner, M., Glage, S., Naumann, R., Behrendt, R. & Roers, A.
384 Mammalian RNase H2 removes ribonucleotides from DNA to maintain genome
385 integrity. *J. Exp. Med*. **209**, 1419-1426 (2012).
386 45. Williams, J. S. & Kunkel, T. A. Ribonucleotides in DNA: Origins, repair and
387 consequences. *DNA Repair* **19**, 27-37 (2014).
388 46. Williams, J. S., Lujan, S. A. & Kunkel, T. A. Processing ribonucleotides incorporated
389 during eukaryotic DNA replication. *Nat. Rev. Mol. Cell Biol.* **17**, 350-363 (2016).
390 47. Tumbale, P., Williams, J. S., Schellenberg, M. J., Kunkel, T. A. & Williams, R. S.
391 Aprataxin resolves adenylated RNA-DNA junctions to maintain genome integrity.
392 *Nature* **506**, 111-115 (2014).
393 48. Wallace, B. D. & Williams, R. S. Ribonucleotide triggered DNA damage and RNA-
394 DNA damage responses. *RNA biology* **11**, 1340-1346 (2014).
395 49. Kaminski, A. et al. Structures of DNA-bound human ligase IV catalytic core reveal
396 insights into substrate binding and catalysis. *Nat. Commun*. **9**, 2642 (2018).

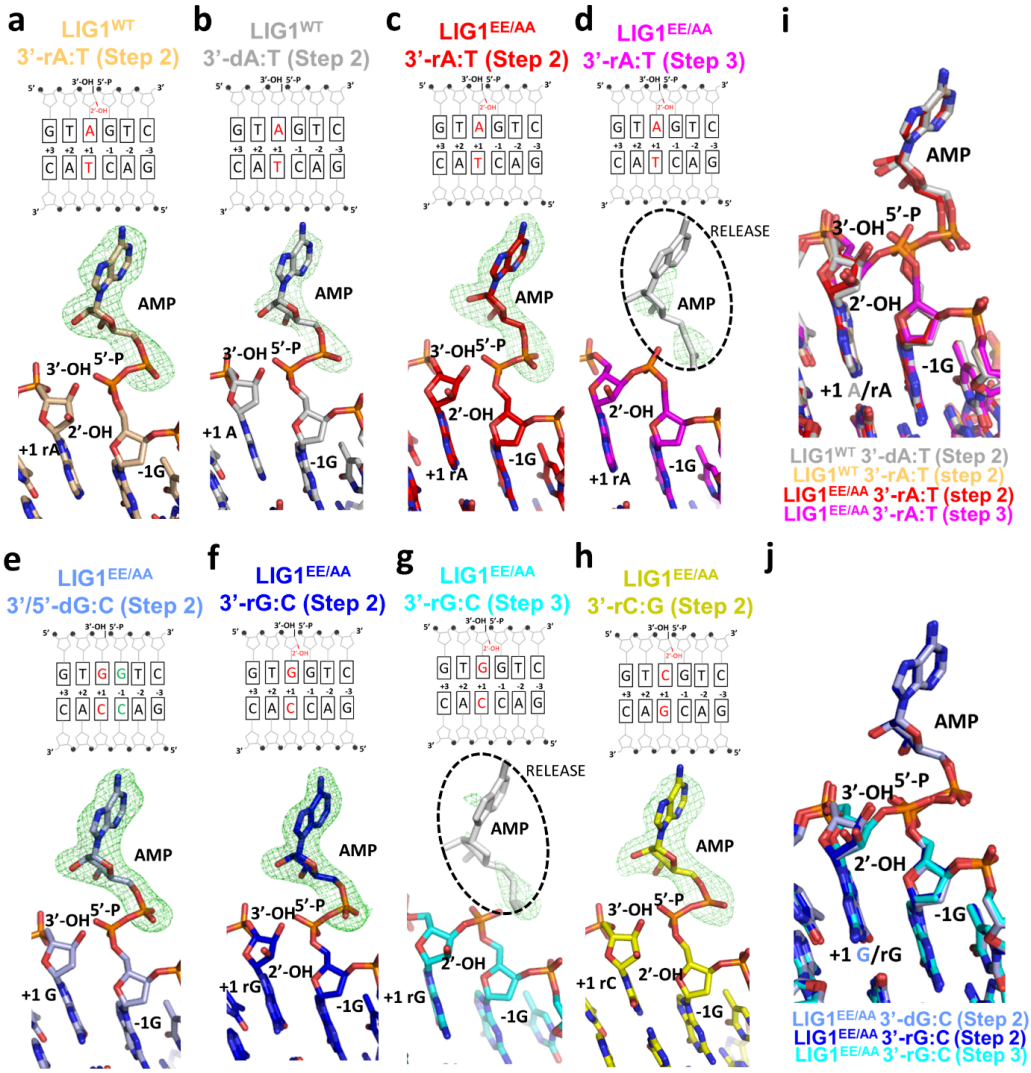
397 50. Williamson, A. & Leiros, H. K. S. Structural intermediates of a DNA-ligase complex
398 illuminate the role of the catalytic metal ion and mechanism of phosphodiester bond
399 formation. *Nucleic Acids Res*. **47**, 7147-7162 (2019).
400 51. Dejaegere, A. P. & Case, D. A. Density functional study of ribose and deoxyribose
401 chemical shifts. *J. Phys. Chem*. **102**, 5280-5289 (1998).
402 52. Çağlayan, M., Batra, V. K., Sassa, A., Prasad, R. & Wilson, S. H. Role of polymerase
403 β in complementing aprataxin deficiency during abasic-site base excision repair. *Nat.*
404 *Struct. Mol. Biol.* **21**, 497-499 (2014).
405 53. Çağlayan, M. et al. Complementation of aprataxin deficiency by base excision repair
406 enzymes. *Nucleic Acids Res*. **43**, 2271-2281 (2015).
407 54. Çağlayan, M. et al. Complementation of aprataxin deficiency by base excision repair
408 enzymes in mitochondrial extracts. *Nucleic Acids Res.* **45**, 10079-10088 (2017).
409 55. Çağlayan, M. et al. Oxidized nucleotide insertion by pol β confounds ligation during
410 base excision repair. *Nat. Commun*. **8**, 14045 (2017).
411 56. Çağlayan, M. Pol β gap filling, DNA ligation and substrate-product channeling during
412 base excision repair opposite oxidized 5-methylcytosine modifications. *DNA Repair*
413 **95**, 102945 (2020).
414 57. Tang, Q., Kamble, P. & Çağlayan, M. DNA ligase I variants fail in the ligation of
415 mutagenic repair intermediates with mismatches and oxidative DNA damage.
416 *Mutagenesis* **35**, 391-404 (2020).
417 58. Kamble, P., Hall, K., Chandak, M., Tang, Q. & Çağlayan, M. DNA ligase I fidelity the
418 mutagenic ligation of pol β oxidized and mismatch nucleotide insertion products in base
419 excision repair. *J. Biol. Chem*. **296**, 100427 (2021).
420 59. McCoy, A. J. et al. Phaser crystallographic software. *J. Appl. Crystallogr.* **40**, 658-674
421 (2007).

60. Emsley, P. et al. Features and development of Coot. *Acta Crystallogr. D. Biol. Crystallogr.* **66**, 486-501 (2010).

61. Adams, P. D. et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D. Biol. Crystallogr.* **66**, 213-221 (2010).

62. Murshudov, G. N. et al. REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallogr. D. Biol. Crystallogr.* **67**, 355-367 (2011).

63. Li, S., Olson, W. K. & Lu, X. J. Web 3DNA 2.0 for the analysis, visualization, and modeling of 3D nucleic acid structures. *Nucleic Acids Res*. **47**, W26-W34 (2019).



433 **Fig. 1. Structures of LIG1 bound to nick DNA duplexes containing 3'-ribonucleotide.**
434 Structures of $LIG1^{WT}$ and $LIG1^{EE/AA}$ in complex with nick DNA containing ribonucleotide or
435 deoxyribonucleotide at the 3'-end are presented for $LIG1^{WT}$/3'-rA:T (step 2, **a**), $LIG1^{WT}$/3'-
436 dA:T (step 2, **b**), $LIG1^{EE/AA}$/3'-rA:T (step 2, **c**), $LIG1^{EE/AA}$/3'-rA:T (step 3, **d**), $LIG1^{EE/AA}$/3'-
437 dG:C (step 2, **e**), $LIG1^{EE/AA}$/3'-rG:C (step 2, **f**), $LIG1^{EE/AA}$/3'-rG:C (step 3, **g**), $LIG1^{EE/AA}$/3'-
438 rC:G (step 2, **h**). Simulated annealing omit maps (Fo-Fc) of AMP are contoured at 3σ.
439 Schematic view of the nick DNA shows the sequence of 3'- and 5'- terminus of the nick DNA
440 substrate used in the LIG1 crystallization. In the step 2 structures, AMPs are built depending
441 on the map, while in the step 3 structures, the AMPs do not exist because the maps of AMPs
442 are incomplete. **i-j**, Differences in the positions of 3'- and 5'- terminus of the nick DNA are
443 presented in the overlays of $LIG1^{WT}$ and $LIG1^{EE/AA}$ structures for the nick DNA duplexes
444 containing 3'-dA:T and 3'-rA:T (i) and $LIG1^{EE/AA}$ structures for the nick DNA duplexes
445 containing 3'-dG:C and 3'-rG:C (j).

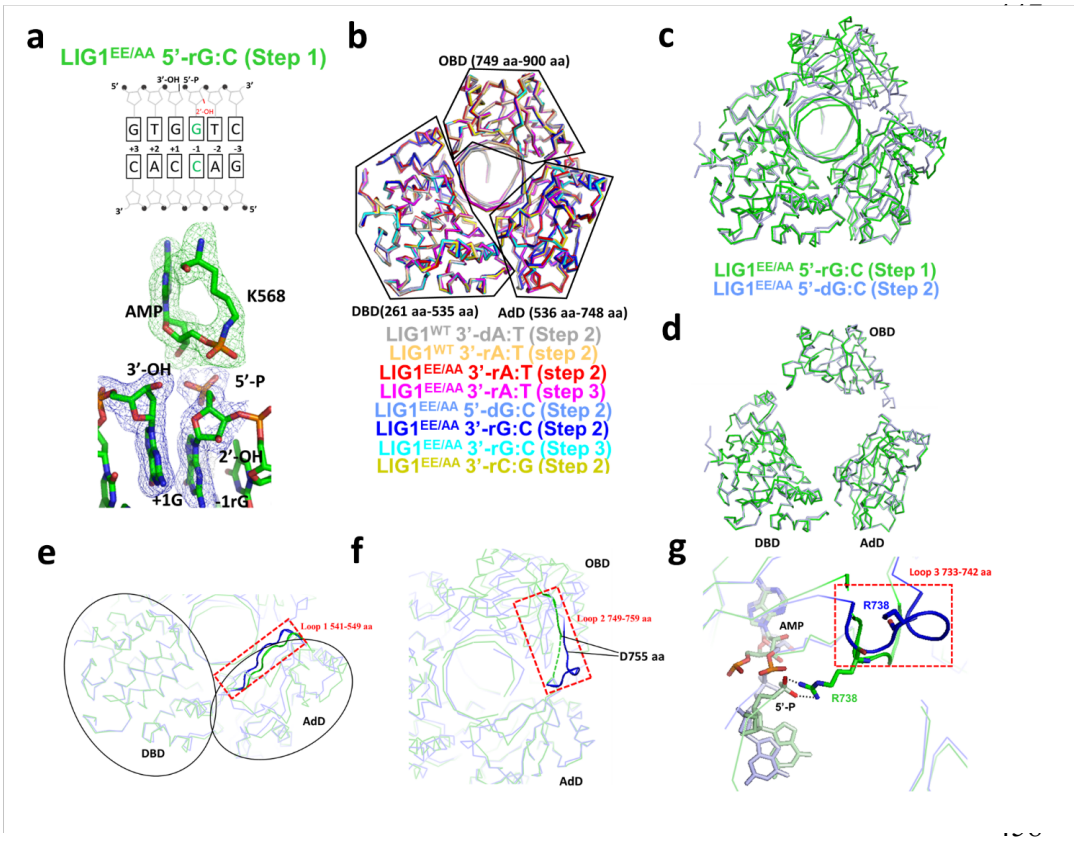
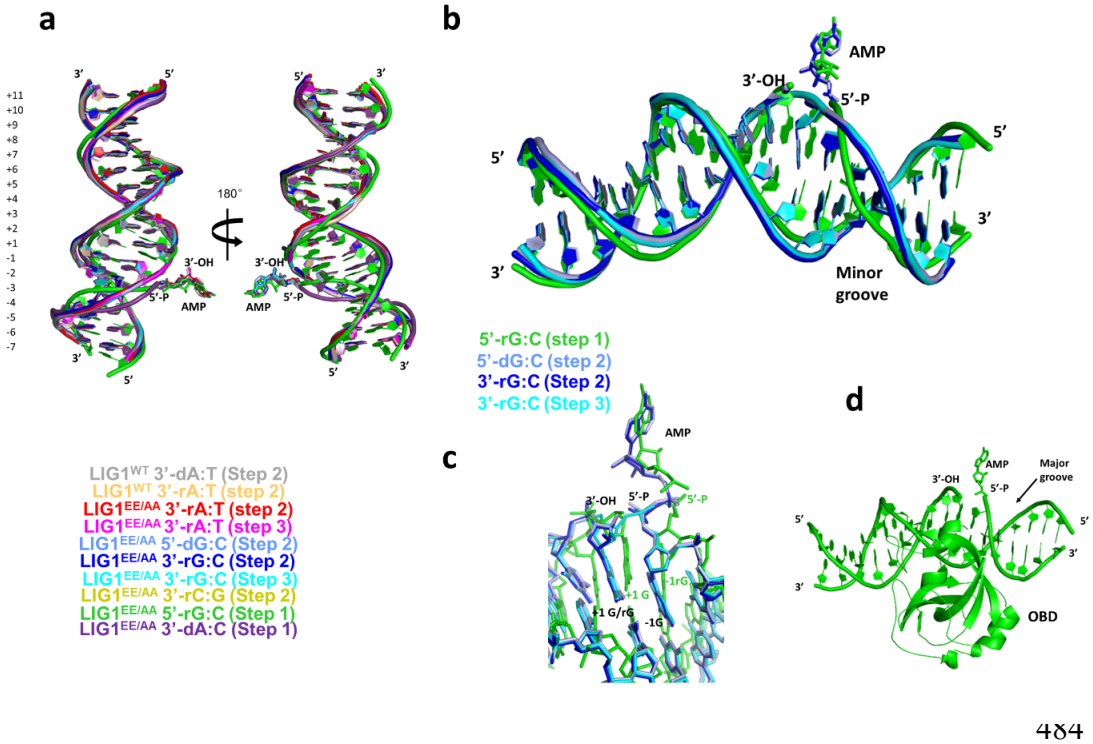


Fig. 2. Structures of LIG1 bound to nick DNA duplexes containing 5'-ribonucleotide. a, Structure of LIG1$^{EE/AA}$/5'-rG:C (LIG1^{K568}-AMP intermediate, step 1). Schematic view of the nick DNA shows the sequence of 3'- and 5'-terminus of the nick DNA substrate. Simulated annealing omit map (Fo-Fc) of AMP is contoured at 3σ (green). 2Fo - Fc density map of 3'- and 5'- terminus of the nick DNAs are contoured at 1σ (blue). **b**, Overlay of LIG1^{WT} and LIG1$^{EE/AA}$ structures for the nick DNA duplexes containing ribonucleotide or deoxyribonucleotide at the 3'-end are presented for the DBD, AdD and OBD domains of LIG1 encircling a nick DNA. **c-g**, Overlays of LIG1$^{EE/AA}$ structures for 5'-rG:C and 5'-dG:C show overall (c) and individual ligase domains (d). The conformational changes in the loop 1 (red circle) residing in the AdD (e), and the loop 2 (red circle) residing in the OBD (f) domains. **g**, Conformational change at the loop 3 (red circle) shows a shift at the R738 side chain that interacts with 5'-P end of nick in the structure of LIG1$^{EE/AA}$/5'-rG:C. The amino acid residue

471 R738, AMP, and nick DNA are shown as a stick mode and all LIG1 are shown as a ribbon

472 mode.

473



484

485 **Fig. 3. Structure of LIG1/nick DNA with a 5'-ribonucleotide uncovers a conformational**

486 **change the minor groove. a**, Superimposition of overall LIG1^{WT} and LIG1$^{EE/AA}$ structures

487 solved in this study shows a difference in the DNA conformation in LIG1$^{EE/AA}$/5'-rG:C

488 structure (green). **b**, Comparison of the nick DNA between LIG1$^{EE/AA}$ structures captured in

489 different steps of the ligation reaction; 5'rG:C (step 1), 3'-dA:C (step 1, 7SX5), 3'-dG:C (step

490 2), and 3'-rG:C (step 3); demonstrates a rotation of the 5'-P downstream of the nick DNA which

491 widens the major groove downstream of the nick DNA and results in a narrowing of the minor

492 groove in LIG1$^{EE/AA}$/5'-rG:C structure. **c**, Overlay of LIG1$^{EE/AA}$ structures for 3'-dG:C, 3'-rG:C,

493 3'-rG:C, and 5'-rG:C shows the conformation change at the 3'- and 5'- terminus of the nick

494 DNA. **d**, Structure of LIG1$^{EE/AA}$/5'-rG:C shows the conformation change in the OBD domain

495 of the ligase interacting with nick DNA. OBD and nick DNA are shown as cartoon mode.

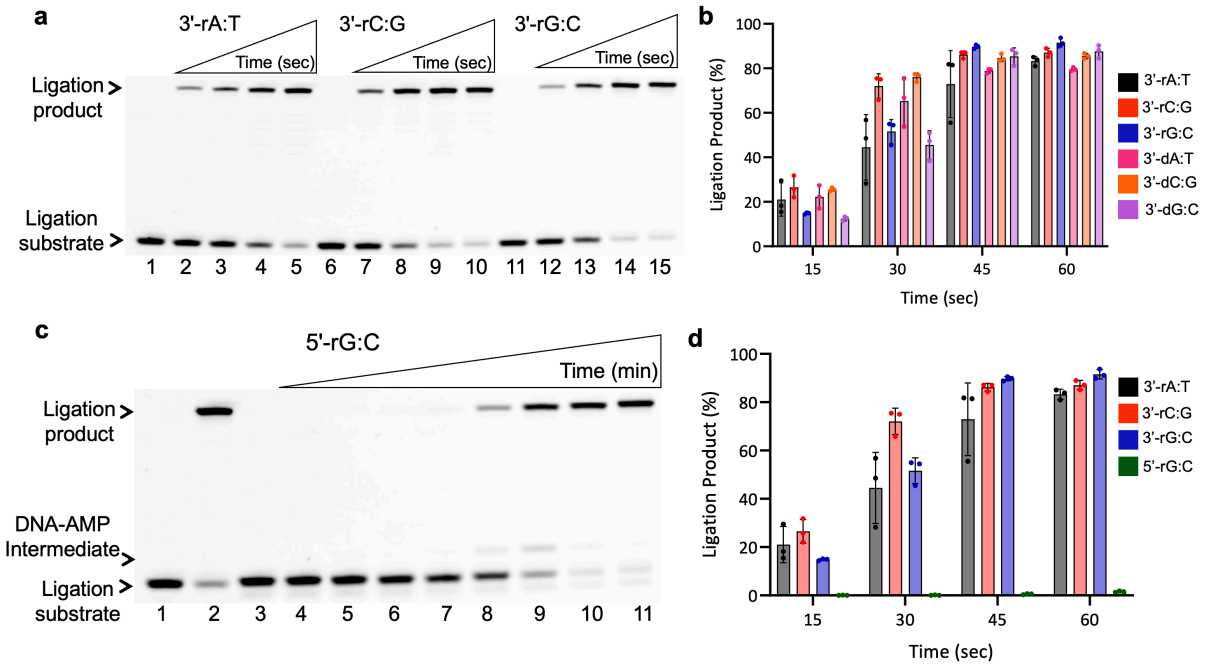


Fig. 4. Ligation of the nick DNA with a ribonucleotide at 3'- or 5'-end by LIG1. a, Lanes 1, 6 and 11 are the negative enzyme controls of the nick DNA substrates with 3'-rA:T, 3'-rC:G, and 3'-rG:C, respectively. Lanes 2-5, 7-10, and 12-15 are the ligation products in the presence of 3'-rA:T, 3'-rC:G, and 3'-rG:C, respectively, and correspond to time points of 10, 30, 45, and 60 sec. **b,** Graph shows time-dependent change in the amount of ligation products for nick DNA substrates containing 3'-ribonucleotides *versus* 3'-deoxyribonucleotides. The data represent the average from three independent experiments ± SD. **c,** Line 1 is the negative enzyme control of the nick DNA substrate with 5'-dG:C, and line 2 is the positive control showing the ligation product in the presence of 5'-dG:C. Line 3 is the negative enzyme control of the nick DNA substrate with 5'-rG:C. Lanes 4-11 are the reaction products in the presence of 5'-rG:C, and correspond to time points of 0.25, 0.5, 0.75, 1, 5, 15, 30, and 60 min. **d,** Graph shows time-dependent change in the amount of ligation products for nick DNA substrates containing 3'-ribonucleotides *versus* 5'-rG:C. The data represent the average from three independent experiments ± SD.