Mechanism and Specificity of Human DNA Ligase I

by

Thomas J. Jurkiw

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Biological Chemistry)
in the University of Michigan
2019

Doctoral Committee:

Associate Professor Patrick J. O'Brien, Chair
Associate Professor Bruce A. Palfey
Professor Janet L. Smith
Associate Professor Raymond C. Trievel
Professor Thomas E. Wilson

Thomas J. Jurkiw

tjurkiw@umich.edu

ORCID iD: 0000-0001-9624-4568

Table of Contents

List of Tables

List of Figures

Abstract

Cellular DNA replication and repair pathways conclude with faithful rejoining of broken phosphodiester bonds by DNA ligases. Therefore, a thorough dissection of the molecular underpinnings of DNA ligation is crucial to a comprehensive understanding of genomic maintenance. To this end, we performed biochemical and biophysical studies detailing the interactions occurring between human DNA ligase 1 (LIG1) and DNA throughout ligation. Our study led to the identification and preliminary characterization of two distinct conformational changes preceding ligation. We posit that the observed steps are distortions of the two DNA ends of the nick site, induced by LIG1 to properly align the ends for ligation. Using crystallographic data, we identified two unique sets of interactions between the DNA ends and LIG1. One such interaction is mediated through a conserved metal-binding site, which our data suggests influences the ability of LIG1 to ligate damaged or mismatched upstream ends. At the downstream end of the nick, we identified a phenylalanine residue positioned near the sugar moiety of the 5'-deoxyribose. Mutagenesis of the residue renders LIG1 unable to discriminate between a downstream deoxyribonucleotide and a downstream ribonucleotide, demonstrating its role in preventing ligation of DNA-RNA hybrids. Both mechanisms appear to influence fidelity by restricting flexibility of the downstream and upstream DNA ends. To understand how positioning of the ends contributes to ligation, we mutagenized conserved residues in the enzyme active site that were identified as putative metal ligands. Biochemical studies of the mutated residues uncovered a cooperative network of magnesium-mediated interactions between the three

conserved active site residues and the DNA substrate. A biochemical characterization of two LIG1 mutations, both identified in patients suffering from broad immunological disorders, illustrates how disruption of this network of interactions can impact human health. Our study reveals that the patient-derived mutations have diminished magnesium affinity, elevating the abortive ligation burden and likely causing significant delays in DNA replication and repair. The work detailed in this thesis significantly expands our knowledge of the complex interactions between DNA ligases and their substrates, furthering our understanding of how DNA ligation fits into the larger context of DNA replication and repair.

Chapter 1
Introduction

First evidenced by the work of Rosalind Franklin[1,2], the structure of the DNA double helix is one of the most well-known scientific discoveries, setting the basis for our understanding of the genetic code. Although DNA is often considered as a stable macromolecule, genomic DNA is susceptible to damage from endogenous and exogenous sources, challenging the cell to constantly maintain genomic integrity. As a consequence of this, tens of thousands of breaks are made throughout the genome[3], which, if allowed to persist, can become mutagenic and likely cytotoxic. Much of molecular biology attempts to address how the genome is maintained despite such challenges, leading to discoveries of conserved DNA repair mechanisms throughout evolution. One such discovery was that of the DNA ligase enzyme[4-7]. First isolated from bacteriophage-infected *E. coli* cells, DNA ligase was shown to be capable of joining adjacent DNA ends and, therefore, assumed to be essential for repairing DNA breaks caused by cellular DNA replication and repair pathways. DNA ligases quickly became ubiquitous in molecular cloning techniques, appreciated for their versatile ability to stitch DNA strands back together. Even as a key enzyme in important cellular pathways and in modern-day biotechnology, much about DNA ligases remains largely unknown. The work documented here serves to fill gaps in our knowledge through extensive study of the structure-function relationship in DNA ligase I (LIG1), the main replicative human DNA ligase. My thesis broadly explores substrate binding and recognition by LIG1, addressing how the enzyme avoids ligation of damaged and mismatched ends, and explores how ligase dysfunction can lead to immunological disease.

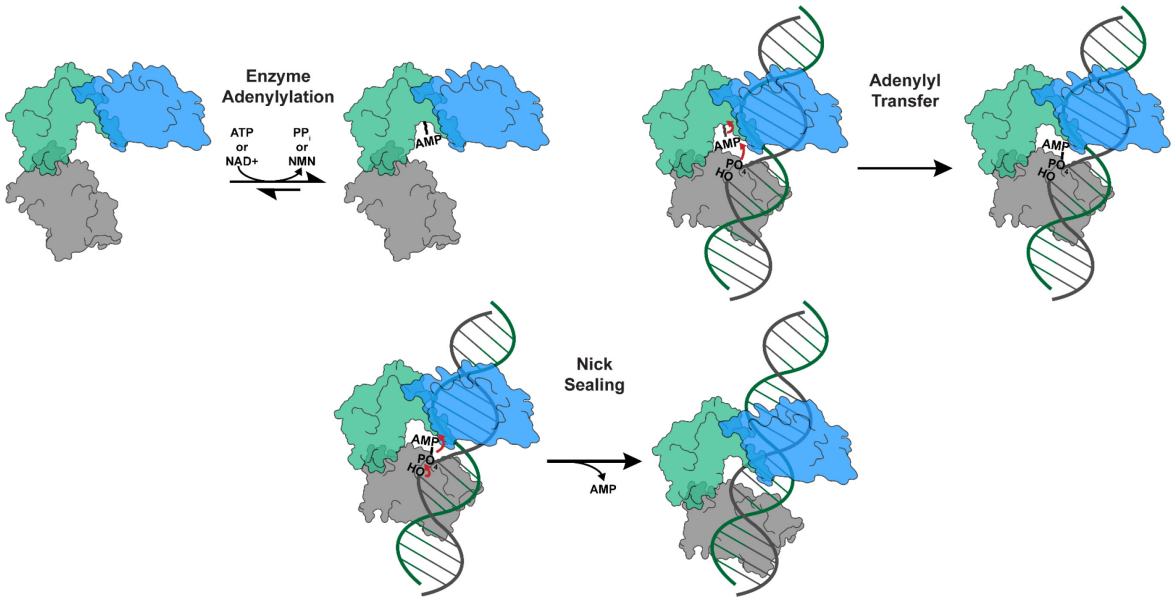


Figure 1.1. Mechanism of DNA ligation. Preceding nick-binding, DNA ligase binds to and utilizes ATP (or NAD+) to form a covalent lysine-AMP intermediate. Following enzyme adenylylation, the DNA ligase is able to bind to DNA and search for breaks in the helix. Once bound to a site of nicked DNA, the AMP group is transferred to the downstream 5'-PO_4, generating an activated DNA end. In the final step of ligation, the upstream 3'-OH attacks the 5'-PO_4, releasing the AMP group and the sealed DNA.

DNA Ligase Function and Structure

Mechanism of DNA Ligation

DNA ligases were first discovered in 1967 as a component of the T4 bacteriophage proteome capable of catalyzing the covalent linkage of DNA strands in *E. coli* cells[4]. Soon after, the *E. coli* homolog was isolated and characterized[8], marking DNA ligases as a common component in both bacteria and viruses. The discovery of T4 and *E. coli* DNA ligases also uncovered evidence of evolutionary divergence, as ligation by T4 DNA ligase is dependent on ATP[4,6], whereas the *E. coli* ligase requires NAD+ for catalytic activity[8,9]. To date, no eukaryotic DNA ligase has been identified to be NAD+-dependent; only prokaryotic and some viral ligases utilize NAD+, with several able to utilize both cofactors[10].

Both ATP and NAD+ are utilized for enzyme adenylylation, where a covalent bond is formed between AMP and the N-ζ of an active site lysine (Figure 1.1). The enzyme-adenylylation step is critical for stimulating the nick-recognition ability of DNA ligases[11-13], although there is evidence pointing to the ability of deadenylylated T4 DNA ligase to bind nicked DNA[14]. The exact mechanism for nick recognition remains unknown, but structural studies have posited that a conformational change triggered by enzyme adenylylation is required to prime the enzyme for DNA-binding[13,15]. A similar conformational change has been documented in mRNA capping enzymes, where a large conformational change after enzyme guanylylation precedes RNA binding [16,17]. Upon nick recognition, the adenylyl group is transferred from the enzyme to the 5'-PO_4 on the downstream strand, activating it for attack by the upstream 3'-OH.

The catalytic activity of DNA ligases, both ATP- and NAD+-dependent, is largely attributed to residues in the conserved nucleotidyl transferase (NTase), also referred to as the adenylylation domain (AdD), and oligonucleotide-binding fold (OB-fold) domains. The two domains—collectively referred to as the catalytic core—are also conserved in mRNA capping enzymes[18,19]. Structural and sequence similarities between DNA ligases and mRNA capping enzymes led to the identification of five motifs in the catalytic core, referred to as motif I, III, IIIa, IV and V, required for ligation or capping activity[20-22]. An additional motif, motif VI, is found in the OB-fold of both ATP-dependent DNA ligases and mRNA capping enzymes. Residues in motif VI provide critical contacts with the β- and γ-phosphates of bound NTPs, stabilizing the pyrophosphate leaving group prior to enzyme adenylylation[20-22]. Consistent with this, motif VI is notably absent from the NAD+-dependent ligases. Crystal structures of both NAD+- and ATP-dependent DNA ligases have revealed a conserved mechanism among DNA

ligases of distorting an upstream region of the DNA helix into a more A-form-like structure, presumably to allow for better positioning of the nick in the enzyme active site[13,15,23-26]

Domain Architecture of NAD+-Dependent Ligases

The structure of NAD+-dependent ligases is best exemplified by the *E. coli* DNA ligase, LigA[23] (Figure 1.3). On either side of the catalytic core, LigA contains structural elements unique to prokaryotic DNA ligases; most notable is the Ia domain located near the N-terminus. Conserved in all NAD+-dependent ligases, the Ia domain is required for binding NAD+ and subsequent formation of the enzyme-adenylylate intermediate[27,28]. However, the Ia domain has been shown to be dispensable for phosphodiester bond synthesis. Indeed, a crystal structures of LigA revealed contacts between the Ia domain and the NMN moiety of NAD+ but not the DNA substrate[23,29].

Downstream of the LigA catalytic core lie the zinc-finger (ZnF), helix-hairpin-helix (HhH) and BRCT domains. Despite not making extensive contacts with the DNA substrate[23], mutations of conserved residues in the ZnF domain of *E. coli* LigA render the enzyme unable to complement DNA ligase-deficient yeast cells and significantly decreases *in vitro* activity[30,31]. Positioned between the OB-fold and HhH domains, the ZnF domain most likely is essential for bridging these two domains and for supporting enzyme stability. The HhH domain, consisting of tandem HhH motifs, functions as a scaffold for the DNA, making numerous contacts along the minor grooves of the helix[23]. Additionally, hydrogen-bonding interactions made between the

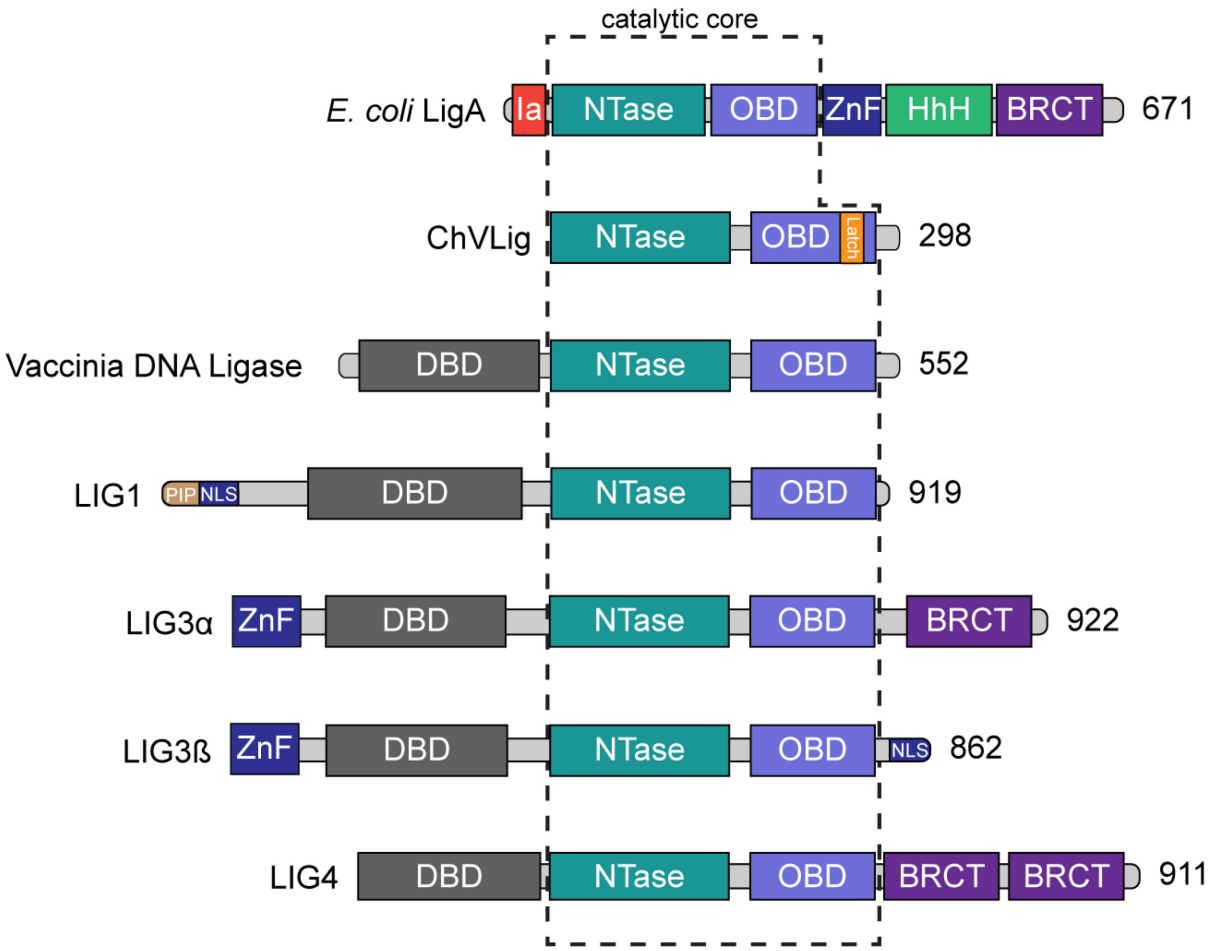


Figure 1.2. Domain architecture of DNA ligases. DNA ligases share a common catalytic core composed of a nucleotidyl-transferase (NTase) and an oligonucleotide-binding fold (OB-fold) domain. *E. coli* LigA represents the prototypical NAD+-dependent DNA ligase, consisting of an N-terminal Ia domain, as well as a zinc-finger (ZnF), helix-hairpin-helix (HhH) and BRCT domain. The DNA ligase from the Chlorella virus (ChVLig) is the smallest of the ATP-dependent DNA ligases known, consisting of the conserved catalytic core and a region in the OB-fold known as the latch region. The latch domain is notably absent in the Vaccinia virus DNA ligase, which instead makes use of an N-terminal DNA-binding domain (DBD). Humans have three families of ATP-dependent DNA ligases—DNA ligase I, III and IV—that all contain a DNA-binding domain towards the N-terminus. DNA ligase I (LIG1) is the simplest of the three ligases with the N-terminal region containing both a nuclear-localization signal (NLS) and a PCNA interaction protein motif (commonly known as a PIP box). DNA ligase III (LIG3) has a more complex architecture, containing an N-terminal ZNF domain and—depending on the splicing isoform—either a BRCT domain (LIG3α) or an NLS (LIG3β) towards the C-terminus. Additional forms of both LIG3α and LIG3β containing a mitochondrial localization sequence (MLS) also exist but are not shown. DNA ligase IV (LIG4) mimics LIG1 with the exception of tandem BRCT domains located near the C-terminus.

HhH and NTase domains appear to help LigA encircle the DNA helix (Figure 1.3). The role of the LigA BRCT domain remains unknown, partially due to the lack of structural information[23,32]. Studies have demonstrated that the BRCT domain contributes to *in vitro* ligation activity and DNA-binding[33], but the mechanism by which it affects both remains unclear. BRCT domains are

often implicated in facilitating protein-protein interactions[34], suggesting that the domain may also play a larger role in how LigA interacts with cellular machinery.

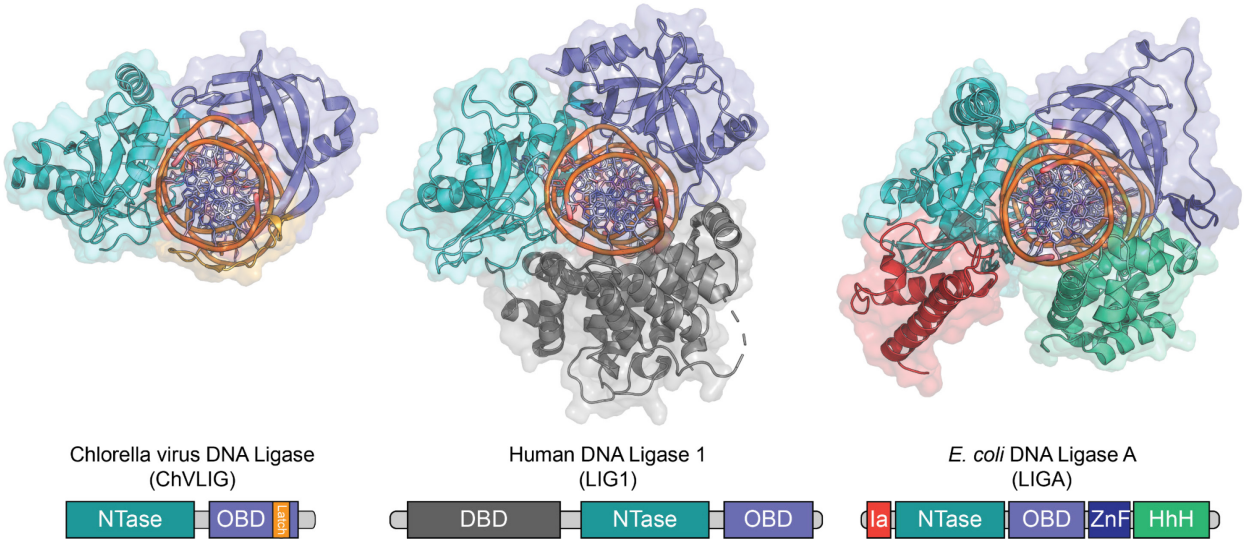


Figure 1.3. DNA ligases encircle the DNA double helix. Crystal structures of ChVLig (left), human LIG1 (center) and *E. coli* LigA (right). All enzymes are complexed with a nicked DNA double helix. The individual domains are color-coded according to the schematics underneath each enzyme structure. Both the N-terminal region of LIG1 and the BRCT domain of LigA were not present in the original X-ray crystallographic experiments. Crystal structure coordinates were obtained from the Protein Data Bank[35] using the following accession IDs: 2Q2T[13] (ChVLig), 1X9N[15] (LIG1) and 2OWO[23] (LigA).

Domain Architecture of ATP-Dependent Ligases

Insight into the biochemistry of ATP-dependent ligases relies heavily on the study of viral DNA ligases, most notably from the Chlorella and Vaccinia viruses. The Chlorella virus DNA ligase, abbreviated as ChVLig, represents the simplest domain architecture among the ATP-dependent ligases[19,36,37]. ChVLig contains the canonical DNA ligase catalytic core but is devoid of any larger structured domains. Early crystal structures of ChVLig in the absence of DNA resemble structures of mRNA capping enzymes, albeit for ~22 amino acids within the OB-fold domain that were not resolved in the early ChVLig structures[12,38]. These residues were previously shown to be protected by limited proteolysis when bound to nicked DNA, suggesting a role in DNA-binding[39]. Subsequent structures of ChVLig•DNA complexes revealed that this unique region—known as the "latch" region—contacts both the DNA substrate and the N-

terminal NTase domain, facilitating the encircling of the DNA substrate by ChVLig[13](Figure 1.3). The latch region is absent from the Vaccinia virus DNA ligase, which instead has an N-terminal DNA-binding domain (DBD)[40] similar to those found in the DNA ligases of higher eukaryotes[36,37]. Although dispensable for the chemistry of nick ligation, the DBD of the Vaccinia virus DNA ligase contributes to its ability to bind to a nicked DNA substrate[40].

Structural Divergence in Mammalian DNA Ligases

Extensive study of viral DNA ligases has given way to the study of the DNA ligases in higher eukaryotes. All higher eukaryotic DNA ligases utilize a large N-terminal DNA-binding domain (DBD) to aide in binding a DNA substrate[37]. The DBD interacts with the C-terminal OB-fold domain, helping to bridge the two domains and encircle the DNA substrate[15,24,25]. While eukaryotic DNA ligases mostly resemble their viral homologs, eukaryotes have evolved to utilize at least two different DNA ligases, while some eukaryotes, such as vertebrates, have a third DNA ligase that is critical for mitochondrial DNA replication.

The three DNA ligase families in mammalian cells are DNA ligase I (LIG1), III (LIG3) and IV (LIG4). While all three of these DNA ligases share the conserved catalytic core and the presence of a variable DBD, the N- and C-terminal regions contain elements that define specific cellular roles. LIG1 represents the simplest of the three, containing only an extended N-terminal region with a PCNA interaction protein motif (PIP box) and a nuclear localization sequence (NLS). While not required for catalytic activity[15,41], the PIP box is essential for tethering LIG1 to PCNA and, therefore, to the replication machinery[42]. In place of a PIP box near the N-terminus, LIG3 contains a zinc-finger domain (ZnF), which is thought to aid in tight binding to nicked DNA substrates[43,44]. The LIG3 gene also undergoes alternative splicing events, generating different forms of the LIG3 protein[45-47]. All mammalian cells contain a LIG3 variant with a N-

terminal mitochondrial localization sequence to guide the enzyme to the mitochondria, where it is the only DNA ligase present. LIG3α contains a single C-terminal BRCT domain adjacent to its OB-fold, which facilitates interactions with XRCC1, a nuclear scaffolding protein[48-51]. LIG3β, which does not contain this BRCT domain, has a shorter C-terminal region containing only a nuclear localization signal. LIG4 contains an extended C-terminal region consisting of two BRCT domains, which help facilitate interactions with the protein XRCC4[52].

Cellular function of human DNA ligases

The diversity of the mammalian DNA ligase N-and C-terminal regions is key to defining the specific cellular roles of the individual enzymes. Since many processes necessitate the breakage of phosphodiester bonds, DNA ligases are ubiquitous in many essential cellular pathways. Many prokaryotic cells have multifunctional DNA ligases, such as LigD, which is able to act as a DNA ligase, polymerase and phosphodiesterase[53]. For simplification, the following section will only focus on the cellular roles of the human DNA ligase families.

DNA ligase function in replication

The discovery of DNA ligases coincided with the first evidence of Okazaki fragment formation, one of the initial studies validating the model of discontinuous DNA replication[54,55]. This turned out to be serendipitous for the study of DNA replication, as the recently discovered DNA ligases quickly became candidates for the cellular component responsible for sealing adjacent Okazaki fragments[54]. Indeed, subsequent studies were able to show that aberrant Okazaki fragment accumulation in cells can be caused by the absence of proper ligase function[55]. Cell cycle mutants were used to identify the ligase responsible for sealing Okazaki fragments in yeast, which resulted in the characterization of Cdc9, a LIG1 homolog, as the main replicative

DNA ligase in yeast[56-58]. The direct participation of human LIG1 in Okazaki fragment maturation, and consequently in replication, was first demonstrated *in vivo*[59], later being demonstrated to associate with the replication machinery in eukaryotic cells[60,61].

Targeting of LIG1 to the nucleus and to the replication fork requires the presence of the N-terminal PIP box, presumably to tether LIG1 to PCNA [62,63]. Direct interactions between LIG1 and PCNA have been demonstrated *in vitro*, confirming the role of the PIP box in facilitating this interaction[64,65]. *In vivo* results have suggested that this interaction is limited to the G_1 and S phases of the cell cycle through a regulation mechanism thought to be controlled by phosphorylation of the LIG1 N-terminal region[66-68]. Due to the biochemical similarity of LIG1 and LIG3, it is often thought that the two enzymes function in a redundant manner. As evidence for this, studies have shown that LIG1-deficient cells are still viable[69-71], implying that another ligase, presumably LIG3, is acting in its place. LIG1 is still considered to be the main replicative DNA ligase, as no direct role for LIG3 in Okazaki fragment maturation has been uncovered[72,73].

DNA ligase function in DNA repair

Defects in DNA ligase activity are often associated with increased sensitivity to DNA damaging agents. Indeed, the major repair pathways in the cell—base excision, nucleotide excision and mismatch repair—all require ligation as the final step of repair (Figure 1.4). Due to the ability of both LIG3 and LIG4 to participate in double-strand break ligation, these two ligases are also found as participants in the cell's double-strand break response—most notably in nonhomologous end-joining (NHEJ) and alternative NHEJ (alt-NHEJ).

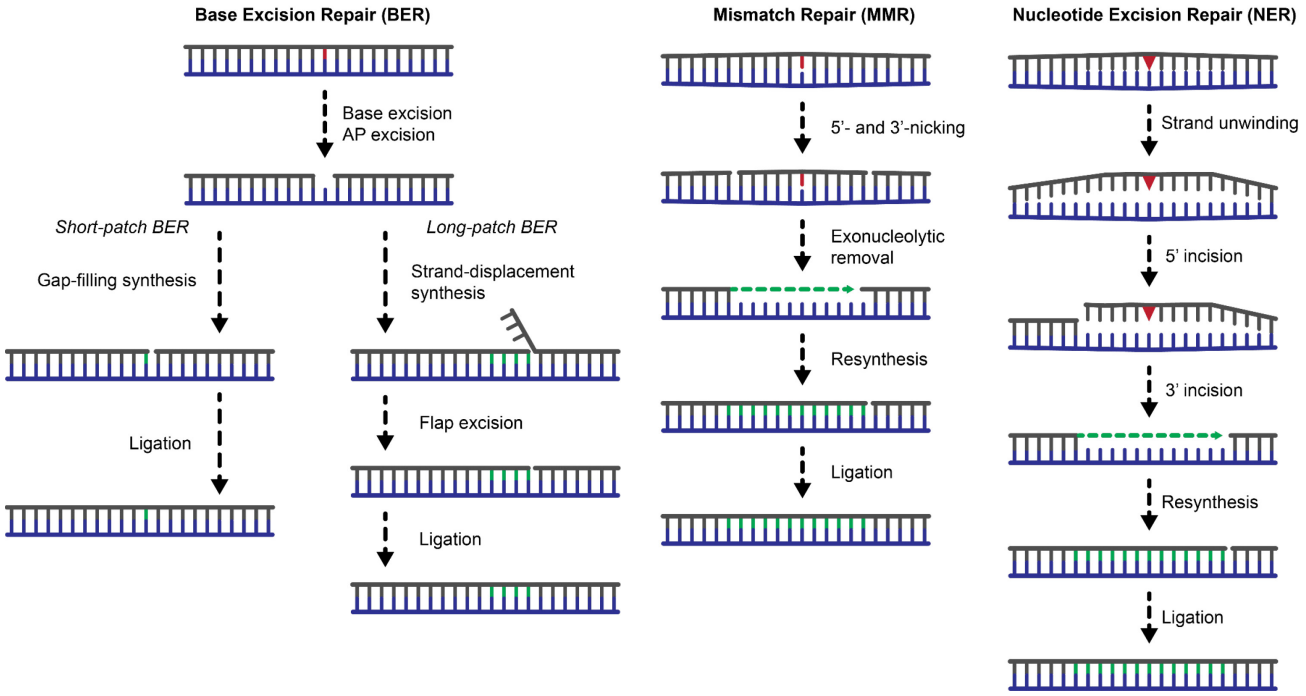


Figure 1.4. Single-strand DNA repair pathways in mammalian cells. Schematic representations of the three major DNA repair pathways for single-strand repair. Base excision repair (BER, left) begins with the recognition and excision of a damaged base by a DNA glycosylase, producing an abasic or apurinic (AP) site[74]. In the short-patch sub-pathway, the resulting 5'-deoxyribose phosphate (dRP) is removed by a dRP lyase and the remaining gap is filled by DNA polymerase β (Polβ), concluding with ligation by either LIG1 or the LIG3α•XRCC1 complex. The long-patch sub-pathway relies on strand-displacement synthesis by Polβ or either of the replicative polymerases (Polδ and Polε). A 5'-flap is created by strand displacement and is subsequently cleaved by FEN1, allowing for LIG1-dependent ligation. The long-patch pathway is also dependent on the presence of PCNA, which tethers the polymerase, FEN1 and LIG1 to the site of damage. Mismatch repair (MMR, middle) begins with recognition of mispaired bases by the MutSα complex, which recruits MutLα[75]. Incisions are made on the nascent DNA strand at sites both upstream and downstream of the mismatch, allowing for the removal of the damaged strand. Resynthesis of the excised DNA is performed by either Polδ or Polε, followed by LIG1- or LIG3-dependent ligation. Nucleotide excision repair (NER, right) is the main pathway for removing bulky DNA lesions in the cell[76,77]. Recognition of the damaged nucleotide can either occur due to RNA polymerase stalling during transcription or through a pathway mediated by XPC-RAD23B, XPA and RPA. The XPD and XPB helicases are recruited to the damaged site and unwind the DNA double helix. Dual excision is performed by XPG at the 3' end and XPF-ERCC1 at the 5' end, allowing for the damaged strand to be removed. Akin to MMR, resynthesis is performed by either of the replicative polymerases, and either LIG1 or LIG3 can perform the final ligation step.

Base excision repair (BER) is the main repair pathway in the cell for the removal of non-helix-distorting base lesions from genomic DNA[74]. Most base lesions recognized and repaired by BER are those resulting from deamination, alkylation and oxidation. BER is initiated by the action of a DNA glycosylase, which recognizes and excises damaged bases, leaving behind an abasic/apurinic site. Single-strand breaks in the phosphodiester bond are then created by AP

endonuclease (APE1) and the resulting 5'-deoxyribose phosphate (5'-dRP) is removed by the action of a dRP lyase. Some glycosylases, such as human OGG1 and NEIL-family glycosylases, are bifunctional and able to catalyze the cleavage of the phosphodiester backbone without APE1 present[74]. Following removal of the 5'-dRP, DNA polymerase β (Polβ) performs gap-filling, allowing for ligation of the adjacent strands by LIG1 or the LIG3•XRCC1 complex[78-80]. This pathway, known as short-patch BER, occurs both in the mitochondria and the nucleus, but is only critical for mitochondrial DNA repair[81]. An alternative BER pathway known as long-patch BER is observed in cells undergoing replication[82-84]. In place of dRP lyase activity, one of the replicative polymerases, Polδ and Polε, performs strand displacement synthesis, creating a 5' DNA flap[85,86]. The flapped DNA strand is then excised by the action of the flap endonuclease FEN1, resulting in a 5'-phosphorylated nick that is ligated by LIG1. Long-patch BER is known to require the presence of PCNA, which facilitates strand-displacement synthesis and tethers both FEN1 and LIG1 to the site of damage[65,87,88].

Mismatch repair (MMR) is tasked with removing mismatched bases that arise during replication[75]. Initiation of MMR is predominantly the role of the Msh2 and Msh6 proteins, which form a complex referred to as MutSα. A second complex known as MutLα is recruited to the site of damage and is activated by PCNA to stimulate 5'- and 3'-incisions in the nascent DNA strand. The incised strand is removed through one of two mechanisms[75], allowing for resynthesis and ligation. Consistent with LIG1 interacting with PCNA and being the major ligase involved in replication, LIG1 is thought to be the main ligase involved in MMR[89].

Nucleotide excision repair (NER) is specialized towards removing entire nucleotides that are damaged by UV-radiation or other environmental mutagens—including chemotherapeutic agents—resulting in bulky DNA lesions[76,77,90]. NER is initiated by the action of a complex of

proteins that recognize the bulky DNA lesion and facilitate the unwinding of the DNA strands at the site of damage. Following unwinding, XPF and XPG catalyze incision at sites both 5' and 3' to the damaged nucleotide. Resynthesis of the incised strand is then performed by a DNA polymerase and the newly synthesized strand is joined to the adjacent strand by the action of a DNA ligase. There is strong evidence showing that this final ligation event is performed primarily by the LIG3α-XRCC1 complex. Indeed, in cells lacking proper LIG1 function, efficient NER is still present[48]. However, the dominance of LIG3 in NER is not as clear in proliferating cells since both LIG1 and the LIG3α-XRCC1 complex colocalize with the NER machinery.

More complex mechanisms are employed in the cell to handle double-strand breaks in the DNA helix. The challenge of repairing double-strand breaks mostly lies within having to accurately detect the two broken ends. While all three DNA ligases can perform double-strand break repair to some extent, only DNA ligase IV is a member of a larger group of end-bridging proteins capable of repairing blunt ends or ends with short or incompatible overhangs[37]. LIG1 and LIG3 only appear to be able to repair double-strand breaks that result in homologous overhangs longer than ~2 bases; this form of end joining is known as alternative end-joining (a-EJ) or microhomology mediated end-joining (MMEJ) and is mainly observed in the absence of LIG4[91-94].

Two more common methods for joining broken DNA ends are homologous recombination and non-homologous end joining (NHEJ). Homologous recombination (HR) is the dominant pathway during the S and G2 phases of the cell cycle and relies on available sister chromatids for precise DSB repair. NHEJ, on the other hand, is active throughout the cell cycle and considered to be error-prone[91,95,96]. NHEJ begins with recognition of either blunt or

overhanging ends by Ku, followed by recruitment of the DNA protein kinase catalytic subunit (DNA-PKcs) to form the DNA-PK complex. Depending on the nature of the two DNA ends, various forms of end processing, such as DNA resectioning, polymerization or dephosphorylation, may be required prior to end joining. In all cases, LIG4-catalyzed ligation is the final step in resealing the DSBs[96-98].

DNA ligase function in DNA recombination

DNA recombination is often thought of as simply the use of a sister chromatid for repairing homologous strands of DNA or as a method for increasing genetic diversity at certain alleles. DNA recombination is also an essential tool for the adaptive immune response, which utilizes two types of somatic recombination, V(D)J and class-switch recombination, to increase the versatility of the immune system. B and T cells undergo V(D)J recombination in the early stages of maturation to diversify the immunoglobulins (Igs) and T cell receptors (TCRs), while class-switch recombination (CSR) changes the type of immunoglobulin produced by a B cell[99,100]. Both processes operate by through the excision of specific regions in DNA, necessitating the rejoining of DNA strands by a DNA ligase.

The primary pathway for both V(D)J recombination and CSR concludes with ligation through the NHEJ pathway, mediated by LIG4. In the absence of core NHEJ components—however—CSR is still observed, pointing to a role for a-EJ in CSR. Cells containing either only LIG1 or only LIG3 as the nuclear ligase exhibit CSR at levels comparable to those in cells containing LIG4[92,94]. V(D)J recombination does not show the same adaptability, however, since disruption of the NHEJ pathway appears to prevent any V(D)J recombination activity[101].

DNA ligases and human disease

LIG1 syndrome

The first reported case of LIG1 syndrome was also the first reported case of a human disease caused by a defective DNA ligase. The afflicted patient was developmentally delayed, susceptible to recurrent infections and had skin hypersensitivity to sunlight exposure[102,103]. The patient was determined to be immunocompromised, exhibiting decreased IgA and IgG levels, as well as poorly proliferative T and B cells. Despite the severity of their symptoms, the patient lived until the age of 19 when they died as the result of pneumonia. Initially, the patient was thought to have Bloom's Syndrome, which is an inherited disease caused by mutations in the helicase RecQ2, but later tests disproved this diagnosis[102,104]. The patient instead had two mutant LIG1 alleles, giving rise to two missense mutations and, consequently, a biallelic loss of function. One of the ligase variants, E566K LIG1, reverses the charge of a conserved residue in the enzyme active site, ablating enzyme activity. Early investigation of the other variant, R771W LIG1, revealed delayed enzymatic activity, resulting in a delay in DNA repair and in Okazaki fragment joining[105]. Novel inherited mutations in the LIG1 gene have recently been identified[106] and are the focus of a biochemical characterization of detailed in Chapter 6.

LIG4 syndrome

LIG4 syndrome is used to describe LIG4 mutations that are known to result in disease. Most of these mutations are in the catalytic core of the enzyme, but patients have mutant LIG4 alleles impacting the DNA-binding and BRCT domains[107]. Like both LIG1 syndrome and AOA1, the prevalence of LIG4 syndrome is low. Patients suffering from LIG4 syndrome are afflicted with a combined immunodeficiency (CID), with some patients suffering from a severe combined immunodeficiency (SCID), consistent with the abrogated levels of V(D)J

recombination often seen in LIG4 syndrome patients[108]. Increased sensitivity to ionizing radiation has been observed in the cells of all reported LIG4 patients, owing to the unique role of LIG4 in double-strand break repair through the NHEJ pathways. Due to the broad symptomology of LIG4 syndrome, no single treatment for the disease exists and symptoms are instead treated on an individual basis.

Ataxia with oculomotor apraxia type 1

AOA1 is a form of ataxia that results in severe neuropathy, as well as abnormal and uncontrolled eye movements, known as oculomotor apraxia. The disease has an early onset, with obvious symptoms of ataxia appearing in early childhood, frequently leaving the afflicted patient wheelchair-bound in early adolescence[109]. AOA1 is not caused by a defect in DNA ligation *per se* but is instead caused by defects in the protein aprataxin (APTX), which repairs the products of abortive ligation[110-112].

Abortive ligation occurs when a DNA ligase fails to complete ligation, instead releasing the adenylylated reaction intermediate before nick sealing can be completed. Upon release of the adenylylated intermediate, the active site lysines of DNA ligases are thought to be rapidly re-adenylylated, preventing the adenylylated enzyme from binding to and ligating the abortive ligation product. APTX can recognize these abortive products and catalyze their deadenylation, either giving the ligase another chance at successful ligation or allowing other proteins to repair the nicked DNA strand[112]. Although the BER pathway is also capable of repairing abortive ligation products[113], APTX is thought to be the predominant repair pathway in quiescent cells, such as neurons. Several mutations in APTX are associated with AOA1, causing a range of biochemical consequences, such as decreased catalytic activity or protein stability[114,115].

Mechanism and specificity of human DNA ligase I

The work detailed in this dissertation centers around expanding our knowledge of DNA ligation through extensive study of human DNA ligase I. Our understanding of human DNA ligase structure and function has relied heavily on insights gained from viral DNA ligases, which do not share some of the complexities of their human counterparts. Additionally, research into the mechanisms of DNA replication and repair have been mostly focused on steps prior to ligation, limiting our knowledge of how the replication and repair pathways are concluded.

Chapter 2 focuses on the investigation of the initial DNA binding steps of the LIG1 catalytic pathway. Although many studies have been aimed at understanding the chemistry behind DNA ligation, the binding and recognition mechanisms utilized by DNA ligases are largely unknown. Using a reporter substrate and intrinsic protein fluorescence, we were able to monitor interactions between LIG1 and its DNA substrate, characterizing previously unknown steps preceding adenylyl transfer by LIG1. As a result, we identified two distinct conformational changes that LIG1 induces in the DNA substrate, helping to position the nick in the active site for catalysis.

We also sought to understand how LIG1 achieves specificity for undamaged nicks with proper Watson-Crick pairing. LIG1 is known to be able to discriminate against nicks containing damaged or mispaired bases. Additionally, ligases have evolved a tolerance for ribonucleotides at the 3' end while heavily discriminating against a ribonucleotide at the 5' end. In pursuit of understanding this inherent discrimination, we have uncovered two novel mechanisms utilized by LIG1 to maintain fidelity. Chapter 3 will primarily focus on mismatch discrimination at the 3' end. X-ray crystallographic data revealed a novel magnesium-binding site that appears to tune the tolerance for 3' mismatches. Mutagenesis of this conserved metal site allows LIG1 to better

accommodate mismatches in the active site. The results of this study suggest that the novel metal site plays a role in stimulating abortive ligation to prevent incorporation of damaged DNA into the genome.

Chapter 4 is focused on the mechanism behind ribonucleotide discrimination at the 5' end of the nick. To investigate this, we performed mutagenesis of a conserved phenylalanine residue positioned near the sugar of the 5' nucleotide. The positioning of the phenylalanine residue suggested that it may prevent the accommodation of the 2'-OH found in ribonucleotides and, therefore, may be key for precluding ligation of downstream RNA. Substituting the phenylalanine with either alanine or leucine significantly impacts all steps of ligation on the canonical ligation substrate. The alanine substitution, however, renders LIG1 unable to discriminate between DNA and RNA at the 5' end, while a leucine substitution appears to enforce stricter discrimination against RNA. In addition, our results suggest that the phenylalanine residue has a role in helping to properly align the 5' end for catalysis.

Chapter 5 is focused on identifying the catalytic metal ligands for ligation. The initial report of the LIG1 structure suggested a network of interactions between conserved carboxylate residues in the active site that coordinate two catalytic magnesium ions. In contrast, only one magnesium ion is observed in coordination with the three residues in more recent high-resolution structures of LIG1. To test the role of these conserved residues, we biochemically characterized mutations at each of the three positions. The mutants exhibited differing degrees of catalytic defects, but all three mutations caused a dramatic decrease in magnesium affinity. Additionally, our evidence suggests that one residue plays an outsized role in stabilizing ATP and AMP in the active site.

Finally, we sought to gain a better understanding of the connection between mutations in the LIG1 gene and the onset of disease. A recent cohort of patients were identified as having novel mutations in their LIG1 gene, causing symptoms like those found in a previous patient suffering from LIG1 syndrome. Using site-directed mutagenesis, we made recombinant versions of two of the patient-derived mutants. Combining biophysical and biochemical techniques, we demonstrated that the two mutations decrease the efficiency of ligation by LIG1. Additionally, we found that the two mutants exhibit dramatic decreases in magnesium affinity, leading to an increase in the amount of abortive ligation. Preliminary crystallographic studies suggest that the biochemical defects result from the inability of the mutants to stably position the DNA substrate for catalysis, which remains to be tested directly.

The work contained in this dissertation represents a significant step toward a more comprehensive understanding of the mechanism and specificity of ligation by LIG1. Collectively, this work defines the molecular interactions between LIG1 and the DNA substrate prior to catalysis, demonstrating how these early interactions help to determine ligation fidelity. These interactions contribute to how the DNA ends are positioned in the enzyme active site, influencing how the catalytic magnesium ion is coordinated. Due to the ubiquity of LIG1 in DNA replication and repair, perturbations of the molecular interactions between LIG1 and its substrates can cause a cascade effect, ultimately resulting in disease. Future work is still necessary to understand the full scope of ligation fidelity and how this may contribute to the overall fidelity of the repair and replication pathways.

Chapter 2
Mechanism of DNA-binding by LIG1

Restoration of phosphodiester bonds necessitates rapid binding and recognition of the break by DNA ligases to avoid deterioration of the exposed DNA ends. While it is well-known that protein partners are often involved in guiding the localization of DNA ligases to sites of replication and repair[49,64], nick recognition is thought to occur by an undefined mechanism that is intrinsic to DNA ligases. Crystal structures of DNA ligases uncovered that the enzymes distort the DNA helix upon binding, compressing the upstream strand into an A-form-like helical arrangement, resembling an intermediate state between an A- and B-form helix[13,15]. This is thought to be important for arranging the 3' end of the nick into a catalytically favorable conformation in the enzyme active site. As detailed in Chapter 3, stabilization of this distorted helix also appears to help enforce ligation fidelity in human DNA ligase 1. Similarly, binding interactions at the 5' end of the nick are thought to contribute to fidelity by enforcing preference for deoxyribonucleotides at this position; further study of this particular mechanism is described in Chapter 4. While studies into the fidelity mechanisms employed by DNA ligases are important for understanding the enzyme in the larger biological context, these studies are limited by a lack of knowledge regarding the binding steps preceding them.

Requirements for nick binding and recognition by DNA ligase have largely been derived from studies of small viral DNA ligases, such as those found in the Vaccinia and Chlorella viruses. An early study of the Vaccinia virus DNA ligase demonstrated discrimination against

binding to DNA nicks lacking a 5'-phosphate[11]. Additionally, deadenylylated forms of the viral DNA ligase are unable to tightly associate with nicked DNA, suggesting that adenylylation of the active site lysine is a prerequisite for nick recognition. Studies of the Chlorella virus DNA ligase have demonstrated a tolerance for ribonucleotides in the upstream strand of a nicked DNA substrate[116], lending some biochemical evidence to the later observation made with LIG1 that DNA ligases enforce an A-form-like helical structure upstream of the nick[15].

More recent studies of the T4 bacteriophage DNA ligase (T4Dnl) used stopped-flow spectroscopy to monitor fluorescence changes associated with DNA-binding and catalysis[14]. This study utilized a nicked DNA substrate containing a deoxythymidine-conjugated fluorescein (TFAM) at the 3' end of the nick, which was shown to be sensitive to interactions between the ligase and the substrate. This modification is well-tolerated by T4Dnl, enabling reactions between the enzyme and the reporter DNA to be observed using real-time fluorescence. Employing the TFAM reporter substrate provided the first evidence of a multi-step binding mechanism for a DNA ligase, which the authors conclude is a transition between two different conformational states of the T4Dnl•DNA complex.

While many questions still linger with respect to the binding mechanism employed by T4Dnl, this study provided us with a viable technique to probe the DNA-binding mechanism of a human DNA ligase. Here we report the characterization of the initial interactions between LIG1 and the nicked DNA substrate, using both TFAM and intrinsic protein fluorescence. We propose a model for DNA binding by LIG1 (Figure 2.1a) wherein LIG1 first associates with the DNA in a non-specific manner. Translocation of the enzyme along the DNA helix halts when a nick is encountered, resulting in the formation of a specific recognition complex. From crystallographic evidence and results from the T4Dnl study, we hypothesize that specific recognition is followed

by a conformational change in LIG1 and/or the substrate, allowing catalysis to occur. The following study represents the first characterization of the full kinetic mechanism of DNA-binding and ligation by a human DNA ligase.

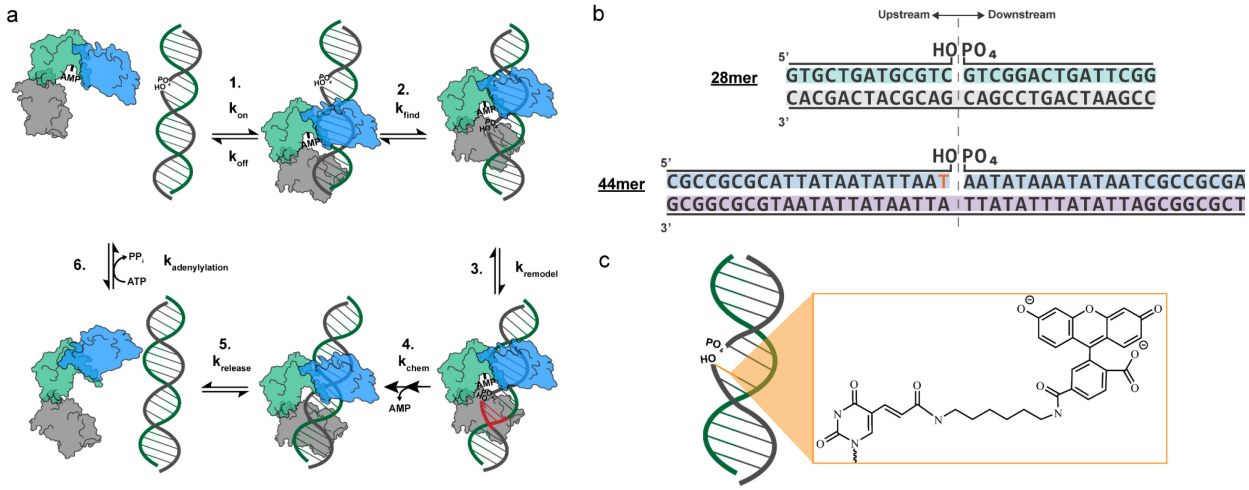


Figure 2.1. Probing the LIG1 DNA-binding mechanism using a reporter substrate. (**a**) Schematic of the hypothetical model of DNA-binding by LIG1. LIG1 interacts in a non-specific manner with the DNA double-helix (1), translocating along the helix until a nick site is found (2). Nick recognition is followed by a hypothetical remodeling step (3), where the DNA helix is distorted to align the nick in the active site for catalysis (4). Following formation of the sealed DNA product, the DNA is released (5), allowing LIG1 to bind and hydrolyze ATP (6) in preparation for a subsequent ligation event. (**b**) DNA substrates used in the determination of the LIG1 binding mechanism. Variations of both the 28mer and the 44mer containing a 3'-FAM label on the downstream strand are used in standard gel-based ligation assays. (**c**) A 44mer variant containing a deoxythymine-linked FAM (TFAM) label at the 3'-end of the nick is used to detect DNA-binding by LIG1. The position of the TFAM label is colored orange in both **b** and **c**.

Materials and Methods

Preparation and purification of materials. The catalytic core of human DNA ligase I (residues 232-919) was expressed in *E. coli* and purified as previously reported[106]. The concentration of the adenylylated enzyme was determined by active site titration against a fixed concentration of the nicked 28mer substrate. Substrate oligonucleotides were synthesized by Integrated DNA Technologies (IDT) and the Keck Biotechnology Resource Laboratory at Yale University. Lyophilized oligos were dissolved in 8 M urea and 1X Tris/Borate/EDTA (TBE) before being gel-purified on denaturing polyacrylamide gels, as previously described[41]. The

oligo sequences used for this study are shown in Figure 2.1b and Figure S2.8. Both the 28mer and 44mer substrates contain a 6-FAM label (IDT) on the 3' end of the downstream oligo. The TFAM-44mer is identical to the 44mer substrate with the exception of the fluorescein reporter linked to the 3' thymine on the upstream strand instead of at the end of the downstream oligo (Figure 2.1c). A double-labeled variant of the TFAM-44mer containing both the internal TFAM label and the downstream 3' 6-FAM was used in single-turnover experiments to follow the adenylylated intermediate species. An additional variant of the TFAM-44mer lacking a 5'-PO_4 on the downstream strand was used in equilibrium-binding studies. An unlabeled 28mer and 28mer variant containing an all ribonucleotide upstream strand were used in stopped-flow experiments monitored using tryptophan fluorescence.

Gel-based ligation assays. All ligation assays were performed at 37°C with a standard reaction buffer containing 50 mM NaMOPS pH 7.5 (adjusted at 25°C), 0.1 mM dithiothreitol, 0.05 mg/mL BSA and varying amounts of NaCl to maintain a constant ionic strength of 150 mM. Steady-state ligation assays were quenched with an equal volume of the gel-loading solution containing 50 mM EDTA/90% formamide. Quenched reactions were heated to 95°C prior to loading on 15-20% (wt/vol) polyacrylamide gels containing 6.6-8 M urea. The fluorescein-labeled species were detected using either a Typhoon Trio^{+} or a Typhoon 5 imager (GE Healthcare) and quantified using the ImageQuant software (GE Healthcare) as previously described[41].

Steady-state ligation assays. DNA-dependences were performed in the presence of either 20 mM $MgCl_2$ and 0.2 mM ATP or 2 mM $MgCl_2$ and 1 mM ATP. Reactions with 2 mM $MgCl_2$ contain 1 mM Mg^{2+}_{free} due to chelation of magnesium by ATP[41]. Reactions contained 0.1-10 nM LIG1 and increasing concentrations of either the 28mer or TFAM-44mer substrate. Initial rates

of reaction were obtained from the slopes of linear fits to the fraction of product as a function of time, which were then plotted as a function of substrate concentration. Values for k_{cat} and K_M were obtained from fits of the data to the Michaelis-Menten equation (Equation 2.1).

Equation 2.1 $$\frac{V_{init}}{[E]} = \frac{V_{max}[S]}{K_M+[S]}$$

Rapid-quench flow ligation assays. Single-turnover assays were carried out using a KinTek RQF-3 apparatus. Ligation reactions were prepared by loading one sample syringe with 1.6 μM LIG1 and the other with 160 nM of one of the three tested DNA substrates. Both syringes contained varying amounts of $MgCl_2$. Ligation was initiated upon mixing of the solutions in a 1:1 ratio, followed by quenching of the reaction at various timepoints with a solution containing 1 N H_2SO_4/0.25% SDS[117]. After quenching, 20 μL of the quenched sample was immediately neutralized in a solution containing 300 mM Tris base, 15 mM EDTA and 0.5X loading buffer. Rates for the adenylyl-transfer and nick-sealing steps were obtained from fits of the data using a two-step irreversible model in Berkeley-Madonna as previously described[41].

Steady-state fluorescence binding assays. Fluorescence intensity and anisotropy measurements were collected using a FluoroMax 3 fluorometer (Horiba) controlled by the DataMax software. Data were collected using excitation and emission wavelengths of 495 nm and 515 nm, respectively. Binding assays were performed at 37°C by titration of increasing amounts of LIG1 against a 2.5 mL solution containing 1-2 nM of the TFAM-44mer and the standard reaction buffer with 0.1 mM EDTA. After addition of LIG1, samples were allowed to equilibrate for 30 seconds to 1 minute before measurements were taken. Anisotropy and relative fluorescence are plotted as a function of added [LIG1] and the data was fit to a hyperbolic binding curve (Equation 2.2) to derive K_D values.

Equation 2.2 $$y = \frac{Y_{max}[LIG1]}{K_D+[LIG1]} + C$$

Where y is either anisotropy or fluorescence intensity, Y_{max} is the anisotropy or fluorescence value at the plateau and C is a constant.

Stopped-flow kinetic experiments. Transient kinetic experiments were performed using the Hi-Tech SF-61DSX2 stopped-flow instrument controlled by the Kinetic Studio (TgK Scientific) software, with a circulating water bath maintaining temperature at 37°C. A minimum of 5-6 shots were performed and averaged for each concentration tested. Fluorescein fluorescence was measured using an excitation wavelength of 495 nm and two 515 nm longpass filters. For experiments looking at tryptophan fluorescence, an excitation wavelength of 295 nm and two 320 longpass filters. All stopped-flow reactions were performed in the standard ligation buffer, except for experiments monitored by tryptophan fluorescence, for which BSA was omitted from the buffer. Data from both PMT detectors were combined and averaged for each shot. The resulting traces were either fit to a single exponential or a series of exponentials, depending on the number of observed phases (Equation 2.3).

Equation 2.3 $$y = a_1 e^{-k_1 * t} + ... a_n e^{-k_n * t} + c$$

Where y is fluorescence intensity, a is the amplitude of the phase and C is a constant. All observed rates (k) were plotted as a function of the concentration of whichever component was varied throughout the experiment. The resulting data plots were fit either by linear regression or with a hyperbola to determine the maximal rates and $K_{1/2}$ values.

The TFAM-44mer was fixed at a constant concentration of 20 nM for measurements of the transient kinetic dependence on the concentration of LIG1. Reactions were performed with a concentration of LIG1 ranging from 0–5 μM and in the standard reaction buffer supplemented with either 0.1 mM EDTA or 20 mM $MgCl_2$. For experiments monitored by tryptophan fluorescence, the concentration of LIG1 was kept constant at either 0.5 or 1 μM. In the absence

of magnesium, the concentration of the 28mer DNA substrate was varied from 1 – 10 μM in the standard reaction buffer supplemented with 0.1 mM EDTA. To characterize the magnesium dependence, LIG1 was kept at a concentration of either 0.5 or 1 μM, with the substrate held at a concentration of either 5 or 10 μM. To measure the rate of product release as a function of ATP concentration, 1.1 μM LIG1 and 1 μM 28mer were mixed prior to loading in the stopped-flow apparatus in the presence of 0.1 mM EDTA to prevent catalysis. Reactions were initiated by rapid-mixing of the LIG1•DNA complex and a solution containing 2 mM $MgCl_2$ and increasing amounts of ATP.

Results

LIG1 performs efficient steady-state ligation of the TFAM reporter substrate

Investigation of the DNA-binding mechanism of LIG1 required a reliable assay for monitoring interactions between LIG1 and a nicked DNA substrate. Previous study of the T4 DNA ligase binding mechanism made use of a substrate containing a deoxythymidine-linked fluorescein (TFAM) at the 3' end of the nick, which did not cause major perturbations to ligation kinetics. This is in contrast to the placement of the TFAM moiety at the 5' end, which strongly inhibited T4 DNA ligase kinetics[14]. Due these findings, we chose to incorporate the TFAM reporter at the 3' end of a 44mer ligation substrate (Figure 2.1b) composed of an AT-rich sequence to avoid fluorescence quenching by guanine bases[118]. The TFAM is situated at the -1 position in the substrate to maximize the chances of observing fluorescence changes due to direct interactions between the protein and DNA (Figure 2.1c).

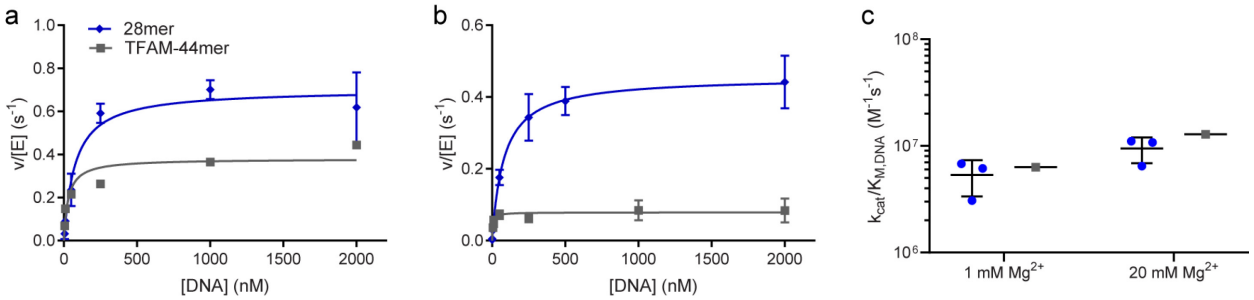


Figure 2.2. TFAM-label impacts catalytic rate but not overall catalytic efficiency. Steady-state substrate dependences were performed with increasing amounts of both the 28mer and TFAM-44mer substrates and 0.1-10 nM LIG1, in the presence of either 20 mM Mg^{2+} and 0.2 mM ATP (**a**) or 1 mM Mg^{2+} and 1 mM ATP (**b**). Data were fit using the Michaelis-Menten equation and best-fit parameters are shown in Table 2.1. (**c**) Plot of the k_{cat}/K_M values for ligation of both the 28mer and TFAM-44mer substrates highlighting that LIG1 ligates both substrates with similar efficiency. Data for the TFAM-44mer are preliminary and only reflects the results of two replicate experiments. The steady-state dependences for the 28mer are adapted from the study detailed in Chapter 3.

To test whether the presence of this fluorophore at the nick perturbs ligation by LIG1, we compared steady-state DNA substrate dependences for the TFAM-44mer substrate and a previously characterized 28mer substrate (Figure 2.2)[41]. At both saturating and physiological concentrations of Mg^{2+}, LIG1 ligates the TFAM-44mer at a slower rate than the 28mer substrate. While the decrease in k_{cat} is only ~2-fold in the presence of 20 mM $MgCl_2$, the decreased rate is more pronounced at low Mg^{2+}, where k_{cat} is ~5-fold slower for the TFAM-44mer than for the 28mer (Table 2.1). This change in k_{cat} is paired with a decrease in $K_{M,DNA}$ for the TFAM-44mer substrate, making the overall efficiency of TFAM-44mer ligation roughly equal to that of the 28mer in the presence of both high and low $MgCl_2$ (Figure 2.2c, Table 2.1).

Sequence-dependent changes in the chemical steps of ligation

Due to the decrease in k_{cat} for the ligation of our reporter substrate, we posited that the DNA-dependent chemical steps—adenylyl transfer and nick sealing—may be impacted by the presence of the TFAM fluorophore. To investigate this, we performed single-turnover experiments to compare the microscopic rates of ligation of the TFAM-44mer and 28mer substrates (Figure 2.3). Single-turnover experiments were carried out in the presence of

increasing amounts of magnesium, using a rapid quench-flow apparatus to observe the formation of the adenylylated ligation intermediate and its subsequent conversion into sealed DNA. The maximal rates and K_{Mg2+} values for the adenylyl-transfer and nick-sealing steps were obtained from hyperbolic fits of the rates at increasing magnesium concentrations (Table 2.1).

The single-turnover characterization of the 28mer substrate is in agreement with previous reports of LIG1-dependent ligation of this substrate[41] (Figure 2.3b-c). The magnesium affinities of the two chemical steps differ, with the K_{Mg2+} value in the sub-millimolar range for adenylyl transfer and ~5 mM for nick sealing (Table 2.1). Despite a tighter magnesium affinity, the maximal rate for adenylyl transfer is roughly 6-fold slower than that of nick sealing, making adenylyl transfer rate-limiting in the presence of magnesium concentrations at or exceeding physiological concentrations (≥ 1 mM)[41]. Meanwhile, the TFAM modification does appear to have significant impacts on the chemical steps of ligation (Figure 2.3b-c). The magnesium dependence on ligation of the TFAM substrate illustrates ~20-fold increases in the K_{Mg2+} values for adenylyl transfer and nick sealing relative to those for ligation of the 28mer. Additionally, the maximal rates of both steps appear to be much faster, although much higher concentrations of $MgCl_2$ would be required to observe these rates for ligation of the TFAM-44mer substrate (Table 2.1).

In addition to the TFAM modification, the sequence composition of the two substrates is also different, raising the question of whether the TFAM modification itself is responsible for these changes or if this behavior is due to the change in sequence context. To test this, we repeated our magnesium dependence assay to look at ligation of the 44mer substrate without the TFAM modification (Figure 2.3b-c). Overall, single-turnover parameters for ligation of the 44mer substrate are noticeably different from those obtained with either the 28mer or TFAM-

44mer substrates (Table 2.1). This divergence between the three substrates seems to indicate that sequence context impacts ligation by LIG1. The sequence composition of the four nucleotides at the nick has been shown to impact ligation by other DNA ligases[119,120]

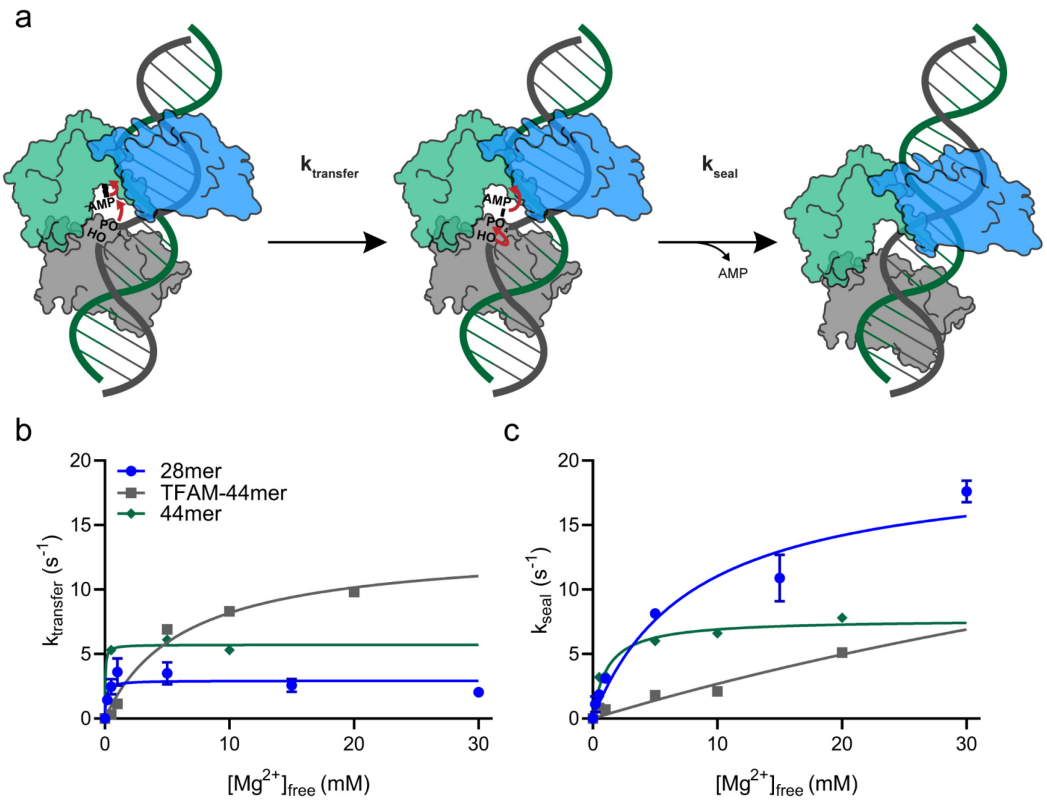


Figure 2.3. The TFAM label significantly impacts the individual chemical steps of ligation. (**a**) Schematic of the DNA-dependent ligation steps. Ligation consists of two chemical steps, starting with transfer of the adenylyl group from LIG1 to the 5'-PO4 at the nick. The adenylylated DNA intermediate is then sealed after attack of the 5'-PO4 by the adjacent 3'-OH, resulting in release of AMP and the sealed DNA. Single-turnover reactions were carried out in the presence of increasing amounts of $MgCl_2$, 800 nM LIG1 and 80 nM the DNA substrate. The individual rates of adenylyl transfer (**b**) and nick sealing (**c**) were derived from fits of the data using a two-step irreversible model in Berkeley-Madonna. Magnesium dependences were determined for ligation of the 28mer, TFAM-44mer and 44mer substrates, which all fit to a hyperbolic dependence on the magnesium concentration. Best-fit parameters for the maximal rates and magnesium affinities are shown in Table 2.1. Data for the 28mer is adapted from the study in Chapter 3, while the data for the TFAM-44mer and 44mer represents a single experimental replicate. Representative single-turnover traces are presented in Figure S2.1.

a

1 mM Mg^{2+}	TFAM-44mer	28mer
k_{cat} (s^{-1})	0.077 ± 0.007	0.45 ± 0.02
K_M (nM)	5.5 ± 2.8	81 ± 15
k_{cat}/K_M ($M^{-1}s^{-1}$)	1.4 ± 0.7 x 10^7	5.6 ± 1.1 x 10^6

20 mM Mg^{2+}	TFAM-44mer	28mer
k_{cat} (s^{-1})	0.38 ± 0.04	0.70 ± 0.04
K_M (nM)	29 ± 16	76 ± 21
k_{cat}/K_M ($M^{-1}s^{-1}$)	1.3 ± 0.7 x 10^7	9.2 ± 2.6 x 10^6

b

	TFAM-44mer	28mer	44mer
$k_{transfer}$ (s^{-1})	13	2.9 ± 0.1	6.6
K_{Mg2+} (mM)	6.3	0.27 ± 0.08	ND
k_{seal}(s^{-1})	32	16 ± 1.6	7.7
K_{Mg2+} (mM)	109	5.2 ± 1.7	0.37

Table 2.1. Best-fit parameters from gel-based ligation assays. (**a**) Steady-state parameters determined from the substrate dependences in Figure 2.2, displayed as the mean ± S.D. Values for the TFAM-44mer are only derived from duplicate experiments. (**b**) Single-turnover parameters determined by hyperbolic fits of the data in Figure 2.3. Values for the two 44mer substrates are based on a single experiment. All data for the 28mer substrate in a and b are adapted from the study detailed in Chapter 3.

The TFAM substrate reports on nick-specific binding

Despite the obvious effects of the TFAM fluorophore on ligation, we reasoned that the reporter may still be a viable tool and sought to test its ability to report on specific interactions between LIG1 and the nick. To this end, we performed equilibrium-binding experiments by titrating increasing amounts of LIG1 against fixed concentrations of either the TFAM-44mer substrate or a variation lacking a 5'-phosphate at the nick site in the absence of magnesium (Figure 2.4a-b). Binding was observed by fluorescence quenching and anisotropy, which were both fit to binding hyperbolas to define the K_D for the two oligos. While we observed a tight interaction between LIG1 and the phosphorylated TFAM-44mer ($K_D \approx 5$ nM), binding to the dephosphorylated nick is much weaker, with a K_D closer to 4 μM (Table 2.2). Our results indicate that the tight binding measured for the phosphorylated substrate is likely a measurement of specific affinity for the nick site.

We next performed stopped-flow spectroscopy experiments to measure the kinetics of binding in the absence of catalytic magnesium. Representative traces resulting from the rapid-mixing of an increasing excess of LIG1 with the reporter substrate are shown in Figure 2.4c. Observed rates constants were obtained from single-exponential fits to the fluorescence traces and are plotted as a function of increasing amounts of LIG1 (Figure 2.4d). Observed rate constants are linearly dependent on the concentration of LIG1, indicating that this fluorescence

decrease corresponds to the bimolecular association with a $k_{on,app}$ of 6.7 ± 1.3 x 10^7 $M^{-1}s^{-1}$ and a $k_{off,app}$ of 29 ± 13 s^{-1}. Using these estimates gives a K_D value close to 0.4 μM, which is substantially higher than what was measured via equilibrium binding. Without further experimentation, it is likely that this discrepancy is due to the presence of a two-step association, in which one step is spectroscopically invisible. If true, then the equilibrium binding measurement would then represent the dissociation constant for both steps combined, while only the first step would be represented in the dissociation constant obtained from stopped-flow experiments.

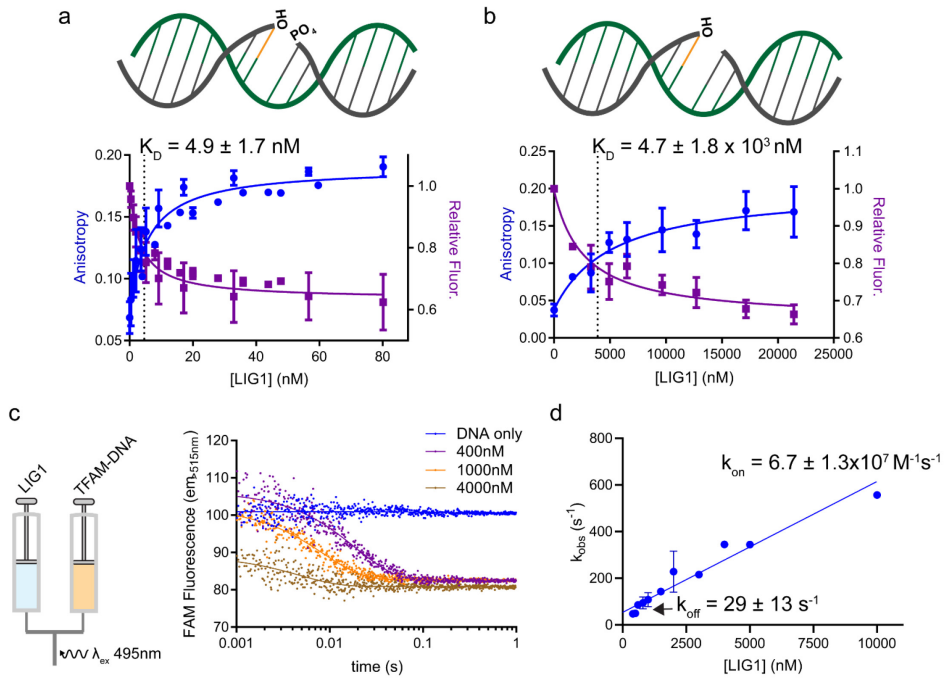


Figure 2.4. TFAM-label reports on nick-specific interactions between DNA and LIG1. Equilibrium-binding experiments were performed by titration of LIG1 against a fixed concentration of either the TFAM-44mer substrate (**a**) or a variation of the substrate lacking the 5'-PO_4 group at the nick (**b**) in the absence of Mg^{2+}. Increases in anisotropy and fluorescence quenching are plotted as a function of the total concentration of added LIG1, which are fit to hyperbolic binding models to determine the binding affinities for each substrate. (**c**) Stopped-flow experiments were performed by rapidly mixing increasing amounts of LIG1 with 20 nM of the TFAM-44mer substrate in the absence of magnesium. Representative traces are shown for experiments with 0-4 μM of the enzyme. All traces were fit using a single-exponential, and the individual rate constants are plotted as a function of [LIG1] in **d**. All values are shown as the mean ± S.D. of at least three replicate experiments and are reported in Table 2.2.

Catalysis on the reporter substrate can be monitored via real-time fluorescence

We next investigated how the presence of magnesium affects the LIG1•DNA interactions by repeating stopped-flow experiments in the presence of 20 mM $MgCl_2$. The resulting fluorescence traces contain an initial fluorescence decrease, followed by two distinct fluorescence recovery phases (Figure 2.5a). Using a four-exponential fit, we were able to obtain rate constants for the observed fluorescence changes (Figure 2.5; Table 2.2). The first two phases—contained within an elongated fluorescence decrease—are dependent on the concentration of added LIG1 (Figure 2.5b-c). The first phase is linearly dependent on enzyme concentration, most likely describing the same initial binding step as seen in the absence of magnesium. Indeed, the slope of the resulting linear fit gives a bimolecular rate constant of 2.0 ± 0.8 x 10^8 $M^{-1}s^{-1}$, which is similar to the value of 6.7 x 10^7 $M^{-1}s^{-1}$ measured in the absence of magnesium, indicating that magnesium is likely not affecting formation of the initial LIG1•DNA complex.

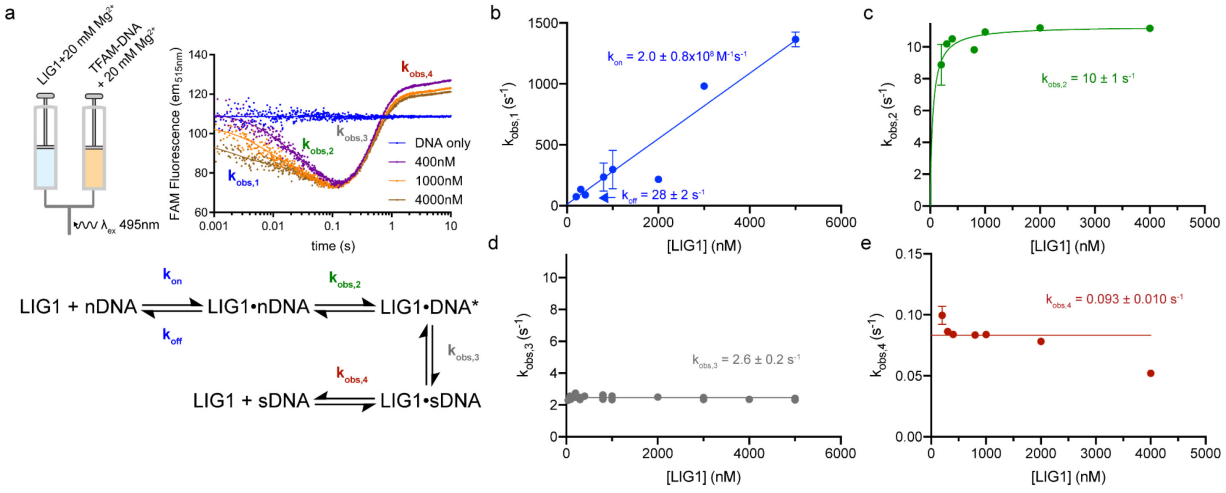


Figure 2.5. Magnesium initiates an unknown step preceding catalysis. (**a**) Stopped-flow experiments were performed by rapidly mixing increasing amounts of LIG1 with 20 nM of the TFAM-44mer substrate in the presence of 20 mM $MgCl_2$. Representative traces are shown for experiments with 0-4 μM of the enzyme. The observed phases are mapped onto the proposed model, beginning with LIG1 binding to the nicked DNA substrate (nDNA). The second phase either reports on adenylyl transfer or an unknown conformational change in the DNA, indicated by DNA*. The third phase matches the single-turnover rate for nick sealing. The final step is similar to the steady-state k_{cat} value for the TFAM substrate and likely represents full product release. All traces were fit using a four-exponential model and the individual rates of each phase are plotted as a function of [LIG1] in **b-d**. $k_{obs,1}$ and $k_{obs,2}$ show a linear and hyperbolic dependence on the concentration of added LIG1, respectively. All values are shown as the mean ± S.D. of at least three replicate experiments and are reported in Table 2.2.

The cause of the second phase of the fluorescence decrease—exhibiting a hyperbolic dependence on enzyme concentration—is less clear (Figure 2.5c). With a k_{obs} of 10 ± 1 s^{-1}, this phase could be attributed to adenylyl transfer, which saturates at ~10 s^{-1} in the presence of 20 mM $MgCl_2$ in our gel-based ligation assays (Table 2.1). Since we have no indication that the 3'-TFAM fluorophore is sensitive to adenylylation of the 5'-PO_4, this phase may instead represent an uncharacterized reversible conformational change in either LIG1 or the DNA. Subsequent experiments designed to test this possibility are described in more detail in the discussion.

The two LIG1-independent phases observed in the latter half of our fluorescence traces likely reflect product formation and release. Across all tested LIG1 concentrations, $k_{obs,3}$ is roughly equal to the rate constant for nick sealing, which, at about 3-5 s^{-1}, is the rate-limiting chemical step for ligation of the TFAM-44mer in the presence of 20 mM $MgCl_2$ (Figure 2.5d). Increased signal is presumably not a direct result of nick sealing itself, but likely results from LIG1 translocating away from the sealed nick and, therefore, away from the TFAM fluorophore. $k_{obs,4}$ is a likely candidate for the final product release step, where LIG1 fully disengages the sealed DNA product (Figure 2.5e). In support of this, we observed that $k_{obs,4}$ is close to the turnover number (k_{cat}) measured in steady-state ligation assays with the TFAM-44mer substrate, which would indicate that product release is rate-limiting for steady-state ligation of the TFAM substrate (Table 2.2).

Intrinsic tryptophan fluorescence reports on the initial binding interactions

Although the TFAM reporter provided us with robust signal with which to characterize the initial binding steps, the peculiar effects of the reporter on catalysis did give us some pause in drawing conclusions from data collected with the substrate. Due to this, we sought to verify our

findings using an alternative method that did not carry the same caveats as the TFAM substrate. Since LIG1 has a tryptophan residue—W742—in its active site pocket (Figure S2.3a-b), we reasoned that internal tryptophan fluorescence may be a viable method for measuring interactions between the nick site and LIG1. It is worth noting that LIG1 contains a total of six tryptophan residues—two in each of the three domains—but only W742 appears as though it would be impacted by interactions between LIG1 and DNA (Figure S2.3c-d).

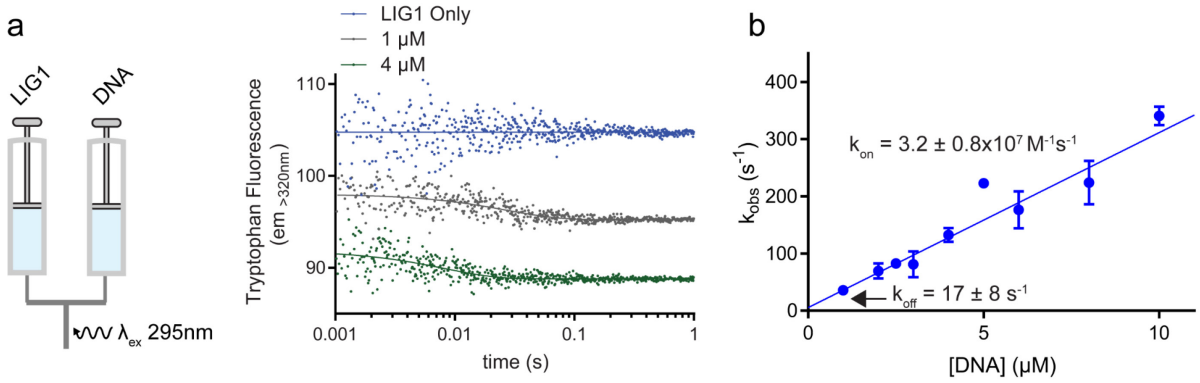


Figure 2.6. Intrinsic tryptophan fluorescence reports on DNA binding. (**a**) Stopped-flow experiments containing an increasing amount of the 28mer DNA substrate and 0.5-1 μM LIG1 were monitored using intrinsic protein tryptophan fluorescence with an excitation wavelength of 295 nm and a 320 nm cutoff longpass filter. Representative traces are shown from the rapid mixing of LIG1 with 0-4 μM DNA. Single exponential fits of the individual traces gave rate constants that are linearly dependent on the concentration of added DNA substrate (**b**). All values are shown as the mean ± S.D. of at least three replicate experiments and are reported in Table 2.2.

The advantage of using tryptophan fluorescence is that it did not require us to alter either the protein or the DNA substrate. Because of this, we were able to repeat stopped-flow experiments with the well-defined 28mer substrate[41]. Instead of monitoring the fluorescence changes as a function of increasing LIG1, we were required to use an increasing excess of the 28mer substrate for observable signal changes. However, multiple turnovers were avoided since ATP was not included in the stopped-flow reaction mixes. In the absence of magnesium, rapid mixing of LIG1 with increasing amounts of the 28mer substrate resulted in single-exponential decreases in tryptophan fluorescence (Figure 2.6). We were able to determine a linear

dependence of the rate constants on substrate concentration, with a k_{on} nearly equal to the value obtained using the TFAM-44mer substrate, indicating that the TFAM fluorophore may not have impacted initial binding by LIG1 (Table 2.2).

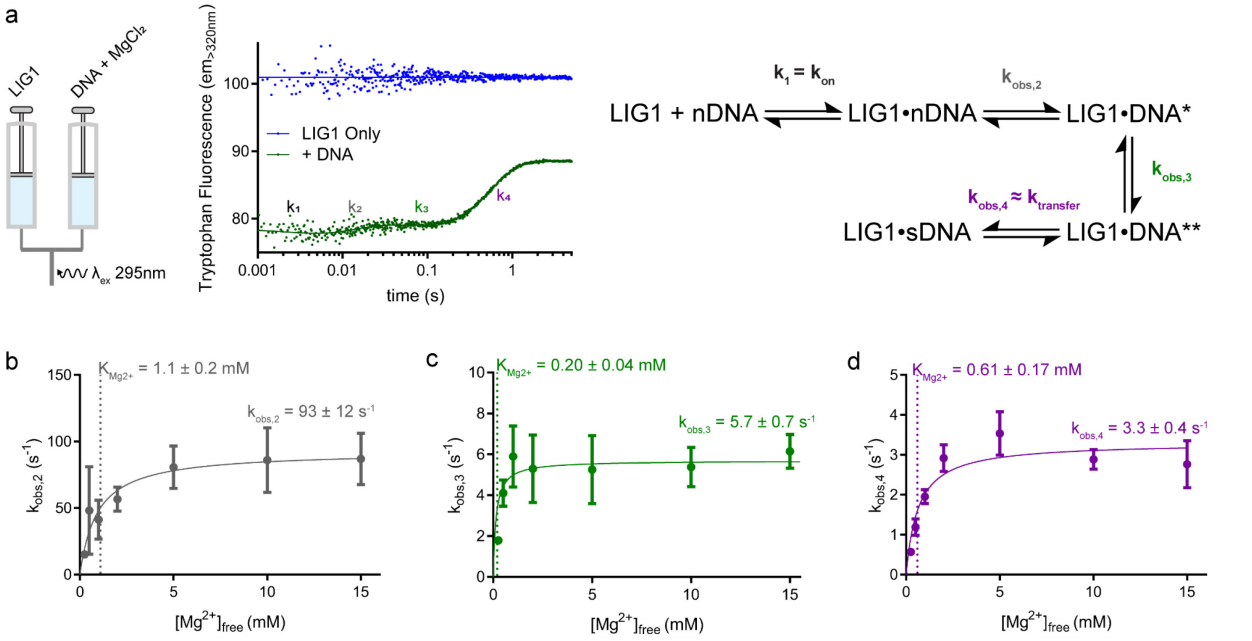


Figure 2.7. Magnesium dependence reveals two magnesium-dependent steps preceding catalysis. (**a**) Tryptophan fluorescence was monitored upon mixing of 1 μM [LIG1] with 5 μM of the 28mer substrate and increasing amounts of $MgCl_2$. Representative traces are shown from mixing of LIG1 either with DNA or buffer in the presence of 10 mM [$MgCl_2$]. Representative traces at multiple concentrations of $MgCl_2$ are shown in Figure S3. The individual phases are mapped onto a hypothetical model. Unlike the model in Figure 2.5, we observe two phases ($k_{obs,2}$, $k_{obs,3}$) that do not match the known chemical steps for ligation and are attributed to two distinct conformational changes. Traces were fit to four-exponentials and the corresponding rate constants are plotted as a function of [Mg^{2+}] in **b-d**, except for $k_{obs,1}$, which was omitted due to poor resolution of the first phase in the fluorescence traces. All observed rate constants have a hyperbolic dependence on [Mg^{2+}] resulting in magnesium affinities in the low millimolar range. All values are shown as the mean ± S.D. of at least three replicate experiments and are reported in Table 2.2.

Intrinsic tryptophan fluorescence reveals putative conformational changes

The amplitudes of the tryptophan fluorescence changes are much smaller than those observed with fluorescein, likely due to increased background signal from other tryptophan residues and an inner filter effect caused by excess DNA. While this did not significantly affect data collection in experiments lacking magnesium, it did hinder our ability to accurately monitor the initial binding event—$k_{obs,1}$—in the presence of magnesium (Figure 2.7a). Due to this, we focused on characterizing the magnesium-dependence of the steps subsequent to initial binding.

The resulting traces from the stopped-flow Mg^{2+}-dependence exhibit a similar behavior to those seen using the TFAM reporter (Figure 2.7a). As stated above, the first phase—corresponding to the initial binding event—is poorly defined but does not appear to be dependent on magnesium concentration (Figure S2.4). The following three phases all show a hyperbolic dependence on the concentration of magnesium. $k_{obs,2}$ was not seen using TFAM fluorescence and is much higher than that of any chemical steps of ligation (Figure 2.7b). The subsequent phase—corresponding to $k_{obs,3}$—is also not seen in the TFAM fluorescence trace and saturates at a low concentration of magnesium, with a K_{Mg2+} value of ~0.1 mM (Figure 2.7c). While further experiments are needed to directly address these two phases, it is plausible that they correspond to changes in the DNA or protein upon formation of the Michaelis complex. It is well-known that DNA ligases distort the helix in order to properly position the nick in the enzyme active site and the active site tryptophan may be sensitive to these events. Since both the 3'-OH and the 5'-PO_4 need to be properly aligned for catalysis, these two phases could represent the independent positioning of the two ends of the nick. Evidence supporting this model is explained further in the discussion.

Plotting $k_{obs,4}$ as a function of magnesium concentration gives a maximal rate constant and a K_{Mg2+} value that both match those of the adenylyl-transfer step of ligation of the 28mer substrate (Figure 2.7c, Figure S2.5). The signal change for this phase likely results from release of AMP following catalysis, with both the rates of nick sealing and AMP release being much faster than adenylyl transfer. While not well-resolved in the magnesium dependence traces, a final phase—$k_{obs,5}$—is observed and is likely a final product release step. Defining this phase is the focus of more recent experiments, discussed below in more detail.

Ligation of upstream ribonucleotides enhances $k_{obs,3}$

Our observations with both TFAM and tryptophan fluorescence demonstrate that multiple steps occur after DNA-binding but prior to ligation by LIG1. One such phase is an unidentified step in our tryptophan fluorescence traces—$k_{obs,3}$—which is magnesium dependent and does not obviously correspond to any of the chemical steps (Figure 2.7a,c). Since the upstream helix is known to be distorted into a more A-form-like helix, we created a variant of the 28mer containing an upstream strand composed of all ribonucleotide bases with the thought that a DNA-RNA hybrid should be more amenable to this type of remodeling by LIG1. Indeed, pilot stopped-flow experiments with this 28mer variant mimicked those with the DNA-only 28mer, except for $k_{obs,3}$, which is roughly 2-fold faster with the DNA-RNA hybrid (Figure 2.8, Table 2.2). We also observe this upstream remodeling step with the TFAM-44mer substrate ($k_{obs,2}$ in Figure 2.5). The AT-rich nature of the TFAM-44mer substrate may lower the barrier for this remodeling to occur, consistent with the faster rate observed with the TFAM-44mer substrate.

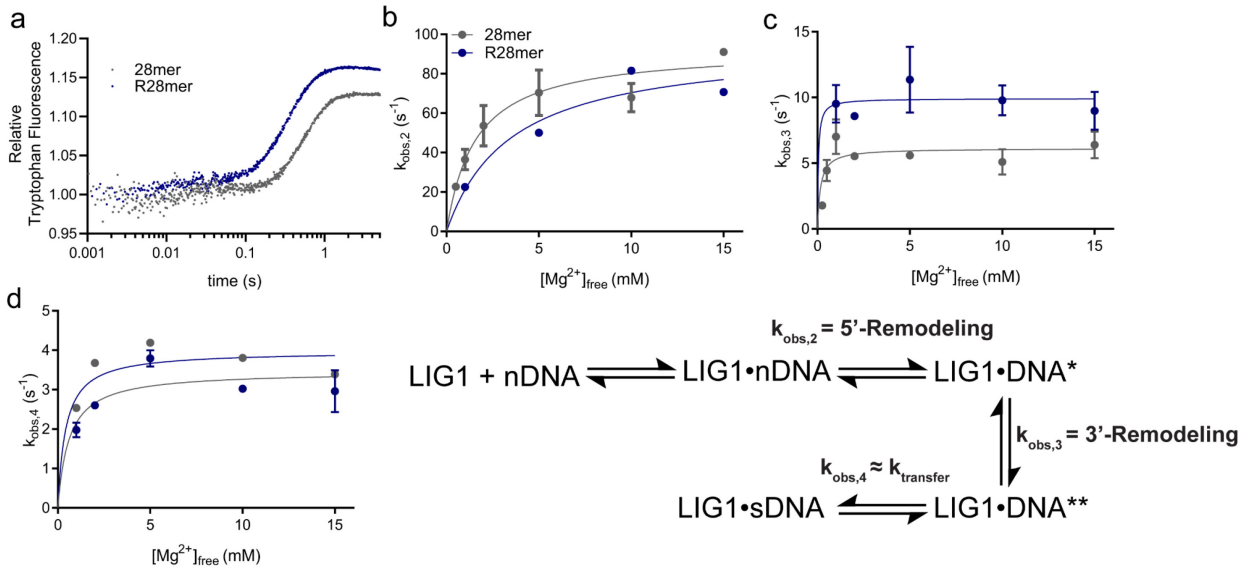


Figure 2.8. Upstream ribonucleotides enhance the rate of a step preceding chemistry. (**a**) Representative stopped-flow traces from experiments with a variant of the 28mer containing an all-ribonucleotide sequence upstream of the nick. Data shown for the DNA-only 28mer are adapted from Figure 2.7. (**b-d**) Observed rate constants from ligation of the 28mer and the upstream ribo-28mer have similar magnesium dependences, except for $k_{obs,3}$, which is 2-fold faster with the ribo-28mer. Data for the ribo-28mer are the result of duplicate experiments and the best-fit values are shown as the mean ± S.D. in Table 2.2.

With this in mind, we believe that another of the phases observed using tryptophan fluorescence, designated as $k_{obs,2}$ (Figure 2.7a-b), could be reporting on interactions between LIG1 and the 5' end of the nick. We have demonstrated that this step is magnesium dependent, exhibiting a low K_{Mg2+} value. Recent crystal structures of LIG1 illustrate the involvement of the 5'-phosphate in the coordination of the catalytic magnesium, so it is plausible that Trp742 would be sensitive to these interactions. The fact that the TFAM reporter seems unable to report on this phase is consistent with this proposed model, as it is unlikely the TFAM reporter on the 3' end of the nick would be sensitive to the changes in the active site or at the 5' end.

Addition of ATP demonstrates the presence of a product-release step

Finally, we wanted to confirm whether the final phase in our stopped-flow traces corresponds to product release. While we did note that the final observed rate in the TFAM fluorescence traces is consistent with k_{cat}, indicating a possible rate-limiting product release, we wanted to probe this step directly. Based on the idea that deadenylylated LIG1 should prefer binding ATP to remaining bound to the sealed DNA product, we expected the addition of ATP to enhance the rate constant for a product-release step. Since tryptophan fluorescence should also be sensitive to the DNA-independent steps of ligation, we wanted to test whether we could observe LIG1 transitioning from a DNA-bound to an ATP-bound state.

We repeated stopped-flow experiments, rapidly mixing the LIG1•DNA complex with $MgCl_2$ and increasing amounts of ATP (Figure 2.9, Table 2.2). To prevent multiple turnovers, we performed these experiments with a slight excess of LIG1. We observed a hyperbolic dependence of the phase on the concentration of ATP, giving a K_{ATP} value of ~10 μM and a maximal rate constant of ~0.4 s^{-1}, both of which are consistent with steady-state $K_{M,ATP}$ and k_{cat}

values[41]. With only the data presented here, it is difficult to know whether this rate constant is a consequence of slow product release or slow ATP-binding, although ATP-binding by other ligases is rapid[121].

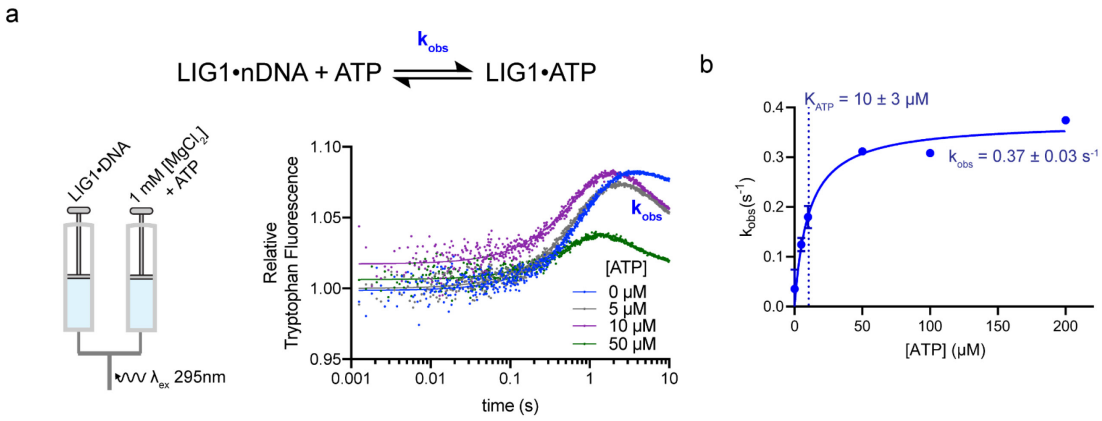


Figure 2.9. ATP enhances the rate of LIG1 product release. (**a**) To probe the product release step, a complex of 1.1 μM [LIG1] and 1 μM of the 28mer substrate was mixed with 1 mM Mg^{2+} and increasing amounts of ATP. Each representative trace was normalized to its initial fluorescence signal for clarity. (**b**) Exponential fits of the final phase in each trace are plotted as a function of increasing ATP, illustrating a hyperbolic dependence on ATP concentration. The data are the result of duplicate experiments and the best-fit values are displayed as the mean ± S.D. and are also in Table 2.2.

a

	TFAM-44mer
$K_{D,specific}$ (nM)	4.9 ± 1.7
$K_{D,nonspecific}$ (nM)	4.7 ± 1.8 x 10^3
k_{on} ($M^{-1}s^{-1}$)	1.2 ± 0.8 x 10^8
k_{off} (s^{-1})	35 ± 18
$k_{3'\text{-}Remodel}$ (s^{-1})	10 ± 1
k_{chem} (s^{-1})	2.6 ± 0.2
$k_{release}$ (s^{-1})	0.093 ± 0.010

b

	Tryptophan Fluorescence	
	28mer	R28mer
k_{on} ($M^{-1}s^{-1}$)	3.2 ± 0.8 x 10^7	ND
k_{off} (s^{-1})	17 ± 8	ND
$k_{5'\text{-}Remodel}$ (s^{-1})	93 ± 12	95
K_{Mg2+} (mM)	1.1 ± 0.2	3.497
$k_{3'\text{-}Remodel}$ (s^{-1})	5.7 ± 0.7	9.9 ± 0.8
K_{Mg2+} (mM)	0.20 ± 0.04	ND
k_{chem} (s^{-1})	3.3 ± 0.4	3.9 ± 0.2
K_{Mg2+} (mM)	0.61 ± 0.17	0.42 ± 0.19
$k_{release}$ (s^{-1})	0.37 ± 0.02	ND
K_{ATP} (mM)	10 ± 3	ND

Table 2.2. Best-fit parameters from real-time and equilibrium fluorescence assays. Parameters obtained from stopped-flow experiments using either TFAM-44mer (**a**) or tryptophan residues (**b**) as the fluorescence reporter. In **a**, ,k_{on} and k_{off} values represent the averages of the values obtained from experiments with and without magnesium. In **b**, the values shown were obtained with either the 28mer and a ribonucleotide derivative of the 28mer (R28mer). Values that were not determined are marked by ND.

Discussion

Since this study remains a work-in-progress, several questions about the binding mechanism of LIG1 remain. However, we have described the first direct observation of nick-binding by a human DNA ligase. Although the deoxythymidine-conjugated reporter appears to adversely affect the chemical steps of ligation, we have demonstrated its potential as a reporter of DNA-binding and catalysis by LIG1. Additionally, we were able to observe a previously-uncharacterized event preceding chemistry, possibly corresponding to a conformational change in the DNA or the protein.

Despite some perturbations to DNA ligation caused by the presence of the TFAM reporter, most of the conclusions drawn from its use were validated using intrinsic tryptophan fluorescence. LIG1 has a single tryptophan residue—W742—in its active site, which we presumed to be sensitive to interactions between the protein and the nicked DNA. Although LIG1 contains additional tryptophan residues, W742 is the only one in proximity to a LIG1•DNA contact. Using tryptophan fluorescence and an unmodified 28mer substrate, we were able to repeat stopped-flow experiments in the absence of any perturbations caused by the TFAM fluorophore. Additionally, tryptophan fluorescence enabled us to monitor the DNA-independent steps of the ligation catalytic cycle. By adding ATP into our stopped-flow experiments, we were able to demonstrate that the final observed phase in our data corresponds to a product-release step, followed by rapid binding to ATP.

Monitoring tryptophan fluorescence traces as a function of increasing magnesium concentrations revealed the existence of two distinct exponential phases in our stopped-flow experiments. While further characterization of these two phases is still necessary, we believe that

this points to a two-step remodeling event, where LIG1 distorts the nick site to position both ends of the nick in the active site for catalysis. Using a modified substrate containing a full ribonucleotide upstream strand, we demonstrated the possibility that at least one of these uncharacterized phases may correspond to a helical distortion at the 3' end of the nick.

Considering the presented data, we believe that our initial model presented in Figure 2.1 is valid, with a few amendments (Figure 2.10 and 2.11). First, we are not able to observe LIG1 translocation across the DNA helix in search of a nick with either TFAM or tryptophan fluorescence. Considering that the LIG1 domains span the length of the entire 18bp DNA substrate in crystal structures of the LIG1•DNA complex, we would need a significantly longer substrate to slow down a putative finding step for observation. In the current study, we can consider the nick-finding step to be very fast, making the first observable event direct association with the nick. Second, we believe that the remodeling step proposed in the initial model can be further expanded. Remodeling likely consists of the alignment of the 5'-PO_4 in the active site pocket, which is followed by a distortion of the upstream helix to position the 3'-OH for catalysis. While experiments with the upstream RNA substrate provide evidence for the presence of the 3' remodeling, we have yet to collect direct evidence validating the 5' remodeling event.

	Initial Binding	5'-Remodeling	3'-Remodeling	Chemistry
Tryptophan Fluorescence	3.2 ± 0.8 x 10^7 $M^{-1}s^{-1}$	87 ± 17 s^{-1}	6.1 ± 0.7 s^{-1}	2.8 ± 0.5 s^{-1}
	LIG1 + nDNA ⇌ LIG1•nDNA ⇌ LIG1•DNA* ⇌ LIG1•DNA** ⇌ LIG1•sDNA			
TFAM Fluorescence	1.2 ± 0.8 x 10^8 $M^{-1}s^{-1}$	Not observed	10 ± 1 s^{-1}	2.6 ± 0.2 s^{-1}

Figure 2.10. Unified model for both TFAM and tryptophan fluorescence experiments. Rate constants for the proposed steps following initial binding are derived from experiments in the presence of 20 mM Mg^{2+} to allow a direct comparison of TFAM and tryptophan fluorescence experiments. Values for the tryptophan fluorescence mechanism were obtained using the 28mer substrate while the TFAM fluorescence values were all obtained using the modified 44mer substrate.

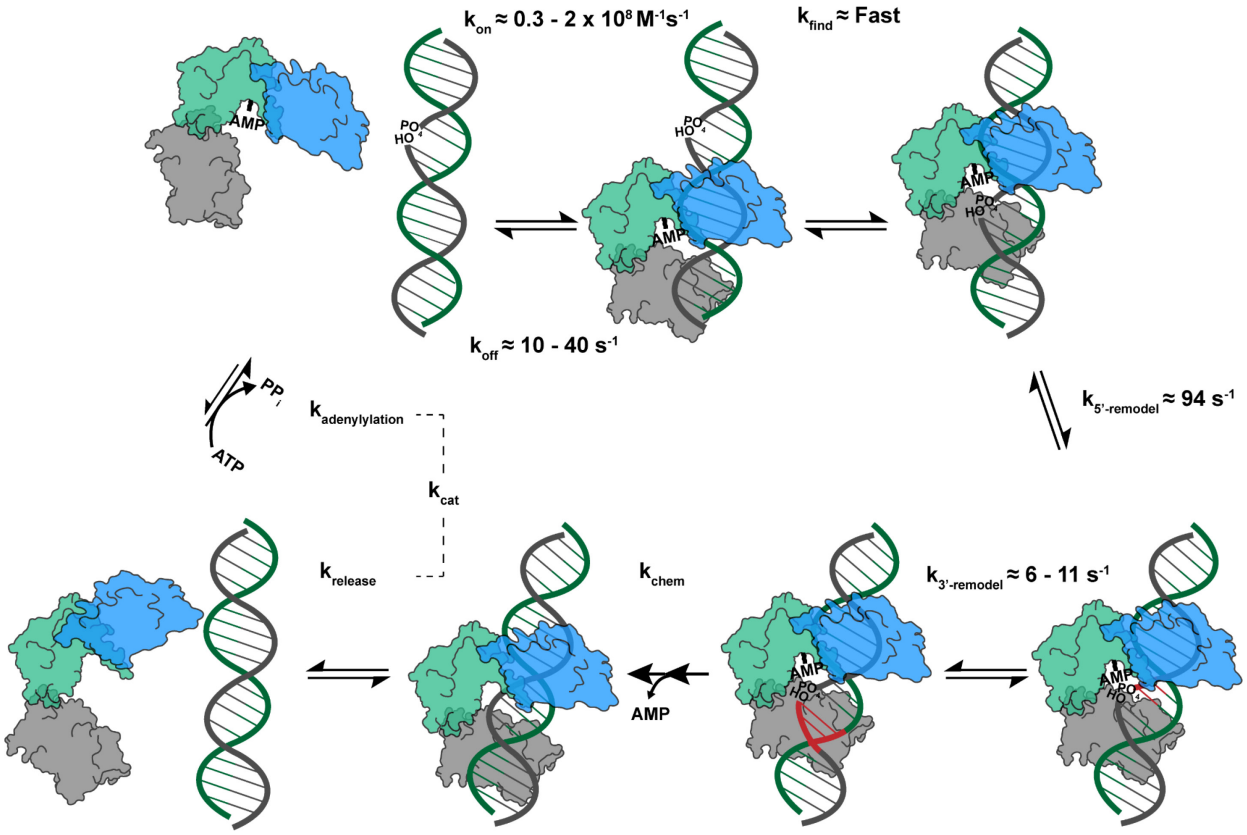


Figure 2.11. Revised model for the full catalytic cycle of LIG1. The estimated rates of each step are derived for the data in Table 2.2 and are ranges of the data collected using either TFAM or tryptophan fluorescence. The model mostly reflects the proposed model in Figure 2.1, except for the remodeling step, which we believe is a two-step process. This model was used in Berkeley Madonna to fit experimental data collected using tryptophan fluorescence and the resulting fit is shown in Figure S2.6.

Chapter 2 Appendix

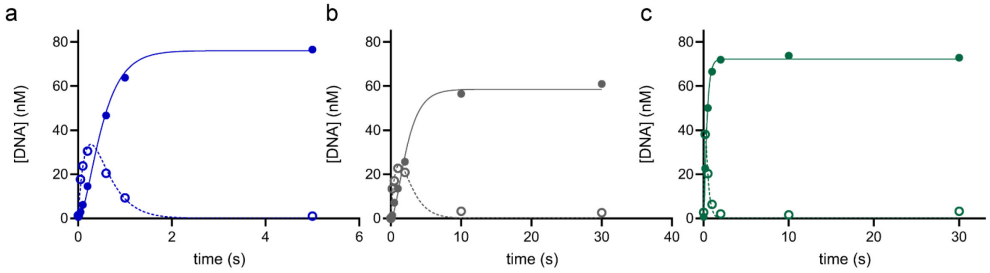


Figure S2.1. Representative single-turnover traces. Traces are shown for ligation of the 28mer (**a**), TFAM-44mer (**b**) and 44mer (**c**) substrates in the presence of 1 mM Mg^{2+}. Product formation is indicated with closed circles, while intermediate formation and decay is represented with open circles. The corresponding fits from Berkeley-Madonna are shown as solid (product) or dashed (intermediate) lines.

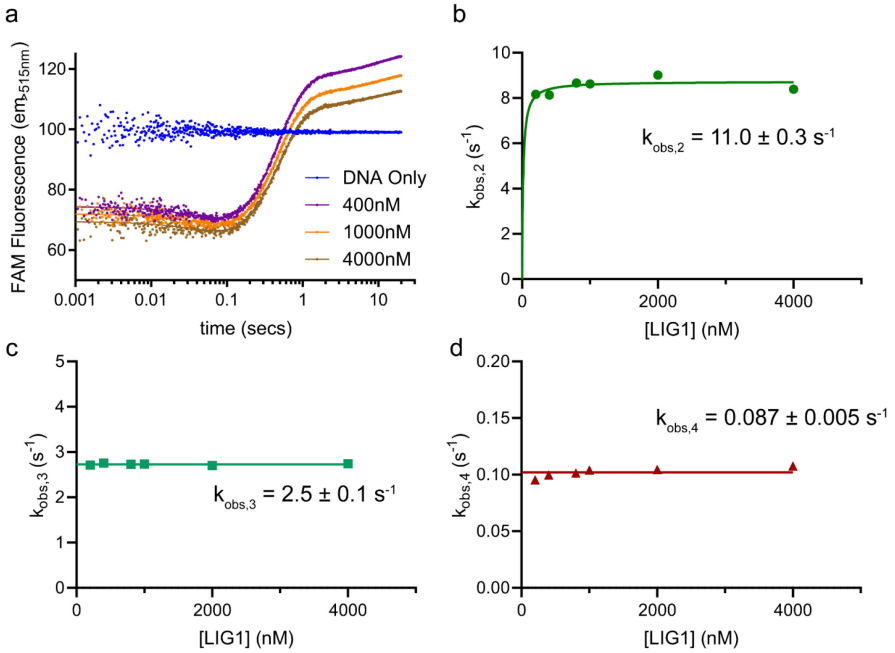


Figure S2.2. Rapid mixing of a preformed LIG1•TFAM-44mer complex with magnesium. Same experiments as in Figure 2.5 except DNA and increasing amounts of LIG1 were incubated together prior to mixing with magnesium. A trace amount of EDTA (0.1 mM) was added to the complex solution to prevent premature catalysis. The resulting traces (**a**) show the same features as in Figure 2.5 except for the absence of the initial binding phase corresponding to $k_{obs,1}$. Traces were fit to three exponentials and the resulting rates are plotted as a function of [LIG1] in **b-d**. The corresponding values were averaged together with the experiments in Figure 2.5 and the resulting means ± S.D. are displayed in Figure 2.5 and in Table 2.

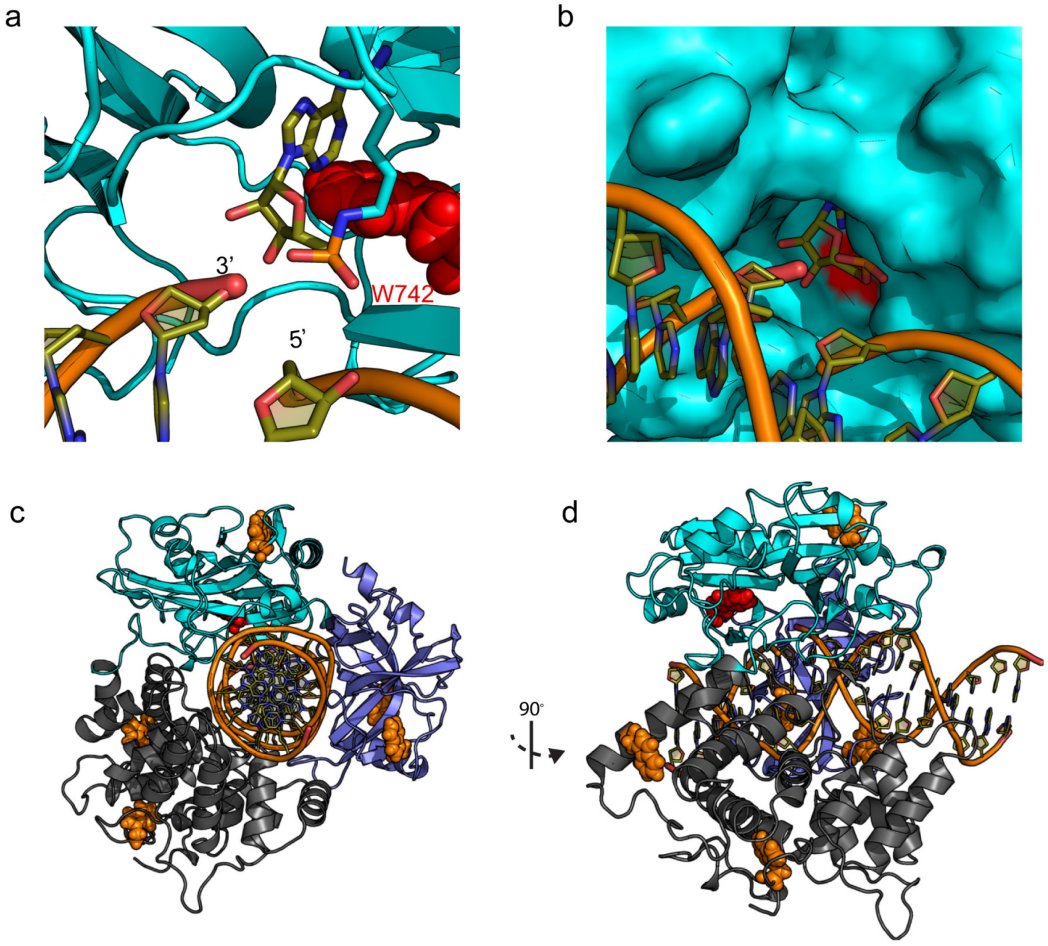


Figure S2.3. LIG1 has a single tryptophan residue in its active site pocket. (**a**) Cartoon representation of the active site of LIG1 highlighting Trp742 (red) and the adenylylated lysine in stick representation. (**b**) Surface representation of the active site with the location of Trp742 highlighted in red. (**c-d**) Cartoon representation of the full LIG1 structure with all tryptophans represented as space-filling models. Trp742 remains highlighted in red while the other five tryptophan residues are shown in orange. Figure adapted from the crystal structure of LIG1 presented in Chapter 3.

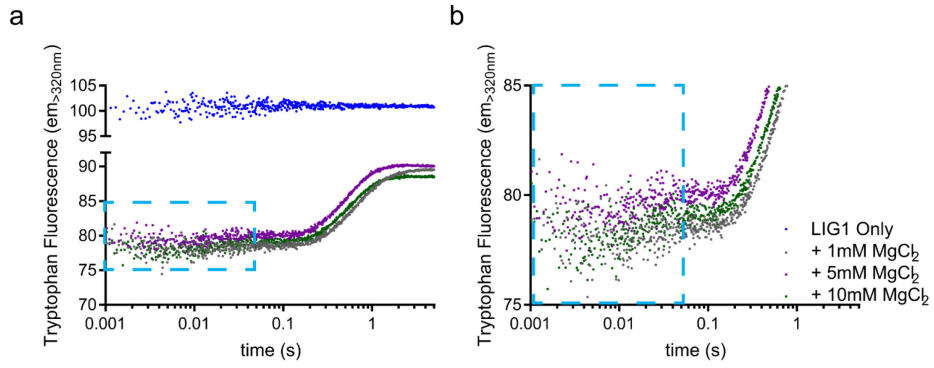


Figure S2.4. Magnesium does not noticeably affect initial binding. (**a**) Same experiment as in Figure 2.7A shown with two other tested concentrations of magnesium. The blue box highlights the region corresponding to $k_{obs,1}$, which does not appear to change in response to an increasing concentration of added magnesium. (**b**) Magnified view of the traces in **a** to highlight the low signal-to-noise of the first phase. The blue box is highlighting the same area as in **a**.

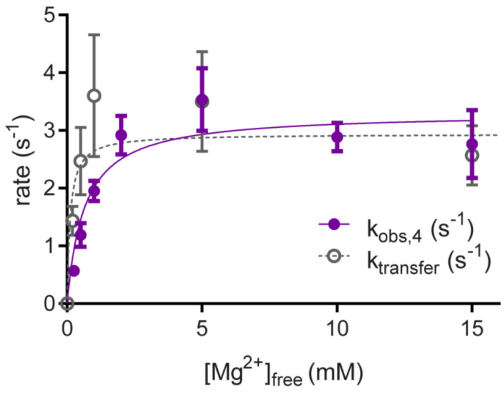


Figure S2.5. Overlay of $k_{obs,4}$ and $k_{transfer}$. Data for $k_{obs,4}$ (Figure 2.7) and for $k_{transfer}$ (Figure 2.3b, 28mer) superimposed to highlight the similarity of the two magnesium dependences.

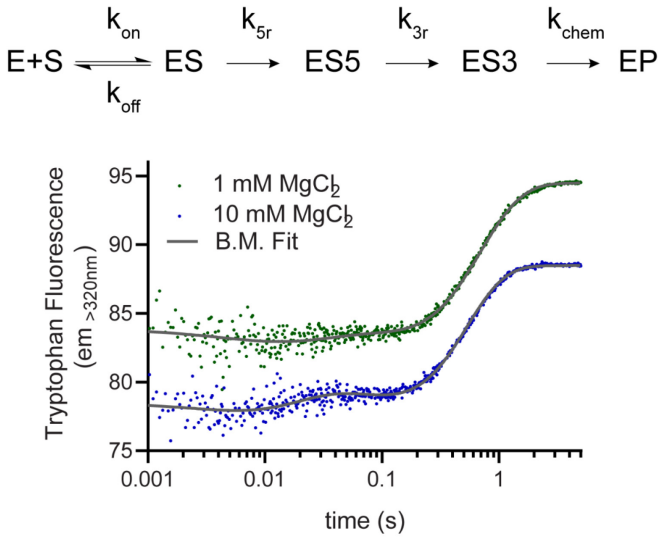


Figure S2.6. Berkeley-Madonna model fitting. Berkeley-Madonna was used to fit traces observed with tryptophan fluorescence to a simplified version of the model in Figure 2.10 (top). Since a nick-finding step was not observed, the model here assumes direct binding to the nick site by LIG1. Additionally, both adenylyl transfer and nick sealing are condensed down to just a single chemical step, since individual signals resulting from the two steps were not observed. Finally, product release is omitted from the model, as it was not observed under the conditions used. A plot of the resulting fit is shown with the experimental data in the graph below. The parameters used for fitting are further detailed in Figure S2.6.

Figure S2.7. Berkeley-Madonna model and parameters used for fits in S2.6.

```
METHOD RK4
STARTTIME = 0.001
STOPTIME=5
DT = 0.02

d/dt(E) = koff*ES - kon*E*S
d/dt(S) = koff*ES - kon*E*S
d/dt(ES) = kon*E*S - koff*ES - k5r*ES
d/dt(ES5) = k5r*ES - k3r*ES5
d/dt(ES3) = k3r*ES5 - kchem*ES3
d/dt(EP) = ES3*kchem

init(E) = 1e-6
init(S) = 5e-6
init(ES) = 0
init(ES5) = 0
init(ES3) = 0
init(EP) = 0

F = FA*E + FB*ES + FC*ES5 + FD*ES3 + FE*EP + os
FA = 8e7
FB = 7.5e7
FC = 7.7e7
FD = 8e7
FE = 8.6e7
os = 10
```

1 mM Mg^{2+} Parameters

```
kon = 3e7
koff = 14
k5r = 40
k3r = 6
kchem = 2
```

10 mM Mg^{2+} Parameters

```
kon = 3e7
koff = 14
k5r = 86
k3r = 6
kchem = 3
```

Figure S2.8. Oligonucleotides used for this study.

Oligo Name	5′ modification	Sequence	3′ modification
Temp28		5' CCGAATCAGTCCGACGACGCATCAGCAC 3'	
Up13OH		5' GTGCTGATGCGTC 3'	
UpRNA13OH		5' rGrUrGrCrUrGrArUrGrCrGrUrC 3'	
P15FAM	Phosphate	5' P-GTCGGACTGATTCGG-FAM 3'	FAM
P15	Phosphate	5' P-GTCGGACTGATTCGG 3'	
Temp44		5' TCGCGGCGATTATATTTATATTATTAATATTATAATGCGCGGCG 3'	
Up22OH		5' CGCCGCGCATTATAATATTAAT 3'	
Up22-TFAM		5' CGCCGCGCATTATAATATTAA-dTFAM 3'	dT-FAM
DownP22FAM	Phosphate	5' P-AATATAAATATAATCGCCGCGA-FAM 3'	FAM
DownP22	Phosphate	5' P-AATATAAATATAATCGCCGCGA 3'	
Down22		5' AATATAAATATAATCGCCGCGA 3'	

Chapter 3
Mechanism for Hi-Fidelity Ligation: Discrimination Against 3' Mismatches

Most efforts to understand the overall fidelity of DNA metabolism have focused on the DNA polymerization steps, proofreading, and sanitization of the nucleotide pool. However, it is known that DNA ligases also contribute to the fidelity of these pathways by discriminating against mismatches and damaged bases at the 3' side of DNA nicks. X-ray crystal structures of the three mammalian DNA ligases (LIG1, LIG3 and LIG4) in complex with DNA have revealed encirclement of the DNA nick by a conserved ligase three-domain core architecture[15,24,25]. Yet, despite these similarities, these ligases display different fidelity profiles. LIG1 and its homologs are tuned as high-fidelity enzymes (e.g., human DNA ligase 1 homologs)[122-124], while LIG4 homologs repair have relaxed stringency for correct DNA base-pairing during DNA nick sealing[125]. However, the molecular underpinnings of ligation fidelity remain unknown.

The ability of LIG1 to discriminate against 3' mismatches has been previously shown to be more stringent than discrimination by LIG3, indicating a feature of LIG1 that allows for high-fidelity ligation[123]. Discrimination by LIG1 has also been shown with the ligation of substrates containing 8-oxo-deoxyguanosine (8oxoG) nucleotides at the 3' end of the nick, resulting in a significant decrease in formation of the sealed DNA product[124]. Ligation of these oxidized 3' ends results in increased abortive ligation behavior by LIG1, which releases the adenylylated intermediate DNA species prior to nick sealing.

The work reported in Chapter 3 is adapted, with permission from the authors, from an unpublished manuscript entitled Two-Tiered Enforcement of High-Fidelity DNA Ligation. The authors are as follows: Percy Tumbale[#], Thomas Jurkiw[#], Matthew J. Schellenberg, Amanda Riccio, Patrick O'Brien*, and R. Scott Williams*. The crystal structures presented here represent preliminary structures obtained by P.T., M.J.S. and A.R. under the guidance of R. Scott Williams at the National Institute of Environmental Health Sciences.

The product of abortive ligation is the substrate for the HIT-family hydrolase, aprataxin (APTX)[112,126]. Loss-of-function mutations in APTX cause a neurodegenerative disease known as ataxia with oculomotor ataxia 1 (AOA1), suggesting that unrepaired abortive ligation products can be cytotoxic[114,115].

In this work we establish a foundation for understanding high-fidelity DNA ligation by LIG1. We report a series of high resolution-structures of wild-type and mutant LIG1 bound to correctly base-paired or damaged DNA. The unexpected finding of a conserved Mg^{2+} binding site forming a junction between DNA ligase domains and the 3'-strand led us to investigate the contribution of this enforced conformation to catalysis and to substrate discrimination. We find that this site is dispensable for both adenylyl transfer and nick sealing, but critical to the discrimination against improper base pairs at the 3' end of the nick. Steady-state and pre-steady state kinetic analysis reveals that this high-fidelity (HiFi) Mg^{2+} architecture allows for discrimination, leading to slower DNA adenylylation and increased abortive ligation, which competes with nick sealing to restrict improper ligation. We show that coupling between LIG1 and aprataxin (APTX) provides a second tier of discrimination by preferentially hydrolyzing AMP-DNA intermediates that contain improper 3' termini. Conservation of the HiFi site in LIG1 homologs across species points to the importance of high-fidelity DNA ligation throughout evolution. Additionally, the absence of the HiFi site in LIG3 and LIG4 could contribute to different specialization of these orthologous DNA ligases.

Materials and Methods

Preparation and purification of synthetic oligonucleotides. The sequences of all oligos used in the study for both crystallization and kinetic experiments are listed in Figure S3.9. Oligonucleotides were purchased from IDT, Midland Certified Reagent Company, or the Keck Center at Yale University and purified by denaturing PAGE as previously described[41]. For purification of 8oxoG-containing oligonucleotides, 10 mM 2-mercaptoethanol was included in the loading buffer and gel.

Preparation and purification of recombinant enzymes. WT and H260N human APTX were expressed and purified as previously described[122]. Expression and purification of the minimal catalytic domain of human LIG1 (232–919) used in all kinetic assays was performed as previously described[106]. The E346A/E592A LIG1 mutant was constructed using site-directed mutagenesis with synthetic primers, and successful mutagenesis was confirmed by sequencing of the coding region. Protein concentrations were estimated using absorbance at 280 nm. Active enzyme concentrations were determined using single-turnover ligation reactions, in which increasing concentrations of LIG1 were incubated with a fixed concentration of the fluorescein-labeled nicked 28mer substrate [41].

LIG1 constructs consisting of either residues 232-916 (Δ232) or 262-904 (Δ262) were used for crystallographic experiments. Mutations in either the Δ232 or Δ262 LIG1 proteins were created using site-directed mutagenesis. *E. coli* Rosetta 2 (DE3) cells containing the expression vectors were grown at 37ºC in Terrific Broth supplemented with ampicillin (100 µg/mL) and chloramphenicol (34 ng/mL) until the OD_{600nm} reached 1, at which 50 µM IPTG was added. Protein expression was carried out at 16ºC overnight. Cells were harvested by centrifugation (5000 rpm, 20 min). Cell pellet was suspended and lysed in 30 mL lysis buffer (50 mM Tris, pH

8.5, 500 mM NaCl, 10 mM imidazole, 0.1 g lysozyme/1L pellet, 1 tablet Roche mini EDTA-free protease inhibitor cocktail) at 4ºC for 30 min, followed by sonication. The cell lysate was fractionated by centrifugation (13000 rpm, 20 min, 4ºC). The soluble fraction was applied to Ni-NTA resins (5 ml packed volume) which has been equilibrated with 15 ml (50 mM Tris, pH 8.5, 500 mM NaCl, 10 mM imidazole). The column was washed with 100 ml (50 mM Tris, pH 8.5, 500 mM NaCl, 10 mM imidazole), 15 ml (50 mM Tris, pH 8.5, 500 mM NaCl, 30 mM imidazole), and the His-tagged protein was eluted in 15 ml (50 mM Tris, pH 7.5, 500 mM NaCl, 300 mM imidazole). The His-tag was removed by TEV protease at 4ºC overnight. The untagged protein was purified on HiLoad 16/600 Superdex 200 gel filtration column in buffer (25 mM Tris, pH 7.5, 150 mM NaCl, 1 mM TCEP, 0.1 mM EDTA), followed by HiTrap SP HP 5 ml cation exchange column (low salt buffer: 20 mM Tris 7.5, 0.2 mM EDTA, 1 mM TCEP; high salt buffer: 20 mM Tris 7.5, 1 M NaCl, 0.2 mM EDTA, 1 mM TCEP). The quality of the purified proteins was analyzed by SDS PAGE. Freshly purified proteins were used immediately in crystallization experiment.

Crystallization of wildtype and mutant LIG1 complexes. Crystals of the LIG1•DNA•EDTA complex were grown by hanging drop method. The protein•DNA complex was formed with 20 mg/mL of either E346A/E592A or wildtype Δ262 LIG1 and the nicked 18mer oligo at a 1.5:1 DNA:LIG1 molar ratio in a solution consisting of 1 mM ATP, 150 mM NaCl, 20 mM Tris-HCl, pH 7.5, and 1 mM TCEP. 1µL of the complex solution was mixed with an equal volume of precipitant solution (100 mM MES, pH 6, 100 mM lithium acetate, 12-15% (w/v) polyethylene glycol 3350) at 20°C. The crystallization drops containing the mutant LIG1 protein were allowed to equilibrate for 30 minutes before micro-crystal seeds prepared from the WT crystals were streaked into the drops.

Crystals of the LIG1 constructs in complex with Mg^{2+} were prepared similarly except the protein•DNA complex solution contained 100 mM $MgCl_2$ and a variant of the 18mer nicked DNA substrate that is blocked at the 3' end of the nick with a dideoxycytidine. Crystals of E346A/E592A LIG1 in complex with the nicked 3'-8oxoG:A substrate were grown in the presence of EDTA and the 8oxoG:A variant of the 18mer substrate. The crystallization drops were allowed to equilibrate for 30 minutes before micro-seeding with crystals of the E346A/E592A LIG1•DNA•EDTA complex. Crystals grew in 1-2 days and were washed in cryo-protectant (40% PEG3350 in precipitant solution) before being flash frozen in liquid nitrogen for data collection. X–ray diffraction data was collected on beamline 22-ID of the Advanced Photon Source at a wavelength of 1.000 Å. X-ray diffraction data were processed and scaled using the HKL2000 suite[127]. WT LIG1 structures were solved by molecular replacement using PDB entry 1X9N as a search model with PHASER[128,129]. The remaining structures were solved by refinement in PHENIX[130] using PDB entry 1X9N as a starting model. Iterative rounds of model building in COOT[131] and refinement with PHENIX were used to produce the preliminary models shown in the study.

Gel-based Ligation Assay. Ligation assays were carried out at 37 °C in a standard reaction buffer consisting of 50 mM NaMOPS at pH 7.5, 1 mM dithiothreitol, 0.05 mg/ml BSA, and sufficient NaCl to maintain a constant ionic strength of 150 mM. Concentrations of ATP, $MgCl_2$, 28mer substrate and LIG1 were varied as indicated below. For reactions stated to be at physiological conditions, final concentrations of 1 mM ATP and 2 mM $MgCl_2$ were used. The free magnesium concentration for these reactions was calculated to be 1 mM using the dissociation constants for the Mg^{2+}•ATP complex[41,132] . Unless noted otherwise, reactions were quenched in the standard loading buffer (50 mM EDTA/90% formamide/0.01% xylene

cyanol/0.01% bromophenol blue). Quenched samples were heated to 95 °C and separated on either a 15 or 20% (w/v) polyacrylamide gel containing 8 or 6.6 M urea, respectively. The fluorescein-labeled oligonucleotide was detected using an Amersham Typhoon 5 imager (GE), and gel images were analyzed using ImageQuant TL software (GE). Rate constants are reported as the average ± standard deviation for at least three independent experiments.

Single-turnover ligation assays. To determine the individual rates of the adenylyl-transfer and nick-sealing steps, single-turnover reactions were performed with 800 nM Δ232 LIG1 and 80 nM DNA substrate. Reactions to determine Mg^{2+} dependence of the chemical steps were performed in a KinTek RQF-3 quench-flow apparatus, where WT or mutant Δ232 LIG1 was loaded in a separate sample loop from the 28mer-G:C substrate, both at double the desired final concentrations to account for dilution upon mixing. Both ligase and DNA solutions were prepared with the standard reaction buffer supplemented with desired concentrations of $MgCl_2$. Drive syringes were loaded with the standard reaction buffer and reactions were quenched using a solution of 1 N H_2SO_4 and 0.25% SDS. Upon quenching, 20 μL of the quenched sample was immediately neutralized in a solution containing 300 mM Tris base, 15 mM EDTA and 0.5X loading buffer. Reactions used to measure the rates of ligation for the 28mer-8oxoG:A substrate were also performed in the RQF-3 but were quenched with the standard quench solution. Reactions containing the 28mer-8oxoG:C substrate were carried out in a heat-block set at 37°C and quenched with the standard quench solution. Rates were obtained from fits to the single-turnover data using a two-step irreversible model in Berkeley-Madonna as previously described[41].

Determining the steady-state dependences of ligation. Steady-state dependences were performed using the standard reaction buffer and 0.1-10 nM WT or mutant Δ232 LIG1. For both

the Mg^{2+} and ATP dependences, 500 nM of the 28mer substrate was used along with either 200 μM ATP or 20 mM $MgCl_2$, respectively. For the DNA substrate dependences at saturating levels of $MgCl_2$ and ATP, reactions contained 20 mM $MgCl_2$ and 200 μM ATP. DNA substrate dependences were also performed under physiological conditions. Initial rates for all steady-state reactions were determined by linear regression, and the rates were fit with the Michaelis-Menten equation (Chapter 2, Equation 2.1) For reactions in which an accumulation of both intermediate and product was observed, the initial rates of total substrate disappearance were used. The fraction of observed abortive ligation events was determined by dividing the rate of intermediate accumulation ($V_{intermediate}$) by the rate of total substrate disappearance ($V_{product} + V_{intermediate}$).

Determining the effect of APTX on ligation. Both steady-state and single-turnover reactions for measuring the effect of APTX on ligation were carried out under physiological ATP and Mg^{2+} concentrations. Steady-state reactions contained 0 or 10 nM APTX, 1 nM WT or mutant Δ232 LIG1, and 500 nM nicked DNA substrate. Using initial rates (V) for total substrate disappearance in the presence and absence of APTX, the percentage of suppressed ligation was calculated as follows:

Equation 3.1 $$\text{Suppressed Ligation (\%)} = \frac{V_{-APTX} - V_{+APTX}}{V_{-APTX}} \times 100\%$$

Single-turnover reactions were initiated by 1:1 mixing of a solution containing 1600 nM WT or mutant Δ232 LIG1, 160 nM DNA substrate and 0.1 mM EDTA with a separate solution containing 0 or 1600 nM H260N APTX and 4 mM MgCl2. Both the ligase and APTX solutions contained the standard reaction buffer supplemented with 1 mM ATP. Reactions were quenched as stated above for single-turnover reactions. Data was fit in Berkeley-Madonna to a modified two-step model, wherein LIG1 partially dissociates and is displaced from the adenylylated

intermediate by H260N APTX (Figure 3.7, S3.8). The rates determined from the fits were used to calculate the catalytic commitment factor as follows:

Equation 3.2 $$\text{Catalytic Commitment} = \frac{k_{seal}}{k_{seal}+k_{off}}$$

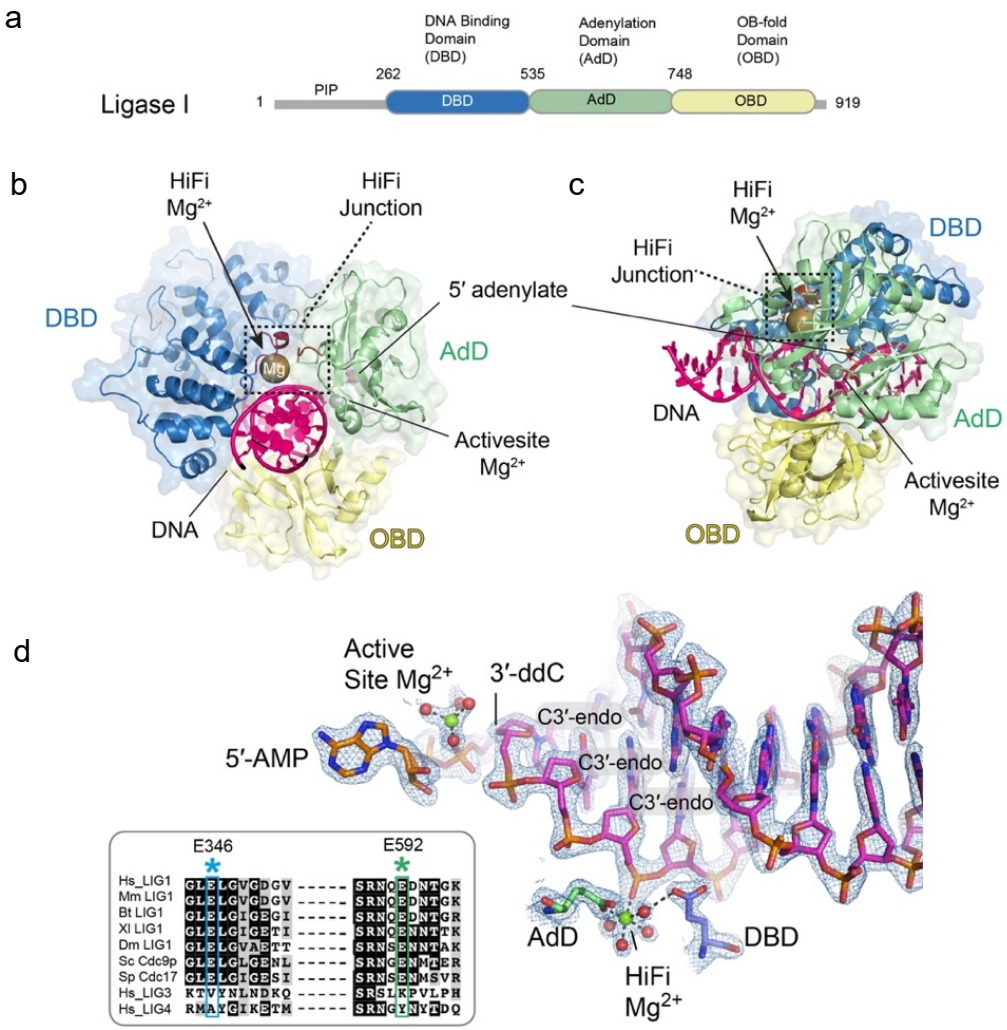


Figure 3.1. X-ray structure of LIG1•Mg2+•adenylylated nicked DNA complex. (**a**)Domain structure of LIG1. The N-terminal domain (gray) contains a nuclear localization signal (NLS) that facilitates nuclear localization, and a conserved PIP box motif that mediates interactions with PCNA. The DNA binding domain (DBD, blue), Adenylation domain (AdD, green), and OB-fold domain (OBD, yellow) comprise the catalytic core, mapping to residues 232-919. (**b**)X-ray structure of LIG1•Mg^{2+}•AMP-DNA complex. The DBD (blue), AdD (green), and OBD (yellow) domains encircle the nicked DNA substrate (pink). The 5′-phosphate of the nick is adenylylated (orange) and a Mg^{2+} (gray) are bound in the active site. The HiFi Mg^{2+} (brown) is bound in the HiFi junction highlighted in a dash-line box. (**c**) A side view of the LIG1•Mg^{2+}•AMP-DNA intermediate. (**d**) Protein sequence alignment of LIG1 and homologs, and human ligases generated by Clustal Omega shows HiFi site residues E346 and E592 are absolutely conversed among LIG1 homologs but diverse in LIG3 and LIG4. Crystal structure of wildtype LIG1 in complex highlights interactions between E346 in the DBD and E592 in the AdD with the Mg^{HiFi}.

Results

Metal-reinforced engagement to the DNA substrate by human LIG1

To define the molecular basis of LIG1 substrate recognition and fidelity we sought conditions suitable for high-resolution structural determination. We tested 16 different nicked DNA duplexes that systematically varied length and nick placement and formed complexes with two protein constructs, Δ232 and Δ262 LIG1. Sparse matrix screens covered more than 36,000 conditions, ultimately identifying orthorhombic crystal forms of LIG1 that diffracted to high resolution. Structure determination revealed that these complexes were trapped as the adenylylated-DNA intermediate, poised for nick sealing. As expected, the enzyme completely envelopes the DNA with the DNA binding domain (DBD), adenylylation domain (AdD), and oligonucleotide binding domain (OBD) encircling the nicked DNA. It was possible to locate individual octahedrally coordinated Mg^{2+} ions in the metal-soaked crystals. In addition to an active site Mg^{2+} (Figure 3.1b-d), 5 additional octahedrally coordinated Mg^{2+} ions could be identified with confidence. Of these non-catalytic Mg^{2+} ions, our attention was drawn to an intriguing Mg^{2+} ion that is coordinated at the three-way juncture of the AdD, DBD, and the DNA phosphodiester backbone. We refer to this metal ion as the high-fidelity Mg^{2+} ion (Mg^{HiFi}) for reasons described herein.

Although the active site Mg^{2+} was sometimes absent from crystals soaked at lower Mg^{2+} concentration, the Mg^{HiFi} was present in all crystal structures, suggesting tight binding. The octahedral coordination of the partially hydrated Mg^{HiFi} involves inner sphere interactions with the phosphodiester backbone between the -3 and -4 nucleotides (relative to the 3'-OH of the nick) as well as two glutamate residues from the protein. The AdD (Glu 346) and DBD (Glu 592) Mg^{HiFi} coordination residues are stringently conserved in LIG1 homologs including budding

yeast Cdc9p and fission yeast cdc17, but they are not found in vertebrate LIG3 or LIG4, suggesting that this site differentiates LIG1 from these other DNA ligases. The location at the junction between the AdD and the DBD and the direct interaction with the DNA implicates the metal ion in binding to DNA, defining the conformation of the protein and DNA substrate in the reactive complex.

Close examination of the Mg^{HiFi} site and its connectivity to the rest of the protein structure provides insight into the structural contributions that this site makes to the engagement of DNA by LIG1. The Mg^{HiFi} ion appears to make key interactions with the 3′-hydroxyl (3′-OH) strand, and these interactions restrict its conformation in the ligase active site. The metal-reinforced DNA binding mode secures the 3′-end in an A-form-like geometry, with the three terminal nucleotides exhibiting C3′-endo sugar puckers (Figure 3.1d). Moreover, analysis of the surface complementarity of binding in the active site shows that Mg^{HiFi}, and DNA binding loops bearing Mg^{HiFi} ligands from the DBD (E346) and AdD (E592) form a network of close contacts with the 3′ strand. These considerations suggested that the Mg^{HiFi} architecture would be necessary to properly recognize the 3′–OH strand of a DNA nick and may play a role in the correct placement of the nucleophilic hydroxyl within the active site. We hypothesized that the metal may be key to controlling the ability of LIG1 to discriminate against nicks adjacent to an upstream damaged or mismatched base.

The Mg^{HiFi} metal site confers intrinsic Lig1 catalytic fidelity

To define the functional contribution of the Mg^{HiFi}, we purified a double alanine mutant (E346A/E592A) of LIG1 that removes both of the Mg^{HiFi} ligands ($LIG1^{EE/AA}$; Figure S3.1a). Using a gel-based ligation assay (Figure 3.2), we determined the concentration of active enzyme

and found that the extent of adenylylation of the purified enzyme was comparable to that of the WT LIG1 (Figure S3.1b-c). We then performed a comprehensive survey of steady-state kinetic experiments with $LIG1^{EE/AA}$ and compared these to the WT enzyme to discern any changes in the biochemical behavior that occur upon disruption of the Mg^{HiFi} site (Figure S3.1d-g). $LIG1^{EE/AA}$ demonstrates an ~3-fold reduction in k_{cat} as compared to WT LIG1, with little or no change in the k_{cat}/K_M value for ATP or in the Mg^{2+} dependence. Remarkably, the K_M value for the DNA substrate is improved by 24-fold for the mutant relative to the WT LIG1 enzyme, resulting in a ~10-fold improvement in catalytic efficiency for $LIG1^{EE/AA}$ (Figures 3.2d and S3.2). Collectively these results suggest that the Mg^{HiFi} ion is dispensable for the ATP-dependent step 1, but it plays a role in tuning the interaction of LIG1 with the DNA substrate.

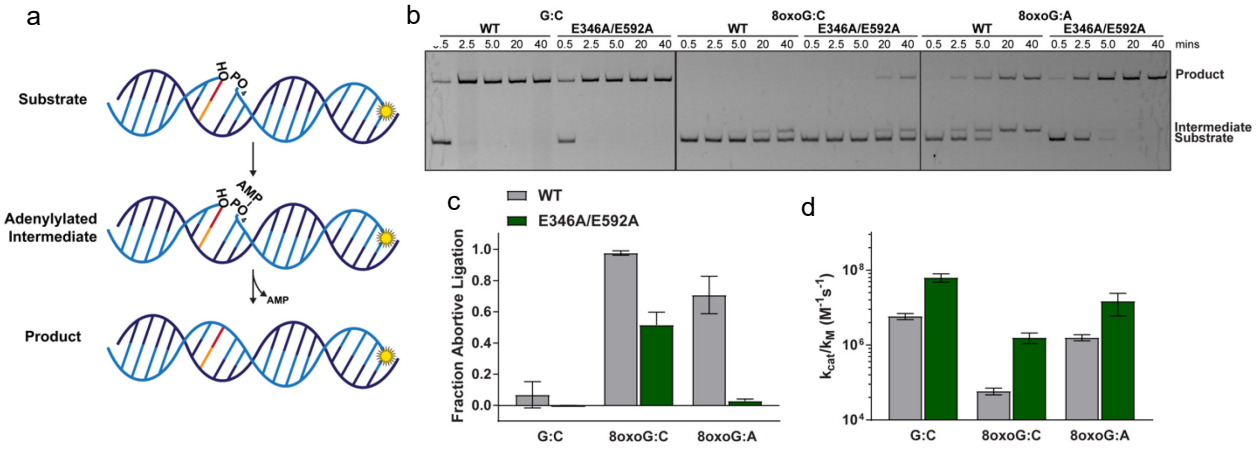


Figure 3.2. Effects of HiFi site on steady-state ligation kinetics. (**a**) Schematic of ligation assay. DNA substrates are fluorescein-labeled at the 3'-end (yellow) and the bases at the -1 nt position are either C or 8oxoG (red) or G, C, or A (orange). (**b**)Representative denaturing gel image of steady-state ligation reactions under physiological conditions. Reactions contained 5 nM of WT or E346A/E592A Δ232 LIG1 and 500 nM DNA substrate with a G:C, 8oxoG:C or 8oxoG:A pair at the -1 nt position. (**c**)The fraction of abortive ligation events determined from steady-state ligation reactions containing 1 or 10 nM LIG1 and 2 μM DNA substrate with 2 mM $MgCl_2$ and 1 mM ATP. (**d**) Relative k_{cat}/K_M values for ligation of the various DNA substrates determined by steady-state ligation dependences (see Figure S3.2) performed under physiological concentrations. All data are reported as the mean ± S.D. of $N \geq 3$ replicates.

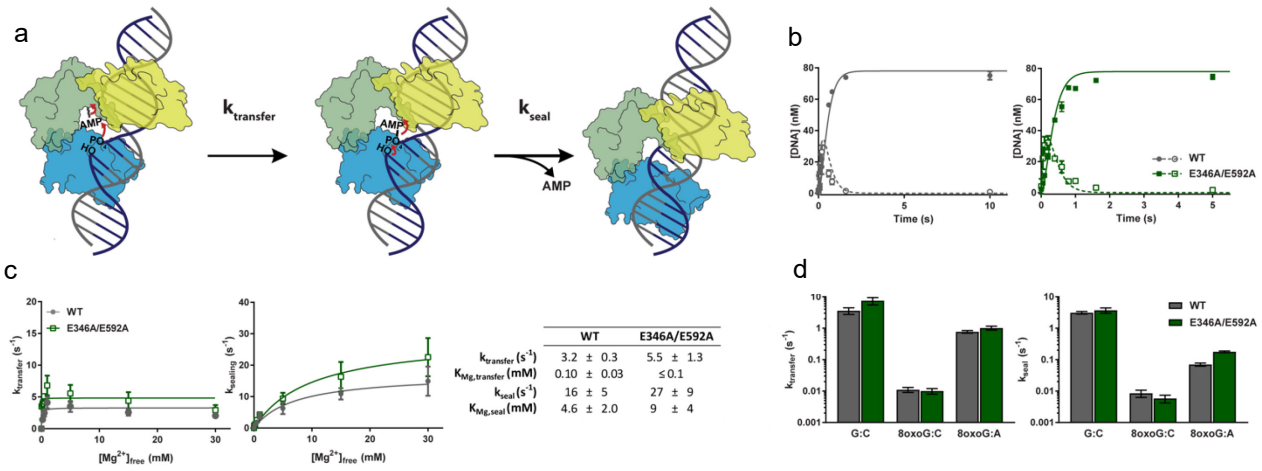
To test whether the Mg^{HiFi} plays a role in substrate discrimination, we investigated the ability of $LIG1^{EE/AA}$ to ligate substrates containing an 8-oxo-dG at the 3' terminal position of the

28mer substrate. It has been previously demonstrated that LIG1 discriminates against both 8-oxoG:C and 8-oxoG:A base pairs at this position[124,133,134]. Consistent with previous reports, LIG is greatly compromised in its ability to complete ligation of the 8-oxoG:C substrate, instead aborting catalysis and leading to an accumulation of the AMP-DNA intermediate (Figure 3.2b-c). Although LIG1 can catalyze ligation of the 8-oxoG:A substrate, it does so slowly, with a significant (~60%) accumulation of the adenylylated DNA intermediate. In contrast, the $LIG1^{EE/AA}$ mutant enzyme has dramatically enhanced activity on the 8-oxoG-containing substrates. This mutant fully ligates the 8-oxoG:A substrate and ligates approximately half of the 8oxoG:C substrate that it engages (Figure 3.2c). Congruent with this decrease in abortive ligation, the $LIG1^{EE/AA}$ mutation substantially increases the catalytic efficiency for ligation of both 8oxoG-containing substrates (Figure 3.2d and S3.2). These observations demonstrate that the Mg^{HiFi} contributes to the discrimination against abnormal base pairs immediately upstream of the nick and raises the question of how disruption of the HiFi site increases the catalytic efficiency of LIG1.

We next characterized the single-turnover ligation kinetics of the $LIG1^{EE/AA}$ mutant using a rapid chemical quench approach. Two chemical steps occur subsequent to DNA binding, adenylyl-transfer and nick sealing (Figure 3.3a). By following the time course for formation of ligated product and for build-up and consumption of the adenylylated DNA intermediate, both rate constants can be determined (Figure 3.3b). LIG1 WT and mutant were purified as the adenylylated intermediate, and therefore no ATP is required in these reactions.

By performing rapid-quenching experiments over a range of Mg^{2+} concentrations with the 28mer-G:C substrate, we were able to obtain the maximal rates of both adenylyl-transfer ($k_{transfer}$) and nick sealing (k_{seal}). Remarkably, both enzymes exhibited similar maximal rates for

both chemical steps (Figure 3.3c). These data demonstrate that the LIG1 EE/AA mutations do not perturb catalysis with a normal DNA substrate and suggests that the slower multiple-turnover ligation by the LIG1$^{EE/AA}$ mutant is due to a step prior to or following the phosphoryl transfer reactions. The transient kinetic experiments were extended to investigate the microscopic rates of ligation for both 8oxoG substrates at a physiological concentration of Mg^{2+}. These data reveal that the WT LIG1 discriminates strongly against 8oxoG:C at both adenylyl transfer and nick sealing steps. Discrimination against 8-oxoG:A also occurs in both steps, but LIG1 has a reduced ability to exclude this damaged substrate as compared to 8oxoG:C, especially in the adenylyl transfer step (Figure 3.3d). The kinetics of adenylyl transfer and nick sealing are not significantly perturbed for the LIG1$^{EE/AA}$ mutant with both 8oxoG-containing substrates (Figure 3.3d). These data demonstrate that the HiFi Mg^{2+} site is dispensable for the chemical steps, but the intact site influences other steps that precede or follow the chemical steps of the ligation reaction.



	WT	E346A/E592A
$k_{transfer}$ (s^{-1})	3.2 ± 0.3	5.5 ± 1.3
$K_{Mg,transfer}$ (mM)	0.10 ± 0.03	≤0.1
k_{seal} (s^{-1})	16 ± 5	27 ± 9
$K_{Mg,seal}$ (mM)	4.6 ± 2.0	9 ± 4

Figure 3.3 Single turnover ligation kinetics of undamaged and 8oxoG-containing substrates. (a) Mechanism of ligation by LIG1. Adenylylated LIG1 transfers an AMP group to the 5'-phosphate ($k_{transfer}$) at the nick and subsequently catalyzes attack of the 3'-hydroxyl to form a new phosphodiester bond (k_{seal}) and release AMP. **(b)** Representative single-turnover reactions for the ligation of the G:C substrate by WT and E346A/E592A LIG1 in the presence of 1 mM $MgCl_2$ (product formation, closed symbols, and intermediate formation, open symbols, are shown). **(c)** Mg^{2+}-dependence of the rates of adenylyl transfer and nick sealing determined by fitting the data with a two-step model in Berkeley-Madonna. The maximum rates and Mg^{2+} affinities of both steps were obtained by hyperbolic fits to the data. Only a limit could be obtained for the Mg^{2+} affinity in the adenylyl transfer step with the E346A/E592A mutant. **(d)** Rate constants for adenylyl transfer and nick sealing were determined for WT and mutant LIG1 with 1 mM Mg^{2+} (see Figure S3 for kinetic data with 8-oxoG substrates). All data are reported as the mean ± S.D. of $n \geq 3$ replicates.

Constrained 3′ strand binding is scaffolded by the HiFi Mg^{2+}

Inspection of the LIG1 nicked DNA complex shows that protein binding widens the major and minor groove around the nick site, inducing the DNA upstream of the nick (the 3′–hydroxyl side) to adopt an A-form like geometry of the 3′-hydroxyl (3′-OH) strand (Figure 3.1d). An extensive network of interactions links the protein-DNA contacts along the 3'-strand to the positioning of the 3'-OH nucleophile in the active site and the coordination with the active site Mg^{2+} ion (Figure 3.4a-c). We were able to determine the crystal structure of the $LIG1^{EE/AA}$ mutant in complex with the ddC-blocked DNA nick (Figure 3.4d-f). The $2F_o$-F_c map verifies the predicted absence of the HiFi Mg^{2+} ion and a large cavity that is created by the EE/AA mutations (Figure 3.4d). The 3'-strand is repositioned, and the network of protein-DNA van der Waals contacts along this strand have been extensively rearranged (Figure 3.4e).

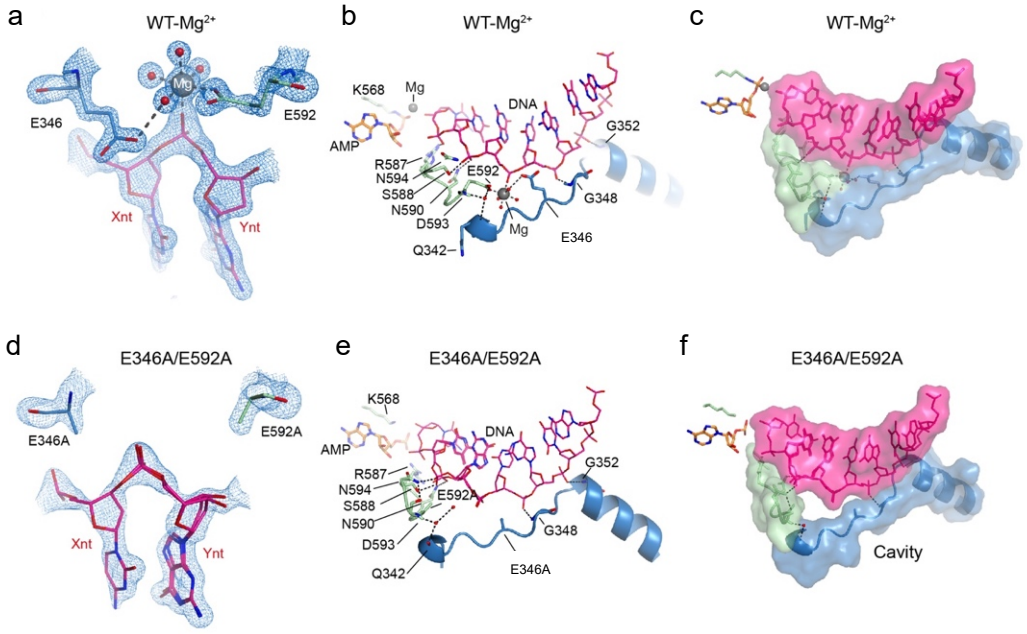


Figure 3.4. X-ray structures of HiFi variant E346A/E592A and structural comparison to WT. (**a**) Omit 2Fo-Fc electron density for the HiFi site in the WT structure illustrates the bound HiFi Mg^{2+}. (**b**) Ribbon and stick representation of protein-DNA contacts at the HiFi site extending to the active site in WT LIG1. DBD is in blue, AdD in green, DNA in pink, adenylate in orange, both active site and HiFi Mg^{2+} are in gray. (**c**) Space filled representation of Figure 2B displays enforcement for correct conformation of bound DNA substrate by complementarity of protein domains and DNA backbone. (**d**) Omit 2Fo-Fc electron density for the HiFi site in E346A/E592A double mutant illustrate absence of bound metal. (**e**) Ribbon and stick representation of protein-DNA contacts at the HiFi site extending to the active site in E346A/E592A double mutant. (**f**) Space filled representation of Figure 3.4E displays a cavity at the HiFi site due to the absence of bound metal.

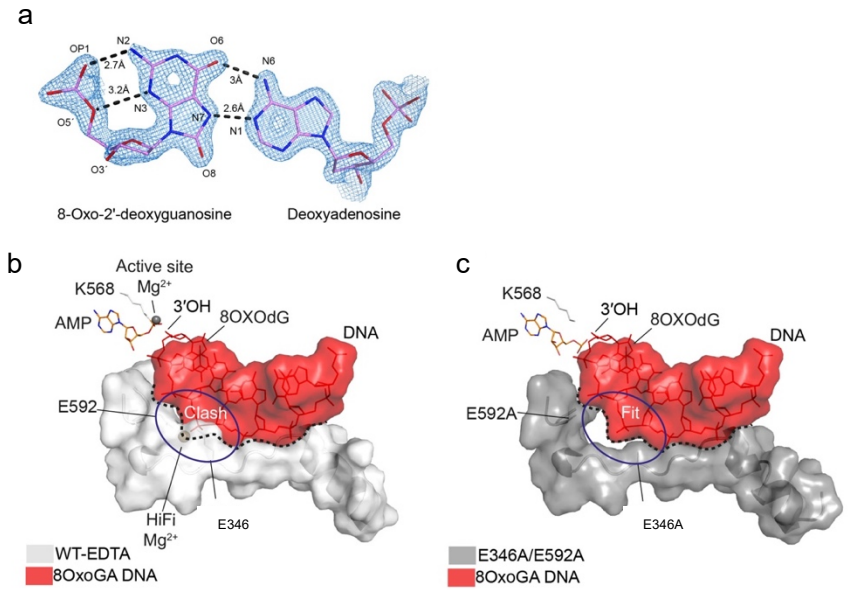


Figure 3.5. X-ray structure of E346A/E592A•8-oxoG:A DNA complex. (a) Omit 2Fo-Fc electron density for the 8-oxoG:A nucleotide pair bound in E346A/E592A double mutant structure shows the pairing is in (A anti: 8-oxoG syn) conformation. The N2 and N3 atoms of the 8-oxoG make additional contacts with the phosphate backbone. **(b)** Modelling of the bound 8-oxoG:A DNA in E346A/592A•8-oxoG:A structure (red) in the WT structure (light gray) shows the conformation of the DNA would create a clash at the HiFi site. **(c)** X-ray structure of E346A/592A•8-oxoG:A complex shows the 8-oxoG:A DNA is accommodated and fits in the metal-free cavity at the HiFi site.

Molecular basis of mutagenic ligation by the HiFi site mutant

Attempts to crystalize WT LIG1 in complex with 8oxoG-containing DNA substrates were not successful, but we obtained a crystal structure of the LIG1$^{EE/AA}$ mutant in complex with 8oxoG:A (Figure 3.5). As predicted by previous structures of DNA polymerases with 8oxoG•A base pairs at their active sites, the 8oxoG:A 3'-terminal pair adopts Hoogsteen base pairing geometry. The *syn* geometry of the 8oxoG is stabilized by hydrogen bonding between the exocyclic 2-amino group and the phosphodiester backbone. The space-filling representation demonstrates how the cavity in the LIG1$^{EE/AA}$ mutant is partially filled by the rearrangements in the 3' strand (Figure 3.5c). The DNA conformation that accommodates the 8oxoG:A pair in the LIG1 active site is predicted to cause a steric clash with the HiFi Mg^{2+} site (Figure 3.5b). The corollary of this prediction is that enforcement of the normal 3'strand geometry is not compatible

with an 8oxoG:A pair in the active site, providing a structural rationale for the discrimination by LIG1.

APTX suppresses LIG1 reactions with 8-oxoG containing substrates

The biochemical characterization of LIG1 3'-end recognition demonstrated the importance of abortive ligation as a bypass mechanism to limit the sealing of damaged/mismatched termini. If adenylylation of an improper nick occurs, then the Mg^{HiFi} site architecture leads to enhanced partitioning away from erroneous ligation and towards abortive ligation. Abortive ligation on damaged nicks could provide additional time for other DNA repair pathways to operate. However, it is necessary for the 5'-AMP group to be removed to restore a 5'-phosphate for ligation after 3'-proofreading and gap filling. Aprataxin (APTX) is a hydrolase that has been shown to catalyze the hydrolysis of 5'-5' adenylylated DNA and RNA. We hypothesized that the interplay between LIG1 and APTX would be especially important for controlling the pathway choice between erroneous ligation and abortive ligation. Therefore, we performed steady-state ligation assays in the presence or absence of human APTX.

Indeed, APTX is able to fully suppress abortive ligation by WT LIG1 with an 8oxoG:C substrate and a 1:10 ratio of APTX:LIG1 (Figure 3.6a, left panel). In contrast, the $LIG1^{EE/AA}$ mutant was only partially suppressed by APTX under these conditions and ligation of the damaged 8oxoG:C continued at ~50% the rate when APTX was present (Figure 3.6a, right panel and 3.6b). The same trend was observed for the 8oxoG:A substrate, with APTX effectively suppressing the reaction catalyzed by WT LIG1, but being unable to affect the reaction catalyzed by the $LIG1^{EE/AA}$ mutant (Figure 3.6b and Figure S3.4). These data suggest that the HiFi Mg^{2+} produces strain in the LIG1•DNA complex that induces increased abortive ligation with

damaged termini, and subsequent processing by APTX. This fidelity comes at a cost to overall ligation efficiency, as APTX causes modest, but readily detectable suppression of ligation by WT LIG1 with the undamaged substrate. In all cases, disruption of this metal site ($LIG1^{EE/AA}$) leads to increased catalytic efficiency and decreased fidelity of ligation.

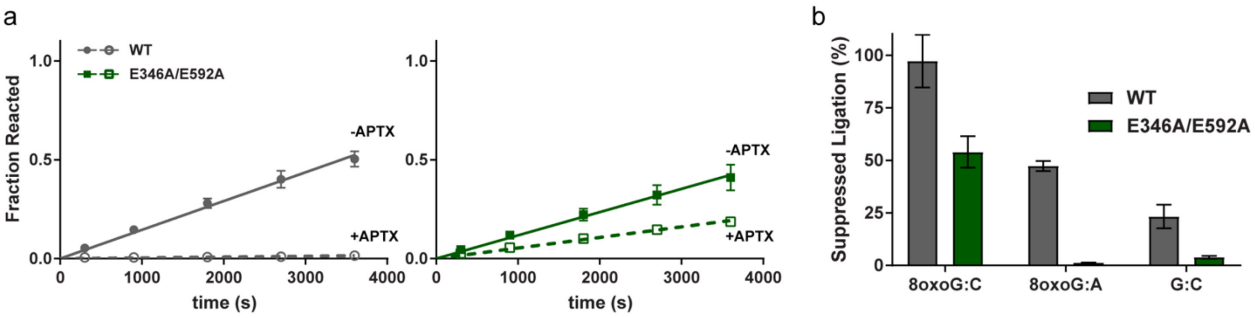


Figure 3.6. APTX suppresses LIG1-catalyzed ligation of 8-oxoG-containing DNA. (a) Representative time courses from steady-state ligation reactions in the absence (solid symbols, solid lines) or presence (open symbols, dashed lines) of APTX. **(b)** The percentage of suppressed ligation was calculated as the percent change in initial rate from steady-state reactions in the absence of APTX to those in the presence of APTX. All data are reported as the mean ± S.D. of N ≥ 3 replicates.

For APTX to engage with the adenylylated intermediate, it must be given access to the intermediate by LIG1, making it important to consider a mechanism for how LIG1 may present the intermediate to APTX. In single-turnover assays with the 8oxoG-containing substrates, we observed that ATP induced an extended lifetime for the adenylylated DNA intermediate, suggesting that the presence of ATP slows down nick sealing (Figure S3.5). This observation led us to believe that LIG1 can transiently bind to ATP during ligation of the damaged substrate. This could serve as a mechanism for abortive ligation, where LIG1 would pause at the damaged nick and transiently bind to ATP. The presence of excess DNA substrate would then increase the likelihood of complete dissociation of LIG1 from the adenylylated intermediate. Such a delay would allow APTX the opportunity to compete with LIG1 for the adenylylated intermediate.

To test the accessibility of the adenylylated intermediate to APTX, we made use of the $APTX^{H260N}$ mutant, which is catalytically-dead but retains the ability to bind adenylylated

ligation intermediates. $ATPX^{H260N}$ was included in single-turnover ligation experiments containing physiological concentrations of both Mg^{2+} and ATP to assess the ability of both $LIG1^{WT}$ and $LIG1^{EE/AA}$ to remain committed to catalysis on the various substrates. In the presence of $APTX^{H260N}$, any adenylylated intermediate that is accessible to the APTX mutant should be sequestered and prevented from re-associating with LIG1 (Figure 3.7a). In control reactions with the 28mer-G:C substrate, both WT LIG1 and $LIG1^{EE/AA}$ remained fully committed to ligating the substrate (Figure 3.7d and S3.6 a,d). With either of the two 8oxoG-containing substrates, however, both ligases performed the initial adenylyl-transfer step but were no longer able to perform nick-sealing to completion due to trapping of the intermediate by $APTX^{H260N}$ (Figure 3.7b-d, Figure S3.6). $LIG1^{EE/AA}$ remained ~80% committed to catalysis with the 8oxoG:A substrate and ~40% committed with the 8oxoG:C substrate.

WT LIG1, however, exhibited a much more dramatic decrease in catalytic commitment with both substrates, with about only 40% of ligation events with the 8oxoG:A substrate and about 20% of ligation events with the 8oxoG:C going to completion. Along with structural data detailing flexibility in the upstream strand when substrate is bound by $LIG1^{EE/AA}$, the ability of $LIG1^{EE/AA}$ to protect ligation intermediates from APTX sequestration implies that the Mg^{HiFi} may be functioning to help enforce a rigid upstream DNA structure. When a damaged nick is encountered, this upstream rigidity may cause LIG1 to sample a more "open" conformation, allowing the enzyme to interact with excess DNA and ATP and for APTX to capture the adenylylated intermediate.

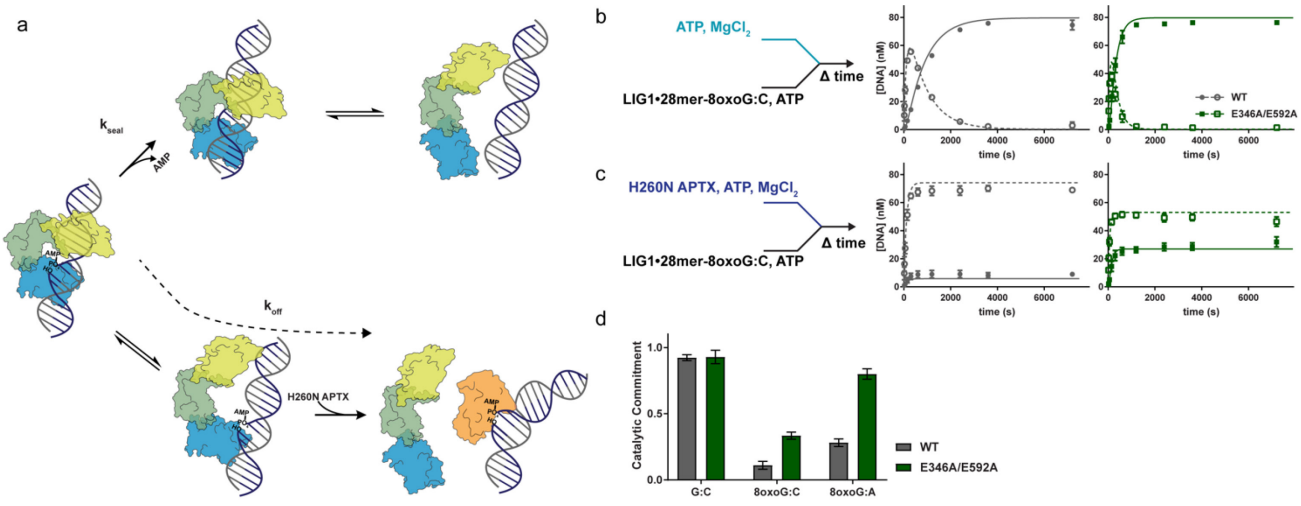


Figure 3.7. Catalytic commitment of LIG1 on the 8-oxoG:C substrate. (a) Kinetic model for trapping of the AMP-DNA intermediate by the H260N APTX catalytically inactive mutant. Partial dissociation of LIG1 gives APTX the opportunity to sequester the intermediate. **(b)** Single-turnover reactions with the 8-oxoG:C substrate were carried out by mixing a solution containing LIG1, 8oxoG:C DNA and ATP with a separate solution containing $MgCl_2$ and ATP. These reactions were also performed with the H260N APTX mutant in the $MgCl_2$-containing solution (**c**). Product formation is represented by closed symbols and intermediate formation is denoted by open symbols. **(d)** Catalytic commitment was determined as the proportioning between nick sealing and dissociation from the AMP-DNA intermediate. Data for the catalytic commitment to ligation of the G:C and 8oxoG:A substrates were also measured (Figure S3.6). These data demonstrate that the E346A/E592A mutant, which disrupts the HiFi Mg^{2+} site, has a greater commitment to nick sealing when bound to an 8oxoG:C substrate. All data are reported as the mean ± S.D. of N ≥ 3 replicates.

Discussion

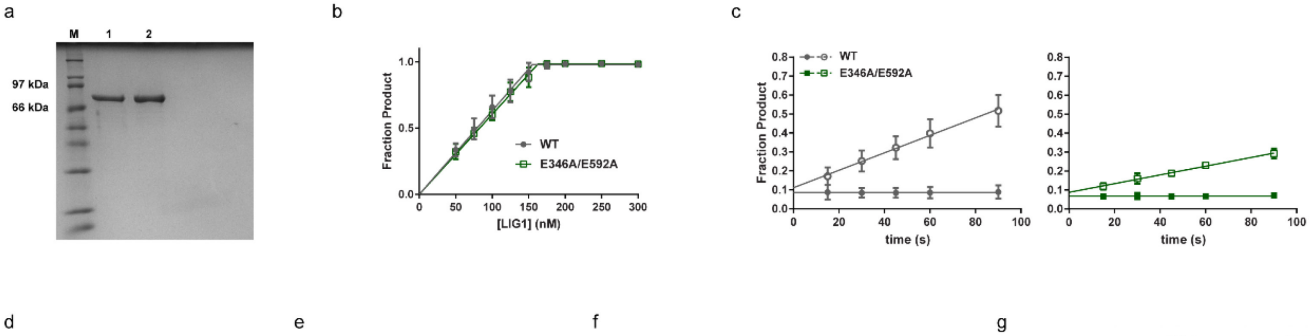
The fidelity of DNA ligation has long been an underappreciated component of the faithful replication and repair of genomic DNA. In this work we have used quantitative biochemical approaches to elucidate the mechanisms by which the replicative DNA ligase, LIG1, discriminates against improper 3' termini at multiple steps of DNA ligation. High resolution crystal structures reveal the existence of a novel Mg^{HiFi} ion that plays an integral role in enforcing recognition of the 3' strand and biases improper nicks towards abortive ligation. We further establish that APTX accesses the stalled AMP-DNA intermediate leading to dissociation of LIG1 and restoration of the 5'-phosphate terminus. Mutagenesis that removes the Mg^{HiFi} ion ($Lig1^{EE/AA}$) significantly reduces the ability of the enzyme to discriminate against improper nicks

and causes an increase in ligation efficiency on both undamaged and damaged nicks. This Mg^{HiFi} site is predicted to be conserved in other eukaryotic replicative DNA ligases, but it is notably absent in the human LIG3 and LIG4 enzymes. Our study suggests that the Mg^{HiFi} architecture of LIG1 is a distinguishing structural feature that is integral to the unique biochemical activities of the different DNA ligases.

Although this study focuses primarily on the characterization of this metal site with respect to discrimination against a damaged end, our preliminary data suggests that this mechanism of discrimination is also applicable to mismatches not arising from oxidative damage. In an attempt to demonstrate the ability of LIG1 to discriminate against an undamaged but mismatched end, we repeated a number of the key experiments with a G:T mismatch at the 3' end of the nick. G:T mismatches are among the more prevalent mismatches found in nature, arising from the deamination of 5-methylcytosine to thymine[135]. Steady-state ligation of a G:T 28mer substrate by both wildtype and EE/AA LIG1 exhibits similar behavior to ligation of the 8oxoG:C substrate (Figure S3.7a-b). In the absence of the Mg^{HiFi} site, we observe efficient ligation of the G:T substrate. Additionally, ligation of the substrate by wildtype LIG1 is susceptible to interruption by APTX in both steady-state and single-turnover assays (Figure S3.7c-f). These results recapitulate our findings with ligation of the 8oxoG-containing substrates, highlighting the importance of the metal site in allowing LIG1 to readily abort ligation of a damaged or mismatched substrate. Our findings here are consistent with the observation of a multi-step remodeling process in Chapter 2, where we believe LIG1 induces a number of conformational changes in the DNA substrate to position the nick for catalysis. Although the Mg^{HiFi} mutant remains to be tested in our DNA-binding assay, we hope that monitoring the

EE/AA mutant as it engages with the DNA substrate will help to further elucidate its role in this discrimination process.

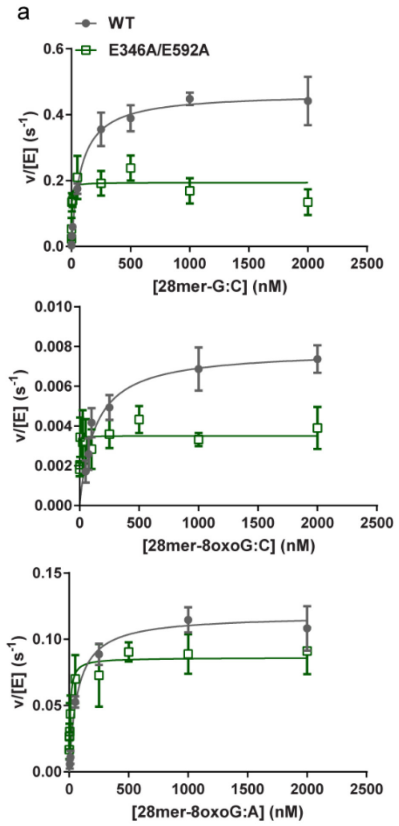
Chapter 3 Appendix



g

	WT			EE/AA		
DNA Dependence						
k_{cat} (s^{-1})	0.83	±	0.15	0.30	±	0.08
K_M (nM)	65	±	33	6.9	±	0.7
k_{cat}/K_M ($s^{-1}M^{-1}$)	1.3	±	0.7 x 10^7	4.4	±	1.2 x 10^7
ATP Dependence						
k_{cat} (s^{-1})	0.96	±	0.02	0.29	±	0.10
K_M (μM)	9.2	±	1.8	26	±	13
k_{cat}/K_M ($s^{-1}M^{-1}$)	1.1	±	0.2 x 10^5	1.9	±	1.1 x 10^4
Mg^{2+} Dependence						
k_{cat} (s^{-1})	0.78	±	0.03	0.27	±	0.06
K_{Mg} (mM)	0.78	±	0.10	0.78	±	0.05

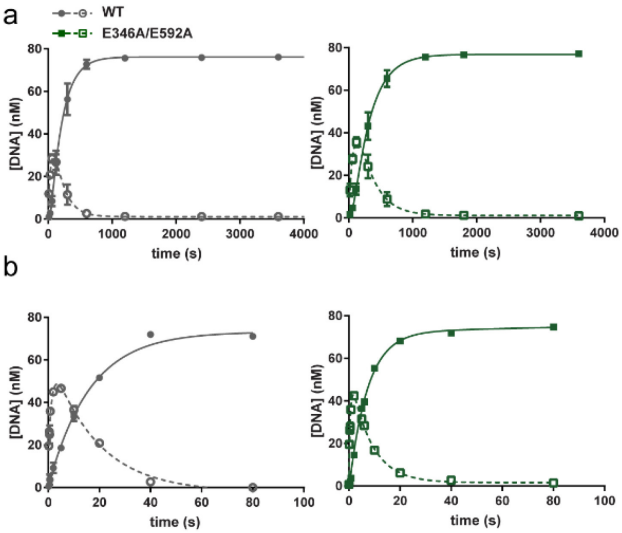
Figure S3.1. Initial characterization of the E346A/E592A LIG1 mutant. (a) Purity of recombinant WT (lane 1) and mutant (lane 2) Δ232 hLIG1 was assessed using SDS-PAGE (~1 μg of LIG1 and a protein ladder (lane M) were loaded on a 15% polyacrylamide/SDS gel). **(b)** The concentration of active, adenylylated Δ232 LIG1 was confirmed by quantifying the amount of ligated DNA in the absence of added ATP. The nicked DNA substrate was fixed at 150 nM and the nominal concentration of LIG1 protein was varied from 0-300 nM. These data gave active concentrations that deviated ~4% and ~8%, respectively, from the estimated concentrations determined via UV-absorbance. **(c)** Pre-steady state burst experiments were performed to assess the adenylylation state of the recombinant enzymes. Mixes containing 50 nM LIG1 and either 0 (closed symbols) or 1 μM (open symbols) ATP were allowed to incubate at 37°C for 5 minutes before reactions were initiated with the addition of 500 nM LIG1 nicked DNA. Differences in the burst amplitude between the ATP-containing and no ATP reactions indicate the presence of deadenylylated enzyme. The ~30% deadenylylated enzyme population was considered when calculating k_{cat} in steady-state reactions. The steady-state Mg^{2+} (**d**), ATP (**e**) and 28mer-G:C DNA dependences (**f**) were determined for WT and E346A/E592A LIG1. Both the ATP and the DNA dependences were performed with 20 mM $MgCl_2$. The values obtained from fitting these dependences with the Michaelis-Menten equation are shown in **g**. All data are reported as the mean ± S.D. of N ≥ 3 replicates.



b

	WT	EE/AA
G:C		
k_{cat} (s^{-1})	0.46 ± 0.04	0.21 ± 0.03
K_M (nM)	79 ± 14	3.3 ± 0.7
k_{cat}/K_M ($s^{-1}M^{-1}$)	5.9 ± 1.1 x 10^6	6.4 ± 1.6 x 10^7
8oxoG:C		
k_{cat} (s^{-1})	7.8 ± 0.8 x 10^{-3}	4.2 ± 0.8 x 10^{-3}
K_M (nM)	132 ± 21	2.6 ± 0.7
k_{cat}/K_M ($s^{-1}M^{-1}$)	5.9 ± 1.2 x 10^4	1.6 ± 0.5 x 10^6
8oxoGA		
k_{cat} (s^{-1})	0.12 ± 0.01	0.078 ± 0.012
K_M (nM)	78 ± 14	5.1 ± 3.0
k_{cat}/K_M ($s^{-1}M^{-1}$)	1.6 ± 0.3 x 10^6	1.5 ± 0.9 x 10^7

Figure S3.2. Steady-state substrate dependences for ligation catalyzed by WT and E346A/E592A LIG1. Steady-state substrate dependences were performed for all 28mer substrates at 1 mM Mg^{2+} (**a**) with WT and E346A/E592A LIG1. The data were fit by the Michaelis-Menten equation, which was used to determine the steady-state constants shown in **b**. All data are reported as the mean ± S.D. of N ≥ 3 replicates.



c

WT			
	28mer	28mer-8oxoG:C	28mer-8oxoG:A
$k_{transfer}$ (s^{-1})	3.6 ± 0.9	0.011 ± 0.002	0.77 ± 0.07
k_{seal} (s^{-1})	3.1 ± 0.3	0.0085 ± 0.0022	0.071 ± 0.008
EE/AA			
	28mer	28mer-8oxoG:C	28mer-8oxoG:A
$k_{transfer}$ (s^{-1})	7.5 ± 2.0	0.010 ± 0.002	1.0 ± 0.1
k_{seal} (s^{-1})	3.7 ± 0.7	0.0058 ± 0.0016	0.18 ± 0.01

Figure S3.3. Single-turnover reactions with 8oxoG-containing substrates. WT and E346A/E592A were assayed under single-turnover conditions with both the 8oxoG:C 28mer (**a**) and the 8oxoG:A 28mer (**b**) substrates. Accumulation of product (closed symbols) and intermediate (open symbols) species are shown as a function of time (mean ± S.D.; N ≥ 3). Rate constants for adenylyl transfer and nick sealing shown in **c** were obtained from fits using a two-step irreversible model in Berkeley-Madonna.

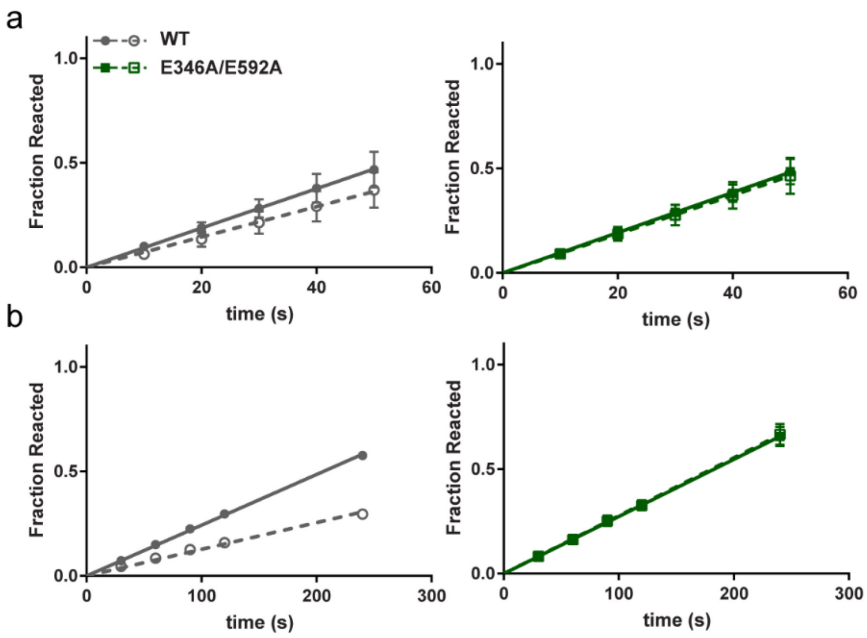


Figure S3.4. Effect of APTX on steady-state ligation. Steady-state ligation reactions contained 10 nM LIG1 and 500 nM of the 28mer substrate with a G:C (**a**) and 8oxoG:A (**b**) pair at the -1nt position. APTX was either absent (closed symbol, solid line) or added at a concentration of 1 nM (open symbol, dashed line). Total reacted substrate for each reaction is plotted as a function of time for reactions containing either WT or E346A/E592A mutant LIG1 (mean ± S.D.; N ≥ 3). The amount of ligation suppressed by APTX activity was calculated as the percent change in the initial ligation rate upon the addition of APTX (Figure 3.6b).

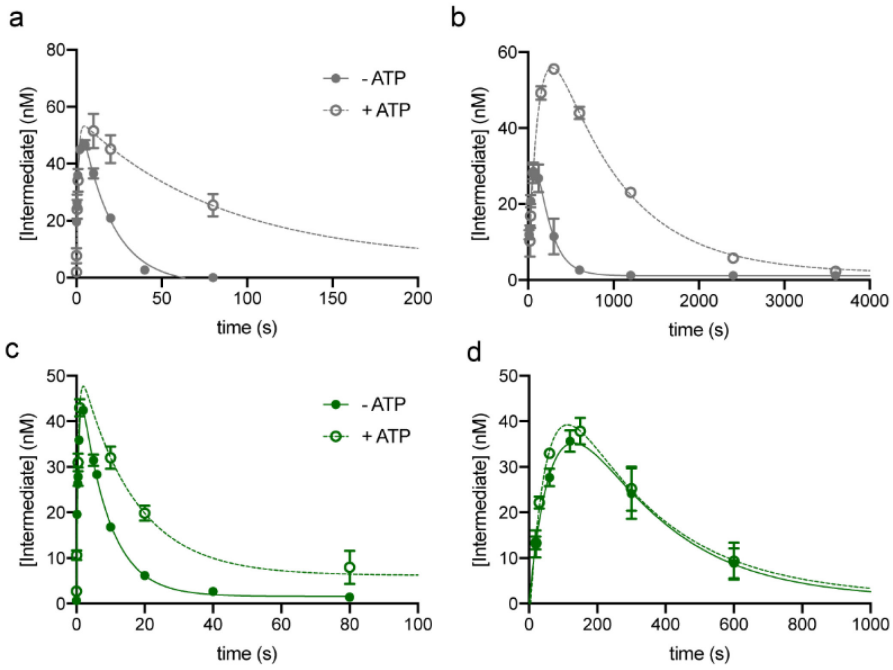


Figure S3.5. Effect of ATP on single-turnover ligation of 8oxoG substrates. Intermediate formation from single-turnover ligation of the 8oxoG:A (**a,c**) and 8oxoG:C (**b,d**) 28mer substrates by WT (gray) or EE/AA (green) LIG1. Traces are shown from single-turnover reactions done in the presence or absence of ATP. Single-turnover data is adapted from Figure 3.7, S3.3 and S3.6.

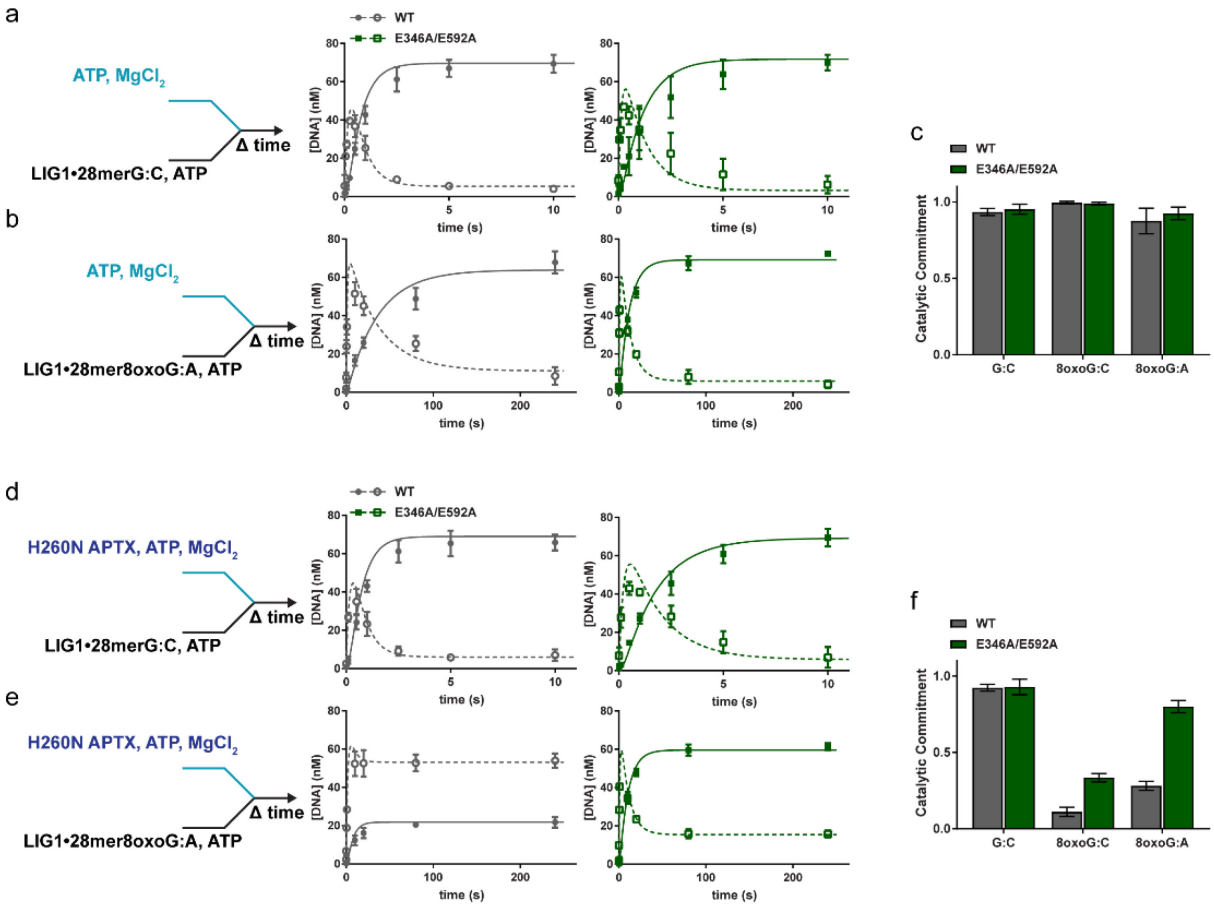


Figure S3.6. Catalytic commitment of WT and mutant LIG1 under single turnover conditions. Single-turnover reactions were carried out by mixing a solution containing either 800nM WT or E346A/E592A LIG1, 80nM 28mer-G:C (**a,d**) or 28mer-8oxoG:A (**b,e**) and with a separate solution containing 2mM $MgCl_2$. Reactions were performed in the presence (**d-e**) or absence (**a-b**) of 800nM H260N APTX to the $MgCl_2$ solution. Product formation is represented by closed symbols, while intermediate formation is denoted by open symbols (mean ± S.D.; N ≥ 3). Curve fits shown for both product (solid line) and intermediate (dashed lines) formation were obtained by fitting the data using Berkeley-Madonna with a two-step irreversible model with an added dissociation step after adenylyl-transfer (Figure 3.7c). The catalytic commitment for both WT and E346A/E592A LIG1 was calculated for reactions without APTX (**c**) and with H260N APTX (**f,** Figure 3.7d).

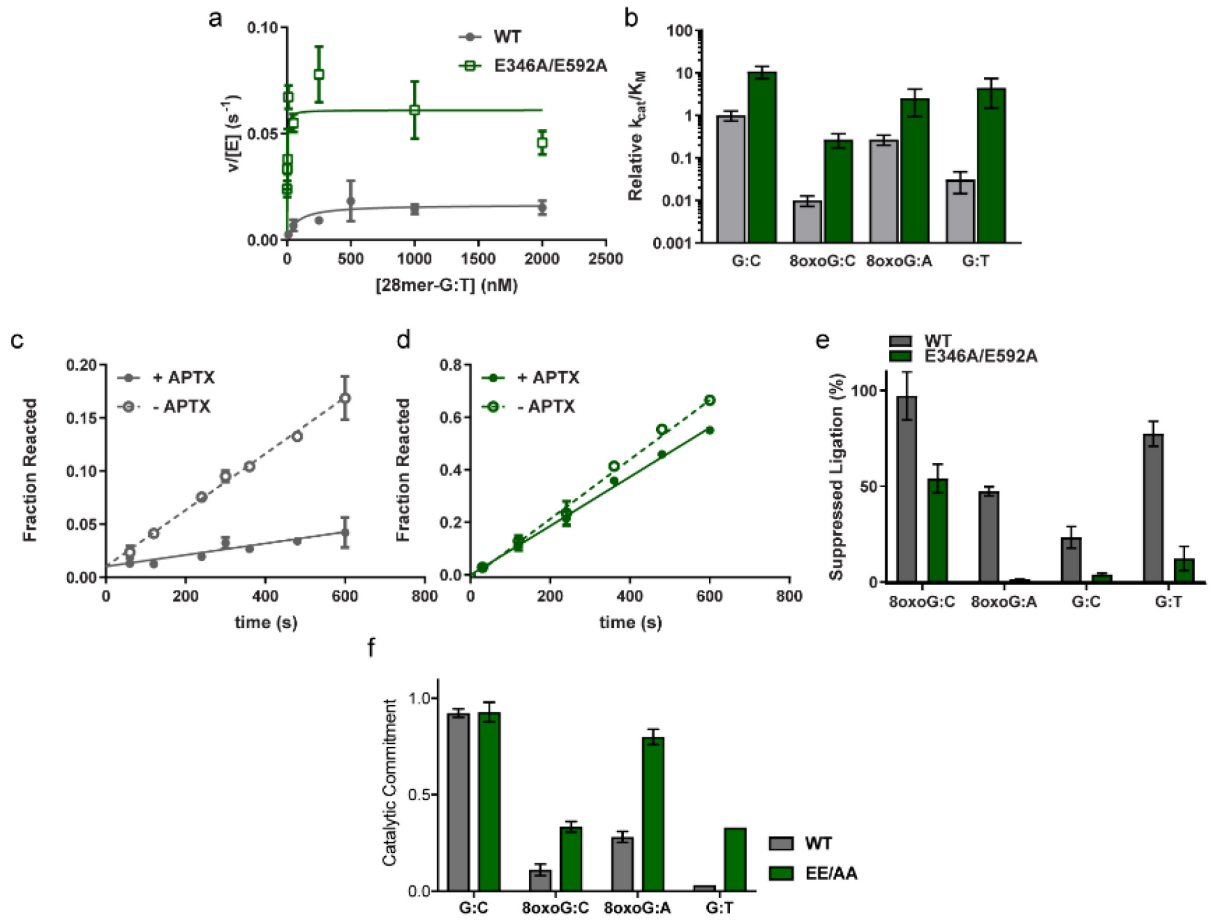


Figure S3.7. Ligation of a G:T mismatched substrate. The key experiments performed for showing discrimination of the 8oxoG-containing substrates were repeated with a 28mer variant containing a G:T mismatch at the 3' end of the nick. (**a-b**) Steady-state substrate dependence demonstrating the effect of the G:T mismatch is similar to what was observed with the 8oxoG:C substrate. (**c-e**) Repeat of the experiments outline in Figure 3.6, demonstrating APTX-dependent suppression of wildtype ligation of the G:T mismatch. All steady-state experiments in **a-e** are shown as the mean ± S.D. of three or more replicates. (**f**) Preliminary single-turnover results showing that ligation of the G:T mismatch is susceptible to interruption by APTX. Data shown is only based on a single replicate.

Figure S3.8. Berkeley Madonna model for single turnover ligation in the presence of H260N APTX

```
METHOD RK4
STARTTIME = 0
STOPTIME=100
DT = 0.02

d/dt(ES) = -ktransfer*ES
d/dt(EI) = ktransfer*ES - EI*(kseal + koff)
d/dt(EI2) = koff*EI
d/dt(EP) =  kseal*EI

init ES = 78
init EI = 0
init EI2 = 0
init EP = 0

ktransfer = 0.01
kseal = 0.008
koff = 0.008

Product = EP
Intermediate = EI + EI2
```

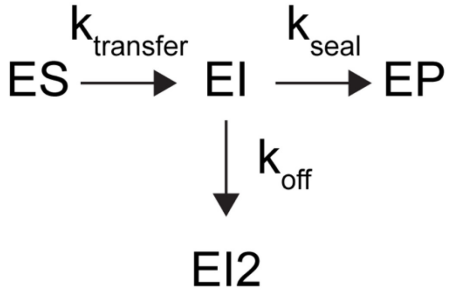


Figure S3.9 Oligos used for crystallization and kinetic assays.

Oligo Name	5′ modification	Sequence	3′ modification
18mer Down		5′ GTCCGACGACGCATCAGC 3′	
11mer 3'ddCUP		5' GCTGATGCGT(ddC) 3'	ddC
7mer 5'PDown	Phosphate	5' P-GTCGGAC 3'	
11mer 3'OHUP		5' GCTGATGCGTC 3'	
18mer 8oxoG:ADown		5' GTCCGACAACGCATCAGC 3'	
11mer 8oxoGUp		5' GCTGATGCGT(8oxoG) 3'	8oxoG
Temp28		5′ CCGAATCAGTCCGACGACGCATCAGCAC 3'	
Temp28-A		5′ CCGAATCAGTCCGAC**A**ACGCATCAGCAC 3'	
Temp28-C		5′ CCGAATCAGTCCGAC**C**ACGCATCAGCAC 3'	
Temp28-T		5′ CCGAATCAGTCCGAC**T**ACGCATCAGCAC 3'	
Up13OH		5' GTGCTGATGCGTC 3'	
Up13OH-8oxoG		5' GTGCTGATGCGT(8oxoG) 3'	8oxoG
Up13OH-G		5' GTGCTGATGCGTG 3'	
P15FAM	Phosphate	5' P-GTCGGACTGATTCGG-FAM 3'	

Chapter 4
Mechanism for Hi-Fidelity Ligation: Discrimination Against 5' Ribonucleotides

Ribonucleotides have long been known to play a large role in DNA replication as the macromolecule used to prime DNA polymerization. Since discontinuous replication requires constant repriming, lagging strand DNA is the source of most incorporated ribonucleotides in cellular DNA[136]. While ribonucleotide priming is essential for initiating DNA replication, ribonucleotides can also be misincorporated into replicating strands by DNA polymerases[137]. Although DNA polymerases strongly discriminate against ribonucleotide insertion *in vitro*, the large abundance of rNTPs in the cell can impede this selectivity[138,139]. Allowing ribonucleotides to persist in DNA presents a significant risk for the cell. The most notable impact of ribonucleotide insertion is on the stability of a DNA strand due to the 2'-OH in ribonucleotides increasing the likelihood of spontaneous hydrolysis of the phosphodiester backbone[140]. Additionally, the difference in sugar pucker makes ribonucleotides more likely to adopt A-form helical structure, potentially impacting how proteins interface with the double helix[141,142].

Due to the potential consequences of ribonucleotide misincorporation, cellular mechanisms have evolved to cleanse DNA strands of ribonucleotides to preserve genomic integrity. The major pathway for removal of ribonucleotides, known as ribonucleotide excision repair (RER), is initiated by the action of RNaseH2[143]. RNaseH2 incises RNA-DNA junctions creating a nick containing a 5'-ribonucleotide monophosphate. The ribonucleotide-containing

---

The work reported in Chapter 4 is a part of an ongoing collaboration with Percy Tumbale and Matthew J. Schellenberg working under the guidance of R. Scott Williams at the National Institute of Environmental Health Sciences. Neha Bokil contributed to the characterization of the F872A LIG1 mutant while working as an undergraduate researcher in the O'Brien lab.

strand is then displaced by the replicative polymerase, allowing for the displaced strand to be fully excised by the flap endonuclease (FEN1). In the absence of RNaseH2, DNA topoisomerase can also catalyze the removal of ribonucleotides from DNA, although this secondary pathway is limited to the leading strand and is sometimes mutagenic[144,145].

The ubiquity of DNA ligases in replication and repair pathways makes them important enzymes to consider with respect to ribonucleotide incorporation. The product of RNase H incision is a potential substrate for DNA ligases, possibly leading to ligation of the site before the ribonucleotide can be fully excised. Studies of DNA/RNA ligation by DNA ligases, however, suggest that DNA ligases have an inherent mechanism for discriminating against 5' ribonucleotides[15,116,122]. Although the exact mechanism behind this type of discrimination has not yet been characterized, human DNA ligase 1 (LIG1) has been demonstrated to ligate DNA-RNA hybrids with a reduced catalytic rate[15,122]. Adenylylated ligation intermediates containing 5' ribonucleotides are readily repaired by aprataxin (APTX), suggesting that discrimination against 5' ribonucleotides involves interplay between LIG1 and APTX, similar to our observations with ligation of 3' damaged or mismatched ends described in Chapter 3.

Though the molecular mechanism behind the abortive ligation seen with a 5' ribonucleotide was previously unknown, the crystal structure of LIG1 provides some possible answers[15]. In the LIG1 structure, a conserved phenylalanine residue—F872—is positioned upstream of the nick, where it lies close to the deoxyribose moiety of the nucleotide at the 5' end (Figure 4.1). The proximity of F872 to the sugar indicates that the presence of a ribonucleotide at this position would cause a steric clash between the residue and the 2'-OH from the ribose. While a past study demonstrated that this residue is important for normal ligation by the Chlorella virus DNA ligase[146], the role of the phenylalanine in ribonucleotide discrimination has

not yet been examined. Here we present the first characterization of the mechanism underlying ribonucleotide discrimination by a DNA ligase. While mutations of F872 cause substantial catalytic deficiencies, substitution of the residue for alanine renders LIG1 unable to discriminate between DNA and RNA at the 5' end of the nick. Additionally, the residue also appears to be integral to positioning the end for adenylyl transfer by LIG1.

Materials and Methods

Preparation and purification of materials. The catalytic domain of human DNA ligase 1 (residues 232-919) was used in all assays and was expressed and purified as detailed previously[106]. The F872A and F872L mutants were constructed using site-directed mutagenesis, which was confirmed through sequencing of the expression vectors. Oligonucleotide substrates were synthesized by Integrated DNA Technologies (IDT), the Midland Certified Reagent Company and the Keck Biotechnology Resource Laboratory at Yale University. All oligonucleotides were purified via gel-purification as previously described (Chapter 2 and 3). The sequences of the oligonucleotides used in this study are shown in Figure S4.4.

Gel-based ligation assays. All ligation assays were performed at 37°C and with the standard ligation buffer (Chapter 2). Reactions were quenched using a quench solution containing 90% formamide and 50 mM EDTA. The different ligation species were separated on 15-20% polyacrylamide gels containing either 8 or 6.6 M urea. Prior to loading, all quenched samples were heated at 95°C for 5 minutes. The individual DNA species were detected using either a Typhoon Trio^{+} or a Typhoon 5 imager (GE Healthcare) and quantified using the ImageQuant software (GE Healthcare), as previously described[41].

Steady-state ligation assays. Steady-state ligation dependences were performed using 0.1-50 nM wildtype or mutant LIG1. For reactions with varying amounts of magnesium, the

concentrations of the 28mer DNA substrate and ATP were fixed at 0.5 μM and 0.2 mM, respectively. The steady-state dependence on the concentration of ATP was performed with varying amounts of ATP and concentrations of the 28mer substrate and $MgCl_2$ fixed at 0.5 μM and 20 mM, respectively. Steady-state DNA dependences were either performed in the presence of 20 mM $MgCl_2$ and 0.2 mM ATP or 2 mM $MgCl_2$ and 1 mM ATP. As stated previously (Chapter 3), reactions containing 2 mM $MgCl_2$ and 1 mM ATP are labeled as having 1 mM Mg^{2+}_{free} due to chelation of magnesium by ATP. Initial rates were determined by linear regression and were plotted as a function of the concentration of whichever component was varied. k_{cat} and K_M values were derived from fitting the initial rates using the Michaelis-Menten equation.

Single-turnover ligation assays. Single-turnover assays were performed with LIG1 held at 0.8 or 1.6 μM and the DNA substrate at a concentration of either 80 or 160 nM. All single-turnover reactions were performed in a heat-block on the benchtop. Reactions were performed in the presence of either 1 mM or 20 mM $MgCl_2$ and in the absence of ATP to prevent multiple-turnovers of the enzyme. The individual rates of adenylyl transfer and nick sealing were derived from fits of the data to a two-step irreversible model in Berkeley-Madonna as described previously[41].

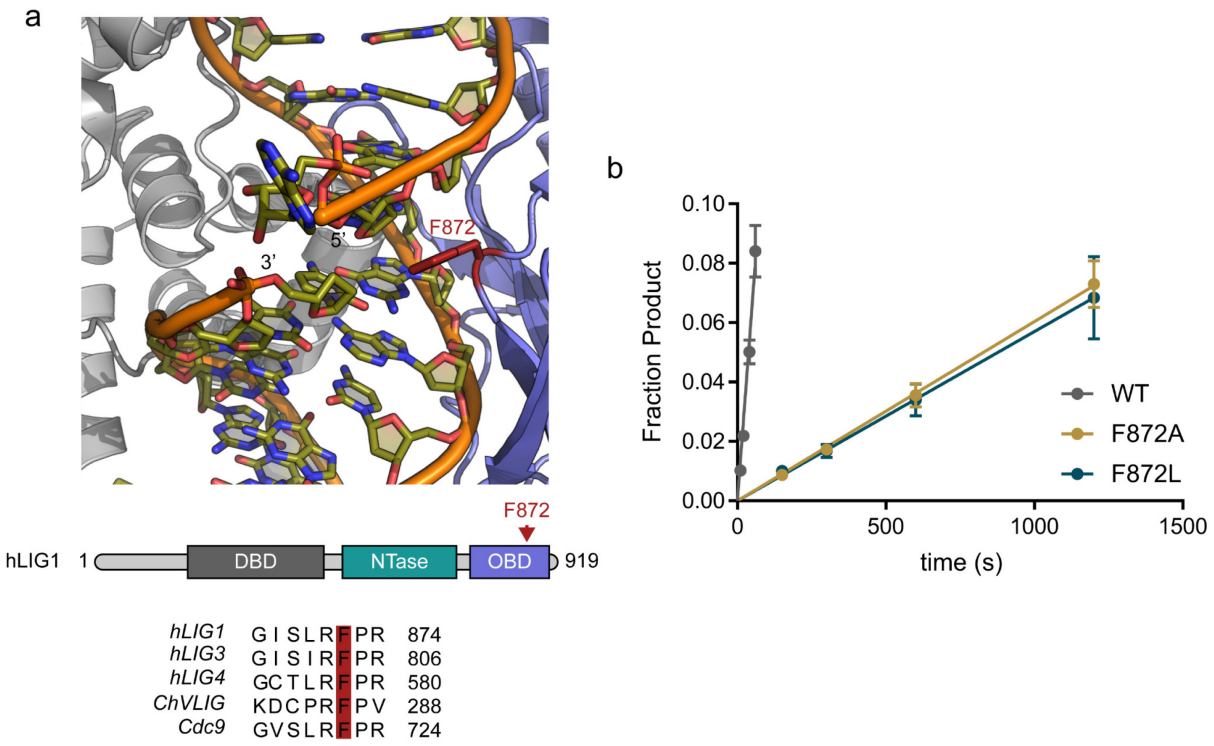


Figure 4.1. F872 is a conserved residue near the 5' nucleotide. (a) Crystal structure of LIG1 (PDB: 1X9N[15]) showing the proximity of the F872 residue (salmon red) to the 5'-adenylylated nucleotide in the nicked DNA intermediate. A sequence alignment of LIG1 to the other human DNA ligases (LIG3 and LIG4), as well as the Chlorella virus DNA ligase (ChVLIG) and the LIG1 homologue in *S. cerevisiae* (Cdc9), indicates that this phenylalanine is well-conserved in ATP-dependent DNA ligases. **(b)** Representative steady-state ligation assay showing the decrease in catalytic rate of LIG1-dependent ligation when the F872 residue is mutated to either an alanine or leucine. Reactions contained 1 nM [LIG1], 500 nM [28mer], 20 mM [Mg^{2+}] and 0.2 mM ATP.

Results

Mutation of the conserved phenylalanine residue causes defects in ligation

To interrogate the contribution of F872 to the overall fidelity of ligation by LIG1, we used site-directed mutagenesis to create F872A and F872L LIG1 mutants. Following expression and purification of the mutants, their concentration and activity were confirmed through active site titrations (Figure S4.1). Preliminary reactions with the two F872 mutants demonstrated a ~20-fold decrease in their steady-state rates of ligation relative to wildtype, indicating a catalytic defect in ligation by the two mutants (Figure 4.1). This effect is unsurprising, as the alanine mutation should allow for more flexibility and, therefore, less stable interaction between the

protein and the 5' nucleotide. In contrast, the presence of a leucine presumably enforces rigidity at this position, hindering efficient alignment of the 5' end in the active site.

To further characterize the defects in the two LIG1 mutants, we performed kinetic experiments to investigate the catalytic efficiency of ATP utilization (Figure 4.2a-b). As we saw in our initial steady-state experiments, the two phenylalanine mutants are ~20-fold slower than wildtype, with the F872A mutant acting about two-fold faster than F872L (Table 4.1). Both mutants have a slightly tighter $K_{M,ATP}$ than seen with wildtype LIG1, which makes the overall change in $k_{cat}/K_{M,ATP}$ less than 10-fold relative to wildtype (Figure 4.2b). The modest change in $K_{M,ATP}$ is consistent with F872 not being positioned near the putative ATP-binding site, implying that the effect on $k_{cat}/K_{M,ATP}$ is likely due to a defect in the DNA-dependent steps of ligation.

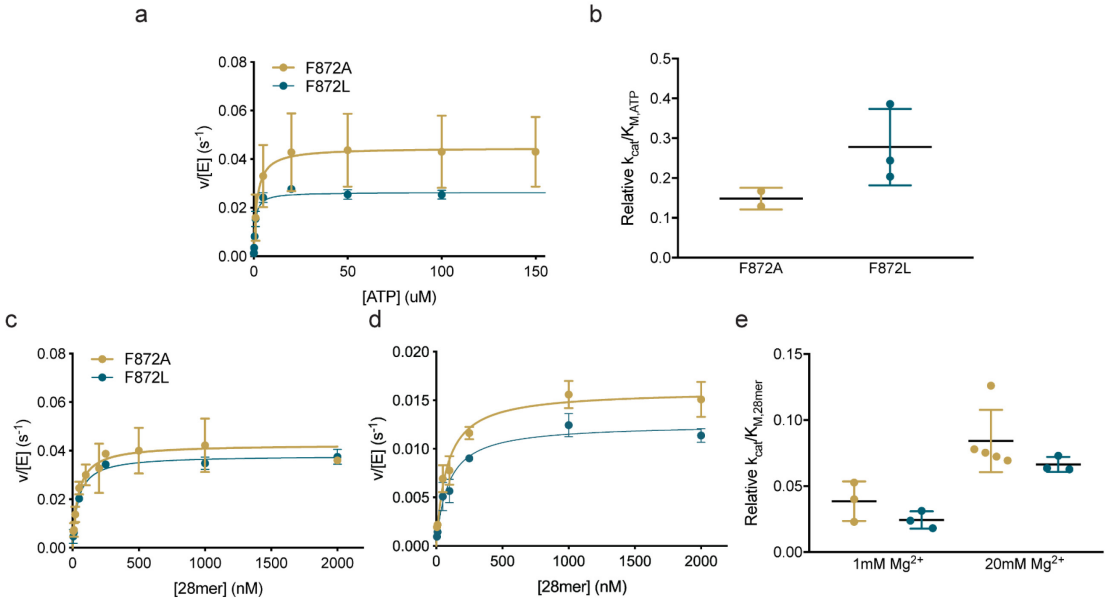


Figure 4.2. Effects of the F872 mutations on ATP- and DNA-dependences. (a) Steady-state ATP-dependence demonstrates that F872A and F872L LIG1 mutants exhibit slower k_{cat}, but no significant defect in $K_{M,ATP}$. Reactions contained 5 nM [LIG1], 500 nM [DNA] and 20 mM [$MgCl_2$] with varying concentrations of ATP. **(b)** Plot showing the relative $k_{cat}/K_{M,ATP}$ of the phenylalanine mutants to wildtype LIG1, highlighting a < 10-fold decrease in overall catalytic efficiency for ATP. Steady-state DNA-dependences were determined for the F872A and F872L mutants in the presence of either 20 mM (**c**) or 1 mM (**d**) Mg^{2+}. Reactions contained 0.1-10 nM [LIG1] and either 0.2 mM (**c**) or 1 mM ATP (**d**). (**e**) Plot showing the relative k_{cat}/K_M for the 28mer substrate of the phenylalanine mutants to wildtype LIG1. All initial rates were fit to the Michaelis-Menten equation and best fit parameters are reported in Table 4.1. Data for wildtype LIG1 used to create relative k_{cat}/K_M plot in c was adapted from the data presented in Chapter 3. All values are reported as the mean ± S.D. of N ≥ 3 replicates.

To assess the ability of the F872 mutants to engage the nicked DNA substrate, we next examined the steady-state DNA substrate dependence of the two mutants. In the presence of saturating amounts of both ATP and Mg^{2+}, there is almost no change in $K_{M,DNA}$ for either mutant relative to wildtype, suggesting the phenylalanine residue has little impact on engagement of the nicked DNA substrate (Figure 4.2c). The same holds true when both ATP and Mg^{2+} are held at 1 mM (Figure 4.2d). Although there is a noticeable drop in the $k_{cat}/K_{M,DNA}$ values for the two mutants relative to wildtype, the absence of a change in $K_{M,DNA}$ suggest that this decrease in catalytic efficiency could be due to the decrease in k_{cat} (Figure 4.2c-e). Interestingly, the relative decrease in k_{cat} is about two-fold worse at the lower concentration of magnesium, suggesting a magnesium-dependent defect.

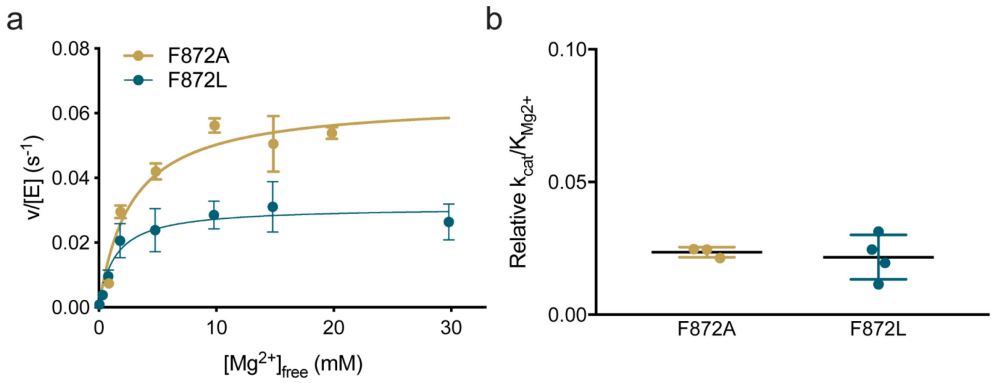


Figure 4.3. F872 mutations adversely affect catalytic efficiency of ligation. (a) The steady-state Mg^{2+}-dependence with the two F872 mutants demonstrates decreased catalytic efficiency, highlighted in the relative k_{cat}/K_{Mg2+} in **b.** All initial rates were fit to the Michaelis-Menten equation and best fit parameters are reported in Table 4.1**.** Data for wildtype LIG1 used to create relative k_{cat}/K_M plots in **b** and **d** was adapted from the data presented in Chapter 3. All values are reported as the mean ± S.D. of $N \geq 3$ replicates.

Crystal structures of LIG1 in complex with both substrate and the adenylylated intermediate reveal coordination of the catalytic magnesium ion by both the protein and the 5'-phosphate (Chapter 5), highlighting the importance of the F872 residue in stabilizing the 5' end. Since both mutants seem to have a significant impact on ligation, we reasoned that this may be due to a weakened Mg^{2+}-affinity resulting from misalignment of the 5' end. To assess how loss

of the phenylalanine residue impacts steady-state magnesium affinity, we performed steady-state Mg^{2+}-dependences with saturating amounts of both DNA and ATP (Figure 4.3a). Both F872A and F872L demonstrated slight defects in magnesium binding compared to wildtype, amounting to ~2-3-fold changes in K_{Mg2+}. The decrease in magnesium affinity results in a nearly 50-fold decrease in k_{cat}/K_{Mg2+}, indicating a weakened catalytic response to magnesium ions (Figure 4.3b).

Table 4.1 Best-fit parameters for steady-state ATP and Mg^{2+} dependences. Values for WT LIG1 are adapted from the data presented in Chapter 3.

ATP Dependence

	WT*	F872A	F872L
k_{cat} (s^{-1})	0.96 ± 0.02	0.045 ± 0.013	0.026 ± 0.002
K_M (μM)	9.2 ± 1.8	1.9 ± 0.76	0.46 ± 0.20
k_{cat}/K_M ($M^{-1}s^{-1}$)	1.0 ± 0.2 x 10^5	2.3 ± 1.1 x 10^4	5.7 ± 2.5 x 10^4

Mg^{2+} Dependence

	WT*	F872A	F872L
k_{cat} (s^{-1})	0.78 ± 0.03	0.064 ± 0.004	0.030 ± 0.006
K_{Mg2+} (mM)	0.78 ± 0.10	2.7 ± 0.33	1.4 ± 0.1
k_{cat}/K_{Mg2+} ($M^{-1}s^{-1}$)	1.0 ± 0.1 x 10^3	24 ± 11	21 ± 5

F872A does not discriminate against a 5' ribonucleotide

Results from the steady-state substrate and magnesium dependences are all consistent with the mutations of F872 causing a misalignment of the 5' end of the nick, whether by increasing or restricting the flexibility of the end. Since DNA ligases are known to discriminate heavily against 5'-ribonucleotides, we believed that the F872 residue is likely acting as a gate, occluding the 2'-hydroxl of a ribose sugar at this position. If this is true, then we believed that the two mutations should have contrasting impacts on the selectivity of the 5' end. To test this idea, we constructed a variant of our 28mer DNA substrate with a single ribonucleotide substitution at the 5'-end of the nick, referred to herein as R28mer.

The R28mer substrate has been previously shown to undergo ligation by LIG1 at a significantly reduced catalytic rate[15,122]. In agreement with these results, we observed a roughly 150-fold slower k_{cat} with the R28mer substrate than observed with the all-DNA 28mer (Figure 4.4a). This decrease in k_{cat} is more pronounced in the presence of physiological magnesium, which results in a ~350-fold decrease in the steady-state rate of ligation (Figure 4.4b). Compared to the 28mer, the R28mer is ligated with a significantly lower catalytic efficiency, suggesting discrimination against the 5'-ribonucleotide (Figure 4.4c).

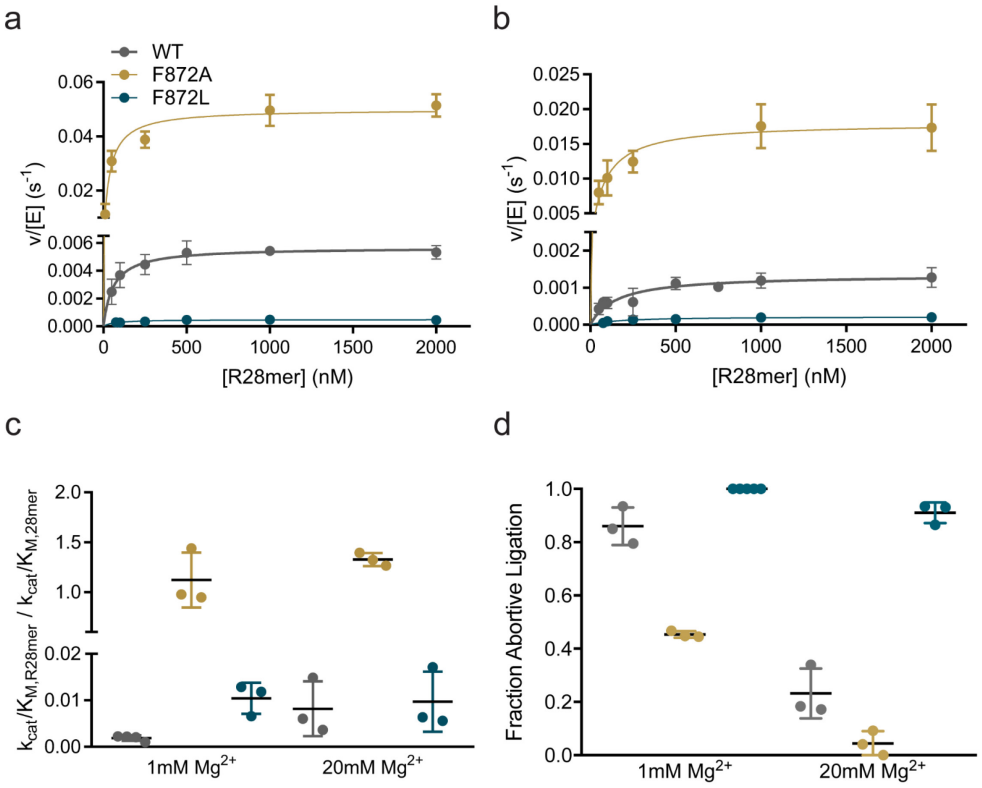


Figure 4.4. F872A LIG1 mutant drastically increases downstream RNA ligation. Steady-state substrate dependences were repeated for wildtype LIG1 and the mutants in the presence of either 20 mM (**a**) or 1 mM (**b**) Mg^{2+} with the R28mer substrate, containing a single ribonucleotide at the nick in 5' position. As in the 28mer dependences, reactions contained 0.1-10 nM [LIG1] and either 0.2 mM (**a**) or 1 mM ATP (**b**). **(c)** Plot showing the relative k_{cat}/K_M for the R28mer substrate to the value for wildtype LIG1. **(d)** Plot showing the amount of abortive ligation observed in the steady-state R28mer dependences. All initial rates were fit to the Michaelis-Menten equation and best fit parameters are reported in Table 4.2. All values are reported as the mean ± S.D. of $N \geq 3$ replicates.

A similar effect is seen with ligation of the R28mer substrate by the F872L mutant. As observed with wildtype LIG1, F872L performs ligation of the R28mer with little change in K_M but with a 100-fold slower catalytic rate as compared to ligation of the canonical 28mer (Figure 4.4a-c). Initial comparison of the relative k_{cat}/K_M values for wildtype and F872L LIG1 indicated that there is no significant change in substrate discrimination caused by the leucine mutation (Figure 4.4c). However, initial velocities are typically calculated as the amount of substrate reacted over time and, therefore, consider both sealed and adenylylated DNA as products. While this is valid for addressing how well a substrate is engaged by the enzyme, it does mask the presence of abortive ligation events where the adenylylated DNA is released by the ligase before being fully converted into sealed DNA. With this in mind, we calculated the fraction of all catalytic events that ended in abortive release of the adenylylated intermediate (Figure 4.4d). This type of analysis illustrates that nearly all collisions between the R28mer substrate and the F872L mutant end with the release of the adenylylated intermediate and the formation of negLIG1ble amounts of the sealed DNA product. This contrasts with wildtype LIG1, which can fully ligate the R28mer substrate in ~15% or ~80% of all catalytic events in the presence of 1 or 20 mM Mg^{2+}, respectively (Figure 4.4d). These results validate our prediction that the added rigidity of the leucine substitution would enforce stricter discrimination against ligation when there is a ribonucleotide present at the 5' end of the nick.

The ability of both wildtype and F872L LIG1 to discriminate against the 5'-ribonucleotide is essentially absent in the alanine mutant. In the presence of saturating amounts of ATP and Mg^{2+}, F872A ligates the R28mer substrate with the same efficiency as it does the 28mer (Figure 4.4a,c). Additionally, there is no observable abortive ligation in the presence of high magnesium (Figure 4.4d). Ligation of the R28mer is similarly unchanged when magnesium

concentrations are lower, although almost half of the catalytic events under this condition end with abortive ligation, which is still less than what is observed with either wildtype or F872L. However, this amount of abortive ligation is also observed in reactions between the mutant and the DNA-only substrate under the same conditions, indicating that abortive ligation of the R28mer substrate by F872A is likely indicative of a global ligation defect at low magnesium and not due to the presence of the ribonucleotide. Altogether, the data suggest that F872 is likely responsible for disfavoring ligation of a 5'-ribonucleotide by LIG1 and can be tuned by mutagenesis to be more selective or more promiscuous.

Table 4.2. Best-fit parameters for steady-state substrate dependences. Data for the wildtype LIG1 28mer dependences at high and low magnesium are adapted from the data in Chapter 3.

28mer Dependence

1 mM Mg^{2+}	WT*	F872A	F872L
k_{cat} (s^{-1})	0.46 ± 0.04	0.016 ± 0.002	0.012 ± 0.0005
K_M (nM)	79 ± 14	82 ± 40	90 ± 17
k_{cat}/K_M ($M^{-1}s^{-1}$)	5.9 ± 1.1 x 10^6	2.0 ± 1.0 x 10^5	1.4 ± 0.3 x 10^5

20 mM Mg^{2+}	WT*	F872A	F872L
k_{cat} (s^{-1})	0.83 ± 0.16	0.043 ± 0.010	0.038 ± 0.0001
K_M (nM)	68 ± 30	40 ± 13	43 ± 1.4
k_{cat}/K_M ($M^{-1}s^{-1}$)	1.3 ± 0.3 x 10^7	1.1 ± 0.3 x 10^6	8.9 ± 0.6 x 10^5

R28mer Dependence

1 mM Mg^{2+}	WT	F872A	F872L
k_{cat} (s^{-1})	1.3 ± 0.2 x 10^{-3}	1.8 ± 0.2 x 10^{-2}	2.1 ± 0.4 x 10^{-4}
K_M (nM)	125 ± 26	75 ± 10	155 ± 45
k_{cat}/K_M ($M^{-1}s^{-1}$)	1.1 ± 0.2 x 10^3	2.4 ± 0.5 x 10^5	1.3 ± 0.5 x 10^3

20 mM Mg^{2+}	WT	F872A	F872L
k_{cat} (s^{-1})	5.2 ± 0.2 x 10^{-3}	5.0 ± 0.3 x 10^{-2}	4.9 ± 0.2 x 10^{-5}
K_M (nM)	68 ± 30	36 ± 5	75 ± 34
k_{cat}/K_M ($M^{-1}s^{-1}$)	1.1 ± 0.5 x 10^5	1.4 ± 0.2 x 10^6	9.0 ± 4.8 x 10^3

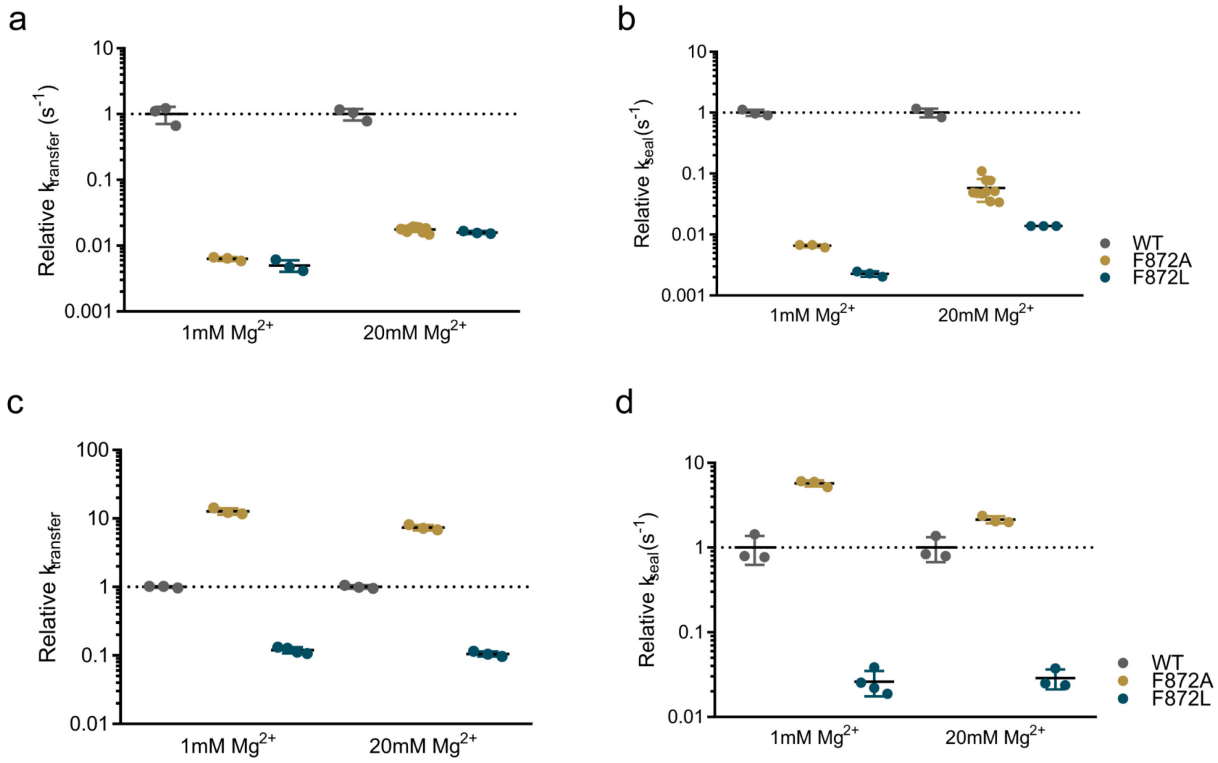


Figure 4.5. Single-turnover characterization of ligation of the 28mer and R28mer. Microscopic chemical rates were determined from single-turnover ligation assays containing 800 nM [LIG1] and 80 nM of either the 28mer (**a-b**) or the R28mer (**c-d**) substrate. The rates of adenylyl transfer and nick sealing are displayed as relative values, with all rates normalized to the values for WT LIG1. Reactions were performed in the presence of both 1 mM and 20 mM Mg^{2+}. Representative plots of the data are shown in Figure S4.3. Reaction traces were fit using a two-step irreversible model in Berkeley-Madonna and the best-fit values are shown in Table 4.3. All values are reported as the mean ± S.D. of N ≥ 3 replicates.

F872A increases the chemical rates of ligation of the R28mer substrate

To gain a deeper insight into how these phenylalanine mutations are affecting ligation of the R28mer substrate, we determined the microscopic rates of the individual chemical steps of ligation. To do this, we performed single-turnover ligation assays, in the absence of ATP, using a rapid-quench apparatus for rapid mixing of the LIG1 enzymes and either the 28mer or R28mer substrate. Unsurprisingly, the F872A and F872L mutants have diminished activity on the 28mer substrate (Figure 4.5a-b). The F872L mutation has a pronounced effect on the individual rates, most notably on the nick-sealing step at low magnesium, where the rate is nearly 500-fold slower

than wildtype LIG1. F872A performs the adenylyl-transfer and nick-sealing steps slightly faster than F872L, but still demonstrates significantly slower ligation kinetics than wildtype LIG1.

The single-turnover assays were then repeated with the R28mer, which severely hinders ligation by F872L—especially in the presence of 1 mM Mg^{2+}—to a similar extent as in the steady-state reactions (Figure 4.5c-d). Wildtype LIG1 also exhibits slower adenylyl transfer and nick sealing rates, although nick sealing is affected slightly less. The impact of the R28mer on adenylyl transfer is ~10-fold more severe than on nick sealing, indicating that the presence of the 2'-OH likely hinders the ability of LIG1 to properly align the 5' phosphate for adenylylation. Consistent with the findings in our steady-state assays, the 5'-ribonucleotide causes no significant impact on ligation by the F872A mutation, with both adenylyl-transfer and nick-sealing rates comparable to those seen in ligation of the unmodified substrate (Figure 4.5, Table 4.3).

Table 4.3. Best-fit parameters for single-turnover ligation assays. Data for the wildtype LIG1 at high and low magnesium are adapted from the data in Chapter 3.

28mer Ligation Rates

1 mM Mg^{2+}	WT*	F872A	F872L
$k_{transfer}$ (s^{-1})	3.6 ± 1.1	0.023 ± 0.001	0.018 ± 0.004
k_{seal} (s^{-1})	3.1 ± 0.4	0.020 ± 0.001	$7.0 \pm 0.7 \times 10^{-3}$

20 mM Mg^{2+}	WT*	F872A	F872L
$k_{transfer}$ (s^{-1})	2.6 ± 0.5	0.045 ± 0.004	0.041 ± 0.002
k_{seal} (s^{-1})	11 ± 2	0.63 ± 0.26	0.15

R28mer Ligation Rates

1 mM Mg^{2+}	WT	F872A	F872L
$k_{transfer}$ (s^{-1})	$2.0 \pm 0.1 \times 10^{-3}$	0.025 ± 0.002	$2.3 \pm 0.2 \times 10^{-4}$
k_{seal} (s^{-1})	$9.1 \pm 3.4 \times 10^{-3}$	0.052 ± 0.004	$2.4 \pm 0.8 \times 10^{-4}$

20 mM Mg^{2+}	WT	F872A	F872L
$k_{transfer}$ (s^{-1})	$5.6 \pm 0.3 \times 10^{-3}$	0.041 ± 0.004	$5.9 \pm 0.4 \times 10^{-4}$
k_{seal} (s^{-1})	0.24 ± 0.01	0.51 ± 0.04	$6.9 \pm 1.8 \times 10^{-3}$

Discussion

To uncover the molecular mechanism underlying the ability of LIG1 to faithfully discriminate against ligation of a nicked end containing a 5' ribonucleotide, we characterized the biochemical effects of two mutations of a conserved phenylalanine residue. While F872 had not previously been implicated in ribonucleotide discrimination, its proximity to the deoxyribose of the 5' nucleotide in LIG1 crystal structures indicated its possible involvement in preventing misligation of a ribonucleotide (Figure 4.1). Our steady-state assays indicate that mutations of this residue adversely affect the catalytic efficiency of LIG1. Slight defects in steady-state ligation indicate that the F872 residue likely plays a general role in DNA ligation. Due to its proximity to the nick site, F872 could be serving to position the 5' nucleotide and, therefore, facilitate interactions between the catalytic magnesium and the 5' phosphate.

To characterize the contribution of F872 to the ability of LIG1 to perform high-fidelity ligation, we performed steady-state and transient kinetic experiments to characterize ligation of a 5' ribonucleotide substrate by wildtype LIG1. While wildtype LIG1 can ligate the R28mer substrate, it does so with a significant decrease in efficiency. Additionally, under conditions of low magnesium, wildtype LIG1 aborts ligation of the R28mer more than it catalyzes the formation of the sealed product, consistent with a prior report[122]. Meanwhile, the F872L variant demonstrates severely diminished ligation of the R28mer substrate. Whereas the abortive ligation burden observed with wildtype LIG1 was partially alleviated when the magnesium concentration was raised to 20 mM, F872L LIG1 almost never catalyzed product formation in steady-state reactions with the R28mer, indicating that the leucine mutation increases selectivity at this

position. Unlike both wildtype and F872L LIG1, the F872A mutant was able to ligate the R28mer substrate with the same efficiency as the 28mer. The lack of selectivity in the F872A mutant was also obvious in single-turnover assays, where no effect was observed in the microscopic ligation rates with the R28mer substrate as compared to the 28mer substrate.

Along with the biochemical experiments presented here, we are working in collaboration for this project with Scott Williams at NIEHS, whose group is working to obtain crystal structures of wildtype and mutant LIG1 in complex with a ribonucleotide-containing substrate. Additionally, alongside Tom Kunkel's lab at NIEHS, we are working to characterize the cellular effects of the phenylalanine mutations in Cdc9, the LIG1 homolog in *S. cerevisiae*. If the phenylalanine residue is indeed allowing LIG1 to discriminate against ribonucleotide ligation, we expect that mutations of the residue could result in genomic instability and, therefore, growth defects in the yeast model. However, it will be important to assess if such an effect is due to the inability to discriminate against ribonucleotide ligation or just due to the slower rate of all ligation events, as observed in our *in vitro* assays. Additionally, we may see no effect from mutation of Cdc9 unless ribonucleotide excision repair is impaired, which will require us to test the mutations in a yeast strain lacking RNase H2.

If the results from these structural and cellular experiments agree with our biochemical characterization, we will have substantial evidence for a fidelity mechanism for enabling DNA ligases to discriminate against ligation of downstream ribonucleotides. This fidelity mechanism is distinct from the mechanism described in Chapter 3 for discrimination against upstream mismatches, suggesting that DNA ligases have multiple intrinsic fidelity mechanisms. The mechanism of ribonucleotide discrimination reported here is similar to the mechanism utilized by DNA polymerases, where a conserved methionine residue is used as a "steric gate" to prevent to

avoid incorporation of ribonucleotide monophosphates (rNMPs) into DNA [147]. Altogether, this demonstrates that multiple steps in DNA repair and replication can prevent the stable incorporation of ribonucleotides. Further characterization of the fidelity mechanism used by the different enzymes involved in repair and replication will allow us to better understand how these enzymes work in concert to maintain genomic fidelity.

Chapter 4 Appendix

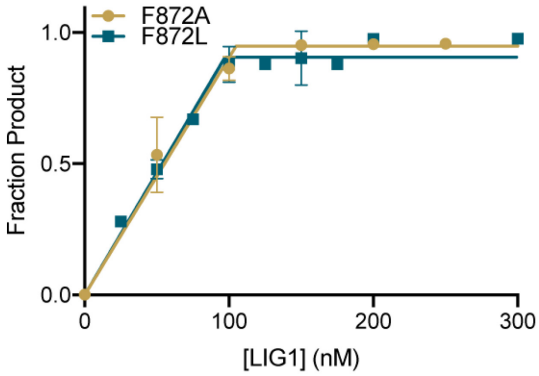


Figure S4.1. Active site titration of the F872A and F872L mutants. The concentration of active protein was estimated from active site titrations, where increasing amounts of LIG1 were titrated against a fixed amount of DNA (150 nM).

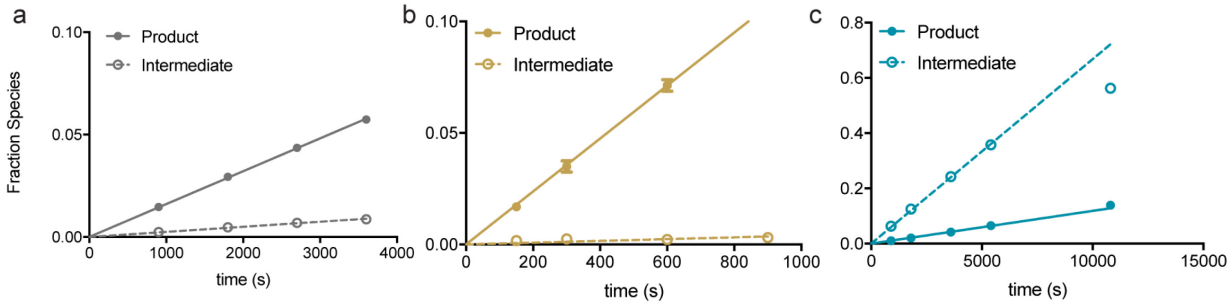


Figure S4.2. Representative traces from steady-state ligation reactions. Steady-state ligations contained either 2 nM [LIG1] (WT and F872A) or 100 nM [LIG1] (F872L) and 500 nM [R28mer]. Reactions were carried out in the presence of 0.2 mM ATP and 20 mM $MgCl_2$. A large amount of the adenylylated ligation intermediate accumulates in reactions with the F872L mutant, indicating a large abortive ligation burden for this enzyme.

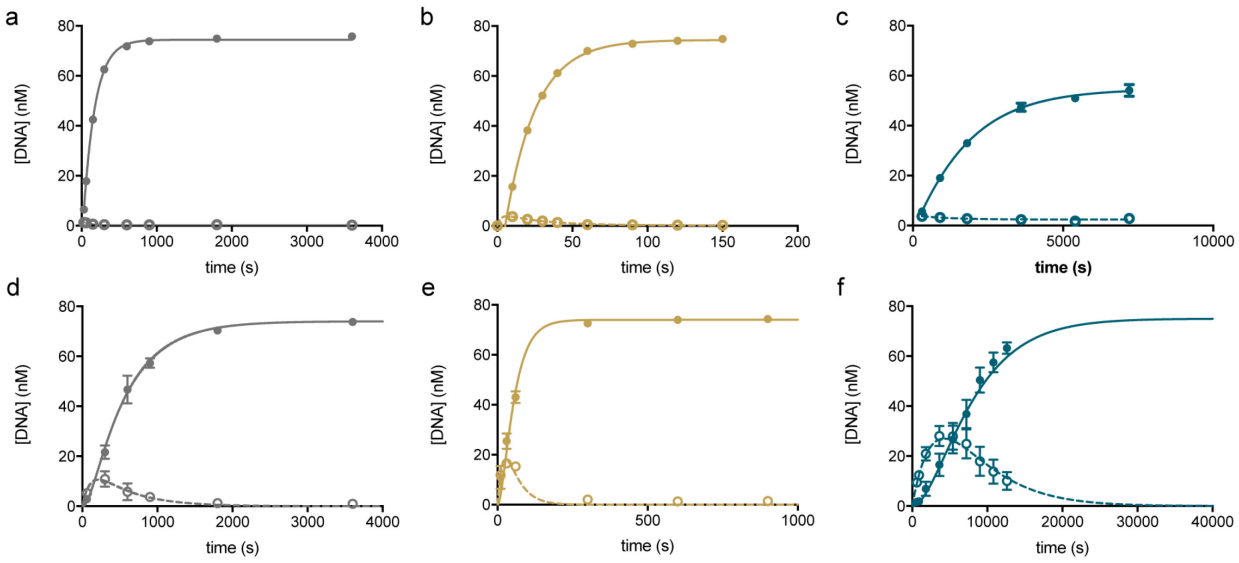


Figure S4.3. Representative traces for single-turnover ligation of the R28mer substrate. Single-turnover ligation reactions between wildtype (gray), F872A (gold) and F872L (green) LIG1 and the R28mer substrate. Reactions contained either 20 mM (**a-c**) or 1 mM (**d-f**) $MgCl_2$. The ligation product and intermediate DNA species are represented by closed and open circles, respectively. Berkeley-Madonna fits are shown for both product formation (solid lines) and the buildup and decay of intermediate (dashed lines).

Figure S4.4. Oligonucleotides used for annealing ligation substrates.

Oligo Name	5′ modification	Sequence	3′ modification
Temp28		5' CCGAATCAGTCCGACGACGCATCAGCAC 3'	
Up13OH		5' GTGCTGATGCGTC 3'	
P15FAM	Phosphate	5' P-GTCGGACTGATTCGG-FAM 3'	FAM
RP15FAM	Ribonucleotide monophosphate	5' P-rGTCGGACTGATTCGG-FAM 3'	FAM

Chapter 5
Identification and Role of Mg^{2+} Ions in DNA Ligation

In considering how DNA ligases achieve tight binding and high fidelity, it is important for us to understand the role of the catalytic magnesium throughout the ligation reaction. The binding mechanism of human DNA ligase 1 (LIG1) outlined in Chapter 2 suggests that magnesium stimulates conformational changes prior to catalysis. Magnesium is also required for all chemical steps of ligation. However, LIG1 has varying magnesium affinities throughout ligation, suggesting that there are unique interactions with Mg^{2+} during the different chemical steps[41].

Early observations of the structure of LIG1 suggested that the enzyme harbors two metal ions in the active site[15]. The approximate locations of these metals are consistent with the chemical steps of ligation, with one metal positioned to be able to activate the nucleophilic 3'-OH and the other near the 5'-PO_4, where it could both activate a hydroxyl group on the phosphate for adenylyl transfer and later stabilize the 5'-AMP leaving group for nick sealing. In contrast to this, the recent high-resolution crystal structures of LIG1 presented in Chapter 3 show a single magnesium coordinated in the LIG1 active site. This lone magnesium ion is observed in the active site when adenylylated LIG1 is bound to the nicked substrate and after the AMP has been transferred to the 5'PO_4 (Figure 5.1), suggesting that a single magnesium ion is used for both adenylyl transfer and nick sealing.

The work reported in Chapter 5 is the result of work done by multiple previous lab members. The project was originally undertaken by Neha Bokil as an undergraduate student working under the direction of Mark Taylor during his graduate studies. April Kaneshiro helped me re-purify some of the mutants presented here and performed early characterizations of the E720Q mutant as a graduate rotation student. Sydney Rosenblum performed many of the experiments for D570N and E720D, also as a graduate rotation student. Structural insights from the work of Percy Tumbale and Matthew Schellenberg in the lab of Scott Williams also contributed to the study presented here.

In both structures, a network of conserved residues is involved in the coordination of magnesium (Figure 5.1). The three central residues, D570, E621 and E720, have all been previously demonstrated to be important for ligation by the Chlorella virus DNA ligase (ChVLIG), in which mutagenesis of the analogous residues led to significant catalytic defects[21,22]. However, these ChVLIG studies predated the first structure of a DNA ligase bound to DNA and, perhaps because of this, did not address if mutation of these conserved residues impacts magnesium binding. In an attempt to confirm these conserved residues as metal ligands, we biochemically characterized mutations of each of the three residues in LIG1. In doing so, we found that mutation of any of the three residues leads to dramatic catalytic defects but differentially impact the magnesium affinity of LIG1. Although none of the mutations tested appear to significantly impact the apparent affinity for the DNA substrate, mutation of E621 drastically decreases the K_M for ATP, suggesting a unique role for this residue in stabilizing the ATP and AMP moieties. Although, we did not find evidence to exclude a one- or two-metal mechanism, we present some structural evidence supporting the one-metal mechanism that we hope to probe with future experiments.

Materials and Methods

Preparation and purification of materials. Oligonucleotide substrates used in this study were synthesized by Integrated DNA Technologies (IDT) and the Keck Biotechnology Resource Laboratory at Yale University. Purification of the oligonucleotides by gel-purification is described in Chapter 2. Mutagenesis of the putative metal ligands was accomplished using site-directed mutagenesis of the Δ232 LIG1 expression vector and the mutants were all purified as described previously[106]. The total concentration of purified protein was estimated using UV absorbance and the active concentration was verified through active site titration[41]. The standard

gel-based ligation assay was used for all kinetic characterizations in this study and is described elsewhere[41]. All assays were performed at 37°C and in solutions with a constant ionic strength held at 150 mM.

Steady-state ligation dependences. The concentration of LIG1 ranged from 0.1-10 nM for all steady-state ligation assays. In order to determine the steady-state magnesium dependence, ATP and DNA substrate concentrations were held at 0.2 mM and 0.5 μM, respectively. To characterize ATP utilization by the various mutants, the concentration of magnesium was held at 20 mM and the DNA substrate was maintained at 0.5 μM. The DNA substrate dependence was performed in the presence of 20 mM $MgCl_2$ and 0.2 mM ATP. In all cases, product formation was plotted as a function of time and initial rates were derived from linear fits to the data. In cases where both product and intermediate ligation species were observed, both species were considered as product formation so that resulting k_{cat} measurements reflect all catalytic events that occurred. All k_{cat} and K_M values were determined using the Michaelis-Menten equation.

Single-turnover ligation assays. All single-turnover assays were performed in the absence of ATP and in the presence of 80 nM 28mer substrate. Single-turnover assays for D570N, E621D and E720D were all performed with the enzyme concentration held at either 0.8 or 1.6 μM. For the preliminary single-turnover assays for the E621Q and E720Q mutants, the enzyme concentrations were held at either 1.6 or 3.2 μM, while the substrate remained at 80 nM. Reactions with the E621Q and E720Q enzymes were performed in a 37°C incubator to maintain a consistent temperature over the long timescale.

Fluorescence anisotropy pilot experiment. Fluorescence anisotropy was measured using a Tecan Safire plate reader with the excitation and emission wavelengths set to 495 and 515 nm, respectively. A master-mix containing the standard ligation buffer, 0.1 mM EDTA and 10 nM of

the TFAM-44mer substrate (Chapter 2) was put into the wells of a black 96-well, flat-bottomed microplate in 20 µL aliquots. Increasing amounts of the enzymes were titrated into each individual well and the plate was gently mixed for 30 seconds in the plate reader before measurements were taken. Resulting fluorescence traces (y) were all fit to a quadratic binding model as shown below.

Equation 5.1. $$y = constant + \ amplitude * \frac{[E]+[S]+K_D-\sqrt{([E]+[S]+K_D)^2-([E][S])}}{2[S]}$$

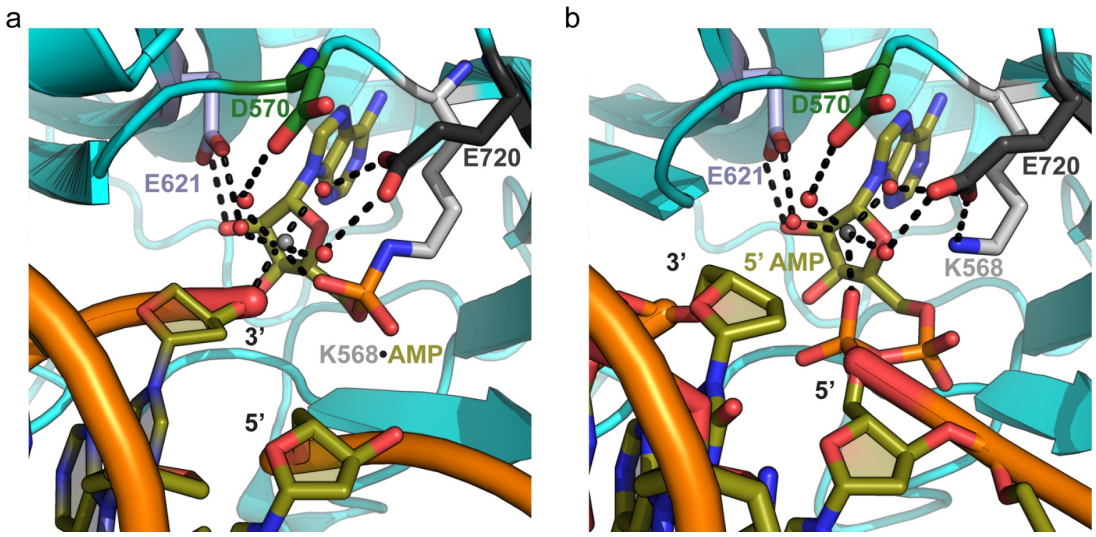


Figure 5.1 Crystal structures of LIG1 in complex with a lone magnesium. Crystal structures from Chapter 3 illustrating the interactions between the conserved carboxylate residues and a single magnesium ion both before (**a**) and after (**b**) adenylyl transfer. D570, E621 and E720 are all involved in water-mediated contacts with the magnesium ion. E621 also contacts the 2'-OH of the AMP both before and after adenylyl transfer.

Mutations of the putative metal ligands result in variable levels of activity

From previous studies of the Chlorella DNA ligase (ChVLIG), mutagenesis of the putative metal ligands to alanine residues seemingly abrogated all ligation activity. With this in mind, we chose to express and purify less deleterious mutations of the three residues. The two glutamate residues—E621 and E720—were each mutated either to aspartate or glutamine, in hopes that either the shorter sidechain or the nitrogen substitution would disrupt metal coordination while retaining measurable activity. Similarly, we substituted an asparagine residue for D570. We chose not to make a D570E mutation since ChVLIG appeared to be largely unaffected by the analogous mutation[148].

With all mutants purified to homogeneity, we verified ligation activity through a series of active site titrations (Figure 5.2a-b). Recombinant wildtype LIG1 predominantly purifies as an adenylylated species, routinely yielding ~90% of the adenylylated enzyme in purified stocks.

Similar behavior was observed for the D570N, E621D, and E720D mutants, with all three enzymes performing ligation in the absence of ATP within 90% of the concentration estimates determined by UV absorbance (Figure 5.2a). Both glutamine mutations—E621Q and E720Q—also showed ligation activity in these assays, albeit on a much slower timescale, requiring the reactions to progress overnight instead of the typical assay timescale of 15-30 minutes (Figure 5.2b). Even following an overnight incubation, E621Q was unable to form more than ~40% product, indicating that the glutamine mutation severely impacts the kinetics of ligation. Indeed, both E621Q and E720Q ligate the nicked DNA substrate at a rate that is >10,000-fold slower than wildtype LIG1 in ligation assays containing a large excess of enzyme and saturating amounts of magnesium (Figure 5.2c). Due to this, we chose to not characterize the glutamine mutants further, instead focusing on the characterization of the three more active mutants. Nonetheless, the dramatic effect of the glutamine mutations indicates that the E621 and E720 residues are likely essential for efficient DNA ligation.

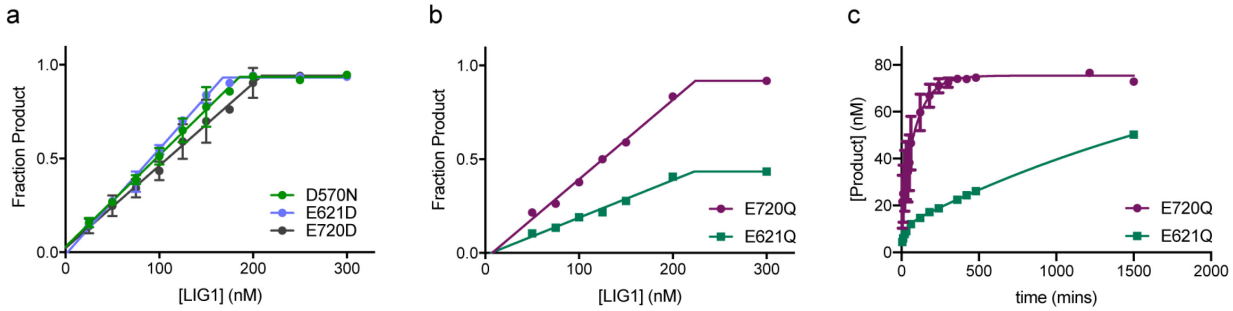


Figure 5.2. Determination of the active enzyme concentration. (**a**) Active site titrations of the D570N, E621D and E720D mutant LIG1 constructs. The active concentration is within 80-90% of the concentration estimated by UV absorbance. The enzymes were titrated against the 28mer substrate held at 150 nM in the absence of ATP and reactions were allowed to proceed to completion, which typically took 30 minutes. (**b**) Active site titrations of the E720Q and E621Q mutants were performed as in a but reactions were allowed to proceed overnight due to severe defects in the catalytic rate. E621Q only catalyzed ~50% product formation in the timeframe tested. The slow kinetics observed here were confirmed in single-turnover reactions (**c**) containing 1.6 or 3.2 μM E621Q or E720Q and 80 nM of the 28mer substrate. All data is shown as the mean ± S.D. of at least three independent replicates, except for **b**, for which only one replicate is shown.

Metal-ligand mutants exhibit magnesium-dependent defects in ligation

We first wanted to assess how the mutations impacted steady-state ligation and metal binding by LIG1. For this, reactions were performed in the presence of increasing amounts of magnesium and an excess of the DNA substrate over LIG1. Additionally, we included ATP—which is necessary to allow for multiple turnovers—at a constant concentration of 0.2 mM, which is a saturating amount for wildtype LIG1 (Figure 5.3). Resulting Michaelis-Menten plots for the magnesium dependences demonstrate a significant decrease in k_{cat} for all three mutants compared to wildtype LIG1. The most significant decrease in catalytic rate is seen with the E621D mutation, which is nearly 60-fold slower than wildtype, while D570N and E720D are only ~30- and ~20-fold slower, respectively. While D570N only has a ~3-fold weaker K_{Mg2+} value, both E621D and E720D have K_{Mg2+} values that are ~14-fold weaker, indicating a significant weakening of the magnesium affinity for these two enzymes. Altogether, these mutations cause a ~100-1000-fold decrease in k_{cat}/K_{Mg2+}—with the E621D mutation causing the most severe impact—indicating that all three mutations result in a decrease in the efficiency of magnesium utilization by LIG1 (Figure 5.3b).

To look at which steps of ligation are specifically affected by the three mutations, we performed ligation assays under single-turnover conditions with varying amounts of Mg^{2+}. Plots showing the formation of the ligation intermediate and its subsequent breakdown into the sealed DNA product were fit by a two-step irreversible model, from which the individual rates of adenylyl transfer and nick sealing were derived (Figure S5.1). Interestingly, the three mutants all perform adenylyl transfer at the same maximal rate, which is ~40-fold slower than wildtype (Figure 5.4a). The magnesium affinity for the adenylyl-transfer step is also similar for all three

mutants, exhibiting a substantial ~40-fold decrease in K_{Mg2+}, reducing the overall $k_{transfer}/K_{Mg2+}$ by more than 1,600-fold (Figure 5.4c, Table 5.1).

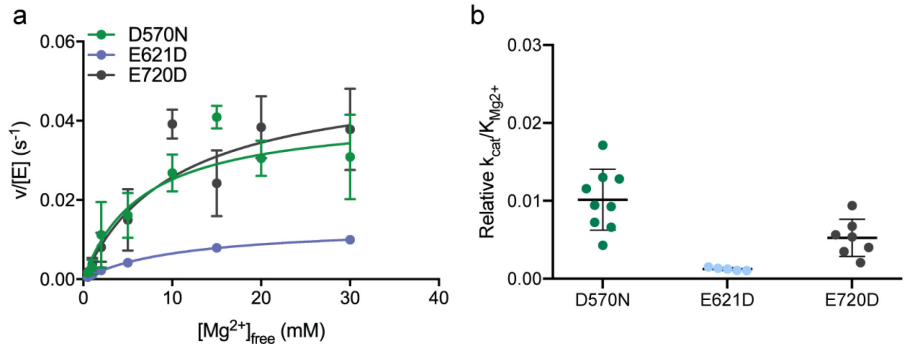


Figure 5.3. D570N, E621D and E720D exhibit large decreases in k_{cat}/K_{Mg2+}. (**a**) Steady-state magnesium dependences for the three mutants. Initial rates are plotted as a function of magnesium concentration and were fit using the Michaelis-Menten equation. The best-fit parameters are shown in Table 5.1. (**b**) The k_{cat}/K_{Mg2+} values for the three mutants relative to wildtype LIG1 were determined using the data for wildtype LIG1 presented in Chapter 3. All three mutants show large decreases in catalytic efficiency. Data is shown as the mean ± S.D. of three or more replicate experiments.

Both the D570N and E720D mutations affect the nick-sealing step to a similar degree, slowing down the maximal rate nearly 100-fold (Figure 5.4b). D570N and E720D catalyze nick sealing with K_{Mg2+} values of 18 and 35 mM, respectively, compared to ~3 mM for the wildtype enzyme (Table 5.1). Since the rate of nick sealing reached a maximum in the presence lowest tested magnesium concentration—0.3 mM—we are only able to place a limit of $\leq$ 0.3 mM for K_{Mg2+} for the E621D mutant (Table 5.1). Additionally, nick sealing is significantly slower than adenylyl transfer for the E621D mutant, leading to a rapid accumulation of the adenylylated DNA intermediate in the single-turnover assays (Figure S5.1b).

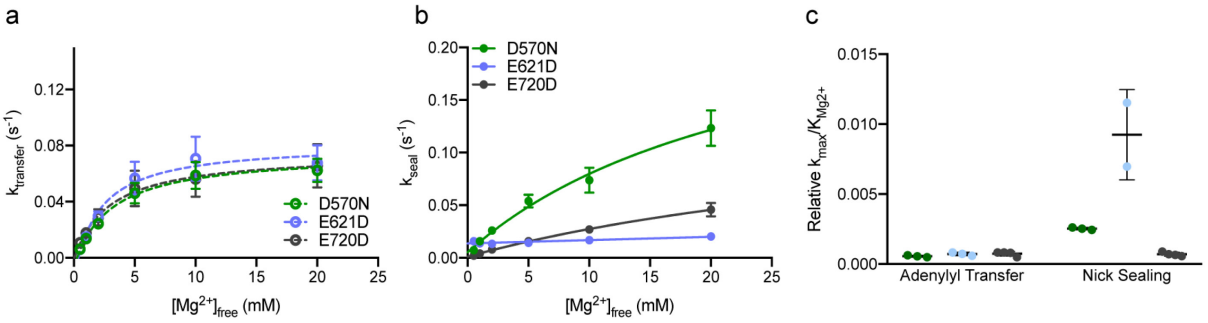


Figure 5.4. Mutation of the putative metal-ligands impacts both chemical steps. Single-turnover reactions were performed over a range of magnesium concentrations and with the mutants held in excess of the 28mer DNA substrate. Individual single-turnover traces were fit using a two-step irreversible model in Berkeley-Madonna, from which we derived the microscopic rates of ligation. The rates of adenylyl transfer (**a**) and nick sealing (**b**) are plotted as a function of magnesium concentration. Maximal rates and magnesium affinities were derived from hyperbolic fits of the data. (**c**) The k_{max}/K_{Mg2+} values for the individual mutants are represented as relative values to the corresponding k_{max}/K_{Mg2+} for wildtype LIG1. Data for wildtype LIG1 is adapted from the study in Chapter 3. The high k_{max}/K_{Mg2+} value for nick sealing by the E261D mutant is due the dramatic decrease in K_{Mg2+} caused by the mutation. Since a true K_{Mg2+} was not measured for nick sealing by E261D, this value only represents a lower limit. All data is shown as the mean ± S.D. of three or more replicate experiments.

Table 5.1 Best-fit parameters for steady-state and single-turnover magnesium dependences.

Mg^{2+} Dependence

Steady-state	D570N	E621D	E720D
k_{cat} (s^{-1})	0.025 ± 0.005	0.0138 ± 0.0004	0.050 ± 0.011
K_{Mg} (mM)	2.4 ± 0.7	11 ± 2	11 ± 5
k_{cat}/K_M ($M^{-1}s^{-1}$)	1.0 ± 0.4	0.12 ± 0.02	0.52 ± 0.22

Single turnover	D570N	E621D	E720D
$k_{transfer}$ (s^{-1})	0.079 ± 0.013	0.08752 ± 0.019	0.076 ± 0.020
$K_{Mg,transfer}$ (mM)	4.0 ± 0.3	3.5 ± 0.5	3.1 ± 0.9
k_{seal} (s^{-1})	0.23 ± 0.05	0.017 ± 0.001	0.1238 ± 0.01
$K_{Mg,seal}$ (mM)	18 ± 4	≤ 0.30	35 ± 3

E621D displays severe defects in ATP utilization

Having confirmed the weakened metal affinity for the three mutants, we were interested in how this potentially affects utilization of the two ligation substrates, ATP and DNA. We first probed the ability of the enzymes to perform enzyme adenylylation by performing steady-state ATP dependences (Figure 5.5a). We held both the nicked DNA substrate and magnesium at concentrations deemed saturating for wildtype LIG1, 500 nM and 20 mM, respectively. Since all three enzymes exhibit a significantly weaker K_{Mg2+}, choosing a single magnesium concentration

for steady-state assays is potentially problematic, as 20 mM Mg^{2+} is saturating for D570N, but not for either E621D or E720D (Table 5.1). However, increasing the Mg^{2+} concentration to a saturating concentration for the E621D and E720D mutants requires a concomitant increase in the ionic strength of the solution, complicating comparisons of the two mutants to D570N or wildtype LIG1.

Michaelis-Menten plots of the ATP-dependence of the three enzymes are shown in Figure 5.5A. The curves for both D570N and E720D follow a similar trend to one another, giving fits with $K_{M,ATP}$ values in the low micromolar range (~1-4 μM). The overall change in ATP utilization caused by the D570N and E720D mutations is a decrease of about 6- or 10-fold, respectively, which could be due solely to the defects in the DNA-dependent steps. The E621D mutation impedes enzyme adenylylation to a much larger degree, causing a 600-fold decrease in the catalytic efficiency of ATP utilization (Figure 5.5a-b). The large effect of the E621D mutation on enzyme adenylylation indicates that the E621 residue may play an outsized role in the initial enzyme adenylylation step.

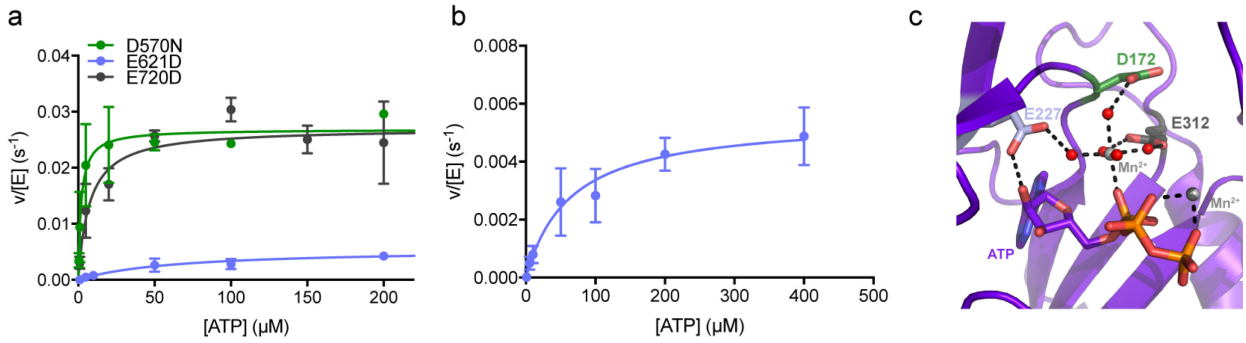


Figure 5.5. E621D severely impacts steady-state ATP affinity. (**a**) Steady-state ligation assays were performed with varying amounts of ATP, 20 mM $MgCl_2$ and 0.5 μM 28mer DNA substrate. Initial rates are plotted as a function of ATP concentration and were fit using the Michaelis-Menten equation. The full ATP-dependence for E621D is shown in **b**. Data is shown as the mean ± S.D. of three or more replicate experiments. (**c**) The crystal structure of an RNA ligase from *Naegleria gruberi* highlighting conserved residues in the active site (PDB: 5COU[149]). D172, E227 and E312 are analogous to D570, E621 and E720 in LIG1. The ligase is ATP-bound and is trapped in a state preceding the enzyme-adenylylation step.

Although no structural information about the enzyme adenylylation step exists for a mammalian DNA ligase, the residues required for this step are conserved in all known ATP-dependent DNA and RNA ligases. Recent structures of an RNA ligase from *Naegleria gruberi* (NgrRnl) captured the enzyme primed for enzyme adenylylation, showing coordination of the ATP substrate in the adenylylation core[149]. Figure 5.5c highlights residues D172, E227 and E312 in the RNA ligase, which correspond to D570, E621 and E720 in LIG1. The carboxylate moiety of E312 moiety coordinates two water-mediated contacts to the catalytic manganese, possibly explaining why mutation of the analogous residue in LIG1—E720—yields a significant decrease in catalytic efficiency for ATP. E227—E621 in LIG1—is also involved in a water-mediated contacts with the metal ion, but more notably contacts the 2'-OH of the bound ATP molecule (Figure 5.5c). This is similar to the contact observed between E621 and the 5'-AMP in the LIG1 crystal structure, indicating that this residue is likely contacting the AMP group throughout the ligase catalytic cycle (Figure 5.1).

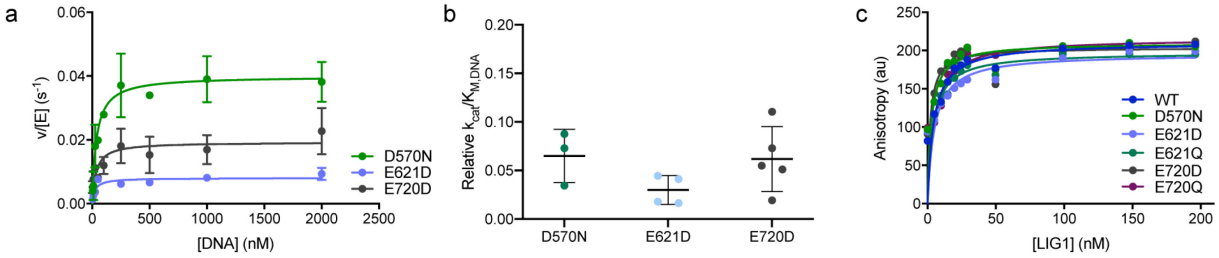


Figure 5.6. Mutations do not appear to cause significant DNA-binding defects. (**a**) Steady-state dependences were performed with 20 mM Mg^{2+} and 0.2 mM ATP. The initial rates were determined over a range of concentrations of the 28mer DNA substrate and were fit using the Michaelis-Menten equation. (b) Relative $k_{cat}/K_{M,DNA}$ values for the individual mutants compared to the values for wildtype determined in Chapter 3 under the same conditions. All values reflect the mean ± S.D. of at least three independent replicates and are shown in Table 5.2. (**c**) Anisotropy values obtained with a titration of increasing amounts of the different LIG1 enzymes against a fixed concentration of the TFAM-44mer reporter substrate (10 nM). All traces were fit using a quadratic binding equation and the corresponding values are shown in Table 5.3.

The metal-ligand mutations do not appear to impact affinity for DNA

Steady-state DNA dependences were performed similar to the ATP dependences, with magnesium and ATP held at concentrations deemed saturating for wildtype LIG1 (Figure 5.6a). Under these conditions, the three mutants exhibit decreased k_{cat} values compared to wildtype LIG1, showing the same ~20-60-fold decreases observed in the other steady-state dependences and leading to an overall decline in catalytic efficiency (Figure 5.6b). Remarkably, all three mutants have low $K_{M,DNA}$ values, ranging from 20 to 40 nM. This is nearly equal to the $K_{M,DNA}$ value obtained with wildtype LIG1 under the same conditions, which is close to 60 nM (Chapter 3)

To directly test for any defects in DNA-binding, we titrated increasing amounts of the LIG1 constructs against a fixed amount of the TFAM-44mer substrate—characterized in Chapter 2—to detect binding via anisotropy (Figure 5.6c). In addition to the three characterized mutants of LIG1, we also measured the binding affinity of the two glutamine mutations—E621Q and E720Q—to test whether the large defects in ligation by the two mutants are partially due to a DNA-binding defect. The resulting binding curves for the different LIG1 enzymes were all fit to a quadratic binding model in order to account for ligand depletion. All of the tested LIG1 enzymes exhibit tight binding affinity for the DNA substrate in the low nanomolar range (Table 5.3). With the exception of E720D and E621D, the mutants show a similar affinity for the substrate as wildtype, with K_D values around 3-5 nM. E621D has a ~3-fold weaker affinity while E720D exhibits a binding affinity that is ~3-fold tighter than that of wildtype. Since the binding data is still preliminary, no definitive conclusions can be drawn from this assay, however it does

appear that mutations of the putative metal ligands do not significantly impact DNA binding by LIG1.

Table 5.2. Best-fit parameters for steady-state substrate dependences

ATP Dependence

	D570N	E621D	E720D
k_{cat} (s^{-1})	0.029 ± 0.002	0.013 ± 0.006	0.020 ± 0.003
K_M (μM)	1.8 ± 0.9	123 ± 59	4.4 ± 2.0
k_{cat}/K_M ($M^{-1}s^{-1}$)	1.6 ± 0.8 x 10^4	1.0 ± 0.7 x 10^2	4.9 ± 2.2 x 10^3

DNA Dependence

	D570N	E621D	E720D
k_{cat} (s^{-1})	0.040 ± 0.007	0.0081 ± 0.0011	0.018 ± 0.005
K_M (nM)	41 ± 19	26 ± 13	21 ± 8
k_{cat}/K_M ($M^{-1}s^{-1}$)	9.7 ± 4.5 x 10^5	3.1 ± 1.6 x 10^5	8.6 ± 3.9 x 10^5

	K_D (nM)	Relative K_D*
WT	5.7	1.2
D570N	2.3	0.5
E621D	17	3.4
E621Q	7.7	1.6
E720D	0.8	0.2
E720Q	5.9	1.2

Table 5.3. Preliminary K_D values for DNA binding by wildtype and mutant LIG1. The relative K_D values were calculated by dividing the obtained K_D values by the value obtained for wildtype LIG1 in the equilibrium-binding assay described in Chapter 2.

Discussion

D570, E621 and E720 have all been recently identified as putative metal ligands in the crystal structure of LIG1 (Chapter 3). Therefore, we sought to biochemically characterize mutants of the three conserved residues to gain insight into the mechanism underlying DNA ligation. Mutagenesis of either E621 or E720 to glutamine resulted in severely diminished enzymatic activity, especially for the E621Q mutant. D570N, E621D and E720D all displayed significant DNA ligation defects but were still able to ligate the 28mer substrate to completion. All three mutations caused significant defects in steady-state utilization of magnesium by LIG1. Additionally, the mutations all similarly impact the adenylyl-transfer step of ligation, resulting in decreases in both the rate of adenylyl transfer and the magnesium affinity for the step. While

both D570N and E720D exhibited significant decreases in their K_{Mg2+} values for nick sealing, the apparent magnesium affinity of E621D was tighter than what is observed with the wildtype enzyme. E621D also catalyzed nick sealing with a much slower rate than either of the other two mutants, suggesting that this residue may play a unique role during the nick-sealing step. E621D also exhibited significantly diminished ATP utilization, consistent with E621 being essential for helping to properly position ATP or AMP in the active site through interactions with the 2'-OH.

During our biochemical characterization of the two mutants, we did not find evidence to suggest that LIG1 preferentially uses one or two magnesium ions for catalysis. While enzyme adenylylation has been demonstrated to require two metal ions, with one of the two coordinated between the β and γ phosphates of ATP (Figure 5.5c)[41,149], the number of metal ions required for the DNA-dependent steps of ligation remains unclear. In both the original crystal structure of LIG1 and the more recent structures described in Chapter 3, a metal ion is observed near the upstream 3'-OH. The absence of a second metal in the recent structures near the downstream 5'-PO_4 suggests that a metal ion may not be required for stabilization of the AMP. Further examination of the LIG1 active site reveals a conserved lysine residue (Lys744) interacting with an oxygen atom from the AMP moiety, both before and after adenylyl transfer (Figure S5.2 and S5.3). Substitution of the analogous residue in ChVLIG to alanine results in a loss of 98% of ligation activity relative to the wildtype enzyme, indicating that the residue likely plays an important role in catalysis[22]. Additionally, the mutation disproportionately affected the final step of ligation, nick sealing, where stabilization of the AMP leaving group would be most important.

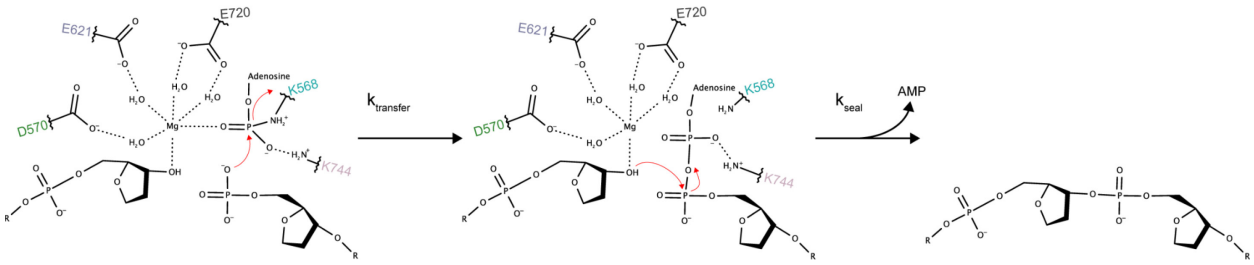


Figure 5.7. Hypothetical one-metal mechanism for DNA ligation. The lone metal ion observed in crystal structures of LIG1 is coordinated by the three residues characterized in this study: D570, E621 and E720. K744 is positioned near the upstream end, where it is seen contacting an oxygen from the AMP group, possibly helping to stabilize AMP as the leaving group during nick sealing. The interaction between E621 and the AMP ribose is omitted for clarity but also appears to be important for the final sealing step.

The ChVLIG studies and positioning of K744 in the LIG1 crystal structures leads us to a hypothetical one-metal mechanism, where the lysine residue would be responsible for stabilizing the AMP leaving group for attack by the 3'-OH (Figure 5.7). Proper positioning of the AMP moiety in the active site could be achieved by E621, which contacts the 2'-OH of the ribose ring throughout the DNA ligation reaction (Figure 5.5c and S5.2). This interaction appears to be important for enzyme adenylylation, as the E621D mutant exhibited a large decrease in efficient ATP utilization, but could also be key for positioning the AMP for interactions with K744. Since the available crystal structures of LIG1 complexed with DNA after the adenylyl-transfer step have all been solved with an upstream dideoxy-block, the exact positioning of the metal prior to nick sealing is unclear. However, it is possible that the active site metal is only necessary for activation of the nucleophilic 3'-OH during nick sealing and that stabilization of the leaving group is accomplished directly by the protein. A deeper characterization of the K744 residue would help to further elucidate the underlying ligation mechanism.

Chapter 5 Appendix

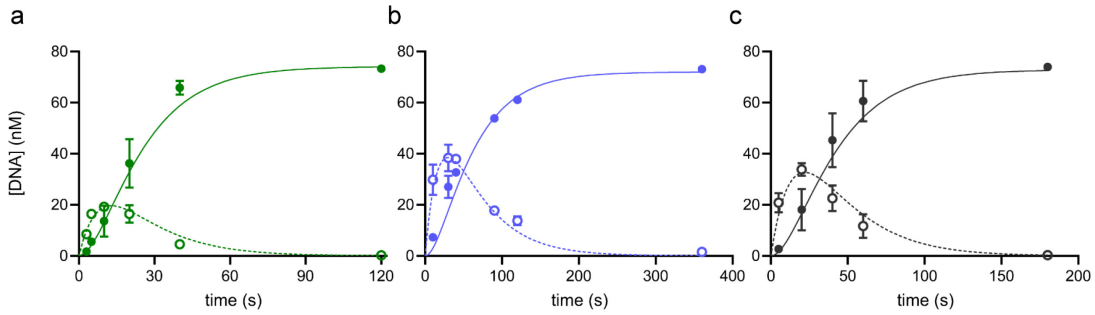


Figure S5.1. Representative single-turnover traces. Representative traces for single-turnover reactions with D570N (**a**), E621D (**b**) and E720D (**c**). Reactions shown here contained 20 mM Mg^{2+} and 80 nM of the 28mer substrate.

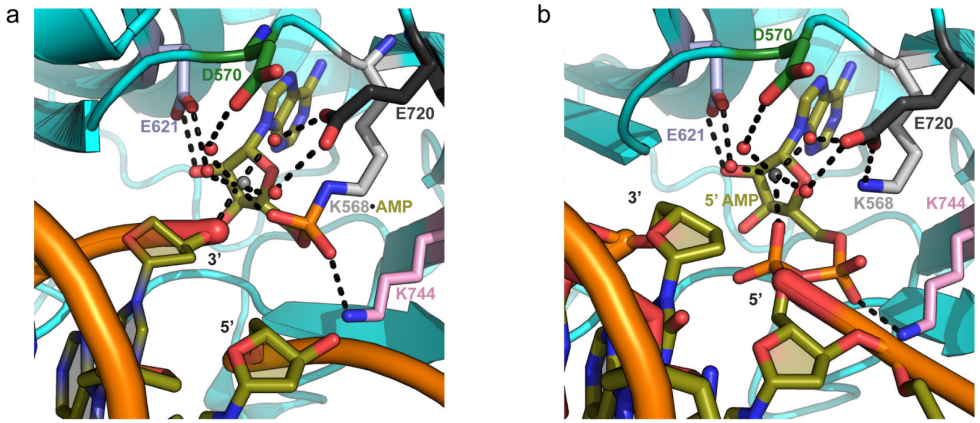


Figure S5.2. Lys744 contacts the AMP moiety before and after adenylyl transfer. Crystal structures from Chapter 3 illustrating the interactions between the conserved carboxylate residues and a single magnesium ion both before (**a**) and after (**b**) adenylyl transfer. K744 is observed contacting an oxygen atom on the 5'-phosphate group on the AMP moiety.

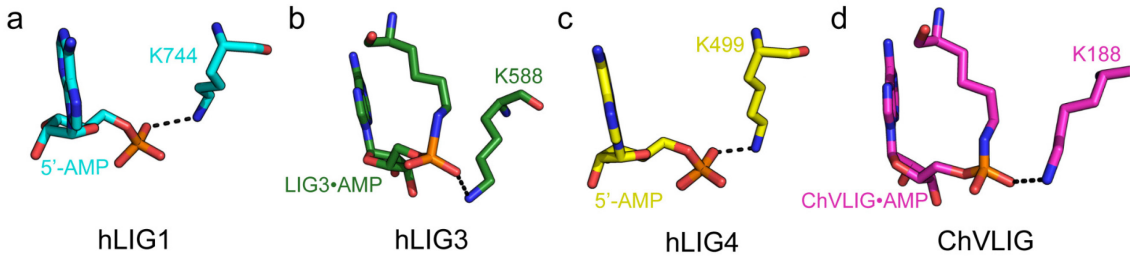


Figure S5.3. Conservation of the lysine-AMP interaction. Crystal structures from LIG1 (**a**), LIG3 (**b**), LIG4 (**c**) and ChVLIG (**d**) show the conservation of a lysine-mediated contact between the ligase and the phosphate moiety of AMP. Structures were adapted from the following PDB entries: 1X9N[15], 3L2P[24], 6BKG[25] and 2Q2T[13].

Chapter 6
Biochemical Characterization of LIG1-Syndrome Mutations

The mammalian DNA ligase family is a group of highly conserved enzymes with essential functions in most DNA-dependent cellular processes. The three mammalian DNA ligases share a conserved catalytic core but have diverse N- and C-termini that define their specific cellular roles[36,37]. DNA ligase I (LIG1) is directed to the DNA repair and replication machinery through interactions between its N-terminus and PCNA, where it engages in single-strand break repair and in Okazaki fragment maturation. DNA ligase 3 (LIG3) is the only DNA ligase found in mammalian mitochondria but is also found in the nucleus where it interacts with the nuclear scaffolding protein XRCC1[49,81,150]. LIG3 participates in both single-strand and double-strand break repair, occasionally acting in a redundant manner with the other two ligases[71,93,94]. Unlike LIG1 and LIG3, DNA ligase IV (LIG4) does not appear to participate in single-strand break repair. Instead, LIG4 is a core member of the non-homologous end joining pathway (NHEJ), which repairs double-strand breaks in the absence of homologous recombination[96,151].

The functions of all three mammalian DNA ligases—DNA ligases I, III and IV—have been implicated in the onset and/or maintenance of cancer, mainly due to their vital role in maintaining genomic integrity[152-154]. DNA ligases I (LIG1) and IV (LIG4) have also been implicated in immunological diseases. For LIG4, the connection to proper immune system

---

Some of my work reported in Chapter 6 was included in the following citation: Maffucci, P. et al. Biallelic mutations in DNA ligase 1 underlie a spectrum of immune deficiencies. *J Clin Invest* 128, 5489-5504, doi:10.1172/JCI99629 (2018).

The crystal structures presented here are preliminary structures adapted from unpublished work by Percy Tumbale and Matthew Schellenberg working under the guidance of Scott Williams at the National Institute of Environmental Health Sciences.

function is obvious, as it plays a key role in the NHEJ pathway that concludes the V(D)J recombination pathway, which is involved in the adaptive immune response[155]. LIG4 syndrome is characterized by a collection of immunodeficiencies caused by disparate mutations in the LIG4 gene[107,108]. Those afflicted by LIG4 syndrome can also exhibit hypersensitivity to UV-radiation and developmental disorders, owing to the role of NHEJ in DNA repair.

LIG1 syndrome was first characterized in the early 90's when a patient exhibiting signs like of developmental defects, UV sensitivity and immunodeficiency was initially misdiagnosed as having Bloom's syndrome[103,104,156]. The cause of these symptoms was identified as deficient LIG1 activity, caused by a E566K mutation in one allele and a R771W mutation in the other[105]. Biochemical studies of the E566K mutation showed that it is a catalytically dead enzyme[157]. A later crystal structure of LIG1 revealed that the E566 residue lies in the enzyme active site[15]. The effect of the R771W mutation on LIG1 activity was unclear at the time, although studies noted delayed Okazaki fragment maturation in the patient-derived 46BR cell line indicating decreased LIG1 activity[105]. Attempts to biochemically characterize the R771W variant were previously unsuccessful due to an apparent instability of the recombinant LIG1 mutant.

Recently, five additional patients harboring biallelic LIG1 mutations were identified as exhibiting similar symptoms as the original patient described in the literature[106]. Exome sequencing showed that two to the patients share the same heterozygous mutations in LIG1, despite no familial connection between the two. One of the shared variants—T415Mfs*10—is a truncated version of LIG1, only containing the N-terminal region and a portion of the DNA-binding domain (DBD). The other shared variant harbors a R641L mutation, which lies in the nucleotidyl-transferase (NTase) domain of LIG1. The other three patients, who all share a maternal grandmother, are homozygous for a LIG1 variant containing a R771W, the same

mutation found in the original LIG1 syndrome patient, and a P529L mutation in the DBD. Initial biochemical characterization of P529L LIG1 indicated a negLIG1ble change in activity for this mutant compared to wildtype[106]. Engineered cell lines containing the patient-derived R641L and T415Mfs*10 mutants of LIG1 exhibited increased DNA damage sensitivity, as did isolated lymphoblastoid β cells from the patients. Additionally, one of the patients harboring both the R641L and T415Mfs*10 mutations, as well as their mother who only has the R641L mutation, exhibited a decrease in somatic hypermutation, consistent with the observed immunodeficiencies.

To contribute to the understanding of a link between aberrant LIG1 activity and the onset of disease, we were interested to investigate the biochemical consequences of these mutations. Due to the severity of the effects of the R641L and R771W mutations in cell-based experiments, we focused solely on these two mutants for biochemical characterization. The two mutations both effect DNA-binding surfaces on the protein. Surprisingly, while we did observe noticeable defects in catalytic efficiency on a nicked DNA substrate, there were no obvious defects in DNA-binding. Instead we observed many abortive ligation events catalyzed by the two mutants and an unexpected decrease in magnesium affinity. Crystal structures of the two mutant LIG1 proteins uncovered many distortions in both the phosphodiester backbone of the substrate and the DNA-binding surfaces of LIG1.

Materials and Methods

Preparation and purification of materials. The catalytic core of human DNA ligase I (residues 232-919) was expressed in *E. coli* and purified as previously reported[106]. Recombinant enzymes harboring the patient-derived mutations were made using site-directed mutagenesis and the mutagenesis was confirmed by sequencing the expression vector. Concentration of the adenylylated enzyme was determined by active site titration against a fixed concentration of the nicked 28mer substrate. Substrate oligonucleotides were synthesized by Integrated DNA Technologies (IDT) and the Keck Biotechnology Resource Laboratory at Yale University. Lyophilized oligos were gel-purified on denaturing polyacrylamide gels and annealed as previously described[41]. The sequences of both the 28mer and the TFAM-44mer substrates are described in Chapter 2.

Gel-based ligation assays. All ligation assays were performed at 37°C with a standard reaction buffer containing 50 mM NaMOPS pH 7.5 (adjusted at 25°C), 0.1 mM dithiothreitol, 0.05 mg/mL BSA and varying amounts of NaCl to maintain a constant ionic strength of 150 mM. Ligation assays were performed at 37ºC and quenched with an equal volume of the gel-loading solution containing 50 mM EDTA/90% formamide. Quenched reactions were heated to 95°C prior to loading on 15-20% (wt/vol) polyacrylamide gels containing 6.6-8 M urea. The fluorescein-labeled species were detected using either a Typhoon Trio^{+} or a Typhoon 5 imager (GE Healthcare) and quantified using the ImageQuant software (GE Healthcare) as previously described[41].

Steady-state ligation assays. Steady-state ligation reactions were typically performed with 0.1-10 nM LIG1. DNA-dependences were performed in the presence of either 20 mM $MgCl_2$ and 0.2 mM ATP or 2 mM $MgCl_2$ and 1 mM ATP. Reactions with 2 mM $MgCl_2$ contain

1 mM Mg^{2+}_{free} due to chelation of magnesium by ATP[41]. For reactions with varying amounts of ATP, the concentrations of $MgCl_2$ and the 28mer substrate were held at 20 mM and 0.5 μM, respectively. For the magnesium dependence, ATP was held at 0.2 mM and the 28mer substrate was held at 0.5 μM. Initial rates of reaction were obtained from the slopes of linear fits to the fraction of product as a function of time, which were then plotted as a function of the concentration of the varied component. Values for k_{cat} and K_M were obtained from fits of the data to the Michaelis-Menten equation.

Rapid-quench flow ligation assays. Single-turnover assays were carried out at 37ºC either in a heat block or by using a KinTek RQF-3 quench-flow apparatus. Ligation reactions in the quench-flow were prepared by loading one sample syringe with 1.6 μM LIG1 and the other with 160 nM of the DNA substrate. Both syringes contained varying amounts of $MgCl_2$. Ligation was initiated upon mixing of the solutions in a 1:1 ratio, followed by quenching of the reaction at various timepoints with the gel-loading solution. For reactions not requiring rapid-quenching, the solutions were made similarly but were mixed 1:1 (v/v) using a micropipette. Rates for the adenylyl-transfer and nick-sealing steps were obtained from fits of the data using a two-step irreversible model in Berkeley-Madonna as previously described[41].

Steady-state fluorescence binding assays. Fluorescence intensity and anisotropy measurements were collected using a FluoroMax 3 fluorometer (Horiba) controlled by the DataMax software. Data was collected using excitation and emission wavelengths of 495 nm and 515 nm, respectively. Binding assays were performed by titration of increasing amounts of LIG1 against a 2.5 mL solution containing 1-2 nM of the TFAM-44mer and the standard reaction buffer with 0.1 mM EDTA. After addition of LIG1, samples were allowed to equilibrate for 30 seconds to 1 minute before measurements were taken. Anisotropy and relative fluorescence are

plotted as a function of added [LIG1] and the data was fit to a hyperbolic binding curve as shown in Equation 2.2 (Chapter 2).

Stopped-flow kinetic experiments. Transient kinetic experiments were performed using the Hi-Tech SF-61DSX2 stopped-flow instrument controlled by the Kinetic Studio (TgK Scientific) software. Data from both PMT detectors were combined and averaged for each shot. A minimum of 5-6 shots were performed and averaged for each concentration tested. Fluorescein fluorescence was measured using an excitation wavelength of 495 nm and two 515 nm longpass filters. For experiments looking at tryptophan fluorescence, an excitation wavelength of 295 nm and two 320 longpass filters. All stopped-flow reactions were performed in the standard ligation buffer, except for experiments monitored by tryptophan fluorescence, for which BSA was omitted from the buffer. The resulting traces were fit to a single exponential (Equation 2.3). All observed rates were plotted as a function of either [LIG1] or [DNA]. The resulting data plots were fit using linear regression to determine k_{on}.

Crystallization of mutant LIG1 complexes. A Δ262 LIG1 construct (amino acids 262-904) and an 18mer substrate were used for crystallization, as described in Chapter 3. Crystals of the LIG1•DNA•EDTA complex were grown by hanging drop method. The protein•DNA complex was formed with 20 mg/mL Δ262 LIG1 and the nicked 18mer oligo at a 1.5:1 DNA:LIG1 molar ratio in a solution consisting of 1 mM ATP, 150 mM NaCl, 20 mM Tris-HCl, pH 7.5, and 1 mM TCEP. 1μL of the complex solution was mixed with an equal volume of precipitant solution (100 mM MES, pH 6, 100 mM lithium acetate, 12-15% (w/v) polyethylene glycol 3350) at 20°C. The crystallization drops containing the mutant LIG1 protein were allowed to equilibrate for 30 minutes before micro-crystal seeds prepared from the WT crystals were streaked into the drops.

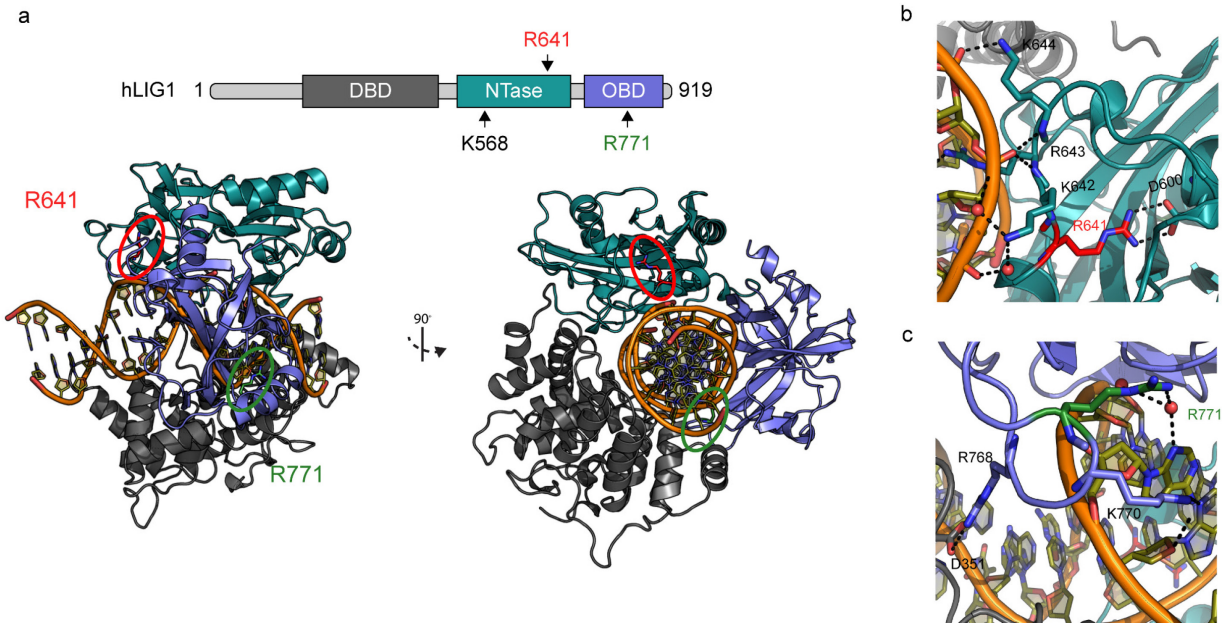


Figure 6.1. R641 and R771 are positioned within DNA-binding loops. (a) Schematic of the human LIG1 domain architecture (top). The approximate locations of the mutated residues (R641 and R771) and the active site lysine (K568) are indicated. The crystal structure of WT LIG1 showing the locations of the R641 (red) and R771 (dark red) residues is shown below. **(b)** Close-up view of the loop in the nucleotidyl transferase domain containing the R641 residue. A salt bridge between R641 and D600 helps to stabilize the loop, allowing for residues 642-644 to make direct and water-mediated contacts with the phosphodiester backbone of the DNA. **(c)** Close-up view of R771 intercalating into the minor groove of the DNA helix. R771 is involved in water-mediated interactions with the DNA, while K770 directly contacts the adjacent nucleotide. R768 forms a salt-bridge with D351 in the DBD.

R641 and R771 appear to be important for LIG•DNA interactions

Both R641 and R771 lie within regions of LIG1 that contact the DNA substrate (Figure 6.1). R641 forms a salt-bridge with D600, stabilizing a looped-region within the nucleotidyl transferase (NTase) domain (Figure 6.1b). A network of contacts between LIG1 and the DNA are present in the R641 loop, with adjacent residues K642, R643 and K644 all making either direct or water-mediated contacts with the DNA. R771 also lies within a loop region—this time in the OB-fold domain—where it engages in water-mediated contacts with an adenine in the major groove (Figure 6.1c). The adjacent K770 forms contacts with both the deoxyribose sugar

and base of a guanine nucleotide. The R771 loop is stabilized by a salt bridge between R768 in the loop and D351 in the DNA-binding domain (DBD).

Based on these observations, we expected that the patient-derived mutants would be defective in the DNA-dependent steps of ligation and may show decreased DNA-affinity. The R641L mutation would eliminate the salt-bridge that stabilizes the loop in the NTase domain, abrogating contacts between the region and the DNA. Similarly, substitution of a tryptophan for R771 will ablate water-mediated interactions between the residue and the DNA, possibly causing a distortion of the loop itself.

Substrate dependences of LIG1 mutants

To determine the biochemical consequences of the patient-derived mutations, we used site-directed mutagenesis to create recombinant Δ232 LIG1 mutants containing either the R641L or R771W mutation. Both mutations resulted in an overall decrease in catalytic activity with the canonical nicked DNA substrate (Table 6.1). The dependence of steady-state ligation on the concentration of ATP demonstrated only modest changes in $k_{cat}/K_{M,ATP}$, which is mostly due to the k_{cat} defect (Table 6.1, Figure S6.1). The absence of a significant defect in ATP utilization is unsurprising due to the proximity of the mutations near the DNA-binding surface and not near the putative ATP-binding residues.

Since we believed the mutants to be deficient in DNA-binding, we sought to characterize the behavior of steady-state ligation over a range of DNA substrate concentrations (Figure 6.2a). Indeed, in the presence of saturating concentrations of ATP and Mg^{2+}, the mutants have a ~2-to-3-fold increase in $K_{M,DNA}$ relative to wildtype. The increase in $K_{M,DNA}$ results in a ~5- and ~10-fold decrease in catalytic efficiency for R641L and R771W, respectively (Figure 6.2C, Table

6.1). Since both enzymes catalyze ligation at a slower rate relative to wildtype, this defect in catalytic efficiency is due in part to a global reduction in catalytic rate and not solely due to weakened $K_{M,DNA}$ values.

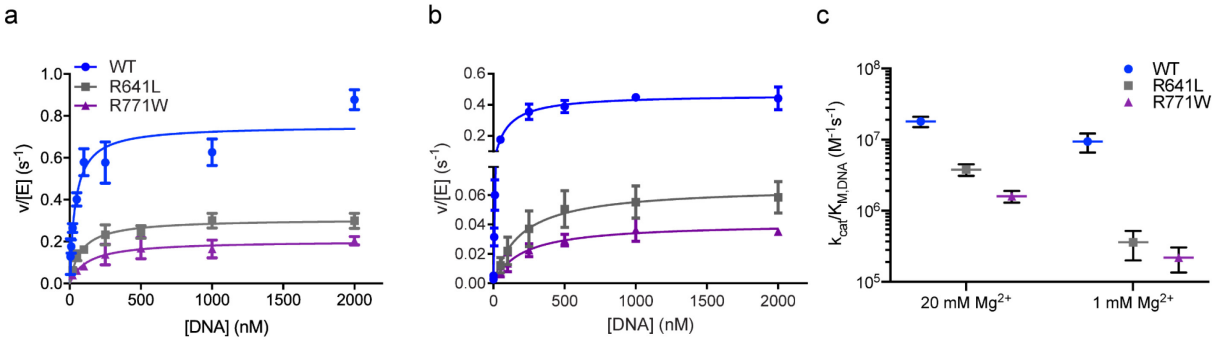


Figure 6.2. R641L and R771W demonstrate reduced catalytic efficiency for nicked DNA. Steady-state DNA dependences were carried out by mixing 0.1-10 nM LIG1 with increasing concentrations of the nicked DNA substrate, ranging from 0-2 µM DNA. Reactions were carried out in the presence of either 20 mM Mg^{2+}_{free} and 0.2 mM ATP (**a**) or 1 mM Mg^{2+}_{free} and 1 mM ATP (**b**). Initial rates are plotted as a function of [DNA] and fit using the Michaelis-Menten equation, from which k_{cat}/K_M values were derived for all three enzymes at either high or low $[Mg^{2+}]_{free}$ (**c**). Data are shown as the mean ± S.D. of at least three independent replications. Data are summarized in Table 6.1.

While we observed an obvious decrease in catalytic efficiency for the DNA substrate, this effect seemed somewhat modest considering the drastic consequences of the mutations *in vivo*[106]. To reconcile this, we repeated the substrate dependence in the presence of more physiological concentrations of both ATP and Mg^{2+}(Figure 6.2b). We define our physiological assay condition as containing 1 mM of both Mg^{2+}_{free} and ATP, based on literature estimates[158]. Under these physiological conditions, the catalytic efficiency of ligation by the wildtype LIG1 enzyme changes very little. However, for both patient-derived mutations, the ligation defects were much more pronounced than under saturating conditions (Figure 6.2c, Table 6.1). Compared to wildtype, we observed a ~30- and 40-fold decrease in catalytic efficiency for R641L and R771W, respectively, corresponding to a 10-fold change in the efficiency compared to the mutants under saturating conditions.

ATP Dependence

	WT	R641L	R771W
k_{cat} (s^{-1})	0.95 ± 0.01	0.27 ± 0.02	0.19 ± 0.02
K_M (μM)	8.0 ± 0.7	5.8 ± 0.3	4.0 ± 0.3
k_{cat} / K_M ($M^{-1}s^{-1}$)	1.2 ± 0.1 x 10^5	4.6 ± 0.3 x 10^4	4.8 ± 0.3 x 10^4

DNA Dependence

20 mM Mg^{2+}

	WT	R641L	R771W
k_{cat} (s^{-1})	0.75 ± 0.01	0.31 ± 0.02	0.20 ± 0.03
K_M (nM)	42 ± 6	82 ± 14	128 ± 10
k_{cat} / K_M ($M^{-1}s^{-1}$)	1.8 ± 0.3 x 10^7	3.8 ± 0.7 x 10^6	1.6 ± 0.3 x 10^6

1 mM Mg^{2+}

	WT	R641L	R771W
k_{cat} (s^{-1})	0.34 ± 0.11	0.066 ± 0.009	0.041 ± 0.002
K_M (nM)	44 ± 27	211 ± 67	220 ± 74
k_{cat} / K_M ($M^{-1}s^{-1}$)	9.4 ± 0.3 x 10^6	3.6 ± 1.6 x 10^5	2.2 ± 0.8 x 10^5

Table 6.1. Best-fit parameters for steady-state substrate dependences. Parameters were obtained from Michaelis-Menten fits to both the steady-state ATP (Figure S6.1) and DNA (Figure 6.2) dependences. Data are shown as the mean ± S.D. of at least three independent replicates.

R641L and R771W do not exhibit significant DNA-binding defects

To look at any changes in DNA affinity more directly, we measured the binding affinity using fluorescence anisotropy. To investigate this, we made use of the reporter substrate containing a fluorescein-tagged thymine base (TFAM) at the site of the nick in the 3' position (Chapter 2)[14]. Binding to the TFAM substrate was monitored by both fluorescence quenching and increases in anisotropy (Figure 6.3a). Utilizing this method provides a K_D of ~6 nM for wildtype LIG1. The measurement is nick-specific, as binding to the TFAM substrate lacking a phosphate group at the 5' end of the nick abolishes tight binding by wildtype LIG1 (Chapter 2). Despite steady-state $K_{M,DNA}$ values for R641L and R771W being significantly higher than that of wildtype, the measured K_D values were nearly identical between all three enzymes, suggesting no apparent defect in DNA-binding affinity (Figure 6.3b).

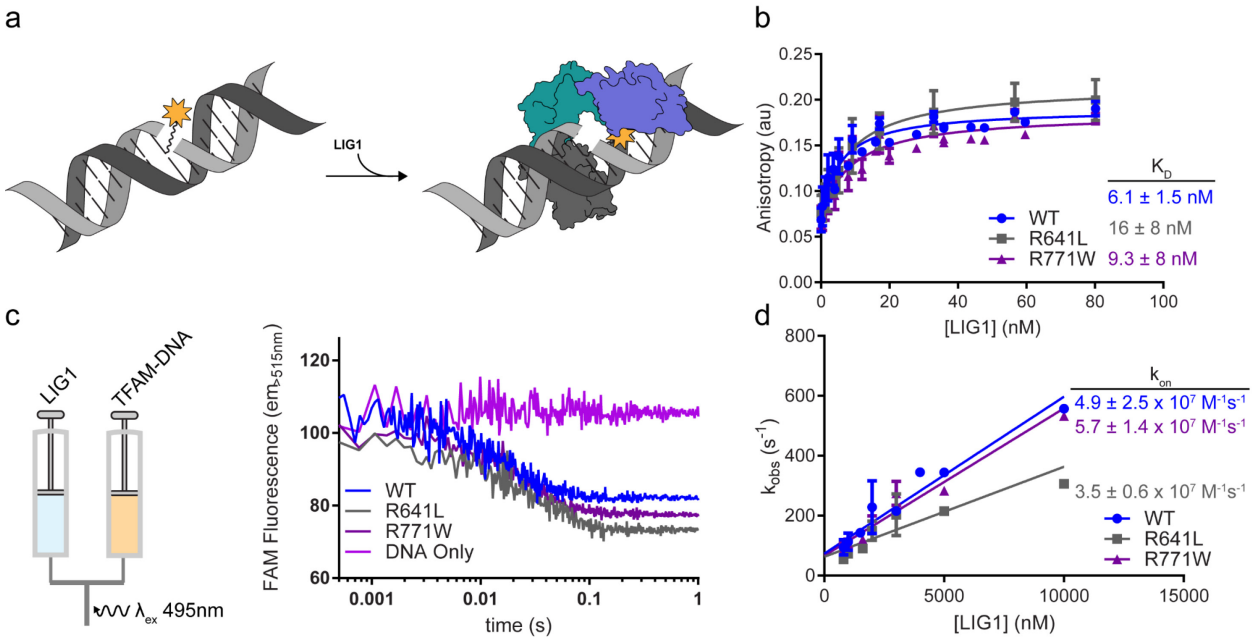


Figure 6.3. Clinical mutations of LIG1 have minimal impact on binding. (a) The reporter substrate used in binding studies makes use of a fluorescein conjugated to a deoxythymidine (TFAM) incorporated at the -1 position of the nicked DNA. The fluorescence signal is quenched upon nick-recognition and binding by LIG1. **(b)** Equilibrium-binding experiments using the TFAM substrate were performed by titration of increasing amounts of the LIG1 constructs against 1 or 2 nM of the TFAM substrate. Data were fit using a saturation binding curve and K_D values are reported as the mean ± SD of ≥ 3 replicates. **(c)** The schematic and representative traces from stopped-flow experiments performed to monitor real-time fluorescence changes resulting from LIG1 binding the TFAM-DNA reporter. Rates obtained from single-exponential fits of the data are plotted as a function of [LIG1] in **d**. The dependence of k_{obs} on [LIG1] was fit using linear regression and the slopes of these fits are reported as the observed on-rate (k_{on}) ± SD of ≥ 3 replicates.

To test for a possible effect on binding kinetics, we monitored binding to the TFAM substrate in real-time using stopped-flow spectroscopy (Figure 6.3c-d). Binding to the reporter substrate by wildtype LIG1 gives a time-dependent fluorescence quenching, which is best fit to a single exponential to obtain $k_{on,obs}$ (Figure 6.3c). The apparent on-rate increases linearly with increasing [LIG1], suggesting a bi-molecular association occurring at ~$5x10^7$ $M^{-1}s^{-1}$ (Figure 6.3d). Enzyme concentration dependences for both the R641L and R771W mutants behaved similarly, with $k_{on,obs}$ values close to $4x10^7$ $M^{-1}s^{-1}$ and $6x10^7$ $M^{-1}s^{-1}$, respectively. To ensure that the absence of an observable binding defect is real and not an artifact of the reporter substrate, we also monitored binding by performing stopped-flow experiments monitoring tryptophan fluorescence (Figure S6.2). LIG1 contains a conserved tryptophan residue in the active site that

we proposed would be sensitive to the enzyme accommodating the nick in the active site. Indeed, experiments with wildtype LIG1 and increasing unlabeled DNA substrate exhibited similar behavior to the TFAM experiments, giving a $k_{on,obs}$ of ~$4x10^7$ $M^{-1}s^{-1}$. Monitoring binding by tryptophan fluorescence for both R641L and R771W also mimicked the behavior seen in the TFAM experiments, suggesting that the mutations do not affect the binding rate (Figure S6.2b,c).

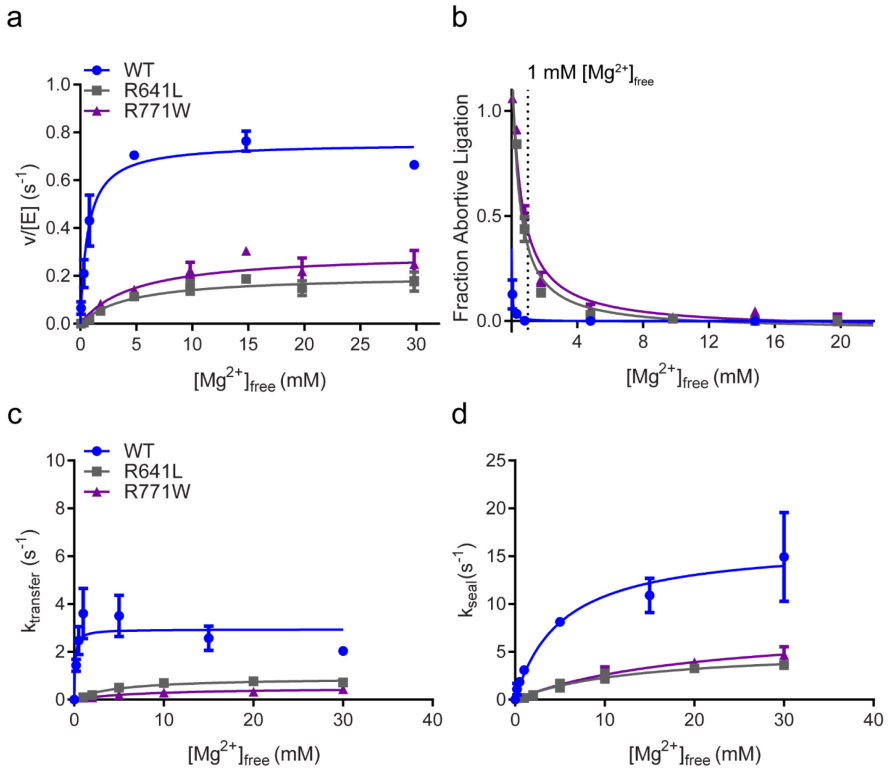


Figure 6.4. Clinical mutations of LIG1 weaken metal affinity. (**a**) Steady-state Mg^{2+} dependence demonstrates that R641L and R771W LIG1 mutants exhibit weakened Mg^{2+} affinity. Reactions contained 1-5 nM LIG1, 500 nM DNA and 0.2 mM ATP with varying concentrations of $MgCl_2$. Initial rates were fit to the Michaelis-Menten equation and best fit parameters are reported in Table 6.2 (**b**) The fraction of abortive ligation events observed in the steady-state reactions from **a**. The physiologically-relevant $[Mg^{2+}]_{free}$ (~1 mM) is marked by a dashed line. Both R641L and R771W exhibit enhanced rates of abortive ligation compared to WT at low $Mg^{2+}_{free.}$ (**c-d**) The rates of adenylyl transfer (**c**) and nick sealing (**d**) were determined from single-turnover experiments containing 800 nM LIG1, 80 nM DNA and increasing [$MgCl_2$]. The maximal rates and K_{Mg} values for both $k_{transfer}$ and k_{seal} were determined by hyperbolic fits to the single-turnover data (Table 6.2)**.** All values are reported as the mean ± S.D. of N ≥ 3 replicates.

R641L and R771W exhibit Mg^{2+}-dependent effects

Despite the absence of an observable defect in DNA-binding, we saw a large reduction in catalytic efficiency in the presence of low [Mg^{2+}] (Figure 6.2c) and posited that there may be a magnesium-dependent defect in the mutants. To test this, we performed a steady-state magnesium dependence in the presence of saturating amounts of ATP and magnesium (Figure 6.4a). Since neither of the mutations lie near the active site or putative metal-binding sites, we did not expect to see a large change in the apparent Mg^{2+} affinity compared to wildtype LIG1. Surprisingly, the mutants exhibited significant decreases in their apparent magnesium affinities compared to wildtype, with both R641L and R771W displaying a 5-fold increase in K_{Mg2+} (Figure 6.3a, Table 6.2). Moreover, at Mg^{2+}_{free} concentrations ≤ 5 mM, there was a noticeable increase in the frequency of abortive ligation events, where the adenylylated intermediate is released before nick-sealing can occur (Figure 6.3b). At physiological concentrations of Mg^{2+}_{free} (1 mM), the ratio of abortive ligation to ligation events is roughly 1:1, with abortive ligation events making up roughly half of all observed catalysis by the two mutants. At the lowest $MgCl_2$ concentration tested—about 50 μM Mg^{2+}_{free}—all catalysis by R641L and R771W resulted in abortive ligation, while only ~10% of catalytic events by wildtype LIG1 ended with release of the adenylylated intermediate, indicating that the mutants undergo abnormally high amounts of abortive ligation (Figure 6.3b).

The surprising magnesium-dependent defects led us to believe that the two chemical steps may be impacted by the two mutations. To characterize the individual chemical rates of adenylyl transfer and nick sealing, we performed single-turnover ligation assays using a rapid-quench apparatus over a range of magnesium concentrations (Figure 6.3c-d). Wildtype LIG1 binds magnesium tightly for adenylyl transfer and has a much weaker association—about 6-

fold—for the catalytic magnesium involved in nick sealing (Figure 6.3c-d, Table S6.2). K_{Mg2+} values for the adenylyl transfer step are also tighter than for nick sealing in the mutants, but this affinity is ~40-50-fold weaker relative to wildtype. Meanwhile, K_{Mg2+} for nick sealing is only ~2-3-fold weaker in the mutants compared to wildtype, while the maximal rates of both chemical steps are 3-6-fold slower. These results indicate that the mutations may have a substantial impact on the initial binding and positioning of the catalytic magnesium ion required, causing observable defects in the DNA-dependent steps of ligation.

Mg^{2+} Dependence

Steady-state

	WT	R641L	R771W
k_{cat} (s^{-1})	0.92 ± 0.22	0.31 ± 0.08	0.22 ± 0.06
K_{Mg2+} (mM)	0.86 ± 0.03	5.3 ± 2.2	5.2 ± 3.0
k_{cat} / K_{Mg2+} ($M^{-1}s^{-1}$)	1.1 ± 0.2 x 10^3	0.63 ± 0.12 x 10^2	0.47 ± 0.10 x 10^2

Single turnover

	WT	R641L	R771W
$k_{transfer}$ (s^{-1})	2.9 ± 0.5	0.8 ± 0.1	0.46 ± 0.07
k_{seal} (s^{-1})	18 ± 5.5	5.3 ± 0.3	7.0 ± 1.4
$K_{Mg,transfer}$ (mM)	0.10 ± 0.03	3.9 ± 0.4	6.1 ± 1.0
$K_{Mg,seal}$ (mM)	6.7 ± 3.7	13 ± 6	17 ± 8

Table 6.2. Best-fit parameters for steady-state and single-turnover magnesium dependences. Steady-state parameters were obtained from magnesium dependences (Figure 6.4a) fit by the Michaelis-Menten equation. Single-turnover parameters were obtained using curve-fitting to a two-step irreversible model in Berkeley-Madonna. Maximal rates and K_{Mg} values were derived from hyperbolic curve fits of the data as a function of magnesium concentration (Figure 6.4c,d). Data are shown as the mean ± S.D. of at least three independent replicates.

DNA-binding interactions are perturbed in R641L and R771W LIG1 crystal structures

Due to the mutants having perturbed $K_{M,DNA}$ and K_{Mg2+} values, we hypothesized that these mutations may be affecting how the DNA helix is oriented for chemistry, especially since the mutated residues reside in DNA-binding regions. To address this, we obtained crystal structures of the two mutant LIG1 enzymes in complex with DNA. For crystallization, a Δ262 LIG1 construct was used to minimize disordered regions in LIG1 observed in the published LIG1 crystal structure. R641L and R771W Δ262 LIG1•DNA complexes were solved in the

presence of EDTA, limiting the scope of any observations gained from the structures due to the absence of the catalytic magnesium ion.

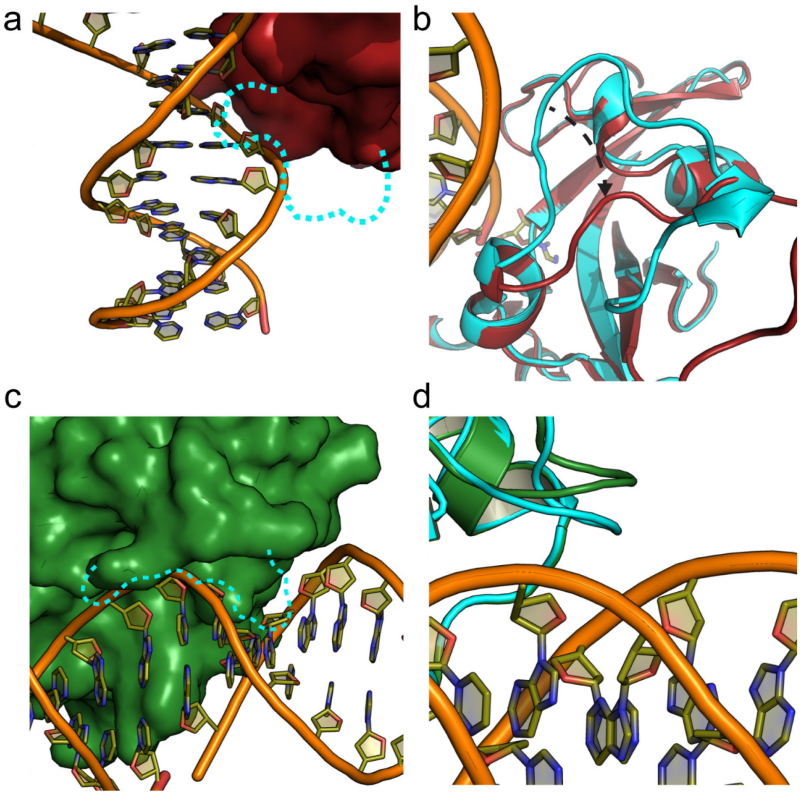


Figure 6.5. R641- and R711-mediated DNA contacts are critical for maintaining helical architecture. Surface and cartoon representations of the R641L (**a,b**) and R771W (**c,d**) crystal structures highlighting the absence of LIG1•DNA contacts observed in the WT structure (cyan). The DNA double helix from the WT LIG1 crystal structure is shown in all panels.

The structures of both the R641L and the R771W mutants reveal several distortions in both the protein and DNA helix. As predicted, both mutants lack extensive interactions seen between wildtype LIG1 and the DNA substrate (Figure 6.5). In the R641L structure, elimination of the salt-bridge interaction between R641 and D600 in the NTase domain loop region causes this region to buckle inwards (Figure 6.5a,b). The entire loop region is shifted ~3 Å inwards towards the interior of the NTase domain, where it can no longer contact the DNA helix. A similar effect is noted in the R771W structure, where a region of the OB-fold is rotated away from the DNA helix, relative to its position in the wildtype structure (Figure 6.5c,d). The R771W mutation causes a nearly 70º rotation of this region away from the DNA helix. Additionally, the

R771W mutation pushes R768 out of the range to form a salt-bridge with D351 in the DBD (Figure 6.6d), which may result in the inability of LIG1 to stably encircle the DNA substrate.

R641L and R771W do not cause apparent active site perturbations

Based on our biochemical results, we posited that the active site of LIG1 is somehow influenced by these mutations. Due to the apparent involvement of the R641 and R771 residues in contacting the DNA helix, we hypothesized that the nicked DNA may not be properly aligned in the active site due to the mutations. To our surprise, there are no apparent rearrangements of any of the known catalytic residues in the active site (Figure 6.6a). Additionally, the positioning of both the 5' and 3' ends of the nick are unchanged in the mutant structures.

In both mutants, the phosphodiester backbone near the mutation kinks outwards, partially filling the void created in the absence of the R641 and R771 loops (Figure 6.5b,d). Distortions in the phosphodiester backbone of the DNA are also observed upstream of the nick, near a known metal-binding site (Figure 6.6b-e). We have previously shown residues E346 and E592 to coordinate a magnesium ion (Chapter 3). Although much remains to be known about this site—known as the HiFi metal site—we believe this metal to be important for proper positioning of the nick for catalysis; mutagenesis of the HiFi site results in deficient substrate discrimination and a distorted phosphodiester backbone near the site. Based on this, we believe that the R641L and R771W mutations can perturb the HiFi metal site, which may cause changes in how LIG1 accommodates a nicked DNA substrate after binding. Although no DNA-binding defect was observed, it is reasonable to believe that these mutants are deficient in a magnesium-dependent step that precedes adenylyl-transfer. Nonetheless, the extensive phosphodiester backbone distortions observed in the two mutants imply that mutations cause improper positioning of the

DNA substrate, possibly explaining the high abortive ligation burden observed in biochemical assays.

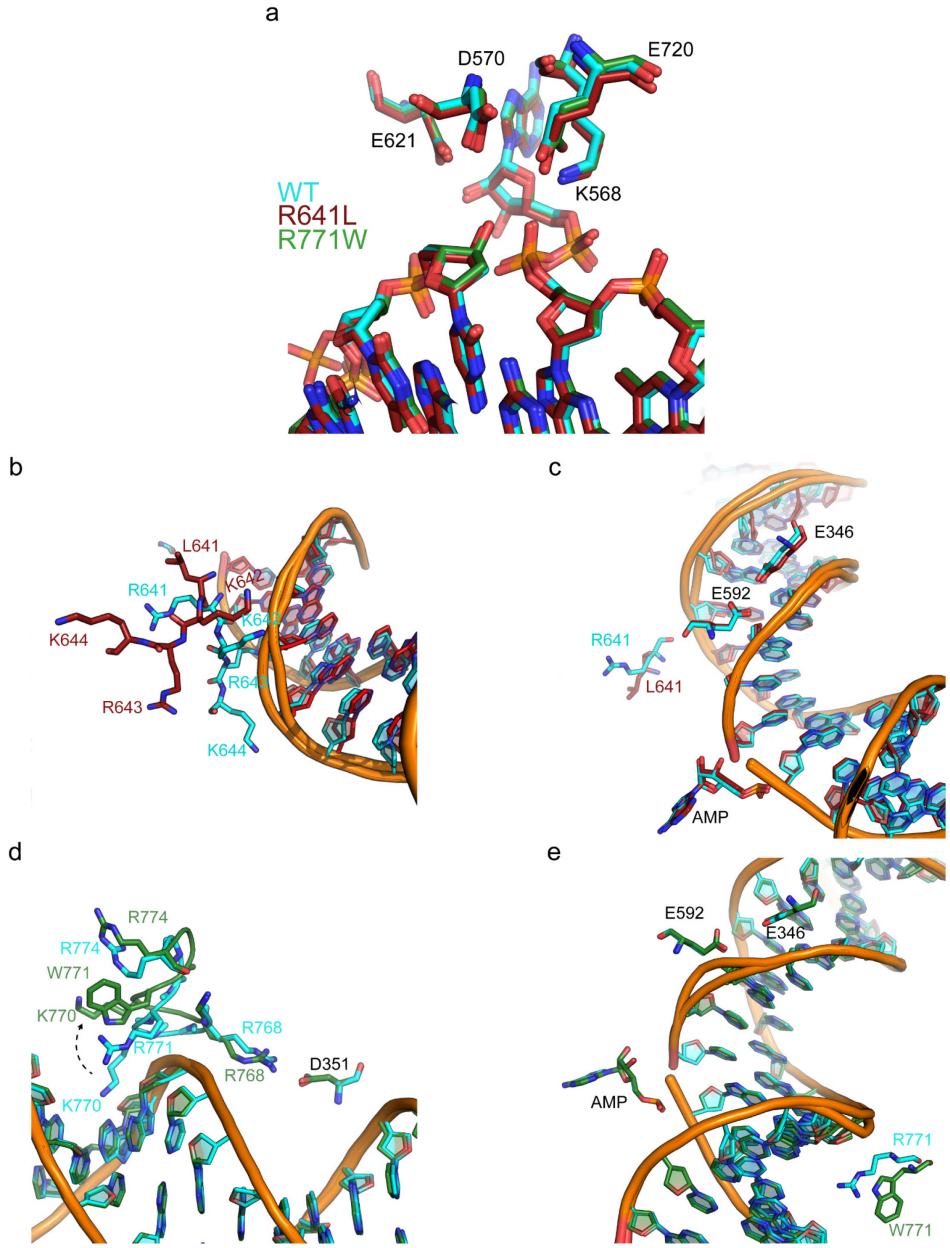


Figure 6.6. R641L and R771W cause extensive structural changes outside of the active site. (**a**) Stick representation of the active site metal ligands in LIG1 (D570, E621 and E720), showing little-to-no change as a result of the R641L or R771W mutations. Additionally, no apparent changes are observed in the positioning of the active site lysine (K568) or the 5'-adenylylated nick. (**b**) Stick representation of the individual residues in the R641 loop in both the WT (cyan) and R641L (red) crystal structures. Contacts to the DNA strand by residues K642, R643 and K644 are disrupted in the R641L structure. The distorted loop causes numerous kinks in the phosphodiester backbone, including one near the HiFi metal ligands, E346 and E592 (**c**). Stick representation of the loop in the OB-Fold domain in both the WT (cyan) and R771W structures (**d**). The R771W mutation causes a rotation of residues 770 and 771 away from the DNA helix. Additionally, R768 is no longer within salt-bridging distance with D531 in the R771W structure. Like the R641L structure, several kinks in the phosphodiester backbone are observed, including one near the E346 and E592 residues (**e**).

Discussion

Much remains to be known about the link between mutations in DNA ligase I and the onset of an immunological disorder. Here, we have addressed some of the key biochemical consequences of two of the most severe LIG1 mutations identified in immunodeficient patients. Due to the positioning of the two mutated residues—R641 and R771—we believed we would uncover a severe DNA-binding defect in the recombinant mutants of LIG1. While we did observe a decrease in catalytic efficiency for the mutants, we did not see any indication of a decrease in binding affinity. By employing our TFAM reporter substrate, we measured $K_{D,DNA}$ values for the R641L and R771W mutants, which are not significantly different than the $K_{D,DNA}$ measured for wildtype LIG1. Additionally, using the TFAM reporter in conjunction with stopped-flow spectroscopy allowed for measurements of k_{on} , which is also similar between wildtype and mutant LIG1 enzymes. This result was verified by monitoring binding using tryptophan fluorescence and an unlabeled DNA substrate.

Unexpectedly, we found a significant decrease in magnesium affinity in the R641L and R771W mutants. Under steady-state conditions, there is a roughly 5-fold decrease in K_{Mg2+} for both mutants compared to wildtype. Along with decreased affinity for magnesium, the mutants appeared to affect the ability of LIG1 to stay committed to catalysis. While catalysis by R641L and R771W LIG1 enzymes in the presence of 20 mM $MgCl_2$ leads to almost complete sealing of nicked DNA, when the concentration of Mg^{2+}_{free} is dropped to near-physiological concentrations, most ligation events are abandoned after adenylyl-transfer.

The increase in abortive ligation could have severe consequences in the cell, where a large buildup of 5’-adenylylated nicks in the genome could overwhelm the DNA repair machinery. While the base excision repair (BER) pathway and aprataxin (APTX) are both able to

remove the 5'-adenylyl group from abortive ligation products, the large burden created by these LIG1 mutants may exceed the ability of these pathways to efficiently recognize and repair the adenylylated nicks. It remains to be known if the abortive products resulting from these LIG1 mutants persist *in vivo*, although initial studies into the consequences of the R771W mutation report delayed Okazaki fragment maturation in the mutated cells[105]. It is worth noting that mutations in APTX or in FEN1—which is essential for long-patch BER—were not observed in any of the recently identified patients, indicating that aberrant LIG1 activity alone is the cause of the disease onset.

Crystal structures of the R641L and R771W LIG1 mutants did not show any significant changes in the enzyme active site due to the mutagenesis. The structures presented here were obtained in the absence of Mg^{2+}, which, due to the Mg^{2+}-dependence of the abortive ligation defect, may limit the conclusions that can be drawn. Nonetheless, we did note several changes outside of the active site. The R641 loop region collapses upon mutation of the residue to leucine, which eliminates the salt-bridging interaction present in the wildtype structure. The R771W mutation causes a complete rotation of the R771 loop region away from the DNA. This results in a loss of interactions between the enzyme and the DNA substrate. Additionally, the R771W mutation seems to affect a salt-bridging interaction between the OB-fold and DNA-binding domains, which could affect the ability of the enzyme to stably encircle the substrate.

Although we did not observe a significant effect on the binding affinity, the crystal structures illustrate that the mutants appear to bind to a distorted DNA helix. The phosphodiester backbones near the regions of the R641L and R771W mutations are unsurprisingly extend outward relative to their position in the wildtype structure, presumably to fill the space vacated by the mutated loop regions. The phosphodiester backbone distortions, however, are not limited

to the backbones adjacent to the mutation. In both structures, a region of the DNA adjacent to a recently-characterized metal binding site in LIG1 is also distorted. Our initial characterization of this metal binding site demonstrated that it has some importance in imposing rigidity in the 3' end of the nick to disallow ligation of mismatched or damaged substrates (Chapter 3). While the exact impact of these mutations on the function of this so-called HiFi site is unclear, perturbing the phosphodiester backbone in this region could lead to destabilization of the LIG1•DNA complex. Along with the collapsed loop in the R641L mutation and the absence of a salt-bridging interaction between the OB-fold and DBD, a dysfunctional HiFi site may be the cause of these mutants being more prone to abortive ligation.

Our understanding of the defects in the clinical mutants necessitates a more thorough understanding of the binding mechanism employed by wildtype DNA ligase I. Using fluorescence signal from either fluorescein or internal tryptophan residues, we have been able to gain some understanding of the Mg^{2+}-dependent steps subsequent to binding but preceding chemistry (Chapter 2). Using these techniques could potentially further our understanding of the clinical mutations. Pilot experiments in the presence of saturating magnesium show that is a significant lag between the initial binding phase and the phase corresponding to chemistry (Figure S6.3a). We also observe a similar change in the behavior of the stopped-flow traces when monitoring catalysis of our canonical substrate by the mutants using internal tryptophan fluorescence, although the effect is less severe (Figure S6.3b). Nonetheless, these initial experiments show potential for helping to further illustrate the biochemical defects of these mutants.

Chapter 6 Appendix

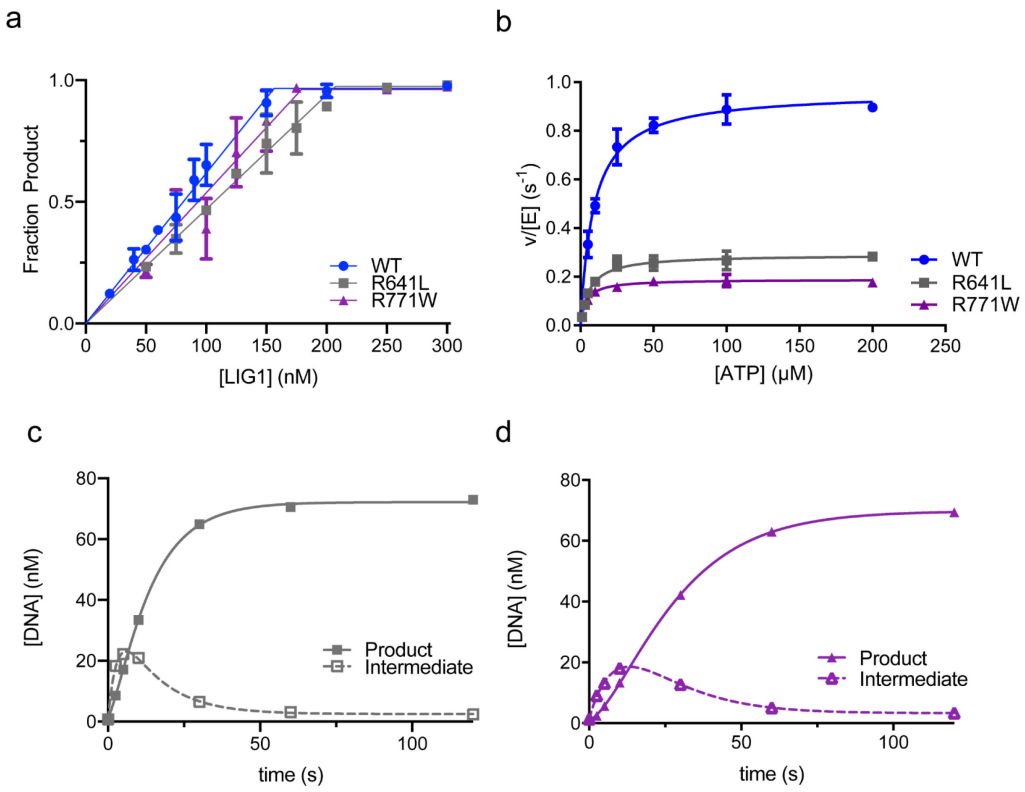


Figure S6.1. Initial biochemical characterization of the R641L and R771W mutants. (**a**) The concentration of active protein was estimated from active site titrations, where increasing amounts of LIG1 were titrated against a fixed amount of DNA. (**b**) Steady-state ATP dependence in the presence of saturating Mg2+ (20 mM) and DNA (500 nM). (**c-d**) Representative single-turnover plots for R641L (**c**) and R771W (**d**).

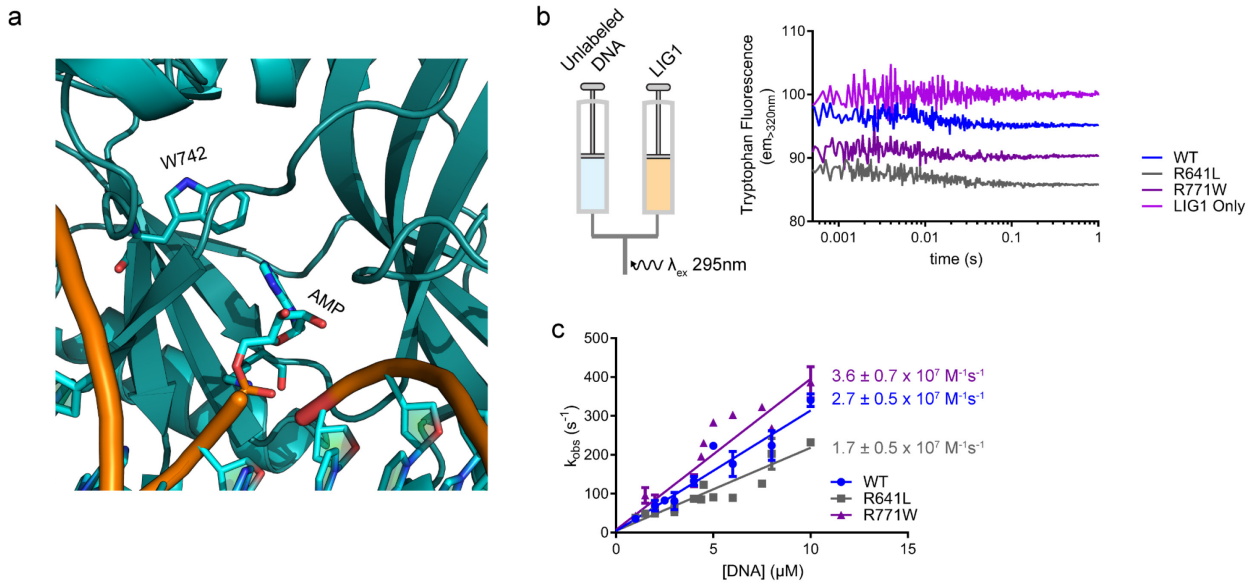


Figure S6.2. Tryptophan fluorescence as a binding reporter. (**a**) Crystal structure showing location of putative reporter tryptophan (W742). (**b**) The schematic and representative traces from stopped-flow experiments performed to monitor real-time fluorescence changes resulting from LIG1 binding the TFAM-DNA reporter. Rates obtained from single-exponential fits of the data are plotted as a function of [LIG1] in **c**. The dependence of k_{obs} on [LIG1] was fit using linear regression and the slopes of these fits are reported as the observed on-rate (k_{on}) ± SD of ≥ 3 replicates.

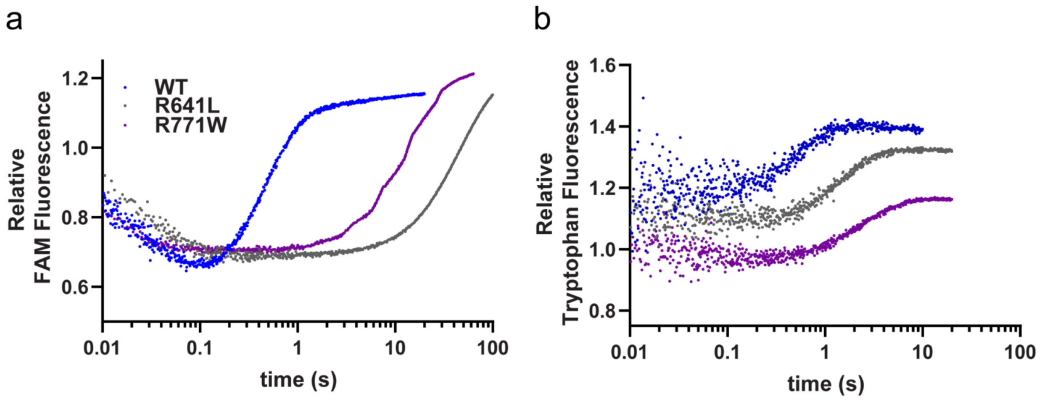


Figure S6.3. Pilot experiments using real-time fluorescence to monitor catalysis. (**a**) Stopped-flow traces resulting from the rapid mixing of 500 nM LIG1 with 20 nM TFAM reporter substrate in the presence of 20 mM Mg^{2+}. As noted in Chapter 2, the large fluorescence increase is largely due to the chemical steps of ligation. Here we note a lag between the initial binding and catalysis with the two patient-derived mutations. (**b**) A similar experiment to **a**, except the reaction is monitored using tryptophan fluorescence. For robust signal, DNA is held in excess at 600 nM, while 200 nM of the LIG1 enzymes are added. ATP is omitted in both **a** and **b** to prevent multiple turnovers.

Chapter 7
Conclusions and Future Directions

The work presented here represents a comprehensive study of the binding and fidelity mechanisms underlying DNA ligation by human DNA ligase 1. Due to the ubiquity of DNA ligases in DNA replication and repair pathways, we sought to better understand how a DNA ligase is able to capture its substrate and faithfully restore phosphodiester bonds. DNA ligases are known to contribute to the overall fidelity of replication and repair, although the mechanisms behind high-fidelity ligation were previously unclear. We uncovered two distinct mechanisms for discriminating against unfavorable nucleotide modifications at both the downstream and upstream ends of a nick. Since magnesium ions are vital to DNA ligation, we also investigated the interactions between LIG1 and magnesium, illustrating a complex network of interactions between magnesium and both the protein and its substrates. Finally, we demonstrated how perturbations of this intricate network caused by patient-derived mutations leads to enzymatic dysfunction, connecting the molecular mechanism of LIG1 to human health.

Binding mechanism of LIG1

In Chapter 2, we describe the first investigation into the mechanism of DNA-binding by a human DNA ligase. In order to study the binding mechanism, we first characterized a binding assay to monitor binding interactions between LIG1 and its nicked DNA substrate. A previous study into the T4 DNA ligase binding mechanism made use of an internally-labeled DNA

substrate[14]. Using the internal fluorescein label, the authors were able to monitor binding and catalysis by T4 DNA ligase using stopped-flow spectroscopy. With this in mind, we used a similarly labeled substrate for investigation of the LIG1 binding mechanism. Although the internal label did impact the kinetics of ligation, as compared to ligation of our canonical substrate, the reporter substrate enabled direct measurements of the initial interactions between LIG1 and the nick site.

Aside from direct measurements of the DNA-binding affinity of LIG1, we uncovered evidence of an unknown step following the initial association but preceding the chemical ligation steps. To rule out possible artifacts caused by the internal label, we performed stopped-flow experiments using intrinsic tryptophan fluorescence. By using tryptophan fluorescence, were able to repeat our binding assays with an unmodified substrate. We found evidence suggesting a two-step remodeling event preceding catalysis. Preliminary results further define these two steps as conformational changes of the upstream and downstream ends of the nick by LIG1.

Molecular mechanisms underlying high-fidelity ligation

In Chapters 3 and 4, we dissected the underlying mechanisms for maintaining ligation fidelity at both ends of the nick. Recent crystal structures of LIG1 revealed a novel magnesium-binding site in LIG1 in close proximity to the upstream DNA strand. Chapter 3 focuses on characterization of this metal site through mutagenesis of the conserved metal ligands observed in the crystal structures. The mutant, referred to as $LIG1^{EE/AA}$, exhibits increased ligation efficiency with our canonical DNA substrate. Additionally, the mutant is able to efficiently ligate oxidatively-damaged upstream ends, containing an 8-oxoguanine base paired with either cytidine or adenine. The metal site mutation appears to affect the ability of LIG1 to perform abortive

ligation, which results in release of the adenylylated ligation intermediate and is considered a mechanism for maintaining ligation fidelity. We demonstrate that this effect is not due to a change in the chemical steps of ligation, as $LIG^{EE/AA}$ catalyzes adenylyl transfer and nick sealing of damaged and undamaged DNA substrates with a similar rate to wildtype LIG1. To connect these findings to the larger cellular context, we performed ligation assays in the presence of the deadenylase aprataxin (APTX). The addition of APTX to steady-state reactions suppresses the activity of wildtype LIG1 on the oxidatively damaged substrates. However, APTX is ineffective at suppressing ligation by $LIG1^{EE/AA}$, indicating that APTX is unable to compete with the mutant for binding to the adenylylated intermediate.

Finally, we observed that inclusion of ATP into single-turnover reactions causes a noticeable decrease in the rate of nick sealing for wildtype LIG1-dependent ligation of the damaged substrates. We posited that this ATP-dependent slowing of ligation may be important for maintaining fidelity. We therefore proposed that, when presented with a damaged substrate, LIG1 is able to catalyze adenylyl transfer, after which free ATP will cause the enzyme to briefly pause and partially dissociate from the DNA. In the absence of excess substrate or another protein, LIG1 can then re-associate with the ligation intermediate to finish nick sealing. In the presence of another repair protein, however, we hypothesize that the adenylylated ligation intermediate gets sequestered away from LIG1 during this partial dissociation. Using a catalytically-deficient form of APTX that retains binding activity, we demonstrated that the adenylylated ligation intermediate is sequestered away from wildtype LIG1 in single-turnover reactions containing ATP and the damaged DNA substrates, while the intermediate is only partially sequestered from the HiFi site mutant. Our results suggest that the HiFi site is critical for allowing LIG1 to disengage the ligation intermediate, where it is quickly bound and

processed by APTX to remove the 5'-AMP. In the cell, this would allow other DNA repair proteins to remove either the damaged or mismatched end to maintain genomic integrity.

Chapter 4 focuses on a different mechanism of substrate discrimination at the downstream end of the nick. Since DNA ligases are known to discriminate against ligation of downstream ribonucleotides[15,122], we sought to define the structural determinants for ribonucleotide discrimination. We identified a highly-conserved phenylalanine residue near the downstream end of the nick as a possible candidate for imposing ribonucleotide discrimination due to its proximity to the deoxyribose of the 5' nucleotide. In order to determine the role of the phenylalanine residue, we mutated the residue to both alanine and leucine. Initial characterization of both mutants demonstrated a significant decrease in catalytic efficiency. The mutations primarily affect k_{cat}, as we only observed small changes in the K_M values for both ATP and DNA.

Under both steady-state and single-turnover conditions, wildtype LIG1 performs ligation of a downstream ribonucleotide at a significantly decreased rate. During ligation of the downstream ribonucleotide by F872L, we observed a large increase in the abortive ligation burden, resulting in almost no product formation. In contrast to both wildtype and F872L LIG1, F872A performs efficient ligation and is unable to discriminate against the ribonucleotide substrate. This data suggests that the conserved phenylalanine serves as a gate for ribonucleotides, preventing ligation of DNA-RNA hybrids.

Characterization of the LIG1 metal ligands

Although magnesium has long been known as an essential cofactor for DNA ligation, the metal ligands in the active site of a DNA ligase have not previously been characterized.

Additionally, the number of metal ions used for the DNA-dependent steps of ligation remains unknown, although the original LIG1 structure proposed a two-metal ligation mechanism[15]. More recent crystal structures of LIG1 show a lone metal ion coordinated by a number of conserved amino acids in the enzyme active site. To address these observations, we performed mutagenesis of the three conserved amino acids to examine the effects of perturbing the interactions between LIG1 and the active site metal. Two of the mutants, E621Q and E720Q, exhibited severe defects in DNA ligation and were not characterized. The three other mutations, E621D, D570N and E720D, also exhibited defective ligation, but retained enough ligation activity to biochemically characterize.

As expected, all three of the mutants displayed decreased magnesium affinity in both steady-state and single-turnover assays. Additionally, the three mutants all performed ligation with a significantly decreased rate. Of the three mutants, E621D caused the most severe impact on the rate of nick sealing, suggesting that it plays an important role during the final step of ligation. While both D570N and E720D did not seem to have significant impacts on the affinity of LIG1 for ATP, the E621D mutation results in dramatic decreases in $K_{M,ATP}$. In support of this finding, structures of an ATP-dependent RNA ligase indicate that the E621 residue likely contacts the 2'-OH of ATP during the enzyme adenylylation step[149]. The E621 residue is also observed contacting the 2'-OH before and after adenylyl transfer in the LIG1 structures (Chapter 3), possibly explaining why the E621D mutation had a larger impact on nick sealing than adenylyl transfer, as the residue may help in stabilizing the AMP leaving group for nick sealing. Finally, none of the tested mutations appear to affect DNA-binding, supporting the finding from Chapter 2 that magnesium binding is not a requirement for the initial binding event.

Biochemical characterization of LIG1 Syndrome mutations

Chapter 6 describes our biochemical characterization of mutations in LIG1 derived from patients suffering from immunological deficiencies. The exact connection between LIG1 dysfunction and the onset of immunodeficiencies has long been a lingering question in the field[102,103,156]. Since the biochemical consequences of patient-derived mutations in LIG1 are unknown, we created recombinant versions of two of the identified LIG1 mutants, R641L and R771W. Due to both of the residues being located near the DNA-binding interface of LIG1, we expected to see defects in the DNA-binding ability of the two mutants. In steady-state ligation assays, we observe a modest decrease in catalytic efficiency for the two mutants with our canonical ligation substrate. However, direct binding assays did not reveal a large effect on the overall DNA-binding affinity of the two mutants.

Under conditions of low magnesium, we observed that the R641L and R771W mutations undergo a large abortive ligation burden. To investigate this further, we performed both steady-state and single-turnover magnesium dependences to probe any magnesium-dependent effects of the mutations. Since the mutations are not near the enzyme active site, we were surprised to observe significant decreases in magnesium affinity for both the R641L and R771W mutations. Preliminary crystal structures of the two mutants revealed that the R641L and R771W mutations cause distortions in the DNA helix, most notably near the HiFi metal site described in Chapter 3. In addition, preliminary stopped-flow experiments demonstrated delayed kinetics of the conformational changes described in Chapter 2, indicating that these mutations may be affecting the ability of LIG1 to properly align the DNA substrate for catalysis.

Future Directions

Mechanism of substrate binding by LIG1

Further study of the binding mechanism is still needed to fully define the remodeling events. To address the upstream remodeling event, we performed our stopped-flow assay using a substrate with a ribonucleotide upstream strand, which resulted in a twofold increase in the rate of one of the two putative remodeling steps. While this experiment suggests one of the two steps is affected by modifications to the upstream strand, we have yet to demonstrate that the other step can be perturbed with changes to the downstream strand. To directly confirm the existence of a downstream remodeling event, a suitable modification to the downstream end would need to be tested. If a downstream reporter is well-tolerated, this could help unify observations made with the TFAM-44mer substrate and those made using intrinsic tryptophan fluorescence.

Although our study is primarily focused on the DNA-binding mechanism of LIG1, we also demonstrated that tryptophan fluorescence can be used to monitor binding of ATP by LIG1. A similar method has been used to define the ATP-binding mechanism for T4 RNA and DNA ligases[121] and would provide the first direct measurements of LIG1 self-adenylylation. Additionally, there is a large conformational change that must occur between enzyme adenylylation and DNA-binding[13,15,149,159,160]. A thorough understanding of the steps before and after this conformational change would help us better understand how this conformational change is triggered. Although it is outside of the scope of this study, determination of the LIG1 structure in an ATP-bound state would also be pivotal to dissecting the conformational transition between adenylylation and DNA-binding.

Substrate specificity of LIG1

The fidelity mechanisms explored in Chapters 3 and 4 were tested using substrates that represent only a small subset of the possible damaged or mismatched ends. To look at fidelity mechanisms in more depth, it would be advantageous to know how broadly these fidelity mechanisms apply. Studies profiling the extent of ligation fidelity have been performed with other DNA ligases, where all possible base combinations were tested at either side of the nick[119,120]. A similar approach could be performed with LIG1, defining how different sequence contexts at and around the nick may impact ligation. We also do not know how relevant these apparent fidelity mechanisms are to ligation in the cell. Repeating the mutagenesis studies in living cells would help to confirm that these fidelity mechanisms are contributing to cell survival. Since the amino acid residues underlying the mechanisms described in Chapters 3 and 4 are conserved in the LIG1 homolog in yeast, Cdc9, the effects of the mutations could be assessed using yeast as a model system.

In addition to F872, another conserved phenylalanine residue is situated near the upstream end, where it is positioned near the sugar moiety of the 3' end. Since upstream ribonucleotides are known to be well-tolerated by DNA ligases, F635 likely plays a different role in than the F872 residue. Since the F635 residue is positioned near the same side of the nick as the HiFi metal site, it is possible that the phenylalanine residue works in concert with the HiFi site to induce and stabilize the upstream A-form-like helix. Cooperativity between the two sites could involve an initial "pushing" of the upstream end by F635, resulting in a conformational change in the upstream strand that is stabilized by interactions between Mg^{HiFi} and the phosphodiester backbone. Mutations of F635 have been characterized in the Chlorella virus

DNA ligase, demonstrating that mutations of the residue ablate the ability of ChVLig to complement a ligase-deficiency in yeast cells, possibly due to significant defects in the DNA-dependent ligation steps[161]. This underscores the importance of the F635 residue, but any possible role the residue plays in substrate discrimination remains untested.

Impact of DNA ligase research

The specific aims of my dissertation research focused on dissecting the molecular mechanisms employed by a human DNA ligase to maintain genomic integrity. However, DNA ligation is only a single step in DNA replication and repair in the cell. Considering how my research fits into the broader context of these pathways is an essential step towards the goal of understanding genomic maintenance. My work highlights unique features of LIG1 that seem to be key in maintaining fidelity during the last step of DNA replication and repair. By tweaking these features in cellular LIG1, future research should be able to elucidate how mis-ligation can impact the cell at-large. Additionally, LIG1 is known to associate with the proliferating cell nuclear antigen (PCNA) in the cell, suggesting that it is important to understand how tethering LIG1 to PCNA impacts the fidelity mechanisms highlighted in my work. Since the three human DNA ligases share a conserved catalytic core, it will also be essential going forward to know if my findings hold true for the other two human DNA ligases. Past research has illustrated the divergent cellular functions of DNA ligases I, III and IV, suggesting that the three DNA ligases may also have different degrees of substrate tolerance. While expanding the scope of this work is a substantial undertaking, it has the potential to alter our current understanding of genomic fidelity and how it is maintained by networks of enzymes in the cell.

Bibliography

1 Franklin, R. E. & Gosling, R. G. Molecular configuration in sodium thymonucleate. *Nature* **171**, 740-741 (1953).

2 Klug, A. Rosalind Franklin and the discovery of the structure of DNA. *Nature* **219**, 808-810 passim (1968).

3 Lindahl, T. Instability and decay of the primary structure of DNA. *Nature* **362**, 709-715, doi:10.1038/362709a0 (1993).

4 Gellert, M. Formation of covalent circles of lambda DNA by E. coli extracts. *Proceedings of the National Academy of Sciences of the United States of America* **57**, 148-155, doi:10.1073/pnas.57.1.148 (1967).

5 Little, J. W., Zimmerman, S. B., Oshinsky, C. K. & Gellert, M. Enzymatic joining of DNA strands, II. An enzyme-adenylate intermediate in the dpn-dependent DNA ligase reaction. *Proceedings of the National Academy of Sciences of the United States of America* **58**, 2004-2011, doi:10.1073/pnas.58.5.2004 (1967).

6 Zimmerman, S. B., Little, J. W., Oshinsky, C. K. & Gellert, M. Enzymatic joining of DNA strands: a novel reaction of diphosphopyridine nucleotide. *Proceedings of the National Academy of Sciences of the United States of America* **57**, 1841-1848, doi:10.1073/pnas.57.6.1841 (1967).

7 Weiss, B. & Richardson, C. C. 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 (1967).

8 Olivera, B. M. & Lehman, I. R. Linkage of polynucleotides through phosphodiester bonds by an enzyme from Escherichia coli. *Proc Natl Acad Sci U S A* **57**, 1426-1433 (1967).

9 Olivera, B. M. & Lehman, I. R. Diphosphopyridine nucleotide: a cofactor for the polynucleotide-joining enzyme from Escherichia coli. *Proc Natl Acad Sci U S A* **57**, 1700-1704 (1967).

10 Shuman, S. DNA Ligases: Progress and Prospects. *Journal of Biological Chemistry* **284**, 17365-17369, doi:10.1074/jbc.R900017200 (2009).

11 Sekiguchi, J. & Shuman, S. Nick sensing by vaccinia virus DNA ligase requires a 5' phosphate at the nick and occupancy of the adenylate binding site on the enzyme. *Journal of virology* **71**, 9679-9684, doi:0022-538X/97/$04.00+0 (1997).

12 Odell, M., Sriskanda, V., Shuman, S. & Nikolov, D. B. Crystal structure of eukaryotic DNA ligase-adenylate illuminates the mechanism of nick sensing and strand joining. *Molecular cell* **6**, 1183-1193, doi:10.1016/S1097-2765(00)00115-5 (2000).

13 Nair, P. a. *et al.* Structural basis for nick recognition by a minimal pluripotent DNA ligase. *Nature structural & molecular biology* **14**, 770-778, doi:10.1038/nsmb1266 (2007).

14 Bauer, R. J., Jurkiw, T. J., Evans, T. C., Jr. & Lohman, G. J. Rapid Time Scale Analysis of T4 DNA Ligase-DNA Binding. *Biochemistry* **56**, 1117-1129, doi:10.1021/acs.biochem.6b01261 (2017).
15 Pascal, J. M., O'Brien, P. J., Tomkinson, A. E. & Ellenberger, T. Human DNA ligase I completely encircles and partially unwinds nicked DNA. *Nature* **432**, 473-478, doi:10.1038/nature03082 (2004).
16 Shuman, S., Liu, Y. & Schwer, B. Covalent catalysis in nucleotidyl transfer reactions: essential motifs in Saccharomyces cerevisiae RNA capping enzyme are conserved in Schizosaccharomyces pombe and viral capping enzymes and among polynucleotide ligases. *Proc Natl Acad Sci U S A* **91**, 12046-12050 (1994).
17 Håkansson, K., Doherty, a. J., Shuman, S. & Wigley, D. B. X-ray crystallography reveals a large conformational change during guanyl transfer by mRNA capping enzymes. *Cell* **89**, 545-553, doi:10.1016/S0092-8674(00)80236-6 (1997).
18 Shuman, S. & Schwer, B. RNA capping enzyme and DNA ligase: A superfamily of covalent nucleotidyl transferases. *Molecular Microbiology* **17**, 405-410, doi:10.1111/j.1365-2958.1995.mmi_17030405.x (1995).
19 Shuman, S. & Lima, C. D. The polynucleotide ligase and RNA capping enzyme superfamily of covalent nucleotidyltransferases. *Current opinion in structural biology* **14**, 757-764, doi:10.1016/j.sbi.2004.10.006 (2004).
20 Sawaya, R. & Shuman, S. Mutational analysis of the guanylyltransferase component of Mammalian mRNA capping enzyme. *Biochemistry* **42**, 8240-8249, doi:10.1021/bi034396d (2003).
21 Sriskanda, V. & Shuman, S. Role of nucleotidyltransferase motifs I, III and IV in the catalysis of phosphodiester bond formation by Chlorella virus DNA ligase. *Nucleic acids research* **30**, 903-911, doi:10.1093/nar/30.4.903 (2002).
22 Sriskanda, V. & Shuman, S. Role of nucleotidyl transferase motif V in strand joining by Chlorella virus DNA ligase. *Journal of Biological Chemistry* **277**, 9661-9667, doi:10.1074/jbc.M110613200 (2002).
23 Nandakumar, J., Nair, P. a. & Shuman, S. Last stop on the road to repair: structure of E. coli DNA ligase bound to nicked DNA-adenylate. *Molecular cell* **26**, 257-271, doi:10.1016/j.molcel.2007.02.026 (2007).
24 Cotner-Gohara, E. *et al.* Human DNA ligase III recognizes DNA ends by dynamic switching between two DNA-bound states. *Biochemistry* **49**, 6165-6176, doi:10.1021/bi100503w (2010).
25 Kaminski, A. M. *et al.* Structures of DNA-bound human ligase IV catalytic core reveal insights into substrate binding and catalysis. *Nat Commun* **9**, 2642, doi:10.1038/s41467-018-05024-8 (2018).
26 Shi, K. *et al.* T4 DNA ligase structure reveals a prototypical ATP-dependent ligase with a unique mode of sliding clamp interaction. *Nucleic Acids Res* **46**, 10474-10488, doi:10.1093/nar/gky776 (2018).
27 Sriskanda, V. & Shuman, S. Conserved residues in domain Ia are required for the reaction of Escherichia coli DNA ligase with NAD+. *J Biol Chem* **277**, 9695-9700, doi:10.1074/jbc.M111164200 (2002).
28 Sriskanda, V., Moyer, R. W. & Shuman, S. NAD+-dependent DNA ligase encoded by a eukaryotic virus. *J Biol Chem* **276**, 36100-36109, doi:10.1074/jbc.M105643200 (2001).

29 Unciuleac, M. C., Goldgur, Y. & Shuman, S. Two-metal versus one-metal mechanisms of lysine adenylylation by ATP-dependent and NAD+-dependent polynucleotide ligases. *Proc Natl Acad Sci U S A*, doi:10.1073/pnas.1619220114 (2017).
30 Sriskanda, V., Schwer, B., Ho, C. K. & Shuman, S. Mutational analysis of Escherichia coli DNA ligase identifies amino acids required for nick-ligation in vitro and for in vivo complementation of the growth of yeast cells deleted for CDC9 and LIG4. *Nucleic Acids Res* **27**, 3953-3963 (1999).
31 Luo, J. & Barany, F. Identification of essential residues in Thermus thermophilus DNA ligase. *Nucleic Acids Res* **24**, 3079-3085 (1996).
32 Lee, J. Y. *et al.* Crystal structure of NAD(+)-dependent DNA ligase: modular architecture and functional implications. *EMBO J* **19**, 1119-1129, doi:10.1093/emboj/19.5.1119 (2000).
33 Wang, L. K., Nair, P. A. & Shuman, S. Structure-guided mutational analysis of the OB, HhH, and BRCT domains of Escherichia coli DNA ligase. *J Biol Chem* **283**, 23343-23352, doi:10.1074/jbc.M802945200 (2008).
34 Gerloff, D. L., Woods, N. T., Farago, A. A. & Monteiro, A. N. BRCT domains: A little more than kin, and less than kind. *FEBS Lett* **586**, 2711-2716, doi:10.1016/j.febslet.2012.05.005 (2012).
35 Berman, H. M. *et al.* The Protein Data Bank. *Nucleic Acids Res* **28**, 235-242 (2000).
36 Tomkinson, A. E., Vijayakumar, S., Pascal, J. M. & Ellenberger, T. DNA ligases: Structure, reaction mechanism, and function. *Chemical Reviews* **106**, 687-699, doi:10.1021/cr040498d (2006).
37 Ellenberger, T. & Tomkinson, A. E. Eukaryotic DNA Ligases: Structural and Functional Insights. *Annual Review of Biochemistry* **77**, 313-338, doi:10.1146/annurev.biochem.77.061306.123941 (2008).
38 Odell, M., Malinina, L., Sriskanda, V., Teplova, M. & Shuman, S. Analysis of the DNA joining repertoire of Chlorella virus DNA ligase and a new crystal structure of the ligase-adenylate intermediate. *Nucleic Acids Research* **31**, 5090-5100, doi:10.1093/nar/gkg665 (2003).
39 Odell, M. & Shuman, S. Footprinting of Chlorella virus DNA ligase bound at a nick in duplex DNA. *Journal of Biological Chemistry* **274**, 14032-14039, doi:10.1074/jbc.274.20.14032 (1999).
40 Sekiguchi, J. & Shuman, S. Domain structure of vaccinia DNA ligase. *Nucleic Acids Res* **25**, 727-734 (1997).
41 Taylor, M. R., Conrad, J. a., Wahl, D. & O'Brien, P. J. Kinetic mechanism of human DNA ligase I reveals magnesium-dependent changes in the rate-limiting step that compromise ligation efficiency. *Journal of Biological Chemistry* **286**, 23054-23062, doi:10.1074/jbc.M111.248831 (2011).
42 Cardoso, M. C. *et al.* Mapping and use of a sequence that targets DNA ligase I to sites of DNA replication in vivo. *J Cell Biol* **139**, 579-587 (1997).
43 Wei, Y. F. *et al.* Molecular cloning and expression of human cDNAs encoding a novel DNA ligase IV and DNA ligase III, an enzyme active in DNA repair and recombination. *Mol Cell Biol* **15**, 3206-3216 (1995).
44 Mackey, Z. B. *et al.* DNA ligase III is recruited to DNA strand breaks by a zinc finger motif homologous to that of poly(ADP-ribose) polymerase. Identification of two

functionally distinct DNA binding regions within DNA ligase III. *J Biol Chem* **274**, 21679-21687 (1999).
45 Lakshmipathy, U. & Campbell, C. The human DNA ligase III gene encodes nuclear and mitochondrial proteins. *Mol Cell Biol* **19**, 3869-3876 (1999).
46 Perez-Jannotti, R. M., Klein, S. M. & Bogenhagen, D. F. Two forms of mitochondrial DNA ligase III are produced in Xenopus laevis oocytes. *J Biol Chem* **276**, 48978-48987, doi:10.1074/jbc.M107177200 (2001).
47 Mackey, Z. B. *et al.* An alternative splicing event which occurs in mouse pachytene spermatocytes generates a form of DNA ligase III with distinct biochemical properties that may function in meiotic recombination. *Mol Cell Biol* **17**, 989-998 (1997).
48 Moser, J. *et al.* Sealing of chromosomal DNA nicks during nucleotide excision repair requires XRCC1 and DNA ligase III alpha in a cell-cycle-specific manner. *Mol Cell* **27**, 311-323, doi:10.1016/j.molcel.2007.06.014 (2007).
49 Caldecott, K. W., McKeown, C. K., Tucker, J. D., Ljungquist, S. & Thompson, L. H. An interaction between the mammalian DNA repair protein XRCC1 and DNA ligase III. *Mol Cell Biol* **14**, 68-76 (1994).
50 Nash, R. A., Caldecott, K. W., Barnes, D. E. & Lindahl, T. XRCC1 protein interacts with one of two distinct forms of DNA ligase III. *Biochemistry* **36**, 5207-5211, doi:10.1021/bi962281m (1997).
51 Dulic, A. *et al.* BRCT domain interactions in the heterodimeric DNA repair protein XRCC1-DNA ligase III. *Biochemistry* **40**, 5906-5913 (2001).
52 Critchlow, S. E., Bowater, R. P. & Jackson, S. P. Mammalian DNA double-strand break repair protein XRCC4 interacts with DNA ligase IV. *Current biology : CB* **7**, 588-598, doi:10.1016/S0960-9822(06)00258-2 (1997).
53 Shuman, S. & Glickman, M. S. Bacterial DNA repair by non-homologous end joining. *Nat Rev Microbiol* **5**, 852-861, doi:10.1038/nrmicro1768 (2007).
54 Okazaki, R., Okazaki, T., Sakabe, K. & Sugimoto, K. Mechanism of DNA replication possible discontinuity of DNA chain growth. *Jpn J Med Sci Biol* **20**, 255-260 (1967).
55 Sugimoto, K., Okazaki, T. & Okazaki, R. 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 (1968).
56 Hartwell, L. H. Saccharomyces cerevisiae cell cycle. *Bacteriol Rev* **38**, 164-198 (1974).
57 Johnston, L. H. & Nasmyth, K. a. Saccharomyces cerevisiae cell cycle mutant cdc9 is defective in DNA ligase. *Nature* **274**, 891-893, doi:10.1038/274891a0 (1978).
58 Johnston, L. H. The DNA repair capability of cdc9, the Saccharomyces cerevisiae mutant defective in DNA ligase. *Mol Gen Genet* **170**, 89-92 (1979).
59 Waga, S. & Stillman, B. Anatomy of a DNA replication fork revealed by reconstitution of SV40 DNA replication in vitro. *Nature* **369**, 207-212, doi:10.1038/369207a0 (1994).
60 Li, C., Goodchild, J. & Baril, E. F. DNA ligase I is associated with the 21 S complex of enzymes for DNA synthesis in HeLa cells. *Nucleic Acids Res* **22**, 632-638 (1994).
61 Wu, Y. *et al.* A 17S multiprotein form of murine cell DNA polymerase mediates polyomavirus DNA replication in vitro. *J Cell Biochem* **54**, 32-46, doi:10.1002/jcb.240540105 (1994).
62 Montecucco, A. *et al.* 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, doi:10.1093/emboj/17.13.3786 (1998).
63 Montecucco, A. *et al.* The N-terminal domain of human DNA ligase I contains the nuclear localization signal and directs the enzyme to sites of DNA replication. *EMBO J* **14**, 5379-5386 (1995).
64 Levin, D. S., Bai, W., Yao, N., O'Donnell, M. & Tomkinson, a. E. An interaction between DNA ligase I and proliferating cell nuclear antigen: implications for Okazaki fragment synthesis and joining. *Proceedings of the National Academy of Sciences of the United States of America* **94**, 12863-12868, doi:10.1073/pnas.94.24.12863 (1997).
65 Levin, D. S., McKenna, A. E., Motycka, T. a., Matsumoto, Y. & Tomkinson, A. E. Interaction between PCNA and DNA ligase I is critical for joining of Okazaki fragments and long-patch base-excision repair. *Current Biology* **10**, 919-922, doi:10.1016/S0960-9822(00)00619-9 (2000).
66 Tom, S., Henricksen, L. A., Park, M. S. & Bambara, R. A. DNA ligase I and proliferating cell nuclear antigen form a functional complex. *J Biol Chem* **276**, 24817-24825, doi:10.1074/jbc.M101673200 (2001).
67 Rossi, R. *et al.* The replication factory targeting sequence/PCNA-binding site is required in G1 to control the phosphorylation status of DNA ligase I. *EMBO Journal* **18**, 5745-5754, doi:10.1093/emboj/18.20.5745 (1999).
68 Koundrioukoff, S. *et al.* A direct interaction between proliferating cell nuclear antigen (PCNA) and Cdk2 targets PCNA-interacting proteins for phosphorylation. *The Journal of biological chemistry* **275**, 22882-22887, doi:10.1074/jbc.M001850200 (2000).
69 Bentley, D. *et al.* DNA ligase I is required for fetal liver erythropoiesis but is not essential for mammalian cell viability. *Nat Genet* **13**, 489-491, doi:10.1038/ng0896-489 (1996).
70 Bentley, D. J. *et al.* DNA ligase I null mouse cells show normal DNA repair activity but altered DNA replication and reduced genome stability. *Journal of cell science* **115**, 1551-1561 (2002).
71 Han, L., Masani, S., Hsieh, C.-L. & Yu, K. DNA ligase I is not essential for Mammalian cell viability. *Cell reports* **7**, 316-320, doi:10.1016/j.celrep.2014.03.024 (2014).
72 Arakawa, H. *et al.* Functional redundancy between DNA ligases I and III in DNA replication in vertebrate cells. *Nucleic Acids Res* **40**, 2599-2610, doi:10.1093/nar/gkr1024 (2012).
73 Mackenney, V. J., Barnes, D. E. & Lindahl, T. Specific function of DNA ligase I in simian virus 40 DNA replication by human cell-free extracts is mediated by the amino-terminal non-catalytic domain. *J Biol Chem* **272**, 11550-11556 (1997).
74 Meira, L. B., Burgis, N. E. & Samson, L. D. in *Genome Instability in Cancer Development* (eds Nathan Back *et al.*) 125-173 (Springer Netherlands, 2005).
75 Kunkel, T. A. & Erie, D. A. Eukaryotic Mismatch Repair in Relation to DNA Replication. *Annu Rev Genet* **49**, 291-313, doi:10.1146/annurev-genet-112414-054722 (2015).
76 Shuck, S. C., Short, E. A. & Turchi, J. J. Eukaryotic nucleotide excision repair: from understanding mechanisms to influencing biology. *Cell Res* **18**, 64-72, doi:10.1038/cr.2008.2 (2008).
77 Scharer, O. D. Nucleotide excision repair in eukaryotes. *Cold Spring Harb Perspect Biol* **5**, a012609, doi:10.1101/cshperspect.a012609 (2013).

78 Caldecott, K. W., Aoufouchi, S., Johnson, P. & Shall, S. XRCC1 polypeptide interacts with DNA polymerase beta and possibly poly (ADP-ribose) polymerase, and DNA ligase III is a novel molecular 'nick-sensor' in vitro. *Nucleic Acids Res* **24**, 4387-4394 (1996).
79 Cappelli, E. *et al.* Involvement of XRCC1 and DNA ligase III gene products in DNA base excision repair. *J Biol Chem* **272**, 23970-23975 (1997).
80 Dianova, II *et al.* XRCC1-DNA polymerase beta interaction is required for efficient base excision repair. *Nucleic Acids Res* **32**, 2550-2555, doi:10.1093/nar/gkh567 (2004).
81 Gao, Y. *et al.* DNA ligase III is critical for mtDNA integrity but not Xrcc1-mediated nuclear DNA repair. *Nature* **471**, 240-244, doi:10.1038/nature09773 (2011).
82 Klungland, A. & Lindahl, T. Second pathway for completion of human DNA base excision-repair: Reconstitution with purified proteins and requirement for DNase IV (FEN1). *EMBO Journal* **16**, 3341-3348, doi:10.1093/emboj/16.11.3341 (1997).
83 Prasad, R. *et al.* Specific interaction of DNA polymerase beta and DNA ligase I in a multiprotein base excision repair complex from bovine testis. *J Biol Chem* **271**, 16000-16007 (1996).
84 Sleeth, K. M., Robson, R. L. & Dianov, G. L. Exchangeability of mammalian DNA ligases between base excision repair pathways. *Biochemistry* **43**, 12924-12930, doi:10.1021/bi0492612 (2004).
85 Stucki, M. *et al.* Mammalian base excision repair by DNA polymerases delta and epsilon. *Oncogene* **17**, 835-843, doi:10.1038/sj.onc.1202001 (1998).
86 Fortini, P. *et al.* Different DNA polymerases are involved in the short- and long-patch base excision repair in mammalian cells. *Biochemistry* **37**, 3575-3580, doi:10.1021/bi972999h (1998).
87 Dovrat, D., Stodola, J. L., Burgers, P. M. J. & Aharoni, A. Sequential switching of binding partners on PCNA during in vitro Okazaki fragment maturation. *Proc Natl Acad Sci U S A* **111**, 14118-14123, doi:10.1073/pnas.1321349111 (2014).
88 Pascal, J. M. *et al.* A Flexible Interface between DNA Ligase and PCNA Supports Conformational Switching and Efficient Ligation of DNA. *Molecular Cell* **24**, 279-291, doi:10.1016/j.molcel.2006.08.015 (2006).
89 Zhang, Y. *et al.* Reconstitution of 5'-directed human mismatch repair in a purified system. *Cell* **122**, 693-705, doi:10.1016/j.cell.2005.06.027 (2005).
90 Costa, R. M., Chigancas, V., Galhardo Rda, S., Carvalho, H. & Menck, C. F. The eukaryotic nucleotide excision repair pathway. *Biochimie* **85**, 1083-1099 (2003).
91 Chang, H. H. Y., Pannunzio, N. R., Adachi, N. & Lieber, M. R. Non-homologous DNA end joining and alternative pathways to double-strand break repair. *Nat Rev Mol Cell Biol* **18**, 495-506, doi:10.1038/nrm.2017.48 (2017).
92 Han, L., Mao, W. & Yu, K. X-ray repair cross-complementing protein 1 (XRCC1) deficiency enhances class switch recombination and is permissive for alternative end joining. *Proc Natl Acad Sci U S A* **109**, 4604-4608, doi:10.1073/pnas.1120743109 (2012).
93 Lu, G. *et al.* Ligase I and ligase III mediate the DNA double-strand break ligation in alternative end-joining. *Proc Natl Acad Sci U S A*, doi:10.1073/pnas.1521597113 (2016).
94 Masani, S., Han, L., Meek, K. & Yu, K. Redundant function of DNA ligase 1 and 3 in alternative end-joining during immunoglobulin class switch recombination. *Proc Natl Acad Sci U S A*, doi:10.1073/pnas.1521630113 (2016).

95 Chiruvella, K. K., Liang, Z., Birkeland, S. R., Basrur, V. & Wilson, T. E. Saccharomyces cerevisiae DNA ligase IV supports imprecise end joining independently of its catalytic activity. *PLoS Genet* **9**, e1003599, doi:10.1371/journal.pgen.1003599 (2013).
96 Wilson, T. E., Grawunder, U. & Lieber, M. R. Yeast DNA ligase IV mediates non-homologous DNA end joining. *Nature* **388**, 495-498, doi:10.1038/41365 (1997).
97 Teo, S. H. & Jackson, S. P. Identification of Saccharomyces cerevisiae DNA ligase .4. Involvement in DNA double-strand break repair. *EMBO Journal* **16**, 4788-4795, doi:10.1093/emboj/16.15.4788 (1997).
98 Marchetti, C. *et al.* Identification of a novel motif in DNA ligases exemplified by DNA ligase IV. *DNA Repair* **5**, 788-798, doi:10.1016/j.dnarep.2006.03.011 (2006).
99 Methot, S. P. & Di Noia, J. M. Molecular Mechanisms of Somatic Hypermutation and Class Switch Recombination. *Adv Immunol* **133**, 37-87, doi:10.1016/bs.ai.2016.11.002 (2017).
100 Arya, R. & Bassing, C. H. V(D)J Recombination Exploits DNA Damage Responses to Promote Immunity. *Trends Genet* **33**, 479-489, doi:10.1016/j.tig.2017.04.006 (2017).
101 Taccioli, G. E. *et al.* Impairment of V(D)J recombination in double-strand break repair mutants. *Science* **260**, 207-210 (1993).
102 Barnes, D. E., Tomkinson, A. E., Lehmann, A. R., Webster, A. D. & Lindahl, T. Mutations in the DNA ligase I gene of an individual with immunodeficiencies and cellular hypersensitivity to DNA-damaging agents. *Cell* **69**, 495-503 (1992).
103 Webster, a. D., Barnes, D. E., Arlett, C. F., Lehmann, a. R. & Lindahl, T. Growth retardation and immunodeficiency in a patient with mutations in the DNA ligase I gene. *Lancet* **339**, 1508-1509, doi:10.1016/0140-6736(92)91266-B (1992).
104 Willis, a. E. & Lindahl, T. Vol. 325 355-357 (1987).
105 Prigent, C., Satoh, M. S., Daly, G., Barnes, D. E. & Lindahl, T. Aberrant DNA repair and DNA replication due to an inherited enzymatic defect in human DNA ligase I. *Molecular and cellular biology* **14**, 310-317, doi:10.1128/MCB.14.1.310.Updated (1994).
106 Maffucci, P. *et al.* Biallelic mutations in DNA ligase 1 underlie a spectrum of immune deficiencies. *J Clin Invest* **128**, 5489-5504, doi:10.1172/JCI99629 (2018).
107 Altmann, T. & Gennery, A. R. DNA ligase IV syndrome; a review. *Orphanet J Rare Dis* **11**, 137, doi:10.1186/s13023-016-0520-1 (2016).
108 O'Driscoll, M. *et al.* DNA ligase IV mutations identified in patients exhibiting developmental delay and immunodeficiency. *Mol Cell* **8**, 1175-1185 (2001).
109 Coutinho, P. & Barbot, C. in *GeneReviews((R))* (eds M. P. Adam *et al.*) (GeneReviews, 1993).
110 Date, H. *et al.* Early-onset ataxia with ocular motor apraxia and hypoalbuminemia is caused by mutations in a new HIT superfamily gene. *Nature genetics* **29**, 184-188, doi:10.1038/ng1001-184 (2001).
111 Moreira, M. C. *et al.* The gene mutated in ataxia-ocular apraxia 1 encodes the new HIT/Zn-finger protein aprataxin. *Nature genetics* **29**, 189-193, doi:10.1038/ng1001-189 (2001).
112 Ahel, I. *et al.* The neurodegenerative disease protein aprataxin resolves abortive DNA ligation intermediates. *Nature* **443**, 713-716, doi:10.1038/nature05164 (2006).
113 Çağlayan, M., Horton, J. K., Prasad, R. & Wilson, S. H. Complementation of aprataxin deficiency by base excision repair enzymes. *Nucleic acids research* **43**, 2271-2281, doi:10.1093/nar/gkv079 (2015).

114 Castellotti, B. *et al.* Ataxia with oculomotor apraxia type1 (AOA1): novel and recurrent aprataxin mutations, coenzyme Q10 analyses, and clinical findings in Italian patients. *Neurogenetics* **12**, 193-201, doi:10.1007/s10048-011-0281-x (2011).
115 Tumbale, P. *et al.* Mechanism of APTX nicked DNA sensing and pleiotropic inactivation in neurodegenerative disease. *EMBO J* **37**, doi:10.15252/embj.201798875 (2018).
116 Sriskanda, V. & Shuman, S. Specificity and fidelity of strand joining by Chlorella virus DNA ligase. *Nucleic Acids Research* **26**, 3536-3541, doi:10.1093/nar/26.15.3536 (1998).
117 Lohman, G. J. S., Chen, L. & Evans, T. C. Kinetic characterization of single strand break ligation in duplex DNA by T4 DNA ligase. *The Journal of biological chemistry* **286**, 44187-44196, doi:10.1074/jbc.M111.284992 (2011).
118 Seidel, C. A. M., Schulz, A. & Sauer, M. H. M. Nucleobase-Specific Quenching of Fluorescent Dyes. 1. Nucleobase One-Electron Redox Potentials and Their Correlation with Static and Dynamic Quenching Efficiencies. *The Journal of Physical Chemistry* **100**, 5541-5553, doi:10.1021/jp951507c (1996).
119 Lohman, G. J. *et al.* A high-throughput assay for the comprehensive profiling of DNA ligase fidelity. *Nucleic Acids Res* **44**, e14, doi:10.1093/nar/gkv898 (2016).
120 Chauleau, M. & Shuman, S. Kinetic mechanism and fidelity of nick sealing by Escherichia coli NAD+-dependent DNA ligase (LigA). *Nucleic Acids Res* **44**, 2298-2309, doi:10.1093/nar/gkw049 (2016).
121 Cherepanov, A. V. & De Vries, S. Kinetic mechanism of the Mg2+-dependent nucleotidyl transfer catalyzed by T4 DNA and RNA ligases. *Journal of Biological Chemistry* **277**, 1695-1704, doi:10.1074/jbc.M109616200 (2002).
122 Tumbale, P., Williams, J. S., Schellenberg, M. J., Kunkel, T. A. & Williams, R. S. Aprataxin resolves adenylated RNA–DNA junctions to maintain genome integrity. *Nature* **506**, 111-115, doi:10.1038/nature12824 (2013).
123 Bhagwat, A. S., Sanderson, R. J. & Lindahl, T. Delayed DNA joining at 3' mismatches by human DNA ligases. *Nucleic Acids Research* **27**, 4028-4033, doi:10.1093/nar/27.20.4028 (1999).
124 Caglayan, M., Horton, J. K., Dai, D. P., Stefanick, D. F. & Wilson, S. H. Oxidized nucleotide insertion by pol beta confounds ligation during base excision repair. *Nat Commun* **8**, 14045, doi:10.1038/ncomms14045 (2017).
125 Wang, Y., Lamarche, B. J. & Tsai, M. D. Human DNA ligase IV and the ligase IV/XRCC4 complex: analysis of nick ligation fidelity. *Biochemistry* **46**, 4962-4976, doi:10.1021/bi0621516 (2007).
126 Rass, U., Ahel, I. & West, S. C. Actions of aprataxin in multiple DNA repair pathways. *Journal of Biological Chemistry* **282**, 9469-9474, doi:10.1074/jbc.M611489200 (2007).
127 Otwinowski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol* **276**, 307-326 (1997).
128 McCoy, A. J. *et al.* Phaser crystallographic software. *J Appl Crystallogr* **40**, 658-674, doi:10.1107/S0021889807021206 (2007).
129 McCoy, A. J. Solving structures of protein complexes by molecular replacement with Phaser. *Acta Crystallogr D Biol Crystallogr* **63**, 32-41, doi:10.1107/S0907444906045975 (2007).
130 Adams, P. D. *et al.* PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr* **66**, 213-221, doi:10.1107/S0907444909052925 (2010).

131 Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* **60**, 2126-2132, doi:10.1107/S0907444904019158 (2004).
132 Taylor, M. R. *The Role of Divalent Metal Ions in Enzymatic DNA Ligation*, University of Michigan, (2014).
133 Hashimoto, K., Tominaga, Y., Nakabeppu, Y. & Moriya, M. Futile short-patch DNA base excision repair of adenine:8-oxoguanine mispair. *Nucleic Acids Res* **32**, 5928-5934, doi:10.1093/nar/gkh909 (2004).
134 Parsons, J. L., Dianova, II & Dianov, G. L. APE1-dependent repair of DNA single-strand breaks containing 3'-end 8-oxoguanine. *Nucleic Acids Res* **33**, 2204-2209, doi:10.1093/nar/gki518 (2005).
135 Modrich, P. DNA mismatch correction. *Annu Rev Biochem* **56**, 435-466, doi:10.1146/annurev.bi.56.070187.002251 (1987).
136 Williams, J. S. & Kunkel, T. A. Ribonucleotides in DNA: origins, repair and consequences. *DNA Repair (Amst)* **19**, 27-37, doi:10.1016/j.dnarep.2014.03.029 (2014).
137 Nick McElhinny, S. A. *et al.* Abundant ribonucleotide incorporation into DNA by yeast replicative polymerases. *Proc Natl Acad Sci U S A* **107**, 4949-4954, doi:10.1073/pnas.0914857107 (2010).
138 Traut, T. W. Physiological concentrations of purines and pyrimidines. *Mol Cell Biochem* **140**, 1-22 (1994).
139 Ferraro, P., Franzolin, E., Pontarin, G., Reichard, P. & Bianchi, V. Quantitation of cellular deoxynucleoside triphosphates. *Nucleic Acids Res* **38**, e85, doi:10.1093/nar/gkp1141 (2010).
140 Li, Y. & Breaker, R. R. Kinetics of RNA Degradation by Specific Base Catalysis of Transesterification Involving the 2‘-Hydroxyl Group. *Journal of the American Chemical Society* **121**, 5364-5372, doi:10.1021/ja990592p (1999).
141 Jaishree, T. N., van der Marel, G. A., van Boom, J. H. & Wang, A. H. J. Structural influence of RNA incorporation in DNA: Quantitative nuclear magnetic resonance refinement of d(CG)r(CG)d(CG) and d(CG)r(C)d(TAGCG). *Biochemistry* **32**, 4903-4911, doi:10.1021/bi00069a027 (1993).
142 Egli, M., Usman, N. & Rich, A. Conformational influence of the ribose 2'-hydroxyl group: Crystal structures of DNA-RNA chimeric duplexes. *Biochemistry* **32**, 3221-3237, doi:10.1021/bi00064a004 (1993).
143 Williams, J. S., Lujan, S. A. & Kunkel, T. A. Processing ribonucleotides incorporated during eukaryotic DNA replication. *Nat Rev Mol Cell Biol* **17**, 350-363, doi:10.1038/nrm.2016.37 (2016).
144 Williams, J. S. *et al.* Topoisomerase 1-mediated removal of ribonucleotides from nascent leading-strand DNA. *Mol Cell* **49**, 1010-1015, doi:10.1016/j.molcel.2012.12.021 (2013).
145 Cho, J. E., Kim, N. & Jinks-Robertson, S. Topoisomerase 1-dependent deletions initiated by incision at ribonucleotides are biased to the non-transcribed strand of a highly activated reporter. *Nucleic Acids Res* **43**, 9306-9313, doi:10.1093/nar/gkv824 (2015).
146 Samai, P. & Shuman, S. Structure-function analysis of the OB and latch domains of chlorella virus DNA ligase. *The Journal of biological chemistry* **286**, 22642-22652, doi:10.1074/jbc.M111.245399 (2011).
147 Nick McElhinny, S. A. *et al.* Genome instability due to ribonucleotide incorporation into DNA. *Nat Chem Biol* **6**, 774-781, doi:10.1038/nchembio.424 (2010).

148 Sriskanda, V. & Shuman, S. Chlorella virus DNA ligase: Nick recognition and mutational analysis. *Nucleic Acids Research* **26**, 525-531, doi:10.1093/nar/26.2.525 (1998).
149 Unciuleac, M.-C., Goldgur, Y. & Shuman, S. Structure and two-metal mechanism of a eukaryal nick-sealing RNA ligase. *Pnas*, 1516536112--1516536112-, doi:10.1073/pnas.1516536112 (2015).
150 Simsek, D. *et al.* Crucial role for DNA ligase III in mitochondria but not in Xrcc1-dependent repair. *Nature* **471**, 245-248, doi:10.1038/nature09794 (2011).
151 Grawunder, U. *et al.* Activity of DNA ligase IV stimulated by complex formation with XRCC4 protein in mammalian cells. *Nature* **388**, 492-495, doi:10.1038/41358 (1997).
152 Sun, D. *et al.* Elevated expression of DNA ligase I in human cancers. *Clinical cancer research* **7**, 4143-4148 (2001).
153 Harrison, C., Ketchen, A. M., Redhead, N. J., O'Sullivan, M. J. & Melton, D. W. Replication failure, genome instability, and increased cancer susceptibility in mice with a point mutation in the DNA ligase I gene. *Cancer Res* **62**, 4065-4074 (2002).
154 Muvarak, N. *et al.* c-MYC Generates Repair Errors via Increased Transcription of Alternative-NHEJ Factors, LIG3 and PARP1, in Tyrosine Kinase-Activated Leukemias. *Mol Cancer Res* **13**, 699-712, doi:10.1158/1541-7786.MCR-14-0422 (2015).
155 Grawunder, U., Zimmer, D., Fugmann, S., Schwarz, K. & Lieber, M. R. DNA ligase IV is essential for V(D)J recombination and DNA double-strand break repair in human precursor lymphocytes. *Mol Cell* **2**, 477-484 (1998).
156 Chan, J. Y., Becker, F. F., German, J. & Ray, J. H. Altered DNA ligase I activity in Bloom's syndrome cells. *Nature* **325**, 357-359, doi:10.1038/325357a0 (1987).
157 Kodama, K., Barnes, D. E. & Lindahl, T. In vitro mutagenesis and functional expression in Escherichia coli of a cDNA encoding the catalytic domain of human DNA ligase I. *Nucleic acids research* **19**, 6093-6099 (1991).
158 Romani, A. M. Cellular magnesium homeostasis. *Arch Biochem Biophys* **512**, 1-23, doi:10.1016/j.abb.2011.05.010 (2011).
159 Kyrieleis, O. J., Chang, J., de la Pena, M., Shuman, S. & Cusack, S. Crystal structure of vaccinia virus mRNA capping enzyme provides insights into the mechanism and evolution of the capping apparatus. *Structure* **22**, 452-465, doi:10.1016/j.str.2013.12.014 (2014).
160 Samai, P. & Shuman, S. Kinetic analysis of DNA strand joining by Chlorella virus DNA ligase and the role of nucleotidyltransferase motif VI in ligase adenylylation. *Journal of Biological Chemistry* **287**, 28609-28618, doi:10.1074/jbc.M112.380428 (2012).
161 Samai, P. & Shuman, S. Functional dissection of the DNA interface of the nucleotidyltransferase domain of chlorella virus DNA ligase. *Journal of Biological Chemistry* **286**, 13314-13326, doi:10.1074/jbc.M111.226191 (2011).