

## SURVEY AND SUMMARY

# Structural insight into DNA joining: from conserved mechanisms to diverse scaffolds

Adele Williamson<sup>1,2,\*</sup> and Hanna-Kirsti S. Leiros<sup>1</sup>

<sup>1</sup>School of Science, University of Waikato, Hamilton 3240, New Zealand and <sup>2</sup>Department of Chemistry, UiT The Arctic University of Norway, Tromsø N-9037, Norway

Received February 18, 2020; Revised April 14, 2020; Editorial Decision April 15, 2020; Accepted April 20, 2020

### ABSTRACT

**DNA ligases are diverse enzymes with essential functions in replication and repair of DNA; here we review recent advances in their structure and distribution and discuss how this contributes to understanding their biological roles and technological potential. Recent high-resolution crystal structures of DNA ligases from different organisms, including DNA-bound states and reaction intermediates, have provided considerable insight into their enzymatic mechanism and substrate interactions. All cellular organisms possess at least one DNA ligase, but many species encode multiple forms some of which are modular multifunctional enzymes. New experimental evidence for participation of DNA ligases in pathways with additional DNA modifying enzymes is defining their participation in non-redundant repair processes enabling elucidation of their biological functions. Coupled with identification of a wealth of DNA ligase sequences through genomic data, our increased appreciation of the structural diversity and phylogenetic distribution of DNA ligases has the potential to uncover new biotechnological tools and provide new treatment options for bacterial pathogens.**

### INTRODUCTION

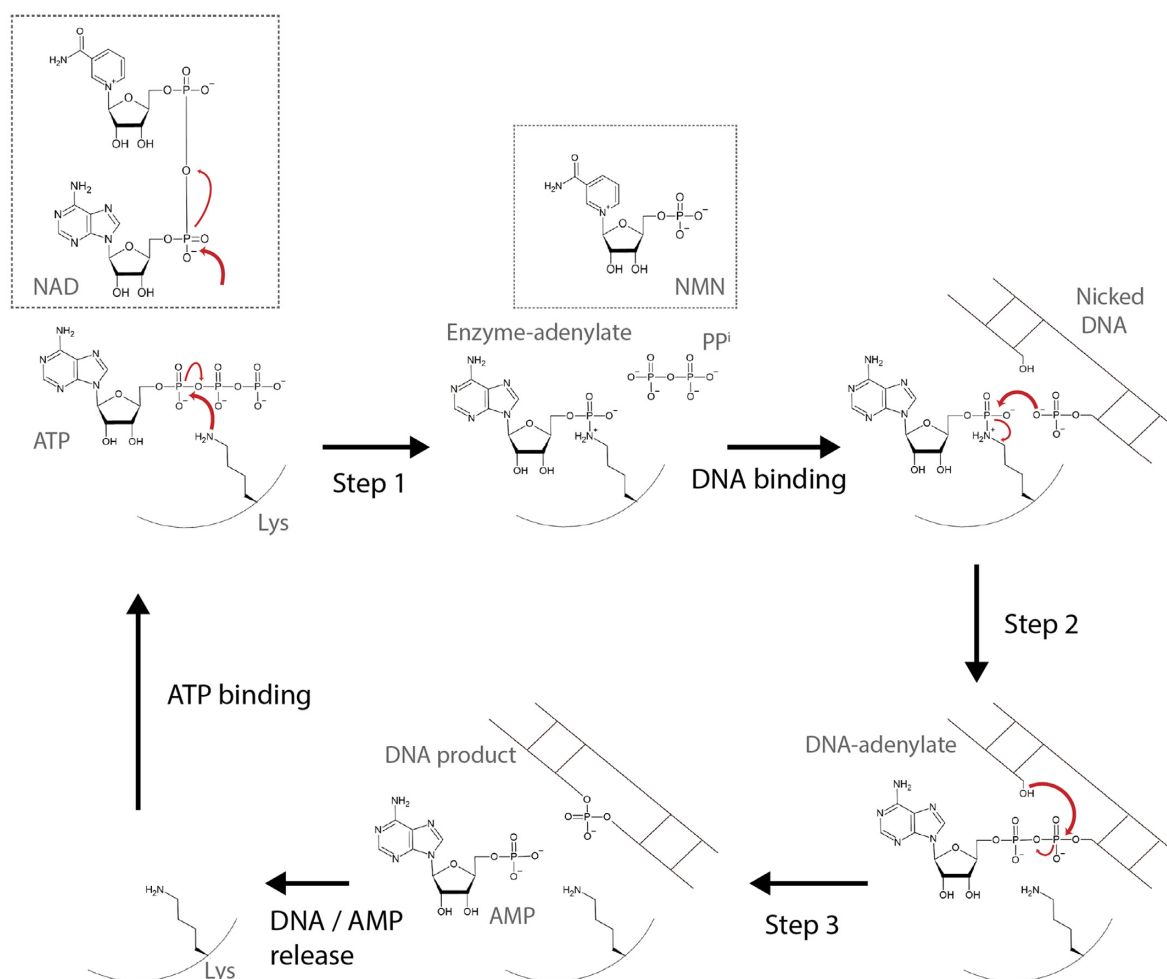
DNA ligases which join breaks in the phosphodiester backbone of double-stranded DNA are essential in all organisms for replication and repair of DNA, and have central roles in many molecular biological applications (1). They are defined as either ATP-dependent (AD-ligases) or NAD-dependent (ND-ligases) depending on the nature of the adenylate-donating cofactor, either ATP or NAD<sup>+</sup>, used to provide energy for ligation of adjacent 5' phosphate and 3' DNA termini. These AD-ligases and ND-ligases are de-

rived from a common ancestor and share a conserved catalytic mechanism and structural features (2).

The ligation reaction involves three discrete catalytic steps (Figure 1): in Step 1 the ligase enzyme in the absence of DNA is covalently adenylated at a conserved lysine residue by the nucleotide donor releasing pyrophosphate (PP<sub>i</sub>) or nicotinamide mononucleotide (NMN), and forming a covalent ligase-(lysyl-N $\zeta$ )-AMP linkage; in Step 2 the AMP moiety is transferred to the 5'P of the nicked DNA, activating it for nucleophilic attack in step 3, which forms the new phosphodiester bond on the DNA backbone and releases AMP. These chemical steps are enacted by conserved nucleotidyl transferase motifs and are accompanied by large-scale domain rearrangements that facilitate sequential cofactor and substrate engagement (Supplementary Figure S1) (3).

The minimal scaffold for effective DNA ligation comprises two domains: the nucleotidyl transferase (NT) domain which is the site of all chemical transactions of the ligation reaction, and the oligonucleotide/oligosaccharide binding (OB) domain which is essential for substrate binding and positioning during catalysis. The NT domain, also referred to as the adenylation domain has a mixed  $\alpha/\beta$  structure, and includes conserved enzyme motifs I, III, IIIa, IV and V, that are common to all members of the nucleotidyl transferases superfamily (4). The smaller OB domain has a  $\beta$  barrel structure that binds the phosphate-rich DNA backbone through its positively-charged concave surface. These domains are joined by a flexible linker region enabling their reorientation between 'open' states where the DNA-binding faces of the AD and OB domains are deflected away from each other exposing the active site, and 'closed' states where binding faces are opposing, usually concomitant with DNA substrate interaction (3,5). This catalytic two-domain core may be decorated N- and C-terminally by additional domains that are involved in enhancing DNA binding during ligation, or that have discrete additional catalytic functions. In the case of the ND-ligases,

\*To whom correspondence should be addressed. Tel: +64 7 838 4688; Email: adelew@waikato.ac.nz



**Figure 1.** Steps of the nick sealing reaction catalyzed by DNA ligase. Step 1 of the main reaction scheme illustrates the reaction of the ATP-dependent DNA ligase isoform using the ATP cofactor and releasing PP<sub>i</sub>, with the NAD cofactor and NMN by-product illustrated in the insets above. The remaining catalytic steps are common to both isoforms.

the appending domains are highly conserved and intrinsically linked to the ligation reaction, while ATP-dependent ligases are structurally diverse and as described below, comprise modular multi-step DNA-processing platforms for DNA repair (6–9).

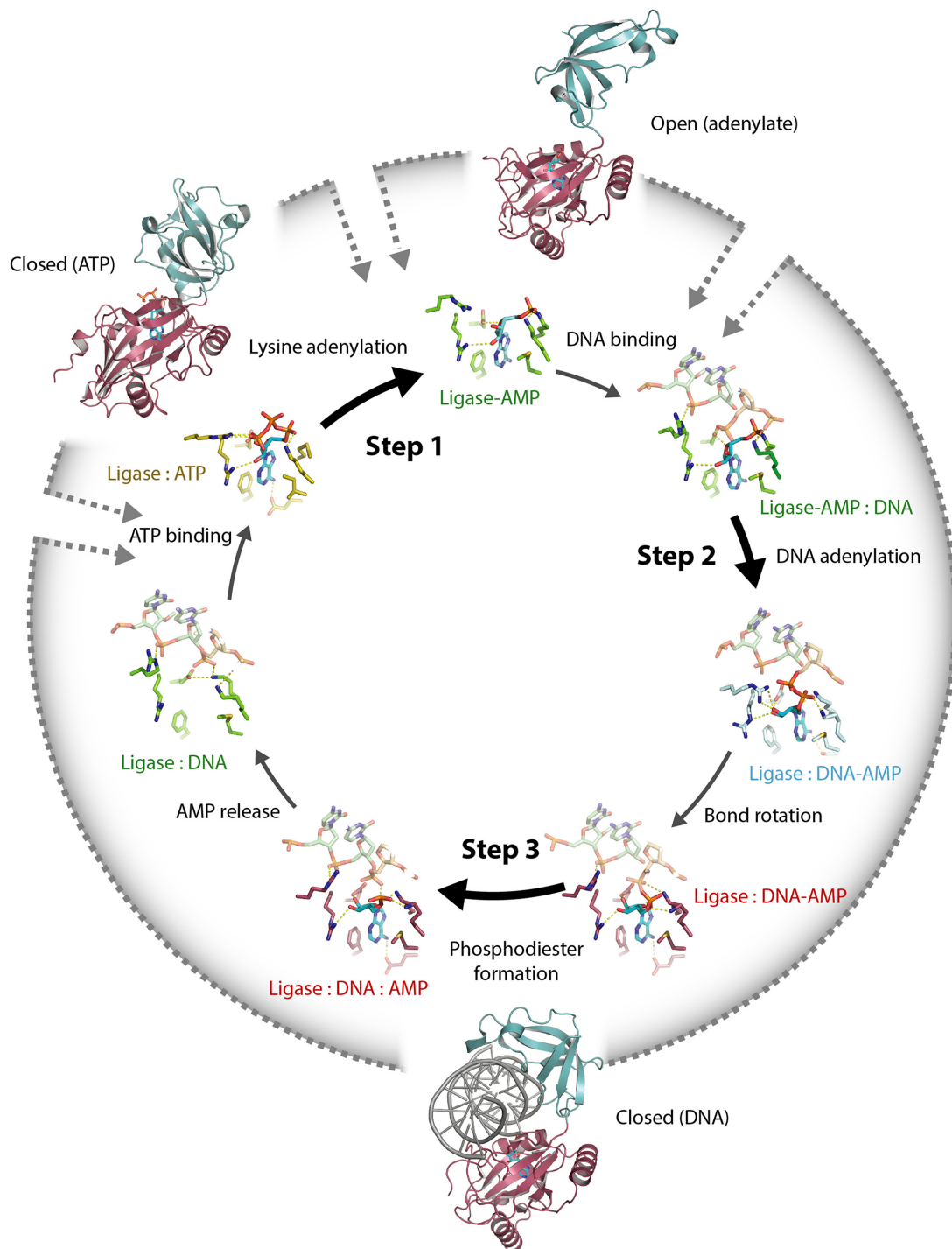
Distribution of AD- and ND-ligases is broadly split along taxonomic lines with eukaryotes and almost all archaea employing the former for both replication and repair. All bacteria studied to date exclusively use ND-ligases for replication processes, although many species possess additional AD-ligases for dedicated repair pathways (4,10,11). Both varieties have been reported in viruses, however the ATP-dependent class have been far more intensely studied, and bacteriophage AD-ligases constitute foundational biotechnological tools in construction of recombinant DNA as well as our biochemical and structural understanding of this class of enzyme (1,12,13).

### MECHANISTIC INSIGHTS INTO DNA JOINING

Recent structural advances have provided considerable new insight into both the mechanistic detail of the three-step catalytic ligation reaction, including binding and orientation

of the nucleotide cofactor, as well as the role of the divalent metal cations that are essential for all catalytic steps.

Step 1, covalent adenylation of the conserved lysine residue located in motif I of the NT domain requires participation of a specialized VI motif located in the OB domain in AD-ligases, and of the N-terminal Ia domain in ND-ligases; both of which function to orient the nucleotide leaving group (14,15). The base of the cofactor, whether ATP or NAD, is buried in a hydrophobic pocket of the NT domain, sandwiched between an aliphatic residue of motif IV and aromatic sidechain from IIIa, while a conserved lysine from motif IV contacts the heterocycle N1, and in many cases an acidic residue from motif I hydrogen bonds with the extracyclic N6 (16–18). Cofactor binding presumably occurs in the open conformation to allow access to the catalytic site, followed by adoption of a closed conformation bringing the relevant catalytic motifs of the OB- and Ia domains into close proximity with PP<sub>i</sub> and NMN (Figure 2 Closed (ATP)). Structures of such productive closed conformations were recently reported in a series of papers which have captured the pre-step1 ligase-cofactor complexes of the ND-ligase from *Escherichia coli* (Eco-Lig) and the catalytic core of Lig-D type AD-ligase from *My-*



**Figure 2.** Structural detail of the catalytic steps of the DNA ligase reaction (inner scheme) and accompanying conformational transitions of the catalytic core domains (outer scheme). Active site structures are: *Mycobacterium tuberculosis* Lig D (yellow) showing the ligase:ATP non-covalent complex (6nhz); *Chlorella* virus (green) showing Ligase-AMP covalent adduct (1fvi), the Ligase-AMP adduct with bound DNA (2q2t) and the Ligase:DNA product (2q2u); *Alteromonas mediterranea* (cyan) Ligase:AMP-DNA enzyme complexed with DNA-adenylate (6gdr); *Prochlorococcus marinus* (dark red) enzyme complexed with DNA-adenylate (6rce) and Ligase-DNA-AMP complex (6rau). Domain conformation structures are: Open (adenylate) *Psychromonas* SP041 (4do5); Closed (DNA) *A. mediterranea* (6gdr); Closed (ATP) catalytic core of *M. tuberculosis* Lig D (6nhz). The NT domain is colored dark red, the OB domain in teal, DNA substrate in grey and the nucleotide cofactor cyan.

*cobacterium tuberculosis* (Mtu-LigD) as well as two closely related RNA ligases (Figure 3 and Supplementary Figure S2) (19–21). In all cases, mutation of the motif I lysine to an isosteric methionine blocked Step 1 chemistry allowing the enzyme to be co-crystallized with both the cofactor and catalytically-essential metal ions. The most salient finding is that while NAD-dependent DNA ligases require only a single ‘catalytic’ metal ion for lysine adenylation, ATP-utilizing ligases (both those acting on DNA as well as RNA substrates) have an additional metal binding site with a structural role (Figure 3A). In both forms, the catalytic metal lowers the  $pK_a$  of the lysine nucleophile and stabilizes the transition state on the  $\alpha$ -phosphate, while in the ATP-dependent form the second structural metal ion orients the  $\beta$ - and  $\gamma$ -phosphates of the leaving group. This configuration was first visualized in the RNA ligase enzymes of *Naegleria gruberi* RNA ligase and T4 RNA ligase1, and reveals that the catalytic metal forms a single direct contact to the  $\alpha$ -phosphate oxygen of the ATP molecule, while the other ligands of the octahedral complex are provided by water-mediated contacts from the side chains of conserved residues in motifs I, III, and IV (19–21). The deprotonated lysine, stabilized by a water-mediated contact to the positively charged metal ion, is positioned for in-line nucleophilic attack on the  $\alpha$ -phosphate, which proceeds via a pentavalent transition state, resulting in an inversion of stereochemistry at this center. The second ‘structural’ metal ion of the ATP-dependent DNA and RNA ligases coordinates oxygens of both the  $\beta$ - and  $\gamma$ -ATP phosphates ensuring their correct orientation. The metal ion is also coordinated by residues from the distinctive C-terminal ‘C domain’ which has no structural homology with the OB domain of the DNA ligases. Interestingly, in the ATP-dependent Mtu-LigD structure there are no direct interactions between motif VI and ATP in the metal-bound complex, thus the precise contribution to this region remains unclear at the present (Supplementary Figure S2) (19). In the ND-ligases, the specialized Ia domain orients the NMN leaving group, analogous to the role of the structural metal ion in the AD-ligases, thus only the single catalytic metal ion is required (Figure 3B). The Ia domain, together with the NT domain, form a binding pocket for the NAD<sup>+</sup> cofactor with the nicotinamide portion adopting an *anti* conformation and forming stacking interactions with a pair of tyrosine residues.

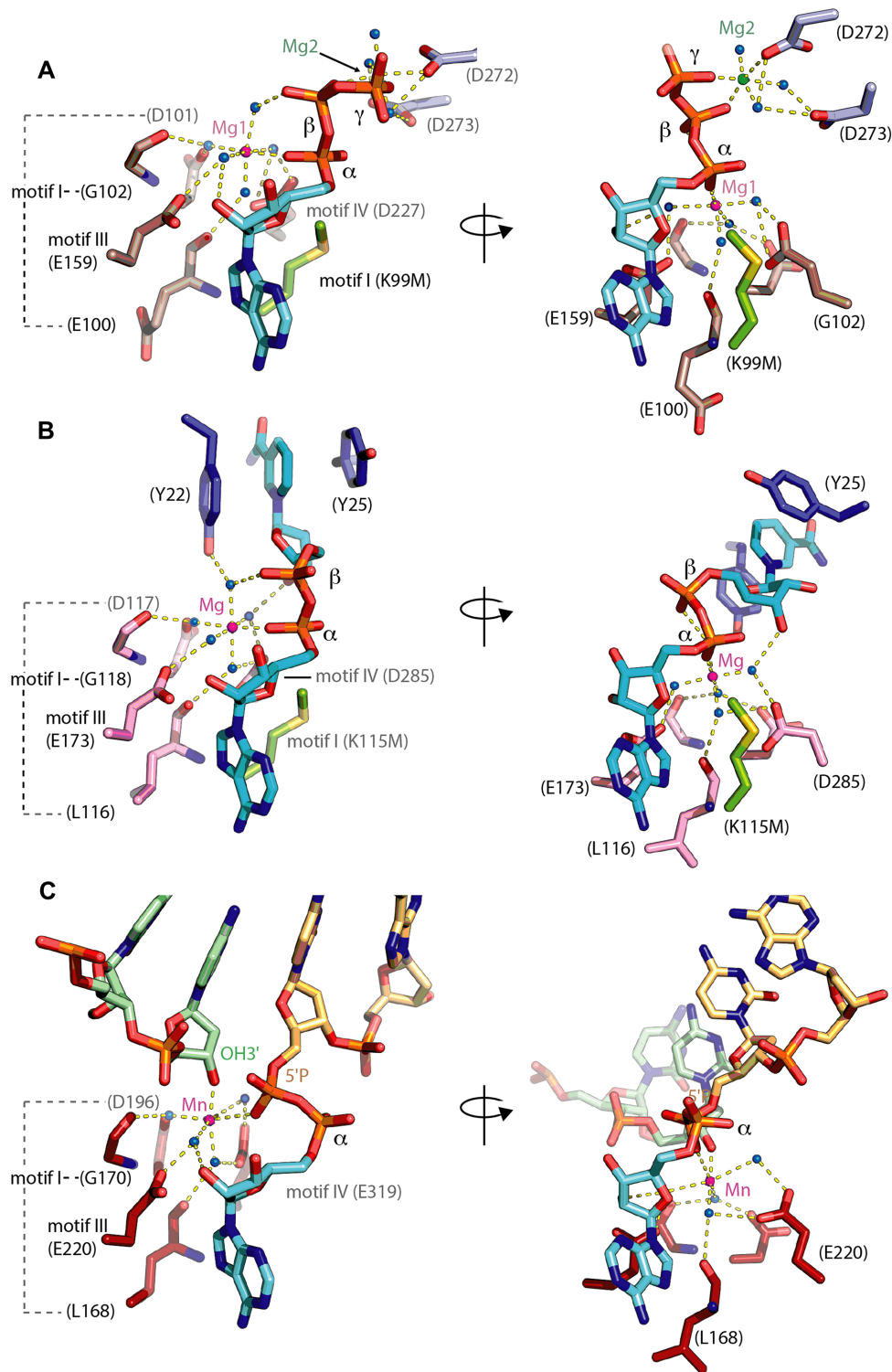
Subsequent to formation of the covalent ligase-(lysyl-N $\zeta$ )-AMP bond during step 1, the leaving group, exits the active site presumably in concert with re-opening of the core domains (Figure 2, Open (adenylate)). Remodeling of the active site post-Step1 is captured by the enzyme-adenylate structure of the *Chlorella* virus AD-ligase ChIV-Lig (17) as well as several subsequent structures from other organisms (16,22,23), where a covalent linkage between the motif I catalytic lysine and the AMP  $\alpha$ -phosphate is clear. The nucleoside now adopts the favourable *anti* conformation, a transition which is accompanied by a change in the hydrogen bonding of the motif I arginine from the ribose O2' pre-step 1 to O3' and loss of the interaction between a second conserved arginine and motif III glutamate from O2' pre-Step1 to O4' of the ribose ring. The enzyme-adenylate now binds and closes around the DNA substrate (Figure 2, Closed

(DNA)). The diverse structural solutions to DNA substrate engagement which result in necessary distortion around the break site are detailed in the following section, but in all cases the OB domain rotates to form a C-shaped clamp about the duplex and remains in this position throughout the remainder of catalysis. DNA binding positions the opposing 5'P and 3'OH ends relative to the adenylated lysine of the NT domain residue and distorts the duplex exposing the ends aligning them optimally for Step 2 catalysis. This bending, is achieved by insertion of the OB domain into the minor groove, widening it and converting the terminal nucleotides of the nick from the typical DNA B form to the RNA-like A configuration with a 3' endo sugar pucker which are induced by a pair of conserved phenylalanine residues. The DNA-bound structure of the ChIV-Lig lysyl-AMP adduct immediately prior to Step 2 represents one of the foundational structures illuminating the mechanism of ligase-DNA engagement, and revealed how the DNA is bound across the active site with the nick positioned above the lysyl-N $\zeta$  adenylate, with the 5' phosphate of the DNA poised for nucleophilic attack (24).

The conformation immediately following handover of the covalent linkage of the AMP  $\alpha$ -phosphate from the catalytic motif I lysine to the 5'P of the nick ligase-bound DNA-adenylate is captured in the structure of the AD-ligase of *Alteromonas mediterranea* (Ame-Lig) (25). Immediately post Step 2, the AMP-DNA phosphoanhydride bond is orientated away from the 3'OH of the nick with the 5' phosphate of the DNA retaining its original position and the  $\alpha$ -phosphate of the AMP forming electrostatic interactions with lysines of motif I and motif V. To be sufficiently close to the 3'OH terminus for nucleophilic attack to occur, the activated phosphate reorients almost 90° about the diphosphate-bond axis, which together with stereochemical inversion during Step 2, returns the AMP nucleoside to a strained *syn* configuration observed in ligase-DNA adenylate structures from several organisms (6,25–30). This reorientation is mediated by the lysine from motif V as well as the catalytic lysine in motif I, while changes in the non-covalent bonding to the AMP ribose hydroxyls again require rearrangement of contacts to the motif I, III and V residues.

The final pre-Step 3 configuration positions the AMP-DNA phosphoanhydride bond for in-line attack by the 3'OH group, which is facilitated by the catalytic metal ion. Recently a series of structures have captured the AD-ligase of *Prochlorococcus marinus* (hereafter Pmar-Lig) at crucial stages of the ligase reaction (27). These include a pre-Step 3 ternary complex where the catalytic metal site is occupied by manganese; the first instance where the metal binding site has been unequivocally identified in a DNA-bound enzyme complex when both nick termini are intact. This single metal ion is coordinated with octahedral geometry by six ligands which include the 3'OH terminus of the nick and one of the non-bridging phosphate oxygens of the 5' terminus as well as water-mediated interactions to residues in motif I, III and IV (Figure 3C). The position of this Mn directly corresponds to the site of the catalytic metal ion in the ATP-bound pre-step 1 complexes of RNA ligases and Ecl-Lig ND-ligase (20,21). The direct Mn contact to the DNA 5' phosphate Pmar-Lig (PreS3-Mn) replaces the interaction with the AMP  $\alpha$ -phosphate seen in the pre-Step





**Figure 3.** (A) Pre step 1 configuration of the T4 RNA ligase1 with the NT domain residues shown in dark salmon, C domain residues shown in light blue and ATP cofactor as cyan. (B) Pre step 1 configuration of the *Escherichia coli* NAD<sup>+</sup> dependent DNA ligase (5tt5) with the NT domain residues shown in light pink, and Ia domain residues shown in dark blue and NAD<sup>+</sup> cofactor as cyan. In both T4 RNA ligase I and *E. coli* ND-ligase, the lysine-to-methionine mutation is shown as green sticks, the catalytic Mg ion as a magenta sphere and the structural Mg ion as a green sphere. (C) Pre-step 3 configuration of ATP-dependent DNA ligase with NT domain residues in firebrick red, AMP in cyan, 3'OH of the nicked DNA in light green and the 5'OH and of the nicked DNA in gold. The catalytic Mn ion is shown as a magenta sphere. In all structures, waters are indicated as blue spheres.

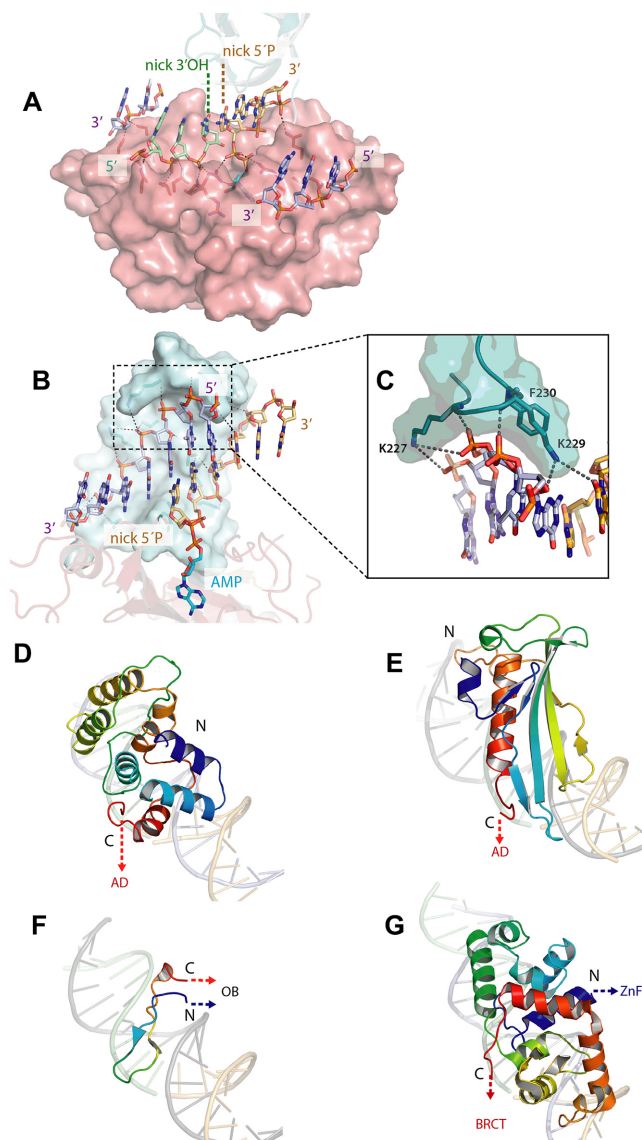
1 complexes, which is consistent with an exchange of location of the chemical transformation as the metal ion is now able to stabilize the pentacoordinate transition state of the 5' phosphate as well as decreasing the  $pK_a$  of the 3'OH nucleophile. It is also likely that the metal centre plays a structural role immediately post Step 2 facilitating bond reorientation for in-line attack.

A second structure of Pmar-Lig immediately after Step 3 catalysis captures the post-ternary complex with the ligase bound to the DNA product, and the AMP cofactor retained in the binding pocket with partial occupancy, revealing that the majority of the protein-nucleoside contacts are retained after formation of the new inter-strand phosphodiester bond (27). Release of the AMP from the binding pocket is presumably spontaneous with loss of its covalent attachment to the DNA, and residues of the vacant active site are then re-set for receipt of a new molecule of ATP upon release of the DNA product. This post-Step 3 state is seen in the enzyme-product complex of ChIV-Lig where the AMP has diffused out of the catalytic site (24). It is likely however that in solution where domain movements are not restrained by crystal packing, AMP ejection is synchronized with return of the catalytic domains to an open conformation and product release described in recent kinetics studies of the T4 DNA ligase (31).

The five recently-published structures of catalytic intermediates with occupied metal sites have been instrumental in understanding the role of divalent cations in steps of ligation; in particular distinguishing between the primary 'catalytic' ion found in all ligases and required for all chemical steps, and the second 'structural' ion required only for step 1 in ATP-utilizing ligases (19–21,27). Two additional non-catalytic Mg binding sites have been identified in the replicative Human ligase I (Hu-LigI) one of which enables discrimination of improper 3'OH termini bearing DNA damages (32). This site enforces correct recognition of 3' termini at both step 2 (DNA adenylation) and step 3 (phosphodiester bond formation) either arresting catalysis or leading to abortive 5' DNA adenylation products which are subsequently resolved by the deadenylase activity of the interaction partner Aprataxin (33).

## STRUCTURALLY-DYNAMIC MODULAR ARCHITECTURES FACILITATE DNA ENGAGEMENT

In the DNA-bound state, the NT and OB domains form a C-shaped clamp about the DNA duplex, with full encirclement often completed by the participation of appending DNA-binding domains (DB-domains) or loop regions (6,24,26,30). Such complete circumferential engagement was previously assumed to be essential for effective ligase activity (3,5), however the recent discovery of a group of AD-ligases from bacteria dubbed 'Lig E' which are able to join a range of DNA substrates including nicks, cohesive ends and mismatches has re-defined the minimal scaffold necessary for effective DNA binding and joining (23,25). Ame-Lig, an example of this group, was captured as a Step 2 intermediate engaging the DNA with partial encirclement limited to the C-shaped configuration using only its NT- and OB-domains (25).



**Figure 4.** (A) Surface of the NT domain showing the position of the bound DNA substrate. (B) Surface of the OB domain showing the position of the bound DNA substrate. In both panels, the structure is the AD ligase of *Alteromonas mediterranea* (6gdr) and the 3'OH strand of the nick is shown in green, the 5'P strand of the nick orange, the complement strand violet. (C) Details of the 'lysine fork' interactions between the OB domain of Lig E enzymes and substrate DNA that facilitate binding. (D) DNA binding domain of the AD-ligase of *Prochlorococcus marinus* (6rce). (E) DNA binding domain of the AD-ligase from African Swine fever virus (6imj). (F) DNA binding latch of the AD-ligase from *Chlorella virus* (2q2t). (G) Helix-turn-Helix DNA binding domain of the ND-ligase from (2owo).

In all DNA ligases, the nicked strand lies across the relatively flat surface of the NT domain with the nick poised over the active site, while the complement runs along the concave center of the OB domain and the nicked strand downstream of the 5'P end tracks the OB-domain periphery (Figure 4A and B). Protein-DNA interactions involve a set of conserved hydrogen bonds and electrostatic interactions about the nick provided by the NT domain, and to the nucleotides opposite the nick by the OB domain. The vast majority of contacts are to oxygens on

the phosphate backbone rather than base or sugar interactions, and many are provided by basic side chains of conserved amino acids. Ame-Lig, has the minimum DNA-binding footprint of any characterized DNA ligase, interacting with 4 bases either side of the nick, and 12 bases on the complement strand (Figure 5A and B). Within these margins, several positions lack any contacts, including a stretch of three consecutive nucleotides on the complement. Ame-Lig, and presumably other members of the LigE group, achieves tight DNA binding via series of conserved positions within the ligase core. These include a lysine ‘fork’ at the tip of the OB domain that straddles the complement strand and interacts with the nucleotides base-paired to the nick termini, as well as residues in the linker region which a swivel, presumably stabilizing open and closed conformations in the presence and absence of DNA (Figure 4C).

In larger DNA ligases, a more extensive binding footprint which often contacts all nucleotides within its margins, is provided by a DNA binding domain or a latch module, with diverse structural strategies giving similar modes of complete DNA encirclement (Figure 5A, C and D). Among AD-ligases, the archetypal DB domain is an N-terminal  $\alpha$ -helical bundle of between 7 and 14  $\alpha$ -helices arranged with a pseudo two-fold symmetry. Reverse turns in inter-helical loops form contacts with the minor groove of both strands, the majority of which involve interactions between polar main-chain atoms and the phosphate backbone (Figure 4D). Examples of the most pared-down seven-helix version of this DB domain are found in Pmar-Lig and the T4 DNA ligase (T4-Lig) which have extremely high structural homology (27,29). There are few clamp closing contacts in either of these enzymes with only a pair of salt bridges observed between the DB- and OB domains in the recently-published structure of T4-Lig (29) and a dispensable hydrogen bond interaction between these domains in Pmar-Lig (27). Larger versions of the helical DB domain are found in the mammalian and archaeal DNA ligases as well as many bacterial AD-ligases and a few viral ligases (22,26,30,34–37). The DB domains of these enzymes share the central seven-helix bundle of Pmar-Lig and T4, but have additional elements adorning their outer margins which form more extensive contacts with DNA as well as the NT and OB domains. The distribution of these  $\alpha$ -helical DB domains among DNA ligases in all kingdoms of life suggests that this composition represents the ancestral form of AD-ligases; however other DNA ligases exhibit structurally unrelated modes of encirclement. The recently-described AD-ligase of African Swine Fever Virus (AsfV-Lig), for example, has a unique N-terminal DB domain with a mixed  $\alpha/\beta$  fold (Figure 4E) (28). This domain comprises a central pair of extended antiparallel  $\beta$ -strands, stacked against a long  $\alpha$ -helix which lies along the major groove interacting with the nick 3’OH strand. Shorter secondary structural elements and loops flank these, providing further interaction sites with the complement. The ChIV-Lig possesses the simplest solution to clamp closing, having a  $\beta$ -hairpin loop that extends from the turn between antiparallel  $\beta$ -strands of the OB domain (Figure 4F) (24). This feature, which is disordered in the absence of DNA becomes structured upon binding, in-

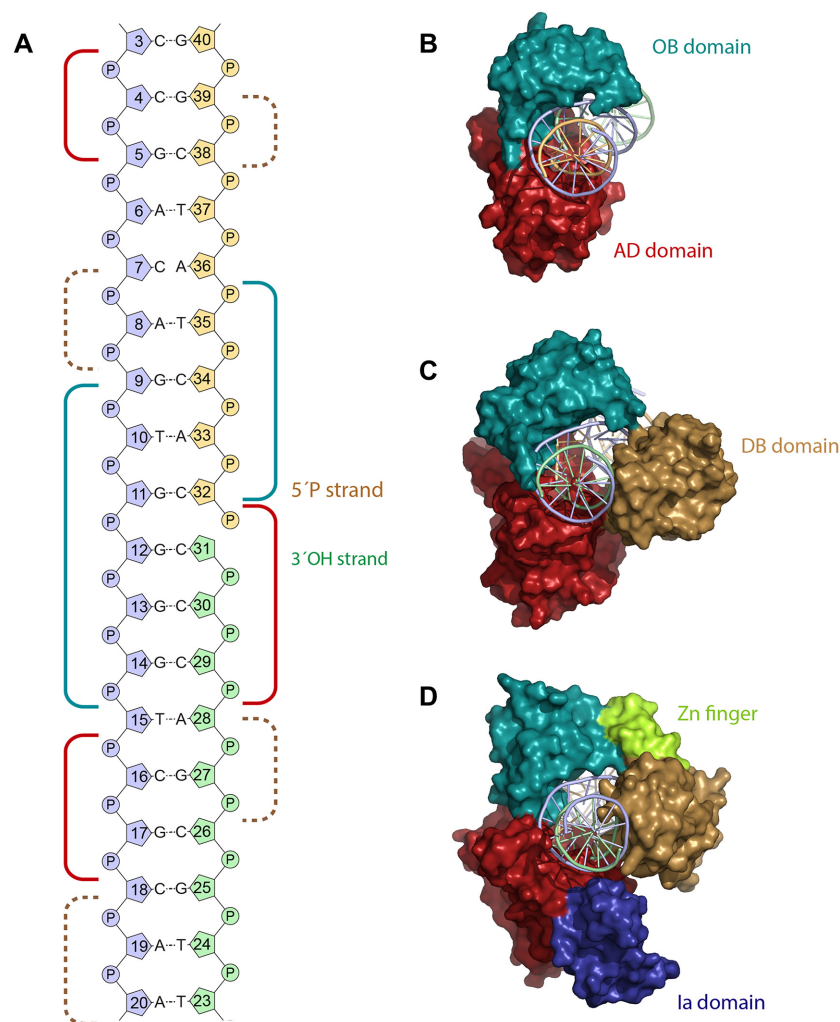
serting into the major groove and forming contacts with the complement strand downstream of the nick 5’ on the DNA as well as kissing contacts with the NT domain.

ND-ligases from various species, unlike their ATP-dependent counterparts, are highly conserved with a common pattern of domains; the N-terminal most Ia domain which is not involved in DNA interaction followed by the NT- and OB domains of the core (Figure 5D). The helix-hairpin-helix (HhH) domain which provides essential DNA binding functions is C-terminal to the OB domain, followed by a zinc finger domain (Zn finger) and BRCA1 C-terminal domain (BRCT). The HhH domain consists of a pair of helix-hairpin-helix motifs symmetrically arranged, and in the closed conformation is analogous in position to the globular DB domains of AD ligases, despite their differing folds and connectivity (Figure 4G) (6). The Zn finger domain functions to position the HhH relative to the OB fold, and clamp-closing interactions are made between the NT- and HhH domains. The BRCT domain has not been resolved in any complete ND-ligase structures to date, however deletion studies indicate it enhances binding suggesting it may function in nick sensing or stabilize the bound conformation indirectly (38,39). As well as ND-ligases, BRCT domains are also found in the AD-ligases Human ligase III and IV (Hu-LigIII and Hu-LigIV).

For those DNA ligases that bind their substrate with complete encirclement, the DB domain is essential for DNA interaction, for example deletion of the N-terminal domains of Asf-Lig, Hu-LigI and T4-Lig almost completely abolished binding and activity (26,28,40). In ChIV-Lig, deletion of the 30 residue DNA-binding latch decreases activity 10-fold and imparts significant salt sensitivity (24). Likewise, point mutations of individual DNA-binding residues embedded within the core scaffold of Ame-Lig significantly impacted interaction, and a quadruple mutant of four key positions in the OB domain that contact the DNA was entirely inactive, despite retaining equivalent secondary structure and thermal stability to the wild-type enzyme (25).

Recently-reported structures of non-catalytic ligase-DNA complexes where the protein is in an extended conformation provide considerable insight into nick sensing and structural transitions on the pathway to complete encirclement (Figure 6). Comparison of DNA-free and DNA-bound extended states of Hu-LigIV demonstrate that initial binding requires minimal rearrangement of the DB- and NT domains, which appear to move as a ridged unit (34). Transition to a fully-closed form is then completed by the almost 180° swivel of the OB domain to position its concave basic surface over the DNA, concomitant with restructuring of the linker region (Figure 6A). Two disordered loops in the NT domain become ordered upon binding, one of which forms bridging interactions with the DB domain. By contrast the extended-state complex of AsfV-Lig with DNA shows interactions with only the DB domain, and significant repositioning of the NT domain is required to orient it for nick binding (Figure 6B). In this extended conformation, the substrate DNA around the nick is in the B-form, and it is not until participation of the NT- and OB domains in binding that the nick residues are distorted to the A-form necessary for catalysis (28). As with other DNA





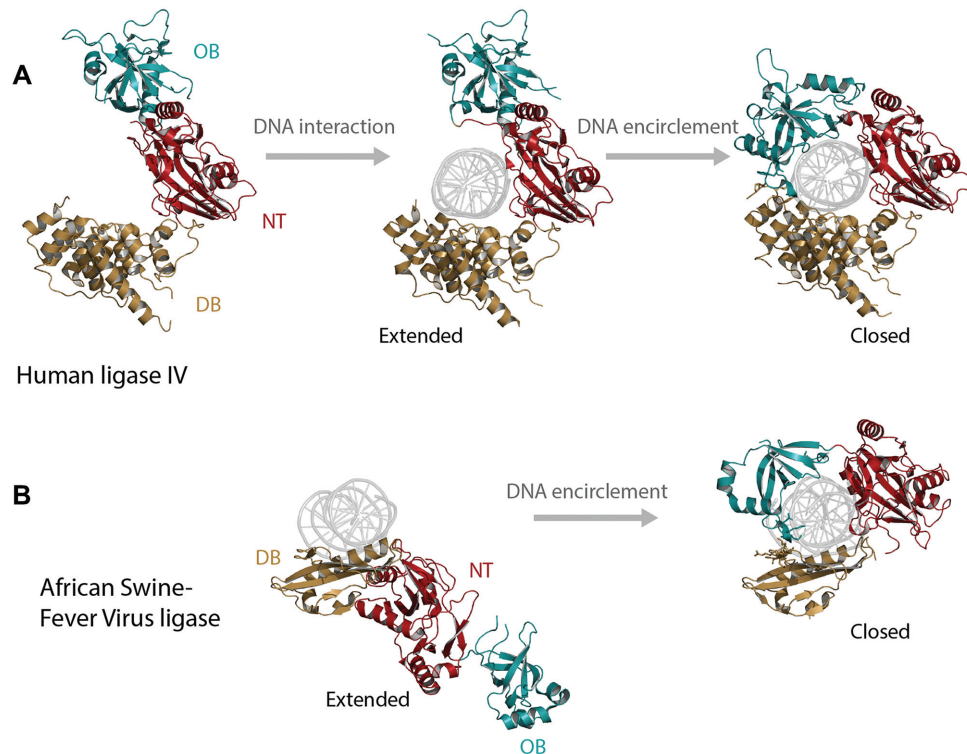
**Figure 5.** (A) DNA binding footprint of DNA ligases on a 21 base-pair double-stranded substrate commonly used in co-crystallization experiments. Solid lines indicate interactions between the core catalytic domains NT domain (red) and OB domain (cyan). Dashed lines indicate the common position of DNA binding domains or latches. (B) Lig-E bacterial AD ligase bound to DNA (*Alteromonas mediterranea*, 6gdr). (C) AD ligase with a typical N-terminal  $\alpha$ -helical DB domain bound to DNA (*Prochlorococcus marinus*, 6rce). (D) ND ligase bound to DNA (*Escherichia coli* 2owo).

ligases, AsfV-Lig  $\Delta$ OB mutants have residual, if diminished binding capacity relative to the wild type, but are unable to support catalysis, likely due to the central role of the OB domain in enforcing duplex distortion (28).

Although no partially-bound complexes of smaller DNA ligases such as ChIV-lig and Ame-Lig in extended complexes with substrate DNA are available, comparison of DNA-free apo-enzyme-adenylate structures (open conformation) with fully DNA-bound (closed conformations) has proved informative. In the case of ChIV-lig, many structural and functional studies have demonstrated the intrinsic nick sensing function of the OB-domain latch (17,24). Comparison of DNA-bound Ame-Lig with structures of close homologs in the absence of DNA implicate a transition between different relative orientations of the NT and OB domains, stabilized by specific interactions between the linker and these domains, in facilitating productive binding (25).

As discussed below, genome sequencing endeavors point to as-yet uncharacterized diversity among DNA ligases, which judging by the variety of structures to date, may yield even more unique forms. This range of conformations for DNA-bound intermediates suggests different binding scenarios may exist between different ligases, with some such as Hu-LigIV proceeding via concerted interaction of the DB-NT unit followed by re-orientation of the OB domain, while the AsfV-Lig DB domain alone appears able to initiate interaction. This is of interest as kinetics studies indicate that domain rearrangement during substrate binding and/or product release represent the rate limiting steps in DNA ligation rather than the catalytic steps themselves (31). It is likely, given the conservation of the catalytic steps, that it is this diversity in binding scenarios which gives rise to the range of rate constants measured for different DNA ligases, and therefore has implications for engineering variants with improved reaction kinetics for specific applications.





**Figure 6.** Structural snapshots of interactions between DNA ligase and substrate. (A) Human Ligase IV without DNA (3w5o), bound to DNA in an extended conformation (6bkf) and in a closed conformation (6bkg). (B) African swine fever virus in an open protein-DNA complex (6imj) and closed protein-DNA complex (6imn).

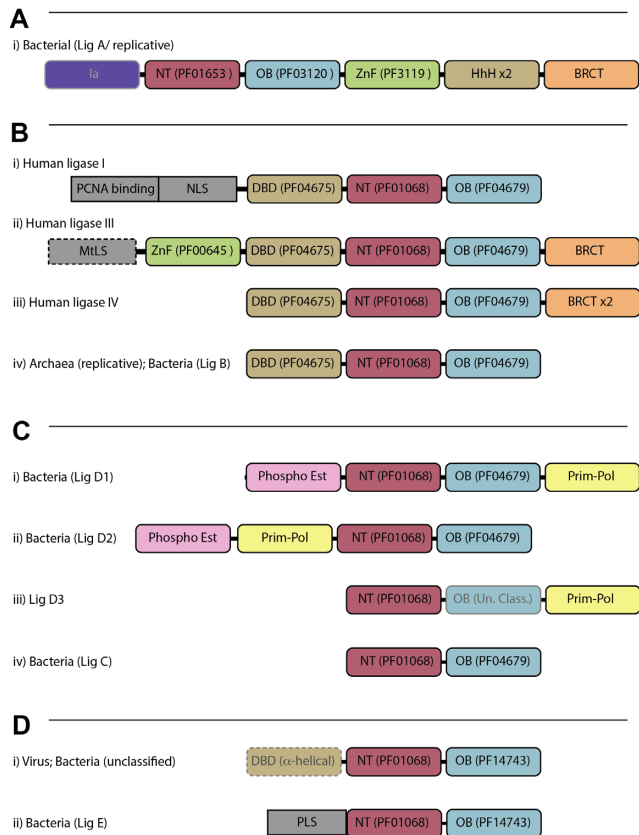
## FUNCTIONS ARE DEFINED BY CATALYTIC MODULES AND TARGETING SEQUENCES

Most DNA ligases characterized to date function as part of multi-step DNA modifying processes to elicit DNA replication and repair which often requires processing of DNA ends by auxiliary enzymes, or recruitment to the site of activity by additional factors. These additional functions may be co-localized with the ligase component in an operonic structure, or fused in same polypeptide expressed as large multi-domain enzymes with multiple modules having independent catalytic function (Figure 7). The best-known examples of such multifunctional ligases are the Lig D proteins which carry out non-homologous end joining in some bacteria during stationary phase. This group, which have been described extensively in previous reviews (8,9,41), possess a primase/polymerase (PrimPol) module either N- or C-terminal to the ligase core (Figure 7C, i-iii, yellow). A phosphoesterase (PE) module is also fused (Figure 7C, I and ii, pink), or may be found as a separate component adjacent to the ligase. Structures of these modules have now been determined, in some cases with bound substrate providing insight into their modes of action (42,43). Multifunctional Lig D acts on double-stranded breaks, often removing nucleotides through its phosphodiesterase activity, or adding nucleotides through its polymerase function (Figure 8D). The final ligation step is carried out by the ligase core, with the resulting product often being mutagenic due the activities of PrimPol and PE at the break site. Effective ligase activity relies on recruitment by the Ku end-binding protein

which synapses the ends of the break, and joining of products containing 3' ribonucleotides is preferable as these are the natural product of the polymerase function (44).

The unique combinations resulting from fusion of independent catalytic domains, such as seen in the Lig D ligases, combined with the presence of different DB domains and localization sequences have defined multiple classes of DNA ligase with distinct functions (Figure 7). The phylogenetic distribution of these groups is described in detail in the subsequent section. While some organisms have a single ligase that is responsible for both replication and repair, many have multiple ligases with dedicated biological roles. In Mammalian cells for example, Lig I (Figure 7B i) is responsible for sealing Okazaki fragments and long-patch base-excision repair (BER), Lig III (Figure 7B ii) carries out short-patch BER, single-strand break repair and mitochondrial DNA maintenance and Lig IV (Figure 7B iii) participates in classical non homologous end joining (NHEJ) of double-stranded breaks and V(D)J recombination in immunoglobulin gene maturation. The division of labor of these eukaryotic proteins and their suite of interaction partners has been reviewed previously (2,45). In the intervening years, knock outs using CRISPR-Cas9 genome editing revealed a surprising level functional redundancy where Lig IV-ablated mouse cell lines either Lig I or Lig III can support some extent of NHEJ via an alternative end joining pathway (46,47).

The diversity and non-ubiquitous distribution of AD-ligases among bacteria is particularly notable with some species possessing as many as five ligase-encoding genes,



**Figure 7.** Schematic of domain arrangements in major classes of DNA ligases characterized to date. (A) NAD-dependent DNA ligases, primarily found in bacteria. (B) ATP-dependent DNA ligases with N-terminal  $\alpha$ -helical DB domains and a common OB domain type (PF04679). (C) ATP-dependent DNA ligases with OB domain type (PF04679) and no DB domains; Lig-D type non-homologous end joining proteins with auxiliary PrimPol domains and Lig C with no auxiliary domains. (D) ATP-dependent DNA ligases with a common OB domain type (PF14734) including viral and bacterial forms that possess or lack an N-terminal  $\alpha$ -helical DB domains, and the bacterial Lig E proteins that have a periplasmic localization sequence (PLS). Rounded boxes indicate folded protein domains, square boxes indicate targeting sequences: PCNA binding (proliferating nuclear antigen binding), NLS (nuclear localization signal), MtLS (mitochondrial localization signal). Grey text indicates domains not recognized/assigned by Pfam.

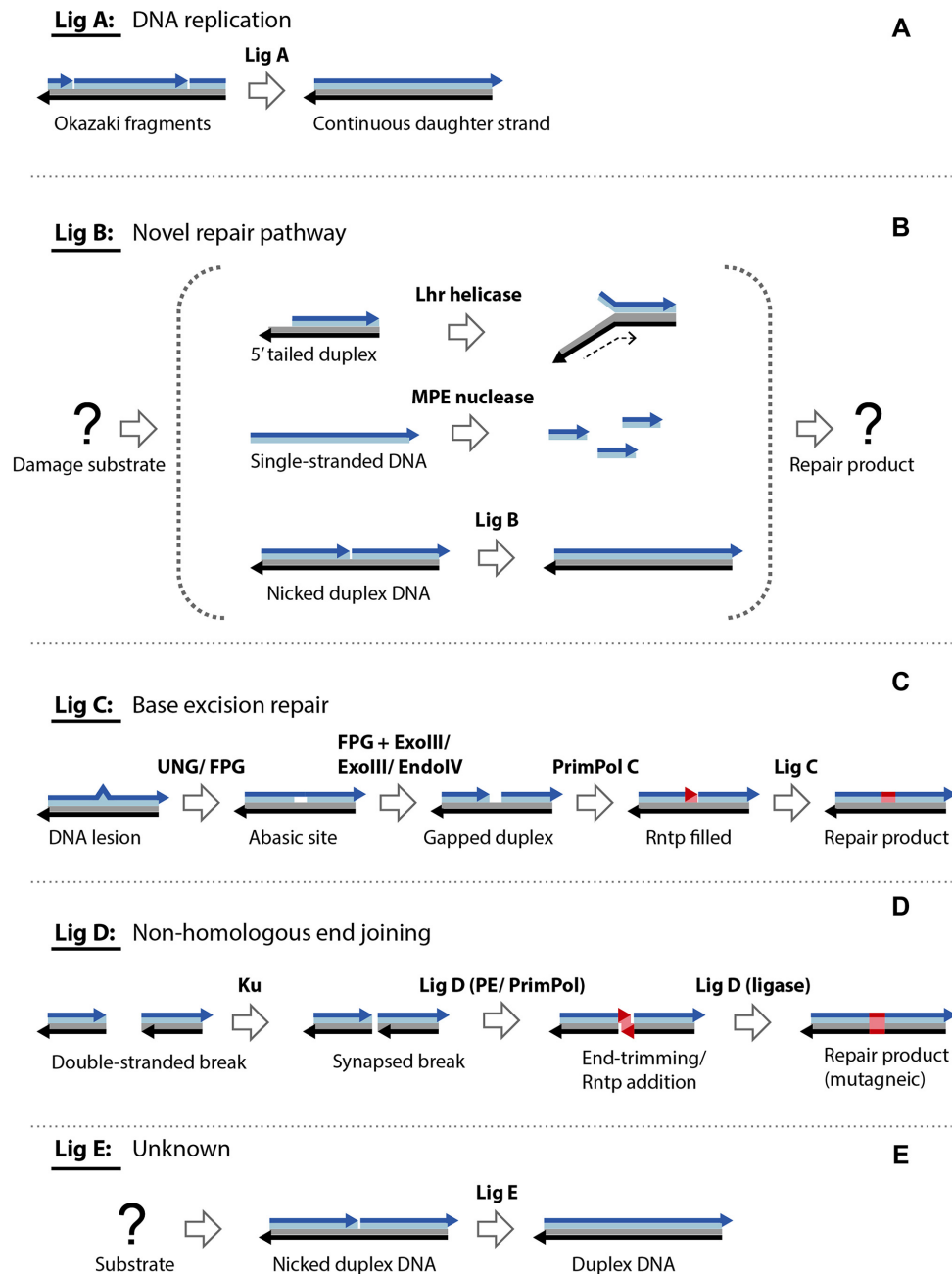
while other, sometimes closely related species harbor none (10). In all cases, these AD-ligases are found in addition to the replicative ND-ligase (Figure 7A, i). So far, four distinct classifications of bacterial AD-ligase isoforms have been delineated based on their structural and functional characteristics: the large Lig D PE-PrimPol-ligase enzymes of NHEJ described above (Figure 7C, i–iii); the Lig B group that closely resemble archaeal replicative ligases with a canonical globular  $\alpha$ -helical domain preceding the NT and OB core domains (Figure 7B, iv), and two groups of minimal AD-ligases, Lig C (Figure 7C, iv) and Lig E (Figure 7D, ii) which include only the catalytic ligase core.

Until recently, the Lig C ligases were thought to act as a ‘backup’ NHEJ enzyme to the primary Lig D-based pathway, and genomic deletions studies indicate that they can substitute to some capacity (48). However, an extensive

structural and functional study by the Doherty group has recently established that the primary role of Lig C is to provide dedicated ligase function in the base-excision repair pathway during stationary phase (Figure 8C) (49). *In vitro* and *in vivo* interaction studies of Mycobacterial Lig C revealed that it interacts with an operonically-associated PrimPolC enzyme, as well as DNA glycosylases NTH, IPG and MPG, and nucleases EndoIV, ExoIII and XthA, forming a hub for lesion processing. PrimPolC, which is a specific Lig C-associated isoform distinct from the NHEJ PrimPol, binds preferentially to gapped substrates which are generated by lesion removal and abasic site processing of DNA glycosylases UNG or FPG together with end-processing nucleases XthA, ExoIII or EndoIV. PrimPolC proceeds to fill the gap, preferentially with ribonucleotides, generating a nicked double-stranded DNA with an RNA 3’OH terminus, which is the preferred substrate for Lig C (50).

Unlike the Lig C and D ligases, both Lig B and Lig E are able to effect high rates of nick sealing in the absence of any accessory enzymes (23,25,51–53). Lig B ligases appear in gene clusters with a novel Lhr-helicase, binuclear metallophosphoesterase (MPE) and putative exonuclease. Lhr is an ATP-dependent 3’ to 5’ helicase which unwinds DNA–DNA and DNA–RNA duplexes preceded by a single-stranded loading segment (54). MPE is a Mn dependent single-strand endonuclease that cleaves both linear and loops of stem-and-loop structures (55). The exonuclease member of the cluster has not yet been biochemically characterized, but it is predicted to be a homolog of the SNM1B/Apollo nuclease which repairs inter-strand crosslinks (54). Although the precise biological substrate and order of activity have not yet been determined, it is likely that these enzymes represent yet another distinct repair pathway in bacteria (Figure 8B).

One of the most enigmatic of the AD-ligases identified to date in bacterial genomes, is the Lig E variant. These minimal ligases which are found in a wide range of Gram-negative bacteria possess a predicted N-terminal signal sequence that targets them to the periplasm, which is intriguing given that both DNA replication and repair are intracellular processes (Figure 8E). Removal of this sequence increases that stability and activity of the enzyme, and both crystal structures of Lig E were obtained by variants where this leader had been truncated during cloning (23,25,53). Investigation of the genomic context of Lig E from various species has failed to reveal a common synteny in gene organization as seen for other bacterial AD ligases, however examination of Lig E-containing genomes reveal that most encode the ComEA proteins for uptake of extracellular DNA and many of these species have been independently shown to be naturally transformable (56), suggesting a possible role for Lig E in competence of Gram negative bacteria. Lig E is unlikely to be essential for DNA uptake *per se*, given that competence is also observed in Gram positive bacteria, but we propose it functions in situations where internalization of long tracts of DNA is desirable. By repairing nicks in double-stranded DNA in the periplasmic space, Lig E would allow longer contiguous sequences to be internalized, given the DNA is rendered single-stranded during translocation across the cytoplasmic membrane (57).



**Figure 8.** DNA modification pathways in bacteria involving DNA ligases. (A) Joining of Okazaki fragments during DNA replication by ND-ligase Lig A. (B) Probable repair pathway involving components of an operon including the AD-ligase Lig B. (C) Participation of AD-ligase Lig C in base-excision repair during stationary-phase. DNA lesions such as deoxyuracil or 8-oxoGuanine are removed by UDG glycosylase or the bifunctional glycosylase FPG. FPG together with 3'phosphatase and/or exonuclease activities generates a gapped duplex which is filled by the primase-polymerase activity of PrimPol using ribonucleotides. This substrate is ligated by Lig C to give an RNA/DNA duplex. (D) Rejoining of double-stranded DNA breaks by multifunctional AD-ligase Lig D. Break ends are synapsed by the Ku end binding protein and the break site is processed through trimming by the Phosphoesterase module and/ or addition of ribonucleotides by the Lig D primase-polymerase module. Resulting ligatable ends are joined by the ligase module often giving a mutated product. (E) The biological function of the AD-ligase Lig E is yet to be determined, however due to its preferential nick-sealing activity, it most likely acts on single-strand nicks in duplex DNA.



## GENOME SEQUENCES REVEAL PHYLOGENETIC DISTRIBUTIONS AND UNIQUE SEQUENCES

The ever-increasing number of genomic sequences, and advances in bioinformatic tools are providing new insight into the diversity and distribution of DNA ligases among both cellular and viral organisms. For example, since the discovery of accessory AD-ligases in the genomes of a few sequenced bacteria (11,58,59), we are now aware that almost half of bacterial phyla possess one or more of these enzymes, and that many species harbor a range of isoforms with different structures and functional roles (10). Analysis of the correlation between taxonomic distribution of these isoforms and habitat specialization may provide further insight into their biological functions. For example the Lig C and Lig D AD-ligases are particularly prevalent among genera such as *Mycobacteria*, *Streptomyces* and *Bacillus* that are subject to periods of desiccation and known to have dormant stages in their lifecycles (9). The putatively periplasmic Lig Es by contrast are mutually exclusive with other AD-ligase types, and are widely distributed among naturally transformable  $\beta$ ,  $\epsilon$  and  $\gamma$ -proteobacteria (10).

To survey the sequence diversity of DNA ligases presently available in public databases, we used the Enzyme Function Initiative's Enzyme Similarity Tool (EFI-EST) to generate Sequence Similarity Networks (SSNs) for protein sequences including the catalytic NT domain of either the AD-ligases (Pfam 01068) or the ND-ligases (Pfam 01653). EFI-EST computes pairwise BLAST scores for all sequences in the set, creating a network of nodes (sequences) connected by edges (BLAST scores) (60). Within the network, clusters of similar sequences were defined by setting a threshold values where edges are retained as 100 (AD-ligases) and 200 (ND-ligases). To decrease the file size, renode networks were downloaded where sequences with greater than 55% identity in aligned regions (AD ligases) or 65% identity (ND ligases) are represented by a single node. SSNs provide a feasible method to analyze sets of evolutionarily related proteins which are too large or diverse for traditional multiple sequence alignment methods to be practically applied. While not as robust for interpreting the evolutionary history of the sequences as a phylogenetic tree, these networks can provide considerable insight into the diversity and similarity of proteins within a given family and have been used successfully in functional assignment of proteins, discovery of novel functions and mapping the evolution of diverse functions from ancestral scaffolds (61–64).

The AD-ligases formed 11 major clusters (>100 nodes in each) and four additional cluster (>35 nodes). More than half of the total dataset found in clusters #1 and #2. To better visualize the taxonomic distribution of these enzymes, the SSNs were colored by super kingdom (Figure 9A), or by different ligase types defined by their Pfam domain composition types as described in the preceding section (Figure 7). Overall, bacterial AD-ligases constitute the majority of the dataset (62%) followed by Eukaryotes (28%) with Archaea and Virus sequences contributing only 7% and 3% respectively. The predominance of bacterial representatives is likely biased due to the greater number of sequenced bacterial genomes compared to other organisms; however, it

serves highlight the importance and widespread prevalence of these 'non-essential' enzymes.

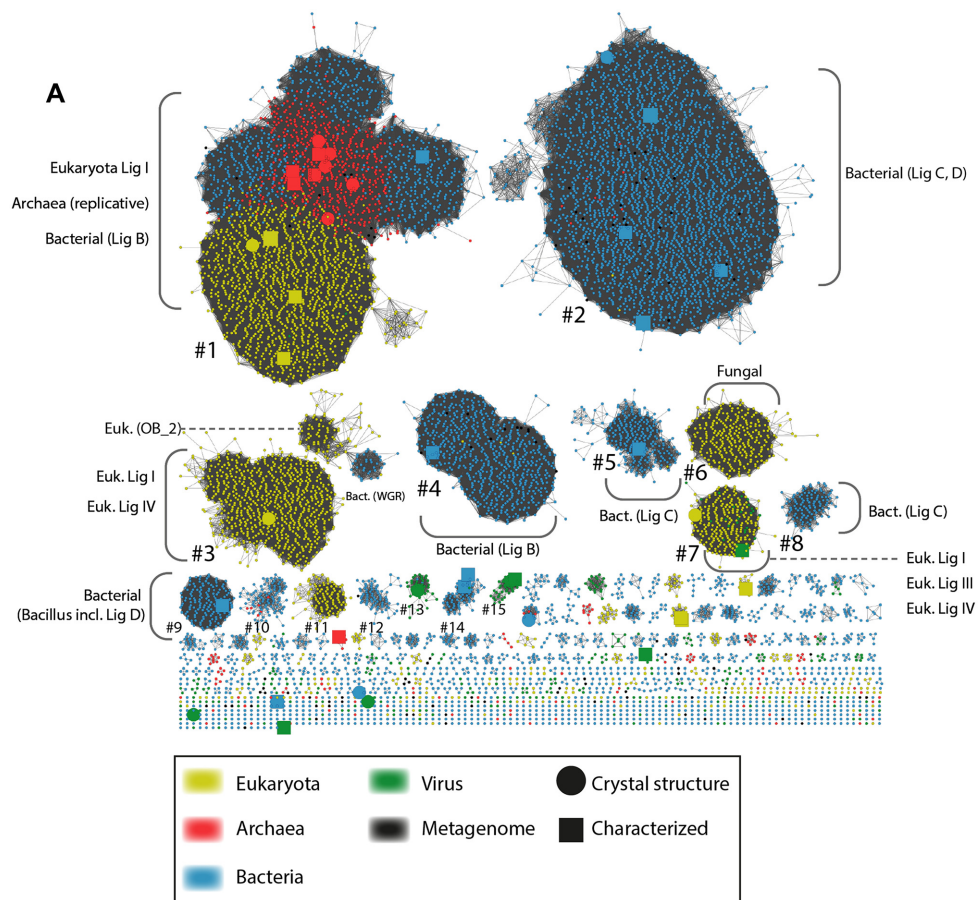
One of the most salient findings of this SSN analysis is that AD-ligases of cellular organisms which possess common complements of protein domains (as described in Figure 7) form common clusters. These clusters form four categories of AD-ligase summarized in Figure 9B, which are defined by their domain configurations, and correlate with their known biochemical activities.

In the first group, are the eukaryotic, archaeal and bacterial Lig B ligases which populate clusters #1, #3, #4 and #7. These share a common central DB-NT-OB arrangement, and all characterized members of these groups possess autonomous ligase activities that do not require additional scaffolding proteins. The taxonomically-diverse Cluster #1 is especially interesting, as it includes all archaeal replicative proteins in the dataset together with the majority of eukaryotic ligase I enzymes and approximately half the bacterial Lig B representatives. This common clustering indicates a shared evolutionary history between these proteins. In the case of the archaeal and eukaryotic (Lig-I) replicative ligases this reflects a broader pattern of conservation in their replicative machinery and is consistent with the evolution of the eukaryotic components from an archaeal-type ancestor (65,66). The co-clustering of many bacterial Lig B proteins, predominantly from *Actinobacteria*, *Acidobacteria* and *Chloroflexi* supports the previously-articulated suggestion that these accessory ligases have been horizontally-acquired from archaea (10). A second Lig B cluster #4 is dominated by Proteobacteria, which is consistent with the previous observation that not all bacterial Lig B proteins are monophyletic and may have arisen from multiple acquisition events (10). The independent clustering of eukaryotic isoforms Lig III and Lig IV (#3 and #7) is consistent with the presence of additional domains in these isoforms; although it is not entirely clear why they are split between two groups rather than a single cluster.

The second category are the modular multi-functional Bacterial Lig D enzymes that possess a Prim/pol domain and require the Ku protein for ligase activity. The majority of these are found in the entirely bacterial cluster #2 which contains almost half of the bacterial sequences in the dataset including those from *M. tuberculosis*, *Pseudomonas aeruginosa* and *Agrobacterium tumefaciens*. The Bacilli Lig D group including the characterized *Bacillus subtilis* NHEJ ligase is found separately in the smaller cluster #9, which is again in line with previous phylogenetic studies that indicate Lig D enzymes form species-specific clades (10).

A third category are the Bacterial Lig C which have only core domains and also require Ku for activity. These predominantly occupy independent clusters #5, #8, #10 #12 and #14, however a small number of Lig C sequences are also present in clusters #1, #2 and #4. While some of these may represent truncations or sequencing artifacts of larger ligases, others such as the *M. tuberculosis* Lig C are *bona fide* enzymes participating in biological processes (49).

The final and least-cohesive category includes bacterial and eukaryotic ligases that have the OB-2 variant of the OB domain (PF14743) which as described previously lacks several of the helical elements found in the Lig B, Lig C and Lig D OB domains (PF04679) (10). Bacterial AD-



**B**

Domain configuration	Ligase type	Cluster
	Eukaryote Lig I Eukaryote Lig III Eukaryote Lig IV Archaea Bacteria Lig B	#1, #3, #7 #7 #3, #7 #1 #1, #4
	Bacteria Lig D	#2, #9
	Bacteria Lig C	#5, #8, #10, #12, #14
	Bacteria Lig E/ OB2 unclassified Eukaryote unclassified	#3*, un-clustered #3*, #6
	Virus	#13, #15, un-clustered

\* Sub-cluster of #3

**Figure 9.** (A) Sequence similarity network (SSN) of ATP-dependent DNA ligases colored by super kingdom. Details of SSN generation are given in the caption of Supplementary Figure S3. (B) Cluster locations of AD-ligases with common domain complements. For the schematic, domains with solid outline/black text are conserved in the configuration shown, while domains with dashed outline/grey text are not present in all instances, and in the case of PrimPol may vary in position relative to the catalytic core.

ligases with the NT-OB-2 combination include the putatively periplasmic Lig E class as well as larger ligases where the appending DNA-binding domain does not have a recognized Pfam annotation; for example, the previously-characterized AD-ligase Pmar-Lig of *P. marinus* (27). The largest clustered group of bacterial NT-OB-2 proteins are a sub-cluster within cluster #3, and includes representatives from  $\delta$ -proteobacteria and Planctomycetia. More than two thirds of these sequences have an annotated WGR domain, which in other proteins functions in RNA binding and is often coupled to the catalytic domain of Poly(ADP-ribose) polymerase (PARP) DNA-repair regulators (67,68). This sub-cluster of #3 also contains a group of mostly fungal eukaryotic enzymes without an annotated DB domain, which intriguingly are connected to the bacterial NT-OB-2-WGR cluster and share the same NT-OB-2 catalytic core. Aside from this sub-cluster, all other bacterial NT-OB-2 sequences are single nodes or in small groups of <30 nodes. Due to the size cutoff applied during SSN generation (Supplementary Figure S3, caption) most Lig E proteins were excluded from the SSN dataset, however Ame-Lig from *A. mediterranea* is found as essentially a singleton sequence, grouped with only other *Ateromonas* ligases, supporting the distinct lineage that the Lig E group seems to occupy. Pmar-Lig is found in a small cluster of 20 sequences, mostly from proteobacterial and cyanobacterial isolates, again reflecting its unique structure among bacterial AD-ligases. Pmar-Lig is one of up to three accessory AD-ligases found in the genomes of *P. marinus* isolates. The Lig B form of *P. marinus* is found together with Proteobacterial forms in cluster #4, while the third AD-ligase, which includes an N-terminal WGR domain is found as a single node. As previously noted, the absence of Ku and other NHEJ components in these genomes suggested the *P. marinus* AD-ligases may perform as-yet undescribed DNA repair functions (10). Also included in this category are Eukaryote-containing clusters #6 and #11 which are exclusively fungal, comprising entirely Ascomycota and Basidiomycota respectively. Most of the sequences in these clusters exceed 600 residues in length indicating they are not fragments and many exceed 1000 residues despite having only the NT and DB domains recognized by Pfam, suggesting that novel functional modules may be present.

In contrast to the large and relatively consistent groupings formed by cellular ligases, >70% of viral sequences are not found in clusters. Two small Virus-only groups, clusters #13 and #15 contain sequences of bacteriophage from *Acinetobacter*, *Yersinia* and several related Enterobacteriaceae, including *Enterobacteria* phages T3, T4 and T7 (18,29,69). A small number of mammalian viral ligases including the *Vaccinia* virus enzyme and related pox viruses are clustered with Mammalian ligases in cluster #7 (70). *Chlorella* Virus and African Swine Fever Virus are both single nodes.

A synopsis of the distribution of bacteria, archaea, eukaryotes and virus among clusters is given in Supplementary Table S1, and a full list of the SSN position of characterized DNA ligases is given in Supplementary Table S2. By mapping structurally and functionally characterized AD ligases onto this SSN, it is clear that while some clusters contain many well-studied representatives, key structural and

biochemical features of other clusters such as are virtually unknown; for example, the fungal clusters #6 and #11 as well as many small bacterial clusters that do not fit any of the presently-classified AD-ligase types based on domain composition. As described below, such novel groups may contain a wealth of medically and biotechnologically important information.

Compared to the AD-ligases, an SSN built from ND-ligases is more homogenous, overwhelmingly comprising bacterial representatives with most sequences grouped into single large cluster (Supplementary Figure S4). Two minor clusters include ND-ligases from Euryarchaeota, the first of which (108 nodes) comprises entirely Halobacteria including *Haloferax volcanii* which seems to have acquired an ND-ligase by lateral gene transfer from a bacterium and uses this as a backup in the case of AD-ligase inactivation, or genotoxic stress (71). The second Archaeal group (64 nodes) includes uncultured isolates and candidate Euryarchaeota. The small number of Viral ND-ligases are single nodes or small groups with fewer than three members. Among the smaller bacterial clusters, a significant group (cluster #2) includes entirely candidate phyla, while two others (clusters #3 and #4) are firmicutes.

Despite the wide range of DNA ligase isoforms and specialized functions described here, these proteins have a common evolutionary origin which is also shared with other members of the nucleotidyl transferase family, including the mRNA capping enzymes and RNA ligases (3,4). These features include the common fold of the NT domain and conservation of carboxylate and basic residues in the active site. Several lines of evidence indicate ND-ligases arose from an ATP-utilizing AD-ligase ancestor including (i) use of ribonucleoside tri-phosphate cofactors by all other members of the superfamily (ATP by the RNA ligases and GTP by the capping enzymes); (ii) the presence of the motif IV necessary for orientation of the triphosphate leaving group during step 1 catalysis in all superfamily members except the ND ligases and (iii) the extremely high structural conservation and limited taxonomic distribution among the ND-ligases (3–5).

## APPLICATIONS AND FUTURE PERSPECTIVES

DNA ligases, in particular the T4 DNA ligase, are one of the foundational enzymatic tools of molecular biology providing the ‘glue’ to introduce foreign genes into DNA backbones during traditional restriction-ligation cloning for production of recombinant DNA (13). In recent times this activity has been used for addition of adapters to input DNA for generation for Next Generation Sequencing libraries, and DNA ligases have also been employed as a key component of the sequencing technology itself (1,72,73). All DNA ligases catalyze some extent of joining on correctly base-paired nicked double-stranded DNA, however their ability to tolerate base-pair mismatches at the nick site varies between different DNA ligases. T4 DNA ligase, for example, is able to join nicks with mismatched bases at either the 5' or the 3' terminus (74). In contrast, the thermophilic ND-ligase from *Thermus aquaticus* exhibits 10–100-fold higher fidelity than T4 on substrates with base mismatches being more discriminating at the 3' end, and



has been exploited in applications such as SNP detection (75,76). The pursuit of thermophilic ligases for high fidelity applications inspired structure-guided engineering efforts of the *Pyrococcus furiosus* AD-ligase, and has been the subject of recent reviews (77–79), while conditions eliciting improved fidelity in T4 and *T. thermophilus* ligases have been identified, improving their efficacy for gene assembly and SNP-detection protocols (80,81). At the other end of the temperature spectrum, recent studies have characterized AD-ligases from psychrophilic bacteria, which may find utility in conditions where low temperature, for example to preserve same integrity, are desirable (52).

A smaller number of DNA ligases are able to act on diverse substrates in addition to nicked DNA duplexes including double-stranded DNA breaks and RNA-DNA hybrids. T4 DNA ligase has the most diverse repertoire of activities, ligating blunt-ended breaks and gapped nicks in double-stranded DNA as well as single-stranded DNA and DNA-RNA hybrid duplexes provided the 3'OH strand is DNA (74,82–84). These activities are exploited in molecular cloning of DNA fragments with either double-stranded 'blunt' ends, or short cohesive overhangs, as well as generation of sequencing libraries by attaching adapter molecules to the sample DNA (1). The ChIV-Lig is able to join single strands of DNA that are splinted by an RNA complement, and is employed as the SplintR ligase in micro RNA detection (85).

The interest in discovering DNA ligases that exhibit novel joining activities or activity optima has in part been driven by the commercial potential of these enzymes. Meanwhile, enzyme engineering endeavors have been undertaken to further tailor the activities of characterized DNA ligases to their biotechnological applications, such as the improvement of double-strand break joining imparted to the T4 DNA ligase by fusion with additional DNA binding domains. As demonstrated by our SSN-based survey of DNA ligases, a wealth of diversity remains to be explored; particularly among the AD-ligases, which are already the most widely-used type in molecular biology. The ever-increasing number of high-resolution crystal structures of DNA ligases, especially in complex with substrate, cofactor and/or metal provides further opportunities for structure-guided engineering to tailor enzyme activities to suit specific applications. This has recently been applied to generate mutants of ChIV-Lig that have increased capacity to ligate DNA duplexes containing xenobiotic nucleic acids. This approach combined docking and molecular dynamics to model substrate binding, and successfully predicted insertion of a glycine into the linker would increase interaction with the artificial substrate (86). The recently-published structure of T4 DNA ligase, which has been the workhorse of DNA joining in molecular biology for almost 50 years, is a particularly important achievement that will doubtless inform structure-guided engineering endeavors in the future (29).

Novel DNA ligases also have potential in CRISPR Cas-based genome-editing biotechnologies by facilitating double-strand break repair. The editing outcome of CRISPR depends on the repair mechanism used by the cell to resolve the Cas 9-induced double-stranded break; in Eukaryotes, direct rejoining by the low-fidelity NHEJ system generates insertions or deletions at the junction cre-

ating gene knock outs, while high-fidelity repair via the homology directed repair pathway can insert exogenously provided segments of DNA with homologous ends (87). The utility of CRISPR Cas 9 editing in bacterial genomes has been limited in part by the lack of ubiquitous NHEJ systems which often renders these double-strand breaks lethal (88). Cross-complementation with Lig D and Ku components have been employed successfully to generate knock-outs in some bacterial strains, and supply of the Lig D component alone has enabled high-efficiency editing in *Streptomyces coelicolor* which possesses a partial NHEJ pathway (89,90). Characterization of NHEJ systems from other species has potential to provide further options for such cross-complementation approaches which could be tailored to the organism in question, while a detailed understanding of the factors influencing indel creation would optimize the frequency of knock-out.

Another area in which DNA ligases are of key importance is the fight against bacterial diseases. Due to their essential function in bacterial DNA replication and absence in eukaryotes, ND-ligases present an attractive target for antibacterial drugs which are highly effective and specific (91). This research, which was the subject of a recent review (92) has identified compounds targeting a hydrophobic tunnel specific to ND ligases, and obtained structures of ND ligases co-crystallized with bound inhibitors. At present C-2 substituted adenosine derivatives provide one of the most promising avenues for inhibitor design, however  $\beta$ -NAD<sup>+</sup> derivatives also hold considerable potential (93). Our increasing insight into the diversity and distribution of AD-ligases among bacteria also highlights their importance in understanding bacterial pathogenesis, and we propose, a potential role in survival of antibiotic treatments and acquisition of antimicrobial resistance. Many of the AD-ligases described above are present in disease-causing bacteria, for example the Lig C and Lig D pathways found in *M. tuberculosis* and *P. aeruginosa* are used to survive assaults on genomic integrity during stationary phase, which would presumably provide an advantage to these bacteria during latent Tuberculosis infection or in antibiotic-tolerant persister cells of *P. aeruginosa*. The low-fidelity nature of Lig D-mediated repair in particular could drive chromosomal mutations which are the predominant means by which *M. tuberculosis* acquires antibiotic resistance (94). Genes encoding Lig E are widespread among naturally-transformable Gram-negative pathogens such as *Vibrio cholerae*, *Campylobacter jejuni*, *Haemophilus influenzae* and *Neisseria gonorrhoeae* (10,58,59), many of which present current or emerging AMR threats (95–97). As we and others have proposed, Lig E may be involved in bacterial competence with the acquired environmental DNA being used in homology-directed DNA repair processes, or facilitating horizontal gene transfer (10,58). We believe the widespread nature of Lig E among significant human pathogens recommends it as a high priority target to elucidate any potential role in disease development and transmission; especially as our recently advanced knowledge of its structure and activity provides a basis for structure-guided development of therapeutic drugs to interfere with its function.

Finally, we propose that an as-yet unstudied group of bacterial AD-ligases found on mobile genetic elements of

Enterobacteriaceae may be important. These AD-ligases, which were identified during our previous phylogenomic study of AD-ligase distribution among bacteria (10), are encoded on mobile genetic elements including lysogenic phage and, phage-like plasmids together with other DNA-metabolizing genes. In some cases, these genes are co-localized with resistance and virulence genes on the same mobile element (98). Organisms include *Yersinia*, *Klebsiella*, *Salmonella* species, as well as *E. coli* strains. As these AD-ligases are not ubiquitous among these groups, it would be interesting to see what, if any, survival advantage they impart, and whether possessing these auxiliary DNA modifying enzymes confers influences resistance or pathogenicity traits.

## CONCLUSIONS

As key gatekeepers of genetic integrity and essential biotechnological tools, determining the structure-activity relationships and biological functions of DNA ligases remains an important area of research. The elucidation of recent high-resolution crystal structures, especially in unveiling details of the catalytic steps of bond formation has considerably increased our understanding of the ligase reaction mechanism. Application of advanced computational methods to model the dynamics of these enzymes will no doubt make further contributions to understanding both the chemical and interaction steps of this process. One aspect of DNA ligase diversity which remains open, is the evolutionary trajectory leading to the diverse forms and distributions of these enzymes. Such questions are intriguing, not only from a fundamental perspective, but also as directed evolution and evolutionarily-informed enzyme-discovery initiatives have potential to deliver and identify new ligase forms for human use. In particular, the present review highlights the diverse structures and unknown functions of bacterial AD-ligases as a high area of interest, both from an applied biotechnological and applied biomedical perspective. An outstanding question regarding the evolution of ND-ligases is why the switch from ATP to NAD-cofactor use arose in bacteria and why this isoform continues to be exclusively and ubiquitously used for bacterial DNA replication. Despite the wide-spread presence of AD-ligases in bacterial genomes, no species of bacteria has been identified to date that relies on the ATP-dependent isoform for survival, or lacks the NAD-dependent isoform. Identification of such organisms, if they exist, would go some way to unravelling this evolutionary enigma.

## SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

## FUNDING

Marsden Fund of New Zealand [18-UOW-034 A.W.]. Funding for open access charge: is University of Tromsø. *Conflict of interest statement.* None declared.

## REFERENCES

- Lohman, G.J., Tabor, S. and Nichols, N.M. (2011) DNA ligases. *Curr. Protoc. Mol. Biol.*, **Chapter 3**, Unit 3.14.
- Tomkinson, A.E., Vijayakumar, S., Pascal, J.M. and Ellenberger, T. (2006) DNA ligases: structure, reaction mechanism, and function. *Chem. Rev.*, **106**, 687–699.
- Pascal, J.M. (2008) DNA and RNA ligases: structural variations and shared mechanisms. *Curr. Opin. Struct. Biol.*, **18**, 96–105.
- Doherty, A.J. and Suh, S.W. (2000) Structural and mechanistic conservation in DNA ligases. *Nucleic Acids Res.*, **28**, 4051–4058.
- Shuman, S. (2009) DNA ligases: progress and prospects. *J Biol Chem.*, **284**, 17365–17369.
- Nandakumar, J., Nair, P.A. and Shuman, S. (2007) Last stop on the road to repair: structure of *E. coli* DNA ligase bound to nicked DNA-adenylate. *Mol. Cell*, **26**, 257–271.
- Wilkinson, A., Smith, A., Bullard, D., Lavesa-Curto, M., Sayer, H., Bonner, A., Hemmings, A. and Bowater, R. (2005) Analysis of ligation and DNA binding by *Escherichia coli* DNA ligase (LigA). *Biochim. Biophys. Acta (BBA) - Proteins Proteomics*, **1749**, 113–122.
- Shuman, S. and Glickman, M.S. (2007) Bacterial DNA repair by non-homologous end joining. *Nat. Rev. Microbiol.*, **5**, 852–861.
- Pitcher, R.S., Brissett, N.C. and Doherty, A.J. (2007) Nonhomologous end-joining in bacteria: a microbial perspective. *Annu. Rev. Microbiol.*, **61**, 259–282.
- Williamson, A., Hjerde, E. and Kahlke, T. (2016) Analysis of the distribution and evolution of the ATP-dependent DNA ligases of bacteria delineates a distinct phylogenetic group ‘Lig E’. *Mol. Microbiol.*, **99**, 274–290.
- Wilkinson, A., Day, J. and Bowater, R. (2001) Bacterial DNA ligases. *Mol. Microbiol.*, **40**, 1241–1248.
- Yutin, N. and Koonin, E.V. (2009) Evolution of DNA ligases of nucleocytoplasmic large DNA viruses of eukaryotes: a case of hidden complexity. *Biol. Direct*, **4**, 51.
- Weiss, B., Jacquemin-Sablon, A., Live, T.R., Fareed, G.C. and Richardson, C.C. (1968) Enzymatic breakage and joining of deoxyribonucleic acid. VI. Further purification and properties of polynucleotide ligase from *Escherichia coli* infected with bacteriophage T4. *J. Biol. Chem.*, **243**, 4543–4555.
- Sriskanda, V. and Shuman, S. (2002) Conserved residues in domain Ia are required for the reaction of *Escherichia coli* DNA ligase with NAD<sup>+</sup>. *J. Biol. Chem.*, **277**, 9695–9700.
- Sriskanda, V. and Shuman, S. (2002) Role of nucleotidyltransferase motifs I, III and IV in the catalysis of phosphodiester bond formation by *Chlorella virus* DNA ligase. *Nucleic Acids Res.*, **30**, 903–911.
- Akey, D., Martins, A., Aniuoku, J., Glickman, M.S., Shuman, S. and Berger, J.M. (2006) Crystal structure and nonhomologous end-joining function of the ligase component of *Mycobacterium* DNA ligase D. *J. Biol. Chem.*, **281**, 13412–13423.
- Odell, M., Sriskanda, V., Shuman, S. and Nikolov, D.B. (2000) Crystal structure of eukaryotic DNA ligase-adenylate illuminates the mechanism of nick sensing and strand joining. *Mol. Cell*, **6**, 1183–1193.
- Subramanya, H.S., Doherty, A.J., Ashford, S.R. and Wigley, D.B. (1996) Crystal structure of an ATP-dependent DNA ligase from bacteriophage T7. *Cell*, **85**, 607–615.
- Unciuleac, M.C., Goldgur, Y. and Shuman, S. (2019) Structures of ATP-bound DNA ligase D in a closed domain conformation reveal a network of amino acid and metal contacts to the ATP phosphates. *J. Biol. Chem.*, **294**, 5094–5104.
- Unciuleac, M.C., Goldgur, Y. and Shuman, S. (2017) Two-metal versus one-metal mechanisms of lysine adenylation by ATP-dependent and NAD(+) dependent polynucleotide ligases. *Proc. Natl. Acad. Sci. U.S.A.*, **114**, 2592–2597.
- Unciuleac, M.C., Goldgur, Y. and Shuman, S. (2015) Structure and two-metal mechanism of a eukaryal nick-sealing RNA ligase. *Proc. Natl. Acad. Sci. U.S.A.*, **112**, 13868–13873.
- Kim, D.J., Kim, O., Kim, H.W., Kim, H.S., Lee, S.J. and Suh, S.W. (2009) ATP-dependent DNA ligase from *Archaeoglobus fulgidus* displays a tightly closed conformation. *Acta Crystallogr. F*, **65**, 544–550.
- Williamson, A., Rothweiler, U. and Leiros, H.-K. (2014) Enzyme-adenylate structure of a bacterial ATP-dependent DNA ligase with a minimized DNA-binding surface. *Acta Crystallogr. D*, **70**, 3043–3056.
- Nair, P.A., Nandakumar, J., Smith, P., Odell, M., Lima, C.D. and Shuman, S. (2007) Structural basis for nick recognition by a minimal pluripotent DNA ligase. *Nat. Struct. Mol. Biol.*, **14**, 770–778.

25. Williamson, A., Grgic, M. and Leiros, H.-K. (2018) DNA binding with a minimal scaffold: structure-function analysis of Lig E DNA ligases. *Nucleic Acids Res.*, **46**, 8616–8629.
26. Pascal, J.M., O'Brien, P.J., Tomkinson, A.E. and Ellenberger, T. (2004) Human DNA ligase I completely encircles and partially unwinds nicked DNA. *Nature*, **432**, 473–478.
27. Williamson, A. and Leiros, H.-K. (2019) Structural intermediates of a DNA-ligase complex illuminate the role of the catalytic metal ion and mechanism of phosphodiester bond formation. *Nucleic Acids Res.*, **47**, 7147–7162.
28. Chen, Y., Liu, H., Yang, C., Gao, Y., Yu, X., Chen, X., Cui, R., Zheng, L., Li, S., Li, X. *et al.* (2019) Structure of the error-prone DNA ligase of African swine fever virus identifies critical active site residues. *Nat. Commun.*, **10**, 387.
29. Shi, K., Bohl, T.E., Park, J., Zasada, A., Malik, S., Banerjee, S., Tran, V., Li, N., Yin, Z., Kurniawan, F. *et al.* (2018) T4 DNA ligase structure reveals a prototypical ATP-dependent ligase with a unique mode of sliding clamp interaction. *Nucleic Acids Res.*, **46**, 10474–10488.
30. Cotner-Gohara, E., Kim, I.-K., Hammel, M., Tainer, J.A., Tomkinson, A.E. and Ellenberger, T. (2010) Human DNA ligase III recognizes DNA Ends by dynamic switching between two DNA-bound states. *Biochemistry*, **49**, 6165–6176.
31. Bauer, R.J., Jurkiw, T.J., Evans, T.C. Jr. and Lohman, G.J. (2017) Rapid time scale analysis of T4 DNA ligase-DNA binding. *Biochemistry*, **56**, 1117–1129.
32. Tumbale, P.P., Jurkiw, T.J., Schellenberg, M.J., Riccio, A.A., O'Brien, P.J. and Williams, R.S. (2019) Two-tiered enforcement of high-fidelity DNA ligation. *Nat. Commun.*, **10**, 5431.
33. Tumbale, P., Appel, C.D., Kraehenbuehl, R., Robertson, P.D., Williams, J.S., Krahn, J., Ahel, I. and Williams, R.S. (2011) Structure of an aprataxin-DNA complex with insights into AOA1 neurodegenerative disease. *Nat. Struct. Mol. Biol.*, **18**, 1189–1195.
34. Kaminski, A.M., Tumbale, P.P., Schellenberg, M.J., Williams, R.S., Williams, J.G., Kunkel, T.A., Pedersen, L.C. and Bebenek, K. (2018) Structures of DNA-bound human ligase IV catalytic core reveal insights into substrate binding and catalysis. *Nat. Commun.*, **9**, 2642.
35. Ochi, T., Gu, X. and Blundell, T.L. (2013) Structure of the catalytic region of DNA ligase IV in complex with an Artemis fragment sheds light on double-strand break repair. *Structure*, **21**, 672–679.
36. Nishida, H., Kiyonari, S., Ishino, Y. and Morikawa, K. (2006) The closed structure of an archaeal DNA ligase from *Pyrococcus furiosus*. *J. Mol. Biol.*, **360**, 956–967.
37. Petrova, T., Bezsudnova, E.Y., Boyko, K.M., Mardanov, A.V., Polyakov, K.M., Volkov, V.V., Kozin, M., Ravin, N.V., Shabalin, I.G., Skryabin, K.G. *et al.* (2012) ATP-dependent DNA ligase from *Thermococcus* sp. 1519 displays a new arrangement of the OB-fold domain. *Acta Crystallogr. F*, **68**, 1440–1447.
38. Wilkinson, A., Smith, A., Bullard, D., Lavesa-Curto, M., Sayer, H., Bonner, A., Hemmings, A. and Bowater, R. (2005) Analysis of ligation and DNA binding by *Escherichia coli* DNA ligase (LigA). *Biochim. Biophys. Acta*, **1749**, 113–122.
39. Wang, L.K., Nair, P.A. and Shuman, S. (2008) Structure-guided mutational analysis of the OB, HhH, and BRCT domains of *Escherichia coli* DNA ligase. *J. Biol. Chem.*, **283**, 23343–23352.
40. Rossi, R., Montecucco, A., Ciarrocchi, G. and Biamonti, G. (1997) Functional characterization of the T4 DNA ligase: a new insight into the mechanism of action. *Nucleic Acids Res.*, **25**, 2106–2113.
41. Gong, C.L., Bongiorno, P., Martins, A., Stephanou, N.C., Zhu, H., Shuman, S. and Glickman, M.S. (2005) Mechanism of nonhomologous end-joining in mycobacteria: a low-fidelity repair system driven by Ku, ligase D and ligase C. *Nat. Struct. Mol. Biol.*, **12**, 304–312.
42. Nair, P.A., Smith, P. and Shuman, S. (2010) Structure of bacterial LigD 3'-phosphoesterase unveils a DNA repair superfamily. *Proc. Natl. Acad. Sci. U.S.A.*, **107**, 12822–12827.
43. Zhu, H., Nandakumar, J., Aniuoku, J., Wang, L.K., Glickman, M.S., Lima, C.D. and Shuman, S. (2006) Atomic structure and nonhomologous end-joining function of the polymerase component of bacterial DNA ligase D. *Proc. Natl. Acad. Sci. U.S.A.*, **103**, 1711–1716.
44. Zhu, H. and Shuman, S. (2005) A primer-dependent polymerase function of *Pseudomonas aeruginosa* ATP-dependent DNA ligase (LigD). *J. Biol. Chem.*, **280**, 418–427.
45. Ellenberger, T. and Tomkinson, A.E. (2008) In: *Annual Review of Biochemistry*, Vol. 77, pp. 313–338.
46. Masani, S., Han, L., Meek, K. and Yu, K. (2016) Redundant function of DNA ligase I and 3 in alternative end-joining during immunoglobulin class switch recombination. *PNAS*, **113**, 1261–1266.
47. Lu, G., Duan, J., Shu, S., Wang, X., Gao, L., Guo, J. and Zhang, Y. (2016) Ligase I and ligase III mediate the DNA double-strand break ligation in alternative end-joining. *Proc. Natl. Acad. Sci. U.S.A.*, **113**, 1256–1260.
48. Bhattarai, H., Gupta, R. and Glickman, M.S. (2014) DNA ligase C1 mediates the LigD-independent nonhomologous end-joining pathway of *Mycobacterium smegmatis*. *J. Bacteriol.*, **196**, 3366–3376.
49. Płociński, P., Brissett, N.C., Bianchi, J., Brzostek, A., Korycka-Machala, M., Dziembowski, A., Dziadek, J. and Doherty, A.J. (2017) DNA Ligase C and Prim-PolC participate in base excision repair in mycobacteria. *Nat. Commun.*, **8**, 1251.
50. Zhu, H. and Shuman, S. (2008) Bacterial nonhomologous end joining ligases preferentially seal breaks with a 3'-OH monoribonucleotide. *J. Biol. Chem.*, **283**, 8331–8339.
51. Gong, C., Martins, A., Bongiorno, P., Glickman, M. and Shuman, S. (2004) Biochemical and genetic analysis of the four DNA ligases of mycobacteria. *J. Biol. Chem.*, **279**, 20594–20606.
52. Berg, K., Leiros, I. and Williamson, A. (2019) Temperature adaptation of DNA ligases from psychrophilic organisms. *Extremophiles*, **23**, 305–317.
53. Williamson, A. and Pedersen, H. (2014) Recombinant expression and purification of an ATP-dependent DNA ligase from *Aliivibrio salmonicida*. *Protein Express. Purif.*, **97**, 29–36.
54. Ejaz, A. and Shuman, S. (2018) Characterization of Lhr-Core DNA helicase and manganese-dependent DNA nuclease components of a bacterial gene cluster encoding nucleic acid repair enzymes. *J. Biol. Chem.*, **293**, 17491–17504.
55. Ejaz, A., Goldgur, Y. and Shuman, S. (2019) Activity and structure of *Pseudomonas putida* MPE, a manganese-dependent single-strand DNA endonuclease encoded in a nucleic acid repair gene cluster. *J. Biol. Chem.*, **294**, 7931–7941.
56. Johnston, C., Martin, B., Fichant, G., Polard, P. and Claverys, J.P. (2014) Bacterial transformation: distribution, shared mechanisms and divergent control. *Nat. Rev. Microbiol.*, **12**, 181–196.
57. Mell, J.C. and Redfield, R.J. (2014) Natural competence and the evolution of DNA uptake specificity. *J. Bacteriol.*, **196**, 1471–1483.
58. Magnet, S. and Blanchard, J.S. (2004) Mechanistic and kinetic study of the ATP-dependent DNA ligase of *Neisseria meningitidis*. *Biochemistry*, **43**, 710–717.
59. Cheng, C.H. and Shuman, S. (1997) Characterization of an ATP-dependent DNA ligase encoded by *Haemophilus influenzae*. *Nucleic Acids Res.*, **25**, 1369–1374.
60. Zallot, R., Oberg, N. and Gerlt, J.A. (2019) The EFI web resource for genomic enzymology tools: leveraging protein, genome, and metagenome databases to discover novel enzymes and metabolic pathways. *Biochemistry*, **58**, 4169–4182.
61. Zallot, R., Oberg, N.O. and Gerlt, J.A. (2018) 'Democratized' genomic enzymology web tools for functional assignment. *Curr. Opin. Chem. Biol.*, **47**, 77–85.
62. Levin, B.J., Huang, Y.Y., Peck, S.C., Wei, Y., Martinez-Del Campo, A., Marks, J.A., Franzosa, E.A., Huttenhower, C. and Balskus, E.P. (2017) A prominent glycol radical enzyme in human gut microbiomes metabolizes trans-4-hydroxy-l-proline. *Science*, **355**, eaai8386.
63. Gerlt, J.A. (2017) Genomic enzymology: web tools for leveraging protein family sequence-function space and genome context to discover novel functions. *Biochemistry*, **56**, 4293–4308.
64. Akiva, E., Copp, J.N., Tokuriki, N. and Babbitt, P.C. (2017) Evolutionary and molecular foundations of multiple contemporary functions of the nitroreductase superfamily. *Proc. Natl. Acad. Sci. U.S.A.*, **114**, E9549–E9558.
65. O'Donnell, M., Langston, L. and Stillman, B. (2013) Principles and concepts of DNA replication in bacteria, archaea, and eukarya. *Cold Spring Harb. Perspect. Biol.*, **5**, a010108.
66. Makarova, K.S. and Koonin, E.V. (2013) Archaeology of eukaryotic DNA replication. *Cold Spring Harb. Perspect. Biol.*, **5**, a012963.
67. Altmeyer, M., Messner, S., Hassa, P.O., Fey, M. and Hottiger, M.O. (2009) Molecular mechanism of poly(ADP-ribosylation) by PARP1 and identification of lysine residues as ADP-ribose acceptor sites. *Nucleic Acids Res.*, **37**, 3723–3738.



68. Huambachano, O., Herrera, F., Rancourt, A. and Satoh, M.S. (2011) Double-stranded DNA binding domain of poly(ADP-ribose) polymerase-1 and molecular insight into the regulation of its activity. *J. Biol. Chem.*, **286**, 7149–7160.
69. Cai, L., Hu, C., Shen, S., Wang, W. and Huang, W. (2004) Characterization of bacteriophage T3 DNA ligase. *J. Biochem.*, **135**, 397–403.
70. Shuman, S. (1995) *Vaccinia virus* DNA ligase: specificity, fidelity, and inhibition. *Biochemistry*, **34**, 16138–16147.
71. Zhao, A., Gray, F.C. and MacNeill, S.A. (2006) ATP- and NAD<sup>+</sup>-dependent DNA ligases share an essential function in the halophilic archaeon *Haloferax volcanii*. *Mol. Microbiol.*, **59**, 743–752.
72. Gansauge, M.T. and Meyer, M. (2019) A method for single-stranded ancient DNA library preparation. *Methods Mol. Biol.*, **1963**, 75–83.
73. Yegnashubramanian, S. (2013) Preparation of fragment libraries for next-generation sequencing on the applied biosystems SOLiD platform. *Methods Enzymol.*, **529**, 185–200.
74. Goffin, C., Bailly, V. and Verly, W.G. (1987) Nicks 3' or 5' to AP sites or to mispaired bases, and one-nucleotide gaps can be sealed by T4 DNA ligase. *Nucleic Acids Res.*, **15**, 8755–8771.
75. Macdonald, S.J. (2007) Genotyping by oligonucleotide ligation assay (OLA). *CSH Protoc.*, **2007**, pdb.prot4843.
76. Tong, J., Cao, W. and Barany, F. (1999) Biochemical properties of a high fidelity DNA ligase from *Thermus* species AK16D. *Nucleic Acids Res.*, **27**, 788–794.
77. Tanabe, M., Ishino, Y. and Nishida, H. (2015) From structure-function analyses to protein engineering for practical applications of DNA ligase. *Archaea*, **2015**, 267570.
78. Chambers, C.R. and Patrick, W.M. (2015) Archaeal nucleic acid ligases and their potential in biotechnology. *Archaea (Vancouver, B.C.)*, **2015**, 170571.
79. Tanabe, M., Ishino, S., Yohda, M., Morikawa, K., Ishino, Y. and Nishida, H. (2012) Structure-based mutational study of an archaeal DNA ligase towards improvement of ligation activity. *ChemBioChem*, **13**, 2575–2582.
80. Potapov, V., Ong, J.L., Kucera, R.B., Langhorst, B.W., Bilotti, K., Pryor, J.M., Cantor, E.J., Canton, B., Knight, T.F., Evans, T.C. Jr. *et al.* (2018) Comprehensive profiling of four base overhang ligation fidelity by T4 DNA ligase and application to DNA assembly. *ACS Synth. Biol.*, **7**, 2665–2674.
81. Lohman, G.J., Bauer, R.J., Nichols, N.M., Mazzola, L., Bybee, J., Rivizzigno, D., Cantin, E. and Evans, T.C. Jr. (2016) A high-throughput assay for the comprehensive profiling of DNA ligase fidelity. *Nucleic Acids Res.*, **44**, e14.
82. Zimmerman, S.B. and Pfeiffer, B.H. (1983) Macromolecular crowding allows blunt-end ligation by DNA ligases from rat liver or *Escherichia coli*. *Proc. Natl Acad. Sci. U.S.A.*, **80**, 5852–5856.
83. Wiaderkiewicz, R. and Ruiz-Carrillo, A. (1987) Mismatch and blunt to protruding-end joining by DNA ligases. *Nucleic Acids Res.*, **15**, 7831–7848.
84. Kuhn, H. and Frank-Kamenetskii, M.D. (2005) Template-independent ligation of single-stranded DNA by T4 DNA ligase. *FEBS J.*, **272**, 5991–6000.
85. Jin, J., Vaud, S., Zhelkovsky, A.M., Posfai, J. and McReynolds, L.A. (2016) Sensitive and specific miRNA detection method using SplintR Ligase. *Nucleic Acids Res.*, **44**, e116.
86. Vanmeert, M., Razzokov, J., Mirza, M.U., Weeks, S.D., Schepers, G., Bogaerts, A., Rozenski, J., Froeyen, M., Herdewijn, P., Pinheiro, V.B. *et al.* (2019) Rational design of an XNA ligase through docking of unbound nucleic acids to toroidal proteins. *Nucleic Acids Res.*, **47**, 7130–7142.
87. Ran, F.A., Hsu, P.D., Wright, J., Agarwala, V., Scott, D.A. and Zhang, F. (2013) Genome engineering using the CRISPR-Cas9 system. *Nat. Protoc.*, **8**, 2281–2308.
88. Barrangou, R. and van Pijkeren, J.P. (2016) Exploiting CRISPR-Cas immune systems for genome editing in bacteria. *Curr. Opin. Biotechnol.*, **37**, 61–68.
89. Su, T., Liu, F., Gu, P., Jin, H., Chang, Y., Wang, Q., Liang, Q. and Qi, Q. (2016) A CRISPR-Cas9 assisted non-homologous end-joining strategy for one-step engineering of bacterial genome. *Sci. Rep.*, **6**, 37895.
90. Tong, Y., Charusanti, P., Zhang, L., Weber, T. and Lee, S.Y. (2015) CRISPR-Cas9 based engineering of actinomycetal genomes. *ACS Synth. Biol.*, **4**, 1020–1029.
91. Korycka-Machala, M., Rychta, E., Brzostek, A., Sayer, H.R., Rumijowska-Galewicz, A., Bowater, R.P. and Dziadek, J. (2007) Evaluation of NAD(+) -dependent DNA ligase of mycobacteria as a potential target for antibiotics. *Antimicrob. Agents Chemother.*, **51**, 2888–2897.
92. Pergolizzi, G., Wagner, G.K. and Bowater, R.P. (2016) Biochemical and structural characterisation of DNA ligases from bacteria and archaea. *Biosci. Rep.*, **36**, e00391.
93. Pergolizzi, G., Wagner, G.K. and Bowater, R.P. (2016) Biochemical and structural characterisation of DNA ligases from bacteria and archaea. *Biosci. Rep.*, **36**, 00391.
94. Gygli, S.M., Borrell, S., Trauner, A. and Gagneux, S. (2017) Antimicrobial resistance in *Mycobacterium tuberculosis*: mechanistic and evolutionary perspectives. *FEMS Microbiol. Rev.*, **41**, 354–373.
95. Lewis, D.A. (2014) Global resistance of *Neisseria gonorrhoeae*: when theory becomes reality. *Curr. Opin. Infect. Dis.*, **27**, 62–67.
96. Obergfell, K.P. and Seifert, H.S. (2015) Mobile DNA in the pathogenic *Neisseria*. *Microbiol. Spectrum*, **3**, MDNA3-0015-2014.
97. Matthey, N. and Blokesch, M. (2016) The DNA-uptake process of naturally competent *Vibrio cholerae*. *Trends Microbiol.*, **24**, 98–110.
98. Octavia, S., Sara, J. and Lan, R. (2015) Characterization of a large novel phage-like plasmid in *Salmonella enterica* serovar *Typhimurium*. *FEMS Microbiol. Lett.*, **362**, fmv044.