


## Review Article

# Nucleic acid joining enzymes: biological functions and synthetic applications beyond DNA

Chelsea Blackstock<sup>1\*</sup>, Caitlin Walters-Freke<sup>1\*</sup>, Nigel Richards<sup>2,3</sup>  and Adele Williamson<sup>1</sup><sup>1</sup>School of Science, University of Waikato, Hamilton, Waikato, 3216, New Zealand; <sup>2</sup>Foundation for Applied Molecular Evolution, Alachua, FL, 32615, U.S.A.; <sup>3</sup>School of Chemistry, Cardiff University, Cardiff, CF10 3AT, U.K**Correspondence:** Adele Williamson (adele.williamson@waikato.ac.nz)

DNA-joining by ligase and polymerase enzymes has provided the foundational tools for generating recombinant DNA and enabled the assembly of gene and genome-sized synthetic products. Xenobiotic nucleic acid (XNA) analogues of DNA and RNA with alternatives to the canonical bases, so-called ‘unnatural’ nucleobase pairs (UBP-XNAs), represent the next frontier of nucleic acid technologies, with applications as novel therapeutics and in engineering semi-synthetic biological organisms. To realise the full potential of UBP-XNAs, researchers require a suite of compatible enzymes for processing nucleic acids on a par with those already available for manipulating canonical DNA. In particular, enzymes able to join UBP-XNA will be essential for generating large assemblies and also hold promise in the synthesis of single-stranded oligonucleotides. Here, we review recent and emerging advances in the DNA-joining enzymes, DNA polymerases and DNA ligases, and describe their applications to UBP-XNA manipulation. We also discuss the future directions of this field which we consider will involve two-pronged approaches of enzyme biodiscovery for natural UBP-XNA compatible enzymes, coupled with improvement by structure-guided engineering.

## Introduction

Enzyme-mediated cutting, joining and amplification of canonical DNA using restriction endonucleases, DNA ligases and polymerases to generate recombinant DNA are the foundational tools of modern molecular biology. Innovations in both the synthesis and assembly steps have led to the synthetic biology revolution, paving the way for semi-synthetic or fully synthetic organisms and other bio-inspired technologies of the future [1,2]. The development of highly efficient synthesis of user-defined polynucleotide sequences has applications ranging from commercial provision of primers and probes bearing fluorescent labels and other functional group modifications to construction of bespoke gene sequences incorporating specific mutations, reading-frame fusions and optimised codons for improved heterologous protein expression [3].

Currently, the most common method for *de novo* polynucleotide synthesis at commercial scale uses nucleoside phosphoramidite chemistry, which has a long and well-documented history and has been the subject of extensive reviews [4–6]. The typical solid-phase approach involves the addition of one 5'-protected dimethoxytrityl (DMT) nucleotide phosphoramidite at a time to a growing strand, that is attached to a solid support (Figure 1A). Use of microarrays increases throughput by enabling the simultaneous synthesis of numerous diverse oligonucleotide strands. However, solid-phase synthesis is limited to products less than 200 nucleotides (nt) in length, because side reactions occurring in each addition cycle negatively impact yields [7]. On the other hand, as reviewed recently [8], enzymatic approaches for the production of longer oligos have proved to be an effective solution to this problem. Template-independent enzymatic oligonucleotide synthesis (TiEOS) methods generate oligonucleotides using the terminal deoxynucleotidyl transferase (TdT) polymerase which incorporates 3'-protected nucleoside 5' triphosphates (NTPs) onto the 3'-terminus of DNA possessing a 3'-overhang (Figure 1B). The use of 3'-capped dNTPs enables the controlled introduction of nt to build sequence-specific oligomers [9]. Additional strategies to generate extended duplexes may be ligation based, where shorter oligonucleotides are synthesised with complementary overhangs and then joined together using a ligase enzyme such as T4 DNA ligase. Alternatively, polymerase cycling assembly, in which overlapping synthetic oligonucleotides are amplified using a DNA polymerase, can produce long pieces of double-stranded DNA (Figure 1C and D) [6]. Both these latter methods, however, rely on base pairing and provision of pre-synthesised

\*These authors contributed equally to this work.

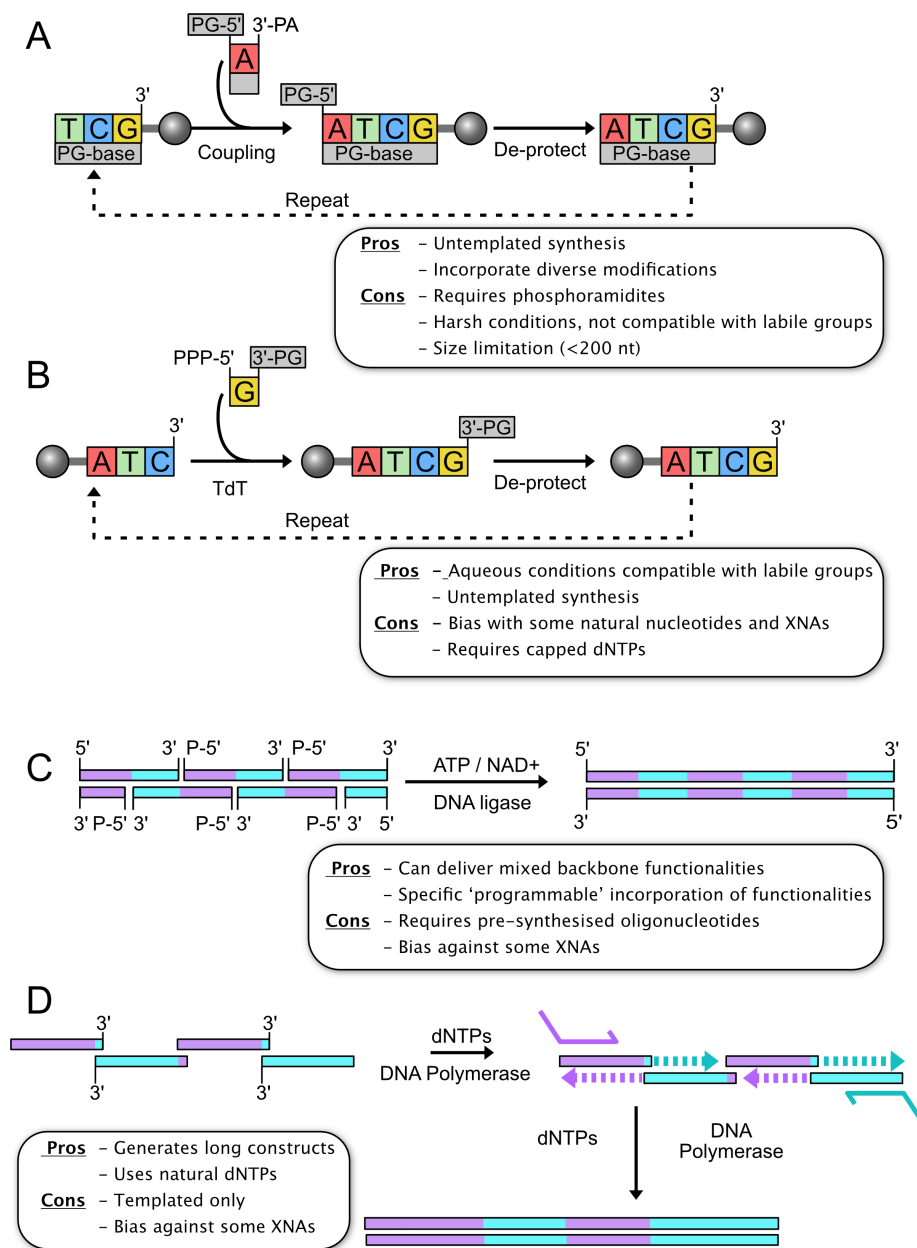
Received: 9 September 2024

Revised: 21 November 2024

Accepted: 4 December 2024

Version of Record

Published: 22 January 2025



**Figure 1: Schematic of different DNA synthesis and assembly methods.**

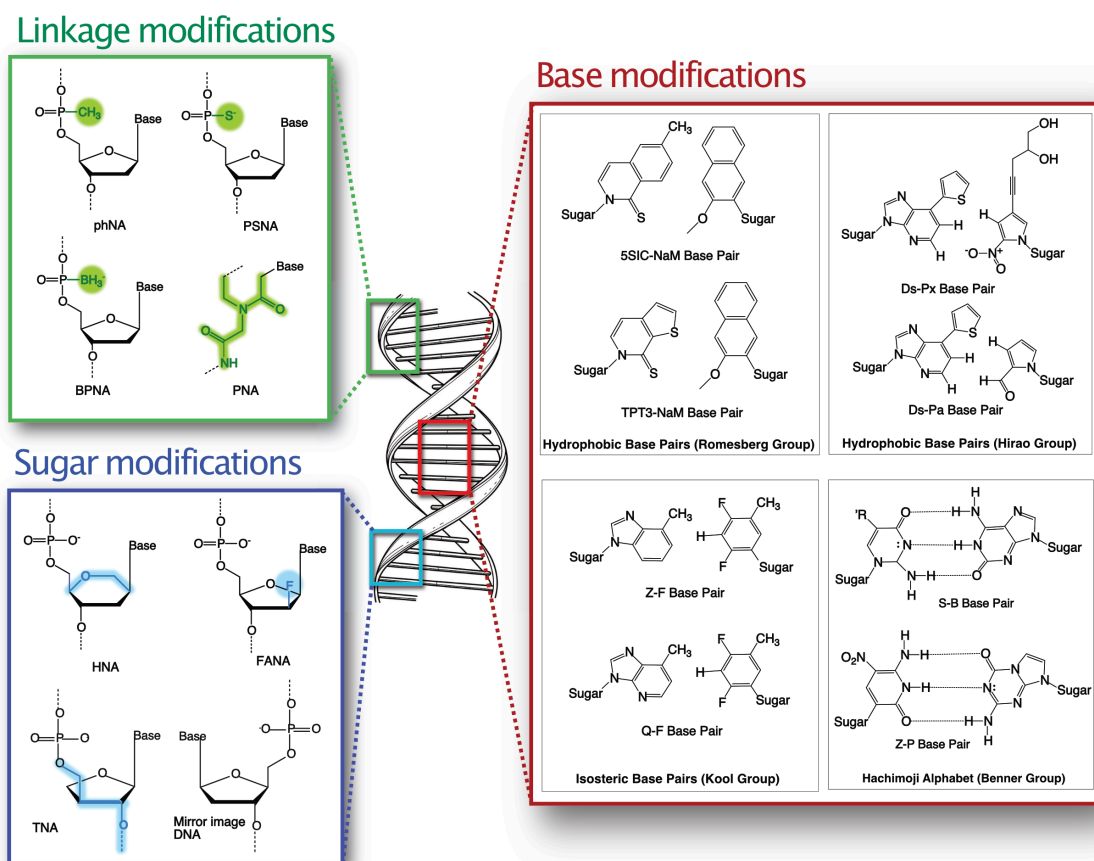
(A) The phosphoramidite synthesis method generates single-stranded nucleic acid by sequentially adding 2' deoxy nucleoside phosphoramidites via the reactive 3' functionality (3' PA) to the 5' OH terminus of the chain, which is immobilised on a bead. Protecting groups prevent unwanted side reactions on the 5' OH terminus of the incoming nucleoside phosphoramidite (PG-5') and on OH and NH<sub>2</sub> groups of the bases (PG-Base). The deprotection step subsequent to nucleoside phosphoramidite addition generates a free 5' OH terminus for the next extension step. (B) TdT synthesis appends single 2' deoxyribonucleotide triphosphates (dNTPs) to the 3' terminus of the immobilised oligonucleotide chain in an enzyme-catalysed step. To ensure only one dNTP is added per extension cycle, the 3' OH of the incoming dNTP is protected (3' PG) and this cap is removed in a deprotection step prior to the next addition to generate a free 3' terminus. (C) DNA-ligase-mediated assembly uses short (<20 nt) pre-synthesised oligonucleotides with complementary overhangs that generate a contiguous nicked duplex with no gaps. All oligonucleotides at internal positions are 5' phosphorylated, and the DNA ligase joins these directly to the adjacent 3' OH terminus to generate the double-stranded product. (D) Polymerase-based synthesis/assembly begins with a series of oligonucleotides with short cohesive overlaps that anneal to give a non-contiguous duplex with long stretches of single-stranded nucleic acid. In the first step, the polymerase extends from the 3' OH of the acceptor strand by the addition of dNTPs, using the complementary strand as a template. Successive cycles of melting, annealing and extension create a complete duplex which spans the entire range and is amplified with external primers in a final step.

oligonucleotides and so do not constitute entirely *de novo* nucleic acid synthesis; however, they can be used to build or amplify larger constructs using oligonucleotides generated from the first two methods.

An emerging area of nucleic acid research is the use of xenobiotic nucleic acids (XNAs) in synthetic biology, therapeutic and biosensing applications [10]. XNAs are nucleic acid analogues of

DNA and RNA which have non-natural substitutions in their sugar moiety, phosphodiester linkages or base pairs (Figure 2) [11]. The exact definition of the abbreviation ‘XNA’ as a general umbrella term for non-biological alternatives to canonical DNA and RNA is currently a matter of discussion. Debate centres on the use of more specific identifiers to delineate the location of the non-canonical moiety (e.g. sugar vs. base) and whether chemically modified variations of natural nucleic acids (e.g. 2’OMe-RNA) should be included. For the purposes of this review article, we use XNA to include all nucleic acids with non-natural substitutions. For a more nuanced account of the proposed redefinitions, the reader is directed to Chaput and Herdewijn [12].

Many XNA applications will eventually require enzymes to produce extended polynucleotide assemblies at the gene, plasmid or even genomic scales. Modern molecular biologists are spoilt by the range of enzymes available for amplification and assembly of natural DNA and RNA nucleic acids. These enzymes have been sourced from the replication, recombination and repair machinery of cellular organisms and viruses, and in many cases, their *in vitro* activity has been improved either by optimising reaction conditions, or by enzyme engineering [13,14]. Currently, the majority of XNA-containing oligonucleotides are synthesised using phosphoramidite methods. In addition to the usual intrinsic limitations on length, this strategy is complicated by the chemical instability of some functionalities and the lack of available phosphoramidite precursors with appropriate protecting groups. Enzymatic synthesis of XNA polynucleotides presents a possible solution to this issue because the mild reaction conditions, typically



**Figure 2: Representative examples of different XNAs.**

Linkage modifications substitute one of the non-bridging oxygen in the phosphodiester backbone, or replace this backbone altogether. In alkyl phosphonate nucleic acid (phNA) this substitution is a methyl group, in phosphorothioate nucleic acid (PS) it is sulphur and in boranophosphate (BPNA) it is a borano group. The peptide nucleic acid (PNA) has a peptide linkage in place of the phosphodiester bond. Sugar modifications include replacement of the 2’ oxygen with fluorine in 2’-fluoro arabinonucleic acid (FANA), or substitution with an alternate sugar as in hexitol nucleic acid (HNA) where 1,5-anhydrohexitol replaces ribose, or threose nucleic acid (TNA) where threose is substituted. In mirror-image DNA, the D-form of the ribose is replaced with the L-ribose enantiomer. Base modifications include substitution of the hydrogen-bonded nitrogenous bases with moieties that pair through hydrophobic interactions, or with isosteres of the natural nitrogenous bases that lack the hydrogen bonds but interact through shape complementarity. The Hachimoji pairs include pyrimidine/purine equivalents of the natural bases, but with rearranged hydrogen bonding patterns.

aqueous and neutral pH, are compatible with labile functional groups. Unfortunately, many of the widely available enzymes used for natural DNA amplification and assembly are only partially functional on XNA substrates and there is typically a compromise in activity, fidelity or a bias against the non-canonical substrate [15–19].

This review summarises the use of different DNA polymerases and DNA ligases in nucleic acid synthesis and their potential for the generation and assembly of XNA-containing oligonucleotides. Ligases and polymerases are both found in structurally and functionally diverse classes, and we begin with a brief account of the biological role of each enzyme class followed by their utility in biotechnology. We then describe recent research into the compatibility of each enzyme type with technologically relevant categories of XNAs that contain non-natural base pairs which are orthogonal to those in canonical DNA. Finally, we discuss future directions in enzyme-mediated XNA synthesis and the role that novel and/or engineered proteins may play in this.

## Unnatural base pair xenobiotic nucleic acids: beyond canonical bases and genetic alphabets

XNAs are analogues of DNA or RNA which have alternative chemical structures in either the sugar-phosphate backbone, or the nucleobases. Non-canonical bases capable of pairing with their natural counterparts have been successfully incorporated into cellular genomes [20,21]. Others have played a pivotal role in the success of mRNA technology [22,23] and key modifications in this field are depicted in Supplementary Figure S1. Where non-canonical bases are able to form novel interactions that are mutually exclusive to natural bases, they are known as unnatural base pairs (UBPs) [24]. The desirable properties of a UBP are that the interaction is both selective and orthogonal, meaning that UBP nt partner exclusively with one another and have minimal mismatches with other canonical bases. In one of several approaches, the creation of hydrogen-bonded UBPs has allowed for the development of Artificially Expanded Genetic Information Systems (AEGIS) in which the genetic alphabet which consists of the canonical A:T/U and G:C base pairs and additional orthogonal base pairs [25,26]. At present, UBPs have been developed which can function in a similar manner to canonical base pairs during replication, transcription and translation, potentially increasing the information content which is able to be encoded in the genes of living organisms [27]. In addition to synthetic biology, other potential applications for these UBP-XNA substrates include polymer-based nanotechnology [28], incorporation into molecular beacons [29] and inclusion in single-stranded nucleic acid aptamers, which bind to targets and can be utilised in drug delivery and therapy [30–33]. Such aptamers incorporating AEGIS bases have been generated using systematic evolution of ligands by exponential enrichment (SELEX). SELEX to generate DNA aptamers involves creating a single-stranded DNA library, then selecting and enriching for particular aptamer sequences which can bind to the molecule [34]. This method has now been used to create aptamers that include non-canonical bases which have increased information density in the target binding region [35]. Such aptamers have been used to develop aptamers against targets such as anthrax protective antigen and T-cell receptor-CD3 $\epsilon$  and have been used for prodrug delivery to cancer cells [33,36,37].

In lifeforms with artificially expanded genetic alphabets, these UBPs must be recognised by enzymatic machinery for efficient replication and transcription, while to have potential in aptamer generation, they must be able to create and maintain stable secondary structures. UBPs that fulfil these criteria fall broadly into three categories depending on how this orthogonality is achieved: rearrangement of the pattern of hydrogen bond donor–acceptor pairs in the bases, introduction of novel pairing interactions through hydrophobicity, or introduction of orthogonal interactions by non-hydrogen-bonded structural complementarity (Figure 2). Of these three broad UBP categories, novel hydrogen-bonded systems hold considerable promise as they preserve Watson-and-Crick like behaviour of natural DNA, are evolvable during replication, and the inclusion of multiple consecutive hydrogen-bonded UBPs has minimal impact on duplex properties [38,39]. Early examples of hydrogen bonding UBPs include isoG (6-amino-2-ketopurine) and isoC (2-amino-4-ketopyrimidine) which are recognised by natural enzymatic machinery in both DNA replication and transcription [40]. The isoG:isoC UBP has been successfully used to encode the unnatural amino acid 3-iodotyrosine, which was incorporated into a peptide by *in vitro* translation using a novel isoGUC anticodon system [41]. Subsequent research has resulted in improvements to the chemical stability of isoC, as well as modification to remove the minor tautomeric enol form of isoG which mispairs with thymine [42]. The synthesis of two additional hydrogen-bonding

bases 'Z' (6-amino-3-(1'- $\beta$ -D-2'-deoxyribofuranosyl)-5-nitro-1*H*-pyridin-2-one), and 'P' (2-amino-8-(1'- $\beta$ -D-2'-deoxyribofuranosyl)-imidazo-[1,2a]-1,3,5-triazin-[8*H*]-4-one) has given rise to the 8-letter Hachimoji alphabet which includes the two non-natural S:B and P:Z UBPs in addition to the A:T and G:C pairs [17]. Duplexes of deoxyribose Hachimoji UBP-XNA exhibit similar thermostability and structural properties to natural DNA, and in an early proof-of-concept were able to be transcribed by RNA polymerase to generate a variant of the spinach fluorescent RNA aptamer which incorporated Hachimoji-base ribonucleotides [39].

Enzymatic replication and crucially biological retention have been achieved with the hydrophobic base pair dNaM-dTPT3 (2-methoxynaphthalene and thieno[2,3 *c*]pyridine-7(6*H*)-thione). This was achieved by co-expression of transporters facilitating the uptake of the non-natural dNTP into the host cells as well as CRISPR-Cas-based elimination of cells which lose the UPB-containing plasmid [43]. The dNaM-dTPT3 UBP is an analogue of the dNaM-d5SICS (6-methyl-2*H*-isoquinoline-1-thione) UBP which was modified to improve polymerase incorporation, highlighting the critical importance of compatibility between non-natural nucleobases and enzymatic replication machinery [44,45]. A second hydrophobic pair 'DS' (7-(2-thienyl)imidazo[4,5-*b*]pyridine) and 'PX' (2-nitro-4-propynylpyrrole) which bears little structural resemblance to dNaM, dTPT3 and d5SICS is also able to be replicated with high fidelity and this UBP is retained during multiple rounds of PCR which has allowed its incorporation into aptamers using SELEX [46]. As with other UBPs, the DS-PX pair represents the outcome of an iterative design process which eliminated mispairings between the novel hydrophobics and the natural bases, while retaining optimal geometry compatible with enzymatic propagation [27].

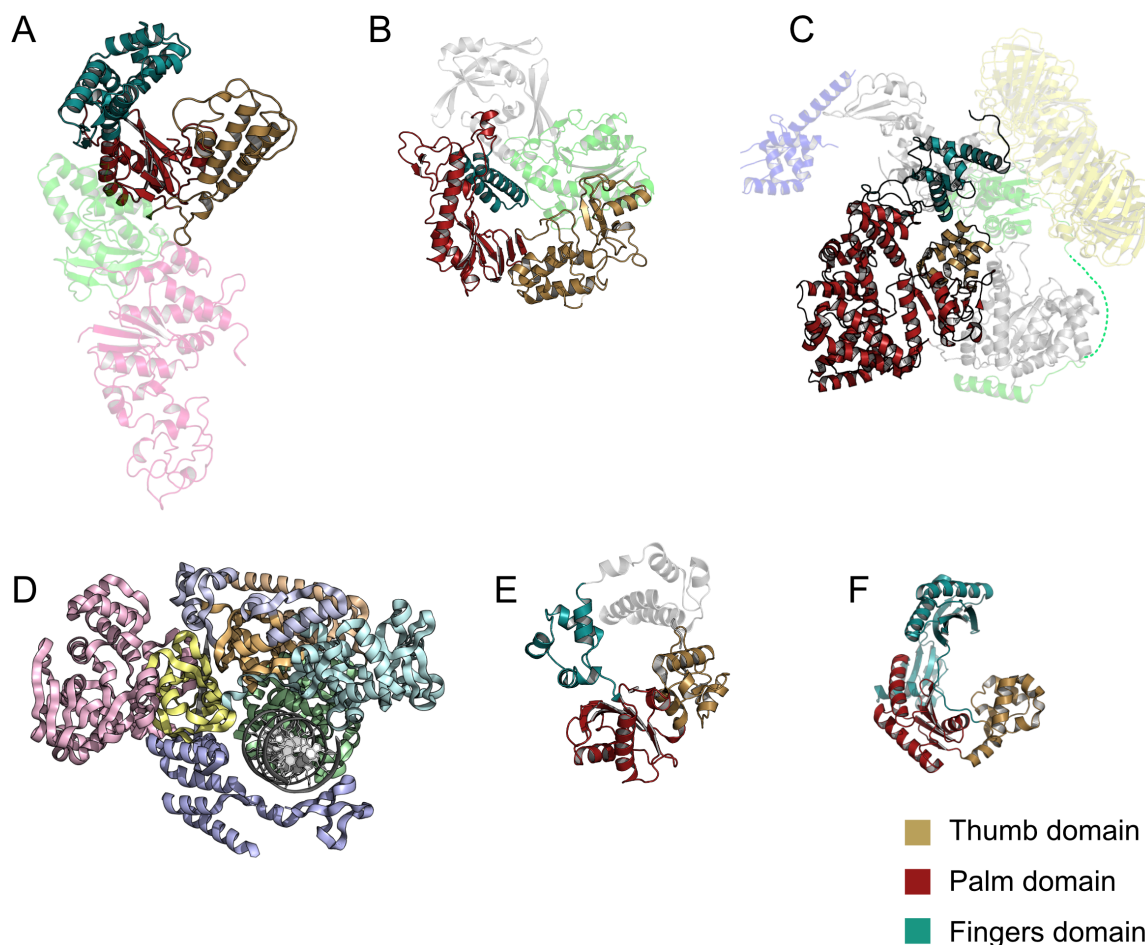
The current review focuses on development of enzymes which are compatible with the UBP-class of XNAs as this is an area where innovation in molecular tools is required, while research on enzymes which function with sugar-modified XNAs is, at present, more complete. The inherent nuclease-resistance of backbone-modified XNAs together with the enhanced duplex-forming properties of many of these variants make them extremely attractive for biotechnology; especially as bio-stable therapeutics [10,11]. As sugar-modified XNAs and their manipulation have been the subject of several excellent reviews in recent years [47–50], they will not be discussed in detail here; however, we point the reader to several key examples where compatibility of polymerase and ligase with backbone-modified XNA has been established [51].

## Nucleic acid polymerases

### Classes of DNA polymerase and their biological functions

DNA polymerases catalyse nucleic acid polymer synthesis by the addition of deoxyribonucleotide triphosphates (dNTPs) to the 3' terminus of a DNA or RNA primer [52]. In almost all cases, synthesis of the new strand is directed by the presence of a complementary antiparallel strand (template) which determines the insertion of new nt based on Watson–Crick base-pairing rules. However, as discussed below, some polymerases are capable of non-templated addition of nt to the 3' end of the primer.

All DNA polymerases, except the archaeal Class D, have a broadly conserved domain architecture which resembles a right hand [53] (Figure 3). The active site is situated in the 'palm' domain and contains two divalent cations, usually magnesium or manganese, coordinated by a pair of carboxylate side chains. Upon binding to the primer-template duplex, the mobile 'thumb' domain hinges into the minor groove, forming a tight binary complex interaction. The incoming dNTP is recruited via base pairing with the template. The 'finger' domains interrogate the geometry of this ternary complex for the correct base pairing and position the incoming dNTP for catalysis, ensuring that only a nucleotide complementary to the template is added [53]. Formation of the new phosphodiester bond between the 3'-hydroxyl of the primer (or preceding nucleotide) and the 5'-phosphate group of the incoming nucleotide begins when the nucleotide-binding metal ion coordinates the triphosphate moiety of the incoming dNTP, positioning it for nucleotide binding. The second catalytic metal ion in the active site lowers the p*K*<sub>a</sub> of the 3'-hydroxyl group on acceptor strand and positions it for inline nucleophilic attack on the  $\alpha$ -phosphate group of the incoming dNTP. The subsequent pentacoordinated bipyrimidal  $\alpha$ -phosphate transition state is resolved by the release of PP<sub>i</sub> and incorporation of the nucleotide [53–56]. Recent time-resolved crystallography studies have demonstrated participation of a third metal ion in some DNA polymerase mechanisms, where it binds immediately prior to phosphoryl transfer, lowering the energy of this step. This metal ion is not



**Figure 3: Representative DNA polymerase enzymes from each family, coloured by functional domains.**

(A) Family A: Okazaki-sealing Polymerase I from *Thermus aquaticus* (1TAQ). (B) Family B DNA polymerase: polymerase from *Thermococcus gorgonarius* (1TG0). (C) Family C DNA: Main replicative Polymerase III from *Escherichia coli* (5FKV). (D) Family D polymerase: DP2 subunit from *Pyrococcus abyssi* (6HMS). (E) Family X: Polymerase lambda from *Homo sapiens* (1XSP). (F) Family Y: DNA repair polymerase Dpo4 from *Sulfolobus solfataricus* (1SOM). Functionally equivalent domains that are found in all families except Family D are ‘fingers’ (cyan), ‘palm’ (red) and ‘thumb’ (gold) while additional regions of the main polymerase enzyme are coloured grey. Other domains with autonomous function include the 3’ to 5’ exonuclease (proofreading) domain of families A, B and C (green), the 5’ to 3’ (primer-removal domain) of Pol I (light pink) and the beta sliding clamp and tau domains of Pol III (yellow and blue respectively). Regions of the structures that do not form part of the ‘right hand’ polymerase scaffold are shown as translucent to better visualise the conserved architecture. The Polymerase D family enzymes lack this ‘right hand’ structure and domains are coloured arbitrarily. Figures were prepared using PyMol from experimentally determined protein structures with coordinates available in the protein data bank (PDB) under the corresponding codes.

coordinated by the polymerase active site, and its recruitment has been suggested to be rate-limiting in catalysis [57,58].

Despite this conserved catalytic mechanism and ‘right hand’ core architecture, there is considerable variation in the amino acid sequences and structural topologies among DNA polymerases which has been reviewed extensively [59–64] and is depicted in Figure 3. Within the Family A, B, C, X and Y classes of DNA polymerase, there is diversity in the folds of the core thumb, palm and fingers domains, as well as the presence of additional domains and subunits. These additional regions determine the enzymes’ processivity; a measure of how many nt are added sequentially per binding event, and its fidelity; the average rate of misincorporation of the incorrect nucleotide in an insertion event. The archaeal PolD polymerases are heterodimeric, containing a small 3’-5’ proof-reading exonuclease, and the larger catalytic subunit, DP2 (Figure 3D). DP2 does not exhibit the typical ‘right hand’ topology of other DNA polymerases and is considered to be representative of a new structural class which actually shares more structural homology with some RNA polymerases [65].

These different structural classes of polymerase also have varied taxonomic distributions and functions as summarised in Table 1 and described in detail in other reviews [59–64]. Briefly, Family A polymerases

include the bacterial polymerase I (PolI) enzyme as well as the polymerase subunit of the mitochondrial Pol  $\gamma$  and many viral polymerases. Family B polymerases comprise replicative polymerases of archaea and eukaryotes as well as bacterial PolII, which is involved in the DNA damage response of bacteria [63,66]. Family C is restricted to bacteria and provides the polymerase functionality to the PolymeraseIII (PolIII) holozyme, while Family D polymerases are unique to archaea where they participate in replication, though the details are not yet fully known [67]. Finally, Families X and Y polymerases primarily mediate translesion synthesis (TLS) and are found in all domains of life [60,68,69]. The wide range of polymerase structures and collection of appending subunits and domains reflect the diversity of biological requirements for DNA synthesis. The polymerase subunit of the replisome conducts the majority of primer extension during DNA replication and has both high processivity and fidelity to ensure accurate copying of genomic material [70,71]. Processivity is imparted by interaction of the polymerase subunit with the ring-shaped- $\beta$ -sliding clamp which encircles the DNA duplex [72,73]. Fidelity is enforced by a combination of accurate nucleotide insertion during catalysis and proofreading [74] by a 3'–5' exonuclease domain which is either encoded on the same polypeptide as the polymerase functionality (Pol I and Pol B Figure 3A and B shown in green) or as a separate subunit (Pol III Figure 3 shown in green) [75,76]. The PolI enzyme which assists with Okazaki fragment maturation as well as participating in high-fidelity DNA repair and recombination exhibits lower processivity than its replisome counterpart PolIII, and also possesses a unique 5' to 3' exonuclease domain for RNA primer removal [71]. In contrast to replicative polymerases, TLS polymerases, which are capable of replicating through damaged DNA typically, have both low fidelity and processivity as a consequence of structural features in the DNA binding region which allow them to bypass lesions [60,77]. Due to the mutagenic potential of this error-prone DNA synthesis, the expression of TLS polymerases is tightly regulated in all organisms [78,79].

## DNA polymerases for nucleic acid amplification

Due to their monomeric structure and high fidelity, most polymerases used in biotechnology are bacterial Family A PolI enzymes and archaeal Family B PolB enzymes [80,81]. Thermophilic polymerases are required for conventional thermocycled PCR applications, and while KlenTaq, the Klenow fragment of *T. aquaticus* PolI is still widely used, higher fidelity enzymes such as PolB from *Pyrococcus furiosus* have superseded it in many synthesis applications [82]. Isothermal amplification, which uses strand displacement by the polymerase rather than heat to melt the DNA duplex, has employed mesophilic polymerases such as the PolI Klenow fragment of *Geobacillus stearothermophilus* (Bst-Pol) [83]. Meanwhile, the ability to synthesise past lesions and the high error rate of TLS polymerases have been utilised for amplification of damaged DNA templates and introduction of random mutations for directed evolution [84].

Protein engineering strategies have been used to further tailor the properties of these natural polymerases. For example, fusion with additional DNA binding domains has been used to overcome the low inherent processivity of PolI and PolB [85,86]. Other properties such as salt and inhibitor tolerance have also been improved by rational design and random mutagenesis [87–89]. The ability of polymerase enzymes to amplify their own genes has made them especially amenable to directed evolution strategies, including compartmentalised self-replication (CSR) which enables the enzyme to replicate only its own encoding gene, giving rise to different variants. This method has produced KlenTaq polymerase mutants with increased thermostability and heparin resistance [90], as well as increased DNA binding affinity and resistance to inhibitors present in blood [91]. Additionally, various *Thermus* DNA polymerases have evolved capable of extending quadruple DNA mismatches and extending past DNA lesions [92]. A subsequent method called short-patch CSR (spCSR), where polymerases self-replicate a short region (or 'patch') of their encoding gene, enables diversification of specific motifs. Using this method, Taq polymerase mutants were obtained which had dual substrate specificity, enabling them to incorporate single ribonucleotides and deoxyribonucleotides with similar rates [93]. Further protein engineering has now evolved Taq variants capable of interconverting between DNA and C2'-OMe-modified oligonucleotides [94].

An emerging area of polymerase-mediated nucleic acid synthesis exploits the non-templated sequential nucleotide addition carried out by the terminal transferase activity of some polymerases. The most well-known of these enzymes are terminal deoxynucleotidyl transferase (TdT), PrimPol and polymerase  $\theta$ . TdT is a Family X DNA polymerase, involved in DNA repair via non-homologous end joining and its structure, function and biotechnological applications are well documented [64,95–97]. Non-templated

**Table 1: Families of DNA polymerases classified by structure.**

Family	Viral	Bacteria	Archaea	Eukaryotes	Example structure
A	T3 DNA pol	DNA pol I <sup>(i,ii)</sup>		DNA pol $\gamma$ <sup>(i, mitochondria)</sup>	1TAQ <i>T. aquaticus</i>
	T5 DNA pol			DNA pol $\theta$ <sup>(iii)</sup>	
	T7 DNA pol			DNA pol $\nu$ <sup>(iii)</sup>	
B	T4 DNA pol	DNA pol II <sup>(ii)</sup>	DNA pol BI <sup>(i,ii)</sup>	DNA pol $\alpha$ <sup>(i)</sup>	1TGO <i>Thermococcus gorgonarius</i>
			DNA pol BII <sup>(i,ii)</sup>	DNA pol $\delta$ <sup>(i)</sup>	
			DNA pol BIII <sup>(i,ii)</sup>	DNA pol $\epsilon$ <sup>(i)</sup> DNA pol $\zeta$ <sup>(iii)</sup>	
C		DNA pol III <sup>(i)</sup>			5FKV <i>E. coli</i>
		DNA pol E <sup>(i, Bacillus subtilis)</sup>			
D			DNA pol D <sup>(i)</sup>		6HMS <i>P. abyssi</i>
X	ASFV <sup>c</sup> DNA pol	DNA pol X <sup>(iv)</sup>	DNA pol X <sup>(iv)</sup>	DNA pol $\beta$ <sup>(ii)</sup>	1XSP <i>Homo sapiens</i>
				DNA pol $\lambda$ <sup>(i,iii)</sup>	
				DNA pol $\mu$ <sup>(iii,iv)</sup>	
				DNA pol $\sigma$ <sup>(iii)</sup>	
				TdT <sup>(iv)</sup>	
Y		DNA pol IV <sup>(iv)</sup>	Dpo4 DNA <sup>(iv)</sup>	DNA pol $\eta$ <sup>(iii)</sup>	1SOM <i>S. solfataricus</i>
		DNA pol V <sup>(iv)</sup>	Dbh DNA <sup>(iv)</sup>	DNA pol $\kappa$ <sup>(iii)</sup> DNA pol $\iota$ <sup>(iii)</sup>	

Representative examples are given for each taxonomic kingdom as well as viruses. For additional information, we direct the reader to comprehensive reviews on structural and taxonomic classifications [59–64]. Polymerases are denoted as being involved in (i) DNA replication, (ii) High-fidelity DNA repair and recombination, (iii) TLS, (iv) Template-independent extension.

DNA synthesis is initiated by TdT binding to a DNA primer and sequentially adding random dNTPs to the 3'-hydroxyl end of the previous nucleotide [64]. A recent innovation uses a multiplexed version of TdT-mediated, template-independent DNA synthesis to 'print' user-defined nt on an inkjet-like platform, presenting the possibility for in-house gene writing in the near future [98]. The use of PrimPol and polymerase  $\theta$  as enzymatic tools for DNA synthesis is less reported on, although both been shown to be able to engage in template-independent transferase activity in the presence of manganese making them interesting potential candidates for use in future DNA synthesis applications [99,100].

## Polymerisation with UBP-XNAs

Polymerase-based amplification of UBP-XNA requires the enzyme to act on UBP-containing substrates with similar incorporation efficiency, fidelity and processivity to natural nucleic acids [101]. PCR reactions with natural bases can have an error rate of  $10^{-5}$  or  $10^{-6}$ , meaning one mistake for every 100,000 to 1,000,000 incorporation events, when using a family A polymerase such as Klen-Taq with an associated 3'-5' exonuclease activity [102]. The ability of natural DNA polymerases to replicate UBP-XNA substrates with varying success has been the subject of reviews which have analysed the structural basis for differences in UBP incorporation [28,103,104]. In general, natural DNA polymerases are more tolerant to base alterations situated in the major groove of the DNA double helix, and if the shape of the UBPs remains similar to that of the canonical base pairs [28,105]. High-fidelity *in vitro* replication of both hydrogen-bonded and hydrophobic UBPs has been described using both Family A and Family B DNA polymerases; however, achieving efficiency, fidelity and processivity at similar levels observed with natural nucleic acid substrates remains elusive [16,17,44,46,106,107]. For the hydrogen-bonded Z:P base pair, fidelity is lowered when using natural DNA polymerases due to the replacement of these UBPs with canonical G:C base pairs over sequential rounds of PCR due to mispairing [17]. With the S:B pair, approximately 10% of S bases



exist in an alternate tautomeric form [108] and so mispair with the canonical base T causing an S to A transition under PCR conditions [42]. Due to their lack of hydrogen bonds, the hydrophobic UBPs do not easily mispair with natural bases; however, natural DNA polymerases are unable to incorporate sequential hydrophobic UBPs when synthesising polynucleotides. For example, at least six natural bases are needed to separate the multiple Ds:Px pairs in order to achieve efficient amplification [46] while for the 5SICS:NAM UBP, incorporation of only two adjacent unnatural bases was possible before the DNA polymerase was challenged [107]. In contrast, stretches of up to four Z:P UBPs can be incorporated by DNA polymerases with high fidelity and efficiency [109].

Engineered versions of DNA polymerases provided some progress in improving UBP-XNA amplification [110]. Evolution of the Family A KlenTaq polymerase increased the incorporation efficiency of dZ when using dP as a template; however, the evolved polymerase was less able to incorporate dP relative to dZ [111]. The nuclease-deficient, N-terminal truncated Taq mutant, TiTaq, was used in PCR reactions with a high ratio of non-natural dS and dB nt to natural nt, which significantly increased correct incorporation of the dS nt and resulted in ~96% fidelity [112]. High resolution crystal structures of polymerases interacting with UBP-XNAs, coupled with computational studies have also provided considerable insight into the requirements of engineering a truly successful UBP-XNA polymerase [101,104]. The crystal structures of a mutant KlenTaq polymerase in a pre-incorporation ternary complex with dZTP opposite dP, and in a post-incorporation binary complex revealed that for native DNA polymerases, it is the post-incorporation product binding, with Z:P in the active site which causes clashes in both the primer and template strands and challenges the DNA polymerase [113]. Therefore, evolved DNA polymerases which can overcome this complication in the post-incorporation step may be beneficial. A crystal structure of native KlenTaq DNA in a ternary complex with the UBP dDs:dPxTP showed that the UBPs pair in the active site in a coplanar conformation, very similar to that of Watson–Crick base pairs; however, the increased width and height of the UBPs result in the O-helix of the finger domain not closing as tightly, which may affect the overall incorporation efficiency [101,114,115]. Molecular dynamic (MD) simulations also offer insight into the effects various mutations could have on WT polymerases' abilities to synthesise UBP polynucleotides. MD simulations of an evolved KlenTaq polymerase variant with four amino acid substitutions showed that compared to the WT KlenTaq, this enzyme exhibited an altered flexibility which allowed it to more appropriately position the P:Z UBP for catalysis [101,114,115]. Strikingly, all four substitutions were located outside of the active site as illustrated in Ouaray et al. [115].

## Nucleic acid ligases

### Classes of DNA ligase and their biological functions

DNA ligases catalyse the formation of a phosphodiester bond between the adjacent 5' phosphate (5'P) and 3' hydroxyl (3' OH) in the backbone of double-stranded DNA. This three-step reaction is powered by a nucleotide cofactor, either adenosine triphosphate (ATP) or nicotinamide adenine dinucleotide (NAD). In the first step, the ligase enzyme self-adenylates at a conserved catalytic lysine residue, forming a covalent linkage between the  $\epsilon$ -amino group of the amino acid side chain and the  $\alpha$ -phosphate of AMP [116]. In the second step, the enzyme–adenylate complex binds to the nicked DNA duplex and the AMP is transferred to the 5'-PO<sub>4</sub> group of the DNA break. In the third and final step, the 3' OH group is positioned for in-line nucleophilic attack on the activated 5'-PO<sub>4</sub> group of the DNA–adenylate which liberates AMP from the active site and forms a new phosphodiester bond in the backbone [116].

DNA ligases are centrally involved in DNA recombination, repair and replication, all of which require the formation of new phosphodiester bonds in the DNA duplex [117]. DNA ligases can be categorised as either NAD-dependent or ATP-dependent based on the cofactor they utilise during step 1 enzyme activation [116]. Both forms have a catalytic core that consists of nucleotidyl transferase (NTase) domain and an oligobinding (OB) domain, joined by a flexible linker that enables conformational changes (Figure 4) [116,118,119]. The NTase domain, also known as the adenylation domain, contains the ligase active site including the catalytic lysine residue in the conserved KxDGxR motif. The OB domain is essential for binding the substrate DNA and rotates 180° around the interdomain linker region during engagement, inserting into the major groove to distort the duplex for catalysis [116,118,120]. NAD-dependent isoforms possess highly conserved appending domains in addition to this conserved core architecture, including a specialised Ia domain for NAD utilisation (Figure 4A, blue) as well as Zinc-finger (Figure 4A, green), BRCT and helix-turn-helix domains for

DNA binding [121]. In contrast, ATP-dependent DNA ligases vary widely in their size and domain composition [118,120,122]. The simplest functional DNA ligases are of the ATP-dependent class and comprise only the core NTase and OB domains including many bacteriophage and viral isoforms [123,124] as well as the recently described Ligase-E class found in a wide range of proteobacteria [118]. The most common ATP-dependent DNA ligase architecture however has an alpha-helical domain N-terminal to the NTase domain, which assists with DNA binding and ligation efficiency (Figure 4B(i) and (ii)).

ATP- and NAD-dependent ligase isoforms have distinct taxonomic distributions with the NAD-dependent Lig-A enzyme being almost entirely restricted to bacteria where it is the essential ligase that joins Okazaki fragments during DNA replication (Table 2). Archaea and eukaryotes by contrast exclusively use ATP-dependent DNA ligases for replication. In some cases, a single ligase isoform is used for both DNA repair and replication; however, many organisms encode one or more specialised DNA repair ligases in their genomes which are usually of the ATP-dependent class. The most notable example of these are the multi-functional ‘Lig-D’ repair enzymes widely distributed among bacteria which possess auxiliary enzymatic domains with autonomous polymerase or nuclease activities that repair damaged DNA ends in advance of their re-ligation (Figure 4C, light blue and pink, respectively).

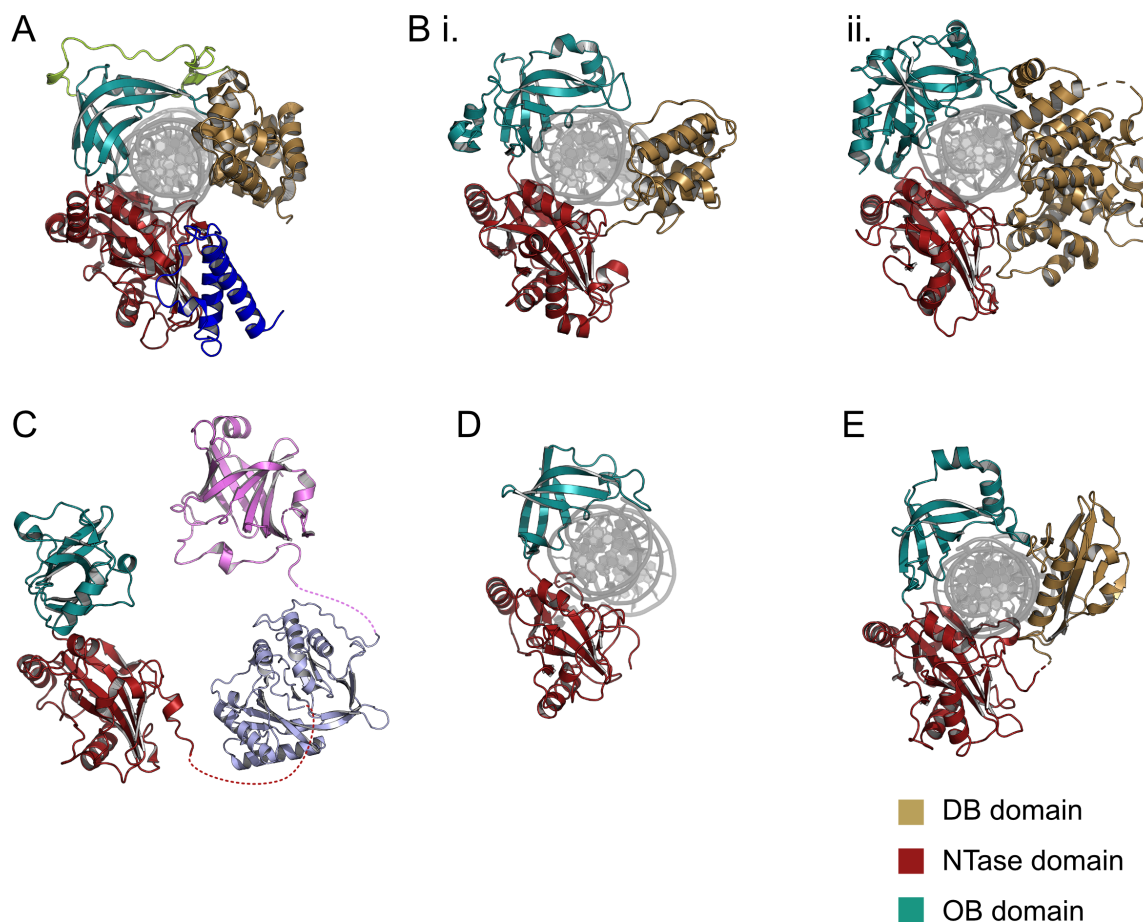
## Use of ligases in DNA assembly and synthesis

The discovery of the DNA ligase was foundational for developing recombinational DNA production via traditional restriction–ligation cloning protocols, with the T4 DNA ligase being the most widely used [131]. DNA ligases remain central to assembly protocols such as Golden-Gate Assembly (Type IIS cloning) and Gibson assembly (isothermal assembly) to join the backbones of annealed fragments. The advantage of both Golden-Gate and Gibson relative to traditional restriction–ligation is that a single reaction can assemble large number of fragments (up to 15 with Gibson and 35 with Golden-Gate) into constructs of several hundred kilobases in a ‘scarless’ manner which avoids disruption of sequences features by vestigial restriction sites [132,133]. Key to the performance of the DNA ligases in both methods is the balance between bias (preference for ligating some sequences over others), fidelity (likelihood of joining mismatched ends) and efficiency [134,135]. For example, comparison of T4 and T7 DNA ligases found that while the latter produced a higher number of correct Golden-Gate assemblies, its bias against joining some sequences and its overall lower activity made the former a more suitable choice in this application [136]. A further variation on these assembly methods, the ligase cycling reaction allows direct splicing of DNA duplexes by employing a thermostable DNA ligase and a short bridging oligonucleotide. This generates a ligatable nicked intermediate when annealed to heat-denatured duplex parts and has the advantage of allowing separate gene blocks to be joined without requiring addition of extensions or homologous overhangs [137,138].

Besides assembly, ligation-mediated de novo synthesis of gene-length DNA duplexes has been explored by isothermal ligation of short octameric synthetic duplexes. The duplexes are generated by annealing 5′ phosphorylated oligonucleotides in a sequence-specific manner that leaves 4 nt ligatable overhangs resulting in products of almost 100 nt [139]. A ligation-based approach has also been used to synthesise functionalised polynucleotides, nt containing varying substituents on their nitrogenous bases. Here, an approach similar to PCR was utilised with an extendable primer annealed to a template strand which directed primer extension. Functionalised trinucleotides with a 5′ phosphate were then ligated to the primer terminus in a template-directed manner, resulting in a functionalised strand built on the natural complement. Further, this functionalised strand could be PCR amplified using DNA polymerase, enabling the enrichment of particular sequences under some selection pressure such as substrate binding [140].

## Ligation with UBP-XNAs

The use of DNA ligases to synthesise or assemble UBP-XNA has been less explored relative to the extensive work done with DNA polymerases with these substrates. Early studies on the feasibility of DNA repair involving hydrogen-bonded UPBs demonstrated that both Taq and DNA T4 ligases were able to join DNA with consecutive S:B base pairs, although the latter was more efficient [141]. T4 DNA ligase has also been shown to join single hydrophobic pairs including the Ds:Pa combination with extremely high selectivity in both a single-stranded nick and a double-stranded cohesive overhang context [142]. Recently, an innovative strategy for reading and writing DNA containing expanded base-pair alphabets was developed which utilised T4-mediated ligation as a key step in XNA synthesis [143]. Here, untemplated tailing using



**Figure 4: Representative DNA ligase enzymes from each family, coloured by functional domains; DNA binding domain (gold), NTase domain (red), oligonucleotide binding domain (teal).**

(A) Ligase A, replicative DNA ligase from *E. coli* (2owo). (B) Helical DNA binding domain: T4 bacteriophage DNA ligase (6dt1). (C) Ligase D: Double-strand break repair ligase; DNA ligase and PrimPol (light blue) subunits from *Mycobacterium tuberculosis* (6NHZ and 2IRU); phosphoesterase subunit (pink) from *Pseudomonas aeruginosa* (3N9B). (D) Minimal structure DNA ligase without DNA binding domain, Ligase E from *Alteromonas mediterranea* (6GDR). (E) Ligases without structural classification; African Swine Fever virus with beta-sheet DNA binding domain (6IMN). Figures were prepared using PyMol from experimentally determined protein structures with coordinates available in the PDB under the corresponding codes.

the small exo fragment of DNA PolI generated cohesive single UBP overhangs which were then ligated to form the new duplex. This approach was used to introduce a range of individual hydrogen-bonded UBPs into a natural sequence context, which were then able to be sequenced using nanopore technology. Crucially, several of the UBPs introduced enzymatically are intractable to oligonucleotide incorporation via phosphoramidite synthesis due to their chemical lability. This process is, in principle, amenable to the addition of multiple UBPs via iterative rounds of cleavage, tailing and re-ligation, provided the enzymes involved will tolerate consecutive tracts of UBPs.

The potential of DNA ligases to deliver XNA with derivatives on sugar moiety has been widely examined and provides an example of the utility of this approach which may also be applied to UBP-XNA [51,144]. Commercially available DNA ligases T3, T4 and T7 were demonstrated to join nicked duplexes containing TNA and 2'OMe RNA, with the latter XNA also being ligated by the *Chlorella* virus (SplintR) enzyme [51,144]. All enzymes showed a marked bias against DNA-XNA joins with the XNA strand in the donor position providing the 5' phosphate to the join relative to the converse XNA-DNA acceptor-donor configuration. Among this small panel, the T4 DNA ligase proved the most effective and was demonstrated to ligate a wide range of backbone-XNA substrates, including 2'-fluoroarabino nucleic acids (FANA), 1,5-dianhydrohexitol nucleic acids (HNA) and locked nucleic acid (LNA) in the presence of crowding agents Pe.g.8000 and betaine [144]. Beyond optimisation of buffer, crowding agents and cofactor concentrations, attempts have been made to improve the enzymatic ligation of modified backbone

**Table 2: Families of DNA ligase classified by structure.**

Family	Viral	Bacteria	Archaea	Eukaryotes	Example structure
NAD-dependent	<i>AmEPV (amsacta moorei</i> entomopoxvirus)	Lig A <sup>(i)</sup>	LigN <sup>(i)</sup> ( <i>Haloferax</i> <i>volcanii</i> )	–	<i>E. coli</i> lig A 2ow0 [121]
Helical DBD	T4 DNA ligase ( <i>Enterobacteria</i> <i>phage T4</i> )	Lig B <sup>(i)</sup>	Replicative ligase <sup>(i)</sup>	Lig I <sup>(i)</sup> Lig III <sup>(i)</sup> Lig IV <sup>(i)</sup>	T4 DNA ligase 6DT1 [125] Human lig I 1 × 9N [126]
Lig D	–	Lig D1 <sup>(ii)</sup> Lig D2 <sup>(ii)</sup> Lig D3 <sup>(ii)</sup>	–	–	Ligase domain from <i>M.</i> <i>tuberculosis</i> 6NHZ [127] Phosphoesterase domain from <i>P. aeruginosa</i> 3N9B [128] PrimPol from <i>M. tuberculosis</i> domain [129] 2IRU
Minimal	<i>Chlorella</i> virus–ligase	Lig C <sup>(ii)</sup> Lig E	–	–	<i>A. mediterranea</i> Lig E 6GDR [118]
Unclassified	Swine fever virus	Lig W ( <i>Prochlorococcus marinus</i> )	–	–	African swine fever virus ligase 6IMN [130]

Representative examples are given for each taxonomic kingdom as well as viruses. For additional information, we direct the reader to comprehensive reviews on structural and taxonomic classifications [116,118–120]. Ligases are denoted as being involved in (i) DNA replication or (ii) DNA repair.

XNAs by rational engineering. The *Chlorella* virus ligase was selected as a scaffold to improve joining of hexitol-based XNA nt (HNA) based on its small size and well-characterised structure. Computational prediction of the XNA–ligase complex indicated that lengthening the inter-domain hinge region would better accommodate the wider diameter of the HNA duplex and guided the insertion of a single glycine residue in the interdomain linker which was then able to ligate various combinations of 2'OMe RNA and HNA, unlike wild-type controls [145]. Other modifications such as fusion of processivity domains to otherwise wild-type enzymes have proved less effective in increasing backbone–XNA ligation [144].

## The future of UBP–XNA-joining enzymes and concluding remarks

Despite the considerable advances in synthesising and characterising XNA with expanded base-pair alphabets, many of the key enzymes in the conventional molecular biologist's toolbox exhibit sub-par performance when challenged with XNA containing UBPs. To provide superior UBP–XNA joining enzymes, we argue for an integrated discovery cycle that explores the UBP–XNA compatibility of DNA ligases and polymerases from existing biological diversity, coupled to further engineering-based improvements.

Current UBP–XNA compatible enzymes have been largely based on conventional molecular biology enzymes; in many examples employing the commercially available version as a source. While this is a logical starting point to build from, it means that much of the structural and functional diversities among these enzymes classes remain unexplored in the context of UBP–XNA enzymology. For example, most polymerases used in molecular biology are derived from the PolA and B families and most UBP–XNA tools are variants of the KlenTaq enzymes; likewise, all XNA ligation (backbone and UBP) has focused on a narrow range of phage enzymes with most work being on the T4–DNA ligase. However, as described above, there is an enormous diversity of ligase scaffolds found in nature, many of which have only been minimally studied, especially in the context of UBP–XNA.

Recombinant enzymes spanning the natural polymerase/ligase diversity need to be empirically and systematically tested with a range of standardised UBP–XNA substrates and the results mapped back to their sequence and structural features. This will be essential to build a better picture of which features promote or inhibit UBP–XNA joining, and inform future discovery initiatives, as well as provide guidelines

to rational engineering endeavours. Catalytic benchmarks should include not only the efficiency of amplification/ligation but also quantify fidelity and bias as these will have a direct bearing on the suitability of the enzymes in different applications. It is important that such benchmarking is undertaken using conditions that reflect those used in the UBP-XNA assembly applications and take account of factors such as sequence bias. In this regard, assays utilising sequence read-out that can be conducted in a multiplex format such as those used to profile DNA ligase fidelity would prove useful [134–136]. The replication machinery of extremophilic microbes represents another possible avenue for discovering novel molecular biology enzymes, which may have a particular benefit to UBP-XNA compatibility. Among the DNA polymerases and ligases described above that join UBP substrates, the majority are meso- and thermophilic, leaving psychrophilic variants relatively under-explored. In part, this reflects the fact that truly psychrophilic enzymes tend to be difficult to characterise in laboratory settings due to their innate instability and heat-lability. However, psychrophilic enzymes also hold biotechnological and industrial value [146,147], especially, in the case of polymerases, if they can be engineered to perform strand separation at low temperatures as this allows them to be used in isothermal amplification protocols [148].

We would also argue that although considerable progress has been made in improving polymerase performance with UBP-XNAs, ligation-based strategies such as those demonstrated for modified backbone XNAs [51,144] should be more thoroughly explored. This could be accomplished either as the sole step of UBP-XNA synthesis using approaches akin to Kawabe et al. [143] or by assembly of longer tracts which have been amplified by a UBP-XNA-compatible polymerase. This would overcome limitations in current polymerase enzymes towards insertion of multiple UBP bases. Implementing such a strategy will, however, require engineering endeavours for DNA ligase enzymes on-par with those directed, to date, more thoroughly at the DNA polymerase enzymes. In the case of the DNA ligases, a structure-centric approach will be necessary as there are currently no simple cell-free directed evolution protocols for DNA ligases that are amenable to incorporation of UBP-XNAs. Ideally, UBP-XNA-ligase engineering will use experimentally determined structures of the enzyme-XNA complex or alternately enzyme-DNA on which the UBP-XNA can be modelled and then analysed in molecular dynamics simulations. Although advances in structural modelling using artificial intelligence have recently been extended to prediction of protein-nucleic acid complexes [149,150], these are unable to model modifications or reaction intermediates, and are known to produce spurious results which could limit their utility in structure-guided engineering.

Finally, to realise the true potential of UBP-XNAs, the compatible enzymology must extend beyond UBP-XNA joining to include other activities as well. For example, incorporating UBPs to generate synthetic/semi-synthetic organisms will require either compatibility of the entire replication machinery including helicases, topoisomerases, initiation and termination factors with the unnatural bases or supplementation of the native enzymes with co-expressed UBP-XNA compatible homologs. Furthermore, incorporation of additional amino acids into an expanded genetic code will require tRNA synthetases capable of charging UBP-XNA-containing anticodons with unnatural amino acids. For stable error-free replication, damage and mismatch repair systems will be necessary and could also improve the accuracy of synthetic UBP-XNA. Arguably one of the most important emerging applications for XNA is the deployment of ribo-XNA for pharmaceutical applications. Although this review has focused on the 2'-deoxy-derivatives of these polymers, the central role of the modified N1-methyl-pseudouridine nucleotide in decreasing immunogenicity of the Covid19 mRNA vaccines [22,23] underscores the utility of XNAs. It also seems likely that enzymes replicating and ligating ribo-XNA will represent the next frontier of this area.

## Funding

This work was supported by the Ministry of Business, Innovation and Employment New Zealand 'Smart Idea' grant (UoWX2205 to CB, CW-F, NR and AW), UKRI Biotechnology and Biological Sciences Research Council Award BB/P018017/1 (NR), UKRI Biotechnology and Biological Sciences Research Council Partnering Award BB/V018094/1 (NR and AW). CB is supported by Waikato Graduate Womens' Educational Trust Doctoral Merit Award, AW is supported by a Rutherford Discovery Fellowship (20-UOW-004).

## CRedit Author Contribution

CB, CW-F and AW conceived the idea for the manuscript, CB and CW-F drafted the initial manuscript, AW and NR reviewed and edited the manuscript. All authors approved the submitted version.

## Abbreviations

AEGIS, Artificially Expanded Genetic Information Systems; ATP, adenosine triphosphate; BPNA, boranophosphate; CSR, compartmentalised self-replication; DMT, dimethoxytrityl; FANA, fluoroarabino nucleic acids; HNA, hexitol nucleic acid; LNA, locked nucleic acid; MD, Molecular dynamic; NAD, nicotinamide adenine dinucleotide; NTase, nucleotidyl transferase; OB, oligobinding; PDB, protein data bank; Poll, polymerase I; PolIII, PolymeraseIII; SELEX, systematic evolution of ligands by exponential enrichment; TLS, translesion synthesis; TNA, threose nucleic acid; TdT, terminal deoxy nucleotidyl transferase; TiEOS, Template-independent enzymatic oligonucleotide synthesis; UBP-XNAs, unnatural nucleobase pairs; XNA, Xenobiotic Nucleic Acid; dNTPs, deoxy ribonucleotide triphosphates; nt, nucleotides; pHNA, phosphonate nucleic acid; spCSR, short-patch.

## References

- 1 Vybornyi, M., Vyborna, Y. and Häner, R. (2019) DNA-inspired oligomers: from oligophosphates to functional materials. *Chem. Soc. Rev.* **48**, 4347–4360 <https://doi.org/10.1039/c8cs00662h>
- 2 Madsen, M. and Gothelf, K.V. (2019) Chemistries for DNA nanotechnology. *Chem. Rev.* **119**, 6384–6458 <https://doi.org/10.1021/acs.chemrev.8b00570>
- 3 Hughes, R.A. and Ellington, A.D. (2017) Synthetic DNA synthesis and assembly: putting the synthetic in synthetic biology. *Cold Spring Harb. Perspect. Biol.* **9**, a023812 <https://doi.org/10.1101/cshperspect.a023812>
- 4 Sproat, B.S. (1995) Chemistry and applications of oligonucleotide analogues. *J. Biotechnol.* **41**, 221–238 [https://doi.org/10.1016/0168-1656\(95\)00012-f](https://doi.org/10.1016/0168-1656(95)00012-f)
- 5 Lönnberg, H. (2017) Synthesis of oligonucleotides on a soluble support. *Beilstein J. Org. Chem.* **13**, 1368–1387 <https://doi.org/10.3762/bjoc.13.134>
- 6 Kosuri, S. and Church, G.M. (2014) Large-scale de novo DNA synthesis: technologies and applications. *Nat. Methods.* **11**, 499–507 <https://doi.org/10.1038/nmeth.2918>
- 7 Efcavitch, J.W. and Heiner, C. (1985) Depurination as a yield decreasing mechanism in oligodeoxynucleotide synthesis. *Nucleos. Nucleot.* **4**, 267–267 <https://doi.org/10.1080/07328318508077883>
- 8 Hoose, A., Vellacott, R., Storch, M., Freemont, P.S. and Ryadnov, M.G. (2023) DNA synthesis technologies to close the gene writing gap. *Nat. Rev. Chem.* **7**, 144–161 <https://doi.org/10.1038/s41570-022-00456-9>
- 9 Pichon, M. and Hollenstein, M. (2024) Controlled enzymatic synthesis of oligonucleotides. *Commun. Chem.* **7**, 138 <https://doi.org/10.1038/s42004-024-01216-0>
- 10 Herdewijn, P. and Marlière, P. (2009) Toward safe genetically modified organisms through the chemical diversification of nucleic acids. *Chem. Biodivers.* **6**, 791–808 <https://doi.org/10.1002/cbdv.200900083>
- 11 Duffy, K., Arangundy-Franklin, S. and Holliger, P. (2020) Modified nucleic acids: replication, evolution, and next-generation therapeutics. *BMC Biol.* **18**, 112 <https://doi.org/10.1186/s12915-020-00803-6>
- 12 Chaput, J.C. and Herdewijn, P. (2019) What Is XNA? *Angew. Chem. Int. Ed.* **58**, 11570–11572 <https://doi.org/10.1002/anie.201905999>
- 13 Kranaster, R. and Marx, A. (2010) Engineered DNA polymerases in biotechnology. *Chembiochem* **11**, 2077–2084 <https://doi.org/10.1002/cbic.201000215>
- 14 Coulther, T.A., Stern, H.R. and Beuning, P.J. (2019) Engineering polymerases for new functions. *Trends Biotechnol.* **37**, 1091–1103 <https://doi.org/10.1016/j.tibtech.2019.03.011>
- 15 Laos, R., Thomson, J.M. and Benner, S.A. (2014) DNA polymerases engineered by directed evolution to incorporate non-standard nucleotides. *Front. Microbiol.* **5**, 565 <https://doi.org/10.3389/fmicb.2014.00565>
- 16 Yamashige, R., Kimoto, M., Takezawa, Y., Sato, A., Mitsui, T., Yokoyama, S. et al. (2012) Highly specific unnatural base pair systems as a third base pair for PCR amplification. *Nucleic Acids Res.* **40**, 2793–2806 <https://doi.org/10.1093/nar/gkr1068>
- 17 Yang, Z., Sismour, A.M., Sheng, P., Puskar, N.L. and Benner, S.A. (2007) Enzymatic incorporation of a third nucleobase pair. *Nucleic Acids Res.* **35**, 4238–4249 <https://doi.org/10.1093/nar/gkm395>
- 18 Sismour, A.M. and Benner, S.A. (2005) The use of thymidine analogs to improve the replication of an extra DNA base pair: a synthetic biological system. *Nucleic Acids Res.* **33**, 5640–5646 <https://doi.org/10.1093/nar/gki873>
- 19 Christensen, T.A., Lee, K.Y., Gottlieb, S.Z.P., Carrier, M.B. and Leconte, A.M. (2022) Mutant polymerases capable of 2' fluoro-modified nucleic acid synthesis and amplification with improved accuracy. *RSC Chem. Biol.* **3**, 1044–1051 <https://doi.org/10.1039/d2cb00064d>
- 20 Marlière, P., Patrouix, J., Döring, V., Herdewijn, P., Tricot, S., Cruveiller, S. et al. (2011) Chemical evolution of a bacterium's genome. *Angew. Chem. Int. Ed. Engl.* **50**, 7109–7114 <https://doi.org/10.1002/anie.201100535>
- 21 Nguyen, H., Abramov, M., Rozenski, J., Eremeeva, E. and Herdewijn, P. (2022) In vivo assembly and expression of DNA containing non-canonical bases in the yeast *Saccharomyces cerevisiae*. *Chembiochem* **23**, e202200060 <https://doi.org/10.1002/cbic.202200060>
- 22 Karikó, K., Buckstein, M., Ni, H. and Weissman, D. (2005) Suppression of RNA recognition by Toll-like receptors: the impact of nucleoside modification and the evolutionary origin of RNA. *Immunity* **23**, 165–175 <https://doi.org/10.1016/j.immuni.2005.06.008>
- 23 Sahin, U., Muik, A., Derhovanessian, E., Vogler, I., Kranz, L.M., Vormehr, M. et al. (2020) COVID-19 vaccine BNT162b1 elicits human antibody and TH1 T cell responses. *Nat. New Biol.* **586**, 594–599 <https://doi.org/10.1038/s41586-020-2814-7>
- 24 Eremeeva, E. and Herdewijn, P. (2019) Non canonical genetic material. *Curr. Opin. Biotechnol.* **57**, 25–33 <https://doi.org/10.1016/j.copbio.2018.12.001>
- 25 Malyshev, D.A. and Romesberg, F.E. (2015) The expanded genetic alphabet. *Angew. Chem. Int. Ed. Engl.* **54**, 11930–11944 <https://doi.org/10.1002/anie.201502890>
- 26 Benner, S.A. (2023) Rethinking nucleic acids from their origins to their applications. *Philos. Trans. R. Soc. Lond., B, Biol. Sci.* **378**, 20220027 <https://doi.org/10.1098/rstb.2022.0027>

- 27 Hamashima, K., Kimoto, M. and Hirao, I. (2018) Creation of unnatural base pairs for genetic alphabet expansion toward synthetic xenobiology. *Curr. Opin. Chem. Biol.* **46**, 108–114 <https://doi.org/10.1016/j.cbpa.2018.07.017>
- 28 Pinheiro, V.B. and Holliger, P. (2014) Towards XNA nanotechnology: new materials from synthetic genetic polymers. *Trends Biotechnol.* **32**, 321–328 <https://doi.org/10.1016/j.tibtech.2014.03.010>
- 29 Morandi, L., Ferrari, D., Lombardo, C., Pession, A. and Tallini, G. (2007) Monitoring HCV RNA viral load by locked nucleic acid molecular beacons real time PCR. *J. Virol. Methods* **140**, 148–154 <https://doi.org/10.1016/j.jviromet.2006.11.014>
- 30 Alshaer, W., Hillaireau, H., Vergnaud, J., Ismail, S. and Fattal, E. (2015) Functionalizing liposomes with anti-CD44 aptamer for selective targeting of cancer cells. *Bioconjug. Chem.* **26**, 1307–1313 <https://doi.org/10.1021/bc5004313>
- 31 Catuogno, S., Di Martino, M.T., Nuzzo, S., Esposito, C.L. and Tassone, P. (2019) An anti-BCMA RNA aptamer for miRNA intracellular delivery. *Mol. Ther. Nucleic Acids* **18**, 981–990 <https://doi.org/10.1016/j.omtn.2019.10.021>
- 32 Alves Ferreira-Bravo, I. and DeStefano, J.J. (2021) Xeno-nucleic acid (XNA) 2'-fluoro-arabino nucleic acid (FANA) aptamers to the receptor-binding domain of SARS-CoV-2 S protein block ACE2 binding. *Viruses* **13**, 1983 <https://doi.org/10.3390/v13101983>
- 33 Biondi, E., Lane, J.D., Das, D., Dasgupta, S., Piccirilli, J.A., Hoshika, S. et al. (2016) Laboratory evolution of artificially expanded DNA gives redesignable aptamers that target the toxic form of anthrax protective antigen. *Nucleic Acids Res.* **44**, 9565–9577 <https://doi.org/10.1093/nar/gkw890>
- 34 Kohlberger, M. and Gadermaier, G. (2022) SELEX: Critical factors and optimization strategies for successful aptamer selection. *Biotechnol. Appl. Biochem.* **69**, 1771–1792, <https://doi.org/10.1002/bab.2244>
- 35 Sefah, K., Yang, Z., Bradley, K.M., Hoshika, S., Jiménez, E., Zhang, L. et al. (2014) In vitro selection with artificial expanded genetic information systems. *Proc. Natl. Acad. Sci. U.S.A.* **111**, 1449–1454 <https://doi.org/10.1073/pnas.1311778111>
- 36 Zhang, L., Wang, S., Yang, Z., Hoshika, S., Xie, S., Li, J. et al. (2020) An aptamer-nanotrain assembled from six-letter DNA delivers doxorubicin selectively to liver cancer cells. *Angew. Chem. Int. Ed. Engl.* **59**, 663–668 <https://doi.org/10.1002/anie.201909691>
- 37 Zumrut, H., Yang, Z., Williams, N., Arizala, J., Batool, S., Benner, S.A. et al. (2020) Ligand-guided selection with artificially expanded genetic information systems against TCR-CD3ε. *Biochemistry* **59**, 552–562 <https://doi.org/10.1021/acs.biochem.9b00919>
- 38 Georgiadis, M.M., Singh, I., Kellett, W.F., Hoshika, S., Benner, S.A. and Richards, N.G.J. (2015) Structural basis for a six nucleotide genetic alphabet. *J. Am. Chem. Soc.* **137**, 6947–6955 <https://doi.org/10.1021/jacs.5b03482>
- 39 Hoshika, S., Leal, N.A., Kim, M.-J., Kim, M.-S., Karalkar, N.B., Kim, H.-J. et al. (2019) Hachimoji DNA and RNA: a genetic system with eight building blocks. *Science* **363**, 884–887 <https://doi.org/10.1126/science.aat0971>
- 40 Switzer, C.Y., Moroney, S.E. and Benner, S.A. (1989) Enzymatic incorporation of a new base pair into DNA and RNA. *J. Am. Chem. Soc.* **111**, 8322–8323 <https://doi.org/10.1021/ja00203a067>
- 41 Bain, J.D., Switzer, C., Chamberlin, A.R. and Benner, S.A. (1992) Ribosome-mediated incorporation of a non-standard amino acid into a peptide through expansion of the genetic code. *Nat. New Biol.* **356**, 537–539 <https://doi.org/10.1038/356537a0>
- 42 Switzer, C.Y., Moroney, S.E. and Benner, S.A. (1993) Enzymatic recognition of the base pair between isocytidine and isoguanosine. *Biochemistry* **32**, 10489–10496 <https://doi.org/10.1021/bi00090a027>
- 43 Zhang, Y., Lamb, B.M., Feldman, A.W., Zhou, A.X., Lavergne, T., Li, L. et al. (2017) A semisynthetic organism engineered for the stable expansion of the genetic alphabet. *Proc. Natl. Acad. Sci. U.S.A.* **114**, 1317–1322 <https://doi.org/10.1073/pnas.1616443114>
- 44 Malyshev, D.A., Seo, Y.J., Ordoukhanian, P. and Romesberg, F.E. (2009) PCR with an expanded genetic alphabet. *J. Am. Chem. Soc.* **131**, 14620–14621 <https://doi.org/10.1021/ja906186f>
- 45 Li, L., Degardin, M., Lavergne, T., Malyshev, D.A., Dhmi, K., Ordoukhanian, P. et al. (2014) Natural-like replication of an unnatural base pair for the expansion of the genetic alphabet and biotechnology applications. *J. Am. Chem. Soc.* **136**, 826–829 <https://doi.org/10.1021/ja408814g>
- 46 Kimoto, M., Kawai, R., Mitsui, T., Yokoyama, S. and Hirao, I. (2009) An unnatural base pair system for efficient PCR amplification and functionalization of DNA molecules. *Nucleic Acids Res.* **37**, e14 <https://doi.org/10.1093/nar/gkn956>
- 47 Taylor, A.I., Houlihan, G. and Holliger, P. (2019) Beyond DNA and RNA: the expanding toolbox of synthetic genetics. *Cold Spring Harb. Perspect. Biol.* **11**, a032490 <https://doi.org/10.1101/cshperspect.a032490>
- 48 Nikoosanzar, A., Chim, N., Yik, E.J. and Chaput, J.C. (2020) Engineering polymerases for applications in synthetic biology. *Q. Rev. Biophys.* **53**, e8 <https://doi.org/10.1017/S0033583520000050>
- 49 Nikoosanzar, A., Dunn, M.R. and Chaput, J.C. (2017) Engineered polymerases with altered substrate specificity: expression and purification. *Curr. Protoc. Nucleic Acid Chem.* **69** <https://doi.org/10.1002/cpnc.33>
- 50 Yik, E.J., Maola, V.A. and Chaput, J.C. (2023) Engineering TNA polymerases through iterative cycles of directed evolution. *Meth. Enzymol.* **691**, 29–59 <https://doi.org/10.1016/bs.mie.2023.04.014>
- 51 McCloskey, C.M., Liao, J.-Y., Bala, S. and Chaput, J.C. (2019) Ligase-mediated threose nucleic acid synthesis on DNA templates. *ACS Synth. Biol.* **8**, 282–286 <https://doi.org/10.1021/acssynbio.8b00511>
- 52 Hübscher, U. (1983) DNA polymerases in prokaryotes and eukaryotes: mode of action and biological implications. *Experientia* **39**, 1–25 <https://doi.org/10.1007/BF01960616>
- 53 Patel, P.H., Suzuki, M., Adman, E., Shinkai, A. and Loeb, L.A. (2001) Prokaryotic DNA polymerase I: evolution, structure, and “base flipping” mechanism for nucleotide selection. *J. Mol. Biol.* **308**, 823–837 <https://doi.org/10.1006/jmbi.2001.4619>
- 54 Xia, S., Wang, M., Blaha, G., Konigsberg, W.H. and Wang, J. (2011) Structural insights into complete metal ion coordination from ternary complexes of B family RB69 DNA polymerase. *Biochemistry* **50**, 9114–9124 <https://doi.org/10.1021/bi201260h>
- 55 Castro, C., Smidansky, E.D., Arnold, J.J., Maksimchuk, K.R., Moustafa, I., Uchida, A. et al. (2009) Nucleic acid polymerases use a general acid for nucleotidyl transfer. *Nat. Struct. Mol. Biol.* **16**, 212–218 <https://doi.org/10.1038/nsmb.1540>
- 56 Beard, W.A. and Wilson, S.H. (2014) Structure and mechanism of DNA polymerase β. *Biochemistry* **53**, 2768–2780 <https://doi.org/10.1021/bi500139h>
- 57 Gao, Y. and Yang, W. (2016) Capture of a third Mg<sup>2+</sup> is essential for catalyzing DNA synthesis. *Science* **352**, 1334–1337 <https://doi.org/10.1126/science.aad9633>
- 58 Chang, C., Lee Luo, C. and Gao, Y. (2022) In crystallo observation of three metal ion promoted DNA polymerase misincorporation. *Nat. Commun.* **13**, 2346 <https://doi.org/10.1038/s41467-022-30005-3>

- 59 Raia, P., Delarue, M. and Sauguet, L. (2019) An updated structural classification of replicative DNA polymerases. *Biochem. Soc. Trans.* **47**, 239–249 <https://doi.org/10.1042/BST20180579>
- 60 Yang, W. and Gao, Y. (2018) Translesion and repair DNA polymerases: diverse structure and mechanism. *Annu. Rev. Biochem.* **87**, 239–261 <https://doi.org/10.1146/annurev-biochem-062917-012405>
- 61 Jain, R., Aggarwal, A.K. and Rechkoblit, O. (2018) Eukaryotic DNA polymerases. *Curr. Opin. Struct. Biol.* **53**, 77–87 <https://doi.org/10.1016/j.sbi.2018.06.003>
- 62 Wu, S., Beard, W.A., Pedersen, L.G. and Wilson, S.H. (2014) Structural comparison of DNA polymerase architecture suggests a nucleotide gateway to the polymerase active site. *Chem. Rev.* **114**, 2759–2774 <https://doi.org/10.1021/cr3005179>
- 63 Kazlauskas, D., Krupovic, M., Guglielmini, J., Forterre, P. and Venclovas, Č. (2020) Diversity and evolution of B-family DNA polymerases. *Nucleic Acids Res.* **48**, 10142–10156 <https://doi.org/10.1093/nar/gkaa760>
- 64 Kuznetsova, A.A., Fedorova, O.S. and Kuznetsov, N.A. (2022) Structural and molecular kinetic features of activities of DNA polymerases. *Int. J. Mol. Sci.* **23**, 6373 <https://doi.org/10.3390/ijms23126373>
- 65 Iyer, L.M., Koonin, E.V. and Aravind, L. (2003) Evolutionary connection between the catalytic subunits of DNA-dependent RNA polymerases and eukaryotic RNA-dependent RNA polymerases and the origin of RNA polymerases. *BMC Struct. Biol.* **3**, 1–23 <https://doi.org/10.1186/1472-6807-3-1>
- 66 Guilliam, T.A. and Yeeles, J.T.P. (2020) An updated perspective on the polymerase division of labor during eukaryotic DNA replication. *Crit. Rev. Biochem. Mol. Biol.* **55**, 469–481 <https://doi.org/10.1080/10409238.2020.1811630>
- 67 Sauguet, L., Raia, P., Henneke, G. and Delarue, M. (2016) Shared active site architecture between archaeal PolD and multi-subunit RNA polymerases revealed by X-ray crystallography. *Nat. Commun.* **7**, 12227 <https://doi.org/10.1038/ncomms12227>
- 68 Ghosh, D. and Raghavan, S.C. (2021) 20 years of DNA Polymerase  $\mu$ , the polymerase that still surprises. *FEBS J.* **288**, 7230–7242 <https://doi.org/10.1111/febs.15852>
- 69 Sale, J.E., Lehmann, A.R. and Woodgate, R. (2012) Y-family DNA polymerases and their role in tolerance of cellular DNA damage. *Nat. Rev. Mol. Cell Biol.* **13**, 141–152 <https://doi.org/10.1038/nrm3289>
- 70 Oakley, A.J. (2019) A structural view of bacterial DNA replication. *Protein Sci.* **28**, 990–1004 <https://doi.org/10.1002/pro.3615>
- 71 Johansson, E. and Dixon, N. (2013) Replicative DNA polymerases. *Cold Spring Harb. Perspect. Biol.* **5**, a012799 <https://doi.org/10.1101/cshperspect.a012799>
- 72 Krishna, T.S., Kong, X.-P., Gary, S., Burgers, P.M. and Kuriyan, J. (1994) Crystal structure of the eukaryotic DNA polymerase processivity factor PCNA. *Cell* **79**, 1233–1243 [https://doi.org/10.1016/0092-8674\(94\)90014-0](https://doi.org/10.1016/0092-8674(94)90014-0)
- 73 Kuriyan, J. and O'Donnell, M. (1993) Sliding clamps of DNA polymerases. *J. Mol. Biol.* **234**, 915–925 <https://doi.org/10.1006/jmbi.1993.1644>
- 74 Fersht, A.R., Knill-Jones, J.W. and Tsui, W.C. (1982) Kinetic basis of spontaneous mutation. Misinsertion frequencies, proofreading specificities and cost of proofreading by DNA polymerases of *Escherichia coli*. *J. Mol. Biol.* **156**, 37–51 [https://doi.org/10.1016/0022-2836\(82\)90457-0](https://doi.org/10.1016/0022-2836(82)90457-0)
- 75 Evans, R.J., Davies, D.R., Bullard, J.M., Christensen, J., Green, L.S., Guiles, J.W. et al. (2008) Structure of PolC reveals unique DNA binding and fidelity determinants. *Proc. Natl. Acad. Sci. U.S.A.* **105**, 20695–20700 <https://doi.org/10.1073/pnas.0809989106>
- 76 Ganai, R.A. and Johansson, E. (2016) DNA replication: a matter of fidelity. *Mol. Cell.* **62**, 745–755 <https://doi.org/10.1016/j.molcel.2016.05.003>
- 77 Goodman, M.F. and Woodgate, R. (2013) Translesion DNA polymerases. *Cold Spring Harb. Perspect. Biol.* **5**, a010363 <https://doi.org/10.1101/cshperspect.a010363>
- 78 Waters, L.S., Minesinger, B.K., Wiltrout, M.E., D'Souza, S., Woodruff, R.V. and Walker, G.C. (2009) Eukaryotic translesion polymerases and their roles and regulation in DNA damage tolerance. *Microbiol. Mol. Biol. Rev.* **73**, 134–154 <https://doi.org/10.1128/MMBR.00034-08>
- 79 Joseph, A.M. and Badrinarayanan, A. (2020) Visualizing mutagenic repair: novel insights into bacterial translesion synthesis. *FEMS Microbiol. Rev.* **44**, 572–582 <https://doi.org/10.1093/femsre/fuaa023>
- 80 Ishino, S. and Ishino, Y. (2014) DNA polymerases as useful reagents for biotechnology - the history of developmental research in the field. *Front. Microbiol.* **5**, 465 <https://doi.org/10.3389/fmicb.2014.00465>
- 81 Aschenbrenner, J. and Marx, A. (2017) DNA polymerases and biotechnological applications. *Curr. Opin. Biotechnol.* **48**, 187–195 <https://doi.org/10.1016/j.copbio.2017.04.005>
- 82 Terpe, K. (2013) Overview of thermostable DNA polymerases for classical PCR applications: from molecular and biochemical fundamentals to commercial systems. *Appl. Microbiol. Biotechnol.* **97**, 10243–10254 <https://doi.org/10.1007/s00253-013-5290-2>
- 83 Ostorbin, I. and Filipenko, M. (2023) Bst polymerase — a humble relative of Taq polymerase. *Comput. Struct. Biotechnol. J.* **21**, 4519–4535 <https://doi.org/10.1016/j.csbj.2023.09.008>
- 84 Berjón-Otero, M., Villar, L. de Vega M., Salas, M. and Redrejo-Rodríguez, M. (2015) DNA polymerase from temperate phage Bam35 is endowed with processive polymerization and abasic sites translesion synthesis capacity. *Proc. Natl. Acad. Sci. U.S.A.* **112**, E3476–84 <https://doi.org/10.1073/pnas.1510280112>
- 85 Wang, Y., Prosen, D.E., Mei, L., Sullivan, J.C., Finney, M. and Vander Horn, P.B. (2004) A novel strategy to engineer DNA polymerases for enhanced processivity and improved performance in vitro. *Nucleic Acids Res.* **32**, 1197–1207 <https://doi.org/10.1093/nar/gkh271>
- 86 Śpibida, M., Krawczyk, B., Zalewska-Piątek, B., Piątek, R., Wysocka, M. and Olszewski, M. (2018) Fusion of DNA-binding domain of *Pyrococcus furiosus* ligase with TaqStoffel DNA polymerase as a useful tool in PCR with difficult targets. *Appl. Microbiol. Biotechnol.* **102**, 713–721 <https://doi.org/10.1007/s00253-017-8560-6>
- 87 Sun, Y., Ko, D.H., Gao, J., Fu, K., Gao, Y., Zhang, Q. et al. (2023) Unraveling the salt tolerance of Phi29 DNA polymerase using compartmentalized self-replication and microfluidics platform. *Front. Microbiol.* **14**, 1267196 <https://doi.org/10.3389/fmicb.2023.1267196>
- 88 Zhai, B., Chow, J. and Cheng, Q. (2019) Two approaches to enhance the processivity and salt tolerance of *Staphylococcus aureus* DNA polymerase. *Protein J.* **38**, 190–198 <https://doi.org/10.1007/s10930-019-09818-7>
- 89 Ostorbin, I.P., Belousova, E.A., Boyarskikh, U.A., Zakabunin, A.I., Khrapov, E.A. and Filipenko, M.L. (2017) Derivatives of Bst-like Gss-polymerase with improved processivity and inhibitor tolerance. *Nucleic Acids Res.* **45**, 9595–9610 <https://doi.org/10.1093/nar/gkx645>



- 90 Ghadessy, F.J., Ong, J.L. and Holliger, P. (2001) Directed evolution of polymerase function by compartmentalized self-replication. *Proc. Natl. Acad. Sci. U.S.A.* **98**, 4552–4557 <https://doi.org/10.1073/pnas.071052198>
- 91 Arezi, B., McKinney, N., Hansen, C., Cayouette, M., Fox, J., Chen, K. et al. (2014) Compartmentalized self-replication under fast PCR cycling conditions yields Taq DNA polymerase mutants with increased DNA-binding affinity and blood resistance. *Front. Microbiol.* **5**, 408 <https://doi.org/10.3389/fmicb.2014.00408>
- 92 d'Abbadie, M., Hofreiter, M., Vaisman, A., Loakes, D., Gasparutto, D., Cadet, J. et al. (2007) Molecular breeding of polymerases for amplification of ancient DNA. *Nat. Biotechnol.* **25**, 939–943 <https://doi.org/10.1038/nbt1321>
- 93 Ong, J.L., Loakes, D., Jaroslowski, S., Too, K. and Holliger, P. (2006) Directed evolution of DNA polymerase, RNA polymerase and reverse transcriptase activity in a single polypeptide. *J. Mol. Biol.* **361**, 537–550 <https://doi.org/10.1016/j.jmb.2006.06.050>
- 94 Chen, T., Hongdilokkul, N., Liu, Z., Adhikary, R., Tsuen, S.S. and Romesberg, F.E. (2016) Evolution of thermophilic DNA polymerases for the recognition and amplification of C2'-modified DNA. *Nat. Chem.* **8**, 556–562 <https://doi.org/10.1038/nchem.2493>
- 95 Ashley, J., Potts, I.G. and Olorunniji, F.J. (2023) Applications of terminal deoxynucleotidyl transferase enzyme in biotechnology. *Chembiochem* **24**, e202200510 <https://doi.org/10.1002/cbic.202200510>
- 96 Motea, E.A. and Berdis, A.J. (2010) Terminal deoxynucleotidyl transferase: the story of a misguided DNA polymerase. *Biochim. Biophys. Acta* **1804**, 1151–1166 <https://doi.org/10.1016/j.bbapap.2009.06.030>
- 97 Sarac, I. and Hollenstein, M. (2019) Terminal deoxynucleotidyl transferase in the synthesis and modification of nucleic acids. *Chembiochem* **20**, 860–871 <https://doi.org/10.1002/cbic.201800658>
- 98 Verardo, D., Adelizzi, B., Rodriguez-Pinzon, D.A., Moghaddam, N., Thomée, E., Loman, T. et al. (2023) Multiplex enzymatic synthesis of DNA with single-base resolution. *Sci. Adv.* **9**, eadi0263 <https://doi.org/10.1126/sciadv.adi0263>
- 99 Keen, B.A., Jozwiakowski, S.K., Bailey, L.J., Bianchi, J. and Doherty, A.J. (2014) Molecular dissection of the domain architecture and catalytic activities of human PrimPol. *Nucleic Acids Res.* **42**, 5830–5845 <https://doi.org/10.1093/nar/gku214>
- 100 Kent, T., Mateos-Gomez, P.A., Sfeir, A. and Pomerantz, R.T. (2016) Polymerase  $\theta$  is a robust terminal transferase that oscillates between three different mechanisms during end-joining. *Elife* **5**, e13740 <https://doi.org/10.7554/eLife.13740>
- 101 Ouaray, Z., Benner, S.A., Georgiadis, M.M. and Richards, N.G.J. (2020) Building better polymerases: engineering the replication of expanded genetic alphabets. *J. Biol. Chem.* **295**, 17046–17059 <https://doi.org/10.1074/jbc.REV120.013745>
- 102 Cline, J., Braman, J.C. and Hogrefe, H.H. (1996) PCR fidelity of pfu DNA polymerase and other thermostable DNA polymerases. *Nucleic Acids Res.* **24**, 3546–3551 <https://doi.org/10.1093/nar/24.18.3546>
- 103 Gerecht, K., Freund, N., Liu, W., Liu, Y., Fürst, M.J.L.J. and Holliger, P. (2023) The expanded central dogma: genome resynthesis, orthogonal biosystems, synthetic genetics. *Annu. Rev. Biophys.* **52**, 413–432 <https://doi.org/10.1146/annurev-biophys-111622-091203>
- 104 Marx, A. and Betz, K. (2020) The structural basis for processing of unnatural base pairs by DNA polymerases. *Chemistry* **26**, 3446–3463 <https://doi.org/10.1002/chem.201903525>
- 105 Loakes, D. and Holliger, P. (2009) Polymerase engineering: towards the encoded synthesis of unnatural biopolymers. *Chem. Commun. (Camb.)* 4619–4631 <https://doi.org/10.1039/b903307f>
- 106 Yang, Z., Chen, F., Chamberlin, S.G. and Benner, S.A. (2010) Expanded genetic alphabets in the polymerase chain reaction. *Angew. Chem. Int. Ed. Engl.* **49**, 177–180 <https://doi.org/10.1002/anie.200905173>
- 107 Malyshev, D.A., Dhama, K., Quach, H.T., Lavergne, T., Ordoukhanian, P., Torkamani, A. et al. (2012) Efficient and sequence-independent replication of DNA containing a third base pair establishes a functional six-letter genetic alphabet. *Proc. Natl. Acad. Sci. U.S.A.* **109**, 12005–12010 <https://doi.org/10.1073/pnas.1205176109>
- 108 Eberlein, L., Beierlein, F.R., Van Eikema Hommes, N.J., RRadadiya, A., Heil, J., Benner, S.A. et al. (2020) Tautomeric Equilibria of Nucleobases in the Hachimoji Expanded Genetic Alphabet. *J. Chem. Theory Comput.* **16**, 2766–2777 <https://doi.org/10.1021/acs.jctc.9b01079>
- 109 Yang, Z., Chen, F., Alvarado, J.B. and Benner, S.A. (2011) Amplification, mutation, and sequencing of a six-letter synthetic genetic system. *J. Am. Chem. Soc.* **133**, 15105–15112 <https://doi.org/10.1021/ja204910n>
- 110 Sun, L., Ma, X., Zhang, B., Qin, Y., Ma, J., Du, Y. et al. (2022) From polymerase engineering to semi-synthetic life: artificial expansion of the central dogma. *RSC Chem. Biol.* **3**, 1173–1197 <https://doi.org/10.1039/d2cb00116k>
- 111 Laos, R., Shaw, R., Leal, N.A., Gaucher, E. and Benner, S. (2013) Directed evolution of polymerases to accept nucleotides with nonstandard hydrogen bond patterns. *Biochemistry* **52**, 5288–5294 <https://doi.org/10.1021/bi400558c>
- 112 Johnson, S.C., Sherrill, C.B., Marshall, D.J., Moser, M.J. and Prudent, J.R. (2004) A third base pair for the polymerase chain reaction: inserting isoC and isoG. *Nucleic Acids Res.* **32**, 1937–1941 <https://doi.org/10.1093/nar/gkh522>
- 113 Singh, I., Laos, R., Hoshika, S., Benner, S.A. and Georgiadis, M.M. (2018) Snapshots of an evolved DNA polymerase pre- and post-incorporation of an unnatural nucleotide. *Nucleic Acids Res.* **46**, 7977–7988 <https://doi.org/10.1093/nar/gky552>
- 114 Betz, K., Kimoto, M., Diederichs, K., Hirao, I. and Marx, A. (2017) Structural basis for expansion of the genetic alphabet with an artificial nucleobase pair. *Angew. Chem. Int. Ed. Engl.* **56**, 12000–12003 <https://doi.org/10.1002/anie.201704190>
- 115 Ouaray, Z., Singh, I., Georgiadis, M.M. and Richards, N.G.J. (2020) Building better enzymes: Molecular basis of improved non-natural nucleobase incorporation by an evolved DNA polymerase. *Protein Sci.* **29**, 455–468 <https://doi.org/10.1002/pro.3762>
- 116 Shuman, S. (2009) DNA ligases: progress and prospects. *J. Biol. Chem.* **284**, 17365–17369 <https://doi.org/10.1074/jbc.R900017200>
- 117 Lehman, I.R. (1974) DNA ligase: structure, mechanism, and function. *Science* **186**, 790–797 <https://doi.org/10.1126/science.186.4166.790>
- 118 Williamson, A. and Leiros, H.S. (2020) Structural insight into DNA joining: from conserved mechanisms to diverse scaffolds. *Nucleic Acids Res.* **48**, 8225–8242 <https://doi.org/10.1093/nar/gkaa307>
- 119 Tomkinson, A.E., Vijayakumar, S., Pascal, J.M. and Ellenberger, T. (2006) DNA ligases: structure, reaction mechanism, and function. *Chem. Rev.* **106**, 687–699 <https://doi.org/10.1021/cr040498d>
- 120 Doherty, A.J. and Suh, S.W. (2000) Structural and mechanistic conservation in DNA ligases. *Nucleic Acids Res.* **28**, 4051–4058 <https://doi.org/10.1093/nar/28.21.4051>
- 121 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 <https://doi.org/10.1016/j.molcel.2007.02.026>
- 122 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 <https://doi.org/10.1111/mmi.13229>

- 123 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 [https://doi.org/10.1016/s0092-8674\(00\)81260-x](https://doi.org/10.1016/s0092-8674(00)81260-x)
- 124 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 [https://doi.org/10.1016/s1097-2765\(00\)00115-5](https://doi.org/10.1016/s1097-2765(00)00115-5)
- 125 Shi, K., Bohl, T.E., Park, J., Zasada, A., Malik, S., Banerjee, S. 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 <https://doi.org/10.1093/nar/gky776>
- 126 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. *Nat. New Biol.* **432**, 473–478 <https://doi.org/10.1038/nature03082>
- 127 Amare, B., Mo, A., Khan, N., Sowa, D.J., Warner, M.M., Tetenych, A. et al. (2021) LigD: a structural guide to the multi-tool of bacterial non-homologous end joining. *Front. Mol. Biosci.* **8**, 787709 <https://doi.org/10.3389/fmolb.2021.787709>
- 128 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 <https://doi.org/10.1073/pnas.1005830107>
- 129 Pitcher, R.S., Brissett, N.C., Picher, A.J., Andrade, P., Juarez, R., Thompson, D. et al. (2007) Structure and function of a mycobacterial NHEJ DNA repair polymerase. *J. Mol. Biol.* **366**, 391–405 <https://doi.org/10.1016/j.jmb.2006.10.046>
- 130 Chen, Y., Liu, H., Yang, C., Gao, Y., Yu, X., Chen, 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 <https://doi.org/10.1038/s41467-019-08296-w>
- 131 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 [https://doi.org/10.1016/S0021-9258\(18\)93227-8](https://doi.org/10.1016/S0021-9258(18)93227-8)
- 132 Gibson, D.G., Young, L., Chuang, R.-Y., Venter, J.C., Hutchison III, C.A. and Smith, H.O. (2009) Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat. Methods* **6**, 343–345 <https://doi.org/10.1038/nmeth.1318>
- 133 Pryor, J.M., Potapov, V., Kucera, R.B., Bilotti, K., Cantor, E.J. and Lohman, G.J.S. (2020) Enabling one-pot Golden Gate assemblies of unprecedented complexity using data-optimized assembly design. *PLoS ONE* **15**, e0238592 <https://doi.org/10.1371/journal.pone.0238592>
- 134 Potapov, V., Ong, J.L., Langhorst, B.W., Bilotti, K., Cahoon, D., Canton, B. et al. (2018) A single-molecule sequencing assay for the comprehensive profiling of T4 DNA ligase fidelity and bias during DNA end-joining. *Nucleic Acids Res.* **46**, e79 <https://doi.org/10.1093/nar/gky303>
- 135 Bilotti, K., Potapov, V., Pryor, J.M., Duckworth, A.T., Keck, J.L. and Lohman, G.J.S. (2022) Mismatch discrimination and sequence bias during end-joining by DNA ligases. *Nucleic Acids Res.* **50**, 4647–4658 <https://doi.org/10.1093/nar/gkac241>
- 136 Potapov, V., Ong, J.L., Kucera, R.B., Langhorst, B.W., Bilotti, K., Pryor, J.M. 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 <https://doi.org/10.1021/acssynbio.8b00333>
- 137 Kok, S.D., Stanton, L.H., Slaby, T., Durot, M., Holmes, V.F., Patel, K.G. et al. (2014) Rapid and reliable DNA assembly via. *ACS Synth. Biol.* **3**, 97–106 <https://doi.org/10.1021/sb4001992>
- 138 Roth, T.L., Milenkovic, L. and Scott, M.P. (2014) A rapid and simple method for DNA engineering using cycled ligation assembly. *PLoS ONE* **9**, e107329 <https://doi.org/10.1371/journal.pone.0107329>
- 139 Horspool, D.R., Coope, R.J. and Holt, R.A. (2010) Efficient assembly of very short oligonucleotides using T4 DNA ligase. *BMC Res. Notes* **3**, 291 <https://doi.org/10.1186/1756-0500-3-291>
- 140 Hili, R., Niu, J. and Liu, D.R. (2013) DNA ligase-mediated translation of DNA into densely functionalized nucleic acid polymers. *J. Am. Chem. Soc.* **135**, 98–101 <https://doi.org/10.1021/ja311331m>
- 141 Moser, M.J. and Prudent, J.R. (2003) Enzymatic repair of an expanded genetic information system. *Nucleic Acids Res.* **31**, 5048–5053 <https://doi.org/10.1093/nar/gkg709>
- 142 Kimoto, M., Soh, S.H.G., Tan, H.P., Okamoto, I. and Hirao, I. (2021) Cognate base-pair selectivity of hydrophobic unnatural bases in DNA ligation by T4 DNA ligase. *Biopolymers* **112**, e23407 <https://doi.org/10.1002/bip.23407>
- 143 Kawabe, H., Thomas, C.A., Hoshika, S., Kim, M.-J., Kim, M.-S., Miessner, L. et al. (2023) Enzymatic synthesis and nanopore sequencing of 12-letter supernumerary DNA. *Nat. Commun.* **14**, 6820 <https://doi.org/10.1038/s41467-023-42406-z>
- 144 Kestemont, D., Renders, M., Leonczak, P., Abramov, M., Schepers, G., Pinheiro, V.B. et al. (2018) XNA ligation using T4 DNA ligase in crowding conditions. *Chem. Commun. (Camb.)* **54**, 6408–6411 <https://doi.org/10.1039/c8cc02414f>
- 145 Vanmeert, M., Razzokov, J., Mirza, M.U., Weeks, S.D., Schepers, G., Bogaerts, A. et al. (2019) Rational design of an XNA ligase through docking of unbound nucleic acids to toroidal proteins. *Nucleic Acids Res.* **47**, 7130–7142 <https://doi.org/10.1093/nar/gkz551>
- 146 Xu, H., Xu, D. and Liu, Y. (2024) Molecular biology applications of psychrophilic enzymes: adaptations, advantages, expression, and prospective. *Appl. Biochem. Biotechnol.* **196**, 5765–5789 <https://doi.org/10.1007/s12010-023-04810-5>
- 147 Cavicchioli, R., Charlton, T., Ertan, H., Mohd Omar, S., Siddiqui, K.S. and Williams, T.J. (2011) Biotechnological uses of enzymes from psychrophiles. *Microb. Biotechnol.* **4**, 449–460 <https://doi.org/10.1111/j.1751-7915.2011.00258.x>
- 148 Piotrowski, Y., Gurung, M.K. and Larsen, A.N. (2019) Characterization and engineering of A DNA polymerase reveals A single amino-acid substitution in the fingers subdomain to increase strand-displacement activity of A-family prokaryotic DNA polymerases. *BMC Mol. Cell Biol.* **20**, 31 <https://doi.org/10.1186/s12860-019-0216-1>
- 149 Abramson, J., Adler, J., Dunger, J., Evans, R., Green, T., Pritzel, A. et al. (2024) Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nat. New Biol.* **630**, 493–500 <https://doi.org/10.1038/s41586-024-07487-w>
- 150 Baek, M., McHugh, R., Anishchenko, I., Jiang, H., Baker, D. and DiMaio, F. (2024) Accurate prediction of protein-nucleic acid complexes using RoseTTAFoldNA. *Nat. Methods.* **21**, 117–121 <https://doi.org/10.1038/s41592-023-02086-5>