Biochemical Investigation and Engineering of a Tardigrade X Family DNA Polymerase for Template-independent DNA Synthesis

Yee-Song Law,[1] Nazreen Abdul Muthaliff,[1] Yifeng Wei,[1] Fu Lin,[1] Huimin Zhao,[2,*] Ee Lui Ang[1, 3, 4,*]

[1] Singapore Institute of Food and Biotechnology Innovation (SIFBI), Agency for Science, Technology and Research (A*STAR), 31 Biopolis Way, #04-01 Nanos Building, Singapore 138669, Republic of Singapore.

[2] Department of Chemical and Biomolecular Engineering, University of Illinois at Urbana-Champaign, Urbana, Illinois 61801, United States.

[3] Institute of Sustainability for Chemicals, Energy and Environment (ISCE2), Agency for Science, Technology and Research (A*STAR), 8 Biomedical Grove, #07-01 Neuros Building, Singapore 138665, Republic of Singapore.

[4] Department of Biochemistry, Yong Loo Lin School of Medicine, National University of Singapore, 8 Medical Drive, Singapore 117596, Republic of Singapore.

* Corresponding authors:

Huimin Zhao (zhao5@illinois.edu)

Ee Lui Ang (ang_ee_lui@sifbi.a-star.edu.sg)

ASTRACT (<300 words):

The X family of DNA polymerases (PolXs) include the well-studied mammalian polymerases Polβ, Polλ, Polμ, and terminal deoxynucleotidyltransferase (TdT). The template-independent DNA polymerase activity of TdT has been harnessed for applications in enzymatic *de novo* DNA synthesis, offering a strategy to overcome the limitations of traditional phosphoramidite-based DNA synthesis methods. Close homologs of the mammalian PolXs are present in other vertebrates, while invertebrate PolXs remain unexplored. In this study, we characterize an invertebrate PolX from the extremotolerant tardigrade *Ramazzottius varieornatus* (RvPolX), and demonstrate that it possesses modest template-independent DNA polymerase activity, despite limited homology to mammalian PolXs (21% sequence identity with TdT). Through a combination of structural modelling, targeted mutagenesis of active site residues, and high-

throughput screening under stringent high salt conditions, we identified a synergistic combination of two mutations (G513A and R522I) that led to a significant increase in activity for the incorporation of all four nucleotides, particularly dATP (~35-fold), yielding a salt-tolerant polymerase with overall higher activity and substrate promiscuity. Under high salt conditions, the engineered RvPolX displays an activity comparable to TdT, and a nucleotide selectivity complementary to TdT. As a template-independent polymerase with low-homology to TdT, RvPolX provides an alternative scaffold for further engineering in various biotechnological applications.

KEYWORDS: *Tardigrade, template-independent, X family DNA polymerase (PolX), enzymatic DNA synthesis, protein engineering*

INTRODUCTION

The X family DNA polymerases (PolXs) are found across all domains of life, and perform a variety of important functions related to DNA repair, including base excision repair (BER), nonhomologous end joining (NHEJ), and V(D)J recombination.[1] The biochemical properties of PolXs have been most extensively investigated in mammals, particularly human and mouse, which contain four PolXs: DNA polymerase beta (Polβ), DNA polymerase lambda (Polλ), DNA polymerase mu (Polμ), and terminal deoxynucleotidyl transferase (TdT).[2] Polβ functions in gap filling during base excision repair (BER), whereas Polλ and Polμ function in NHEJ.[1, 3] Meanwhile, Polμ and TdT are primarily expressed in lymphoid tissue, where they play crucial roles in affinity maturation of antibodies and T-cell receptors.[1] The mammalian PolXs display a spectrum of fidelity and template-dependency. Template-dependent polymerase activity is observed for Polβ, Polλ, and Polμ but not TdT, while template-independent polymerase activity is not observed for Polβ, and increases in the order of Polλ < Polμ < TdT.[1, 2] The exclusive template-independent polymerase activity of TdT is unique among DNA polymerases, and is linked to its function in creating junctional diversity during V(D)J recombination.[1, 4]

A recent renewal of interest in PolXs has been driven by potential applications in enzymatic *de novo* DNA synthesis. The cost and quality of chemically synthesized DNA presents a constraint to the rapidly developing field of synthetic biology and nucleic acid therapeutics, and poses a barrier to harnessing the high data density and long-term stability of DNA as a

medium for data storage.[5] Enzymatic DNA synthesis offers a potential solution to the challenges associated with current phosphoramidite-based solid-phase DNA synthesis methods, including accuracy limitations, length restrictions, high cost of materials, use of toxic reagents and solvents, and unintended side-reactions during the elongation and deprotection steps. A recently described method for DNA digital data storage involves enzymatic synthesis of long DNA sequences using TdT and apyrase.[5-7] While the method allows encoding and decoding of information in DNA, it lacks the capability for synthesis with single-base resolution, which is required to minimize redundancy and maximize data storage density.[6] Two other recent methods involve incorporation of 3′-protected deoxynucleotide triphosphates (dNTPs) by an engineered TdT,[8] and DNA synthesis using TdT-dNTP conjugates,[9] which enable synthesis of short DNA sequences with single-base resolution. Proprietary enzymatic DNA synthesis methods, employed by companies such as DNA Script and Camena Bio, have reported coupling efficiency exceeding that of phosphoramidite synthesis.[10] At present, the practical application of template-independent polymerases in DNA digital data storage remains a topic of ongoing exploration. Continued progress will necessitate the development of robust polymerases that are compatible with various automation procedures, and retain activity under diverse reaction and workup conditions, including variations in temperature, pH levels, and ionic concentrations.

Despite the widespread distribution of PolXs, relatively few have been experimentally characterized. A recent study by Lu et al. [8] described assays of selected vertebrate (mammalian, fish, amphibian, reptile and avian) TdT homologs, and engineering of the *Zonotrichia albicollis* (white-throated sparrow) TdT for incorporation of 3′-ONH_2-modified dNTPs. Apart from the vertebrate enzymes, other PolXs that have been studied include Arabidopsis Polλ[11] and yeast Pol4[12], thought to be functional counterparts of the mammalian Polλ. Trypanosomatids contain homologs of Polβ.[13] The viral PolX from African swine fever virus is an unusually compact enzyme (20 kDa), has relatively low fidelity in DNA synthesis, and plays a role in viral DNA repair.[14, 15] The bacterial PolX from *Thermus thermophilus*[16] and *Bacillus subtilis*[17] are thought to function in gap filling. In addition, recent studies of a noncanonical bacterial PolX from *Deinococcus radiodurans* showed that it lacks polymerase activity, but function as an exonuclease in proofreading and DNA repair.[18] One promising approach for identifying PolXs with favorable biochemical properties involves investigating PolX homologs in extremotolerant organisms. Tardigrades, commonly known as water bears, have gained notoriety for their ability to endure harsh environmental conditions, including extreme

temperatures and pressures, and exposure to radiation and desiccation, while in a cryptobiotic state.[19, 20] The terrestrial tardigrade *Ramazzottius varieornatus* stands out as one of the most stress-tolerant tardigrade species, and has served as a model organism for astrobiological studies.[21] Previous investigations of *R. varieornatus* revealed expansion of genes involved in the repair of DNA double-strand breaks, and have led to the discovery of a unique DNA-associating damage suppressor protein that protects against radiation-induced DNA damage.[19] Thus, *R. varieornatus* and other extremotolerant tardigrades could serve as a potential source of robust enzymes for applications in DNA biotechnology.[22]

In this study, we investigate the biochemical properties of a PolX enzyme from *R. varieornatus* (RvPolX), and demonstrate that RvPolX catalyzes template-independent DNA polymerization, retaining its enzymatic activity at high salt concentrations (up to 1000 mM) and temperatures (45 °C). Targeted engineering of the RvPolX active site increased its catalytic activity at high salt concentrations, and substrate promiscuity towards dNTP substrates, yielding a polymerase with distinctive biochemical characteristics for *de novo* DNA synthesis applications.

RESULTS

Phylogenetic Analysis of Eumetazoan PolXs

PolXs are present in the genomes of nearly all animals (kingdom Eumetazoa). To identify PolX homologs divergent from the well-studied mammalian for further investigation, we conducted a phylogenetic analysis of sequences belonging to the DNA polymerase X family (InterPro family IPR002054), in Eumetazoa, within the UniProt database. To reduce redundancy, UniRef50 clusters were used (UniRef50 contains sequences clustered at ≥50% identity, Dataset S1). Four previously studied non-metazoan PolXs, including *Saccharomyces cerevisiae* DNA polymerase IV (ScPolIV), *Schizosaccharomyces pombe* DNA polymerase IV (SpPolIV), *Oryza sativa* DNA polymerase lambda (OsPolλ), and *Arabidopsis thaliana* DNA polymerase lambda (AtPolλ), were included in the analysis. A maximum likelihood phylogenetic tree was constructed (**Figure 1**), and the position of the root was inferred using the plant PolXs as an outgroup. In the resulting tree, the PolXs can be classified into six major groups: TdT/Polμ, Polλ, and Polβ, which contain sequences largely from the superphylum Deuterostomes (including mammals), and TdT/Polμ-like, Polλ-like, and Polβ-like polymerases, which include sequences from protostomes (including tardigrades, from the superphylum Ecdysozoa).

The mammalian TdT, Polμ, and Polλ contain an N-terminal BRCT domain, which serves to interact with the core NHEJ factors in the repair of DNA double stranded breaks.[23, 24] The majority of the TdT/Pol, Polλ, TdT/Polμ-like and Polλ-like proteins also contain this domain, suggesting that they may share a related function in DNA repair. Conversely, the majority of Polβ and Polβ-like proteins lack this domain, suggesting a different function, such as gap filling. The TdT/Polμ, Polβ, and Polλ groups contain 21 SwissProt reviewed Eumetazoan PolX sequences (**Table S1**), of which eight have an associated PDB structure. By contrast, the TdT/Polμ-like, Polβ-like, and Polλ-like groups do not contain any characterized proteins. The tardigrade *R. varieornatus* contains three PolXs, of which two belong to the Polλ-like group (UniProt IDs: A0A1D1UV65 and A0A1D1UJX5, 34 and 38% sequence identities to human Polλ), and one belongs to the Polβ-like group (A0A1D1W154, 32% sequence identity to human Polβ). Their distant homology to the characterized mammalian enzymes suggests that the *R. varieornatus* PolXs may warrant further investigation as potential catalysts for template-independent DNA synthesis.

RvPolX Displays Template-independent DNA Polymerization Activity and High Salt Tolerance

To facilitate biochemical investigations of the three *R. varieornatus* PolXs, we attempted to recombinantly produce and purify them from *E. coli*. However, soluble expression could only be achieved for one Polλ-like enzyme (UniProt ID: A0A1D1UV65, designated as RvPolX), which became the focus of our ensuing experiments described below. The domain structure of RvPolX consists of an N-terminal BRCT domain and a C-terminal catalytic domain, similar to TdT, Polμ, and Polλ (**Figure 2A**). A multiple sequence alignment of the catalytic domains of RvPolX and the mammalian PolXs suggest a conserved fold, which includes the 8 kDa, fingers, palm, and thumb subdomains. The TdT crystal structure contains a specific loop (loop1) that obstructs template strand binding, and is reported to contribute to its template-independent polymerase activity. Loop1 is also present in Polμ, and although this loop is disordered in the Polμ crystal structures, it is nevertheless also thought to play a role during template-independent polymerization (**Figure S1**). Deletion of loop1 was reported to decrease template-independent activity and increase template-dependent activity in both TdT and Polμ.[25] An equivalent loop1 region is absent in Polβ, Polλ, and is also absent in the sequence of RvPolX (**Figure 2A**, **Figures S1** and **S2**).

To investigate whether RvPolX can catalyze template-independent DNA polymerization, RvPolX (4 μM) was incubated with a 14-mer single-stranded DNA (ssDNA14, 0.05 μM; see **MATERIALS AND METHODS** for sequence) and different dNTPs (0.3 mM). Despite lacking the loop1 element, the assays revealed that RvPolX displays modest template-independent DNA polymerization activity (**Figures 2B-E**). RvPolX activity increased with increasing pH (pH 5.0 to pH 8.8, **Figure 2B**), and decreased with increasing NaCl concentration (0-700 mM, **Figure 2C**). The isoelectric point of RvPolX was calculated to be 6.80, with a net charge of -11.1 at pH 8.8 (**Figure S3**). A high negative surface charge density is found in many halophilic proteins, and could contribute to the stability of RvPolX under the reaction conditions. Activity was detectable up to 700 mM NaCl for incorporation of dGTP, dTTP, and dCTP, and up to 500 mM NaCl for incorporation of dATP, suggesting that RvPolX might be suitable for further development as a salt-tolerant polymerase. In assays with various divalent metal ions (Ca^{2+}, Co^{2+}, Mg^{2+}, Mn^{2+}, and Zn^{2+}), RvPolX activity was highest with Mn^{2+} as the metal cofactor, followed by Mg^{2+} and Co^{2+} (**Figure S4**). In assays with different combinations of divalent metal ions, activity was highest when both Mg^{2+} and Mn^{2+} were present (**Figure 2D**). Subsequent assays were carried out at pH 8.8, with Mg^{2+} (5 mM) and Mn^{2+} (1 mM).

Next, we examined the effect of the oligonucleotide substrate sequence on RvPolX activity, by employing 15-mer oligonucleotide substrates (ssDNA14A, ssDNA14T, ssDNA14G, and ssDNA14C, **Table S2**) composed of ssDNA14 with a single base (A, T, G, or C) added at the 3' end. The assays show that RvPolX is able to extend ssDNA14 with all four 3' bases, with its activity showing a limited dependence on the specific 3' base (**Figure 2E**). Conversely, across a range of reaction conditions, RvPolX exhibits a preference for the dNTP substrate in the following order: dGTP > dTTP > dATP = dCTP (**Figures 2B-E**). Collectively, the assays demonstrate that RvPolX catalyzes template-independent DNA polymerization, and might be suitable for further development as a salt-tolerant polymerase.

The RvPolX N-terminal Region Contributes to its Stability

To investigate the possible contributions of the RvPolX N-terminal region to its activity and stability, we constructed a RvPolX N-terminal 258-amino acid deletion mutant (Δ1-258aa) containing only the catalytic domain, and compared its activity to that of RvPolX wild-type (WT). For assays in the absence of NaCl, the activity of WT was constant at 25 °C and 40 °C,

but decreased at 45 °C (**Figure 3A**). Activity of Δ1-258aa is comparable to WT at 25 °C and 40 °C, but significantly less that WT at 45 °C. For assays in the presence of 500 mM NaCl, the activity of WT decreases between 25 °C and 40 °C, and is eliminated at 45°C. Activity of Δ1-258aa is significantly less that WT at 25 °C and 40 °C, and is also eliminated at 45°C. The assays demonstrate that deletion of the RvPolX N-terminal region decreases its resilience to high temperatures, particularly under high salt concentrations. Thus, although the RvPolX N-terminal region is not essential for template-independent DNA polymerase activity, it plays a role in preserving activity under high temperature and salt conditions.

We next investigated whether the RvPolX N-terminal region contributes to its structural stability under high salt concentrations. Far-UV circular dichroism (CD) spectroscopy experiments of WT and Δ1-258aa were conducted at a range of temperatures (25 °C to 50 °C), in the presence of 500 mM NaCl (**Figures 3B-C**). At 25-35 °C, the CD spectra of both WT and Δ1-258aa exhibited characteristic features of α-helices. At 40 °C and above, drastic changes to the CD spectra were observed above for Δ1-258aa, indicating a loss of α-helical secondary structure. A corresponding drastic change was not observed in WT, suggesting that deletion of the N-terminal amino acids (1-258aa) adversely affects the stability of the RvPolX secondary structure. Subsequent experiments were thus carried out with full-length RvPolX.

Protein Engineering of RvPolX

Protein engineering of RvPolX was carried out, involving targeted mutagenesis of active site residues, aimed at enhancing its salt-tolerant template-independent DNA polymerase activity. A structural model of the ternary complex of RvPolX and its substrates was constructed by molecular docking of the AlphaFold structure of RvPolX with ssDNA14, two Mn^{2+} ions, and each of the four dNTPs, followed optimization by molecular dynamics simulation (**Figure 4A**, details provided in the **MATERIALS AND METHODS** section). From the structural model, 12 amino acid residues located in and around the dNTP binding pocket (R276, E281, C425, S427, R432, G513, W514, M518, Y519, R525, R522, and D537, **Figure 4B**) were selected for saturation mutagenesis and high throughput activity screening. A total of 240 clones from the mutant library were individually expressed and purified using Dynabeads™, and screened for template-independent DNA polymerase activity, with reaction products analyzed by high-performance liquid chromatography (HPLC). Four mutants G513A, R522I, R522M, and

R522V displayed increased activity for dATP incorporation relative to the WT (**Figure S5A** and **Dataset S2**), and were thus selected for further development.

To explore the potential synergistic effects of the G513A and R522I/M/V mutations, the double mutants were generated and assayed for template-independent DNA polymerization. Among three double mutants, the combination of G513A and R522I (G513A+R522I) led to the highest activity (**Figure S5B**). The extent of conversion of the oligonucleotide substrate for WT and selected mutants, in the presence of 500 mM and 1000 mM NaCl, was estimated by quantifying the ssDNA14 and extension product bands in each gel lane. For assays in the presence of 500 mM NaCl, G513A+R522I displayed higher conversion of the oligonucleotide substrate relative to WT, especially an increase from 4% to 32% extension product for the incorporation of dATP (7.9-fold) and from 46% to 83% extension product for the incorporation of dGTP (1.8-fold). The enhancements were further amplified in the presence of 1000 mM NaCl, for incorporation of dTTP (1.7-fold), dCTP (1.5-fold), dGTP (2.3-fold), and dATP (increase in extension product from undetectable to 8%). Overall, the G513A+R522I exhibits an enhanced salt-tolerant polymerase activity, and increased promiscuity towards dNTP substrates, with a particularly marked increase in proficiency for incorporation of dATP. In addition, both WT and G513A+R522I also displayed limited activity for incorporation of the protected 3'-ONH_2 dNTPs (**Figure S6**).

Biochemical Investigation of RvPolX G513A+R522I

For a more quantitative comparison of the activity of the G513A+R522I with RvPolX WT, the assays were repeated for each of the four dNTPs in the absence of salt. As shown in **Figures 5A** and **5B**, the oligonucleotide chain elongation products were measured over a course of 300 or 1800 s, and the chain growth polymerization kinetics modelled as a Poisson distribution.[26] For each of the assays, we selected an oligonucleotide product band for which the time-dependent rise and subsequent fall in band intensity could be observed during the course of the assay. The apparent rate of chain elongation (V_{obs}) was estimated by fitting the corresponding time-dependent accumulation and subsequent depletion of the product to a Poisson probability distribution function (**Figure 5C**). For dATP, we were unable to select a product band containing the rise and fall phases for both WT and G513A+R522I. Therefore, we monitored the +3 extended product band, which contains the rise and fall phases for G513A+R522I, and the rise and plateau phases for WT. We also monitored the +1 band containing the rise and fall

phases for WT, showing that the rates calculated for the +1 and +3 bands are comparable. As shown in **Table 1**, for each of the four dNTPs, the calculated V_{obs} for G513A+R522I was significantly higher than those for WT (2-, 3-, 7-, and 36-fold higher for G, C, T, and A, respectively).

To examine the possible mechanistic origin for the enhancement of catalytic activity of the G513A+R522I, the dissociation constants for binding of the four dNTPs to RvPolX WT and G513A+R522I were determined using label-free microscale thermophoresis (MST, **Figure 5C**). For both WT and G513A+R522I, the lowest K_d (corresponding to the highest affinity) was observed for dATP, followed by dTTP, dCTP, and then dGTPs. The K_d for binding of dGTP to G513A+R522I was ~5-fold lower than WT, while the K_d for binding of the other dNTPs to the G513A+R522I was comparable to WT. The absence of correlation between dNTP binding affinity and enhancement of catalytic activity indicates the need for alternative explanations, particularly for the non-conservative R522I mutation, which has not been previously reported in relation to template-independent activity. The residue corresponding to R522 is conserved in Polλ, Polμ, and TdT, but replaced by lysine in Polβ (**Figure S2**). The crystal structure of human Polμ gap filling pre- and post-catalytic complexes revealed that this residue does not directly interact with the dNTP substrate, but interacts with the phosphodiester backbone of the template DNA strand.[27] In the template-independent DNA polymerase assays, the oligonucleotide substrate ssDNA14 could bind to either the primer elongation site or the template binding site. The DNA binding analysis revealed that WT had slightly stronger affinity for ssDNA14 than G513A+R522I, but the difference was not significant (**Figure S7**). This finding suggests that the R522I mutation has potential to destabilize non-productive binding of ssDNA14.

DISCUSSION

Among the X family DNA polymerases, the mammalian Polβ, Polλ, Polμ, and TdT are the four representative enzymes that have been subjects of in-depth studies. In particular, TdT has garnered significant attention for *de novo* DNA synthesis applications, especially in the context of DNA-based data storage, due to its high template-independent polymerase activity.[6, 8, 9, 28] Our research on RvPolX, from the extremotolerant tardigrade *R. varieornatus*, expands the scope of biochemically characterized PolX enzymes to invertebrates. We demonstrate that

RvPolX possesses modest template-independent DNA polymerase activity, despite sharing only 21% sequence identity with TdT. A side-by-side comparison of TdT and RvPolX variants under the same conditions revealed that both enzymes retained catalytic activity at 1000 mM KCl (**Figure S8**). TdT exhibited higher activity for the incorporation of dGTP and dCTP, while RvPolX showed greater activity for the incorporation of dATP and dTTP. In addition, RvPolX variants displayed higher activity compared to commercial calf thymus TdT for incorporation of the protected 3'-ONH_2 dNTPs (**Figure S9**), suggesting that they may complement each other in future applications such as DNA data storage. Salt tolerant versions of the template-dependent DNA polymerase from *Bacillus* phage phi29 have previously been employed in nanopore devices for DNA sequencing[29-31], and the use of phi29 polymerase-nanopore conjugates for monitoring template-dependent DNA synthesis was previously described.[32] The development of salt-tolerant template-independent polymerases could enable applications in nanopore-based devices for monitoring of *de novo* DNA synthesis.

Previous structural studies showed that loop1 of TdT adopts a lariat-like conformation, impeding the binding of a continuous template DNA strand,[33] a feature believed to be important in its template-independent DNA polymerase activity. Deletion of loop1 or substitution of amino acid residues within the loop reduces or completely abolishes template-independent DNA polymerise activity.[25] Interestingly, loop1 is absent in both RvPolX and the well-studied Polλ, which has also been reported to have modest template-independent DNA polymerase activity. During template-independent polymerization, Polλ exhibits a preference for incorporating pyrimidine nucleotides,[34] while wild-type RvPolX exhibits a preference for dGTP, with lower activities for pyrimidine nucleotides, and limited activity for dATP. Nevertheless, our experiments demonstrated that the activity of RvPolX can be enhanced by engineering of the dNTP binding pocket (**Figure 4**). By employing a strategy of involving computational modelling of the active site, mutagenesis of the dNTP binding pocket, and HPLC-based activity screening under stringent reaction conditions with 350 mM NaCl, we identified two mutations (G513A and R522I) with increased activity. Combining the two mutations led to a synergistic increase in activity for incorporation of all four dNTPs, particularly dATP (~35-fold), yielding a polymerase with overall higher activity and substrate promiscuity (**Table 1**). Modifications to this procedure, such as mutagenesis of the oligonucleotide-binding pocket or screening under alternative stringent reaction conditions, can be adapted to further enhance the activity and robustness of this enzyme.

The metal-dependence of mammalian PolX family enzymes has been studied in detail,[35] and their catalytic mechanism is known to involve two metal ions (metal A and metal B). Metal A is coordinated by the dNTP α-phosphate and the primer strand 3'OH and serves as a Lewis acid to activate the 3'OH group for nucleophilic attack, while metal B is chelated by the dNTP β and γ phosphates and serves to stabilize the pyrophosphate leaving group. Metal B is typically Mg^{2+}, while maximal template-independent polymerase activity of TdT and Polμ is observed when metal A is a divalent transition metal such as Mn^{2+}, Co^{2+} or Zn^{2+}. The similar metal-dependence of RvPolX, with maximal activity obtained with a combination of Mg^{2+} and Mn^{2+}, is consistent with a common reaction mechanism with other PolX enzymes (**Figure S10**). Analysis of the phylogenetic tree of PolX in animals (Eumetazoa) reveals that numerous diverse sequences remain to be explored, particularly in invertebrates. Like RvPolX, many of the invertebrate PolX variants, contain an N-terminal BRCT domain associated with NHEJ, where template-independent DNA polymerase activity has been proposed to provide flexibility in the repair of various types of double-strand breaks.[36] Thus, a systematic investigation of diverse invertebrate PolX enzymes, particularly those containing a BRCT domain, has the potential to enrich the collection of template-independent polymerases with unique characteristics suitable for various DNA synthesis applications.

In RvPolX, G513 is part of a GW motif, which is also present in TdT and Polμ, but is replaced with a YF motif in Polλ and Polβ, thought to underlie the different nucleotide selectivity of these enzymes.[37-39] The Y residue in Polλ and Polβ is positioned near the dNTP 2' carbon and is proposed to act as a steric gate, preventing the binding and incorporation of rNTPs.[1, 40] Substitution of Y with G in TdT and Polμ expands the nucleotide binding pocket, allowing the binding and incorporation of both dNTPs and rNTPs.[38] Interestingly, RvPolX is the only Polλ-like enzyme containing the GW motif, while the others contain YF, HY or other motifs (**Figure S11**). Given its proximity to the dNTP substrate, the increased activity of the G513A mutation may result from the binding of the dNTP in a more reactive conformation, or through rigidification of the dNTP binding pocket.

In conclusion, our research underscores the potential for discovering new template-independent polymerases from various organisms within the broader PolX family, which may offer properties distinct from those of the well characterized mammalian TdT. Our work on RvPolX led to the development of an enzyme with salt-tolerant template-independent DNA polymerase activity, adding to the enzymatic toolkit available for *de novo* DNA synthesis. The properties of RvPolX may relate to its expected role in the NHEJ pathway of DNA repair in

the presence of various stressors *R. varieornatus*. For applications in enzymatic DNA synthesis, further work is required to engineer RvPolX for single base incorporation, by increasing its activity for incorporation of 3′-protected dNTPs,[8] or by developing enzyme-dNTP conjugates as demonstrated for TdT.[9]

MATERIALS AND METHODS

Phylogenetic Analysis

Sequences belonging to DNA-directed DNA polymerase X (InterPro accession: IPR002054), in Eumetazoa, were retrieved from the UniProt database (accessed on 15 July 2022).[41] To reduce redundancy, the sequences were mapped onto 230 UniRef50 clusters (UniRef50 contains sequences clustered at ≥50% identity, **Dataset S1**).[42] Four previously studied PolXs, including *Saccharomyces cerevisiae* (strain ATCC 204508 / S288c) DNA polymerase IV (ScPolIV), *Schizosaccharomyces pombe* (strain 972 / ATCC 24843) DNA polymerase type-X family protein (SpPolIV), *Oryza sativa* subsp. *japonica* DNA polymerase lambda (OsPolλ), and *Arabidopsis thaliana* DNA polymerase lambda (AtPolλ), were included in the analysis. A multiple sequence alignment was constructed using ClustalW,[43] and the phylogenetic tree was constructed using MEGA 11 software,[44] with the maximum likelihood method.[45]

Expression Constructs

Synthesis of the codon-optimized, full-length gene sequence of RvPolX (UniProt accession number: A0A1D1UV65), in a pET-28a (+) vector (Novagen) with a cleavable N-terminal His-SUMO tag, was carried out by Twist Bioscience. To prepare mutant for CD spectroscopy studies, mutant with Δ1-258 was generated. For protein engineering studies, the selected residues (R276, E281, C425, S427, R432, G513, W514, M518, Y519, R525, R522, and D537) located at the incoming nucleotide binding pocket were subjected to mutations. Twelve single residues were chosen, and each was mutated to 19 different amino acids using the NEBuilder HiFi DNA Assembly master kit (New England Biolabs), following the manufacturer's instructions.

Recombinant Protein Production

The experiment was conducted following a previously reported protocol by Law et al. [46]. The constructs were transformed into *Escherichia coli* Rosetta 2 (DE3). The bacterial culture

carrying the plasmid of interest was grown in LB medium supplemented with antibiotics at 37 °C with shaking at 250 rpm. Upon reaching OD_{600} reach around 0.8, 0.5 mM isopropyl-1-thio-β-D-galactopyranoside (IPTG) was introduced, and the culture was then incubated at 16 °C with shaking at 200 rpm overnight. The bacterial cultures were subsequently harvested, lysed, and the supernatant proteins were purified using TALON® Metal Affinity Resin (Takara). To remove the N-terminal His-SUMO tag, the purified proteins were incubated with SUMO protease overnight at 4 °C with an approximate molar ratio of 1:25. The purified proteins were further subjected to purification via HiTrap™ Heparin HP affinity columns (GE Healthcare Life Sciences). Finally, 4-20% mini-PROTEAN® TGX™ precast protein gel (Bio-Rad) was used to analyze the purified recombinant proteins.

DNA Polymerase Activity Assays

The experiment was conducted following a previously reported protocol by Loc'h et al. [47], with minor modifications. A 5'-Alexa Fluor 488-labeled 14-mer single stranded DNA (Alexa-ssDNA14) oligonucleotide (5'-Alexa488-TACGCATTAGCCTG- 3') and ssDNA14A/T/G/C were used as the DNA primer (**Table S2**). Assay mixtures were prepared on ice. The reaction was initiated by adding 0.3 mM of the respective dNTP, incubated at 25 °C for 5 to 30 minutes, and inactivated at 95 °C for 2 minutes using a thermocycler. The products were analyzed by gel electrophoresis on an 18% acrylamide gel with 8 M urea (urea-PAGE). The gel bands were visualised through the fluorescence of the Alexa488 fluorophore using ChemiDoc MP Imaging System (Bio-Rad), and the band intensities were quantified using Image Lab software (Bio-Rad).

To assay the primer elongation by RvPolX under different pH conditions, a reaction mixture containing 4 μM RvPolX, 50 nM Alexa-ssDNA14, 1 mM Tris(2-carboxyethyl)phosphine hydrochloride (TCEP), 1% glycerol, and 1 mM $MnCl_2$ was prepared in various pH buffers, including 50 mM MES pH 6.0, 50 mM Tris-HCl pH 6.8, 50 mM 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES) pH 7.4, 50 mM Tris-HCl pH 8.0, or 50 mM Tris-HCl pH 8.8. The reaction was initiated by adding nucleotide.

To assay the primer elongation by RvPolX in the presence of different divalent ions, a reaction mixture containing 50 mM Tris–HCl pH 8.8, 4 μM RvPolX, 50 nM Alexa-ssDNA14, 1 mM TCEP, and 1% glycerol was prepared. Various divalent ions, including 5 mM $CaCl_2$, 5 mM $LiCl_2$, 5 mM $MgCl_2$, 1 mM $MnCl_2$, or 1 mM $ZnCl_2$, were added individually. To assay the

primer elongation by RvPolX in the presence of a mixture of two different divalent ions, 1 mM $MnCl_2$ and 5 mM $MgCl_2$ were added. The reaction was initiated by adding nucleotide, and terminated after 5 minutes.

To assay the primer elongation by RvPolX at different temperatures and salt concentration, various concentrations of NaCl (ranging from 0 to 700 mM) were added to the reaction mixture containing 50 mM Tris-HCl pH 8.8, 4 μM RvPolX, 50 nM Alexa-ssDNA14, 1 mM TCEP, 1 mM $MnCl_2$, 5 mM $MgCl_2$, and 1% glycerol. The reaction was initiated by adding nucleotide solution containing various concentrations of NaCl (ranging from 0 to 700 mM), followed by rapid stirring. The reaction was terminated after 30 minutes.

To assay primer elongation by RvPolX at different temperatures, a reaction mixture containing 50 mM Tris–HCl pH 8.8, 4 μM RvPolX, 50 nM Alexa-ssDNA14, 1 mM TCEP, 1 mM $MnCl_2$, 5 mM $MgCl_2$, and 1% glycerol was prepared. The reaction mixture and nucleotide solutions were pre-incubated separately at different temperatures for 5 minutes. The reaction was then initiated by adding nucleotide solution, followed by rapid stirring. The reaction was terminated after 5 minutes.

Circular Dichroism Spectroscopy

The experiment was conducted following a previously reported protocol.[48] The protein solutions were diluted to a final concentration of 3 μM in a buffer containing 25 mM HEPES at pH 7.4, 10% glycerol, and 500 mM NaCl. Prior to the measurements, all samples were incubated at room temperature for 5 minutes. A quartz cuvette with a path length of 0.1 cm, filled with 300 μL of the sample, was used for the measurements. Circular dichroism (CD) measurements were conducted using a Jasco J-815 Spectrometer (Jasco), at a wavelength range of 200 to 250 nm, with the incubation temperature gradually increased from 25 to 50 °C at intervals of 0.1 nm, and 10 accumulations recorded for each reading. The resulting data was analyzed using the web-based server BeStSel (Beta Structure Selection).[49]

Computational Modelling

The protein sequence of RvPolX was retrieved from the UniProt database (assession number: A0A1D1UV65). The whole modelling protocol is as follows: (1) three initial models of RvPolX were firstly predicted by AlphaFold v2.0 program;[50] (2) the best model with the top ranked score was selected and its N-terminus (1-234aa) containing a long-disordered loop was

truncated; (3) this pruned model was subjected to the sidechain optimization in Sybyl-X package (v2.1),[51] which afforded diverse models; (3) the model with the lowest energy was selected for the subsequent modelling; (4) two coordinated metal ions (e.g., Mn^{2+}) in the catalytic site were modelled with the following procedure: one ion was positioned in the center of two carboxyl groups of D437 and D439, while the other was placed in the center of three carboxyl groups of D437, D439, and D498; subsequently, two metal ions and the sidechains of D437, D439, and D498 were optimized in Sybyl-X package to relax their conformations, while the remaining part of enzyme was fixed. (5) a single-strand DNA (ssDNA, 5'-AGCCTG-3') was placed in the binding pocket. More specifically, the coordinates of ssDNA backbone were initialized from the crystal structure (PDB ID: 4I27)[35] of its close homolog TdT with ssDNA by aligning the structure of RvPolX to that of TdT; and then the missing bases were added by tleap program in the AmberTool package;[52] only the ssDNA was optimized in the binding pocket by the Sybyl package, while the whole enzyme and two metal ions were restrained during the optimization; both ssDNA and its neighbouring sidechains on the RvPolX were fully relaxed with the restraint of the remaining system; (6) the sidechains of RvPolX and bases of ssDNA were optimized with the restraint of backbone; (7) the whole system including RvPolX, ssDNA, and two metal ions is fully optimized for subsequent molecular dynamics simulation. The complex model from the previous protocol was further optimized by molecular dynamics simulation in the explicit water environment with the following steps. (1) the complex model was solvated by TIP3P water model[53] and neutralized by counter ions with the NaCl concentration of 150 mM; (2) the whole system was minimized with position restraint of the heavy atoms by exerting a harmonic force constant of 5.0 kcal mol^{-1} $Å^{-2}$, and subsequently fully relaxed without any restraint; (3) the whole system was gradually heated to 298.15K with the position restraint of 5.0 kcal mol^{-1} $Å^{-2}$ for 25ps; (4) the whole system was equilibrated in the NVT ensemble for 25ps; (5) the whole system was subjected to three rounds of 500ps equilibration in NPT ensemble with the successive decrease of harmonic force constants (5.0, 1.0, and 0.0 kcal mol^{-1} $Å^{-2}$, respectively); (6) the whole system was subjected to the production simulation without any restraint at the pressure of 1 atm and temperature of 298.15K; Monte-Carlo barostat[54] and Langevin thermostat[55] were employed to control the pressure and temperature, respectively; ff14SB[56] and OC15[57] force field parameters were applied for RvPolX and ssDNA, respectively; the production simulation time was set to 50ns; Smooth Particle Mesh Ewald (SPME) method[58] was employed to calculate the long-range electrostatic interaction with the cutoff of 9 Å; the whole simulation was conducted for three times in the AMBER22 program.[59]

Five typical models were retrieved from the previous molecular dynamics simulations by K-Mean cluster analysis, minimized, and processed in the Sybyl package to remove water molecules and counter ions. Only two models with an open binding site exposure to the incoming nucleotide were kept for subsequent molecular docking simulations. The incoming nucleotide dGTP were modelled and optimized in Sybyl package (v2.1) and subsequently docked into the binding site of RvPolX/ssDNA complex model by the GOLD program v2018.[60] The docking parameters were provided as follows: scoring function is ChemScore; population size is 400; selection pressure is 1.1; number of operations is 500,000; number of islands is 5; niche size is 2; crossover frequency is 95%; mutation frequency is 95%; migration frequency is 10%. After docking simulations, all the sampled conformations for each incoming nucleotide were further filtered by the following criteria: (1) the base of the incoming nucleotide can form favourable π-π stacking with the ssDNA (5'-AGCCTG-3'); (2) the phosphate group(s) of the incoming nucleotide can form the salt bridge with the nearby residues such as lysine or arginine; (3) the phosphate group(s) of the incoming nucleotide can form the coordination interaction with a metal ion in the catalytic pocket. According to this filtering, a reasonable conformation (**Figure 4**) with the incoming nucleotide dGTP was achieved, minimized in the Sybyl package, and finally rescored in the GOLD program.

Protein Engineering and Screening of RvPolX Mutants

Plasmids encoding mutations to the RvPolX dNTP binding pocket residues were transformed into Rosetta™ 2(DE3) *E. coli* cells. For protein expression, a single colony of each mutant was selected and grown in 4 mL LB cultures. The cells were harvested by centrifugation at 3,300 *g* for 1 minute and then lysed by resuspending them in 500 µL of B-PER™ Bacterial Protein Extraction Reagent (Thermo Scientific), followed by incubation at room temperature for 30 minutes. After centrifugation at 16,000 *g* for 20 minutes at 4 °C, the supernatant was incubated with 20 µL of pre-washed Dynabeads™ (Thermo Scientific, US) in 4 °C for 1 hour. After that, the beads were separated and washed 3 times with wash buffer (25 mM HEPES pH 7.4, 500 mM NaCl, 0.01% Tween-20, 10 mM imidazole). The immobilized protein on the Dynabeads™ was incubated with a reaction buffer containing 25 mM HEPES pH 7.4, 1 µM ssDNA14, 1 mM dCTP, 1 mM TCEP, 1 mM $MnCl_2$, 5 mM $MgCl_2$, and 1% glycerol at 25 °C and 1000 rpm for 60 minutes. After centrifugation, 30 µL of supernatant was analyzed using a UHPLC system (Shimadzu, Japan). Each analysis involved injecting 10 µL of the sample onto a Clarity 2.6 µm Oligo-XT 100 Å, LC Column 100 x 4.6 mm (Phenomenex). Mobile phase A (15 mM

trimethylamine, 400 mM 1,1,1,3,3,3-hexafluoro-2-propanol, 10% methanol, pH 8.0) and mobile phase B (100% methanol) were prepared. A 15-minute gradient program ranging from 16% to 36% mobile phase B was employed to analyze the extended products.

Microscale Thermophoresis

Microscale thermophoresis (MST) study was conducted following the protocol described by Kuznetsova et al. [61] to investigate the interaction between RvPolX and a mutant with four dNTPs. This study utilized the Monolith^{TM} NT.labelFree system by NanoTemper Technologies. A series of samples containing nucleotides (at concentrations ranging from 0.01 to 0.15 μM) and enzyme (0.8 μM) in a reaction buffer consisting of 50 mM HEPES at pH 7.4, 100 mM NaCl, 5 mM $MgCl_2$, 1 mM $MnCl_2$, and 2 mM TCEP were incubated at room temperature for 10 minutes before being loaded individually into Monolith^{TM} NT.labelFree capillaries. The measurements were conducted at room temperature with 20% LED power and 20% MST power, and the resulting data were analyzed using MO Affinity Analysis to determine the dissociation constants.

Kinetic Analysis

The experiment was conducted following a previously reported protocol.[61] To monitor the time-dependent primer elongation, a reaction mixture containing 4 μM RvPolX, 50 mM Tris–HCl pH 8.8, 50 nM Alexa-ssDNA14, 1 mM TCEP, 1 mM $MnCl_2$, 5 mM $MgCl_2$, and 1% glycerol was prepared. After a 5 minutes pre-incubation at 25 °C, the reaction was initiated by adding nucleotide. The reaction was halted at specific time intervals and analyzed using 18% denaturing urea-PAGE, visualized using Chemidoc (BioRad), and the bands quantified using Image Lab (Bio-Rad). Curve fitting was performed using the Graphpad Prism 9. The oligonucleotide chain growth polymerization kinetics was modelled as a Poisson distribution. For each of the assays, a specific oligonucleotide product band (k) that allowed us to observe the time-dependent accumulation and subsequent depletion of the oligonucleotide product, was selected for analysis. The apparent rate of chain elongation (V_{obs}) was estimated by fitting the changes in band intensity (I) over time (t) to a Poisson probability distribution function, with a normalization constant (A).

552 $$I = At^k e^{-V_{obs}t}$$

553

554 **Supporting Information:**

555 Protein production of *Homo sapiens* terminal deoxynucleotidyl transferase (HsTdT); DNA
556 polymerase activity assay using modified nucleotides; DNA polymerase activity assay of
557 G513A+R522I and HsTdT; DNA-binding assay (note); domain structure and 3D structures of
558 PolX family members (Figure S1); multiple sequence alignment of the catalytic domains of
559 RvPolX and members of the PolX family (Figure S2); net charge and pH of RvPolX using
560 Peptide Property Calculator from NovoPro (Figure S3); denaturing urea-PAGE analysis of the
561 RvPolX template-independent DNA polymerase assays with four different dNTPs and
562 different divalent metal ions (Figure S4); protein engineering of RvPolX (Figure S5);
563 denaturing urea-PAGE analysis of template-independent DNA polymerase assays using 3′-
564 ONH_2 modified nucleotides (Figure S6); DNA-binding analysis of WT and G513A+R522I
565 through fluorescence anisotropy (Figure S7); denaturing urea-PAGE analysis of the
566 G513A+R522I and HsTdT template-independent DNA polymerase assays (Figure S8);
567 denaturing urea-PAGE analysis of template-independent DNA polymerase assays using 3′-
568 ONH_2 modified nucleotides with G513A+R522I and commercial calf thymus TdT (Figure S9);
569 denaturing urea-PAGE analysis of the G513A+R522I template-independent DNA polymerase
570 assays with different divalent metal ions (Figure S10); WebLogo presentation of the GW motif
571 and the R522 residue within the thumb domain (Figure S11); list of SwissProt reviewed
572 Eumetazoan PolX sequences (Table S1); list of sequences of DNA primer used for DNA
573 polymerase activity assay (Table S2).

574

575 **ACKNOWLEDGMENTS** We thank the Agency for Science, Research, and Technology of
576 Singapore Visiting Investigator Program Grant 1535j00137 and the Advanced Manufacturing
577 and Engineering Programmatic Grant A18A9b0060. We also thank the computational
578 resources support from A*STAR BII and NSCC (13002460, 13002658, and 13003556) and
579 appreciate the software support and kind help from Dr Hao Fan in A*STAR BII. We extend
580 our gratitude to Dr. Xixian Chen and Mr. Jing Sen Ong for their assistance with the automation
581 system. We are thankful to Dr. Dahai Luo for providing SUMO protease as a gift. Additionally,
582 we appreciate Dr. Zebin Hong for providing technical advice on microscale thermophoresis.
583 We also appreciate Ms. Hui Pen Tan, Dr. Michiko Kimoto, and Dr. Ichiro Hirao for providing

technical advice on CD spectroscopy. Special thanks to Dr. Tse How Jason Tan for his valuable scientific advice.

Author contributions Y-S.L., H.Z., and E.L.A. conceived the project and designed the experiments. Y-S.L. N.A.M., Y.W., and F.L. conducted the wet and dry lab experiments. Y-S.L., and Y.W. wrote the paper with input from all authors. All authors read and approved the final manuscript.

Notes: The authors declare the following competing financial interest(s): This work has been included in patent application with the Intellectual Property Office of Singapore (application number: 10202400148X) by the Singapore Institute of Food and Biotechnology Innovation, Agency for Science, Technology, and Research (A∗STAR).

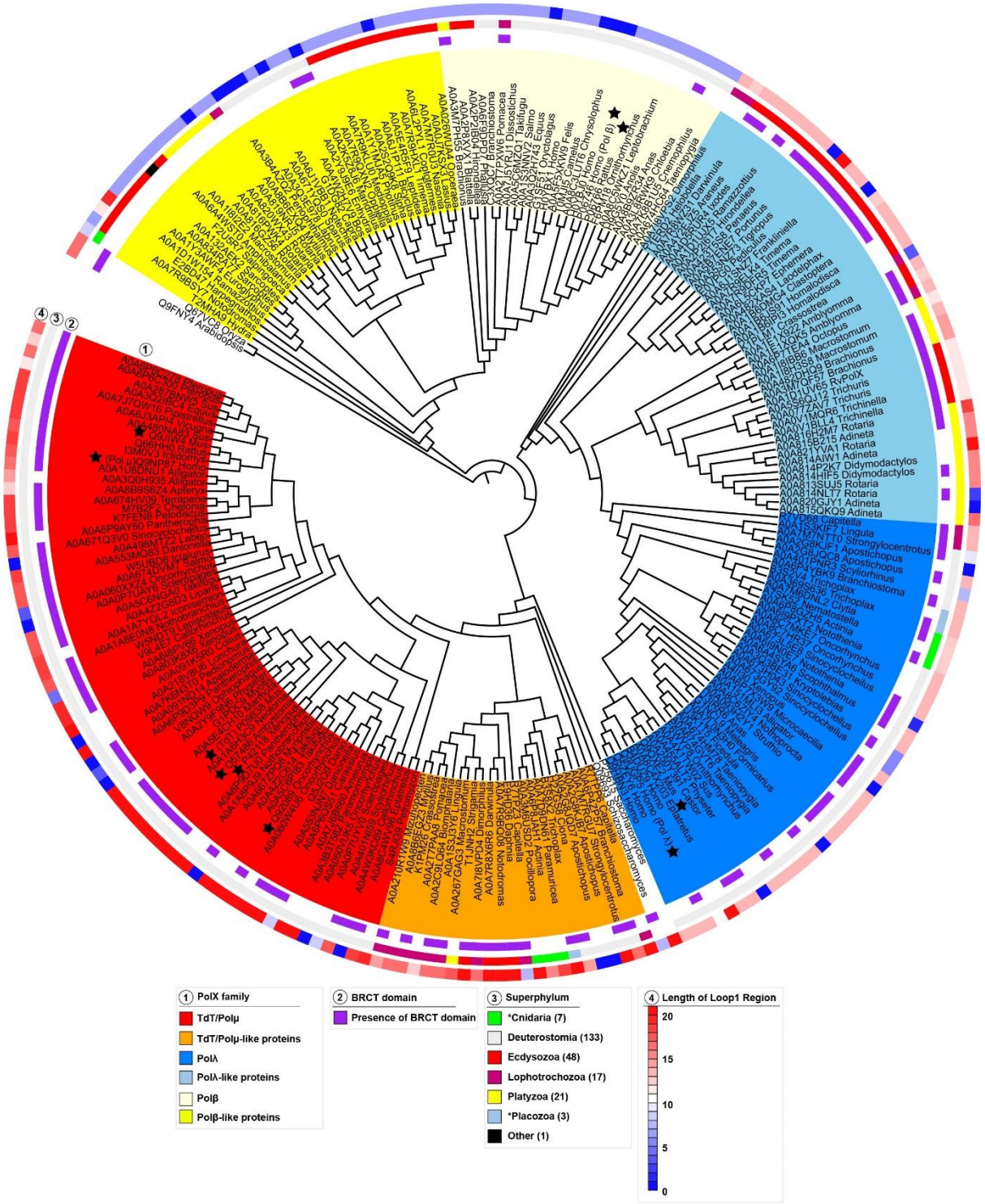

Figure 1. Maximum likelihood phylogenetic tree of DNA polymerase X family proteins in Eumetazoa. The 230 PolX family proteins can be divided into six groups: TdT/Polμ, TdT/Polμ-like proteins, Polλ, Polλ-like proteins, Polβ, and Polβ-like proteins. SwissProt reviewed sequences are denoted with black stars. The leaf label states the UniProt ID and genus. Numbers in circle are annotated according to the legend.

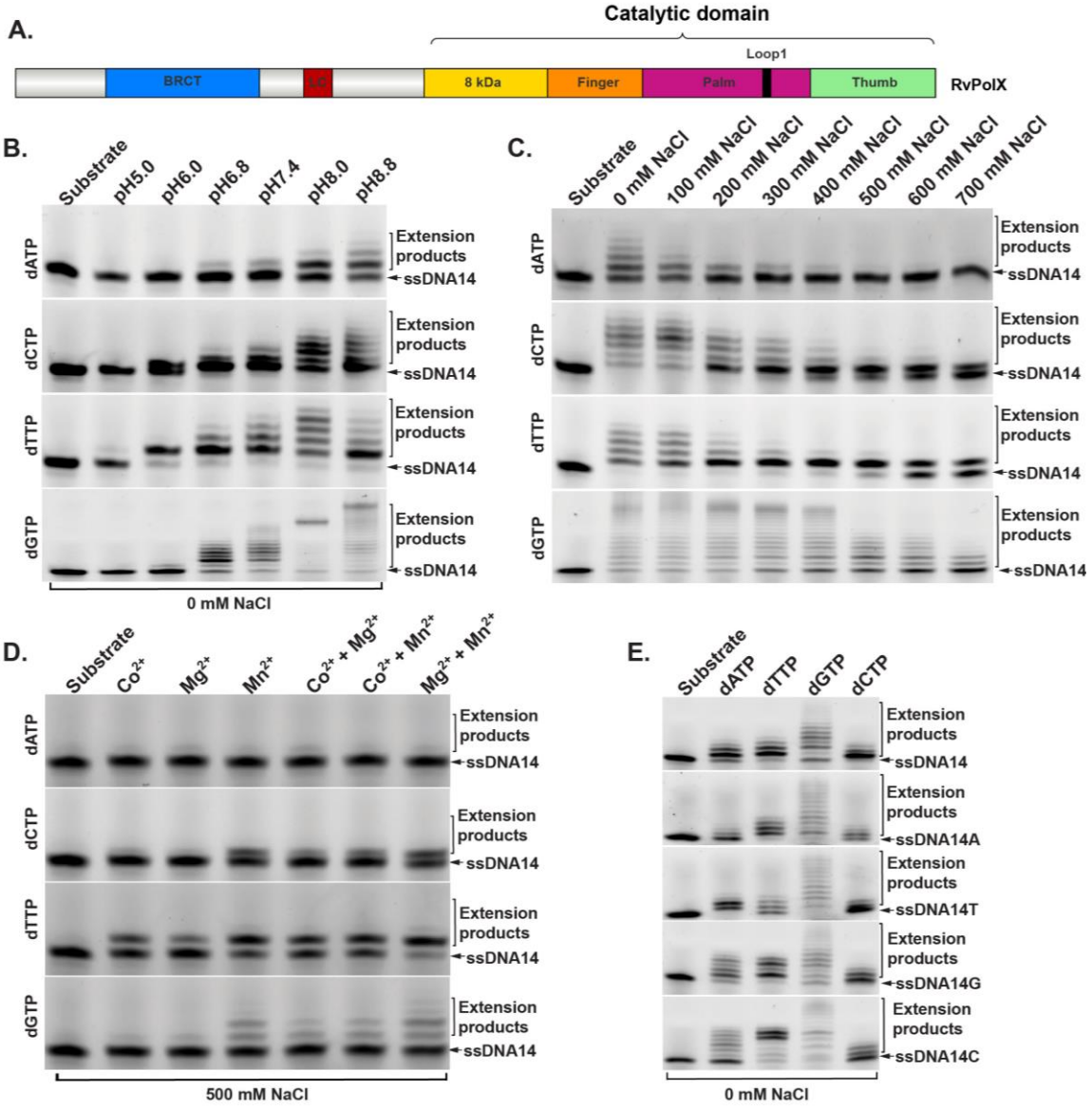

Figure 2. Domain structure and template-independent DNA polymerase activity of RvPolX. (A) Domain structure of RvPolX, with the BRCA1 C-terminal (BRCT) domain in blue, and the low-complexity (LC) region in red. The catalytic domain consists of the 8 kDa domain (yellow), fingers domain (orange), palm domain (pink), and thumb domain (green). The loop1 region within the palm domain is colored in black. Denaturing urea-PAGE assay of RvPolX-catalyzed template-independent extension of a 14-mer single-stranded DNA (ssDNA14) under varying conditions, including various (B) pH buffers, (C) salt concentrations, (D) divalent metal ions, and (E) ssDNA substrates (the substrates ssDNA14A, ssDNA14T, ssDNA14G, and ssDNA14C consist of ssDNA14 extended by A, T, G or C, respectively).

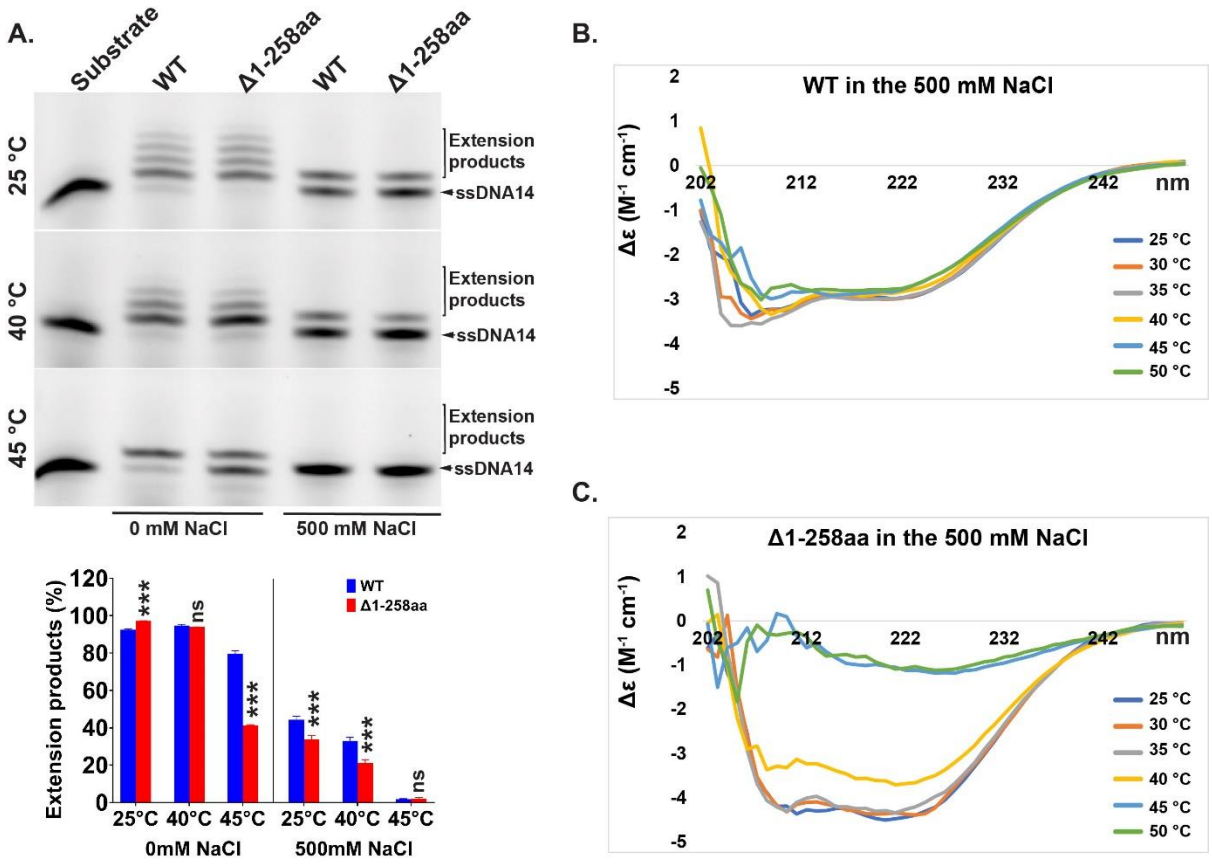

Figure 3. Contribution of the RvPolX N-terminal region to its activity and stability. (A) Denaturing urea-PAGE analysis of the DNA polymerase assays of WT and Δ1-258aa performed at 25 °C, 40 °C, and 45 °C, in the presence or absence of 500 mM NaCl. Each assay was conducted in triplicate, and the band intensities of the extension products were quantified and plotted on the graph. Error bars reflect the standard deviations, and significance was determined by one-way analysis of variance (ANOVA) with Dunnett's posttest comparing mutants to WT (vehicle control). *, P<0.05; **, P<0.01; ***, P < 0.001; ns, not significant. The far-UV CD spectra analysis of (B) WT and (C) Δ1-258aa, was conducted at a range of temperatures (25 °C to 50 °C), in the presence of 500 mM NaCl.

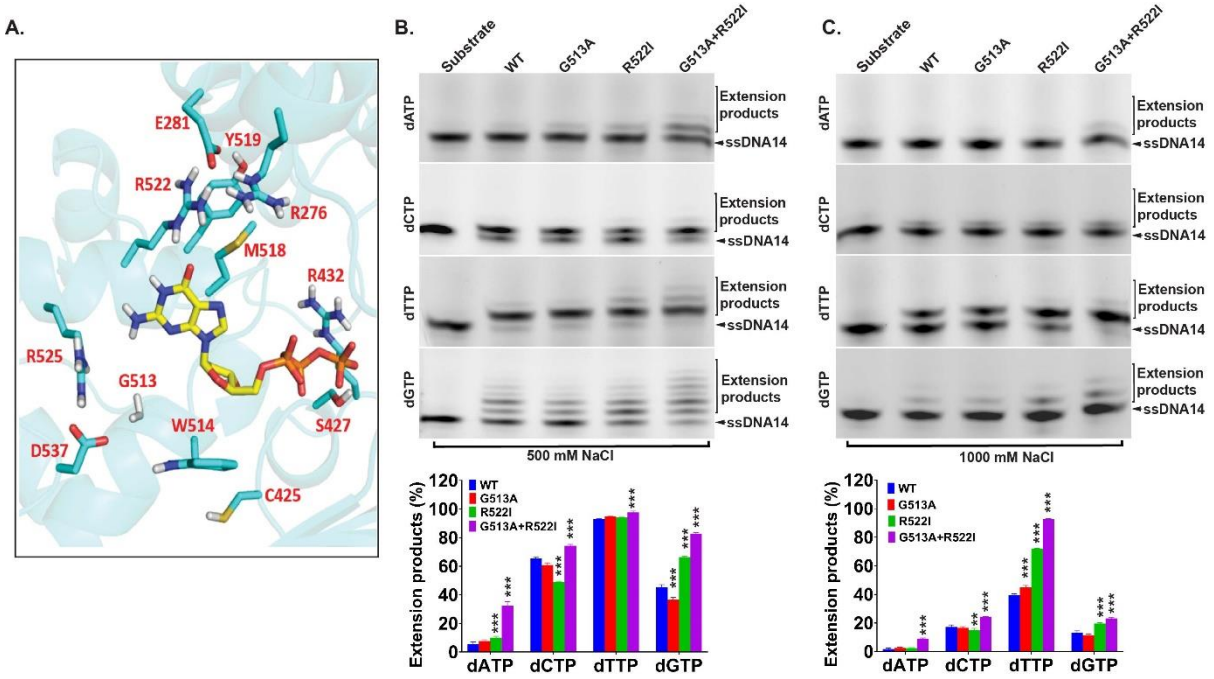

Figure 4. Protein engineering of RvPolX. (A) Twelve amino acid residues around the dNTP binding pocket were selected for saturation mutagenesis. Denaturing urea-PAGE analysis shows the effect of RvPolX mutations on template-independent DNA polymerase activity at (B) 500 mM NaCl and (C) 1000 mM NaCl, with four different dNTPs. Each assay was conducted in triplicate, and the band intensities of the extension products were quantified and plotted on the graph. Error bars reflect the standard deviations, and the significance was determined by one-way analysis of variance (ANOVA) with Dunnett's posttest comparing mutants to WT (vehicle control). *, $P<0.05$; **, $P<0.01$; ***, $P < 0.001$; ns, not significant.

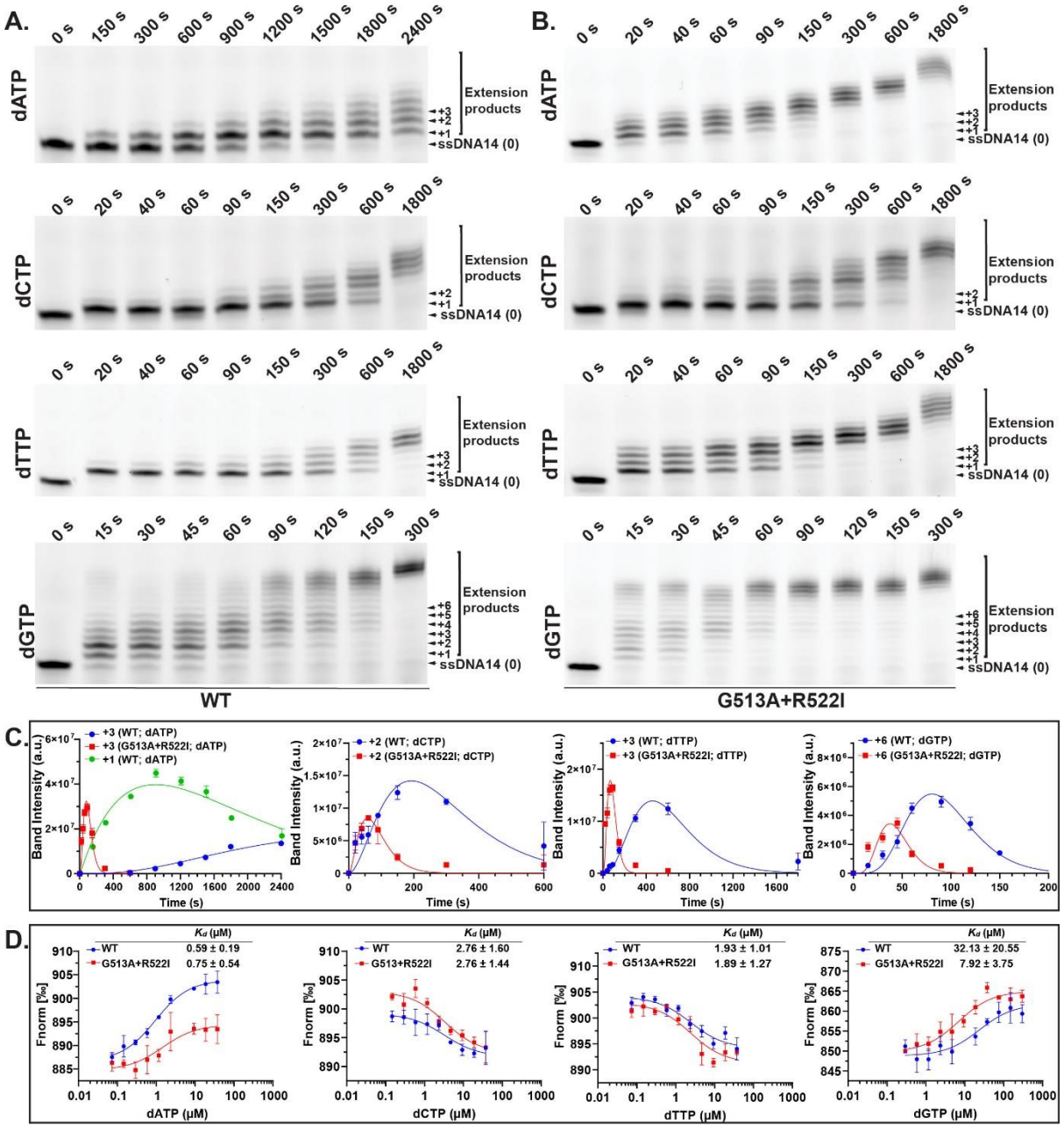

Figure 5. Kinetic analysis and microscale thermophoresis (MST) studies of WT and G513A+R522I. Denaturing urea-PAGE analysis shows time-dependent DNA polymerization catalyzed by (A) WT and (B) G513A+R522I with four different dNTPs. The labels on the products bands (+1, +2, +3, *etc*.) indicate ssDNA14 extended by a corresponding number of nucleotides. (C) Time-dependent changes in intensity for selected product bands. Each assay was conducted in triplicate, and the data are presented as mean values, with error bars reflecting the standard deviations, and the solid lines representing the exponential fits. (D) MST titration curves for the interactions between WT and G513A+R522I with four different dNTPs, used to determine the dissociation constants (K_d) for each dNTP. Each assay was conducted in

triplicate, and the data are presented as mean values, with error bars reflecting the standard errors.

Table 1. The apparent rates of chain elongation (V_{obs}) of WT and G513A+R522I.

	Band of extension product	V_{obs} of WT (s^{-1})	V_{obs} of G513A+R522I (s^{-1})
dATP	**3**	**0.0009 ± 0.0001**	**0.0380 ± 0.0014**
dATP	**1**	**0.0011 ± 0.0000**	**N.A.**
dCTP	**2**	**0.0104 ± 0.0007**	**0.0356 ± 0.0018**
dTTP	**3**	**0.0066 ± 0.0002**	**0.0439 ± 0.0020**
dGTP	**6**	**0.0744 ± 0.0010**	**0.1597 ± 0.0065**

REFERENCES

(1) Loc'h, J.; Delarue, M. Terminal deoxynucleotidyltransferase: the story of an untemplated DNA polymerase capable of DNA bridging and templated synthesis across strands. *Current opinion in structural biology* **2018**, *53*, 22-31.

(2) Uchiyama, Y.; Takeuchi, R.; Kodera, H.; Sakaguchi, K. Distribution and roles of X-family DNA polymerases in eukaryotes. *Biochimie* **2009**, *91* (2), 165-170. DOI: 10.1016/j.biochi.2008.07.005.

(3) Ma, Y.; Lu, H.; Tippin, B.; Goodman, M. F.; Shimazaki, N.; Koiwai, O.; Hsieh, C. L.; Schwarz, K.; Lieber, M. R. A biochemically defined system for mammalian nonhomologous DNA end joining. *Mol Cell* **2004**, *16* (5), 701-713. DOI: 10.1016/j.molcel.2004.11.017.

(4) Loc'h, J.; Rosario, S.; Delarue, M. Structural Basis for a New Templated Activity by Terminal Deoxynucleotidyl Transferase: Implications for V(D)J Recombination. *Structure* **2016**, *24* (9), 1452-1463. DOI: 10.1016/j.str.2016.06.014.

(5) Doricchi, A.; Platnich, C. M.; Gimpel, A.; Horn, F.; Earle, M.; Lanzavecchia, G.; Cortajarena, A. L.; Liz-Marzan, L. M.; Liu, N.; Heckel, R.; et al. Emerging Approaches to DNA Data Storage: Challenges and Prospects. *ACS Nano* **2022**, *16* (11), 17552-17571. DOI: 10.1021/acsnano.2c06748.

(6) Lee, H. H.; Kalhor, R.; Goela, N.; Bolot, J.; Church, G. M. Terminator-free template-independent enzymatic DNA synthesis for digital information storage. *Nat Commun* **2019**, *10* (1), 2383. DOI: 10.1038/s41467-019-10258-1.

(7) Xu, C.; Zhao, C.; Ma, B.; Liu, H. Uncertainties in synthetic DNA-based data storage. *Nucleic Acids Research* **2021**, *49* (10), 5451-5469. DOI: 10.1093/nar/gkab230.

(8) Lu, X.; Li, J.; Li, C.; Lou, Q.; Peng, K.; Cai, B.; Liu, Y.; Yao, Y.; Lu, L.; Tian, Z. Enzymatic DNA synthesis by engineering terminal deoxynucleotidyl transferase. *ACS Catalysis* **2022**, *12* (5), 2988-2997.

(9) Palluk, S.; Arlow, D. H.; de Rond, T.; Barthel, S.; Kang, J. S.; Bector, R.; Baghdassarian, H. M.; Truong, A. N.; Kim, P. W.; Singh, A. K.; et al. De novo DNA synthesis using polymerase-nucleotide conjugates. *Nature Biotechnology* **2018**, *36* (7), 645-650. DOI: 10.1038/nbt.4173.

(10) Eisenstein, M. Enzymatic DNA synthesis enters new phase. *Nature biotechnology* **2020**, *38* (10), 1113-1116.

(11) Roy, S.; Choudhury, S. R.; Singh, S. K.; Das, K. P. AtPollambda, a homolog of mammalian DNA polymerase lambda in Arabidopsis thaliana, is involved in the repair of UV-B induced DNA damage through the dark repair pathway. *Plant Cell Physiol* **2011**, *52* (2), 448-467. DOI: 10.1093/pcp/pcr002.

(12) Leem, S. H.; Ropp, P. A.; Sugino, A. The yeast Saccharomyces cerevisiae DNA polymerase IV: possible involvement in double strand break DNA repair. *Nucleic Acids Res* **1994**, *22* (15), 3011-3017. DOI: 10.1093/nar/22.15.3011.

(13) Genois, M. M.; Paquet, E. R.; Laffitte, M. C.; Maity, R.; Rodrigue, A.; Ouellette, M.; Masson, J. Y. DNA repair pathways in trypanosomatids: from DNA repair to drug resistance. *Microbiol Mol Biol Rev* **2014**, *78* (1), 40-73. DOI: 10.1128/MMBR.00045-13.

(14) Garcia-Escudero, R.; Garcia-Diaz, M.; Salas, M. L.; Blanco, L.; Salas, J. DNA polymerase X of African swine fever virus: insertion fidelity on gapped DNA substrates and AP lyase activity support a role in base excision repair of viral DNA. *J Mol Biol* **2003**, *326* (5), 1403-1412. DOI: 10.1016/s0022-2836(03)00019-6.

(15) Redrejo-Rodriguez, M.; Rodriguez, J. M.; Suarez, C.; Salas, J.; Salas, M. L. Involvement of the reparative DNA polymerase Pol X of African swine fever virus in the maintenance of viral genome stability in vivo. *J Virol* **2013**, *87* (17), 9780-9787. DOI: 10.1128/JVI.01173-13.

(16) Nakane, S.; Nakagawa, N.; Kuramitsu, S.; Masui, R. Characterization of DNA polymerase X from Thermus thermophilus HB8 reveals the POLXc and PHP domains are both required for 3'-5' exonuclease activity. *Nucleic Acids Res* **2009**, *37* (6), 2037-2052. DOI: 10.1093/nar/gkp064.

(17) Banos, B.; Lazaro, J. M.; Villar, L.; Salas, M.; de Vega, M. Characterization of a Bacillus subtilis 64-kDa DNA polymerase X potentially involved in DNA repair. *J Mol Biol* **2008**, *384* (5), 1019-1028. DOI: 10.1016/j.jmb.2008.09.081.

(18) Prostova, M.; Shilkin, E.; Kulikova, A. A.; Makarova, A.; Ryazansky, S.; Kulbachinskiy, A. Noncanonical prokaryotic X family DNA polymerases lack polymerase activity and act as exonucleases. *Nucleic Acids Res* **2022**, *50* (11), 6398-6413.

(19) Hashimoto, T.; Horikawa, D. D.; Saito, Y.; Kuwahara, H.; Kozuka-Hata, H.; Shin-i, T.; Minakuchi, Y.; Ohishi, K.; Motoyama, A.; Aizu, T. Extremotolerant tardigrade genome and improved radiotolerance of human cultured cells by tardigrade-unique protein. *Nature communications* **2016**, *7* (1), 12808.
(20) Yoshida, Y.; Koutsovoulos, G.; Laetsch, D. R.; Stevens, L.; Kumar, S.; Horikawa, D. D.; Ishino, K.; Komine, S.; Kunieda, T.; Tomita, M.; et al. Comparative genomics of the tardigrades Hypsibius dujardini and Ramazzottius varieornatus. *PLoS Biol* **2017**, *15* (7), e2002266. DOI: 10.1371/journal.pbio.2002266.
(21) Horikawa, D. D.; Kunieda, T.; Abe, W.; Watanabe, M.; Nakahara, Y.; Yukuhiro, F.; Sakashita, T.; Hamada, N.; Wada, S.; Funayama, T.; et al. Establishment of a rearing system of the extremotolerant tardigrade Ramazzottius varieornatus: a new model animal for astrobiology. *Astrobiology* **2008**, *8* (3), 549-556. DOI: 10.1089/ast.2007.0139.
(22) Kasianchuk, N.; Rzymski, P.; Kaczmarek, L. The biomedical potential of tardigrade proteins: A review. *Biomed Pharmacother* **2023**, *158*, 114063. DOI: 10.1016/j.biopha.2022.114063.
(23) Mueller, G. A.; Moon, A. F.; DeRose, E. F.; Havener, J. M.; Ramsden, D. A.; Pedersen, L. C.; London, R. E. A comparison of BRCT domains involved in nonhomologous end-joining: introducing the solution structure of the BRCT domain of polymerase lambda. *DNA repair* **2008**, *7* (8), 1340-1351.
(24) DeRose, E. F.; Clarkson, M. W.; Gilmore, S. A.; Galban, C. J.; Tripathy, A.; Havener, J. M.; Mueller, G. A.; Ramsden, D. A.; London, R. E.; Lee, A. L. Solution structure of polymerase μ's BRCT domain reveals an element essential for its role in nonhomologous end joining. *Biochemistry* **2007**, *46* (43), 12100-12110.
(25) Romain, F.; Barbosa, I.; Gouge, J.; Rougeon, F.; Delarue, M. Conferring a template-dependent polymerase activity to terminal deoxynucleotidyltransferase by mutations in the Loop1 region. *Nucleic Acids Res* **2009**, *37* (14), 4642-4656. DOI: 10.1093/nar/gkp460.
(26) Flory, P. J. Molecular Size Distribution in Ethylene Oxide Polymers. *J Am Chem Soc* **2002**, *62* (6), 1561-1565. DOI: 10.1021/ja01863a066.
(27) Kaminski, A. M.; Pryor, J. M.; Ramsden, D. A.; Kunkel, T. A.; Pedersen, L. C.; Bebenek, K. Structural snapshots of human DNA polymerase mu engaged on a DNA double-strand break. *Nat Commun* **2020**, *11* (1), 4784. DOI: 10.1038/s41467-020-18506-5.
(28) Lee, H.; Wiegand, D. J.; Griswold, K.; Punthambaker, S.; Chun, H.; Kohman, R. E.; Church, G. M. Photon-directed multiplexed enzymatic DNA synthesis for molecular digital data storage. *Nat Commun* **2020**, *11* (1), 5246. DOI: 10.1038/s41467-020-18681-5.
(29) Wang, S.; Haque, F.; Rychahou, P. G.; Evers, B. M.; Guo, P. Engineered nanopore of Phi29 DNA-packaging motor for real-time detection of single colon cancer specific antibody in serum. *ACS Nano* **2013**, *7* (11), 9814-9822. DOI: 10.1021/nn404435v.
(30) Haque, F.; Wang, S.; Stites, C.; Chen, L.; Wang, C.; Guo, P. Single pore translocation of folded, double-stranded, and tetra-stranded DNA through channel of bacteriophage phi29 DNA packaging motor. *Biomaterials* **2015**, *53*, 744-752. DOI: 10.1016/j.biomaterials.2015.02.104.
(31) Gao, Y.; He, Y.; Chen, L.; Liu, X.; Ivanov, I.; Yang, X.; Tian, H. Chimeric Phi29 DNA polymerase with helix-hairpin-helix motifs shows enhanced salt tolerance and replication performance. *Microb Biotechnol* **2021**, *14* (4), 1642-1656. DOI: 10.1111/1751-7915.13830.
(32) Fuller, C. W.; Kumar, S.; Porel, M.; Chien, M.; Bibillo, A.; Stranges, P. B.; Dorwart, M.; Tao, C.; Li, Z.; Guo, W. Real-time single-molecule electronic DNA sequencing by synthesis using polymer-tagged nucleotides on a nanopore array. *Proceedings of the National Academy of Sciences* **2016**, *113* (19), 5233-5238.
(33) Delarue, M.; Boule, J. B.; Lescar, J.; Expert-Bezancon, N.; Jourdan, N.; Sukumar, N.; Rougeon, F.; Papanicolaou, C. Crystal structures of a template-independent DNA polymerase: murine terminal deoxynucleotidyltransferase. *EMBO J* **2002**, *21* (3), 427-439. DOI: 10.1093/emboj/21.3.427.
(34) Ramadan, K.; Maga, G.; Shevelev, I. V.; Villani, G.; Blanco, L.; Hubscher, U. Human DNA polymerase lambda possesses terminal deoxyribonucleotidyl transferase activity and can elongate RNA primers: implications for novel functions. *J Mol Biol* **2003**, *328* (1), 63-72. DOI: 10.1016/s0022-2836(03)00265-1.

(35) Gouge, J.; Rosario, S.; Romain, F.; Beguin, P.; Delarue, M. Structures of intermediates along the catalytic cycle of terminal deoxynucleotidyltransferase: dynamical aspects of the two-metal ion mechanism. *Journal of Molecular Biology* **2013**, *425* (22), 4334-4352.
(36) Lieber, M. R. Pol X DNA polymerases contribute to NHEJ flexibility. *Nat Struct Mol Biol* **2023**, *30* (1), 5-8. DOI: 10.1038/s41594-022-00904-6.
(37) Nick McElhinny, S. A.; Ramsden, D. A. Polymerase mu is a DNA-directed DNA/RNA polymerase. *Mol Cell Biol* **2003**, *23* (7), 2309-2315. DOI: 10.1128/mcb.23.7.2309-2315.2003.
(38) Ruiz, J. F.; Juárez, R.; García-Díaz, M.; Terrados, G.; Picher, A. J.; González-Barrera, S.; Fernández de Henestrosa, A. R.; Blanco, L. Lack of sugar discrimination by human Pol mu requires a single glycine residue. *Nucleic Acids Res* **2003**, *31* (15), 4441-4449. DOI: 10.1093/nar/gkg637.
(39) Moon, A. F.; Pryor, J. M.; Ramsden, D. A.; Kunkel, T. A.; Bebenek, K.; Pedersen, L. C. Structural accommodation of ribonucleotide incorporation by the DNA repair enzyme polymerase Mu. *Nucleic acids research* **2017**, *45* (15), 9138-9148.
(40) Cavanaugh, N. A.; Beard, W. A.; Wilson, S. H. DNA polymerase beta ribonucleotide discrimination: insertion, misinsertion, extension, and coding. *J Biol Chem* **2010**, *285* (32), 24457-24465. DOI: 10.1074/jbc.M110.132407.
(41) Consortium, U. UniProt: a worldwide hub of protein knowledge. *Nucleic Acids Res* **2019**, *47* (D1), D506-D515.
(42) Suzek, B. E.; Wang, Y.; Huang, H.; McGarvey, P. B.; Wu, C. H.; UniProt, C. UniRef clusters: a comprehensive and scalable alternative for improving sequence similarity searches. *Bioinformatics* **2015**, *31* (6), 926-932. DOI: 10.1093/bioinformatics/btu739.
(43) Thompson, J. D.; Gibson, T. J.; Higgins, D. G. Multiple sequence alignment using ClustalW and ClustalX. *Curr Protoc Bioinformatics* **2003**, (1), 2.3. 1-2.3. 22.
(44) Kumar, S.; Stecher, G.; Li, M.; Knyaz, C.; Tamura, K. MEGA X: molecular evolutionary genetics analysis across computing platforms. *Molecular biology and evolution* **2018**, *35* (6), 1547.
(45) Hasegawa, M.; Kishino, H.; Saitou, N. On the maximum likelihood method in molecular phylogenetics. *J Mol Evol* **1991**, *32* (5), 443-445. DOI: 10.1007/BF02101285.
(46) Law, Y.-S.; Utt, A.; Tan, Y. B.; Zheng, J.; Wang, S.; Chen, M. W.; Griffin, P. R.; Merits, A.; Luo, D. Structural insights into RNA recognition by the Chikungunya virus nsP2 helicase. *Proceedings of the National Academy of Sciences* **2019**, *116* (19), 9558-9567. DOI: 10.1073/pnas.1900656116.
(47) Loc'h, J.; Gerodimos, C. A.; Rosario, S.; Tekpinar, M.; Lieber, M. R.; Delarue, M. Structural evidence for an in trans base selection mechanism involving Loop1 in polymerase μ at an NHEJ double-strand break junction. *Journal of Biological Chemistry* **2019**, *294* (27), 10579-10595.
(48) Takahashi, M.; Takahashi, E.; Joudeh, L. I.; Marini, M.; Das, G.; Elshenawy, M. M.; Akal, A.; Sakashita, K.; Alam, I.; Tehseen, M. Dynamic structure mediates halophilic adaptation of a DNA polymerase from the deep-sea brines of the Red Sea. *The FASEB Journal* **2018**, *32* (6), 3346.
(49) Micsonai, A.; Wien, F.; Bulyaki, E.; Kun, J.; Moussong, E.; Lee, Y. H.; Goto, Y.; Refregiers, M.; Kardos, J. BeStSel: a web server for accurate protein secondary structure prediction and fold recognition from the circular dichroism spectra. *Nucleic Acids Res* **2018**, *46* (W1), W315-W322. DOI: 10.1093/nar/gky497.
(50) Jumper, J.; Evans, R.; Pritzel, A.; Green, T.; Figurnov, M.; Ronneberger, O.; Tunyasuvunakool, K.; Bates, R.; Žídek, A.; Potapenko, A.; et al. Highly accurate protein structure prediction with AlphaFold. *Nature* **2021**, *596* (7873), 583-589. DOI: 10.1038/s41586-021-03819-2.
(51) SYBYL, X. Tripos International, 1699 South Hanley Rd. *St. Louis, Missouri, 63144, USA*.
(52) Salomon‐Ferrer, R.; Case, D. A.; Walker, R. C. An overview of the Amber biomolecular simulation package. *Wiley Interdisciplinary Reviews: Computational Molecular Science* **2013**, *3* (2), 198-210.
(53) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of simple potential functions for simulating liquid water. *The Journal of chemical physics* **1983**, *79* (2), 926-935.
(54) Åqvist, J.; Wennerström, P.; Nervall, M.; Bjelic, S.; Brandsdal, B. O. Molecular dynamics simulations of water and biomolecules with a Monte Carlo constant pressure algorithm. *Chemical Physics Letters* **2004**, *384* (4), 288-294. DOI: https://doi.org/10.1016/j.cplett.2003.12.039.

(55) Hoover, W. G.; Ladd, A. J.; Moran, B. High-strain-rate plastic flow studied via nonequilibrium molecular dynamics. *Physical Review Letters* **1982**, *48* (26), 1818.
(56) Maier, J. A.; Martinez, C.; Kasavajhala, K.; Wickstrom, L.; Hauser, K. E.; Simmerling, C. ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *J Chem Theory Comput* **2015**, *11* (8), 3696-3713. DOI: 10.1021/acs.jctc.5b00255.
(57) Galindo-Murillo, R.; Robertson, J. C.; Zgarbova, M.; Sponer, J.; Otyepka, M.; Jurecka, P.; Cheatham III, T. E. Assessing the current state of amber force field modifications for DNA. *Journal of chemical theory and computation* **2016**, *12* (8), 4114-4127.
(58) Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. A smooth particle mesh Ewald method. *The Journal of chemical physics* **1995**, *103* (19), 8577-8593.
(59) Case, D. A.; Cheatham III, T. E.; Darden, T.; Gohlke, H.; Luo, R.; Merz Jr, K. M.; Onufriev, A.; Simmerling, C.; Wang, B.; Woods, R. J. The Amber biomolecular simulation programs. *Journal of computational chemistry* **2005**, *26* (16), 1668-1688.
(60) Jones, G.; Willett, P.; Glen, R. C.; Leach, A. R.; Taylor, R. Development and validation of a genetic algorithm for flexible docking11Edited by F. E. Cohen. *Journal of Molecular Biology* **1997**, *267* (3), 727-748. DOI: https://doi.org/10.1006/jmbi.1996.0897.
(61) Kuznetsova, A. A.; Tyugashev, T. E.; Alekseeva, I. V.; Timofeyeva, N. A.; Fedorova, O. S.; Kuznetsov, N. A. Insight into the mechanism of DNA synthesis by human terminal deoxynucleotidyltransferase. *Life science alliance* **2022**, *5* (12).