Terminal deoxynucleotidyltransferase: the story of an untemplated DNA polymerase also capable of DNA bridging and templated synthesis across strands

Jérôme Loc'h, Marc Delarue

Terminal deoxynucleotidyltransferase: the story of an untemplated DNA polymerase also capable of DNA bridging and templated synthesis across strands

Jérôme Loc'h and Marc Delarue*

Unit of Structural Dynamics of Biological Macromolecules and UMR 3528 du CNRS, Institut Pasteur, 75015 Paris, France.

* Corresponding author

marc.delarue@pasteur.fr

Summary

Terminal deoxynucleotidyltransferase (TdT) is a member of the polX family which is involved in DNA repair. It has been known for years as an untemplated DNA polymerase used during V(D)J recombination to generate diversity at the CDR3 region of immunoglobulins and T-cell receptors. Recently, however, TdT was crystallized in the presence of a complete DNA synapsis made of two double-stranded DNA (dsDNA), each with a 3' protruding end, and overlapping with only one micro-homology base-pair, thus giving structural insight for the first time into DNA synthesis across strands. It was subsequently shown that TdT indeed has an *in trans* template-dependent activity in the presence of an excess of the downstream DNA duplex. A possible biological role of this dual activity is discussed.

Bullet-points

-TdT has been known for more than 50 years as an untemplated DNA polymerase

-However, it was recently shown to be able to bind to a DNA synapsis in solution

-It was subsequently shown that TdT has *in-trans* templated polymerase activity

-New crystal structures show TdT caught in the act of DNA synthesis across strands

-This dual activity of TdT can be understood in the context of its biological role.

Introduction

Terminal deoxynucleotidyltransferase (TdT) was one of the the first eukaryotic DNA polymerases purified in the early 1960s [1], from calf thymus extracts. However, instead of the expected classical *templated* polymerase activity, the biochemical characterization of TdT revealed an efficient *untemplated* polymerase (nucleotidyltransferase) activity [2,3], especially in the presence of divalent transition metal ions [4]. *In vivo*, *the* function of TdT was only fully understood in the eighties [5–7], after the discovery of V(D)J recombination [8–10]. During this process, TdT adds random nucleotides (N-segments) at the V-D and D-J junctions in heavy chains of immunoglobulins (Ig) and T-cell receptors, thereby contributing significantly to the diversity of the immune repertoire [11,12]. Subsequently, it was revealed that the V(D)J uses the same machinery [13,14] as the one of Non-Homologous End Joining (NHEJ) that repairs DNA double-strand breaks (DSB). This machinery includes a recognition complex (Ku heterodimer, DNA-PKcs), DNA end-processing enzymes such as a nuclease (Artemis or Metnase) and a DNA polymerase (pol X), as well as a ligation complex (Lig IV, XRCC4, XLF) [15]. The DNA polymerase is a member of the family polX that includes not only TdT, but also pol λ [16] and pol μ [17,18], the last two participating to both NHEJ and V(D)J recombination [19–23]. All three polymerase domains X-ray structures have been determined to high resolution [24–26] but the only one that was crystallized in a DNA-bridging context is TdT [27,28]. Here we focus on TdT and on the biological implications of these new structures.

Structural and biochemical features of pol X family polymerases

The X-family DNA polymerase (polX) is specialized in DNA repair. This family is composed of four different DNA polymerases: pol β, pol λ, pol μ and TdT. Only

three members of the polX family possess an N-terminal BRCT (BRCA1 carboxy-terminal) domain (**Figure 1A**) that is essential for NHEJ activity both *in vitro* and *in vivo* [29,30]. Pol β participates only in base excision repair (BER) [31,32] and is devoid of this small domain (around 11 kDa), which mediates protein-protein or protein-DNA interactions [33]. Three individual structures of a BRCT domain from the polX family were solved by NMR [34,35]. BRCT domain is typically an αβα sandwich made up of a central antiparallel β-sheet flanked by three α-helices [36–38]. Although the structures of all known BRCT domains are highly conserved, their interaction mode with ligands is greatly variable and their role in the sequential recruitment of the different proteins during NHEJ repair is not completely understood [39]. Ligase IV contains 2 BRCT domains whose peptide junction interacts with XLF, XRCC4 [40,41] and which interact with Ku heterodimer [42]. The BRCT domain of pol μ binds DNA [43] but the precise mode of binding of the BRCT domain of either pol μ or TdT to Ku heterodimer remains to be elucidated, as well as its orientation with respect to the catalytic polymerase domain.

Sequence analysis of the polymerase domain of polX polymerases allows to divide this family into two sub-groups [44]. Pol β is closer to pol λ (34% of pairwise sequence identity) and TdT is closer to pol μ (44% of pairwise sequence identity). The structures of all four polymerases belonging to X-family have been solved by X-ray crystallography. Pol β was the first polX to be solved, alone and with different binary and ternary complexes [45–47]. The overall structure of the catalytic domain shares the same general architecture (but not the topology) of all DNA polymerases, namely a finger domain, a palm domain and a thumb domain (**Figure 1**). The additional 8-kDa domain in pol β and pol λ contains a deoxyribose phosphate (dRP) lyase activity required in base excision repair (BER) of oxidative DNA damage

(**Figure 1**). The amino acids necessary for dRP lyase activity are not conserved in TdT and pol μ, which do not participate in BER.

One striking feature common to all polX is the high degree of conservation of the catalytic site, with three strictly conserved Aspartates (**Figure 1B**) that coordinate two essential metal ions, involved in the so-called two-metal ions mechanism, first described in [48] and later shown to be present also in DNA polymerases [47]. Metal A activates the 3'OH of the last nucleotide to allow the attack of the alpha phosphate while Metal B that comes in with the incoming nucleotide triphosphate stabilizes the leaving group (PPi) (**Figure 1B**). In TdT, the coordination geometry of the divalent metal ions was studied in atomic detail during a full catalytic site, including transition metal ions such as Mn^{++}, Co^{++} and Zn^{++} which are known to be more efficient than Mg^{++} for the nucleotidyltransferase activity. It was concluded that Metal A has to leave and be replaced by Na^{+} in order to allow translocation of the newly extended primer strand into a catalytically competent position for a new addition [49]. A movie of the reaction cycle, based on thirteen different structures was built [49]. This scenario was also described in even greater details for pol β where it was also found that binding of Na^{+} in Metal A binding site, after nucleotide incorporation, is a key step for DNA translocation [50]. In addition, time-lapse crystallography showed that in pol β, there is an additional divalent ion (a third $Mg^{++,}$ Metal C) that comes during the reaction to counter-balance the apparition of a charge on the beta phosphate, and then leaves before the end-state is reached [50,51]. This situation was also observed for pol μ, in the case of Mn^{++} ions [52].

Common features of TdT, pol μ and pol λ

TdT, pol μ and pol λ remain in a closed conformation throughout their catalytic cycle, contrary to pol β [24,53,54]. One possible explanation for this observation is that they have traded fidelity (which requires open-to-closed transition) for a very tight binding of the DNA synapsis, a very fragile structure. One unique feature revealed when the first x-ray structure of TdT was solved [24], is a specific Loop (Loop1), composed of 20 amino acids (382-401), located between the β3 and β4 strands (**Figure 1B**), that prevents the binding of a 5' overhang of the template strand. More than 30 structures of TdT (wild-type or mutants in different complexes) are available on PDB and in all of them, Loop1 adopts the same lariat-like conformation that prevents the binding of an *uninterrupted* template strand on TdT (**Figure 2**). Perhaps somewhat deceptively, in all known structures of pol μ, Loop1 (also about 20 amino acids long) is invisible in the electron density map, meaning that this region is disordered (**Figure 2 and 3A**). In pol λ, Loop1 is comparatively shorter (8 aa), but still longer than in pol β (**Figure 3**). However, pol λ has an additional Loop, called Loop3, that, interestingly, is located precisely where another form of TdT resulting from alternative splicing has an insertion of 20 additional residues [55,56] (**Figure 3B**) and where it is ideally placed to control bulges or insertions just before the *in-trans* templating base [28] (**Figure 3A**). Mutations experiments have consistently shown the importance of Loop1 for the substrate specificity not only in TdT [27,57] but also in pol μ [58–60], as well as pol λ [61].

Early sequence comparisons in a structural context helped to define two important regions for the specificity of TdT vs pol μ [24], later named SD1 and SD2 [57] (**Figure 3**): they are located at the C-terminal border of Loop1, and in a β-turn-β structure close to Loop1, which can also bind an extra Zn^{++} ion [49] but the precise role of this additional divalent ion is currently unknown. Mutations in these two

regions profoundly affect the activity of both TdT [27,28] and pol μ [60]. Mutation of only one amino acid in SD1 region (F401A) confers to Tdt an *in cis* templated polymerase activity, even in the presence of Co^{++} [57].

A remaining puzzle concerns the conformation of Loop1 in pol μ, and its role in binding the DNA synapsis substrate. Experiments are currently underway in our lab using a TdT-pol μ chimera to determine the conformation of Loop1 of pol μ and they indeed suggest that Loop1 plays the crucial role of a gate that can be open or closed (disordered or ordered) when pol μ searches for a micro-homology region across a DNA synapsis (Loc'h et al., submitted).

Nucleotidyltransferase activity of TdT

Extensive biochemical experiments have demonstrated that TdT can add random deoxyribonucleotides (dNTPs) on a ssDNA primer, which has to be at least three nucleotides long, in a template-independent manner [2]. *In vitro* experiments show that TdT can use all four natural dNTPs with a preferential incorporation of dCTP and dGTP compared to dATP and dTTP [3]. Pol μ also has a significant nucleotidyltransferase activity in the presence of Mn^{++} [57,62,63]. Interestingly, TdT nucleotide binding site can accommodate both deoxyribo-nucleotide and ribo-nucleotide triphosphates (dNTPs and rNTPs). TdT shares this property with pol μ, but not with pol λ and pol β (**Figure 1A**). Indeed, the presence of a YF sequence motif, the so-called steric-gate at the vicinity of the 2' OH of the nucleotide, prevents the binding of rNTPs in pol λ and pol β, whereas a GW sequence motif in TdT and pol μ (**Figure 1B**) increases the size of the nucleotide binding pocket, allowing binding of both dNTPs and rNTPs [64,65]. However, addition of rNTPs by TdT stops after few incorporations on ssDNA [66,67]. This can be interpreted by noting that the path of

the primer is constrained in a B-DNA form by the protein, especially through a Na^+ ion coordinated by the HhH2 motif (**Figure 1B**) at the level of the penultimate phosphate of the primer, and this B-DNA form is not suitable for an RNA backbone.
TdT can also incorporate efficiently various un-natural bases [68,69]. An interesting consequence of the large tolerance of TdT on the incoming nucleotide is to use it for making polymers of un-natural DNA, using nucleotides modified either in the sugar moiety or the base moiety [70] or for click chemistry [71]. Also a recent application for FISH experiments and the design of RNA capture probes can be found in [72].

In trans-templated polymerase activity across strands breaks

In 2015-2016, half a century after its first biochemical characterization as a template-independent polymerase, it was shown that TdT can i) assemble a DNA synapsis by itself, optimally with one micro-homology base-pair between strands [27] and ii) perform a template-dependent nucleotide incorporation across strands breaks [28] in the presence of an excess of downstream dsDNA with a 3' protruding end. Because this template-dependent activity of TdT is achieved by using an *in trans* template strand, instead of the usual *in cis* template strand, we refer to it as the *in trans* activity. Interestingly, this *in trans* templated activity was also described for pol μ [60,73], but without the need of an excess of the downstream DNA duplex [28]. *In vitro* biochemical experiments on *chimeric* constructs of TdT, involving substitution of Loop1 by pol μ's sequence and/or reconstitution of the 5'-phosphate binding site (**Figure 1**), show an activity similar to pol μ, with a protein/DNA ratio of 1:1 [28]. The existence of templated synthesis across strand breaks has recently been described *in vivo* for both pol μ and pol λ [74–77].

The structure of TdT in complex with a DSB-DNA substrate [28] is the first of its kind to be solved for a polX (**Figure 2**)**.** It looks as if TdT was designed to "isolate" a mini-helix made of only two base pairs to stabilize and establish a fragile bridge between the upstream and downstream duplexes (Figure 1). The two upstream and downstream dsDNA are in B-DNA conformation, while the micro-homology (MH) mini-helix between them is in A-DNA conformation. L398 in Loop1 is crucial to break the helical path from the upstream dsDNA to the MH-mini helix and its role has been verified by site-directed mutagenesis [27].

Function of polX during V(D)J recombination

Expression of TdT is only observed in the primary lymphoid organs, thymus and bone marrow where V(D)J recombination is active [78]. Indeed, expression of TdT is only detected during heavy chain rearrangements, but is absent from the next step where light chain rearrangements occurs [79] (**Figure 4**). The expression of TdT is also tightly regulated in time as it is not expressed in fetal or neo-natal life. One way this regulation is done is probably through ubiquitylation [80]*.* Interestingly, two TdT interacting factors (TdIF1 and TdIF2) have been identified and characterized to inhibit Tdt activity [81,82]. Pol μ participates in light chain rearrangements during V(D)J recombination, whereas pol λ participates only in heavy chain rearrangements [83] (**Figure 4**). It should be noted that when TdT is made to express in non-lymphoid cells, it participates in NHEJ DNA repair [84]. Also, when expressed constitutively in B-cells, it generates N-regions in both heavy and light chains [85].

It was found that inhibiting TdT in some cancer cells can kill them, such as in acute lymphoblastic leukemia cells [86]. Chemical compounds were designed and

synthesized using TdT as a drug target [87,88] and the structure of some of these compounds was solved in their bound form [89].

The length of the N-segments incorporated by TdT ranges from 2 to about 15-20, with two clearly different regimes in its probability distribution function: a rising phase with a peak at 4, followed by a decreasing phase [90,91]. Loc'h and colleagues proposed that the dual activity of TdT may correspond to these two regimes [28]: after the addition of a few random nucleotides on the 3' end DNA, the downstream DNA is finally reached/sensed, at which point TdT switches to an *in-trans* template-dependent synthesis (**Figure 4**). This synthesis will stop if the micro-homology base pair is of Watson-Crick type and continue otherwise, which occurs in three out of four possible cases; strikingly, this quantitatively explains the size-distribution law of the N-regions of the second phase, that exponentially decreases with a slope of -¼ [91].

After incorporation of random nucleotides by TdT during heavy-chain rearrangements, both TdT and pol λ may perform *in-trans* polymerase activity (in unknown proportions), whereas synthesis of the complementary strand can only be achieved by pol λ using its gap-filling activity, which TdT lacks because of its Loop1 (**Figure 4**). In light-chain gene rearrangements Pol μ can perform not only template-independent, but also *in-trans* polymerase and gap-filling activities (**Figure 4**).

Perspectives and Conclusion

Due to its ability to add random sequences to a DNA primer, TdT is an intrinsically "unpredictable" polymerase (rather than a "misguided" one [12]). This explains its tight regulation both in time and space, in order to restrict its use to V(D)J recombination. Here we suggest that not only its recruitment but also its gradual stop

is programmed, by switching to a previously unsuspected templating mode across strand breaks after the addition of 4-5 random nucleotides. This would be due to the spatial constraints of the architecture of the whole NHEJ apparatus, a very active field in structural biology that achieved impressive progress recently, first for the structure of the ligation complex [92], and more recently for the structure of the huge loading complex of NHEJ [93–95]. It is known that TdT interacts with Ku heterodimer through its BRCT domain, as does pol μ [13,14]. If the interaction of the BRCT domain with Ku heterodimer could be mapped, then it would be possible to place the polX with respect to the DNA-PKcs-DNA complex and thereby to shed light on spatial constraints at work. On the evolutionary level it would be interesting to do it for both pol mu and TdT, so as to assess how similar are the positioning of TdT and pol μ in this integrated view of the NHEJ complex.

From Figure 1, it appears that TdT has lost just the 5' phosphate binding site of pol mu and that its Loop1 is of the same length, but with a different sequence. Regarding Loop1 and its vexing property of escaping structural characterisation in pol mu, we expect that the structure of the TdT chimera containing Loop1 of pol mu will inform us on its conformation and also allow, eventually, a comparison with the LigD polymerase that performs NHEJ in bacteria.

Acknowledgements

JL was supported by an ARC post-doctoral fellowship.

References

1. Bollum FJ: **Calf thymus polymerase**. *J Biol Chem* 1960, **235**:2399–2403.

2. Kato KI, Gonçalves JM, Houts GE, Bollum FJ: **Deoxynucleotide-polymerizing enzymes of calf thymus gland. II. Properties of the terminal deoxynucleotidyltransferase**. *J Biol Chem* 1967, **242**:2780–2789.

3. Chang LM, Bollum FJ: **Molecular biology of terminal transferase**. *CRC Crit Rev Biochem* 1986, **21**:27–52.

4. Chang LM, Bollum FJ: **Multiple roles of divalent cation in the terminal deoxynucleotidyltransferase reaction**. *J Biol Chem* 1990, **265**:17436–17440.

5. Alt FW, Baltimore D: **Joining of immunoglobulin heavy chain gene segments: implications from a chromosome with evidence of three D-JH fusions**. *Proc Natl Acad Sci USA* 1982, **79**:4118–4122.

6. Landau NR, Schatz DG, Rosa M, Baltimore D: **Increased frequency of N-region insertion in a murine pre-B-cell line infected with a terminal deoxynucleotidyl transferase retroviral expression vector**. *Mol Cell Biol* 1987, **7**:3237–3243.

7. Gilfillan S, Dierich A, Lemeur M, Benoist C, Mathis D: **Mice lacking TdT: mature animals with an immature lymphocyte repertoire**. *Science* 1993, **261**:1175–1178.

8. Tonegawa S, Steinberg C, Dube S, Bernardini A: **Evidence for Somatic Generation of Antibody Diversity**. *Proc Natl Acad Sci U S A* 1974, **71**:4027–4031.

9. Brack C, Hirama M, Lenhard-Schuller R, Tonegawa S: **A complete immunoglobulin gene is created by somatic recombination**. *Cell* 1978, **15**:1–14.

10. Saito H, Kranz DM, Takagaki Y, Hayday AC, Eisen HN, Tonegawa S: **A third rearranged and expressed gene in a clone of cytotoxic T lymphocytes**. *Nature* 1984, **312**:36–40.

11. Fowler JD, Suo Z: **Biochemical, Structural, and Physiological Characterization of Terminal Deoxynucleotidyl Transferase**. *Chemical Reviews* 2006, **106**:2092–2110.

12. Motea EA, Berdis AJ: **Terminal deoxynucleotidyl transferase: the story of a misguided DNA polymerase**. *Biochim Biophys Acta* 2010, **1804**:1151–1166.

13. Mahajan KN, Gangi-Peterson L, Sorscher DH, Wang J, Gathy KN, Mahajan NP, Reeves WH, Mitchell BS: **Association of terminal deoxynucleotidyl transferase with Ku**. *Proc Natl Acad Sci USA* 1999, **96**:13926–13931.

14. Mahajan KN, McElhinny SAN, Mitchell BS, Ramsden DA: **Association of DNA Polymerase μ (pol μ) with Ku and Ligase IV: Role for pol μ in End-Joining Double-Strand Break Repair**. *Mol Cell Biol* 2002, **22**:5194–5202.

15. Lieber MR: **The mechanism of double-strand DNA break repair by the**

nonhomologous DNA end-joining pathway. *Annu Rev Biochem* 2010, **79**:181–211.

16. * García-Díaz M, Domínguez O, López-Fernández LA, de Lera LT, Saníger ML, Ruiz JF, Párraga M, García-Ortiz MJ, Kirchhoff T, del Mazo J, et al.: **DNA polymerase lambda (Pol lambda), a novel eukaryotic DNA polymerase with a potential role in meiosis**. *J Mol Biol* 2000, **301**:851–867.
Discovery of pol λ and description of its specific functional properties.

17. * Aoufouchi S, Flatter E, Dahan A, Faili A, Bertocci B, Storck S, Delbos F, Cocea L, Gupta N, Weill JC, et al.: **Two novel human and mouse DNA polymerases of the polX family**. *Nucleic Acids Res* 2000, **28**:3684–3693.
First of two papers that describe the discovery of pol μ and the characterisation of its functional properties.

18. * Domínguez O, Ruiz JF, Laín de Lera T, García-Díaz M, González MA, Kirchhoff T, Martínez-A C, Bernad A, Blanco L: **DNA polymerase mu (Pol mu), homologous to TdT, could act as a DNA mutator in eukaryotic cells**. *EMBO J* 2000, **19**:1731–1742.
Second (independent) paper on the discovery of pol μ and description of its functional properties.

19. Nick McElhinny SA, Ramsden DA: **Sibling rivalry: competition between Pol X family members in V(D)J recombination and general double strand break repair**. *Immunol Rev* 2004, **200**:156–164.

20. Moon AF, Garcia-Diaz M, Batra VK, Beard WA, Bebenek K, Kunkel TA, Wilson SH, Pedersen LC: **The X family portrait: structural insights into biological functions of X family polymerases**. *DNA Repair (Amst)* 2007, **6**:1709–1725.

21. Yamtich J, Sweasy JB: **DNA polymerase Family X: Function, structure, and cellular roles**. *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics* 2010, **1804**:1136–1150.

22. Ramsden DA: **Polymerases in Nonhomologous End Joining: Building a Bridge over Broken Chromosomes**. *Antioxid Redox Signal* 2011, **14**:2509–2519.

23. Waters CA, Strande NT, Wyatt DW, Pryor JM, Ramsden DA: **Nonhomologous end joining: a good solution for bad ends**. *DNA Repair (Amst)* 2014, **17**:39–51.

24. * Delarue M, Boulé JB, Lescar J, Expert-Bezançon N, Jourdan N, Sukumar N, Rougeon F, Papanicolaou C: **Crystal structures of a template-independent DNA polymerase: murine terminal deoxynucleotidyltransferase**. *EMBO J* 2002, **21**:427–439.
X-ray structure of mouse TdT alone or in different binary complexes, with either the incoming dNTP or the single-stranded primer DNA strand. The catalytic metal binding sites are similar to what is seen in pol beta and the two-metal ion mechanism is probably conserved. All structures are in a closed state, suggesting there is no open/closed transition during the catalytic cycle. The critical conformation of Loop1 is described as well as its role in the template-independent activity of TdT by physically excluding the 5' end of the template strand. Stretches of sequences that maximally distinguish pol mu and Tdt in a multialignment are highlighted.

25. * Garcia-Diaz M, Bebenek K, Krahn JM, Blanco L, Kunkel TA, Pedersen LC: **A structural solution for the DNA polymerase lambda-dependent repair of DNA gaps with minimal homology**. *Mol Cell* 2004, **13**:561–572.
First X-ray structure of pol λ bound to a gapped template-primer duplex. The polymerase is in a closed conformation even in the absence of dNTP binding.

26. * Moon AF, Garcia-Diaz M, Bebenek K, Davis BJ, Zhong X, Ramsden DA, Kunkel TA, Pedersen LC: **Structural insight into the substrate specificity of DNA Polymerase μ**. *Nature Structural & Molecular Biology* 2007, **14**:45–53.
First X-ray structure of pol μ with a gap-filling DNA substrate and a correct incoming nucleotide.

27. ** Gouge J, Rosario S, Romain F, Poitevin F, Béguin P, Delarue M: **Structural basis for a novel mechanism of DNA bridging and alignment in eukaryotic DSB DNA repair**. *EMBO J* 2015, **34**:1126–1142.
First X-ray structure of DNA polX (TdT) with a DSB-DNA mimic substrate, supplemented by a demonstration that a TdT-primer complex is also able to bind a downstream DNA duplex in solution. The presence of a DNA mini-helix composed of one microhomology(MH)-base pair and the nascent base pair provides a rationale for the statistics of sequences incorporated by TdT.

28. ** 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**:1452–1463.
X-ray structure of TdT with a complete DSB DNA substrate. A new templated activity of TdT is demonstrated in vitro using an excess of the downstream duplex DNA, leading to in trans templated activity. The relevance of this dual TdT activity during V(D)J recombination is described.

29. * Ma Y, Lu H, Tippin B, Goodman MF, Shimazaki N, Koiwai O, Hsieh C-L, Schwarz K, Lieber MR: **A biochemically defined system for mammalian nonhomologous DNA end joining**. *Molecular cell* 2004, **16**:701–713.
First description of a reconstituted NHEJ machinery that is biochemically active in vitro.

30. Nick McElhinny SA, Havener JM, Garcia-Diaz M, Juárez R, Bebenek K, Kee BL, Blanco L, Kunkel TA, Ramsden DA: **A gradient of template dependence defines distinct biological roles for family X polymerases in nonhomologous end joining**. *Mol Cell* 2005, **19**:357–366.

31. Matsumoto Y, Kim K: **Excision of deoxyribose phosphate residues by DNA polymerase beta during DNA repair**. *Science* 1995, **269**:699–702.

32. Singhal RK, Prasad R, Wilson SH: **DNA polymerase beta conducts the gap-filling step in uracil-initiated base excision repair in a bovine testis nuclear extract**. *J Biol Chem* 1995, **270**:949–957.

33. Gerloff DL, Woods NT, Farago AA, Monteiro ANA: **BRCT domains: A little more than kin, and less than kind**. *FEBS Lett* 2012, **586**:2711–2716.

34. DeRose EF, Clarkson MW, Gilmore SA, Galban CJ, Tripathy A, Havener JM,

Mueller GA, Ramsden DA, London RE, Lee AL: **Solution Structure of Polymerase μ's BRCT Domain Reveals an Element Essential for Its Role in Nonhomologous End Joining**. *Biochemistry* 2007, **46**:12100–12110.

35. Mueller GA, Moon AF, DeRose EF, Havener JM, Ramsden DA, Pedersen LC, London RE: **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**:1340–1351.

36. Zhang X, Moréra S, Bates PA, Whitehead PC, Coffer AI, Hainbucher K, Nash RA, Sternberg MJ, Lindahl T, Freemont PS: **Structure of an XRCC1 BRCT domain: a new protein-protein interaction module**. *EMBO J* 1998, **17**:6404–6411.

37. Derbyshire DJ, Basu BP, Serpell LC, Joo WS, Date T, Iwabuchi K, Doherty AJ: **Crystal structure of human 53BP1 BRCT domains bound to p53 tumour suppressor**. *EMBO J* 2002, **21**:3863–3872.

38. Clapperton JA, Manke IA, Lowery DM, Ho T, Haire LF, Yaffe MB, Smerdon SJ: **Structure and mechanism of BRCA1 BRCT domain recognition of phosphorylated BACH1 with implications for cancer**. *Nat Struct Mol Biol* 2004, **11**:512–518.

39. Leung CCY, Glover JNM: **BRCT domains: easy as one, two, three**. *Cell Cycle* 2011, **10**:2461–2470.

40. Sibanda BL, Critchlow SE, Begun J, Pei XY, Jackson SP, Blundell TL, Pellegrini L: **Crystal structure of an Xrcc4-DNA ligase IV complex**. *Nat Struct Biol* 2001, **8**:1015–1019.

41. Hammel M, Yu Y, Fang S, Lees-Miller SP, Tainer JA: **XLF regulates filament architecture of the XRCC4·Ligase IV complex**. *Structure* 2010, **18**:1431–1442.

42. Costantini S, Woodbine L, Andreoli L, Jeggo PA, Vindigni A: **Interaction of the Ku heterodimer with the DNA ligase IV/Xrcc4 complex and its regulation by DNA-PK**. *DNA Repair (Amst)* 2007, **6**:712–722.

43. Matsumoto T, Go K, Hyodo M, Koiwai K, Maezawa S, Hayano T, Suzuki M, Koiwai O: **BRCT domain of DNA polymerase μ has DNA-binding activity and promotes the DNA polymerization activity**. *Genes to Cells* 2012, **17**:790–806.

44. * Bienstock RJ, Beard WA, Wilson SH: **Phylogenetic analysis and evolutionary origins of DNA polymerase X-family members**. *DNA Repair (Amst)* 2014, **22**:77–88.
An up-to-date paper on phylogeny and sequence multialignment of members of the polX family.

45. * Sawaya MR, Pelletier H, Kumar A, Wilson SH, Kraut J: **Crystal structure of rat DNA polymerase beta: evidence for a common polymerase mechanism**. *Science* 1994, **264**:1930–1935.
First of two papers describing the crystal structure of pol β in the apo form.

46. * Davies JF, Almassy RJ, Hostomska Z, Ferre RA, Hostomsky Z: **2.3 A crystal structure of the catalytic domain of DNA polymerase beta**. *Cell* 1994, **76**:1123–1133.

Second (independent) paper describing the crystal structure of pol β in the apo form.

47. ** Sawaya MR, Prasad R, Wilson SH, Kraut J, Pelletier H: **Crystal structures of human DNA polymerase beta complexed with gapped and nicked DNA: evidence for an induced fit mechanism**. *Biochemistry* 1997, **36**:11205–11215.
First structure of a ternary complex of a polX with a gapped-template DNA substrate and the incoming nucleotide, which sets up the stage for all polX polymerases and describes for the first time in atomic detail the coordination of the two essential catalytic ions through strictly conserved Aspartates. An open-to-closed transition is described. A landmark paper for polX.

48. Steitz TA, Steitz JA: **A general two-metal-ion mechanism for catalytic RNA**. *Proc Natl Acad Sci USA* 1993, **90**:6498–6502.

49. * 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**. *J Mol Biol* 2013, **425**:4334–4352.
Binary and ternary structures of TdT in precatalytic, post-catalytic and competent states. It is demonstrated that some of the transitions between these states can occur in the crystal. The necessity of metal A departure for DNA translocation and new nucleotide binding is deduced and a movie of the catalytic cycle is built from 13 different structures. A change of the coordination of $Cobalt^{++}$ from tetrahedral to octahedral during the reaction is described. The presence of an additional Zn^{++} near Loop1 is also demonstrated.

50. ** Freudenthal BD, Beard WA, Shock DD, Wilson SH: **Observing a DNA polymerase choose right from wrong**. *Cell* 2013, **154**:157–168.
Time-lapse crystallography of pol beta demonstrates the transient presence of a third Mg^{++} ion during the catalytic cycle.

51. * Freudenthal BD, Beard WA, Wilson SH: **Watching a DNA polymerase in action**. *Cell Cycle* 2014, **13**:691–692.
Comment on [50].

52. * Jamsen JA, Beard WA, Pedersen LC, Shock DD, Moon AF, Krahn JM, Bebenek K, Kunkel TA, Wilson SH: **Time-lapse crystallography snapshots of a double-strand break repair polymerase in action**. *Nature Communications* 2017, **8**:253.
Time-lapse crystallography with pol μ using a gap-filling DNA substrate in the presence of magnesium or manganese as catalytic ions. A third metal ion that binds the product state is observed, but only in the case of manganese.

53. Garcia-Diaz M, Bebenek K, Krahn JM, Kunkel TA, Pedersen LC: **A closed conformation for the Pol lambda catalytic cycle**. *Nat Struct Mol Biol* 2005, **12**:97–98.

54. * Moon AF, Pryor JM, Ramsden DA, Kunkel TA, Bebenek K, Pedersen LC: **Sustained active site rigidity during synthesis by human DNA polymerase μ**. *Nat Struct Mol Biol* 2014, **21**:253–260.
Structure of the apo form of pol μ, with little change compared to the gap-filling complex.

55. Doyen N, d'Andon MF, Bentolila LA, Nguyen QT, Rougeon F: **Differential splicing in mouse thymus generates two forms of terminal deoxynucleotidyl transferase**. *Nucleic Acids Res* 1993, **21**:1187–1191.

56. Bentolila LA, Fanton d'Andon M, Nguyen QT, Martinez O, Rougeon F, Doyen N: **The two isoforms of mouse terminal deoxynucleotidyl transferase differ in both the ability to add N regions and subcellular localization**. *EMBO J* 1995, **14**:4221–4229.

57. * 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**:4642–4656.
Attempts to destabilize Loop1 by site-directed mutagenesis or deletions are not enough to convert TdT to a template-dependent polymerase. But a Tdt chimera that contains Loop1 of pol μ has an activity comparable to pol μ, as tested for *in cis*-templated polymerase activity. SD1 and SD2 regions are formally defined and mutations in these two regions are systematically explored. The F401 mutation (SD1) converts TdT into an *in-cis* templated polymerase.

58. * Juárez R, Ruiz JF, Nick McElhinny SA, Ramsden D, Blanco L: **A specific loop in human DNA polymerase mu allows switching between creative and DNA-instructed synthesis**. *Nucleic Acids Res* 2006, **34**:4572–4582.
Loop1 in pol μ strongly influences the balance between its template-dependent versus template-independent capacities. Pol μ chimera (with Loop1 of TdT) mimicks the behavior and substrate preferences of TdT. In the discussion part, the authors suggest that Loop1 in pol μ could adopt different conformations for different substrates.

59. Esteban V, Martin MJ, Blanco L: **The BRCT domain and the specific loop 1 of human Polμ are targets of Cdk2/cyclin A phosphorylation**. *DNA Repair (Amst)* 2013, **12**:824–834.

60. * Martin MJ, Blanco L: **Decision-making during NHEJ: a network of interactions in human Polμ implicated in substrate recognition and end-bridging**. *Nucleic Acids Res* 2014, **42**:7923–7934.
Site-directed Mutagenesis experiments on Loop1 in pol μ as well as in SD1 and SD2 regions. The authors propose a « Loop1 interaction network » that regulates the balance between terminal transferase and NHEJ activities, which awaits a crystal structure characterization.

61. * Bebenek K, Garcia-Diaz M, Zhou R-Z, Povirk LF, Kunkel TA: **Loop 1 modulates the fidelity of DNA polymerase lambda**. *Nucleic Acids Res* 2010, **38**:5419–5431.
Loop1 of pol lambda (9 residues) is replaced by Loop1 of pol beta (4 residues).

62. * Andrade P, Martín MJ, Juárez R, López de Saro F, Blanco L: **Limited terminal transferase in human DNA polymerase mu defines the required balance between accuracy and efficiency in NHEJ**. *Proc Natl Acad Sci USA* 2009, **106**:16203–16208.
Demonstration that R387 in pol μ (K403 in TdT) regulates the rate of TdT-like activity in pol μ, while H329 (H342 in TdT) is essential for pol μ activity. The limited TdT-like activity in pol μ could be used for DNA bridging in the presence of an *in trans* tremplating base.

63. * Gu J, Lu H, Tippin B, Shimazaki N, Goodman MF, Lieber MR: **XRCC4:DNA ligase IV can ligate incompatible DNA ends and can ligate across gaps**. *EMBO J* 2007, **26**:1010–1023.
In addition to the XRCC4-ligase IV functional studies, it is shown that Pol μ can add random nucleotides under physiological conditions such that the subset of generated ends that gain

some terminal microhomology can then be ligated. The importance of creating a bridge between the usptream and the downstream DNA is highlighted.

64. Ruiz JF, Juárez R, García-Díaz M, Terrados G, Picher AJ, González-Barrera S, Fernández de Henestrosa AR, Blanco L: **Lack of sugar discrimination by human Pol μ requires a single glycine residue**. *Nucleic Acids Res* 2003, **31**:4441–4449.

65. Moon AF, Pryor JM, Ramsden DA, Kunkel TA, Bebenek K, Pedersen LC: **Structural accommodation of ribonucleotide incorporation by the DNA repair enzyme polymerase Mu**. *Nucleic Acids Res* 2017, **45**:9138–9148.

66. Roychoudhury R: **OLIGODEOXYNUCLEOTIDES WITH ONE 3'-TERMINAL RIBONUCLEOTIDE AS PRIMERS FOR POLYDEOXYNUCLEOTIDE SYNTHESIS**. *Journal of Biological Chemistry* 1972, **247**:3910–3917.

67. Boulé JB, Rougeon F, Papanicolaou C: **Terminal deoxynucleotidyl transferase indiscriminately incorporates ribonucleotides and deoxyribonucleotides**. *J Biol Chem* 2001, **276**:31388–31393.

68. Motea EA, Lee I, Berdis AJ: **A Non-natural Nucleoside with Combined Therapeutic and Diagnostic Activities against Leukemia**. *ACS Chemical Biology* 2012, **7**:988–998.

69. Jarchow-Choy SK, Krueger AT, Liu H, Gao J, Kool ET: **Fluorescent xDNA nucleotides as efficient substrates for a template-independent polymerase**. *Nucleic Acids Res* 2011, **39**:1586–1594.

70. Tauraitė D, Jakubovska J, Dabužinskaitė J, Bratchikov M, Meškys R: **Modified Nucleotides as Substrates of Terminal Deoxynucleotidyl Transferase**. *Molecules* 2017, **22**.

71. Winz M-L, Linder EC, André T, Becker J, Jäschke A: **Nucleotidyl transferase assisted DNA labeling with different click chemistries**. *Nucleic Acids Res* 2015, **43**:e110.

72. Gaspar I, Wippich F, Ephrussi A: **Enzymatic production of single-molecule FISH and RNA capture probes**. *RNA* 2017, **23**:1582–1591.

73. Martin MJ, Juarez R, Blanco L: **DNA-binding determinants promoting NHEJ by human Polμ**. *Nucleic Acids Res* 2012, **40**:11389–11403.

74. Martin MJ, Garcia-Ortiz MV, Gomez-Bedoya A, Esteban V, Guerra S, Blanco L: **A specific N-terminal extension of the 8 kDa domain is required for DNA end-bridging by human Polμ and Polλ**. *Nucleic Acids Res* 2013, **41**:9105–9116.

75. * Martin MJ, Garcia-Ortiz MV, Esteban V, Blanco L: **Ribonucleotides and manganese ions improve non-homologous end joining by human Polμ**. *Nucleic Acids Research* 2013, **41**:2428–2436.
No steric gate for nucleotide selection in pol μ (G433). Physiological concentration of manganese ions increases Pol μ efficiency at no fidelity cost during NHEJ. Moreover, insertion of ribonucleotides is achieved with high fidelity by Pol μ during NHEJ.

76. * Pryor JM, Waters CA, Aza A, Asagoshi K, Strom C, Mieczkowski PA, Blanco L, Ramsden DA: **Essential role for polymerase specialization in cellular nonhomologous end joining**. *Proc Natl Acad Sci USA* 2015, **112**:E4537-4545.
Mouse Embryonic Fibroblast cells are used to understand the substrate preference between pol μ and pol λ. The authors identified contexts where only Pol μ or Pol λ are active and clarify the mechanistic basis for their specific activities. A model is proposed to explain the mechanism by which specialized polymerases contribute to NHEJ**.**

77. * Davis BJ, Havener JM, Ramsden DA: **End-bridging is required for pol μ to efficiently promote repair of noncomplementary ends by nonhomologous end joining**. *Nucleic Acids Research* 2008, **36**:3085–3094.
Early demonstration of the importance of end-bridging interactions in pol μ's ability to promote NHEJ. In the discussion, the authors suggest that DNA-bridging through gap recognition may be a feature common to NHEJ-associated polymerases, including in bacteria.

78. Bollum FJ: **Terminal deoxynucleotidyl transferase as a hematopoietic cell marker**. *Blood* 1979, **54**:1203–1215.

79. Li Y-S, Hayakawa K, Hardy RR: **The regulated expression of B lineage associated genes during B cell differentiation in bone marrow and fetal liver.** *Journal of Experimental Medicine* 1993, **178**:951–960.

80. Maezawa S, Fukushima R, Matsushita T, Kato T, Takagaki Y, Nishiyama Y, Ando S, Matsumoto T, Kouda K, Hayano T, et al.: **Ubiquitylation of terminal deoxynucleotidyltransferase inhibits its activity**. *PLoS ONE* 2012, **7**:e39511.

81. Kubota T, Maezawa S, Koiwai K, Hayano T, Koiwai O: **Identification of functional domains in TdIF1 and its inhibitory mechanism for TdT activity**. *Genes Cells* 2007, **12**:941–959.

82. Fujita K, Shimazaki N, Ohta Y, Kubota T, Ibe S, Toji S, Tamai K, Fujisaki S, Hayano T, Koiwai O: **Terminal deoxynucleotidyltransferase forms a ternary complex with a novel chromatin remodeling protein with 82 kDa and core histone**. *Genes Cells* 2003, **8**:559–571.

83. * Bertocci B, De Smet A, Weill J-C, Reynaud C-A: **Nonoverlapping functions of DNA polymerases mu, lambda, and terminal deoxynucleotidyltransferase during immunoglobulin V(D)J recombination in vivo**. *Immunity* 2006, **25**:31–41.
Demonstration that pol lambda is recruited during heavy chain rearrangement at a step that precedes the action of TdT. Previously, the same authors had demonstrated that pol mu is implicated in light chain rearrangements.

84. Boubakour-Azzouz I, Bertrand P, Claes A, Lopez BS, Rougeon F: **Terminal deoxynucleotidyl transferase requires KU80 and XRCC4 to promote N-addition at non-V(D)J chromosomal breaks in non-lymphoid cells**. *Nucleic Acids Research* 2012, **40**:8381–8391.

85. Bentolila LA, Wu GE, Nourrit F, Fanton d'Andon M, Rougeon F, Doyen N: **Constitutive expression of terminal deoxynucleotidyl transferase in transgenic mice is sufficient for N region diversity to occur at any Ig locus throughout B cell**

differentiation. *J Immunol* 1997, **158**:715–723.

86. McCaffrey R, Bell R, Lillquist A, Wright G, Baril E, Minowada J: **Selective killing of leukemia cells by inhibition of TdT**. *Haematol Blood Transfus* 1983, **28**:24–27.

87. Locatelli GA, Di Santo R, Crespan E, Costi R, Roux A, Hübscher U, Shevelev I, Blanca G, Villani G, Spadari S, et al.: **Diketo hexenoic acid derivatives are novel selective non-nucleoside inhibitors of mammalian terminal deoxynucleotidyl transferases, with potent cytotoxic effect against leukemic cells**. *Mol Pharmacol* 2005, **68**:538–550.

88. Di Santo R, Maga G: **Human terminal deoxynucleotidyl transferases as novel targets for anticancer chemotherapy**. *Curr Med Chem* 2006, **13**:2353–2368.

89. * Costi R, Crucitti GC, Pescatori L, Messore A, Scipione L, Tortorella S, Amoroso A, Crespan E, Campiglia P, Maresca B, et al.: **New nucleotide-competitive non-nucleoside inhibitors of terminal deoxynucleotidyl transferase: discovery, characterization, and crystal structure in complex with the target**. *J Med Chem* 2013, **56**:7431–7441.
Structures of TdT in complex with a non-nucleoside selective inhibitors that compete with the dNTP binding site suggest how to combine them to get more potent compounds.

90. Mora T, Walczak AM, Bialek W, Callan CG: **Maximum entropy models for antibody diversity**. *Proc Natl Acad Sci USA* 2010, **107**:5405–5410.

91. * Murugan A, Mora T, Walczak AM, Callan CG: **Statistical inference of the generation probability of T-cell receptors from sequence repertoires**. *Proc Natl Acad Sci USA* 2012, **109**:16161–16166.
Reconstitution by maximum entropy of the probability distribution of the lengths of N-regions in T-cell receptors. Figure 3 gives the reconstructed probability distribution of the lengths of these N-regions.

92. * Ochi T, Wu Q, Blundell TL: **The spatial organization of non-homologous end joining: from bridging to end joining**. *DNA Repair (Amst)* 2014, **17**:98–109.
Review on the recent results on an integrated structural view of the NHEJ machinery.

93. ** Yin X, Liu M, Tian Y, Wang J, Xu Y: **Cryo-EM structure of human DNA-PK holoenzyme**. *Cell Research* 2017, **27**:1341.
Third of a series of three exciting papers on the DNA-PK holoenzyme bound to DNA duplex at 6.6 Angstrom in a 1:1:1:1 stoeichiometry for Ku 70/80: DNA PKcs: DNA (650 KDa).

94. ** Sharif H, Li Y, Dong Y, Dong L, Wang WL, Mao Y, Wu H: **Cryo-EM structure of the DNA-PK holoenzyme**. *Proc Natl Acad Sci USA* 2017, **114**:7367–7372.
Second paper on the DNA-PK holoenzyme, at a resolution of 5.8 Angstrom, with the full Ku 70/80 hetero-duplex but not with DNA.

95. ** Sibanda BL, Chirgadze DY, Ascher DB, Blundell TL: **DNA-PKcs structure suggests an allosteric mechanism modulating DNA double-strand break repair**. *Science* 2017, **355**:520–524.
The first of a series of three papers that describe the structure of the DNA-PK holoenzyme. In this case, only a C-terminal peptide of Ku 80 is present, and the structure was solved by x-ray crystallography at 4.3 Angstroms.

Legends of Figures

Figure 1

A. Domain organization of eukaryotic polX family; the pol IV of *S. cerevisiae* and *S. pombe* have been omitted. B. Three-dimensional structure of TdT bound to a DNA synapsis [28]. The color code for essential specific features (HhH motif, 5'-phosphate binding site, Loop1, steric gate) is the same as in panel A. The catalytic aspartates are shown in ball-and-stick representation, as well as residues from the steric gate (GW) and the HhH2 motif (GVG). The DNA synapsis is in cyan, the 3' end of the primer strand is in pink, the two catalytic divalent ions, Metal A and Metal B, are shown as a CPK sphere and a dotted circle, respectively (in green), the Na^+ ion bound to the HhH2 motif and a phosphate of the primer strand is in blue.

Figure 2

Schematic representation of all available structures for TdT, pol μ, pol λ and pol β: the apo form, complexes with the incoming dNTP, with a downstream (D/S) dsDNA substrate, a gap-filling DNA substrate or a true DNA synapsis. Loop1 conformation is highlighted, when visible in the electron density map. The DNA substrates are explained in the top panel.

Figure 3

A. Superposition of the DNA synapsis region in the four known polX-DNA complexes, highlighting the conformations of Loop1 and Loop3, as well as the position of the motifs SD1 (purple) and SD2 (dark blue). TdT is in green, pol μ is in blue, pol λ is in brown and pol β is in red.

B. Multialignment of murine sequences of TdT (with its two main forms resulting from alternative splicing), pol μ, pol λ and pol β, in the region of Loop1 and Loop3. The essential Sequence Determinants motifs [24], that maximally discriminate TdT and pol μ [27,57] are highlighted and referred to as SD1 and SD2.

Figure 4

Proposed scenario of the different DNA polymerase activities occurring at a V(D)J junction site for light-chain rearrangements (right panels) and heavy-chain rearrangements (left panels). From top to bottom: template-independent activity, *in trans* template-dependent activity and gap-filling activity. A putative arrangement of the DNA-PK complex is shown in the background, merely to indicate the scales. The ligation step by Ligase IV is omitted for clarity. The definition of the V, D, J regions, constant and variable regions in light chains and heavy chains is recalled in the central panel.