

DNA expansions generated by human Pol μ on iterative sequences

Ana Aza, Maria Jose Martin, Raquel Juarez, Luis Blanco* and Gloria Terrados

Centro de Biología Molecular Severo Ochoa (CSIC-UAM), 28049 Madrid, Spain

Received November 21, 2011; Revised October 8, 2012; Accepted October 10, 2012

ABSTRACT

Pol μ is the only DNA polymerase equipped with template-directed and terminal transferase activities. Pol μ is also able to accept distortions in both primer and template strands, resulting in misinsertions and extension of realigned mismatched primer terminus. In this study, we propose a model for human Pol μ -mediated dinucleotide expansion as a function of the sequence context. In this model, Pol μ requires an initial dislocation, that must be subsequently stabilized, to generate large sequence expansions at different 5'-P-containing DNA substrates, including those that mimic non-homologous end-joining (NHEJ) intermediates. Our mechanistic studies point at human Pol μ residues His³²⁹ and Arg³⁸⁷ as responsible for regulating nucleotide expansions occurring during DNA repair transactions, either promoting or blocking, respectively, iterative polymerization. This is reminiscent of the role of both residues in the mechanism of terminal transferase activity. The iterative synthesis performed by Pol μ at various contexts may lead to frameshift mutations producing DNA damage and instability, which may end in different human disorders, including cancer or congenital abnormalities.

INTRODUCTION

Maintaining the integrity of the DNA sequence is essential for all living cells, which is notably allowed by maximizing fidelity during DNA replication and performing accurate repair of damaged DNA (1). Those processes require a large number of proteins including specialized DNA polymerases. DNA polymerases have been categorized into four different groups attending to their biochemical properties and to the biological processes in which they are involved. They are grouped by their primary sequence homology into family A, B, Y and X. Among them, only

X family DNA polymerases (PolX) are devoted to DNA repair, being evolutionarily conserved in prokaryotes, eukaryotes and archaea (2). DNA polymerases of the X family, which in mammals include DNA polymerase beta (Pol β), lambda (Pol λ), mu (Pol μ) and terminal deoxynucleotidyl transferase (TdT), are structurally related enzymes specialized in DNA repair pathways involving gaps and double-strand breaks (DSBs) [reviewed in (3)].

Human Pol μ , consisting of 494 amino acids, has 41% identity to TdT, its closest homologue in the family. The structural similarities with TdT include a nuclear localization signal at the N-terminus, followed by a BRCT domain and the conserved Pol β core (4). Regarding the biochemical properties, Pol μ displays both terminal transferase and DNA-dependent DNA polymerization activities (4,5). The strong enhancement of these two activities by manganese ions, and the lack of proofreading, make Pol μ a low-fidelity polymerase with a strong mutator behaviour. This mutator activity is further enhanced by its lack of sugar discrimination, allowing the use of both dNTPs and rNTPs (6,7). Moreover, two hallmarks of Pol μ activity are: (i) the capacity to induce/accept template distortions, in order to realign imperfectly paired DNA primers (8,9); and (ii) the capacity to bridge DNA ends with minimal or null complementary, contributing to the efficiency of end-joining by performing either templated or untemplated insertions at the 3' end termini (10,11). Likewise, Pol μ is able to perform DNA synthesis despite the presence of mismatched nucleotides near the primer terminus (12). The combination of all these properties make Pol μ well-suited for a role in the non-homologous end-joining (NHEJ) DNA repair mechanism, as it was early proposed (5,9), and strongly supported by the demonstration of direct interactions of Pol μ with NHEJ factors (13–15), and by analysis of Pol μ deficiency in various *in vivo* systems (13,16–20).

On the other hand, this enzymatic versatility regarding the use and interactions with DNA and nucleotide substrates are the basis for Pol μ to exhibit a high misincorporation rate, being one of the most unfaithful polymerases known in higher eukaryotes (4). The strong

*To whom correspondence should be addressed. Tel: +34 91 1964685; Fax: +34 91 1964799; Email: lblanco@cbm.uam.es

Present address:

Gloria Terrados. Institute of Molecular Biology and Tumor Research (Philipps-Universität Marburg), 35032 Marburg, Germany.

substrate, 50 μ M of the dNTP or the mix of dNTPs indicated in each case, 2 mM $MgCl_2$ and amounts of DNA polymerase as indicated in the figure legends. After the incubation indicated in each case, reactions were stopped by adding gel loading buffer [95% (v/v) formamide, 10 mM EDTA, 0.1% (w/v) xylene cyanol and 0.1% (w/v) bromophenol blue]. Samples were analysed by 8 M urea/20% PAGE and either autoradiography or Phosphorimager scanning.

3D-modeling

The different conformations of selected residues of Pol μ were analysed by using the Swiss PDB-Viewer (<http://www.expasy.ch/spdbv/>) and MacPymol (The PyMOL Molecular Graphics System, Schrödinger, LLC, <http://www.pymol.org/>).

RESULTS

Pol μ , but neither Pol λ nor Pol β , produces large sequence expansions when a specific repeated trinucleotide sequence is used as a template

Pol μ can act as a mutator polymerase, based on its ability to realign and dislocate DNA chains during polymerization, whereas Pol λ and Pol β , belonging to the same family as Pol μ , are not so versatile and promiscuous in the use of DNA substrates (36). All three polymerases have a marked preference for 5'-P gapped molecules, as the 8 kDa domain strongly interacts with this group at the 5'-end of the downstream chain (35,37,38), improving binding stability. Thus, we addressed if any of these DNA polymerases could generate triplet repeat expansions in the context of a gap. According to the substrate preferences of these polymerases, four different DNA substrates containing a 6-nt gap were designed, with different trinucleotide sequences as templates, each one repeated twice (see scheme in Figure 1). Expansion of triplet repetitions of the sequence 3'-GAC (or its complementary chain 3'-GTC) are associated to Huntington's disease; triplet repeat expansions of 3'-AAG (or its complementary sequence 3'-CTT) are related to Friedrich's ataxia. Polymerization reactions were assayed on these molecules, in the presence of either Pol μ , Pol λ or Pol β , and the complementary dNTPs needed in each case (Figure 1). In general, Pol λ and Pol β filled the gap and then polymerization stopped (+6 product). In the case of 3'-(GTC) $_2$, Pol β displayed a significant strand-displacement capacity and efficiently continued adding nucleotides after filling the gap (+18 product). This imprecise gap-filling by Pol β has been already described (39). Interestingly, Pol μ was eventually able to continue polymerization, producing extra additions that could be also originated via strand displacement. However, this behaviour is very remarkable in the case of 3'-(CTT) $_2$, where Pol μ fills the gap and then efficiently generates a very large DNA expansion as a final product. The main difference between this sequence and the others is the presence of a dinucleotide (TT) in the template strand at the end of the gap.

Pol μ requires a dinucleotide at the end of the template sequence to produce large sequence expansions

Taking into account the dislocation capacity of Pol μ , strongly driven by the presence of dinucleotides in the template strand, we wanted to confirm if the presence of a repeated nucleotide at the end of the gap determines the capability of Pol μ to produce sequence expansions. For this, two new DNA substrates containing 6-nt gaps were chosen: the first, with the template sequence 3'-(CAG) $_2$, with no repeated nucleotides, and the second 3'-(GAA) $_2$, which maintains the presence of a duplicated nucleotide at the end of the gap (Figure 2). We compared the behaviour of Pol μ and Pol λ on these substrates, providing the complementary nucleotides needed in each case. As shown in Figure 2, in both cases Pol λ filled the 6-nt gap and then polymerization stopped. Conversely, Pol μ was able to continue polymerization beyond the expected +6 product in both cases, producing a large sequence expansion only in the case of confronting a repeated nucleotide (AA) at the end of the gap (Figure 2). Thus, this iteration is a relevant requirement for Pol μ in order to produce promiscuous elongation.

Next we addressed if the repetition of the trinucleotide was specifically needed for Pol μ to produce the expansion. For that, Pol μ was assayed on a DNA substrate having only one repetition of the triplet sequence, forming a 3-nt gap containing the sequence 3'-GAA, that maintains the requirement of a repeated nucleotide at the end of the gap. As shown in Figure 2, when complementary nucleotides (dC and dT) were provided, Pol μ generated again a large sequence expansion, demonstrating that the dinucleotide is required only once at the end of the template strand.

An initial dislocation and remaining distortion are necessary for the generation of large sequence expansions by Pol μ

By using additional sequence contexts, Pol μ was confirmed to be able to produce large sequence expansions of the four different duplicated nucleotides (AA, TT, CC and GG) at the end of a 3-nt gap (Figure 3A). Pol λ was used as negative control of expansion also in this experiment. Interestingly, in the 3'-CCC gapped substrate Pol μ had a different behaviour than on the other 3 DNA substrates: no large sequence expansion was produced, but only a few more additions were observed once the gap had been filled. Strikingly, the behaviour appears to be the opposite in the case of Pol λ , which produced some significant expansion only when copying the sequence 3'-CCC, perhaps the most prone to facilitate slippage after gap-filling. Thus, the modest outcome of the reaction on the 3'-CCC substrate indicates a new specific feature needed for the generation of large expansions by Pol μ : the nucleotide at the first (n+1) template position must differ from that forming the repetition in order to obtain the maximal sequence expansion. This demand at the trinucleotide sequence suggests that the expansion reaction begins with an initial dislocation reaction, a very particular mechanism described for Pol μ (8,9), producing a -1 frameshift during the initial step of the gap-filling reaction: Pol μ 's propensity to dislocate

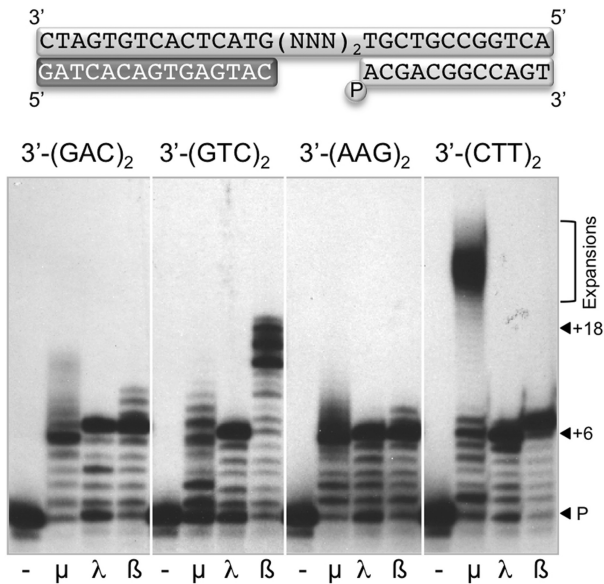


Figure 1. Pol μ generates a large sequence expansion on a DNA gap with a specific trinucleotide sequence. In the scheme, the template sequence indicated (NNN) corresponds to the trinucleotides shown below. Subindex '2' indicates that each trinucleotide sequence is twice repeated at the gap. Polymerization reactions (described in 'Materials and Methods' section) were performed in the presence of 4 nM of the indicated DNA substrate in each case, 2 mM MgCl₂, 270 nM of each polymerase and 50 μ M mix dNTPs (using the complementary nucleotides to the gap sequence in each case). After 1 h incubation at 30°C, polymerization products were analysed by electrophoresis on 20% polyacrylamide/8 M urea gels and autoradiography. P indicates the unextended primer; +6 the elongated product upon complete gap-filling; +18 the fully elongated product after gap-filling and strand displacement; the products of nucleotide expansion are indicated with a bracket.

the template strand is such that it preferentially inserts the nucleotide complementary to the $n+2$ position of the template, both in an open template/primer and in a 2-nt gap (Supplementary Figure S1), either in the conditions that favour a mechanism of primer slippage (Supplementary Figure S1A) or when this is not possible due to the sequence context (Supplementary Figure S1B). This second mechanism, 'dNTP-selection-mediated', is not as efficient as the 'slippage-mediated' alternative, but is considerably improved when the substrate is a small 5'-P-containing gap (Supplementary Figure S1C).

To further understand the requirements for the special mechanism of dinucleotide expansion by Pol μ , we evaluated both the impact of the sequence at the first position of the gap, and also the necessity of providing either the 2 nt complementary to the sequence of the gap, or only the one complementary to the dinucleotide (the only one needed if dislocation occurs). We used four different 3-nt gapped DNA substrates (Figure 3B), with the same terminal repetition (AA), but different nucleotide at the $n+1$ position of the gap. As shown in Figure 3B, a large expansion is obtained in the case of the 3'-GAA substrate, and no difference is obtained when providing either dC+dT or only dT, indicating that in most of the cases the first templating base (dG) is not copied. Such a preferential dislocation is compatible with

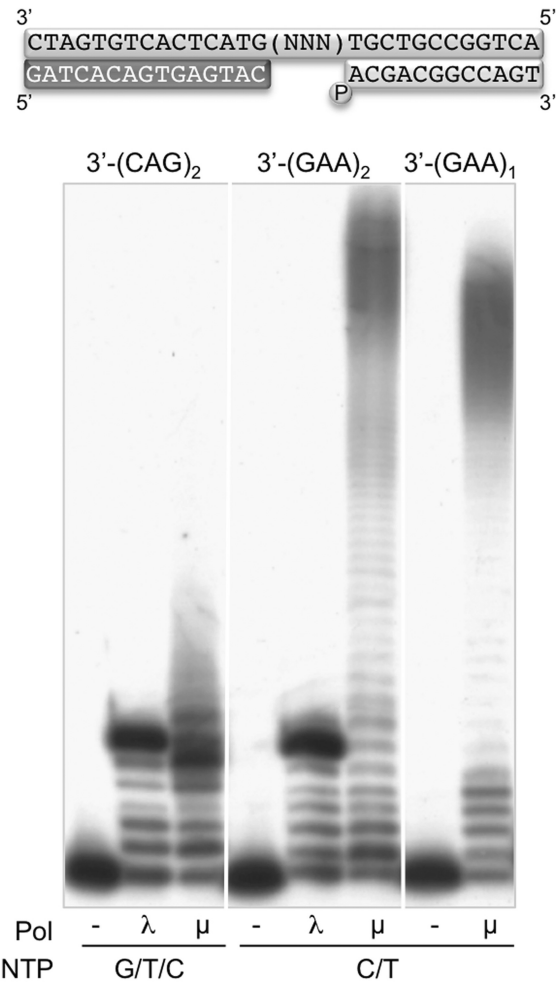


Figure 2. A repeated nucleotide at the end of a gap is required to generate large sequence expansions. In the scheme, the template sequence indicated (NNN) corresponds to the trinucleotides shown below, that could be present twice (subindex 2) or only once (subindex 1). Polymerization reactions (described in 'Materials and Methods' section) were performed in the presence of 4 nM of the indicated DNA substrate in each case, 2 mM MgCl₂, 270 nM of either Pol λ or Pol μ and 50 μ M mix dNTPs (using the complementary nucleotides to the gap sequence in each case). After 1 h incubation at 30°C, polymerization products, including sequence expansions, were analysed by electrophoresis on 20% polyacrylamide/8 M urea gels and autoradiography.

a slippage-mediated mechanism, since the 3'-terminus of the primer strand (dC) could be realigned and matched to the first templating base (dG) [(40); see also Supplementary Figure S1A]. Moreover, large expansions on the gap sequences 3'-TAA and 3'-CAA are also obtained by providing only dTTP. In these two cases, the dTTP insertion event triggering expansion would be compatible with a dNTP-selection-mediated dislocation [(8); see Supplementary Figure S1B]. This mechanism dominates insertion in the 3'-TAA gap, as the addition of dATP+dTTP has minimal, if any, effect on expansion. Conversely, expansions on the 3'-CAA gap are significantly inhibited by the simultaneous addition of dGTP and dTTP. These results point to the efficiency of dislocation as an important factor determining the expansion

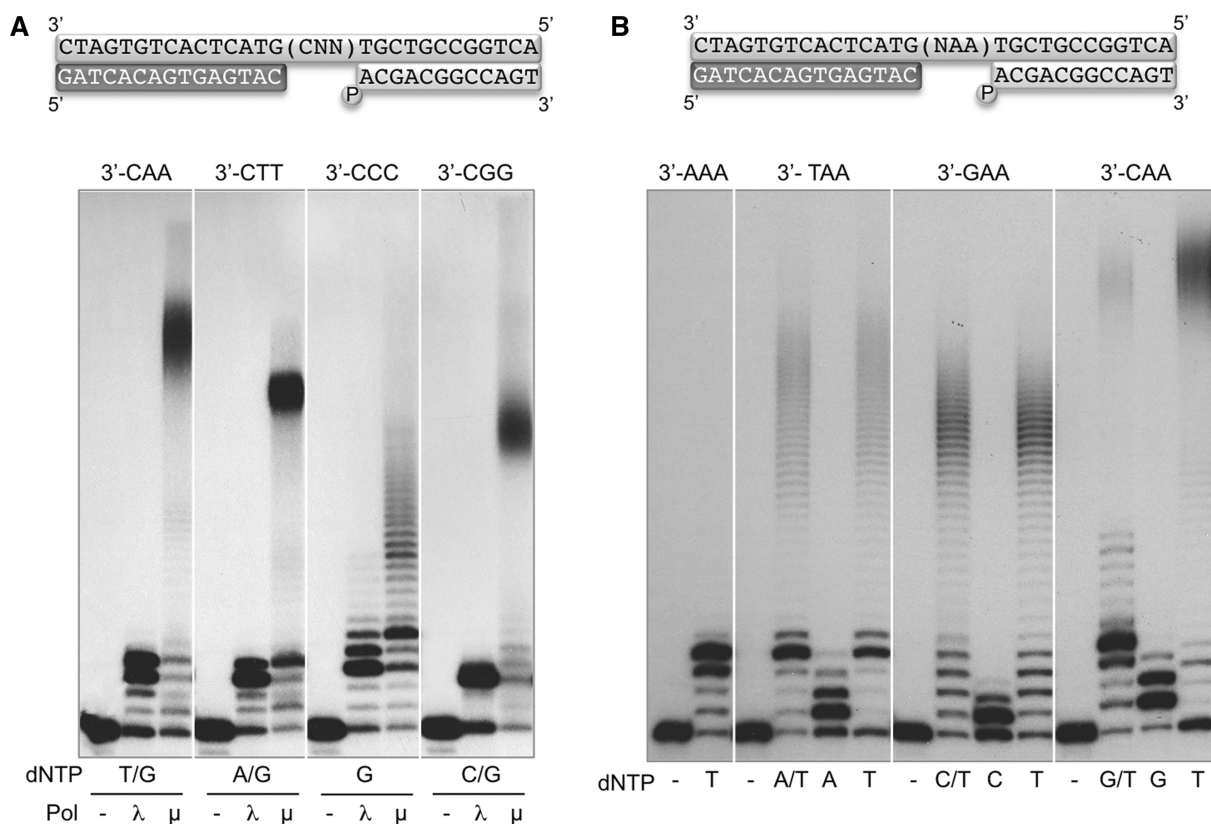


Figure 3. A distortion upstream to the dinucleotide is a prerequisite for the formation of Pol μ -mediated sequence expansions during gap-filling. (A) In the scheme, the template sequence indicated (CNN) corresponds to the four trinucleotides shown below. The first base at the trinucleotide is always dC, followed by the four different homo-dinucleotides. Polymerization reactions (described in 'Materials and Methods' section) were performed in the presence of 4 nM of the indicated DNA substrate in each case, 2 mM MgCl₂, 270 nM of either Pol λ or Pol μ and 50 μ M mix dNTPs (using the complementary nucleotides to the gap sequence in each case). (B) Importance of the first nucleotide of the triplet (NAA) for the efficiency of expansion by Pol μ . The gap sequence used and the different combinations of nucleotides provided, were as indicated. Polymerization reactions were performed as in (A). After 1 h incubation at 30°C, polymerization products were analysed by electrophoresis on 20% polyacrylamide/8 M urea gels and autoradiography.

capability, which could be also affected by an unbalanced concentration of the deoxynucleotide precursors.

In agreement with our previous observations using the 3'-CCC substrate (Figure 2), Pol μ did not produce a significant expansion on the 3'-AAA substrate. Assuming that an initial dislocation event can also occur, an important difference in these two cases is that, after complete gap-filling, the nascent chain could be completely realigned, producing a perfectly matched +3 product. In this case, expansion seems to be precluded, as evidenced by the minimum extension of the primer over the size of the gap. Therefore, an initial dislocation of the template only triggers dinucleotide expansion if the distortion remains after gap-filling. Further analysis showed that the necessary distortion that allows generation of the observed sequence expansions might be present in the substrate prior to the arrival of the polymerase, triggering dinucleotide expansions even at single nicks in DNA (Supplementary Figure S2). As it will be evaluated later in this section, that is particularly relevant considering the *in vivo* situations where Pol μ deals with substrates containing distortions and/or misalignments, caused by microhomology search during the bridging of two DNA ends in NHEJ reactions. The presence of a 5'-P group at

the downstream strand of the gap is also a prerequisite for the generation of these nucleotide expansions, as shown in Supplementary Figure S3. This observation was expected since the 5'-P is a main anchor point for Pol μ on the DNA substrate (38).

Specific residues regulating the expansion of repeated sequences

Pol μ is an exceptional enzyme since it is the only DNA polymerase able to display template-independent (terminal transferase) and template-dependent activities (4,5). Recent structure-function studies have shed light on the molecular basis for the terminal transferase activity: on the one hand, Pol μ contains a flexible piece, Loop 1, which is able to undergo conformational changes, acting as a pseudo-template that allows incorporation of nucleotides in the absence of template information (41), or during NHEJ of some incompatible ends (15). Moreover, the crystal structure of Pol μ in ternary complex with gap DNA and deoxynucleotide (42) allowed to infer that a specific histidine residue (His³²⁹ in Pol μ and His³⁴² in TdT; absent in Pol β and Pol λ) could play an important role to overcome the rate-limiting step of untemplated polymerization, allowing terminal transferase activity

to occur. In fact, this residue, involved in the proper positioning of the primer terminus and the incoming nucleotide (Figure 4A), was shown to be critical during template-independent polymerization but also in template-directed reactions associated to NHEJ of short incompatible ends (10,42). Furthermore Arg³⁸⁷ is a specific DNA-binding ligand (Figure 4A) that limits untemplated nucleotide additions and is thus responsible for the lower terminal transferase activity of Pol μ in comparison to TdT (10).

To study the relevance of both residues in the expansion of repeated sequences we used two mutants: H329G (with a strongly reduced terminal transferase activity) and R387K (displaying an augmented terminal transferase activity). The selected mutants, obtained and purified as described (10), were tested on the set of gapped DNA substrates that maintain the same repetition (AA) at the 5'-end, but differ in the nucleotide at the $n+1$ position in the gap. As shown in Figure 4B, the wild-type Pol μ produced an expansion pattern similar to that described before (Figure 3B). Strikingly, mutant H329G was only

able to fill the different gaps, displaying very minor or even undetectable levels of sequence expansions on these substrates. Mutant R387K, on the other hand, produced a level of expansions similar in most cases to that of the wild-type enzyme (Figure 4B), but, remarkably, was also able to produce a significant expansion in the case of the 3'-AAA substrate. This was striking, as the latter substrate does not allow formation of the distortion that was an obligatory requirement for producing expansions by the wild-type enzyme. These results support our initial hypothesis that Arg³⁸⁷ has a constitutive role in preventing slippage of the primer in each round of the catalytic cycle, through a direct contact between this residue and the -2 position of the primer strand that can be observed in the ternary complex of Pol μ (Figure 4A). Mutation of this arginine to lysine favours expansion in the absence of distortions through a mechanism that facilitates the backwards translocation of the primer strand. In the case of distortion-mediated expansions, where the primer strand is not properly oriented, Arg³⁸⁷ cannot exert its 'braking'

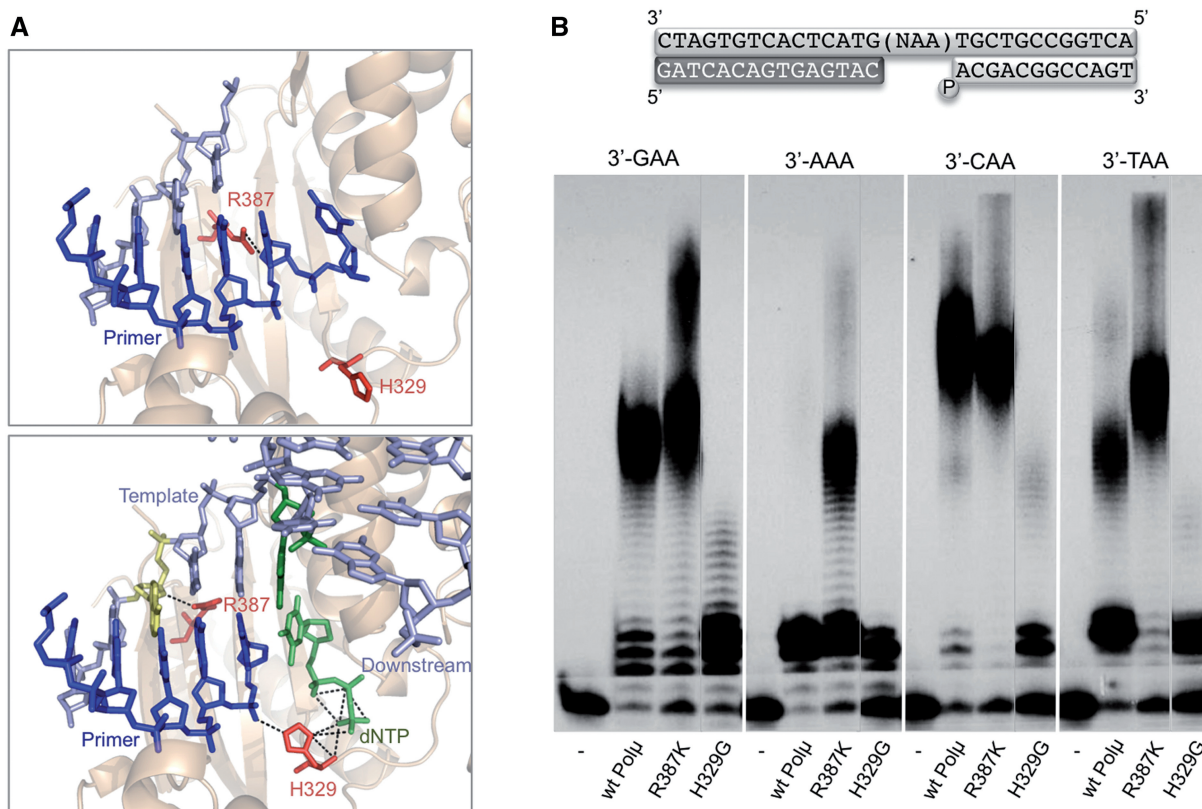


Figure 4. His³²⁹ allows while Arg³⁸⁷ limits sequence expansions by Pol μ . (A) *Top*: Model of Pol μ bound to a template/primer substrate in which the 3'-protruding primer terminus is in an unproductive position, occupying the incoming dNTP site. Residue Arg³⁸⁷, in a conformation modelled to match that of the lysine present in TdT in a similar structure, is contacting the primer impeding its backwards translocation. His³²⁹, modelled in the conformation observed for the same residue in the crystal of TdT bound to an ssDNA primer, is not making any contacts with the DNA substrate. *Bottom*: Crystal structure of Pol μ bound to a gapped DNA substrate and incoming dNTP. His³²⁹ has rotated and is contacting both incoming dNTP and primer terminus, helping to reposition the latter. Arg³⁸⁷ is now contacting the template strand ($n-3$ position; indicated in yellow), having allowed the movement backwards of the primer. DNA substrates are indicated in dark (primer strand) and light (template and downstream strand) blue. Incoming dNTP is indicated in green. (B) In the scheme, the template sequence indicated (NAA) corresponds to the four trinucleotides shown below. The two last bases always form the dinucleotide AA, preceded by any of the four different nucleotides. Polymerization reactions (described in Materials and Methods section) were performed in the presence of 4 nM of the indicated DNA substrate in each case, 2 mM MgCl₂, 270 nM Pol μ and 50 μ M mix dNTPs (using the complementary nucleotide to the dinucleotide AA). After 1 h incubation at 30°C, polymerization products were analysed by electrophoresis on 20% polyacrylamide/8 M urea gels and autoradiography.

role. Mutation of arginine to lysine in this context does not increase the sequence expansions any further since the control mechanism is already compromised.

Impact of the sequence expansions during end-joining reactions

As noted before, dinucleotide expansion could be generated by Pol μ as a by-product of its physiological role of repairing DSBs by the NHEJ pathway. We wanted to check whether the sequence expansion capacity exhibited by Pol μ in the context of DNA-gapped substrates is also demonstrated during NHEJ reactions that include the formation of short gaps, and if the requirements detected during gap-filling (i.e. the formation of a distortion upstream of the polymerization point, and the presence of a dinucleotide in the template strand) also need to be met during end-joining of two DNA ends.

For this we used a tailored set of 3'-protruding NHEJ substrates whose protrusions provide, once bridged by the polymerase, a microhomology of 3 base-pairs, the flipping-out of a nucleotide at the -1 position of the primer strand, and the formation of two 1-nt gaps at both sides of the connection (see scheme in Figure 5A). Radioactive labelling of one of the 3'-protruding strands allows detection of nucleotide incorporation on this end, using the template information provided *in trans* by the other 3'-protrusion, that contains a 5'-P-group and will thus be used as a template/downstream structure (through binding of the phosphate by the 8 kDa domain of Pol μ). In order to evaluate if the reaction is trans-directed, the templating base (X) on this second DNA substrate was changed to A, C, G or T (Figure 5). Our results showed that Pol μ is able to perform an efficient and mostly template-directed trans-polymerization on this kind of NHEJ substrates, since the polymerase incorporated preferentially the nucleotide complementary to the templating base (white asterisks and box in Figure 5A). As expected, when the templating base is a G and thus the template strand contains a dinucleotide (GG), large sequence expansions were produced when providing the nucleotide (dCTP) complementary to the dinucleotide. The reiterative additions of dGTP that are also catalysed by Pol μ in every case can be considered pure terminal transferase incorporations, since control experiments in which the template-providing end is not present, also showed this outcome (Supplementary Figure S4). This is in accordance with the preference for nucleotide incorporation displayed by Pol μ during untemplated additions (41).

Strikingly, in the context of NHEJ we were able to observe the formation of large sequence expansions even in the absence of a provided distortion upstream of the polymerization site (Figure 5B), thus limiting the requirements of this reaction only to the presence of a dinucleotide, a prerequisite that still needs to be met in this context. The amount of nucleotide that Pol μ requires to produce these expansions is low (20 μ M), indicating that this process is highly efficient (Supplementary Figure S5A and B). Again, dGTP is being inserted as a result of pure terminal transferase additions, as demonstrated by

a control experiment, in which the template-providing end is not present (Supplementary Figure S4). We also detected expansion during NHEJ when the dinucleotide was formed by a pair of adenines (Supplementary Figure S5E and F), indicating that this mechanism is independent of the sequence context. Taken together, our results indicate that if the sequence context is favourable to the expansions (i.e., iteration of nucleotides at the template strand), the polymerase itself may generate the required upstream distortion by adjusting the bridging of the two ends, a scenario that emphasizes the importance of the mutagenic potential of Pol μ during the NHEJ pathway, specifically regarding nucleotide expansions.

DISCUSSION

Indels (insertions and deletions) are common errors produced during DNA replication and repair, related to with different human pathologies including cancer and diseases associated with expansion of repeats. All polymerases studied to date generate indels during DNA synthesis *in vitro* (43), but with very different frequency. Thus, although X family members Pol β , Pol λ and Pol μ all generate single-base deletions during synthesis (9,21,22), Pol λ has a much higher deletion rate, whose structural basis has been proposed (44). The first hypothesis explanatory of the production of indels was introduced by Streisinger *et al.* (40): these frameshift mutations were described as products of strand slippage in repetitive DNA sequences. Other two models have been proposed since then, namely 'direct misincorporation misalignment', in which a polymerase introduces an initial mismatch that causes the primer *terminus* to be subsequently realigned (44,45) and 'dNTP-stabilized misalignment', in which incorporation of the correct dNTP occurs in front of a complementary downstream template base (46,47).

The results presented here demonstrate that human Pol μ can catalyse large nucleotide expansions when copying a repeated templating base in the vicinity of a 5'-P. Based on our findings with different sequence contexts, we propose a specific model for Pol μ -mediated generation of the expansions during gap-filling that requires: (i) initial dislocation of the template strand; (ii) generation of a mismatch/distortion, that will trigger nucleotide expansion (Figure 6B and C). Initial template dislocation can be either facilitated by slippage, when the primer-*terminus* is complementary to the first nucleotide at the gap (Figure 6B), or stabilized by the incoming nucleotide (dNTP-mediated; Figure 6C); in both cases, after initial dislocation (stage 1), the gap is filled and a mismatch is left behind (stage 2), and then expansion occurs (stage 3). However, such expansions are not simply the result of Streisinger's 'strand slippage': if the sequence to be copied is formed by the three same nucleotides (Figure 6A), the expansion hardly happens because the mismatch/distortion that could trigger further nucleotide incorporation beyond gap-filling is not allowed to occur.

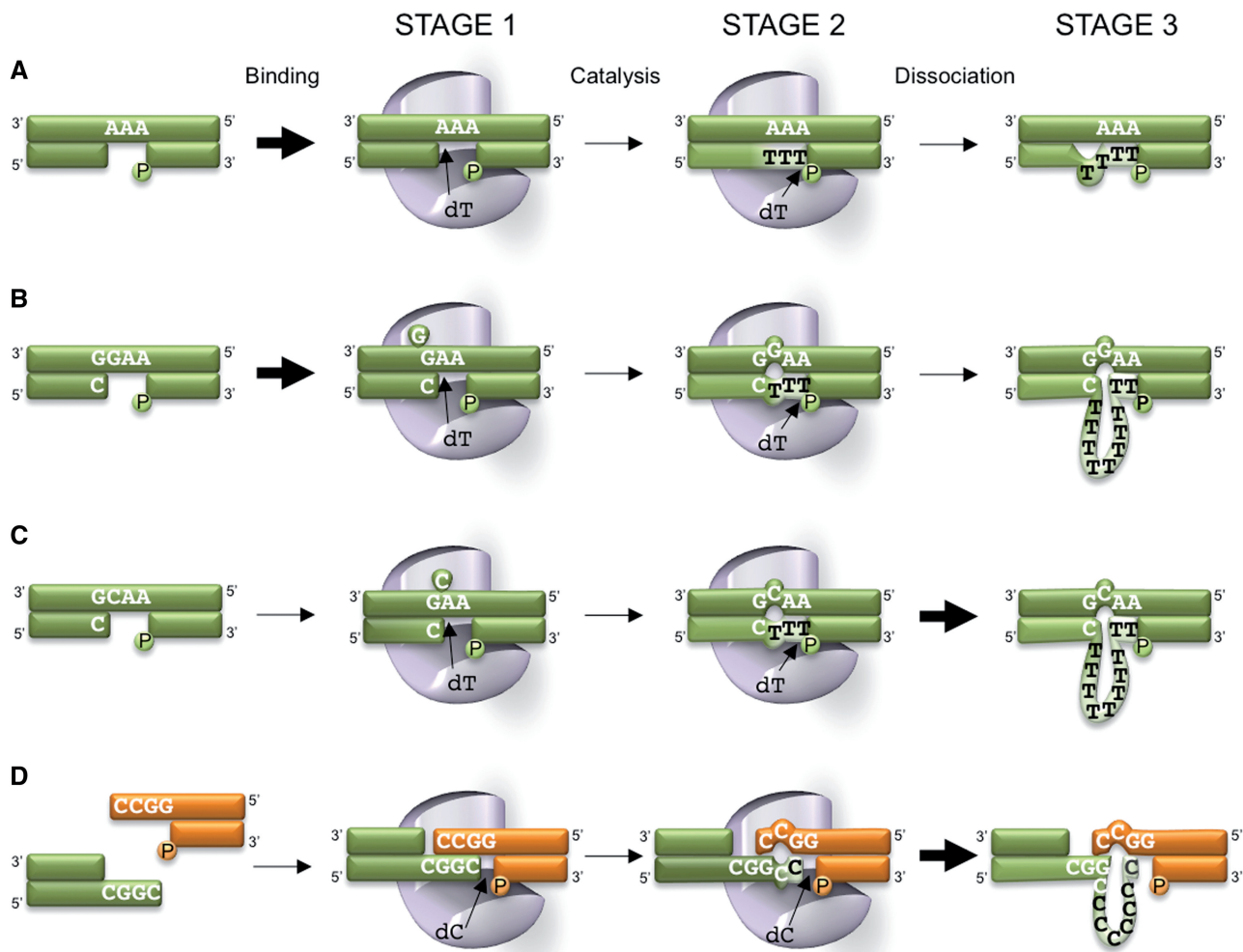


Figure 6. Mechanistic model for dinucleotide expansions generated by human Pol μ . Thick arrows indicate more efficient reactions, whereas thin arrows are related to slower or less favourable reactions. (A) Trinucleotide conformed by the same nucleotide. In this case, no distortion associated to gap-filling is generated, thus precluding a large expansion. (B) After polymerase binding and realignment of the primer terminus, a 'slippage-mediated' dislocation is formed, creating a template distortion. In this situation, a large expansion reaction is observed. (C) In this case the distortion is induced by a 'dNTP-selection-mediated' dislocation of the template strand, again resulting in the generation of large sequence expansions. (D) In a NHEJ context, a repeated nucleotide neighbour to the 5'-P can induce nucleotide expansions by Pol μ , although in this case, a pre-existing stable distortion or impairing is not strictly required.

for NHEJ of non-complementary ends, has an unexpected downside, since it also allows Pol μ to generate large sequence expansions in the context of repair reactions.

Pol μ , a candidate to generate mono- and dinucleotide expansions *in vivo*

The genome of most organisms thus far examined contains many tracts of repetitive DNA called microsatellites. The discovery that a number of human diseases are the direct consequence of mutations within such repeats has triggered considerable interest in the mechanisms that change the number of copies of repeated DNA sequences. DNA expansions in mono- and dinucleotide repeats are more likely to be deleterious to the cell by causing not only addition mutations but also frameshift mutations. Current models of DNA repeat instability involve DNA polymerase slippage at these iterative tracts, which are normally

associated with mutation hot-spots. Which polymerases are responsible for expansions? Initial efforts were oriented to measure DNA polymerase-catalysed 'reiterative replication' of repeat sequences with replicative polymerases (34,49–52), but evidence soon indicated that those lacking proofreading and also strand displacement capabilities are better candidates (53–55). Moreover, other results indicate that sequence expansions could be linked to DNA damage (32), a process causatively related with ageing (56,57). That could be a vicious cycle, as it is quite possible that repetitive DNA is a better target for DNA damage than normal DNA. Repair substrates as short gaps are produced *in vivo* during base excision repair, as well as during the final steps of nucleotide excision repair and post-replication MMR. More specifically, DNA polymerases as Pol μ are able to configure gap-like substrates during NHEJ. In all these substrates, the presence of a downstream strand would lead to the

'stalling' of the polymerization reaction (54), thus providing sufficient time for realigning the primer strand and triggering expansive nucleotide insertion.

Our results suggest that Pol μ can generate the expansion of mononucleotide tracts during these repair reactions *in vivo*, given the appropriate sequence context. It is not yet known whether homopolymeric runs are more prone to DSBs than non-iterative sequences, but the possibility can be easily envisaged, due to the ssDNA-containing secondary structures that these sequences can adopt. As derived from our work, in the case of a DSB occurring at a site containing an iterative sequence as short as a single dinucleotide, the outcome of a Pol μ -mediated repair reaction would result in the generation of frameshifts and sequence expansions. This could in turn lead to an increased risk of microsatellite and genome instability, events that have been related to cancer and other illnesses such as neurodegenerative syndromes.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online: Supplementary Figures 1–5.

FUNDING

Ministerio de Ciencia y Tecnologia [BFU2009-10085 and CONSOLIDER CSD2007-00015]; Fundacion Ramon Areces (to Centro de Biologia Molecular "Severo Ochoa"); Ministerio de Educacion y Ciencia (to A.A.); Comunidad Autonoma de Madrid (to M.J.M.). Funding for open access charge: Comunidad de Madrid grant [S2011/BMD-2361].

Conflict of interest statement. None declared.

REFERENCES

- Hoeijmakers, J.H. (2001) DNA repair mechanisms. *Maturitas*, **38**, 17–22, discussion 22–13.
- Burgers, P.M., Koonin, E.V., Bruford, E., Blanco, L., Burtis, K.C., Christman, M.F., Copeland, W.C., Friedberg, E.C., Hanaoka, F., Hinkle, D.C. *et al.* (2001) Eukaryotic DNA polymerases: proposal for a revised nomenclature. *J. Biol. Chem.*, **276**, 43487–43490.
- Moon, A.F., Garcia-Diaz, M., Batra, V.K., Beard, W.A., Bebenek, K., Kunkel, T.A., Wilson, S.H. and Pedersen, L.C. (2007) The X family portrait: structural insights into biological functions of X family polymerases. *DNA Repair*, **6**, 1709–1725.
- Dominguez, O., Ruiz, J.F., Lain de Lera, T., Garcia-Diaz, M., Gonzalez, M.A., Kirchhoff, T., Martinez, A.C., Bernad, A. and Blanco, L. (2000) DNA polymerase mu (Pol mu), homologous to TdT, could act as a DNA mutator in eukaryotic cells. *EMBO J.*, **19**, 1731–1742.
- Ruiz, J.F., Dominguez, O., Lain de Lera, T., Garcia-Diaz, M., Bernad, A. and Blanco, L. (2001) DNA polymerase mu, a candidate hypermutase? *Philos. Trans. R. Soc. Lond. B Biol. Sci.*, **356**, 99–109.
- Nick McElhinny, S.A. and Ramsden, D.A. (2003) Polymerase mu is a DNA-directed DNA/RNA polymerase. *Mol. Cell. Biol.*, **23**, 2309–2315.
- Ruiz, J.F., Juarez, R., Garcia-Diaz, M., Terrados, G., Picher, A.J., Gonzalez-Barrera, S., Fernandez de Henestrosa, A.R. and Blanco, L. (2003) Lack of sugar discrimination by human Pol mu requires a single glycine residue. *Nucleic Acids Res.*, **31**, 4441–4449.
- Ruiz, J.F., Lucas, D., Garcia-Palomero, E., Saez, A.I., Gonzalez, M.A., Piris, M.A., Bernad, A. and Blanco, L. (2004) Overexpression of human DNA polymerase mu (Pol mu) in a Burkitt's lymphoma cell line affects the somatic hypermutation rate. *Nucleic Acids Res.*, **32**, 5861–5873.
- Zhang, Y., Wu, X., Yuan, F., Xie, Z. and Wang, Z. (2001) Highly frequent frameshift DNA synthesis by human DNA polymerase mu. *Mol. Cell. Biol.*, **21**, 7995–8006.
- Andrade, P., Martin, M.J., Juarez, R., Lopez de Saro, F. and Blanco, L. (2009) Limited terminal transferase in human DNA polymerase mu defines the required balance between accuracy and efficiency in NHEJ. *Proc. Natl Acad. Sci. USA*, **106**, 16203–16208.
- Davis, B.J., Havener, J.M. and Ramsden, D.A. (2008) End-bridging is required for pol mu to efficiently promote repair of noncomplementary ends by nonhomologous end joining. *Nucleic Acids Res.*, **36**, 3085–3094.
- Duvauchelle, J.B., Blanco, L., Fuchs, R.P. and Cordonnier, A.M. (2002) Human DNA polymerase mu (Pol mu) exhibits an unusual replication slippage ability at AAF lesion. *Nucleic Acids Res.*, **30**, 2061–2067.
- Capp, J.P., Boudsocq, F., Besnard, A.G., Lopez, B.S., Cazaux, C., Hoffmann, J.S. and Canitrot, Y. (2007) Involvement of DNA polymerase mu in the repair of a specific subset of DNA double-strand breaks in mammalian cells. *Nucleic Acids Res.*, **35**, 3551–3560.
- Mahajan, K.N., Nick McElhinny, S.A., Mitchell, B.S. and Ramsden, D.A. (2002) Association of DNA polymerase mu (pol mu) with Ku and ligase IV: role for pol mu in end-joining double-strand break repair. *Mol. Cell. Biol.*, **22**, 5194–5202.
- Nick McElhinny, S.A., Havener, J.M., Garcia-Diaz, M., Juarez, R., Bebenek, K., Kee, B.L., Blanco, L., Kunkel, T.A. and Ramsden, D.A. (2005) A gradient of template dependence defines distinct biological roles for family X polymerases in nonhomologous end joining. *Mol. Cell*, **19**, 357–366.
- Bertocci, B., De Smet, A., Weill, J.C. and Reynaud, C.A. (2006) Nonoverlapping functions of DNA polymerases mu, lambda, and terminal deoxynucleotidyltransferase during immunoglobulin V(D)J recombination *in vivo*. *Immunity*, **25**, 31–41.
- Chayot, R., Danckaert, A., Montagne, B. and Ricchetti, M. (2010) Lack of DNA polymerase mu affects the kinetics of DNA double-strand break repair and impacts on cellular senescence. *DNA Repair (Amst)*, **9**, 1187–1199.
- Chayot, R., Montagne, B. and Ricchetti, M. (2012) DNA polymerase mu is a global player in the repair of non-homologous end-joining substrates. *DNA Repair (Amst)*, **11**, 22–34.
- Gozalbo-Lopez, B., Andrade, P., Terrados, G., de Andres, B., Serrano, N., Cortegano, I., Palacios, B., Bernad, A., Blanco, L., Marcos, M.A. *et al.* (2009) A role for DNA polymerase mu in the emerging DJH rearrangements of the postgastrulation mouse embryo. *Mol. Cell. Biol.*, **29**, 1266–1275.
- Lucas, D., Escudero, B., Ligos, J.M., Segovia, J.C., Estrada, J.C., Terrados, G., Blanco, L., Samper, E. and Bernad, A. (2009) Altered hematopoiesis in mice lacking DNA polymerase mu is due to inefficient double-strand break repair. *PLoS Genet.*, **5**, e1000389.
- Bebenek, K., Garcia-Diaz, M., Blanco, L. and Kunkel, T.A. (2003) The frameshift infidelity of human DNA polymerase lambda. Implications for function. *J. Biol. Chem.*, **278**, 34685–34690.
- Kunkel, T.A. (1985) The mutational specificity of DNA polymerase-beta during *in vitro* DNA synthesis. Production of frameshift, base substitution, and deletion mutations. *J. Biol. Chem.*, **260**, 5787–5796.
- Covo, S., Blanco, L. and Livneh, Z. (2004) Lesion bypass by human DNA polymerase mu reveals a template-dependent, sequence-independent nucleotidyl transferase activity. *J. Biol. Chem.*, **279**, 859–865.
- Bowater, R.P. and Wells, R.D. (2001) The intrinsically unstable life of DNA triplet repeats associated with human hereditary disorders. *Prog. Nucleic Acid Res. Mol. Biol.*, **66**, 159–202.
- Wells, R.D., Dere, R., Hebert, M.L., Napierala, M. and Son, L.S. (2005) Advances in mechanisms of genetic instability related to

- hereditary neurological diseases. *Nucleic Acids Res.*, **33**, 3785–3798.
26. Mirkin, S.M. (2005) Toward a unified theory for repeat expansions. *Nat. Struct. Mol. Biol.*, **12**, 635–637.
 27. Mirkin, S.M. (2007) Expandable DNA repeats and human disease. *Nature*, **447**, 932–940.
 28. Pearson, C.E., Nichol Edamura, K. and Cleary, J.D. (2005) Repeat instability: mechanisms of dynamic mutations. *Nat. Rev. Genet.*, **6**, 729–742.
 29. Manley, K., Shirley, T.L., Flaherty, L. and Messer, A. (1999) Msh2 deficiency prevents in vivo somatic instability of the CAG repeat in Huntington disease transgenic mice. *Nat. Genet.*, **23**, 471–473.
 30. van den Broek, W.J., Nelen, M.R., Wansink, D.G., Coerwinkel, M.M., te Riele, H., Groenen, P.J. and Wieringa, B. (2002) Somatic expansion behaviour of the (CTG)_n repeat in myotonic dystrophy knock-in mice is differentially affected by Msh3 and Msh6 mismatch-repair proteins. *Hum. Mol. Genet.*, **11**, 191–198.
 31. Bogdanov, M.B., Andreassen, O.A., Dedeoglu, A., Ferrante, R.J. and Beal, M.F. (2001) Increased oxidative damage to DNA in a transgenic mouse model of Huntington's disease. *J. Neurochem.*, **79**, 1246–1249.
 32. Kovtun, I.V., Liu, Y., Bjoras, M., Klungland, A., Wilson, S.H. and McMurray, C.T. (2007) OGG1 initiates age-dependent CAG trinucleotide expansion in somatic cells. *Nature*, **447**, 447–452.
 33. Liu, Y., Prasad, R., Beard, W.A., Hou, E.W., Horton, J.K., McMurray, C.T. and Wilson, S.H. (2009) Coordination between polymerase beta and FEN1 can modulate CAG repeat expansion. *J. Biol. Chem.*, **284**, 28352–28366.
 34. Sinden, R.R., Potaman, V.N., Oussatcheva, E.A., Pearson, C.E., Lyubchenko, Y.L. and Shlyakhtenko, L.S. (2002) Triplet repeat DNA structures and human genetic disease: dynamic mutations from dynamic DNA. *J. Biosci.*, **27**, 53–65.
 35. Garcia-Diaz, M., Bebenek, K., Sabariego, R., Dominguez, O., Rodriguez, J., Kirchhoff, T., Garcia-Palomero, E., Picher, A.J., Juarez, R., Ruiz, J.F. *et al.* (2002) DNA polymerase lambda, a novel DNA repair enzyme in human cells. *J. Biol. Chem.*, **277**, 13184–13191.
 36. Picher, A.J., Garcia-Diaz, M., Bebenek, K., Pedersen, L.C., Kunkel, T.A. and Blanco, L. (2006) Promiscuous mismatch extension by human DNA polymerase lambda. *Nucleic Acids Res.*, **34**, 3259–3266.
 37. Sawaya, M.R., Prasad, R., Wilson, S.H., Kraut, J. and Pelletier, H. (1997) Crystal structures of human DNA polymerase beta complexed with gapped and nicked DNA: evidence for an induced fit mechanism. *Biochemistry*, **36**, 11205–11215.
 38. Martin, M.J., Juarez, R. and Blanco, L. (2012) DNA-binding determinants promoting NHEJ by human Pol μ . *Nucleic Acids Res.*, **40**, 11389–11403.
 39. Singhal, R.K. and Wilson, S.H. (1993) Short gap-filling synthesis by DNA polymerase beta is processive. *J. Biol. Chem.*, **268**, 15906–15911.
 40. Streisinger, G., Okada, Y., Emrich, J., Newton, J., Tsugita, A., Terzaghi, E. and Inouye, M. (1966) Frameshift mutations and the genetic code. This paper is dedicated to Professor Theodosius Dobzhansky on the occasion of his 66th birthday. *Cold Spring Harb. Symp. Quant. Biol.*, **31**, 77–84.
 41. Juarez, R., Ruiz, J.F., Nick McElhinny, S.A., Ramsden, D. and Blanco, L. (2006) A specific loop in human DNA polymerase mu allows switching between creative and DNA-instructed synthesis. *Nucleic Acids Res.*, **34**, 4572–4582.
 42. Moon, A.F., Garcia-Diaz, M., Bebenek, K., Davis, B.J., Zhong, X., Ramsden, D.A., Kunkel, T.A. and Pedersen, L.C. (2007) Structural insight into the substrate specificity of DNA Polymerase mu. *Nat. Struct. Mol. Biol.*, **14**, 45–53.
 43. Kunkel, T.A. (2004) DNA replication fidelity. *J. Biol. Chem.*, **279**, 16895–16898.
 44. Bebenek, K. and Kunkel, T.A. (1990) Frameshift errors initiated by nucleotide misincorporation. *Proc. Natl Acad. Sci. USA*, **87**, 4946–4950.
 45. Kunkel, T.A. and Soni, A. (1988) Mutagenesis by transient misalignment. *J. Biol. Chem.*, **263**, 14784–14789.
 46. Bloom, L.B., Chen, X., Fyngenson, D.K., Turner, J., O'Donnell, M. and Goodman, M.F. (1997) Fidelity of Escherichia coli DNA polymerase III holoenzyme. The effects of beta, gamma complex processivity proteins and epsilon proofreading exonuclease on nucleotide misincorporation efficiencies. *J. Biol. Chem.*, **272**, 27919–27930.
 47. Efrati, E., Tocco, G., Eritja, R., Wilson, S.H. and Goodman, M.F. (1997) Abasic translesion synthesis by DNA polymerase beta violates the "A-rule". Novel types of nucleotide incorporation by human DNA polymerase beta at an abasic lesion in different sequence contexts. *J. Biol. Chem.*, **272**, 2559–2569.
 48. Romain, F., Barbosa, I., Gouge, J., Rougeon, F. and Delarue, M. (2009) Conferring a template-dependent polymerase activity to terminal deoxynucleotidyltransferase by mutations in the Loop1 region. *Nucleic Acids Res.*, **37**, 4642–4656.
 49. Lyons-Darden, T. and Topal, M.D. (1999) Effects of temperature, Mg²⁺ concentration and mismatches on triplet-repeat expansion during DNA replication in vitro. *Nucleic Acids Res.*, **27**, 2235–2240.
 50. Petruska, J., Hartenstine, M.J. and Goodman, M.F. (1998) Analysis of strand slippage in DNA polymerase expansions of CAG/CTG triplet repeats associated with neurodegenerative disease. *J. Biol. Chem.*, **273**, 5204–5210.
 51. Canceill, D., Viguera, E. and Ehrlich, S.D. (1999) Replication slippage of different DNA polymerases is inversely related to their strand displacement efficiency. *J. Biol. Chem.*, **274**, 27481–27490.
 52. Sinden, R.R. and Wells, R.D. (1992) DNA structure, mutations, and human genetic disease. *Curr. Opin. Biotechnol.*, **3**, 612–622.
 53. da Silva, E.F. and Reha-Krantz, L.J. (2000) Dinucleotide repeat expansion catalyzed by bacteriophage T4 DNA polymerase in vitro. *J. Biol. Chem.*, **275**, 31528–31535.
 54. Hartenstine, M.J., Goodman, M.F. and Petruska, J. (2002) Weak strand displacement activity enables human DNA polymerase beta to expand CAG/CTG triplet repeats at strand breaks. *J. Biol. Chem.*, **277**, 41379–41389.
 55. Kunkel, T.A., Patel, S.S. and Johnson, K.A. (1994) Error-prone replication of repeated DNA sequences by T7 DNA polymerase in the absence of its processivity subunit. *Proc. Natl Acad. Sci. USA*, **91**, 6830–6834.
 56. Stadtman, E.R. (1992) Protein oxidation and aging. *Science*, **257**, 1220–1224.
 57. Cardozo-Pelaez, F., Song, S., Parthasarathy, A., Hazzi, C., Naidu, K. and Sanchez-Ramos, J. (1999) Oxidative DNA damage in the aging mouse brain. *Mov. Disord.*, **14**, 972–980.