DNA polymerase θ promotes CAG•CTG repeat expansions in Huntington’s disease via insertion sequences of its catalytic domain

Kara Y. Chan1,3, Xueying Li, Janice Ortega1, Liya Gu, and Guo-Min Li1,2,*

From the 1Department of Radiation Oncology, University of Texas Southwestern Medical Center, Dallas, Texas, USA; 2Department of Toxicology and Cancer Biology, University of Kentucky College of Medicine, Lexington, Kentucky, USA

Edited by Patrick Sung

Huntington’s disease (HD), a neurodegenerative disease characterized by progressive dementia, psychiatric problems, and chorea, is known to be caused by CAG repeat expansions in the HD gene HTT. However, the mechanism of this pathology is not fully understood. The translesion DNA polymerase θ (Polθ) carries a large insertion sequence in its catalytic domain, which has been shown to allow DNA loop-outs in the primer strand. As a result of high levels of oxidative DNA damage in neural cells and Polθ’s subsequent involvement in base excision repair of oxidative DNA damage, we hypothesized that Polθ contributes to CAG repeat expansion while repairing oxidative damage within HTT. Here, we performed Polθ-catalyzed in vitro DNA synthesis using various CAG•CTG repeat DNA substrates that are similar to base excision repair intermediates. We show that Polθ efficiently extends (CAG)n/(CTG)n hairpin primers, resulting in hairpin retention and repeat expansion. Polθ also triggers repeat expansions to pass the threshold for HD when the DNA template contains 35 repeats upward. Strikingly, Polθ depleted of the catalytic insertion fails to induce repeat expansions regardless of primers and templates used, indicating that the insertion sequence is responsible for Polθ’s error-causing activity. In addition, the level of chromatin-bound Polθ in HD cells is significantly higher than in non-HD cells and exactly correlates with the degree of CAG repeat expansion, implying Polθ’s involvement in triplet repeat instability. Therefore, we have identified Polθ as a potent factor that promotes CAG•CTG repeat expansions in HD and other neurodegenerative disorders.

The expansion of trinucleotide repeats (TNRs) is associated with a number of neurodegenerative disorders, including Huntington’s disease (HD) and myotonic dystrophy (1, 2). Once the expansion exceeds a certain threshold, for example, 35 CAG repeats in HD, it inactivates the expression and/or alters the function of the affected genes, which leads to disease onset. Despite extensive studies, how the TNR expands is still not fully understood. Elucidating the mechanisms of TNR expansion will significantly impact the therapeutic approaches for diseases caused by this process (3, 4).

Several DNA metabolic pathways, including DNA replication and repair, have been implicated in TNR instability (1, 2, 5–10). A common feature associated with all DNA metabolic reactions is DNA breaks or free DNA ends, which can induce the formation of DNA hairpins within TNRs via strand slippage (6, 11, 12). DNA hairpin formation results in TNR expansion (1, 2, 10–12) when these hairpins are not removed (13, 14). Because cell division does not occur in the human brain, CAG/CTG repeat expansions in HD and other neurodegenerative disorders may not be related to DNA replication, but to DNA repair.

Reactive oxygen species (ROS) are a major source of DNA damage in the human brain, and guanine, which is enriched in CAG/CTG repeats, is the favored target of ROS (15). The repair of oxidative adduct 8-oxo-guanine (8-oxoG) by the 8-oxo-guanine DNA glycosylase 1 (OGG1) has been implicated in CAG/CTG repeat expansions, as depleting Ogg1 abolishes age-dependent CAG repeat expansion in HD mouse models (16). OGG1-initiated base excision repair (OGG1-BER) involves lesion recognition, flipping the lesion from the DNA double helix into the base-binding pocket (active site) of OGG1, and site-specific changes in the DNA structure (17, 18). As a bifunctional DNA glycosylase, OGG1 excises the oxidized guanine base to generate an abasic (AP) site by using its glycosylase activity (21). Alternatively, flap-endonuclease 1, which is required for long-patch BER, can remove the 5′-dRP to...
produce an ssDNA gap with 1 to 4 nucleotides, depending on the DNA polymerase involved in long-patch BER (19). Regardless, the ssDNA gap generated by BER promotes hairpin formation within CAG repeats via strand slippage (16, 22). Subsequently, Polβ fills the DNA gap by efficiently utilizing a CAG hairpin structure as a primer for DNA synthesis, which leads to CAG repeat expansion (23). Consistently, there is a close association between age-dependent somatic CAG repeat expansion and oxidative DNA damage in HD mouse models (24, 25), and both OGG1 and Polβ have been shown to promote CAG repeat expansions (16, 22, 23, 26, 27).

In addition to Polβ, DNA polymerase θ (Polθ), a low-fidelity family A DNA polymerase (28), participates in BER of oxidative DNA damage (29). Polθ is known to add single nucleotides to homopolymeric runs at a high rate (30). This is probably related to Polθ’s large insertion loops (Fig. 1A), particularly insertion 2 in the polymerase thumb domain (31). Recent structural studies have revealed that the presence of insertion 2 creates a big flexible cavity in the DNA-binding surface (32, 33). Strikingly, Polθ can extend a loop-out–containing primer (32), an activity also associated with Polβ (34). Therefore, we hypothesized that Polθ is a major polymerase that promotes CAG/CTG repeat expansions during DNA synthesis.

We performed Polθ-catalyzed DNA synthesis using a series of BER intermediates containing various numbers of CAG/CTG repeats with or without a hairpin. Like Polβ, Polθ can effectively extend hairpin primers to stabilize hairpin structures, and it induces large expansions when copying a DNA template that contains CAG/CTG repeats. However, a Polθ mutant that lacks insertion 2 does not induce CAG/CTG repeat expansions regardless of the primers and templates used, which suggests that insertion 2 is responsible for Polθ’s error-prone activity. We also found that Polθ’s chromatin level is significantly higher in HD cells than in non-HD cells and closely correlates with the degree of CAG repeat expansion, which implies the involvement of Polθ in CAG repeat expansion. Therefore, this study has identified Polθ as a potent factor that promotes the CAG/CTG repeat expansions that cause HD and other neurodegenerative diseases.

Results

Polθ extends CAG/CTG hairpin primers in vitro

To determine whether Polθ can extend a CAG or CTG hairpin primer, we conducted in vitro DNA synthesis using a limited amount of HeLa nuclear extracts supplemented with a purified 90-kD catalytic Polθ polypeptide (Fig. S1) (32), a (CAG)_5 or (CTG)_5 hairpin primer that anneals to ssM13mp18(CTG)_10 or ssM13mp18(CAG)_10 (Fig. 1B), respectively, as described previously (23). The reaction products were fractionated by denaturing PAGE, followed by Southern blot analysis using a probe that specifically
recognizes the downstream sequence near the BsrBI site of the newly synthesized strand (Fig. 1B, red bar). Thus, whether the (CAG)_5 or (CTG)_5 hairpin is removed or retained can be readily determined based on its mobility during gel electrophoresis (Fig. 1B).

Consistent with our previous observations (23), incubating the (CAG)_5 hairpin substrate with HeLa nuclear extracts supplemented with the proofreading-active Polθ, which removes the hairpin structure, generated a major product that migrates similarly to the (CAG)_10-containing DNA fragment (Fig. 1C, lane 4). Supplementing the reaction with Polβ yielded two major products: a hairpin-retained (CAG)_15 product and a hairpin-removed (CAG)_10 product (Fig. 1C, lane 6). When purified Polθ replaced Polβ or Polδ in the synthesis reaction, most yielded products were the same as those observed in the Polβ reaction (Fig. 1C, lane 5). This suggests that, like Polβ, Polθ can effectively use a CAG hairpin as a primer for DNA synthesis. In addition, a slowly migrated minor product (product I) was also observed in both the Polθ- and the Polβ-containing reactions (Fig. 1C, lanes 5 and 6), which indicates that both polymerases promote CAG repeat expansions beyond the (CAG)_5 hairpin size.

We then tested Polθ’s ability to extend a (CTG)_5 hairpin primer by using the same in vitro DNA synthesis assay. The results revealed that, in addition to the hairpin-retained product, that is, the band migrating at the same place as the (CTG)_15 band, a slowly migrated product merged as the major product, that is, the band migrating at the same place as the results revealed that, in addition to the hairpin-retained primer by using the same two major products: a hairpin-retained (CAG)_15 product and a hairpin-removed (CAG)_10 product (Fig. 1C, lane 6). When purified Polθ replaced Polβ or Polδ in the synthesis reaction, most yielded products were the same as those observed in the Polβ reaction (Fig. 1C, lane 5). This suggests that, like Polβ, Polθ can effectively use a CAG hairpin as a primer for DNA synthesis. In addition, a slowly migrated minor product (product I) was also observed in both the Polθ- and the Polβ-containing reactions (Fig. 1C, lanes 5 and 6), which indicates that both polymerases promote CAG repeat expansions beyond the (CAG)_5 hairpin size.

We next examined Polθ’s ability to synthesize CAG/CTG repeats by conducting in vitro DNA synthesis using a defined DNA synthesis system (Fig. 3A) that contains purified Polθ, proliferating cellular nuclear antigen (PCNA) and replication factor C (RFC), a 5’-32P-labeled nonhairpin primer, and M13mp18 derivatives with various lengths of CAG/CTG repeats in the presence or absence of replication protein A, an ssDNA-binding protein that protects ssDNA from nuclease attack and secondary structure formation (36). We observed no expanded DNA products in reactions with high-fidelity polymerases Polδ and T7, regardless of the DNA templates used, that is, (CAG)_20, (CTG)_20, (CAG)_25, (CTG)_25, (CAG)_35, (CTG)_35, or random DNA sequences (Fig. 3, B–D). PolθΔi2-generated products were essentially the same as those produced by Polδ or T7 polymerase in all reactions (Fig. 3, B–D), which further suggests that insertion 2 is indeed responsible for Polθ-mediated CAG/CTG repeat expansions.

**Polθ’s insertion 2 is responsible for CAG/CTG repeat expansion**

Polθ’s large insertion 2 is responsible for the polymerase’s error-prone nature during DNA synthesis (30, 32, 35). To explore the impact of insertion 2 on the CAG/CTG expansion by Polθ that we observed, we generated an insertion 2–deleted Polθ mutant, as described (35). We examined the mutant Polθ for its ability to synthesize (CAG)_5- or (CTG)_5-primed DNA. Remarkably, no product I was generated by the insertion 2–deleted Polθ (PolθΔi2), regardless of whether a (CAG)_5 or (CTG)_5 hairpin primer was used (Fig. 2A, lanes 6 and 13). In addition, the amount of the hairpin-retained product, that is, (CAG)_15 or (CTG)_15, was much lower in the reactions that contained PolθΔi2 (Fig. 2A). PolθΔi2 only generated 8.9% of the (CAG)_5 hairpin–retained and 14.9% of the (CTG)_5 hairpin–retained products produced by Polθ (Fig. 2A, compare lane 5 with lane 6, and lane 12 with lane 13). Quantitative analysis shows that this reduction in the hairpin-retained and expanded products in PolθΔi2–catalyzed reactions is highly significant (Fig. 2B). These results suggest that insertion 2 is indeed responsible for Polθ-mediated CAG/CTG repeat expansions.

**Polθ-mediated CAG/CTG repeat expansion depends on the repeat length**

We next examined Polθ’s ability to synthesize CAG/CTG repeats by conducting in vitro DNA synthesis using a defined DNA synthesis system (Fig. 3A) that contains purified Polθ, proliferating cellular nuclear antigen (PCNA) and replication factor C (RFC), a 5’-32P-labeled nonhairpin primer, and M13mp18 derivatives with various lengths of CAG/CTG repeats in the presence or absence of replication protein A, an ssDNA-binding protein that protects ssDNA from nuclease attack and secondary structure formation (36). We observed no expanded DNA products in reactions with high-fidelity polymerases Polδ and T7, regardless of the DNA templates used, that is, (CAG)_20, (CTG)_20, (CAG)_25, (CTG)_25, (CAG)_35, (CTG)_35, or random DNA sequences (Fig. 3, B–D). PolθΔi2–generated products were essentially the same as those produced by Polδ or T7 polymerase in all reactions (Fig. 3, B–D), which further suggests that insertion 2 is mutagenic.
Polymerase θ promotes CAG•CTG repeat expansion

However, Polθ-catalyzed products differed from those produced by Polθ, T7, and PolθΔι2. First, in addition to full-length products, Polθ generated a product that is one nucleotide larger than the full-length band in essentially all reactions (see red asterisks in Fig. 3). This one-nucleotide–larger product appears to be unrelated to CAG/CTG repeats, as also seen in reactions with the M13mp18 template. This product is probably derived from Polθ’s terminal deoxynucleotidyl transferase (TdT) activity (37), which catalyzes the incorporation of single deoxynucleotides into the 3’-OH terminus of ssDNA or dsDNA. Second, when M13mp18–(CAG)35 or M13mp18–(CTG)35 was used as a template for DNA synthesis, we detected several expanded products in Polθ-catalyzed reactions (Fig. 3D), especially for the M13mp18(CTG)35 template (Fig. 3D, red bracket). This is probably because a CTG hairpin is easier to form and more stable than a CAG hairpin (11, 23, 38).

To determine the nature of the expansion, we recovered a major Polθ-expanded band (Fig. 3D, red arrow) and a T7-generated full-length band (Fig. 3D, red circle) from the gel. The DNA samples were PCR-amplified, cloned into a vector, and transfected into a bacterial strain. Plasmid DNA samples were isolated from six clones derived from the T7-catalyzed product and 11 clones derived from the Polθ-catalyzed band before DNA-sequencing analysis. The results reveal that in six T7-derived clones, four of them showed original 35 repeats (Fig. S2A), and two exhibited 33 repeats, but in 11 Polθ-expanded clones, one clone contained 40 repeats (Fig. S2B), seven showed 35 repeats, and three demonstrated 34 repeats. These results are consistent with a previous study, which shows that while the majority of (CTG)180 clones undergo contractions after replicating for 100 generations, ~20% of these clones still remain the full-length repeats (39). We therefore conclude that the 40 repeats represent the original length of the Polθ-expanded band, that is, an expansion of five repeats by Polθ during DNA synthesis (Fig. S2C). We did not observe this repeat length–dependent expansion by Polθ in reactions with PolθΔι2 (Fig. 3D), which suggests that insertion 2 is also responsible for Polθ-induced repeat length–dependent expansion. However, the mechanism by which Polθ induces repeat expansions in a repeat length–dependent manner is unclear. Notably, DNA substrates used in this assay were gap-containing molecules, and the primer contained no repeat sequences and no hairpin. Thus, Polθ used an error-free primer to initiate DNA synthesis. When it fills a short repeat gap, for example, 15 or 20 repeats, Polθ may have sufficient processivity to quickly finish synthesizing the repeats and reach a nonrepeat template, which provides less of a chance for Polθ to make errors. When Polθ encounters a template with a large repeat number, such as 35 repeats, which is challenging for all DNA polymerases, it may have to pause a couple of times, and this may allow a repeat hairpin to form, leading to repeat expansion. Future studies are required to test these possibilities.

Mn²⁺ stimulates Polθ’s error-prone activity for repeat expansion

Divalent metal ion Mn²⁺ alters the structural flexibility of polymerase active sites in favor of error-prone synthesis (40–43). To further determine whether the catalytic flexibility of the Polθ active site is responsible for the observed CAG/CTG repeat instability, we performed an in vitro DNA synthesis assay using oligonucleotide templates that contained 20 or 35 CTG repeats in the presence of Mg²⁺ or Mn²⁺. Although Polθ’s TdT activity could only add one extra nucleotide when the (CTG)20 template was used for synthesis in the presence of Mg²⁺ (Fig. 4A, lane 12, red asterisk), it incorporated multiple nucleotides in the presence of Mn²⁺ (Fig. 4A, lanes 15 and 16, red bracket). We found that Mn²⁺ could also induce Polθ to generate a product that is one nucleotide (1-nt) bigger than the full-length product (Fig. 4A, lanes 7 and 8). We observed the
Mn$^{2+}$ stimulates Pol$\theta$ generated very few expanded products in the presence of Mn$^{2+}$. To conduct repeat expansions, we synthesized long triplet repeats (Figs. 3 and 4A). To confirm this, we examined Pol$\theta$’s ability to replicate the (CTG)$_{35}$ template in the presence of Mg$^{2+}$ or Mn$^{2+}$. The results show that, although Pol$\delta$ could generate a limited amount of full-length product in the presence of Mg$^{2+}$ (Fig. 4B, lanes 9–12), it failed to produce any products in the presence of Mn$^{2+}$ (Fig. 4B, lanes 13–16). Taken together, these results suggest that large repeat expansions can only be carried out by polymerases with a large insertion in their catalytic domain, such as Pol$\theta$, and that Mn$^{2+}$ further enhances the error-prone property of Pol$\theta$, but not of polymerases with no insertion in their active sites, such as Pol$\delta$ and Pol$\theta$Δi2.

**Polymerase $\theta$ promotes CAG•CTG repeat expansion**

The data presented above strongly suggest that Pol$\theta$ could be an important driving factor for CAG/CTG repeat expansion in HD. To explore this possibility, we set out to determine Pol$\theta$’s expression and chromatin binding in cells derived from individuals with and without HD. We obtained seven fibroblast cell lines derived from patients with HD (GM04208, GM04212, GM04210, GM04220, GM04230, GM21756, and GM09197) and two non-HD cell lines (GM04204 and GM02153) from the Coriell Institute for Medical Research. The first five HD lines were from the same family, with GM04204 being an unaffected family member (44). To ensure the cell lines’ HD status, we amplified their $HTT$ exon 1 sequence that contained CAG repeat sequences by PCR and analyzed the resulting products in a polyacrylamide gel (Fig. 5A). We then calculated the CAG repeat numbers based on the migration distances of known CAG repeat numbers in HD cell lines, as the CAG repeat number ($y$) is a linear function of the logarithm of its migration distance ($x$) in gel electrophoresis:

$$y = -193.37x + 191.44$$

The known CAG repeat numbers are 18/17 for GM04204, 45/19 for GM04230, 44/21 for GM04208, 70/15 for GM21756, 180/19 for GM09197, and 32/16 for GM02153 (Table S1). This analysis indeed confirmed the cell lines’ HD origin, with GM04212, GM04210, and GM04220 carrying an allele of at least 40 CAG repeats (Fig. 5B and Table S1). We noted that the 180-CAG repeat allele of GM09197 was not detectable in Figure 5A. This is probably because the large repeat number was hard to amplify, as only a small amount of the 70-CAG repeat product was obtained (Fig. 5A).

We then determined Pol$\theta$ expression in these cell lines by Western blotting. In general, all cells expressed abundant Pol$\theta$ (Fig. 5C, upper panel), but Pol$\theta$ levels in whole-cell lysates (WCLs) in several HD cells were significantly higher than in non-HD control GM04204 (Fig. 5D). Strikingly, all HD cells exhibited a chromatin-bound Pol$\theta$ level significantly higher than that in non-HD control GM04204 (Fig. 5C, lower panel, and Fig. 5E), which suggests that Pol$\theta$ is efficiently recruited to chromatin in HD cells. For example, although GM04204 (Fig. 5C, lane 1) displayed an overall Pol$\theta$ level slightly lower than that of GM04210 (Fig. 5C, lane 4) and GM04220 (Fig. 5C, lane

**Pol$\theta$ is highly chromatin enriched in cell lines derived from patients with HD**
Polymerase θ promotes CAG•CTG repeat expansion

Figure 5. Polθ’s chromatin levels correlate with HTT CAG repeats. A, fractionation of the PCR products of HTT CAG repeat numbers of HD and non-HD cell lines by denaturing PAGE. PCR products were visualized by Southern blot analysis using a 32P-labeled (CTG)5 probe. B, determination of CAG repeat numbers by using a unique linear function (y = −193.37x + 191.44) between the CAG repeat number (y) and the logarithmic value of its migration distance. Cells lines with known or calculated CAG repeat numbers are shown in blue and red, respectively. C, determination of Polθ levels in whole-cell lysates (WCL) and chromatin by Western blotting. D and E, quantification of Polθ levels in WCL (D) and chromatin (E) in HD and non-HD cell lines. *p < 0.05, **p < 0.01, and ***p < 0.001. HD, Huntington’s disease; Polθ, DNA polymerase θ.

Discussion

We show here that Polθ promotes CAG repeat instability in multiple ways. Polθ can extend a CAG repeat hairpin, which results in hairpin retention, and, in turn, CAG repeat expansion (Figs. 1 and 2). Polθ also induces large expansions when it copies a DNA template that contains 35 CAG repeats. This error-prone property of Polθ is due to its insertion 2, as PolθΔi2 cannot induce repeat expansions regardless of the primers or templates used (Fig. 3). Mn2+, which induces flexibility in polymerases‘ active sites, further stimulates Polθ’s error-prone activity (Fig. 4). Therefore, we have identified Polθ as a potent factor in promoting CAG repeat expansions in HD and other neurodegenerative diseases.

Polθ’s involvement in CAG/CTG repeat expansion appears to be well justified. First, Polθ participates in BER (29), a DNA repair pathway implicated in CAG repeat instability (16). The brains of patients with HD are characterized by high levels of ROS (47), which can induce 8-oxoG adducts in guanine-rich CAG repeats (15). Repairing 8-oxoG requires the OGG1 glycosylase, which, together with the AP endonuclease, generates a small gap in the damaged DNA strand. This gap is then filled by Polθ or Polβ because both enzymes are required to repair oxidative damage (29). Because CAG repeats are prone to hairpin formation in the presence of a strand break/gap, a

lane 5), the former showed much less chromatin-bound Polθ than the latter ones (Fig. 5E). These results indicate a close association between Polθ chromatin binding and CAG repeat expansion. Interestingly, although GM02153 only carried an HTT with 32 CAG repeats, its chromatin-bound Polθ level is essentially the same as in several HD cell lines (Fig. 5C). This may explain why some patients display HD symptoms although their CAG repeat number is lower than the threshold 35 (45, 46). Thus, Polθ’s chromatin level could be a reliable hallmark for HD diagnosis.
hairpin formed within the repeats can be utilized as a primer for DNA synthesis by Polθ (this study) or Polβ (23, 34), which leads to hairpin retention and large repeat expansion. In addition, Polθ possesses a large insertion 2 in the thumb domain (31), which creates a large flexible cavity in the DNA-binding surface (32, 33). This cavity facilitates misalignments between the primer and template during DNA synthesis (32), which leads to the addition of nucleotides to homopolymeric runs (30, 32). We showed that Polθ uses its large catalytic domain for CAG/CTG hairpin-priming synthesis (Figs. 1 and 2) and for large repeat expansions without a pre-existing hairpin structure in a repeat length–dependent manner (Fig. 3). Depleting insertion 2 makes Polθ a relatively high-fidelity polymerase, as PolθΔi2 no longer synthesizes error-containing DNA (Figs. 3 and 4). Finally, Polθ is highly chromatin-enriched in HD cells (Fig. 5). This gives Polθ the advantage of participating in DNA metabolic reactions, including the repair of oxidative DNA damage, so it can stimulate CAG repeat expansions in HD cells. However, how Polθ is preferentially recruited to chromatin in HD cells remains to be investigated.

Reports about whether people with an HTT gene that contains 27 to 35 CAG repeats are at risk for developing HD are controversial. Although CAG repeats in the 27 to 35 range are highly stable (48), some patients had the repeats extend into the HD onset range (>36), and some displayed HD symptoms although their CAG repeats remained unchanged (45, 46). This suggests that a portion of individuals with 27 to 35 intermediate CAG repeats can develop HD with or without the repeats expanding to ≥36. However, biomarkers that indicate HD onset and mechanisms that extend the repeats beyond the HD threshold are unknown. Our data presented here provide possible explanations for both questions. We showed that cells derived from patients with HD had significantly higher levels of Polθ in chromatin than cells from non-HD controls (Fig. 5). Surprisingly, we also found that, despite carrying an intermediate allele of 32 CAG repeats and being considered a non-HD control, the GM02153 cell line showed chromatin-bound Polθ levels similar to those of HD cells, rather than those of non-HD cells (Fig. 5). This suggests that the patient from whom GM02153 was derived might exhibit clinical features of HD. Furthermore, because Polθ carries out repeat length–dependent expansion, the polymerase can expand the intermediate (i.e., 27–35) repeats to 36 and beyond. Therefore, Polθ concentration, particularly the chromatin-bound level, could be an important marker to determine the risk of HD for individuals with intermediate HTT alleles.

Based on published data and the results presented here, we propose a model to elucidate the mechanism by which Polθ promotes CAG repeat expansions in nondividing neural cells during OGG1-mediated repair of 8-oxoG (Fig. 6). Once BER generates a strand break or a small gap within the CAG repeats, a CAG hairpin structure forms via strand slippage.

**Figure 6. Model of Polθ-catalyzed CAG repeat expansion.** OGG1-mediated repair of oxidative lesions within CAG repeats generates a strand break, which triggers the formation of a hairpin via strand slippage. Repair DNA synthesis is initiated when Polθ is recruited to the hairpin or nick. In the latter case, Polθ utilizes the hairpin as a primer for extension, leading to hairpin retention and repeat expansion. In the former case, Polθ incorporates nucleotides to the 3' end of the nick. However, the newly incorporated nucleotides are trapped into the packet of insertion 2 (Ins2) to form a loop-out or a hairpin, which can grow into a big hairpin or shift out of insertion 2. The continuous synthesis by Polθ can trap a new hairpin formation in insertion 2, resulting in large expansion of the CAG repeats. OGG1, 8-oxo-guanine DNA glycosylase 1; Polθ, DNA polymerase θ.
Polymerase θ promotes CAG•CTG repeat expansion

Polθ can use the resulting hairpin as a primer for error-prone synthesis. Alternatively, Polθ incorporates nucleotides to the 3′ end of the nick/gap, and the newly incorporated nucleotides enter into the insertion 2-formed packet to form a loop-out, which can then easily convert into a (CAG)n hairpin. The hairpin can either grow further or move out of the insertion 2 packet as the polymerase continues to synthesize. In the latter case, additional hairpins can form, which leads to variously expanded CAG repeats. However, the model, which is mainly based on our biochemical studies, requires thorough biological investigations in model organisms. For example, gene knockdown or KO of Polθ, Polβ, and/or other DNA polymerases may provide definitive evidence of how these polymerases promotes CAG repeat instability.

Experimental procedures

Cells and cell culture

Human fibroblast cell lines GM04204, GM04210, GM04230, GM04212, GM04208, GM04220, GM21756, GM09197, and GM02153 were purchased from the Coriell Institute for Medical Research and cultured in Minimum Essential Medium (v/v) ice-cold 0.2 M HCl for 10 min on ice to denature and disassociate proteins from chromatin. The acidic solution was neutralized with 1:1 (v/v) of 1 M Tris-HCl (pH 8.0):1 U/μl Benzonase Nuclease (Sigma) and incubated on ice for 1 h to allow DNA digestions. The digested genomic DNA and insoluble proteins were removed by centrifugation. WCLs were prepared by resuspending cells in 1:1 (v/v) ice-cold 0.2 M HCl treated as chromatin fractions. Lysates were clarified by high-speed centrifugation (21,000g) for 15 min at 4 °C. Protein concentrations were determined by using the Bio-Rad Bradford Protein Assay reagent. After electrophoresis, chromatin-binding proteins and WCLs were analyzed for Polθ by Western blotting using a Polθ-specific antibody (Novus). Polθ bands were visualized and quantified by using a Bio-Rad ChemiDoc Imaging system.

Nuclear extract and protein preparation

HeLa nuclear extracts were prepared as previously described (23). DNA sequences coding for human Polθ polymerase catalytic domain (Fig. 1A) and its insertion 2-deleted derivative PolθΔi2 were cloned into pLEXm and designated pLEXm-Polθ and pLEXm-PolθΔi2, respectively. pLEXm-PolθΔi2 was derived from pSUMO3-PolθΔi2 (a generous gift from Dr Richard T. Pomerantz, Temple University) by inserting the (23). DNA sequences coding for human PolθΔi2, respectively. pLEXm-PolθΔi2 was expressed in HEK 293T cells and HEK 293GnTI cells were grown in suspension in FreeStyle 293 Expression Medium (Gibco) supplemented with 1% FBS. All human cell lines were cultivated at 37 °C in a humidified atmosphere containing 5% CO₂.

Hairpin primer extension assay

Unless otherwise mentioned, hairpin primer extension was assayed by Southern blot hybridization as described previously (23). Oligonucleotides that contained 15 CTG or 15 CAG repeats were annealed with ssDNA of M13mp18-(CAG)₁₀ or M13mp18-(CTG)₁₀ (51) to form CTG and a CAG hairpin substrate, respectively (Fig. 1B). Individual hairpin substrates were incubated with 30 μg HeLa nuclear extracts in the presence or absence of a polymerase (δ, θ, or θΔi2) for DNA synthesis at 37 °C for 30 min in a 40-μl reaction containing 110 mM KCl, 20 mM Tris-HCl, pH 7.6, 5 mM MgCl₂, 1.5 mM ATP, 0.1 mM of various dNTP, and 0.05 mg/ml BSA. In the purified system, each reaction contained RFC (110 fmol) and PCNA (2 pmol) in addition to the indicated polymerase (0.1 μM). The resulting products were digested with BsrBI before electrophoresis through a 10% denaturing polyacrylamide gel, followed by Southern blotting using a 32P-labeled probe as described (23). The products were detected by an Amersham Typhoon phosphor imager.

PCR amplification of CAG repeats and DNA sequencing analysis

PCR amplification of CAG repeats located in exon 1 of HTT in HD cell lines used forward (5′-ATGAGGGCCTTC-GAGTCCCTCAAGTCCTTC-3′) and reverse (5′-CGGAGG-CAGGCAGGG CTG TGCCCTGC-3′) primers. PCRs were performed in a volume of 25 μl containing 100 ng genomic DNA, 1.6 mM of each dNTP, 4 pmol of each primer, and 0.5 U Q5 DNA polymerase (New England Biolabs). After an initial denaturation of 4 min at 98 °C, 40 cycles of 45 s at 98 °C, 1 min at 68 °C, and 3 min at 72 °C were carried out, followed by a final extension of 10 min. PCR products were resolved in a 7.5% (wt/vol) denaturing polyacrylamide gel, followed by Southern blot analysis using a 32P-labeled (CTG)₅ oligonucleotide probe. PCR products were then visualized by an Amersham Typhoon phosphor imager as described (49).

To analyze the expanded CTG repeats derived from Polθ-catalyzed primer extension, DNA bands (32P-labeled) were excised and eluted from gels, then PCR-reamplified for 35 cycles by REDTaq DNA polymerase (Sigma-Aldrich) using primers 5′-ACGTGTTGAAACAGCAGGGCCA-3′ (forward) and 5′-CATGATTACGAATTC-3′ (reverse), essentially as described above. PCR products were cloned into a pGEM-T vector (Promega) and transfected into E. coli DH5-Alpha.
Polymerase θ promotes CAG•CTG repeat expansion

2. Mirkin, S. M. (2007) Expandable DNA repeats and human disease. Nature 447, 932–940
3. Kristich, A. N., and Mirkin, S. M. (2020) On the wrong DNA track: Molecular mechanisms of repeat-mediated genome instability. J. Biol. Chem. 295, 4134–4170
4. Schmidt, M. H., and Pearson, C. E. (2016) Disease-associated repeat instability and mismatch repair. DNA Repair 38, 117–126
5. Liu, G., Chen, X., Bissler, J. I., Sinden, R. R., and Lefkaj, M. (2010) Replication-dependent instability at (CTG)x (CAG) repeat hairpins in human cells. Nat. Chem. Biol. 6, 652–659
6. 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
7. McMurray, C. T. (2010) Mechanisms of trinucleotide repeat instability during human development. Nat. Rev. Genet. 11, 786–799
8. Usdin, K., and Grabczyk, E. (2000) DNA repeat expansions and human disease. Cell Mol. Life Sci. 57, 914–931
9. Tian, L., Hou, C., Tian, K., Holcomb, N. C., Gu, L., and Li, G. M. (2009) Mismatch recognition protein MutSbeta does not hijack (CAG)n hairpin repair in vitro. J. Biol. Chem. 284, 20452–20456
10. Guo, J., Chen, L., and Li, G. M. (2017) DNA mismatch repair in trinucleotide repeat instability. Sci. China Life Sci. 60, 1087–1092
11. Gacy, A. M., Goellner, G., Juranic, N., Macura, S., and McMurray, C. T. (1995) Trinucleotide repeats that expand in human disease form hairpin structures in vitro. Cell 81, 533–540
12. Gacy, A. M., and McMurray, C. T. (1998) Influence of hairpins on template renaturation at trinucleotide repeat duplexes: A model for slipped DNA. Biochemistry 37, 9426–9434
13. Hou, C., Chan, N. L., Gu, L., and Li, G. M. (2009) Incision-dependent and error-free repair of (CAG)(n)/(CTG)(n) hairpins in human cell extracts. Nat. Struct. Mol. Biol. 16, 869–875
14. Panighali, G. B., Lau, R., Montgomery, S. E., Leonard, M. R., and Pearson, C. E. (2005) Slipped (CTG)•(CAG) repeats can be correctly repaired, escape repair or undergo error-prone repair. Nat. Struct. Mol. Biol. 12, 654–662
15. Madahushi, R., Pan, L., and Tsai, L. H. (2014) DNA damage and its links to neurodegeneration. Neuron 83, 266–282
16. Kottuv, 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
17. Bruner, S. D., Norman, D. P., and Verdine, G. L. (2000) Structural basis for recognition and repair of the endogenous mutagen 8-oxoguanine in DNA. Nature 403, 859–866
18. Li, G. M. (2010) Novel molecular insights into the mechanism of GO removal by MutM. Cell Res. 20, 116–118
19. Friedberg, E. C., Walker, G. C., Siede, W., Wood, R. D., Schultz, R. A., and Ellenberger, T. (2006) DNA Repair and Mutagenesis. ASM Press, Washington, DC.
20. Krishnamurthy, N., Haraguchi, K., Greenberg, M. M., and David, S. S. (2008) Efficient removal of formamidopyrimidines by 8-oxoguanine glycosylases. Biochemistry 47, 1083–1093
21. Prasad, R., Shock, D. D., Beard, W. A., and Wilson, S. H. (2010) Substrate channeling in mammalian base excision repair pathways: Passing the baton. J. Biol. Chem. 285, 40479–40488
22. 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
23. Chan, N. L., Guo, J., Zhang, T., Mao, G., Hou, C., Yuan, F., Huang, J., Zhang, Y., Wu, J., Gu, L., and Li, G. M. (2013) Coordinated processing of 3′ slipped (CAG)n/(CTG)n hairpins by DNA polymerases beta and delta preferentially induces repeat expansions. J. Biol. Chem. 288, 15015–15022
24. 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
25. Kennedy, L., and Shelbourne, P. F. (2000) Dramatic mutation instability in HD mouse striatum: Does polyglutamine load contribute to cell-specific vulnerability in Huntington’s disease? Hum. Mol. Genet. 9, 2539–2544
Polymerase θ promotes CAG•CTG repeat expansion

26. Crespan, E., Hubscher, U., and Maga, G. (2015) Expansion of CAG triplet repeats by human DNA polymerases lambda and beta in vitro, is regulated by flap endonuclease 1 and DNA ligase 1. *DNA Repair* 29, 101–111

27. Liu, Y., and Wilson, S. H. (2012) DNA base excision repair: A mechanism of trinucleotide repeat expansion. *Trends Biochem. Sci.* 37, 162–172

28. Wood, R. D., and Doublet, S. (2016) DNA polymerase theta (POLQ), double-strand break repair, and cancer. *DNA Repair* 44, 22–32

29. Yoshimura, M., Kohzaki, M., Nakamura, J., Asagoshi, K., Sonoda, E., Hou, E., Prasad, R., Wilson, S. H., Tano, K., Yasui, A., Lan, L., Seki, M., Wood, R. D., Arakawa, H., Bauerstedde, I. M., et al. (2006) Vertebrate POLQ and POLbeta cooperate in base excision repair of oxidative DNA damage. *Mol. Cell* 24, 115–125

30. Arana, M. E., Seki, M., Wood, R. D., Rogozin, I. B., and Kunkel, T. A. (2008) Low-fidelity DNA synthesis by human DNA polymerase theta. *Nucleic Acids Res.* 36, 3847–3856

31. Hogg, M., Seki, M., Wood, R. D., Doublet, S., and Wallace, S. S. (2011) Lesion bypass activity of DNA polymerase theta (POLQ) is an intrinsic property of the pol domain and depends on unique sequence inserts. *J. Biol. Chem.* 286, 642–652

32. Lee, Y. S., Gao, Y., and Yang, W. (2015) How a homolog of high-fidelity DNA polymerases conducts mutagenic DNA synthesis. *Nat. Struct. Mol. Biol.* 22, 298–303

33. Zahn, K. E., Averill, A. M., Aller, P., Wood, R. D., and Doublet, S. (2015) Human DNA polymerase theta grasps the primer terminus to mediate DNA repair. *Nat. Struct. Mol. Biol.* 22, 304–311

34. Guo, J., Gu, L., Leffak, M., and Li, G. M. (2016) MutSbeta promotes trinucleotide repeat expansion by recruiting DNA polymerase beta to nascent (CAG)n or (CTG)n hairpins for error-prone DNA synthesis. *Cell Res.* 26, 775–786

35. Kent, T., Chandramouly, G., McDevitt, S. M., Ozdemir, A. Y., and Pomerantz, R. T. (2015) Mechanism of microhomology-mediated end-joining by human polymerase theta. *Nat. Struct. Mol. Biol.* 22, 230–237

36. Ramilo, C., Gu, L., Guo, S., Zhang, X., Patrick, S. M., Turchi, J. J., and Li, G. M. (2002) Partial reconstitution of human DNA mismatch repair in vitro: Characterization of the role of human replication protein A. *Mol. Cell. Biol.* 22, 2037–2046

37. Kent, T., Mateos-Gomez, P. A., Sleir, A., and Pomerantz, R. T. (2016) Polymerase theta is a robust terminal transferase that oscillates between three different mechanisms during end-joining. *Elife* 5, e13740

38. Pearson, C. E., Tam, M., Wang, Y. H., Montgomery, S. E., Dar, A. C., Cleary, J. D., and Nichol, K. (2002) Slipped-strand DNAs formed by long (CAG)n repeats: Slipped-out repeats and slip-out junctions. *Nucleic Acids Res.* 30, 4534–4547

39. Jaworski, A., Rosche, W. A., Gellibolan, R., Kang, S., Shimizu, M., Bowater, R. P., Sinden, R. R., and Wells, R. D. (1995) Mismatch repair in Escherichia coli enhances instability of (CTG)n triplet repeats from human hereditary diseases. *Proc. Natl. Acad. Sci. U. S. A.* 92, 11019–11023

40. Poranen, M. M., Salgado, P. S., Koivunen, M. R., Wright, S., Barnford, D. H., Stuart, D. I., and Grimes, J. M. (2008) Structural explanation for the role of Mn2+ in the activity of phiλ RNA-dependent RNA polymerase. *Nucleic Acids Res.* 36, 6633–6644

41. Vashistha, A. K., Wang, J., and Konigsberg, W. H. (2016) Different divalent cations alter the kinetics and fidelity of DNA polymerases. *J. Biol. Chem.* 291, 20869–20875

42. Frank, E. G., and Woodgate, R. (2007) Increased catalytic activity and altered fidelity of human DNA polymerase iota in the presence of manganese. *J. Biol. Chem.* 282, 24689–24696

43. Wu, W.-I., Yang, W., and Tsai, M.-D. (2017) How DNA polymerases catalyse replication and repair with contrasting fidelity. *Nat. Rev. Chem.* 1, 0068

44. Gussella, J. F., Wexler, N. S., Conneally, P. M., Naylor, S. L., Anderson, M. A., Tanzi, R. E., Watkins, P. C., Ottina, K., Wallace, M. R., Sakaguchi, A. Y., Young, A. B., Shoulson, I., Bonilla, E., and Martin, J. B. (1983) A polymorphic DNA marker genetically linked to Huntington's disease. *Nature* 306, 234–238

45. Ha, A. D., and Jankovic, J. (2011) Exploring the correlates of intermediate CAG repeats in Huntington disease. *Postgrad. Med.* 123, 116–121

46. Semaka, A., Collins, J. A., and Hayden, M. R. (2010) Unstable familial transmissions of Huntington disease alleles with 27–35 CAG repeats (intermediate alleles). *Am. J. Med. Genet. B Neuropsychiatr. Genet.* 153B, 314–320

47. Kumar, A., and Ratan, R. R. (2016) Oxidative stress and Huntington's disease: The good, the bad, and the ugly. *J. Huntington's Dis.* 5, 217–237

48. Brocklebank, D., Gayan, J., Andresen, J. M., Roberts, S. A., Young, A. B., Snodgrass, S. R., Penney, J. B., Ramos-Arroyo, M. A., Cha, J. J., Rosas, H. D., Hersch, S. M., Feigin, A., Cherry, S. S., Wexler, N. S., Housman, D. E., et al. (2009) Repeat instability in the 27–39 CAG range of the HD gene in the Venezuelan kindreds: Counseling implications. *Am. J. Med. Genet. B Neuropsychiatr. Genet.* 150B, 425–429

49. Ortega, J., Lee, G. S., Gu, L., Yang, W., and Li, G. M. (2021) Mispair-bound human MutS-MutL complex triggers DNA incisions and activates mismatch repair. *Cell Res.* 31, 542–553

50. Zhang, Y., Yuan, F., Presnell, S. R., Tian, K., Gao, Y., Tomkinson, A. E., Gu, L., and Li, G. M. (2005) Reconstitution of 5'-directed human mismatch repair in a purified system. *Cell 122*, 693–705

51. Zhang, T., Huang, J., Gu, L., and Li, G. M. (2012) *In vitro* repair of DNA hairpins containing various numbers of CAG/CTG trinucleotide repeats. *DNA Repair* 11, 201–209

52. Fang, J., Huang, Y., Mao, G., Yang, S., Rennert, G., Gu, L., Li, H., and Li, G. M. (2018) Cancer-driving H3K34V/R/D mutations block H3K36 methylation and H3K36me3-MutSalpha interaction. *Proc. Natl. Acad. Sci. U. S. A.* 115, 9598–9603

53. Cohen-Carmon, D., Sorek, M., Lerner, V., Divya, M. S., Nissim-Rafinia, M., Yaron, Y., and Mesheror, E. (2020) Progerin-induced transcriptional changes in Huntington's disease human pluripotent stem cell-derived neurons. *Mol. Neurobiol.* 57, 1768–1777

54. de Mezer, M., Wojciechowska, M., Napierala, M., Sobczak, K., and Krzyziowski, W. J. (2011) Mutant CAG repeats of Huntington transcript fold into hairpins, form nuclear foci and are targets for RNA interference. *Nucleic Acids Res.* 39, 3852–3863

55. Victor, M. B., Richner, M., Olsen, H. E., Lee, S. W., Monteys, A. M., Ma, C., Huh, C. J., Zhang, B., Davidson, B. L., Yang, X. W., and Yoo, A. S. (2018) Striatal neurons directly converted from Huntington disease: The good, the bad, and the ugly. *Elife* 7, e13740

56. Victor, M. B., Richner, M., Olsen, H. E., Lee, S. W., Monteys, A. M., Ma, C., Huh, C. J., Zhang, B., Davidson, B. L., Yang, X. W., and Yoo, A. S. (2018) Striatal neurons directly converted from Huntington's disease patient fibroblasts recapitulate age-associated disease phenotypes. *Nat. Neurosci.* 21, 341–352

57. Ferlazzo, M. L., Sonzogni, L., Granzotto, A., Bodgi, L., Lartin, O., Devic, C., Vogin, G., Pereira, S., and Foray, N. (2014) Mutations of the Huntington disease protein impact on the ATM-dependent signaling and repair pathways of the radiation-induced DNA double-strand breaks: corrective effect of statins and bisphosphonates. *Mol. Neurobiol.* 49, 1200–1211

58. Mollica, P. A. (2015) DNA Repair Deficiency in Huntington's Disease Fibroblasts and Induced Pluripotent Stem Cells. Doctor of Philosophy (PhD), Dissertation, Biological Sciences, Old Dominion University