The Mechanism of Replication

Stalling and Recovery within Repetitive DNA

Corella S Casas-Delucchi¹, Manuel Daza-Martin¹, Sophie L Williams and Gideon
Coster*

Genome Replication lab, Division of Cancer Biology, Institute of Cancer Research,
Chester Beatty Laboratories, 237 Fulham Road, London SW3 6JB, UK

¹ Contributed equally

* Lead contact: Gideon.coster@icr.ac.uk

CosterLab.com | Twitter: [@CosterLab](https://twitter.com/CosterLab) | +442034377582

ORCID: [0000-0002-1449-6655](https://orcid.org/0000-0002-1449-6655)

SUMMARY

Accurate chromosomal DNA replication is essential to maintain genomic stability. Genetic evidence suggests that certain repetitive sequences impair replication, yet the underlying mechanism is poorly defined. Replication could be directly inhibited by the DNA template or indirectly, for example by DNA-bound proteins. Here, we reconstituted replication of mono-, di- and trinucleotide repeats in vitro using eukaryotic replisomes assembled from purified proteins. We found that structure-prone repeats are sufficient to impair replication. Whilst template unwinding was unaffected, leading strand synthesis was inhibited, leading to fork uncoupling. Synthesis through hairpin-forming repeats relied on replisome-intrinsic mechanisms, whereas synthesis of quadruplex-forming repeats required an extrinsic accessory helicase. DNA-induced fork stalling was mechanistically similar to that induced by leading strand DNA lesions, highlighting structure-prone repeats as an important potential source of replication stress. Thus, we propose that our understanding of the cellular response to replication stress also applies to stalling induced by repetitive sequences.

KEY WORDS

DNA replication; Genome stability; Replisome; Polymerase; Repetitive DNA; DNA secondary structures; Replication fork uncoupling

INTRODUCTION

Faithful and accurate chromosomal DNA replication is a fundamental process that is required to maintain genome stability and is performed by a multi-protein complex termed the replisome (Bell and Labib, 2016). The replisome encounters various types of challenges, including DNA damage, DNA-bound proteins, collisions with the transcriptional machinery, RNA-DNA hybrids (R-loops), topological stress and limiting dNTPs (Gaillard et al., 2015). Under unperturbed conditions, leading strand synthesis is coupled to unwinding, and this contributes to maximal fork rates (Yeeles et al., 2017). However, when synthesis is stalled, CMG can continue to unwind at a reduced rate, a scenario termed helicase-polymerase uncoupling (Devbhandari and Remus, 2020; Taylor and Yeeles, 2018, 2019).

In addition to exogenous factors, certain DNA sequences can intrinsically pose a challenge to the replisome, in terms of both fidelity and dynamics. Most of our understanding of how DNA affects its own replication stems from studies of expansion-prone repeats which drive nearly 50 different neurodegenerative diseases (Lopez Castel et al., 2010; McMurray, 2010; Mirkin, 2007). Roughly half of these conditions are caused by expansion of just three repeat classes – (CGG/CCG)_n, (GAA/TTC)_n and (CTG/CAG)_n (hereafter referred to as (CGG)_n, (GAA)_n and (CTG)_n). In these diseases, the number of repeat units is highly variable within the general population. When repeats expand to an intermediate range, individuals may exhibit partial phenotypes. Further expansion, usually within one generation, leads to a full mutation. For example, Fragile X syndrome is caused by expansion of (CGG)_n repeats within the 5' UTR of the FMR1 locus. Unaffected individuals harbour 6-52 (CGG)_n repeats, an intermediate range is 53-250, whereas patients harbour 250 to 2,000 repeat units.

One of the earliest proposed mechanisms for contractions or expansions of repeats was replication slippage ([Hartenstine et al., 2000](#); [Petruska et al., 1998](#)), a process by which the template and nascent strands reanneal out of register due to the repetitive nature of the template ([Kunkel, 1990](#)). However, large scale contractions and expansions cannot easily be explained by slippage. Furthermore, slippage can occur in any repetitive sequence, yet only some repeats undergo large scale expansions. Replication, transcription and various repair pathways have been implicated in large scale expansions ([Khristich and Mirkin, 2020](#)), but the exact underlying mechanisms are not yet fully understood. Current models are based on the finding that repeat expansions correlate with the propensity of sequences to fold into unusual DNA secondary structures.

Several types of non B-DNA secondary structures have been characterised, including (i) hairpins, (ii) G-quadruplexes (G4s), (iii) intercalated motifs (i-motifs) and (iv) triplexes. Hairpins are simple intramolecular fold-back structures that rely on classic Watson-Crick base pairing. Inverted repeats and palindromes can form perfectly annealed hairpins while (CNG)_n repeats can form mismatch-containing hairpins. G4s are four stranded intra- or intermolecular structures formed by Hoogsteen base pairing between guanine residues ([Gellert et al., 1962](#)). Four guanines can form a planar arrangement termed a G-quartet and stacking of multiple G-quartets yields a G4. While G-rich sequences can form G4s, C-rich sequences can form a four-stranded structure called an i-motif, where pairs of hemi-protonated cytosines form Hoogsteen base pairing in a criss-cross pattern ([Gehring et al., 1993](#)). Hairpins, G4s and i-motifs can all form locally within a stretch of single stranded DNA (ssDNA). In contrast, triplex DNA requires a donor duplex DNA, with a third strand annealing via Hoogsteen base pairing ([Frank-Kamenetskii and Mirkin, 1995](#)).

Triplexes can arise from homopurine-homopyrimidine mirror repeats, such as (GAA)_n repeats, and their formation is favoured by negative supercoiling ([Wells, 2008](#)).

The first evidence that repeats can stall fork progression in vivo was the detection of replication intermediates of plasmids containing (CGG)_n repeats in bacteria by two-dimensional (2D) gel electrophoresis ([Samadashwily et al., 1997](#)). Stalling was observed in both orientations and was later also detected in budding yeast and mammalian cells ([Pelletier et al., 2003](#); [Voineagu et al., 2009](#)). In contrast, stalling by (GAA)_n repeats in yeast only occurs when they are on the lagging strand template ([Kim et al., 2008](#); [Krasilnikova and Mirkin, 2004](#); [Shishkin et al., 2009](#)), whereas stalling by (CTG)_n repeats is significantly weaker and is orientation-independent ([Pelletier et al., 2003](#); [Viterbo et al., 2016](#)). More recent evidence of perturbed replication in patient-derived cells seem inconsistent with findings in yeast. Fiber labelling of individual replication forks in the CGG-expanded FMR1 locus from Fragile X syndrome cells revealed very little stalling ([Gerhardt et al., 2014](#)). Interestingly, replication forks progressed in either direction in cells from unaffected individuals, whereas almost all forks in patient cells replicated (CGG)_n as the leading strand template. Similar experiments with cells from Friedreich's ataxia patients showed pronounced stalling in the GAA-expanded FXN locus. Fork directionality was also altered, positioning the (GAA)_n repeats on the leading strand template ([Gerhardt et al., 2016](#)), which is the exact opposite orientation that generates stalls in budding yeast ([Kim et al., 2008](#); [Krasilnikova and Mirkin, 2004](#); [Shishkin et al., 2009](#)). The reasons for these discrepancies are unclear. Furthermore, the underlying mechanism of repeat-induced stalling is poorly defined. Stalling could be induced indirectly, for example by DNA-bound proteins or R-loops. In the case of (CTG)_n repeats, stalling was suggested to be driven by binding of mismatch repair factors to mismatched

hairpins (Viterbo et al., 2016). This raises the question of whether the DNA template by itself is sufficient to stall the replisome. If so, which sequences stall and what is the underlying mechanism? Finally, how does the replisome recover from such blocks?

Studies of repeat replication in vitro have thus far been limited to primer extension assays and have shown that polymerases are impeded by (CGG)_n, (CTG)_n and (GAA)_n repeats (Delagoutte et al., 2008; Gacy et al., 1998; Kang et al., 1995; Murat et al., 2020; Ohshima and Wells, 1997; Usdin and Woodford, 1995). Most studies employed bacterial or viral polymerases, with very little work done with all three eukaryotic replicative polymerases. One study compared yeast pol δ with human pols α and ϵ , all of which were stalled by (CGG)₇ (Kamath-Loeb et al., 2001). One limitation of such assays is the use of ssDNA templates that are pre-folded into structures. Whether sufficient ssDNA could be exposed for structures to form during unperturbed coupled leading strand synthesis is unknown. Another caveat is the lack of any additional replisome components. Reconstituted E. coli replisomes are not affected by (CTG)_n repeats but are stalled by short (CCG)_n repeats and inverted repeats (Lai et al., 2016; Le et al., 2015). To date, studies of repeat replication with reconstituted eukaryotic replisomes are lacking.

In this study we set out to determine the molecular events that transpire when the eukaryotic replisome encounters repetitive templates. Using reconstituted replisomes assembled from purified budding yeast proteins, we found that certain repeats induce leading strand stalling. Since these experiments lack components from other pathways, they indicate that DNA alone can cause replication fork stalling. We tested a wide range of mono-, di- and trinucleotide repeats and found that stalling correlated most with structure-forming capacity. Mechanistically, the CMG helicase was able to continue unwinding but synthesis was inhibited, resulting in helicase-

polymerase uncoupling, thereby resembling events induced by a leading strand DNA lesion. We found that the two major replicative polymerases, pols δ and ε , exhibit different inherent capacities to synthesize through hairpin-forming repeats and uncovered a role for pol δ in rescuing DNA-induced leading strand stalling. Moreover, fork recovery mechanisms differed by the type of secondary structure that repeats can form. Leading strand synthesis through hairpin-forming repeats was modulated by various replisome-intrinsic aspects, including the presence of pol δ , synthesis rate by pol ε and levels of dNTPs. In contrast, quadruplex-forming repeats were not affected by any of these factors, but instead required the extrinsic accessory helicase Pif1 for efficient replication. Altogether, these results provide a mechanistic understanding of how the eukaryotic replisome copes with challenging repetitive templates and highlights certain sequences as an important potential source of endogenous replication stress.

RESULTS

(CGG)_n repeats induce leading strand stalling

To investigate the effect of repeats on the eukaryotic replisome, we constructed a set of substrates for in vitro replication assays, whereby eukaryotic replisomes are assembled using purified budding yeast proteins (Yeeles et al., 2015; Yeeles et al., 2017). (CTG)_n, (GAA)_n and (CGG)_n repeats were cloned 3 kb downstream of the replication origin (Fig. 1A) of a 9.8 kb substrate that supports origin-specific replication initiation (Taylor and Yeeles, 2018). Short oligonucleotides were used for initial cloning, followed by a PCR-free approach which involved iterative steps of controlled expansion of repeats to yield substrates with up to 161 uninterrupted repeats (Scior et

al., 2011). Given the potentially unstable nature of certain repeats during propagation in bacteria, we validated that our final preparations contained the correct insert size and sequence (**Fig. S1**). Since replication initiates from a defined position, we can assign which sequences serve as the leading and lagging strand templates. When describing insert sequences throughout this manuscript, we refer to sequences that reside on the leading strand template. For example, the (CGG)₆₁ substrate contains 61 CGG repeats on the leading strand template, and therefore 61 CCG repeats on the lagging strand template.

To avoid the confounding effects of two replication forks converging on a circular template, we first performed reactions on linear templates. Plasmids were linearised with a restriction enzyme (AhdI) such that the replication origin was positioned 1.5 kb from one end, and 8.2 kb from the other, with the repeats located within the 8.2 kb fragment. Enzymes required for Okazaki fragment maturation were omitted to simplify analysis. As expected, analysis of the control replication reaction by denaturing alkaline gel electrophoresis produced three main products: the leftward moving 1.5 kb leading strand, the rightward moving 8.2 kb leading strand, and a heterogeneous population of smaller unligated lagging strand Okazaki fragments (**Fig. 1B**, lane 1). Replication of substrates containing (CTG)₁₆₁ or (GAA)₁₆₁ did not differ from the empty vector control (**Fig. 1B**, lanes 1-3). However, a very faint 3kb stall band was reproducibly detected with (CGG)₁₆₁ (**Fig. 1B**, lane 4). The intensity of this stall was increased when reactions were performed without pol δ (**Fig. 1C**), suggesting a role for pol δ in preventing or rescuing leading strand stalls induced by (CGG)₁₆₁. Since these experiments lack components from other pathways, we conclude that the DNA template itself can induce fork stalling, and that this is modulated by polymerase usage.

Stalling threshold is 17 (CGG)_n repeats and is orientation-dependent

To establish the threshold for (CGG)_n stalling we replicated a set of substrates with increasing repeat units in the absence of pol δ . This revealed that as few as 17 repeats were sufficient to induce some stalling, which was further enhanced with 21 and 41 repeats, and saturated with 61 repeats or more (**Fig. 1D**). Similar results were obtained with circular plasmids in the presence of topoisomerase I (**Fig. S2A, B, C**), indicating that stalling is neither promoted nor prevented by a topologically closed template or by topoisomerase activity. When compared to a stall driven by a site-specific leading strand DNA lesion (a cyclobutane pyrimidine dimer; CPD), even the longest (CGG)_n inserts produced a partial stall, also evident by the larger proportion of full length 8.2 kb products (**Fig. 1D**, compare lanes 9 and 10). Consistent with the accumulation of stalled forks, large replication intermediates were observed by native gel electrophoresis (**Fig. S2D**), mirroring the pattern seen by alkaline denaturing analysis. We note that (CGG)_n inserts containing 81 repeat units or more were not completely stable in bacteria (**Fig. S1B**, lanes 7-10, seen as smearing below the main band). We therefore chose to use (CGG)₆₁ in all subsequent experiments as it drove maximal stalling but was genetically stable.

If, as suggested by genetic evidence, the orientation of repeats relative to replication origins plays a role, one might expect to observe a difference in stalling as a function of orientation. To test this idea, we reversed the orientation of these repeats to yield (CAG)_n, (TTC)_n and (CCG)_n templates. While we were able to clone (CAG)₁₆₁ and (TTC)₁₆₁, we were only able to obtain stable clones of up to 61 CCG repeat units, as longer CCG repeats are unstable in this orientation in bacteria ([Hirst and White, 1998](#); [Shimizu et al., 1996](#)). Nonetheless, in contrast to (CGG)_n templates, replication of all (CCG)_n substrates produced no detectable stalls (**Fig. 1E**), even when compared

side-by-side (**Fig. S2E**). Replication of (CAG)₁₆₁ and (TTC)₁₆₁ produced no stalling with either linear or circular templates (**Fig. S2F, G**). In summary, as many as 161 (CTG)_n or (GAA)_n repeats do not induce stalling in either orientation, whereas 17 (CGG)_n repeats or more stall replication forks, but only when positioned on the leading strand template.

Short (CG)_n repeats also induce leading strand stalling

The fact that (CGG)_n produced a stall, yet other trinucleotide repeats did not, suggested that it is not simply their repetitive nature that causes a stall. We considered the possibility that stalling is caused by DNA secondary structures. While all (CNG)_n repeats can fold into hairpins, the thermal stability of (CGG)_n hairpins is significantly higher ([Gacy and McMurray, 1998](#)), possibly explaining the stalling observed only with (CGG)_n. This raises the prediction that other G-rich hairpin-forming repeats may also stall the replisome. To test this, we cloned and replicated a range of dinucleotide repeats. Of these, stalling was only observed with (CG)_n repeats (**Fig. 2A** and **Fig. S3A**), which are indeed G-rich and form hairpins in solution ([Murat et al., 2020](#)). Relative to (CGG)_n, much shorter stretches of (CG)_n dinucleotides produced a strong stall (**Fig. 2A**), with a lower threshold of only 10 repeat units. Similar to that observed with (CGG)_n templates, analysis of (CG)_n replication products on a native gel revealed accumulation of replication intermediates (**Fig. S3B**) and the stalling threshold was similar with circular templates (**Fig. S3C**). Therefore, these results provide further evidence that hairpin-forming repeats can stall the replisome. To further support this interpretation, we generated scrambled sequences with the same length, base pair composition and strand bias as (CGG)₂₁ or (CG)₂₄. For each repeat type we chose two randomly generated sequences which contain minimal stretches of consecutive CG or CGG repeats, thereby interrupting continuous base-pairing within the predicted

hairpins. All of the scrambled sequences were replicated without any stalling (**Figure 2B, C**). Altogether, these results indicate that the nucleotide composition and strand bias of (CGG)_n and (CG)_n repeats do not account for their ability to stall leading strand synthesis. Rather, stalling is most consistent with their structure-forming potential.

The replisome is affected by quadruplex-forming homopolymers

The leading strand stalling we observed correlated with the ability of sequences to fold into hairpin structures. We reasoned that repeats that form other types of DNA secondary structures may also impede replication. We therefore tested the effect of guanine and cytosine homopolymers, which can fold into a G4 or i-motif, respectively (**Murat et al., 2020**). Leading strand stalls were indeed observed, with a threshold of 20 and 30 repeat units for (G)_n and (C)_n, respectively (**Fig. 2D, E**). This difference in threshold was also seen when compared side-by-side within the same experiment (**Fig. S3D**) and was maintained on circular templates (**Fig. S3E, F**). In contrast, stretches of over 200 consecutive adenine or thymine residues, which are not predicted to form stable secondary structures, did not cause a significant stall (**Fig. 2F**). Altogether, we conclude that hairpin- and quadruplex-forming repeats can stall the replisome.

Pol δ drives recovery from hairpin-forming, but not quadruplex-forming repeats

Our results thus far highlight four different types of repeats that induce leading strand stalling – (CGG)_n, (CG)_n, (C)_n and (G)_n. Given our initial observation that pol δ can assist replication through (CGG)₁₆₁ (**Fig. 1B, C**), we next asked whether this holds true for the other sequences. While replication of (CGG)₆₁ and (CG)₂₄ was improved by the presence of pol δ, stalling by (C)₅₀ and (G)₅₀ was essentially unaffected (**Fig. 3A** and

Fig. S3G). Thus, the ability of pol δ to synthesise past these sequences correlates with the type of secondary structure that they can form.

To assess if stalling is terminal or transient we performed pulse-chase experiments, in which nascent DNA was labelled with dATP for the first 10 minutes and chased with excess unlabelled dATP, allowing us to follow the fate of forks labelled within the pulse without detection of new initiation events. In the absence of pol δ , stalling by (CGG)₆₁ was persistent for at least two hours, indicating that pols α and ϵ are unable to resolve this stall (**Fig. 3B**). In contrast, in the presence of pol δ stalling at the earliest time point was weaker, gradually resolved over time, and was barely discernible by 40 minutes (**Fig. 3C**). A similar pattern was observed with (CG)₂₄ (**Fig. 3D, E**). These results indicate that pol δ does not prevent the formation of stalls induced by (CGG)_n and (CG)_n but rather resolves them. Pulse chase experiments with (C)₅₀ and (G)₅₀ revealed persistent stalling regardless of the presence of pol δ (**Fig. S4**), further supporting our earlier observation that pol δ cannot support replication through these two sequences (**Fig. 3A**). In summary, hairpin-forming sequences induce persistent stalls in the absence of pol δ , but these are resolved over time when pol δ is present. In contrast, G4- and i-motif-forming sequences generate persistent stalls that cannot be resolved by pol δ .

The ability of pol δ to rescue certain leading strand stalls could either require its continued presence within the replisome during stalling or could occur behind the fork. To test whether pol δ could rescue pre-existing stalls we carried out pulse-chase experiments in which stalls were pre-formed during the pulse, and pol δ was only introduced in the chase. Stalling at (CGG)₆₁ and (CG)₂₄ was evident after 10 min and remained unaltered in the absence of pol δ (**Fig. 3F**, lanes 4 vs 5 and 7 vs 8). However,

pol δ was able to resolve most of these stalls within 10 min (**Fig. 3F**, compare lanes 5 vs 6 and 8 vs 9). We note that we have observed some degree of variation in pol δ -dependent rescue efficiency, with synthesis through (CG)₂₄ typically being less efficient than (CGG)₆₁. Leading strand rescue by pol δ was largely dependent on RFC/PCNA (**Fig. S5**), suggesting that PCNA is either retained or reloaded on the leading strand template after stalling occurs. We conclude that pol δ can rescue pre-existing leading strand stalls in a PCNA-dependent fashion.

DNA-induced stalls trigger helicase-polymerase uncoupling

Replication forks could either stall due to impaired unwinding by the CMG helicase or inhibition of synthesis by pol ϵ , which would trigger uncoupled unwinding downstream of the stall. The fact that pol δ could rescue pre-existing stalls (**Fig. 3F**) supports the latter, as it strongly argues for the presence of a free primer-template junction and an available exposed template downstream. Previous work revealed that repriming past a leading strand CPD by pol α is inefficient, and that an exogenously added primer allows resumption of leading strand synthesis ([Taylor and Yeeles, 2018](#)). Primer annealing only occurs if ssDNA is exposed, thereby serving as an indirect measure of uncoupled CMG unwinding. We therefore asked whether a primer that anneals 265 nt downstream of the insert would promote the formation of a restart product. Indeed, addition of this primer, but not a scrambled control primer, led to the appearance of a 5 kb restart product for all four stall-forming repeats, to an extent similar to that seen with a leading strand CPD template (**Fig. 4A**). This result strongly suggests that stalling is not a consequence of CMG arrest, but is rather due to lack of synthesis by pol ϵ . Interestingly, while pol δ resolved the 3 kb stall products induced by (CGG)₆₁ and (CG)₂₄, 5 kb restart products were still evident (**Fig. 4B**, lanes 8 and 9). Therefore,

CMG continued to unwind at least 265 nt beyond the repeats in both cases. Thus, although pol δ can resolve certain leading strand stalls, it cannot completely prevent uncoupling.

Additional evidence for helicase-polymerase uncoupling was seen upon closer inspection of replication products analysed on native gels, whereby faster migrating species accumulated. These species were previously shown to correspond to uncoupled products, in which CMG has unwound to the end of the template but without any synthesis (Taylor and Yeeles, 2018). This was especially clear with the (CG)_n templates, where uncoupled products accumulated at levels similar to those observed with a CPD containing template (Fig. S3B). To increase the fraction of uncoupled products, we truncated substrates with EcoRV so that CMG has to unwind only 1.6 kb beyond the insert rather than 5 kb. When analysed on a native gel, uncoupled products were observed for all four classes of sequences (Fig. 4C), but were not observed for (CGG)₆₁ and (CG)₂₄ when pol δ was present (Fig. 4D), which is consistent with resumption of synthesis. Altogether, these results show that structure-forming repeats can trigger helicase-polymerase uncoupling and that pol δ limits the extent of uncoupling by rescuing leading strand synthesis at (CGG)₆₁ and (CG)₂₄, but not at (C)₅₀ or (G)₅₀.

Read-through of (CGG)_n and (CG)_n is facilitated by pol ϵ variants or elevated dNTPs

The observation that pol ϵ could not synthesise past (CGG)_n or (CG)_n, yet pol δ could, may be explained by their different enzymatic properties. More specifically, the weak strand displacement activity of pol ϵ relative to pol δ might preclude it from coping with hairpin-forming repeats. This activity can be mildly enhanced by inactivating the

exonuclease domain of pol ϵ (Ganai et al., 2016). In addition, modelling of the most frequent cancer-associated pol ϵ mutation (P286R) in budding yeast (P301R) revealed a hyperactive enzyme in which DNA entry into the exonuclease domain is blocked, allowing it to synthesise past a hairpin structure more efficiently than an exonuclease-dead mutant (Parkash et al., 2019; Xing et al., 2019). We therefore wondered whether these pol ϵ variants might be able to resolve leading strand stalls even in the absence of pol δ . Leading strand stalls induced by either (CGG)₆₁ or (CG)₂₄ were significantly weaker in reactions carried out with pol ϵ P301R (Fig. 5A, compare lanes 1,2 vs 3,4 and 9,10 vs 11,12), while an exonuclease-dead pol ϵ produced an intermediate phenotype. Neither of these pol ϵ variants were able to replicate past (G)₅₀ or (C)₅₀ (Fig. S6A) and similar results were obtained when pol δ was present (Fig. S6B). Pol ϵ P301R was able to rescue pre-existing stalls produced by WT pol ϵ (Fig. S6C), and this was largely dependent on RFC/PCNA (Fig. S6D). These observations are almost identical to those obtained with pol δ (Fig. 3F and S5), suggesting that pol ϵ P301R and pol δ employ a similar mechanism to rescue leading strand stalls.

Inactivation of the exonuclease domain of pol ϵ shifts the balance from proofreading to synthesis, leading to an overall increase in synthesis rate. Other factors that enhance synthesis rate could also play a role. We therefore asked whether increased dNTPs could ameliorate DNA-induced stalling. We performed pulse-chase experiments in which dATP was the labelled nucleotide, and chased with either excess unlabelled dATP alone, or an excess of all four dNTPs (raised from 30 μ M to 400 μ M). In the absence of pol δ , elevated dNTPs significantly improved replication past (CGG)₆₁ but not (CG)₂₄ (Fig. 5B, compare lanes 5 vs 6 and 8 vs 9). In the presence of pol δ , excess dNTPs also improved synthesis past (CG)₂₄ (Fig. 5B, compare lanes

17 vs 18). However, there was no effect on replication of (G)₅₀ or (C)₅₀, regardless of pol δ (**Fig. S7A**). Thus, increased concentrations of dNTPs improve the ability of both replicative polymerases to resolve stalls induced by hairpin-forming repeats. Combined with the results obtained with pol δ and pol ϵ variants, we conclude that the replisome can cope with hairpin-forming repeats by a variety of replisome-intrinsic mechanisms.

Pif1 resolves DNA-induced stalls

In contrast to hairpin-forming repeats, none of the conditions or enzyme variants we tried thus far allowed the replisome to cope with stalls induced by (G)₅₀ and (C)₅₀, both of which can form quadruplex structures. We considered that the ssDNA binding protein RPA may play a protective role as it has been demonstrated to unfold G4 structures ([Fan et al., 2009](#); [Ray et al., 2013](#); [Salas et al., 2006](#)). However, stalled products were observed across a broad range of RPA concentrations (10-200 nM) with all tested sequences, regardless of pol δ (**Fig. S7B, C**). Therefore, in this context, RPA does not prevent or resolve DNA-induced leading strand stalls.

Several accessory helicases have been implicated in replication of repetitive or structure-prone DNA ([Anand et al., 2012](#); [Sauer and Paeschke, 2017](#)). In budding yeast, Pif1 has been shown to play an important role in allowing efficient replication past G4 sequences in vivo ([Dahan et al., 2018](#); [Lopes et al., 2011](#); [Paeschke et al., 2011](#)) and in vitro ([Byrd et al., 2018](#); [Maestroni et al., 2020](#); [Paeschke et al., 2013](#); [Ribeyre et al., 2009](#); [Sparks et al., 2019b](#)). We therefore assayed the ability of purified Pif1 to rescue DNA-induced stalled forks. Strikingly, not only was Pif1 able to fully rescue replication past (G)₅₀, it also accelerated replication through all of the other sequences (**Fig. 6A**). Importantly, an ATPase active site mutant of Pif1 (K264A) which

cannot unwind DNA (**Fig. S8A**), was unable to perform any of these tasks (**Fig. 6**), indicating an essential requirement for its helicase motor function. For comparison, we also tested the nuclease-helicase Dna2, but found it had no effect on DNA-induced stalling despite showing robust nuclease activity (**Fig. S8B, C**). Pif1 was previously shown to directly bind PCNA ([Buzovetsky et al., 2017](#)) and to collaborate with pol δ and PCNA in break induced replication (BIR) ([Saini et al., 2013](#); [Wilson et al., 2013](#)) and in stimulating strand displacement during lagging strand maturation ([Koc et al., 2016](#); [Osmundson et al., 2017](#); [Rossi et al., 2008](#)). However, our results show that the ability of Pif1 to resolve DNA-induced stalls is distinct from these functions, as it did not require pol δ (**Fig. 6B**) or PCNA (**Fig. S8D**). Altogether, we conclude that Pif1 is a general-purpose accessory helicase that accelerates recovery from a variety of leading strand DNA-induced stalls.

DISCUSSION

We have reconstituted repeat replication with eukaryotic replisomes and have found that DNA alone is sufficient to cause significant leading strand stalling. Therefore, certain DNA sequences are an important source of endogenous replication stress. Mechanistically, stalling induced by DNA repeats and leading strand DNA lesions is similar – CMG unwinding is unaffected and inhibition of synthesis triggers helicase-polymerase uncoupling. Furthermore, we demonstrate that the two major replicative polymerases exhibit different inherent capacities to cope with repetitive templates, with pol δ showing more robust activity than pol ϵ , allowing it to rescue leading strand stalls caused by hairpin-forming repeats. The replisome could recover from stalls induced by hairpin-forming sequences by employing a variety of replisome-intrinsic

mechanisms, including pol δ , hyperactive pol ϵ or elevated dNTPs. In contrast, stalls induced by quadruplex-forming sequences required extrinsic support, revealing a general role for the Pif1 helicase in accelerating recovery from a variety of DNA-induced stalls. These results invoke several interesting and important questions, including the root cause of stalling and the emergence of different recovery mechanisms.

It is evident that only certain sequences induce leading strand stalling, yet the underlying reason is unclear. Our results show stalling cannot be easily explained by the repetitive nature of sequences, their base pair composition or their strand bias. Rather, fork stalling is best correlated with the ability of sequences to fold into stable DNA secondary structures. Although (CGG)_n repeats have been shown to fold into a G4 structure (Fry and Loeb, 1994) or Z-DNA (Renčiuk et al., 2011) in vitro, this only occurs under non-physiological conditions (Amrane and Mergny, 2006; Fojtík et al., 2004). Thus, hairpins are the most physiologically likely structures formed by (CGG)_n, with stretches of over 12 repeats suggested to form branched hairpins (Amrane and Mergny, 2006). The stall threshold we observed (n=17) was surprisingly low, meaning that in most normal FMR1 alleles (n=5-63) local uncoupling may occur, providing a plausible mechanism for small scale expansions. Given that even short (CGG)_n repeats are inherently difficult to replicate, it would be tempting to speculate that Pif1 and pol δ become more important in efficient and accurate replication of expanded Fragile X syndrome alleles.

Out of all the sequences we tested, stalling by relatively short (CG)_n repeats exhibited the highest proportion of stalled forks. Recent NMR analysis shows that in solution (CG)_n repeats form hairpins (Murat et al., 2020). Although (CG)_n repeats could in theory also form cruciforms ahead of the fork, this does not happen even in

negatively supercoiled plasmids ([Singleton et al., 1982](#)) because CG-rich DNA inhibits cruciform nucleation ([Sinden, 1994](#)). Interestingly, (CG)_n repeats are extremely rare - not only in the human genome, but across the entire tree of life - constituting less than 1% of all dinucleotides in most species ([Srivastava et al., 2019](#)). Methylation of cytosine within CpG increases its rate of deamination, resulting in C to T transitions. This has been proposed as the main evolutionary mechanism for genomic suppression of (CG)_n dinucleotides ([Pfeifer, 2006](#)). However, trinucleotides such as (CGG)_n do not show such remarkable genomic depletion, despite harbouring the same CpG sequences. This suggests that (CG)_n sequences undergo negative selection. We propose that the capacity of (CG)_n to efficiently stall replication serves as a selective force that leads to their genomic suppression. Nonetheless, the human reference genome contains over 50 loci with 10 or more consecutive (CG)_n repeats, with the longest one being a single instance of (CG)₁₄. Since the minimal stalling threshold we observed was 10 repeats, these genomic loci may be hot spots of fork stalling and uncoupling and may be more dependent on the activities of pol δ and Pif1.

The lack of stalling by (GAA)_n repeats may seem unexpected, as these repeats induce robust stalling in vivo in multiple organisms ([Gerhardt et al., 2016](#); [Kim et al., 2008](#); [Krasilnikova and Mirkin, 2004](#); [Shishkin et al., 2009](#)). However, the fact that stalling is observed in opposite orientations in yeast and human patient derived cells strongly points to additional factors being involved. One possible factor could be sequence context. Analysis of SV40-based (GAA)_n plasmids by electron microscopy revealed the formation of unusual fork structures such as reversed forks ([Follonier et al., 2013](#)). Interestingly, in this context only weak and transient stalling was observed. Triplex structures were also observed, and these formed between the (GAA)_n repeats and other GA-rich regions within the plasmid. This is in agreement with earlier studies

showing that in bacteria, two tracts of (GAA)_n can form triplexes, whereas a single tract cannot (Vetcher et al., 2002). It is therefore possible that our substrates lack a sufficiently long second GA-rich array to serve as a dsDNA donor. An alternative explanation was raised by a recent study carried out in DT40 cells, where replication stalling by relatively short (GAA)_n tracts was suggested to occur due to R-loops (Šviković et al., 2019). Altogether, we conclude that within our experimental conditions, (GAA)_n repeats by themselves do not cause significant leading strand stalling.

Our results with guanine homopolymers are consistent with previous analysis of the effects of G4 forming sequences on replication and the role of Pif1 in resolving stalling (Paeschke et al., 2011). While past work supports the idea that G4 structures impede replication, the evidence is conflicting with regards to the effect of their orientation relative to replication origins. Loss of epigenetic information in avian DT40 cells due to uncoupling can be induced by a single G4 forming sequence, but only when positioned on the leading strand template (Sarkies et al., 2010). Similarly, genetic instability of G4-forming human minisatellites in budding yeast is only induced when the G-rich strand is positioned on the leading strand template (Lopes et al., 2011). In contrast, live cell imaging of fluorescent arrays in budding yeast detected delays in replisome progression only when G4 sequences were positioned on the lagging strand template (Dahan et al., 2018). Our results show that cytosine homopolymers also induce leading strand stalling and NMR spectroscopy analysis directly demonstrated that (C)₂₂ forms an i-motif (Murat et al., 2020). It is therefore possible that for some G4-forming sequences the C-rich strand produces a stall due to an i-motif structure, whereas in other cases the G-rich strand does so due to a G4 structure.

In the context of replication, DNA secondary structures could either form ahead of the fork or behind the fork. Our current working model is that DNA secondary structures form behind CMG, since we observe efficient uncoupling. However, we cannot rule out the possibility that our substrates contain pre-existing structures, and that these are bypassed intact by CMG. Recent work in *Xenopus* egg extracts revealed that CMG is able to bypass a large protein cross-linked to the leading strand template (Sparks et al., 2019a), although this required generation of ssDNA downstream by the accessory helicase Rtel1, and may require additional factors. Single molecule studies revealed that yeast CMG possesses an Mcm10-dependent “gate” that allows it to transition from ssDNA to dsDNA (Wasserman et al., 2019), which could perhaps allow it to bypass certain structures on the leading strand. Alternatively, CMG may unwind and dismantle pre-existing structures. While the capacity of CMG to unwind various DNA secondary structures is unknown, our results imply that if CMG does unwind structures, these must form again behind it to inhibit synthesis. But how could structures arise on the leading strand if unwinding and synthesis are coupled? Although pol ϵ directly binds CMG, its catalytic domain is tethered via a flexible linker (Zhou et al., 2017). This raises the possibility that stochastic disengagement of pol ϵ from the leading strand template leads to local uncoupling and exposure of short stretches of ssDNA, thereby allowing structures to form. However, fork stalling induced by (CG)₂₄ was extensive, which would require such a stochastic event to be very frequent. Another option is that structures form on the ssDNA stretch that runs between the exit channel of CMG and the active site of pol ϵ . At present there is no exact information on the length of exposed leading strand template during coupled synthesis. Current estimates are at least 16 nt, based on a recent structure of pol ϵ bound to CMG (Yuan et al., 2020). Importantly, the minimum

length required to form a three stacked G4 or i-motif is 15 nt, whereas hairpins could nucleate from even shorter sequences. Very recent super resolution imaging of individual replication forks in human cells have detected G4 structures behind CMG, but not in front of it (Lee et al., 2021), providing support for our model that structures form as a consequence of replication.

We have discovered that the replisome can intrinsically resolve stalls induced by hairpin-forming sequences through multiple mechanisms, with pol δ playing a major role. In contrast, stalls induced by quadruplex-forming sequences require the extrinsic support of the accessory helicase Pif1. Our results are in strong agreement with a recent high-throughput primer extension assay that tested the ability of T7 polymerase to extend through all possible 1-6 nt long repeats, as well as a large library of hairpin, G4 and i-motif sequences (Murat et al., 2020). Synthesis by this model polymerase gradually progressed through hairpins, with more stable hairpins taking longer to resolve, but was terminally stalled at either G4s or i-motifs. It thus seems that quadruplexes are a more robust block to synthesis by many polymerases. In contrast, we found that the two major eukaryotic replicative polymerases exhibit varying intrinsic capacities to synthesise through hairpins. The strand displacement activity of pol δ most likely evolved for the purpose of Okazaki fragment maturation. However, this comes with the added benefit of allowing pol δ to rescue leading strand stalls caused by hairpin-forming sequences.

Replication fork uncoupling leads to exposure of ssDNA on the leading strand template, threatening genetic and epigenetic stability. It is therefore essential to minimize these events. Although pol δ was able to resume synthesis of hairpin-forming repeats on the leading strand, local uncoupling was not completely prevented. Several types of DNA lesions on the leading strand template induce events similar to those we

observed here, including inhibition of synthesis and uncoupling of synthesis from unwinding. Interestingly, similar to its ability to synthesise past hairpin-forming sequences, pol δ could also rescue leading strand synthesis past 8-oxoguanine and thymine glycol (Guilliam and Yeeles, 2021). In contrast, replication past an abasic site or a CPD could not be carried out by any of the replicative polymerases (Taylor and Yeeles, 2018, 2019). However, translesion synthesis by pol η could perform synthesis past a CPD (Guilliam and Yeeles, 2020). This requirement of an external factor is very much akin to the role of Pif1 in rescuing replication of quadruplex-forming sequences. Thus, the molecular events that underlie DNA-induced stalling could be mechanistically analogous to those induced by leading strand DNA lesions, exhibiting both intrinsic and extrinsic recovery pathways.

The exposure of ssDNA by uncoupled CMG generates a checkpoint response by binding of RPA and subsequent activation of the ATR kinase. Drugs that induce replication stress, such as hydroxyurea and aphidicolin, as well as ATR inhibition, induce DNA double strand breaks in non-random locations across the genome. Genome wide studies have identified structure forming repeats, inverted repeats, quasi-palindromes (Shastri et al., 2018) and short poly(dA:dT) homopolymers (Tubbs et al., 2018) as hot spots for breakage. These observations are broadly in agreement with our findings and strengthen the notion that the ability of certain sequences to interfere with replication is dictated by their structure-forming potential. In addition, recent work revealed that in tumours which exhibit Microsatellite Instability (MSI), large scale expansions of (AT)_n repeats generate genome-wide cleavage sites for the Mus81-Eme1 nuclease. Furthermore, the WRN helicase plays a major role in resolving such structures and preventing breakage (van Vietmarschen et al., 2020). Despite the clear ability of G4s and i-motifs to stall replication in our experiments, these were not

identified as hot spots in such studies. One possibility is that Pif1 and other related helicases such as FancJ and Rtel1 resolve these structures very efficiently, thereby preventing uncoupling and breakage. Another possibility is that these structures are refractory to cleavage by structure-specific endonucleases. Nonetheless, it would be tempting to speculate that repetitive sequences that can inherently stall replication, such as those identified in this study, are also hotspots for fork collapse or breakage.

In summary, we have shown that repetitive DNA is an important potential source of endogenous replication stress and have revealed how the eukaryotic replisome is able to cope with difficult-to-replicate sequences. The response of the replisome to certain repetitive sequences is mechanistically similar to events driven by leading strand DNA lesions. We therefore propose that repetitive sequences can also induce the checkpoint response to replication stress. Thus, our broad knowledge and understanding of the cellular response to replication stress and DNA damaging agents may now be extended to encompass DNA-induced replication stalling.

ACKNOWLEDGMENTS

This work was funded by a Wellcome Trust and Royal Society Sir Henry Dale fellowship (210470/Z/18/Z) as well as internal funding from the Institute of Cancer Research. We would like to thank Max Douglas for reagents, experimental support and critical reading of the manuscript. We would also like to thank Jonathon Pines, Wojciech Niedzwiedz, Sebastian Guettler, Marco Di Antonio, Christian Zierhut and Allison McClure for critical reading of the manuscript.

AUTHOR CONTRIBUTIONS

Methodology and Investigation, C.S.C.D, M.D.M, S.W and G.C; Conceptualization, supervision, writing and funding acquisition, G.C.

DECLARATION OF INTERESTS

The authors declare no competing interests.

FIGURE TITLES AND LEGENDS

Figure 1 - (CGG)_n repeats induce orientation-dependent leading strand stalling

(A) Schematic of the replication substrate used in this study. The ARS306 origin of replication drives sequence-specific initiation events. Two forks emanating from the origin would generate a leftward moving 1.5 kb leading strand (“Left leading”) and a rightward moving 8.2 kb leading strand (“Right leading”). Various repeats were cloned 3 kb downstream such that only the rightward moving fork would encounter them. Thus, the left leading product serves as an internal control.

(B and C) Analysis of in vitro replication reactions by denaturing gel electrophoresis, using linear substrates with different types of leading strand repeats, in the presence (B) or absence (C) of pol δ.

(D and E) Replication reactions carried out in the absence of pol δ with a series of substrates containing increasing numbers of either (CGG)_n (D) or (CCG)_n (E) leading strand repeats, as well as a comparison to a site-specific leading strand CPD.

Figure 2 - (CG)_n, (G)_n and (C)_n also induce leading strand stalling

(A) Replication reactions carried out in the absence of pol δ with a series of (CG)_n substrates, as well as a comparison to a site-specific leading strand CPD.

(B and C) Replication reactions carried out in the absence of pol δ, comparing two randomly generated scrambled sequences with the same base-pair composition and strand bias as either (CG)₂₄ (B) or (CGG)₂₁ (C).

(D and E) Replication reactions carried out in the absence of pol δ with a series of substrates containing increasing numbers of either guanine (D) or cytosine (E) leading strand homopolymers.

(F) Replication reactions carried out in the absence of pol δ with a series of substrates containing different leading strand homopolymeric templates as indicated.

Figure 3 - Pol δ drives recovery from (CG)₂₄ and (CGG)₆₁ stalls, but not (G)₅₀ or (C)₅₀

(A) Replication reactions carried out with the indicated templates in the absence or presence of pol δ .

(B and C) Pulse-chase experiments carried out with the (CGG)₆₁ template in the absence (B) or presence (C) of pol δ . Reactions were initiated with radiolabelled dATP for 10 min, chased with excess 'cold' dATP and samples taken at the indicated time points.

(D and E) Same as in (B and C) but with the (CG)₂₄ template.

(F) Pulse-chase experiments carried out with the indicated templates. Reactions were initiated without pol δ to generate a pre-existing stall. After a 10 min pulse, pol δ was either added with the chase or not and samples taken after another 10 min.

Figure 4 - DNA-induced stalls trigger helicase-polymerase uncoupling

(A and B) Replication reactions carried out with the indicated templates in the presence of a primer that anneals 265 nt downstream of the repeats or a scrambled control primer in the absence (A) or presence (B) of pol δ .

(C and D) Replication reactions carried out with the indicated templates in the absence (C) or presence (D) of pol δ analysed by native gels. The insets show a longer exposure of the dashed boxes to better visualize uncoupled products.

Figure 5 - Read-through of (CGG)_n and (CG)_n is facilitated by pol ϵ variants or elevated dNTPs

(A) Replication reactions carried out with the indicated templates with different pol ϵ variants in the absence of pol δ .

(B) Pulse-chase experiments carried out with the indicated templates in the absence (left panel) or presence (right panel) of pol δ . Reactions were initiated with radiolabelled dATP for 10 min and chased for another 10 min with either excess 'cold' dATP alone (dA) or with excess of all four dNTPs (dN).

Figure 6 - Pif1 resolves DNA-induced stalls

(A and B) Pulse-chase experiments carried out with the indicated templates in the presence (A) or absence (B) of pol δ . Reactions were initiated with radiolabelled dATP. After a 10 min pulse, either WT or ATPase-dead (K264A) pif1 was added with the chase and samples taken after another 10 min.

STAR METHODS

Cloning

All replication templates are based on the 9.8 kb pZN3 plasmid ([Taylor and Yeeles, 2018](#)), in which a new linker was inserted 3 kb downstream from the ARS306 origin, yielding pGC504. Repeats were cloned step-wise using a previously described method for expansion of repeats ([Scior et al., 2011](#)). Briefly, repeats were first cloned using annealed oligonucleotides. For the first expansion step, annealed duplexes were used as a source of insert. In subsequent steps, each resulting template was used both as a source of insert and as a target vector. The use of type IIS restriction enzymes (BsaI and Esp3I) allowed seamless cloning of uninterrupted repeats. Because of the unstable nature of some repeats, we first cloned repeats into a pSMART derivative in which we removed a BsaI site and introduced a new linker. Although this vector has been designed to better support unstable inserts, we found that repeats were overall more stable in the pZN3 backbone. We therefore removed two BsaI sites from pGC504, to generate pGC542, and from that point onward cloned all repeats directly into pGC542. To clone repeats in the reverse orientation we replaced the linker in pGC542 so that the PacI and NotI sites were reversed, yielding pGC558. See Table 1 for complete annotation of all plasmids used and generated in this study and Table 2 for a list of oligonucleotides.

Protein expression and purification

The expression and purification of most proteins used in this study have been described before ([Baretić et al., 2020](#); [Coster et al., 2014](#); [Deegan et al., 2019](#); [Douglas et al., 2018](#); [Frigola et al., 2013](#); [Goswami et al., 2018](#); [Hill et al., 2020](#); [On et al., 2014](#); [Yeeles et al., 2015](#); [Yeeles et al., 2017](#)). For full details see Table 3. To

generate the pol ϵ P301R expression strain (ySW1), a synthetic gene fragment spanning part of pol2 which contains the desired mutation (ordered as a gBlock, IDT) was cloned using HiFi assembly (New England Biolabs, E2621S) to replace the corresponding WT sequence in plasmid pAJ6, yielding plasmid pSW62. Plasmid pSW62 was linearised with Bsu36I, transformed into yeast strain yAE94 (Yeeles et al., 2015) and positive transformants were selected for on plates lacking TRP. Integration was confirmed by PCR of genomic DNA as described (Coster et al., 2014). WT and mutant pol ϵ variants were purified as previously described (Yeeles et al., 2015) except that yeast cultures were not synchronized.

Pif1 and Pif1 K264A were expressed and purified as described (Deegan et al., 2019) with the following modifications: imidazole concentrations were 0 mM during lysis, 15 mM during washes and 300 mM for elution. The eluate from the HIS pulldown was diluted 1:2 to reduce salt and loaded on a monoS column. Pif1 containing fractions were concentrated and loaded onto a 24 mL Superdex 200 column equilibrated in 0.15 mM NaCl. Pif1 was concentrated with a 30 kDa Amicon to 2 μ M.

For Sld2 expression, pGC441 was transformed into BL21 bacteria and grown overnight in a starter culture of 250 ml LB-broth with 100 μ g/ml ampicillin and 37 μ g/ml chloramphenicol at 37°C. The next day 20 ml per litre starter was added to 12 litre of LB with ampicillin and chloramphenicol, incubated at 37°C until OD(600) reached 0.5, then cooled on ice for 20 min, and IPTG was added to 0.2 mM IPTG. Induction took place at 16°C overnight. Cells were harvested by centrifugation and pellets resuspended in buffer S [25 mM HEPES pH 7.6, 10% glycerol, 0.02% NP-40, 0.1% Tween, 1 mM EDTA, 1 mM DTT] + 0.5 M NaCl and protease inhibitors, incubated with 0.1 mg/ml lysozyme for 20 mins at 4°C and sonicated for 4 mins (5 sec on / 5 sec off) on ice. The lysate was cleared by centrifugation at 15,000 rpm, 15 min, 4°C using a

JA-25.50 rotor. The cleared lysate was incubated 1 hour at 4°C with 2.4 ml of 20% glutathione agarose slurry pre-washed in lysis buffer. The beads were washed extensively with buffer S + 0.5 M NaCl and finally resuspended in 3 ml wash buffer with 200 µg PreScission protease and incubated on rotating wheel 2 hours at 4°C. The eluate was collected, the beads washed with 3 x 1 ml buffer S + 0.5 M NaCl and all fractions were pooled and diluted 1:2 in buffer S without salt. The sample was applied to a HiTrap SP FF 1 ml column equilibrated in buffer S + 250 mM NaCl. After washing with 20 CV of equilibration buffer Sld2 was eluted in 0.5 ml fractions with 100% buffer S + 700 mM NaCl for 10 CV. Sld2-containing fractions were pooled and concentrated with a 10 kDa Amicon to 1.1 µM.

Preparation of templates for replication assays

All plasmids were maintained in NEB Stable E. coli cells (New England Biolabs, C3040H) and purified using the HiSpeed Plasmid Maxi kits (Qiagen, 12663) from bacteria grown at 30°C to minimize loss or rearrangements of unstable inserts. We sometimes observed variability in the overall efficiency of in vitro replication between substrates, presumably due to a contaminant. This variability was eliminated by further purifying templates in batch using PlasmidSelect Xtra resin (VWR, 28-4024-02). Plasmid DNA was diluted fivefold in 3 M ammonium sulfate, added to 300 µl bead slurry pre-washed with 2.3 M ammonium sulfate and incubated for 30 min rotating. The beads were washed four times with 1.9 M ammonium sulfate. Supercoiled plasmid was eluted with 2 x 1 ml of 1.5 M ammonium sulfate. Both fractions were pooled and dialysed for 3 h and ON against 1 L of 0.1xTE buffer in the dark using in a D-Tube Dialyzer Mini, MWCO 6-8 kDa (Merck, 71504). The dialysed DNA sample was concentrated to 400 µl with a 100 kDa Amicon, precipitated with 1 ml 100% ethanol and 90 mM NaCl and resuspended in TE. All steps except for the ethanol precipitation

were performed at RT. All ammonium sulfate buffers contained 100 mM Tris HCl pH 7.5 and 10 mM EDTA.

In vitro replication assays

For MCM loading and phosphorylation, 3 nM plasmid DNA template was incubated with 5 mM ATP, 75 nM Cdt1/Mcm2-7, 45 nM Cdc6, 20 nM ORC, 50 nM DDK in 25 mM HEPES-KOH (pH 7.6), 100 mM potassium glutamate, 0.01% NP-40-S, 1 mM DTT, 10 mM Mg(OAc)₂, 0.1 mg/ml BSA (1x reaction buffer) and 80 mM KCl at 24°C for 10 min. For replication of linearised templates 0.6 U/μl AhdI was added during MCM loading. For truncated templates in Figure 4C and D, 2 U/μl EcoRV was also added. Loading was stopped by adding 120 nM S-CDK for further 5 min. The loaded reaction was diluted in 1x reaction buffer so that the final dilution in the replication reaction was 6-fold. To start replication the following components were added to the reaction: 200 μM CTP, GTP, UTP, 30 μM dATP, dCTP, dGTP, dTTP, 33 nM α-[³³P]-dATP, 5 mM ATP, 10 nM S-CDK, 30 nM Dpb11, 100 nM GINS, 40 nM Cdc45, 20 nM Pol ε, 10 nM Mcm10, 40 nM RPA, 20 nM Csm3/Tof1, 20 nM Mrc1, 30 nM RFC, 40 nM PCNA, 10 nM Topol (for circular reactions), 40 nM Pol α, 5 nM Pol δ (where indicated), 20 nM Sld3/7, 20 nM Sld2, and the mix was incubated at 30°C for the indicated time. For samples loaded on denaturing gels, 0.5 U/μl SmaI was added 5 min before the end of the reaction, which eliminates product length heterogeneity which stems from variable initiation sites (Taylor and Yeeles, 2018). Reactions were stopped with 100 mM EDTA. For pulse-chase experiments, unlabeled deoxyribonucleotide concentrations were adjusted during the pulse to 30 μM dCTP, dTTP, dGTP and 2.5 μM dATP, or 7.5 μM dATP for experiments without RFC/PCNA. After a 10 min pulse, the chase was performed by adding 200 μM unlabelled dATP, or for Figure 5B 400 μM of dATP alone or dCTP, dTTP, dGTP and dATP, as indicated. Reactions were

stopped at the indicated time point by addition of EDTA to 100 mM. For repriming experiments: oligonucleotides were added to 60 nM (molecules) before starting the replication reaction.

Post-reaction sample processing

For samples to be analysed on denaturing gels, alkaline loading dye (0.5 M NaOH, 10% sucrose, xylene cyanol in water) was added at 1/10 volume. Samples were loaded in denaturing 0.8% agarose gels run at 32 V overnight in 30 mM NaOH, 2 mM EDTA.

For reactions to be loaded on native gels, SDS (to 0.1%) and proteinase K (1/100 volumes) were added and incubated at 37°C for 20 min. The sample volume was increased to 25 µl with TE and DNA was extracted with phenol:chloroform:isoamyl alcohol 25:24:1 (Sigma-Aldrich, P2069). The extracted sample was mixed with 5x Invitrogen™ Novex™ High-Density TBE Sample Buffer and loaded on a 1% agarose / TAE gel.

Substrate preparation for helicase assays

Complementary oligonucleotides containing a 5' overhang were resuspended to 10 µM in 10 mM Tris pH-8.0. One oligo was labelled in a reaction containing 5 pmol of DNA, 1X PNK buffer, 1U of PNK enzyme (NEB, M0201S), and γ-P32-ATP (0.03 mCi). The reaction was incubated for 30 min at 37°C and subsequently heat inactivated for 20 min at 65°C. Excess γ-P32-ATP was then cleared using a G50 column (GE healthcare, 2753002) and volume adjusted to 100 µl (=50 nM). To generate duplex DNA 1 pmol of labelled oligo was mixed with 1.5 pmol of unlabelled oligo and incubated at 90°C for 5 min in a thermal cycler. The mix was then gradually cooled down to room temperature over 2 hours. Duplex DNA was stored at -20°C.

Helicase assays

Helicase assays were carried out using 0.5 nM γ -P32-ATP labelled duplex with a 5' overhang in buffer containing 25 mM Hepes 7.6, 2 mM MgOAc, 0.1 mg/ml BSA and 2 mM ATP. Reactions were assembled on ice, equilibrated to room temperature and the respective helicases (Pif1 or Dna2) added to 50 nM final concentration. Reactions were incubated for 30 min at 30°C and samples collected at different time points (5, 10 and 20 min). Reactions were stopped by addition of 0.5% SDS and 200 mM EDTA. The samples were supplemented with Novex Hi-Density TBE Sample buffer (ThermoFisher Scientific, LC6678) and analysed on 10% Novex TBE gels (ThermoFisher Scientific, EC62755BOX) at 150V for 1 hour in 1X TBE. Gels were dried onto filter paper, autoradiographed with phosphoscreens imaging plates (Fujifilm) and developed on a Typhoon phosphorimager (GE Healthcare).

CPD Substrate

Preparation of a substrate containing site-specific DNA damage (CPD) was prepared as previously described ([Taylor and Yeeles, 2018](#)) with several modifications. An oligonucleotide containing a CPD (AfIII CPD, HPLC-purified; TriLink Biotechnology) was synthesised and stored in 10 mM Tris-HCl (pH 8.0), 1 mM EDTA at -20°C. To introduce the oligo into the plasmid of interest (pGC504), 4 x 200 μ g of the relevant plasmid was cut with 15 μ l (150U) of Nt.BbvCI (NEB, R0632) in a 200 μ l final volume reaction at 37°C for 3 hours. The reaction was stopped by adding 50 mM EDTA. Following digestion, competitor oligonucleotide (AfIII competitor, IDT) was added to 1000-fold molar excess over plasmid concentration (27 μ L from 1 mM Stock). The mix was incubated at 50°C for 20 min, then transferred to 37°C and SDS added to 0.1%. After 5 min, 1/100 volumes of proteinase K (New England Biolabs P8107S) was added and incubated at 37°C for a further 15 min. All tubes were then pooled and the gapped

plasmid purified. Excess oligo was separated from gapped plasmid using High prep PCR magnetic beads (Magbio, AC-60050) with a ratio of 1.8 μ l of bead slurry/ μ l of sample and binding for 30 min at room temperature. Bound fractions were washed 3 times with a mix containing 70% EtOH and 0.02% NP-40 and then eluted in 1X TE. DNA was pooled and concentration measured. This step usually yielded around 60% of input material.

100 μ g of gapped plasmid was collected per oligonucleotide ligation. Complementary oligonucleotide containing a CPD (AflII CPD) was added at a 20-fold molar excess and incubated at 50°C for 15 min before gradually letting it cool down to room temperature. 100 μ g of DNA was ligated in 1X T4 DNA ligase buffer (NEB: B0202S) and T4 ligase (100U/ μ g) (NEB: M0202M) plus 2 mM Mg(OAc)₂ overnight at 16°C in the dark. The following day, SDS (to 0.1%) and proteinase K (1/100 volumes) were added and incubated at 37°C for 20 min. The ligated plasmid was then subjected to CsCl gradients as in (Taylor and Yeeles, 2018) to specifically purify fully ligated supercoiled substrates. Following the CsCl gradient DNA was dialyzed against two changes of 2 L TE over 16 h total in a D-Tube Dialyzer Midi, MWCO 6-8 kDa (Merck 71507) at 4°C in the dark to remove all traces of CsCl. The DNA was collected and subjected to ethanol precipitation using 0.3 M NaCl + 2.8 volumes ice cold 100% ethanol in dry ice. The pellet was harvested, washed with room temperature 70% ethanol, harvested, air-dried and resuspended in 50 μ L TE. As a control, the exact same procedure was also carried out with an undamaged oligo (AflII undamaged) and the resulting template replicated in the same manner as the parental template, indicating that the observed stalling was induced by the CPD and not due to the process itself.

Analytical digestion of substrates

Substrates of interest were subjected to enzymatic digestion to verify the length of the repetitive sequence. Briefly, 100 ng of plasmid was digested with 0.5U of NotI (NEB, R0189L) and PaeI (NEB, R0547L) in 1X Cutsmart buffer (NEB, B7204S) at 37°C for 30 min. Reactions were stopped by adding 50 mM EDTA. The samples were then supplemented with Novex Hi-Density TBE Sample buffer (ThermoFisher Scientific, LC6678) and analysed on 10% Novex TBE gels (ThermoFisher Scientific, EC62755BOX) at 150V for 1 hour in 1X TBE. Gels were then stained with SYBR™ Gold Nucleic Acid Gel Stain (Invitrogen, S11494) for 20 min at room temperature in the dark and imaged on a Typhoon phosphorimager (GE Healthcare).

REFERENCES

- Amrane, S., and Mergny, J.L. (2006). Length and pH-dependent energetics of (CCG)_n and (CGG)_n trinucleotide repeats. *Biochimie* 88, 1125-1134.
- Anand, R.P., Shah, K.A., Niu, H., Sung, P., Mirkin, S.M., and Freudenreich, C.H. (2012). Overcoming natural replication barriers: differential helicase requirements. *Nucleic Acids Res* 40, 1091-1105.
- Baretić, D., Jenkyn-Bedford, M., Aria, V., Cannone, G., Skehel, M., and Yeeles, J.T.P. (2020). Cryo-EM Structure of the Fork Protection Complex Bound to CMG at a Replication Fork. *Molecular Cell* 78, 926-940.e913.
- Bell, S.P., and Labib, K. (2016). Chromosome Duplication in *Saccharomyces cerevisiae*. *Genetics* 203, 1027-1067.
- Buzovetsky, O., Kwon, Y., Pham, N.T., Kim, C., Ira, G., Sung, P., and Xiong, Y. (2017). Role of the Pif1-PCNA Complex in Pol delta-Dependent Strand Displacement DNA Synthesis and Break-Induced Replication. *Cell Reports* 21, 1707-1714.
- Byrd, A.K., Bell, M.R., and Raney, K.D. (2018). Pif1 helicase unfolding of G-quadruplex DNA is highly dependent on sequence and reaction conditions. *Journal of Biological Chemistry* 293, 17792-17802.
- Coster, G., Frigola, J., Beuron, F., Morris, E.P., and Diffley, J.F. (2014). Origin licensing requires ATP binding and hydrolysis by the MCM replicative helicase. *Mol Cell* 55, 666-677.
- Dahan, D., Tsirkas, I., Dovrat, D., Sparks, M.A., Singh, S.P., Galletto, R., and Aharoni, A. (2018). Pif1 is essential for efficient replisome progression through lagging strand G-quadruplex DNA secondary structures. *Nucleic Acids Research* 46, 11847-11857.
- Deegan, T.D., Baxter, J., Ortiz Bazán, M.Á., Yeeles, J.T.P., and Labib, K.P.M. (2019). Pif1-Family Helicases Support Fork Convergence during DNA Replication Termination in Eukaryotes. *Molecular Cell*.
- Delagoutte, E., Goellner, G.M., Guo, J., Baldacci, G., and McMurray, C.T. (2008). Single-stranded DNA-binding protein in vitro eliminates the orientation-dependent impediment to polymerase passage on CAG/CTG repeats. *J Biol Chem* 283, 13341-13356.
- Devbhandari, S., and Remus, D. (2020). Rad53 limits CMG helicase uncoupling from DNA synthesis at replication forks. *Nat Struct Mol Biol* 27, 461-471.
- Douglas, M.E., Ali, F.A., Costa, A., and Diffley, J.F.X. (2018). The mechanism of eukaryotic CMG helicase activation. *Nature* 555, 265.
- Fan, J.-H., Bochkareva, E., Bochkarev, A., and Gray, D.M. (2009). Circular Dichroism Spectra and Electrophoretic Mobility Shift Assays Show That Human Replication Protein A Binds and Melts Intramolecular G-Quadruplex Structures. *Biochemistry* 48, 1099-1111.
- Fojtík, P., Kejnovská, I., and Vorlíčková, M. (2004). The guanine-rich fragile X chromosome repeats are reluctant to form tetraplexes. *Nucleic Acids Research* 32, 298-306.
- Follonier, C., Oehler, J., Herrador, R., and Lopes, M. (2013). Friedreich's ataxia-associated GAA repeats induce replication-fork reversal and unusual molecular junctions. *Nature Structural & Molecular Biology* 20, 486-494.
- Frank-Kamenetskii, M.D., and Mirkin, S.M. (1995). Triplex DNA structures. *Annu Rev Biochem* 64, 65-95.
- Frigola, J., Remus, D., Mehanna, A., and Diffley, J.F.X. (2013). ATPase-dependent quality control of DNA replication origin licensing. *Nature* 495, 339-343.
- Fry, M., and Loeb, L.A. (1994). The fragile X syndrome d(CGG)_n nucleotide repeats form a stable tetrahelical structure. *Proceedings of the National Academy of Sciences* 91, 4950-4954.
- Gacy, A.M., Goellner, G.M., Spiro, C., Chen, X., Gupta, G., Bradbury, E.M., Dyer, R.B., Mikesell, M.J., Yao, J.Z., Johnson, A.J., et al. (1998). GAA Instability in Friedreich's Ataxia Shares a Common, DNA-Directed and Intraallelic Mechanism with Other Trinucleotide Diseases. *Molecular Cell* 1, 583-593.
- Gacy, A.M., and McMurray, C.T. (1998). Influence of hairpins on template reannealing at trinucleotide repeat duplexes: a model for slipped DNA. *Biochemistry* 37, 9426-9434.

- Gaillard, H., García-Muse, T., and Aguilera, A. (2015). Replication stress and cancer. *Nat. Rev. Cancer* 15, 276-289.
- Ganai, R.A., Zhang, X.-P., Heyer, W.-D., and Johansson, E. (2016). Strand displacement synthesis by yeast DNA polymerase ϵ . *Nucleic acids research* 44, 8229-8240.
- Gehring, K., Leroy, J.-L., and Guéron, M. (1993). A tetrameric DNA structure with protonated cytosine-cytosine base pairs. *Nature* 363, 561-565.
- Gellert, M., Lipsett, M.N., and Davies, D.R. (1962). Helix formation by guanylic acid. *Proceedings of the National Academy of Sciences* 48, 2013-2018.
- Gerhardt, J., Bhalla, Angela D., Butler, Jill S., Puckett, James W., Dervan, Peter B., Rosenwaks, Z., and Napierala, M. (2016). Stalled DNA Replication Forks at the Endogenous GAA Repeats Drive Repeat Expansion in Friedreich's Ataxia Cells. *Cell Reports* 16, 1218-1227.
- Gerhardt, J., Tomishima, Mark J., Zaninovic, N., Colak, D., Yan, Z., Zhan, Q., Rosenwaks, Z., Jaffrey, Samie R., and Schildkraut, Carl L. (2014). The DNA Replication Program Is Altered at the FMR1 Locus in Fragile X Embryonic Stem Cells. *Molecular Cell* 53, 19-31.
- Goswami, P., Abid Ali, F., Douglas, M.E., Locke, J., Purkiss, A., Janska, A., Eickhoff, P., Early, A., Nans, A., Cheung, A.M.C., *et al.* (2018). Structure of DNA-CMG-Pol epsilon elucidates the roles of the non-catalytic polymerase modules in the eukaryotic replisome. *Nature Communications* 9, 5061.
- Guilliam, T.A., and Yeeles, J.T. (2021). The eukaryotic replisome tolerates leading-strand base damage by replicase switching. *The EMBO Journal* n/a, e107037.
- Guilliam, T.A., and Yeeles, J.T.P. (2020). Reconstitution of translesion synthesis reveals a mechanism of eukaryotic DNA replication restart. *Nature Structural & Molecular Biology* 27, 450-460.
- Hartenstine, M.J., Goodman, M.F., and Petruska, J. (2000). Base stacking and even/odd behavior of hairpin loops in DNA triplet repeat slippage and expansion with DNA polymerase. *J Biol Chem* 275.
- Hill, J., Eickhoff, P., Drury, L.S., Costa, A., and Diffley, J.F.X. (2020). The eukaryotic replisome requires an additional helicase to disarm dormant replication origins. *bioRxiv*, 2020.2009.2017.301366.
- Hirst, M.C., and White, P.J. (1998). Cloned human FMR1 trinucleotide repeats exhibit a length- and orientation-dependent instability suggestive of in vivo lagging strand secondary structure. *Nucleic Acids Research* 26, 2353-2358.
- Kamath-Loeb, A.S., Loeb, L.A., Johansson, E., Burgers, P.M.J., and Fry, M. (2001). Interactions between the Werner Syndrome Helicase and DNA Polymerase Delta Specifically Facilitate Copying of Tetraplex and Hairpin Structures of the d(CGG)_n Trinucleotide Repeat Sequence. *Journal of Biological Chemistry* 276, 16439-16446.
- Kang, S., Ohshima, K., Shimizu, M., Amirhaeri, S., and Wells, R.D. (1995). Pausing of DNA Synthesis in Vitro at Specific Loci in CTG and CGG Triplet Repeats from Human Hereditary Disease Genes. *Journal of Biological Chemistry* 270, 27014-27021.
- Khristich, A.N., and Mirkin, S.M. (2020). On the wrong DNA track: Molecular mechanisms of repeat-mediated genome instability. *Journal of Biological Chemistry* 295, 4134-4170.
- Kim, H.M., Narayanan, V., Mieczkowski, P.A., Petes, T.D., Krasilnikova, M.M., Mirkin, S.M., and Lobachev, K.S. (2008). Chromosome fragility at GAA tracts in yeast depends on repeat orientation and requires mismatch repair. *EMBO J* 27, 2896-2906.
- Koc, K.N., Singh, S.P., Stodola, J.L., Burgers, P.M., and Galletto, R. (2016). Pif1 removes a Rap1-dependent barrier to the strand displacement activity of DNA polymerase δ . *Nucleic Acids Research* 44, 3811-3819.
- Krasilnikova, M.M., and Mirkin, S.M. (2004). Replication stalling at Friedreich's ataxia (GAA)_n repeats in vivo. *Mol Cell Biol* 24.
- Kunkel, T.A. (1990). Misalignment-mediated DNA synthesis errors. *Biochemistry* 29.
- Lai, P.J., Lim, C.T., Le, H.P., Katayama, T., Leach, D.R.F., Furukohri, A., and Maki, H. (2016). Long inverted repeat transiently stalls DNA replication by forming hairpin structures on both leading and lagging strands. *Genes to Cells* 21, 136-145.

Le, H.P., Masuda, Y., Tsurimoto, T., Maki, S., Katayama, T., Furukohri, A., and Maki, H. (2015). Short CCG repeat in huntingtin gene is an obstacle for replicative DNA polymerases, potentially hampering progression of replication fork. *Genes to Cells* 20, 817-833.

Lee, W.T.C., Yin, Y., Morten, M.J., Tonzi, P., Gwo, P.P., Odermatt, D.C., Modesti, M., Cantor, S.B., Gari, K., Huang, T.T., *et al.* (2021). Single-molecule imaging reveals replication fork coupled formation of G-quadruplex structures hinders local replication stress signaling. *Nature Communications* 12, 2525.

Lopes, J., Piazza, A., Bermejo, R., Kriegsman, B., Colosio, A., Teulade-Fichou, M.-P., Foiani, M., and Nicolas, A. (2011). G-quadruplex-induced instability during leading-strand replication. *The EMBO Journal* 30, 4033-4046.

Lopez Castel, A., Cleary, J.D., and Pearson, C.E. (2010). Repeat instability as the basis for human diseases and as a potential target for therapy. *Nat Rev Mol Cell Biol* 11, 165-170.

Maestroni, L., Audry, J., Luciano, P., Coulon, S., Géli, V., and Corda, Y. (2020). RPA and Pif1 cooperate to remove G-rich structures at both leading and lagging strand. *Cell Stress* 4, 48-63.

McMurray, C.T. (2010). Mechanisms of trinucleotide repeat instability during human development. *Nat Rev Genet* 11, 786-799.

Mirkin, S.M. (2007). Expandable DNA repeats and human disease. *Nature* 447, 932-940.

Murat, P., Guilbaud, G., and Sale, J.E. (2020). DNA polymerase stalling at structured DNA constrains the expansion of short tandem repeats. *Genome Biology* 21, 209.

Ohshima, K., and Wells, R.D. (1997). Hairpin formation during DNA synthesis primer realignment in vitro in triplet repeat sequences from human hereditary disease genes. *Journal of Biological Chemistry* 272, 16798-16806.

On, K.F., Beuron, F., Frith, D., Snijders, A.P., Morris, E.P., and Diffley, J.F. (2014). Prereplicative complexes assembled in vitro support origin-dependent and independent DNA replication. *EMBO J* 33, 605-620.

Osmundson, J.S., Kumar, J., Yeung, R., and Smith, D.J. (2017). Pif1-family helicases cooperatively suppress widespread replication-fork arrest at tRNA genes. *Nature Structural & Molecular Biology* 24, 162-170.

Paeschke, K., Bochman, M.L., Garcia, P.D., Cejka, P., Friedman, K.L., Kowalczykowski, S.C., and Zakian, V.A. (2013). Pif1 family helicases suppress genome instability at G-quadruplex motifs. *Nature* 497, 458-462.

Paeschke, K., Capra, John A., and Zakian, Virginia A. (2011). DNA Replication through G-Quadruplex Motifs Is Promoted by the *Saccharomyces cerevisiae* Pif1 DNA Helicase. *Cell* 145, 678-691.

Parkash, V., Kulkarni, Y., ter Beek, J., Shcherbakova, P.V., Kamerlin, S.C.L., and Johansson, E. (2019). Structural consequence of the most frequently recurring cancer-associated substitution in DNA polymerase ϵ . *Nature Communications* 10, 373.

Pelletier, R., Krasilnikova, M.M., Samadashwily, G.M., Lahue, R., and Mirkin, S.M. (2003). Replication and expansion of trinucleotide repeats in yeast. *Mol Cell Biol* 23, 1349-1357.

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.

Pfeifer, G.P. (2006). Mutagenesis at Methylated CpG Sequences. In *DNA Methylation: Basic Mechanisms*, W. Doerfler, and P. Böhm, eds. (Berlin, Heidelberg: Springer Berlin Heidelberg), pp. 259-281.

Ray, S., Qureshi, Mohammad H., Malcolm, Dominic W., Budhathoki, Jagat B., Çelik, U., and Balci, H. (2013). RPA-Mediated Unfolding of Systematically Varying G-Quadruplex Structures. *Biophysical Journal* 104, 2235-2245.

Renčiuk, D., Kypr, J., and Vorlíčková, M. (2011). CGG repeats associated with fragile X chromosome form left-handed Z-DNA structure. *Biopolymers* 95, 174-181.

- Ribeyre, C., Lopes, J., Boulé, J.-B., Piazza, A., Guédin, A., Zakian, V.A., Mergny, J.-L., and Nicolas, A. (2009). The Yeast Pif1 Helicase Prevents Genomic Instability Caused by G-Quadruplex-Forming CEB1 Sequences In Vivo. *PLOS Genetics* 5, e1000475.
- Rossi, M.L., Pike, J.E., Wang, W., Burgers, P.M.J., Campbell, J.L., and Bambara, R.A. (2008). Pif1 Helicase Directs Eukaryotic Okazaki Fragments toward the Two-nuclease Cleavage Pathway for Primer Removal. *Journal of Biological Chemistry* 283, 27483-27493.
- Saini, N., Ramakrishnan, S., Elango, R., Ayyar, S., Zhang, Y., Deem, A., Ira, G., Haber, J.E., Lobachev, K.S., and Malkova, A. (2013). Migrating bubble during break-induced replication drives conservative DNA synthesis. *Nature* 502, 389-392.
- Salas, T.R., Petruseva, I., Lavrik, O., Bourdoncle, A., Mergny, J.-L., Favre, A., and Saintomé, C. (2006). Human replication protein A unfolds telomeric G-quadruplexes. *Nucleic Acids Research* 34, 4857-4865.
- Sarkies, P., Reams, C., Simpson, L.J., and Sale, J.E. (2010). Epigenetic Instability due to Defective Replication of Structured DNA. *Molecular Cell* 40, 703-713.
- Sauer, M., and Paeschke, K. (2017). G-quadruplex unwinding helicases and their function in vivo. *Biochemical Society Transactions* 45, 1173-1182.
- Scior, A., Preissler, S., Koch, M., and Deuerling, E. (2011). Directed PCR-free engineering of highly repetitive DNA sequences. *BMC Biotechnology* 11, 87.
- Shastri, N., Tsai, Y.-C., Hile, S., Jordan, D., Powell, B., Chen, J., Maloney, D., Dose, M., Lo, Y., Anastassiadis, T., *et al.* (2018). Genome-wide Identification of Structure-Forming Repeats as Principal Sites of Fork Collapse upon ATR Inhibition. *Molecular Cell* 72, 222-238.e211.
- Shimizu, M., Gellibolian, R., Oostra, B.A., and Wells, R.D. (1996). Cloning, Characterization and Properties of Plasmids Containing CGG Triplet Repeats from the FMR-1 Gene. *Journal of Molecular Biology* 258, 614-626.
- Shishkin, A.A., Voineagu, I., Matera, R., Cherng, N., Chernet, B.T., Krasilnikova, M.M., Narayanan, V., Lobachev, K.S., and Mirkin, S.M. (2009). Large-scale expansions of Friedreich's ataxia GAA repeats in yeast. *Mol Cell* 35, 82-92.
- Sinden, R.R. (1994). Chapter 4 - Cruciform Structures in DNA. In *DNA Structure and Function*, R.R. Sinden, ed. (San Diego: Academic Press), pp. 134-178.
- Singleton, C.K., Klysik, J., Stirdivant, S.M., and Wells, R.D. (1982). Left-handed Z-DNA is induced by supercoiling in physiological ionic conditions. *Nature* 299, 312-316.
- Sparks, J.L., Chistol, G., Gao, A.O., Räsche, M., Larsen, N.B., Mann, M., Duxin, J.P., and Walter, J.C. (2019a). The CMG Helicase Bypasses DNA-Protein Cross-Links to Facilitate Their Repair. *Cell* 176, 167-181.e121.
- Sparks, M.A., Singh, S.P., Burgers, P.M., and Galletto, R. (2019b). Complementary roles of Pif1 helicase and single stranded DNA binding proteins in stimulating DNA replication through G-quadruplexes. *Nucleic Acids Research* 47, 8595-8605.
- Srivastava, S., Avvaru, A.K., Sowpati, D.T., and Mishra, R.K. (2019). Patterns of microsatellite distribution across eukaryotic genomes. *BMC Genomics* 20, 153.
- Šviković, S., Crisp, A., Tan-Wong, S.M., Guillian, T.A., Doherty, A.J., Proudfoot, N.J., Guilbaud, G., and Sale, J.E. (2019). R-loop formation during S phase is restricted by PrimPol-mediated repriming. *The EMBO Journal* 38, e99793.
- Taylor, M.R.G., and Yeeles, J.T.P. (2018). The Initial Response of a Eukaryotic Replisome to DNA Damage. *Molecular Cell* 70, 1067-1080.e1012.
- Taylor, M.R.G., and Yeeles, J.T.P. (2019). Dynamics of Replication Fork Progression Following Helicase–Polymerase Uncoupling in Eukaryotes. *Journal of Molecular Biology* 431, 2040-2049.
- Tubbs, A., Sridharan, S., van Wietmarschen, N., Maman, Y., Callen, E., Stanlie, A., Wu, W., Wu, X., Day, A., Wong, N., *et al.* (2018). Dual Roles of Poly(dA:dT) Tracts in Replication Initiation and Fork Collapse. *Cell* 174, 1127-1142.e1119.
- Usdin, K., and Woodford, K.J. (1995). CGG repeats associated with DNA instability and chromosome fragility form structures that block DNA-synthesis in-vitro. *Nucleic Acids Research* 23, 4202-4209.

- van Wietmarschen, N., Sridharan, S., Nathan, W.J., Tubbs, A., Chan, E.M., Callen, E., Wu, W., Belinky, F., Tripathi, V., Wong, N., *et al.* (2020). Repeat expansions confer WRN dependence in microsatellite-unstable cancers. *Nature* **586**, 292-298.
- Vetcher, A.A., Napierala, M., Iyer, R.R., Chastain, P.D., Griffith, J.D., and Wells, R.D. (2002). Sticky DNA, a Long GAA·GAA·TTC Triplex That Is Formed Intramolecularly, in the Sequence of Intron 1 of the Frataxin Gene. *Journal of Biological Chemistry* **277**, 39217-39227.
- Viterbo, D., Michoud, G., Mosbach, V., Dujon, B., and Richard, G.-F. (2016). Replication stalling and heteroduplex formation within CAG/CTG trinucleotide repeats by mismatch repair. *DNA Repair* **42**, 94-106.
- Voineagu, I., Surka, C.F., Shishkin, A.A., Krasilnikova, M.M., and Mirkin, S.M. (2009). Replisome stalling and stabilization at CGG repeats, which are responsible for chromosomal fragility. *Nat Struct Mol Biol* **16**, 226-228.
- Wasserman, M.R., Schauer, G.D., O'Donnell, M.E., and Liu, S. (2019). Replication Fork Activation Is Enabled by a Single-Stranded DNA Gate in CMG Helicase. *Cell* **178**, 600-611.e616.
- Wells, R.D. (2008). DNA triplexes and Friedreich ataxia. *Faseb Journal* **22**, 1625-1634.
- Wilson, M.A., Kwon, Y., Xu, Y., Chung, W.-H., Chi, P., Niu, H., Mayle, R., Chen, X., Malkova, A., Sung, P., *et al.* (2013). Pif1 helicase and Polδ promote recombination-coupled DNA synthesis via bubble migration. *Nature* **502**, 393-396.
- Xing, X., Kane, D.P., Bullock, C.R., Moore, E.A., Sharma, S., Chabes, A., and Shcherbakova, P.V. (2019). A recurrent cancer-associated substitution in DNA polymerase ε produces a hyperactive enzyme. *Nature Communications* **10**, 374.
- Yeeles, J.T., Deegan, T.D., Janska, A., Early, A., and Diffley, J.F. (2015). Regulated eukaryotic DNA replication origin firing with purified proteins. *Nature* **519**, 431-435.
- Yeeles, J.T.P., Janska, A., Early, A., and Diffley, J.F.X. (2017). How the Eukaryotic Replisome Achieves Rapid and Efficient DNA Replication. *Molecular Cell* **65**, 105-116.
- Yuan, Z., Georgescu, R., Schauer, G.D., O'Donnell, M.E., and Li, H. (2020). Structure of the polymerase ε holoenzyme and atomic model of the leading strand replisome. *Nature Communications* **11**, 3156.
- Zhou, J.C., Janska, A., Goswami, P., Renault, L., Abid Ali, F., Kotecha, A., Diffley, J.F.X., and Costa, A. (2017). CMG–Pol epsilon dynamics suggests a mechanism for the establishment of leading-strand synthesis in the eukaryotic replisome. *Proceedings of the National Academy of Sciences* **114**, 4141-4146.

Figure 1

Figure 2

Figure 3

Figure 4

Figure 5

A

B

Figure 6