The Response of the Eukaryotic Replisome to G-quadruplex- and i-motif-forming Sequences

Sophie L. Williams[1], Corella S. Casas-Delucchi[1], Federica Raguseo[2, 3], Dilek Guneri[4], Masashi Minamino[5], Emma Elisabeth Fletcher[6], Joseph T. P. Yeeles[6], Zoë A. E. Waller[4], Marco Di Antonio[2, 3, 7], Gideon Coster[1]

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

[2] Imperial College London, Chemistry Department, MSRH, 82 Wood Lane, London W12 0BZ, UK

[3] Institute of Chemical Biology, MSRH, 82 Wood Lane, London W12 0BZ, UK

[4]UCL, School of Pharmacy, 29-39 Brunswick Square, London WC1N 1AX, UK

[5] Francis Crick Institute, Chromosome Segregation Lab, 1 Midland Rd, London NW1 1AT, UK

[6]MRC Laboratory of Molecular Biology, Cambridge, UK

[7] Francis Crick Institute, 1 Midland Rd, London NW1 1AT, UK

Abstract

DNA replication is a tightly regulated, complex process that ensures the entire genome is inherited accurately in each cell division. Secondary DNA structures such as G-quadruplexes (G4s) and intercalated-Motifs (i-motifs) can pose a challenge to the replication machinery and in turn threaten genome stability. Multiple lines of evidence suggest G4s interfere with replication, but the underlying mechanism remains unclear. Moreover, there is a lack of evidence of how i-motifs affect the replisome. Here, we have reconstituted replication of physiologically derived structure-forming sequences and found that a single G4 or i-motif is sufficient to arrest DNA replication. This ability to stall replication correlates with the stability of the secondary structure under physiological conditions. Both G4s and i-motifs trigger-helicase polymerase uncoupling, which can only be rescued by a specialised helicase, Pif1. This study provides a potential mechanism for quadruplex structure formation and resolution during replication and highlights G4s and i-motifs as endogenous sources of replication stress, which may explain their genomic instability and mutation frequencies in cancer.

Introduction

Eukaryotic DNA replication is a highly regulated process carried out by a complex molecular machine known as the replisome[1]. Parental duplex DNA is unwound by the replicative helicase, CMG, and nascent DNA is subsequently synthesised by DNA polymerase ε on the leading strand, or DNA polymerase δ on the lagging strand. The replisome must replicate all regions of the genome accurately while encountering various obstacles such as DNA damage, protein barriers and transcription-replication collisions. All of these can lead to replication stress, which poses a challenge to genome integrity[2].

Another potential barrier to replisome progression is the DNA template itself. Aside from the canonical B-DNA conformation, certain DNA sequences can fold into secondary structures, particularly from ssDNA exposed during replication. Examples of well-characterised secondary structures include hairpin[3,4], G-quadruplex (G4)[5] and intercalated-Motif (i-motif)[6] structures, which are thought to act as barriers to replication. For example, hairpin-forming repeats cause replication stalling in yeast[7]. Moreover, two unbiased studies have mapped sites of replication fork collapse *in vivo*, highlighting poly(dA) sites[8] and a variety of structure-forming repeats[9]. Therefore, secondary structures may be responsible for the fact that repetitive sequences are unstable and give rise to genomic instability (reviewed in [10]).

G4 structures are formed from guanine-rich sequences through the stacking of G quartets formed by Hoogsteen base pairing[11]. Their abundance increases during S-phase, suggesting they arise as a consequence of replication[12]. G4-forming sequences are defined by a motif consisting of four consecutive tracts of at least three guanines (G-tetrads), separated by one to seven non-G nucleotides ($G_3N_{1-7}G_3N_{1-7}G_3N_{1-7}G_3$)[13] which are referred to as 'loops'. The length and composition of these loops affects the stability of the G4 structure. For example, shorter loops composed of thymine residues correlate with more thermally stable structures, while longer loops result in less stable structures[14]. G4s can exist in a variety of structural topologies, depending on the orientation of the G-rich strands relative to one another. Where all strands are oriented in the same direction, structures are defined as parallel. In contrast, anti-parallel structures consist of strands oriented in opposite directions. Where a structure consists of a combination of parallel and anti-parallel strands, it is

defined as a hybrid[15]. G4-forming sequences are abundant in the human genome and are known to have a range of physiological roles. Firstly, they have been associated with telomere maintenance[16]. Secondly, they are enriched in promoter regions[17] where they can affect the transcriptional state of genes (reviewed in [18]). Importantly, these structures are also implicated in human diseases. For example, G4s correlate with telomere fragility and induce both genetic and epigenetic instability[19-21] and have been shown to disrupt repressive chromatin structures[22]. More recently, ChIP-Seq of G4-structures on a genome-wide scale revealed that these structures mark sites of actively transcribed genes[23]. Moreover, G4s, such as those found at the *c-myc* gene promoter, are enriched at mutation hotspots in an array of cancers[24,25].

i-motifs arise within cytosine-rich DNA and therefore can often be found on the opposite strand to G4s, such as at telomeres. These structures are held together by hemi-protonated cytosine-cytosine base pairs (C:C+) and are stabilised by low pH[6]. Although the existence of i-motifs has long been established, there has been speculation about their biological relevance due to their apparent requirement for acidic conditions. However, recent studies have shown that certain i-motifs can form at neutral pH[26]. Under physiological conditions, longer tracts of cytosines are required for stable structure formation. Therefore, some examples of i-motifs capable of forming at physiological pH have been defined by sequences comprised of four consecutive tracts of at least five cytosines, separated by one to nineteen non-C nucleotides ($(C_5(N_{1-19}C_5)_3)$)[26]. Similar to G4s, a longer tract of Cs and a shorter loop length creates a more stable i-motif. In contrast to G4s, i-motifs can only adopt an anti-parallel conformation. i-motifs can be found at transcription start sites, and in the promoters of oncogenes such as *BCL2*[27], *HRAS*[28], *KRAS*[29], *c-MYC*[30] and *VEGF*[31].

Many studies have indicated that G4s can impact replication *in vivo.* In the absence of specialised helicases, such as Pif1, G4s positioned on the leading strand template can undergo rearrangements and mutations[32]. While the replication machinery stalls at G4-forming sequences in a variety of organisms, there is conflicting evidence of stalling on the leading versus lagging strand template[22,32-34]. In addition, a recent super-resolution microscopy study demonstrated that replication-coupled G4s interfere with replication stress signalling[35]. Despite the abundance of evidence that G4s can interfere with replication, the underlying trigger remains unknown. There are at least three possible causes: (i) The DNA sequence itself, (ii) the secondary structure

formed or (iii) bound proteins. These possibilities are not mutually exclusive. Moreover, it remains unclear what the mechanism of stalling is, and which component of the replisome is affected.

Primer extension assays have demonstrated that structure-forming sequences inhibit synthesis. A variety of G4-forming sequences are sufficient to inhibit a range of polymerases *in vitro,* including the replicative polymerases ε and δ[36-39]. Interestingly, there is some evidence that polymerase inhibition is affected not only by the stability of the G4, but also its topology[40]. DNA synthesis through these G4s can be aided by accessory factors such as specialised G4-unwinding helicases, including Pif1, REV1 and FANCJ[39,41-43]. However, these studies were carried out with pre-folded G4 structures. Whether such structures can form in the context of dsDNA unwound by CMG is unknown. Moreover, these assays employ isolated polymerases. How G4s affect the complete eukaryotic replisome remains unclear.

Very little work has been done on how i-motif structures affect DNA replication. Primer extension assays have demonstrated that i-motifs can inhibit DNA synthesis *in vitro*[38,40,44]. This is the only existing evidence describing replication of physiological i-motif sequences.

Recent work from our lab demonstrated that a variety of repetitive sequences are able to stall the eukaryotic replisome. These include tracts of poly$(dG)_n$ and poly$(dC)_n$ sequences[45]. However, these were the only G-rich or C-rich sequences tested. Poly$(dG)_n$ and poly$(dC)_n$ sequences are unique types of G4 or i-motif-forming sequences, and the stability of the structures they form remains unclear. Moreover, it is not clear whether stalling by these sequences is due to structure formation or their homopolymeric nature. As such, these sequences may not be representative of physiological G4 or i-motif-forming sequences. Yeast telomeric DNA alone does not affect replisome progression *in vitro*, and replication stalling is only seen when the telomeric binding protein Rap1 is present[46]. However, yeast telomeres are relatively weak G4-forming sequences. Whether other G4s in the absence of G4-binding proteins are sufficient to stall replication remains unclear. Recent work from the Remus lab demonstrated that G4s in the *Airn* locus can stall the eukaryotic replisome when initially stabilised by an R-loop. This study also highlighted the ability of a pre-formed G4 (generated from the synthetic sequence $(GGGT)_4$) to inhibit unwinding by the CMG

helicase[47]. However, this study only observed replisome stalling at G4s when the region had firstly been transcribed in order to generate an R-loop. This would promote G4 formation due to the exposure of ssDNA on the displaced strand and the lack of a complementary C-rich sequence opposite the G4 due to the presence of a DNA:RNA hybrid. As such, G4 structures were likely pre-formed before undergoing replication. Although stalling was still observed when the R-loop was subsequently removed, it remains unclear whether R-loop formation is initially required for G4s to form and subsequently stall the replisome. Moreover, whether G4s from a range of other loci have the same effect on replisome progression is yet to be understood.

Here, we carry out an extensive study on a variety of physiologically derived G4 and i-motif sequences and investigate how they impact eukaryotic replication. Using a reconstituted budding yeast DNA replication system, we find that a single G4 or i-motif-forming sequence is sufficient to cause replisome stalling. The ability of these sequences to stall replication correlated with their ability to form stable structures, strongly suggesting that secondary structure formation was the underlying trigger for replication stalling. Interestingly, CMG was able to unwind past these sequences, albeit more slowly, while DNA synthesis by polymerase ε was inhibited, leading to helicase-polymerase uncoupling and exposure of ssDNA. Conditions that enhanced replication fork uncoupling led to an increased proportion of replication fork stalling. Remarkably, stalling could only be rescued by the G4-unwinding helicase Pif1, highlighting the specificity of this enzyme. Ultimately, this study describes the response of the eukaryotic replisome to a variety of physiological structure-forming motifs and highlights their ability to induce replication stress. This mechanism may provide a potential explanation as to why these sequences are both genetically and epigenetically unstable and may explain their high mutational frequencies in cancer.

Results

The replisome stalls at a single G4 or i-motif sequence

To establish how quadruplex-forming sequences affect the eukaryotic replisome, we cloned several well-characterised G4- and i-motif-forming sequences into a 9.8kb plasmid. These were used to generate substrates for *in vitro* eukaryotic replication using purified budding yeast proteins[48,49]. The G4-forming sequences tested include one of the most well-characterised G4s found in the human genome, *c-MYC* Pu22.

This is a 22-nt-long segment derived from the human *c-MYC* promoter region[50] which is frequently mutated in cancers. Other G4-forming sequences tested include (i) the human telomeric repeat sequence $(TTAGGG)_4$, (ii) the Bu1 +3.5 G4 motif (derived from the avian DT40 genome) which induces replication-dependent epigenetic silencing[20], (iii) the CEB25 L111(T) motif derived from the naturally occurring human CEB25 mini-satellite, where all three loops have been modified to a single thymine, leading to a thermally stable G4 structure[14], (iv) the GGGGCC repeat of the C9orf72 gene which is associated with familial amyotrophic lateral sclerosis[51], (v) a synthetic sequence $(GGGT)_4$, and (vi) a tract of poly$(dG)_{16}$ which has previously been shown to induce replication stalling[45] (Table 1).

i-motif sequences were selected based on their ability to form secondary structures at physiological pH. These included those from the promoter regions of (i) DAP, (ii) DUX4L22, (iii) SNORD112, (iv) AC017019.1, (v) PIM1 and (vi) ZBTB7B (Table 2). All of these sequences are derived from the human genome and form i-motifs at neutral pH[26]. Sequences were cloned 3kb downstream of the ARS306 origin, from which site-specific replisome loading and firing occurs (Fig. 1A). As replication initiates from a defined site in the template, we can infer the identity of the leading and lagging nascent strands. The structure-forming sequences we refer to throughout the manuscript are positioned on the leading strand template. The templates were linearised by digestion with AhdI before replication to avoid confounding effects of converging replisomes on circular replication templates. Upon replication initiation on the parental control substrate, two replication forks proceed from the origin in either direction, generating one longer leading strand product of 8.2kb, and one shorter leading strand product of 1.5kb. However, if the replisome stalls at a structure-forming sequence, a 3kb band appears at the expense of the longer 8.2 kb leading strand product (Fig. S1). Lagging strand maturation factors have been omitted in these experiments. Therefore, lagging strand products remain as Okazaki fragments and run as a smear on denaturing agarose gels (Fig. 1B, lane 1).

As expected, the replicated control template generated two major leading strand products (8.2kb and 1.5kb) and Okazaki fragments (Fig. 1B, lane 1). In contrast, a faint but reproducible 3kb band appeared in the presence of some of the G4-forming sequences tested, indicating replisome stalling (Fig. 1B, lanes 2-8). The intensity of the 3kb stall band is quantified in Fig. 1C. A first thermal difference spectrum (TDS)

was performed for each sequence to observe the characteristic G4-profile. The thermal stability of the G4s was assessed via UV-vis melting which revealed a positive correlation between melting temperature (T_m) and stalling intensity (Fig. 1D). This is highlighted in Table 1, which depicts the melting temperatures in order of increasing thermal stability. Some of the G4s we tested, such as $(GGGGCC)_6$, could not be assigned a melting temperature due to aggregation issues, likely due to formation of multimolecular structures. Similarly, poly$(dG)_{16}$ could not be accurately characterised. These sequences were therefore excluded from further studies.

Replication stalling was also reproducibly induced by a range of physiologically derived single i-motif sequences, although to a lesser degree than G4s (Fig. 1E, lanes 4-9 and Fig. 1F). In contrast to G4s, the intensity of replication stalling did not correlate with their relative thermal stability of i-motifs (Fig. S2A) but was marginally correlated with their transitional pH (Fig. 1G, Table 2). This, together with the thermal stability data for the G4s, strongly argues that formation and stabilisation of G4 and i-motif structures is a key determinant of replication stalling. Intriguingly, upon replication of i-motif sequences, a novel 5kb replication product accumulated (Fig. 1E, lanes 4-9). Although this product was sometimes weakly visible with G4-forming sequences, it was consistently prominent upon replication of i-motifs. This product corresponds to the length of leading strand 1 from the site of the structure-forming sequence downstream to the end of the template (Fig. S1). This could either be a result of intrinsic repriming events, or DNA breakage at the site of the i-motif. These possibilities are currently under investigation. Interestingly, the intensity of stalling seen at a single G4-forming sequence was stronger than at a single i-motif-forming sequence (Fig. 1E, compare lane 3 to lanes 4-9). We anticipate that this may be due to the fact that i-motifs typically have a lower melting temperature at physiological pH than G4s (Tables 1 and 2).

Stalling is dependent on structure formation

The positive correlations we observed between stalling and biophysical properties of the structures suggest that both G4-forming and i-motif-forming sequences induce replisome stalling due to structure formation. To further substantiate this hypothesis, we inserted sequence mutations to abrogate or disrupt structure formation (Fig. S2B). These included removing a G-tract from $(GGGT)_4$ to $(GGGT)_3$ and mutating the central

guanine to an adenine in each G-tetrad of *c-MYC* Pu22. As an intermediate experiment, we mutated the loop regions in CEB25 L111(T) from thymine to adenine residues, which still allows for G4 formation but reduces its thermal stability[14]. Consistent with our hypothesis, stalling was reduced with the mutated sequences (Fig. 2A, 2B). The reduction in stalling was less significant for CEB25 L111(A), likely due to the fact it is still able to form a structure, albeit weaker. Importantly, biophysical characterisation demonstrated that these sequences either form weak structures, or none at all (Table 1). As expected, the melting temperatures of the mutated G4 sequences were reduced relative to the wildtype sequences (Fig. 2C), resulting in a positive correlation between stalling and thermal stability (Fig. S2D). Although the melting temperature for CEB25 L111(A) remained relatively high as it was still able to form a G4, the stalling intensity was reduced when compared to CEB25 L111(T) which is consistent with its weakened thermal stability (Fig. 2B, 2C). Similarly, when the i-motif-forming sequences DUX4L22 and SNORD112 were disrupted for their structure-forming ability (Fig. S2C), stalling was reduced (Fig. 2D, 2E). The reduction in stalling intensity for the mutated i-motif sequences correlated with a reduction in the transitional pH (Fig. 2F, S2E, Table 2) but did not show as much of a correlation with the melting temperature (Fig. S2F, Table 2).

Replisome stalling is dependent on the probability of structure formation

Stalling induced by a single G4- or i-motif-forming sequence was persistent, although relatively weak when compared to other types of structure-prone sequences[45]. Previous studies have shown that expanded repetitive sequences pose a greater challenge to the replisome[45]. Therefore, larger arrays of consecutive quadruplex-forming sequences may exacerbate replisome stalling and result in a greater proportion of stalled replication forks. To test this, we replicated templates containing an increasing number of consecutive G4 or i-motif-forming sequences (Fig. 3A, 3B). Stalling was only marginally increased and did not exhibit a linear relationship with the number of consecutive G4 or i-motif sequences. Interestingly, the 5kb product previously noted for the i-motif-containing substrates was directly proportional to the level of stalling, suggesting it forms as a consequence of replisome stalling (Fig. 3B).

Stalling at a large number of G4-forming sequences was not as strong as that previously seen with a poly$(dG)_{60}$ sequence, even with sixteen consecutive G4-forming

sequences present (Fig. 3A). Thus, stalling efficiency cannot be simply explained by thermal stability or the total number of G4 structures. We hypothesised that in addition to structure stability, stalling efficiency is dictated by the probability of structure formation. In the case of an uninterrupted tract of guanines, a G4 could form from any guanines within the sequence. This is in contrast to other distinct G4-forming sequences we tested (Fig. 3A), where structure formation is constrained to a defined window of G-tracts. To test this possibility, we interrupted the poly$(dG)_{60}$ sequence such that it could still support G4-formation but constrained to specific G-tracts. The poly$(dG)_{60}$ sequence was interrupted every fifth or tenth guanine with either an adenine or thymine residue and tested for its effects on replication (Fig. 3C). Leading strand stalling was reduced when the sequence was interrupted, and this was more prominent when interruptions were more frequent (Fig. 3C, compare lane 2 to lanes 3-6). This indicates that the polymeric, uninterrupted nature of the poly$(dG)_{60}$ sequence contributes to its ability to efficiently stall replication. We made similar observations with the poly$(dC)_{60}$ sequence, where interrupting the tract of consecutive cytosines reduced the level of stalling (Fig. 3D).

All observations made thus far have been in reference to the response of the replisome to a quadruplex-forming sequence on the leading strand template. We next wanted to determine how a quadruplex-forming sequence on the lagging strand template affects leading strand replication. To this end, we cloned G4-forming sequences such that they were positioned on either the leading or lagging strand template and observed leading strand replication products (Fig. 3E). As previously, we observed leading strand stalling when the G4-forming sequence was on the leading strand template (Fig. 3E, lanes 3 and 5). However, a G4-forming sequence on the lagging strand template had no effect on leading strand replication or replisome progression (Fig. 3E, lanes 4 and 6). Although we did not analyse lagging strand synthesis, previous work has demonstrated that the replication machinery skips over a lagging strand G4, resulting in a small gap, about the size of an Okazaki fragment, on the nascent lagging strand[47].

We performed similar experiments to determine how an i-motif-forming sequence on the lagging strand template affects leading strand synthesis (Fig. 3F). Interestingly, leading strand stalling also occurred when the i-motif-forming sequence was on the lagging strand template (Fig. 3F, lanes 4, 6 and 8). We speculate that this is due to

G4 formation by the complementary G-rich sequence which is on the leading strand template. These G-rich sequences are predicted to form stable G4 structures due to their long tracts of guanines and short loops. In contrast, the C-rich sequences which are complementary to the G4-forming sequences tested in Figure 3C are not able to form very stable i-motif structures (Table 2, both have a low pH_T ~5.8). This may explain why we do not observe leading strand replication stalling when these sequences are positioned on the leading strand template (Fig. 3E, lanes 4 and 6).

CMG is partially inhibited by a pre-formed quadruplex structure

Having observed consistent replisome stalling at quadruplex-forming sequences, we next wanted to determine whether the stall was transient or persisted over time. To determine this, we carried out pulse-chase experiments, where newly synthesised DNA is labelled with radiolabelled dATP for 10 minutes, after which an excess of unlabelled dATP is added. This prevents labelling of newly initiating replication forks and allows specific analysis of forks that have initiated in the first 10 minutes of the reaction. Replication forks stalled at G4 and i-motif-forming sequences were not resolved over time and persisted for up to two hours (Fig. 4A). This persistent arrest could either be due to blocked unwinding by the CMG helicase or lack of synthesis by polymerase ε. To address the first possibility, we carried out unwinding assays to determine the ability of CMG to unwind pre-formed G4 or i-motif structures. We chose sequences which induced the strongest replisome arrest – namely $(GGGT)_4$ and $(SNORD112)_1$. These sequences were located on the translocating strand (Fig. 4B). To favour structure formation, we inserted a $poly(dT)_{19}$ stretch on the opposite strand[52]. Consistent with previous work[47], time course analysis revealed that relative to a duplex substrate, CMG unwinding was initially inhibited by a G4 structure, evident within the first 5 to 10 minutes of the reaction (Fig. 4C, compare middle panel to top panel). Similar inhibition was also induced by an i-motif structure, but to a lesser degree (Fig. 4C, bottom panel). However, at later time points, CMG was able to unwind substrates with G4 or i-motif structures to a slightly lesser extent than duplex (Fig. 4C). We next assessed the effect of a G4-stabilising ligand, PhenDC3[53]. In the presence of a low concentration of PhenDC3 (0.25 μM), we observed no effect on duplex unwinding. In contrast, unwinding of the G4 substrate was inhibited (Fig. 4D). This suggests that CMG may be able to partially bypass a pre-existing G4 structure by

dismantling or unwinding the structure, rather than 'hopping' over the structure and leaving it intact.

Replisome stalling at G4s and i-motifs leads to helicase-polymerase uncoupling

The second possibility that may explain persistent fork stalling for at least 120 min (Fig. 4A) is inhibition of synthesis by polymerase ε, while CMG continues to unwind downstream. This phenomenon, termed helicase-polymerase uncoupling, occurs in response to various leading strand DNA damage lesions[54] and other types of repetitive sequences[45,47] and leads to the exposure of ssDNA. To determine if the same response occurs at G4-forming sequences, we utilised a previously established approach whereby an exogenous primer complementary to the region 264 nt downstream of the G4 sequence is added to the reaction[45,54]. This primer can only anneal if ssDNA is exposed, leading to restart of DNA synthesis, thereby serving as a readout for helicase-polymerase uncoupling. As shown in Figure 5A, addition of this primer led to the appearance of 5kb restart products for all G4-forming sequences tested, but these were not seen with a scrambled primer. Importantly, this restart product was absent from the empty vector (Fig. 5A, lane 1) but was evident with a template containing a CPD lesion (Fig. 5A, lane 3), which acted as a positive control. This is strong evidence that unwinding by the CMG helicase is able to continue through the G4 sequence, but synthesis by polymerase ε is inhibited. Analysis of the mechanism of stalling at i-motif sequences using this method is more complex, due to the presence of the intrinsic 5kb band seen in the absence of any exogenous primer upon replication of i-motif-forming sequences. However, analysis of these replication products on native gels resulted in the appearance of a population of faster-migrating species which has been shown to correspond to uncoupled products. This is clear for both the G4 and i-motif-forming sequences (Fig. 5B, lanes 2-5). Interestingly, we also observed a second population of faster-migrating species, which has been suggested to correspond to broken forks[55].

Given that CMG is able to unwind past G4 and i-motif sequences, yet synthesis by the polymerase is inhibited, we considered the possibility that structures arise on the ssDNA exposed by CMG unwinding. If this were true, then increasing the amount of ssDNA would enhance structure formation and consequently result in more polymerase stalling. Exposure of excess ssDNA is typically limited as polymerase ε is

tightly coupled to CMG through a direct interaction[56]. As pol ε is essential for replication initiation, it cannot be omitted from replication reactions. However, deletion of its catalytic domain, which completely eliminates its polymerase activity, does not impact replication initiation. Under these conditions, polymerase δ carries out leading strand synthesis. Since pol δ does not directly interact with CMG, this results in discontinuous synthesis that is not coupled to unwinding. Replication reactions with this polymerase ε mutant resulted in more replisome stalling when compared to the wildtype protein (Fig. 5C, compare lanes 3, 5, 7, 9, 11 and 13 to lanes 4, 6, 8, 10, 12, and 14). This was true for both G4 and i-motif-forming sequences and is in agreement with our hypothesis that increased replication fork uncoupling leads to more fork stalling. We propose that this is due to the increased opportunity for secondary structures to form on the stretch of ssDNA behind CMG, resulting in inhibition of polymerase ε, and thus replication fork stalling.

Stalling at G4s and i-motifs can be rescued by a specialised helicase

Previous work in our lab had revealed that polymerase δ can rescue stalling at hairpin-forming sequences, but not at poly$(dG)_n$ or poly$(dC)_n$[45]. Consistent with this, polymerase δ was not able to rescue stalling at any G4 (Fig. S3A) or i-motif-forming sequences (Fig. S3B). Since increased fork uncoupling enhanced replisome stalling (Fig. 5C), we considered the possibility that improved coupling might reduce stalling. CTF18-RFC is an alternative PCNA clamp-loader to the canonical RFC1-RFC and has been proposed to increase the coupling of DNA synthesis with unwinding during replication by directly binding the catalytic subunit of polymerase ε[57,58]. We hypothesised that a more tightly coupled replisome would reduce structure-formation and in turn decrease replication stalling. However, addition of CTF18-RFC, in the absence or presence of RFC1-RFC, did not affect stalling at either G4s or i-motifs (Fig. S3C).

We next considered whether a high concentration of dNTPs could rescue stalling at quadruplex-forming sequences. Previous work had demonstrated that a high concentration of dNTPs could resolve stalls at hairpin-forming sequences, such as $(CG)_n$ and $(CGG)_n$ repeats, but not poly$(dG)_n$ or poly$(dC)_n$ repeats[45]. Consistent with the quadruplex-forming potential of poly$(dG)_n$ and poly$(dC)_n$, an excess of any dNTP alone, or in combination, was not able to resolve stalling at either G4 (Fig. S4A) or i-

motif (Fig. S4B) forming sequences. As these were pulse-chase experiments, an excess of dATP was added to outcompete radiolabelled dATP and prevent labelling of new replication forks. Therefore, when an excess of another dNTP was added, it was the same concentration as dATP. We considered the possibility that the relative ratio of dNTPs may be more important than their absolute concentration. This raised the prediction that an increased proportion of dCTP relative to dA/dG/dTTP might be able to rescue stalling at G4-forming sequences. However, stalled forks were not rescued by a further excess of dCTP (2-fold over dATP and 26-fold over dGTP and dTTP (Fig. S4C)). Altogether, this suggests that once the replisome stalls at a quadruplex-forming sequence, the stall is persistent and requires a specialised activity to overcome it.

We next considered whether a G4-unwinding helicase might be able to rescue stalling at quadruplex-forming sequences. There are a number of helicases that have been shown to bind and unwind G4 structures both *in vitro* and *in vivo.* Examples include Sgs1, Chl1 and Pif1. Sgs1 is a RecQ family helicase shown to preferentially unwind G4 DNA[59]. Chl1 is the yeast homolog of human ChlR1 (also called DDX11). ChlR1 directly unwinds G4 structures and is proposed to help process G4s during DNA replication[60,61]. However, it has been suggested that Chl1 has stronger unwinding activity on certain types of G4s, namely anti-parallel G4s, and less so on the more common parallel conformation of G4s[60]. Pif1 is a well-characterised G4-unwinding helicase shown to play a vital role in enabling the efficient replication of G4s both *in vivo* and *in vitro*[32-34,39,62-65]. To test the role of these helicases, we carried out pulse-chase experiments for 10 minutes, during which a persistent stall occurred at G4 or i-motif-forming sequences. We then added a candidate helicase and allowed replication to continue for a further 10 minutes to determine if the stall could be resolved. Neither Sgs1 (Fig. 6A) nor Chl1 (Fig. 6B) were able to rescue stalling at either G4 or i-motif-forming sequences (Fig. S5A, S5B). Importantly, both of these enzymes were active on a model substrate (Fig. S5C, S5D). Chl1 is recruited to replisomes via an interaction with Ctf4[66]. Although Ctf4 is usually excluded from our reactions as it has minimal effect on budding yeast replication[48], we have included Ctf4 in our experiments with Chl1, so we can infer that the lack of rescue by Chl1 is not due to its inability to be recruited to replication forks.

Previous work has shown that Pif1 can rescue forks stalled at poly$(dG)_n$ and poly$(dC)_n$ sequences[45]. Consistent with this, Pif1 was also able to rescue forks stalled at both G4s and i-motifs (Fig. 6C, 6D). The extent of rescue was less evident for expanded sequences such as $(GGGT)_{28}$ and $(SNORD112)_4$ which may indicate that consecutive G4 or i-motif sequences pose a greater challenge. This rescue was dependent on its helicase activity as no rescue was observed with the Pif1 ATPase mutant K264A. The fact that Pif1, but not Sgs1 or Chl1, was able to rescue forks stalled at quadruplex-forming sequences demonstrates the specificity of Pif1, and highlights that only specialised helicases are able to resolve fork stalling at G4 and i-motif sequences.

Discussion

We have assessed the response of the eukaryotic replisome to a variety of G4 and i-motif-forming sequences and found that a single quadruplex-forming sequence alone is able to stall DNA replication. This is a significant finding as these types of sequences are highly prevalent in the human genome. Current estimates are between 370,000[13,67] and 700,000[17] G4-forming sequences and 5,125 i-motif-forming sequences[26]. These must all be replicated accurately in order to maintain genome integrity. Importantly, we have observed that the encounter of a replisome with a single quadruplex forming sequence can lead to the same mechanistic response as triggered in response to DNA damage. This raises the possibility that there are many physiological DNA sequences within the human genome that have the potential to threaten genome stability.

The ability of a G4-forming sequence to stall DNA replication correlates with its structure-forming potential. Although we have not been able to directly detect structures, we have observed a consistent trend between the thermal stability of structures and the level of replisome stalling induced. Consistent with this, mutated sequences which cannot form structures do not impact replisome progression. Together, these observations provide strong evidence that it is a G4 structure that causes replication stalling, and not the sequence itself.

Importantly, we have characterised the effects of i-motifs on replisome progression, which had not been previously well investigated. Here, we observed consistent replisome stalling at a variety of physiological i-motif sequences. Although the response of the replisome to i-motifs appears largely similar to G4s, we do observe

some differences. Firstly, replication stalling at an i-motif was generally weaker than the stalling we observed at a G4. We speculate that this may be because all of the i-motifs examined here are weaker structures than those formed by the G4s. This could mean they are less likely to form during replication, although once formed they appear to be a persistent and stable block to replisome progression. Secondly, the ability of i-motifs to stall replication seems to be less influenced by melting temperature than G4s, and may correlate more with transitional pH. This is in agreement with the fact that i-motif structures, but not G4s, are strongly influenced by pH. The transitional pH is defined as the pH at which the sequence is 50% folded. Although we see a minor role for the transitional pH here, i-motif replication stalling appears to be less impacted by predicted structural stability than G4s. This may be because weak i-motif structures are more likely to fold than weak G4 structures, so even when the predicted thermal stability is low, an i-motif structure may still fold. However, the i-motif sequences tested here were selected based on their ability to form i-motifs at neutral pH and thus all display similar pH_T values. Subsequently, we have not been able to compare a wide range of pH_T values with replication stalling intensity and make definitive correlations. Despite this, previous studies have found that i-motifs with a higher transitional pH have a higher potential for i-motif formation *in vivo* and in turn are more likely to undergo mutations and deletions in human cells[68]. Moreover, the biophysical characterisation of these sequences was done on short oligonucleotides, and the actual thermal stability of these structures within the context of duplexed DNA may in fact be higher. Still, the fact that these structures are more greatly influenced by pH than G4s may explain why we observe a weaker stall in our assays carried out under physiological pH conditions. The fact that stalling occurs at i-motifs despite these conditions strongly supports the idea that i-motifs are indeed able to form and can be a robust block to the replisome under physiological conditions, highlighting their biological relevance.

The evidence we have gathered thus far points towards G4 and i-motif structures forming during replication. Given that we observe helicase-polymerase uncoupling, we suggest that structures found behind the helicase inhibit synthesis by polymerases and result in replication stalling. It is possible that these structures may be pre-formed before replication as a result of other processes such as transcription within a physiological context. Consistent with this idea, previous work on the effects of R-

loops on DNA replication found that a G4 on the strand opposite the DNA:RNA hybrid was able to stall replication[47]. In this scenario, the G4 would be stabilised by the R-loop as the complementary C-rich sequence would not be available for duplex formation. However, although we currently do not have the tools available to directly detect where these structures are forming, it is unlikely that our replication substrates contain significant levels of pre-formed structures. One reason is that our substrates contain the complementary DNA sequence on the opposite strand. In this scenario, duplex is thermodynamically favoured over secondary structure formation. Therefore, for secondary structures to form, ssDNA would need to be exposed to allow structures to fold. As our substrates are not pre-treated to induce structure formation, we favour a model whereby structures form on ssDNA exposed behind the CMG helicase. This is consistent with a recent super-resolution microscopy study which detected the presence of G4 structures between the CMG helicase and either PCNA or nascent DNA[35]. Although CMG and polymerase ε are usually tightly coupled in the replisome, it is not altogether clear what the size of the gap is between the exit channel of CMG and the active site of polymerase ε. The most recent structural models of the eukaryotic replisome suggest a gap of at least 16-nt[69]. This would be sufficient ssDNA to allow G4 formation, or to nucleate i-motif formation, and in turn inhibit DNA synthesis by polymerase ε.

Although we observed consistent replisome stalling at G4 and i-motif-forming sequences, we never observed a complete block to all replication forks, and only saw a proportion of replication forks stalling. This was true even in the presence of up to sixteen consecutive G4-forming sequences. Therefore, we believe that replication fork stalling is determined by the likelihood of structure formation as opposed to the number of structures present. This would explain why we observed only a marginal increase in the proportion of replication forks stalling with an increasing number of consecutive G4s or i-motifs. In the presence of a larger number of structure-forming repeats, a structure is more likely to fold due to sequence availability, but once a single structure has formed it is sufficient to block the replisome. Similarly, the fact that we consistently observed a greater proportion of replication forks stalling at poly(dG)$_{60}$ and poly(dC)$_{60}$ may be due to the fact that a structure can form in any given window and is not constrained by loop sequences. Therefore, structures may be more likely to form in this context during replication as ssDNA exposed in any region of this sequence can

form a G4 or i-motif. This is consistent with previous biophysical characterisations of poly$(dC)_n$ sequences, which found that the optimum transitional pH peaks at poly$(dC)_{28}$ and gets lower as the number of cytosines increases. The longest tract of cytosines that was tested was poly$(dC)_{40,}$ and this did not have a transition[70]. This would suggest that the i-motif structures formed by longer tracts of cytosines are not inherently more stable, but rather are more likely to have the opportunity to fold and in turn stall replication. However, once a structure has formed, this is sufficient to be a robust and permanent block to the replisome and cannot be resolved by any of the replicative polymerases. In addition, structures which fold more quickly may be more likely to fold and in turn stall replication. Although unclear, folding kinetics usually correlate with structure stability, which may explain why we observe stronger stalling for more thermally stable G4 structures. In addition to the dynamics of the structure itself, the dynamics of how tightly each replication fork is coupled will determine whether there is sufficient time and space for structures to form.

In some cases, G4s and i-motifs could form from the same sequence but on complementary strands. In a physiological context, a stable G4 requires four tracts of three guanines. However, the complementary C-rich sequence would not form a stable i-motif at a physiological pH. Therefore, a stable G4 does not necessarily equate to a stable i-motif on the opposite strand. However, a stable i-motif requires four tracts of five cytosines at a physiological pH[26]. The complementary G-rich sequence would conform to the requirement of a stable G4, and as such it is more conceivable that a stable G4 structure would be found opposite an i-motif. This may explain why we observed an orientation-dependent stalling for G4s, but an orientation-independence for i-motifs.

It is unsurprising that we observe rescue of replication fork stalling at G4s by Pif1, given the breadth of data describing its role as a G4-unwinding helicase[32-34,39,62-65]. However, the fact that Pif1 is also able to resolve forks stalled at i-motifs points towards the broad specificity of Pif1 as a helicase. This is in line with its ability to also resolve forks stalled at hairpins[45]. Our observation that rescue of replisome stalling was less efficient with an array of consecutive G4 and i-motif sequences may indicate that multiple structures are forming within these arrays. As such, multiple consecutive secondary structures may pose a greater challenge to the replisome and require more extensive helicase activities.

Interestingly, none of the other helicases tested were able to rescue fork stalling at G4s or i-motifs, despite their demonstrated ability to unwind G4s[59,60]. This may reflect the fact that these helicases show a preference for certain types of G4s. For example, Chl1 has been shown to have stronger unwinding activity on anti-parallel G4s[60], while the G4 sequences we tested all form parallel G4s which is a more common conformation. This may also explain the large number of quadruplex-unwinding helicases that appear to have some level of redundancy, as they may each have a role in resolving structures in different scenarios. For example, BLM helicase has been shown to suppress recombination at G4s in transcribed genes[71]. The fact that only an accessory helicase was able to resolve stalls at G4s and i-motifs may also explain why mutations in helicases such as these lead to genome instability diseases, such as Bloom's Syndrome[72], Werner's syndrome[73] and Fanconi Anaemia[74].

Similar to previous work investigating the effects of repetitive sequences on eukaryotic replication[45], we observe uncoupling between helicase unwinding and DNA synthesis in response to both i-motifs and G4s. This is consistent with previous work which demonstrates an enrichment of G4-forming sequences within 6kb of uncoupled forks in tumour cells[75]. This response is akin to the response to leading strand DNA damage[54] and leads to the exposure of ssDNA. In a physiological context, exposure of large amounts of ssDNA as a result of replication stress can lead to the activation of checkpoint pathways[76]. Further studies are required to explore if such processes occur in response to G4s and i-motifs during replication. The ssDNA exposed during replisome uncoupling can also be a substrate for recombination and mutation events, which may explain why G4s and i-motifs are frequent mutation hotspots and undergo rearrangements[32]. In addition to this, further work should address how replisome stalling at G4s and i-motifs affects the fidelity of replication. Although replication can restart with the assistance of accessory helicases, it remains to be seen whether this process leaves any lesions in its wake. A recent high-throughput primer extension assay using T7 DNA polymerase demonstrated that polymerase stalling at structure prone sequences generates point mutations with a higher frequency than slippage events. Consistent with their higher propensity to inhibit synthesis, G4s displayed higher mutation rates than i-motifs and were often mutated in the loop regions[38]. Whether this same phenomenon occurs with eukaryotic polymerases in the context of the complete replisome remains unclear. Many of these structure-forming sequences

have been found at cancer mutation hotspots or breakpoints[24,25,77], and this replication stalling could be a potential explanation for mutational events.

We have demonstrated that CMG is able to bypass a G4 or i-motif sequence and established that synthesis by polymerase ε is inhibited. However, it is not clear whether stalling occurs in the same position each time. Higher resolution studies could map the position of replisome stalling within the G4 or i-motif and decipher whether the replisome consistently stalls at the base of the structure, or if it is able to progress some distance through.

Recent advances in the field have provided the tools to reconstitute human DNA replication *in vitro*[78]. Although the core replication machinery is conserved from budding yeast to human, it remains to be seen whether G4s and i-motifs have the same impact on progression of the human replisome. Using this system to study replication of structure-forming sequences would also enable one to study the roles of other human proteins in this process, such as an array of human helicases including RecQ helicases such as BLM[79], WRN[80] and FANCJ[81]. Moreover, the human replisome appears to be more dependent on certain factors than the yeast replisome. For example, although PCNA stimulates replication rates in yeast replisomes, it is not essential for fast and efficient leading strand replication, whereas it is an essential requirement in the human replisome. Moreover, AND-1, the human orthologue of the yeast protein Ctf4, stimulates leading strand replication in the human replisome, while Ctf4 has minimal effect on yeast replisomes[78]. These factors may affect how the replisome deals with quadruplex-forming sequences. AND-1 also directly interacts with polymerase α[82], which may affect the ability of the replisome to re-prime downstream of forks stalled at secondary structures. In addition to this, human cells make use of a primase known as PrimPol, which is absent in budding yeast[83]. PrimPol has both priming and polymerase activity and has been shown to facilitate replication fork progression by repriming and thus reinitiating replication downstream of DNA damage lesions[84]. Indeed, PrimPol is required to prevent epigenetic instability at G4 loci in vertebrate cells by binding and repriming downstream of G4s[85]. It remains to be seen whether this mechanism occurs directly at stalled replisomes and whether PrimPol has the same ability to reprime downstream of i-motif-forming sequences. However, repriming by PrimPol generates ssDNA gaps which may be less desirable than resolving secondary structures and generating continuous nascent strands.

Therefore, repriming by PrimPol may act as a backup pathway if there is a failure to recruit specialised helicases in time to resume replication. Beyond this, there is scope to explore how structure-forming sequences impact replication forks in the broader context of cells, to determine if the same mechanistic response is triggered.

Altogether, we discovered that a range of physiological G4 and i-motif-forming sequences can stall eukaryotic DNA replication. This study provides further insight as to why these sequences provide a barrier to DNA replication and suggests a potential mechanism for structure formation and resolution during replication. We therefore propose that endogenous DNA secondary structures are a source of replication stress, which may explain their genomic instability and mutation frequencies in cancer.

Materials and Methods

Constructing repeat-containing replication substrates

Substrates for replication assays were constructed by inserting repetitive sequences into a 9.8kb parental plasmid (pGC542) containing a single synthetic yeast replication origin (ARS306) which has been previously described[45]. Repetitive sequences were cloned into the MCS 3kb downstream of the origin and expanded using a strategy that employs synthetic oligos and type IIS restriction enzymes[86]. All oligos were ordered from Integrated DNA Technologies (IDT), the sequences of which can be found in Table 3.

Preparing templates for replication assays

Plasmids (Table 4) were transformed into NEB Stable Competent *E. coli* cells (#C3040I) which are ideal for propagation of repeat-containing plasmids. Cultures were grown at 30°C to reduce potential recombination and mutation events. Plasmids were purified using a QIAGEN HiSpeed Maxi Kit. Subsequently, supercoiled plasmids were isolated from nicked plasmids by PlasmidSelect clean up. DNA samples were diluted 6-fold in 100 mM Tris-HCl (pH 7.5), 10 mM EDTA and 3 M $(NH_4)_2SO_4$ (final concentration = 2.5 M) before incubation with 300 µL PlasmidSelect Xtra slurry (pre-washed with 100 mM Tris-HCl (pH 7.5), 10 mM EDTA and 2.3 M $(NH_4)_2SO_4$) for 30 minutes to bind. Following binding, nicked plasmids were eluted initially with 1 ml buffer consisting of 100 mM Tris-HCl (pH 7.5), 10 mM EDTA and 1.9 M $(NH_4)_2SO_4$. This was repeated twice. Then, 1 ml buffer was added to the beads and allowed to incubate for 10 minutes at RT. Subsequently, supercoiled plasmids were eluted with 100 mM Tris-HCl (pH 7.5), 10 mM EDTA and 1.5 M $(NH_4)_2SO_4$ by incubation for 10 minutes at RT. This was repeated once. Samples were de-salted by dialysis against 0.1xTE for 3 hours and overnight. DNA was concentrated using a 100 kDa Amicon concentrator followed by ethanol precipitation and resuspended in 1xTE.

CPD substrate

The substrate containing a site-specific DNA damage (CPD lesion) was prepared as has been previously described[45].

Protein purification

All protein expression strains, and expression and purification steps were carried out as previously described[45].

CMG

Saccharomyces cerevisiae CMG was expressed from the budding yeast strain yJCZ3[56]. All purification steps were carried out as previously described[87], with the following exceptions. After elution from Calmodulin Sepharose 4B (GE healthcare), the eluate was applied to a MonoQ 5/50 GL column (GE Healthcare) equilibrated in 25 mM Tris-HCl pH 7.2, 10% glycerol, 0.005% TWEEN 20, 0.5 mM DTT, 150 mM KCl. The protein was eluted with a 20 CV gradient from 150-1000 mM KCl and peak fractions were dialysed overnight against 500 ml 25 mM HEPES-KOH pH 7.6, 40 mM KOAc, 40 mM K-glutamate, 2 mM Mg(OAc)2, 0.25 mM EDTA, 0.5 mM DTT, 40% glycerol. The protein was subsequently concentrated through a 0.5 ml 30K MWCO Pierce™ Protein Concentrator (Thermo Scientific, 88502), snap-frozen and stored at -80°C.

Pol ε-Δcat

The DNA polymerase ε mutant with a deleted catalytic domain (Pol ε-Δcat) was expressed from the budding yeast strain yAJ25, which has been previously described[49], and purified as per wildtype pol ε.

Sgs1

Purified recombinant yeast Sgs1 was expressed and purified as previously described[88].

Chl1

Chl1 expression strain

Y5562 MAT**a**, ade2-1, ura3-1, his3-11,15, trp1-1, leu2-3,112, can1-100, pep4::HIS3, ade2::ADE2-pRSII402-Gal4-pGal1/10-Chl1- FLAG- Protein A

Chl1 purification

Budding yeast cells overexpressing Chl1 were grown in YP medium containing 2% raffinose as the carbon source to an optical density of 1.0 at 30°C. 2% galactose was then added to the culture to induce the protein expression, and the cells were further grown for 90 minutes. Cells were collected by centrifugation, washed with deionized water and suspended in Chl1 buffer (50 mM Tris-HCl pH 7, 10% glycerol, 2 mM $MgCl_2$, 0.5 mM TCEP) containing 0.1% Triton X-100, 500mM NaCl, 0.5mM Pefabloc, as well as the cOmplete-EDTA protease inhibitor cocktails. The cell suspension was frozen in liquid nitrogen, then cells were broken in a cryogenic freezer mill. The cell powder was thawed on ice, and further Chl1 buffer containing 0.1% Triton X-100 and 500 mM NaCl, and protease inhibitors was added. The lysate was clarified by centrifugation at 20,000 x g for 1 hour. The clarified lysate was transferred to pre-equilibrated IgG agarose beads. 8 µg/ml RNase A (Merck) was added, and incubated for 2 hours. The resin was washed with Chl1 buffer containing 0.1% Triton X-100 and 500 mM NaCl and then incubated in Chl1 buffer containing 0.1% Triton X-100, 500 mM NaCl, 10 mM $MgCl_2$ and 1 mM ATP for 15 minutes. The resin was washed again with Chl1 buffer containing 0.1% Triton X-100 and 500 mM NaCl and incubated overnight in the same buffer containing 10 µg/ml PreScission protease. The eluate was collected, and Chl1 dilution buffer (50 mM Tris-HCl pH7, 10% glycerol, 2 mM $MgCl_2$, 0.5 mM TCEP, 10 mM NaCl) was added to adjust the salt concentration to 160 mM NaCl. The diluted sample was loaded onto a HiTrap Heparin (Cytiva) column, equilibrated with Chl1 buffer containing 160 mM NaCl. The column was developed with a linear gradient from 160 mM to 1 M NaCl in Chl1 buffer. The peak fractions were pooled and loaded onto a Superdex 200 Increase (Cytiva) gel filtration column that was equilibrated and developed with Chl1 gel filtration buffer (20 mM Tris-HCl pH 7.5, 150 mM NaCl, 10% Glycerol, 0.5 mM TCEP). The peak fractions were concentrated by ultrafiltration.

RFC-Ctf18

RFC-CTF18 expression strain: yEF4

MATa ade2-1 ura3-1 his3-11,15 trp1-1 leu2-3,112 can1-100
bar1::Hyg
pep4::KanMX
ura3::URA3pRS306/Rfc2, CBP-Rfc3
trp1::TRP1pRS304/Rfc4, Rfc5

his3::HIS3pRS303/Ctf18
leu2::LEU2pRS305/Ctf8, Dcc1

Expression and purification steps were carried out as per RFC purification[45].

In vitro replication assays

MCM loading was carried out for 10 minutes at 24°C on 3nM circular DNA that was linearised during loading with 0.3 µL AhdI in a buffer containing 25 mM HEPES (pH 7.6), 10 mM magnesium acetate, 100 mM potassium glutamate, 1 mM DTT, 0.01% NP-40-S, 0.1 mg/ml BSA, 80 mM potassium chloride, 5 mM ATP, 20 nM ORC, 45 nM Cdc6, 75 nM Cdt1-Mcm2-7 and 50 nM DDK. Loading was stopped by the addition of 120 nM S-CDK for 5 minutes at 24°C. Following loading, samples were diluted in a buffer containing 25 mM HEPES (pH 7.6), 10 mM magnesium acetate, 100 mM potassium glutamate, 1 mM DTT, 0.01% NP-40-S, 0.1 mg/ml BSA to dilute the final contribution of chloride to 14 mM. A nucleotide mix was added to give final concentrations of 200 µM ATP, CTP, GTP and UTP; 30 µM dATP, dCTP, dGTP and dTTP; and 132 nM α-P^{33}-dATP. Subsequently, to initiate replication, a master mix of proteins was added to give final concentrations of 100 nM GINS, 10 nM S-CDK, 10 nM Mcm10, 40 nM Csm3/Tof1, 20 nM Pol ε, 30 nM Dpb11, 40 nM Cdc45, 40 nM Mrc1, 60 nM RPA, 40 nM RFC, 120 nM PCNA, 5 nM Pol δ, 50 nM Pol α, 20 nM Sld3/7 and 20 nM Sld2. Reactions were incubated at 30°C for 40 minutes. For samples to be run on denaturing gels, 0.5 µl SmaI was added to each 10 µl reaction in the final 10 minutes of the reaction, which cleaves products approx. 100 bp from the origin of replication. This removes heterogeneity in the length of leading strand products arising due to variability in the exact location where synthesis of leading strands begins, despite origin specificity[54]. Reactions were quenched by adding EDTA to a final concentration of 100 mM.

Pulse-chase experiments were carried out as previously described[45]. During the pulse, unlabelled deoxyribonucleotide concentrations were as follows: 30 µM dCTP, dTTP, dGTP and 2.5 µM dATP. To carry out the chase, unlabelled dATP was added to a final concentration of 400 µM, with the addition of 400 µM of dGTP, dCTP or dTTP, or 800 µM dCTP where indicated.

Repriming experiments were carried out by the addition of 60 nM oligonucleotide after loading and before initiation of replication[45].

Post-replication sample processing

For denaturing gels, replication products were separated on 0.8% alkaline agarose gels in 30 mM NaOH, 2 mM EDTA at 32V for 16 hours. Subsequently, products were fixed on denaturing agarose gels by incubation in 5% TCA for 40 minutes at room temperature.

For native gels, replication products were treated with 0.1% SDS and 1/100 volumes of Proteinase K at 37°C for 20 minutes. Sample volumes were topped up to 20 µl with TE before DNA extraction using phenol:chloroform:isoamyl alcohol 25:24:1 (Sigma-Aldrich, P2069). 5× Invitrogen™ Novex™ High-Density TBE Sample Buffer was added to samples before separation on a 0.8% agarose/TAE gel in modified TAE (50 mM Tris-HCl (pH 7.9), 40 mM NaOAC, 1 mM EDTA (pH 8.0))[54] at 30V for 20 hours at 4°C.

Gels were dried before being exposed to a Storage Phosphor Screen (GE Healthcare, BAS-IP MS 2025) and imaged on a Typhoon scanner (Cytiva). Images were analysed and quantified in ImageJ. Quantifications of stalling intensities were calculated from an average of three replicates in separate experiments. To calculate intensities, the background for each lane was subtracted from each measurement. The 3kb stalling intensity band was normalised to the intensity of 'leading strand 2' in each lane to account for variation in the efficiencies of reactions for each substrate.

Preparing templates for helicase assays

Forked DNA substrates were prepared as previously described[52]. Briefly, dried oligos from IDT were resuspended to 10 µM in 10 mM Tris pH 8.0. The bottom strand of the substrate (GC340, SW051 or SW105) was end labelled with γ-^{32}P-ATP in a reaction containing 5 pmol oligo, 1X PNK buffer, 1U of PNK enzyme (NEB, M0201S), and γ-^{32}P-ATP (0.03 mCi). The reaction was incubated for 1 hour at 37°C, followed by heat inactivation of PNK for 20 minutes at 80°C and then 10 minutes at 90°C. Unincorporated radiolabelled nucleotides were removed by passing the sample through a G50 column (GE healthcare, 2753002) equilibrated in 10mM Tris pH 8.0. The concentration of K+ was then adjusted to 50 mM by the addition of 1 M KOAc. To

anneal the complementary top strand (GC339 or SW114), 1 µl of a 10 µM stock of oligonucleotide was added and the reaction was incubated at 95°C for 5 minutes, followed by a slow cooling to 10°C at a rate of -1°C/minute. Annealed products were run on 10% TBE gels (ThermoFisher Scientific, EC62755BOX) in 0.5xTBE at 150 V for 45 minutes. Fully annealed substrates were isolated from the gel using a crush-soak method as described in Batra et al., 2022.

For Chl1 and Sgs1 unwinding assays, substrates containing a 5' overhang (for Chl1) and a 3' overhang (for Sgs1) were generated as previously described[45].

Sequences of all oligonucleotides used to generate helicase unwinding substrates are detailed in Table 3.

Helicase assays

CMG unwinding assays were carried out using 0.5 nM of labelled substrate and 20 nM of purified CMG. Reactions were assembled in a buffer containing 25 mM HEPES pH 7.6, 10 mM MgOAc, 30 mM NaCl, 0.1 mg/ml BSA and 0.1 mM AMP PNP. Reactions were incubated at 30°C for 30 minutes to allow CMG to pre-load onto the template, before addition of ATP to a final concentration of 2 mM to stimulate unwinding. At this point, 65 nM of the unlabelled version of the labelled oligo (GC340, SW051 or SW105) was added to trap any unwound oligos and prevent substrate re-annealing. Reactions were incubated at 30°C and time points taken as indicated. Reactions were stopped by the addition of 0.5% SDS and 200 mM EDTA, supplemented with Novex Hi-Density TBE Sample buffer (ThermoFisher Scientific, LC6678) and analysed on 10% Novex TBE gels (ThermoFisher Scientific, EC62755BOX) in 1× TBE at 90 V for 90 minutes. Gels were exposed to a Storage Phosphor Screen (GE Healthcare, BAS-IP MS 2025) and imaged in a Typhoon (Cytiva). Images were analysed and quantified in ImageJ.

For Chl1 and Sgs1 unwinding assays, reactions were carried out using 0.5 nM of labelled duplex and 50 nM of purified recombinant protein as previously described[45].

Biophysical characterisation of G4 sequences

DNA Annealing Step

DNA sequences (Table 1) were purchased in their lyophilised form with standard desalting purification from Integrated DNA Technology (IDT) and re-dissolved in MilliQ water to reach a stock concentration of 100 μM. The sequences were then further diluted to 10 μM in the annealing buffer (25 mM HEPES, 10 mM $MgCl_2$, 110 mM KCl/LiCl), heated to 95℃ for 15 minutes and left to slowly cool down to room temperature overnight.

Thermal difference spectra (TDS)

100 μL of each of the 10 μM DNA sequences (annealed as above) was transferred into a High Precision Cell (quartz glass) Light Path 10 mm (HellmaAnalytics) and covered with 200 μL of mineral oil to prevent evaporation. The cuvettes were sealed with a plastic lid and transferred into the Agilent Cary 3500 UV-Vis Multicell Peltier spectrometer. A first scan was run at 25℃ (Scan range = 800-200 nm | Averaging time (s) = 0.02 | Data Interval (nm) = 1 | Scan rate (nm/min) = 3000 | Spectral bandwidth (nm) = 2). The samples were then heated to 95℃ and left to equilibrate for 7 mins before running a second scan. Each TDS curve was obtained by subtracting the absorbance spectra (25°C) by the absorbance spectra (95 °C).

UV-vis melting curve

100 μL of each of the 10 μM DNA sequences (annealed as above) was transferred into a High Precision Cell (quartz glass) Light Path 10 mm (HellmaAnalytics) and covered with 200 μL of mineral oil to prevent evaporation. The cuvettes were sealed with a plastic lid and transferred into the Agilent Cary 3500 UV-Vis Multicell Peltier spectrometer. The melting curve was obtained by imputing the following parameters (Wavelengths: 295 nm | Averaging time (s) = 0.1 | Spectral bandwidth (nm) = 2) and the heating protocol in Table 5. Data was collected only for stages 5 and 6 for each run.

Stage	Data Interval (℃)	Rate (℃/min)	Temperature Start/End (℃)
1	1	5	25/95
2	1	5	95/25

3	1	5	25/95
4	1	5	95/25
5	0.1	0.5	25/95
6	0.1	0.5	95/25

Table 5. Heating protocol for Uv-Vis melting curve of DNA sequences.

T_m was extrapolated for each sample by fitting the relative melting curve as shown by Mergny and Lacroix[89] with the Python script provided by Giacomo Fabrini[90].

Biophysical characterisation of i-motif sequences

Oligonucleotides

DNA sequences (Table 2) were supplied by Eurogentec (Belgium), synthesized on a 1000 nmol scale and purified by reverse phase HPLC. All DNA sequences were dissolved in ultra-pure water to give 100 μM final concentrations, which were confirmed using a Nanodrop. For all experiments, ODNs were diluted in buffer containing 10 mM sodium cacodylate and 100 mM potassium chloride at the indicated pH. DNA samples were thermally annealed by heating in a heat block at 95°C for 5 minutes and cooled slowly to room temperature overnight.

Circular Dichroism

CD spectra were recorded on a Jasco J-1500 spectropolarimeter using a 1 mm path length quartz cuvette. ODNs were diluted to 10 μM (total volume: 100 μL) in buffer at pH increments of 0.25 or 0.5 pH units from 4.0 to 8.0, depending on the sequence. Spectra were recorded at 20 °C between 200 and 320 nm. Data pitch was set to 0.5 nm and measurements were taken at a scanning speed of 200 nm/min, response time of 1 s, bandwidth of 2 nm and 100 mdeg sensitivity; each spectrum was the average of four scans. Samples containing only buffer were also scanned according to these parameters to allow for blank subtraction. Transitional pH (pH_T) for each i-motif was calculated from the inflection point of fitted ellipticity at 288 nm.

UV Absorption Spectroscopy

UV spectroscopy experiments were performed on a Jasco J-750 equipped with a Julabo F-250Temperature Controller and recorded using low volume masked quartz

cuvettes (1 cm path length). Annealed DNA samples (250 µL) were transferred to a cuvette and covered with a stopper to reduce evaporation of the sample. The absorbance of the DNA was measured at 295 nm and 260 nm as the temperature of the sample was held for 10 min at 4°C, heated to 95°C at a rate of 0.5°C per min, then held at 95°C for 10 min before the process was reversed; each melting/annealing process was repeated three times. Data were recorded every 1°C during both melting and annealing and melting temperatures (T_m) were determined using the first derivative method. TDS were obtained by subtracting the spectrum of the folded structure between 220 and 320 nm at 4°C from that of the unfolded structure at 95°C. The data was normalized and maximum change in absorption was set to +1 as previously described[91].

Data Analysis

Final analysis and presentation of the data was performed using GraphPad Prism version 9.0. All sets of data passed Shapiro-Wilk normality test, p-values were calculated by One way ANOVA followed by Holm-Sidak posthoc analysis for the melting temperature and thermodynamic data collected from the triplicate measurements for each oligonucleotide.

Author Contributions Statement

SLW and CSCD performed all experiments, except for biophysical characterisations of structures. GC conceptualised and supervised the research and acquired funding. SLW wrote the manuscript with input from GC, MDA, FR, ZAEW and DG. MDA and FR have designed and performed biophysical characterisation of G4-structures. ZAEW and DG have designed and performed biophysical characterisation of i-motif-structures. MM expressed and purified recombinant Chl1. EEF and JTPY generated the expression strain for recombinant RFC-CTF18.

Acknowledgements

This work was funded by a Wellcome Trust and Royal Society Sir Henry Dale fellowship for Gideon Coster (210470/Z/18/Z) as well as internal funding from the Institute of Cancer Research. Marco Di Antonio is supported by a Biotechnology and Biological Sciences Research Council (BBSRC) David Phillips Fellowship [BB/R011605/1] and a Lister Prize Fellowship. Federica Raguseo is supported by a Leverhulme Trust, Cellular Bionics scholarship [EP/S023518/1]. We would like to thank Petr Cejka (Institute for Research in Biomedicine, Switzerland) for providing published purified recombinant protein. We would like to thank Max Douglas for reagents and experimental support, and Manuel Daza Martin for critical reading of the manuscript. We are thankful to Giacomo Fabrini and Lorenzo Di Michele for support on Tm analysis.

References

1 Bell, S. P. & Labib, K. Chromosome Duplication in Saccharomyces cerevisiae. *Genetics* **203**, 1027-1067 (2016). https://doi.org:10.1534/genetics.115.186452

2 Gaillard, H., Garcia-Muse, T. & Aguilera, A. Replication stress and cancer. *Nat Rev Cancer* **15**, 276-289 (2015). https://doi.org:10.1038/nrc3916

3 Gacy, A. M., Goellner, G., Juranic, N., Macura, S. & McMurray, C. T. Trinucleotide repeats that expand in human disease form hairpin structures in vitro. *Cell* **81**, 533-540 (1995).

4 Nadel, Y., Weisman-Shomer, P. & Fry, M. The fragile X syndrome single strand d(CGG)n nucleotide repeats readily fold back to form unimolecular hairpin structures. *J Biol Chem* **270**, 28970-28977 (1995). https://doi.org:10.1074/jbc.270.48.28970

5 Fry, M. & Loeb, L. A. The fragile X syndrome d(CGG)n nucleotide repeats form a stable tetrahelical structure. *Proc Natl Acad Sci U S A* **91**, 4950-4954 (1994). https://doi.org:10.1073/pnas.91.11.4950

6 Gehring, K., Leroy, J. L. & Gueron, M. A tetrameric DNA structure with protonated cytosine.cytosine base pairs. *Nature* **363**, 561-565 (1993). https://doi.org:10.1038/363561a0

7 Voineagu, I., Narayanan, V., Lobachev, K. S. & Mirkin, S. M. Replication stalling at unstable inverted repeats: interplay between DNA hairpins and fork stabilizing proteins. *Proc Natl Acad Sci U S A* **105**, 9936-9941 (2008). https://doi.org:10.1073/pnas.0804510105

8 Tubbs, A. *et al.* Dual Roles of Poly(dA:dT) Tracts in Replication Initiation and Fork Collapse. *Cell* **174**, 1127-1142.e1119 (2018). https://doi.org:10.1016/j.cell.2018.07.011

9 Shastri, N. *et al.* Genome-wide Identification of Structure-Forming Repeats as Principal Sites of Fork Collapse upon ATR Inhibition. *Mol Cell* **72**, 222-238.e211 (2018). https://doi.org:10.1016/j.molcel.2018.08.047

10 Brown, R. E. & Freudenreich, C. H. Structure-forming repeats and their impact on genome stability. *Curr Opin Genet Dev* **67**, 41-51 (2021). https://doi.org:10.1016/j.gde.2020.10.006

11 Sen, D. & Gilbert, W. Formation of parallel four-stranded complexes by guanine-rich motifs in DNA and its implications for meiosis. *Nature* **334**, 364-366 (1988). https://doi.org:10.1038/334364a0

12 Biffi, G., Tannahill, D., McCafferty, J. & Balasubramanian, S. Quantitative visualization of DNA G-quadruplex structures in human cells. *Nat Chem* **5**, 182-186 (2013). https://doi.org:10.1038/nchem.1548

13 Huppert, J. L. & Balasubramanian, S. Prevalence of quadruplexes in the human genome. *Nucleic Acids Res* **33**, 2908-2916 (2005). https://doi.org:10.1093/nar/gki609

14 Piazza, A. *et al.* Short loop length and high thermal stability determine genomic instability induced by G-quadruplex-forming minisatellites. *EMBO J* **34**, 1718-1734 (2015). https://doi.org:10.15252/embj.201490702

15 Ou, T. M. *et al.* G-quadruplexes: targets in anticancer drug design. *ChemMedChem* **3**, 690-713 (2008). https://doi.org:10.1002/cmdc.200700300

16 Maiti, S. Human telomeric G-quadruplex. *FEBS J* **277**, 1097 (2010). https://doi.org:10.1111/j.1742-4658.2009.07468.x

17 Chambers, V. S. *et al.* High-throughput sequencing of DNA G-quadruplex structures in the human genome. *Nat Biotechnol* **33**, 877-881 (2015). https://doi.org:10.1038/nbt.3295
18 Robinson, J., Raguseo, F., Nuccio, S. P., Liano, D. & Di Antonio, M. DNA G-quadruplex structures: more than simple roadblocks to transcription? *Nucleic Acids Res* **49**, 8419-8431 (2021). https://doi.org:10.1093/nar/gkab609
19 Papadopoulou, C., Guilbaud, G., Schiavone, D. & Sale, J. E. Nucleotide Pool Depletion Induces G-Quadruplex-Dependent Perturbation of Gene Expression. *Cell Rep* **13**, 2491-2503 (2015). https://doi.org:10.1016/j.celrep.2015.11.039
20 Schiavone, D. *et al.* Determinants of G quadruplex-induced epigenetic instability in REV1-deficient cells. *EMBO J* **33**, 2507-2520 (2014). https://doi.org:10.15252/embj.201488398
21 Vannier, J. B. *et al.* RTEL1 is a replisome-associated helicase that promotes telomere and genome-wide replication. *Science* **342**, 239-242 (2013). https://doi.org:10.1126/science.1241779
22 Sarkies, P., Reams, C., Simpson, L. J. & Sale, J. E. Epigenetic instability due to defective replication of structured DNA. *Mol Cell* **40**, 703-713 (2010). https://doi.org:10.1016/j.molcel.2010.11.009
23 Hansel-Hertsch, R. *et al.* G-quadruplex structures mark human regulatory chromatin. *Nat Genet* **48**, 1267-1272 (2016). https://doi.org:10.1038/ng.3662
24 Wang, G. & Vasquez, K. M. Effects of Replication and Transcription on DNA Structure-Related Genetic Instability. *Genes (Basel)* **8** (2017). https://doi.org:10.3390/genes8010017
25 De, S. & Michor, F. DNA secondary structures and epigenetic determinants of cancer genome evolution. *Nat Struct Mol Biol* **18**, 950-955 (2011). https://doi.org:10.1038/nsmb.2089
26 Wright, E. P., Huppert, J. L. & Waller, Z. A. E. Identification of multiple genomic DNA sequences which form i-motif structures at neutral pH. *Nucleic Acids Res* **45**, 2951-2959 (2017). https://doi.org:10.1093/nar/gkx090
27 Kendrick, S. *et al.* The dynamic character of the BCL2 promoter i-motif provides a mechanism for modulation of gene expression by compounds that bind selectively to the alternative DNA hairpin structure. *J Am Chem Soc* **136**, 4161-4171 (2014). https://doi.org:10.1021/ja410934b
28 Miglietta, G., Cogoi, S., Pedersen, E. B. & Xodo, L. E. GC-elements controlling HRAS transcription form i-motif structures unfolded by heterogeneous ribonucleoprotein particle A1. *Sci Rep* **5**, 18097 (2015). https://doi.org:10.1038/srep18097
29 Kaiser, C. E. *et al.* Insight into the Complexity of the i-Motif and G-Quadruplex DNA Structures Formed in the KRAS Promoter and Subsequent Drug-Induced Gene Repression. *J Am Chem Soc* **139**, 8522-8536 (2017). https://doi.org:10.1021/jacs.7b02046
30 Simonsson, T., Pribylova, M. & Vorlickova, M. A nuclease hypersensitive element in the human c-myc promoter adopts several distinct i-tetraplex structures. *Biochem Biophys Res Commun* **278**, 158-166 (2000). https://doi.org:10.1006/bbrc.2000.3783
31 Guo, K., Gokhale, V., Hurley, L. H. & Sun, D. Intramolecularly folded G-quadruplex and i-motif structures in the proximal promoter of the vascular endothelial growth factor gene. *Nucleic Acids Res* **36**, 4598-4608 (2008). https://doi.org:10.1093/nar/gkn380

32 Lopes, J. *et al.* G-quadruplex-induced instability during leading-strand replication. *EMBO J* **30**, 4033-4046 (2011). https://doi.org:10.1038/emboj.2011.316
33 Paeschke, K., Capra, J. A. & Zakian, V. A. DNA replication through G-quadruplex motifs is promoted by the Saccharomyces cerevisiae Pif1 DNA helicase. *Cell* **145**, 678-691 (2011). https://doi.org:10.1016/j.cell.2011.04.015
34 Dahan, D. *et al.* Pif1 is essential for efficient replisome progression through lagging strand G-quadruplex DNA secondary structures. *Nucleic Acids Res* **46**, 11847-11857 (2018). https://doi.org:10.1093/nar/gky1065
35 Lee, W. T. C. *et al.* Single-molecule imaging reveals replication fork coupled formation of G-quadruplex structures hinders local replication stress signaling. *Nat Commun* **12**, 2525 (2021). https://doi.org:10.1038/s41467-021-22830-9
36 Edwards, D. N., Machwe, A., Wang, Z. & Orren, D. K. Intramolecular telomeric G-quadruplexes dramatically inhibit DNA synthesis by replicative and translesion polymerases, revealing their potential to lead to genetic change. *PLoS One* **9**, e80664 (2014). https://doi.org:10.1371/journal.pone.0080664
37 Lormand, J. D. *et al.* DNA polymerase δ stalls on telomeric lagging strand templates independently from G-quadruplex formation. *Nucleic Acids Res* **41**, 10323-10333 (2013). https://doi.org:10.1093/nar/gkt813
38 Murat, P., Guilbaud, G. & Sale, J. E. DNA polymerase stalling at structured DNA constrains the expansion of short tandem repeats. *Genome Biol* **21**, 209 (2020). https://doi.org:10.1186/s13059-020-02124-x
39 Sparks, M. A., Singh, S. P., Burgers, P. M. & Galletto, R. Complementary roles of Pif1 helicase and single stranded DNA binding proteins in stimulating DNA replication through G-quadruplexes. *Nucleic Acids Res* **47**, 8595-8605 (2019). https://doi.org:10.1093/nar/gkz608
40 Takahashi, S., Brazier, J. A. & Sugimoto, N. Topological impact of noncanonical DNA structures on Klenow fragment of DNA polymerase. *Proc Natl Acad Sci U S A* **114**, 9605-9610 (2017). https://doi.org:10.1073/pnas.1704258114
41 Fan, J. H., Bochkareva, E., Bochkarev, A. & Gray, D. M. 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 (2009). https://doi.org:10.1021/bi801538h
42 Ray, S. *et al.* RPA-mediated unfolding of systematically varying G-quadruplex structures. *Biophys J* **104**, 2235-2245 (2013). https://doi.org:10.1016/j.bpj.2013.04.004
43 Salas, T. R. *et al.* Human replication protein A unfolds telomeric G-quadruplexes. *Nucleic Acids Res* **34**, 4857-4865 (2006). https://doi.org:10.1093/nar/gkl564
44 Catasti, P. *et al.* Cystosine-rich strands of the insulin minisatellite adopt hairpins with intercalated cytosine+.cytosine pairs. *J Mol Biol* **272**, 369-382 (1997). https://doi.org:10.1006/jmbi.1997.1248
45 Casas-Delucchi, C. S., Daza-Martin, M., Williams, S. L. & Coster, G. The mechanism of replication stalling and recovery within repetitive DNA. *Nat Commun* **13**, 3953 (2022). https://doi.org:10.1038/s41467-022-31657-x
46 Douglas, M. E. & Diffley, J. F. X. Budding yeast Rap1, but not telomeric DNA, is inhibitory for multiple stages of DNA replication in vitro. *Nucleic Acids Res* **49**, 5671-5683 (2021). https://doi.org:10.1093/nar/gkab416

47 Kumar, C., Batra, S., Griffith, J. D. & Remus, D. The interplay of RNA:DNA hybrid structure and G-quadruplexes determines the outcome of R-loop-replisome collisions. *Elife* **10** (2021). https://doi.org:10.7554/eLife.72286
48 Yeeles, J. T., Deegan, T. D., Janska, A., Early, A. & Diffley, J. F. Regulated eukaryotic DNA replication origin firing with purified proteins. *Nature* **519**, 431-435 (2015). https://doi.org:10.1038/nature14285
49 Yeeles, J. T. P., Janska, A., Early, A. & Diffley, J. F. X. How the Eukaryotic Replisome Achieves Rapid and Efficient DNA Replication. *Mol Cell* **65**, 105-116 (2017). https://doi.org:10.1016/j.molcel.2016.11.017
50 Dai, J., Carver, M., Hurley, L. H. & Yang, D. Solution structure of a 2:1 quindoline-c-MYC G-quadruplex: insights into G-quadruplex-interactive small molecule drug design. *J Am Chem Soc* **133**, 17673-17680 (2011). https://doi.org:10.1021/ja205646q
51 Thys, R. G. & Wang, Y. H. DNA Replication Dynamics of the GGGGCC Repeat of the C9orf72 Gene. *J Biol Chem* **290**, 28953-28962 (2015). https://doi.org:10.1074/jbc.M115.660324
52 Batra, S., Devbhandari, S. & Remus, D. CMG helicase activity on G4-containing DNA templates. *Methods Enzymol* **672**, 233-260 (2022). https://doi.org:10.1016/bs.mie.2022.02.020
53 De Cian, A., Delemos, E., Mergny, J. L., Teulade-Fichou, M. P. & Monchaud, D. Highly efficient G-quadruplex recognition by bisquinolinium compounds. *J Am Chem Soc* **129**, 1856-1857 (2007). https://doi.org:10.1021/ja067352b
54 Taylor, M. R. G. & Yeeles, J. T. P. The Initial Response of a Eukaryotic Replisome to DNA Damage. *Mol Cell* **70**, 1067-1080 e1012 (2018). https://doi.org:10.1016/j.molcel.2018.04.022
55 Taylor, M. R. G. & Yeeles, J. T. P. Dynamics of Replication Fork Progression Following Helicase-Polymerase Uncoupling in Eukaryotes. *J Mol Biol* **431**, 2040-2049 (2019). https://doi.org:10.1016/j.jmb.2019.03.011
56 Zhou, J. C. *et al.* CMG-Pol epsilon dynamics suggests a mechanism for the establishment of leading-strand synthesis in the eukaryotic replisome. *Proc Natl Acad Sci U S A* **114**, 4141-4146 (2017). https://doi.org:10.1073/pnas.1700530114
57 Grabarczyk, D. B., Silkenat, S. & Kisker, C. Structural Basis for the Recruitment of Ctf18-RFC to the Replisome. *Structure* **26**, 137-144 e133 (2018). https://doi.org:10.1016/j.str.2017.11.004
58 Stokes, K., Winczura, A., Song, B., Piccoli, G. & Grabarczyk, D. B. Ctf18-RFC and DNA Pol form a stable leading strand polymerase/clamp loader complex required for normal and perturbed DNA replication. *Nucleic Acids Res* **48**, 8128-8145 (2020). https://doi.org:10.1093/nar/gkaa541
59 Huber, M. D., Lee, D. C. & Maizels, N. G4 DNA unwinding by BLM and Sgs1p: substrate specificity and substrate-specific inhibition. *Nucleic Acids Res* **30**, 3954-3961 (2002). https://doi.org:10.1093/nar/gkf530
60 Wu, Y., Sommers, J. A., Khan, I., de Winter, J. P. & Brosh, R. M., Jr. Biochemical characterization of Warsaw breakage syndrome helicase. *J Biol Chem* **287**, 1007-1021 (2012). https://doi.org:10.1074/jbc.M111.276022
61 Lerner, L. K. *et al.* Timeless couples G-quadruplex detection with processing by DDX11 helicase during DNA replication. *EMBO J* **39**, e104185 (2020). https://doi.org:10.15252/embj.2019104185

62 Byrd, A. K., Bell, M. R. & Raney, K. D. Pif1 helicase unfolding of G-quadruplex DNA is highly dependent on sequence and reaction conditions. *J Biol Chem* **293**, 17792-17802 (2018). https://doi.org:10.1074/jbc.RA118.004499
63 Ribeyre, C. *et al.* The yeast Pif1 helicase prevents genomic instability caused by G-quadruplex-forming CEB1 sequences in vivo. *PLoS Genet* **5**, e1000475 (2009). https://doi.org:10.1371/journal.pgen.1000475
64 Paeschke, K. *et al.* Pif1 family helicases suppress genome instability at G-quadruplex motifs. *Nature* **497**, 458-462 (2013). https://doi.org:10.1038/nature12149
65 Maestroni, L. *et al.* RPA and Pif1 cooperate to remove G-rich structures at both leading and lagging strand. *Cell Stress* **4**, 48-63 (2020). https://doi.org:10.15698/cst2020.03.214
66 Samora, C. P. *et al.* Ctf4 Links DNA Replication with Sister Chromatid Cohesion Establishment by Recruiting the Chl1 Helicase to the Replisome. *Mol Cell* **63**, 371-384 (2016). https://doi.org:10.1016/j.molcel.2016.05.036
67 Huppert, J. L. & Balasubramanian, S. G-quadruplexes in promoters throughout the human genome. *Nucleic Acids Res* **35**, 406-413 (2007). https://doi.org:10.1093/nar/gkl1057
68 Martella, M. *et al.* i-Motif formation and spontaneous deletions in human cells. *Nucleic Acids Res* **50**, 3445-3455 (2022). https://doi.org:10.1093/nar/gkac158
69 Yuan, Z., Georgescu, R., Schauer, G. D., O'Donnell, M. E. & Li, H. Structure of the polymerase epsilon holoenzyme and atomic model of the leading strand replisome. *Nat Commun* **11**, 3156 (2020). https://doi.org:10.1038/s41467-020-16910-5
70 Fleming, A. M. *et al.* 4n-1 Is a "Sweet Spot" in DNA i-Motif Folding of 2'-Deoxycytidine Homopolymers. *J Am Chem Soc* **139**, 4682-4689 (2017). https://doi.org:10.1021/jacs.6b10117
71 van Wietmarschen, N. *et al.* BLM helicase suppresses recombination at G-quadruplex motifs in transcribed genes. *Nat Commun* **9**, 271 (2018). https://doi.org:10.1038/s41467-017-02760-1
72 Cunniff, C., Bassetti, J. A. & Ellis, N. A. Bloom's Syndrome: Clinical Spectrum, Molecular Pathogenesis, and Cancer Predisposition. *Mol Syndromol* **8**, 4-23 (2017). https://doi.org:10.1159/000452082
73 Yu, C. E. *et al.* Positional cloning of the Werner's syndrome gene. *Science* **272**, 258-262 (1996). https://doi.org:10.1126/science.272.5259.258
74 Brosh, R. M., Jr. & Cantor, S. B. Molecular and cellular functions of the FANCJ DNA helicase defective in cancer and in Fanconi anemia. *Front Genet* **5**, 372 (2014). https://doi.org:10.3389/fgene.2014.00372
75 Amparo, C. *et al.* Duplex DNA from Sites of Helicase-Polymerase Uncoupling Links Non-B DNA Structure Formation to Replicative Stress. *Cancer Genomics & Proteomics* **17**, 101-115 (2020).
76 MacDougall, C. A., Byun, T. S., Van, C., Yee, M. C. & Cimprich, K. A. The structural determinants of checkpoint activation. *Genes Dev* **21**, 898-903 (2007). https://doi.org:10.1101/gad.1522607
77 Bacolla, A., Ye, Z., Ahmed, Z. & Tainer, J. A. Cancer mutational burden is shaped by G4 DNA, replication stress and mitochondrial dysfunction. *Prog Biophys Mol Biol* **147**, 47-61 (2019). https://doi.org:10.1016/j.pbiomolbio.2019.03.004

78 Baris, Y., Taylor, M. R. G., Aria, V. & Yeeles, J. T. P. Fast and efficient DNA replication with purified human proteins. *Nature* **606**, 204-210 (2022). https://doi.org:10.1038/s41586-022-04759-1
79 Sun, H., Karow, J. K., Hickson, I. D. & Maizels, N. The Bloom's syndrome helicase unwinds G4 DNA. *J Biol Chem* **273**, 27587-27592 (1998). https://doi.org:10.1074/jbc.273.42.27587
80 Fry, M. & Loeb, L. A. Human werner syndrome DNA helicase unwinds tetrahelical structures of the fragile X syndrome repeat sequence d(CGG)n. *J Biol Chem* **274**, 12797-12802 (1999). https://doi.org:10.1074/jbc.274.18.12797
81 Wu, Y., Shin-ya, K. & Brosh, R. M., Jr. FANCJ helicase defective in Fanconia anemia and breast cancer unwinds G-quadruplex DNA to defend genomic stability. *Mol Cell Biol* **28**, 4116-4128 (2008). https://doi.org:10.1128/MCB.02210-07
82 Kilkenny, M. L. *et al.* The human CTF4-orthologue AND-1 interacts with DNA polymerase alpha/primase via its unique C-terminal HMG box. *Open Biol* **7** (2017). https://doi.org:10.1098/rsob.170217
83 Garcia-Gomez, S. *et al.* PrimPol, an archaic primase/polymerase operating in human cells. *Mol Cell* **52**, 541-553 (2013). https://doi.org:10.1016/j.molcel.2013.09.025
84 Kobayashi, K. *et al.* Repriming by PrimPol is critical for DNA replication restart downstream of lesions and chain-terminating nucleosides. *Cell Cycle* **15**, 1997-2008 (2016). https://doi.org:10.1080/15384101.2016.1191711
85 Schiavone, D. *et al.* PrimPol Is Required for Replicative Tolerance of G Quadruplexes in Vertebrate Cells. *Mol Cell* **61**, 161-169 (2016). https://doi.org:10.1016/j.molcel.2015.10.038
86 Scior, A., Preissler, S., Koch, M. & Deuerling, E. Directed PCR-free engineering of highly repetitive DNA sequences. *BMC Biotechnol* **11**, 87 (2011). https://doi.org:10.1186/1472-6750-11-87
87 Baretic, D. *et al.* Cryo-EM Structure of the Fork Protection Complex Bound to CMG at a Replication Fork. *Mol Cell* **78**, 926-940 e913 (2020). https://doi.org:10.1016/j.molcel.2020.04.012
88 Cejka, P. & Kowalczykowski, S. C. The full-length Saccharomyces cerevisiae Sgs1 protein is a vigorous DNA helicase that preferentially unwinds holliday junctions. *J Biol Chem* **285**, 8290-8301 (2010). https://doi.org:10.1074/jbc.M109.083196
89 Mergny, J. L. & Lacroix, L. Analysis of thermal melting curves. *Oligonucleotides* **13**, 515-537 (2003). https://doi.org:10.1089/154545703322860825
90 Fabrini, G. UV-Vis_Melting_Analysis. *https://github.com/GiacomoFabrini/UV-Vis_Melting_Analysis/blob/main/UVVis_MeltingCurveAnalysis.ipynb* (2022).
91 Mergny, J. L., Li, J., Lacroix, L., Amrane, S. & Chaires, J. B. Thermal difference spectra: a specific signature for nucleic acid structures. *Nucleic Acids Res.* **33**, e138 (2005). https://doi.org:10.1093/nar/gni134

Figure Legends

Figure 1 – A single G4- or i-motif-forming sequence can stall the replisome

(**A**) Schematic of replication templates used in this study. Position of the origin of replication is marked as 'ori' from which replication initiates. The multiple cloning site (indicated by a red star) is positioned 3kb downstream of the origin and was used to insert various G4 or i-motif-forming sequences into the template.

(**B**) *In vitro* replication reactions using linear substrates containing a single G4-forming sequence. Products were analysed by denaturing gel electrophoresis.

(**C**) Quantification of stall intensities from three independent experiments of substrates, as shown in (B). The 3kb stalling intensity band was normalised to the intensity of 'leading strand 2' in each lane to account for variation in the efficiencies of reactions for each substrate. The mean stalling intensity is plotted with error bars representing the standard error.

(**D**) Melting temperatures of the G4s formed by sequences tested in (B) versus their stalling intensities as calculated in (C).

(**E**) Analysis of products of *in vitro* replication reactions on linear substrates containing a G4-forming sequence (indicated in green) or a single i-motif-forming sequence (indicated in blue). Products were analysed on denaturing agarose gels.

(**F**) Quantification of stall intensities induced by substrates in (E) from three independent experiments. Stalling intensities were normalised to the intensity of 'leading strand 2' in each lane to account for variation in reaction efficiencies. Mean ± standard error is plotted.

(**G**) The transitional pH (pH_T) of the i-motifs formed by the sequences tested in (E) versus their stalling intensities as presented in (F).

Figure 2 – Replication stalling at G4- and i-motif-forming sequences is due to structure formation

(**A**) Replication reactions of linear substrates containing *wildtype* or mutated G4 sequences depicted in Fig. S2B. Wildtype sequences are indicated in green and mutated sequences are indicated in red. Products are visualised on a denaturing agarose gel.

(**B**) Stalling intensities of substrates as tested in (A) from three independent experiments. Stalling intensities are normalised to the intensity of 'leading strand 2' in each lane, and the mean ± standard error is plotted.

(**C**) Melting temperatures of the secondary structures formed by sequences depicted in (A).

(**D**) Analysis of replication products of linear *wildtype* i-motif substrates or mutated versions as shown in Fig. S2C on a denaturing agarose gel. Wildtype sequences are indicated in blue and mutated sequences are indicated in red.

(**E**) Quantifications of stalling intensities induced by substrates as depicted in (D) from three independent experiments. 3kb stall band intensities were normalised to the intensity of the 1.5kb 'leading strand 2' product to account for variation in reaction efficiencies. Mean is plotted and error bars represent standard error of the mean.

(**F**) The transitional pH (pH_T) of the structures formed by the sequences depicted in (D).

Figure 3 – Replisome stalling is dependent on the opportunity for structure-formation and is affected by orientation

(**A**) Replication of G4 substrates containing poly$(dG)_{60}$ or an increasing number of consecutive (*c-MYC* Pu22)$_n$ or $(GGGT)_n$ repeats. Products have been analysed on a denaturing agarose gel.

(**B**) Replication of i-motif substrates containing poly$(dC)_{60}$ or an increasing number of consecutive $(DUX4L22)_n$ or $(SNORD112)_n$ repeats. Products have been analysed on a denaturing agarose gel.

(**C**) Substrates containing a tract of poly$(dG)_{60}$, or poly$(dG)_{60}$ interrupted every 5th guanine with either a thymine $(G_4T)_{12}$ or adenine $(G_4A)_{12}$, or poly$(dG)_{60}$ interrupted every 10th guanine either a thymine $(G_9T)_6$ or adenine $(G_9A)_6$ were replicated. Replication products were analysed on a denaturing agarose gel.

(**D**) Substrates containing a tract of poly$(dC)_{60}$, or poly$(dC)_{60}$ interrupted every 5th cytosine with either an adenine $(C_4A)_{12}$ or thymine $(C_4T)_{12}$, or poly$(dC)_{60}$ interrupted every 10th cytosine either an adenine $(C_9A)_6$ or thymine $(C_9T)_6$ were replicated. Replication products were analysed on a denaturing agarose gel.

(**E**) Replication products from substrates containing a tract of poly$(dG)_{60}$, or (c-MYC Pu22)$_8$ or $(GGGT)_{28}$ on either the leading strand template (lead) or the lagging strand template (lag).

(**F**) Replication products from substrates containing a tract of poly$(dC)_{60}$, or $(DAP)_4$, $(DUX4L22)_4$, or $(SNORD112)_4$ on either the leading strand template (lead) or the lagging strand template (lag).

Figure 4 – The response of CMG to a pre-formed G4 or i-motif

(**A**) Pulse-chase time course experiment with (*c-MYC* Pu22)$_4$ or (SNORD112)$_4$ substrates. Reactions were initiated with radiolabelled dATP for 10 min, chased with excess 'cold' dATP and samples taken at the indicated time points.

(**B**) Schematic of substrates generated for CMG unwinding assays. The bottom strand of either duplex (top), G4 (middle) or i-motif (bottom) substrates were radioactively end-labelled. The addition of CMG and ATP activates unwinding and produces a single-stranded radiolabelled bottom strand, which can be visualised on a gel.

(**C**) CMG unwinding assays on substrates containing duplex (top), a pre-formed G4 (middle) or a pre-formed i-motif (bottom) as depicted in (B). CMG unwinding was stimulated by the addition of 2 nM ATP following CMG loading in the presence of ATPγS. Samples were taken at the indicated time points. Products were run on 10% TBE gels. Input and boiled substrates were used as controls to visualise where original and unwound substrates run on the gel. The proportion of template unwound was calculated by measuring the intensity of the 'unwound' product band as a proportion of the total product intensity for each lane.

(**D**) CMG unwinding assays on substrates containing duplex (top) or a pre-formed G4 (bottom) as depicted in (B). Reactions were carried out as in (C) but with the addition of 0.25 μM PhenDC3 where indicated.

Figure 5 – Stalling at a G4 or i-motif leads to helicase-polymerase uncoupling, and is exacerbated by uncoupled forks

(**A**) Replication reactions carried out on G4-containing templates in the presence of a primer that anneals 264 nt downstream of the G4 (rp) or a scrambled control primer (scr).

(**B**) Replication products of reactions using substrates containing the indicated G4 or i-motif-forming sequences analysed by native gel. The inset shows an increased contrast of the region indicated by the dashed box to better visualize uncoupled products.

(**C**) Replication of G4 or i-motif substrates with either *wildtype* pol ε (wt) or a pol ε mutant with a deleted catalytic domain (Δcat). Substrates were digested with EcoRV prior to replication in order to reduce the size of the longer leading strand (leading strand 1). Products were analysed on a denaturing agarose gel.

Figure 6 – Pif1, but not Sgs1 or Chl1, is able to rescue stalling at both G4s and i-motifs

(**A** to **D**) Pulse-chase experiments carried out with the indicated templates. Reactions were initiated with radiolabelled dATP. After a 10 min pulse, either Sgs1 (**A**), Chl1 (**B**) or *wildtype* or ATPase-dead K264A (mut) pif1 (**C** and **D**) was added with the chase and samples taken after another 10 min.

Figure 1

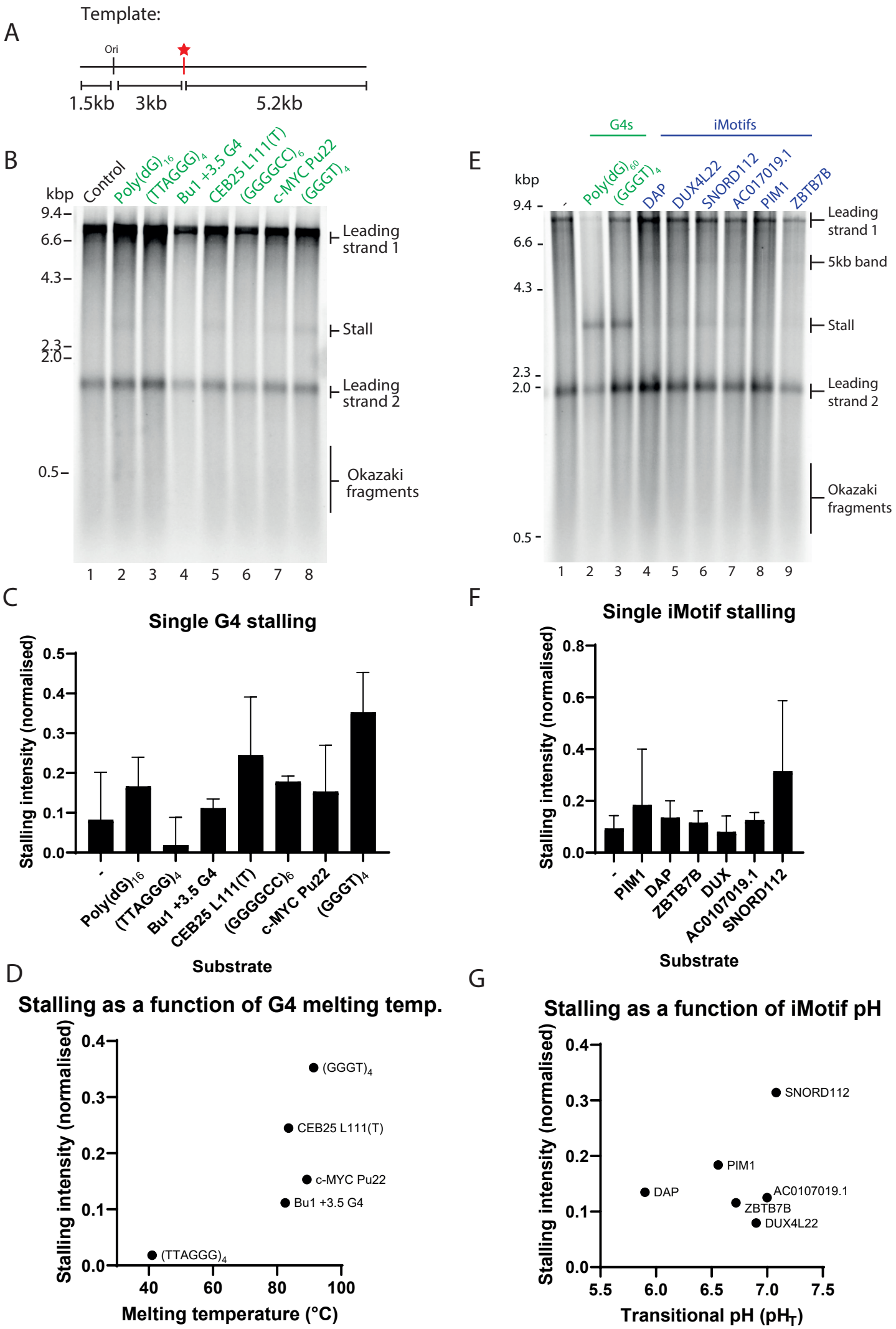
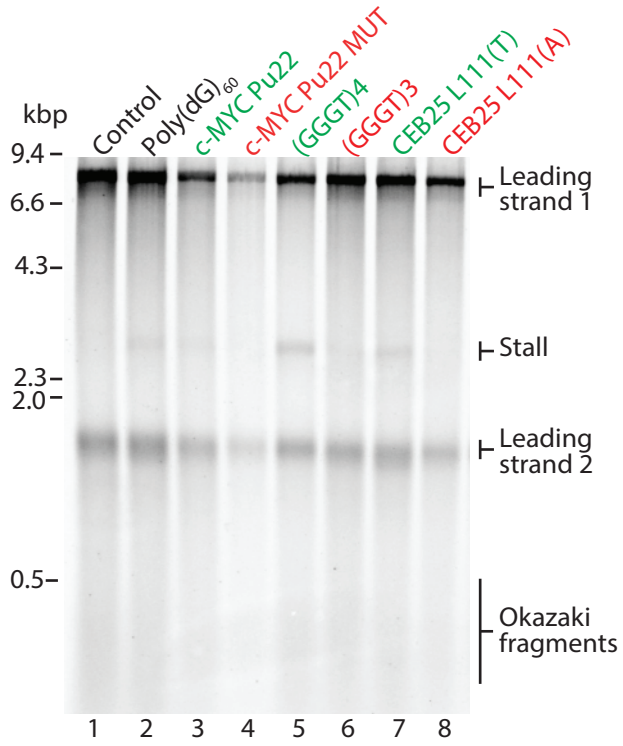
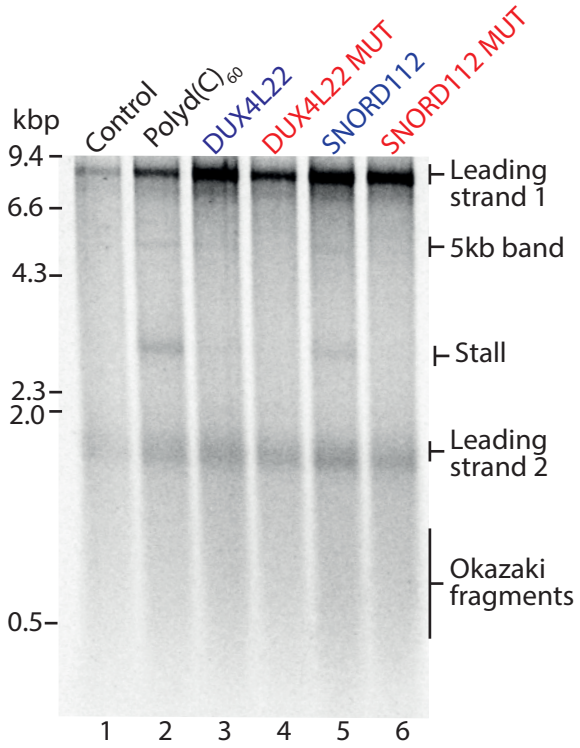

A
Template:
Ori
1.5kb
3kb
5.2kb
B
kbp
Control
Poly(dG)16
(TTAGGG)4
Bu1 +3.5 G4
CEB25 L111(T)
(GGGGCC)6
c-MYC Pu22
(GGGT)4
9.4
6.6
4.3
2.3
2.0
0.5
Leading strand 1
Stall
Leading strand 2
Okazaki fragments
1 2 3 4 5 6 7 8
E
G4s
iMotifs
kbp
-
Poly(dG)60
(GGGT)4
DAP
DUX4L22
SNORD112
AC017019.1
PIM1
ZBTB7B
9.4
6.6
4.3
2.3
2.0
0.5
Leading strand 1
5kb band
Stall
Leading strand 2
Okazaki fragments
1 2 3 4 5 6 7 8 9
C
Single G4 stalling
Stalling intensity (normalised)
0.5
0.4
0.3
0.2
0.1
0.0
-
Poly(dG)16
(TTAGGG)4
Bu1 +3.5 G4
CEB25 L111(T)
(GGGGCC)6
c-MYC Pu22
(GGGT)4
Substrate
F
Single iMotif stalling
Stalling intensity (normalised)
0.8
0.6
0.4
0.2
0.0
-
PIM1
DAP
ZBTB7B
DUX
AC0107019.1
SNORD112
Substrate
D
Stalling as a function of G4 melting temp.
Stalling intensity (normalised)
0.4
0.3
0.2
0.1
0.0
(GGGT)4
CEB25 L111(T)
c-MYC Pu22
Bu1 +3.5 G4
(TTAGGG)4
40
60
80
100
Melting temperature (°C)
G
Stalling as a function of iMotif pH
Stalling intensity (normalised)
0.4
0.3
0.2
0.1
0.0
SNORD112
PIM1
DAP
AC0107019.1
ZBTB7B
DUX4L22
5.5
6.0
6.5
7.0
7.5
Transitional pH (pHT)

Figure 2

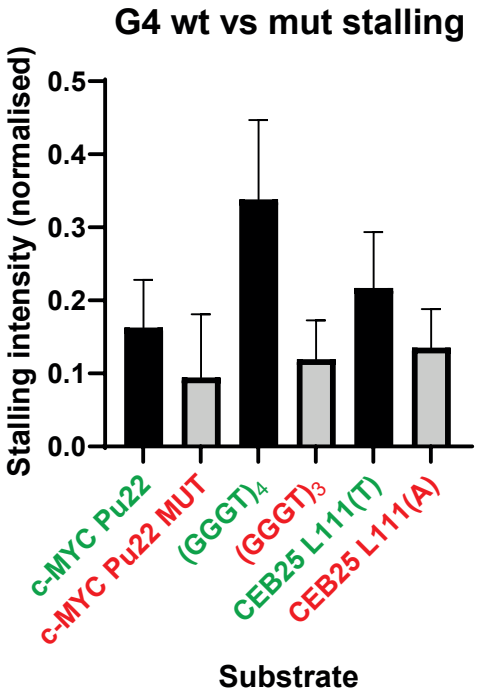
A



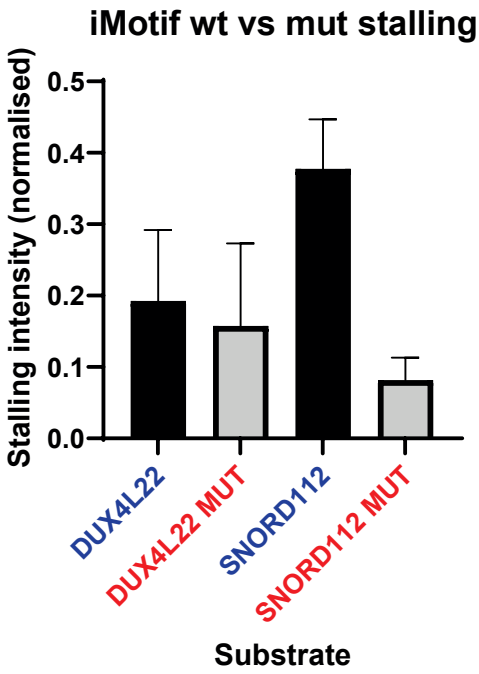
D



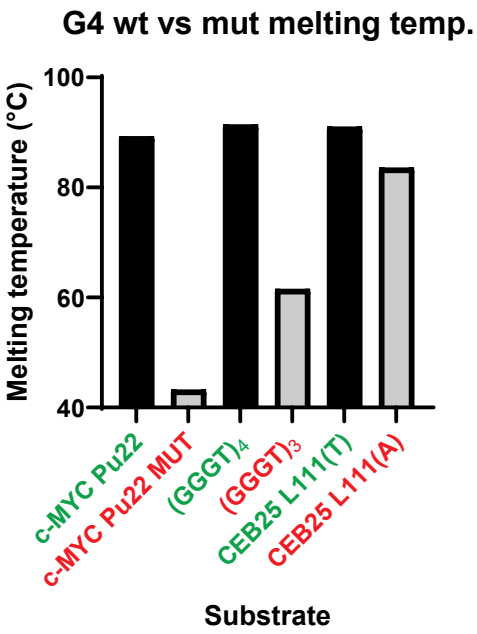
B



E



C



F

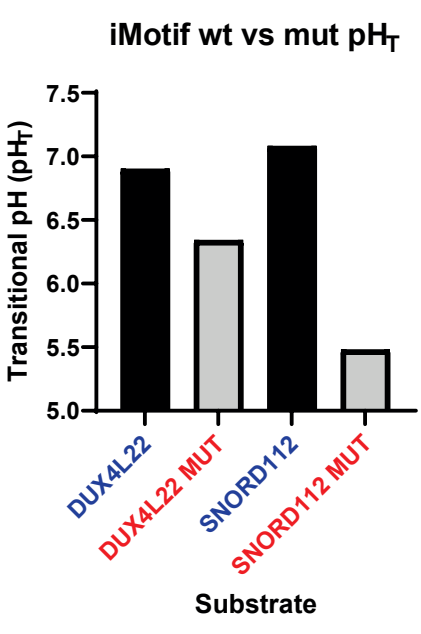


Figure 3

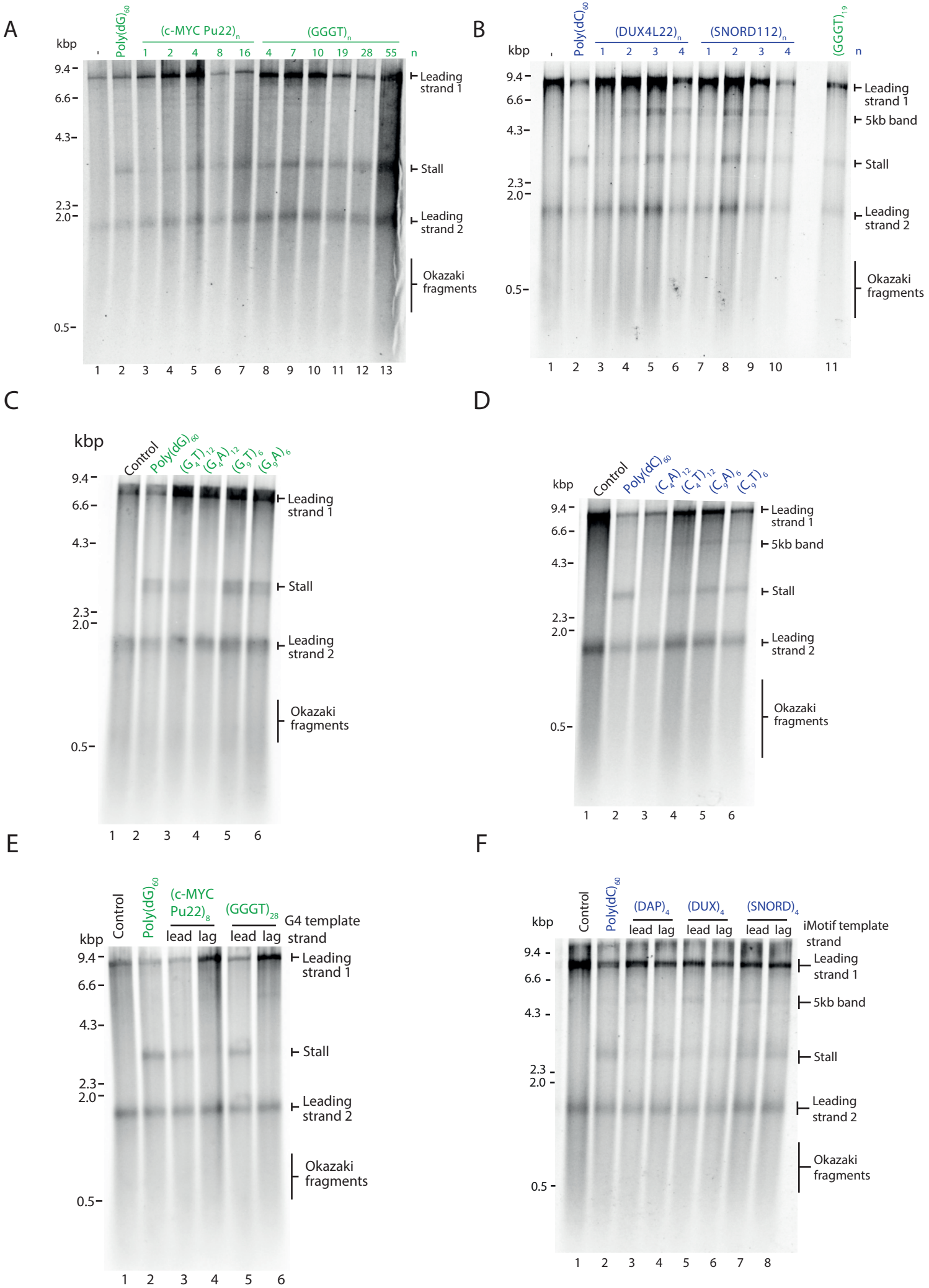

A
kbp
Poly(dG)60
(c-MYC Pu22)n
(GGGT)n
1 2 4 8 16 4 7 10 19 28 55 n
9.4
6.6
4.3
2.3
2.0
0.5
Leading strand 1
Stall
Leading strand 2
Okazaki fragments
1 2 3 4 5 6 7 8 9 10 11 12 13
B
kbp
Poly(dC)60
(DUX4L22)n
(SNORD112)n
(GGGT)19
1 2 3 4 1 2 3 4 n
9.4
6.6
4.3
2.3
2.0
0.5
Leading strand 1
5kb band
Stall
Leading strand 2
Okazaki fragments
1 2 3 4 5 6 7 8 9 10 11
C
kbp
Control
Poly(dG)60
(G4T)12
(G4A)12
(G9T)6
(G9A)6
9.4
6.6
4.3
2.3
2.0
0.5
Leading strand 1
Stall
Leading strand 2
Okazaki fragments
1 2 3 4 5 6
D
kbp
Control
Poly(dC)60
(C4A)12
(C4T)12
(C9A)6
(C9T)6
9.4
6.6
4.3
2.3
2.0
0.5
Leading strand 1
5kb band
Stall
Leading strand 2
Okazaki fragments
1 2 3 4 5 6
E
kbp
Control
Poly(dG)60
(c-MYC Pu22)8
(GGGT)28
lead lag lead lag
G4 template strand
9.4
6.6
4.3
2.3
2.0
0.5
Leading strand 1
Stall
Leading strand 2
Okazaki fragments
1 2 3 4 5 6
F
kbp
Control
Poly(dC)60
(DAP)4
(DUX)4
(SNORD)4
lead lag lead lag lead lag
iMotif template strand
9.4
6.6
4.3
2.3
2.0
0.5
Leading strand 1
5kb band
Stall
Leading strand 2
Okazaki fragments
1 2 3 4 5 6 7 8

Figure 4

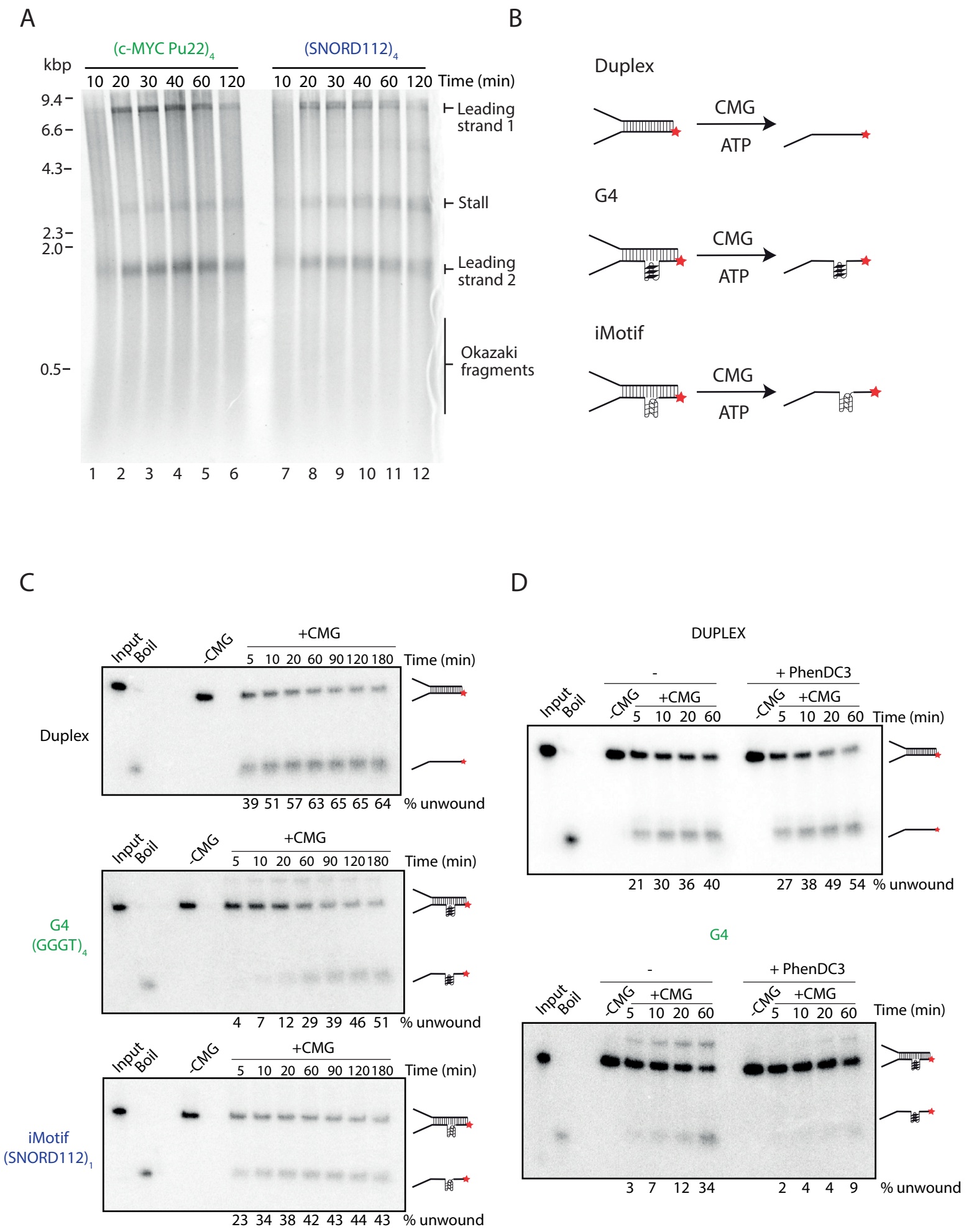

A
(c-MYC Pu22)4
(SNORD112)4
kbp
10 20 30 40 60 120
10 20 30 40 60 120 Time (min)
9.4
6.6
4.3
2.3
2.0
0.5
Leading strand 1
Stall
Leading strand 2
Okazaki fragments
1 2 3 4 5 6 7 8 9 10 11 12
B
Duplex
CMG
ATP
G4
CMG
ATP
iMotif
CMG
ATP
C
Input
Boil
-CMG
+CMG
5 10 20 60 90 120 180 Time (min)
Duplex
39 51 57 63 65 65 64 % unwound
G4
(GGGT)4
4 7 12 29 39 46 51 % unwound
iMotif
(SNORD112)1
23 34 38 42 43 44 43 % unwound
D
DUPLEX
-
+ PhenDC3
5 10 20 60 Time (min)
21 30 36 40 27 38 49 54 % unwound
G4
3 7 12 34 2 4 4 9 % unwound

Figure 5

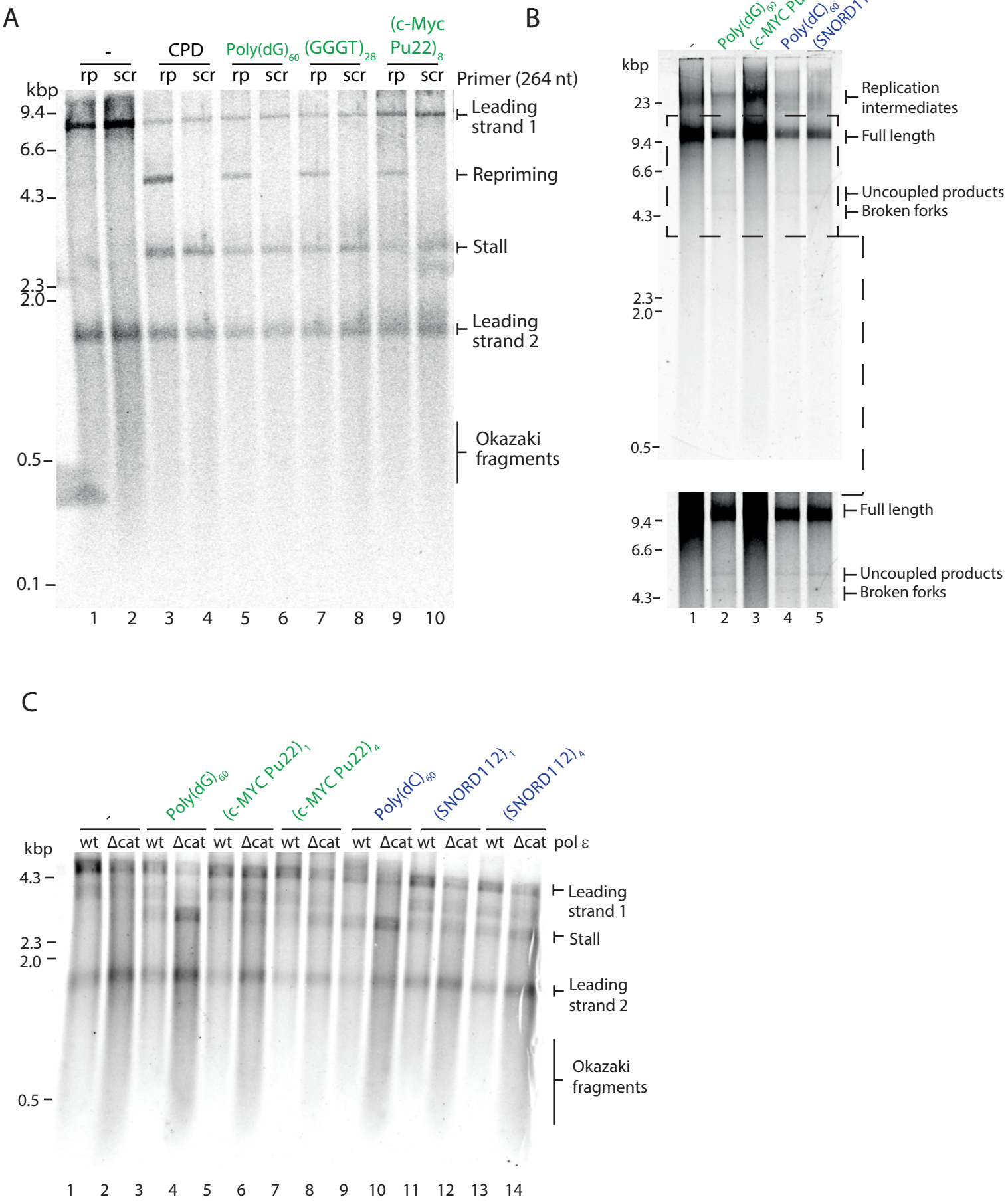
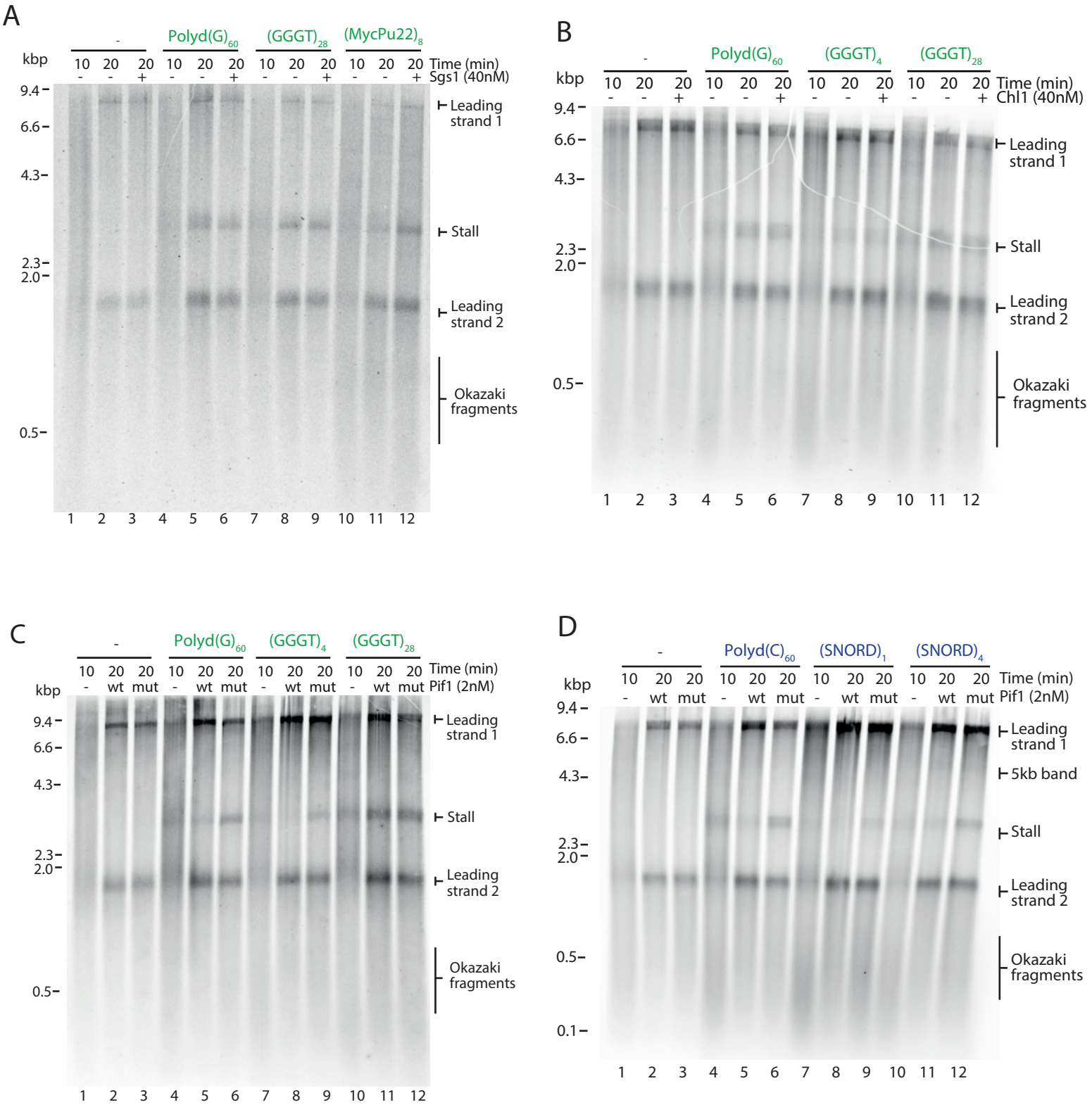

A
-
CPD
Poly(dG)60
(GGGT)28
(c-Myc Pu22)8
rp scr rp scr rp scr rp scr rp scr
Primer (264 nt)
kbp
9.4
6.6
4.3
2.3
2.0
0.5
0.1
Leading strand 1
Repriming
Stall
Leading strand 2
Okazaki fragments
1 2 3 4 5 6 7 8 9 10
B
-
Poly(dG)60
(c-MYC Pu22)4
Poly(dC)60
(SNORD112)4
kbp
23
9.4
6.6
4.3
2.3
2.0
0.5
Replication intermediates
Full length
Uncoupled products
Broken forks
Full length
Uncoupled products
Broken forks
1 2 3 4 5
C
-
Poly(dG)60
(c-MYC Pu22)1
(c-MYC Pu22)4
Poly(dC)60
(SNORD112)1
(SNORD112)4
kbp
wt Δcat wt Δcat wt Δcat wt Δcat wt Δcat wt Δcat wt Δcat pol ε
4.3
2.3
2.0
0.5
Leading strand 1
Stall
Leading strand 2
Okazaki fragments
1 2 3 4 5 6 7 8 9 10 11 12 13 14

Figure 6


A
-
Polyd(G)60
(GGGT)28
(MycPu22)8
kbp
10 20 20 10 20 20 10 20 20 10 20 20 Time (min)
- - + - - + - - + - - + Sgs1 (40nM)
9.4
6.6
4.3
2.3
2.0
0.5
Leading strand 1
Stall
Leading strand 2
Okazaki fragments
1 2 3 4 5 6 7 8 9 10 11 12
B
-
Polyd(G)60
(GGGT)4
(GGGT)28
kbp
10 20 20 10 20 20 10 20 20 10 20 20 Time (min)
- - + - - + - - + - - + Chl1 (40nM)
9.4
6.6
4.3
2.3
2.0
0.5
Leading strand 1
Stall
Leading strand 2
Okazaki fragments
1 2 3 4 5 6 7 8 9 10 11 12
C
-
Polyd(G)60
(GGGT)4
(GGGT)28
kbp
10 20 20 10 20 20 10 20 20 10 20 20 Time (min)
- wt mut - wt mut - wt mut - wt mut Pif1 (2nM)
9.4
6.6
4.3
2.3
2.0
0.5
Leading strand 1
Stall
Leading strand 2
Okazaki fragments
1 2 3 4 5 6 7 8 9 10 11 12
D
-
Polyd(C)60
(SNORD)1
(SNORD)4
kbp
10 20 20 10 20 20 10 20 20 10 20 20 Time (min)
- wt mut - wt mut - wt mut - wt mut Pif1 (2nM)
9.4
6.6
4.3
2.3
2.0
0.5
0.1
Leading strand 1
5kb band
Stall
Leading strand 2
Okazaki fragments
1 2 3 4 5 6 7 8 9 10 11 12